Collected Papers 1979–1995

Richard G. Pierse

Collected Papers 1979–1995

British Library Cataloguing-in-Publication Data
A catalogue record of this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data
A catalogue record of this book is available from the Library of Congress
Richard G Pierse
p. cm. - (Collected Papers)

ISBN 978-0-1234567-8-9

Published by
Richard Pierse Publications Ltd
36 Cyprus Road
Cambridge, CB1 3QA, UK
http://www.winsolve.co.uk

Dedicated to the memory of my father
Reginald Frederick Pierse 1923–2012

Contents

List of Figures

List of Tables

Preface

The earliest of these papers (Chapter 1) is the dissertation I submitted in 1979 in part fulfillment of the requirement for the MSc in Econometrics and Mathematical Economics at the London School of Economics. My supervisor was David F. Hendry who had introduced me to an (at the time) unpublished paper by Jean-François Richard on exogeneity and suggested that I could use it as the basis of a dissertation. Richard's paper was very difficult but rewarding.

Later, I got to meet Jean-François and to collaborate with him and Michel Lubrano on a paper applying Bayesian techniques to exogeneity testing (Chapter 2). By this time I was doing a Ph.D. at the LSE on exogeneity and had built a model of money demand to look at the impact of Competition and Credit Control on the UK money market. Jean-François and Michel needed an application for their Bayesian analysis so we decided to collaborate on a joint paper. We tried very hard but ultimately unsuccessfully to get this published in *Econometrica* but in the end it was fast-tracked into a special econometrics issue of the *Review of Economic Studies*. Because of the last-minute rush of its final publication and the fact that its three authors lived in three different countries, the published paper has a number of typographical errors that I have aimed to correct in re-setting the paper for this book.

I had only a two-year grant for my Ph.D at LSE so in 1981–2, in order to stay on for another year, I took a part-time Research Assistant job working with Andrew Harvey. The main outcome of this year was my first publication, in 1984 in the *Journal of the American Statistical Association* on temporal aggregation and missing observations (Chapter 3).

In October 1982 I moved to Cambridge to join the Cambridge Growth Project at the Department of Applied Economics. Here I met Hashem Pesaran though it was a few years before the start of our collaboration. By 1985, funding for the Growth Project (and several other macroeconomic modelling groups) had come under threat as the Economic and Social Research Council (ESRC) began to take notice of the criticisms (mainly coming from the USA) of macroeconomic models, especially large ones. I began to think about whether it was possible to prove statistically that disaggregated macroeconomic modelling was worthwhile. I was thinking along the lines of formal tests for aggregation restrictions but it became obvious that the conditions for valid aggregation are very stringent and will almost never be satisfied. I happened to mention what I was working on to Hashem who then got interested in alternative approaches such as the Grunfeld-Griliches criterion.

The result, after a long and painful struggle, was a publication in *Econometrica* in 1989 (Chapter 5). Angus Deaton was the journal editor and he and Hashem had long fights over revisions that Angus demanded and Hashem didn't want to make. The final fight was over an appendix which proved the asymptotic validity of a test for perfect aggregation for a special case. Angus wanted it cut; Hashem insisted it should remain. Angus won that one by pointing out that, as he was soon due to step down as editor, if we didn't agree to cut it, he would hand the paper over to his successor and we would be back to square one. Hashem was forced to concede but he had the last word as we managed to get the appendix accepted as a separate paper in *Economics Letters* (Chapter 6), published in time for a reference to it to be included in the *Econometrica* paper.

The presence of the name of a third author on the *Econometrica* paper, Mohan Kumar, needs some explanation. When Hashem and I first started working on the paper, we were planning an illustrative application to UK wage equations. The policy of the Growth Project (enforced by its Director, Terry Barker) was that research into any macroeconomic area should always be done in collaboration with the person on the Project who was allocated to work on that area. Although our paper was mainly a theoretical one and the application was solely for illustration of the new tests and criteria developed, since Mohan Kumar was in charge of wage equations in the Growth Project, he was drafted onto the team. However, Hashem and Mohan did not get on and soon Hashem and I were working without him and I ended up doing all the applied work on my own. By the time that the paper appeared, Mohan Kumar had long departed from Cambridge to the IMF, the Growth Project was defunct (its ESRC funding having been cut), and the application in the paper was to employment equations rather than wage equations. Despite all this, we did not think of removing Mohan's name from the paper until the last minute, when it was too late. Mohan had no idea that the paper even existed until I rang him up to tell him he had a forthcoming paper in probably the most prestigious journal in economics. The fantasy that one day this might happen to me still haunts me.

The final irony concerning the *Econometrica* paper is that, while it included as a co-author someone who had no part in the paper, it failed to acknowledge the part played, in the final stages of preparation of the paper, by another colleague, Kevin Lee. Kevin joined the Growth Project in 1985 and, by the time that the *Econometrica* paper was published, he was already working with Hashem and me on other papers on disaggregation. I believe that he actually wrote the Data Appendix in the *Econometrica* paper yet he does not even merit a mention in the acknowledgements.

Kevin, Hashem and I worked together for several years, producing papers on aggregation bias in the *Economic Journal* in 1990 (Chapter 7), in the book *Disaggregation in Econometric Modelling* (Barker and Pesaran (1990)) also in 1990 (Chapter 8), and in the *Journal of Econometrics* in 1993 (Chapter 10) and papers on persistence in disaggregated models in the *Economic Journal* in 1992 (Chapter 9) and in the *Journal of Business & Economic Statistics* in 1994 (Chapter 11).

After a break of several years, I returned to issues of temporal aggregation in the context of unit root testing in a paper written with Andy Snell and published in the *Journal of Econometrics* in 1995 (Chapter 4). Like many of these papers, this one had a very long gestation period. We conceived it in the late 1980s when Andy and I were both working in Cambridge. In 1991, when Andy had moved to the University of Edinburgh but I was still at Cambridge, we won an ESRC research grant to develop it. We were still working on revisions in 1994 by which time I was at the London Business School and, by the time it was published, I had moved to the University of Surrey. Initially, the paper was planned to be mainly a Monte Carlo exercise. However, this was partly pre-empted by the appearance of a paper in *Economics Letters* in 1995, Shiller and Perron (1985), that did something close to what we had been doing. We were forced to refocus our paper in a more theoretical direction as well as adding an application to real data.

Acknowledgements

Most of these papers were jointly authored and I am indebted to all my co-authors: Andrew Harvey, Kevin Lee, Michel Lubrano, Hashem Pesaran, Jean-François Richard and Andy Snell. Finally, I would like to thank my family and friends for their encouragement and tolerance.

Richard Pierse, Cambridge, October 2016.

Part I

Exogeneity and Switching Regimes

Chapter 1

Multiple Regime Models with Switches in Exogeneity

1.1 Introduction

Mutiple regime models have been used by econometricians in several different fields, a major application being to models of market disequilibrium (see Maddala and Nelson (1974)). One interesting field of application is the controlled economy where the policy-maker switches between control instruments at different points in time, these switchpoints being generally known (unlike the situation in market disequlibrium models). A characteristic of such models is that the partitioning between "endogenous" and "exogenous" variables changes between regimes.

Multiple regime models with this characteristic have recently been analysed by Jean-François Richard in a forthcoming paper (Richard (1979)). He shows that it is possible to test for the exogeneity of the control variables in the different regimes and hence to test whether or not the economy was being controlled.

This study is an attempt to implement the analysis of Richard for the FIML estimation and testing of multiple regime models and apply it to model the change in regime in the U.K. monetary sector of the introduction of 'Competition and Credit Control' in October 1971. An existing FIML computer program written by David Hendry was generalised to maximise the likelihood function of multiple regime models and to incorporate the tests proposed by Richard. The application involved the estimation of a money demand equation; in developing this equation a single equation approach was used and the general to simple methodology of Hendry and Mizon (1979) was adopted. The resulting equation in seasonally unadjusted data provides direct comparison with the equations in Hendry and Mizon (1979).

The rest of this paper is organised as follows:

- In Section 1.2 some of the key concepts are developed for some simple models
- Section 1.3 derives Richard's exogeneity test for the case of a complete simultaneous model with a single regime

[0] Dissertation submitted in partial fulfillment of the requirements of the MSc in Econometrics and Mathematical Economics, London School of Economics, June 1979. The author is grateful to his supervisor, David Hendry for his help and encouragement.

- Section 1.4 deals with the FIML estimation of multiple regime models and describes a testing procedure for these models
- Section 1.5 presents an application to the estimation of a simple two-equation model of the monetary sector. Results and conclusions are presented.

A note on exogeneity.
We define exogeneity for our purposes as follows: a variable is exogenous if we can run the analysis conditional on it without loss of information. This is a weak definition of exogeneity. In particular we do not require that lagged values of endogenous variables do not enter the determination of exogenous variables. Let x_t be an exogenous variable and y_t an endogenous variable. Then in the lag formulation

$$a(L)x_t = b(L)y_t + w_t$$

where L is the lag operator, we do not require that $b(L) = 0$ (although b_0 must be zero). Thus y_t may in fact be "causing" x_t in the Granger sense of causality (see Granger (1969)).

1.2 Exogeneity in Simple Models

Suppose that we have two variables y_{1t} and y_{2t} which are jointly normally distributed with density

$$f(y_{1t}, y_{2t}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1.2.1}$$

or explicitly

$$f(y_{1t}, y_{2t}) = (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\begin{pmatrix} y_{1t}-\mu_1 \\ y_{2t}-\mu_2 \end{pmatrix}' \boldsymbol{\Sigma}^{-1} \begin{pmatrix} y_{1t}-\mu_1 \\ y_{2t}-\mu_2 \end{pmatrix}\right) \tag{1.2.2}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{pmatrix}.$$

The marginal distributions for y_{1t} and y_{2t} are given by

$$f(y_{1t}) = \mathrm{N}(\mu_1, \sigma_{11}) \quad \text{and} \quad f(y_{2t}) = \mathrm{N}(\mu_2, \sigma_{22}).$$

Defining $s_1 = \frac{\sigma_{11}\sigma_{22}-(\sigma_{12})^2}{\sigma_{11}}$ and $s_2 = \frac{\sigma_{11}\sigma_{22}-(\sigma_{12})^2}{\sigma_{22}}$ we have

$$\sigma^{11} = \frac{\sigma_{22}}{\sigma_{11}} s_1^{-1} \quad , \quad \sigma^{12} = -\frac{\sigma_{12}}{\sigma_{11}} s_1^{-1} \quad , \quad \sigma^{22} = s_1^{-1}$$

and the conditional distribution of y_{2t} given y_{1t} is

$$f(y_{2t}|y_{1t}) = f(y_{1t}, y_{2t})/f(y_{1t})$$

$$\begin{aligned} &= (2\pi)^{-\frac{1}{2}} \sigma_{11}^{\frac{1}{2}} \left(\sigma_{11}\sigma_{22} - (\sigma_{12})^2\right)^{-\frac{1}{2}} \exp(-\frac{1}{2s_1}[\frac{\sigma_{22}}{\sigma_{11}}(y_{1t}-\mu_1)^2 \\ &- 2(y_{1t}-\mu_1)(y_{2t}-\mu_2) + (y_{2t}-\mu_2)^2 - \left(\frac{\sigma_{11}\sigma_{22}-(\sigma_{12})^2}{\sigma_{11}^2}\right)(y_{1t}-\mu_1)^2]) \end{aligned} \tag{1.2.3}$$

$$= (2\pi s_1)^{-\frac{1}{2}} \exp\left(-\frac{1}{2s_1}\left[(y_{2t}-\mu_2)^2 - 2\frac{\sigma_{12}}{\sigma_{11}}(y_{1t}-\mu_1)(y_{2t}-\mu_2) + \left(\frac{\sigma_{12}}{\sigma_{11}}\right)^2 (y_{1t}-\mu_1)^2\right]\right)$$

which is a univariate distribution

$$f(y_{2t}|y_{1t}) = \mathrm{N}\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(y_{1t}-\mu_1), s_1\right). \tag{1.2.4}$$

Similarly we can derive the conditional density for y_{1t} given y_{2t} which is

$$f(y_{1t}|y_{2t}) = \mathrm{N}\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(y_{2t}-\mu_2), s_2\right). \tag{1.2.5}$$

From (1.2.4) we get the regression model

$$y_{2t} = a + by_{1t} + u_t \quad , \quad u_t \sim \mathrm{NI}(0, \delta^2) \tag{1.2.6}$$

with $\mathrm{Cov}(y_{1t}, u_t) = 0$, where $a = \mu_2 - b\mu_1$, $b = \frac{\sigma_{12}}{\sigma_{11}}$ and $\delta^2 = s_1$ and from (1.2.5) the regression model

$$y_{1t} = c + dy_{2t} + v_t \quad , \quad v_t \sim \mathrm{NI}(0, \psi^2) \tag{1.2.7}$$

with $\mathrm{Cov}(y_{2t}, v_t) = 0$, $c = \mu_1 - d\mu_2$, $d = \frac{\sigma_{12}}{\sigma_{22}}$ and $\psi^2 = s_2$.

As long as there are no prior cross-restrictions on the five parameters of the joint distribution $(\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})$ we lose no information on (a, b, δ^2), the parameters of the conditional distribution of y_{2t}, by treating (μ_1, σ_{11}) as nuisance parameters and running the analysis conditional on y_{1t}. Equally, if we choose to treat (μ_2, σ_{22}) as nuisance parameters no information is lost by running the analysis conditional on y_{2t}. Thus (1.2.6) and (1.2.7) are equally valid parameterisations of the model (1.2.1) where we choose to treat (μ_1, σ_{11}) or (μ_2, σ_{22}) respectively as nuisance parameters.

Now consider the two-regime model

$$\begin{aligned} y_{1t} &= a_1 + b_1 y_{2t} + u_{1t} \quad , \quad t \in I_1 = \{1, \cdots, T1-1\} \\ y_{2t} &= a_2 + b_{21} y_{1t} + u_{2t} \quad , \quad t \in I_2 = \{T1, \cdots, T\} \end{aligned} \tag{1.2.8}$$

where $u_{1t} \sim \mathrm{NI}(0, \delta_1^2)$, $u_{2t} \sim \mathrm{NI}(0, \delta_2^2)$, $\mathrm{Cov}(y_{2t}, u_{1t}) = 0$ $t \in I_1$ and $\mathrm{Cov}(y_{1t}, u_{2t}) = 0$ $t \in I_2$. The switching time $T1$ is assumed to be known. There appears to be a switch in the exogeneity of the variables in this model at $T1$ with y_{2t} exogenous for $t \in I_1$ and y_{1t} exogenous for $t \in I_2$. However, it is clear from our previous analysis that this parameterisation is inadequate to describe such a switch in exogeneity. Let us assume that the joint distribution of (y_{1t}, y_{2t}) in regime i is

$$f(y_{1t}, y_{2t}) = \mathrm{N}(\boldsymbol{\mu}^i, \boldsymbol{\Sigma}^i) \quad , \quad t \in I_1 \quad , \quad i = 1, 2. \tag{1.2.9}$$

Then indeed (1.2.8) follows together with

$$(a_i, b_i, \delta_i^2) = \left(\mu_i^i - \frac{\sigma_{12}^i}{\sigma_{jj}^i}\mu_j^i, \frac{\sigma_{12}^i}{\sigma_{jj}^i}, \frac{\sigma_{11}^i \sigma_{22}^i - (\sigma_{12}^i)^2}{\sigma_{jj}^i}\right) \tag{1.2.10}$$

for $i = 1, 2$, $j = 3 - i$ where we treat (μ_j^i, σ_{jj}^i) as the nuisance parameters. Equally, however, we have the parameterisation

$$\begin{gathered} y_{1t} = a_3 + b_3 y_{2t} + u_t, \quad t = 1, \cdots, T \\ y_{2t} \sim \mathrm{NI}(0, \phi_1^2) \quad t \in I_1, \quad y_{2t} \sim \mathrm{NI}(0, \phi_2^2) \quad t \in I_2 \end{gathered} \tag{1.2.11}$$

with $u_t \sim \mathrm{NI}(0, \delta_3^2)$ where

$$(a_3, b_3, \delta_3^2) = \left(\mu_1^i - \frac{\sigma_{12}^i}{\sigma_{22}^i}\mu_2^i, \frac{\sigma_{12}^i}{\sigma_{22}^i}, \frac{\sigma_{11}^i\sigma_{22}^i - (\sigma_{12}^i)^2}{\sigma_{22}^i}\right) \quad i = 1, 2 \tag{1.2.12}$$

and $(\mu_2^i, \sigma_{22}^i) = (0, \phi_i^2)$ are treated as the nuisance parameters. (Note that we have the parameter correspondences $(a_1, b_1, \delta_1^2) = (a_3, b_3, \delta_3^2)$ for the first regime and $(a_2, b_2, \delta_2^2) = (-a_3b_3\lambda, b_3\lambda, \delta_3^2\lambda)$ for the second regime where $\lambda = \sigma_{22}^2/\sigma_{11}^2$.)

Equations (1.2.8) and (1.2.11) are both valid parameterisations of the data generation process (1.2.9) yet (1.2.8) exhibits a switch in exogeneity between regimes whereas in (1.2.11) the same variable y_{2t} is treated as exogenous over the whole period. The choice of exogenous variables in these models is quite arbitrary and the statement that y_{2t} is exogenous in (1.2.11) is no more than a statement that (μ_2, σ_{22}) are nuisance parameters. It is not subject to testing since it forces no restrictions on the joint distribution of the observable variables (1.2.9).

Suppose that, returning to the single regime model, we now do have a cross-restriction on the parameters of the joint distribution (1.2.1). We introduce a behavioural hypothesis

$$\mu_{2t} = c_0 + c_1\mu_{1t}. \tag{1.2.13}$$

This hypothesis says that the expectation of variable y_{2t} at time t depends on the expectation of y_{1t} at time t. We can interpret it as a behavioural rule followed by agents who take the joint distribution (1.2.1) as given. Looking at the conditional distribution of y_{2t} given y_{1t} we have from (1.2.4)

$$\mathrm{E}(y_{2t}|y_{1t}) = \mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(y_{1t} - \mu_1) \tag{1.2.14}$$

but from (1.2.13) this implies

$$\mathrm{E}(y_{2t}|y_{1t}) = c_0 + \left(c_1 - \frac{\sigma_{12}}{\sigma_{11}}\right)\mu_{1t} + \frac{\sigma_{12}}{\sigma_{11}}y_{1t}. \tag{1.2.15}$$

In general this expression will now involve μ_{1t} and we will lose information on the parameters of the conditional distribution by running the regression model (1.2.6) treating y_{1t} as exogenous. Only if

$$c_1 = \frac{\sigma_{12}}{\sigma_{11}} \tag{1.2.16}$$

so that

$$\mathrm{E}(y_{2t}|y_{1t}) = c_0 + \frac{\sigma_{12}}{\sigma_{11}}y_{1t} \tag{1.2.17}$$

do we lose no information by treating y_{1t} as exogenous. Condition (1.2.16) then is a direct test for the exogeneity of y_{1t}. An alternative intuitive explanation of this result may be helpful. Let us write

$$\mu_{1t} = \mathrm{E}(y_{1t}|\ell_{t-1}) \quad , \quad \mu_{2t} = \mathrm{E}(y_{2t}|\ell_{t-1})$$

where ℓ_{t-1} is the information set available at time t. Then the exogeneity condition (1.2.16) is equivalent to the condition that $y_{1t} \in \ell_{t-1}$ in which case $\mu_{1t} = y_{1t}$ and (1.2.15) becomes

$$\mathrm{E}(y_{2t}|y_{1t}) = c_0 + c_1y_{1t}. \tag{1.2.17$'$}$$

Similarly from the conditional distribution of y_{1t} given y_{2t} the hypothesis (1.2.13) gives

$$\mathrm{E}(y_{1t}|y_{2t}) = \left(1 - \frac{\sigma_{12}}{\sigma_{22}}c_1\right)\mu_{1t} - \frac{\sigma_{12}}{\sigma_{22}}c_0 + + \frac{\sigma_{12}}{\sigma_{22}}y_{2t} \tag{1.2.18}$$

which leads to the exogeneity condition

$$c_1 = \frac{\sigma_{22}}{\sigma_{12}} \tag{1.2.19}$$

in which case

$$\mathrm{E}(y_{1t}|y_{2t}) = -\frac{c_0}{c_1} + \frac{\sigma_{12}}{\sigma_{22}}y_{2t} \tag{1.2.20}$$

which valididates the parameterisation (1.2.7).

Treating y_{1t} (y_{2t}) as exogenous is only valid when the appropriate restriction ((1.2.16) or (1.2.19)) is satisfied and the two conditions cannot hold jointly (otherwise $\boldsymbol{\Sigma}$ would be singular). When y_{1t} is exogenous and if it is controlled then agents adjust their behaviour by revising their expectations according to the behavioural rule (1.2.13).

This framework provides a basis for examining multiple regime models with possible switches in exogeneity. We have seen that we need to specify the joint distribution of all the relevant variables in each regime, whether or not they are being controlled, so that the hypothesis of exogeneity can be tested. We now go on to develop this approach for the full information estimation of complete simultaneous equations models.

1.3 Exogeneity in a Single Regime

In the simple models considered in section 1.2 there were no predetermined variables. We now want to extend the analysis to allow for lagged dependent variables; we may also allow for some variables to be purely exogenous without proposing to test this. The general form of model for a single regime is

$$\boldsymbol{A}\boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{y}_t + \boldsymbol{C}\boldsymbol{z}_t = \boldsymbol{u}_t \quad , \quad \boldsymbol{u}_t \sim \mathrm{NI}(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{1.3.1}$$

with reduced form

$$\boldsymbol{y}_t = \boldsymbol{\Pi}\boldsymbol{z}_t + \boldsymbol{v}_t \quad , \quad \boldsymbol{v}_t \sim \mathrm{NI}(\boldsymbol{0}, \boldsymbol{\Omega}) \tag{1.3.2}$$

where $\boldsymbol{y}_t$ is now an $(n \times 1)$ vector of observations on n endogenous variables and $\boldsymbol{z}_t$ is an $(m \times 1)$ vector of observations on predetermined variables. We assume that (1.3.1) is a complete model so that $\boldsymbol{B}$ is a square, non-singular $(n \times n)$ matrix and $\boldsymbol{\Pi} = -\boldsymbol{B}^{-1}\boldsymbol{C}$. The matrices $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{\Sigma}$ are functions of a vector of parameters $\boldsymbol{\theta}$.

The joint distribution of $\boldsymbol{y}_t$ is given by

$$f(\boldsymbol{y}_t|\boldsymbol{z}_t, \boldsymbol{\theta}) = \mathrm{N}(\boldsymbol{\Pi}\boldsymbol{z}_t, \boldsymbol{\Omega}). \tag{1.3.3}$$

In formulating the data density of $\boldsymbol{Y}' = (\boldsymbol{y}_1 \cdots \boldsymbol{y}_T)$ we must take into account that $\boldsymbol{z}_t$ includes lagged $\boldsymbol{y}_t$. If $\boldsymbol{y}_t^0$ are the lagged endogenous variables in $\boldsymbol{z}_t$ and $\boldsymbol{z}_t^*$ are the purely exogenous variables then the data density, conditional with respect to initial conditions, is

$$f(\boldsymbol{y}_1 \cdots \boldsymbol{y}_T|\boldsymbol{y}_1^0, \boldsymbol{z}_1^* \cdots \boldsymbol{z}_T^*, \boldsymbol{\theta}) = \prod_{t=1}^{T} f(\boldsymbol{y}_t|\boldsymbol{y}_1^0, \boldsymbol{z}_t^*, \boldsymbol{\theta}). \tag{1.3.4}$$

From now on, for notational convenience, we will no longer make explicit the conditionality on $\boldsymbol{y}_1^0$, $\boldsymbol{z}_t^*$ but write

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{t=1}^{T} f(\boldsymbol{y}_t|\boldsymbol{z}_t, \boldsymbol{\theta}). \tag{1.3.5}$$

Consider partitioning $\boldsymbol{y}_t$ into

$$\boldsymbol{y}_t' = (\boldsymbol{y}_{1t}' : \boldsymbol{y}_{2t}') \quad , \quad \boldsymbol{Y} = (\boldsymbol{Y}_1 : \boldsymbol{Y}_2) \tag{1.3.6}$$

where $\boldsymbol{y}_{1t} = (n_1 \times 1)$, $\boldsymbol{y}_{2t} = (n_2 \times 1)$ and $n_1 + n_2 = n$. $\boldsymbol{Y}_2$ is now defined to be exogenous if and only if there exists a partitioning of $\boldsymbol{\theta}$ into $\boldsymbol{\theta} = (\boldsymbol{\theta}_1 : \boldsymbol{\theta}_2)$ such that

$$f(\boldsymbol{y}_t|\boldsymbol{z}_t, \boldsymbol{\theta}) = f(\boldsymbol{y}_{1t}|\boldsymbol{y}_{2t}, \boldsymbol{z}_t, \boldsymbol{\theta}_1) f(\boldsymbol{y}_{2t}|\boldsymbol{z}_t, \boldsymbol{\theta}_2) \tag{1.3.7}$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are not subject to any cross-restrictions and $\boldsymbol{\theta}_2$ may be regarded as nuisance parameters. Under these conditions the likelihood function $LL(\boldsymbol{\theta}|\boldsymbol{Y})$ factorises as

$$LL(\boldsymbol{\theta}|\boldsymbol{Y}) = \prod_{t=1}^{T} f(\boldsymbol{y}_{1t}|\boldsymbol{y}_{2t}, \boldsymbol{z}_t, \boldsymbol{\theta}_1) \prod_{t=1}^{T} f(\boldsymbol{y}_{2t}|\boldsymbol{z}_t, \boldsymbol{\theta}_2) = LL_1(\boldsymbol{\theta}_1) LL_2(\boldsymbol{\theta}_2) \tag{1.3.8}$$

and no relevant sample information is lost if we drop the factor $LL_2(\boldsymbol{\theta}_2)$. Note that in general the joint density (1.3.5) will not factorise since $\boldsymbol{y}_{2t}$ will not be independent of lagged values of $\boldsymbol{y}_{1t}$ in $\boldsymbol{z}_t$.

Let us partition the model (1.3.1) into two subsets of equations

$$\begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_{1t} \\ \boldsymbol{y}_{2t} \end{pmatrix} + \begin{pmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \end{pmatrix} \boldsymbol{z}_t = \begin{pmatrix} \boldsymbol{u}_{1t} \\ \boldsymbol{u}_{2t} \end{pmatrix}. \tag{1.3.9}$$

We want to know the conditions under which $\boldsymbol{y}_{2t}$ is exogenous and the likelihood function factorises. Richard has proved (Richard, 1979, Theorem 3.1) that for the general case of models which may be incomplete, the conditions (in our notation) are

$$\boldsymbol{B}_{21} = \boldsymbol{0} \tag{1.3.10}$$

and

$$\boldsymbol{B}_{11}\boldsymbol{\Omega}_{12} + \boldsymbol{B}_{12}\boldsymbol{\Omega}_{22} = \boldsymbol{0} \tag{1.3.11}$$

where

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}$$

is the reduced form covariance matrix. For the case of complete models we can translate the condition (1.3.11) into a condition on $\boldsymbol{\Sigma}$, the covariance matrix of the structural form since

$$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}' = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{12}' & \boldsymbol{\Omega}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}_{11}' & \boldsymbol{B}_{21}' \\ \boldsymbol{B}_{12}' & \boldsymbol{B}_{22}' \end{pmatrix} \tag{1.3.12}$$

so that

$$\begin{aligned} \boldsymbol{\Sigma}_{12} &= \boldsymbol{B}_{11}\boldsymbol{\Omega}_{11}\boldsymbol{B}_{21}' + \boldsymbol{B}_{12}\boldsymbol{\Omega}_{12}'\boldsymbol{B}_{21}' + \boldsymbol{B}_{11}\boldsymbol{\Omega}_{12}\boldsymbol{B}_{22}' + \boldsymbol{B}_{12}\boldsymbol{\Omega}_{22}\boldsymbol{B}_{22}' \\ &= (\boldsymbol{B}_{11}\boldsymbol{\Omega}_{12} + \boldsymbol{B}_{12}\boldsymbol{\Omega}_{22})\boldsymbol{B}_{22}' \end{aligned}$$

by (1.3.10). Thus (1.3.11) becomes for complete models

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = \boldsymbol{0}. \tag{1.3.13}$$

The conditions (1.3.10) and (1.3.13) are equivalent to the condition that the model (1.3.9) is block recursive.

The log-likelihood function for the model is

$$L(\boldsymbol{A},\boldsymbol{\Sigma}) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}| + T\log||\boldsymbol{B}|| - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}'\boldsymbol{U}) \tag{1.3.14}$$

where

$$\boldsymbol{U}' = (\boldsymbol{u}_1 \cdots \boldsymbol{u}_T) = \begin{pmatrix} \boldsymbol{U}_1' \\ \boldsymbol{U}_2' \end{pmatrix}$$

$$\begin{aligned} L(\boldsymbol{A},\boldsymbol{\Sigma}) = &-\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log\begin{vmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{\Sigma}_{22} \end{vmatrix} + T\log\left\|\begin{matrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{0} & \boldsymbol{B}_{22} \end{matrix}\right\| \\ &- \frac{1}{2}\operatorname{tr}\left(\begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{U}_1'\boldsymbol{U}_1 & \boldsymbol{U}_1'\boldsymbol{U}_2 \\ \boldsymbol{U}_2'\boldsymbol{U}_1 & \boldsymbol{U}_2'\boldsymbol{U}_2 \end{pmatrix}\right) \end{aligned} \tag{1.3.15}$$

$$= \left\{\frac{-n_1T}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}_{11}| + T\log||\boldsymbol{B}_{11}|| - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{U}_1'\boldsymbol{U}_1)\right\} \tag{1.3.16}$$

$$+ \left\{\frac{-n_2T}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}_{22}| + T\log||\boldsymbol{B}_{22}|| - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{U}_2'\boldsymbol{U}_2)\right\}$$

$$= L_1(\boldsymbol{B}_{11},\boldsymbol{B}_{12},\boldsymbol{C}_1,\boldsymbol{\Sigma}_{11}) + L_2(\boldsymbol{B}_{22},\boldsymbol{C}_2,\boldsymbol{\Sigma}_{22}) \tag{1.3.17}$$

which does indeed factorise as in (1.3.8). It follows that the two factors can be estimated independently of each other, and if the parameters $(\boldsymbol{B}_{22},\boldsymbol{C}_2,\boldsymbol{\Sigma}_{22})$ are nuisance parameters the factor L_2 can be dropped.

The conditions (1.3.10) and (1.3.13) give us a straightforward test for the exogeneity of the $\boldsymbol{y}_2$ variables. First we estimate the unrestricted model (1.3.14). We can concentrate out the $\boldsymbol{\Sigma}$ matrix in the usual way to give a concentrated log-likelihood function

$$L^*(\boldsymbol{A},\boldsymbol{\Sigma}) = -\frac{nT}{2}\log(2\pi) - \frac{nT}{2} - \frac{T}{2}\log\left|\frac{\boldsymbol{AX}'\boldsymbol{XA}'}{T}\right| + T\log||\boldsymbol{B}|| \tag{1.3.18}$$

together with

$$\boldsymbol{\Sigma} = \frac{\boldsymbol{AX}'\boldsymbol{XA}'}{T} \tag{1.3.19}$$

and then maximise (1.3.18). Let the maximum be L^+. We then reestimate the model imposing the restrictions (1.3.10) amd (1.3.13). Note that we can concentrate $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ out of (1.3.16)

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\Sigma}_{11}^{-1}} &= \frac{T}{2}\boldsymbol{\Sigma}_{11} - \frac{1}{2}\boldsymbol{A}_1\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}_1' = \boldsymbol{0} \\ &\Longrightarrow \boldsymbol{\Sigma}_{11} = \frac{1}{T}\boldsymbol{A}_1\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}_1' \end{aligned} \tag{1.3.20}$$

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\Sigma}_{22}^{-1}} &= \frac{T}{2}\boldsymbol{\Sigma}_{22} - \frac{1}{2}\boldsymbol{A}_2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}_2' = \boldsymbol{0} \\ &\Longrightarrow \boldsymbol{\Sigma}_{22} = \frac{1}{T}\boldsymbol{A}_2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}_2' \end{aligned} \tag{1.3.21}$$

giving the concentrated log-likelihood function

$$L^{**}(\boldsymbol{A}_1, \boldsymbol{A}_2; \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22}) = -\frac{nT}{2}\log(2\pi) - \frac{nT}{2} - \frac{T}{2}\log\left|\frac{\boldsymbol{A}_1\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}_1'}{T}\right| + T\log||\boldsymbol{B}_{11}|| - \frac{T}{2}\log\left|\frac{\boldsymbol{A}_2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}_2'}{T}\right| + T\log||\boldsymbol{B}_{22}|| \tag{1.3.22}$$

where

$$\begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{C}_1 \\ \boldsymbol{0} & \boldsymbol{B}_{22} & \boldsymbol{C}_2 \end{bmatrix}.$$

Let L^{++} be the maximum of (1.3.22). Then on the null hypothesis that $\boldsymbol{y}_{2t}$ is exogenous

$$\lambda = 2(L^{+} - L^{++}) \sim_a \chi^2(q_1 + q_2) \tag{1.3.23}$$

where q_1 is the number of unrestricted parameters in $\boldsymbol{B}_{21}$ in model (1.3.18) and q_2 is the number of parameter restrictions corresponding to the condition (1.3.13) $= n1 \cdot n2$. (Not $2 \cdot n1 \cdot n2$ since $\boldsymbol{\Sigma}$ is already restricted to be symmetric). In fact often we may want to impose $\boldsymbol{B}_{21} = \boldsymbol{0}$ to identify the unrestricted model even before we test for exogeneity. In this case $q_1 = 0$ and we test only the $\boldsymbol{\Sigma}$ restrictions.

We can now go on to consider testing for switches in exogeneity in a multiple regime model using this test. If there are no constant parameters in different regimes then each may be estimated and tested separately. However, it is more reasonable to suppose that a subset of equations remain constant across regimes in which case we have cross-regime parameter restrictions. Then we must set up the joint likelihood function for all the regimes and maximise this subject to these parameter restrictions between regimes. Then we can test the exogeneity restrictions for each regime separately. This is the topic of the next section.

However, we want first to give an interpretation of the model we have been developing. We can consider it as an economy which is controlled by a policymaker. Then the first set of structural equations describes the economy on which control is forced and the second set the control process itself. The equations for $\boldsymbol{y}_{2t}$ are in fact reaction functions for the control instruments and by condition (1.3.10) the control authorities must make their plans solely on the basis of predetermined variables. Further, if the instruments are to be exogenous, condition (1.3.13) must also be satisfied. Switching between control instruments leads to switches in the exogeneity/endogeneity of variables.

The simplest illustration is an equilibrium model of a market where the supply of the commodity is under the control of the authorities. The authorities have two possible alternatives: they can either control the commodity price letting the market determine the quantity traded or they can control the supply and let the market determine the clearing price. We assume that the demand relationship is not changed by this regime switch. Then, taking a simple linear demand function we have, for the first regime, the demand equation

$$q_t = a_0 + a_1 p_t + v_t \quad , \quad \mathrm{E}(v_t^2) = \sigma_v^2 \tag{1.3.24}$$

together with the authority reaction function for price

$$p_t = a_2 + a_3 z_t + w_t. \tag{1.3.25}$$

This reaction function depends solely on the single predetermined variable z_t.

In the second regime the authorities switch to controlling the commodity supply and so the demand relationship (1.3.24) now determines the price and we renormalise it as

$$p_t = b_0 + b_1 q_t + \varepsilon_t \quad , \quad \mathrm{E}(\varepsilon_t^2) = \sigma_\varepsilon^2 \tag{1.3.26}$$

and suppose for simplicity that the same predetermined variable z_t continues to determine the new reaction function which is then

$$q_t = b_2 + b_3 z_t + \eta_t. \tag{1.3.27}$$

The constancy of the demand relation across regimes means that we have restrictions between the parameters in each regime which are

$$(b_0, b_1, \sigma_\varepsilon^2) = \left(-\frac{a_0}{a_1}, \frac{1}{a_1}, \frac{1}{(a_1)^2}\sigma_v^2 \right). \tag{1.3.28}$$

Full information maximum likelihood estimation requires maximising the joint likelihood function of the two regimes subject to all prior restrictions on the parameters which include these non-linear cross-regime restrictions.

1.4 FIML Estimation

In this section we consider the FIML estimation of multiple regime models in which we want to be able to test for switches in exogeneity. We have r regimes which we assume, without loss of generality, to have been operating sequentially. For regime i the model is

$$A^i(\theta)x_t = u_t \quad , \quad t \in I_i. \tag{1.4.1}$$

The switchpoints between regimes are assumed to be known. the log-likelihood for regime i is

$$\begin{aligned} L_i(\boldsymbol{\theta}) = -\frac{nT_i}{2}\log(2\pi) - \frac{T_i}{2}\log\left|\boldsymbol{\Sigma}^i(\boldsymbol{\theta})\right| + T_i \log\left|\left|\boldsymbol{B}^i(\boldsymbol{\theta})\right|\right| \\ -\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^i(\boldsymbol{\theta})^{-1}\boldsymbol{A}^i(\boldsymbol{\theta})\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{A}^i(\boldsymbol{\theta})') \end{aligned} \tag{1.4.2}$$

where

$$\boldsymbol{X}' = (\boldsymbol{x}_1 \cdots \boldsymbol{x}_T) = (\boldsymbol{X}_1' : \cdots : \boldsymbol{X}_r'),$$

and the joint log-likelihood for all r regimes is

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{r} L_i(\boldsymbol{\theta}). \tag{1.4.3}$$

The elements of the coefficient matrix $\boldsymbol{A}^i$ and the covariance matrix $\boldsymbol{\Sigma}^i$ of each regime are functions of a common vector of parameters $\boldsymbol{\theta}$. This is because we hypothesise that some relationships are common to more than one regime so that there are cross-restrictions between some elements of $\boldsymbol{A}^i$ and $\boldsymbol{\Sigma}^i$ for different regimes. When there are no cross-restrictions we can partition $\boldsymbol{\theta}$ so that

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{r} L_i(\boldsymbol{\theta}_i) \tag{1.4.4}$$

and the r vectors $\boldsymbol{\theta}_i$ have no elements in common. It follows that each regime can be estimated separately by standard FIML programs so these models are not of interest to us.

From (1.4.3) we have that

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{r} \frac{\partial L_i}{\partial \boldsymbol{\theta}} \quad , \quad \text{and} \quad \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{i=1}^{r} \frac{\partial^2 L_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}. \tag{1.4.5}$$

If the $\boldsymbol{\Sigma}^i$ are unrestricted we can concentrate them out of the log-likelihood as usual to get

$$L^*(\boldsymbol{\theta}^*) = \sum_{i=1}^{r} L_i^*(\boldsymbol{\theta}^*) \tag{1.4.6}$$

where

$$\begin{aligned} L_i^*(\boldsymbol{\theta}^*) = & -\frac{nT_i}{2}\log(2\pi) - \frac{nT_i}{2} - \frac{T_i}{2}\log\left|\frac{1}{T_i}\boldsymbol{A}^i(\boldsymbol{\theta}^*)\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{A}^i(\boldsymbol{\theta}^*)'\right| \\ & + T_i \log\left|\left|\boldsymbol{B}^i(\boldsymbol{\theta}^*)\right|\right| \end{aligned} \tag{1.4.7}$$

and

$$\widetilde{\boldsymbol{\Sigma}}^i = \frac{1}{T_i}\boldsymbol{A}^i\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{A}^{i\prime}. \tag{1.4.8}$$

We have specified nothing, so far, about the nature of the mappings from $\boldsymbol{\theta}$ to the elements of $\boldsymbol{A}^i$ and $\boldsymbol{\Sigma}^i$. If these mappings are non-linear then we say that the model is non-linear in parameters. We now show that by choosing the normalisation of the model in each regime we can keep the model linear in parameters.

We want to ensure that the parameters of each regime are identified before we consider imposing cross-regime restrictions. This is because we will later want to test these extra restrictions. We will consider only exclusion restrictions on $\boldsymbol{A}^i$ (the only available form of restriction in many standard FIML packages) where some elements of $\boldsymbol{A}^i$ are set a priori to zero. Let $\boldsymbol{\Phi}^k$ be a diagonal matrix defined by

$$\begin{aligned} \phi_{jj}^k &= 1 \text{ if } a_{kj}^i = 0 \\ &= 0 \text{ otherwise.} \end{aligned} \tag{1.4.9}$$

Then a necessary and sufficient condition for the kth equation to be identified is that

$$\operatorname{rank}(\boldsymbol{A}^i\boldsymbol{\Phi}^k) = n - 1 \tag{1.4.10}$$

where n is the number of equations in the system. For the complete model to be identified condition (1.4.10) must be satisfied for each equation ($k = 1, \cdots, n$) of each regime ($i = 1, \cdots, r$).

Condition (1.4.10) uniquely identifies the parameters of each equation only up to multiplication by a scalar (see for example Schmidt (1976)). To eliminate the remaining indeterminacy we need to specify in addition some normalisation rule. Conventionally we normalise along the diagonal setting

$$B_{kk}^i = -1 \quad (k = 1, \cdots, n). \tag{1.4.11}$$

However, this particular normalisation rule is quite arbitrary and we are free to normalise each equation on any variable with a non-zero coefficient. FIML procedures are invariant to the rule chosen.

In fact by choosing a different model normalisation in different regimes we can make each regime linear in parameters $\boldsymbol{\theta}$. We can illustrate this with the two equation model developed at the end of the last section. The two regimes with the conventional normalisation can be written as

$$\begin{pmatrix} -1 & a_1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} + \begin{pmatrix} a_0 & 0 \\ a_2 & a_3 \end{pmatrix} \begin{pmatrix} c \\ z_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^1_{1t} \\ \boldsymbol{u}^1_{2t} \end{pmatrix} \tag{1.4.12}$$

$$\begin{pmatrix} -1 & 0 \\ b_1 & -1 \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} + \begin{pmatrix} b_2 & b_3 \\ b_0 & 0 \end{pmatrix} \begin{pmatrix} c \\ z_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^2_{1t} \\ \boldsymbol{u}^2_{2t} \end{pmatrix} \tag{1.4.13}$$

with covariance matrices

$$\boldsymbol{\Sigma}^1 = \begin{pmatrix} \sigma^1_{11} & \sigma^1_{12} \\ \sigma^1_{21} & \sigma^1_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^2 = \begin{pmatrix} \sigma^2_{11} & \sigma^2_{12} \\ \sigma^2_{21} & \sigma^2_{22} \end{pmatrix} \quad \text{respectively.}$$

The cross-restrictions between parameters are then

$$(b_0, b_1, \sigma^2_{11}) = \left(-\frac{a_0}{a_1}, \frac{1}{a_1}, \frac{1}{(a_1)^2}\sigma^1_{11} \right) \tag{1.4.14}$$

which are non-linear. However, if we keep the demand relation normalised on q_t in both regimes, instead of (1.4.13) we have

$$\begin{pmatrix} -1 & b_1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} + \begin{pmatrix} b_0 & 0 \\ b_2 & b_3 \end{pmatrix} \begin{pmatrix} c \\ z_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^2_{1t} \\ \boldsymbol{u}^2_{2t} \end{pmatrix} \tag{1.4.15}$$

and the cross-restrictions are simply

$$(b_0, b_1, \sigma^2_{11}) = (a_0, a_1, \sigma^1_{11}). \tag{1.4.16}$$

In general, by preserving the normalisation of equations which remain constant across regimes, which means changing the normalisation of the system from the conventional diagonal normalisation, we avoid non-linear restrictions between the regimes. This proves to be convenient. There is no inherent difficulty in treating non-linearity in parameters if we wish to do so (and we might wish to allow for other non-linear restrictions apart from those arising from the constancy of some relationships across regimes) but this is not necessary. From now on we will assume the mapping from $\boldsymbol{\theta}$ to A^i and $\boldsymbol{\Sigma}^i$ to be linear for each regime.

The numerical problem to be solved, then, is the maximisation of (1.4.3) with $L_i(\boldsymbol{\theta})$ defined by (1.4.2) and a linear mapping from $\boldsymbol{\theta}$ to A^i and $\boldsymbol{\Sigma}^i$. Clearly the simplest approach would have been to write a special program for the problem. However, it seemed more useful to attempt to generalise an existing FIML program, to find out how difficult it might prove to implement the necessary modifications within an existing framework. We have seen that cross-regime restrictions may involve the $\boldsymbol{\Sigma}^i$ matrices and in any case we need to impose block diagonality of $\boldsymbol{\Sigma}^i$ for Richard's test for exogeneity. For these reasons the program GENRAM, written by David Hendry (Hendry (1978a)), was chosen because it already included the option of some restrictions on the covariance matrix.

When we have cross-restrictions on $\boldsymbol{\Sigma}^i$ we cannot in general concentrate the matrix out of the likelihood function and we have to deal with the problem that the elements of $\boldsymbol{\Sigma}^i$ are already constrained by the restrictions of symmetry and positive-definiteness.

For the numerical optimisation routine the parameters must be unconstrained so that we cannot work directly with the elements of $\boldsymbol{\Sigma}^i$. We make use of the fact that we can factorise $\boldsymbol{\Sigma}^i$ as

$$\boldsymbol{\Sigma}^i = \boldsymbol{H}^i \boldsymbol{H}^{i\prime} \tag{1.4.17}$$

where $\boldsymbol{H}^i$ is a lower triangular matrix with $n(\frac{n+1}{2})$ non-zero elements which are unconstrained. Now if we wish to impose restrictions on $\boldsymbol{\Sigma}^i$ these must be translated into restrictions on $\boldsymbol{H}^i$.

We wish to consider two types of restriction. Firstly we have cross-regime covariance restrictions arising from the constancy of some relationships across regime shifts. We assume that the first block of n_1^* equations remain constant over all regimes. Let us partition $\boldsymbol{\Sigma}^i$ and $\boldsymbol{H}^i$ for each regime as

$$\boldsymbol{\Sigma}^i = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^i & \boldsymbol{\Sigma}_{12}^i \\ \boldsymbol{\Sigma}_{21}^i & \boldsymbol{\Sigma}_{22}^i \end{pmatrix} \quad , \quad \boldsymbol{H}^i = \begin{pmatrix} \boldsymbol{H}_{11}^i & \mathbf{0} \\ \boldsymbol{H}_{21}^i & \boldsymbol{H}_{22}^i \end{pmatrix} \tag{1.4.18}$$

where we note that $\boldsymbol{H}_{11}^i$ and $\boldsymbol{H}_{22}^i$ are square lower triangular matrices. Then the covariance restriction is that $\boldsymbol{\Sigma}_{11}^i = \boldsymbol{\Sigma}_{11}^j$, $\forall i, j$ which implies $\boldsymbol{H}_{11}^i = \boldsymbol{H}_{11}^j$, $\forall i, j$.

Secondly we want to apply Richard's test for the exogeneity of the control process for each regime. We saw in section 1.3 that this implies block diagonality of $\boldsymbol{\Sigma}^i$. There for a single regime we were able to analytically concentrate the likelihood but when we also have to deal with cross-regime restrictions this is no longer possible and we must deal with the restrictions explicitly. Let (1.4.18) now correspond to a partitioning into blocks of n_1 and n_2 equations ($n_1^* \leq n_1$) where the last block represents the equations of the control process, Then Richard's covariance restriction $\boldsymbol{\Sigma}_{12}^i = \boldsymbol{\Sigma}_{21}^i = \mathbf{0}$, $\forall i$ becomes $\boldsymbol{H}_{21}^i = 0$, $\forall i$.

When there are no restrictions the number of unconstrained $\boldsymbol{\Sigma}$ parameters to be estimated is $rn(\frac{n+1}{2})$. Cross-regime restrictions subtract $(r-1)n_1^*(\frac{n_1^*+1}{2})$ parameters and exogeneity restrictions another rn_1n_2 parameters. However, from the point of view of the order of the numerical problem to be solved, imposing $\boldsymbol{\Sigma}$ restrictions increases the size by

$$\frac{r}{2}\left[n_1(n_1+1) + n_2(n_2+1)\right] - \frac{r-1}{2}\left[n_1^*(n_1^*+1)\right] \tag{1.4.19}$$

extra parameters. This will become prohibitively expensive for all but small values of n and r.

We can now summarise the extra information needed by the modified program. Let

$$\boldsymbol{\psi}^i = \text{vec}\, \boldsymbol{A}^i \quad \text{and} \quad \boldsymbol{\phi}^i = \boldsymbol{S}^i \boldsymbol{\psi}^i \tag{1.4.20}$$

where $\boldsymbol{S}^i$ denotes a selection matrix picking out unrestricted elements only. Note that since there are no restrictions between the parameters of $\boldsymbol{A}^i$ and $\boldsymbol{\Sigma}^i$ we can partition $\boldsymbol{\theta}$ into $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1' : \boldsymbol{\theta}_2')$ and write $\boldsymbol{A}^i$ $(\boldsymbol{\theta}_1)$ and $\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_2)$. Then the additional information is:

1. (i) The number of regimes r.
   (ii) Their periods of operation I_i, $i = 1, \cdots, r$.
   (iii) The size of the block of constant equations n_1^*.
   (iv) The size of the block of control equations n_2.

2. For each regime:

(i) The mapping from $\boldsymbol{\psi}^i$ to $\boldsymbol{\phi}^i$ (i.e. which elements of $\boldsymbol{A}^i$ are unrestricted)

(ii) The mapping from $\boldsymbol{\phi}^i$ to $\boldsymbol{\theta}_1$ (i.e. which element of $\boldsymbol{\theta}_1$ corresponds to each unrestricted element of $\boldsymbol{A}^i$)

(iii) The model normalisation (i.e. which elements of $\boldsymbol{A}^i$ are preset to -1)

3. Initial values for the parameter vector $\boldsymbol{\theta}_1$.

The program generates its own initial values for $\boldsymbol{\theta}_2$, the parameters of $\boldsymbol{\Sigma}^i$ and also defines the mapping from $\boldsymbol{\Sigma}^i$ to $\boldsymbol{\theta}_2$ on the basis of 1(iii) and 1(iv).

Now we want to develop tests for the validity of the restrictions we impose on the model. The most general model we consider is a multiple regime model with no cross-regime restrictions and no restrictions on the covariance matrices. We can then find a sequential testing procdure of two nested hypotheses. First we test the hypothesis of the constancy of some chosen sub-block of equations over the regime shifts. Then if this test does not reject the null hypothesis we can go on to test the hypothesis of the exogeneity of the control equations in each regime using Richard's test developed in section 1.3. This second test then becomes a test of whether or not the economy was controlled.

If, however, the first test rejects the hypothesis of a constant sub-block of equations then each regime is a completely separate model. Of course we can still test for exogeneity in each regime but even if this test fails to reject the null in each case we cannot really interpret this as the authorities switching between instruments in their control of the economy, since the economy being controlled changes with the regime.

The general unrestricted model can be estimated easily using the concentrated log-likelihood function defined by (1.4.6) – (1.4.8). Let the maximum of this log-likelihood be L^+. Then we reestimate the model imposing all the cross-regime restrictions. Initial values for the parameters of the covariance matrices can be derived from the solution of the unrestricted model. Let L^{++} be the maximum of this log-likelihood and let p be the number of unrestricted parameters in the first n_1^* rows of the $\boldsymbol{A}^i$ matrix. The first test is then

$$2(L^+ - L^{++}) \sim_a \chi^2_{((r-1)(p+n_1^*(\frac{n_1^*+1}{2})))}. \tag{1.4.21}$$

Lastly we reestimate the second model imposing the block diagonality on the $\boldsymbol{\Sigma}^i$ matrices for the exogeneity test. The joint test for exogeneity in each regime is

$$2(L^{++} - L^{+++}) \sim_a \chi^2_{(rn_1n_2)} \tag{1.4.22}$$

where L^{+++} is the maximum log-likelihood of the third model. We can also apply the exogeneity test separately to each regime. Let L_i^{++} and L_i^{+++} be the corresponding maxima for regime i. Then

$$2(L_i^{++} - L_i^{+++}) \sim_a \chi^2_{(n_1n_2)}. \tag{1.4.23}$$

1.5 A Model of the Money Market

As an illustration of the general approach already outlined it was decided to estimate a simple two-equation model of the money market to try and model the change in regime of the introduction of Competition and Credit Control (CCC) in October 1971.

Firstly we must explain the economic background to the change. (For a good account of British monetary policy before 1971 see the chapter by Goodhart in Goodhart (1980a). Goodhart (1980b) gives a very interesting inside view of the Bank of England thinking behind the change and early experience of operating the new regime.) Before 1971 the control of short term interest rates was the main concern of the monetary authorities. The key to their control was the Bank Rate which was an institutional rate set directly by the Bank of England. Other short term interest rates moved closely in line with it; in particular a Clearing banks cartel existed which linked the interest rates on Advances and time deposits directly to Bank Rate. Changes in the Bank Rate came to have a strong "announcement effect" signalling the direction of monetary policy. On the other hand there was no attempt to directly control the main monetary aggregates (which only came to be regarded by the Bank as important indicators towards the end of the 1960s) although starting in 1964 there was some attempt to reduce bank lending to the private sector, especially persons, in a series of formal "Requests" to the Clearing banks. However, these were quantitative ceilings imposed on the market rather than operating through it. We can distinguish four main targets of monetary policy: the balance of payments, the rate of unemployment, the rate of inflation and, in the longer term, the growth rate. Of these the balance of payments was perhaps the most important. (For a statement of the aims of monetary policy by the authorities themselves see the Radcliffe Report Report (1959)).

The effect of the 1971 changes was to switch emphasis to control of the monetary aggregates. In 1972 the Bank Rate was replaced by a Minimum Lending Rate (MLR) which is determined by the market (The Bank sets it at 0.5 above the current Treasury bill rate). The Clearing bank cartel was abolished and the banks encouraged to compete for funds, thus allowing interest rates to move freely in response to market presures. In this way control over interest rates was relaxed in favour of control over the monetary aggregates, although the mechanism of control was viewed as being through portfolio adjustment in response to changes in asset relative prices. (In the new spirit of free competition the formal ceilings on lending were abandoned although in 1973 they were reintroduced in a new guise as the Supplementary Special Deposit scheme or "corset".) The targets of monetary policy since 1971 have been essentially the same as before, with perhaps less emphasis on the balance of payments, although the managed floating of the exchange rate in 1972 has not alleviated all external balance problems.

We attempted to model this complicated policy change in a simple two regime model of two equations: the demand for money equation and a reaction function for the control variable (interest rate in the first regime and money supply in the second) although clearly, with such an oversimplified model, our results can only be very tentative. Previous work on reaction functions for the instruments of monetary policy for the U.K. has been done by Douglas Fisher in Fisher (1968, 1970) and also by Goodhart (in his appendix to Goodhart (1980a)). The reaction function treats a policy control variable as a function of the target variables which are the objects of control. The coefficients then measure the response of the control variable to changes in the target variables. For our policies, in order to test for the exogeneity of the control variables, our reaction functions must be purely predetermined and we may interpret this as reflecting the necessary lags in the reaction of the authorities.

Rather than try to estimate the model from the start by FIML it was decided to do preliminary single equation estimation to choose the dynamic specification of the equations. This allowed access to the much fuller diagnostic tests available in programs such as GIVE (Hendry (1978b)), especially tests for serial correlation, which was important since we wanted to avoid dealing with this problem in the full simultaneous estimation of the two regime model. Also the preliminary estimates obtained provided us with initial values to use in the FIML program. In determining the dynamic structure the methodology of Hendry and Mizon (1979) was adopted. Economic theory gives us little help in choosing dynamic specification and, rather than impose the dynamics a priori, we follow a consistent data-based approach. Starting from the most general hypothesis we are prepared to accept (generally a maximum lag length on all variables) we can then test for parameter restrictions in a series of Wald tests until we reach the most parsimonious description consistent with the data. There is no unique sequence for testing and the particular restrictions we test may often be governed by what seems intuitively plausible or appealing. However, this procedure should prevent us from imposing invalid restrictions.

One objection sometimes raised against this approach is the problem of multicollinearity in parameter estimates in the most general equation forms. In practice this did not greatly hinder the simplification process since, although t-ratios in these cases were in general quite small using conventional significance levels, the relative sizes of t-ratios were found to be still quite a good indication of the relative importance of lags. A more critical practical problem was degrees of freedom (especially for the reaction functions) which restricted the maximum lag length which could reasonably be estimated.

Seasonally unadjusted data was used. This avoided the problem of possible distortion of the true dynamic structure arising from using series separately seasonally adjusted using different filters (see Wallis (1974)). For the money demand equation it also gave us an interesting comparison with the results of Hendry and Mizon who used adjusted data.

The two endogenous variables used in the analysis were personal sector M3 and the Local Authority short term interest rate. It is sometimes argued that the appropriate monetary indicator is narrow money M1 rather than M3 which includes an interest bearing asset, time deposits. However, the view of the Bank at the time (see for example Goodhart (1980b)) was that, because of the ease of shifting between time deposit and current accounts, M1 could not be controlled, so that M3 was the appropriate indicator for control. Total M3 was tried but this was not very successful (even in an unrestricted specification with up to 8 quarter lags on all variables there was significant residual autocorrelation) and personal sector M3 was found to perform much better. This accords with the results of studies by the Bank of England (see Hacche (1974)). Finding a single interest rate with the dual property of being both monetary control variable and a measure of the true opportunity cost of holding money proved difficult. Several short rates were tried but finally the Local Authority rate was chosen as the best proxy.

The general unrestricted money demand equation took the form

$$\ln M_t = \sum_{j=0}^{J} (\alpha_j \ln Y_{t-j} + \beta_j \ln P_{t-j} + \gamma_j \ln r_{t-j} + \delta_j \ln M_{t-j-1}) + c_0 + \sum_{i=1}^{3} c_i q_{t-i} + e_t \tag{1.5.1}$$

Table 1.1: Equation 1.5.1 with $J = 5$. OLS 5603–7802

$j \backslash Vble$	$\ln M_{t-j-1}$	$\ln r_{t-j}$	$\ln Y_{t-j}$	$\ln P_{t-j}$
0	1.07 (.13)	.011 (.013)	.33 (.11)	.85 (.33)
1	−.28 (.18)	−.019 (.019)	−.15 (.13)	−.95 (.45)
2	.06 (.18)	.015 (.020)	.06 (.13)	−.03 (.47)
3	.20 (.19)	−.017 (.019)	−.13 (.13)	−.08 (.45)
4	−.26(.18)	.027 (.020)	.04 (.13)	.59 (.45)
5	.06 (.13)	−.012 (.015)	−.01 (.12)	−.29 (.29)
$\sum_j$	.85	.005	.14	.09
$c_0 = -.53\,(.25)$	$c_1 = -.001\,(.011)$	$c_2 = -.003\,(.010)$	$c_3 = .008\,(.011)$	
$R^2 = .9992$	$\widehat{\sigma} = .0159$	$\chi^2_{12} = 3.84$	$\chi^2_6 = 8.18$	$\chi^2_1 = 1.28$

(standard errors in parentheses)

where

$$\begin{aligned} M_t &= \text{personal sector M3} \\ Y_t &= \text{personal disposable income at constant 1975 prices} \\ P_t &= \text{the deflator for } Y_t \\ r_t &= \text{the Local Authority interest rate as a percentage} \end{aligned}$$

and q_{t-i}, $i = 1, \cdots, 3$ are the usual seasonal dummies. Data was available quarterly over the period 5501-7802. Estimating (1.5.1) by OLS over the full data period with $J = 5$ gave the results in Table 1.1. The three χ^2 statistics quoted are tests for residual autocorrelation: χ^2_{12} is the Box-Pierce random residual correlogram test (see Pierce (1971)), χ^2_6 is a Lagrange multiplier test for serial correlation up to the 6th order, χ^2_1 is the squared Durbin h-statistic for a first order AR process (Durbin (1970)). These statistics show no evidence of dynamic misspecification or serial correlation in the residuals, confirming that choice of $J = 5$ is adequately general (as might be expected for quarterly data). Perhaps the most striking feature of the equation is the strong affirmation of unit price elasticity suggesting a reparameterisation with $\Delta \ln(M/P)$ as the dependent variable. Comparison of Table 1.1 with Table 1 in Hendry and Mizon (estimated with $J = 4$ over a much shorter period 6301-7503) shows much similarity in the lag patterns suggesting that the seasonal bias in using adjusted data may not be very important in this case. In particular there is the same $\Delta \ln r_t$ term with perverse sign, although in general the interest rate terms in our equation are much less significant (Hendry and Mizon used the consol yield)

To guard against possible simultaneous equation bias in using OLS, equation (1.5.1) was reestimated using instrumental variables for r_t. Note that in the complete two equation model if $\mathbf{\Sigma}$ is not diagonal then OLS on M_t will be inconsistent but if $\mathbf{\Sigma}$ is diagonal (Richard's condition satisfied) then OLS is maximum likelihood. Full 2SLS using as instruments all predetermined variables in the most general specification of both equations (i.e. including all lags up to $J = 5$) would have exceeded the program dimension limitations but it was possible to use $j = 1, \cdots, 4$ on all the additional instruments (giving 39 instruments in all). However, the results were virtually identical to Table 1.1 with the interest rate terms coming out even less significant.

Before attempting to simplify the unrestricted equation (1.5.1) it was decided to test its stability over the regime change in October 1971. It is well known that the introduction

Table 1.2: Solved Coefficients from Equation 1.5.2. OLS 5603-7802

$j \backslash Vble$	$\ln M_{t-j-1}$	$\ln r_{t-j}$	$\ln Y_{t-j}$	$\ln P_{t-j}$
0	1.10	.011	.27	.65
1	$-.14$	$-.011$	$-.23$	$-.75$
2	0.0	0.0	0.0	$-.41$
3	.26	0.0	0.0	.15
4	$-.26$	.009	0.0	.54
5	0.0	0.0	0.0	$-.14$
$\sum_j$	.96	.009	.04	.04

of CCC was followed by a very large expansion in M3 (most especially in company holdings but also to a lesser extent in personal sector holdings), an expansion not matched in the narrow money definition M1, and Hacche (1974) and Goodhart (1980b) reported a complete breakdown in the Bank of England's own forecasting equations. Since one of the effects of CCC was to encourage banks to compete for finance through the rates paid on deposit accounts, Hacche argued convincingly that this breakdown could be explained by the omission of own-rate terms from the Bank's demand equations, terms which would not have mattered before 1971 when the differential between deposit rates and the Bank Rate was constant. However, own-rate terms were not found to be significant either by Hacche or Artis and Lewis (1974) who explained the expansion of M3 as a disequilibrium excess supply of money. (In an earlier equation for total M3 we tested the importance of own-rates by including interest differential terms of the form $\ln(r^*/r)_{t-j}$ where r^* was the time deposit rate. However, these terms came up with perverse negative signs).

A Chow test was carried out on equation 1.5.1 reestimated over the period ending 7103. The statistic was $F_{(27,60)} = 2.62$, the 5 significance level for $F_{(24,60)}$ being 1.70, so that the hypothesis of parameter stability was rejected. This probably reflects the clearly inadequate proxying of interest effects in the equation.

Simplifying equation (1.5.1) we obtained

$$\begin{array}{llll} \Delta\ln(M/P)_t = & 27\Delta\ln Y_t & +.14\Delta\ln(M/P)_{t-1} & -.35\Delta\ln P_t \\ & (.08) & (.10) & (.20) \\ & -.55\Delta\ln P_{t-2} & -.40\Delta^2\ln P_{t-2} & +.26\Delta\ln(M/P)_{t-4} \\ & (.20) & (.19) & (.10) \\ & +.011\Delta\ln r_t & +.009\ln r_{t-4} & -.04\ln(M/PY)_{t-1} \\ & (.011) & (.008) & (.02) \end{array}$$

$$c_0 = -.17\ (.07) \quad c_1 = -.01\ (.01) \quad c_2 = -.01\ (.01) \quad c_3 = -.003\ (.005)$$
$$R^2 = .644 \quad \widehat{\sigma} = .0151 \quad \chi^2_{12} = 6.86 \quad \chi^2_6 = 4.51 \quad \chi^2_1 = .002. \tag{1.5.2}$$

The χ^2 tests still reject the hypothesis of serial correlation in the residuals and the F-ratio test for the restrictions imposed on (1.5.1) is $F_{(15,60)} = .533$. (The 5 significance level for $F_{(15,60)}$ is 1.84). The solved coefficients from (1.5.2) are recorded in Table 1.2 and these quite closely correspond with the coefficients on the unrestricted equation. However, like Hendry and Mizon's equation 21, the specification (1.5.2) was primarily chosen so as to include decision variables with sensible economic interpretations. In the short-run inflation has a negative influence on money demand through three inflation terms , including a term for the rate of acceleration of inflation $\Delta^2\ln P_{t-2}$. However, in long run equilibrium when all variables are growing along steady-state growth paths the equation exhibits unit elasticity with respect to both prices and income through the inverse velocity levels term $\ln(M/PY)_{t-1}$. This term also serves as an error feeedback correction mechanism in the short run, allowing agents to adjust from previous disequilibrium in the relationship betwen their real income and money holdings.

In the long run we would expect some levels interest rate effect. However, the coefficient on $\ln r_{t-4}$ in (1.5.2) has the "wrong" sign and is statistically insignificant, its value being approximately equal to the sum of the interest rate coefficients on the unrestricted equation. This again seems to reflect the unsatisfactory nature of the Local Authority rate (perhaps any single rate) as a proxy for the opportunity cost to the personal sector of holding money. It is clear that to have gone on to try and produce a fully satisfactory equation would have required much further work experimenting with various interest rates (possibly including several different rates) in an attempt to model the portfolio choice properly. For the purpose of this exercise it was decided to accept (1.5.2), however unsatisfactory, as the best equation we had been able to find.

In approaching the estimation of the reaction functions we had immediately to face the problem that with only 27 observations for the second regime (7104-7802) it was quite impossible to estimate an unrestricted function. Instead we chose the dynamic specification for the first regime on the basis of an unrestricted equation and then were forced to assume that this specification remained the same after the regime switch. We note that this assumption is a very strong one since we might expect the reaction lags of different control instruments to be very different.

The general hypothesis for the first regime was

$$\begin{aligned}\ln r_t = &\sum_{j=0}^{J}(a_j \ln(B/P)_{t-j} + b_j \ln U_{t-j} + d_j \ln P_{t-j} + e_j \ln E_{t-j})\\ &+\sum_{k=1}^{K}(f_k \ln r_{t-k} + g_k \ln M_{t-k}) + c_0 + \sum_{i=1}^{3} c_i q_{t-i} + \varepsilon_t \end{aligned} \tag{1.5.3}$$

where

U_t = the percentage rate of unemployment
B_t = the level of reserves
E_t = the Eurodollar rate.

The Eurodollar rate was included on the hypothesis that a long run objective of interest rate control would be to keep U.K. interest rates in line with foreign rates. Equation (1.5.3) was estimated over the period 5602-7103 taking $J = K = 4$ giving the results in Table 1.3. The χ^2 tests show no evidence of residual autocorrelation. Having little a priori theory to guide us to a sensible reparameterisation the equation was simplified by choosing the most important lags from Table 1.3. The resulting equation was

$$\begin{aligned}\Delta \ln r_t = \; & \underset{(2.10)}{2.17\Delta \ln P_t} \quad \underset{(2.35)}{-3.78\Delta \ln P_{t-3}} \quad \underset{(.16)}{-.53\Delta \ln U_{t-1}}\\ & \underset{(.17)}{-.37\Delta \ln U_{t-3}} \quad \underset{(.13)}{-.47\Delta \ln r_{t-2}} \quad \underset{(.10)}{-.41\Delta_3 \ln(B/P)_{t-1}}\\ & \underset{(.81)}{-.19\Delta \ln(M/P)_{t-1}} \quad \underset{(.008)}{+.008 \ln(r/E_{-1})_{t-1}}\end{aligned}$$

$$\begin{gathered} c_0 = .07\ (.04) \quad c_1 = -.05\ (.06) \quad c_2 = -.05\ (.04) \quad c_3 = -.07\ (.07)\\ R^2 = .526 \quad \widehat{\sigma} = .0954 \quad \chi^2_{12} = 12.28 \quad \chi^2_6 = 2.96 \quad \chi^2_1 = .19. \end{gathered} \tag{1.5.4}$$

Table 1.3: Equation 1.5.3 with $J = K = 4$. OLS 5602–7103

$j\backslash Vble$	$\ln r_{t-j-1}$	$\ln M_{t-j-1}$	$\ln P_{t-j}$	$\ln U_{t-j}$	$\ln(B/P)_{t-j}$	$\ln E_{t-j}$
0	.83 (.18)	−.22 (1.19)	6.01 (3.2)	−.03 (.24)	−.09 (.19)	.17 (.18)
1	−.35 (.22)	.89 (1.36)	−7.51 (4, 4)	−.35 (.39)	−.42 (.25)	−.10 (.29)
2	.19 (.23)	.70 (1.39)	−.07 (4.2)	.49 (.41)	.01 (.23)	.21 (.31)
3	.27 (.17)	−1.20 (1.05)	−5.24 (4.0)	−.72 (.43)	−.07 (.25)	−.37 (.29)
4	—	—	6.22 (3.3)	.48 (.26)	.33 (.24)	.24 (.19)
$\sum_j$	.94	.17	−.59	−.13	−.24	.15

$c_0 = 1.40\,(2.47)$ $\quad c_1 = -.11\,(.13)$ $\quad c_2 = -.14\,(.12)$ $\quad c_3 = -.02\,(.14)$

$R^2 = .916$ $\quad \widehat{\sigma} = .1006$ $\quad \chi^2_{12} = 4.79$ $\quad \chi^2_6 = 6.14$ $\quad \chi^2_1 = .03$

Table 1.4: Solved Coefficients from Equation 1.5.4. OLS 5602–7103

$j\backslash Vble$	$\ln r_{t-j-1}$	$\ln M_{t-j-1}$	$\ln P_{t-j}$	$\ln U_{t-j}$	$\ln(B/P)_{t-j}$	$\ln E_{t-j}$
0	1.008	−.19	2.17	0.0	0.0	0.0
1	−.47	.19	−1.98	−.53	−.41	−.008
2	.47	0.0	−.19	.53	0.0	0.0
3	0.0	0.0	−3.78	−.37	0.0	0.0
4	0.0	0.0	3.78	.37	.41	0.0
$\sum_j$	1.008	0.0	0.0	0.0	0.0	−.008

The F-test on the imposed restrictions gave $F_{(20,30)} = .746$ (5 significance level $F_{(20,30)} = .1.93$). Although this test failed to reject the restrictions imposed in (1.5.4) the solved coefficients in Table 1.4 are rather different from those in Table 1.3. This is especially true of the inflation terms which have changed considerably in magnitude. The last term in (1.5.4) was a levels effect intended to capture a long run relationship between U.K. and foreign interest rates. However, it was found to be insignificant.

The specification (1.5.4) was reestimated over the period 7104-7802 with $\Delta\ln(M/P)$ as the dependent variable. This gave the equation

$$\begin{array}{llll}
\Delta\ln(M/P)_t = & -.95\Delta\ln P_t & -.29\Delta\ln P_{t-3} & -.04\Delta\ln U_{t-1} \\
 & (.18) & (.19) & (.03) \\
 & -.05\Delta\ln U_{t-3} & -.004\Delta\ln r_{t-2} & -.01\Delta_3\ln(B/P)_{t-1} \\
 & (.03) & (.01) & (.005) \\
 & +.11\Delta\ln(M/P)_{t-1} & -.02\ln(r/E_{-1})_{t-1} & \\
 & (.16) & (.008) &
\end{array}$$

$$c_0 = .06\ (.01) \quad c_1 = -.03\ (.008) \quad c_2 = .0004\ (.007) \quad c_3 = -.02\ (.009)$$

$$R^2 = .912 \quad \widehat{\sigma} = .0098 \quad \chi^2_{12} = 16.72. \tag{1.5.5}$$

Caution must be exercised in the interpretation of this equation because of the number of degrees of freedom (only 15). This means that the very good fit (high R^2, small $\widehat{\sigma}$) is probably spurious. Also the χ^2_{12} statistic is large compared with our previous equations, although still not significant. (The other χ^2 tests were unfortunately not available for this equation). Comparing with (1.5.4) we see that, although most of the coefficients have changed sign (as expected) a few (as with equation (1.5.4)) seem to have perverse signs.

These reaction functions are not very satisfactory, in particular because of the ad hoc way in which the dynamic specification was chosen from the unrestricted equation (1.5.3). Nevertheless, our purpose being illustrative only, the specifications (1.5.4) and (1.5.5) were used, together with the demand equation specification (1.5.2) for the purpose of a FIML estimation of the complete model to illustrate the application of Richard's test for exogeneity.

The complete model was estimated three times to carry out the two likelihood ratio tests described at the end of section 1.4. Firstly, the log-likelihood was maximised without any cross-restrictions, allowing the coefficients on the money demand equation to take different values in the two regimes. (This corresponds to L^+ in section 1.4). Then the model was reestimated constraining the coefficients on the demand equation to be the same for both regimes. (This corresponds to L^{++}). Finally, the second model was reestimated imposing Richard's exogeneity restrictions which in this model implies constraining the two covariance matrices $\mathbf{\Sigma}^i$ to be diagonal. (This corresponds to L^{+++}). The exact log-likelihoods obtained were

$$L^+ = 292.06 \quad , \quad L^{++} = 276.69 \quad , \quad L^{+++} = 267.23 \tag{1.5.6}$$

so that the two χ^2 tests corresponding to (1.4.21) and (1.4.22) were

$$\chi^2_{(9)} = 30.74 \quad \text{and} \quad \chi^2_{(2)} = 18.92 \tag{1.5.7}$$

respectively.

The first test convincingly rejects the hypothesis of a stable money demand function, confirming the result of the Chow test done on the single equation. (This test has only 9 degrees of freedom because the additional cross-restriction on the covariance terms that $\sigma^1_{11} = \sigma^2_{11}$ was not imposed). We have argued above that rejection of the null on this first test means that we cannot regard Richard's joint exogeneity test as a test of whether or not the economy was being controlled. Nevertheless, it is still interesting that Richard's test strongly rejects the hypothesis that the instruments of control were exogenous for this model imposing a stable money demand function. The individual χ^2 tests on each regime (corresponding to (1.4.23)) were

$$\chi^2_{(1)} = 13.52 \quad \text{and} \quad \chi^2_{(1)} = 5.40 \tag{1.5.8}$$

respectively, showing that the rejection of the null was much stronger for the first regime.

We also applied Richard's test for exogeneity to each regime estimately separately (allowing a shift in the money demand equation between the two regimes). This gave the statistics

$$\chi^2_{(1)} = 2.46 \quad \text{and} \quad \chi^2_{(1)} = .114 \tag{1.5.9}$$

showing remarkably different results from the previous tests. Taking these results at face value the implication is clear: in both regimes separately the control instruments appear to have been exogenous but the regime change itself was accompanied by a shift in the money demand function. We should be cautious about putting too much weight on this conclusion without further evidence (in particular it would have been interesting to test the exogeneity of the money supply in the first regime and the interest rate in the second). Before accepting the hypothesis of a shift in the money demand function we would like to have done more work with different relative interest rates to model the portfolio decision more carefully. Nevertheless, the hypothesis of a change in the behaviour of the personal sector cannot be ruled out, especially since one explicit purpose of CCC was to bring about a change in the behaviour of the banks.

The FIML estimates for the second model (with the demand equation constrained to be the same in both regimes) were as follows:

$$\begin{array}{llll} \Delta\ln(M/P)_t = & 31\Delta\ln Y_t & -.003\Delta\ln(M/P)_{t-1} & +.17\Delta\ln P_t \\ & (.08) & (.02) & (.22) \\ & +.015\Delta\ln P_{t-2} & -.02\Delta^2\ln P_{t-2} & +.28\Delta\ln(M/P)_{t-4} \\ & (.027) & (.03) & (.11) \\ & +.11\Delta\ln r_t & +.003\ln r_{t-4} & +.03\ln(M/PY)_{t-1} \\ & (.03) & (.01) & (.03) \end{array} \tag{1.5.10}$$

$$\begin{array}{llll} \Delta\ln r_t = & .18\Delta\ln P_t & +.24\Delta\ln P_{t-3} & -.50\Delta\ln U_{t-1} \\ & (.17) & (.19) & (.13) \\ & +.01\Delta\ln U_{t-3} & -.24\Delta\ln r_{t-2} & -.12\Delta_3\ln(B/P)_{t-1} \\ & (.17) & (.11) & (.08) \\ & +.23\Delta\ln(M/P)_{t-1} & -.02\ln(r/E_{-1})_{t-1} & \\ & (.12) & (.05) & \end{array} \tag{1.5.11}$$

for regime 1 and

$$\begin{array}{llll} \Delta\ln(M/P)_t = & 1.83\Delta\ln P_t & +.10\Delta\ln P_{t-3} & -.17\Delta\ln U_{t-1} \\ & (.14) & (.11) & (.11) \\ & -.03\Delta\ln U_{t-3} & +.006\Delta\ln r_{t-2} & +.007\Delta_3\ln(B/P)_{t-1} \\ & (.08) & (.04) & (.02) \\ & +.20\Delta\ln(M/P)_{t-1} & -.06\ln(r/E_{-1})_{t-1} & \\ & (.07) & (.03) & \end{array} \tag{1.5.12}$$

for regime 2, with covariance matrices

$$\boldsymbol{\Sigma}^1 = \begin{pmatrix} .0005 & -.0014 \\ -.0014 & .0103 \end{pmatrix} \quad , \quad \boldsymbol{\Sigma}^2 = \begin{pmatrix} .0005 & .0004 \\ .0004 & .0014 \end{pmatrix}. \tag{1.5.13}$$

Comparison with the single equation estimates ((1.5.2), (1.5.4) and (1.5.5)) shows considerable differences in the significance as well as the magnitude of individual coefficients.

Experience with the FIML program highlighted the importance of good initial values in the iteration routine if the speed of convergence was to be increased. Two algorithms were used: Gill-Murray-Pitfield which was fast from good initial values but sometimes would fail to converge if started at all far away from a maximum, and the Powell conjugate directions algorithm which was much more robust but considerably slower (sometimes taking up to 40 iterations or 15-20 seconds on the CDC 7600 computer).

Chapter 2

Stability of a U. K. Money Demand Equation: a Bayesian Approach to Testing Exogeneity

The paper analyses an M3 demand for money equation for the United Kingdom. Attention is paid to the policy change that occurred in 1971 with the introduction of the measure known as Competition and Credit Control. Classical and Bayesian single equation instrumental variables procedures are developed to investigate the exogeneity of the short-term interest rate and the constancy of the parameters of the underlying relationships. The parameters of the short-term equation have changed as well as the exogeneity status of the interest rate variable but the parameters of the long-run equation appear to be less affected by the policy change.

2.1 Introduction

A large number of demand for money equations in the United Kingdom exhibit parameter instability across the major policy change that occurred in 1971 with the introduction of the measures known as Competition and Credit Control (CCC).[1] Their instability has been attributed to a structural break on the (implicit) justification that the CCC changes were specifically directed to changing the competitive structure of the banking system. We note, however, that these equations are often estimated by Ordinary Least Squares

[0] Published in *Review of Economic Studies* (1986), Vol. 53, pp. 603–634. Co-authors M. Lubrano and J.-F. Richard. This paper has benefitted from numerous discussions with L. Bauwens, J. H. Drèze, J. P. Florens, V. Ginsburgh, G. E. Mizon and M. Mouchart. Three referees have made a number of insightful and constructive comments. Special thanks are due to D. F. Hendry for his constant willingness to comment on our findings and to suggest new routes of investigation. (some of which are yet to be tested!) and also to B. Govaerts for her invaluable assistance in the development of the Bayesian numerical algorithms we have been using. Obviously we claim full responsibility for errors and shortcomings. The support of "Les Services de la Programmation de la Politique Scientifique" of the Belgian Government through the "Projet d'Action Concertée No. 80.85-12" is gratefully acknowledged. Part of the work was done when M. Lubrano and R. Pierse visited CORE whose support is gratefully acknowledged.

[1] A noticeable exception is the M_1 demand for money equation estimated by Hendry (1980) and updated in Hendry and Richard (1983), whose coefficients are stable over the period 1963(i)–1980(ii).

(OLS) though it has occasionally been argued that the interest rate should be treated as an endogenous variable— see e.g. Artis and Lewis (1976). Furthermore, as we shall see below, the interest rate setting process was indeed fundamentally modified with the introduction of CCC. Following Engle et al. (1983) these are precisely the circumstances under which an invalid exogeneity assumption entails instability of OLS estimators. This issue, which is central to our paper, is developed further in Section 2.3.1 below.

The main object of our paper is, therefore, to develop operational classical and Bayesian Instrumental Variables (IV) procedures for analysing the exogeneity of a variable in a single structural equation. These procedures atre then used to investigate whether the instability of a demand for money equation has been induced by the invalid assumption that the interest rate is exogenous or whether it corresponds to a genuine structural break. These alternativesd whose policy implications are quite different are formalised in Section 2.3.1 below.

The paper is organised as follows. In Section 2.2 we discuss the specification of an M3 demand for money equation that is first estimated by OLS under the working assumption that the interest rate is exogenous; multiplicative dummies are then introduced and the equation is reestimated by Weighted Least Squares (WLS) in order to obtain a parsimonious description of parameter insdtability; in Section 2.3 we develop calssical and Bayesian IV procedures for investigating the exogeneity of a variable within a bivariate linear model; these procedures are described in general terms in Section 2.3.1 while Sections 2.3.2 to 2.3.5 regroup the more technical material; in Section 2.4 our money demand equation is imbedded with an interest rate equation in a two-equation model and the exogeneity of the interest variable is then formally analysed; conclusions are drawn in Section 2.5 and the technical details are presented in an appendix.

Sections 2.2 and 2.3 are largely autonomous with respect to each other and the reader may consider skipping Sections 2.3.2 to 2.3.5 that contain the more technical material. Those whose interest lies in the algebra of an exogeneity analysis may wish to read first Section 2.3 that provides the theoretical background for the empirical analysis.

2.2 Single Equation Analysis of the Demand for Money

2.2.1 The institution background: competition and credit control

The introduction of the measures known as Competition and Control (CCC) in October 1971 was an attempt by the monetary authorities to move from a regime where the primary objective was to restrain movements in short term interest rates, to a regime in which control over the monetary aggregates could be achieved through the free operation of market forces. To this end, restrictive practices in the banking sector were swept away, the clearing bank cartel (which had previously linked the rates on Advances and time deposits to the administered "Bank Rate") was abolished and the banks were encouraged to compete for funds by offering competitive rates. In 1972 the Bank Rate was replaced by a "Minimum Lending Rate" that was market determined, being related to the Treasury bill rate.

Part of the rationale for the policy change was evidence from published studies of the demand for monet in the U.K. (Fisher (1968), Laidler and Parkin (1970), Goodhart and Crockett (1970)) of a stable behavioural relationship that could be exploited to achieve the objectives of monetary control. However, the immediate result of the switch to the new regime weas an upsurge in holdings of interest earning deposits that was unpredicted by these demand functions and Hacche (1974) reported a complete breakdown of the Bank of England's own forecasting equations for M3 after 1971. The operation of the CCC regime proved difficult as Goodhart (1980b) describes and some of the direct control mechanisms abolished in 1971 were later reintroduced before the regime finally came to an end on 20th August 1981.

The breakdown of M3 equations with CCC have several alternative explanations. One possibility is that the equations were misspecified because of omitted variables and the obvious candidate here is the own-rate of interest on money. We discuss this further below. However, money demand equations estimated by OLS implicitly assume that it is legitimate to treat the rate of interest as an exogenous variable. Since the introduction of CCC has moved the economy from a regime of administered interest rates to a regime where interest rates are market determined, we have to consider the possibility that the exogeneity status of the interest rate has changed. Finally, we must also consider the possibility that a demand for money function may not be invariant to a change in the process generating interest rates even if, within each regime, the interest rate is a valid exogenous variable. The Lucas (1976) critique would suggest that agents might modify their behaviour in response to an attempt by the authorities to control it in this way.

2.2.2 Specification search

Our discussion of the institutional background suggests that we should, ideally, conduct a joint search for the money demand and for the interest rate equations. Such a search would, however, prove computationally ery demanding and, anyway, hard to conduct given the difficulties encountered in the specification of the interest rate equation. Therefore, on grounds of tractability, we have adopted the following sequential procedure: in this section, we conduct a single equation specification search for the money demand equatioon equation by means of OLS and WLS estimation under the working assumption that interest rates, as well as the other current dated regressors are weakly exogenous in the terminology of Engle et al. (1983); the specification that emerges from this analysis is then used as such in Section 2.4, where it is embedded within a bivariate model for the purpose of investigating the exogeneity of the interest rate.

As discussed further in Section 2.5, the empirical evidence relative to the present application seems to suggest that our final conclusions about the exogeneity of the interest rate are unlikely to be severely biased by the adoption of this operational stepwise seach procedure.

2.2.3 The data

The data consists of 79 seasonally adjusted quarterly observations (1961(iv)–1981(ii)), for which the first five are used for the initialisation of the lagged variables and the last four for predictive tests. The remaining 70 observations are eventually divided into the subperiods $A\{$1963(i)–1971(iii)$\}$ and B $\{$1971(iv)–1980(ii)$\}$. The variables relevant to our analysis are:

- M: the M3 personal sector monetary aggregate;
- Y: real personal disposable income;
- P: the deflator of Y

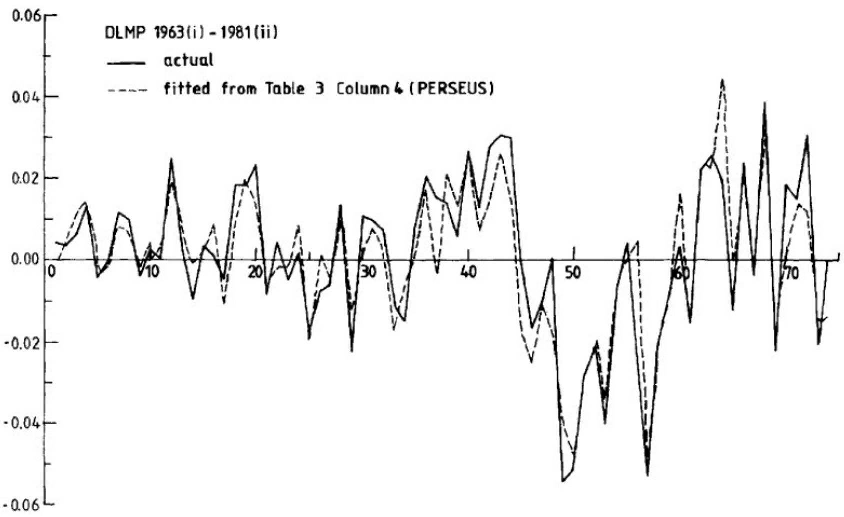

Figure 2.1: Actual and fitted (Table 2.3 Col 4) values of $\Delta \ln(M/P)$, 1963(i)–1981(ii)

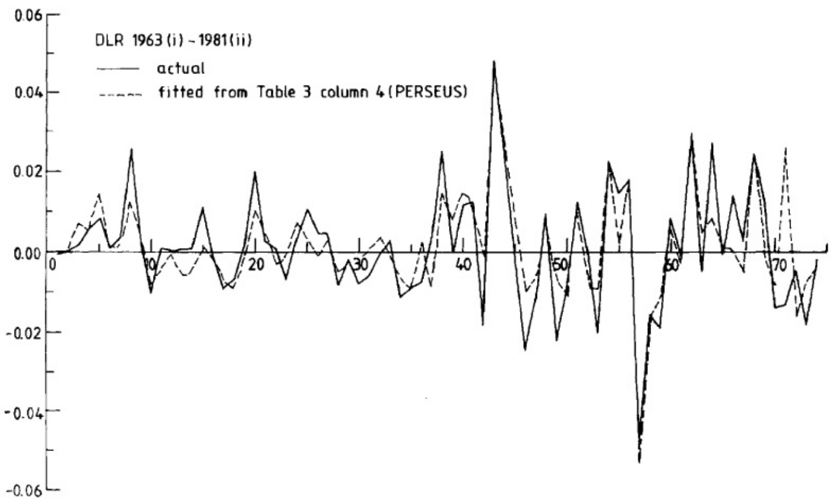

Figure 2.2: Actual and fitted (Table 2.3 Col 4) values of $\Delta \ln(1 + R_t)$, 1963(i)–1981(ii)

R: the local authorities short-term interest rate

The sources are described in Appendix .1. Graphs of $\Delta \ln(M/P)$, $\Delta \ln(1 + R_t)$ and $\ln(M/(PY))$ are reproduced in Figures 2.1–2.3. Our choice of variables calls for a number of comments.

The choice of personal sector M3 as the approproiate money aggregate follows from our discussion of the institutional background. Some initial work was done using total M3 but proved unsatisfactory in many respects probably reflecting different behaviour by individuals and companies. That led us to look at the sectoral disaggregation of M3 following, thereby, the Bank of England practice. A complete study of M3 would then require the specification of separate personal and company sectors demand equations typically depending on different interest rates. An exogeneity analysis within this joint context would prove computationally very demanding, espaecially within a Bayesian

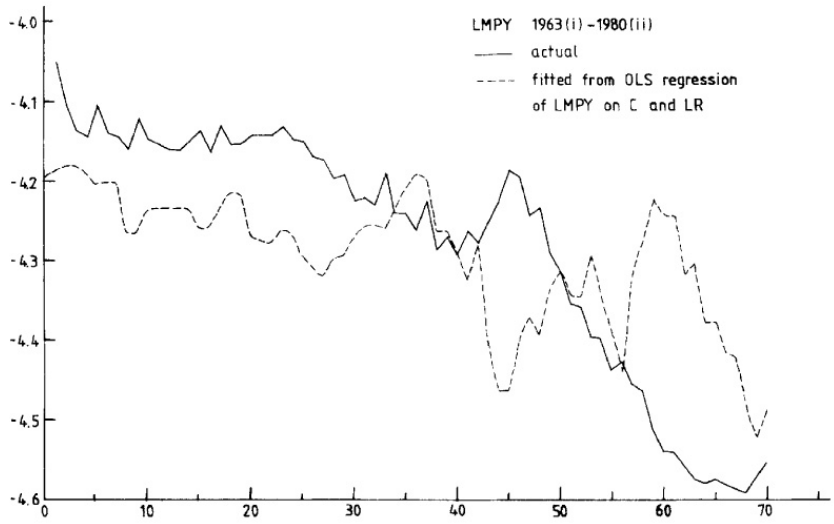

Figure 2.3: Actual and fitted (Table 2.3 Col 4) values of $\ln(M/(PY)$, 1963(i)–1981(ii)

framework and goes beyond the objectives of the present paper. Therefore, we restricted ourselves to looking at the personal sector only.

The choice of the interest rate variables raises a number of issues. On theoretical grounds our equation should include the opportunity cost of holding M3, a substantial part of which is non-interest bearing. It should, therefore, include the differential between an outside interest arte and the own-rate on the interedst bearing component of M3 as well as that outside rate itself. Before 1971, since all short-term interest rates were closely liked to the Bank Rate, the own-rate differential is essentially constant so that its impact is only estimible in the second subperiod. The Local Authority rtae was chosen as a representative outside rate. WE included both rates in initial empirical work over the whole period but found the own rate wholly insignificant.[2] This seems to indicate that the motivation for holding bank time-deposit accounts instead of such substitutes as Building Socity accounts, etc., lies elsewhere than in the interest rate differential. We therefore kept only the Local Authority rate, while wishing to stress that the issue of the substitution effect of interest rates is not thereby closed.

2.2.4 Notation

The following mnemonics are used throughout the paper: MP for M_t/P_t, MPY for $M_t/(P_tY_t)$, D for the difference operator (Δ when conventional notation is used), i for the i-th lag operator, D_i for the i-th difference operator (Δ_i), DD for the squared difference operator (Δ^2). L for natural logarithms (ln) and R for $1+R_t$. For example , $DDLP2$ reads as $\Delta^2 \ln P_{t-2}$, $D4LY$ as $\Delta_4 \ln Y_t$, $LR5$ as $\ln(1+R_{t-5})$, and so on. Also C stands for the constant term and C_i for the i-th quarter seasonal dummy ($i = 1, 2, 3$).

Other notations are: SSR for the sum of squared residuals (these sums are instrumental in the computation of several F-test statistics), SDR for the unbiased standard deviation of the regression error, R^2 for the unadjusted squared multiple correlation co-

[2] An F-test of the specification corresponding to column 1 in Table 2.1 against a specification that included in addition lags of the own interest rate up to 5th gave an F value of 0.24 with degrees of freedom 6 and 37.

efficient and DW for the conventional Durbin-Watson statistic. Following Kiviet (1985) no formal significance is attached to the DW statistic. However, values well within the d_u critical interval are at least not worrying.

The following statistics are reported when available:
$\eta_1(4)$ is a forecast test (see Hendry (1979)) asymptotically distributed as χ^2_4 on the null of no predictive failure;
$\eta_2(4, k)$ is the Chow (1960) test for parameter constancy for four periods approximately distributed as $F_{4,k}$ on the null of parameter constancy;
$\eta_3(1)$ is the squared h-test for first-order autocorrelation, asymptotically distributed as χ^2_1 on the null of serial independence;
$\eta_4(8)$ is the Box and Pierce (1970) test statistic for 8th-order residual autocorrelation, asymptotically distributed as χ^2_8 on the null of serial independence;
$\eta_5(4)$ is a Lagrange multiplier test for 4th-order autocorrelation and
$\eta_5(4, j)$ is an F-version thereof (see Godfrey (1978)), approximately distributed as χ^2_4 and $F_{4,j}$ on the null of serial independence;
$\eta_7(1)$ is the test for first-order ARCH (see Engle (1982a)) asymptotically distributed as χ^2_1 on the null of no ARCH effect.

2.2.5 Results

The specification search has been conducted along the principles described in Hendry and Richard (1982, 1983) and consists of three main steps.

Step 1. Independent specification searches over the pre- and post-1971 periods, each of which consists of 35 observations only, cannot be envisaged because of lack of degrees of freedom. Therefore, the starting point of our analysis is an unrestricted OLS regression over the 70 observations of $DLMP$—i.e. $\Delta \ln(M_t/P_t)$—on 27 regressors consisting of a constant term, three seasonal dummies, current and lagged values up to the fifth-order of LP, LY, LR and lagged values of LM. THis equation serves essentially to calibrate the error standard deviation, which equals here 0.0095 and to ensure that the error process is a mean innovation process (MIP) relative to our data base. The individual coefficient values are of little interest and are not reported here.

Successsive simplifications lead to equation (2.2.1)

$$\begin{aligned}\Delta \ln(M/P)_t = \beta_0 + \beta_1 \Delta \ln(M/P)_{t-1} + \beta_2 \Delta \ln P_t + \beta_3 \Delta^2 \ln P_{t-2} + \beta_4 \ln(M/(PY))_{t-5} \\ + \beta_5 \Delta_4 \ln Y_t + \beta_6 \Delta \ln(1 + R_t) + \beta_7 \Delta^2 \ln(1 + R_{t-3}) + \beta_8 \ln R_{t-5} + u_t\end{aligned} \tag{2.2.1}$$

or, in our notation,

$$\begin{aligned}DLMP = C + \beta_1 DLMP1 + \beta_2 DLP + \beta_3 DDLP2 + \beta_4 LMPY5 \\ + \beta_5 D4LY + \beta_6 DLR + \beta_7 DDLR3 + \beta_8 LR5\end{aligned}$$

whose coefficients are reported in column 1 of Table 2.1. In short, equation (2.2.1) takes the form of an error correction mechanism (ECM) for real money balance. Its steady-state equilibrium solution is characterised by a constant velocity of circulation of money. The disequilibrium feedback coefficient ($LMPY5$) exhibits an unusually long lag of 15 months though the time-lag is in fact poorly identified and equation (2.2.1) is only marginally better than those in which $LMPY5$ is replaced by any one of the other $LMPYi$variables.[3] The coefficient of the interest rate DLR has the "wrong" sign according to conventional wisdom, an issue to which we shall pay further attention below.

Step 2. Equation (2.2.1) is then reestimated over the subperiods A and B separately. The results are found in columns 2A and 2B of Table 2.1. Four salient features emerge from the comparison between those two regressions:

(i) The sample variance is substantially larger in period B than it is in period A with a variance ratio of about 3. While this could be a symptom of model misspecification arising from changes in omitted variables orthogonal to those included, we know of no such variables and have treated the problem as one of changes in the market structure;

(ii) the coefficient of DLR changes sign with the introduction of CCC and the explanation for the overall positive coefficient of DLR in equation (2.2.1) lies in the post-1971 period. Though at this stage of our analysis the difference is not yet statistically significant, it will prove critical for our purpose and has obvious policy implications, some of which are discussed in Section 2.5;

(iii) The coefficient of $LR5$, which determines the direction of the long-run impact of interest rate on the velocity of circulation of money, has the "right" sign and is remarkably constant across the change of regimes;

(iv) None of the differences between the other coefficients appear to be statistically significant suggesting that we can impose common coefficient restrictions across the two regimes, gaining thereby precision on the point estimates.

Step 3. Equation (2.2.1) is finally reestimated by WLS over the entire sample period with multiplicative dummies accounting for the major coefficient changes. The results are reported in columns 3 and 4 of Table 2.1. In both columns common coefficients for the variables $DLMP1$, DLP, $DDLP2$, $LMPY5$, $D4LY$ and $LR5$ have been imposed while in column 4 we have also imposed common seasonal coefficients and have deleted $DDLR3$ in the second subperiod.

The specification in Column 4 is the one that will be used for the exogeneity analysis. It contains 13 unrestricted coefficients (namely the 9 reported in Table 2.1 Column 4, togther with constant term $C = -0.15(0.05)$ and seasonals $C1 = -0.03(0.004)$, $C2 = m - 0.001(0.004)$ and $C3 = -0.01(0.003)$) leaving 57 degrees of freedom. Our analysis does not seem to provide significant statistical evidence against this equation. Despite data, period and adjustment differences, the actual values for M3 are in accordance with those in Hendry and Mizon (1979), except for the tiny ECM coefficient and for the long term unit elasticity of income (Hendry and Mizon (1979) found 1.6 while in the course of our specification search we have set it equal to 1). A discussion of some of the intriguing features of our equation is postponed until Section 2.5 since we first have to investigate whether our results suffer from simultaneous biases.

[3] In connection with this issue of time lag, note also the significance of the coefficient $LR5$.Throughout the simplification search we have run auxiliary regressions to test for the inclusion of additional lags but none has turned out significant.

2.3 Bivariate Instrumental Variables Analysis and Exogeneity

2.3.1 Introduction

Let us first indicate how the framework developed in Engle et al. (1983) applies within the present context. For the sake of simplicity, we can restrict our attention to a stylised version[4] of the bivariate model we shall construct below for the variables $\dot{m}_t = \Delta(M/P)_t$ and $\dot{r}_t = \Delta \ln(1 + R_t)$. It consists of the money demand equation

$$\dot{m}_t = \beta \dot{r}_t + u_t \tag{2.3.1}$$

paired with the interest rate reaction function

$$\dot{r}_t = \rho z_t + v_t \tag{2.3.2}$$

where z_t is an exogenous (instrumental) variable. It is assumed further that (u_t, v_t) are jointly identically and independently normally distributed with mean zero and covariance matrix

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim \mathrm{IN}(\mathbf{0}, \mathbf{\Sigma}) \quad \text{with} \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{vu} & \sigma_v^2 \end{pmatrix}. \tag{2.3.3}$$

Let $\widehat{\beta}$ denote the OLS estimator of β in (2.3.1).

The joint distribution of $(\dot{m}_t, \dot{r}_t | z_t)$ factorises[5] into the product of the marginal distribution of $(\dot{r}_t | z_t)$, as characterised by (2.3.1), and the conditional distribution of $(\dot{m}_t | \dot{r}_t, z_t)$, which is normal with conditional expectation

$$\mathrm{E}(\dot{m}_t | \dot{r}_t, z_t) = (\beta + \mu)\dot{r}_t - \mu\rho z_t \tag{2.3.4}$$

with $\mu = \sigma_{uv}\, \sigma_v^2$ and conditional variance $\tau^2 = \sigma_u^2 - \mu^2 \sigma_v^2$. Therefore, if $\sigma_{uv} = 0$ and if, furthermore, β and (ρ, σ_v^2) are not subject to cross-restrictions, then the conditional distribution of $(\dot{m}_t | \dot{r}_t, z_t)$ is fully characterised by the structural equation (2.3.1) on its own and the OLS estimator $\widehat{\beta}$ is BLUE. Equally, importantly, as long as β is invariant with respect to intervention affecting (ρ, σ_v^2) or, more generally, the distribution of z_t, the OLS estimator $\widehat{\beta}$ is not affected by these interventions.

The situation changes dramatically if σ_{uv} differs from zero since the sampling distribution of $\widehat{\beta}$ depends heavily on the distribution of $\dot{r}_t$ and z_t. To take the simplest case, let us assume that z_t is identically and independently normally distributed with zero mean and variance σ_z^2. In such a case, the distribution of $\dot{m}_t | \dot{r}_t$ (which is marginalised with respect to z_t since we are discussing the properties $\widehat{\beta}$, the OLS estimator of $\dot{m}_t$ on $\dot{r}_t$ only) is normal with conditional expectation

$$\mathrm{E}(\dot{m}_t | \dot{r}_t) = (\beta + \mu_*)\dot{r}_t \tag{2.3.5}$$

[4] All the regressors that are inessential to the argument are deleted for notational convenience. Therefore we are left with the simple model described by equations (2.3.1) and (2.3.2), which is, however, meant to be interpreted as a stylised version of a *short run dynamic model* (not to be confused with the static long run solution of the model we shall discuss in Section 2.5 below). Our discussion of weak exogeneity specifically refers to a property of (short term) dynamic model.

[5] Our more general model being dynamic it is essential to view this factorisation as a sequential one $(t : 1 \to T)$ in the sense that, at time t, it is conditional on the past of all the variables in the model. Therefore, the concept under consideration here is that of weak exogeneity. Strong exogeneity requires in addition that $\dot{m}_t$ does not Granger-cause $\dot{r}_t$ (e.g. through z_t).

with $\mu_* = \sigma_{uv}(\sigma_v^2 + \rho^2\sigma_z^2)^{-1}$ and conditional variance $\tau_*^2 = \sigma_u^2 - \mu_*^2(\sigma_v^2 + \rho^2\sigma_z^2)$. In such a case, interventions affecting the "nuisance" parameters $(\rho, \sigma_v^2, \sigma_z^2)$ will induce changes in the sampling properties of $\widehat{\beta}$, even when the underlying "structural" coefficient is invariant with respect to these interventions.

In our application we are confronted with the empirical finding that $\widehat{\beta}$ has changed with the introduction of CCC in 1971. Our analysis suggests immediately three possiible explanations for the lack of invariance:

(i) β itself has changed, a possibility that cannot be ruled out since a declared objective of the policy change was precisely to modify the market structure and since, more generally, the Lucas (1976) critique obviously requires our attention in the present context;

(ii) $\sigma_{uv} \neq 0$ and the interest rate setting process has changed;

(iii) the interest rate setting process has not changed but σ_{uv} has (the institutional background suggests, in particular, that σ_{uv} might be zero before the introduction of CCC and non-zero after that).

Obviously, these three possibilities are not mutually exclusive.

The analysis of our empirical findings relies heavily upon a correct interpretation of the "behavioural" content of the weak exogeneity assumption $\sigma_{uv} = 0$. Let, therefore, equation (2.3.1) be reformulated in terms of expectations, as in Florens et al. (1974, 1979):

$$\mathrm{E}(\dot{m}_t|z_t) = \beta\,\mathrm{E}(\dot{r}_t|z_t). \tag{2.3.6}$$

It appears that the condition $\sigma_{uv} = 0$ is necessary and sufficient for the equivalence of equation (2.3.4) as

$$\mathrm{E}(\dot{m}_t|\dot{r}_t, z_t) = \beta\rho z_t + (\beta + \mu)(\dot{r}_t - \rho z_t) \tag{2.3.7}$$

where ρz_t and $\dot{r} - \rho z_t$ are the "anticipated" and "unanticipated" components of $\dot{r}_t$. This indicates that when $\sigma_{uv} = 0$ $(\mu = 0)$ economic agents treat in exactly the same way the anticipated and unanticipated components of $\dot{r}_t$. We shall describe such a situation as one of "effective" control to be contrasted with situations where economic agents might find ways of countering the (restrictive) measures that are enforced upon them.

Having set the basic framework, let us now outline the algebra of the exogeneity analysis for the simple model (2.3.1)–(2.3.3). Doing so enables us to motivate the introduction of the auxiliary parameters on which our analysis focuses and leaves the reader with the possibility of skipping the more technical details in the sections that follow. Unless we restrict our attention to deriving Lagrange Multiplier (LM) test-statistics for weak exogeneity, e.g. as in Engle (1982b), we need an operational factorisation of the llikelihood function that works even when $\sigma_{uv} \neq 0$ and that, as much as possible, enables us to deal analytically with the nuisance parameters (ρ, σ_v^2) in equation (2.3.2) and to draw inference on σ_{uv} or on appropriate functions thereof.

As is often the case with likelihood functions, it proves convenient to set the factorisation in terms of the distribution of the unobservable disturbance terms (u_t, v_t). This can be done by using either of the following two auxiliary regression functions:

(i) the regression of v_t on u_t:

$$v_t = \lambda u_t + \epsilon_{1t}, \quad \epsilon_{1t} \sim \mathrm{IN}(0, \omega^2) \tag{2.3.8}$$

with $\lambda = \sigma_{uv}\,\sigma_u^{-2}$ and $\omega^2 = \sigma_v^2 - \lambda^2\sigma_u^2$; or

(ii) the regression of u_t on v_t:

$$u_t = \mu v_t + \epsilon_{2t}, \quad \epsilon_{2t} \sim \mathrm{IN}(0, \tau^2) \tag{2.3.9}$$

where μ and τ^2 are defined as in (2.3.4).

The correspondence between $\mathbf{\Sigma}$ and the two sets of parameters $(\sigma_u^2, \lambda, \omega^2)$ and $(\sigma_v^2, \mu, \tau^2)$ is one-to-one and is characterised by the identitites

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 & \lambda\sigma_u^2 \\ \lambda\sigma_u^2 & \omega^2 + \lambda^2\sigma_u^2 \end{pmatrix} = \begin{pmatrix} \tau + \mu^2\sigma_v^2 & \mu\sigma_v^2 \\ \mu\sigma_v^2 & \sigma_v^2 \end{pmatrix}. \tag{2.3.10}$$

Starting with Wu (1973) most exogeneity tests are based on the auxiliary regression (2.3.9), whether implicitly or explicitly when, in a Lagrange Multiplier (LM) framework as described e.g. in Engle (1982b), they amount to including an estimated residual $\widehat{v}_t$ as an additional regessor in (2.3.1) and testing for its significance. In fact, as indicated by (2.3.4), the associated factorisation coincides with the sequential factorisation of the joint distribution of $(\dot{m}_t, \dot{r}_t|z_t)$ into the conditional distribution of $(\dot{m}_t|\dot{r}_t, z_t)$ and the marginal distribution of $(\dot{r}_t|z_t)$. However, when $\sigma_{uv} \neq 0$ this factorisation does not meet our requirements since, in particular, the nuisance parameters (ρ, σ_u^2) appear on both sides. This is not the case in the factorisation associated with the auxiliary regression (2.3.8). This explains why our subsequent analysis is based on the following factorisation of the likelihood function

$$L(\mathbf{Y}; \boldsymbol{\theta}) = L_1(\mathbf{Y}; \boldsymbol{\theta}_1) \cdot L_2(\mathbf{Y}; \boldsymbol{\theta}_2), \tag{2.3.11}$$

with

$$L_1(\mathbf{Y}; \boldsymbol{\theta}_1) = \prod_{t=1}^{T} f_N^1(u_t|0, \sigma^2) \tag{2.3.12}$$

$$L_2(\mathbf{Y}; \boldsymbol{\theta}_2) = \prod_{t=1}^{T} f_N^1(v_t|\lambda u_t, \omega^2) \tag{2.3.13}$$

where $\mathbf{Y}$ denotes the $T \times 2$ matrix of observations on $(\dot{m}_t, \dot{r}_t)$, u_t and v_t are given in (2.3.1) and (2.3.2) respectively, $\boldsymbol{\theta}_1{}' = (\beta, \sigma^2)$ and $\boldsymbol{\theta}_2{}' = (\beta, \rho, \lambda, \omega^2)$. Also $f_N^1(x|\mu, \nu^2)$ denotes a univariate normal density function with mean μ and variance ν^2, Its expression is given in Appendix .2.

Conditionally on β, the submodel (2.3.13) takes the form of a standard regression model to which we can apply the usual classical and Bayesian techniques in order to derive *analytical* expressions for the conditional point estimates and posterior distributions of $(\rho, \lambda, \omega^2)$. They are to be marginalised with respect to β at the final stage of the analysis. Technical details are provided in Sections 2.4.2 and 2.4.3. Note that $\lambda = 0$ if and only if $\sigma_{uv} = 0$ so that inference on the exogeneity of r_t is a direct byproduct of our analysis.

The covariance matrix of $(\dot{m}_t, \dot{r}_t|z_t)$, say $\mathbf{V}$, is related to β and $\mathbf{\Sigma}$ through the following identity:

$$\mathbf{V} = \mathbf{Q}'\mathbf{\Sigma}\mathbf{Q} \quad \text{with } \mathbf{Q} = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} = (\mathbf{b} : \mathbf{s}) \text{ say} \tag{2.3.14}$$

so that λ, as defined via (2.3.8), may be written as

$$\lambda = \mathbf{b}'\mathbf{V}\mathbf{s}(\mathbf{b}'\mathbf{V}\mathbf{b})^{-1}. \tag{2.3.15}$$

The question then arises of deciding whether prior information on λ should be thought in terms of $\mathbf{\Sigma}$ or in terms of β and $\mathbf{V}$. Technically, it makes little difference, in the present case at least, since, conditionally on β, the correspondence between Σ and $\mathbf{V}$ is one-to-one and bilinear and since such prior densities as the inverted-Wishart are functionally invariant with respect to such transformations, Simply $\mathbf{\Sigma}$ and $\mathbf{V}$ cannot be both *a priori* independent of β. We have a definite preference for reasoning in terms of $\mathbf{V}$ since, from a statistical point of view, disturbances are merely "derived" unobservable quantities which *de facto* regroup all factors that have been omitted from the equations under consideration.[6] It seems, therefore, difficult to assume for example, prior independence between β and $\mathbf{\Sigma}$, while we have no conceptual problems in doing so between β and $\mathbf{V}$.

Three important issues remain to be clarified before we can concentrate on the more technical issues.

1. The interest rate equation (2.3.2), as well as the more general equation we shall introduce below, takes the form of an Instrumental Variables (IV) equation whereby the current value of money $\dot{m}_t$ is excluded from the list of regressors. The point is that we lack economic theories to support a complete specification search towards a genuine "structural" equation for the interest rate and that we are faced with limited sample sizes (35 observations in each regime). Considerations of robustness against the specification of the interest rate equation are, therefore, critical since the latter is only instrumental in the construction of the exogeneity tests.

2. Conventional Limited Information Maximum Likelihood (LIML) procedures, or approximations thereof such as Two-Stage Least Squares (2SLS), require that *all* the predetermined variables in the money demand equation should be included in z_t, viewing thereby equation (2.3.2) as an "unrestricted reduced form" equation. This requirement will not be imposed here since it renders a parsimonious selection of instruments impossible taking into account the facts that z_t already includes 12 variables and that sample size is limited. Also hypotheses of interest such as the non-causality of money on interest rate cannot be dealt within an LI framework since, as we have seen, the money demand equation includes lagged values of money. In the present application instruments will, therefore, be selected on their own merits.

3. As discussed e.g. in Leamer (1978) there are a number of ways in which a Bayesian can approach the problem of "testing" a (point) hypothesis. A central issue is that of whether or not he should use continuous density functions whereby zero prior and posterior probabilities are attached to zero measure subsets of the parameter space. One route consists of attaching non-zero prior (discrete) probabilities to the hypothesis of interest and in analysing how the corresponding "prior odds" are revised into "posterior odds" in the light of sample evidence. In the specific context of exogeneity tests, this route has been adopted e.g. by Reynolds (1982) within an LI framework. It is our view that this approach can occasionally lead to questionable empirical results for example when it produces posterior odds which are much more extreme than one should be willing to accept on the basis of limited sample evidence. In fact, as argued

[6] Furthermore, as discussed e.g. in Florens et al. (1979) or Richard (1984), the distribution of the disturbances no longer uniquely characterises the distribution of the observables as soon as the number of relationships under consideration is strictly less than the number of endogenous variables as is naturally the case with general linear models such as errors-in-variables models or so-called "incomplete" simultaneous equations models.

e.g. by Kiefer and Richard (1979), it can easily lead to paradoxes when the (informative) prior odds are paired with prior densities which are otherwise "non-informative" within each hypothesis. We shall adopt here a "smoother" procedure whereby we rely upon continuous prior densities. Prior beliefs that a variable might be exogenous are then expressed in the form of an informative prior density for λ which is centred around zero. Sample evidence will then either tighten the corresponding posterior density around zero (confirmation) or shift it away from zero (refutation). The complete posterior distribution of λ is obviously far more informative than the scalar posterior odds and, for example, one can always examine whether or not an appropriate 95% posterior probability interval for λ contains the origin.

Evidently conducting inference about a quantity such as λ requires the choice of a metric whereby one can attach meaning to a non-zero value of λ. We see two ways of approaching this problem within the present context. Note first that, following its definition in (2.3.8), λ is subject to the inequality constraint $|\lambda\,(\sigma_u/\sigma_v)| < 1$. Though λ and σ_u/σ_v are not independent this inequality can serve to have a rough appreciation of how far λ is from zero. (The ratio of the OLS point estimates of σ_u and σ_v is of the order of 1.1 to 1.2 in both regimes.) We shall follow an alternative route, which seems more relevant to the object of our paper, whereby we shall compute the prior and posterior correlations bewteen β and λ since these enable us to translate approximately shifts in λ into shifts in β within a metric of standard deviations.

2.3.2 Sampling theory analysis

The money demand equation we have obtained in Section 2.2.4 is rewritten as

$$\mathbf{b}'\mathbf{y}_t + \mathbf{c}'\mathbf{x}_t + u_t \tag{2.3.16}$$

where $\mathbf{b}' = (1 : \beta)$, ${\mathbf{y}_t}' = (\dot{m}_t : \dot{r}_t)$ and $\mathbf{x}_t \in \mathbb{R}^m$ regroups all the other variables entering the equation including lagged $\mathbf{y}$'s. The interest rate IV equation is written as

$$\dot{r}_t + \mathbf{p}'\mathbf{z}_t + v_t \tag{2.3.17}$$

where $\mathbf{z}_t \in \mathbb{R}^k$ represents the set of instruments. In order to single out the variables that are common to $\mathbf{x}_t$ and $\mathbf{z}_t$ let the corresponding data matrices be partitioned as

$$\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2), \quad \mathbf{Z} = (\mathbf{X}_2 : \mathbf{X}_3) \tag{2.3.18}$$

where $\mathbf{X}_1$ is $T \times m_1$ with $m_1 + m_2 = m$ and $m_2 + m_3 = k$. Under a bivariate normality assumption the model consisting of equations (2.3.16) and (2.3.17) is rewritten as

$$\mathbf{y}_t | \mathbf{x}_t, \mathbf{z}_t \sim \mathrm{N}(\boldsymbol{\xi}_t, \mathbf{V}) \tag{2.3.19}$$

$$\mathbf{b}'\boldsymbol{\xi}_t + \mathbf{c}'\mathbf{x}_t = 0 \tag{2.3.20}$$

$$\mathbf{s}'\boldsymbol{\xi}_t + \mathbf{p}'\mathbf{z}_t = 0, \quad t = 1 \to T \tag{2.3.21}$$

where $\mathbf{s}' = (0 : 1)$ has been introduced for notational convenience. The model (2.3.19)–(2.3.21) belongs to a class of linear models discussed in Florens et al. (1974, 1979), Lubrano and Richard (1981) and Richard (1984) whose derivations serve as the basis of our analysis. The likelihood function associated with the model (2.3.19)–(2.3.21) factorises as in (2.3.11)–(2.3.13) except that u_t and v_t are now given by (2.3.16) and (2.3.17) respectively. Also θ_1 and θ_2 now include the additional parameter vector $\mathbf{c}$. Let $\mathbf{c}'$ be partitioned into $(\mathbf{c}_1{}' : \mathbf{c}_2{}')$ conformably with X in (2.3.18). Additional notation is:

$$\mathbf{M}_2 = \mathbf{I}_T - \mathbf{X}_2(\mathbf{X}_2{}'\mathbf{X}_2)^{-1}\mathbf{X}_2{}', \quad \mathbf{M}_Z = \mathbf{I}_T - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \tag{2.3.22}$$

$$\mathbf{u}_1 = \mathbf{Y}\mathbf{b} + \mathbf{X}_1\mathbf{c}_1. \tag{2.3.23}$$

The superscripts $\widetilde{}$ and $\widehat{}$ denote IV and OLS estimators respectively.

The concentrated log-likelihood function of $(\beta, \mathbf{c}_1)$ and the corresponding stepwise IVML estimator of λ are derived in Appendix .3:

$$L^*_{IV}(\mathbf{Y}; \beta, \mathbf{c}_1) = -\frac{T}{2}\log\left[\frac{\mathbf{u}_1{}'\mathbf{M}_2\mathbf{u}_1}{\mathbf{u}_1{}'\mathbf{M}_z\mathbf{u}_1}\left|(\mathbf{u}_1 : \mathbf{Y}\mathbf{s})'\mathbf{M}_z(\mathbf{u}_1 : \mathbf{Y}\mathbf{s})\right|\right] \tag{2.3.24}$$

$$\widetilde{\lambda}(\beta, \mathbf{c}_1) = (\mathbf{u}_1{}'\mathbf{M}_z\mathbf{u}_1)^{-1}\mathbf{u}_1{}'\mathbf{M}_z\mathbf{Y}\mathbf{s} \tag{2.3.25}$$

where use has been made of the following identities:

$$\mathbf{M}_z\mathbf{u} = \mathbf{M}_z\mathbf{u}_1 \quad \text{and} \quad \min_{\epsilon_2} \mathbf{u}'\mathbf{u} = \mathbf{u}_1{}'\mathbf{M}_2\mathbf{u}_1. \tag{2.3.26}$$

Numerical optimisation of (2.3.24) yields the IVML estimators of $(\beta, \mathbf{c}_1)$ and, by subsitution in (2.3.25), that of λ. Under the null hypothesis of $\lambda = 0$, the equations (2.3.16) and (2.3.17) are estimated by OLS independently of each other and the corresponding log-likelihood function is given by

$$L^*_{OLS} = -\frac{T}{2}\log\left(\widehat{\mathbf{b}}'\mathbf{Y}'\mathbf{M}_x\mathbf{Y}\widehat{\mathbf{b}} \cdot \mathbf{s}'\mathbf{Y}'\mathbf{M}_z\mathbf{Y}\mathbf{s}\right) \tag{2.3.27}$$

with $\widehat{\mathbf{b}}' = (1 : \widehat{\beta})$. The log-likelihood ratio (LR) test statistic for the null hypothesis $\lambda = 0$ is

$$\eta_8(1) = 2\left[L^*_{IV}(\mathbf{Y}; \widetilde{\beta}, \widetilde{\mathbf{c}}_1) - L^*_{OLS}\right] \xrightarrow[\lambda=0]{\mathscr{L}} \chi^2_1. \tag{2.3.28}$$

2.3.3 Bayesian analysis

As usual we have to find a compromise between flexibility and tractability in the choice of a prior density. We wish to specify a prior density which is information on $(\mathbf{V}, \beta)$ and, thereby on λ. We might also think of useful prior information as regards $\mathbf{c}$ although, as discussed below, taking it into account would substantially increase the computational burden. It is anyway convenient to think of $\mathbf{V}$ and $(\beta, \mathbf{c})$ as being a priori independent.[7] We have little grounds for assessing an informative prior density in the form

$$D(\beta, \mathbf{c}, p, \mathbf{V}) = D(\beta, \mathbf{c}) \cdot D(p|\mathbf{V}) \cdot D(\mathbf{V}) \tag{2.3.29}$$

where $D(\beta, \mathbf{c})$ is left unspecified at the moment, $D(\mathbf{V})$ is an inverted-Wishart density

$$D(\mathbf{V}) = f^2_{iw}(\mathbf{V}|\mathbf{V}_0, \nu_0) \tag{2.3.30}$$

[7] We could accomodate prior dependence between $\mathbf{V}$ and $(\beta, \mathbf{c})$ by letting $\mathbf{V}_0$ in (2.3.30) be a function of β and $\mathbf{c}$ since the analytical derivations in our analysis are mostly conditional on them. In particular, an independent prior density on $\mathbf{\Sigma}$ leads to replacing $\mathbf{Q}'\mathbf{V}_0\mathbf{Q}$ in (2.3.32) by, say, $\mathbf{\Sigma}_0$. The final numerical analysis of the posterior density of $(\beta, \mathbf{c}, \lambda)$ as well as the elicitation procedure described in Section 2.3.4 would have to be adapted in consequence.

whose functional expression is given in Appendix .2 and $D(p|\mathbf{V})$ is a limiting non-informative natural conjugate prior density

$$D(p|\mathbf{V}) \propto \omega^{-k} \tag{2.3.31}$$

with ω^2 being the variance associated with the partial likelihood function (2.3.13). A number of alternative forms of the prior densities (2.3.30) and (2.3.31) are discussed in Lubrano and Richard (1981).

Conditionally on β, the prior density of $\mathbf{\Sigma}$ is also inverted-Wishart

$$D(\mathbf{\Sigma}|\beta) = f_{iw}^2(\mathbf{\Sigma}|\mathbf{Q}'\mathbf{V}_0\mathbf{Q}, \nu_0) \tag{2.3.32}$$

where $\mathbf{Q}$ is defined in (2.3.14). The conditional distribution of λ given β is, therefore, a univariate t-density

$$D(\lambda|\beta) = f_t^1\left(\lambda|\lambda_0, \frac{h_0}{\omega_0^2}, \nu_0\right) \tag{2.3.33}$$

whose functional expression in given in Appendix .2. The hyperparameters λ_0, h_0 and ω_0^2 are functions of β and are defined by the identity

$$\mathbf{Q}'\mathbf{V}_0\mathbf{Q} = \begin{pmatrix} h_0 & h_0\lambda_0 \\ h_0\lambda_0 & \omega_0^2 + h_0\lambda_0\lambda_0^2 \end{pmatrix}. \tag{2.3.34}$$

The posterior densities of $(\beta, \mathbf{c})$ and $(\lambda|\beta, \mathbf{c}_1)$ are derived in Appendix .4

$$D(\beta, \mathbf{c}|\mathbf{Y}) \propto \left(\frac{h_*}{\omega_*^2}\right)^{(1/2)(\nu-1)} |\mathbf{\Omega}|^{-(1/2)\nu_*} \cdot D(\beta, \mathbf{c}) \tag{2.3.35}$$

$$D(\lambda|\beta, \mathbf{c}_1, \mathbf{Y}) = f_t^1\left(\lambda \middle| \lambda_*, \frac{h_*}{\omega_*^2}, \nu_*\right) \tag{2.3.36}$$

where $\nu_* = \nu_0 + T$ and

$$\sigma_*^2 = \mathbf{b}'\mathbf{V}_0\mathbf{b} + \mathbf{u}'\mathbf{u} \tag{2.3.37}$$

$$\mathbf{\Omega}_* = \begin{pmatrix} h_* & h_*\lambda_* \\ h_*\lambda_* & \omega_*^2 + h_*\lambda_*^2 \end{pmatrix} = \mathbf{Q}'\mathbf{V}_0\mathbf{Q} + (\mathbf{u}_1 : \mathbf{Y}_s)'\mathbf{M}_Z(\mathbf{u}_1 : \mathbf{Y}_s). \tag{2.3.38}$$

If $D(\beta, \mathbf{c})$ is a mutivariate Student, then the posterior density (2.3.35) belongs to a class of so-called 3-1 poly-t densities for which, as discussed in Richard and Tompa (1980), there exist efficient numerical methods of analysis. The evaluation of the *marginal posterior* denity (2.3.36) *jointly* with respect to β and $\mathbf{c}_1$ proves tedious to implement. An operational alternative consists first in multiplying together the posterior densities (2.3.35) and (2.3.36), obtaining thereby the joint posterior density of β, $\mathbf{c}$ and λ and taking advantage of a number of cancellations in the product. The evaluation of the posterior density of (β, λ) at any given point then requires numerical integration with respect to $\mathbf{c}$ but, dince $D(\mathbf{c}|\beta, \lambda)$ is also poly-t, can be organised in such a way that the cost of computation does not critically depend on the dimension of $\mathbf{c}$. Finally, the marginal posterior densities of β and λ are obtained by means of coneventional bivariate numerical integration procedures paying attention to the fact that these densities can be extremely skewed. The details of this implementation are given in Appendix .5 where it is also shown that the use of a non-informative prior density on $\mathbf{c}$

$$D(\mathbf{c}|\beta) \propto 1 \tag{2.3.39}$$

results in a major reduction of the cost of computation. In contrast we can be fully flexible in the choice of $D(\beta)$.

2.3.4 Elicitation of the prior density

The prior density of $\mathbf{V}$ and β has to be assessed in such a way that it reflects ones prior beliefs on the exogeneity of $\dot{r}_t$. the first and second order moments of $(\lambda|\beta)$ are central to this discussion. Following (2.3.33) they can be written as

$$\mathrm{E}(\lambda|\beta) = \lambda_0 = \phi_0 f_1(\beta\phi_0, \rho_0), \quad \nu_0 > 1 \tag{2.3.40}$$

$$\mathrm{V}(\lambda|\beta) = \frac{1}{\nu_0 - 2}\frac{\omega_0^2}{h_0} = \frac{1}{\nu_0 - 2}[\phi_0 \cdot f_2(\beta\phi_0, \rho_0)]^2, \quad \nu_0 > 2 \tag{2.3.41}$$

together with

$$f_1(x, \rho) = (\rho - x) \cdot (1 - 2\rho x + x^2)^{-1} \tag{2.3.42}$$

$$f_2(x, \rho) = (1 - \rho^2)^{1/2} \cdot (1 - 2\rho x + x^2)^{-1} \tag{2.3.43}$$

$$\phi_0 = (\nu_{22}^0/\nu_{11}^0)^{1/2}, \quad \rho_0 = \nu_{12}^0(\nu_{11}^0 \cdot \nu_{22}^0)^{-1/2}. \tag{2.3.44}$$

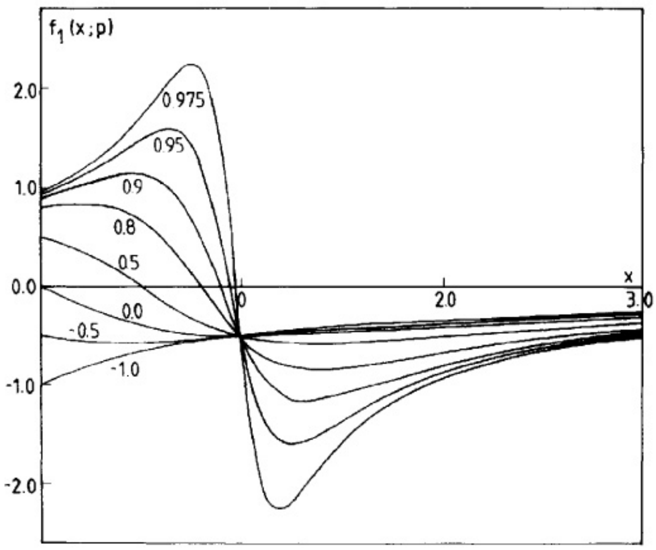

Figure 2.4: Function $f_1(x, \rho)$ in (2.3.42) for different ρ.

Figure 2.4 and 2.5 reproduce charts of the functions f_1 and f_2 for different values of ρ and $x > 0$. Their values for $x < 0$ are obtained by symmetry since $f_1(-x, \rho) = -f_1(x, \rho)$ and $f_2(-x, \rho) = -f_2(x, \rho)$. We note that f_1 and f_2 are *bounded* functions of x for any given ρ such that $|\rho| < 1$.

$$|f_1(x, \rho)| < \frac{1}{2}(1 - \rho^2)^{-1/2} \quad \text{and} \quad 0 < f_2(x, \rho) < (1 - \rho^2)^{-1/2}. \tag{2.3.45}$$

It follows that the marginal prior and posterior moments of λ are *finite* (up to the order ν_0 and ν_* respectively) on the sole condition that the prior distribution of β is integrable even if the prior and posterior moments of β themselves do not exist, as with the Cauchy prior used below. In contrast the existence of prior and posterior moments of μ, as defined in (2.3.9), require sharper *prior* information on β (as discussed e.g. in Dréze and Richard (1983), the sample information itself typically does not contribute to

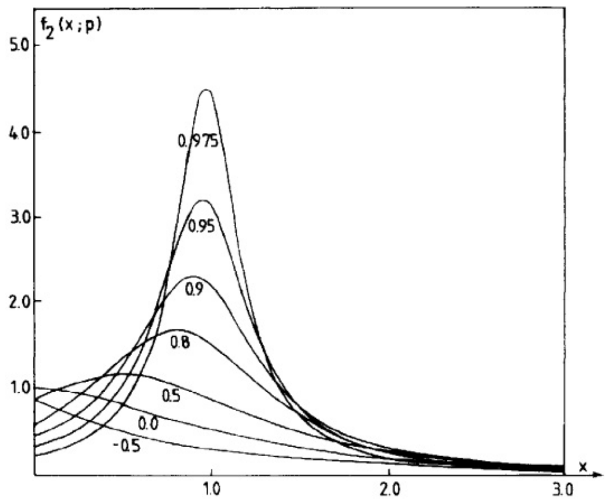

Figure 2.5: Function $f_2(x, \rho)$ in (2.3.43) for different ρ.

the existence of moments for β). This is, in our view, a major argument for conducting inference on the exogeneity of $\dot{r}_t$ in terms of λ instead of μ.

The above discussion suggests the following procedure for specifying a prior density on $\mathbf{V}$ and β which approximately reflects our prior beliefs on the exogeneity of $\dot{r}_t$:

1. We first specify a *proper* prior density $D(\beta)$, e.g. in the form of a Cauchy density or of a more "informative" t-density;

2. The prior expectations of v_{11} and v_{22}, the diagonal elements of $\mathbf{V}$, are then elicitated on such heuristic considerations as the expected "goodness of fit" of our model. The choice of ν_0 determines the prior squared variation coefficient $\mathrm{E}^2(v_{ii})/\operatorname{Var}(\sigma_{ii})$ for $i = 1, 2$;

3. ρ_0 is then selected in such a way that $\mathrm{E}(\lambda)$ takes the desired value, possibly at the cost of trying different values and computing the corresponding $\mathrm{E}(\lambda)$. If, in particular, $\rho_0 = \phi_0 M(\beta)$, where $M(\beta)$ denotes the prior median of β, we expect $\mathrm{E}(\lambda)$ to be near zero given the symmetry of $D(\beta)$ and the shape of f_1, as depicted in Figure 2.4.

2.3.5 Shifts of regime

The above analysis can be applied as such to the pre- and post-1971 periods separately. However, hypotheses about the constancy of the coefficients of the demand for money equation over the complete sample period are of major interest to us. Joint tests for the exogeneity of subsets of variables can be conducted within a sampling theory framework along the lines discussed in Richard (1980) by means of the computer program PERSEUS developed by Pierse (1982). The sampling theory results which are reported in Section 2.4 have been computed with PERSEUS.

However, PERSEUS has no Bayesian counterpart since it is obvious from the discussion in Section 2.3.3 that the analysis of posterior densities combining together sample information from the two subperiods would prove analytically tedious and numerically very costly. This explains why the Bayesian results which are reported in Section 2.4 have been computed for each period separately.

2.4 Bivariate Analysis of the Demand for Money

2.4.1 Specification of the reaction functions

We have already mentioned in Section 2.3.1 the difficulties we encountered in the specification of the interest rate reaction function. For each subperiod, the selection of instruments has been conducted by OLS estimation. The choice is restricted to lagged values of money (LM) and interest rate (LR) together with current and lagged values of prices (LP), reserves (LB) and unemployment (LU) since these are likely targets of monetary policies. The results which are reported in Table 2.2 are less than fully satisfactory though the signs are generally in accordance with common sense. Neither of the two equations has a constant growth long-run solution and, in line with our description of the institutional background, money does not enter significantly into the first period reaction function.

2.4.2 IVML estimation and exogeneity tests

The IVML estimators of the coefficients of the demand for money equation have been obtained with PERSEUS. Some of our empirical findings in section 2.3 have been reexamined within this framework including evaluating the constancy of several coefficients, particularly those of $LMPY5$ and $LR5$. the main results are reported in Table 2.3 except for the seasonal coefficients and are numbered conformably with their OLS equivalents in Table 2.1.

These results clearly indicate that the shift in the OLS estimate of β is not caused by simultaneity biases since the IVML estimate of β exhibits an even larger shift with the introduction of CCC! Also the weak exogenity of the interest rate suffers a borderline rejection at the 5% level in the first period while it is accepted in the second period. We shall elaborate upon these results in our conclusions. In the meantime we should take due account of the fact that the small sample properties of the LR test statistic (2.3.26) for weak exogeneity are largely unknown. We might of course use degree of freedom adjustments as in Kiviet (1985) but the application of the Bayesian procedures we have developed in Section 2.3.3 and 2.3.4 should provide us with more useful information as regards the exact (finite sample) information content of our data set.

2.4.3 Elicitation of the prior densities

The elicitation procedure described in Section 2.3.4 is now applied separately to the pre- and post-1971 period. In both cases, several specifications, including "non-informative" ones, are considered in order to conduct a sensitivity analysis. All the informative prior densities are constructed in such a way that $\mathrm{E}(\lambda) = 0$.

The period 1963(i)–1971(iii)

Our prior beliefs are that β, the short-term elasticity of $\dot{m}_t$ with respect to $\dot{r}_t$, probably lies between -1.0 and 0. Since inferences on λ are likely to be sensitive to the choice of $D(\beta)$—see Figures 2.4 and 2.5—two different specifications are considered:

(i) The Cauchy density:

$$D(\beta) \propto [1.0 + 4.0(\beta + 0.5)^2]^{-1} \tag{2.4.1}$$

is invariant with respect to the normalisation of the money demand equation and is relatively "non-informative" with $\Pr(-1 \le \beta \le 0) = 0.5$;

(ii) The Student density:

$$D(\beta) \propto [1.0 + 0.75(\beta + 0.5)^2]^{-1/2} \tag{2.4.2}$$

is more informative with standard deviation $\sigma_\beta \simeq 0.41$ and $\Pr(-1 \le \beta \le 0) = 0.8$. The prior means of v_{11} and v_{22}, the conditional variances of $(\dot{m}_t, \dot{r}_t | x_t, z_t)$, can usefully be thought of as fractions of the corresponding unconditional variances. For the first period, $\mathrm{E}(v_{11})$ is set equal at 20% of the sampling variance of $\dot{m}_t$ and $\mathrm{E}(v_{22})$ at 40% of the sampling variance of $\dot{r}_t$. The corresponding numerical values are:

$$\mathrm{E}(v_{11}) = 0.243 \times 10^{-4}, \quad \mathrm{E}(v_{22}) = 0.233 \times 10^{-4}. \tag{2.4.3}$$

We have little grounds on which to select ν_0, which can be interpreted as the size of the "hypothetical sample" on which prior beliefs are based. Three different values will be considered: $\nu_0 = 0$ (non-informative on $\mathbf{V}$), $\nu_0 = 15$ and $\nu_0 = 30$. Note that ϕ_0 and ρ_0, as defined in (2.3.45) are invariant with respect to the choice of $\nu_0 > 0$ and so is $\mathrm{E}(\lambda|\beta)$ in (2.3.40).

The discussion in Section 2.3.4 suggests taking $\rho_0 = -0.5\,\phi_0$ so that, following (2.3.42), $v_{12}^0 = -0.5\,v_{22}^0$. This completes the first period elicitation of $\mathbf{V}_0$ which is set at zero if $\nu_0 = 0$ and is otherwise given by

$$\mathbf{V}_0 = (\nu_0 - 2)\begin{pmatrix} 0.243 & -0.117 \\ -0.117 & 0.233 \end{pmatrix} \times 10^{-4}, \quad \nu_0 = 15, 30 \tag{2.4.4}$$

Numerical integration of the bivariate prior density $D(\lambda, \beta)$—with $\nu_0 > 0$—reveals that in all cases $|\,\mathrm{E}(\lambda)| < 0.01\sigma(\lambda)$ as intended (see Table 2.4).

The period 1971(iv)–1980(ii)

The fact that we already know that β has changed sign after the introduction of CCC creates an obvious problem in our assessment of the "prior" density of β. In order to cope with this problem two different sets of prior densities are introduced.

(i) It is unlikely that in 1971 many economists would have predicted the change in the sign of β. The prior densities (2.4.1) and (2.4.2) are taken as representative of such "pre-1971" prior beliefs:

(ii) As an alternative we can put ourselves in the in the position of an economist who would have correctly inferred the positive sign of β after 1971, e.g. on the grounds that the initial impact of an *unexpected* rise in interest rates will be to increase money holdings if money is a buffer financial asset and if agents take time to adjust towards long-run equilibrium. Changing the sign of the median of β in the prior densities (2.4.1) and (2.4.2) yields densities which are representative of such "post-1971" prior beliefs.

For the rest the elicitation procedure is conducted as in Section 2.4.3, except that $\mathrm{E}(v_{11})$ is now set equal at 10% of the sampling variance of $\dot{m}_t$. $\mathbf{V}_0$b is then given by

$$\mathbf{V}_0 = (\nu_0 - 2)\begin{pmatrix} 0.647 & 0.716\mathscr{S} \\ 0.716\mathscr{S} & 1.432 \end{pmatrix} 10^{-4}, \quad \nu_0 = 15, 30 \tag{2.4.5}$$

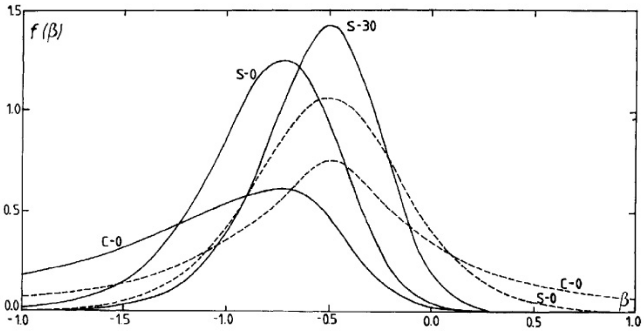

Figure 2.6: First period prior and posterior densities of β

with $\mathscr{S} = \mathrm{sign}(M(\beta))$. In the rest of the paper we use a two-character notation to identify the prior sign on β and $\mathbf{V}$: the first character refers to the prior density on β (C for a Cauchy-density and S for a t-density) and the second one indicates the value of ν_0.

2.4.4 Posterior densities

The period 1963(i)-1971(iii)

Two sets of posterior densities have been computed. In the first set, $DLM2$ is included in the reaction function and it is, therefore the weak exogeneity of $\dot{r}_t$ which is under investigation. In the second set $DLM2$ is excluded on the basis of the results given by the OLS specification search, in which case the weak exogeneity of $\dot{r}_t$ implies its strong exogeneity. The posterior means of β and λ are reported in Table 2.4 together with prior moments. Graphs of the prior and posterior densities of β and λ are reproduced in Figures 2.6 and 2.7. The results obtained under a non-informative prior density for $\mathbf{V}$ ($\nu_0 = 0$) confirm the rejection of the exogeneity of $\dot{r}_t$. The C-O graph of the posterior densities of β reveals that the (marginal) likelihood function is highly skewed towards large negative values of β. Note, furthermore that β and λ are negatively correlated. the introduction of prior information on β in the form of the t-density (2.4.2) reduces the skewness and its impact increases with ν_0.

The period 1971(iv)-1980(ii)

The posterior means and variances of β and λ under the two sets of prior densities we have introduced in Section 2.4.3 are reported in Tables 2.5(a) and 2.5(b) together with prior moments. Graphs of the posterior densities of β and λ are found in Figures 2.8 to 2.11.

The results in Table 2.5(a) are essentially unambiguous and lead to the acceptance of the weak exogeneity of $\dot{r}_t$ under minimal prior information. We note simply that the second period sample is comparatively more informative on β and less informative on λ than the first period sample. Also the negative correlation between β and λ is more pronounced.

The results in Table 2.5(b) have been derived under a prior density which is in conflict with the sample evidence. The introduction of a prior density for β centred around -0.5 shifts the posterior density of β towards negative values and the posterior density of λ towards positive values (in accordance with the negative correlation between β and λ).

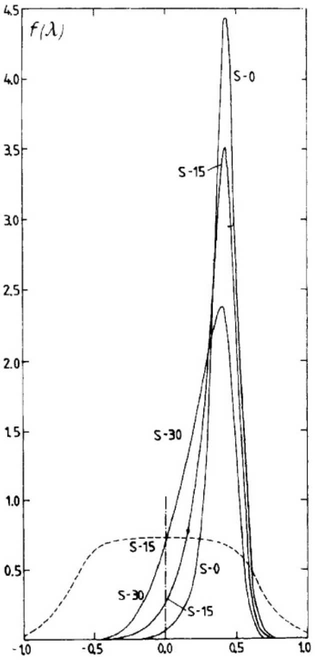

Figure 2.7: First period prior and posterior densities of λ

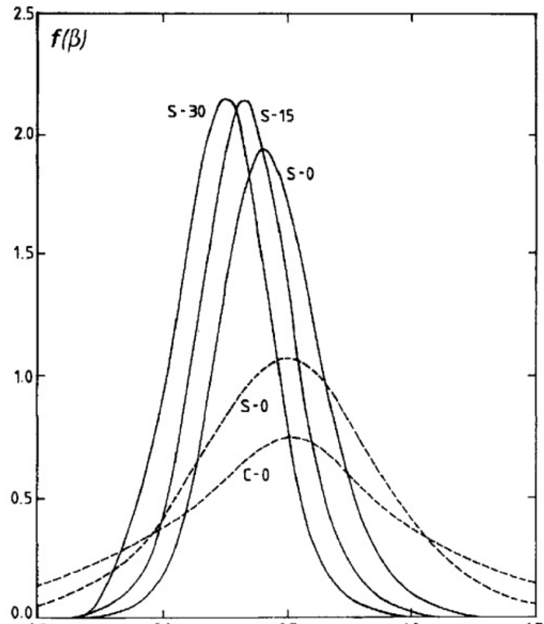

Figure 2.8: Second period prior and posterior densities of β ($\beta_0 = 0.5$)

It is, therefore, a mere coincidence that the posterior expectation of λ is close to zero when $\nu_0 = 0$. The cases where $\nu_0 = 15$ or 30 clearly indicate that a conflict between the prior and sample information can totally distort the evidence relative to exogeneity.

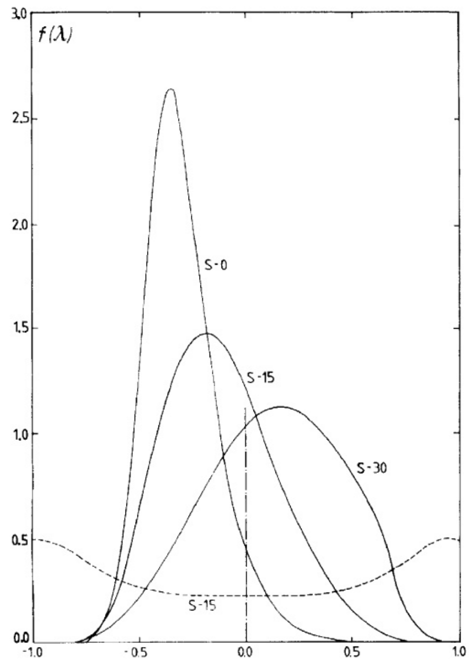

Figure 2.9: Second period prior and posterior densities of λ ($\beta_0 = 0.5$)

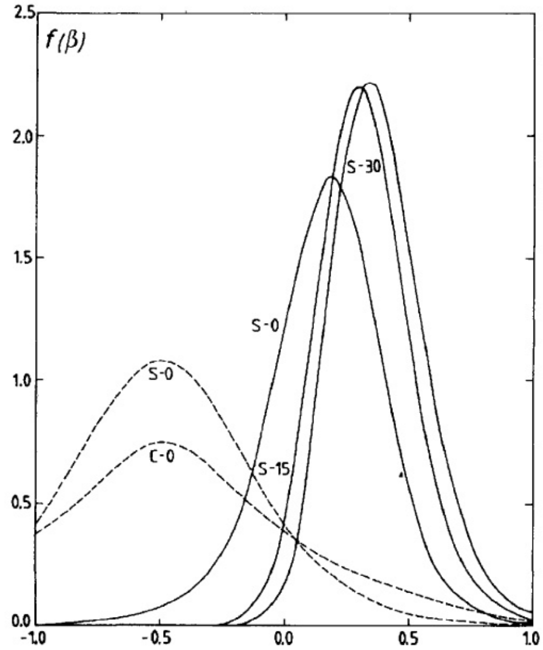

Figure 2.10: Second period prior and posterior densities of β ($\beta_0 = -0.5$)

2.5 Conclusions

A first set of conclusions concerns the applicability of the Bayesian methods to the class of problems we have discussed and, more specifically, to inference on exogeneity. The limited information maximum likelihood framework whereby all the explanatory variables in the structural equation of interest have to be included in the list of "instrumental

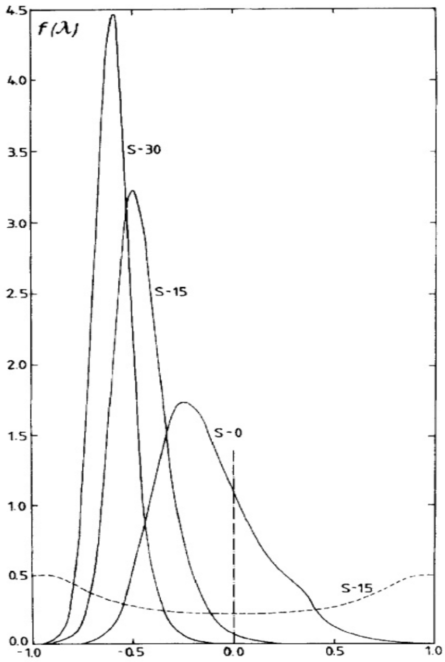

Figure 2.11: Second period prior and posterior densities of λ ($\beta_0 = -0.5$)

variables" has proved inpractical for the sample sizes we were confronted with. This has led us to develop instead an "instrumental variables" approach whereby the instrumental variables are selected solely on their own contribution to the reaction function. We have demonstrated that this more general approach remains fully tractable. Prior information on the exogeneity of a variable is easily taken into account.

The second set of conclusions concerns our money demand equation itself.

1. The evidence on the weak exogeneity of the interest rate is, at first sight, counterintuitive since it leads to rejection in the first period and acceptance in the second period while we might have expected just the opposite in the light of our description of the institutional background. However, as discussed in Section 2.3.1, weak exogeneity measures the *effectiveness* of a control policy. Situations in which agents find ways of bypassing the restrictions which are enforced upon them naturally lead to a rejection of the weak exogeneity assumption either by linking together the coefficients in the relevant equations or by inducing a significant correlation between the corresponding disturbances. Such may well have been the case in the pre-CCC period where changes in the Bank Rate were rare and carried important signalling effects (hence the presence of a term such as $DDLR3$ in our money demand equation) and where banks could probably find ways of countering the restrictive "requests" they were confronted with. In contrast the more erratic behaviour of interest rates after the introduction of CCC might have made it more difficult for the economic agents to react differently to anticipated and non-anticipated variations in the interest rates. These are, however, mere conjectures that would have to be supported by a more detailed analysis of the economic background.

2. In Section 2.3.1 we proposed three alternative explanations for the shift in the OLS estimate of β, the impact coefficient of the interest rate on money. Sample evidence unambiguously suggests that the introduction of CCC has jointly induced a large shift

in the structural coefficient β, a shift from a significantly positive σ_{uv} to a moderately negative one and a substantial change in the interest equation itself. The last two effects combine together in such a way that the OLS estimate underestimates the shift in β! It is, however, comforting to discover that simultaneity biases do not seem to have much effect on the other coefficients in the money demand equation, including those of $LMPY5$ and $LR5$ which determines the long-term impact of interest rate on the velocity of circulation of money. This empirical finding also seems to suggest that the preliminary specification search based on OLS estimation in Section 2.2 is unlikely to have severely biased our choice of a functional form for the money demand equation and, therefore, to have distorted the evidence on the exogeneity of the interest rate variable.

3. Our money demand equation presents a number of intriguing features which might deserve further investigation. Two which are specific to the second subperiod are the positive sign of the impact coefficient of interest rate, for which we have ventured a possible explanation in the course of Section 2.4.3, and the lack of significance of the own rate for which we have no explanation. A number of problems probably hinge around the existence of a long-term solution characterised by a constant velocity of circulation of money: the long lag associated with the disequilibrium feedback variable $LMPY$ (though the precise lag is essentially unidentified), the tiny ECM coefficient of $LMPY5$ itself and possibly also the unit income elasticity[8]

In fact we suspect that the two subperiods are probably more distinct than our analysis seems to suggest. The functional form we have selected originates from an overall specification search which may have been heavily influenced by the second subperiod, hence the overall positive sign of β. It did prove convenient for our purposes to have a common functional form across the two regimes for the ease of comparison and, more importantly, in order to gain degrees of freedom that were critically needed. With larger sample sizes we might have conducted independent specification searches over the two regimes but CCC has been abolished since 1981 and we might well be faced since with new coefficient changes (though the 1982 data and our reading of recent economic indicators seem to confirm the positive impact coefficient of interest rate on M3).

We would guess that the problem lies mostly with the CCC regime and that we need an ECM formulation which is coherent both in level, as the present one is, and in differences, the latter requirement being critical with such money aggregates as M3, a substantial part of which is now bearing interest.[9] Also the concept of cointegrability recently developed by Granger and Engle (1985) provides us with another route of investigation worth exploring for the second regime.[10]

The comforting message in our analysis is that such additional investigation can probably be conducted by means of OLS estimation if one's attention is restricted to the CCC

[8] We did compute t-test statistics for the addition of $LY5$ to the first equations in Table 2.1. The results are

Column	1	2A	2B	3	4
t-value:	2.18	0.22	0.53	1.72	1.89

This variable $LY5$ is clearly not significant for runs on separate periods but the imposition of coefficient restrictions across the two periods increases its significance. This might contribute towards explaining the non-unit elasticity found by Hendry and Mizon (1979) and suggests (ex post) that we might usefully consider deleting the common coefficient restrictions for $D4LY$. Such a modificatoion would, however, not affect our findings as regards the exogeneity of interest rates (compare cases 2 and 4 in Table 2.3).

[9] We are grateful to D. F. Hendry for this suggestion.

[10] Though as illustrated in figure 2.3 a "long-run" OLS regression of $LMPY$ on LR and a constant does not seem to support an hypothesis of cointegrability of M and R.

regime as it should be at this level of investigation. Also the procedures we have developed in our paper are now fully operational and could easily be applied in these and other contexts.

.1 The Data Sources

All data are quarterly and seasonally unadjusted. The following abbreviations are used:

ETAS	Economic Trends Annual Supplement (1982 edition)
BESA1	Bank of England Statistical Abstract No. 1 (1970)
BESA2	Bank of England Statistical Abstract No. 2 (1975)
FS	Financial Statistics (various issues)

M	Personal Sector M3. Cumulated from changes from the Flow of Funds accounts. Source: *BESA2* (1963–1973), *FS* (1974–1981)
R	Local Authority 3 month deposit rate (last working day). Source: *BESA1* (1963–1973), *BESA2* (1970–1974), *FS* (1975–1981)
Y	Real Personal Disposable Income (£million. 1975 prices. Source: *ETAS*
P	Implied deflator for Personal Disposable Income. Souce: *ETAS*
U	Unemployment rate (Total unemployed / Working population). Source: *ETAS*
B	Real value of UK Official Reserves (£million. 1975 prices). Source: *ETAS*

.2 Notation for Density Functions

The properties of the distribution which are presented here are found e.g. in Zellner (1971) or in Dréze and Richard (1983).

1. Multivariate Normal Distribution

$$f_N^n(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-(1/2)n}|\boldsymbol{\Sigma}|^{-1/2}\exp\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}).$$

2. Multivariate t-Distribution

$$f_t^n(\mathbf{x}|\boldsymbol{\mu},\mathbf{H},\nu) = \pi^{-(1/2)n}\Gamma\left(\frac{\nu+n}{2}\right)/\Gamma\left(\frac{\nu}{2}\right)|\mathbf{H}|^{1/2}[1+(\mathbf{x}-\boldsymbol{\mu})'\mathbf{H}(\mathbf{x}-\mu)]^{-\mathbf{(1/2)}(\nu+\mathbf{p})}.$$

3. Inverted Gamma Distribution

$$f_{i\gamma}(\sigma^2|s^2,\nu) = \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1}\left(\frac{s^2}{2}\right)^{\nu/2}(\sigma^2)^{-(1/2)(\nu+2)}\exp-\frac{1}{2}\frac{s^2}{\sigma^2}.$$

4. Inverted Wishart Distribution

$$f_{iW}^n(\boldsymbol{\Sigma}|\mathbf{S},\nu) = \left[2^{(1/2)\nu}\pi^{(1/4)n(n-1)}\prod_{i=1}^n\Gamma\left(\frac{\nu+1-i}{2}\right)\right]^{-1} \cdot|\mathbf{S}|^{(1/2)\nu}|\boldsymbol{\Sigma}|^{-(1/2)(\nu+q+1)}\exp-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S}.$$

.3 Derivation of Formulae (2.3.24) and (2.3.25)

The likelihood fiunction of our model is given in (2.3.11)–(2.3.13) with respect to $(\mathbf{c}_2, \sigma^2)$ and $(\rho, \lambda, \omega^2)$ respectively yields the following expressions:

$$\widetilde{\mathbf{c}}_2\ (\beta, \mathbf{c}_1) = -\mathbf{X}_2{}'\mathbf{X}_2)^{-1}\mathbf{X}_2{}'\mathbf{u}_1 \tag{.3.1}$$

$$\widetilde{\sigma}^2(\beta, \mathbf{c}_1) = \frac{1}{T}\mathbf{u}_1{}'\mathbf{M}_2\mathbf{u}_1 \tag{.3.2}$$

$$\widetilde{\rho}\,(\beta, \mathbf{c}_1) = -(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathbf{Y}_s - \widetilde{\lambda}(\beta, \mathbf{c}_1)\cdot \mathbf{M}_2\mathbf{u}_1] \tag{.3.3}$$

$$\widetilde{\lambda}\,(\beta, \mathbf{c}_1) = (\mathbf{u}_1{}'\mathbf{M}_Z\mathbf{u}_1)^{-1}\mathbf{u}_1{}'\mathbf{M}_Z\mathbf{Y}_s \tag{.3.4}$$

$$\begin{aligned}\widetilde{\omega}^2\,(\beta, \mathbf{c}_1) &= \frac{1}{T}\left[\mathbf{s}'\mathbf{Y}'\mathbf{M}_Z\mathbf{Y}_s - \mathbf{u}_1{}'\mathbf{M}_Z\mathbf{u}_1\cdot\widetilde{\lambda}^2(\beta, \mathbf{c}_1)\right] \\ &= \left[\widehat{\theta}\,(\beta, \mathbf{c}_1)\right]^{-1}\cdot\left|\widetilde{\boldsymbol{\Omega}}\,(\beta, \mathbf{c}_1)\right|\end{aligned} \tag{.3.5}$$

together with

$$\widetilde{h}\,(\beta, \mathbf{c}_1) = \frac{1}{T}\mathbf{u}_1{}'\mathbf{M}_Z\mathbf{u}_1 \tag{.3.6}$$

$$\widetilde{\Omega}\,(\beta, \mathbf{c}_1) = \frac{1}{T}(\mathbf{u}_1 : \mathbf{Y}_s)'\mathbf{M}_Z(\mathbf{u} : \mathbf{Y}_s). \tag{.3.7}$$

The concentrated log-likelihood function is then given by

$$\begin{aligned}L^*_{IV}(\mathbf{Y}; \beta, \mathbf{c}_1) &\propto -\frac{T}{2}\log\left[\widetilde{\sigma}^2(\beta, \mathbf{c}_1)\cdot\widetilde{\omega}^2(\beta, \mathbf{c}_1)\right] &(.3.8)\\ &\propto -\frac{T}{2}\log\left[\frac{\mathbf{u}_1{}'\mathbf{M}_2\mathbf{u}_1}{\mathbf{u}_1{}'\mathbf{M}_Z\mathbf{u}_1} - |(\mathbf{u}_1 : \mathbf{Y}_s)'\mathbf{M}_Z(\mathbf{u}_1 : \mathbf{Y}_s)|\right]. &(.3.9)\end{aligned}$$

More detail and generalisations to systems of equations are found in Richard (1984).

.4 Derivation of Formulae (2.3.25) and (2.3.26)

By application of the properties of inverted-Wishart densities as described e.g. in (Dréze and Richard, 1983, Appendix) the prior densities of $(\sigma^2|\beta)$ and $(\lambda, \delta^2|\beta)$ as derived from (2.3.32) are given by

$$D(\sigma^2|\beta) = f_{i\gamma}(\sigma^2|\sigma_0^2, \nu_0 - 1) \tag{.4.1}$$

$$D(\lambda, \omega^2|\beta) = f_N^1(\lambda|\lambda_0, \omega^2 h_0^{-1})\cdot\ f_{i\gamma}(\omega^2|\omega_0^2, \nu_0) \tag{.4.2}$$

where $(h_0, \lambda_0, \omega_0^2)$ are defined in (2.3.34) and $\sigma_0^2 = h_0$ (a distinct notation is used for σ_0^2 and h_0 because their posterior counterparts σ_*^2 and h_*, as defined below, differ and because it proves notationally convenient to have common functional forms for the prior and posterior moments of $(\sigma^2, \lambda, \omega^2|\beta)$). Let $l_i(\mathbf{Y}; \beta, \mathbf{c})$ for $i = 1, 2$ denote the "marginalised" likelihood function as derived from $L_i(\mathbf{Y}; \boldsymbol{\theta}_i)$ under the relevant prior density.

Combining together the partial likelihood function (2.3.12) and the prior density (.4.1) yields the following expressions

$$D(\sigma^2|\mathbf{Y}, \beta, \mathbf{c}) = f_{i\gamma}(\sigma^2|\sigma_*^2, \nu_* - 1) \tag{.4.3}$$

$$l_1(\mathbf{Y};\beta,\mathbf{c}) \propto (\sigma_*^2)^{-1/2(\nu_*-1)}, \tag{.4.4}$$

with

$$\sigma_*^2 = \sigma_0^2 + \mathbf{u}'\mathbf{u} + \mathbf{b}'\mathbf{V}_0\mathbf{b} + \mathbf{u}'\mathbf{u} \quad \text{and} \quad \nu_* = \nu_0 + T. \tag{.4.5}$$

The product of the partial likelihood function (2.3.13) and the prior densities (2.3.21) and (.4.2) is handled in a smilar way, except that it proves convenient to derive sequentially the posterior densities of $(p|\lambda,\omega^2,\cdot)$ and of $(\lambda,\omega^2|\cdot)$ which are respectively

$$D(p|\mathbf{Y},\lambda,\omega^2,\beta,\mathbf{c}) = f_N^k(p|p_*,\omega^2(\mathbf{Z}'\mathbf{Z})^{-1}) \tag{.4.6}$$

$$D(\lambda,\omega^2|\mathbf{Y},\beta,\mathbf{c}) = f_N^1(\lambda|\lambda_*,\omega^2 h_*^{-1}) \cdot\ f_{i\gamma}(\omega^2|\omega_*^2,\nu_*) \tag{.4.7}$$

where

$$p_* = -(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Y}s - \lambda\mathbf{u}) \tag{.4.8}$$

and, in parallel with (2.3.34),

$$\mathbf{\Omega}_* = \begin{pmatrix} h_* & h_*\lambda_* \\ h_*\lambda_* & \omega_*^2 + h_*\lambda_*^2 \end{pmatrix} = \mathbf{\Omega}_0 + (\mathbf{u}_1 : \mathbf{Y}\mathbf{s})'\mathbf{M}_Z(\mathbf{u}_1 : \mathbf{Y}\mathbf{s}). \tag{.4.9}$$

Also $l_2(\mathbf{Y};\beta,\mathbf{c})$ is given by

$$l_2(\mathbf{Y};\beta,\mathbf{c}) = l_2(\mathbf{Y};\beta,\mathbf{c}_1) \propto (\omega_*^2)^{-(1/2)\nu_*} h_*^{-1/2}. \tag{.4.10}$$

The posterior density of $(\beta,\mathbf{c})$ is given by the product of $D(\beta,\mathbf{c})$ and of the two marginalised likelihood functions (.4.4) and (.4.10) and may be rewritten as (2.3.35).

More detail and generalisations to systems of equations are found in Richard (1984).

.5 Implementation of the Bayesian Analysis in Section 2.3.3

Combining together formulae (2.3.35), (2.3.36) and (.4.2), and taking advantage of the identity $|\mathbf{\Omega}_*| = h_* \cdot \omega_*^2$ we can write the joint posterior density of β, $\mathbf{c}$ and λ as

$$D(\beta,\mathbf{c},\lambda|\mathbf{Y}) \propto D(\beta,\mathbf{c}) \cdot\ (\sigma_*^2)^{-(1/2)(\nu_*-1)} \left[\omega_*^2 + h_*(\lambda-\lambda_*)^2\right]^{-1/2(\nu_*+1)}. \tag{.5.1}$$

Throughout the rest of the discussion it is assumed that $D(\beta,\mathbf{c})$ is a t-density. In all generality $D(\mathbf{c}\,|\,\beta,\lambda,\mathbf{Y})$ is then a product of three kernels of t-densities, i.e. a so-called 3-0 poly-t density whose evaluation requires a bivariate numerical integration on an auxiliary random variable—see Richard and Tompa (1980) for details. All together the analysis of the posterior density (.5.1) requires, therefore a four-dimensional numerical integration. For the integration of β and λ we use a bivariate iterative Simpson procedure, as described in Tompa (1973), which has proved far more reliable than the other methods we have tested (such as Gaussian rules) given that the posterior density of β and λ can be extremely skewed. It implies, however, that the algorithm has to be run twice to obtain the marginal densities of β and λ since the use of a bivariate iterative Simpson rule is essentially incompatible with a complete analysis of the marginal density associated with the inner integration loop. In compensation this repetition provides a very useful check of numerical accuracy since the integrating constants and the moments are evaluated twice on different grid points. For a relative precision of the order of 1% a complete run of computation may require up to 200 minutes of CPU time on a DGMV/8000 mini computer equipped with a floating point accelerator.

The cost of computation can be divided by a factor of 20 if we use the non-informative prior density (2.3.39) on $\mathbf{c}$. In such a case σ_*^2 is the sole factor in (.5.1) which still depends on $\mathbf{c}_2$. It can be rewritten as

$$\sigma_*^2 = \mathbf{b}'\mathbf{V}_0\mathbf{b} + \mathbf{u}_1{}'\mathbf{M}_2\mathbf{u}_1 + (\mathbf{c}_2 - \mathbf{c}_2^*)'\mathbf{Z}_2{}'\mathbf{Z}_2(\mathbf{c}_2 - \mathbf{c}_2^*) \tag{.5.2}$$

with

$$\mathbf{c}_2^* = (\mathbf{Z}_2{}'\mathbf{Z}_2)^{-1}\mathbf{Z}_2{}'\mathbf{u}_1. \tag{.5.3}$$

Therefore, the conditional posterior density of $(\mathbf{c}_2|\beta, \mathbf{c}_1)$ is Student whence

$$D(\beta, \mathbf{c}_1, \lambda) \propto D(\beta)(\sigma_0^2 + \mathbf{u}_1{}'\mathbf{M}_2\mathbf{u}_1)^{-1/2(\nu_* - m_2 - 1}[\omega_*^2 + h_*(\lambda - \lambda_*)^2]^{-1/2(\nu_*+1)}. \tag{.5.4}$$

The numerical analysis of $D(\lambda, \beta, \mathbf{c}_1|\mathbf{Y})$ is then based on the following identities

$$\begin{aligned} \sigma_0^2 + \mathbf{u}_1{}'\mathbf{M}_2\mathbf{u}_1 = {} & (\mathbf{c}_1 - \mathbf{c}_{1a}^*)'\mathbf{X}_1{}'\mathbf{M}_2\mathbf{X}_1(\mathbf{c}_1 - \mathbf{c}_{1a}^*) \\ & + \mathbf{b}'[\mathbf{V}_0 + \mathbf{Y}'\mathbf{M}_2\mathbf{Y} - \mathbf{Y}'\mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1{}'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1{}'\mathbf{M}_2\mathbf{Y}]\mathbf{b} \end{aligned} \tag{.5.5}$$

$$\begin{aligned} \omega_*^2 + h_*(\lambda - \lambda_*)^2 = {} & \lambda^2(\mathbf{c}_1 - \mathbf{c}_{1b}^*)'\mathbf{X}_1{}'\mathbf{M}_Z\mathbf{X}_1(\mathbf{c}_1 - \mathbf{c}_{1b}^*) \\ & + \boldsymbol{\phi}'[\mathbf{V}_0 + \mathbf{Y}'\mathbf{M}_Z\mathbf{Y} - \mathbf{Y}'\mathbf{M}_Z\mathbf{X}_1(\mathbf{X}_1{}'\mathbf{M}_Z\mathbf{X}_1)^{-1}\mathbf{X}_1{}'\mathbf{M}_Z\mathbf{Y}]\boldsymbol{\phi} \end{aligned} \tag{.5.6}$$

together with

$$\mathbf{c}_{1a}^* = -(\mathbf{X}_1{}'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1{}'\mathbf{M}_2\mathbf{Y}\mathbf{b} \tag{.5.7}$$

$$\mathbf{c}_{1b}^* = -\frac{1}{\lambda}(\mathbf{X}_1{}'\mathbf{M}_Z\mathbf{X}_1)^{-1}\mathbf{X}_1{}'\mathbf{M}_Z\mathbf{Y}\boldsymbol{\phi} \tag{.5.8}$$

$$\boldsymbol{\phi} = \lambda\mathbf{b} - \mathbf{s}. \tag{.5.9}$$

It follows that the conditional posterior density of $(\mathbf{c}_1|\beta, \lambda)$ is a product of two kernels of Student densities, i.e. a so-called 2-0 poly-t density whose evaluation requires only a one-dimensional numerical integration on an auxiliary random variable. The analysis of the posterior density (.5.4) then require altogether a tridimensional numerical integration. The numerical procedure we have just described has proved reliable and numerically efficient. It is now part of a Bayesian Regression Computer Program (BRP) developed at CORE.

We mention finally that, as discussed in Lubrano and Richard (1981), equally efficient numerical procedures apply to the case where the independent prior density $D(\beta, \mathbf{c})$ in (2.3.29) is replaced by a conditional prior density $D(\beta, \mathbf{c}|\sigma^2)$ where σ^2 is the variance of u_t, in the form of a conventional natural conjugate prior density for the parameters of the sole equation (2.3.16). The posterior density of $(\beta, \mathbf{c}, \lambda)$, as given in (.5.1), then takes a simpler expression in that $D(\beta, \mathbf{c})$ is incorporated within σ_*^2 in the form of an additional quadratic term in (2.3.37) and a non-informative prior density on $\mathbf{c}$ is no longer required to obtain an expression similar to (.5.4). We decided, however, against using such a conditional prior density which suffers the major drawback of imposing a spurious dependence between $(\beta, \mathbf{c})$ and σ^2.

Table 2.1: OLS and WLS estimators for the UK money demand equation
Dependent variable: $DLMP$

Case	1	2A	2B	3	4
estimator	OLS	OLS	OLS	WLS	WLS
periods	A+B	A	B	A+B	A+B
$DLMP1$	0.28(0.10)	0.11(0.16)	0.36(0.15)	0.27(0.10)	0.29(0.10)
DLP	−0.36(0.13)	−0.48(0.24)	−0.44(0.22)	−0.45(0.15)	−0.38(0.13)
$DDLP2$	−0.16(0.12)	−0.36(0.20)	−0.16(0.18)	−0.21(0.13)	−0.23(0.12)
$LMPY5$	−0.04(0.01)	−0.09(0.06)	−0.03(0.02)	−0.03(0.02)	−0.04(0.01)
$D4LY$	0.08(0.05)	0.11(0.06)	0.01(0.08)	0.07(0.05)	0.08(0.05)
DLR	0.12(0.10)	−0.08(0.18)	0.17(0.15)	-0.17 (0.18) 0.16 (0.13)	-0.13 (0.17) 0.16 (0.12)
$DDLR3$	−0.06(0.07)	−0.28(0.15)	−0.04(0.10)	-0.30 (0.15) -0.01 (0.09)	-0.34 (0.14) —
$LR5$	−0.17(0.08)	−0.13(0.11)	−0.12(0.13)	−0.14(0.07)	−0.16(0.07)
SSR	0.0586	0.00115	0.00368	0.00405	0.00419
SDR	0.0100	0.0071	0.0126	0.0088	0.0086
R^2	0.78	0.73	0.83	0.77	0.77
DW	1.92	2.45	2.00	2.11	2.06
$F_{p,q}$	1.50	—	—	0.89	0.65
p, q	$15, 43$	—	—	$6, 46$	$11, 46$
$\eta_1(4)$	5.69	2.77	4.03	2.96	3.48
$\eta_2(4, k)$	0.98	0.54	0.54	0.53	0.64
k	58	23	23	52	57
$\eta_3(1)$	0.15	8.06	0.10	0.12	0.01
$\eta_4(8)$	5.63	28.78	4.81	11.43	10.67
$\eta_5(4)$	2.48	18.98	1.07	5.38	3.71
$\eta_6(4, j)$	0.50	5.63	0.15	1.00	0.74
j	54	19	19	48	53
$\eta_7(1)$	0.007	3.292	0.593	0.907	0.245

Notes

[1] The numbers in parentheses are standard errors (corrected for degrees of freedom)

[2] Joint F-test of linear restrictions against the following alternatives: OLS: unrestricted initial equation. WLS: no common coefficient across the two subperiods (conditional on a variance ratio equal to 2.8).

Table 2.2: OLS estimators of the interest rate equations
Dependent variable DLR

Period A			
$DLR2$:	$-0.28\,(0.16)$	$D2LU$:	$-0.013\,(0.007)$
$DLR4$:	$-0.21\,(0.18)$	$DDLU3$:	$-0.035\,(0.013)$
$DDLP2$:	$0.27\,(0.18)$	$D4LB$:	$-0.017\,(0.007)$
$DLM2$:	$0.10\,(0.16)$	$DDLB1$:	$-0.021\,(0.011)$
$SSR = 0.00083$ $\quad SDR = 0.0060$ $\quad R^2 = 0.59$ $\quad DW = 1.67$			
$\eta_1(4) = 32.13$ $\quad \eta_2(4, 23) = 3.33$ $\quad \eta_3(1) = 1.65$ $\quad \eta_4(8) = 9.09$			
$\eta_5(4) = 3.44$ $\quad \eta_6(4, 19) = 0.52$ $\quad \eta_7(1) = 3.66$			
Period B			
$D4LR1$:	$-0.13\,(0.10)$	$DLPU$:	$0.74\,(0.26)$
$LR2$:	$-0.19\,(0.10)$	$DLM1$:	$0.60\,(0.21)$
$DLR3$:	$0.28\,(0.15)$	DLB :	$-0.028\,(0.008)$
$DDLP$:	$-0.38\,(0.15)$	$DLB4$:	$-0.012\,(0.007)$
$DLP2$:	$-0.31\,(0.21)$	$D2LU1$:	$-0.051\,(0.019)$
$SSR = 0.00257$ $\quad SDR = 0.111$ $\quad R^2 = 0.80$ $\quad DW = 2.22$			
$\eta_1(4) = 16.60$ $\quad \eta_2(4, 23) = 1.66$ $\quad \eta_3(1) = 2.24$ $\quad \eta_4(8) = 7.47$			
$\eta_5(4) = 6.78$ $\quad \eta_6(4, 19) = 1.02$ $\quad \eta_7(1) = 2.80$			

Table 2.3: FIML (PERSEUS) estimation

Case	2_A	2_B	3		4	
Period	A	B	A	B	A	B
			Demand for money equation			
$DLMP1$	0.04 (0.088)	0.29 (0.085)	0.19 (0.061)		0.17 (0.061)	
DLP	−0.49 (0.131)	−0.41 (0.124)	−0.44 (0.093)		−0.39 (0.082)	
$DDLP2$	−0.29 (0.131)	−0.11 (0.105)	−0.15 (0.084)		−0.18 (0.076)	
$LMPY5$	−0.06 (0.034)	−0.04 (0.010)	−0.04 (0.010)		−0.04 (0.008)	
$D4LY$	0.17 (0.036)	0.01 (0.048)	0.11 (0.029)		0.11 (0.028)	
DLR	−0.51 (0.125)	0.30 (0.099)	−0.55 (0.125)	0.24 (0.092)	−0.41 (0.120)	0.23 (0.088)
$DDLR3$	−0.25 (0.088)	−0.25 (0.076)	−0.28 (0.089)	−0.03 (0.058)	−0.32 (0.086)	—
$LR5$	−0.13 (0.062)	−0.12 (0.044)	−0.15 (0.045)		−0.17 (0.044)	
			Reaction functions			
DLR2	−0.25 (0.081)		−0.23 (0.082)		−0.25 (0.085)	
DLR4	−0.18 (0.094)		−0.18 (0.094)		−0.18 (0.097)	
DDLP2	0.26 (0.101)		0.31 (0.093)		0.32 (0.095)	
D2LU	−0.02 (0.004)		−0.02 (0.004)		−0.02 (0.004)	
DDLU3	−0.03 (0.007)		−0.03 (0.007)		−0.04 (0.007)	
D4LB	−0.01 (0.004)		−0.02 (0.004)		−0.02 (0.004)	
DDLB1	−0.01 (0.006)		−0.01 (0.006)		−0.01 (0.006)	
DLM2	0.03 (0.075)		0.05 (0.077)		0.06 (0.081)	
D4LR1		−0.14 (0.055)		−0.13 (0.054)		−0.13 (0.054)
LR2		−0.19 (0.055)		−0.19 (0.054)		−0.20 (0.054)
DLR3		0.28 (0.083)		0.29 (0.083)		0.29 (0.080)
DDLP		−0.43 (0.078)		−0.42 (0.077)		−0.42 (0.077)
DLP2		−0.26 (0.112)		−0.24 (0.110)		−0.23 (0.110)
DLP4		0.71 (0.138)		0.70 (0.137)		0.70 (0.136)
DLM1		0.62 (0.111)		0.62 (0.109)		0.62 (0.109)
DLB		−0.03 (0.004)		−0.03 (0.004)		−0.03 (0.004)
DLB4		−0.01 (0.004)		−0.01 (0.004)		−0.01 (0.004)
D2LU1		−0.06 (0.010)		−0.06 (0.010)		−0.06 (0.010)
Log-likelihood	−48.29		−51.78		−53.39	
Joint ex. tests	5.3		6.0		5.1	
Indiv. ex. tests	3.5	1.8	5.0	1.0	3.9	1.2
$\|\mathbf{\Sigma}\|$	$7.1E-10$	$7.3E-9$	$8.0E-10$	$8.1E-9$	$8.3E-10$	$8.8E-9$
σ_u	0.0066	0.0105	0.0067	0.0110	0.0064	0.0114
σ_v	0.0050	0.0087	0.0050	0.0087	0.0050	0.0087
σ_{uv}	$1.9E-5$	$-3.2E-5$	$1.8E-5$	$-3.4E-5$	$1.4E-5$	$-3.5E-5$

Note: The numbers in parentheses are asymptotic standard errors (*uncorrected for degrees of freedom*)

Table 2.4: First period prior and posterior moments of β and λ

$D(\beta)$	ν_0	β	λ	ρ_{β_λ}
Prior				
S	15	−0.50	0.00	−0.74
		(0.41)	(0.44)	
S	30	−0.50	0.00	−0.84
		(0.41)	(0.39)	
Posterior ($DLM2$ included)				
C	0	—	0.37	—
			(0.11)	
S	0	−0.80	0.41	−0.25
		(0.35)	(0.11)	
S	15	−0.67	0.35	−0.60
		(0.33)	(0.15)	
S	30	−0.56	0.27	−0.77
		(0.31)	(0.19)	
Posterior ($DLM2$ excluded)				
S	0	−0.67	0.41	−0.31
		(0.25)	(0.11)	
S	15	−0.61	0.36	−0.60
		(0.24)	(0.14)	
S	30	−0.54	0.28	−0.75
		(0.23)	(0.17)	

Table 2.5: Second period prior and posterior moments of β and λ

(a): $\beta_0 = 0.5$					(b): $\beta_0 = -0.5$				
$D(\beta)$	ν_0	β	λ	ρ_{β_λ}	$D(\beta)$	ν_0	β	λ	ρ_{β_λ}
Prior					Prior				
S	15	0.50	0.00	−0.76	S	15	−0.50	0.00	−0.76
		(0.41)	(0.97)				(0.41)	(0.97)	
S	30	0.50	0.00	−0.81	S	30	−0.50	0.00	−0.81
		(0.41)	(0.91)				(0.41)	(0.91)	
Posterior					Posterior				
C	0	—	−0.29						
			(0.18)						
S	0	0.45	−0.30	−0.54	S	0	0.16	−0.12	−0.73
		(0.21)	(0.17)				(0.26)	(0.26)	
S	15	0.34	−0.12	−0.78	S	15	0.33	−0.45	−0.49
		(0.20)	(0.26)				(0.20)	(0.14)	
S	30	0.25	0.13	−0.84	S	30	0.38	−0.59	−0.22
		(0.20)	(0.31)				(0.19)	(0.10)	

Part II

Temporal Aggregation

Chapter 3

Estimating Missing Observations in Economic Time Series

Two related problems are considered. The first concerns the maximum likelihood estimation of the parameters in an ARIMA model when some of the observations are missing or subject to temporal aggregation. The second concerns the estimation of the missing observations. Both problems can be solved by setting up the model in state space form and applying the Kalman filter.

Key Words: Autoregressive-integrated-moving average processes; Kalman filter; Maximum likelihood estimation; Missing observations; Smoothing; Temporal aggregation.

3.1 Introduction

It is not unusual to encounter economic time series that are currently published at monthly or quarterly intervals but are only available on an annual basis in earlier periods. For a stock variable, such as the money supply, this means that there are missing observations in the first part of the series, while for a flow, like investment, it means that the early observations are subject to temporal aggregation. Note that a stock is the quantity of something at a particular point in time, while a flow is a quantity that accumulates over a given period of time. The relevance of these concepts is obviously not confined to economics.

This article considers two related problems: the estimation of an autoregressive-integrated-moving average (ARIMA) model based on *all* the available observations and

[0] Published in *Journal of the American Statistical Association* (1984), Vol. 79, pp. 125–131. Co-author A. C. Harvey. The work was carried out while R. G. Pierse was a Research Assistant on the program in Methodology, Inferrence and Modelling in Econometrics at the London School of Economics. Financial support from the SSRC is gratefully acknowledged. The authors would like to thank an Associate Editor of this journal and a referee for helpful comments on an earlier draft. In addition they have benefited from valuable discussions with P. Pereira and with several of the participants in the Symposium on Time Series Analysis of Irregularly Observed Data held at Texas A&M University in February 1983. Any errors remain the authors' responsibility.

the estimation of the missing values in the first part of the series. The solution of both problems lies in finding a suitable state space representation of the ARIMA model. Maximum likelihood estimation is then possible via the prediction error decomposition, and once this has been done the missing observations can be estimated by smoothing.

Section 3.2 introduces the state space methodology and shows how it can be applied to series that can be modelled by stationary ARMA processes. The application of this technique to stock variables is already fairly well known (see, e.g. Jones (1980)), but it does not seem to have been used in connection with flows. The extension to the more relevant case of an ARIMA model raises some nontrivial problems and has not been dealt with before, even for a stock variable. Two approaches to estimating ARIMA models with missing observations are described in Section 3.3. Section 3.4 deals with the prediction of future observations and describes how the missing observations are estimated by smoothing. The additional complications caused to ARIMA modelling by the logarithmic transformation are considered in Section 3.5. Some examples of the application of the techniques are given in Section 3.6, and Section 3.7 sets out a general solution to the problem considered by Chow and Lin (1971, 1976), namely the estimation of missing observations by regressing on a related series.

3.2 Stationary Processes

Since this article is only concerned with univariate time series, attention can be focused on a special case of the state space model. Let $\boldsymbol{\alpha}_t$ be an $m \times 1$ state vector that obeys the transition equation

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{R}\epsilon_t, \quad t = 1, \dots, T, \tag{3.2.1}$$

where $\mathbf{T}$ is a fixed matrix of dimension $m \times m$, $\mathbf{R}$ is an $m \times 1$ vector, and ϵ_t is a sequence of normally distributed independent random variables with mean zero and variance σ^2; that is $\epsilon_t \sim \mathrm{NID}(0, \sigma^2)$. The state vector is related to a single series of observations by the measurement equation

$$y_t = \mathbf{z}_t{}'\boldsymbol{\alpha}_t + \zeta_t, \quad t = 1, \dots, T, \tag{3.2.1b}$$

where y_t is the observed value, $\mathbf{z}_t$ is a fixed $m \times 1$ vector and ζ_t is a sequence of normally distributed independent random variables with mean zero and variance $\sigma^2 h_t$.

Let $\mathbf{a}_{t-1}$ denote the optimal or minimum mean squared estimator (MMSE) based on all the information available at time $t-1$. Let $\sigma^2\mathbf{P}_{t-1}$ denote the covariance matrix of $\mathbf{a}_{t-1} - \boldsymbol{\alpha}_{t-1}$. Given $\mathbf{a}_{t-1}$ and $\mathbf{P}_{t-1}$, the MMSE of $\boldsymbol{\alpha}_t$, $\mathbf{a}_{t|t-1}$, together with its associated covariance matrix, $\mathbf{P}_{t|t-1}$ is obtained by applying the prediction and updating equations of the Kalman filter (see, e.g., Anderson and Moore (1979), or Harvey (1981b)).

3.2.1 State space formulation of an ARMA model

A stationary ARMA(p,q) model for a sequence of normally distributed variables $y_1^\dagger, \dots, y_T^\dagger$ can be written as

$$\begin{aligned} y_t^\dagger = {} & \phi_1 y_{t-1}^\dagger + \cdots + \phi_p y_{t-p}^\dagger + \epsilon_t \\ & + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}, \quad t = 1, \dots, T, \end{aligned} \tag{3.2.2}$$

where $\phi, \ldots, \phi_p$ are the AR parameters, $\theta_1, \ldots, \theta_q$ are the MA parameters and $\epsilon_t \sim \mathrm{NID}(0, \sigma^2)$. The model can be put in state space form that obeys a transition equation of the form (3.2.1). The transition matrix $\mathbf{T}$, has ϕ_i, $i = 1, \ldots, p$ as the ith element in its first column, unity as element $(j, j+1), j = 1, \ldots, m-1$ and all other elements zero. The $m \times 1$ vector $\mathbf{R}$ is defined as $\mathbf{R} = (1, \theta_1, \ldots, \theta_{m-1})'$, where $\theta_{q+1}, \ldots, \theta_{m-1}$ are zero if $m > q+1$. This particular matrix and vector will be denoted $\mathbf{\Phi}$ and $\boldsymbol{\theta}$ in future sections. Given these definitions the first element in $\boldsymbol{\alpha}_t$ is identically equal to $y_t^{\dagger}$. Thus in the measurement equation, (3.2.1b), $\mathbf{z}_t = (1, 0, \ldots, 0)'$ for $t = 1, \ldots, T$, and if $y_t^{\dagger}$ is observed without error, $y_t = y_t^{\dagger}$ and $h_t = 0$.

Because the model is stationary, the initial conditions for the Kalman filter are given by $\mathbf{a}_{1|0} = \mathbf{0}$ and $\mathbf{P}_{1|0} = \sigma^{-2}\,\mathrm{E}(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t{}')$. Given that an observation is available in every time period, the Kalman filter produces a set of T prediction errors or innovations, $\nu_t = y_t - \mathbf{z}_t{}'\mathbf{a}_{t|t-1}$ for $t = 1, \ldots, T$. These can be used to construct the likelihood function by the prediction error decomposition; that is,

$$\log L(y_1, \ldots, y_T; \mathbf{\Phi}, \boldsymbol{\theta}, \sigma^2) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\log f_t - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\nu_t^2/f_t. \quad (3.2.3)$$

The quantities $f_1, \ldots, f_T$ each of which is proportional to the variance of the corresponding innovation, are also produced by the Kalman filter. The parameter σ^2 does not appear in the Kalman filter, and it can be concentrated out of (3.2.3).

Careful programming of the Kalman filter recursions leads to a very efficient algorithm for evaluating the exact likelihood function of an ARMA model; compare the evidence presented in Gardner et al. (1980). Various modifications of the Kalman filter can also be used for this purpose (see, e.g. Pearlman (1980)).

3.2.2 Missing observations on a stock variable

A missing observation on a stock variable can be handled very easily simply by bypassing the corresponding updating equation. Skipping the missing observations in this way makes no difference to the validity of the prediction error decomposition provided that when an observation is missing the corresponding $\log f_t$ term is omitted from the likelihood. Thus the likelihood function is of the form given in (3.2.3) but with the summations covering only those values of t for which the variable is actually observed. The T appearing in the first two terms is replaced by the number of observations.

3.2.3 Temporal aggregation of a flow variable

Let n denote the maximum number of time periods over which a flow variable is aggregated and let y_{t-1}^* be the $(n-1) \times 1$ vector $(y_{t-1}^{\dagger}, \ldots, y_{t-n+1}^{\dagger})'$. Define an $(m+n-1) \times 1$ augmented state vector, $\boldsymbol{\alpha}_t = (\overline{\boldsymbol{\alpha}}_t' \;\; y_{t-1}^*)'$, where $\overline{\boldsymbol{\alpha}}_t$ is the state vector appropriate to the ARMA model for $y_t^{\dagger}$. The transition equation for the augmented state vector is

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \mathbf{\Phi} & & \mathbf{0} \\ 1 & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_{n-2} & \mathbf{0} \end{bmatrix} \boldsymbol{\alpha}_{t-1} + \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix} \epsilon_t, \quad t = 1, \ldots, T \quad (3.2.4)$$

(When $q = 0$, a more economical state space representation is obtained by redefining m as $\max(p, n)$ and letting the state vector be $\boldsymbol{\alpha}_t = (y_t^{\dagger}, \ldots, y_{t-m+1}^{\dagger})'$. The matrix $\mathbf{T}$ in the transition equation is then the transpose of $\mathbf{\Phi}$.) If at time t, the aggregate y_t of the previous $n(t)$ terms in the series is observed, the measurement equation is

$$y_t = (1, \mathbf{0}_{m-1}{}', \mathbf{i}_{n(t)-1}{}', \mathbf{0}_{n-n(t)}{}')\boldsymbol{\alpha}_t = \sum_{j=0}^{n(t)-1} y^{\dagger}_{t-j}, \quad 1 \leq t \leq T, \tag{3.2.5}$$

where $\mathbf{i}$ is an $(n(t)-1) \times 1$ vector of ones. In periods when there is no observation the updating equation can be skipped in the same way as for a stock variable. Note that the definition of the vector $\mathbf{z}_t$ changes as the basis upon which the variable is aggregated changes.

The starting values for the augmented state space model are $\mathbf{a}_{1|0} = 0$ and $\mathbf{P}_{1|0} = \sigma^{-2}\,\mathrm{E}(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t{}')$. However, if a run of disaggregated observations is available at the end of the series, the calculations can be simplified by working backwards. This is quite legitimate since if $y^{\dagger}_t$, $t = 1, \ldots, T$ is generated by an ARMA $(p,\, q)$ process, the observations taken from $t = T$ to $t = 1$ can be regarded as being generated by exactly the same process; see (Box and Jenkins, 1976, pp. 197–198). The calculations begin with the state model appropriate to $y^{\dagger}_t$ and only when the agregate observations start to arrive is a transfer to the augmented model made. The advantage of this approach is that the initial $m \times m$ matrix $\widehat{\mathbf{P}}_{1|0} = \sigma^{-2}\,\mathrm{E}(\widehat{\boldsymbol{\alpha}}_t \widehat{\boldsymbol{\alpha}}_t{}')$ can be evaluated using standard algorithms. When the run of disaggregate observations comes to an end, the MMSE of the augmented state vector and the associated $\mathbf{P}_t$ matrix can be formed immediately since all the observations in the vector corresponding to y^*_{t-1} are known. Note that when the observations are processed in reverse, the aggregate observations must be regarded as arising at the beginning of the period of aggregation rather than at the end.

3.3 Nonstationary Process

In general, economic time series are nonstationary and the usual approach is to fit an ARIMA model. Thus if Δ denotes the first difference operator and Δ_s denotes the seasonal difference operator (for a season of s periods), the series $\Delta^d \Delta_s^D y^{\dagger}_t = w^{\dagger}_t$ is modelled as a stationary (seasonal) ARMA process.

There are basically two ways of constructing the likelihood function for an ARIMA model with missing observations. The first approach formulates the state space model in such a way that the observations are in levels, while the second has the observations in differences. The choice between them depends on the pattern of missing observations. If they are missing at regular intervals, the algorithm based on differenced observations may be preferrable. The levels formulation is, however, more flexible. In addition it forms the basis for the smoothing algorithm.

3.3.1 Levels formulation

Let L be the lag operator and let $-\delta_j$ be the coefficient of L^j in the expansion of $\Delta^d \Delta_s^D = (1-L)^d (1-L^s)^D$. Let $\overline{\boldsymbol{\alpha}}_t$ be the state vector in the state space model for the stationary ARMA$(p,\, q)$ process, $w^{\dagger}_t$, and define the augmented state vector, $\boldsymbol{\alpha}_t = (\overline{\boldsymbol{\alpha}}_t{}' \quad y^*_{t-1}{}')'$, where $y^*_{t-1} = (y^{\dagger}_{t-1}, \ldots, y^{\dagger}_{t-d-sD})'$. The transition equation for the augmented state vector is

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \overline{\boldsymbol{\alpha}}_t \\ y^*_{t-1} \end{bmatrix} = \begin{bmatrix} & \boldsymbol{\Phi} & \mathbf{0}' & \\ 1 & 0\cdots 0 & \delta_1\cdots\delta_k & \\ & \mathbf{0}' & \mathbf{I}_{k-1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{\alpha}}_{t-1} \\ y^*_{t-2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix} \epsilon_t, \quad t = 1, \ldots, T, \qquad (3.3.1)$$

where $k = d + sD$.

If $y^\dagger_t$ is observed for all $t = 1, \ldots, T$, the measurement equation is

$$y_t = (1, \mathbf{0}_{m-1}{}', \delta_1 \cdots \delta_k)\boldsymbol{\alpha}_t, \quad t = 1, \ldots, T, \qquad (3.3.2)$$

and the Kalman filter can be initialised at $t = k$ with $\mathbf{a}_{k+1|k} = (\mathbf{0}', y^*_{k+1}{}')'$ and

$$\mathbf{P}_{k+1|k} = \begin{bmatrix} \overline{\mathbf{P}}_{1|0} & 0 \\ 0 & 0 \end{bmatrix}, \qquad (3.3.3)$$

where $\overline{\mathbf{P}}_{1|0} = \sigma^{-2}\,\mathrm{E}(\overline{\boldsymbol{\alpha}}_t\overline{\boldsymbol{\alpha}}_t{}')$. It is not difficult to show that the likelihood function constructed from the Kalman filter is identical to the likelihood function that would result from applying the Kalman filter to the differenced observations $\Delta^d\Delta^D_s y_t$; compare a similar argument in Harvey (1981a).

For stock variables, missing observations can be handled in the same way as described in Section 2. For a flow variable, the measurement equation for an aggregate observation of the form

$$y_t = \sum_{j=0}^{n(t)-1} y^\dagger_{t-j}$$

is

$$y_t = (1, \mathbf{0}_{m-1}{}', \delta_1 + 1, \ldots, \delta_{n(t)-1} + 1, \delta_{n(t)}, \ldots, \delta_k)'\boldsymbol{\alpha}_t. \qquad (3.3.4)$$

This assumes $n-1 \leq d+sD$; if this is not the case, y^*_{t-1} must be redefined as $(y^\dagger_{t-1}, \ldots, y^\dagger_{t-n+1})'$.

The only problem with the levels formulation concerns starting values since in most applications the complete set of values $y^\dagger_1, \ldots, y^\dagger_k$ will not be available. However, if at least k consecutive observations are available at the end of the series the problem can be solved by reversing the order of the observations.

3.3.2 Difference formulation

Suppose that observations on a stock variable are available every n time period. An immediate difficulty arises with an ARIMA model because it may not be possible to construct the differenced observations $\Delta^d\Delta^D_s y^\dagger_t$, from such a sequence. Thus, for example, first differences cannot be formed if the variable in question is only observed every other time period. The solution to the problem is to construct a differenced series that can be observed. If $s \geq n$ and s/n is an integer, the observable differenced series is

$$y_t = \Delta^d_n \Delta^D_s y^\dagger_t, \quad t = nd + SD, n(d+1) + sD, \ldots, \qquad (3.3.5)$$

where

$$\Delta^d_n = (1 - L^n)^d = (1 - L)^d(1 + L + \cdots + L^{n-1})^d. \qquad (3.3.6)$$

Expression (3.3.5) becomes

$$y_t = (1 + L + \cdots + L^{n-1})^d w^\dagger_t, \qquad (3.3.7)$$

where $w_t^\dagger$ is the underlying ARMA process, $\Delta^d \Delta_s^D y_t^\dagger$. Thus although $y_t^\dagger$ is a stock, the observable differenced series, y_t, is a flow when considered in terms of $w_t^\dagger$. The techniques described for flow variables in Section 2 can be applied directly, although if $d > 1$ the weights for different lags of $w_t^\dagger$ are not the same, and the measurement equation must be amended accordingly. Similar methods can be applied when the original variable is itself a flow.

3.4 Predicting Future Observations and Estimating Missing Observations

Once the parameters of the ARIMA model have been estimated, optimal predictions of future observations, together with their conditional mean squared errors (MSE's), can be made by repeated application of the Kalman filter prediction equations. Similarly, MMSE's of the missing observations can be computed by smoothing. The levels form of the model will normally be appropriate for both purposes.

There are a number of smoothing algorithms available. The best known is probably the fixed interval algorithm; see (Anderson and Moore, 1979, pp. 187–190) or (Harvey, 1981b, Ch. 4). It is based on a set of backward recursions starting at time T, but it has two drawbacks in the present context. The first is that it requires the storage of a large number of covariance matrices, the $\mathbf{P}_t$ and $\mathbf{P}_{t|t-1}$'s computed from the initial pass through the Kalman filter. The second is that the inverse of $\mathbf{P}_{t|t-1}$ is needed and if some of the elements in $\boldsymbol{\alpha}_t$ are known at time $t-1$, this matrix will be singular. Neither of these problems is insurmountable, but for the kind of situations with which we are concerned, a fixed-point smoothing algorithm is more attractive.

Fixed-point smoothing can be applied by proceeding with the Kalman filter and augmenting the state vector each time a missing observation is encountered. Once all the observations have been processed, the components added to the state vector will contain the MMSE's of the missing observations. The corresponding MSE's can be obtained directly from the associated augmented covariance matrix.

The recursions for the augmented parts of the state vector can, in fact, be separated from the Kalman filter recursions for the original state vector. This is extremely useful since it means that a new series of recursions can simply be started off with each missing observation, leaving the basic Kalman filter undisturbed. The form of these recursions is as follows. Suppose that the underlying variable $y_t^\dagger$, is not observed at time $t = \tau$. The state vector is augmented by $y_\tau^\dagger$, which is, it should be noted, a linear combination of the elements of $\boldsymbol{\alpha}_\tau$; that is,

$$y_\tau^\dagger = \mathbf{z}'\boldsymbol{\alpha}_\tau, \tag{3.4.1}$$

where $\mathbf{z}$ is constant throughout the series. (Note that for a flow variable $\mathbf{z}_t$ will not be the same as $\mathbf{z}$ when there is temporal aggregation. However, for a stock variable, $\mathbf{z}_t = \mathbf{z}$ whenever there is an observation.) Modifying the formulas in (Anderson and Moore, 1979, 172–173) to take account of the fact that only a linear combination of $\boldsymbol{\alpha}_\tau$ is to be estimated leads to the smoothing recursions

$$\overline{y}_{\tau|t} = \overline{y}_{\tau|t-1} + \mathbf{p}_{\tau|t-1}{}'\mathbf{z}_t f_t^{-1} \nu_t, \quad t = \tau, \ldots, T \tag{4.2a}$$

and

$$\mathbf{p}_{\tau|t} = \mathbf{T}(\mathbf{I} - \mathbf{g}_t \mathbf{z}_t{}')\mathbf{p}_{\tau|t-1}, \quad t = \tau, \ldots, T, \tag{4.2b}$$

where $\mathbf{g}_t = \mathbf{P}_{t|t-1}\mathbf{z}_t f_t^{-1}$. The initial values are $\overline{y}_{\tau|\tau-1} = \mathbf{z}'\mathbf{a}_{\tau|\tau-1}$ and $\mathbf{p}_{\tau|\tau-1} = \mathbf{P}_{\tau|\tau-1}\mathbf{z}$. The quantities f_t, ν_t, and $\mathbf{g}_t$ are all produced by the Kalman filter for the original state space model, the vector $\mathbf{z}_t$ being used to define the measurement equation. If there is no observation at time t, then (4.2a) and (4.2b) collapse to $\overline{y}_{\tau|t} = \overline{y}_{\tau|t-1}$ and $\mathbf{p}_{\tau|t} = \mathbf{T}\mathbf{p}_{\tau|t-1}$, respectively. The MSE of $\overline{y}_{\tau|T}$ is given by $\sigma^2 f_{\tau|T}$, where $f_{\tau|T}$ is obtained from the recursion

$$f_{\tau|t} = f_{\tau|t-1} - \mathbf{p}_{\tau|t-1}{}'\mathbf{z}_t f_t^{-1}\mathbf{z}_t{}'\mathbf{p}_{\tau|t-1}, \quad t = \tau, \ldots, T \tag{3.4.3}$$

with $f_{\tau|\tau-1} = z'\mathbf{P}_{\tau|\tau-1}\mathbf{z}$.

When set up in this way, the fixed-point smoothing algorithm is extremely efficient. The storage requirements are negligible and in a typical application, the time taken to run the augmented Kalman filter is usually less than twice the time taken for a normal run. This is trivial compared with the time taken to compute the ML estimators of the unknown parameters.

3.5 Logarithmic Transformations

It is very common to take logarithms of a variable before fitting an ARIMA model. This creates no difficulties whatsoever for a stock variable. For a flow variable, however, an immediate problem arises because it is the sum of the original variables that is observed and the logarithm of a sum is not equal to the sum of the logarithms. Assuming that the logarithms of the aggregated variables are normally distributed is then inconsistent with the assumption that the corresponding disaggregated variables are normal. Notwithstanding this point, one way to proceed is to assume that the logarithms of all variables actually observed are normally distributed. The logarithm of the observed aggregate at time t is

$$y_t = \log \sum_{j=0}^{n(t)-1} \exp(y_{t-j}^{\dagger}), \tag{3.5.1}$$

where $y_t^{\dagger}$ is the underlying ARIMA process. Adopting the notation of (3.3.4), but assuming that $n = k$ for simplicity, the measurement equation can be written as

$$y_t = \log[\exp\{(1, \mathbf{0}_{m-1}{}', \delta_1, \ldots, \delta_n)\boldsymbol{\alpha}_t\} + \sum_{j=m+1}^{m+n} \exp(\alpha_{jt})], \tag{3.5.2}$$

where α_{jt} is the jth element in $\boldsymbol{\alpha}_t$. this equation is obviously nonlinear but by using the extended Kalman filter, as in (Anderson and Moore, 1979, pp. 193–195), an approximation to the likelihood function can be computed by the prediction error decomposition.

3.6 Example

The airline passenger data given in (Box and Jenkins, 1976, p. 531) consists of 144 monthly observations on the number of passengers carried by airlines over the period 1949 to 1960. It is highly seasonal and (Box and Jenkins, 1976, Ch. 9) fitted the following model to the logarithms of the observations:

$$\Delta\Delta_{12} y_t = (1 + \theta_1 L)(1 + \theta_{12} L^{12})\epsilon_t. \tag{3.6.1}$$

The above model was estimated using four variations of the data set: (i) all 144 observations; (ii) the observations from January to November deleted for the last six years; (iii) the logarithms of the observations of each of the last six years aggregated and assigned to December; (iv) the raw observations for each of the last six years aggregated and assigned to December. The second data set represents an example of missing observations, with the variable treated as though it were a stock observed annually, rather than monthly, for part of the sample period. The third and fourth data sets are examples of temporal aggregation. Data set (iii) would be relevant if the observations used in the ARIMA model were original observations rather than logarithms. Note that placing the missing or temporally aggregated observations at the end of the series, whereas in practice they might come at the beginning, is unimportant. As already noted, the order of the observations can always be reversed without affecting the underlying ARIMA model.

Table 3.1: Maximum Lkelihood Estimates of Parameters for Airline Passenger Model, (3.6.1)

	Parameters[a]	
Data set	θ_1	θ_{12}
(i) Full Data	$-.402$	$-.557$
	(.090)	(.073)
(ii) Missing Observations	$-.457$	$-.758$
	(.121)	(.236)
(iii) Temporal Aggregation (logs)	$-.475$	$-.741$
	(.114)	(.223)
(iv) Temporal Aggregation (raw data)	$-.477$	$-.738$
	(.114)	(.221)

[a] Figures in parentheses are asymptotic standard errors.

The computer program we wrote is a fairly general one that can handle both missing observations and temporal aggregation. The pattern of missing observations need not be regular. The only proviso is that there should be a run of $d + sD$ observed values at either the beginning or the end of the series. In writing the program, considerable care was taken to devise a computationally efficient routine for evaluating the likelihood function. Particular attention was paid to the evaluation of $\overline{\mathbf{P}}_{1|0}$, the matrix used to initialise the Kalman filter, and the algorithm adopted is described in some detail in our original research report (Harvey and Pierse (1982)). Maximisation of the likelihood function was carried out by one of the Gill-Murray-Pitfield numerical optimisation routines in the UK NAG library, E04JBF. This is a Quasi-Newton algorithm that uses numerical derivatives and allows simple bounds to be placed on the parameters. By choosing a suitable parameterisation, we were able to devise a very effective procedure in which we were able to constrain the roots of the MA polynomial to lie outside or on the unit circle; again see (Harvey and Pierse, 1982, Appendix B). The reason for allowing strictly noninvertible MA processes is set out in Harvey (1981b).

The results of exact ML estimation are shown in Table 3.1. The estimates obtained with data sets (ii), (iii) and (iv) are quite close to the estimates obtained with the full set of observations. The higher asymptotic standard errors are a reflection of the smaller number of observations.

Table 3.2: Estimates of Logarithms of Missing Observations and Associated Root Mean Squared Errors for 1957

	Month											
Data set	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
(ii) Missing Observations	5.733	5.738	5.893	5.850	5.843	5.951	6.051	6.055	5.938	5.8123	5.680	—
	(.045)	(.049)	(.052)	(.054)	(.055)	(.055)	(.055)	(.054)	(.052)	(.049)	(.045)	
(iii) Temporal Aggregation (logs)	5.770	5.778	5.937	5.896	5.890	5.997	6.094	6.093	5.971	5.839	5.700	5.818
	(.041)	(.040)	(.039)	(.038)	(.037)	(.037)	(.037)	(.037)	(.038)	(.039)	(.040)	(.041)
(iv) Temporal Aggregation (raw data)	5.772	5.779	5.939	5.848	5.893	6.001	6.098	6.099	5.976	5.844	5.704	5.823
	(.041)	(.041)	(.039)	(.038)	(.037)	(.036)	(.036)	(.036)	(.037)	(.039)	(.041)	(.041)
Actual Values	5.753	5.707	5.875	5.852	5.872	6.045	6.146	6.146	6.001	5.849	5.720	5.817

Table 3.2 shows the estimates of the missing observations for 1957 computed by the smoothing algorithm described in Section 3.4. The root mean squared errors associated with each estimate are

$$\mathrm{RMSE}(\overline{y}_{t|T}) = \widetilde{\sigma}\sqrt{f_{t|T}},$$

where $f_{t|T}$ is given by (3.4.3) and $\widetilde{\sigma}$ is the square root of the ML estimator of σ^2. For data sets (ii) and (iii) the estimates are all remarkably close to the actual values. Similar results were obtained for the other years in which there were missing observations. The theoretical justification for the estimates obtained when there is temporal aggregation but the model is in logarithms—case (iv)— is somewhat weaker because of the approximation involved in the use of the extended Kalman filter. However, the results presented in Table 3.2 lend some support to the use of this device in smoothing, as well as estimation. The figures presented for data set (iv) are virtually indistinguishable from those derived for data set (iii).

The results in Table 3.2 refer to estimates of the logarithms of the missing observations. If x_t denotes the original observation, a direct estimate of a missing x_t is given by $\overline{x}_{t|T} = \exp(\overline{y}_{t|T})$. However, the relationship between the normal and lognormal distributions suggests the modified estimator

$$\overline{x}_{t|T}{}^{*} = \exp\left\{\overline{y}_{t|T} + \frac{1}{2}\,\mathrm{MSE}(\overline{y}_{t|T})\right\}. \tag{3.6.2}$$

The estimator is unbiased in the sense that the expectation of $x_t - \overline{x}_{t|T}{}^{*}$ is zero when the parameters of the underlying ARIMA model are known. For the airline passenger data, the use of the modification in (3.6.2) made very little difference. In the case of (ii), for example, the direct and modified estimates for May 1957 were 344.8 and 345.4 respectively. The true value is 355. For the same data point, the 95% prediction interval was 309.5 to 384.1.

3.7 Regression

Chow and Lin (1971, 1976) approach the problem of estimating missing observations by assuming that $y_t^{\dagger}$ is related to a set of k nonstochastic variables that are observed in all time periods. They assume a linear relationship of the form

$$y_t^{\dagger} = \mathbf{x}_t{}'\boldsymbol{\beta} + u_t, \quad t = 1, \ldots, T, \tag{3.7.1}$$

where $\mathbf{x}_t$ is the $k \times 1$ vector of related variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters, and u_t is a stationary stochastic disturbance term.

Given the covariance matrix of the disturbances, finding estimates of the missing observations is basically an exercise in best linear unbiased estimation and prediction. However, it does require the construction and inversion of the covariance matrix associated with the variables (aggregates in the flow case) actually observed. Furthermore the covariances between the missing values and the observed values must also be found. In solving the problem in this way Chow and Lin concentrate on situations where the observations are missing at regular intervals and the disturbances are either serially uncorrelated or generated by an AR(1) process.

The Kalman filter can be applied directly to (3.7.1) by using the techniques described in Section 3.2, with $y_t^\dagger$ replaced by $y_t^\dagger - \mathbf{x}_t{}'\boldsymbol{\beta}$. The likelihood function must then be maximised nonlinearly with respect to $\boldsymbol{\beta}$ as well as the ARMA parameters and any estimates of MSE's obtained in a subsequent smoothing operation will only be conditional on $\boldsymbol{\beta}$. This constitutes a possible disadvantage compared to the Chow-Lin approach although it does avoid the necessity for repeated inversions of relatively large covariance matrices. However, it is possible to conserve the advantages of the Chow-Lin approach while applying the Kalman filter by redefining the state space model to include $\boldsymbol{\beta}$ in the state vector (cf. Harvey and Phillips (1979)). This formulation yields the BLUE of $\boldsymbol{\beta}$ for given values of the ARMA parameters and enables $\boldsymbol{\beta}$ to be concentrated out of the likelihood function. (If consistent estimators of the ARMA parameters can be obtained, this estimator of $\boldsymbol{\beta}$ will be asymptotically efficient under suitable regularity conditions.) Note that it will normally be necessary to process the observations backwards to obtain starting values.

The above techniques can also be used if the regression model has been framed in first differences as suggested by Denton (1971) and Fernandez (1981). In this case the model is

$$\Delta y_t^\dagger = (\Delta \mathbf{x}_t)'\boldsymbol{\beta} + u_t, \quad t = 2, \ldots, T. \tag{3.7.2}$$

The Kalman filter can handle models in which u_t is an ARMA process rather than the serially uncorrelated process assumed in the references just cited. In fact even when u_t is serially uncorrelated, the Kalman filter may still be advantageous since the inversion of the $(T-1) \times (T-1)$ matrix in expressions (3) and (4) of the paper by Fernandez is avoided.

Finally, it is worth noting that one suggestion made by Chow and Lin is to use a time trend and seasonal dummies as the explanatory variables in (3.7.1). Computing estimates of the missing values from such a model can also be carried within the ARIMA framework described in Sections 3.3 and 3.4. If the disturbance in (3.7.1) is a serially uncorrelated process, ϵ_t, the model is a special case of (3.6.1) in which $\theta_1 = \theta_{12} = -1$. Although this model is strictly noninvertible, starting off the Kalman filter in the manner suggested in Section 3.3 ensures that estimates of missing values and predictions of future observations are exactly the same as if the model had been estimated within the regression framework of Chow and Lin (cf. Harvey (1981a)).

3.8 Conclusion

The results reported in Section 3.6 show that maximum likelihood estimation of ARIMA models can be carried out efficiently when there are missing or temporally aggregated observations. Furthermore, mimimum mean squared estimates of the missing observations together with their conditional root mean squared errors, can be computed at vey little extra cost. Additional complications arise with temporal aggregation when the ARIMA model is based on logarithms, but an approximate solution can be obtained by the extended Kalman filter. This solution is not altogether satisfactory from the theoretical point of view, although it does seem to give quite reasonable results with the airline data.

Although the approach developed here can handle most configurations of missing values, it does need an unbroken run of observations at the beginning or end of the series. One way of relaxing this requirement is by modifying an algorithm given in Rosenberg (1973). An indication of how this may be done can be found in Harvey and McKenzie (1983).

Chapter 4

Temporal Aggregation and the Power of a Unit Root

The asymptotic local power of unit root tests with the same data span is shown to be independent of sampling frequency. A measure of the power trade-off between sampling frequency and time span for distinct alternatives is derived using an approximate slopes approach. Only small span increases are generally required to maintain power when reducing sampling frequency. Monte Carlo results support the asymptotic analysis for finite samples. An application is made to a consumption function for the UK. Cointegration of consumption and wealth is rejected with quarterly data but convincingly accepted with a longer span of annual data.

Key Words: Unit roots; Power; Sampling; Temporal aggregation
JEL Classification: C12; C15; C22

4.1 Introduction and summary

The effects of sampling frequency on estimation and inference in economic models have been explored in the literature in a number of different ways. Nijman and Palm (1990), for example, analyse the gain in forecasting efficiency obtained by using monthly instead of quarterly models to forecast monthly series and weigh this against the cost of collecting monthly National Accounts data. Lippi and Reichlin (1991) assess the effect of temporal aggregation on estimated measures of persistence in the context of trade cycle analysis while Hotta et al. (1993) look at problems arising from overlapping aggregation caused by commonly used smoothing procedures.

This paper focuses on the effect of temporal aggregation on the power of tests for a unit root in simple time series models, following the line of previous work by Shiller and Perron (1985) and Perron (1989b, 1991). The topic of testing for the presence of a unit root in economic time series has received much attention in the literature following the influential paper by Nelson and Plosser (1982), and various test statistics have been

[0] Published in *Journal of Econometrics* (1995), Vol. 65, pp. 333–345. Co-author Andy Snell. The authors would like to thank Len Gill and Steve Satchell for helpful comments. Financial support from ESRC grant no. R000232912 is gratefully acknowledged.

proposed by Dickey and Fuller (1979, 1981), Phillips (1987a) and Phillips and Perron (1988) among others. Unit root tests also form the basis of the tests for cointegration between economic time series proposed by Engle and Granger (1987) and the power of these tests to reject the null hypothesis of a unit root (implying no cointegration) is of considerable importance. In this paper the relationship between power and sampling frequency is explored. This is an issue of direct practical relevance to the applied researcher who often has the option of choosing between data sets of different frequencies covering different time spans.[1] The macroeconomist, for example, typically has available either quarterly observations since the war or annual data since the beginning of the century. By contrast, in the study of financial markets, it is possible to collect some data on an almost continuous basis over a short period of time or use the published sources which cover a much longer time span but at weekly or monthly intervals.

There is a widely held view in empirical work that long data spans are important for identifying mean reversion in slowly decaying processes (see, for example, Diebold *et al.* 1991)). This view is supported by the Monte Carlo results of Shiller and Perron (1985) and Perron (1989b) who find that 'over a substantial range of parameter values . . . [power depends more] on the span of the data rather than the number of observations' (*op. cit.* p.381). At the analytical level, Perron (1989b) finds that with *fixed* alternatives, tests for a unit root are only consistent when the time span rises with the number of observations.

In Section 4.2 of the paper, the view that it is the data span rather than the number of observations that affects power in the context of testing for unit roots is expressed more forecfully than previously by showing that, when comparing models with the same data span, the asymptotic local power of a one-sided unit root test is *independent* of the frequency of sampling. In Section 4.3, we turn to lookat power against fixed alternatives using an approach due to Geweke (1981). Using the almost sure limit of the ratio of competing test statistics (comparing a model in 'basic' time units with a corresponding temporally aggregated model) as a measure of relative power, a theorem is stated that allows this ratio to be interpreted as the increase in span required to maintain power fixed at some arbitrary level when moving from the basic model to the aggregated model. This provides a measure of the trade-off between sampling frequency and time span when the interest is in power against a distinct rather than a local alternative. In most cases it is found that the increase in span that is required to keep power fixed is very small relative to the order of temporal aggregation. In order to help the investigator assess the absolute power of a unit root test given a particular data set, an approximation to the power function is proposed that is easy to compute. Section 4.4 presents the results of some Monte Carlo simulations for some simple models. These show that the asymptotic analytical results of Sections 4.2 and 4.3 perform well in finite samples. Finally, in Section 4.5 the results of the paper are applied to a model of consumer behaviour estimated by Molana (1991). Although Molana was unable to reject the null hypothesis that consumption and wealth are not cointegrated, using the short span of available quarterly data, when his model is reestimated using annual data over a longer span, the null hypothesis is convincingly rejected.

[1] Harvey and Pierse (1984) show how data sets of different frequencies can sometimes be combined using the Kalman filter. However, this requires knowledge of the order of integration of the series which must be determined in some way, such as by a unit root test.

4.2 Asymptotic local power and temporal aggregation

Let y_t be a variable generated by the discrete first order process[2]

$$y_t = \rho y_{t-1} + u_t \quad t = 1, 2, \ldots, n_b \tag{4.2.1}$$

where n_b is the number of observations in 'basic' time units and u_t is an error which is assumed to follow a finite order stationary $ARMA(p, q)$ process whose largest AR root has modulus less than ρ.

Eq. (4.2.1) represents a model in 'basic' time units which is to be compared with a counterpart aggregated over a time interval of m (where m is a finite integer). The form of the aggregation will depend on whether y_t is a stock or a flow counterpart. Aggregating over a time interval of mh (where m is a finite integer) gives us the mth order temporally aggregated variable. In the former case the aggregated variable, denoted as y_t^* is simply y_t observed at the points $t = m, 2m, 3m$, and so on. In the case of a flow variable,

$$y_t^* = (1 + L + L^2 + \cdots + L^{m-1})y_t,$$

where L is the lag operator.

In either case, the aggregated form of model (4.2.1) is given by

$$y_t^* = \rho^* y_{t-1}^* + u_t^*, \quad t = m, 2m, \ldots, n_a m, \tag{4.2.2}$$

where u_t^* is a finite-order stationary $ARMA(p, q^*)$ process, $n_a = [n_b/m]$ is the number of temporally aggregated observations (where $[x]$ denotes the integer part of x), and

$$\rho^* = \rho^m. \tag{4.2.3}$$

Amemiya and Wu (1972) and Brewer (1973) show that, for the stock variable case, $q^* = [((p+1)(m-1)+q)/m]$, and for the flow variable case, $q^* = [((p+2)(m-1)+q)/m]$. As $m \to \infty$, then for the stock case, $q^* \to p+1$ for $q \geq p+1$, otherwise $q^* \to p$. IN the flow case, $q^* \to p+2$ for $q \geq p+2$, otherwise $q^* \to p+1$.

Consider testing the unit root hypothesis

$$H_0 : \rho = 1 \quad \text{against} \quad H_1 : \rho = e^{-c/n_b}, \tag{4.2.4}$$

using some appropriate statistic, t. H_1 represents a one-sided local alternative to the null hypothesis for $c > 0$.[3]

Let the statistic calculated using n_b observations of the basic sampling frequency be denoted by t_b^n and that using n_a observations of the aggregated sampling frequency by t_a^n. Now consider a sequence of tests for increasing values of s, $s = 1, 2, \ldots$, where

$$n_a = s \quad \text{and} \quad n_b = sm. \tag{4.2.5}$$

Note that s is the time span measured in temporally aggregated time units. We now prove that the asymptotic local power of a class of unit root tests is independent of the degree of temporal aggregation.

---

[2]It is possible to derive (4.2.1) from an underlying model formulated in continuous time by integrating over some time interval (see Bergstrom (1984) for details). Whether the underlying economic decision process being modelled is more appropriately viewed as discrete or continuous is an open question and does not affect the validity of the discrete representation (4.2.1).

[3]In the terminology of Phillips (1987b), y_t is said to be near integrated under H_1.

Proposition 1. *Any test of (4.2.4) that is asymptotically independent of nuisance parameters under both H_0 and H_1 has a limiting distribution under both null and local alternative that is independent of the frequency of sampling, m.*

Proof: The null[4] and alternative for the temporally aggregated model are

$$H_0 : \rho^* = 1 \quad \text{against} \quad H_1 : \rho^*(= \rho^m) = e^{-c/n_a}\,. \tag{4.2.6}$$

t_b^n and t_a^n are similar tests so that under a common null they have the same limiting distribution. On the alternative hypotheses H_1 in (4.2.4) and (4.2.6), consider a sequence of tests for increasing values of s where n_a and n_b are defined y (4.2.5). Since by assumption the statistics t_b^n and t_a^n are asymptotically independent of the parameters of their respective error processes, it follows that as $s \to \infty$, t_b^n and t_a^n have the same limiting distribution and therefore the same asymptotic local power. **Q.E.D.**

We note that the conditions of Proposition 1 are weak and are met by most of the various different tests for unit roots[5] in the literature such as those proposed by Dickey and Fuller (1979, 1981), Phillips (1987a) and Phillips and Perron (1988). In practice, provided that data spans are long enough, Proposition 1 implies that the power of a unit root test against alternatives close to unity is going to be largely unaffected by the frequency of observation. The intuition behind the result is that the power loss from discarding $((m-1)/m)$ths of the data points is made up by the increased separation of H_0 from H_1 resulting from temporal aggregation.

4.3 Power against fixed alternatives

The previous section looked at power against a sequence of local alternatives. When we turn to consider fixed alternatives then in order to maintain the same power with a temporally aggregated model, the span of the data has to increase. Here an approach of Geweke (1981) is followed to derive a large-sample approximation to the required increase in span. First some definitions are needed.

Let t_b be a statistic based on the basic data $y_t, t = 1, 2h, \ldots, n_b$ and let t_a be a statistic based on the temporally aggregated data $y_t, t = m, 2m, \ldots, n_a m$. Let t_i^* be a critical value such that H_0 is rejected in favour of H_1 if $t_i > t_i^*$, $(i = a, b)$, and let $\beta(n_i, t^*)$, be the Type II error associated with a common critical value, t^* with sample size n_i. Finally, let $\overline{n}_i$ be the smallest integer for which $\beta(\overline{n}_i, t^*) < \bar{\beta}$, for some Type II error $\bar{\beta} \in [0, 1]$.

Assumption 1. *Let t_b and t_a be statistics with identical asymptotic distributions under H_0 and let them both be consistent under H_1, such that, for some appropriate power, $k > 0$,*

$$t_i/(n_i)^k \xrightarrow{\text{a.s.}} c_i < \infty \quad as \quad n_i \to \infty.$$

[4] It is clear that, for *even* values of m, H_0 in (4.2.6) is also consistent with the hypothesis that $\rho = -1$. There is thus, in principle, an identification problem here. In practice, however, the possibility of negative unit roots in economic time series is virtually never entertained, so that it is reasonable to restrict consideration to $\rho \geq 0$.

[5] This still applies when the tests are being used on residuals from a cointegrating regression to test a null of no cointegration. A proof of this is available from the authors on request.

Proposition 2. *Under the conditions of Assumption 3:*

$$\lim\{(\bar{n}_b/\bar{n}_a)^k\} = c_a/c_b\,.$$

This result is a straightforward extension of (Geweke, 1981, p. 1431) to which the reader is referred for a proof.

In Proposition 2 as t^* tends to infinity, the sample sizes n_b and n_a must also tend to infinity because the tests are consistent. Both tests have the same asymptotic distribution under H_0 so the ratio in the proposition may be interpreted as being evaluated at some constant power $1-\overline{\beta}$ for a common asymptotic test size of zero. The conditions for the proposition are weak and are satisfied by the commonly used tests.

To see how Proposition 2 may be used to assess the power-preserving span increase referred to above, take an example. Suppose that u_t in (4.2.1) is white noise. Let $\widehat{\rho}$ be the OLS estimate of ρ in (4.2.1) and $\widehat{\rho}^*$ the OLS estimate of ρ^* in (4.2.2) and consider the test statistics $\widehat{t_b} = n_b(\widehat{\rho}-1)$ and $\widehat{t_a} = n_a(\widehat{\rho}^*-1)$. For this case $k=1$ and under the H_1 of $\rho<1$, the almost sure limits of t_b/n_b and t_a/n_b are $\rho-1$ and ρ^m-1, respectively. The ratio of sample sizes (n_a to n_b) required to maintain power at the arbitrary level $1-\overline{\beta}$ for a critical value t^* is, therefore, $(1-\rho)/(1-\rho^m)$ and the corresponding relative span is simply $m(1-\rho)/(1-\rho^m) = m/\sum i = 0^{m-1}\rho^i$. For example, if quarterly and annual data are being compared and if at quarterly frequency $\rho = 0.9$, then the proposition indicates that a 16.3% increase in span is required so that 40 quarterly observations should give the same power as 12 annual ones. If the Dickey-Fuller t-ratio is used, then the formula for the ratio of spans is $\{m(1-\rho)(1+\rho^m)/((1+\rho)(1-\rho^m))\}^{0.5}$ which gives a required increase in span for $\rho = 0.9$ of 0.7%.

More complicated cases are the Z_α and the ADF t-ratio tests when y_t is an $ARMA(1,1)$ flow variable given by

$$y_t = \rho y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}.$$

In this case the process remains $ARMA(1,1)$ after temporal aggregation. For the Z_α test of Phillips (1987a), $k=1$ and

$$\operatorname{plim}\left(\frac{t_b}{n_b}\right) \equiv \operatorname{plim}\left(\frac{Z_\alpha}{n_b}\right) = -\left\{1-\rho-\frac{\theta(1-\rho^2)}{(1+\theta^2+2\theta\rho)}\right\}^2 \times \left\{\frac{1+\theta^2+2\theta\rho+\theta(1+\rho^2)}{(1+\theta^2+2\theta\rho)(1-\rho)}\right\}.$$

Replacing θ and ρ in this expression with θ^* and ρ^* from the aggregated model gives $\operatorname{plim}(t_a/n_a)$, and the power-preserving span ratio is simply $m\{\operatorname{plim}(t_b/n_b)/\operatorname{plim}(t_a/n_a)\}$. For $m=2$, $\rho=0.9$, and $\theta=0.25$, the required span increase is 3.6%.

In the case of the ADF t-test, the MA error is approximated by the addition of lagged differences to the regression. For low values of θ, an ADF of order one may be sufficient to approximate the process. For this case $k=\frac{1}{2}$ and

$$\operatorname{plim}\left(\frac{t_b}{\sqrt{n_b}}\right) = \frac{-(1+\theta^2+\theta\rho)}{\sqrt{2(1+\theta+\theta\rho+\theta^2)/(1-\rho)}}.$$

Again, replacing θ and ρ with θ^* and ρ^* from the aggregated model gives a corresponding form for $\operatorname{plim}(t_a/\sqrt{n_a})$, and the power-preserving span ratio is $\sqrt{m}\{\operatorname{plim}(t_b/\sqrt{n_b})/\operatorname{plim}(t_a/\sqrt{n_a})\}$. For $m=2$, $\rho=0.9$, and $\theta=0.25$, the formula predicts a required span increase of 3.3%. If $\theta=0.6$, then this figure rises to 17.6%.

Proposition 2 provides a measure of 'efficiency' in terms of the relative power under H_1. This must be distinguished from the more usual notion of efficiency which is in terms of the relative variances of parameter estimates. The latter has been examined by Palm and Nijman (1994) who find that the increase in span required to achieve equal efficiency of parameter estimates for the AR(1) model with $m = 2$ is $2\rho^2/(1+\rho^2)$. This compares with our measure of power-preserving span for the $n(\widehat{\rho}-1)$ statistic of $2/(1+\rho)$ which is strictly larger. Our measure will of course vary according to the test statistic used.[6]

As well as having a measure of relative power of basic and aggregated samples under H_1, it is also interesting to gauge absolute power in each case. The power, P, of the unit root test of sample size n is given by

$$P(\gamma, n, t_n^*) = \int_{-\infty}^{t_n^*} f_n(x; \gamma, n)\, dx \tag{4.3.1}$$

where f_n is the pdf of the test statistic, γ is a vector of nuisance parameters and t_n^* is the critical value for sample size n. The pdf in (4.3.1) is in general unknown but a change of variables is usually possible such that

$$P(\gamma, n, t_n^*) = \int_{-\infty}^{g(t_n^*, \gamma, n)} f_n^*(x; \gamma, n)\, dx \tag{4.3.2}$$

where the sequence of functions $f_n^*(x; \gamma, n)$ converges to the asymptotic distribution $f(x)$ independent of γ and n as $n \to \infty$. An approximation to (4.3.2) is, therefore,

$$P(\gamma, n, t_n^*) \simeq \int_{-\infty}^{g(t_n^*, \gamma, n)} f(x)\, dx\,. \tag{4.3.3}$$

To illustrate, consider the Dickey-Fuller t-ratio test for the simplest case of (4.2.1) where u_t is white noise. In this case, $f(x)$ is the standard normal distribution function and

$$g(t_n^*, \gamma, n) = t_n^* - \frac{\sqrt{n}(\rho - 1)}{\sqrt{1-\rho^2}}. \tag{4.3.4}$$

In the same model, the test statistic based on $n(\hat{\rho}-1)$ also has an asymptotic standard nortmal distribution, but here

$$g(t_n^*, \gamma, n) = \frac{t_n^*}{\sqrt{n(1-\rho^2)}} + \frac{\sqrt{n}(1-\rho)}{\sqrt{1-\rho^2}}. \tag{4.3.5}$$

Finally, consider the Augmented Dickey-Fuller t-ratio test for (4.2.1), where u_t is now ARMA(1,0) so that y_t is ARMA(2,0). Again $f(x)$ is standard normal with

$$g(t_n^*, \gamma, n) = t_n^* + \frac{\sqrt{n}(1-\rho_1-\rho_2)\sqrt{1+\rho_1-\rho_2}}{\sqrt{2\{(1-\rho_1^2-\rho_2^2)(1-\rho_2) - 2\rho_1^2\rho_2\}}}. \tag{4.3.6}$$

[The critical values t_n^* for all these tests are tabulated in (Fuller, 1976, Tables 8.5.1 and 8.5.2).] It should be clear that both the relative power measure in Proposition 2 and the approximation to the power function in (4.3.3) can be computed and in most practical circumstances given values for γ and ρ.

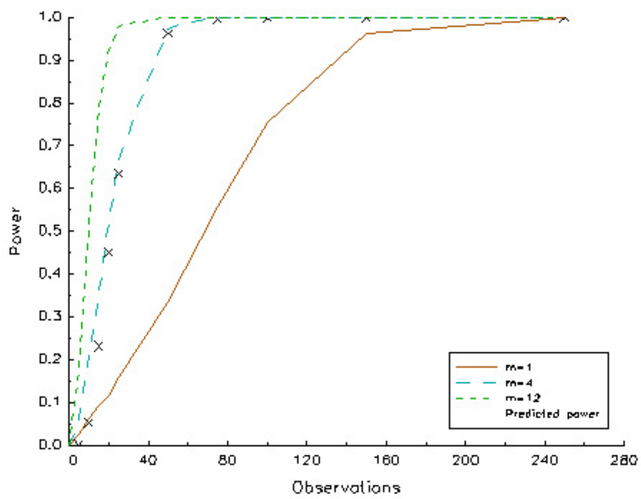

Figure 4.1: Power functions: ARMA(1,0) stock case, $n(\rho-1)$ test, $\rho = 0.9$.

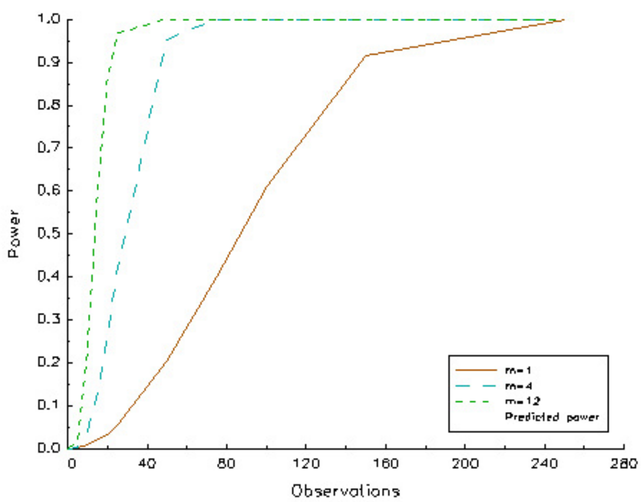

Figure 4.2: Power functions: ARMA(1,1) flow case, Z_α test, $\rho = 0.9$, $\theta = 0.25$.

4.4 Monte Carlo simulation

Both Propositions 1 and 2 and the approximations in (4.3.3) of Section 4.3 are based on asymptotic results. To assess how useful these are in finite samples, Monte Carlo simulations were conducted for a number of simple models for y_t and illustrating the use of different unit root tests. Simulations are reported for three models of the y_t process:[7] the ARMA(1,0) stock variable case and two ARMA(1,1) flow models, one with a low value of the moving average parameter ($\theta = 0.25$) and one with a high value ($\theta = 0.6$).

For each model a different test statistic is illustrated: the Dickey-Fuller $n(\rho-1)$ test for model 1, the Phillips (1987a) Z_α test for model 2, and the ADF t-ratio test for model

[6]In the case of estimation ML is asymptotically efficient, and so is a natural choice for an estimator. In the context of test against distinct alternatives there is no such natural candidate.

[7] These were selected as being representative of a large number of simulations that were run, all programmed using the GAUSS econometric programming language.

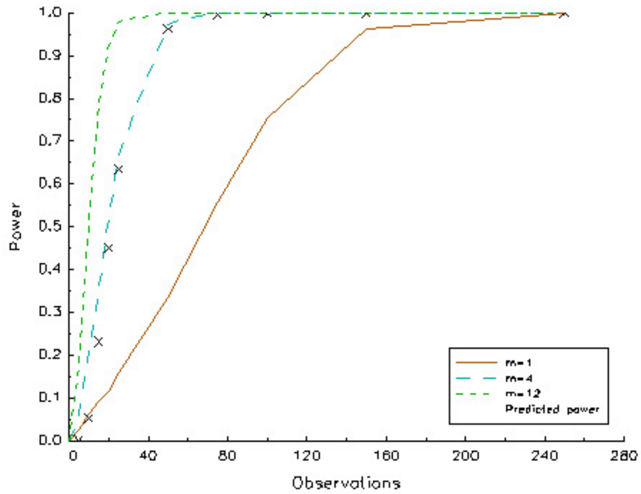

Figure 4.3: Power functions: ARMA(1,1) flow case, ADF test. $\rho = 0.9$, $\theta = 0.6$.

3. the ADF t-ratio test was computed following the procedure recommended by Said and Dickey (1984), where the number of lagged differences included in the regression is increased with the number of observations, n, but at a slower rate [a rate of order$O(n^{1/4})$ was used]. The Z_α test was implemented using a Parzan lag window[8] for the weights on the correction factor, with the lag length increasing with n, again at rate $O(n^{1/4})$.

The results, based on 5000 replications, are presented in Figs. 1-3, where power curves are graphed for values of $m = 1, 4, 12$, where $\rho = 0.9$ in the 'basic ' ($m = 1$) model. These curves illustrate the gain in power of moving from quarterly ($m = 1$) to annual ($m = 4$) or from monthly ($m = 1$) to annual ($m = 12$) data. the dotted lines in the figures plot the power for $m = 2$ against the number of observations scaled by the power-preserving span ratio for the test statistic as derived in Section 4.3. In Proposition 2 were to hold *exactly* in finte samples, then this dotted line would lie exactly on top of the $m = 1$ curve. The closeness of the dotted lines to the $m = 1$ curves in all three figures thus shows how remarkably well Proposition 2 holds for all our models, even in very small samples.

Finally, the power approximation to the power function based on (4.3.3) was computed for the $m = 4$ curves for the Dickey-Fuller $n(\rho - 1)$ test and the ADF t-ratio test, and is plotted by the crosses in Figs. 1 and 3. For the DF test in Fig. 1 it can be seen that the approximation, using Eq. (4.3.5), while it slightly underestimates power at very small sample sizes, holds very well at all other sample sizes. For the ADF t-ratio test in Fig. 3 the approximations were computed as in Section 4.3, using (4.3.3) and the formula

$$g(t_n^*, \gamma, n) = t_n^* + \sqrt{n} z_n,$$

where z_n is the plim of the estimated AR($O(n^{1/4})$) process under the truer ARMA(1,1) model. It can be seen from the figure that the approximation overestimates power at small sample sizes and underestimates it at large sample sizes. However, taken overall the approximation, while obviously not as good as that for the simpler model, is still not bad, and can serve as a rough guide to finite-sample power.

[8]This was the choice in the simulations reported by Phillips and Perron (1988).

4.5 The cointegration of consumption and wealth

In a recent paper, Molana (1991) extends the intertemporal model of consumer behaviour of Hall (1978) to allow for credit constraints and derives a specification relating (non-durable) consumption to the stock of wealth. However, on the basis of an augmented Dickey-Fuller (*ADF*) test using quarterly observations from 1966(4) to 1981(4), he was unable to reject the null hypothesis that consumption and wealth are not cointegrated. Although quarterly data on wealth are only available from 1966(4), annual data go back to 1957. We repeated Molana's exercise[9] using the 25 annual data points between 1957 and 1981 and our results are compared to his in Table 4.1.

Table 4.1: Tests for the cointegration of non-durable consumption and wealth

Span	Frequency	$CRDW$	ADF	Z_t	$CRDW^c$	ADF^c	Z_t^c
1966-1981	Quarterly	0.83	−1.22	—	0.11	−3.18	−3.15
1966-1981	Annual	0.63	−2.53	−1.99	0.13	−3.47	−3.34
1957-1981	Annual	0.81	−3.56	−3.02	0.17	−3.46	−3.13

$CRDW$ is the cointegrating regression Durbin-Watson statistic, ADF the cointegrating regression augmented Dickey-Fuller statistic, Z_t the Phillips Z_t test statistic and superscript ^{c} denotes the appropriate 5% critical values (obtained by simulation).

The cointegrating regression was $\ln C_t = \alpha + \beta \ln W_t + u_t$ where C_t is real nondurable consumption and W_t is real net household wealth (seasonals were included in the quarterly model). $CRDW$, ADF, and Z_t are the cointegrating regression Durbin-Watson, augmented Dickey-Fuller and Phillips Z_t test statistics respectively, and the superscript ^{c} denotes the appropriate 5% critical value. Critical values were estimated by numerical simulation on the assumption that $\ln C$ and $\ln W$ follow independent random walks with drift given by the average quarterly or annual growth rate of the respective series.

The quarterly roots in the ADF auxiliary regression were not reported in Molana (1991). In the annual data however, where an AR(2) auxiliary regression proved adequate, the largest of the two roots was found to be 0.552. An estimate of the largest root in the quarterly cointegrating regression is therefore given by $0.552^{\frac{1}{4}} = 0.862$ (see for example Amemiya and Wu (1972)). Propositions 1 and 2 indicate that with a root so close to unity, we should suffer little loss in power in moving from quarterly to annual data with the same time span. The first two rows of Table 1 show that despite relatively high $CRDW$ statistics, the ADF and Z_t tests based on the 1966-1981 span and using either frequency of data observation, fail to reject noncointegration by a large margin. Using the formula for the power approximation derived in (4.3.6) in Section 4.3, (consistent) estimates of ρ_1 and ρ_2 from the annual data, and setting $T = 15$ gives a power estimate of 0.08.[10] In the light of this, Molana's own failure to reject is unsurprising. The third row shows the gain from increasing the span; the ADF now rejects noncointegration and although Z_t remains insignificant, it only does so by the smallest of margins. Finally, updating

[9] In Molana's work the use of quarterly data was necessary to explore dynamics etc. However, this does not preclude the use of annual data to establish cointegration as the first stage of a two-step procedure.

[10] The corresponding figure for the 1957-1981 span is 0.26, a threefold increase.

the span to 1987 gives highly significant $CRDW$, ADF and Z_t values of 0.78, −4.09 and −3.31, respectively. We may conclude therefore that consumption and wealth in the UK are indeed cointegrated and that Molana's failure to find a cointegrating relationship can be explained by the low power of unit root tests when applied to data of a short time span.

Part III

Sectoral Disaggregation

Chapter 5

Econometric Analysis of Aggregation in the Context of Linear Prediction Models

This paper deals with the problem of aggregation where the focus of the analysis is whether to predict aggregate variables using macro or micro equations. The Grunfeld-Griliches prediction criterion for choosing between aggregate and disaggregate equations is generalised to allow for contemporaneous covariances between the disturbances of micro equations and the possibility of different parametric restrictions on the equations of the disaggregate model. A new test is proposed of the hypothesis of 'perfect aggregation' which tests the validity of aggregation either through coefficient equality or through the stability over time of the composition of the regressors across the micro units. The tools developed in the paper are then applied to employment demand functions for the UK economy disaggregated by 40 industries. Firstly a set of unrestricted log-linear dynamic specifications are estimated for the disaggregate equations and then linear parameter restrictions are imposed as appropriate. Corresponding unrestricted and restricted aggregate equations are estimated. Two different levels of aggregation are considered: aggregation over the 23 manufacturing industries and aggregation over all 40 industries of the economy. In both cases the hypothesis of perfect aggregation is firmly rejected. For the manufacturing industries the prediction criterion marginally favors the aggregate equation but over all industries the disaggregated equations are strongly preferred.

Key Words: Aggregation; Linear prediction models, Labour demand.

5.1 Introduction

The problem of aggregation over micro units has been approached in the empirical literature from a number of different viewpoints. In the case of linear models one important

[0] Published in *Econometrica* (1989), Vol. 57, pp. 861–888. Co-authors M. H. Pesaran and M. S. Kumar. This is a substantially revised version of the paper 'On the Problem of Aggregation in Econometrics', presented at the European Meeting of the Econometric Society, Budapest, 1986. The authors are grateful to Angus Deaton, Arnold Zellner, Ron Smith, Clive Granger, and the referees for their helpful comments. Partial support from the ESRC is gratefully acknowledged.

issue addressed in this literature is the problem of 'aggregation bias', defined by the deviation of the macro parameters from the average of the corresponding micro parameters. (See, for example, Theil (1954), Boot and de Wit (1960), Orcutt et al. (1968), Edwards. and Orcutt (1969), Barker (1970), Gupta (1971), Sasaki (1978), and Winters (1980)).[1] Another closely related issue is the prediction problem originally discussed by Grunfeld and Griliches (1960), where the focus of the analysis is whether to predict aggregate variables using macro or micro equations. Our primary concern in this paper is with the prediction problem in the context of linear models. We present a generalization of the Grunfeld-Griliches (GG) prediction criterion which allows for contemporaneous covariances between the disturbances of the micro equations, and the possibility of different linear parametric restrictions on the equations of the disaggregate model. We also develop a formal statistical test of the hypothesis of 'perfect aggregation' which, unlike the test proposed by Zellner (1962) in the context of the seemingly unrelated regression model, does not necessitate the requirement that all coefficients across the equations of the disaggregated model be the same. The proposed test allows for the possibility of valid aggregation either through coefficient equality or through the invariance of the composition of the regressors across the micro units over time. The choice criterion and the test of perfect aggregation developed in the paper are then applied to two alternative specifications of employment functions for the UK economy disaggregated by 40 industries, and for the manufacturing sector disaggregated by 23 industries. As far as the choice criterion is concerned, the empirical results show that for the economy as a whole the disaggregate model fits better than the aggregate specification, while the reverse is true for the manufacturing industries taken as a group. The slightly better fit obtained for the aggregate model in the case of the manufacturing industries should not, however, be taken to mean that there are no aggregation problems at this level. In fact the application of the test of perfect aggregation to the employment functions provides strong evidence in favor of rejecting the hypothesis of perfect aggregation both for the economy as a whole, and for the manufacturing sector. Our results also suggest serious upward bias in the estimates of output and real wage elasticities of aggregate employment demand obtained for the UK in the literature using aggregate relations. The slightly better within-sample performance of the aggregate specification in the case of the manufacturing industries is best interpreted as an indication of the misspecification of the disaggregate equations. The plan of the paper is as follows. Section 5.2 sets out the basic econometric framework. Section 5.3 examines the small sample bias of the GG prediction criterion. Section 5.4 generalises the basic model so that different specifications for the micro equations are possible, and derives a goodness-of-fit criterion for discrimination between aggregate and disaggregate models that does not suffer from the small sample problem. Section 5.5 considers alternative methods of testing for the errors of aggregation, and develops a new test of the hypothesis of perfect aggregation. Section 5.6 deals with the problem of misspecification of the disaggregate model and the implications that this has for the use of the proposed choice criterion. Section 5.7 contains a detailed application of the econometric methods developed in the paper to the UK employment functions.

[1] On the problem of aggregation across nonlinear micro equations see, for example, Ando (1971), Kelejian (1980), Stoker (1984, 1986a), and the references cited therein.

5.2 The Basic Econometric Framework

We start with the micro model analyzed by Theil (1954), and subsequently by Grunfeld and Griliches (1960), and others, and suppose that the n observations of the m micro units $\{y_i,\ i = 1, 2, \ldots, m;\ t = 1, 2, \ldots, n\}$ are generated according to the following linear specifications:

$$y_{it} = \sum_{j=1}^{k} \beta_{ij} x_{i,jt} + u_{it} \quad (i = 1, 2, \ldots, m;\ t = 1, 2, \ldots, n),$$

or in matrix notations (Kloek (1961))

$$H_d: \quad \underset{n\times 1}{\mathbf{y}_i} = \underset{n\times k}{\mathbf{X}_i}\ \underset{k\times 1}{\boldsymbol{\beta}_i} + \underset{n\times 1}{\mathbf{u}_i} \tag{5.2.1}$$

In the above specification it is assumed that the variations in dependent variables of all micro units can be explained by means of linear combination of the same set of k explanatory variables. This assumption will be relaxed in the next section. Writing (5.2.1) as a system of seemingly unrelated equations (SURE), following Zellner (1962) we have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{5.2.2}$$

where $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_m')'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_m')'$, $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_m')'$, and $\mathbf{X}$ is an $mn \times mk$ block-diagonal matrix of full column rank with matrix $\mathbf{X}_i$ as its ith block. We also make the following assumption:

Assumption 2. *The $mn \times 1$ disturbance vector $\mathbf{u}$ is distributed independently of $\mathbf{X}$, has mean zero and the variance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$, where $\boldsymbol{\Sigma} = (\sigma_{ij})$, and $\mathbf{I}_n$ is the identity matrix of order n.*

The problem of aggregation can arise when an investigator interested in the behaviour of the macro variable $\mathbf{y}_a = \sum_{i=1}^{m} \mathbf{y}_i$, considers the single macro equation

$$H_a: \quad \underset{n\times 1}{\mathbf{y}_a} = \underset{n\times k}{\mathbf{X}_a}\ \underset{k\times 1}{\mathbf{b}} + \underset{n\times 1}{\boldsymbol{\nu}_a} \tag{5.2.3}$$

where $\mathbf{X}_a = \sum_{i=1}^{m} \mathbf{X}_i$, instead of the m micro equations in (5.2.1). Following Grunfeld and Griliches (1960) we examine the question of whether to predict $\mathbf{y}_a$ using the macro equation (5.2.3), or the micro equations (5.2.1).

5.3 The Small Sample Bias of the Grunfeld-Griliches Criterion

The GG prediction (or more accurately the within-sample goodness-of-fit) criterion for the discrimination between the disaggregate model, H_d, and the aggregate model, H_a can be written as:

$$\text{Choose } H_d \text{ if } \mathbf{A}_d'\mathbf{e}_d < \mathbf{e}_a'\mathbf{e}_a, \text{ otherwise choose } H_a,$$

where $\mathbf{e}_d$ and $\mathbf{e}_a$ are the estimates of the errors in predicting $\mathbf{y}_a$ under H_d and H_a respectively. The estimates employed by GG for $\mathbf{e}_d$ and $\mathbf{e}_a$ are based on the ordinary least squares (OLS) method and are given by

$$\mathbf{e}_a = \mathbf{M}_a\mathbf{y}_a, \quad \mathbf{M}_a = \mathbf{I}_n - \mathbf{X}_a(\mathbf{X}_a'\mathbf{X}_a)^{-1}\mathbf{X}_a' = \mathbf{I}_n - \mathbf{A}_a, \tag{5.3.1}$$

and

$$\mathbf{e}_d = \sum_{i=1}^{m} \mathbf{M}_i \mathbf{y}_i, \quad \mathbf{M}_i = \mathbf{I}_n - \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i' = \mathbf{I}_n - \mathbf{A}_i. \tag{5.3.2}$$

It is important to note that in general $\mathbf{e}_d$ is not an efficient estimator of $\mathbf{u}_d = \mathbf{y}_a - \sum_{i=1}^{m} \mathbf{X}_i \boldsymbol{\beta}_i$, unless the disturbances of the micro equations are contemporaneously uncorrelated (i.e. $\sigma_{ij} = 0$, for $i \neq j$), or when $\mathbf{X}_i$ can be written as exact linear functions of $\mathbf{X}_a$. The problem of efficient estimation of $\boldsymbol{\beta}_i$, and hence $\mathbf{u}_d$, and the effect that this has for the GG criterion will be discussed later. For the moment we assume that the GG criterion, as specified above, is applied even in the case where the micro equation disturbances are contemporaneously correlated, and investigate the small sample bias that such a procedure entails. Like the justification offered for Theil's $\overline{R}^2$ criterion, the rationale behind the use of the GG criterion must lie in the fact that if the micro equations are correctly specified, then 'on average' the fit of $\mathbf{y}_a$ from the macro equation should not be any better than that obtained from the micro equations. That is we should have

$$\mathrm{E}_d(\mathbf{e}_d'\mathbf{e}_d) \leq \mathrm{E}_d(\mathbf{e}_a'\mathbf{e}_a), \tag{5.3.3}$$

where $\mathrm{E}_d(\cdot)$ represents the mathematical expectations operator under H_d. However, using (5.3.1) and (5.3.2) it is easily seen that[2]

$$\mathrm{E}_d(\mathbf{e}_d'\mathbf{e}_d) - \mathrm{E}_d(\mathbf{e}_a'\mathbf{e}_a) = -\,\mathrm{E}(\boldsymbol{\xi}'\mathbf{M}_a\boldsymbol{\xi}) - 2\sum_{s=1}^{k}\sum_{i>j}^{m} \sigma_{ij}\left\{1 - \mathrm{E}\left(\rho_{s,ij}^2\right)\right\},$$

where $\boldsymbol{\xi} = \sum_{i=1}^{m} \mathbf{X}_i\boldsymbol{\beta}_i - \mathbf{X}_a\mathbf{b}$, and $\rho_{s,ij}$ is the sth canonical correlation coefficient between the explanatory variables of the ith and the jth micro equations. Therefore, in general the inequality condition (5.3.3) need not be satisfied even if H_d is correctly specified. There are, however, two circumstances under which the GG criterion satisfies the inequality relationship (5.3.3):

(i) when $\mathbf{X}_i$ can be written as *exact* linear functions of $\mathbf{X}_j$, for all i and j. In this case $\rho_{s,ij}^2 = 1$, and irrespective of the values of σ_{ij} we have $\mathrm{E}_d(\mathbf{e}_d'\mathbf{e}_d) - \mathrm{E}_d(\mathbf{e}_a'\mathbf{e}_a) = -\,\mathrm{E}(\boldsymbol{\xi}'\mathbf{M}_a\boldsymbol{\xi})$.

(ii) when the micro disturbances are all contemporaneously uncorrelated ($\sigma_{ij} = 0, i \neq j$). In general the direction of the bias involved in the use of the GG criterion in small samples depends on the signs of σ_{ij} for $i \neq j$.

The finite sample bias in the use of the GG criterion will not disappear even when $\boldsymbol{\beta}_i$ are estimated efficiently by the SURE method. Consider the simple case where $\boldsymbol{\Sigma}$ is known. The SURE estimator of $\mathbf{u}_d$, which we denote by $\mathbf{e}_s$, will be

$$\mathbf{e}_s = \mathbf{S}(\mathbf{I}_{nm} - \mathbf{A})\mathbf{y}$$

where $\mathbf{S}$ stands for the $n \times nm$ summation matrix

$$\mathbf{S} = \left[\mathbf{I}_n : \mathbf{I}_n : \cdots : \mathbf{I}_n\right], \tag{5.3.4}$$

[2] In deriving this result we have also made use of the relation $k - \mathrm{tr}(\mathbf{A}_i\ A_j) = \sum_{s=1}^{k}(1 - \rho_{s,ij}) \geq 0$. See, for example, (Rao, 1973, pp. 582-587).

and

$$\mathbf{A} = \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}. \tag{5.3.5}$$

Under H_d, $\mathbf{e}_s = \mathbf{S}(\mathbf{I}_{nm} - \mathbf{A})\mathbf{u}$, and hence

$$\mathrm{E}_d(\mathbf{e}_s'\mathbf{e}) - \mathrm{E}_d(\mathbf{e}_a'\mathbf{e}_a) = k\sigma_a^2 - \mathrm{E}(\boldsymbol{\xi}'\mathbf{M}_a\boldsymbol{\xi}) - \mathrm{E}\left\{\mathrm{tr}\left[(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}'\mathbf{S}\mathbf{X}\right]\right\},$$

where $\sigma_a^2 = \sum_{i,j=1}^{m}\sigma_{ij}$. Again leaving the case where $\mathbf{X}_i$ are exact linear functions of $\mathbf{X}_a$ to one side, the strict inequality $\mathrm{E}_d(\mathbf{e}_s'\mathbf{e}_s) \le \mathrm{E}_d(\mathbf{e}_a'\mathbf{e}_a)$ holds only in the special case where $\sigma_{ij} = 0$, for $i \neq j$.

5.4 A Generalised Goodness-of-fit Criterion for Discriminating Between Aggregate and Disaggregate Models

From the results of the previous section it is now a straightforward matter to derive a choice criterion for discrimination between the disaggregate and the aggregate models that does not suffer from the finite sample bias of the GG criterion. But it is first important to extend the econometric framework of Section 5.2, so that different specifications for the micro equations can be considered. Such a generalization is particularly important when the primary purpose of the disaggregation is to achieve a better explanation of the macro variables. Accordingly, we consider the following specifications for the disaggregate and the aggregate models:

$$\begin{aligned}
\widetilde{H}_d: \quad & \underset{n\times 1}{\mathbf{y}_i} = \underset{n\times k_i}{\mathbf{X}_i}\ \underset{k_i\times 1}{\boldsymbol{\beta}_i} + \underset{n\times 1}{\mathbf{u}_i} \quad (i = 1, 2, \ldots, m),\\
\widetilde{H}_a: \quad & \underset{n\times 1}{\mathbf{y}_a} = \underset{n\times k_a}{\mathbf{X}_a}\ \underset{k_a\times 1}{\mathbf{b}} + \underset{n\times 1}{\boldsymbol{\nu}_i}
\end{aligned}$$

where $\mathrm{rank}(\mathbf{X}_i) = k_i$, and $\mathrm{rank}(\mathbf{X}_a) = k_a$. In this formulation there are no restrictions on the number of columns of $\mathbf{X}_i$, or what these columns may represent. The micro equations under H_d can also be viewed as a restricted version of the equations under H_d, with each micro equation having its own specific linear parametric restrictions. In this way a wide range of different specifications across the micro equations can be allowed for. The specification of the macro equation is also generalised so that the investigator can specify a restricted form of the macro equation defined in (5.2.3).

Consider now the following 'adjusted' goodness-of-fit criteria for the aggregate and the disaggregate models:

$$s_a^2 = \mathbf{e}_a'\mathbf{e}_a/(n - k_a), \tag{5.4.1}$$

and

$$s_d^2 = \sum_{i,j=1}^{m}\widehat{\sigma}_{ij}, \tag{5.4.2}$$

where

$$\widehat{\sigma}_{ij} = \{n - k_i - k_j - \mathrm{tr}(\mathbf{A}_i\mathbf{A}_j)\}^{-1}\mathbf{e}_i'\mathbf{e}_j, \tag{5.4.3}$$

with $\mathbf{e}_a$ and $\mathbf{e}_i$ being respectively the OLS residual vectors of the regressions under $\widetilde{H}_a$ and $\widetilde{H}_d$, and $\mathbf{A}_i = \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'$. The use of s_d^2 as a measure of the goodness-of-fit of the disaggregate model is justified on the grounds that it represents an unbiased (and consistent) estimator of $\sigma_a^2 = \mathrm{V}(\sum_{i=1}^m u_{it})$, the population variance of the error of predicting $\mathbf{y}_a$ from the disaggregate model. It is now easily seen that under $\widetilde{H}_d$,

$$\mathrm{E}_d(s_d^2) - \mathrm{E}_d(s_a^2) = -(n-k_a)^{-1}\,\mathrm{E}(\boldsymbol{\xi}'\mathbf{M}_a\boldsymbol{\xi}) \leq 0, \tag{5.4.4}$$

where $\boldsymbol{\xi}$ is now defined by

$$\boldsymbol{\xi} = \sum_{i=1}^m \mathbf{X}_i\boldsymbol{\beta}_i - \mathbf{X}_a\mathbf{b}. \tag{5.4.5}$$

Therefore, as required we have $\mathrm{E}_d(s_d^2) \leq \mathrm{E}_d(s_d^2)$, and unlike the GG criterion, the use of the proposed goodness-of-fit criteria s_a^2 and s_d^2 will 'on average' result in the choice of the disaggregate model in finite samples, assuming, of course, that the disaggregate model is correctly specified. In situations where the disaggregate model fits worse than the aggregate model (i.e. $s_d^2 > s_a^2$) it is likely that the disaggregate model is misspecified. The implications for the above choice criterion when the disaggregate model is subject to errors of specification will be discussed below. Here, for comparison purposes it is worth considering the following decomposition of the s_d^2 criterion:

$$\begin{aligned} s_d^2 &= (n-k_a)^{-1}\mathbf{e}_d'\mathbf{e}_d + (n-k_a)^{-1}\sum_{i=1}^m (k_i - k_a)\widehat{\sigma}_{ii} \\ &\quad + 2(n-k_a)^{-1}\sum_{i>j}^m \{\phi_{ij}/(1-\phi_{ij})\}\,\mathbf{e}_i'\mathbf{e}_j, \end{aligned} \tag{5.4.6}$$

where

$$\mathbf{e}_d = \sum_{i=1}^m \mathbf{e}_i, \quad \text{and}$$

$$\phi_{ij} = (n-k_a)^{-1}\{K_i + k_j - k_a - \mathrm{tr}(\mathbf{A}_i\mathbf{A}_j)\}.$$

The GG prediction criterion focuses on the first term on the right-hand side of (5.4.6) and ignores the asymptotically negligible second and third terms. The second term represents the contribution to the s_d^2 criterion arising out of the possible differences in the number of estimated coefficients between the aggregate and the disaggregate models. The third term in (5.4.6) captures the effect of the contemporaneous correlation amongst the disturbances of the micro equations.

5.5 Tests of Aggregation

In studying the aggregation problem our emphasis so far has been on the model selection procedures. An alternative approach would be to employ classical hypothesis testing procedures and develop a statistical test of the conditions necessary for valid aggregation. In the context of the generalised disaggregate model $\widetilde{H}_d$, the necessary condition for perfect aggregation is given by $\boldsymbol{\xi} = 0$, where $\boldsymbol{\xi}$ is defined in (refeq45). Under the hypothesis of 'perfect aggregation'

$$H_\xi: \quad \boldsymbol{\xi} = \sum_{i=1}^m \mathbf{X}_i\boldsymbol{\beta}_i - \mathbf{X}_a\mathbf{b} = 0,$$

it readily follows from (5.4.4) that $\mathrm{E}_d^2(s_d) = \mathrm{E}_d(s_a^2) = \sigma_a^2$, and as far as the fit of $\mathbf{y}_a$ is concerned we should not expect to gain from disaggregation.[3]

Before developing a formal test of H_ξ, it is important to note that the condition $\boldsymbol{\xi} = \mathbf{0}$ can be given a meaningful interpretation only in the context of the basic model (5.2.1) where $\boldsymbol{\beta}_i$ are of the same dimension and refer to the same type of variables across the micro equations. In this case the condition $\boldsymbol{\xi} = \mathbf{0}$ is clearly satisfied under the 'micro-homogeneity' hypothesis,[4]

$$H_\beta : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_m.$$

This is not, however, the only situation where H_ξ holds. Another hypothesis of interest which yields $\boldsymbol{\xi} = \mathbf{0}$, is the 'compositional stability' hypothesis

$$H_x : \quad \mathbf{X}_i = \mathbf{X}_a \mathbf{C}_i, \quad (i = 1, 2, \dots, m),$$

where $\mathbf{C}_i$ are $k \times k$ nonsingular matrices of fixed constants, such that $\sum_{i=1}^m \mathbf{C}_i = \mathbf{I}_k$. The 'compositional stability' hypothesis represents a set of restrictions on the joint probability distribution of the regressors and states that the composition of the regressors across micro units remain fixed over time. This condition for valid aggregation in linear models has been discussed in the econometric literature by Klein (1953) and Wold and Jurlen (1953). Distributional assumptions on the regressors have also been employed in the literature, for example, by Ando (1971), McFadden and Reid (1975), Kelejian (1980), and more recently by Stoker (1984) to connect the aggregate function to the underlying micro equations in the context of nonlinear models. Under H_x, the macro coefficient vector $\mathbf{b}$, is defined in terms of the micro coefficients through the identity $\mathbf{b} = \sum_{i=1}^m \mathbf{C}_i \boldsymbol{\beta}_i$. The condition $\boldsymbol{\xi} = \mathbf{0}$ will also be met under the mixed hypothesis[5]

$$\begin{aligned} 4H_{\beta x} : \quad & \mathbf{X}_i = \mathbf{X}_a \mathbf{C}_i, \quad (i = 1, \dots, s; s < m), \\ & \boldsymbol{\beta}_{s+1} = \boldsymbol{\beta}_{s+2} = \cdots = \boldsymbol{\beta}_m = \mathbf{b}_1, \end{aligned}$$

where in this case $\mathbf{X}_a = \sum_{i=1}^s \mathbf{X}_i$, $\sum_{i=1}^s \mathbf{C}_i = \mathbf{I}_k$ and $\mathbf{b}_i = \sum_{i=1}^s \mathbf{C}_i \boldsymbol{\beta}$. The test proposed by Zellner (1962) for aggregation bias is a test of the micro homogeneity hypothesis, H_β, and is not necessarily relevant as a test of $H_\xi : \boldsymbol{\xi} = \mathbf{0}$. The Zellner test can therefore be unduly restrictive. Rejection of H_β does not necessarily imply that the perfect aggregation hypothesis H_ξ should also be rejected. What is needed is a direct test of $\boldsymbol{\xi} = \mathbf{0}$. In what follows we develop such a test in the case of the basic disaggregated model (5.2.1) and the aggregate model (5.2.3). Although our results can be extended to the generalised model $\widetilde{H}_d$, we have chosen not to do this here, since we do not think that the perfect aggregation condition $\boldsymbol{\xi} = \mathbf{0}$ can be given a plausible interpretation under $\widetilde{H}_d$. In the case of the generalised model neither the micro homogeneity hypothesis nor the compositional stability hypothesis can be maintained.

[3] For the basic disaggregated model (5.2.1), the hypothesis $H_x i$ is equivalent to the n-covariance condition discussed in Theil (1954) and Lancaster (1966), in the special case where the number of regressors is equal to one.

[4] Notice that this hypothesis cannot hold under the generalised disaggregated model H_d.

[5] The aggregation condition is also met by an alternative mixed hypothesis where the k regressors $\mathbf{X}_i$ can be partitioned into two subsets, one of which satisfies the compositional stability hypothesis and the other has an associated parameter vector satisfying the micro homogeneity hypothesis.

5.5.1 A test of perfect aggregation

To help clarify the nature of the test that we are proposing, we first develop the test in the case where $\boldsymbol{\Sigma}$, the covariance matrix of the micro disturbances, is known. A computationally feasible version of the test will then be discussed. The idea behind the test is straightforward and asks whether the estimator of $\boldsymbol{\xi}$ is significantly different from zero. When $\boldsymbol{\Sigma}$ is known an efficient estimator of $\boldsymbol{\xi}$t is given by

$$\widetilde{\boldsymbol{\xi}} = \S\mathbf{X}\widetilde{\boldsymbol{\beta}} - \mathbf{X}_a\widehat{\mathbf{b}}, \tag{5.5.1}$$

where $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}$ are the SURE and the OLS estimators of the parameters of the disaggregate and the aggregate equations respectively, and $\S$ is the summation matrix defined by (5.3.4). Substituting $\widetilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$, $\widehat{\mathbf{b}} = (\mathbf{X}_a'\mathbf{X}_a)^{-1}\mathbf{X}_a'\mathbf{y}_a$ in (5.5.1) now yields $\boldsymbol{\xi} = \mathbf{H}\mathbf{y}$ where $\mathbf{H} = \S\mathbf{A} - \mathbf{A}_a\S$. The matrices $\mathbf{A}_a$ and $\mathbf{A}$ are already defined by (5.3.1) and (5.3.5), respectively. On the null hypothesis that $\boldsymbol{\xi} = \sum_{i=1}^m \mathbf{X}_i\boldsymbol{\beta}_i - \mathbf{X}_a\mathbf{b} = \mathbf{0}$, we have $\boldsymbol{\xi} = \mathbf{H}\mathbf{u}$. Therefore, under the assumption that $\mathbf{u}$ is normally distributed with zero means and a known nonsingular variance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}\otimes\mathbf{I}_n$,

$$\widetilde{\boldsymbol{\xi}}'(\mathbf{H}\Omega\mathbf{H}')^{-1}\widetilde{\boldsymbol{\xi}} \sim \chi^2_n.$$

A necessary condition for $\mathbf{H}\boldsymbol{\Omega}\mathbf{H}'$ to have a full rank can be obtained in the following manner: since, by assumption $\boldsymbol{\Omega}$ is a nonsingular matrix, then $\mathrm{rank}(\mathbf{H}\boldsymbol{\Omega}H') = \mathrm{rank}(\mathbf{H})$. But,

$$\begin{aligned}
&\mathrm{rank}(\mathbf{H}) \le \mathrm{rank}(\mathbf{S}\mathbf{A}) + \mathrm{rank}(\mathbf{A}_a\mathbf{S}),\\
&\mathrm{rank}(\mathbf{A}_a\mathbf{S}) = \mathrm{rank}(\mathbf{A}_a) = k,\\
&\mathrm{rank}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}) = mk, \quad \mathrm{rank}(\mathbf{S}) = n,
\end{aligned}$$

and

$$\mathrm{rank}(\mathbf{S}\mathbf{A}) \le \min(n, mk).$$

Consequently, $\mathrm{rank}(\mathbf{H}) \le k + \min(n, mk)$, and for matrix $\mathbf{H}$ to have full rank equal to n, it is necessary that $k + \min(n, mk) \ge n$, or

$$k(m+1) \ge n. \tag{5.5.2}$$

This rank condition is clearly satisfied when m is large relative to n/k. But in situations where the number of micro equations is relatively large, the computational burden of obtaining the SURE estimates, $\widetilde{\boldsymbol{\beta}}$ in (5.5.1), can be considerable. One possibility would be to construct a test of H_ξ based on the OLS estimates of ,8 instead of the SURE estimates. The estimate of t based on the OLS estimators is given by

$$\boldsymbol{\xi} = \sum_{i=1}^m \mathbf{X}_i\widetilde{\boldsymbol{\beta}}_i - \mathbf{X}_a\widehat{\mathbf{b}} = \mathbf{e}_a - \mathbf{e}_d,$$

where $\mathbf{e}_a$ and $\mathbf{e}_d$ are already defined by (5.3.1) and (5.3.2), respectively. Under H_d, and on the assumption that the hypothesis of perfect aggregation H_ξ holds, we have

$$\widehat{\boldsymbol{\xi}} = \sum_{i=1}^m (\mathbf{A}_i - \mathbf{A}_a)\mathbf{u}_i = \sum_{i=1}^m bfH_i\mathbf{u}_i. \tag{5.5.3}$$

Now assuming that $\mathbf{u}_i$ are normally distributed, then conditional on $\mathbf{X}_i$, we have

$$m^{-1/2}\widehat{\boldsymbol{\xi}}|\mathbf{X}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}_m),$$

where

$$\boldsymbol{\Psi}_m = m^{-1}\sum_{i,j=1}^{m} \sigma_{ij}\mathbf{H}_i\mathbf{H}_j. \tag{5.5.4}$$

Therefore, assuming that $\boldsymbol{\Psi}_m$ is a nonsingular matrix,[6] we arrive at the result

$$m^{-1}(\mathbf{e}_a - \mathbf{e}_d)'\boldsymbol{\Psi}_m^{-1}(\mathbf{e}_a - \mathbf{e}_d) \sim \chi^2_n, \tag{5.5.5}$$

which is the OLS counterpart of (5.5.1).

When $\boldsymbol{\Sigma} = (\sigma_{ij})$ is unknown, it is still possible to obtain an 'approximate' test of the perfect aggregation hypothesis by replacing σ_{ij} in (5.5.1) or (5.5.5) with their SURE or OLS estimates. Here, we focus on the latter and consider testing H_ξ by means of the statistic

$$a_m = m^{-1}(\mathbf{e}_a - \mathbf{e}_d)'\widehat{\boldsymbol{\Psi}}_m^{-1}(\mathbf{e}_a - \mathbf{e}_d), \tag{5.5.6}$$

where

$$\widehat{\boldsymbol{\Psi}}_m = m^{-1}\sum_{i,j=1}^{m} \widehat{\sigma}_{ij}\mathbf{H}_i\mathbf{H}_j, \tag{5.5.7}$$

$$\widehat{\sigma}_{ij} = \{n - 2k + \mathrm{tr}(\mathbf{A}_i\mathbf{A}_j)\}^{-1} e_i'e_j. \tag{5.5.8}$$

We shall refer to a test of H_ξ based on (5.5.6) as the perfect aggregation test, or the a-test for short.

It seems reasonable to suppose that the distribution of a_m, on the null hypothesis of perfect aggregation will tend towards a χ^2_n as $m \to \infty$, although at this stage we are not able to present a proof.[7]

5.6 Disaggregation and Specification Error

The model selection criterion and the aggregation test developed in this paper are based on the assumption that the disaggregate model is correctly specified. In reality, however, both the disaggregate and the aggregate models may suffer from errors of specification, with the latter also being subject to the additional problem of aggregation error. In such a circumstance the issue of whether disaggregation is useful for the study of macro phenomena and the extent of the gain that may be expected from disaggregation depends very much on the relative importance of the two types of errors of specification and aggregation. In this section the implications that errors of specification may have for the use of our proposed choice criterion will be examined.

Let the correctly specified disaggregate model be

$$\underset{n\times 1}{\mathbf{y}_i} = \underset{n\times k_i}{\mathbf{X}_i}\ \underset{k_i\times 1}{\boldsymbol{\beta}_i} + \underset{n\times s_i}{\mathbf{W}_i}\ \underset{s_i\times 1}{\boldsymbol{\gamma}_i} + \underset{n\times 1}{\mathbf{u}_i} \quad (i = 1, 2, \ldots, m) \tag{5.6.1}$$

[6] Notice that a necessary condition for $\boldsymbol{\Psi}_m$ to be invertible is given by (5.5.2).

[7] A proof of this result for the special case where the disturbances are independently distributed is given in Pesaran and Pierse (1989).

which in a stacked form can also be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}_i\boldsymbol{\gamma} + \mathbf{u}, \tag{5.6.2}$$

where $\mathbf{X}$ is now an $mn \times \widetilde{k}$ ($\widetilde{k} = \sum_{i=1}^{m} k_i$) block diagonal matrix with $\mathbf{X}_i$ as its ith block, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \dots, \boldsymbol{\gamma}_m')'$, and $\mathbf{W}$ is an $mn \times \widetilde{s}$, ($\widetilde{s} = \sum_{i=1}^{m} s_i$) block diagonal matrix with $\mathbf{W}_i$ on its ith block. The other notations are as in relation (5.2.2). Suppose now that a researcher misspecifies this model by omitting the variables in $\mathbf{W}$, and continues to employ the model selection criterion based on s_a^2 and s_d^2, defined by (5.4.1) and (5.4.2) respectively. Clearly, the result $\mathrm{E}_d(s_d^2) \leq \mathrm{E}_d(s_a^2)$, which provided the rationale for the choice criterion, need no longer hold.

Stacking the OLS residuals $\mathbf{e}_i = \mathbf{M}_i\mathbf{y}_i$ in the vector $\mathbf{e} = (\mathbf{e}_1', \mathbf{e}_2', \dots, \mathbf{e}_m')'$, s_d^2 can also be written as $s_d^2 = \mathbf{e}'\mathbf{L}\mathbf{e}$ in which $= (\boldsymbol{\Lambda} \otimes \mathbf{I}_n)$, and $\boldsymbol{\Lambda}$ is an $m \times m$ matrix with a typical element equal to $[\mathrm{tr}(\mathbf{M}_i\mathbf{M}_j)]^{-1}$. Now under the correctly specified model (5.6.2),

$$\begin{aligned} \mathbf{e} &= \mathbf{M}\mathbf{y}, \quad \mathbf{M} = \mathbf{I}_{mn} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \\ &= \mathbf{M}\mathbf{W}\boldsymbol{\gamma} + \mathbf{M}\mathbf{u}. \end{aligned}$$

Hence

$$\mathrm{E}_d(s_d^2|\mathbf{X}, \mathbf{W}) = \sigma_a^2 + \boldsymbol{\gamma}'\mathbf{W}'\mathbf{M}\mathbf{L}\mathbf{M}\mathbf{W}\boldsymbol{\gamma}. \tag{5.6.3}$$

Since in general $\mathbf{L}$ may not be a positive semi-definite matrix, without further information about the nature of the specification error, it will not be possible to say whether misspecification leads to an upward or a downward bias in the application of the choice criterion. Expanding (5.6.3) in terms of the misspecification of the individual micro equations, we have

$$\begin{aligned} \mathrm{E}_d(s_d^2|\mathbf{X}, \mathbf{W}) = \sigma_a^2 + (n - k_a)^{-1} \sum_{i=1}^{m} \mathbf{d}_i'\mathbf{d}_i \\ + 2\sum_{i>j}^{m} \left\{ \mathbf{d}_i'\mathbf{d}_j / \mathrm{tr}(\mathbf{M}_i\mathbf{M}_j) \right\}, \end{aligned} \tag{5.6.3'}$$

where $\mathbf{d}_i = \mathbf{M}_i\mathbf{W}_i\boldsymbol{\gamma}_i$, and $\mathrm{tr}(\mathbf{M}_i\mathbf{M}_j) = n - k_i - k_j + \mathrm{tr}(\mathbf{A}_i\mathbf{A}_j)$. The direction of the bias resulting from misspecification clearly depends on the sign of the cross-equation terms $\mathbf{d}_i'\mathbf{d}_j$, $i \neq j$, and their quantitative importance relative to the equation-specific terms $\mathbf{d}_i'\mathbf{d}_i$. In practice, however, it is reasonable to expect that $\mathrm{E}_d(s_d^2) > \sigma_a^2$.

Now turning to the s_a^2 criterion, under (5.6.1) we obtain

$$\mathrm{E}_d(s_a^2|\mathbf{X}, \mathbf{W}) = \sigma_a^2 + (n - k_a)^{-1}\boldsymbol{\xi}'\mathbf{M}_a\boldsymbol{\xi} \geq \sigma_a^2, \tag{5.6.4}$$

where

$$\boldsymbol{\xi} = \sum_{i=1}^{m} \mathbf{X}_i\boldsymbol{\beta}_i + \sum_{i=1}^{m} \mathbf{W}_i\boldsymbol{\gamma}_i = \boldsymbol{\xi}_a + \boldsymbol{\xi}_s. \tag{5.6.5}$$

Comparing (5.6.3) and (5.6.4) it is clear that in general it is not possible to say whether $\mathrm{E}_d(s_a^2)$ exceeds $\mathrm{E}_d(s_d^2)$. The result depends on the relative importance of the specification error and the aggregation error for the explanation of the macro variable $\mathbf{y}_a$. In their work, Grunfeld and Griliches (1960) consider a special case of some interest where there are micro specification errors that cancel out in the aggregate. In the context of model (5.6.1) this can arise either when there are, for example, errors of measurement in the micro variables that cancel out exactly in the aggregate[8] (i.e. $\boldsymbol{\xi}_s = \sum_{i=1}^{m} \mathbf{W}_i \boldsymbol{\gamma}_i = 0$), or when the micro specification errors involve omission of macro variables already included in the aggregate model,[9] (i.e. $\mathbf{M}_a \boldsymbol{\xi}_s = 0$). In such a case, using (5.6.4), we have

$$\mathrm{E}_d(s_a^2 | \mathbf{X}, \mathbf{W}) = \sigma_a^2 + (n - k_a)^{-1} \boldsymbol{\xi}' \mathbf{M}_a \boldsymbol{\xi}_a,$$

and only aggregation errors ($\boldsymbol{\xi}_a \neq 0$) cause the expectations of s_a^2 to exceed the true error variance of the aggregate model. However, even in this special case it is not possible to say whether it is better to use the aggregate model. The answer still depends on the relative importance of the micro specification errors in the disaggregate model and the aggregation error in the aggregate model for the explanation and prediction of macro behavior. The issue of whether one should choose the aggregate or the disaggregate model cannot be resolved by *a priori* reasoning alone and has to be settled with respect to particular problems and in the context of specific models.

5.7 Applications: Employment Demand Functions in the UK

In this section the methods described in the preceding sections will be applied to the annual estimates of disaggregate and aggregate employment demand functions for the UK economy. Although our emphasis will be on the aggregation problem, it is hoped that the disaggregate results are of some interest in their own right.

Our empirical analysis is based on the Cambridge Growth Project Databank and uses a consistent set of data on man-hours (EH_i), outputs (Y_i), and real product wages (W_i) across 41 industry groups. Details of the data and the sources are given in the Data Appendix (.1). For the employment equation at the industry level we have adopted the following fairly general log-linear dynamic specification:

$$\begin{aligned} LEH_{it} = {} & \beta_{i1}/m + \beta_{i2}(T_t/m) + \beta_{i3} LEH_{i,t-1} + \beta_{i4} LEH_{i,t-2} \\ & + \beta_{i5} LY_{it} + \beta_{i6} LYi,t-1 + \beta_{i7} LW_{it} + \beta_{i8} LW_{i,t-1} \\ & + \beta_{i9}(SLYT_t/m) + \beta_{i,10}(SLYT_{t-1}/m) + u_{it}, \\ & \qquad (i = 1, 2, 3, 5, 6, \ldots, 41; t = 1956, 1957, \ldots, 1984), \end{aligned} \tag{5.7.1}$$

where LEH_{it} is log of man-hours employed in industry i at time t, T_t is time trend ($T_{1980} = 0$), LY_{it} is log of industry i output at time t, LW_{it} is log of average real wage rate per man-hour employed in industry i at time t, and $SLYT_t = \sum_{i=1, i \neq 4}^{41} LY_{it}$. Industry 4 (Mineral oil and natural gas) is excluded from the analysis, on the grounds that output and employment in this industry were negligible before 1975.

The above specification for the employment demand function can be justified theoretically when employment decisions are made at the industry level by cost minimizing firms with identical production functions and the same given demand and factor price expectations. In this framework the inclusion of lagged employment variables can be justified on

---

[8] The problem of measurement errors in a disaggregate model in the special case where $m = k = 2$ is discussed by Aigner and Goldfeld (1974).

[9] It is beyond the scope of the present paper to go into the reasons for the importance of macro variables in the explanation of micro behavior. In general they may arise because individual micro behavioral relations are not independent but are influenced or constrained by outcomes (or expectations of outcomes) of the market as a whole.

the grounds of inertia in revision of expectations, adjustment costs involved in hiring and firing of workers, or aggregation over different labour types. (See, for example, Sargent (1978) and Nickell (1984).) The variable $SLYT_t$, which measures the level of aggregate output (in logs), is a proxy measure intended to capture changes in demand expectations arising from the perceived interdependence of demand in the economy by the firms in the industry.[10] The time trend is included in the specification in order to allow for the effect of neutral technical progress on labour productivity.[11] Ideally, we would have liked to avoid using a simple time trend as a proxy for the trend productivity. But, unfortunately, direct reliable observations on technical change, especially at the industry level are not available.[12] The use of time trends in regression equations with nonstationary variables also poses a number of important econometric problems and, as shown by Nelson and Kang (1983), Mankiw and Shapiro (1985, 1986), and Durlauf and Phillips (1986), can result in biased inferences.[13] In view of these measurement and econometric problems it is not clear how one should proceed to allow for trend changes in labour productivity on employment demand functions.[14] Here, in the absence of direct measures of trend productivity at the industry level we estimate (5.7.1) with a time trend, but also briefly report on the effects of omitting the time trends.[15]

For the aggregate employment function we adopted the following dynamic specification:

$$\begin{aligned} SLET_t = b_1 + b_2T_t + b_3SLET_{t-1} + b_4SLET_{t-2} + b_5SLYT_t \\ + b_6SLYT_{t-1} + b_7SLWT_t + b_8SLWT_{t-1} + u_t \\ (t = 1956, 1957, \ldots, 1984), \end{aligned} \tag{5.7.2}$$

where

$$SLET_t = \sum_{i=1,\, i\neq 4}^{41} LEH_{it}, \quad \text{and} \quad SLWT_t = \sum_{i=1,\, i\neq 4}^{41} LW_{it}.$$

Here we are assuming that the purpose of the study is to explain $SLET_t$, which is the sum of the logarithms of industry employment (in man-hours). This is clearly different from the more usual practice of specifying aggregate employment functions in terms of the

[10] Apart from the aggregate variable $SLYT_t$, the employment function (5.7.1) is similar to the equations estimated by Peterson (1988), as a part of the Cambridge Multisectoral Dynamic Model of the UK economy. (See Barker and Peterson (1988).

[11] Notice that, for the ease of comparison of the aggregate and the disaggregate parameter estimates, the time trend and the aggregate output variable that are common to all the micro equations are specified in the 'average' form. Clearly this has no effect on the overall fit of the equations for a fixed level of disaggregation.

[12] In their work on aggregate employment demand functions, (?, p. 168) use a production function approach to obtain an index of labour-augmenting technical progress as a 'residual'. This approach requires time series data on capital stock and the share of capital which are not readily available at the industry level. Moreover, since their measure of technical progress is constructed using actual employment, including it as a regressor in the employment demand function can lead to biased estimates.

[13] Notice, however, that in the case of the test of perfect aggregation where the test is justified asymptotically for a fixed sample size but with an increasing number of micro units, the inclusion of time trends in the micro equations does not affect the validity of the test.

[14] However, see Harvey et al. (1986) where a stochastic specification (a random walk with a drift) is advanced for trend productivity. In their formulation T_t, is modelled as $T_t = a + T_{t-1} + \epsilon_t$, where ϵ_t is a white-noise process.

[15] The effect of replacing the time trend by other proxies such as distributed lag functions of gross investment as a way of modelling endogenous technical change à la Kaldor Kaldor (1957, 1961) is discussed in Lee et al. (1989).

logarithm of the sum of industry employment. For our purposes the specification (5.7.2) has the advantage that it fits directly within the theoretical framework of the paper, and as is pointed out, for example, by Lovell (1973), it also satisfies the Klein-Nataf consistency conditions. A theoretical analysis of the alternative methods of aggregating micro specifications such as (5.7.1), and an econometric investigation of the relative merits of such aggregation methods, are beyond the scope of the present paper.

5.7.1 Results for the economy as a whole

The estimates of the unrestricted version of the industry demand functions (5.7.1) for the 40 industry groups $1, 2, 3, 5, 6, \ldots, 41$, over the sample period 1956–84 are set out in Table 5.1. The estimates of the standard errors of the regression coefficients are given in the brackets. The Table also includes the adjusted multiple correlation coefficient ($\overline{R}^2$), the equations' standard errors ($\widehat{\sigma}$), the maximised values of the log-likelihood function (LLF), the Durbin-Watson statistic (DW), the Lagrange multiplier statistic for testing against second order residual autocorrelation (χ^2_{SC}).

Table 5.1: Disaggregate Employment Demand Functions (Unrestricted) (1956–1984)

Industry groups	INPT/40	T/40	LY_{it}	$LY_{i,t-1}$	$LEH_{i,t-1}$	$LEH_{i,t-2}$	LW_{it}	$LW_{i,t-1}$	$SLYT_t/40$	$SLYT_{t-1}/40$	$\hat{\sigma}$	$\overline{R}^2$	LLF	DW	$\chi^2_{SC}(2)$
1. Agriculture,	56.4355	0.0264	0.2538	0.1743	0.5162	0.0152	−0.4196	−0.0073	−0.2408	−0.1762	0.0148	0.9981	87.21	1.9633	0.93
Forestry and Fishing	(80.5624)	(80.5624)	(0.1580)	(0.1722)	(0.1926)	(0.2165)	(0.1479)	(0.0899)	(0.1315)	(0.1091)	(0.1358)				
2. Coal Mining	−46.3870	−0.3462	0.2753	−0.4222	1.1160	−0.1770	−0.2095	−0.0670	0.0901	0.0709	0.0159	0.9986	85.11	1.6715	1.07
	(60.2401)	(0.0742)	(0.0362)	(0.0600)	(0.1186)	(0.1076)	(0.0356)	(0.0615)		(0.1188)	(0.1315)				
3. Coke	−334.1719	−1.4903	−0.0190	0.5195	0.1201	−0.1380	−0.3478	0.0845	0.9039	0.2270	3.3499	0.9763	54.48	2.4823	0.05
	(89.0939)	(0.3678)	(0.1169)	(0.2007)	(0.1636)	(0.1091)	(0.0762)	(0.0998)	(0.3324)	(0.3801)					
4. Mineral Oil and	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Natural Gas															
5. Petroleum	−438.0316	−0.7008	0.3845	−0.4844	0.5455	0.1274	−0.2769	−0.0819	0.7001	0.7214	0.0580	0.9136	47.53	1.9463	0.68
Products	(286.8173)	(0.2349)	(0.3395)	(0.3300)	(0.2211)	(0.2427)	(0.1085)	(0.1570)	(0.7530)	(0.7879)					
6. Electricity, etc.	91.2396	−0.2074	0.2053	0 : 2029	1.1513	−0.6629	−0.1465	−0.0588	−0.1838	−0.1144	0.0205	0.9855	77.73	2.0344	0.41
	(69.2323)	(0.2179)	(0.2592)	(0.3127)	(0.2277)	(0.2021)	(0.0937)	(0.1039)	(0.2523)	(0.2579)					
7. Public Gas	−103.6365	−0.4472	−0.1297	0.1423	0.4114	0.1617	−0.2605	0.1188	0.2816	0.3240	0.0329	0.9707	63.97	2.2360	2.99
Supply	(116.3665)	(0.2799)	(0.2341)	(0.2165)	(0.2288)	(0.1842)	(0.0962)	(0.1212)	(0.2945)	(0.2682)					
8. Water Supply	42.9711	0.0492	0.6751	−0.8120	0.8817	−0.0969	−0.3652	0.3937	−0.5947	0.7993	0.0448	0.9149	55.07	1.8170	0.90
	(76.4264)	(0.3303)	(0.5110)	(0.5967)	(0.2256)	(0.1862)	(0.1664)	(0.1806)	(0.4200)	(0.3770)					
9. Minerals and Ores	197.0285	−0.0037	0.2785	0.0091	0.4998	0.1898	−0.1741	0.0437	−0.5639	−0.0421	0.0342	0.9722	62.89	2.0457	0.72
	(119.2942)	(0.1311)	(0.1428)	(0.1517)	(0.2283)	(0.1899)	(0.0814)	(0.0971)	(0.4006)	(0.3775)					
10. Iron and Steel	−349.9418	−0.9245	0.1361	0.0248	0.4620	−0.0029	−0.4526	0.0666	0.9679	0.2003	0.0279	0.9925	68.77	1.9940	0.85
	(110.1168)	(0.3780)	(0.1031)	(0.0976)	(0.2354)	(0.1512)	(0.1087)	(0.1491)	(0.3587)	(0.4190)					
11. Non-Ferrous	−58.6696	−0.4029	0.1387	−0.2285	1.3339	−0.5140	−0.0696	0.0623	0.8235	−0.4050	0.0251	0.9863	71.84	2.2252	1.69
Metals	(38.5798)	(0.2149)	(0.1359)	(0.1414)	(0.1682)	(0.1418)	(0.0549)	(0.0592)	(0.2501)	(0.2939)					
12. Non-Metallic	−389.8347	−0.5395	0.3985	−0.2419	0.5945	0.0854	−0.3173	−0.1961	0.4744	0.4906	0.0179	0.9933	81.72	2.3446	3.62
Mineral Products	(116.1591)	(0.2589)	(0.1690)	(0.1680)	(0.2070)	(0.1785)	(0.1212)	(0.1348)	(0.3233)	(0.3838)					
13. Chemicals and	−159.1132	−0.0882	0.0983	0.1058	0.2440	0.2676	−0.2988	−0.1215	0.3032	0.2011	0.0158	0.9789	85.22	2.2664	5.28
Manmade Fibres	(71.0291)	(0.1744)	(0.1607)	(0.1596)	(0.2123)	(0.1780)	(0.0923)	(0.1146)	(0.3051)	(0.2983)					
14. Metal Goods	−31.2352	−0.2375	0.2866	0.0653	0.6460	−0.0954	−0.1761	0.0345	0.3585	−0.2214	0.0207	0.9858	77.48	2.1717	5.66
	(54.8630)	(0.2189)	(0.1359)	(0.1618)	(0.2287)	(0.1537)	(0.1164)	(0.1288)	(0.3055)	(0.3215)					
15. Mechanical	−131.8896	0.1976	0.5334	−0.2266	0.4505	−0.1080	−0.2575	−0.3748	−0.0908	0.5055	0.0144	0.9912	87.94	2.0471	3.81
Engineering	(56.4251)	(0.1567)	(0.1187)	(0.1038)	(0.2068)	(0.1457)	(0.1193)	(0.1400)	(0.1860)	(0.1974)					
16. Office Machinery,	−397.5477	0.2032	0.2159	−0.0564	1.0066	0.1224	−0.6515	−0.1085	0.3853	−0.0189	0.0303	0.9166	66.37	1.7819	2.05
etc.	(147.5041)	(0.3070)	(0.0947)	(0.1499)	(0.2501)	(0.2832)	(0.1613)	(0.2020)	(0.2938)	(0.3014)					
17. Electrical	11.0264	−0.3711	0.4461	−0.2587	1.0101	−0.2373	−0.4120	0.4282	0.1741	−0.1295	0.0179	0.9678	81.67	2.2173	2.73
Engineering	(35.8630)	(0.2243)	(0.1248)	(0.1853)	(0.2172)	(0.1463)	(0.1355)	(0.1440)	(0.2172)	(0.2073)					
18. Motor Vehicles	−210.7119	−0.2799	0.5063	−0.3633	0.8391	−0.1636	−0.0568	−0.1675	0.5151	0.1640	0.0192	0.9867	79.67	2.3117	2.34
	(61.6410)	(0.1336)	(0.0718)	(0.1230)	(0.1883)	(0.0946)	(0.1000)	(0.1043)	(0.2088)	(0.2397)					
19. Aerospace	200.2420	−0.7488	0.0710	0.0468	0.5991	−0.3688	−0.0085	−0.1422	−0.2225	−0.2250	0.0284	0.9864	68.28	2.0370	1.17
Equipment	(98.6992)	(0.1808)	(0.0821)		(0.2238)	(0.2001)	(0.0921)	(0.0879)	(0.2480)	(0.2600)					
20. Ships and Other	−159.2346	0.2741	0.6650	−0.3609	1.1618	−0.1246	−0.0186	−0.0490	0.6358	−0.5276	0.0302	0.9840	66.47	2.2101	2.48
Vessels	(63.9773)	(0.2704)	(0.1660)	(0.1540)	(0.2077)	(0.2075)	(0.0808)	(0.0921)	(0.2322)	(0.2384)					
21. Other Vehicles	−127.8464	−0.4292	0.2594	0.0648	0.8089	−0.0679	−0.1583	0.0360	0.1202	0.1451	0.0258	0.9969	71.09	1.9994	1.32
	(84.5230)	(0.2006)	(0.1060)	(0.1115)	(0.2445)	(0.2152)	(0.0656)	(0.0669)	(0.1989)	(0.2126)					

(continued)

Table 5.1: Continued

Industry groups	INPT/40	T/40	LY_{it}	$LY_{i,t-1}$	$LEH_{i,t-1}$	$LEH_{i,t-2}$	LW_{it}	$LW_{i,t-1}$	$SLYT_t/40$	$SLYT_{t-1}/40$	$\widehat{\sigma}$	$\overline{R}^2$	LLF	DW	$\chi^2_{SC}(2)$
22. Instrument	−102.7548	−0.2292	−0.0485	−0.1541	0.6941	−0.1319	−0.2264	0.1475	0.5565	0.2893	0.0246	0.9311	72.37	1.7674	0.81
Engineering	(86.1619)	(0.1516)	(0.2108)	(0.1699)	(0.2304)	(0.2037)	(0.1196)	(0.1373)	(0.2681)	(0.3691)					
23. Manufactured	−209.2242	−0.4900	0.7273	0.0656	0.2394	0.2501	−0.1577	−0.0753	0.0340	0.0808	0.0174	0.9816	82.49	1.7024	4.77
Food	(138.5611)	(0.2336)	(0.3465)	(0.2334)	(0.2303)	(0.1746)	(0.0869)	(0.1211)	(0.1695)	(0.1486)					
24. Alcoholic Drinks,	101.6376	−0.6943	0.5685	−0.0005	0.8466	−0.2612	−0.0878	0.0948	−0.2328	−0.2029	0.0282	0.9152	68.45	2.0663	2.53
etc.	(121.3637)	(0.3435)	(0.4859)	(0.4296)	(0.2624)	(0.2854)	(0.1068)	(0.1051)	(0.4551)	(0.4300)					
25. Tobacco	−300.3142	−0.2605	0.7063	−0.2275	0.8345	0.4788	0.0238	−0.0569	−0.6541	0.8452	0.0496	0.8803	52.10	2.3769	5.28
	(126.6923)	(0.4339)	(0.4676)	(0.5169)	(0.2947)	(0.3493)	(0.0730)	(0.0907)	(0.5131)	(0.4385)					
26. Textiles	−127.5754	0.1293	0.4654	−0.1376	0.5979	0.0464	−0.4302	0.0292	0.1772	−0.0316	0.0186	0.9979	80.59	2.2309	2.33
	(62.4817)	(0.5631)	(0.1663)	(0.1515)	(0.2123)	(0.1390)	(0.0970)	(0.1446)	(0.3237)	(0.2984)					
27. Clothing and	−59.4303	−0.1340	0.4585	0.0217	0.5867	−0.0346	−0.4004	0.1067	−0.0342	0.0194	0.0118	0.9981	93.85	2.0674	0.26
Footwear	(26.3298)	(0.2073)	(0.1154)	(0.1572)	(0.2160)	(0.1428)	(0.0833)	(0.1092)	(0.1568)	(0.1683)					
28. Timber and	33.7633	−0.5183	0.2993	−0.0213	0.3362	−0.1005	−0.2491	0.1186	0.1765	0.1864	0.0140	0.9859	88.71	1.9694	1.53
Furniture	(44.1269)	(0.1714)	(0.0895)	(0.1248)	(0.2314)	(0.1190)	(0.0853)	(0.0892)	(0.2467)	(0.2876)					
29. Paper and Board	−17.8043	−0.2271	0.5324	0.3381	0.1236	0.1281	−0.2353	−0.0844	−0.1617	−0.0889	0.0200	0.9921	78.36	2.2984	4.73
	(37.0048)	(0.1687)	(0.1528)	(0.1761)	(0.2513)	(0.1320)	(0.0722)	(0.1120)	(0.3221)	(0.2940)					
30. Books, etc.	106.4095	0.0419	0.3518	−0.0926	1.2912	−0.6039	−0.0640	−0.0395	−0.1222	−0.1749	0.0124	0.9296	92.29	2.2104	3.23
	(41.1675)	(0.0468)	(0.1167)	(0.1381)	(0.2215)	(0.1718)	(0.0611)	(0.0649)	(0.1968)	(0.1937)					
31. Rubber and Plastic	−124.2511	−0.6223	0.1943	−0.0152	0.4938	0.1046	−0.2650	0.2121	0.5905	−0.0211	0.0176	0.9811	82.18	2.1726	7.97
Products	(52.6895)	(0.2588)	(0.2485	(0.2093)	(0.2227)	(0.1469)	(0.1171)	(0.1475)	(0.4566)	(0.3819)					
32. Other Manufactures	202.2515	−0.4204	0.1757	0.1393	0.5916	−0.0512	0.0166	0.1451	0.3999	−0.7045	0.0134	0.9921	90.08	1.9558	0.90
	(80.7245)	(0.1658)	(0.0858)	(0.1148)	(0.1755)	(0.0985)	(0.0939)	(0.0967)	(0.1678)	(0.1813)					
33. Construction	84.3552	−0.0906	0.3618	−0.2488	1.1406	−0.3369	−0.2854	0.3684	0.1373	−0.2078	0.0174	0.9709	82.53	1.5347	1.84
	(48.6187)	(0.0723)	(0.1301)	(0.1453)	(0.1571	(0.1199)	(0.1122)	(0.1105)	(0.1798)	(0.1962)					
34. Distribution, etc.	111.7644	0.5543	0.0254	0.5177	0.6887	−0.1943	−0.2772	−0.1398	−0.0449	−0.5522	0.0142	0.9587	88.44	2.2835	2.14
	(46.6165)	(0.2857)	(0.1910)	(0.2715)	(0.2281)	(0.1634)	(0.1209)	(0.1452)	(0.1750)	(0.2230)					
35. Hotels and	131.717	0.1650	0.2394	0.1746	0.6033	−0.1807	−0.3824	0.2140	−0.0718	−0.1783	0.0209	0.9077	77.22	1.9588	0.30
Catering	(132.1092)	(0.1157)	(0.2469)	(0.3314)	(0.2542)	(0.2362)	(0.1426)	(0.1300)	(0.2039)	(0.1910)					
36. Rail Transport	9.2868	−0.0880	0.0969	0.311	0.8301	−0.0399	−0.0821	0.0953	−0.0469	−0.1113	0.0253	0.9952	71.59	1.9033	3.80
	(135.1609)	(0.1796)	(0.1543)	(0.1906)	(0.2087)	(0.2156)	(0.1427)	(0.1388)	(0.2911)	(0.2803)					
37. Other Land	141.1240	−0.4169	0.0524	0.1751	0.9730	−0.5159	−0.0200	0.0269	0.1628	−0.1715	0.0170	0.9724	83.11	2.4060	5.29
Transport	(74.6539)	(0.1293)	(0.1457)	(0.1818)	(0.2248)	(0.2020)	(0.0599)	(0.0638)	(0.1645)	(0.1545)					
38. Sea, Air, and Other	8.3655	−0.112	0.3135	−0.3582	1.1912	−0.4875	−0.2868	0.1941	−0.1525	0.4478	0.0216	0.9254	76.26	2.2168	1.04
	(132.0896)	(0.1340)	(0.1866)	(0.1944)	(0.1884)	(0.2454)	(0.1370)	(0.1291)	(0.2147)	(0.2436)					
39. Communications	72.2461	−0.4312	0.6876	−0.5043	0.7816	−0.2629	−0.1278	0.2150	0.0045	0.2006	0.0178	0.9392	81.85	2.3222	2.15
	(57.0163)	(0.2596)	(0.2987)	(0.2480)	(0.1689)	(0.1616)	(0.0851)	(0.0879)	(0.2004)	(0.1648)					
40. Business Services	−0.2212	0.0139	0.9929	88.96	.1808	1.79									
	(131.8402)	(0.3506)	(0.1452)	(0.1485)	(0.2703)	(0.2248)	(0.0839)	(0.0783)	(0.1145)	(0.1257)					
41. Miscellaneous	41.4060	0.2004	0.2764	−0.2928	0.8375	0.0023	−0.1728	0.0470	−0.0622	0.0803	0.0240	0.9429	73.10	1.6954	2.48
Services	(241.2260)	(0.3980)	(0.2071)	(0.2149)	(0.2620)	(0.2789)	(0.1434)	(0.1433)	(0.2152)	(0.1956)					

Notes. For source of data see the Appendix (.1). Standard errors in brackets. $\widehat{\sigma}$ is equation standard error, $\overline{R}^2$ is adjusted multiple correlation coefficient, LLF is the maximised value of the log-likelihood function, DW is the Durbin-Watson statistic, and $\chi^2_{SC}(2)$ is the Lagrange multiplier test against second order residual serial correlation.

The results are in general quite satisfactory: the equations fit reasonably well, and the value of $\overline{R}^2$ for the majority of the industries is well above 0.95. Only in the case of the tobacco industry does it fall below 0.90. With the exception of the estimates for industry 31 (Rubber and Plastic Products), the results do not show significant evidence of residual serial correlation. The parameter estimates, when statistically significant, have signs that are *a priori* plausible. The short run elasticities of employment with respect to real wages and output are generally well determined and have the correct signs. The (current) real wage variable is significant at the five percent level in 23 out of the 40 industry groups, and the (current) output variable is significant in 17 of the industries. Notice also that the few incorrectly signed estimates obtained for the real wage and the output variables are not statistically significant, even at the 10 percent level of significance using a one-tailed test. Overall the results provide further evidence in support of the view that both the demand and the product wage variables are significant determinants of changes in employment, although, as is already stressed by Peterson (1988), in the case of most industries changes in demand have been historically more important than changes in product wages in the explanation of employment changes.

Table 5.2: Disaggregate Employment Demand Functions (Restricted) (1956–1984)

Industry groups	INPT/40	T/40	LY_{it}	$LY_{i,t-1}$	$LEH_{i,t-1}$	$LEH_{i,t-2}$	LW_{it}	$LW_{i,t-1}$	$SLYT_t/40$	$SLYT_{t-1}/40$
1. Agriculture,	52.1517	—	0.2687	0.1752	0.5312	—	−0.4211	—	−0.2437	−0.1729
Forestry and Fishing	(64.9478)		(0.1375)	(0.1121)	(0.0589)		(0.0821)		(0.0981)	(0.1088)
2. Coal Mining	41.2000	−0.3502	0.2734	−0.4181	1.1604	−0.2648	−0.2018			
	(14.3416)	(0.0670)	(0.0345)	(0.0589)	(0.0811)	(0.0765)	(0.0296)			
3. Coke	−351.5712	−1.3100	—	0.6330	—	—	−0.3005	—	1.0448	—
	(44.6561)	(0.1752)		(0.1471)			(0.0418)		(0.1564)	
4. Mineral Oil and	—	—	—	—	—	—	—	—	—	—
Natural Gas										
5. Petroleum	−70.7959	−0.5087	0.3640	—	0.5185		−0.3144	—	—	—
Products	(71.7711)	(0.1297)	(0.1324)		(0.1348)		(0.0869)			
6. Electricity, etc.	18.5225	—	0.1614	—	1.2739	−0.5958	−0.1732	—	—	—
	(14.6976)		(0.0798)		(0.1744)	(0.1563)	(0.0687)			
7. Public Gas	−47.1096	−0.6014	—	0.0611	0.4191	—	−0.1507	—	0.5379	—
Supply	(97.2188)	(0.1995)		(0.0659)	(0.1524)		(0.0496)		(0.1827)	
8. Water Supply	8.1676	—	0.6536	−0.6536	0.8112	—	−0.4027	0.4027	−0.6415	0.7906
	(18.9241)		(0.4042)	(0.4042)	(0.0785)		(0.1086)	(0.1086)	(0.3085)	(0.3064)
9. Minerals and Ores	172.9158	—	0.2655	—	0.6931	—	−0.1494	—	−0.5337	—
	(79.1246)		(0.1265)		(0.0790)		(0.0622)		(0.2560)	
10. Iron and Steel	−349.9558	−0.9045	0.1083	—	0.4978	—	−0.3873		1.1803	—
	(58.8686)	(0.2732)	(0.0893)		(0.0832)		(0.0777)		(0.2928)	
11. Non-Ferrous	−84.8257	−0.5749	0.1817	−0.3091	1.2461	−0.4796	−0.0756	0.0756	0.5854	—
Metals	(30.7245)	(0.1517)	(0.1286)	(0.1273)	(0.1458)	(0.1229)	(0.0481)	(0.0481)	(0.1789)	
12. Non-Metallic	−280.5702	−0.3729	0.3101	—	0.6919	—	−0.2356	−0.2214	0.5170	—
Mineral Products	(60.6439)	(0.2148)	(0.1511)		(0.0877)		(0.1075)	(0.0959)	(0.2901)	
13. Chemicals and	−125.0557	—	—	—	0.6205	—	−0.2810	—	0.6049	—
Manmade Fibres	(23.8339)				(0.0693)		(0.0337)		(0.0773)	
14. Metal Goods	−32.2448	−0.1231	0.4365	—	0.5798	—	−0.1671	—	—	—
	(25.5280)	(0.0976)	(0.0444)		(0.0542)		(0.0817)			
15. Mechanical	−149.7049	—	0.4122	−0.1779	0.3215	—	−0.3100	−0.2725	—	0.6080
Engineering	(38.5546)		(0.0584)	(0.0977)	(0.1093)		(0.0868)	(0.1104)		(0.1488)
16. Office Machinery	−3.4674	—	0.1694	−0.1694	1.2748	−0.3244	−0.3884	0.3123	—	—
etc.	(22.7537)		(0.0865)	(0.0865)	(0.2004)	(0.1800)	(0.1379)	(0.1344)		
17. Electrical	2.5709	−0.3785	0.5239	−0.2827	0.9582	−0.1929	−0.4143	0.4027	—	—
Engineering	(32.7774)	(0.2110)	(0.0757)	(0.1276)	(0.1935)	(0.1228)	(0.1259)	(0.1295)		
18. Motor Vehicles	−184.6112	−0.2365	0.4908	−0.3811	0.9237	−0.1783	—	−0.1843	0.5856	—
	(50.0625)	(0.1093)	(0.0629)	(0.1093)	(0.1610)	(0.0897)		(0.0713)	(0.1774)	
19. Aerospace	200.3920	−0.6788	0.0732	—	0.7560	−0.4659	—	−0.1252	—	—
Equipment	(53.1219)	(0.1586)	(0.0654)		(0.1659)	(0.1440)	(0.0674)			
20. Ships and Other	−0.7667	—	0.4809	−0.4809	1.4717	−0.4717	—	—	0.5103	−0.5103
Vessels	(0.3086)		(0.1171)	(0.1171)	(0.1543)	(0.1543)			(0.2000)	(0.2000)
21. Other Vehicles	−132.1537	−0.4754	0.3130	—	0.7270	—	−0.1432	—	—	0.2845
	(54.3892)	(0.1730)	(0.0729)		(0.0884)		(0.0462)			(0.1069)

Table 5.2: Continued

Industry groups	INPT/40	T/40	LY_{it}	$LY_{i,t-1}$	$LEH_{i,t-1}$	$LEH_{i,t-2}$	LW_{it}	$LW_{i,t-1}$	$SLYT_t/40$	$SLYT_{t-1}/40$
22. Instrument	−11.3576	−0.3580	0.3611	—	0.5319	—	−0.2624	—	—	—
Engineering	(44.4947)	(0.1353)	(0.1005)		(0.1253)		(0.1134)			
23. Manufactured	−172.1572	−0.4510	0.6697	—	0.3177	0.2237	−0.1962	—	—	0.1157
Food	(76.0519)	(0.1973)	(0.1734)		(0.1742)	(0.1560)	(0.0645)			(0.1233)
24. Alcoholic Drinks,	−15.1802	−0.4844	0.2933	—	0.7283	—	−0.0945	0.0591	—	—
etc.	(73.4889)	(0.1411)	(0.1167)		(0.1239)		(0.0919)	(0.0882)		
25. Tobacco	−213.3698	−0.3959	0.7424	—	0.7367	0.2633	—	—	—	—
	(80.8449)	(0.1161)	(0.2840)		(0.2225)	(0.2225)				
26. Textiles	−68.1499	—	0.5278	−0.1236	0.5880	—	−0.3428	—	—	—
	(10.0202)		(0.0546)	(0.0754)	(0.0600)		(0.0465)			
27. Clothing and	−68.9489	—	0.4514	—	0.5364	—	−0.3756	—	—	—
Footwear	(11.9600)		(0.0372)		(0.0411)		(0.0284)			
28. Timber and	60.3105	−0.3017	0.3769	—	0.4312	—	−0.2460	0.1493	—	—
Furniture	(20.9479)	(0.0788)	(0.0365)	—	(0.0582)	—	(0.0662)	(0.0740)		
29. Paper and Board	−44.7394	−0.3259	0.4680	0.1585	0.3644	—	−0.2503	—	—	—
	(13.2869)	(0.1040)	(0.0652)	(0.0925)	(0.0842)		(0.0433)			
30. Books, etc.	58.9250	—	0.2973	−0.2575	1.4842	−0.7029	−0.0454	—	—	—
	(20.8186)		(0.0583)	(0.0592)	(0.1686)	(0.1518)	(0.0482)			
31. Rubber and Plastic	−64.4432	−0.3192	0.5398	−0.1401	0.6844	—	−0.1820	—	—	—
Products	(14.2846)	(0.1872)	(0.0588)	(0.0963)	(0.0818)		(0.1007)			
32. Other Manufactures	60.3555	−0.3233	0.2345	—	0.6028	—	—	—	0.4274	−0.4274
	(20.0274)	(0.0653)	(0.0435)		(0.0933)				(0.1287)	(0.1287)
33. Construction	7.2409	—	0.5490	−0.4527	1.0813	−0.2453	−0.4434	0.3376	—	—
	(20.5598)		(0.0828)	(0.0863)	(0.1559)	(0.1135)	(0.0822)	(0.1096)		
34. Distribution, etc.	109.9863	0.3892	—	0.5034	0.5641	—	−0.3036	—	—	−0.5578
	(43.3346)	(0.2057)		(0.1964)	(0.0884)	—	(0.1187)	—	—	(0.1655)
35. Hotels and	−58.7494	—	0.3544	—	0.7096	—	−0.3876	0.1959	—	—
Catering	(44.4425)		(0.1150)		(0.1022)		(0.1191)	(0.1094)		
36. Rail Transport	−65.1073	—	—	0.4070	0.8047	—	−0.0729	—	—	—
	(26.2802)			(0.0979)	(0.0532)		(0.0531)			
37. Other Land	146.4317	−0.4542	—	0.2451	0.9023	−0.4855	—	—	—	—
Transport	(37.8129)	(0.1047)		(0.0701)	(0.1931)	(0.1838)				
38. Sea, Air, and Other	48.5900	−0.1921	0.1924	—	1.1919	−0.5542	−0.0853	—	—	—
	(104.9126)	(0.1054)	(0.1634)		(0.1741)	(0.2189)	(0.0683)			
39. Communications	14.3221	−0.6566	0.9014	−0.4533	0.8261	−0.2785	−0.1686	0.1565	—	—
	(41.3966)	(0.2354)	(0.1808)	(0.1966)	(0.1727)	(0.1579)	(0.0822)	(0.0807)		
40. Business Services	209.6513	—	0.3108	—	0.6781	−0.3104	—	—	—	−0.1633
	(49.1545)		(0.0718)		(0.1759)	(0.1680)				(0.0486)
41. Miscellaneous	−39.9043	—	0.2123	—	0.8264	—	−0.1408	—	—	—
Services	(33.3057)		(0.0790)		(0.0970)		(0.0747)			

For source of data see the Data Appendix (.1). The standard errors are in brackets. The relevant summary and diagnostic statistics are given in Table 5.3.

As far as the time trends are concerned they are significant at the five percent level only in 12 of the industry estimates, and there are no cases where the coefficient of the time trend is positive and statistically significant. In fact omitting the time trend variable from the analysis in general proved to have only a marginal effect on the coefficient estimates and the significance of the real wage and the output variables.[16] The results in Table 5.1 are, however, subject to two important shortcomings: in many cases they seem to be over-parameterised, and the estimates for the industries 16 (Office Machinery, etc.), 20 (Ships and Other Vessels), and 25 (Tobacco) are unstable.[17] To deal with these shortcomings we estimated a restricted version of the industry employment functions by imposing suitable linear restrictions on the coefficients of (5.7.1). The coefficient estimates of this 'restricted' specification and their estimated standard errors are summarised in Table 5.2. The chi-squared statistics for testing the validity of the restrictions together with a number of important diagnostic statistics for tests of misspecification arising from residual serial correlation, functional form, nonnormal errors, and heteroscedasticity are given in Table 5.3. These results are generally more satisfactory than the unrestricted versions. The parameter restrictions cannot be rejected, and only in the case of a very few of the industries do diagnostic statistics indicate that the regression equations are likely to be misspecified.[18] Also note that the restricted estimates for the industries 16, 20, and 25 are no longer unstable, although the equations for the latter two industries are specified in first differences and do not possess long run solutions. The long run elasticities of employment with respect to output and real wages for the 38 industries that do have long run solutions are displayed graphically in Figures 5.1 and 5.2, respectively. Although

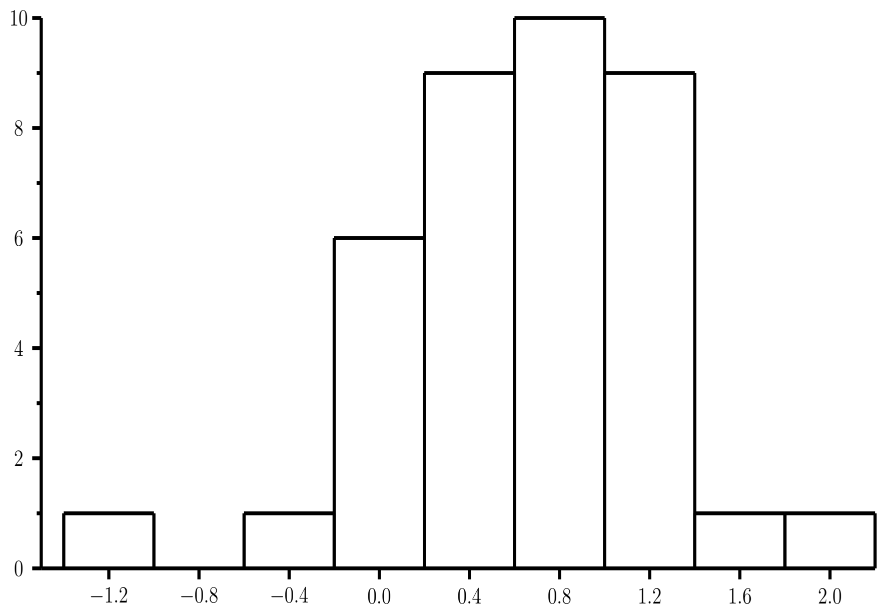

Figure 5.1: Histogram of the long run elasticities of employment with respect to output in different industries.

[16] The effects of omitting the time trend variable on the coefficient estimates were particularly marked only in the case of industries 2, 3, 5, 10, 24, 32, and 37.

[17] The autoregressive parts of the regressions for these three industries have unstable roots.

[18] The results in Table 5.2 are also of some interest insofar as they show evidence of significant aggregate output effects on employment demand at the industry level. See footnote 10.

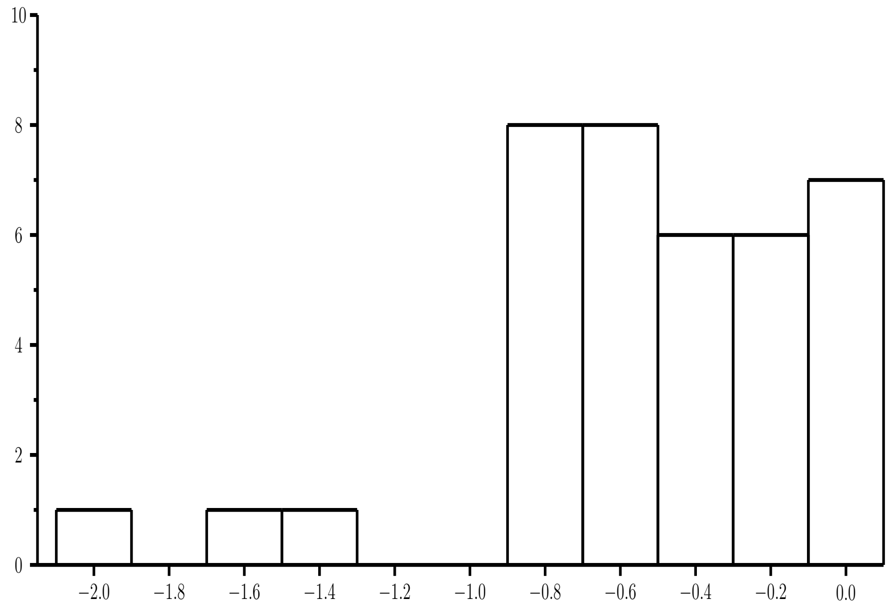

Figure 5.2: Histogram of the long run elasticities of employment with respect to real wages in different industries.

there is still a great deal more room for improving the results by, for example, including 'industry specific' variables in the employment demand functions, we believe that the results obtained so far provide a reasonable basis for the application of the methods developed in this paper to the restricted and unrestricted disaggregate results and those that can be obtained by the direct estimation of the aggregate specification (5.7.2). For the unrestricted estimate of (5.7.2) we obtained

$$
\begin{aligned}
SLET_t = {} & \underset{(51.47)}{-136.50} - \underset{(0.0861)}{0.0217}T_t + \underset{(0.2274)}{0.5862}SLET_{t-1} \\
& + \underset{(0.1833)}{0.0819}SLET_{t-2} + \underset{(0.0670)}{0.4817}SLYT_t + \underset{(0.1253)}{0.0088}SLYT_{t-1} \\
& \underset{(0.0799)}{-0.3508}SLWT_t - \underset{(0.0955)}{0.0334}SLWT_{t-1} + \widehat{u}_{it}, \\
& LLF - 7.77, \quad \overline{R}^2 = 0.9958, \quad \widehat{\sigma} = 0.3717, \\
& DW = 2.06, \quad n = 29, \quad \chi^2_{SC}(1) = 0.88, \\
& \chi^2_{FF}(1) = 0.6279, \quad \chi^2_N(2) = 4.86, \quad \chi^2_H(1) = 1.48.
\end{aligned}
\tag{5.7.3}
$$

The notations are as before, and the test statistics χ^2_{SC}, χ^2_{FF}, χ^2_N, and χ^2_H are already defined at the foot of Table 5.3. This aggregate specification passes all the tests and has reasonable short run and long run properties. However, it is again over-parameterised. The coefficients of T_t, $SLET_{t-1}$, $SLYT_{t-1}$, and $SLWT_{t-1}$ are statistically insignificant whether considered singly or jointly. The chi-squared statistic for the joint test of zero restrictions on the coefficients of these variables was equal to 0.53. So we also estimated the following restricted version of (5.7.2):

$$\begin{aligned} SLET_t = & -134.07 + 0.6956 SLET_{t-1} + 0.4611 SLYT_t \\ & \quad (15.22) \quad (0.0417) \qquad (0.0457)(51.47) \\ & -0.3718 SLWT_t + \widehat{u}_{it}, \\ & \quad (0.0354) \\ & LLF = -8.39, \quad \overline{R}^2 == 0.9963, \quad \widehat{\sigma} = 0.3481, \\ & DW = 2.27, \quad n = 29, \quad \chi^2_{SC}(1) = 0.65, \\ & \chi^2_{FF}(1) = 0.86, \quad \chi^2_N(2) = 5.53, \quad \chi^2_H(1) = 2.20. \end{aligned} \tag{5.7.4}$$

The coefficient estimates are all well determined and imply long run elasticities of aggregate employment with respect to output and real wages of 1.52 and −1.22, respectively.[19] The long run real wage elasticity is only marginally different from the value of −0.92 reported recently by **?** and (Nickell, 1984, p. 177) for the U.K. This similarity is especially striking considering the differences that exist between the two analyses as far as the aggregation procedure, the specification of employment function, and the estimation periods are concerned.

We are now in a position to compare the disaggregate and the aggregate results. As far as the in-sample 'predictive' performance of the aggregate and the disaggregate models is concerned, we computed the s_d^2 criterion (as defined by (5.4.2) for the unrestricted and the restricted versions of the disaggregate model. These were 0.1091 and 0.1000 respectively, thus providing evidence of a slightly better fit for the restricted version of the disaggregate model.[20] The value of the goodness-of-fit criterion for the aggregate equations (5.7.3) and (5.7.4) were equal to 0.1382 and 0.1211, respectively. These results are summarised in Table 5.4, where the uncorrected GG criterion (the first term on the

---

[19] To check for the possible effect of the simultaneous determination of output, employment, and real wages on the OLS estimates, we also estimated (5.7.4) by the instrumental variable method using $z_t = \{1, SLET_{t-1}, SLET_{t-2}, SLYT_{t-1}, SLYT_{t-2}, SLWT_{t-1}, SLWT_{t-2}\}$ as instruments. We obtained the following results:

$$\begin{aligned} SLET_t = & -137.01 + +0.6840 SLET_{t-1} + 0.4745 SLYT_t - 0.3830 SLWT_t + \widehat{u}_{it}, \\ & (20.70) \quad (0.0569) \qquad\qquad (0.0708) \qquad (0.0540) \\ & \overline{R}^2 = 0.9963, \quad \widehat{\sigma} = 0.3487, \quad DW = 2.25, \quad n = 29, \\ & \chi^2_{SC}(1) = 0.56, \quad \chi^2_{FF}(1) = 0.07, \quad \chi^2_N(2) = 4.15, \quad \chi^2_H(1) = 2.18, \end{aligned}$$

which differ only marginally from the OLS results. In fact the Wu-Hausman statistic (T_2 statistic in Wu (1973)), for the test of the 'exogeneity' of $SLYT_t$ and $SLWT_t$ in (5.7.4), using z_t as the instruments, was equal to 0.112, which is well below the 5 percent critical value of the F distribution with 2 and 23 degrees of freedom.

[20] Notice that in general there is no reason to believe that the restricted model should perform better than the unrestricted model as far as the s_d^2 criterion is concerned. Although it is true that the imposition of statistically 'acceptable' linear restrictions on the parameters of the micro equations, such as omitting one or more variables from the micro equations whose t or F values are less than unity, lowers the estimates of σ_{ii}, the same is not true of the estimates of the contemporaneous covariances, σ_{ij}, $i \neq j$. As a result the effect of parameter restrictions on

$$s_d^2 = \sum_{i=1}^{m} \widehat{\sigma}_{ii} + 2 \sum_{i>j} \widehat{\sigma}_{ij}$$

will, in general, be ambiguous.

Table 5.3: Summary and Diagnostic Test Statistics for the Restricted Employment Equations
(1956–1984)

Industry groups	$\overline{R}^2$	χ_r^2	$\widehat{\sigma}$	$\chi_{SC}^2(1)$	$\chi_{FF}^2(1)$	$\chi_N^2(2)$	$\chi_H^2(1)$
1. Agriculture, Forestry and Fishing	.9983	0.04(3)	0.0137	0.01	7.25	0.39	2.25
2. Coal Mining	.9986	3.72(3)	0.0158	0.84	0.91	0.32	0.05
3. Coke	.9771	5.20(5)	0.0449	0.24	0.67	0.27	1.87
4. Mineral Oil and Natural Gas	—	—	—	—	—	—	—
5. Petroleum Products	.9178	4.89(5)	0.0566	0.48	0.01	1.83	0.85
6. Electricity, etc.	.9876	2.19(5)	0.0190	0.17	1.26	0.18	0.12
7. Public Gas Supply	.9719	3.97(4)	0.0322	1.29	0.00	4.86	1.42
8. Water Supply	.9279	0.73(4)	0.0412	1.67	0.00	0.47	1.05
9. Minerals and Ores	.9760	2.40(5)	0.0318	1.36	0.16	32.70	0.00
10. Iron and Steel	.9933	2.49(4)	0.0265	0.08	0.19	1.42	0.43
11. Non-Ferrous Metals	.9864	2.63(2)	0.0250	0.01	3.47	0.20	1.89
12. Non-Metallic Mineral Products	.9935	3.44(3)	0.0177	1.11	0.23	0.76	3.15
13. Chemicals and Manmade Fibers	.9795	6.27(6)	0.0156	3.51	1.80	0.96	1.14
14. Metal Goods	.9877	2.37(5)	0.0192	0.09	0.27	0.38	1.00
15. Mechanical Engineering	.9913	3.64(3)	0.0143	0.93	0.10	0.02	0.73
16. Office Machinery, etc.	.8922	10.47(4)	0.0345	0.05	2.68	7.24	5.05
17. Electrical Engineering	.9698	1.02(2)	0.0173	0.33	0.11	2.19	2.29
18. Motor Vehicles	.9874	1.29(2)	0.0186	1.55	8.92	3.89	0.01
19. Aerospace Equipment	.9878	2.21(4)	0.0268	0.90	0.30	1.81	1.30
20. Ships and Other Vessels	.9817	9.70(6)	0.0323	0.45	0.61	0.40	4.46
21. Other Vehicles	.9973	1.69(4)	0.0241	0.01	0.81	0.17	0.04
22. Instrument Engineering	.9250	7.92(5)	0.0257	0.47	3.07	0.01	0.84
23. Manufactured Food	.9837	0.85(3)	0.0164	1.69	2.78	1.33	4.38
24. Alcoholic Drinks, etc.	.9232	2.56(4)	0.0269	1.32	0.02	0.94	2.06
25. Tobacco	.8796	7.09(6)	0.0497	0.25	8.22	0.65	7.62
26. Textiles	.9981	3.18(5)	0.0175	0.05	4.46	0.74	5.09
27. Clothing and Footwear	.9984	3.76(6)	0.0110	0.36	1.92	0.62	0.03
28. Timber and Furniture	.9864	4.24(4)	0.0138	0.00	2.43	1.34	0.30
29. Paper and Board	.9927	2.86(4)	0.0192	1.09	1.33	1.74	4.41
30. Books, etc.	.9306	4.69(4)	0.0123	1.70	0.01	0.14	0.44
31. Rubber and Plastic Products	.9818	4.73(4)	0.0173	0.21	1.59	0.96	1.03
32. Other Manufactures	.9917	7.18(5)	0.0137	0.37	0.21	1.12	0.00
33. Construction	.9689	5.54(3)	0.0179	5.00	2.34	1.62	1.00
34. Distribution, etc.	.9580	5.44(4)	0.0143	0.49	0.02	0.94	2.06
35. Hotels and Catering	.9169	3.49(5)	0.0198	0.58	1.88	0.45	0.63
36. Rail Transport	.9960	2.36(6)	0.0230	0.28	0.00	1.27	1.98
37. Other Land Transport	.9747	4.04(5)	0.0163	0.02	1.23	0.64	2.71
38. Sea, Air, and Other	.9155	7.87(4)	0.0229	0.27	4.31	0.39	2.75
39. Communications	.9351	4.42(2)	0.0184	1.56	0.48	0.14	1.81
40. Business Services	.9940	1.81(5)	0.0128	0.98	2.01	1.98	0.17
41. Miscellaneous Services	.9512	3.21(6)	0.0222	0.06	0.47	0.39	1.91

Notes: χ_r^2 is the chi-squared statistic for the test of r linear restrictions on the parameters of unrestricted employment equations (see Table 5.1). The value of r is given in brackets after the statistic. $\chi_{SC}^2(1)$ is the first order LM test of residual serial correlation. $\chi_{FF}^2(1)$ is Ramsey's RESET test of order 1. $\chi_N^2(2)$ is a test of normality of the errors. $\chi_H^2(1)$ is a heteroscedasticity test of order 1. $\widehat{\sigma}$ is equation's standard error. $\overline{R}^2$ is the adjusted multiple correlation coefficient. The underlying regressions and the test statistics reported in this table are computed on Data-FIT package. For details of relevant algorithms and references, see Pesaran and Pesaran (1987a).

right-hand side of (5.4.6) is also reported in brackets. On the basis of the proposed choice criterion the restricted as well as the unrestricted versions of the disaggregate model are preferable to the aggregate equation. The computation of the statistic for the test of perfect aggregation defined by (5.5.7) also provided additional support in favour of the disaggregate model. In the case of the unrestricted version the value of this test statistic was equal to 81.66, which is approximately distributed as a χ^2_{29}, thus firmly rejecting the hypothesis of perfect aggregation. This is also clearly reflected in the estimates of the long run elasticities obtained from the disaggregate and the aggregate results. For example, concentrating on the restricted versions of the employment functions, the long run elasticity of aggregate employment with respect to output based on the disaggregate results (Table 5.2) turned out to be equal to 0.724 as compared with the figure of 1.52 obtained using the aggregate specification (5.7.4).[21] Similarly the long run elasticity of aggregate employment with respect to real wages based on the disaggregate results was equal to -0.4551 as compared with the estimate of -1.22 based on the aggregate specification (5.7.4). These results clearly suggest the existence of important upward bias in the estimates of output and real wage elasticities of employment demand obtained in the literature using an economy wide aggregate specification.

5.7.2 Results for the Manufacturing Industries

Having rejected the aggregate employment function in favour of the disaggregate model, the question of what the appropriate level of disaggregation should be naturally arises. One possibility would be to repeat the above analysis for all possible levels of disaggregation. Here in the way of illustration we only consider the problem in the case of the manufacturing industries. The disaggregate results for this industry grouping are given by the industries labelled 10 to 32 inclusive in Tables 5.1 and 5.2. We also obtained the following estimates of the unrestricted and the restricted employment demand functions for the manufacturing sector as a whole:

$$\begin{aligned} SLEM_t = {} & \underset{(22.19)}{-65.58} - \underset{(0.0565)}{0.0039}T_t + \underset{(0.2211)}{0.7491}SLEM_{t-1} \\ & - \underset{(0.1655)}{0.0162}SLEM_{t-2} + \underset{(0.0531)}{0.4933}SLYM_t - \underset{(0.1170)}{0.0897}SLYM_{t-1} \\ & - \underset{(0.0659)}{0.2979}SLWM_t - \underset{(0.0837)}{0.0148}SLWM_{t-1} + \widehat{u}_{it}, \\ & LLF7.20, \quad \overline{R}^2 = 0.9968, \quad \widehat{\sigma} = 0.2218, \quad DW = 2.21, \quad n = 29, \\ & \chi^2_{SC}(1) = 4.56, \quad \chi^2_{FF}(1) = 0.01, \quad \chi^2_N(2) = 0.24, \quad \chi^2_H(1) = 1.70, \end{aligned} \tag{5.7.5}$$

[21] The estimates of the long run elasticities for the disaggregate model were computed from the simple averages of the micro coefficients.

and

$$
\begin{aligned}
SLEM_t = & -66.82 + 0.7407SLEM_{t-1} + 0.4004SLYM_t \\
& \quad (6.65) \quad (0.0417) \quad (0.0434) \\
& +0.0906\Delta SLEM_t - 0.3137SLWM_t + \widehat{u}_{it}, \\
& \quad (0.0551) \qquad\qquad (0.0343) \\
& LLF7.13, \quad \overline{R}^2 = 0.9972, \quad \widehat{\sigma} = 0.2080, \quad DW = 2.19, \quad n = 29, \\
& \chi^2_{SC}(1) = 0.34, \quad \chi^2_{FF}(1) = 0.00, \quad \chi^2_N(2) = 0.21, \quad \chi^2_H(1) = 1.47,
\end{aligned} \tag{5.7.6}
$$

where

$$
SLEM_t = \sum_{i=10}^{32} LEH_{it}, \quad SLYM_t = \sum_{i=10}^{32} LY_{it}, \quad SLWM_t = \sum_{i=10}^{32} LW_{it}.
$$

The restricted version (5.7.6) clearly cannot be rejected against the unrestricted version (5.7.5).[22] In this application the values of the goodness-of-fit criterion (s_d^2) for the unrestricted and the restricted models were 0.0506 and 0.0437 respectively, indicating that the restricted version of the disaggregate model has a better in-sample performance insofar as predicting the aggregate employment variable $SLEM_t$ is concerned. The goodness-of-fit criterion for the aggregate specifications (5.7.5) and (5.7.6) are given by 0.0492 and 0.0433 respectively. (See also Table 5.4.) Hence on the basis of the choice criterion, for the manufacturing industries the aggregate models give a marginally better fit than either of the disaggregate models. This, of course, does not mean that the aggregate model is not subject to the aggregation error problem. In fact the application of the test of perfect aggregation to this example resulted in the value of 69.92 for the a_m statistic which is well in excess of the $5(= n)$ degrees of freedom.

The rejection of the perfect aggregation hypothesis is also reflected in the large differences that exist between the estimates of the long run real wage and output elasticities of the manufacturing employment based on the disaggregate and the aggregate results. In the case of the restricted models, the estimates of the long run real wage elasticity based on the aggregate and the disaggregate models were -1.21 and -0.509, respectively. The corresponding figures for the long run real output elasticities were 1.54 and 0.763, respectively. The better performance of the aggregate model should be interpreted as an important indication that the disaggregate employment functions are misspecified. This suggests the need for a much more detailed analysis of employment demand at the industry level, which may involve including 'industry specific' variables in employment equations, experimenting with a different choice of functional forms across industries, or searching for new industry-specific explanatory variables, or even compiling a more reliable set of micro data.

5.8 Concluding Remarks

In this paper our primary concern has been with the problem of choice between macro and micro regression equations for the purpose of predicting macro variables. The test of

[22] We also estimated the restricted version (5.7.6) by the IV method using $z_t = (1, SLEM_{t-1}, SLEM_{t-2}, SLYM_{t-1}, SLYM_{t-2}, SLWM_{t_1}, SLWM_{t-2})$ as instruments and obtained very similar results.

Table 5.4: Relative Predictive Performance of the Aggregate and the Disaggregate Employment Functions (1956–1984)

	Aggregate Equations		Disaggregate Equations[3]	
	Unrestricted[1]	Restricted[2]	Unrestricted	Restricted
All industries[4]	0.1382	0.1211	0.1091 (0.0846)[6]	0.1000 (0.0859)
Manufacturing[5]	0.0492	0.0433	0.0506 (0.0439)	0.0437 (0.0389)

[1] See equations (5.7.3) and (5.7.5).
[2] See equations (5.7.4) and (5.7.6).
[3] See the results in Tables 5.1 and 5.2.
[4] Excluding Industry 4, Mineral Oil and Natural Gas.
[5] Industries 10 to 32 inclusive.
[6] Bracketed figures refer to the degrees-of-freedom uncorrected measure of the choice criterion, given by the first term in the expression for s_d^2 defined in (5.4.6).

perfect aggregation developed in the paper also addresses the macro prediction problem; although as our application to the UK employment demand functions shows, it has some bearing on the problem of aggregation bias as well. In using the goodness-of-fit criterion and the test of perfect aggregation it is, however, important to note that these methods, like most other methods of inference in econometrics, suffer from the fact that they may have little to say on the validity of the aggregation conditions outside the estimation period. In the case of aggregation across micro units this problem is especially serious as the extension of the results of aggregation tests to the post estimation period requires stability of the micro coefficients as well as the stability of the industrial composition of the economy.

.1 Data Appendix: Data Sources and Definitions

The data used in the empirical analysis in Section 5.7 are annual observations on 41 industry groups for the UK obtained from the Cambridge Growth Project Databank. The data on industry man-hours, employment, wages and salaries, and employers' contributions were originally provided by the Institute for Employment Research at the University of Warwick. The data on industry output were obtained from the Central Statistical Office's commodity flow accounts adjusted for our industrial classification. The data on producer price indices of industry output were obtained from a number of published sources including the Department of Trade and Industry's publication "British Business", the CSO's publications, the "Annual Abstract of Statistics" and the "Monthly Digest of Statistics", and the Department of Energy's "Energy Trends".

Some of the 41 industry groups are identical to the 'groups' distinguished in the 1980 Standard Industrial Classification. However, in view of the significant differences between

Table 0.5: Classification of Industry Groups
(In Terms of the 1980 Standard Industrial Classification)

Industry	Division, Class or Group
1. Agriculture, Forestry, and Fishing	0
2. Coal Mining	1113, 1114
3. Coke	1115, 1200
4. Mineral Oil and Natural Gas	1300
5. Petroleum Products	140
6. Electricity, etc.	1520, 1610, 1630
7. Public Gas Supply	1620
8. Water Supply	1700
9. Minerals and Ores n.e.s.	21, 23
10. Iron and Steel	2210, 2220, 223
11. Non-Ferrous Metals	224
12. Non-Metallic Mineral Products	24
13. Chemicals and Manmade Fibers	25, 26
14. Metal Goods n.e.s.	31
15. Mechanical Engineering	32
16. Office Machinery, etc.	33
17. Electrical Engineering	34
18. Motor Vehicles	35
19. Aerospace Equipment	3640
20. Ships and Other Vessels	3610
21. Other Vehicles	3620,363,3650
22. Instrument Engineering	37
23. Manufactured Food	41, 4200, 421, 422, 4239
24. Alcoholic Drinks, etc.	4240, 4261, 4270, 4283
25. Tobacco	4290
26. Textiles	43
27. Clothing and Footwear	45
28. Timber and Furniture	46
29. Paper and Board	4710, 472
30. Books, etc.	475
31. Rubber and Plastic Products	48
32. Other Manufactures	44, 49
33. Construction	5
34. Distribution, etc.	61, 62, 63, 64, 65, 67
35. Hotels and Catering	66
36. Rail Transport	71
37. Other Land Transport	72
38. Sea, Air, and Other	74, 75, 76, 77
39. Communications	79
40. Business Services	81, 82, 83, 84, 85
41. Miscellaneous Services	94, 98, 923, 95, 96, 97.

them in a large number of cases, the groups are listed in Table 0.5, using as a reference the Division, Class or Group of the 1980 Standard Industrial Classification. In the analysis of the manufacturing sector groups 10 to 32 inclusive are included.

For empirical estimation, the man-hours employed (EH_t) are defined as a product of the actual hours worked per week and the numbers employed in each of 41 industries, including self employed ('000s) in these industries. Industry output (Y) is gross value added by industry in 1980 prices (m). Average real wage rate (W) is a measure of the real product wage by industry. It is obtained by first deflating an industry's total labour costs including both employees' wages and salaries and employers' national insurance contributions (m) by the price index of industry output (1980 = 1.00). This is then divided by the man-hours employed in that industry to obtain the average real wage rate. All the data are annual covering the period 1954–1984 with both the aggregate and disaggregate equations estimated over the period 1956–1984. These data, and the computer programs used both in estimation and in the computation of the choice criterion and the statistics for the test of perfect aggregation, are available on request from the authors.

Chapter 6

A proof of the asymptotic validity of a test for perfect aggregation

An asymptotic proof is presented for a test of perfect aggregation in linear models developed in Pesaran et al. (1989a). The limiting distribution is derived by letting the degree of disaggregation increase without bound for a fixed sample size.

6.1 Introduction

In a recent paper, Pesaran et al. (1989a), henceforth PPK, propose a new test of the hypothesis of perfect aggregation in the context of a linear disaggregate model. It is shown there that, when the covariance matrix of the disaggregate model is known, then this test statistic is distributed as χ^2_n, where n is the number of observations. However, when the covariance matrix is unknown, the exact distribution of the statistic is not easily computable. The purpose of this paper is to show that in this case the test is still valid asymptotically. Since the test statistic has dimension n, the usual asymptotic theory which lets n, the sample size, tend to infinity is clearly not applicable. Instead we derive a limiting distribution by allowing the degree of disaggregation denoted by m to increase without bound. Similar large m-asymptotics have been used previously by Powell and Stoker (1985) and Granger (1987). The perfect aggregation test is derived in section6.2. Section 6.3 presents a proof of the asymptotic validity of the test for the special case where the covariance matrix is diagonal.

6.2 The perfect aggregation test

Let the disaggregate model be written as

$$H_d: \quad \mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{u}_i, \quad i = 1, 2, \ldots, m. \tag{6.2.1}$$

where $\mathbf{y}_i$ is an $n \times 1$ vector of n observations on the ith micro-unit, $\mathbf{X}_i$ is an $n \times k$ matrix of n observations on k regressors for the ith micro-unit, $\mathbf{u}_i$ is an $n \times 1$ vector of associated

[0] Published in *Economic Letters* (1989), Vol 30, pp. 41–47. Co-author M. H. Pesaran. The authors are grateful to Jon Breslaw and James MacKinnon for helpful comments.

disturbances. The disturbances $\mathbf{u}_i$ are assumed to be distributed independently of $\mathbf{X}_i$ with $\mathrm{E}(\mathbf{u}_i) = \mathbf{0}$ and $\mathrm{E}(\mathbf{u}_i\mathbf{u}_j') = \sigma_{ij}\mathbf{I}_n$.

The aggregate model is given by

$$H_a: \quad \mathbf{y}_a = \mathbf{X}_a\mathbf{b} + \boldsymbol{\nu}_a, \tag{6.2.2}$$

where

$$\mathbf{y}_a = \sum_{i=1}^{m} \mathbf{y}_i \quad \text{and} \quad \mathbf{X}_a = \sum_{i=1}^{m} \mathbf{X}_i.$$

Then the hypothesis of perfect aggregation is defined by

$$H_\xi: \quad \xi = \sum_{i=1}^{m} \mathbf{X}_i\boldsymbol{\beta}_i - \mathbf{X}_a\mathbf{b} = \mathbf{0}, \tag{6.2.3}$$

in which case the regression functions of (6.2.1) and (6.2.2) coincide. It can be seen that this hypothesis encompasses the two special cases of 'micro-homogeneity' ($\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_m$) and 'compositional stability' ($\mathbf{X}_i = \mathbf{X}_a\mathbf{C}_i, i = 1, 2, \ldots, m$) but it can also be satisfied in more general cases (see PPK for details).

A test of the hypothesis (6.2.3) can be constructed based on OLS estimates of $\boldsymbol{\beta}_i$. Let

$$\widehat{\xi} = \sum_{i=1}^{m} \mathbf{X}_i\widehat{\boldsymbol{\beta}}_i - \mathbf{X}_a\widehat{\mathbf{b}} = \mathbf{e}_a - \mathbf{e}_d, \tag{6.2.4}$$

where the hat denotes OLS estimates,

$$\mathbf{e}_a = (\mathbf{I}_n - \mathbf{A}_a)\mathbf{y}_a,$$

$$\mathbf{e}_d = \sum_{i=1}^{m}(\mathbf{I}_n - \mathbf{A}_i)\mathbf{y}_i = \sum_{i=1}^{m}\mathbf{e}_i,$$

$$\mathbf{A}_a = \mathbf{X}_a(\mathbf{X}_a'\mathbf{X}_a)^{-1}\mathbf{X}_a',$$

$$\mathbf{A}_i = \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i' = \mathbf{I}_n - \mathbf{M}_i.$$

The test statistic is then given by

$$\underline{a}_m = m^{-1}(\mathbf{e}_a - \mathbf{e}_d)'\widehat{\boldsymbol{\Psi}}_m^{-1}(\mathbf{e}_a - \mathbf{e}_d),$$

$$\widehat{\boldsymbol{\Psi}}_m = m^{-1}\sum_{i,j=1}^{m}\widehat{\sigma}_{ij}\mathbf{H}_i\mathbf{H}_j, \tag{6.2.5}$$

$$\mathbf{H}_i = \mathbf{A}_i - \mathbf{A}_a,$$

$$\widehat{\sigma}_{ij} = \{n - 2k + \mathrm{tr}(\mathbf{A}_i\mathbf{A}_j)\}^{-1}\mathbf{e}_i'\mathbf{e}_j.$$

It is shown in PPK that $\widehat{\sigma}_{ij}$ is an unbiased estimator of the covariance element σ_{ij}.

6.3 A proof of the asymptotic validity of the test

In this section a proof is presented of the asymptotic validity of the test of perfect aggregation for the special case where the disturbances u_{it} are distributed independently across equations so that $\sigma_{ij} = 0, i \neq j$. The framework adopted is to let the degree of disaggregation, m, increase without bound while keeping the sample size, n, fixed.

The following assumptions are made:

Assumption 1. The standardised micro-disturbances $\nu_{it} = u_{it}/\sqrt{\sigma_{ii}}$ are identically distributed, independently both across time periods and across equations, with zero means, unit variances and finite third-order moments.[1]

Assumption 2. The average matrix $\overline{\mathbf{X}}_m = m^{-1}\mathbf{X}_a$, and the aggregate projection matrix $\overline{\mathbf{X}}_m(\overline{\mathbf{X}}'_m\overline{\mathbf{X}}_m)^{-1}\overline{\mathbf{X}}'_m$, converge (in probability) to finite limits.

Assumption 3. The elements of the disaggregate projection matrices, $\mathbf{A}_i = \mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i$, remain bounded in absolute value as $m \to \infty$. Notationally, we write $|\mathbf{A}_i| < \mathbf{P} < \infty$.

Assumption 4. The elements of the variance matrix $\mathbf{\Sigma} = (\sigma_{ij})$ remain bounded $m \to \infty$. Namely, $|\sigma_{ij}| < \tau^2 < \infty, \forall i, j$.

Assumption 5. The variance matrix $\mathbf{\Psi}_m$ defined by

$$\mathbf{\Psi}_m = m^{-1}\sum_{i,j=1}^{m}\sigma_{ij}\mathbf{H}_i\mathbf{H}_j,$$

tends to a non-singular matrix $\mathbf{\Psi}$, as $m \to \infty$.

Theorem 1. Under *Assumptions 1—5* and conditional on $\mathbf{X}$, the statistic

$$\underline{a}_m = (\mathbf{e}_a - \mathbf{e}_d)'\left(\sum_{i=1}^{m}\widehat{\sigma}_{ii}\mathbf{H}_i^2\right)^{-1}(\mathbf{e}_a - \mathbf{e}_d)$$

will be asymptotically distributed as a χ^2_n variate on the null hypothesis of perfect aggregation (6.2.3), as $m \to \infty$.

Proof. Let

$$\mathbf{g}_m = \left(\sum_{i=1}^{m}\widehat{\sigma}_{ii}\mathbf{H}_i^2\right)^{-1/2}(\mathbf{e}_a - \mathbf{e}_d). \tag{6.3.1}$$

Then the test statistic in the theorem can be written as

$$\underline{a}_m = \mathbf{g}'_m\mathbf{g}_m. \tag{6.3.2}$$

Consider now the probability limit of $\widehat{\mathbf{\Psi}}_m = m^{-1}\sum_{i=1}^{m}\widehat{\sigma}_{ii}\mathbf{H}_i^2$, as $m \to \infty$. Under (1) we obtain

$$\widehat{\mathbf{\Psi}}_m = [m(n-k)]^{-1}\sum_{i=1}^{m}(\mathbf{u}'_i\mathbf{M}_i\mathbf{u}_i)\mathbf{H}_i^2. \tag{6.3.3}$$

[1] The assumption that ν_{it} have finite third-order moments can be replaced by the slightly weaker assumption that, for some positive δ, $\mathrm{E}\,|\nu_{it}|^{2+\delta}$ is uniformly bounded. See, for example, (White, 1984, Theorem 5.10).

But, since $\mathbf{M}_i$ is an idempotent matrix of rank $n-k$, we can also write

$$\sigma_{ii}^{-1}\mathbf{u}_i'\mathbf{M}\mathbf{u}_i = \sum_{t=1}^{n-k}\epsilon_{it}^2, \quad i = 1, 2, \ldots, m, \tag{6.3.4}$$

where ϵ_{it} represents scalar random variables distributed independently across i and t with zero means and unit variances. Substituting (6.3.4) in (6.3.3) yields

$$\widehat{\mathbf{\Psi}}_m = (n-k)^{-1}\sum_{t=1}^{n-k} m^{-1}\left(\sum_{i=1}^{m}\sigma_{ii}\epsilon_{it}^2\mathbf{H}_i^2\right). \tag{6.3.5}$$

But, noting that $\mathbf{H}_i = \mathbf{A}_i - \mathbf{A}_a$, we have

$$m^{-1}\sum_{i=1}^{m}\sigma_{ii}\epsilon_{it}^2\mathbf{H}_i^2 = f_m\mathbf{A}_a + \mathbf{F}_m - \mathbf{F}_m\mathbf{A}_a - \mathbf{A}_a\mathbf{F}_m, \tag{6.3.6}$$

where

$$f_m = m^{-1}\sum_{i=1}^{m}\sigma_{ii}\epsilon_{it}^2, \quad \mathbf{F}_m = m^{-1}\sum_{i=1}^{m}\sigma_{ii}\epsilon_{it}^2\mathbf{A}_i.$$

Now, under Assumption 4, it readily follows that

$$\operatorname*{plim}_{m\to\infty}(f_m) \leq \tau^2 \operatorname*{plim}_{m\to\infty}\left(m^{-1}\sum_{i=1}^{m}\epsilon_{it}^2\right),$$

and since ϵ_{it} are identically and independently distributed random variables then, by the law of large numbers, $m^{-1}\sum_{i=1}^{m}\epsilon_{it}^2 \xrightarrow{p} 1$, and

$$\operatorname*{plim}_{m\to\infty}(f_m) \leq \tau^2 < \infty. \tag{6.3.7}$$

Similarly, under Assumptions 3 and 4, we have

$$\operatorname*{plim}_{m\to\infty}(\mathbf{F}_m) \leq \tau^2\mathbf{P} < \infty, \tag{6.3.8}$$

where $\mathbf{P}$ is already defined by Assumption 3. The results (6.3.7) and (6.3.8) establish the existence of the probability limits of f_m and $\mathbf{F}_m$, as $m \to \infty$, and this in turn establishes [using (6.3.6) and noting that, by Assumption 2, matrix $\mathbf{A}_a$, has a finite limit as $m \to \infty$] that

$$\operatorname*{plim}_{m\to\infty}\left(m^{-1}\sum_{i=1}^{m}\sigma_{ii}\epsilon_{it}^2\mathbf{H}_i^2\right) = \lim_{m\to\infty}\left(m^{-1}\sum_{i=1}^{m}\sigma_{ii}\mathbf{H}_i^2\right).$$

Using this result in (6.3.5), we finally obtain

$$\widehat{\mathbf{\Psi}}_m = m^{-1}\sum_{i=1}^{m}\widehat{\sigma}_{ii}\mathbf{H}_i^2 \xrightarrow{p} \lim_{m\to\infty}\left(m^{-1}\sum_{i=1}^{m}\sigma_{ii}\mathbf{H}_i^2\right) = \mathbf{\Psi}. \tag{6.3.9}$$

Therefore, asymptotically we have[2]

$$\mathbf{g}_m \overset{a}{\sim} \mathbf{\Psi}^{-1/2}m^{-1/2}(\mathbf{e}_a - \mathbf{e}_d).$$

[2] Note that, by Assumption 5, matrix $\mathbf{\Psi}$ is non-singular.

But, under (1) and on the assumption that $H_\xi : \sum_{i=1}^{m} \mathbf{X}_i \widehat{\boldsymbol{\beta}}_i = \mathbf{X}_a \widehat{\mathbf{b}}$ holds,

$$m^{-1/2}(\mathbf{e}_a - \mathbf{e}_d) = m^{-1/2} \sum_{i=1}^{m} \mathbf{H}_i \mathbf{u}_i.$$

Hence,

$$\mathbf{g}_m \overset{a}{\sim} m^{-1/2} \sum_{i=1}^{m} \mathbf{z}_i, \tag{6.3.10}$$

in which

$$\mathbf{z}_i = \sqrt{\sigma_{ii}} \boldsymbol{\Psi}^{-1/2} \mathbf{H}_i \boldsymbol{\nu}_i,$$

and $\boldsymbol{\nu}_i = \mathbf{u}_i / \sqrt{\sigma_{ii}}$. We now show that under the assumptions of the theorem, as $m \to \infty$, the sum $\mathbf{s}_m = m^{-1/2} \sum_{i=1}^{m} \mathbf{z}_i$ tends to a multivariate normal distribution with mean zero and covariance matrix $\mathbf{I}_n$, an identity matrix of order n. For this purpose, it is sufficient to demonstrate that for any fixed vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)'$, the limiting distribution of $\boldsymbol{\lambda}' \mathbf{s}_m$ is $N(0, \boldsymbol{\lambda}'\boldsymbol{\lambda})$.

Let

$$d_m = \boldsymbol{\lambda}' \mathbf{s}_m = m^{-1/2} \sum_{i=1}^{m} w_i, \tag{6.3.11}$$

in which

$$w_i = \sqrt{\sigma_{ii}} \boldsymbol{\lambda}' \boldsymbol{\Psi}^{-1/2} \mathbf{H}_i \boldsymbol{\nu}_i, \quad i = 1, 2, \ldots, m \tag{6.3.12}$$

is now a scalar random variable. We have, for all i,

$$\mathrm{E}(w_i) = 0,$$

$$\mathrm{V}(w_i) = \sigma_{ii} \boldsymbol{\lambda}' \boldsymbol{\Psi}^{-1/2} \mathbf{H}_i^2 \boldsymbol{\Psi}^{-1/2} \boldsymbol{\lambda} > \mathbf{0}.$$

Setting $\boldsymbol{\mu} = \boldsymbol{\Psi}^{-1/2} \boldsymbol{\lambda}$, then

$$C_m^2 = \sum_{i=1}^{m} \mathrm{V}(w_i) = \boldsymbol{\mu}' \left(\sum_{i=1}^{m} \sigma_{ii} \mathbf{H}_i^2 \right) \boldsymbol{\mu}. \tag{6.3.13}$$

Denoting the (s,t) element of matrix $\mathbf{H}_i$ by $h_{i,st}$, we also have [using (6.3.12)]

$$w_i = \sqrt{\sigma_{ii}} \sum_{t=1}^{n} \left(\sum_{s=1}^{n} \mu_s h_{i,st} \right) \nu_{it}.$$

Therefore, since by assumption $\boldsymbol{\Psi}$ is non-singular and $h_{i,st}$ are bounded in absolute value for all i, then

$$|w_i| \leq n\kappa \sqrt{\sigma_{ii}} \left| \sum_{t=1}^{n} \nu_{it} \right|,$$

where $|\mu_s h_{i,st}| < \kappa < \infty$. Consequently,

$$\mathrm{E}\,|w_i|^3 \leq n^3 \kappa^3 \sigma_{ii}^{3/2}\, \mathrm{E} \left| \sum_{t=1}^{n} \nu_{it} \right|^3.$$

However, since the random variables ν_{it} are i.i.d. with finite third-order moments, $\mathrm{E}\,|\sum_{t=1}^{n}\nu_{it}|^3 \leq n\theta^3$, where $\theta^3 = \mathrm{E}\,|\nu_{it}|^3$, and

$$\mathrm{E}\,|w_i|^3 \leq n^4\kappa^3\theta^3\sigma_{ii}^{3/2}. \tag{6.3.14}$$

We are now in a position to apply the Liapunov Central Limit Theorem to the sum d_m defined by (6.3.11).[3] Setting

$$B_m^3 = \sum_{i=1}^{m} \mathrm{E}\,|w_i|^3,$$

then using (6.3.14) it follows that

$$B_m^3 \leq (n^4\kappa^3\theta^3)\sum_{i=1}^{m}\sigma_{ii}^{3/2},$$

which together with (6.3.13) yields[4]

$$\lim_{m\to\infty}\left[\frac{B_m}{C_m}\right] \leq \left[\frac{n^{4/3}\kappa\theta}{(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{1/2}}\right]\lim_{m\to\infty} m^{-1/2}\left[\sum_{i=1}^{m}\sigma_{ii}^{3/2}\right]^{1/3}.$$

But, under Assumption 4,

$$\lim_{m\to\infty} m^{-1/2}\left[\sum_{i=1}^{m}\sigma_{ii}^{3/2}\right]^{1/3} \leq \lim_{m\to\infty}(m^{-1/6}\tau) = 0,$$

and for a fixed n, we have $\lim(B_m/C_m) = 0$, as $m \to \infty$, and the condition of the Liapunov theorem will be met. Hence,

$$\mathbf{g}_m \overset{a}{\sim} \mathbf{s}_m \overset{a}{\sim} N(\mathbf{0}, \mathbf{I}_n).$$

Now, using (6.3.2), we have

$$\underline{a}_m = \mathbf{g}_m'\mathbf{g}_m \overset{a}{\sim} \chi_n^2.$$

Q.E.D.

[3] See, for example, (Rao, 1973, p. 127).

[4] Notice that $\lim_{m\to\infty}\{\boldsymbol{\mu}'(m^{-1}\sum_{i=1}^{m}\sigma_{ii}\mathbf{H}_i^2)\boldsymbol{\mu}\} = \boldsymbol{\mu}'\boldsymbol{\Psi}\boldsymbol{\mu} = \boldsymbol{\lambda}'\boldsymbol{\lambda}$.

Chapter 7

Testing for Aggregation Bias in Linear Models

The problem of aggregation over micro units has had a long tradition in the econometrics literature, stretching back to the pioneering work of Theil (1954). In this literature two issues in particular have attracted attention. The first concentrates on the prediction problem of choosing whether to use macro or micro equations to predict aggregate variables. This issue was raised by Grunfeld and Griliches (1960) and is further addressed in a recent paper by Pesaran et al. (1989b) (PPK). In PPK a generalised prediction criterion and a formal statistical test of the hypothesis of perfect aggregation are developed. The present paper considers the second strand in this literature which is concerned with the problem of 'aggregation bias' defined by the deviation of the macro parameters from the average of the corresponding micro parameters. (See for example Theil (1954), Boot and de Wit (1960), Orcutt et al. (1968), Gupta (1971) and Sasaki (1978).) In this paper we develop direct tests of aggregation bias in contrast to the indirect test proposed by Zellner (1962) which tests the hypothesis that all the disaggregated coefficients are equal. We also derive generalised versions of the tests for the case where the parameters of interest are subsets or (possibly non- linear) functions of the full parameter vector. This is particularly relevant when the focus of the analysis is on the long run properties of the aggregate and disaggregate models. Since the tests of the aggregation bias, whether of the type discussed here or the one proposed in Zellner (I962), assume the disaggregate model is correctly specified, in this paper we also develop a Durbin-Hausman type misspecification test of the disaggregate model. Section 7.1 sets out the statistical framework and assumptions. Section 7.2 develops the aggregation bias tests. Section 7.3 derives the Durbin-Hausman type misspecification test of the disaggregate model. Section 7.4 applies these tests to a disaggregate model of employment demand for the United Kingdom taken from PPK.

[0] Published in *Economic Journal* (1990), Vol. 100 (Conference 1990), pp. 137–150. Co-authors K. C. Lee and M. H. Pesaran.

7.1 Framework and Assumptions

In order to develop the tests we consider the following general disaggregate model:

$$H_d: \quad \mathbf{y}_t = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i, \quad i = 1, 2, \ldots, m \tag{7.1.1}$$

where $\mathbf{y}_i$ is the $n \times 1$ vector of observations on the dependent variable for the ith unit, $\mathbf{X}_i$ is the $n \times k$ matrix of observations on the regressors in (7.1.1) for the ith unit, $\boldsymbol{\beta}_i$ is the $k \times 1$ vector of the coefficients associated with columns of $\mathbf{X}_i$, and $\mathbf{u}_i$ is the $n \times 1$ vector of disturbances for the ith unit. The corresponding aggregate equation that satisfies the Klein-Nataf consistency requirement is given by [1]

$$H_a: \quad \mathbf{y}_a = \mathbf{X}_a\mathbf{b}_a + \mathbf{v}, \tag{7.1.2}$$

where

$$\mathbf{y}_a = \sum_{i=1}^{m} \mathbf{y}_i, \quad \mathbf{X}_a = \sum_{i=1}^{m} \mathbf{X}_i,$$

and $\mathbf{b}_a$ is the $k \times 1$ vector of macro parameters. The $n \times 1$ disturbance vector $\mathbf{v}$, will be equal to $\mathbf{u}_a = \sum_{i=1}^{m} \mathbf{u}_i$, only if the 'perfect aggregation' condition

$$H_\xi: \boldsymbol{\xi} = \sum_{i=1}^{m} \mathbf{X}_i\boldsymbol{\beta}_i - \mathbf{X}_a\mathbf{b}_a = \mathbf{0}, \tag{7.1.3}$$

discussed in detail in PPK, is satisfied. Here we focus on the problem of aggregation bias and develop alternative methods of analysing and formally testing the extent of this bias in economic applications. In what follows we adopt the following assumptions:

Assumption 3. *The n elements of the disturbance vector $\mathbf{u}_i = \{u_{it}\}$, have zero means, constant variances and are serially independently distributed. They also satisfy the moment condition*

$$\mathrm{E}\left|u_{it}\right|^{2+\delta} < \Delta < \infty, \ \textit{for some } \delta > 0, \ \textit{and all } t.$$

Assumption 4. *The disturbance vectors $\mathbf{u}_i$ are distributed independently of $\mathbf{X}_i$, and $\mathrm{E}(\mathbf{u}_i\mathbf{u}_j{}') = \sigma_{ij}\mathbf{I}_n$, for all i and j, ($\sigma_{ii} > 0$).*

Assumption 5. *The matrices $\mathbf{X}_i$ have full rank, the probability limits*

$$\operatorname*{plim}_{n\to\infty}(n^{-1}\mathbf{X}_i{}'\mathbf{X}_i) = \boldsymbol{\Sigma}_{ij}, \quad i, j = a, 1, 2, \ldots, m,$$

exist, and the $k \times k$ matrices $\boldsymbol{\Sigma}_{ii}$, $i = a, 1, 2, \ldots, m$ are non-singular.

We also base our tests on the OLS estimates

$$\widehat{\mathbf{b}} = (\mathbf{X}_a{}'\mathbf{X}_a)^{-1}\mathbf{X}_a{}'\mathbf{y}_a, \quad \widehat{\boldsymbol{\beta}}_i = (\mathbf{X}_i{}'\mathbf{X}_i)^{-1}\mathbf{X}_i{}'\mathbf{y}_i, \quad i = 1, 2, \ldots, m,$$

although, in principle, the tests proposed below can also be constructed using the more efficient SURE (Seemingly Unrelated Regression Equations) estimators of $\boldsymbol{\beta}_i$, due to Zellner (1962).

[1] See Lovell (1973), and the discussion in PPK (p. 25).

7.2 Direct Tests of Aggregation Bias

The problem of 'aggregation bias', as originally discussed by Theil (1954) is defined in terms of the deviations of macro parameters from the averages of the corresponding micro parameters.[2] In the context of the linear disaggregate and aggregate models (7.1.1) and (7.1.2), the vector of aggregation bias is defined by

$$\boldsymbol{\eta}_\beta = \mathbf{b} - \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{\beta}_i, \tag{7.2.1}$$

A test of aggregation bias then involves testing the hypothesis $H_0 : \boldsymbol{\eta}_\beta = \mathbf{0}$. In testing this hypothesis the case where $\mathbf{b}$ is given *a priori* (for example by a 'consensus' view) should be distinguished from the case where $\mathbf{b}$ is defined as the pseudo true value of $\widehat{\mathbf{b}}$ assuming that the disaggregate model is correctly specified. In the former case the relevant statistic for testing the hypothesis $H_0 : \boldsymbol{\eta}_\beta = \mathbf{0}$ is given by

$$q_1 = \left(\mathbf{b} - \frac{1}{m}\sum_{i=1}^{m}\widehat{\boldsymbol{\beta}}_i\right)' \widehat{\boldsymbol{\Omega}}_n^{-1} \left(\mathbf{b} - \frac{1}{m}\sum_{i=1}^{m}\widehat{\boldsymbol{\beta}}_i\right), \tag{7.2.2}$$

where $\widehat{\boldsymbol{\Omega}}_n$ represents a consistent estimator of $\boldsymbol{\Omega} = m^{-2}\sum_{i,j=1}^{m}\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\beta}}_j)$.[3] Under assumptions 3–5 it is easily seen that q_1 is asymptotically distributed as χ^2_k The statistic q_1 takes $\mathbf{b}$ as a fixed vector, and tests for the deviation of the average of micro parameters from this fixed vector on the assumption that H_d holds. In practice, however, it is rare that a 'consensus' value for $\mathbf{b}$ or some of its elements is available, and $\mathbf{b}$ needs to be chosen in light of the knowledge of the disaggregate model. When H_d holds the pseudo true value of $\mathbf{b}$ is given by

$$\mathbf{b} = \operatorname*{plim}_{n\to\infty}(\widehat{\mathbf{b}}|H_d) = \sum_{i=1}^{m}\mathbf{C}_i\boldsymbol{\beta}_i, \tag{7.2.3}$$

where

$$\mathbf{C}_i = \boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ai}, \quad i = 1, 2, \ldots, m, \tag{7.2.4}$$

satisfy the condition $\sum_{i=1}^{m}\mathbf{C}_i = \mathbf{I}_k$. $\mathbf{I}_k$ is an identity matrix of order k.) The matrices $\mathbf{C}_i$ are the probability limits of the coefficients in the OLS regressions of the columns of $\mathbf{X}_i$ on $\mathbf{X}_a$; the 'auxiliary' equations in Theil's terminology. Notice that result (7.2.3) holds only when H_d is correctly specified. We will use this result later as the basis of a Durbin-Hausman type test of misspecification of the disaggregate model. For the time being, however, we assume that the disaggregate model H_d is correctly specified and write H_0 as

$$H_0; \quad \sum_{i=1}^{m}\left(\mathbf{C}_i - \frac{1}{m}\mathbf{I}_k\right)\boldsymbol{\beta}_i = \mathbf{0}. \tag{7.2.5}$$

An indirect, albeit familiar method of testing (7.2.5), originally proposed by Zellner (1962), is to test the micro-homogeneity hypothesis

$$H_\beta : \quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_m.$$

[2] For empirical analysis of aggregation bias see, for example, the papers by Boot and de Wit (1960), Gupta (1971) and Sasaki (1978).

[3] Note that $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\beta}}_j) = \sigma_{ij}(\mathbf{X}_i{}'\mathbf{X}_i)^{-1}(\mathbf{X}_i{}'\mathbf{X}_j)(\mathbf{X}_j{}'\mathbf{X}_j)^{-1}$.

Testing H_β as a method of testing H_0 is however rather too restrictive. Although H_β implies H_0, the reverse is not true. It is possible for $\boldsymbol{\eta}_\beta = \mathbf{0}$ to hold even when the micro-homogeneity hypothesis is rejected. Here we propose a direct test of H_0 based on the OLS estimate of $\boldsymbol{\eta}_\beta$, namely

$$\widehat{\boldsymbol{\eta}}_\beta = \widehat{\mathbf{b}} - \frac{1}{m}\sum_{i=1}^{m} \widehat{\boldsymbol{\beta}}_i. \tag{7.2.6}$$

Under H_0, $\widehat{\boldsymbol{\eta}}_\beta$ is given by

$$\widehat{\boldsymbol{\eta}}_\beta = \sum_{i=1}^{m} \mathbf{P}_i \mathbf{u}_i, \tag{7.2.7}$$

where

$$\mathbf{P}_i = (\mathbf{X}_a{}'\mathbf{X}_a)^{-1}\mathbf{X}_a{}' - \frac{1}{m} - (\mathbf{X}_i{}'\mathbf{X}_i)^{-1}\mathbf{X}_i{}'. \tag{7.2.8}$$

This suggests basing a test of H_0 on the statistic

$$q_2 = n^{-1}\widehat{\boldsymbol{\eta}}_\beta{}'\widehat{\boldsymbol{\Phi}}_n^{-1}\widehat{\boldsymbol{\eta}}_\beta, \tag{7.2.9}$$

$$\widehat{\boldsymbol{\Phi}}_n = n^{-1}\sum_{i,j=1}^{m} \widehat{\sigma}_{ij}\mathbf{P}_i\mathbf{P}_j{}', \tag{7.2.10}$$

and $\widehat{\sigma}_{ij}$ is a consistent estimator of σ_{ij}.[4] Notice that except for the extreme case where $\mathbf{X}_i = m^{-1}\mathbf{X}_a$, matrix $\widehat{\boldsymbol{\Phi}}_n$ will in general be non-singular.

Theorem 1. *Suppose*

(i) The disaggregate model H_d is correctly specified;

(ii) Assumptions 3–5 hold;

(iii) The matrix $\widehat{\boldsymbol{\Phi}}_n$ defined by (7.2.10) and the matrix $n^{-1}(\mathbf{P}_i\mathbf{P}_i{}')$ both are non-singular and also converge in probability to non-singular matrices.

Then on the hypothesis of no aggregation bias, H_0, the statistic q_2 defined in (7.2.9) is asymptotically distributed as a chi-squared variate with k degrees of freedom.

Proof. See the Mathematical Appendix.

This theorem provides an asymptotic justification for the use of q_2 in testing the null hypothesis of no aggregation bias, and holds for $\sigma_{ij} \neq 0$ and $m \geq 2$, but requires n, the sample size, to be sufficiently large. This contrasts the asymptotic framework underlying the perfect aggregation test proposed in PPK where n is fixed but m is allowed to increase without bounds.

The test statistics q_1 and q_2 are applicable when the focus of the analysis is on all the elements of $\boldsymbol{\beta}_i$. In practice, it is often the case that the parameters of interest are subsets or, more generally, (non-linear) functions of $\boldsymbol{\beta}_i$. To deal with such cases we now consider a generalisation of (7.2.1) and write the null hypothesis of no aggregation bias as

$$\boldsymbol{\eta}_g = \mathbf{g}(\mathbf{b}) - \frac{1}{m}\sum_{i=1}^{m} \mathbf{g}(\boldsymbol{\beta}_i), \tag{7.2.11}$$

[4] In small samples we suggest using the unbiased (and consistent) estimator of σ_{ij} proposed in PPK. (See equation (5.9) in PPK.)

where $\mathbf{g}(\boldsymbol{\beta}_i)$ is an $s \times 1$ $(s \leq k)$ vector of known functions of $\boldsymbol{\beta}_i$.

Denoting the $s \times k$ derivative matrix $\partial \mathbf{g}(\boldsymbol{\beta}_i)/\partial \boldsymbol{\beta}_i{}'$ by $\mathbf{G}(\boldsymbol{\beta}_i)$ and assuming that $\text{rank}[\mathbf{G}(\boldsymbol{\beta}_i)] = s$, the relevant statistic for the test of $\boldsymbol{\eta}_g = \mathbf{0}$ when $\mathbf{b}$ is set *a priori* is given by

$$q_1^* = \left[\mathbf{g}(\mathbf{b}) - \frac{1}{m}\sum_{i=1}^{m} \mathbf{g}(\widehat{\boldsymbol{\beta}}_i)\right]' \widehat{\boldsymbol{\Omega}}_n^{-1} \left[\mathbf{g}(\mathbf{b}) - \frac{1}{m}\sum_{i=1}^{m} \mathbf{g}(\widehat{\boldsymbol{\beta}}_i)\right] \tag{7.2.12}$$

where $\widehat{\boldsymbol{\Omega}}_n$ is now defined by

$$\widehat{\boldsymbol{\Omega}}_n = \frac{1}{m^2}\sum_{i,j=1}^{m} \widehat{\mathbf{G}}_i \widehat{\text{Cov}}(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\beta}}_j) \widehat{\mathbf{G}}_j{}', \tag{7.2.13}$$

and $\widehat{\mathbf{G}}_i = \mathbf{G}(\widehat{\boldsymbol{\beta}})$. (The expression for $\text{Cov}(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\beta}}_j)$ is given in footnote 3.) Then on the null hypothesis of $\boldsymbol{\eta}_g = \mathbf{0}$ (with $\mathbf{b}$ set *a priori*), $q_1^* \overset{a}{\sim} \chi_s^2$.

Turning to the case where $\mathbf{b}$ is defined by (7.2.3), Theorem 1 continues to hold with this difference that the appropriate statistic is now given by

$$q_2^* = n^{-1} \widehat{\boldsymbol{\eta}}_g' \widehat{\boldsymbol{\Phi}}_n^{-1} \widehat{\boldsymbol{\eta}}_g \overset{a}{\sim} \chi_s^2, \tag{7.2.14}$$

where

$$\widehat{\boldsymbol{\eta}}_g = \mathbf{g}(\widehat{\mathbf{b}}) - \frac{1}{m}\sum_{i=1}^{m} \mathbf{g}(\widehat{\boldsymbol{\beta}}_i), \tag{7.2.15}$$

and $\widehat{\boldsymbol{\Phi}}_n$ is defined by (7.2.10), although in this more general case $\mathbf{P}_i$ is now given by

$$\mathbf{P}_i = \widehat{\mathbf{G}}_a (\mathbf{X}_a{}'\mathbf{X}_a)^{-1}\mathbf{X}_a{}' - m^{-1}\widehat{\mathbf{G}}_i (\mathbf{X}_i{}'\mathbf{X}_i)^{-1}\mathbf{X}_i{}', \tag{7.2.16}$$

in which $\widehat{\mathbf{G}}_a = \mathbf{G}(\widehat{\mathbf{b}})$ and $\widehat{\mathbf{g}}_i = \mathbf{g}(\widehat{\boldsymbol{\beta}}_i)$. Notice, also that under $\boldsymbol{\eta}_g = \mathbf{0}$, the asymptotic distribution of q_2^* will be a chi-squared with $s(\leq k)$ degrees of freedom. The statistics q_1^* and q_2^* are direct generalisations of q_1 and q_2 and will reduce to them in the case where $\mathbf{g}(\boldsymbol{\beta}_i) = \boldsymbol{\beta}_i$.

So far, we have limited attention to aggregation bias of the type discussed by Theil (1954) where the bias is defined in terms of the deviations of macro parameters from the simple average of the corresponding micro parameters, as in (7.2.1). It is possible that in some circumstances the macro parameters of interest are derived from the micro parameters via a more general function than the average expression $(1/m)\sum \mathbf{g}(\boldsymbol{\beta}_i)$. An obvious example is when the macro parameters are defined as weighted averages of the corresponding micro parameters. To deal with this and other more complicated averaging schemes, we adopt a generalisation of (7.2.11) and consider aggregation bias defined as

$$\boldsymbol{\eta}_h = \mathbf{g}(\mathbf{b}) - \mathbf{h}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) \tag{7.2.17}$$

where $\mathbf{h}(\mathbf{b}, \ldots, \mathbf{b}) = \mathbf{g}(\mathbf{b})$. As before, aggregation bias is zero under the micro homogeneity hypothesis, H_β, but zero aggregation bias (i.e. $\boldsymbol{\eta}_h = \mathbf{0}$) does not necessarily imply H_β.

The relevant statistics for the test of $\boldsymbol{\eta}_h = \mathbf{0}$ are given by

$$q_1^* = \left[\mathbf{g}(\mathbf{b}) - \mathbf{h}(\widehat{\boldsymbol{\beta}})\right]' \widehat{\boldsymbol{\Omega}}_n^{-1} \left[\mathbf{g}(\mathbf{b}) - \mathbf{h}(\widehat{\boldsymbol{\beta}})\right] \tag{7.2.12$'$}$$

and

$$q_2^* = \left[\mathbf{g}(\widehat{\mathbf{b}}) - \mathbf{h}(\widehat{\boldsymbol{\beta}})\right]' \widehat{\boldsymbol{\Phi}}_n^{-1} \left[\mathbf{g}(\widehat{\mathbf{b}}) - \mathbf{h}(\widehat{\boldsymbol{\beta}})\right] \tag{7.2.13$'$}$$

where $\mathbf{h}(\boldsymbol{\beta}) = \mathbf{h}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_m)$. The covariance matrices ($\boldsymbol{\Phi}_n$ and $\boldsymbol{\Omega}_n$ have the same form as before and are given by (7.2.10) and (7.2.13) respectively, with the difference that the matrix $\widehat{\mathbf{G}}_i$ in (7.2.13) need now be replaced by $\widehat{\mathbf{H}}_i = \partial \mathbf{h}(\widehat{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}$, and the matrix $\mathbf{P}_i$ by

$$\mathbf{P}_i = \widehat{\mathbf{G}}_n(\mathbf{X}_a{}'\mathbf{X}_a)^{-1}\mathbf{X}_a{}' - \widehat{\mathbf{H}}_i(\mathbf{X}_i{}'\mathbf{X}_i)^{-1}\mathbf{X}_i{}'. \tag{7.2.18}$$

Once more, $q_1^* \overset{a}{\sim} \chi_s^2$ under the null hypothesis that $\boldsymbol{\eta}_h = \mathbf{0}$, and is the appropriate statistic where $\mathbf{g}(\mathbf{b})$ is given *a priori*. The statistic q_2^* is relevant when $\mathbf{g}(\mathbf{b})$ is estimated from an aggregate equation and is also asymptotically distributed as χ_s^2 under the null hypothesis.

In the application of the above tests to cases where the general functions $\mathbf{g}(\boldsymbol{\beta}_i)$ or $\mathbf{h}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m)$ are non-linear in the parameters, special care needs to be exercised in the way the nonlinear restrictions $\boldsymbol{\eta}_g = \mathbf{0}$ or $\boldsymbol{\eta}_h = \mathbf{0}$ are formulated. As has been discussed in the recent literature,[5] when the Wald statistic is used for testing nonlinear restrictions, the value of the test statistic depends on the form of the nonlinear restrictions used in the formulation of the null hypothesis. Although asymptotically this does not matter, in finite samples it is possible to obtain very different values for the Wald statistic by parameterising the hypothesis to be tested in different ways. A simple example which is directly relevant to the empirical application that follows in Section 7.4 helps clarify some of these points. Suppose, for example that we are interested in testing the hypothesis for a single sector i that the long run elasticity of y_{it} with respect to x_{it} in the simple model

$$\log y_{it} = \beta_{i0} + \beta_{i1} \log y_{it-1} + \beta_{i2} \log x_{it} + u_{it}, \tag{7.2.19}$$

is equal to, say c_i. A usual way of formulating this hypothesis is by means of the nonlinear restriction

$$d_1(\boldsymbol{\beta}_1) = \beta_{i2}/(1 - \beta_{i1}) - c_i = 0. \tag{7.2.20}$$

This is not, however, the only way that the hypothesis can be formulated. An alternative and in many ways much more satisfactory formulation of this hypothesis is the linear restriction

$$d_2(\boldsymbol{\beta}_1) = \beta_{i2} + c_1\beta_{i1} - c_i = 0. \tag{7.2.21}$$

Although the Wald tests of (7.2.20) and (7.2.21) are equivalent asymptotically, in small samples, depending on how different $\widehat{c}_i = \widehat{\beta}_{12}/(1 - \widehat{\beta}_{i1})$ is from c_i, they can lead to very different results. In this particular example, the linearity of the restriction (7.2.21) recommends it over the nonlinear formulation (7.2.20),[6] but in general, the choice between alternative parameterisations of nonlinear restrictions is not a straightforward matter.

Similar considerations also apply to our Wald tests of the aggregation bias. Suppose we are interested in testing the hypothesis that the macro long run elasticity of $Y_t = \sum \log y_{it}$ with respect to $X_t = \sum \log x_{it}$ is equal to, say, c. When the micro homogeneity hypothesis does not hold, there is no unique method of defining the macro long run elasticity in terms of the micro parameters β_{i1} and β_{i2}. Here we consider two possible

[5] See, for example, Gregory and Veall (1985, 1987), Lafontaine and White (1986), Breusch and Schmidt (1985).

[6] Specifically, in calculating the Wald statistic in the two cases, $\mathrm{Var}(d_2)$ involves the known hypothesised value of c_i, while $\mathrm{Var}(\widehat{d}_1)$ involves $\widehat{c}_i$, and so becomes less reliable under H_0 as $\widehat{c}_i$ deviates from c_i.

methods, the first of which is based on the average of the micro long run elasticities $g(\boldsymbol{\beta}_i) = \beta_{12}/(1 - \beta_{i1}$, namely

$$\epsilon_x^1 = \frac{1}{m}\sum_{i=1}^{m} g(\boldsymbol{\beta}_i) = \frac{1}{m}\sum_{i=1}^{m} \frac{\beta_{i2}}{1-\beta_{i1}},$$

and the second of which is based on the averages of the micro parameters, namely

$$\epsilon_x^2 = \mathbf{h}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) = \frac{1}{m}\sum_{i=1}^{m} \beta_{i2}\left(1 - \frac{1}{m}\sum_{i=1}^{m}\beta_{i1}\right)^{-1}.$$

Depending on which of these two definitions are adopted, the null hypothesis of interest can be written as

$$\eta_g = c - \epsilon_x^1 = 0, \tag{7.2.22}$$

or

$$\eta_h = c - \epsilon_x^2 = 0. \tag{7.2.23}$$

As they stand both restrictions are non-linear in the micro parameters and the application of the Wald test to them will be subject to the type of small sample problems emphasised by Gregory and Veall (1985, 1987). Notice, however, that restriction (7.2.23) has the advantage that it can be written in linear form:

$$\eta_h' = c - \frac{1}{m}\sum_{i=1}^{m}\beta_{i2} + \frac{1}{m}\sum_{i=1}^{m} c\beta_{i1} = 0, \tag{7.2.24}$$

which is the appropriate form to use in the application of the Wald test. Unfortunately, in general the same is not true of the nonlinear restriction (7.2.22). The significance of these issues will be illustrated in Section 7.4.

7.3 A Misspecification Test of the Disaggregate Model

The tests of aggregation bias advanced above are based on the assumption that the disaggregate model H_d is correctly specified. In particular the tests based on the q_2 and q_2^* statistics assume that estimating the macro-parameters directly from the regression of $\mathbf{y}_a$ on $\mathbf{X}_a$, or indirectly by utilising the expression $\sum_{i=1}^{m} \mathbf{C}_i\boldsymbol{\beta}_i$ should not make any difference asymptotically, in the sense that both give consistent estimators of $\mathbf{b}$ under H_d. This implication of the disaggregate model can be tested by means of a Durbin-Hausman type misspecification test and suggests basing a test of H_d on the statistic

$$\widehat{\boldsymbol{\eta}}_s = \widehat{\mathbf{b}} - \sum_{i=1}^{m} \widehat{\mathbf{C}}_i\widehat{\boldsymbol{\beta}}_i, \tag{7.3.1}$$

where $\widehat{\mathbf{C}}_i$ represents a consistent estimator of $\mathbf{C}_i$ defined by (7.2.4).[7] Using the least squares estimates $\widehat{\mathbf{C}}_i = (\mathbf{X}_a{}'\mathbf{X}_a)^{-1}\mathbf{X}_a{}'\mathbf{X}_i$, $(i = 1, 2, \ldots, m)$, we have

$$\widehat{\boldsymbol{\eta}}_s = (\mathbf{X}_a{}'\mathbf{X}_a)^{-1}\mathbf{X}_a{}'\mathbf{e}_d, \tag{7.3.2}$$

[7] See Durbin (1954) and Hausman (1978). Also see Ruud (1984), and Pesaran and Smith (1989) for a unified treatment of misspecification tests in the context of simultaneous equation models.

where

$$\mathbf{e}_d = \sum_{i=1}^{m}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_i) = \sum_{i=1}^{m}\mathbf{M}_i\mathbf{y}_i, \tag{7.3.3}$$

and

$$\mathbf{M}_i = \mathbf{I}_n - \mathbf{X}_i(\mathbf{X}_i{}'\mathbf{X}_i)^{-1}\mathbf{X}_i{}'.$$

Since $(\mathbf{X}_a{}'\mathbf{X}_a)$ is by assumption a non-singular matrix, a test based on $\widehat{\boldsymbol{\eta}}_s$ and $\mathbf{X}_a{}'\mathbf{e}_d$ will be equivalent and for simplicity we use the latter. Suppose now $\mathbf{X}_a$ and $\mathbf{X}_i$ have p variables in common and write[8]

$$\mathbf{X}_a = (\mathbf{X}_{a1}|\mathbf{X}_{a2}); \quad \mathbf{X}_i = (\mathbf{X}_{i1}|\mathbf{X}_{i2}), \quad \text{for all } i,$$

where the $n \times p$ matrix $\mathbf{X}_{a1}$ contains the observations on the common set of variables. It is now easily seen that

$$\mathbf{X}_a{}'\mathbf{e}_d[\underset{p\times 1}{\mathbf{0}} : \underset{(k-p)\times 1}{\mathbf{X}_{a2}{}'\mathbf{e}_d}],$$

and the appropriate statistics on which to base the misspecification test are the non-zero components of $\mathbf{X}_a{}'\mathbf{e}_d$, namely $\mathbf{X}_{a2}{}'\mathbf{e}_d$. Under H_d, we have

$$\mathbf{X}_{a2}{}'\mathbf{e}_d = \sum_{i=1}^{m}\mathbf{X}_{a2}{}'\mathbf{M}_i\mathbf{u}_i, \tag{7.3.4}$$

which suggests the following theorem.

Theorem 2. *Suppose*

(i) Assumptions 3–5 hold;

(ii) The matrices $n^{-1}(\mathbf{X}_{a2}{}'\mathbf{M}_i\mathbf{X}_{a2})$ are non-singular in finite samples, and also converge in probability to non-singular matrices;

(iii) The matrix

$$\widehat{\mathrm{V}}_n = n^{-1}\sum_{i,j=1}^{m}\widehat{\sigma}_{ij}(\mathbf{X}_{a2}{}'\mathbf{M}_i\mathbf{M}_j\mathbf{X}_{a2}), \tag{7.3.5}$$

is non-singular for a finite n, and converges in probability to the non-singular matrix, $\mathbf{V}$.

Then on the hypothesis that the disaggregate model is correctly specified the test statistic

$$q_3 = n^{-1}\mathbf{e}_d{}'\mathbf{X}_{a2}\widehat{\mathbf{V}}_n^{-1}\mathbf{X}_{a2}{}'\mathbf{e}_d, \tag{7.3.6}$$

is asymptotically distributed as a χ^2 variate with $k-p$ degrees of freedom.

Proof. See the Mathematical Appendix.

This theorem complements Theorem 1 and in a sense precedes it. Since Theorem 1 assumes the validity of the disaggregate specification, it is important that the misspecification test of Theorem 2 is carried out before testing for aggregation bias. It is also worth noting that since in general $\sum_{i=1}^{m}\widehat{\mathbf{C}}_i\widehat{\boldsymbol{\beta}}_i$ is not necessarily a more efficient estimator of $\mathbf{b} = \sum_{i=1}^{m}\mathbf{C}_i\boldsymbol{\beta}_i$ than $\widehat{\mathbf{b}}$, the familiar Hausman formula for the covariance of $\widehat{\boldsymbol{\eta}}_s$, namely

[8] Examples of such variables include the intercept term, time trends and seasonal dummies.

$\mathrm{Cov}(\sum_{i=1}^m \widehat{\mathbf{C}}_i \widehat{\boldsymbol{\beta}}_i)$ is not valid. However, when $\boldsymbol{\beta}_i$ are estimated by the SURE method, the resultant estimators, say $\widetilde{\boldsymbol{\beta}}_i$ will be efficient and the covariance difference formula

$$\mathrm{Cov}(\widetilde{\boldsymbol{\eta}}_s) = \mathrm{Cov}(\widehat{\mathbf{b}}) - \mathrm{Cov}\left(\sum_{i=1}^m \widetilde{\mathbf{C}}_i \widetilde{\boldsymbol{\beta}}_i\right) \geq 0,$$

applies. But even in this case to avoid some of the computational problems that arise because of the possible singularity of $\mathrm{Cov}(\widehat{\mathbf{b}}) - \mathrm{Cov}(\sum_{i=1}^m \widetilde{\mathbf{C}}_i \widetilde{\boldsymbol{\beta}}_i)$, a direct derivation of the variance of $\widetilde{\boldsymbol{\eta}}_s$, along the above lines seems to be more desirable.

7.4 An Application

In this section we apply the tests developed in this paper to the annual estimates of aggegate and disaggregate employment demand functions for the U.K. economy presented in PPK. The general log-linear dynamic specification used in the analysis is as follows

$$\begin{aligned} LE_{it} &= \beta_{i1}/m + \beta_{i2}(T_t/m) + \beta_{i3} LE_{i,t-1} + \beta_{i4} LE_{i,t-2} + \beta_{i5} LY_{it} \\ &\quad + \beta_{i6} LY_{i,t-1} + \beta_{i7} LW_{it} + \beta_{i8} LW_{i,t-1} + \beta_{i9} \overline{LY}_{at} \\ &\quad + \beta_{i10} \overline{LY}_{a,t-1} + u_{it}, \quad i = 1, 2, 3, 5, 6, \ldots, 41 \quad t = 1956, 1957, \ldots, 1984, \end{aligned} \tag{7.4.1}$$

where

LE_{it} = log of man-hours employed in sector i at time t;
T_t = time trend ($T_{1980} = 0$);
LY_{it} = log of sector i output at time t;
LW_{it} = log of average product real wage rate per man-hours employed in sector i at time t;
$\overline{LY}_{at}$ = average of LY_{it} over the 40 sectors;
m = number of sectors, ($m = 40$).

The data cover the whole of the private sector, excluding the Mineral Oil and Natural Gas sector (sector 4) for which the sample size is too short to permit estimation. The rationale behind the above disaggregate model and full details of sources and definitions can be found in PPK.

In order to check the overall validity of the disaggregate specification we first computed the Durbin-Hausman type misspecification test statistic given by (7.3.6) in Section 7.3. We obtained a value of 15.9 for this statistic which is distributed as $\chi^2(7)$; this result is just significant at the 5% level and indicates that the disaggregate model may be misspecified. The specification of the disaggregate model requires further consideration, and the following results therefore need to be treated with some caution.

For the purposes of this paper the parameters of interest from the disaggregate model (7.4.1) are the long run elasticities with respect to wages and output given respectively by:[9]

$$\epsilon_{iw} = \frac{\beta_{i7} + \beta_{i8}}{1 - \beta_{i3} - \beta_{i4}}, \quad \text{and} \quad \epsilon_{iy} = \frac{\beta_{i5} + \beta_{i6} + \beta_{i9} + \beta_{i10}}{1 - \beta_{i3} - \beta_{i4}}. \tag{7.4.2}$$

[9] The formula for the output elasticity allows for the long run effect of the sectoral changes on employment of the ith sector both directly through the terms LY_{it} and $LY_{i,t-1}$, and indirectly through the aggregate output effects $\overline{LY}_{at}$ and $\overline{LY}_{a,t-1}$.

Table 7.1: Long Run Elasticities from Restricted Employment Equations* (1956–84)

Industrial sector	Wage		Output	
1 Agriculture, forestry and fishing	−0.8981	(0.2679)	0.0581	(0.2904)
2 Coal Mining	−1.9336	(1.3890)	−1.3866	(0.4410)
3 Coke	−0.3005	(0.0418)	1.6778	(0.1438)
4 Mineral Oil and Natural Gas	—	—	—	—
5 Petroleum products	−0.6530	(0.2947)	0.7560	(0.3552)
6 Electricity etc.	−0.5379	(0.4028)	0.5015	(0.4090)
7 Public gas supply	−0.2594	(0.1128)	1.0311	(0.5060)
8 Water supply	0.0	—	0.7899	(0.7082)
9 Minerals and Ores nes.	−0.4870	(0.2788)	−0.8741	(0.6483)
10 Iron and steel	−0.7712	(0.2483)	2.5657	(0.8473)
11 Non-ferrous metals	0.0	—	1.9619	(1.2480)
12 Non-metallic mineral products	−1.4832	(0.3596)	2.6847	(1.1330)
13 Chemicals and manmade fibres	−0.7405	(0.2146)	1.5938	(0.4538)
14 Metal goods nes.	−0.3976	(0.1987)	1.0368	(0.2219)
15 Mechanical engineering	−0.8587	(0.2457)	1.2415	(0.3497)
16 Office machinery etc.	−1.5343	(2.1250)	0.0	—
17 Electrical engineering	−0.0495	(0.6733)	1.0277	(0.8828)
18 Motor vehicles	−0.7238	(0.4988)	2.7303	(1.6720)
19 Aerospace equipment	−0.1763	(0.1042)	0.1031	(0.0953)
†20 Ships and other vessels	—	—	—	—
21 Other vehicles	−0.5247	(0.2978)	2.1886	(0.6949)
22 Instrument engineering	−0.5607	(0.3201)	0.7715	(0.3876)
23 Manufactured food	−0.4277	(0.1530)	1.7126	(0.8334)
24 Alcoholic drinks etc.	−0.1302	(0.3108)	1.0793	(0.6737)
†25 Tobacco	—	—	—	—
26 Textiles	−0.8320	(0.2335)	0.9812	(0.2690)
27 Clothing and footwear	−0.8101	(0.1323)	0.9737	(0.1454)
28 Timber and furniture	−0.1700	(0.1214)	0.6627	(0.1240)
29 Paper and board	−0.3938	(0.0968)	0.9856	(0.2567)
30 Books etc.	0.2074	(0.2758)	0.1818	(0.1955)
3i Rubber and plastic products	−0.5767	(0.3846)	1.2662	(0.5789)
32 Other manufactures	0.0	—	0.5903	(0.2431)
33 Construction	−0.6453	(1.0200)	0.5872	(0.9331)
34 Distribution etc.	−0.6965	(0.3259)	−.01248	(0.1323)
35 Hotels and catering	−0.6602	(0.4499)	1.2205	(0.7663)
36 Rail transport	−0.3735	(0.3632)	2.0843	(0.7845)
37 Other land transport	0.0	—	0.4203	(0.1983)
38 Sea, air and other	−0.2354	(0.1985)	0.5309	(0.4275)
39 Communications	−0.0267	(0.1905)	0.9905	(0.6643)
40 Business services	0.0	—	0.2333	(0.1294)
41 Miscellaneous services	−0.8108	(0.7865)	1.2228	(1.0810)
Mean of long run elasticities	−0.5233		0.9489	
Standard deviation of elasticities	0.4437		0.8867	
Median of long run elasticities	−0.5247		0.9812	

* The estimates reported in this table are based on the results in table 2 of PPK. The bracketed figures are the estimated standard errors.

† These industries are excluded from the analysis. See the text for further explanation.

Table 7.1 presents estimates for these elasticities derived from the set of restricted disaggregate employment equations estimated by PPK (table 2), with asymptotically valid standard errors in parentheses.[10] For two sectors, 20 and 25, the employment equations estimated by PPK do not possess long run solutions so that there are no corresponding elasticities in Table 7.1. These two industries are excluded from the subsequent analysis. For a few sectors, PPK found no significant response with respect to the real wage variable, and in one sector no response was found with respect to the output variable. In these cases the estimates of the long run elasticity in the table are set equal to zero and no standard errors are given. The last three rows of Table 7.1 present the mean, standard deviation and the median of the distribution of the estimates of the elasticities across the sectors. In the case of both sets of elasticity estimates the mean and median are approximately equal showing that the distributions are close to being symmetric. Both of the standard deviations are large highlighting the considerable variation in the employment responses between sectors. This in itself can be viewed as an argument for the use of disaggregated analysis. We now consider the application of the tests of aggregation bias developed in Section 7.2 (namely the q_1^* and the q_2^* tests) to the disaggregated long run elasticities of Table 7.1 and those of the corresponding restricted aggregate equation given by:[11]

$$
\begin{aligned}
LE_{at} &= \underset{(17.177)}{-140.93} + \underset{(0.0424)}{0.6935} LE_{a,t-1} + \underset{(0.0488)}{0.4665} LY_{at} - \underset{(0.0387)}{0.3948} LW_{at} + \widehat{u}_{at}, \qquad (7.4.3)\\
&\overline{R}^2 = 0.9954, \quad \widehat{\sigma}^2 = 0.3666, \quad n = 29 \ (1954\text{–}84),\\
&\chi^2_{SC}(1) = 1.28, \quad \chi^2_{FF}(1) = 2.45, \quad \chi^2_{N}(2) = 4.41, \quad \chi^2_{H}(1) = 3.45,
\end{aligned}
$$

where LE_{at}, LW_{at} and LY_{at} are the sums of LE_{it}, LW_{it}, and LY_{it} over the 38 sectors respectively. $\overline{R}^2$ is the adjusted R^2, $\widehat{\sigma}^2$ is the estimated standard error of the regression, and χ^2_{SC}, χ^2_{FF}, χ^2_N and χ^2_H are respectively the chi-squared statistics for residual serial correlation, functional form misspecification, normality, and homoskedasticity of the disturbances.[12] The estimates of the long run real wage and output elasticities based on (7.4.3) are -1.2880 (0.2947) and 1.5221 (0.3386) respectively. The numbers in parentheses are asymptotically valid standard errors. It is clear that these results are consistent with the hypothesis of wage and output elasticities of -1 and $+1$ respectively. The relevant statistic for testing the hypothesis that the average of the disaggregate elasticities of Table 7.1 is equal to unity is given by q_1^*, (7.2.12). In this case $\mathbf{g}(\mathbf{b}) = --1$, $\mathbf{g}(\widehat{\boldsymbol{\beta}}_i) = \widehat{\epsilon}_{iw}$ for the real wage variable and $\mathbf{g}(\mathbf{b}) = 1$, $\mathbf{g}(\widehat{\boldsymbol{\beta}}_i) = \widehat{\epsilon}_{iy}$ for the output variable where the long run elasticities ϵ_{iw} and ϵ_{iy} are already defined in (7.4.2). The hypothesis of a unit average long run output elasticity can not be rejected even at the 10% level, since in this case q_1^* equals 0.104 based on an estimated value for the average disaggregate output elasticity of 0.9489. In contrast, the value of -0.5233 obtained for the average disaggregate wage elasticity is significantly different from -1 with q_1^* taking the value of 19.27 in this case.[13] The estimated q_2^* statistics reinforce the finding that the aggregate and disaggregate results differ significantly. The values of this statistic for the wage and output elasticities

[10] Details of the exclusion restrictions imposed for each sector and the diagnostic statistics computed for each equation can also be found in PPK.

[11] This result corresponds to equation (7.4) in PPK, reestimated to exclude sectors 20 and 25.

[12] See Pesaran and Pesaran (1987b) for further details and the relevant algorithms.

[13] In the present application, q_1^* and q_2^* are both distributed asymptotically as χ^2_1 under the null hypothesis.

are 10.95 and 4.97 respectively, rejecting the null hypothesis of no aggregation bias in both cases.[14]

We also considered the alternative aggregate restrictions involving the responsiveness of employment to real wage and output changes corresponding to (7.2.23) in the simple model of Section7.2. As noted there, these restrictions can also be written in a linear form as in (7.2.24) and the q_1^* statistic given by (7.2.12)$'$ was computed here using both linear and nonlinear forms of the restrictions. This allows us to examine the practical importance of the issue of parameterisation of the nonlinear restrictions in the case of the Wald tests discussed in Section 7.2. For wage responsiveness, the values of the q_1^* statistic were 166.88 and 66.01 for the restriction forms (7.2.23) and (7.2.24) respectively, both forms of the test massively rejecting the null hypothesis of a long run real wage elasticity of minus unity. For output responsiveness, the values of the q_1^* statistic were 0.129 and 0.126 respectively, so that the null hypothesis of a unit long run output elasticity cannot be rejected in the case of either formulation.[15] Clearly, the alternative parameterisations considered here have a considerable effect on the value of the statistic obtained for the wage restriction although the results are unaffected qualitatively.

In conclusion, our estimates of the disaggregate labour demand relationships show that there is considerable variation across sectors so that important information may be lost in working with aggregate figures. This is confirmed by the application of the tests developed in the paper. Significant aggregation biases are found in the estimates of a variety of measures of the responsiveness of employment to real wage and output changes based on our aggregate and disaggregate employment equations. The problem of aggregation bias seems, however, to be much more serious for the estimates of the long run real wage elasticity as compared to the estimates of the long run output elasticity.

.1 Mathematical Appendix

Proof of Theorem 1. Under H_0 defined by (7.2.5), the statistic q_2 in (7.2.7) can be written as

$$q_2 = \mathbf{d}_n{}'\mathbf{d}_n, \tag{.1.1}$$

where

$$\mathbf{d}_n = \sum_{i=1}^{m} \mathbf{z}_{in}, \tag{.1.2}$$

and

$$\mathbf{z}_{in} = n^{-\frac{1}{2}} \widehat{\mathbf{\Phi}}_n^{-1} \mathbf{P}_i \mathbf{u}_i, \quad i = 1, 2, \ldots, m. \tag{.1.3}$$

The matrices $\widehat{\mathbf{\Phi}}_n$ and $\mathbf{P}_i$ are defined by (7.2.10) and (7.2.8) in the text, respectively. The proof we offer here has two stages: we first show that for each i and for any real $k \times 1$ vector $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$, $\boldsymbol{\lambda}'\mathbf{z}_{in} \overset{a}{\sim} \mathrm{N}(0, \phi_{ii})$ where $\phi_{ii} > 0$. Using this result in (.1.2), we then show that $\boldsymbol{\lambda}'\mathbf{d}_n \overset{a}{\sim} \boldsymbol{\lambda}'\mathbf{d}$, where $\mathbf{d} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_k)$. From this it follows that $\mathbf{d}_n \overset{a}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{I}_k)$, and $\mathbf{d}_n{}'\mathbf{d}_n \overset{a}{\sim} \chi_k^2$. See proposition 5.1 in White (1984).

[14] The q_2^* statistics are calculated using (7.2.14) replacing $\mathbf{g}(\widehat{\mathbf{b}})$ and $\mathbf{g}(\widehat{\boldsymbol{\beta}}_i)$ by their corresponding aggregate and disaggregate long run elasticity estimates.

[15] The estimates of the aggregate long run elasticities of output and real wages underlying the restriction form (7.2.23) are 0.9770 and $-0 - 4551$ respectively, as compared to the estimates 0.9909 and -0.7850 underlying the restriction form (7.2.24).

Under Assumptions 3–5 it readily follows that

$$\underset{n\to\infty}{\mathrm{plim}}(\widehat{\sigma}_{ij}) = \sigma_{ij}, \quad \underset{n\to\infty}{\mathrm{plim}}(\mathbf{\Phi}_n) = \mathbf{\Phi},$$

where

$$\mathbf{\Phi} = \sum_{i,j=1}^{m} \sigma_{ij}\mathbf{Q}_{ij},$$

and the matrices $\mathbf{Q}_{ij}$ defined by

$$\begin{aligned}\mathbf{Q}_{ij} &= \underset{n\to\infty}{\mathrm{plim}}(n^{-1}\mathbf{P}_i\mathbf{P}_j{}') \\ &= \mathbf{\Sigma}_{aa} - \frac{1}{m}\mathbf{\Sigma}_{aa}^{-1}\mathbf{\Sigma}_{aj}\mathbf{\Sigma}_{jj}^{-1} - \frac{1}{m}\mathbf{\Sigma}_{ii}^{-1}\mathbf{\Sigma}_{ia}\mathbf{\Sigma}_{aa}^{-1} + \frac{1}{m^2}\mathbf{\Sigma}_{ii}^{-1}\mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1},\end{aligned}$$

are finite for all i and j and are non-singular for $i = j$. Now noting that by assumption $\mathbf{\Phi}$ is also non-singular we have

$$\boldsymbol{\lambda}'\mathbf{z}_{in} \overset{a}{\sim} n^{-\frac{1}{2}}\boldsymbol{\mu}'\mathbf{P}_i\mathbf{u}_i = n^{-\frac{1}{2}}\sum_{t=1}^{n}\delta_{it}u_{it}, \tag{.1.4}$$

where $\boldsymbol{\mu} = \mathbf{\Phi}^{-1}\boldsymbol{\lambda}$, and δ_{it} stands for a typical element of vector $\mathbf{P}_i{}'\boldsymbol{\mu}$. It is now easily seen that under assumptions of the theorem, the conditions for the application of the version of Liapounov's Theorem cited in (White, 1984, theorem 5.10) to the right hand side of (.1.4), which is a sum of independently, but non-identically distributed random variables, are met and

$$\boldsymbol{\lambda}'\mathbf{z}_{in} \overset{a}{\sim} \mathrm{N}(0, \phi_{ii}), \quad \text{where} \quad \phi_{ii} = \sigma_{ii}\boldsymbol{\mu}'\mathbf{Q}_{ii}\boldsymbol{\mu} > 0.$$

Therefore, asymptotically $\boldsymbol{\lambda}'\mathbf{d}_n = \sum_{i=1}^{m}\boldsymbol{\lambda}'\mathbf{z}_{in}$ is distributed as a linear function of m normal variates and itself will be distributed normally with zero mean and variance[16]

$$\lim_{n\to\infty}\mathrm{V}(\boldsymbol{\lambda}'\mathbf{d}_n) = \boldsymbol{\lambda}'\boldsymbol{\lambda} = 1.$$

Hence, for a finite m, $\mathbf{d}_n \overset{a}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{I}_k)$, and $\mathbf{d}_n{}'\mathbf{d}_n \overset{a}{\sim} \chi^2_k$. **Q.E.D.**

Proof of Theorem 2. The proof is similar to that presented for Theorem 1. Under H_d the statistic q_3 defined by (7.3.6) can be written as $q_3 = \mathbf{d}_n{}'\mathbf{d}_n$ where $\mathbf{d}_n$ is defined by (.1.2), but $\mathbf{z}_{in}$ is now given by $z_{in} = n^{-\frac{1}{2}}\widehat{\mathbf{V}}_n^{-\frac{1}{2}}\mathbf{X}_{a2}{}'\mathbf{M}_i\mathbf{u}_i$. Since by assumption $\widehat{\mathbf{V}}_n$ converges in probability to a non-singular matrix, say $\mathbf{V}$, we also have

$$\boldsymbol{\lambda}'\mathbf{z}_{in} \overset{a}{\sim} n^{-\frac{1}{2}}\boldsymbol{\lambda}\mathbf{V}^{-\frac{1}{2}}\mathbf{X}_{a2}{}'\mathbf{M}_i\mathbf{u}_i = n^{-\frac{1}{2}}\boldsymbol{\mu}'\mathbf{X}_{a2}{}'\mathbf{M}_i\mathbf{u}_i,$$

where $\boldsymbol{\lambda}$ is now a $(k-p)\times 1$ vector of constants such that $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$, and $\boldsymbol{\mu} = \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\lambda}$. Denoting the tth element of $\mathbf{M}_i\mathbf{X}_{a2}\boldsymbol{\mu}$ by η_{it} we now have

$$\boldsymbol{\lambda}'\mathbf{z}_{in} \overset{a}{\sim} n^{-frac12}\sum_{t=1}^{n}\eta_{it}u_{it}, \tag{.1.5}$$

[16] Notice that since $\lim_{n\to\infty}\mathrm{V}(\boldsymbol{\lambda}'\mathbf{z}_{in}\mathbf{z}_{jn}{}'\boldsymbol{\lambda}) = \sigma_{ij}\boldsymbol{\mu}'\mathbf{Q}_{ij}\boldsymbol{\mu} = \phi_{ij}$, then $\lim_{n\to\infty}\mathrm{V}(\boldsymbol{\lambda}'\mathbf{d}_n) = \lim_{n\to\infty}\mathrm{V}(\sum_{i=1}^{m}\boldsymbol{\lambda}'\mathbf{z}_{in}) = \sum_{i,j=1}^{m}\phi_{ij} = \boldsymbol{\mu}'(\sum_{i,j=1}^{m}\sigma_{ij}\mathbf{Q}_{ij})\boldsymbol{\mu} = \boldsymbol{\mu}'\mathbf{\Phi}^{-1}\boldsymbol{\mu} = \boldsymbol{\lambda}'\mathbf{\Phi}^{-\frac{1}{2}}\mathbf{\Phi}\mathbf{\Phi}^{-\frac{1}{2}}\boldsymbol{\lambda} = 1$.

which is a sum of independently, but non-identically distributed random variables. As in the proof of Theorem 1, it is easily seen that under assumptions of Theorem 2, the Liapounov's theorem ((White, 1984, theorem 5.10)) is applicable to (.1.5) and

$$\boldsymbol{\lambda}'\mathbf{z}_{in} \overset{a}{\sim} \mathrm{N}(0, \psi_{ii}),$$

where

$$\psi_{ii} = \boldsymbol{\mu}' \left[\underset{n\to\infty}{\mathrm{plim}} (n^{-1}\mathbf{X}_{a2}{}'\mathbf{M}_i\mathbf{X}_{a2}) \right] \boldsymbol{\mu} > 0.$$

Hence, by a similar reasoning as in the proof of Theorem 1, we have

$$\boldsymbol{\lambda}'\mathbf{d}_n \overset{a}{\sim} \mathrm{N}(0, 1), \quad \text{and} \quad \mathbf{d}_n \overset{a}{\sim} N(bfzero, \mathbf{I}_{k-p}),$$

which establishes that

$$q_3 = \mathbf{d}_n{}'\mathbf{d}_n \overset{a}{\sim} \chi^2_{k-p}.$$

Q.E.D.

Chapter 8

Aggregation Bias in Labour Demand Equations for the UK Economy

Introduction

The responsiveness of employment to changes in real wages is an issue of considerable importance, particularly for policy analysis, and over the past decade a number of studies have been devoted to this issue in the UK. Notable examples include the papers by Nickell (1984), Symons (1985), Wren-Lewis (1986), and Burgess (1988) for the manufacturing sector, and by Beenstock and Warburton (1984), ?Layard and Nickell (1985a) for the private sector and the economy as a whole. In contrast to the earlier work by Godley and Shepherd (1964), Brechling (1965), and Ball and Cyr (1966), these recent studies find a significant and quantitatively important effect for real wages on employment. The point estimates of the long-run wage elasticity obtained in these studies vary widely depending on the coverage of the data (whether the data set used is economy-wide or just manufacturing), and on the specification of the estimated equations. A recent review of these studies by Treasury (1985) concludes that the estimate of long-run wage elasticity most likely falls in the region -0.5 to -1 although, under the influence of Layard and Nickell's important contributions, for the economy as a whole the 'consensus' estimate of this elasticity in the UK currently seems to centre on the figure of -1.[1] All these studies are, however, carried out using highly aggregated data, either at the level of the whole economy or the manufacturing sector, and given the significance of their results for macroeconomic policy it is important that the robustness of their results to the level of aggregation chosen are carefully investigated.

This paper extends the empirical work described in Pesaran et al. (1989b) (PPK), and examines the effect of aggregation on the estimates of long-run wage and output elasticities of demand for employment in the UK. The aggregate and the disaggregate employment functions analysed in this paper differ from those in PPK in two respects. First, the functions allow for a longer lagged effect of output on employment. Second, in order to deal with some of the econometric difficulties associated with the use of the

[0]Published in T. S. Barker and M. H. Pesaran (eds.) *Disaggregation in Econometric Modelling* (1990), Routledge, London, Chapter 6, pp. 113–149. Co-authors K. C. Lee and M. H. Pesaran. We are grateful to Ed Learner and Franco Peracchi for helpful comments and suggestions. Financial support from the ESRC and the Newton Trust is gratefully acknowledged.

[1] This consensus estimate is also the same as the figure obtained by Beenstock and Warburton (1984) for their extended data set.

time trend as a proxy for technical change in estimating the employment functions,[2] the time trend will be replaced by a measure of embodied technological change based on the current and past movements of *gross* investment, à la Kaldor (1957, 1961). This measure of technological change is both statistically less problematic than a simple time trend and more satisfactory from a theoretical standpoint.

The paper also applies the statistical methods recently developed for the analysis of aggregation by PPK and Lee et al. (1990b) (LPP) to employment equations for the UK. Specifically, the aggregation bias in the estimates of the long-run wage and output elasticities will be tested statistically, and the possibility of misspecification of the disaggregate employment equations will be investigated by means of the Durbin-Hausman type test developed in LPP. The adequacy of the aggregate model (relative to the disaggregate specification) will also be investigated by means of the goodness-of-fit criteria and the test of perfect aggregation proposed in PPK.

The plan of the paper is as follows. Section 8.1 sets out the disaggregate employment functions and discusses the theoretical rationale that underlies them. Section 8.2 motivates the use of a distributed lag function in gross investment as a proxy far technological change. Section 8.3 reviews the various statistical methods to be applied. Section 8.4 presents the empirical results, and the final section provides a summary of the main findings of the paper.

8.1 Industrial employment functions: theoretical considerations

In specifying the employment demand functions we follow the literature on derivation of dynamic factor demand models and suppose that the employment decision is made at the industry level by identical cost minimizing firms operating under uncertainty in an environment where adjustment can be costly. We assume that in the absence of uncertainty and adjustment costs the industry's employment function is given by

$$h_t^* = f(w_t, y_t, a_t) + v_t, \tag{8.1.1}$$

where

h_t^* = the desired level of man-hours employment (in logs),
w_t = the real wage rate (in logs),
y_t = the expected level of real demand (in logs),
a_t = an index of technological change,
v_t = mean zero serially uncorrelated productivity shocks.

The actual level of employment, h_t, measured in logarithms of man-hours employed in the industry is then set by solving the following optimisation problem

$$\min_{h_t, h_{t+1}, \ldots} \mathrm{E}\left\{\sum_{\tau=0}^{\infty} \beta^\tau \left[(h_{t+\tau} - h_{t+\tau}^*)^2 + \frac{1}{2}\phi_1(\Delta h_{t+\tau})^2 + \frac{1}{2}\phi_2(\Delta^2 h_{t+\tau})^2\right] |\Omega_t\right\} \tag{8.1.2}$$

[2] The econometric problems involved in the use of time trends in regression equations containing non-stationary variables are discussed, for example, by Mankiw and Shapiro (1985, 1986), and Durlauf and Phillips (1986).

where $\Omega_t = (h_t, h_{t-1}, \ldots, w_t, w_{t-1}, \ldots, y_t, y_{t-1}, \ldots, a_t, a_{t-1} \ldots, u_t, u_{t-1}, \ldots,)$ represents the information set of the firm at time t, Δ is the first difference operator, and $0 \leq \beta < 1$ is the real discount factor. The first term in (8.1.2) measures the cost of being out of equilibrium, and the second and the third terms stand respectively for the costs of changing the level and the speed with which changes in employment are put into effect. The inclusion of the last term in (8.1.2) is proposed in Pesaran (1988b) and generalises the familiar adjustment cost-rational expectations models discussed, for example, by Sargent (1978) and Kennan (1979), and is of some interest as it provides a theoretical justification for the inclusion of h_{t-2} in the employment function.[3] In practice, the speed of adjustment coefficients ϕ_1 and ϕ_2 could vary with the state of the labour market as argued, for example, by Smyth (1984) and Burgess (1988). Here, however, we shall assume that they are fixed. The unique solution to the above optimisation problem is derived in Pesaran (1988b) and is given by

$$h_t = \psi_1 h_{t-1} + \psi_2 h_{t-2} + \sum_{j=0}^{\infty} \theta_j \, \mathrm{E}(h^*_{t+j} | \Omega_t) \tag{8.1.3}$$

where

$$\psi_1 = \mu_1' + \mu_2' > 0, \quad \psi_2 = -\mu_1' \mu_2' < 0, \quad \theta_j = (\mu_1^{-j-1} - \mu_2^{-j-1}) / [\phi_2(\mu_2 - \mu_1)]$$

and μ_1, μ_2, μ_1' and μ_2' are the roots of

$$a_2 x^2 + a_1 x + \lambda_1 x^{-1} + \lambda_2 x^{-2} = 1.$$

The reduced-form parameters a_1, a_2, λ_1 and λ_2 are defined in terms of the structural parameters, β, ϕ_1 and ϕ_2 (see, Pesaran (1988b)). It is important to note that for plausible values of the structural parameters the theory suggests a negative value for the coefficient of h_{t-2} in (8.1.3). Adopting a linear approximation for (8.1.1), and assuming that conditional expectations of w_{t+j}, y_{t+j} and a_{t+j} with respect to Ω_t are formed rationally on the basis of an rth order vector autoregressive (VAR) system, the decision rule (8.1.3) becomes

$$h_t = \text{ intercept } + \psi_1 h_{h-1} + \psi_2 h_{t-2} + \mathbf{c}_{t-1}{}'(L) \mathbf{z}_t + u_t \tag{8.1.4}$$

where $u_t = (1 - \psi_1 - \psi_2)(1 - \psi_1/\beta - \psi_2/\beta^2) v_t$, $\mathbf{z}_t = (a_t, y_t, w_t)'$, and $\mathbf{c}_{t-1}(L) = \sum_{i=1}^{r} c_i L^{i-1}$ is a 3×1 vector of lag polynomials of order $r-1$ in the lag operator L. In the case where the variables y_t, w_t and a_t have univariate AR(r_i), $i = y, w, a$ representations, (8.1.4) simplifies to

$$h_t = \text{ intercept } + \psi_1 h_{h-1} + \psi_2 h_{t-2} + \left(\sum_{i=1}^{r_y} \gamma_{iy} L^{i-1} \right) y_t + \left(\sum_{i=1}^{r_w} \gamma_{iw} L^{i-1} \right) w_t + \left(\sum_{i=1}^{r_a} \gamma_{ia} L^{i-1} \right) a_t + u_t, \tag{8.1.5}$$

which is a generalisation of the aggregate employment function (7.2) in PPK.[4] Under the rational expectations hypothesis (REH), the coefficients c_i in (8.1.4), and γ_{iy}, γ_{iw},

[3] The inclusion of first or higher order lags of h_t in the employment function can also be justified by appeal to aggregation over different types of labour or firms with different adjustment costs (Nickell (1984)).

[4] To derive (7.2) in PPK from (8.1.5), let $r_y = r_w = 2$, and notice that when a simple linear trend is used as a proxy for a_t, then $a_t = a_{t-1} + b$, where b is a fixed constant, and

$$\left(\sum_{i=1}^{r_a - 1} \gamma_{ia} L^{i-1} \right) a_t = \left(\sum_{i=1}^{r_a - 1} \gamma_{ia} \right) a_t - b \sum_{i=1}^{r_a - 1} (i-1) \gamma_{ia} = \gamma_a a_t + \text{ constant}.$$

γ_{ia} in (8.1.5) will be subject to $3r-4$ and $(r_y+r_w+r_a)-4$ cross-equation restrictions, respectively. However, given our concern with the problem of aggregation, in the present study we do not consider imposing these restrictions, and employ instead the unrestricted version of (8.1.5) as our maintained hypothesis.[5] We then choose the orders of the lag polynomials on h_t, y_t, w_t, and a_t empirically. The validity of the RE restrictions at the industry level and the problem of aggregation bias in the context of RE models is beyond the scope of the present paper.

8.2 Modelling and measurement of technological change

In the empirical analysis of labour demand, technological change, broadly defined to include new scientific, engineering, and electronic discoveries and inventions, is generally assumed to occur exogenously, evolving independently of market conditions and government policy interventions. It is inferred either indirectly as a residual using a production function approach, or is represented by linear, piece-wise linear, or non-linear functions of time. Neither procedure is satisfactory. The former approach, employed, for example, by **?**, assumes an *a priori* knowledge of the production possibilities and involves circular reasoning, while the latter is devoid of a satisfactory theoretical rationale and is adopted by most researchers as a 'practical' method of dealing with a very difficult problem (Arrow (1962)).[6]

Ideally, what we need are direct reliable measures of technological change, and there are some data such as expenditure on research and development (R&D) and the number of patents and product designs granted over a given period that can be used. In the absence of suitable direct measures of technological change, here we adopt an indirect approach and following Kaldor (1957, 1961) postulate a distributed lag relationship between the a_t, the technological change index, and the rate of *gross* investment, GI_t,

$$a_t = \text{ intercept } + \sum_{j=0}^{\infty} \lambda_j \log(GI_{t-j}). \tag{8.2.1}$$

A static version of this relationship when used in a linear version of (8.1.1) yields a log linear approximation to Kaldor's 'technical progress function', which relates the rate of change of productivity per worker to the rate of change of gross investment.[7] According to this model technological progress is 'embodied' in the process of capital accumulation and takes place primarily through gross capital formation by the infusion of new equipment and machines, embodying the most up-to-date technology into the economy. The formulation (8.2.1) can also be justified along the lines suggested by Arrow (1962) in his seminal paper on 'learning by doing'. (Arrow, 1962, p. 157) himself uses cumulative gross investment as an index of experience, which is closely related to the distributed lag function in (8.2.1).

The technological progress function (8.2.1) is more than a theoretical postulate. It is also based on direct empirical support. Schmookler (1966) in his pioneering work, using patents as a disaggregate measure of technological change, showed there exist strong

[5] This is similar to the research strategy followed by Nickell (1984) and Burgess (1988).

[6] Notice that the use of time trends in regression equations containing integrated stochastic processes is also subject to important econometric pitfalls and as argued in Mankiw and Shapiro (1985, 1986), and Durlauf and Phillips (1986) can lead to spurious inference.

[7] See in particular (Kaldor and Mirlees, 1962, pp. 176–7). Notice, however, that Kaldor's formulation abstracts from the effect of real wages on labour productivity, while ours does not.

positive correlations between gross investment and patents in railroads, petroleum refining, and building industries over the period 1873–1940. He also obtained similar results using cross-section data. While there is some doubt about the direction of causation in Schmookler's findings, there is little dispute about the existence of a close relationship between gross investment and technological change.[8] Since our aim here is not to explain the causes of technological change but to estimate its impact on employment demand, we feel that the controversy over the causality of the investment-patents relationship has little bearing on our analysis.

The coefficients $\lambda_j, j = 1, 2, \ldots$ measure the impact of past investments on the current state of technological advance, and it is reasonable to assume that they are a decreasing function of the lag length, $j = 1, 2, \ldots$. The likely rate of decline of λ_j depends on the importance of the learning-by-doing component of a_t. Under a pure learning story, $\{\lambda_j\}$ will be fixed or show a very slow rate of decline. The rate of decline of $\{\lambda_j\}$ is likely to be much higher if one adopts Kaldor's idea. Here, for the purpose of empirical analysis we assume the following geometrically declining pattern for λ_j

$$\lambda_j = \alpha(1-\lambda)\lambda^j, \quad j = 0, 1, 2, \ldots \quad \alpha, \lambda > 0$$

and write (8.2.1) as

$$a_t = \text{ intercept } + \alpha d_t(\lambda) \tag{8.2.2}$$

where $d_t(\lambda)$ satisfies the following recursive formula

$$d_t(\lambda) = \lambda d_{t-1}(\lambda) + (1-\lambda)\log(GI_t). \tag{8.2.3}$$

Substituting (8.2.2) in (8.1.5) now yields

$$h_t = \text{ intercept } + \psi_1 h_{t-1} + \psi_2 h_{t-2} + \gamma_y(L)y_t + \gamma_w(L)w_t + \alpha\gamma_a(L)d_t(\lambda) + u_t \tag{8.2.4}$$

where $\gamma_y(L)$, $\gamma_w(L)$, and $\gamma_a(L)$ are lag operator polynomials of orders $r_y - 1$, $r_w - 1$ and $r_a - l$, respectively. It is clear that in general α is not identifiable, although the decay coefficient, λ, can in principle be estimated from the data. We shall return to the issue of the estimation of (8.2.4) in section 8.4, but first we briefly review the econometric issues concerning testing for aggregation bias and the relative predictive performance of aggregate and disaggregate models.

8.3 The aggregation problem: econometric considerations

Suppose that, for a given value of the decay parameter λ, the variables in (8.2.4), namely h_t, y_t, w_t, and $d_t(\lambda)$, are observed over the period $t = 1, 2, \ldots, n$ for each of the m firms (industries), $i = 1, 2, \ldots, m$. Then the disaggregate employment equations can be written in matrix notation as

$$H_d: \quad \mathbf{h}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i, \quad i = 1, 2, \ldots, m \tag{8.3.1}$$

[8] For a review of more recent evidence see, for example Beggs (1984) and Baily and Chakrabarti (1985). Notice, however, that Beggs uses wage expenditures as a surrogate for investment data and his results may not be directly comparable to those obtained by Schmookler. On this see the comments by Schankerman (1984) on Beggs's paper.

where $\mathbf{h}_i$ is the $n \times 1$ vector of observations on the log of manhours employment in the ith firm (industry), $\mathbf{X}_i$ is the $n \times k$ $(k = r_y + r_w + r_a + 3)$ matrix of observations on the regressors in (8.2.4) for the ith firm (industry). $\boldsymbol{\beta}$ is the $k \times 1$ vector of the coefficients associated with columns of $\mathbf{X}_i$, and $\mathbf{u}_i$ is the $n \times 1$ vector of disturbances for the ith firm (industry). The aggregate equation associated with (8.2.4) is given by

$$H_a: \quad \mathbf{h}_a = \mathbf{X}_a \mathbf{b}_a + \mathbf{v} \tag{8.3.2}$$

where

$$\mathbf{h}_a = \sum_{i=1}^{m} \mathbf{h}_i. \quad \mathbf{X}_a = \sum_{i=1}^{m} \mathbf{X}_i$$

and $\mathbf{b}_a$ is the $k \times 1$ vector of macro parameters.

The aggregation problem arises when the disaggregate model (8.3.1) holds but the investigator decides to base his/her analysis on the aggregate specification (8.3.2). The econometric implications of aggregation in linear models have been discussed in the literature in some detail.[9] The principal issues concern the accuracy of predictions and the bias in the parameter estimates. For the analysis of the predictive performance of models (8.3.1) and (8.3.2), PPK propose using a modified version of the Grunfeld and Griliches criterion which compares the sums of squared errors of predicting $\mathbf{h}_a$ using the aggregate and disaggregate models, adjusting for the differences in the degrees of freedom (see section 4 of PPK). They also propose a test of perfect aggregation which tests the hypothesis that

$$\boldsymbol{\xi} = \sum_{i=1}^{m} \mathbf{X}_i \boldsymbol{\beta}_i - \mathbf{X}_a \mathbf{b} = \mathbf{0}.$$

To test for aggregation bias, two approaches are possible. The first is the method employed in Zellner (1962) and involves testing the micro homogeneity hypothesis

$$H_\beta: \quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_m.$$

However, as is pointed out in LPP, as a test of aggregation bias this approach is unduly restrictive. Instead they propose testing the hypothesis of zero aggregation bias directly by comparing an average of the estimates of the micro coefficients, or a function thereof, with the aggregate counterpart. In the case of the present study, the parameters of interest are the long-run output and wage elasticities, which, assuming $r_y = r_w = 2$ in (8.3.1), are given (in terms of the elements of $\boldsymbol{\beta}_i$) for the ith industry by

$$\epsilon_{iy} = \frac{\beta_{i4} - \beta_{i5}}{1 - \beta i2 - \beta_{i3}}$$

and

$$\epsilon_{iw} = \frac{\beta_{i6} + \beta_{i7}}{1 - \beta i2 - \beta_{i3}}$$

respectively.[10] The null hypothesis we wish to test is that aggregation bias is zero, i.e.

$$\eta_g = \mathbf{g}(\mathbf{b}) = -\frac{1}{m} \sum_{i=1}^{m} \mathbf{g}(\boldsymbol{\beta}_i) = \mathbf{0} \tag{8.3.3}$$

[9] See for example Theil (1954), Grunfeld and Griliches (1960), Boot and de Wit (1960), Zellner (1962), Gupta (1971), and Sasaki (1978).

[10] From (8.1.5) note that $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{ik})' = (\text{intercept}, \psi_{i1}, \psi_{i2}, \gamma_{1y}, \gamma_{2y}, \gamma_{1w}, \gamma_{2w}, \gamma_{1a}, \gamma_{2a})$.

where $\mathbf{g}(\boldsymbol{\beta}_i)$ is an $s \times 1$ vector of parameters of interest from the disaggregate model (8.3.1) and $\mathbf{g}(\mathbf{b})$ is the corresponding vector from the aggregate model. In the case of our application

$$\mathbf{g}(\boldsymbol{\beta}_i) = (\epsilon_{iy}, \epsilon_{iw})'. \tag{8.3.4}$$

Following LPP we distinguish two situations: (i) where $\mathbf{g}(\mathbf{b})$ is given *a priori* (for example by a consensus view) and (ii) where $\mathbf{g}(\mathbf{b})$ is estimated from the aggregate model (8.3.2).

Two corresponding statistics are derived

$$q_1^* = \left[\mathbf{g}(\mathbf{b}) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{g}(\widehat{\boldsymbol{\beta}}_i)\right]' \widehat{\boldsymbol{\Omega}}_n^{-1} \left[\mathbf{g}(\mathbf{b}) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{g}(\widehat{\boldsymbol{\beta}}_i)\right] \quad \overset{a}{\sim} \quad \chi_s^2 \tag{8.3.5}$$

and

$$q_2^* = n^{-1}\widehat{\boldsymbol{\eta}}_g{}'\widehat{\boldsymbol{\Phi}}_n^{-1}\widehat{\boldsymbol{\eta}}_g \quad \overset{a}{\sim} \quad \chi_s^2 \tag{8.3.6}$$

where

$$\widehat{\boldsymbol{\eta}}_g = \mathbf{g}(\widehat{\mathbf{b}}) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{g}(\widehat{\boldsymbol{\beta}}_i) \tag{8.3.7}$$

and $\widehat{\boldsymbol{\Omega}}_n$ and $\widehat{\boldsymbol{\Phi}}_n$ are estimated covariance matrices defined in LPP.[11] These tests of aggregation bias assume that the disaggregate model H_d holds and it is important that this assumption is also tested. To this end LPP derive a Durbin-Hausman-type misspecification test which examines the statistical significance of the difference between the estimates of the parameters of the aggregate model based on the disaggregate and aggregate specifications respectively. If this difference turns out to be significant then it is likely that the disaggregate model is misspecified and the aggregation bias tests may be misleading.

8.4 Empirical results

In this section the theoretical considerations on employment functions of sections 8.1 and 8.2 and the statistical methods outlined in section 8.3 are brought together in the estimation of disaggregate and aggregate employment functions for the UK and the analysis of aggregation bias. The data employed are taken from the Cambridge Growth Project Databank, and full details are provided in appendix .1. Figures are available annually for the period 1954–84 and, except for some public sector services, the whole of the UK economy is covered, with data provided on a 41-industry basis. As in PPK, industry 4 (mineral oil and natural gas) is excluded from the analysis, and both the disaggregate and the aggregate specifications are based on the remaining 40 industry groups (i.e. $m = 40$). Although our data set starts in 1954, all the equations are estimated over the period 1956–84, and the data for the years 1954 and 1955 are used to generate the lagged values of employment, output, and real wages that are included in the employment function (see equation 8.2.4). For the technical change variable $d_t(\lambda)$, we employed the recursive formula given by (8.2.3), for $t = 1955, 1956, \ldots, 1984$ and experimented with different methods of initializing the recursive process. We also experimented with different estimates of the decay rate, λ.

[11] It is recognized that the application of the proposed Wald test may be problematic in finite samples when the restriction set is non-linear since the test is not invariant to the parameterisation of the restrictions. See Gregory and Veall (1985) and LPP.

8.4.1 Initialization of the $d_t(\lambda)$ process

We tried two methods for generating the initial value, $d_{1954}(\lambda)$. In one set of experiments we derived $d_{1954}(\lambda)$ assuming that the process generating $\log(GI_t)$ in the pre-1954 period can be characterized by a random walk and that on average $\mathrm{E}[\log(GI_{1954})] = \mathrm{E}[\log(GI_{1953})] = \ldots = \log(\overline{GI})$, where we estimate $\overline{GI}$ by the average of gross investment over the 1954–8 period. Under these assumptions, the estimate of $d_{1954}(\lambda)$, which we denote by $\widehat{d}_{01}$ is given by[12]

$$\widehat{d}_{01} = \log(\overline{GI}). \tag{8.4.1}$$

As an alternative procedure we followed the backward forecasting procedure proposed in Pesaran (1973), and derived the following alternative estimate for $d_{1954}(\lambda)$

$$\widehat{d}_{02} = \left\{ \frac{\widehat{\rho}\lambda}{\widehat{\rho} - (1-\lambda)} \right\} \log(GI_{1954}). \tag{8.4.2}$$

This estimate assumes that in the pre-1954 period $\log(GI_t)$ follows the first-order autoregressive process

$$\log(GI_t) = \rho \log(GI_{t-1}) + \epsilon_t, \quad t = 1954, 1953, \ldots$$

and that ρ can be estimated consistently by the OLS method using data over the period 1954–84.

8.4.2 Estimation of the decay rate parameter, λ

In the initial experiments we assumed a decay rate of $\lambda = 0.10$ and estimated the employment equations under both methods of initializing the $d_t(\lambda)$ process described above. We found that the technological variable, $d_t(\lambda)$ showed significantly in about half of the industries, and of these the majority demonstrated the better fit using $\widehat{d}_{01}$, (i.e. had the larger log likelihood value, LLF) as opposed to $\widehat{d}_{02}$. The difference between LLF obtained in most industries was well below 1, and in only two cases did the difference exceed 2. In both of these $\widehat{d}_{01}$ proved to be the more satisfactory measure. In view of these preliminary results we decided to initialise the $d_t(\lambda)$ process with $\widehat{d}_{01}$. However, we note that, apart from the size of the coefficient on the constant in the estimated equations, there was little qualitative difference between results obtained using either of the two initialization methods.

Using $\widehat{d}_{01}$, we also estimated the industrial employment equations by the grid search method, for values of λ in the range $(0.0, 0.30)$. Again restricting attention to those industries with significant technological change effects, we found for about half of these industries the maximum likelihood estimates of λ fell within this interval, with many of the rest located on the $\lambda = 0.0$ bound. In general, however, we found the results to be

[12] Notice that

$$d_{1954}(\lambda) = \lambda \sum_{i=0}^{\infty} (1-\lambda)^i \log(GI_{1954-i})$$

and under $\mathrm{E}[\log(GI_{1954})] = \mathrm{E}[\log(GI_{1953})] = \ldots = \log(\overline{GI})$, we have

$$d_{01} = \mathrm{E}[d_{1954}(\lambda)] = \log(\overline{GI}).$$

qualitatively robust to the choice of the decay parameter in the range $(0.0, 0.30)$. In the absence of any strong evidence of a more appropriate estimate of λ, therefore, we decided to maintain our original choice of $\lambda = 0.10$ in the remainder of the empirical work.

8.4.3 The estimated equations

The most general set of equations that we considered are presented in Table 8.1. This includes among the explanatory variables two lagged dependent variables, h_{t-1} and h_{t-2}, and current and lagged values of industry output, wages, and technological change (y_t, y_{t-1}, y_{t-2}, w_t, w_{t-1}, d_t, d_{t-1}).[13] This equation follows from the theoretical discussion of sections 8.1 and 8.2, by setting $r_y = 3$ and $r_w = r_a = 2$ in (8.2.4). Also included in the list of explanatory variables are current and lagged aggregate output measures, $\overline{y}_{ta}$ and $\overline{y}_{t-1,a}$ ($\overline{y}_{ta} = \frac{1}{m}\sum_{i=1}^{m} y_{it}$). These variables were shown to be important in the empirical work of PPK, and it is clearly necessary to consider their influence here also. Their inclusion can be justified on the grounds that agents could use this aggregate information in the formation of their conditional expectations of y_{t+j}, w_{t+j}, which we have shown to be important in explaining current employment. This unrestricted model differs from that in PPK by excluding the time trend, and by including y_{t-2}, d_t and d_{t-1}. Replacing the time trend by d_t and d_{t-1} alone caused a serious deterioration in the performance of many of the industrial equations, and in particular many became unstable. The inclusion of a second lagged output term remedied this in most of the equations, however, and Table 8.1 represents a satisfactory set of results. The fit of most of the equations is satisfactory, with $\overline{R}^2$ falling below 0.90 only for industry 5 (Petroleum products). Short-run elasticities of employment with respect to wages, employment, and technological change are generally of the expected sign, although as the standard errors of the coefficients (shown in brackets) indicate, the equations are in many cases overparameterised.

For this reason, a specification search was carried out on these equations to obtain a more parsimonious set of results, and these are presented in Tables 8.2 and 8.3. Coefficients with t-values less than one (in absolute value) were omitted. Some *a priori* incorrectly signed coefficients were also constrained to zero where the constraints were not violated by the data. Specifically, we expect the coefficients on h_{t-2}, and the long-run wage and technological change effects to be negative. The χ^2 statistic for testing the validity of linear restrictions imposed on the parameters of the unrestricted equations to obtain the results of Table 8.2 are given in the second column in Table 8.3. It can be seen that the imposed restrictions are not rejected for any industry, at the conventional levels of significance.

The overall performance of the equations in Table 8.2 is good and in line with those of PPK. Real wages show up significantly (and negatively) in most industries, with no long-run wage effect found only in industries 22, 33, 37, and 40. The output variable also performed well, showing significantly and positively in all but three industries (6, 20, 38), the last one of which shows a strong positive aggregate output influence. Only fifteen of the industries failed to demonstrate any technological change effects, although there are problem industries (10, 11, 26, 36, and 39) for which the technological change variables are (in sum) incorrectly signed. Other industries with *a priori* implausible parameter estimates include 23, 26, and 31, in which an unexpected positive second lagged dependent variable appears, and industry 25 which is unstable. Industry 20 also

[13] For each industry the technological variable $d_t(\lambda)$, or d_t for short, is computed using (8.2.3), with the initial value given by (8.4.1) and the decay parameter, $\lambda = 0.10$.

remains problematic: the differenced form reported in the PPK paper could not be improved upon, and this equation is retained here.

The histograms in figure 8.1–8.3 illustrate the long-run elasticities of employment with respect to industrial output, wages, and technological change as obtained from the results in Table 8.2. Figure 8.1 shows the long-run output elasticities for thirty-eight industries, omitting industries 20 and 25; the two industries with incorrectly signed output effects show to the left of the vertical axis, while industry 21 demonstrates an implausibly high positive output elasticity (the equation for this industry has a coefficient on the

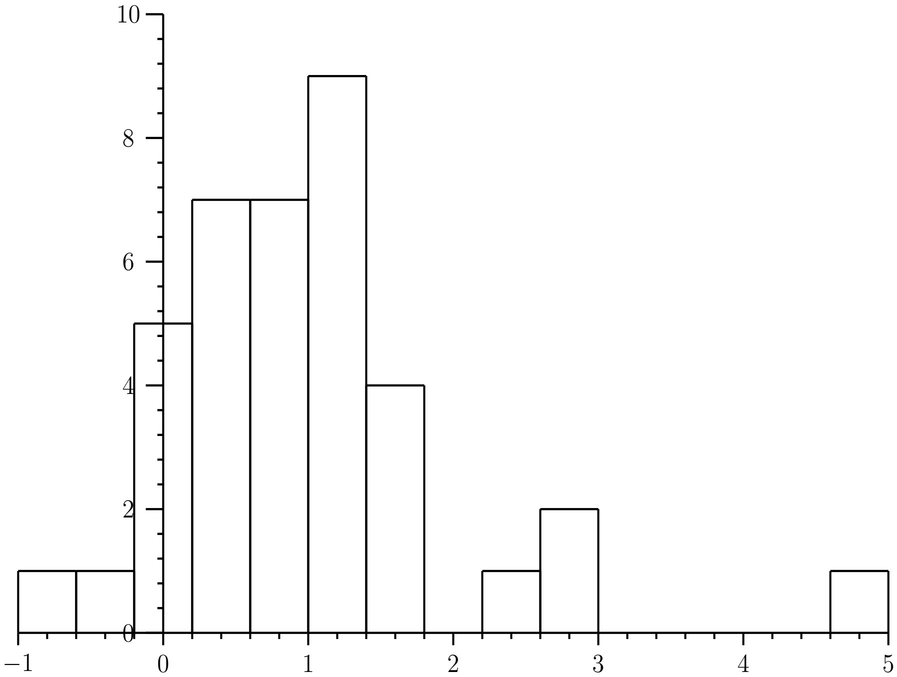

Figure 8.1: Long-run industry output elasticities from Table 8.2.

lagged dependent variable in excess of 0.9). The output elasticities for the bulk of the industries, however, lie within the interval $(0, 1.5)$ and the mean long-run output elasticity is 0.97. The histogram in figure 8.1 provides a clear illustration of the variability in the responsiveness of employment to output changes across industries, and this is confirmed by a standard deviation around the mean of 0.86. Similar observations can be made on the long-run real wage and technological change elasticities, which have means (standard deviations in brackets), $-0.68(0.66)$, and $-0.41(0.81)$ respectively.

The preferred equations set out in Table 8.2 show the technological change variable d_t to be an adequate replacement for the time trend in some industries, but not all. Of the twenty-four industrial equations in which a significant time trend was found in PPK, fifteen are improved upon, in terms of the equation standard errors, by their equivalent estimate in Table 8.2, while nine fit less well in the absence of the time trend. Moreover, there are a further eleven equations which did not previously involve a time trend whose standard error is lower in Table 8.2 than that in PPK, demonstrating the extra explanatory power of the additional lagged output and technological change variables. Since we prefer to replace a time trend with a variable with a more satisfactory theoretical basis, and given that the fit of this new set of equations is generally higher, these results, taken as a whole, can be seen as an improvement over those obtained previously.

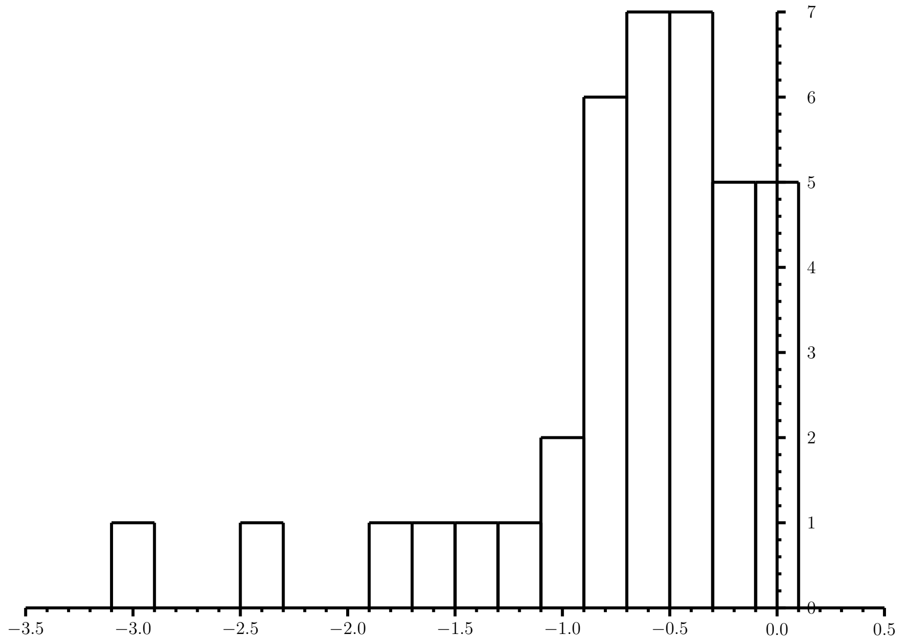

Figure 8.2: Long-run industry real wage elasticities from Table 8.2.

Table 8.1: Unrestricted industrial labour demand equations

	inpt/40	y_t	y_{t-1}	y_{t-2}	h_{t-1}	h_{t-2}	w_t	w_{t-1}	$\overline{y}_{ta}$	$\overline{y}_{(t-1)a}$	d_t	d_{t-1}	$\overline{R}^2$ (LLF)	$\widehat{\sigma}$
1 Agriculture,etc.	−8.3291	0.3996	0.3135	−0.2215	0.5188	0.0008	−0.5179	0.0399	−0.1477	−0.1409	0.4305	0.3869	0.9982	0.0141
	(101.8650)	(0.1641)	(0.1996)	(0.1442)	(0.2661)	(0.1542)	0.1006	0.1531	(0.1740)	(0.1714)	(0.4127)	(0.3553)	(90.1369)	
2 Coal mining	−99.6098	0.3380	−0.5043	0.1258	1.3944	−0.3702	−0.2331	−0.0772	0.0075	0.1290	−0.4265	0.3679	0.9986	0.0160
	(66.1970)	(0.0499)	(0.0953)	(0.1254)	(0.2044)	(0.2355)	(0.0369)	(0.0777)	(0.1254)	(0.1402)	(0.2022)	(0.2374)	(86.5871)	
3 Coke	−426.9027	0.4256	0.2996	0.1716	0.0496	0.1848	−0.6430	0.0465	0.2449	0.1604	−1.3583	1.8793	09710	0.0906
	(183.1692)	(0.1479)	(0.2329)	(0.2395)	(0.2288)	(0.1285)	(0.1066)	(0.1225)	(0.3889)	(0.4312)	(0.4749)	(0.5582)	(53.1491)	
4 Mineral oil and														
natural gas														
5 Petroleum	69.7436	0.7333	−0.0163	0.0576	0.6440	−0.0224	−0.4099	0.0223	11.1563	−1.0622	0.5526	−0.6812	0.8841	0.0672
products	(233.7634)	(0.4544)	(0.4349)	(0.2395)	(0.2651)	(0.2815)	(0.1412)	(0.1804)	(0.8680)	(0.8383)	(0.3451)	(0.3579)	(44.8866)	
6 Electricity, etc.	−26.1434	0.0030	0.3339	−0.2057	0.8984	−0.3624	−0.1186	−0.1291	0.0347	0.3182	0.6020	−0.7048	0.9917	0.0155
	(51.5244)	(0.1887)	(0.2605)	(0.1605)	(0.2033)	(0.1964)	(0.0768)	(0.0845)	(01893)	(0.2049)	(0.2725)	(0.2477)	(87.3926)	
7 Public gas	184.9195	−0.1753	0.6582	−0.6325	0.5349	0.0739	−0.3584	0.2904	−0.2904	0.1382	0.2513	−0.1382	0.9779	0.0286
supply	(131.9965)	(0.2299)	(0.2748)	(0.2016)	(0.1630)	(0.2029)	(0.0857)	(0.0858)	(0.2853)	(0.2625)	(0.1922)	(0.2368)	(69.6787)	
8 Water supply	−187.1010	1.0830	0.2506	−0.4169	0.5300	00286	−0.4472	0.3289	−0.0699	0.1324	−2.8566	2.0447	0.9594	0.0309
	(93.4028)	(0.4370)	(0.5354)	(0.4807)	(0.1454)	((1.1480)	(01069)	(0.1258)	(0.3251)	(0.2887)	((1.7126)	(0.5450)	(67.4242)	
9 Minerals and	79.1062	0.2078	−0.0659	−0.1183	0.5591	0.2587	−0.1445	0.0580	−0.5481	0.4380	-0_2124	0.1637	0.9706	0.11351
ores n.e.s.	(171.5604)	(0.1627)	(0.1775)	(0.1319)	(0.2462)	(0.1974)	(0.0952)	(0.1142)	(0.4423)	(0.6235)	(0.3888)	(0.4321)	(63.71109)	
10 Iron and steel	−114.7884	0.3066	0.2143	−0.1191	0.5330	0.1296	−0.1959	−0.0661	0.2340	−0.2408	0.4557	−0.2938	0.9923	0.0284
	(76.2388)	(0.1046)	(0.1108)	(0.1014)	(0.2617)	(0.1655)	(0.1471)	(0.1674)	(0.3665)	(0.3609)	(0.1928)	(0.2185)	(69.8891)	
11 Non-ferrous	−31.6300	0.2543	−0.1371	−0.0952	1.0529	−0.2393	−0.1006	−0.0500	0.6953	−0.6961	0.2762	−0.0547	0.9906	0.0208
metals	(33.2585)	(0.1174)	(0.1294)	(0.1154)	(0.1748)	(0.1712)	(0.0465)	(0.0576)	(0.2225)	(0.1985)	(0.1125)	(0.1306)	(78.8431)	
12 Non-metallic	−412.1530	0.3213	0.0723	−0.1101	0.4784	0.4562	−0.3170	−0.1929	0.5393	0.2921	−0.0472	−0.2831	0.9932	0.0181
mineral products	(163.0320)	(0.1825)	(0.1916)	(0.1668)	(02423)	(0.2947)	(0.1253)	(0.1451)	(0.3488)	(0.4147)	(0.4862)	(0.4049)	(82.9853)	
13 Chemicals and	−185.8193	0.0643	0.0698	−0.0238	0.0445	0.4134	−0.3640	−0.0953	0.3115	0.2586	0.4584	−0.3184	0.9819	0.0146
mm fibres	(45.5295)	(0.1694)	(0.1161)	(0.0925)	(0.2309)	(0.1742)	(0.0978)	(0.1229)	(0.30118)	(0.2523)	(0.2036)	(0.2028)	(89.0752)	
14 Metal goods	54.6916	0.2179	0.0515	−0.1951	0.5656	0.1800	−02192	0.0742	0.1296	−0.2402	1.1805	−1.0864	09893	0.01794
n.e.s.	(125.6380)	(0.1241)	(0.1540)	(01097)	(0.2116)	(0.1728)	(0.1143)	(0.1190)	(0.3236)	(0.3611)	(0.6468)	(0.4716)	(832485)	
15 Mech.	−61.0479	0.4429	−0.2544	0.0449	0.4550	−0.0847	−0.1735	−0.3027	−0.1211	0.4252	0.6855	−0.5910	0.9917	0.0140
engineering	(69.0331)	(0.1756)	(0.1849)	(0.1384)	(0.2231)	(0.2248)	(0.1317)	(0.1531)	(0.1878)	(0.2083)	(0.4794)	(0.5108)	(90.3325)	
16 Office	−469.2709	0.3130	−0.1861	0.2028	1.0064	0.1025	−0.6552	−0.0961	0.4731	0.2497	−0.7454	0.3020	0.9331	0.0272
machinery, etc.	(238.0055)	(0.1262)	(0.1643)	(01074)	(0.2490)	(0.2770)	(0.2125)	(0.2181)	(0.3203)	(0.3034)	(1.1230)	(1.1174)	(71.1734)	
17 Elect.	116.3987	0.3691	−0.0526	0.0530	0.1906	0.1288	−0.2311	0.0905	−0.0759	0.5089	0.5176	−1.3098	0.9894	0.0102
engineering	(26.1750)	(0.0762)	(0.1130)	(0.0950)	(0.1719)	(0.1140)	(0.0838)	(00842)	(0.1322)	(0.1589)	(0.2558)	(0.2954)	(99.4433)	
18 Motor vehicles	−26.1567	0.5670	−0.2497	−0.0747	0.8065	0.0150	0.1033	−0.2076	0.2030	−0.0142	−0.0786	−0.2601	0.9868	0.0191
	(88.6237)	(0.0721)	(0.1452)	(0.1808)	(0.1991)	(0.2676)	(0.1286)	(0.1413)	(0.2125)	(0.2216)	(0.3168)	(0.3650)	(81.3391)	
19 Aerospace	222.3672	0.0106	0.0719	−0.0980	0.8612	−0.3314	−0.0477	−0.1448	−0.2439	02592	−0.0317	−0.6019	0.9837	0.0310
equipment	(112.9182)	(0.0956)	(0.0983)	(0.0900)	(0.2301)	(0.2567)	(0.1001)	(0.1029)	(0.2800)	(0.2864)	(0.5831)	(0.4519)	(67.3423)	
20 Ships and	−234.0763	0.5828	−0.1889	−0.3240	1.0843	0.1554	0.0276	0.0009	0.8235	−0.4509	−0.8088	0.8595	0.9881	0.0261
other vessels	(75.5925)	(0.1078)	(0.1700)	(0.1530)	(0.1868)	(02188)	(0.0654)	(0.0839)	(0.2530)	(0.2490)	(0.5227)	(0.3407)	(72.3533)	
21 Other vehicles	−168.3526	0.2968	0.1114	−0.0248	1.0604	−0.1252	−0.1545	0.1153	0.2878	−0.1187	0.9568	−0.8510	0.9968	0.0262
	(135.2829)	(0.1062)	(0.1206)	(0.1278)	(0.2377)	(0.2950)	(0.0741)	(0.0652)	(0.2332)	(0.2443)	(0.4453)	(0.3900)	(72.1643)	

Table 8.1: Unrestricted industrial labour demand equations (contd.)

	inpt/40	y_t	y_{t-1}	y_{t-2}	h_{t-1}	h_{t-2}	w_t	w_{t-1}	$\overline{y}_{ta}$	$\overline{y}_{(t-1)a}$	d_t	d_{t-1}	$\overline{R}^2$ (LLF)	$\widehat{\sigma}$
22 Instr.	467.8993	0.0551	0.1219	0.3384	0.2943	−0.2474	−0.1954	0.2810	−0.0956	0.1420	0.9770	−2.4352	0.9727	0.0155
engineering	(122.7449)	(0.1390)	(0.1263)	(0.1302)	(0.1752)	(0.1310)	(0.0850)	(0.0897)	(0.2179)	(0.2458)	(0.4345)	(0.5578)	(87.4193)	
23 Manufactured	−96.1864	0.7245	−0.3829	0.1395	0.4446	0.2049	−0.1728	−0.0398	−0.0032	0.2371	0.6892	−1.0381	0.9862	0.0151
food	(138.7636)	(0.3199)	(0.2337)	(0.2045)	(0.1915)	(0.1648)	(0.0757)	(0.1034)	(0.1593)	(0.1614)	(0.4081)	(0.3510)	(88.2687)	
24 Alcoholic	−106.2884	0.4993	0.0207	−0.3969	1.0974	−0.1643	−0.1113	0.0734	0.0755	0.2879	−0.5152	0.2831	0.9033	0.0301
drinks, etc.	(190.7833)	(0.5409)	(0.5076)	(0.5117)	(0.2355)	(0.3246)	(0.1279)	(0.1178)	(0.6244)	(0.4363)	(0.6288)	(0.5310)	(68.1601)	
25 Tobacco	150.5910	1.3370	−0.4851	−0.5259	0.6519	1.0220	0.0061	−0.1805	−2.0127	0.0303	1.6853	−0.6736	0.8996	0.0454
	(238.0243)	(0.4483)	(04499)	(0.4688)	(0.2606)	(0.3253)	(0.0687)	(0.0846)	(0.5811)	(0.5378)	(0.9638)	(0.7384)	(56.2608)	
26 Textiles	−163.5445	0.4008	0.0199	−0.3391	0.3040	0.1939	−0.4269	−0.1452	0.2241	0.1107	−0.1934	0.5008	0.9984	0.0160
	(110.9788)	(0.1809)	(0.1746)	(0.1335)	(0.2162)	(0.1579)	(0.0869)	(0.1585)	(0.2388)	(0.3882)	(0.3831)	(0.3380)	(86.4858)	
27 Clothing and	−50.6793	0.5050	0.0437	−0.0218	0.5797	−0.0361	−0.4216	0.0527	−0.1067	−0.0706	−0.0193	0.1237	0.9980	0.0123
footwear	(62.0796)	(0.1060)	(0.1899)	(0.1111)	(0.2890)	(0.1783)	(0.0871)	(0.1505)	(0.1944)	(0.1993)	(0.3359)	(0.2419)	(94.1785)	
28 Timber and	57.3350	0.2824	0.0633	0.0163	0.3392	0.0730	−0.2734	0.0580	0.0904	−0.0689	0.3085	−0.4176	0.9851	0.0145
furniture	(83.7562)	(0.1185)	(0.1490)	(0.1127)	(0.2469)	(0.1428)	(0.0935)	(0.0981)	(0.2622)	(0.3858)	(0.3788)	(0.3563)	(89.4445)	
29 Paper and	92.5460	0.1240	0.1082	0.0645	0.3571	0.3316	−0.1795	0.1637	0.4807	0.1567	0.1453	−1.5894	0.9942	00171
board	(49.1811)	(0.2041)	(0.1692)	(0.7207)	(0.2419)	(0.1362)	(0.0769)	(0.1279)	(03626)	(0.3013)	(0.4445)	(0.4768)	(84.5039)	
30 Books, etc.	116.8745	0.2547	0.0013	−0.0353	0.8079	−0.2086	−0.0864	−0.0626	−0.1167	−0.2457	0.7865	−0.5611	0.9427	0.0112
	(38.6891)	(0.1161)	(0.1355)	(0.0777)	(0.2747)	(0.2297)	(0.0590)	(0.0617)	(0.1828)	(0.1840)	(0.3149)	(0.2934)	(96.8764)	
31 Rubber and	−114.0693	0.3664	−0.0228	−0.3178	0.4959	0.3564	−0.2941	0.0733	0.0953	0.0950	0.1452	0.0405	0.9795	0.0183
plastic pr.	(68.7654)	(0.2515)	(0.2248)	(0.1387)	(0.2956)	(0.2238)	(0.2183)	(0.1284)	(0.4141)	(0.4318)	(0.7177)	(06053)	(82.5569)	
32 Other	157.9064	0.3169	0.0061	0.0167	0.6974	−0.0339	−0.1025	0.0520	0.1378	−0.6710	0.5953	−0.4945	0.9918	0.0136
manufactures	(71.7964)	(0.0739)	(0.1356)	(0.1194)	((0.1986)	(0.2011)	(0.0938)	(0.0909)	(0.2044)	(0.2084)	(0.2537)	(0.1999)	(91.1621)	
33 Construction	0.1708	03346	−0.4336	0.16s0	1.1032	−0.2506	−0.3106	0.4361	0.3935	0.0579	−0.4806	0.1353	0.9804	0.0142
	(72.0577)	(0.1097)	(0.1464)	(0.0958)	(0.1785)	(0.1212)	(0.0893)	(0.1073)	(0.1970)	(0.1980)	(0.3877)	(0.3364)	(89.9104)	
34 Distribution. etc.	165.8094	−0.0730	0.6536	−0.1662	0.7386	−0.2110	−0.1118	−0.0537	−0.0828	−0.5416	0.6452	−0.4648	0.9548	0.0148
	(89.2530)	(0.2295)	(0.3158)	(0.1553)	(0.2320)	(0.1739)	(0.1285)	(0.1296)	(0.1893)	(0.2338)	(0.5798)	(0.4992)	(88.7516)	
35 Hotels and	42.4162	0.3120	0.3603	−0.2291	0.5370	0.0360	−0.3282	0.1610	−0.0757	−0.1791	−0.1462	0.2173	0.8996	0.0218
catering	(100.6322)	(0.2517)	(0.4124)	(0.3271)	(0.3275)	(0.2834)	(0.1593)	(0.1738)	(0.2217)	(0.2135)	(0.3011)	(0.2850)	(77.6065)	
36 Rail transport	120.7537	0.3027	0.4311	−0.0013	0.4013	−0.1819	−0.1381	0.0703	−0.1976	−0.3762	1.1548	−0.7263	0.9979	0.0170
	(103.6322)	(0.1210)	(0.1312)	(0.1354)	(0.1979)	(0.1413)	(0.1133)	(0.1043)	(0.2297)	(0.2246)	(0.2405)	(0.1962)	(84.8173)	
37 Other land	191.4749	−0.0685	0.2417	−0.2581	0.9714	−0.3262	0.0266	0.0380	0.0908	−0.0147	0.3518	−0.4643	0.9740	0.0165
transport	(77.5060)	(0.1657)	(0.1952)	(0.1712)	(0.2405)	(0.2238)	(0.0591)	(0.0650)	(0.1671)	(0.1733)	(0.2886)	(0.2605)	(85.5580)	
38 Sea, air and other	63.4575	0.2006	−0.5254	−0.1053	1.1557	−0.5003	−0.2569	0.2130	−0.0463	0.6873	−0.3834	0.3880	0.9196	0.0224
	(132.3929)	(0.1897)	(0.2355)	(0.1519)	(0.2143)	(0.2906)	(0.1420)	(0.1521)	(0.2573)	(0.3391)	(0.4332)	(0.3686)	(76.7924)	
39 Communications	259.2154	0.5147	−0.4403	−0.2152	0.6941	−0.3138	−0.1371	0.1954	−0.1709	0.1549	0.7280	−0.5472	0.9416	0.0174
	(52.3347)	(0.2695)	(0.4696)	(0.3093)	(0.2071)	(0.2094)	(0.1147)	(01024)	(0.2188)	(0.1711)	((0.3443)	(0.2974)	(84.0346)	
40 Business services	354.3580	0.0441	0.1689	−0.1074	0.3361	−0.1646	0.0769	−0.1400	−0.0283	−0.2971	0.9987	−0.7970	0.9942	0.0125
	(95.7801)	((0.1496)	(0.1514)	(0.1404)	(0.2678)	(0.2481)	(0.0859)	(0.0039)	(0.1137)	(0.1173)	(0.3955)	(0.3127)	(93.5970)	
41 Miscell. services	−248.5543	0.4159	−0.2700	0.2990	1.0218	0.0167	−0.2817	0_1537	0.0955	0.1960	−0.1665	−0.1150	0.9483	0.0229
	(145.5570)	((0.1826)	(0.1973)	(0.1843)	(0.2370)	(0.2849)	(0.1605)	(0.1468)	(0.2302)	(0.2065)	(0.4956)	(0.4013)	(76.1713)	

Table 8.2: Restricted industrial labour demand equations

	inpt/40	y_t	y_{t-1}	y_{t-2}	h_{t-1}	h_{t-2}	w_t	w_{t-1}	$\overline{y}_{ta}$	$\overline{y}_{(t-1)a}$	d_t	d_{t-1}
1 Agriculture, etc.	5.9025	0.4080	0.3215	−0.1981	0.4782		−0.5134		−0.3886			
	(57.5851)	(0.0338)	(0.1440)	(0.1169)	(0.0669)		(0.0836)		(0.0480)			
2 Coal mining	−24.6379	0.2845	−0.4268	1.3624	−0.4332	−0.2194					−0.1226	
	(10.1993)	(0.0499)	(0.0716))	(0.0957)	(0.1359)	(0.1680)	(0.0297)				(0.0326)	
3 Coke	−81.9147	0.3733	0.5137		0.1876		−0.4294		−0.1354		−14606	1.4606
	(66.9629)	(0.1351)	(0.2122)		(0.1151)		(0.0790)		(0.1158)		(0.4225)	(0.4225)
4 Mineral oil and nat. gas												
5 Petroleum products	−92.6973	0.3475			0.7915		−0.2882				0.3281	−0.5234
	(88.2755)	(0.1847)			(0.1253)		(0.1086)				(0.2698)	(0.2702)
6 Electricity, etc.	42.7381		0.5112	−0.3648	1.1202	−0.4619	−0.1490				0.2793	−0.3007
	(26.8282)		(0.1342)	(0.1475)	(0.1862)	(0.1803)	(0.0746)				(0.2408)	(0.2702)
7 Public gas supply	112.7709		0.4662	−0.5338	0.6934		−0.2975	0.2300				
	(44.4602)		(0.1803)	(0.1653)	(0.1060)		(0.0681)	(0.0774)				
8 Water supply	−167.9418	1.4846			0.4752		−0.4094	0.1846			−3.0775	2.4681
	(62.5607)	(0.3476)			(0.1039)		(0.0976)	(0.1085)			(0.6623)	(0.5508)
9 Minerals and ores nes	172.9158	0.2655			0.6931		−0.1494		−0.5337			
	(79.1246)	(0.1265)			(0.0790)		(0.0622)		(0.2560)			
10 Iron and steel	−155.7089	0.3796	0.1997		0.5533		−0.3402					0.1489
	(20.5070)	(0.0562)	(0.0711)		(0.0854)		(0.0596)					(0.0636)
11 Non-ferrous metals	−21.3448	0.2912	−0.1825		1.0929	−0.3288	−0.1061	−0.0520	0.6320	−0.7019	0.2278	
	(27.8342)	(0.1040)	(0.1138)		(0.1544)	(0.1229)	(0.0439)	(0.0496)	(0.2043)	(0.1920)	(0.0572)	
12 Non-metallic min. pr.	−236.8126	0.3842			0.8437		−0.2505	−0.1262	0.3674			−0.2609
	(65.3042)	(0.1125)			(0.0450)		(0.1029)	(0.1188)	(0.2118)			(0.1336)
13 Chemicals and mm fibres	−99.1032				0.5831		−0.2735		0.5764		0.2720	−0.2720
	(28.1020)				(0.0710)		(0.0329)		(0.0769)		(0.1678)	(0.1678)
14 Metal goods nes	−51.0504	0.2663			0.6006		−0.2335		0.2716		0.4410	−0.6038
	(67.1158)	(0.1006)			(0.0653)		(0.0763)		(0.2663)		(0.3831)	(0.3240)
15 Mech. engineering	−101.2153	0.4909	−0.3408	0.1184	0.5996	−0.2925	−0.1979	−0.3348		0.4802		
	(51.8451)	(0.0750)	(0.1484)	(0.0960)	(0.2087)	(0.1792)	(0.1133)	(0.1163)		(0.1746)		
16 Office machinery etc.	−67.1178			0.2278	0.8571		−0.2389		0.2548		1.0366	−1.5206
	(75.6691)			(0.0600)	(0.0721)		(0.1049)		(0.1399)		(0.5489)	(0.5314)
17 Elect. engineering	106.1219	0.3463			0.3886		−0.2592	0.1351		0.3684	0.2694	−0.9762
	(25.5818)	(0.0462)			(0.0646)		(0.0771)	(0.0637)		(0.1099)	(0.1979)	(0.2051)
18 Motor vehicles	−74.0164	0.5451	−0.2618		0.8395	−0.1165		−0.2102	0.2625		−0.2471	
	(48.8105)	(0.0470)	(0.1197)		(0.1614)	(0.0909)		(0.0666)	(0.1099)		(0.0892)	
19 Aerospace equipment	246.9802				1.1388	−0.6421		−0.1737			−0.6598	
	(75.2084)				(0.1495)	(0.1727)		(0.0720)			(0.2197)	
20 Ships and other vessels	−0.7667	0.4809	−0.4809		1.4717	−0.4717			0.5103	−0.5103		
	(0.3086)	(0.1171)	(0.1171)		(0.1543)	(0.1543)			(0.2000)	(0.2000)		
21 Other vehicles	−165.4705	0.3896			0.9154		−0.1680	0.1064	0.2078		0.7687	−0.7687
	(58.7346)	(0.0706)			(0.0419)		(0.0552)	(0.0564)	(0.1235)		(0.3109)	(0.3109)

(continued)

Table 8.2: Restricted industrial labour demand equations (contd.)

	inpt/40	y_t	y_{t-1}	y_{t-2}	h_{t-1}	h_{t-2}	w_t	w_{t-1}	$\overline{y}_{ta}$	$\overline{y}_{(t-1)a}$	d_t	d_{t-1}
22 Instr. engineering	423.2630		0.2115	0.3539	0.3177	−0.2486	−0.2377	0.2377			1.2396	−2.6097
	(55.7285)		(0.0825)	(0.0987)	(0.1267)	(0.1064)	(0.0581)	(0.0581)			(0.3043)	(0.3537)
23 Manufactured food	−172.9101	0.5982			0.4844	0.3037	−0.2337				0.6483	−0.8005
	(63.0103)	(0.1660)			(0.1266)	(0.1327)	(0.0498)				(0.3462)	(0.3177)
24 Alcoholic drinks, etc.	−96.0446				1.1312	−0.3072	−0.0956		0.4479			−0.1047
	(69.2247)				(0.1750)	(0.2359)	(0.0705)		(0.1457)			(0.0868)
25 Tobacco	155.5488	1.3459	−0.4827	−0.5240	0.6473	1.0209		−0.1781	−2.0136		1.7244	−0.6987
	(182.3643)	(0.4049)	(0.4058)	(0.3982)	(0.2205)	(0.2796)		(0.0668)	(0.5337)		(0.6965)	(0.6201)
26 Textiles	−69.3333	0.3637		−0.2759	0.5652		−0.4337		0.0882			0.2960
	(39.2933)	(0.0675)		(0.0871)	(0.0657)		(0.0753)		(0.1207)			(0.1265)
27 Clothing and footware	−68.9489	0.4514			0.5364		−0.3756					
	(11.9600)	(0.0372)			(0.0411)		(0.0284)					
28 Timber and furniture	27.5732	0.3925			0.5144		−0.2885	0.1108			−0.1174	
	(13.7004)	(0.0352)			(0.0572)		(0.0595)	(0.0681)			(0.0282)	
29 Paper and board	39.4291	0.4375	0.1880		0.4661		−0.2130					−0.5252
	(31.8083)	(0.0640)	(0.0929)		(0.0659)		(0.0477)					(0.1523)
30 Books, etc.	96.1742	0.4094	−0.2040		1.3273	−0.6121	−0.0578		−0.2125			
	(32.3071)	(0.0946)	(0.0681)		(0.1955)	(0.1602)	(0.0477)		(0.1434)			
31 Rubber and plastic pr.	−81.7005	0.4581		−0.2463	0.5365	0.3160	−0.2692					
	(16.8687)	(0.0427)		(0.0729)	(0.1198)	(0.1195)	(0.0700)					
32 Other manufactures	56.9103	0.2992			0.7367		−0.0805		0.4048	−0.6459		
	(51.0746)	(0.0602)			(0.0902)		(0.0810)		(0.1692)	(0.1125)		
33 Construction	3.3516	0.3475	−0.3710	0.1345	0.9814	−0.2355	−0.3435	0.3435	0.3916		−0.2830	
	(27.4087)	(0.0858)	(0.1167)	(0.0835)	(0.0957)	(0.0955)	(0.0704)	(0.0704)	(0.1500)		(0.0705)	
34 Distribution. etc.	141.2744		0.7842	−0.2726	0.6360		−0.0409			−0.5717		
	(41.2202)		(0.1615)	(0.1207)	(0.0917)		(0.0356)			(0.1527)		
35 Hotels and catering	−58.7494	0.3544			0.7096		−0.3876	0.1959				
	(44.4425)	(0.1150)			(0.1022)		(0.1191)	(0.1094)				
36 Rail transport	−50.9886	0.1307	0.3372		0.5187		−0.2545				0.8608	−0.6958
	(25.3141)	(0.0894)	(0.1055)		(0.0978)		(0.0718)				(0.2211)	(0.1953)
37 Other land transport	118.4359		0.2123	−0.2016	1.003	−0.2638					0.4509	−0.5162
	(34.6439)		(0.1302)	(0.1434)	(0.1803)	(0.1884)					(0.2304)	(0.1937)
38 Sea, air and other	67.2451	0.1608	−0.3506		1.2952	−0.6002	−0.2432	0.1835		0.3148		
	(92.1103)	(0.1269)	(0.1263)		(0.1460)	(0.1863)	(0.1204)	(0.1255)		(0.1521)		
39 Communications	309.7212				0.5777	−0.4744	−0.0937	0.0860			1.0539	−0.9139
	(57.1105)				(0.1677)	(0.1616)	(0.0631)	(0.0664)			(0.2010)	(0.1745)
40 Business services	209.6513	0.3108			0.6781	−0.3104				−0.1633		
	(49.1545)	(0.0718)			(0.1759)	(0.1680)				(0.0486)		
41 Miscell. services	−39.9043	0.2123			0.8264		−0.1408					
	(33.3057)	(0.0790)			(0.0970)		(0.0747)					

Notes: See notes to Table 8.1.

Table 8.3: Summary and diagnostic statistics for the restricted employment equations of Table 8.2

Industry	$\overline{R}^2$	χ^2_r	$\widehat{\sigma}$	$\chi^2_{SC}(1)$	$\chi^2_{FF}(1)$	$\chi^2_N(2)$	$\chi^2_H(1)$
1 Agriculture, etc.	0.9983	5.00 (5)	0.0136	0.08	5.96	0.02	0.06
2 Coal mining	0.9987	4.02 (4)	0.0155	0.00	0.25	1.18	0.32
3 Coke	0.9656	10.10 (5)	0.0551	0.00	13.82	0.50	2.96
4 Mineral oil and nat. gas	—	—	—	—	—	—	—
5 Petroleum products	0.8874	6.94 (6)	0.0663	0.00	1.89	1.93	0.33
6 Electricity, etc.	0.9909	7.59 (4)	0.0163	1.08	1.26	0.22	2.58
7 Public gas supply	0.9773	8.14 (6)	0.0290	4.60	1.11	4.12	5.42
8 Water supply	0.9520	10.06 (5)	0.0336	0.01	0.38	0.71	0.00
9 Minerals and ores nes	0.9760	3.84 (7)	0.0318	1.36	0.16	32.70	0.00
10 Iron and steel	0.9919	8.63 (6)	0.0291	0.20	4.59	2.08	0.77
11 Non-ferrous metals	0.9912	1.31 (2)	0.0202	4.25	2.19	2.75	0.05
12 Non-metallic min. pr.	0.9937	4.82 (5)	0.0174	1.82	4.70	2.09	4.48
13 Chemicals and mm fibres	0.9808	9.67 (7)	0.0151	4.76	3.53	1.65	1.29
14 Metal goods nes	0.9898	5.39 (5)	0.0174	0.21	5.96	1.69	4.74
15 Mech. engineering	0.9916	4.60 (3)	0.0141	0.17	0.21	0.02	1.09
16 Office machinery, etc.	0.9291	7.85 (5)	0.0280	0.38	1.10	0.76	4.48
17 Elect. engineering	0.9893	5.82 (4)	0.0103	1.79	0.01	0.41	1.12
18 Motor vehicles	0.9887	1.43 (4)	0.0176	4.16	0.88	1.55	1.58
19 Aerospace equipment	0.9847	7.20 (7)	0.0301	3.56	1.04	0.06	2.78
20 Ships and other vessels	0.9818	16.13 (8)	0.0323	0.45	0.61	0.40	4.46
21 Other vehicles	0.9972	2.85 (5)	0.0243	0.90	2.10	0.33	2.59
22 Instr. engineering	0.9759	2.44 (4)	0.0146	1.57	1.29	2.14	1.02
23 Manufactured food	0.9856	7.50 (5)	0.0154	0.43	0.48	6.76	2.46
24 Alcoholic drinks, etc.	0.9119	5.49 (6)	0.0288	0.10	1.43	0.71	3.57
25 Tobacco	0.9101	0.02 (2)	0.0430	4.95	16.07	0.09	0.04
26 Textiles	0.9985	5.09 (5)	0.0155	4.21	2.46	2.20	2.03
27 Clothing and footwear	0.9984	4.32 (8)	0.0110	0.36	1.92	0.62	0.03
28 Timber and furniture	0.9873	3.76 (6)	0.0133	0.08	3.73	0.91	0.85
29 Paper and board	0.9932	10.81 (6)	0.0186	5.30	0.14	1.73	6.16
30 Books, etc.	0.9341	9.52 (5)	0.0120	1.76	0.00	0.00	1.13
31 Rubber and plastic pr.	0.9834	2.51 (6)	0.0165	0.14	2.45	0.50	0.01
32 Other manufactures	0.9908	9.82 (6)	0.0144	0.22	0.37	1.11	0.02
33 Construction	0.9819	2.42 (3)	0.0137	0.02	0.10	0.73	0.27
34 Distribution, etc.	0.9603	4.63 (6)	0.0139	0.32	1.96	0.23	2.47
35 Hotels and catering	0.9169	4.17 (7)	0.0198	0.58	1.88	0.45	0.63
36 Rail transport	0.9975	9.84 (5)	0.0183	0.03	0.38	2.66	0.86
37 Other land transport	0.9766	4.12 (5)	0.0157	0.13	0.98	1.02	0.02
38 Sea, air and other	0.9278	2.87 (4)	0.0212	2.23	7.95	1.88	0.91
39 Communications	0.9388	7.59 (5)	0.0178	0.53	1.51	1.16	4.09
40 Business services	0.9940	9.26 (7)	0.0128	0.98	2.01	1.98	0.17
41 Miscell. services	0.9512	8.14 (8)	0.0222	0.06	0.47	0.39	191

Notes:

$\widehat{\sigma}$ is the equation standard error. $\overline{R}^2$ is the adjusted multiple correlation coefficient.

χ^2_r is the chi-squared statistic for the test of r linear restrictions on the parameters of unrestricted employment equations (see Table 8.1). The value of r is given in brackets after the statistic.

Table 8.4: Composite restricted industrial labour demand equations

	inpt/40	y_t	y_{t-1}	y_{t-2}	h_{t-1}	h_{t-2}	w_t	w_{t-1}	$\overline{y}_{ta}$	$\overline{y}_{(t-1)a}$	d_t	d_{t-1}	T_t
1 Agriculture, etc.	5.9025	0.4080	0.3215	−0.1981	0.4782		−0.5134		−0.3886				
	(57.5851)	(0.0338)	(0.1440)	(0.1169)	(0.0669)		(0.0836)		(0.0480)				
2 Coal mining	−24.6379	0.2845	−0.4268	1.3624	−0.4332	−0.2194					−0.1226		
	(10.1993)	(0.0499)	(0.0716))	(0.0957)	(0.1359)	(0.1680)	(0.0297)				(0.0326)		
3 Coke	−351.5712		0.6330				−0.3005		1.0448				−1.3100
	(44.6561)		(0.1471)				(0.0418)		(0.1564)				(0.1752)
4 Mineral oil and nat. gas													
5 Petroleum products	−70.7059	0.3640			0.5185		−0.3144						−0.5087
	(71.7711)	(0.1324)			(0.1348)		(0.0869)						(0.1297)
6 Electricity, etc.	42.7381		0.5112	−0.3648	1.1202	−0.4619	−0.1490				0.2793	−0.3007	
	(26.8282)		(0.1342)	(0.1475)	(0.1862)	(0.1803)	(0.0746)				(0.2408)	(0.2702)	
7 Public gas supply	−47.7381		0.0611		0.4191		−0.1507		0.5379				−0.6014
	(97.2188)		(0.0659)		(0.1524)		(0.0496)		(0.1827)			(0.1995)	
8 Water supply	−167.9418	1.4846			0.4752		−0.4094	0.1846			−3.0775	2.4681	
	(62.5607)	(0.3476)			(0.1039)		(0.0976)	(0.1085)			(0.6623)	(0.5508)	
9 Minerals and ores nes	172.9158	0.2655			0.6931		−0.1494		−0.5337				
	(79.1246)	(0.1265)			(0.0790)		(0.0622)	(0.2560)					
10 Iron and steel	−349.9558	0.1083			0.4978		−0.3873		1.1803				−0.9045
	(58.8686)	(0.0893)			(0.0832)		(0.0777)		(0.2928)				(0.2732)
11 Non-ferrous metals	−84.8257	0.1817	−0.3091		1.2461	−0.4796	−0.0756	0.0756	0.5854				−0.5749
	(30.7245)	(0.1286)	(0.1273)		(0.1458)	(0.1229)	(0.0481)	(0.0481)	(0.1789)				(0.1517)
12 Non-metallic min. pr.	−280.5702	0.3101			0.6919		−0.2356	−0.2214	0.5170				−0.3729
	(60.6439)	(0.1511)			(0.0877)		(0.1075)	(0.0959)	(0.2901)				(0.2148)
13 Chemicals and mm fibres	−125.0557				0.6205		−0.2810		0.6049				
	(23.8339)				(0.0693)		(0.0337)		(0.0773)				
14 Metal goods nes	−32.2448	0.4365			0.5798		−0.1671						−0.1231
	(25.5280)	(0.0444)			(0.0542)		(0.0817)						(0.0976)
15 Mech. engineering	−101.2153	0.4909	−0.3408	0.1184	0.5996	−0.2925	−0.1979	−0.3348		0.4802			
	(51.8451)	(0.0750)	(0.1484)	(0.0960)	(0.2087)	(0.1792)	(0.1133)	(0.1163)		(0.1746)			
16 Office machinery etc.	−67.1178			0.2278	0.8571		−0.2389		0.2548		1.0366	−1.5206	
	(75.6691)			(0.0600)	(0.0721)		(0.1049)		(0.1399)		(0.5489)	(0.5314)	
17 Elect. engineering	106.1219	0.3463			0.3886		−0.2592	0.1351		0.3684	0.2694	−0.9762	
	(25.5818)	(0.0462)			(0.0646)		(0.0771)	(0.0637)		(0.1099)	(0.1979)	(0.2051)	
18 Motor vehicles	−74.0164	0.5451	−0.2618		0.8395	−0.1165		−0.2102	0.2625		−0.2471		
	(48.8105)	(0.0470)	(0.1197)		(0.1614)	(0.0909)		(0.0666)	(0.1099)		(0.0892)		
19 Aerospace equipment	200.3920	0.0732			0.7560	−0.4659		−0.1252					−0.6788
	(53.1219)	(0.0654)			(0.1659)	(0.1440)		(0.0674)					(0.1586)
20 Ships and other vessels	−0.7667	0.4809	−0.4809		1.4717	−0.4717			0.5103	−0.5103			
	(0.3086)	(0.1171)	(0.1171)		(0.1543)	(0.1543)			(0.2000)	(0.2000)			
21 Other vehicles	−165.4705	0.3896			0.9154		−0.1680	0.1064	0.2078		0.7687	−0.7687	
	(58.7346)	(0.0706)			(0.0419)		(0.0552)	(0.0564)	(0.1235)		(0.3109)	(0.3109)	

(continued)

Table 8.4: Composite restricted industrial labour demand equations (contd.)

	inpt/40	y_t	y_{t-1}	y_{t-2}	h_{t-1}	h_{t-2}	w_t	w_{t-1}	$\overline{y}_{ta}$	$\overline{y}_{(t-1)a}$	d_t	d_{t-1}	T_t
22 Instr. engineering	423.2630		0.2115	0.3539	0.3177	−0.2486	−0.2377	0.2377			1.2396	−2.6097	
	(55.7285)		(0.0825)	(0.0987)	(0.1267)	(0.1064)	(0.0581)	(0.0581)			(0.3043)	(0.3537)	
23 Manufactured food	−172.1572	0.6697			0.3177	0.2237	−0.1962			0.1157		−0.4510	
	(76.0517)	(0.1734)			(0.1742)	(0.1560)	(0.0645)			(0.1233)			(0.1973)
24 Alcoholic drinks, etc.	−15.1802	0.2933			0.7283		−0.0945	0.0591					−0.4844
	(73.4889)	(0.1167)			(0.1239)		(0.0919)	(0.0882)					(0.1411)
25 Tobacco	−213.3698	0.7424			0.7367	0.2633							−0.3959
	(80.8449)	(0.2840)			(0.2225)	(0.2225)							(0.1161)
26 Textiles	−69.3333	0.3637		−0.2759	0.5652		−0.4337		0.0882			0.2960	
	(39.2933)	(0.0675)		(0.0871)	(0.0657)		(0.0753)		(0.1207)			(0.1265)	
27 Clothing and footware	−68.9489	0.4514			0.5364		−0.3756						
	(11.9600)	(0.0372)			(0.0411)		(0.0284)						
28 Timber and furniture	27.5732	0.3925			0.5144		−0.2885	0.1108			−0.1174		
	(13.7004)	(0.0352)			(0.0572)		(0.0595)	(0.0681)			(0.0282)		
29 Paper and board	−44.7394	0.4680	0.1585		0.3644		−0.2503						−0.3259
	(13.2869)	(0.0652)	(0.0925)		(0.0842)		(0.0433)						(0.1040)
30 Books, etc.	96.1742	0.4094	−0.2040		1.3273	−0.6121	−0.0578		−0.2125				
	(32.3071)	(0.0946)	(0.0681)		(0.1955)	(0.1602)	(0.0477)		(0.1434)				
31 Rubber and plastic pr.	−64.4432	0.5998	−0.1401		0.6844		−0.1820						−0.3192
	(14.2846)	(0.0588)	(0.0963)		(0.0818)		(0.1007)						(0.1872)
32 Other manufactures	60.3555	0.2345			0.6028				0.4274	−0.4274			−0.3233
	(20.0274)	(0.0435)			(0.0933)				(0.1287)	(0.1287)			(0.0653)
33 Construction	3.3516	0.3475	−0.3710	0.1345	0.9814	−0.2355	−0.3435	0.3435	0.3916		−0.2830		
	(27.4087)	(0.0858)	(0.1167)	(0.0835)	(0.0957)	(0.0955)	(0.0704)	(0.0704)	(0.1500)		(0.0705)		
34 Distribution. etc.	141.2744		0.7842	−0.2726	0.6360		−0.0409			−0.5717			
	(41.2202)		(0.1615)	(0.1207)	(0.0917)		(0.0356)			(0.1527)			
35 Hotels and catering	−58.7494	0.3544			0.7096		−0.3876	0.1959					
	(44.4425)	(0.1150)			(0.1022)		(0.1191)	(0.1094)					
36 Rail transport	−50.9886	0.1307	0.3372		0.5187		−0.2545				0.8608	−0.6958	
	(25.3141)	(0.0894)	(0.1055)		(0.0978)		(0.0718)				(0.2211)	(0.1953)	
37 Other land transport	118.4359		0.2123	−0.2016	1.003	−0.2638					0.4509	−0.5162	
	(34.6439)		(0.1302)	(0.1434)	(0.1803)	(0.1884)					(0.2304)	(0.1937)	
38 Sea, air and other	67.2451	0.1608	−0.3506		1.2952	−0.6002	−0.2432	0.1835		0.3148			
	(92.1103)	(0.1269)	(0.1263)		(0.1460)	(0.1863)	(0.1204)	(0.1255)		(0.1521)			
39 Communications	14.3221	0.9014	−0.4533		0.8261	−0.2785	−0.1686	0.1565					−0.6566
	(41.3966)	(0.1808)	(0.1966)		(0.1727)	(0.1579)	(0.0822)	(0.0807)					(0.2354)
40 Business services	209.6513	0.3108			0.6781	−0.3104				−0.1633			
	(49.1545)	(0.0718)			(0.1759)	(0.1680)				(0.0486)			
41 Miscell. services	−39.9043	0.2123			0.8264		−0.1408						
	(33.3057)	(0.0790)			(0.0970)		(0.0747)						

Notes: See notes to Table 8.1.

Having made these points, however, closer comparison of the results in Tables 8.2 and 8.3 with those in PPK reveals that in some cases the diagnostic test statistics on the new set of equations are less reasonable than those previously found, and in all sixteen industries have a preferable specification in the PPK paper. The superiority of the original equations in so many industries cannot of course be ignored, and for this reason we present a third set of industrial equations in Tables 8.4 and 8.5 which are an amalgamation of the results in Table 8.2 and those in PPK. (The PPK results are labelled *.) These

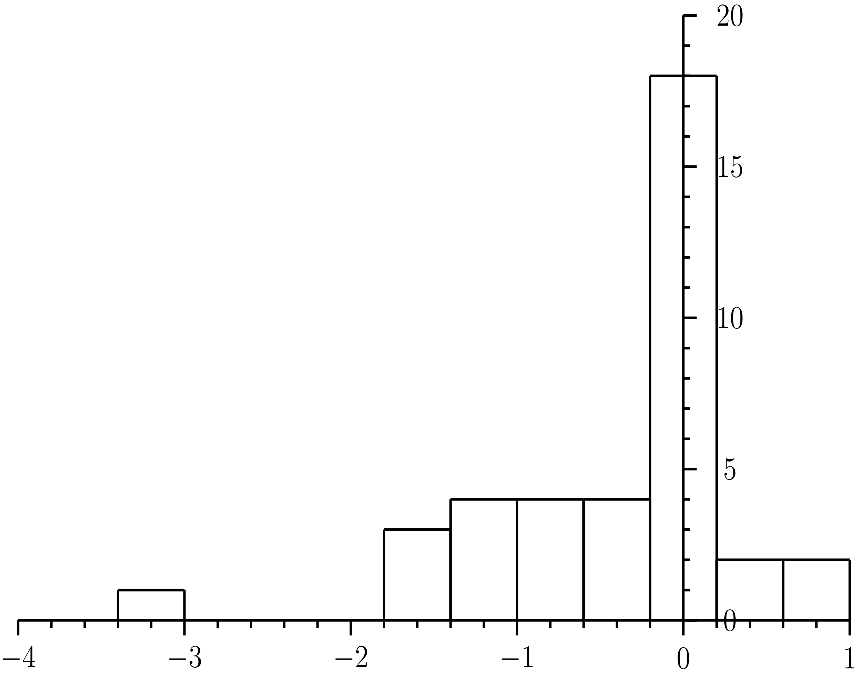

Figure 8.3: Long-run industry productivity elasticities from Table 8.2.

results represent the most satisfactory set of equations that we have been able to obtain for explaining employment at the industrial level in the UK. As before, the long-run coefficients are represented diagrammatically in the histograms of figures 8.4–8.6. Estimated coefficients are once again largely of the expected sign, and of a reasonable magnitude. The mean and standard deviation (in brackets) of the plotted long-run elasticities are 0.86(0.88), −0.54(0.58), and −0.27(0.72) for output, wages, and technological change respectively, confirming the considerable variability of long-run estimates across the industries and providing a reasonable *a priori* case for the use of disaggregated analysis.

8.4.4 Comparison with the aggregate relations

The following unrestricted and restricted aggregate employment equations, corresponding to the results discussed above, were also estimated:

Table 8.5: Summary and diagnostic statistics for the restricted employment equations of Table 8.4

Industry	$\overline{R}^2$	χ_r^2	$\widehat{\sigma}$	$\chi_{SC}^2(1)$	$\chi_{FF}^2(1)$	$\chi_N^2(2)$	$\chi_H^2(1)$
1 Agriculture, etc.	0.9983	5.21 (6)	0.0136	0.08	5.96	0.02	0.06
2 Coal mining	0.9987	6.84 (5)	0.0155	0.00	0.25	1.18	0.32
3 Coke (*)	0.9771	10.15 (8)	0.0449	0.24	0.67	0.27	1.87
4 Mineral oil and nat. gas	—	—	—	—	—	—	—
5 Petroleum products (*)	0.9178	13.40 (8)	0.0566	0.48	0.01	1.83	0.85
6 Electricity, etc.	0.9909	14.25 (5)	0.0163	1.08	1.26	0.22	2.58
7 Public gas supply (*)	0.9719	23.36 (7)	0.0322	1.29	0.00	4.86	1.42
8 Water supply	0.9520	10.18 (6)	0.0336	0.01	0.38	0.71	0.00
9 Minerals and ores nes	0.9760	3.90 (8)	0.0318	1.36	0.16	32.70	0.00
10 Iron and steel (*)	0.9933	12.88 (7)	0.0265	0.08	0.19	1.42	0.43
11 Non-ferrous metals (*)	0.9864	13.33 (5)	0.0250	0.01	3.47	0.20	1.89
12 Non-metallic min. pr. (*)	0.9935	12.21 (6)	0.0177	1.11	0.23	0.76	3.15
13 Chemicals and mm fibres (*)	0.9795	11.78 (9)	0.0156	3.51	1.80	0.96	1.14
14 Metal goods nes (*)	0.9877	12.31 (8)	0.0192	0.09	0.27	0.38	1.00
15 Mech. engineering	0.9916	4.63 (4)	0.0141	0.17	0.21	0.02	1.09
16 Office machinery, etc.	0.9291	9.49 (6)	0.0280	0.38	1.10	0.76	4.48
17 Elect. engineering	0.9893	11.58 (5)	0.0103	1.79	0.01	0.41	1.12
18 Motor vehicles	0.9887	4.82 (5)	0.0176	4.16	0.88	1.55	1.58
19 Aerospace equipment (*)	0.9878	6.31 (7)	0.0268	0.90	0.30	1.81	1.30
20 Ships and other vessels	0.9818	16.91 (9)	0.0323	0.45	0.61	0.40	4.46
21 Other vehicles	0.9972	12.07 (6)	0.0243	0.90	2.10	0.33	2.59
22 Instr. engineering	0.9759	2.46 (5)	0.0146	1.57	1.29	2.14	1.02
23 Manufactured food (*)	0.9837	13.89 (6)	0.0164	1.69	2.78	1.33	4.38
24 Alcoholic drinks, etc.	0.9232	15.42 (7)	0.0269	1.32	0.02	0.94	2.06
25 Tobacco (*)	0.8796	16.63 (9)	0.0497	0.25	8.22	0.65	7.62
26 Textiles	0.9985	5.59 (6)	0.0155	4.21	2.46	2.20	2.03
27 Clothing and footwear	0.9984	4.37 (9)	0.0110	0.36	1.92	0.62	0.03
28 Timber and furniture	0.9873	6.05 (7)	0.0133	0.08	3.73	0.91	0.85
29 Paper and board (*)	0.9927	12.00 (7)	0.0192	1.09	1.33	1.74	4.41
30 Books, etc.	0.9341	9.55 (6)	0.0120	1.76	0.00	0.00	1.13
31 Rubber and plastic pr. (*)	0.9818	8.01 (7)	0.0173	0.21	1.59	0.96	1.03
32 Other manufactures (*)	0.9917	13.49 (8)	0.0137	0.37	0.21	1.12	0.00
33 Construction	0.9819	2.58 (4)	0.0137	0.02	0.10	0.73	0.27
34 Distribution, etc.	0.9603	13.35 (7)	0.0139	0.32	1.96	0.23	2.47
35 Hotels and catering	0.9169	5.59 (8)	0.0198	0.58	1.88	0.45	0.63
36 Rail transport	0.9975	11.38 (6)	0.0183	0.03	0.38	2.66	0.86
37 Other land transport	0.9766	8.48 (6)	0.0157	0.13	0.98	1.02	0.02
38 Sea, air and other	0.9278	6.26 (5)	0.0212	2.23	7.95	1.88	0.91
39 Communications (*)	0.9351	8.03 (5)	0.0184	1.56	0.48	0.14	1.81
40 Business services	0.9940	9.33 (8)	0.0128	0.98	2.01	1.98	0.17
41 Miscell. services	0.9512	8.20 (9)	0.0222	0.06	0.47	0.39	191

Notes: See notes to Table 8.3.

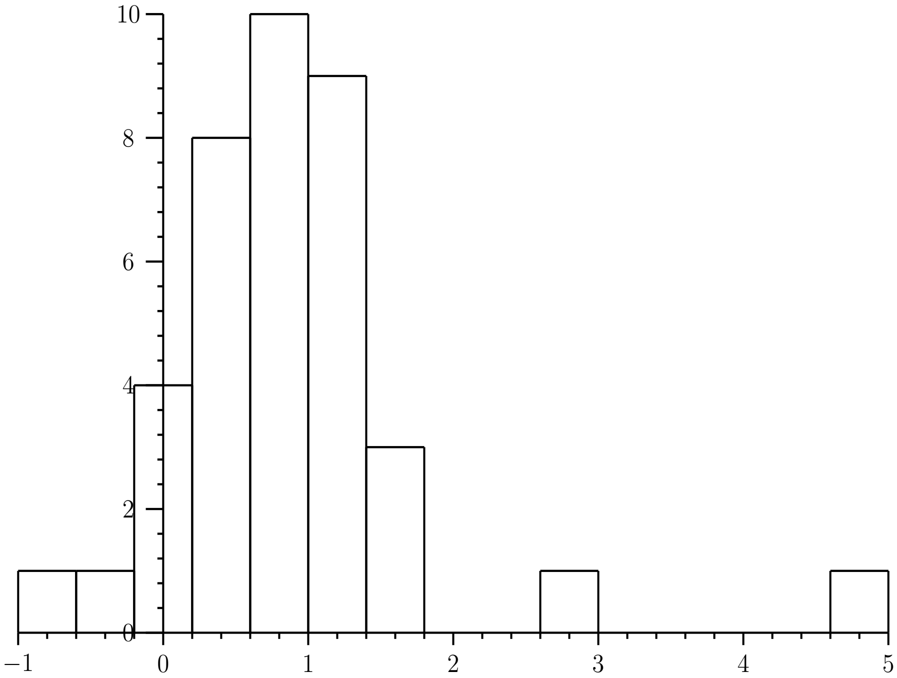

Figure 8.4: Long-run industry output elasticities from Table 8.3.

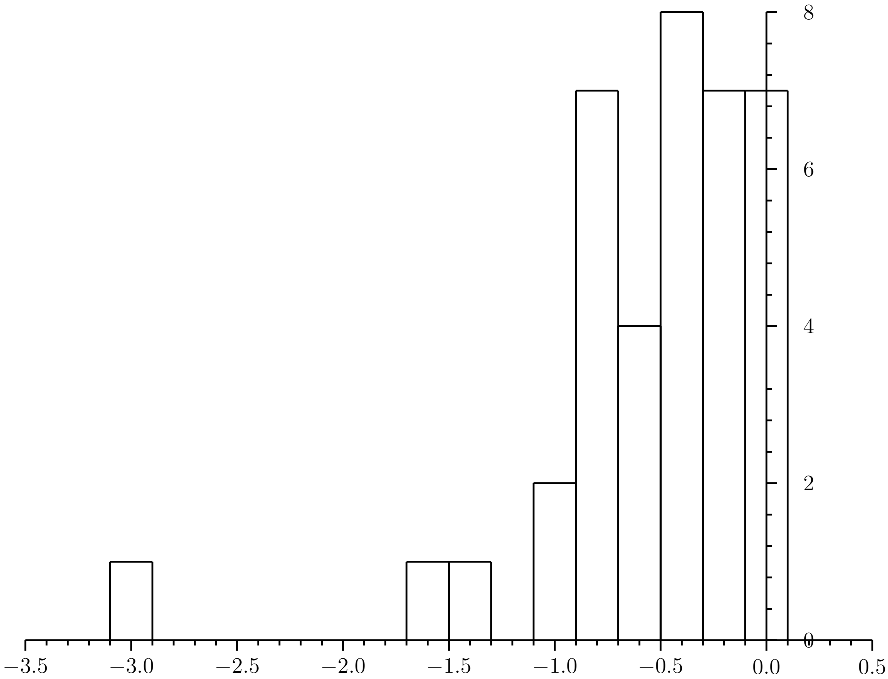

Figure 8.5: Long-run industry real wage elasticities from Table 8.3.

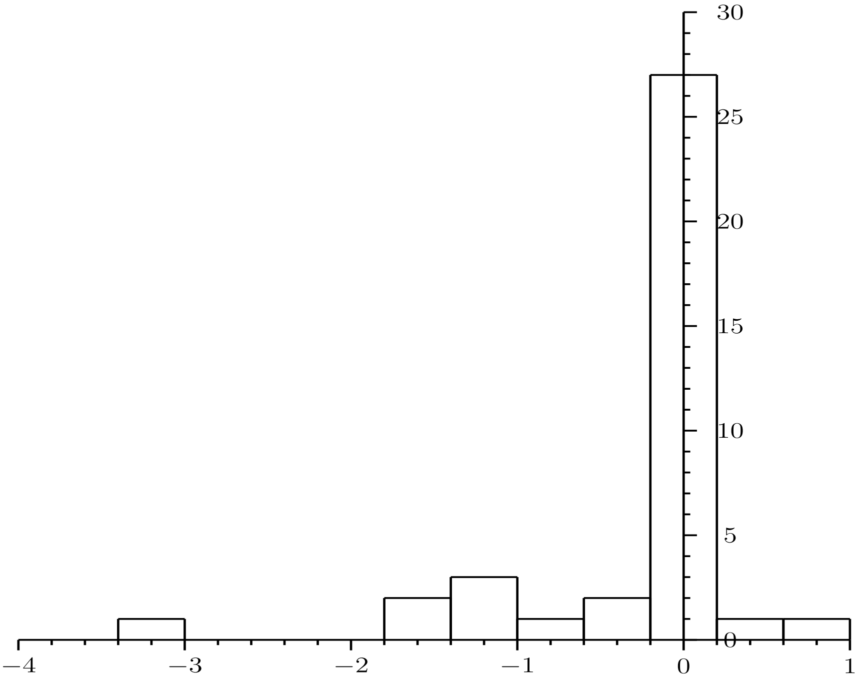

Figure 8.6: Long-run industry productivity elasticities from Table 8.3.

Unrestricted aggregate equation

$$
\begin{aligned}
h_{ta} = &-137.45 + 0.49689\, y_{ta} + 0.19565\, y_{(t-1)a} + 0.11375\, y_{(t-2)a} \\
&\quad (3.77) \quad (6.39) \qquad\qquad (1.32) \qquad\qquad (1.03) \\
&+ 0.33110\, h_{(t-1)a} + 0.18986\, h_{(t-2)a} - 0.34110\, w_{ta} \\
&\quad (1.48) \qquad\qquad (1.17) \qquad\qquad (-5.13) \\
&- 0.087043\, w_{(t-1)a} - 0.046365\, d_{ta} - 0.23405\, d_{(t-1)a} \qquad (8.4.3) \\
&\quad (-1.00) \qquad\qquad (-0.13) \qquad\qquad (-0.82) \\
&\overline{R}^2 = 0.998, \quad \widehat{\sigma} = 0.3316, \quad n = 29 \ (1956\text{–}1984) \\
&\chi^2_{SC}(1) = 3.23, \quad \chi^2_{FF}(1) = 1.27, \quad \chi^2_{N}(1) = 0.67, \quad \chi^2_{H}(1) = 3.51.
\end{aligned}
$$

The figures in brackets are t-ratios, $\widehat{\sigma}$ is the standard error of the regression, $\overline{R}^2$ is the adjusted R^2, n is the number of observations. $\chi^2_{SC}(1)$, $\chi^2_{FF}(1)$, $\chi^2_{N}(2)$, $\chi^2_{H}(1)$ are diagnostic statistics distributed approximately as chi-squared variates (with degrees of freedom in parentheses), for tests of residual serial correlation, functional form misspecification, non-normal errors, and heteroscedasticity, respectively. (For more details about these test statistics and their computations see Pesaran and Pesaran (1987b).)

Restricted aggregate equation

$$\begin{aligned} h_{ta} = & -99.28 + 0.49854\, y_{ta} + 0.67897\, h_{(t-1)a} \\ & \quad (-4.84) \quad (11.06) \qquad (17.37) \\ & -0.31216\, w_{ta} - 0.12049\, d_{(t-1)a} \\ & \quad (-7.53) \qquad (-2.33) \end{aligned} \tag{8.4.4}$$

$$\begin{aligned} & \overline{R}^2 = 0.997, \quad \widehat{\sigma} = 0.3209, \quad n = 29 \text{ (1956–1984)} \\ & \chi^2_{SC}(1) = 1.83, \quad \chi^2_{FF}(1) = 2.46, \quad \chi^2_N(1) = 1.68, \quad \chi^2_H(1) = 5.29. \\ & \text{LM test on exclusion of } (y_{(t-1)a}, y_{(t-2)a}, h_{(t-2)a}, w_{(t-1)a}, d_{ta}) = 4.47, \text{ cf } \chi^2(5) \\ & \text{LM test on exclusion of } (y_{(t-1)a}, y_{(t-2)a}, h_{(t-2)a}, w_{(t-1)a}, d_{ta}, T_t) = 5.90, \text{ cf } \chi^2(6) \end{aligned}$$

where h_{ta}, w_{ta}, and d_{ta} are the aggregate measures of employment, wages, and technological change derived from the industrial figures, and T_t, is a linear time trend ($T_{1980} = 0$).

To check for the possible effect of the simultaneous determination of output, employment, and real wages on the OLS estimates, we also estimated the restricted aggregate equations using the instrumental variable method. With $\mathbf{z}_t = \{1, h_{(t-1)a}, h_{(t-2)a}, y_{(t-1)a}, y_{(t-2)a}, w_{(t-1)a}, w_{(t-2)a}, d_{(t-1)a}\}$ as instruments, we obtained

$$\begin{aligned} h_{ta} = & -86.86 + 0.4708\, y_{ta} + 0.702\, h_{(t-1)a} \\ & \quad (-3.36) \quad (7.11) \qquad (13.48) \\ & -0.2783\, w_{ta} - 0.1365\, d_{(t-1)a} \\ & \quad (-4.69) \qquad (-2.21) \end{aligned} \tag{8.4.5}$$

$$\begin{aligned} & \overline{R}^2 = 0.997, \quad \widehat{\sigma} = 0.3258 \\ & \text{Sargan's misspecification statistic } = 4.40 \text{ cf } \chi^2(3) \\ & \chi^2_{SC}(1) = 1.53, \quad \chi^2_{FF}(1) = 0.48, \quad \chi^2_N(1) = 3.83, \quad \chi^2_H(1) = 5.23. \end{aligned}$$

These clearly differ only marginally from the OLS results in (8.4.4).

The parameter estimates in (8.4.3) and (8.4.4) imply long-run elasticities with respect to aggregate output, real wages, and technological change of (1.68, −0.89, −0.59) for the unrestricted equation and (1.55, −0.97, −0.38) for the restricted equation.

8.4.5 Predictive performance and aggregation bias

Table 8.6 presents the prediction criteria developed in PPK for the aggregate equations (8.4.3) and (8.4.4) and the disaggregate equations of Tables 8.1, 8.2, and 8.4. In each case the disaggregate model outperforms the aggregate equation. The superiority (in terms of predictive performance) of the specifications in Table 8.4 over those in Table 8.2 can also be seen in the estimates presented in Table 8.6. The computation of the statistic for the test of perfect aggregation also provides evidence in favour of the disaggregate model. In the case of the unrestricted version the value of the test statistic is 89.6 which is approximately distributed as $\chi^2(29)$. This strongly rejects the null hypothesis of perfect aggregation.

Bearing this finding in mind, we applied the tests of aggregation bias discussed in section 8.3 to the aggregate and disaggregate employment equations. The results obtained are summarised in Table 8.7. The first row of this table shows the statistics, q_1^*, for testing the hypothesis that the average of the long-run wage elasticities across industries is equal

Table 8.6: Relative predictive performance of the aggregate and the disaggregate employment functions*

	Unrestricted specifications	Restricted specifications	
	(Table 8.1)	(Table 8.2)	(Table 8.4)
Disaggregate criterion	0.1007	0.0856	0.0737
Aggregate criterion	0.1100^a	0.1030^b	0.1030^b

Notes: *Results exclude industry 4 (Mineral oil and Natural gas).
[a]Corresponds to the unrestricted aggregate equation (8.4.3).
[b]Corresponds to the restricted aggregate equation (8.4.4).

to −1. As discussed in the introduction, much policy debate has centred around the extent to which aggregate employment in the UK is influenced by real wage levels. The unit long-run wage elasticity has emerged as the consensus view from this debate and it is for this reason that we use this *a priori* value in our test. The average of the estimated long-run wage elasticities obtained on the basis of the disaggregate results in the three Tables 8.1, 8.2, and 8.4 is −0.66, −0.68, and −0.54 respectively, and these were each compared

Table 8.7: Tests of aggregation bias†

	Unrestricted specifications	Restricted specifications	
	(Table 8.1)[a]	(Table 8.2)[b]	(Table 8.4)[b]
$q_1^*(1)$ [wages]	0.63	5.13	17.20
$q_2^*(1)$ [wages]	0.32	2.46	5.18
$q_2^*(1)$ [output]	0.00	1.68	1.66
$q_2^*(1)$ [technology]	0.07	0.05	0.34

Notes: †Square brackets indicate variables over which restrictions are imposed; figures in round brackets show the number of restrictions imposed, s. Test statistics are compared to $\chi^2(s)$. The $q_1^*$and q_2^* statistics are computed using the results (8.3.5) and (8.3.6), respectively.
[a]Results compared to unrestricted aggregate equation (8.4.3).
[b]Results compared to restricted aggregate equation (8.4.4).

to the consensus value of −1. As is clear, the hypothesised unit elasticity is accepted in the case of the unrestricted specifications, but when a more precisely determined set of results are considered, as in Tables 8.2 and 8.4, the hypothesis is firmly rejected.

Since the q_1^* statistic does not take account of the sampling variation in the consensus estimate that it uses, a more appropriate test of aggregation bias is that based on the q_2^* statistic (see LPP and section 8.3). This test is based on a pseudo true aggregate elasticity obtained through estimation of the aggregate relation, and has the advantage that the same data and the same general specification are used in estimating the aggregate and the disaggregate elasticities. Row 2 of Table 8.7 presents the results of this test for the specifications in Tables 8.1, 8.2, and 8.4. The average wage elasticity across industries in Table 8.1 is compared to −0.89 (the estimated long-run wage elasticity of equation (8.4.3)), while the averages from Tables 8.2 and 8.4 are compared to −0.97 (obtained from the restricted aggregate equation (8.4.4)). Once again, the poorly determined set of equations in Table 8.1 provide no evidence of aggregation bias. A similar conclusion is also obtained from the results of Table 8.2. However, the hypothesis of no aggregation bias based on the more satisfactory estimates in Table 8.4 is firmly rejected at the 5 per cent level, providing strong evidence in support of the claim that the aggregate relation overstates the responsiveness of employment to changes in wages. Similar tests on aggregation bias are reported in rows 3 and 4, for the long-run output and technological change elasticities in turn. Here, average output elasticities of 1.63, 1.23, and 1.24 are obtained from Tables 8.1, 8.2, and 8.4 respectively, while the corresponding average estimates for the technological change elasticity were -0.46, -0.41, and -0.27. These estimates are compared to long-run output and technological change elasticities of 1.68 and -0.59 from the unrestricted equation (8.4.3), and of 1.55 and -0.38 from the restricted aggregate equation (8.4.4). In none of these tests is there any evidence of aggregation bias in the estimated coefficients.

Finally, to check the robustness of the above tests to the specification of the disaggregate model, we computed the Durbin-Hausman misspecification test statistic, q_3, as developed in LPP. For the set of unrestricted disaggregated results of Table 8.1 and the unrestricted aggregate equation (8.4.3) we obtained a value of 48.56 which is distributed as a $\chi^2(7)$) since there are three regressors common to the aggregate and the disaggregate specifications (namely the intercept term, y_{ta} and $y_{(t-1)a}$). This result implies strong rejection of the orthogonality of the disaggregate residuals to the aggregate variables and sheds some doubt on the results of the aggregation bias tests. The misspecification of the disaggregate model might be due to the omission of industry-specific variables, measurement errors, functional form, or dynamic misspecification. It is therefore important that further research is carried out on the specification of the disaggregate employment equations and on the importance of aggregation bias in estimating long-run wage and output elasticities for the economy as a whole.

8.5 Concluding remarks

The application of the statistical methods recently developed by the authors to the study of employment equations in the UK provides some important insights for academics and policy makers alike. The estimated industrial employment equations show that there is a wide diversity in the responsiveness of labour demand to different influences across industries, illustrated most clearly by the histograms discussed in the previous section. In itself, this provides strong support for employing disaggregated analysis rather than aggregate analysis, since the latter cannot capture the structural detail that clearly exists.[14] The result of the test for perfect aggregation confirms that this detail is important even if we are interested only in the prediction of aggregate employment levels, discounting the possibility that errors in disaggregate relations might be offsetting ones. Further, the results of the aggregation bias tests show that the emphasis of policy makers on the importance of wage restraint in attempts to reduce unemployment may be misplaced. These tests confirm the view put forward in PPK that labour demand equations estimated at the aggregate level significantly overstate the extra employment that might be achieved

[14] Indeed, the relatively poor diagnostic statistics obtained in the case of some of the industrial equations indicate that there is likely to be scope for further structural detail in the form of industry-specific variables, and the use of different functional forms across industries.

through wage reductions, however these are achieved. In fact, a wage elasticity of around -0.6 is suggested by the disaggregate results, considerably less than the unit elasticity that has become the consensus view in the UK and which is supported by our own aggregate estimates.[15] The results do not, however, provide any evidence of aggregation bias in the long-run estimates of output and technological change elasticities. Taken together, therefore, these results provide an illustration of the gains to be made from disaggregate analysis, and of the dangers involved in aggregation.

.1 Appendix

With the exception of data on industrial investment, the data used in this study are the same as those employed in PPK, and are taken from the Cambridge Growth Project (CGP) Databank. For the sources of the data and the classifications of industry groups see the data appendix and table A in PPK. For convenience, table A is reproduced in this Appendix (Table A1).

Data on industrial investment in vehicles, in plant and machinery, and in buildings are available separately for the period 1954–84, from which total gross investment is constructed. There is not a one-to-one correspondence between the Blue Book (BB) industrial classifications for which the data are published and our own, however. Where the BB data are more disaggregated, this causes no problem, since we simply amalgamate the appropriate industries. There remain six areas in which the BB data are more aggregated than our own. These are listed in Table A2.

In these cases, we have made the simplifying assumption that the investment reported by BB classification can be divided equally over the (more disaggregate) CGP industrial groups. This procedure is satisfactory if the CGP industry groups within the BB classifications show similar investment growths over the 1954–84 period. This is likely to be the case for the Coal and Coke Industries, but is less likely to hold in the case of the BB industry classifications 13 and 17.

[15] Of course, estimated wage elasticities obtained in unconditional labour demand equations would be somewhat higher as reduced wage inflation helps encourage higher output levels.

Table A1: Classification of industry groups (in terms of the 1980 standard industrial classification)

Industry groups (CGP classification)	Division, class, or group
1 Agriculture, forestry, and farming	0
2 Coal mining	1113, 1114
3 Coke	1115, 1200
4 Mineral oil and natural gas	1300
5 Petroleum products	140
6 Electricity, etc.	1520, 1610, 1630
7 Public gas supply	1620
8 Water supply	1700
9 Minerals and ores nes	21,23
10 Iron and steel	2210, 2220, 223
11 Non-ferrous metals	224
12 Non-metallic mineral products	24
13 Chemicals and man-made fibres	25,26
14 Metal goods nes	31
15 Mechanical engineering	32
16 Office machinery, etc.	33
17 Electrical engineering	34
18 Motor vehicles	35
19 Aerospace equipment	3640
20 Ships and other vessels	3610
21 Other vehicles	3620, 363, 3650
22 Instrument engineering	37
23 Manufactured food	41, 4200, 421, 422, 4239
24 Alcoholic drinks, etc.	4240, 4267, 4270, 4283
25 Tobacco	4290
26 Textiles	43
27 Clothing and footwear	45
28 Timber and furniture	46
29 Paper and board	4710, 472
30 Books, etc.	475
31 Rubber and plastic products	48
32 Other manufactures	44, 49
33 Construction	5
34 Distribution, etc.	61, 62, 63, 64, 65, 67
35 Hotels and catering	66
36 Rail transport	71
37 Other land transport	72
38 Sea, air, and other	74, 75, 76, 77
39 Communications	79
40 Business services	81, 82, 83, 84, 85
41 Miscellaneous services	94, 98, 923, 95, 96, 97

Table A2: Blue Book and Cambridge Growth Project industrial classifications

CCP classification	BB classification
2 Coal	2 Coal and coke
3 Coke	
9 Minerals and ores nes	8 Metals
10 Iron and steel	9 Other minerals
11 Non-ferrous metals	
12 Non-metallic mineral products	
16 Office machinery	13 Electrical and instrument engineering
17 Electrical engineering	
22 Instrument engineering	
19 Aerospace equipment	15 Transport, other than motor vehicles
20 Ships	
21 Other vehicles	
24 Drink	17 Drink and tobacco
25 Tobacco	
29 Paper and board	21 Paper, printing, and publishing
30 Books	

Chapter 9

Persistence of Shocks and their Sources in a Multisectoral Model of UK Output Growth

The extent to which the effects of shocks to the economy persist over time has been the subject of extensive investigation over the past few years. Following the seminal paper by Nelson and Plosser (1982), it has now become a widely held view that aggregate output is best represented by a first-difference stationary process, rather than by a stationary process around a deterministic trend. This has the important implication that macroeconomic shocks can have effects on output levels which continue into the indefinite future; an isolated recessionary shock may cause output growth to be only temporarily lower than usual, but this would be reflected by a time path for the level of output which is permanently lower than what it would have been in the absence of the shock. The size of the long run response of output to a unit shock, known as the persistence of shocks to output, is an empirical issue, and several studies have now been conducted to estimate the persistence measure for the real gross national product in the United States and elsewhere.[1] The evidence presented in these papers is mixed and inconclusive however, largely reflecting the difficulties involved in determining the long run properties of the output series from the relatively short data set available over the post war period. In a recent paper, Pesaran et al. (1991) (PPL), we advocated the use of sectoral output data in order to bring extra information to bear on the analysis of persistence at the aggregate level. In that paper, we noted that the information contained in the relationships between sectoral growth rates, and in the correlations that exist between innovations in output growths of different sectors can be fruitfully utilised to obtain a more reliable estimate of the persistence measure for aggregate output using a multisectoral model of output growths than can be obtained through a univariate model. We presented empiri-

[0] Published in *Economic Journal* (1992), Vol. 102, pp. 342–356. Co-authors K. C. Lee and M. H. Pesaran. An earlier version of this paper was presented at the RES Conference in Warwick, 8-11 April, 1991. We are grateful to Simon Potter and two anonymous referees for helpful comments and suggestions. Financial support from the ESRC and the Isaac Newton Trust of Trinity College, Cambridge, is also gratefully acknowledged.

[1]See, for example, Campbell and Mankiw (1987a, 1989), Harvey (1985), Clark (1987a), Watson (1986), Cochrane (1988), Christiano and Eichenbaum (1989), Shapiro and Watson (1988), Evans (1989), Blanchard and Quah (1989), Demery and Duck (1990), and Mills (1991).

cal support for this approach by analysing output growth in the United States using data disaggregated according to a ten sector classification. The point estimate of the aggregate persistence measure based on this disaggregated model was 0.67, with a standard error of 0.072. This estimate is somewhat lower, and is considerably more precisely estimated, than the estimate based on the aggregate univariate models.

The proposed disaggregated framework of PPL also allows us to decompose the persistence of shocks to aggregate or sectoral outputs into those generated by particular 'macro' shocks and those generated by 'other', possibly sector- specific, shocks. In the PPL paper, we focused on the persistence effect of 'monetary' shocks, finding these to be statistically significant although relatively unimportant in their contribution to the overall persistence measure.[2] In fact, under certain identifying assumptions, this advantage of the disaggregated model could be further exploited, so that the contribution to aggregate persistence of various sources of shocks can be identified and assessed. Of course, 'macro' shocks of different kinds are unlikely to be independent of each other, as innovations in one sphere engender unanticipated changes elsewhere (e.g. governments may react to an unexpected rise in oil prices, say, by engaging in an unexpectedly expansionary monetary policy). In such a case we might wish to measure both the direct long run effect of a shock to a particular macro variable, obtained in the absence of any shocks occurring elsewhere, and the overall long run effect of the shock on the economy, taking into account the feedbacks that have been present historically amongst the different types of shocks in the economy. In this paper, we aim to develop the multisectoral framework set out in PPL further, focusing in particular on the relative contribution of shocks generated from different sources to the overall measure of persistence. In Section 9.1, we develop a multisectoral model in which the effects of different shocks to output in different sectors can be analysed explicitly, and discuss the measurement of persistence effects of the different types of shocks in such a model. In Section 9.2, this framework is applied in an analysis of output growth in the UK economy disaggregated by eight industrial sectors. Estimates of persistence of shocks in each sector and for the economy as a whole are presented, using quarterly data covering the period 1960q1-1989q4. Four types of macro shocks are considered; innovations in money supply growth, in stock returns, in exchange rates, and in oil prices. This is by no means a comprehensive list of all possible sources of macro shocks relevant to the United Kingdom, but represents some of the more interesting ones. Moreover, this provides a first step in the direction away from the univariate approach of Campbell and Mankiw (i987, I989) towards a more behavioural approach to time series modelling of the persistence of shocks in macroeconomics. In so doing, the empirical work suggests which of the shocks are the more important sources of cyclical fluctuations in-sectoral and aggregate outputs, and this may be of practical relevance to policymakers.[3]

[2] The factor analysis of sectoral output growths carried out in Long and Plosser (1987) gives results for the United States which are in line with those in PPL. However, Long and Plosser do not consider the decomposition of macro shocks into its various components.

[3] In view of the difficulties involved in obtaining precise measures of the long run properties of a series on the basis of relatively short data sets, throughout the present work, we give estimates of the standard errors for the various persistence measures, which should serve as an indicator of the degree of uncertainty that surrounds the interpretation of the results.

9.1 Measuring the persistence effects of different types of shocks in a multisectoral model

Let $\mathbf{y}_t$ be an m x 1 vector of sectoral outputs, and suppose that $\mathbf{y}_t$ can be represented by a first-difference stationary linear process. Then a general multisectoral model of output growths may be written as:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{D}(L)\boldsymbol{\nu}_t + \mathbf{A}(L)\boldsymbol{\epsilon}_t, \tag{9.1.1}$$

where $\boldsymbol{\alpha}$ is an m x 1 vector of constants representing sector-specific mean growth rates, $\boldsymbol{\nu}_t$ is a p x 1 vector of innovations in macroeconomic variables $\mathbf{x}_t$, and $\boldsymbol{\epsilon}_t$ is an m x 1 vector of sector-specific innovations with mean zero and the covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$. $\mathbf{A}(L)$ and $\mathbf{D}(L)$ are matrix polynomials of the form

$$\mathbf{A}(L) = \sum_{i=0}^{\infty} \mathbf{A}_i L^i, \quad \mathbf{D}(L) = \sum_{i=0}^{\infty} \mathbf{D}_i L^i \tag{9.1.1b}$$

where the $\mathbf{A}_i$'s and $\mathbf{D}_i$'s are m x m and m x p matrices of fixed coefficients, $\mathbf{A}_0 = \mathbf{I}_m$, and the (i, j)th element of $\mathbf{A}(L)$ is denoted by the lag polynomials $a_{ij}(L)$. The vector of innovations $\boldsymbol{\nu}_t$ are defined by

$$\mathbf{x}_t = \boldsymbol{\Gamma}\mathbf{z}_t + \boldsymbol{\nu}_t$$

where $\boldsymbol{\Gamma}$ is a p x k matrix of fixed coefficients, and $\mathbf{z}_t$ is a k x 1 vector of predetermined variables. This formulation is fairly general and includes the vector autoregressive specification as a special case. The innovations $\boldsymbol{\nu}_t$ are assumed to be white noise processes with mean zero and the constant covariance matrix $\boldsymbol{\Psi} = (\psi_{ij})$. These innovations correspond to macroeconomic shocks such as unexpected changes in oil prices, money supply or exchange rates. The sector-specific innovations, $\boldsymbol{\epsilon}_t$, represent the residual variability in $\Delta\mathbf{y}_t$ not associated with the p identified macroeconomic shocks, $\boldsymbol{\nu}_t$. In order to ensure that the equation system (9.1.1) is identified we assume that $\boldsymbol{\nu}_t$ and $\boldsymbol{\epsilon}_t$ are uncorrelated.[4]

The system described by (9.1.1) provides time series representations of output growth over m sectors, where each sector is subject to p identified 'macro' shocks and a residual, possibly sector-specific, shock. The aggregate level of output, Y_t, is defined by

$$Y_t = \mathbf{w}'\mathbf{y}_t \tag{9.1.2}$$

where $\mathbf{w} = (w_1, \cdots, w_m)'$ is an m x 1 vector of fixed positive weights.[5] Under the multisectoral model, (9.1.1), aggregate output growth can be written as

$$\Delta \mathbf{Y}_t = \mathbf{w}'\boldsymbol{\alpha} + \mathbf{w}'\mathbf{D}(L)\boldsymbol{\nu}_t + \mathbf{w}'\mathbf{A}(L)\boldsymbol{\epsilon}_t. \tag{9.1.3}$$

Notice that even for simple specifications of univariate sectoral output growths, (9.1.3) can have a very high order ARMA representation. This in itself provides *a priori* rationale for the use of disaggregated data in the analysis of aggregate persistence, since with relatively short time series available, the estimation of high order ARMA processes for aggregate

[4] On this, see section 3.3 in PPL.

[5] In the empirical analysis, we set $\mathbf{w} = (1, 1, \cdots, 1)'$, and hence use the sum of the logs of the sectoral outputs as the measure of 'aggregate output'. This measure differs from the log of the sum of sectoral outputs used in the literature, but in our work we find that the two measures follow each other very closely and have very similar autocorrelation functions.

output growth may not be desirable or even feasible (see PPL for further discussion). Various statistics have been suggested in the literature as measures of persistence in univariate models, although, as is shown in PPL, all of these alternative measures are based on the spectral density function of $\Delta\mathbf{Y}_t$, evaluated at zero frequency, $f_{\Delta Y}(0)$, and differ only in the way they are scaled. The persistence measure for aggregate output, P_Y, in the multisectoral model, (1a), is based on the spectral density of $\Delta Y_t = \mathbf{w}'\mathbf{y}_t$ at zero frequency scaled by the conditional variance of ΔY_t as follows:

$$P_Y^2 = \frac{2\pi f_{\Delta Y}(0)}{\mathrm{V}(\Delta Y_t|\mathbf{\Omega}_{t-1})} = \frac{\mathbf{w}'\mathbf{D}(1)\mathbf{\Psi}\mathbf{D}(1)'\mathbf{w} + \mathbf{w}'\mathbf{A}(1)\mathbf{\Sigma}\mathbf{A}(1)'\mathbf{w}}{\mathrm{V}(\Delta Y_t|\mathbf{\Omega}_{t-1})}, \tag{9.1.4}$$

where $\mathrm{V}(\Delta Y_t|\mathbf{\Omega}_{t-1}) = \mathbf{w}'\mathbf{D}(0)\mathbf{\Psi}\mathbf{D}(0)'\mathbf{w} + \mathbf{w}'\mathbf{\Sigma}\mathbf{w}$.[6] This measure can be decomposed into a component due to the identified macroeconomic shocks, P_S, and a component due to 'other shocks', P_O, as follows:

$$P_Y^2 = \lambda P_S^2 + (1-\lambda)P_O^2 \tag{9.1.5}$$

where

$$P_S^2 = \frac{\mathbf{w}'\mathbf{D}(1)\mathbf{\Psi}\mathbf{D}(1)'\mathbf{w}}{\mathbf{w}'\mathbf{D}(0)\mathbf{\Psi}\mathbf{D}(0)'\mathbf{w}}, \quad P_O^2 = \frac{\mathbf{w}'\mathbf{A}(1)\mathbf{\Sigma}\mathbf{A}(1)'\mathbf{w}}{\mathbf{w}'\mathbf{A}(0)\mathbf{\Sigma}\mathbf{A}(0)'\mathbf{w}}, \quad \lambda = \frac{\mathbf{w}'\mathbf{D}(0)\mathbf{\Psi}\mathbf{D}(0)'\mathbf{w}}{\mathbf{w}'\mathbf{D}(0)\mathbf{\Psi}\mathbf{D}(0)'\mathbf{w} + \mathbf{w}'\mathbf{A}(0)\mathbf{\Sigma}\mathbf{A}(0)'\mathbf{w}}.$$

Moreover, the component due to the macroeconomic shocks can be further decomposed:

$$P_S^2 = \sum_{j=1}^{p} \mu_j^2 (P_{Sj}^2 + P_{SXj}) \tag{9.1.6}$$

where

$$P_{Sj}^2 = \frac{\mathbf{w}'\mathbf{d}_j(1)\psi_{jj}\mathbf{d}_j(1)'\mathbf{w}}{\mathbf{w}'\mathbf{d}_j(0)\psi_{jj}\mathbf{d}_j(0)'\mathbf{w}}, \quad P_{SXj} = \frac{\sum_{k=1,k\neq j}^{p} \mathbf{w}'\mathbf{d}_j(1)\psi_{jk}\mathbf{d}_k(1)'\mathbf{w}}{\mathbf{w}'\mathbf{d}_j(0)\psi_{jj}\mathbf{d}_j(0)'\mathbf{w}}, \quad \mu_j^2 = \frac{\mathbf{w}'\mathbf{d}_j(0)\psi_{jj}\mathbf{d}_j(0)'\mathbf{w}}{\mathbf{w}'\mathbf{D}(0)\mathbf{\Psi}\mathbf{D}(0)'\mathbf{w}}.$$

In the above expressions, $\mathbf{d}_j(L)$ denotes the jth column of the matrix $\mathbf{D}(L)$; $\mathbf{d}_j(0)$ measures the immediate impacts of the jth shock on the m sectoral output growths, while $\mathbf{d}_j(1)$ measures its long run effects. The components $P_{Sj}, j = 1, \cdots, p$ provide p measures of persistence due to the *direct* effects of shocks to each of the p identified macroeconomic variables assuming all the other shocks are set to zero, and their contribution to P_S is determined by the relative size of the shocks, as represented by the weights μ_j. The *overall* long run impact of the jth shock on Y_t go beyond the direct effects, and include also the interaction terms, P_{SXj} These terms capture the effect of correlations that exist between different shocks on the overall persistence measure. For any macroeconomic shock, comparison of the direct persistence measure, P_{Sj}, and the overall measure, $(P_{Sj} + P_{SXj})$, provides an indication of the extent to which the persistence effects of the jth shock, ν_{jt}, are offset or compounded by associated shocks in other macroeconomic variables.

[6] In the univariate case where the sources of shocks are not explicitly identified, we can write $\Delta Y_t = \alpha + a(L)u_t$, where α is a scalar constant, $a(L)$ is a polynomial in the lag operator, and u_t are mean zero, serially uncorrelated shocks with common variance σ_u^2. Here, (9.1.4) collapses to

$$P_Y^2 = \frac{2\pi f_{\Delta Y}(0)}{\mathrm{V}(\Delta Y_t|\mathbf{\Omega}_{t-1})} = \frac{\sigma_u^2 a^2(1)}{\sigma_u^2} = a^2(1)$$

which is the measure of persistence popularised by Campbell and Mankiw (1987a).

If we are interested not only in persistence of shocks at the aggregate level, but also in the long term effects of shocks on output of a particular sector, then in place of (9.1.4), we can consider the matrix of cross-sectoral persistence measures, $\mathbf{P}$, with its *(i, j)*th element given by

$$P_{ij} = \frac{\mathbf{e}_i'\mathbf{D}(1)\mathbf{\Psi}\mathbf{D}(1)'\mathbf{e}_j + \mathbf{e}_i'\mathbf{A}(1)\mathbf{\Sigma}\mathbf{A}(1)'\mathbf{e}_j}{\mathbf{e}_j'\mathbf{D}(0)\mathbf{\Psi}\mathbf{D}(0)'\mathbf{e}_j + \mathbf{e}_j'\mathbf{\Sigma}\mathbf{e}_j} \tag{9.1.7}$$

and $\mathbf{e}_i$ is a selection vector with unity on its ith element and zeros elsewhere. These provide measures of the long-term effects of shocks in sector j on the level of output in sector i. Sector-specific measures of persistence can be obtained from the diagonal elements of $\mathbf{P}$, and these can be decomposed as in (9.1.5) and (9.1.6), replacing $\mathbf{w}$ with $\mathbf{e}_i$ for sector i. The relationship between the cross-sectoral persistence measures and the aggregate persistence measure, P_y, is in general a complicated one, and is affected by the cointegrating properties of the sectoral output series (see PPL for more details).

9.2 Empirical results; an analysis of sectoral output growth in the UK economy

In this section we apply the multisectoral framework developed above to an analysis of output growths across eight industrial sectors of the UK economy using quarterly data over the period 1960q1-1989q2. The eight sectors cover the whole of UK industrial production, and correspond closely to the main divisions of the 1980 Standard Industrial Classification. The macroeconomic shocks that we investigate explicitly in the analysis include unexpected changes in (nominal) oil prices, in stock returns, in exchange rates, and in the money stock (detailed definitions of the measurement of these series, and further information on the sectoral classification employed, are provided in the Data Appendix of Lee et al. (1991). Of course, other macroeconomic aggregates could be included, but we believe these four types of shock represent some of the more interesting ones in the case of the United Kingdom.

The first stage in the analysis is to obtain an overview of the time series properties of the sectoral output data. Augmented Dickey-Fuller (ADF) statistics for a variety of different lag lengths computed over the sample period do not provide statistically significant evidence in favour of rejecting the unit root hypothesis for the sectoral output series,[7] and this remains true even if we allow for a different trend growth path before and after the first oil shock in 1973q4.[8] Using the ADF procedure, we also tested the hypothesis of a unit root in sectoral output growth rates and found that it was rejected in the case of all the eight sectors. These test results suggest that it is reasonable to proceed with the assumption that sectoral output growth rates are stationary.[9] We also applied the maximum likelihood procedure of Johansen (1988, 1989) to investigate the cointegrating properties of the eight sectoral output series, and found evidence of either

[7] The ADF statistics are based on regressions including an intercept, and lags of various lengths in sectoral and aggregate output growths. The inclusion of the lagged values of aggregate output growth in the ADF regressions for the sectoral output growths does not alter the asymptotic properties of the ADF test, but can improve efficiency by reducing the residual serial correlation which may arise because of the inter-relationships of the output growths in different sectors.

[8] This is the 'changing growth' model of Perron (I989).

[9] A more complete description of the ADF tests carried out, together with the relevant tables summarising the ADF statistics, are given in Lee et al. (1991).

one or two cointegrating vectors, depending on whether we used 'trace' or 'maximal eigenvalue' statistics.[10] The relatively small number of cointegrating vectors found indicates that there are a relatively large number of independent sources of shocks to output, thus providing some evidence of the importance of sector-specific shocks in generating cyclical fluctuations.

The ADF results presented above indicate that the multisectoral model described in (9.1.1) is an appropriate framework with which to analyse persistence in the UK economy. As a preliminary exercise in obtaining estimates of the persistence measures, however, we consider a simplified version of ((9.1.1) in which the macroeconomic shocks are not explicitly identified. Persistence measures are still provided by (9.1.5), setting $\lambda = 0$, and interpreting P_O as the overall measure of persistence. In the empirical analysis, we consider the following simplified versions of (9.1.1):

$$\begin{aligned}
M_1: &\quad \Delta y_{it} = a_i + \sum_{s=1}^{r} c_{s,ii}\Delta y_{i,t-s} + \sum_{j=1,j\neq i}^{r}\sum_{s=1}^{r} c_{s,ij}\Delta y_{j,t-s} + u_{it}, \quad i = 1, \cdots, m \\
M_2: &\quad \Delta y_{it} = a_i + \sum_{s=1}^{r} c_{s,ii}\Delta y_{i,t-s} + \sum_{s=1}^{r} b_{s,i}\Delta y_{-i,t-s} + u_{it}, \quad i = 1, \cdots, m \\
M_3: &\quad \text{a restricted version of } M_2\text{, where variables with coefficients having a} \\
&\quad \text{t-ratio less than unity (in absolute terms) are excluded from the model} \\
M_4: &\quad \Delta y_{it} = a_i + i + u_{it}, \quad i = 1, \cdots, m.
\end{aligned}$$

Model M_1 is an unrestricted VAR, and explains output growth in sector i, Δy_{it}, in terms of lagged output growth in all sectors, including sector i, lagged by up to r quarters. M_2 imposes $rm(m-2)$ restrictions on M_1, and explains Δy_{it} in terms of lagged output growths in sector i and lags in aggregate output growth in the rest of the economy (denoted by $\Delta y_{-i,t,} = \sum_{j=1,j\neq i}^{m} \Delta y_{j,t,}$. Model M_3 imposes further restrictions on M_2 to exclude insignificant variables, and M_4 represents the most simple model considered in which (log) output in each sector is described by a random walk with drift.

The four models were estimated for our eight sectors of the UK economy including up to five lags in sectoral and aggregate output growth, using the Full Information Maximum Likelihood (FIML) method over the period 1961q4- 1989q2. With $m = 8$ and $r = 5$, model M_1 contains 328 parameters, not counting the parameters of the variance covariance matrix. This model is clearly overparameterised and is entertained here as a benchmark. Imposing the 240 restrictions that underlie model M_2 reduces the number of parameters to be estimated to 88. The likelihood ratio statistic for the test of these restrictions is given by 126.03 (277.1) which is well below its 95% critical value given in brackets. The imposition of a further 50 restrictions on model M_2, setting coefficients equal to zero where t-ratios are less than unity in absolute value, cannot be rejected either, since the likelihood ratio statistic for this test is 23.80 (67.50). However, model M_4 is readily rejected against model M_3 as the relevant likelihood ratio statistic is equal to 217.58 (43.77).

[10] Test statistics based on the maximal eigenvalue of the stochastic matrix suggest that there is precisely one cointegrating vector, while those based on the trace of the stochastic matrix suggest that there are two. These findings were robust to the choice of the specification of the underlying model: similar test results were obtained on the basis of VAR models of order 2, 3, and 4, either allowing for or excluding the possibility of a time trend in, the underlying data generation process. The computation of the Johansen test statistics are carried out on Microfit 3-0. See Pesaran and Pesaran (1991a).

Table 9.1: Sectoral and Aggregate Persistence Measures

	Models		
Sectors	M_l	M_2	M_3
1. Agriculture	2.53	1.75	1.58
	(1.24)	(0.14)	(0.05)
2. Construction	1.23	0.93	0.96
	(0.43)	(0.04)	(0.01)
3. Durables	2.01	1.28	1.24
	(0.99)	(0.08)	(0.01)
4. Non-durables	1.59	1.16	.22
	(0.59)	(0.06)	(0.01)
5. Transport	1.28	1.06	0.97
	(0.50)	(0.06)	(0.01)
6. Energy	1.45	1.01	0.91
	(0.82)	(0.07)	(0.01)
7. Distribution	1.60	0.97	1.10
	(0.60)	(0.05)	(0.01)
8. Services	1.17	0.93	1.15
	(0.59)	(0.06)	(0.06)
Aggregate output	1.33	1.11	1.07
	(0.42)	(0.29)	(0.11)

Notes: Sectoral persistence measures, P_i, are estimated using (9.1.5) setting $\lambda = 0$, and using the selection vector $\mathbf{e}_i$ in place of $\mathbf{w}$. The aggregate persistence measure, P_y, uses $\mathbf{w}$. Bracketed figures are asymptotic standard errors. These are calculated using analytic derivatives. The formulae used are given in Appendix B of PPL.

Estimates of sectoral and aggregate persistence measures based on models M_1, M_2 and M_3 are provided in Table 9.1. As is to be expected, persistence measures based on the more parsimonious models M_2 and M_3 are much more precisely determined than the estimates based on the unrestricted model M_1. The aggregate persistence measure obtained from model M_3 is estimated to be 1.07 (0.11), with the standard error of the estimate given in brackets, which is somewhat lower than that obtained using models M_1 or M_2. The persistence measures obtained from model M_3 show considerable variability across sectors; persistence measures in the Agriculture and Manufacturing sectors (1, 3 and 4) are well in excess of unity, and are also rather larger than those obtained for the Construction and Service sectors (2, 5 to 8).

It is of interest to compare the results obtained from the multisectoral models M_1, M_2 and M_3 with those obtained from a univariate model, and we therefore also calculated persistence measures for aggregate output estimated on the basis of various ARMA models fitted to aggregate output growth over the same sample period. The most general specification considered for the aggregate series was an ARMA(5,4) model, although the maximised values of the log likelihood function obtained for this model and for lower order ARMA processes were very close, indicating that the process for aggregate output may be adequately characterised by a random walk with drift (for which the persistence measure is equal to unity). We have already noted that model M_4, in which output in each sector follows a random walk with drift, is rejected by the data, so that the univariate result is consistent with the multisectoral findings only under particular restrictions on the size of the ARMA coefficients in the sectoral equations and on the correlations between sectoral shocks. There are, however, no *a priori* reasons for the validity of such aggregation restrictions, and consequently, these results raise the possibility of aggregation bias in models estimated at the economy-wide level, suggesting that caution should be exercised in the use of aggregate data.[11] It is noted that the unit measure of persistence associated with the univariate model is lower than that obtained for aggregate output based on the multisectoral model M_3. However, given the estimated standard errors, there is no inconsistency in these results, which are in line with those provided in the literature on persistence of shocks to aggregate output in the United Kingdom.[12]

We now turn our attention to the primary concern of this paper, which is to identify the contribution of different types of shocks to the total persistence measure. The following version of the complete multisectoral model, (9.1.1), is therefore considered:

$$\widetilde{M}_2: \quad \Delta y_{it} = a_i + \sum_{s=1}^{r} c_{s,ii}\Delta y_{i,t-s} + \sum_{s=1}^{r} b_{s,i}\Delta y_{-i,t-s} + \sum_{j=1}^{p}\sum_{s=0}^{r} \gamma_{i,js}\nu_{j,t-s} + u_{it}, \quad i = 1, \cdots, 8.$$

The model contains up to four lags of sectoral and aggregate output growth rates, as well as current and four lagged values of the macroeconomic shocks. Model $\widetilde{M}_2$ to be completed with equations for the p types of macroeconomic shocks, $\nu_{jt}(j = i, \cdots, p)$. Here, we consider four types of shocks; namely, (i) unexpected changes in the money supply ('money shocks'), (ii) unexpected changes in excess returns on stocks ('stock market shocks'), (iii) unexpected changes in Sterling exchange rate ('foreign exchange shocks'), and (iv) unexpected changes in nominal oil prices ('oil price shocks'). The specification of the four equations that are used to determine these shocks are shown in Table 9.2. For each of the four equations, the most general specification that was considered included among the explanatory variables values of the four dependent variables lagged by up to four periods. For the first two of the macro equations, these were further supplemented by (lagged) measures of growth in Governmnent expenditure, and by an unemployment variable U (for the money equation), and by measures of changes in interest rates, of the dividend yield, and of the rate of price inflation for the excess returns equation.[13] (Precise variable definitions are provided in the Data Appendix of Lee et al. (1991)). A specification search was carried out on the OLS estimates of the equations to obtain the specifications chosen in Table 9.2. This involved dropping those variables whose coefficients had t-values which were less than unity (in absolute value), ensuring that none of the variables thus excluded were jointly significant. The results of Table 9.2 show that the inclusion of the additional behavioural variables in the first two equations is an important exercise, with the estimated coefficients of the additional variables showing significantly,

---

[11] See Lee et al. (1990a,b) for a more complete discussion of the problems of aggregation bias in the context of linear regression models.

[12] Campbell and Mankiw (1989) report a (bias-corrected) estimate of 0.94 for $a(i)$, based on 60 autocorrelation coefficients using data covering the period 1957q1-1986q2; and Mills (1991) reports a value of unity based on annual data over the post-war period.

[13] These specifications therefore incorporate more behavioural content. The choice of the additional variables included in the money equation is justified in Barro (1977) and Pesaran (1991b), while those in the stock returns equation are discussed in Pesaran and Timmerman (1990).

and taking their expected signs.[14] The results also indicate that simple AR representations cannot be improved upon for the exchange rate and the oil price variables. Given that the residuals from these equations are to be employed as expectational errors, it is important that they do not contain a systematic element, and we note that we could not reject the hypothesis of no serial correlation in any of the four equations.

Having established the form of the macro equations to be used to identify macroeconomic shocks, the system of equations M_2, with eight equations explaining sectoral output growth plus four equations identifying the different macroeconomic shocks, was estimated jointly by the FIML method.[15] Table 9.3 provides Wald statistics for tests of two null hypotheses, H_1 and H_2:

$$H_1: \quad \gamma_{i,j0} = \gamma_{i,j1} = \gamma_{i,j2} = \gamma_{i,j3} = \gamma_{i,j4} = 0, \qquad j = 1, \cdots, 4, \quad i = 1, \cdots, 8$$

$$H_2: \quad \sum_{s=0}^{4} \gamma_{i,js} = 0, \qquad j = 1, \cdots, 4, \quad i = 1, \cdots, 8.$$

Under H_1, the macro shocks have no effect on output growths (whether short- run or long-run), while under H_2, macro shocks are allowed to have short run effects, but no long run impact on output growths. Clearly H_1 implies H_2, but not vice versa. Both hypotheses are rejected only in a small number of cases (5 out of 32), although the more restrictive hypothesis, H_1, is rejected in 11 cases. Imposing the restrictions H_1 where they were not rejected,[16] and excluding variables with t-values less than unity in absolute terms, we obtained a new restricted model, $\widetilde{M}_3$.[17] The sectoral and aggregate persistence measures derived on the basis of this model are presented in Table 9.4. Total sectoral and aggregate persistence measures are given in column (i) of this table, and the decomposition of these totals into the component due to 'macro' shocks and that due to 'other' shocks is given in columns (3) and (2), respectively. The further decomposition of the persistence measure due to the four 'macro' shocks, as defined by equation (6), is given in columns (4) to (8). In terms of the measures of the total persistence of shocks, the results in column (i) of Table 4 are similar to those presented in Table i. Point estimates of the persistence measures for the Agriculture, Services, and the Manufacturing sectors are the largest obtained, although the first two of these are relatively poorly determined. The point estimate of the total persistence measure for aggregate output is o-8833 (0-067), which is lower than that obtained in Table I and is significantly less than unity.

In almost all sectors, the contribution of the 'macro' shocks to total persistence is relatively small. This is clearly reflected in the similar estimates obtained for P_Y and P_O. This seems to be primarily due to the fact that 'macro' shocks over the period under

[14] The insignificance of the interest rate variable in the equation explaining excess returns may be because the variable used does not adequately reflect the most up-to-date information available in the market. A more elaborate investigation of effects of using different interest measures might also be worthwhile. Given the purpose of this paper, these possible further refinements will not be pursued here, however.

[15] The use of the FIML method enables us to avoid the generated regressor problem highlighted in Pagan (1984). See also (Pesaran, 1987, ch. 7).

[16] We did not impose the restriction in H_2 even if the hypothesis was not rejected so as not to eliminate the effects of macroeconomic shocks which may show significantly in a more parsimonious model in which coefficient values are more precisely estimated.

[17] The unrestricted model $\widetilde{M}_2$ contains 250 parameters. Model $\widetilde{M}_3$ is obtained through the imposition of 168 restrictions, and the corresponding likelihood ratio statistic for the joint test of the validity of these restrictions is equal to 160.12, cf. $\chi^2(168)$.

consideration have been small in size as compared with the 'other' shocks, and this is reflected by the small weight given to the macro shocks in the total persistence measure in (9.1.5).[18] The exceptions to this observation are the Durable and Non-durable Manufacturing sectors, for which the component due to 'macro' shocks is relatively large and significantly different from zero. Columns (4) to (8) of Table 9.4 consider the further decomposition of the sectoral and aggregate persistence measures for macroeconomic shocks, and for each type of macroeconomic shock give the contribution of the direct and overall measures of persistence of shocks to P_S, (i.e, $\mu_j P_{Sj}$, $(j = 1, \cdots, 4)$, $\sum_{j=1}^{4} \mu_j^2 P_{SXj}$ respectively, as in (9.1.6)). To a large extent, these measures reflect the results of Table 9.3, with only a small number of measures being significantly different from zero.[19] However, the results indicate that among the four 'macro' shocks, it is the foreign exchange shocks which have the largest persistence effects on aggregate output, primarily exerted through their effect on the Durable and Non-Durable Manufacturing sectors. The contribution of oil price shocks and stock market shocks are smaller, although still significant, while money shocks appear to be the least important in contributing to the persistence measure due to 'macro' shocks.[20]

In every case where more than one type of macroeconomic shock is included in the output equation, the contribution of $\sum_{j=1}^{4} \mu_j^2 P_{SXj}$ is negative, indicating that in general an unanticipated change in one of the macroeconomic variables is associated with offsetting unanticipated changes in the other macro variables such that the overall impact of the shock on the persistence measure is much reduced. As an illustration of this phenomenon, consider the direct and the overall impact of oil price shocks on UK output growth. In practice, it is reasonable to assume that oil price shocks are exogenous to the UK economy, but we cannot rule out the possibility that oil price shocks generate unanticipated movements in other macro variables, namely money supply, exchange rates, etc. In fact, we find the direct long run effects of oil price shocks on aggregate output, denoted by P_{S1}^2 in (9.1.6), takes the value of 0. 8812 (0.2487), while the overall measure, given by $(P_{S1}^2 + P_{SX1})$ takes the value of 0.7374 (0.4306). This can be interpreted as providing evidence that the responses of the monetary authorities and the foreign exchange and stock markets serve partially to offset the long run impact of oil price shocks.

The results described above provide some justification for the use of sectoral data not only in the analysis of the persistence of shocks to sectoral output levels, but also in the analysis of aggregate persistence. As is the case with the results presented in PPL for the United States, the estimate of the aggregate persistence measure based on the multisectoral model is lower than that obtained from a univariate model, raising the question of whether there is an element of aggregation bias introduced in the aggregate model.[21] However, although the measure of persistence of shocks to aggregate output in

[18] We should recall here that the measures of persistence due to 'other' shocks include the effects of all macroeconomic shocks which are independent of those explicitly accommodated within the analysis, as well as purely sector-specific shocks.

[19] The presence of non-zero persistence measures in Table 9.4 which have turned out to be not significantly different from zero are generally due to the indirect effect of shocks in other sectors operating through the lagged aggregate output terms included in the model.

[20] In contrast to the United Kingdom, we did find a significant persistence effect due to money shocks in the United States (see PPL). However, it would be interesting to check the robustness of the US results to the explicit inclusion of more macro shocks in the model. This is particularly important in the light of the evidence provided in Hamilton (I983) on the effects of oil price shocks on the US economy.

[21] It would be interesting to investigate whether the relatively high estimated persistence measures obtained in many countries by Campbell and Mankiw (1989) using univariate models for aggregate output are similarly affected by such aggregation bias, and if so, whether the large variability in measures found

the UK based on the multisectoral model is less than the unit value obtained using the univariate models, it is clear that the measure is substantially larger than zero, so that output levels seem to be permanently affected by shocks.

The results also suggest that sectoral analysis is a useful exercise in its own right. Certainly, there are some sectors in which shocks have larger long term effects than in others, and it is clear that certain types of shocks are more significant in some sectors than in others. Moreover, subject to the qualifications elaborated above, the results can be interpreted as providing evidence that sector-specific shocks have a more permanent impact on sectoral and aggregate output than macroeconomic shocks. A partial explanation for the relative unimportance of macro shocks is that a macro shock of one type may result in offsetting macro shocks of another type, and indeed, we found evidence that the persistence effects of an oil price shock are partially offset by the (unanticipated) reactions of the monetary authorities, foreign exchange market, and stock market. However, a large element of the total persistence of shocks was identified to be due to 'other' shocks. The cointegration test results reported in the paper also suggest that there are a relatively large number of independent sources of shocks in the UK economy, and one interpretation of this result is that shocks which are most important in generating persistence effects are the sector-specific ones. This is an important conclusion and deserves further investigation. The sensitivity of the results to the inclusion of other types of macro shocks, the choice of sample period, lag lengths and the level of disaggregation need to be further studied. It would also be interesting to see whether results obtained for the United Kingdom can be replicated for other countries.

across the countries would be reduced by the use of sectoral data.

Table 9.2: Estimates of the Equations used in the Derivation of Macro Shocks

Money supply growth equation

$$DLM = 0.0246 + 0.1108\ DLM(-1) + 0.11858\ DLM(-2) + 0.1858\ DLM(-3) + 0.4879\ DLM(-4$$

(2 747) (1.1335) (2.220) (2-279) (5 967)

$$+0.0194\ DLG(-1) + 0.1252\ DLG(-2) + 0.1009\ DLG(-3) + 0.0040\ U(-1) + \widehat{\nu}_{1t}$$

(0 302) (2-022) (1.1548) (2 036)

$R^2 = 0.4650$, S.E. equation = 0.01382, LLF = 328.5408, SC = 4.7343

Excess returns equation

$$ST = \text{-}0.1804\ \text{-}0.0078\ DI(-1) + 0.0388\ DIVY(-1)\ \text{-}1.6704\ PI(-1) + \widehat{\nu}_{2t}$$

(4 705) (1.196) (4.388) (2 504)

$R^2 = 0.1678$, S.E. equation = 0.09119, LLF = 110.3585 SC = 4.7743

Exchange rate equation

$$DER = \text{-}0.4465\text{-}0.1866\ DER(-1) + \widehat{\nu}_{3t}$$

(1.288)(1.977)

$R^2 = 0.0346$, S.E. equation = 3.6174, LLF = -299.2134, SC = 3.4339

Oil price equation

$$DLP = 0.0154 + 0.4544\ DLP(-1)\ \text{-}0.2036\ DLP(-2) + \widehat{\nu}_{4t}$$

(1.290) (4.818) (2.127)

$R^2 = 0.1776$, S.E. equation = 0.1242, LLF = 75.5793, SC = 3.7554

Notes: The estimates presented in this table are computed using the OLS method. However, the estimates of the persistence measures reported in Tables 9.3 and 9.4 are computed by the joint estimation of the sectoral output growth equations and the macro equations by the FIML method.

DLM refers to changes in the (log) money stock, DLG refers to changes in the (log) real Government final consumption, U refers to the unemployment variable used in Barro (I977), ST refers to excess returns on stocks, DI refers to the change in the rate of interest on 91 day Treasury Bills, $DIVY$ refers to the dividend yield (%), PI refers to the rate of price inflation, DER refers to changes in the logarithm of the Sterling exchange rate, and DLP refers to the changes in the (log) nominal oil prices (See Data Appendix of Lee et al. (1991) for further details).

'(-i)' indicates that the variable is lagged i periods. Figures in brackets are (absolute) t-ratios. LLF is the maximised log likelihood, SC is the Lagrange Multiplier statistic for testing residual serial correlation (cf. $\chi^2(4)$).

Table 9.3: Wald Test Statistics on the Coefficients of the Macroeconomic Shocks

	Oil price shocks		Money shocks		Foreign Exchange shocks		Stock Market shocks	
	H_1	H_2	H_1	H_2	H_1	H_2	H_1	H_2
1. Agriculture	5.33	0.21	10.69*	0.04	2.31	0.98	3.12	2.01
2. Construction	15.39*	13.46*	3.00	0.27	1.54	0.03	4.80	0.11
3. Durable Manufacturing	12.49*	2.89	6.12	0.59	12.44*	6.60*	16.37*	2.38
4. Non-Durable Manufacturing	8.24	2.74	6.55	0.94	13.64*	10.23*	13.35*	7.36*
5. Transport	1.53	0.63	0.72	0.10	0.05	0.00	3.23	1.46
6. Energy	16.29*	1.83	5.89	2.19	12.60*	3.78	2.42	0.09
7. Distribution	11.89*	10.50*	4.02	0.45	3.65	0.25	9.21	1.46
8. Services	8.21	1.56	4.85	0.02	18.64*	1.01	4.38	0.78

Notes: Results relate to model R_3 described in the text. For each of the sectors and macroeconomic shock, Wald statistics are computed for the test of the hypotheses:

$$H_1: \quad \gamma_{i,j0} = \gamma_{i,j1} = \gamma_{i,j2} = \gamma_{i,j3} = \gamma_{i,j4} = 0, \qquad (j = 1, \cdots, 4)$$

$$H_2: \quad \sum_{s=0}^{4} \gamma_{i,js} = 0, \qquad (j = 1, \cdots, 4).$$

Wald statistics for test of H_1, (H_2) are to be compared with the critical values of the chi-squared distribution with five (one) degree(s) of freedom. '*' indicates that the statistic is significant at the 5% level.

Table 9.4: Decomposition of Sectoral and Aggregate Persistence Measures for Macroeconomic Shocks

(FIML estimates, 1961q4-1989q2)

Sectors	(1) Total P_Y	(2) Other P_O	(3) Macro P_S	(4) DLP $\mu_1 P_{S1}$	(5) DLM $\mu_2 P_{S2}$	(6) DER $\mu_3 P_{S3}$	(7) ST $\mu_4 P_{S4}$	(8) Interaction $\sum_{j=1}^{4} \mu_j^2 P_{SXj}$
1. Agriculture	1.6107	1.6176	1.1472	0.0000	1.1472	0.0000	0.0000	0.0000
	(0.2453)	(0.2442)	(1.7964)		(1.7964)			
2. Construction	0.9010	0.8097	4.0146	4.0146	0.0000	0.0000	0.0000	0.0000
	(0.0854)	(0.0581)	(3.1310)	(3.1310)				
3. Durables	1.2156	1.0751	3.1429	0.0000	0.0000	2.9701	1.0868	-0.1248
	(0.1353)	(0.1151)	(1.1280)			(1.0021)	(0.8700)	
4. Non-Durables	1.0731	0.8312	4.1267	0.0000	0.0000	3.1920	2.6779	-0.3306
	(0.0943)	(0.0526)	(1.6133)			(1.3099)	(1.2566)	
5. Transport	0.8392	0.8392	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0788)	(0.0788)						
6. Energy	0.8464	0.8586	0.7713	0.4055	0.0172	0.6786	0.1051	-0.0416
	(0.0802)	(0.0671)	(0.4163)	(0.2494)	(0.0347)	(0.4781)	(0.0587)	
7. Distribution	0.7893	0.7916	0.6331	0.3050	0.0276	0.5420	0.1697	-0.0156
	(0.0656)	(0.0641)	(0.4238)	(0.2186)	(0.0569)	(0.3738)	(0.1254)	
8. Services	1.2195	1.2662	0.3663	0.1323	0.0120	0.3411	0.0736	-0.0052
	(0.2644)	(0.2762)	(0.5122)	(0.1054)	(0.0250)	(0.5615)	(0.0576)	
Aggregate	0.8833	0.8062	1.8538	0.8930	0.0808	1.5873	0.4969	-0.1337
output	(0.0671)	(0.0566)	(0.5254)	(0.2530)	(0.1599)	(0.5741)	(0.1988)	

Notes: Results relate to model $\widetilde{M_3}$ described in the text. The number of estimated coefficients, N, is 82. The maximised log-likelihood value, LLF, is 2627.72. Sectoral persistence measures, P_{ii}, are estimated using (9.1.7) in the text. The aggregate persistence measure, P_Y, is estimated using (9.1.4) in the text. The decomposition of total persistence into P_S and P_O, ie persistence due to 'Macroeconomic' and 'Other' shocks, is described by (9.1.5) in the text. The table also shows for each type of macroeconomic shock the contribution of the direct and overall measures of persistence of shocks to P_S as described by (9.1.6) in the text, using $\mathbf{w} = (1, \cdots, 1)'$ to obtain the aggregate output persistence measures, and using the selection vector $\mathbf{e}_i$ in place of $\mathbf{w}$ to obtain sectoral persistence measures. Definitions of DLP, DLM, DER and ST are provided in the Notes to Table 9.2.

Bracketed figures are asymptotic standard errors. These are calculated using analytic derivatives. The formulae used are given in Appendix B of PPL.

Chapter 10

Persistence, cointegration, and aggregation: a disaggregated analysis of output fluctuations in the U.S. economy

A framework is developed for measuring the persistence of shocks to aggregate output in the context of a multisectoral model. It is argued that persistence coefficients can be estimated more precisely using a disaggregated model of output growths rather than univariate representations. The effect of cointegration among sectoral output series on the persistence measure is also analysed, and a decomposition of the persistent effect of output innovations into 'monetary' and 'other' shocks provided. The framework is applied to U.S. data, and although 'money' shocks are shown to be statistically significant, their contribution to the total persistence of output fluctuations is found to be relatively unimportant.

10.1 Introduction

Whether the effect of supply or demand shocks on output is temporary or long lasting is an issue of utmost importance in macroeconomics which has attracted a great deal of attention over the past few years. The traditional view of business cycle decomposes the variations in aggregate output into a deterministic trend and a stationary cyclic component, so that the effect of innovations in output are transitory, having no influence on output levels in the long run. But following the influential work of Nelson and Plosser (1982), many economists have come to view the variations in aggregate output in terms of first-difference stationary processes, thus arguing that shocks have a permanent effect

[0] Published in *Journal of Econometrics* (1993), Vol. 56, pp. 57–88. Co-authors M. H. Pesaran and K. C. Lee. Partial support from the ESRC and the Isaac Newton Trust of Trinity College, Cambridge is gratefully acknowledged. We would like to thank Steven Durlauf, two anonymous referees of this Journal and the workshop participants at the ESRC Econometric Study Group Conference, Warwick, the European University Institute, and the University of Alberta for constructive com- ments and suggestions on an earlier draft of the paper. Correspondence to: M. H. Pesaran, Trinity College, Cambridge CB2 ITQ, United Kingdom.

on the level of output. One important issue in this literature is the sire of the long-run response of aggregate output to a unit shock, commonly referred to as the *persistence* of shocks to output.

Several studies have provided estimates of the persistence measure for the real gross national product (GNP) in the United States. These estimates vary considerably depending on the data set used and the estimation procedure adopted. On the basis of low-order ARIMA models estimated on the quarterly U.S. data over the period 1947-85. Campbell and Mankiw (1987a) conclude that 'a 1 percent innovation to real GNP should change one's forecast of GNP over a long horizon by over 1 percent'. Harvey (1985) obtains a similar result using an unobserved component model applied to annual data over the period 1948-70. However, Clark (1987a) and Watson (1986) have obtained substantially lower estimates of persistence using an unobserved component model estimated on a quarterly data set comparable to that employed by Campbell and Mankiw. In these studies a 1 percent shock would lead to around a 0.6 percent change in output in the long run, Cochrane (1988), using a nonparametric procedure also finds little evidence of persistence in GNP. The evidence on the persistence of aggregate output fluctuations in the U.S. is mixed and inconclusive, and as argued in Christiano and Eichenbaum (1989) the issue of whether real GNP is trend or difference-stationary may be very difficult to resolve on the basis of the available post-war quarterly data.

The situation is not, however, as hopeless as it may appear at first. All the studies cited above and reviewed in Christiano and Eichenbaum (1989) base the estimation of the persistence measure on *univariate* representations of real GNP, and therefore ignore the information contained in other variables such as aggregate consumption and investment in the estimation of the persistence of output fluctuations. A number of recent studies have followed this line of argument and have used multivariate models, containing variables in addition to the real GNP, to obtain a more reliable estimate of the size of the random walk component in GNP. These include the studies by Campbell and Mankiw (1987b), Clark (1987b), Shapiro and Watson (1988), Evans (1989) and Blanchard and Quah (1989). Evans (1989), for example, argues that due to the presence of feedback between output growth and unemployment and the strong negative contemporaneous correlation between output growth and unemployment innovations, the unemployment data contain important information for the analysis of output persistence. A similar argument is also made in King et al. (1987) with respect to consumption and investment.

In this paper we consider using a different type of additional information, namely sectoral output growth rates, in the analysis of output persistence at the aggregate level. We develop a suitable framework for the measurement of persistence of shocks to aggregate output in the context of a multisectoral model of output growths. We will be arguing that the information contained in the relations between sectoral growth rates, and the correlations that exist between innovations in output growth of different sectors, enables us to obtain a more precise estimate of the persistence of shocks to aggregate output via the disaggregated model than can be obtained through a direct analysis of the aggregate series.

The disaggregated framework also allows us to decompose the persistence effect of output innovations into 'macro' and 'other' possibly sector-specific shocks, under the identifying assumption that the two types of shocks are contemporaneously uncorrelated. The same assumption is also made by Blanchard and Quah (1989). However, unlike Blanchard and Quah's analysis which imposes the additional restriction of a zero long-run impact of 'demand' shocks on output, our approach does not require any further identifying restrictions.

The plan of the paper is as follows: Section 10.2 presents a brief overview of the literature on measurement of persistence in univariate models. This material prepares the ground for our multisectoral generalization in section 10.3, where we propose a general measure of persistence based on the spectra1 density of first differences, evaluated at zero frequency. We investigate the effect of cointegration among the sectoral outputs on the sectoral and cross-sectoral persistence measures (section 10.3.1) and consider the effect that aggregation may have on persistence (section 10.3.2). Section 10.3.3 of the paper is concerned with the measurement of persistence in multisectoral models with macroeconomic shocks. Finally, section 10.4 applies the disaggregated framework developed in the paper to the analysis of output growth in the U.S. economy, using data disaggregated by ten industrial sectors, and provides estimates of persistence for each sector and for the economy as a whole. We also present separate estimates of the persistence measures for the 'monetary' and 'other' shocks, under the identifying restriction that these two types of shocks are uncorrelated. The results show that the estimate of the aggregate persistence measure based on the multisectoral model is appreciably below that obtained from the aggregate series directly, thus providing further evidence on the upward bias of the estimates of persistence measures obtained on the basis of low order univariate ARIMA specifications. We also present results on the statistical significance of the short-term and the long-term effect of unanticipated monetary growth of sectoral outputs, and show that in five out of the ten sectors studied 'money' shocks are statistically significant and their effects do not die out in the long term. Despite this, due to the relatively unimportant nature of 'money' shocks as compared to the other shocks affecting the economy over the period 1955–87, the contribution of 'money' shocks to the total persistence of output fluctuations in the economy turned out to be rather small.

10.2 Persistence measures in univariate models

In this section we give a brief overview of the different approaches taken to analyze the problem of persistence in univariate models. The aim here is to prepare the ground for our multisectoral generalization of the persistence concept and its measurement in the next section.[1]

Suppose that y_t follows the genera1 first-difference linear stationary process:

$$\Delta y_t = \mu + a(L)\epsilon_t. \tag{10.2.1}$$

where Δ is the first difference operator,

$$a(L) = a_0 + a_1 L + a_2 L^2 + \cdots \tag{10.2.2}$$

is a polynomial in the lag operator L, and μ is a scalar constant. The ϵ_t are mean zero, serially uncorrelated shocks with common variance σ_ϵ^2. The trend-stationary process

$$y_t = \gamma t + b(L)\epsilon_t \tag{10.2.3}$$

is a limiting case of (10.2.1) and arises if $\mu = \gamma$, and the lag polynomial in (10.2.1) has a unit root, namely if $a(1) = 0$. In general, the extent to which the first-difference

[1] Useful surveys of the persistence literature are already available in Diebold. and Nerlove (1989), Stock and Watson (1988) and Christiano and Eichenbaum (1989).

process (10.2.1) deviates from the trend-stationary process (10.2.3) is clearly related to the magnitude of $a(1)$. A natural measure of the size of the random walk, or the unit root component of (10.2.1), is therefore given by $a(1)$, and this is in fact that measure proposed by Campbell and Mankiw (1987a).

The importance of the stochastic trend can also be measured directly in terms of the size of the stochastic variability of the trend component of y_t. Using the Beveridge and Nelson (1981) decomposition, $y_t = \tau_t + z_t$, whereτ_t is the stochastic trend and z_t is the cyclical component, the magnitude of the random walk component can be measured by

$$\mathrm{V}(\tau_t|\Omega_{t-1}) = \sigma_\epsilon^2 a^2(1). \tag{10.2.4}$$

Alternatively, Cochrane (1988) has proposed using the variance of the long differences of y_t, $\mathrm{V} = \lim_{s\to\infty}(\mathrm{V}_s)$, as a measure of persistence, where

$$\mathrm{V}_s = \mathrm{V}(y_t - y_{t-s})/s\,\mathrm{V}(\Delta y_t),$$

$\mathrm{V}(\cdot)$ denotes the variance operator, and $\mathrm{V}(y_t - y_{t-s})$ is the variance of s-differences of y_t. The relationships between $a(1)$, $\mathrm{V}(\tau|\Omega_{t-1})$, V as measures of persistence can be motivated by noting that all the three measures represent different methods of scaling the spectral density of Δy_t at zero frequency. Let $f_{\Delta y}(\omega)$ be the spectral density function of Δy_t. Then under (10.2.1),

$$2\pi f_{\Delta y}(\omega) = \sigma_\epsilon^2 a(e^{i\omega})a(e^{-i\omega}), \quad -\pi \le \omega < \pi,$$

and it is easily seen that

$$a^2(1) = 2\pi f_{\Delta y}(0)/\sigma_\epsilon^2, \tag{10.2.5}$$

$$\mathrm{V} = 2\pi f_{\Delta y}(0)/\sigma_{\Delta y}^2, \tag{10.2.6}$$

where $\sigma_\epsilon^2 = \mathrm{V}(y_t|\Omega_{t-1})$ is the conditional variance of Δy_t and $\sigma_{\Delta y}^2 = \mathrm{V}(\Delta y_t)$ is the unconditional variance of Δy_t. Therefore, the problem of measurement and estimation of persistence of fluctuations in y_t reduces to the problem of estimating the spectral density of Δy_t at zero frequency.[2] This estimate can then be deflated by the conditional or the unconditional variance of Δy_t to obtain scale-free measures of persistence.

10.3 Measurement of persistence in a multisectoral model

Consider the following multivariate generalization of (10.2.1):

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}(L)\boldsymbol{\epsilon}_t, \tag{10.3.1}$$

where $\Delta \mathbf{y}_t$ denotes ther $m \times 1$ vector of output growths $\{\Delta y_{it}$, $\boldsymbol{\mu}$ is an $m \times 1$ vector of constants representing sector-specific mean growth rates, and $\boldsymbol{\epsilon}_t$ is an $m \times 1$ vector of

[2] Some authors, notably Durlauf (1989, 1990) have argued that it is more appropriate to base the analysis of persistence or testing for unit roots on the spectral density (or the spectral distribution function) of first differences at *all* frequencies rather than only at the zero frequency. This seems a useful extension of the spectral density approach to the analysis of persistence, but will not be followed in this paper.

white noise innovations with mean zero and covariance matrix $\mathbf{\Sigma} = \{\sigma_{ij}\}$. The matrix polynomial

$$\mathbf{A}(L) = \sum_{i=0}^{\infty} \mathbf{A}_i L^i$$

is assumed to be absolutely summable. The $\mathbf{A}_i$'s are $m \times m$ matrices of fixed parameters and $\mathbf{A}_0 = \mathbf{I}_m$ (the $m \times m$ identity matrix). We denote the (i, j) element of $\mathbf{A}(L)$ by the lag polynomials $a_{ij}(L)$.

Almost all the multivariate models analysed in the persistence literature are nested within (10.3.1). The bivariate models of Evans (1989) and Blanchard and Quah (1989) and the equilibrium real-business-cycle model discussed in Long and Plosser (1987) can be readily cast in the form of (10.3.1). Although the model is specified in the first-difference-stationary form, it allows for one or more of the variables in $\mathbf{y}_t$ to be trend-stationary.[3]

In the above multisectoral model, shocks originating from sector j can influence the long-run level of output in sector i both directly through the lag filter $a_{ij}(L)\epsilon_{jt}$ and indirectly through their correlations with shocks in the other sectors. In computing the persistence measures, it is therefore important that the effect of output fluctuations through both of these channels are taken into account. With this in mind, consider the blind application of the Campbell and Mankiw procedure to (10.3.1). This yields

$$\lim_{s \to \infty} \{\partial \, \mathrm{E}(\mathbf{y}_{t+s} | \Omega_t) / \partial \boldsymbol{\epsilon}_t\} = \mathbf{A}(1),$$

which clearly ignores the effects of cross-correlations that may exist between shocks in different sectors and is appropriate only in the orthogonal case where $\sigma_{ij} = 0$ for $i \neq j$. In an attempt to render the Campbell and Mankiw approach generally applicable to multivariate systems, one can make use of the Choleski decomposition $\mathbf{\Sigma} = \mathbf{T}^{-1}\mathbf{D}\mathbf{T}'^{-1}$, where $\mathbf{T}$ is lower triangular with unit diagonal elements and $\mathbf{D}$ is a diagonal matrix. Then (10.3.1) can also be written as

$$\Delta \mathbf{y}_t = \mu + [\mathbf{A}(L)\mathbf{T}^{-1}]\mathbf{u}_t, \tag{10.3.1$'$}$$

where $\mathbf{u}_t = \mathbf{T}\boldsymbol{\epsilon}_t$. In this representation the u_{it}'s (the elements of $\mathbf{u}_t$) are contemporaneously uncorrelated, and it may therefore seem appropriate to measure the persistence of shocks by the following multivariate generalization of the Campbell and Mankiw measure.[4]

$$\lim_{s \to \infty} \{\partial \, \mathrm{E}(\mathbf{y}_{t+s} | \Omega_t) / \partial \mathbf{u}_t\} = \mathbf{A}(1)\mathbf{T}_{-1}. \tag{10.3.2}$$

This measure is, however, subject to two main criticisms. Firstly, the Choleski decomposition is not unique and there exist many other orthogonal transformations of $\boldsymbol{\epsilon}_t$. Secondly, the (i, j) element of $\mathbf{A}(1)\mathbf{T}^{-1}$ refers to the long-run response of y_{it} to changes in u_{jt}, which is composed of a linear combination of all the sectoral innovations, ϵ_{ij}, $i = 1, 2, \ldots, m$, and does not necessarily correspond to the persistence effect of a shock originating in a *particular* sector.[5]

[3] The conditions for y_{it} to be trend-stationary are given by $a_{ij}(1) = 0$, for $j = 1, 2, \ldots, m$.

[4] This is in fact the persistence measure used by (Evans, 1989, p. 220) in his bivariate model of output and unemployment.

[5] In their bivariate model, Blanchard and Quah (1989) manage to avoid these difficulties by first identifying 'supply' and 'demand' shocks with the composite disturbances such as those in $\mathbf{u}_t$, and secondly by imposing the identifying restriction that only 'supply' shocks have a long-run impact on output. This restriction allows them to uniquely determine the Choleski factor of $\mathbf{\Sigma}$, defined by $\mathbf{S} = \mathbf{T}^{-1}\mathbf{D}^{1/2}$.

The spectral density approach to the measurement of persistence is not, however, subject to the above-mentioned shortcomings, and can be easily adapted to derive persistence measures both at the level of individual sectors and at the aggregate level. The (unscaled) sectoral measures of persistence are given by the spectral density of matrix $\Delta\mathbf{y}_t$ evaluated at zero frequency; namely,

$$2\pi f_{\Delta y}(0) = \mathbf{A}(1)\mathbf{\Sigma}\mathbf{A}(1)'. \tag{10.3.3}$$

As in the univariate case, this result can also be rationalised directly as measuring the size of the random walk components of the $\mathbf{y}_t$ process. To see this, consider the following multivariate version of the Beveridge-Nelson decomposition:

$$\mathbf{y}_t = \boldsymbol{\tau}_t + \mathbf{z}_t,$$

$$\boldsymbol{\tau}_t = \boldsymbol{\mu} + \boldsymbol{\tau}_{t-1} + \mathbf{A}(1)\boldsymbol{\epsilon}_t,$$

$$\mathbf{z}_t = \sum_{i=0}^{\infty} \mathbf{C}_i \boldsymbol{\epsilon}_{t-i}, \quad \mathbf{C}_i = -\sum_{j=i+1}^{\infty} \mathbf{A}_j,$$

where $\boldsymbol{\tau}_t$ is the $m \times 1$ vector of the (stochastic) trend components and $\mathbf{z}_t$ is the $m \times 1$ vector of the transitory components. Therefore,

$$\mathrm{V}(\boldsymbol{\tau}_t|\Omega_{t-1}) = \mathbf{A}(1)\mathbf{\Sigma}\mathbf{A}(1)',$$

which is identical to the expression in (10.3.3). The (i, j) element in this matrix can now be scaled either by the conditional variance of Δy_{jt}, $\mathrm{V}(\Delta y_{jt}|\Omega_{t-1}) = \sigma_{jj}$, or by its unconditional variance, $\mathrm{V}(\Delta_{jt})$, to obtain scale-free measures of persistence of output fluctuations in sector i caused by a unit shock in sector j. The former method of scaling yields a multisectoral generalisation of Campbell and Mankiw's univariate measure $(a(1)|)$, and the latter gives a generalisation of the Cochrane measure (V). While in principle there is little to choose between the two scaling methods, in practice most researchers have focussed on the Campbell and Mankiw type measure which in the univariate case can be interpreted as the long-run response of y_t to shocks. In what follows we shall also contine our analysis to persistence measures scaled by the conditional variance of first differences.

Let $\mathbf{e}_t$ be a selection vector which has unity on its ith element and zeros elsewhere. Then the cross-sectoral persistence measures, being the long-term effects of shocks in sector j on the level of output in sector i, can be written as

$$P_{ij} = \mathbf{e}_i'\mathbf{A}(1)\mathbf{\Sigma}\mathbf{A}(1)'\mathbf{e}_j / \mathbf{e}_i\mathbf{\Sigma}\mathbf{e}_j, \quad i, j = 1, 2, \ldots, m. \tag{10.3.4}$$

The sector-specific measures of persistence, which we denote by P_i (> 0), can be obtained from (10.3.4) and are given by $P_i = \sqrt{P_{ii}}$. It is clear that in a univariate model P_i reduces to $A(1)$, which is the familiar Campbell and Mankiw measure. Also, in the case where y_{it} is trend-stationary, it easily follows that $P_i = 0$, as it should.[6]

[6] When y_{it} is trend-stationary, all the elements in the ith row and in the ith column of matrix $\mathbf{A}(1)\mathbf{\Sigma}\mathbf{A}(1)'$ will be identically equal to zero (see footnote 3).

10.3.1 Cointegration and persistence

In this section we briefly consider the effect that cointegration among the elements of y_t may have for the cross-sectoral measures of persistence. Here we assume that all components of $\mathbf{y}_t$ are first-difference-stationary, and that there exists an $m \times r$ matrix $\mathbf{a}$ of rank $r(< m)$, such that $\mathbf{a}'\mathbf{y}_t$ is stationary.[7] The matrix $\boldsymbol{\alpha}$ is called the cointegrating matrix and its columns the cointegrating vectors of $\mathbf{y}_t$. The necessary and sufficient conditions for cointegration are given by [see, for example, Engle and Granger (1987)] :[8]

$$\boldsymbol{\alpha}'\mathbf{A}(1) = 0 \text{ and } \boldsymbol{\alpha}'\boldsymbol{\mu} = 0. \tag{10.3.5}$$

The condition $\boldsymbol{\alpha}'\mathbf{A}(1) = 0$, which plays a central role in the analysis of cointegrated systems, also implies that $\mathbf{A}(1)\boldsymbol{\Sigma}\mathbf{A}(1)'$ is singular and there will therefore be some exact linear relationships between the cross-sectoral persistence measures, P_{ij} defined in (10.3.4). These relationships are given by $\boldsymbol{\alpha}'\mathbf{P} = 0$, where $\mathbf{P} = \{P_{ij}\}$ is the matrix of cross-sectoral persistence measures. The conditions $\boldsymbol{\alpha}'\mathbf{P} = 0$ and $\boldsymbol{\alpha}'\mathbf{A}(1) = 0$ are in fact mathematically equivalent. Also, $\boldsymbol{\alpha}'\mathbf{P} = 0$ implies that the matrix of persistence measures, $\mathbf{P}$, has rank $m - r$, and there are, therefore, only $m - r$ independent sources of random variations that can have persistence effects on the level of sectoral outputs. This is in line with the Stock and Watson (1988) characterisation of cointegrated systems in terms of common trends and represents an alternative formalization of the cointegration property in terms of independent sources of random variations that have persistence effects.

10.3.2 Aggregation and persistence

Suppose now we are interested in measuring the persistence effect of shocks at the level of aggregate output, Y_t, defined by

$$Y_t = \sum_{i=1}^{m} w_i y_{it} = \mathbf{w}'\mathbf{y}_t, \tag{10.3.6}$$

where $\mathbf{w}' = (w_1, w|_2, \ldots, w_m)$, is an $m \times 1$ vector of positive fixed weights. Under the multisectoral model (10.3.1) we have

$$\Delta Y_t = \mathbf{w}'\boldsymbol{\mu} + \mathbf{w}'\mathbf{A}(L)\boldsymbol{\epsilon}_t. \tag{10.3.7}$$

This specification is directly comparable to the *univariate* ARIMA models used in the literature for the measurement of persistence at the aggregate level. The main advantage of using (10.3.7) over the univariate aggregate models lies in the fact that by exploiting even very simple univariate time series specifications at the disaggregate level, we are still able to arrive at very-high-order ARIMA specification for the aggregate output, Y_t. For example, it is possible to obtain an ARIMA(m,$m-1$) specification for Y_t, even if the sectoral output growths are specified to follow independent AR(1) processes.[9] Thus, given

[7] To ensure that $\mathbf{a}'\mathbf{y}_t$ has bounded variance we also assume that $\mathbf{A}(L)$ is I-summable, namely that $\sum_{i=1}^{\infty} i|\mathbf{A}_i| < \infty$. This condition is satisfied when $\mathbf{A}(L)$ is the lag polynomial matrix in the Wold representation of a vector ARMA specification of $\Delta\mathbf{y}_t$.

[8] We are assuming that the variance matrix of the innovations, $\boldsymbol{\Sigma}$, is nonsingular.

[9] See, for example, Granger and Morris (1976) and the review of small scale aggregation in Granger (1990).

the difficulties involved in obtaining accurate estimates of high-order ARIMA processes using available aggregate time series, one possible way out would be to base the estimation of the aggregate persistence measure on (10.3.7) instead of relying on low-order ARIMA specifications of the aggregate output directly.[10]

Let P_y be the persistence measure of aggregate output obtained on the basis of the disaggregate specification (10.3.1). Applying the spectral density approach to (10.3.7) and using the conditional variance of ΔY_t as the scaling factor, we have

$$P_y^2 = \mathbf{w}'\mathbf{A}(1)\boldsymbol{\Sigma}\mathbf{A}(1)'\mathbf{w}/\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}. \tag{10.3.8}$$

This measure is directly comparable to the Campbell and Mankiw measure $a(1)$, derived using a univariate time series specification of aggregate output.[11]

It is important to note that the 'aggregate persistence measure', P_y, is valid irrespective of whether one or more of the sectoral outputs in y_t are trend- stationary. For example, in the case where output of all sectors except the output of sector i are trend-stationary we have

$$P_y = \{\sigma_{ii}w_i^2/\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}\}^{1/2}P_i,$$

where P_i is the persistence measure of sector i.

Clearly, $P_y = 0$ if $P_i = 0$, $i = 1, 2, \ldots, n$. The reverse is not, however, true. In principle we could have $P_i = 0$, even if none of the sectoral persistence measures is equal to zero. This arises when $\mathbf{y}_t$ is cointegrated and the aggregating $\mathbf{w}$ is proportional to one of the cointegrating vectors in $\boldsymbol{\alpha}$.[12] In general, however, the effect of cointegration among the sectoral outputs on P_y is complex and involves all the cross-sectoral persistence measures, P_{ij}. Using (10.3.4) and (10.3.8), we have

$$(\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w})P_y^2 = \sum_{i,j=1}^{m} w_i w_j \sigma_j j P_{ij}, \tag{10.3.9}$$

where P_{ij} are related through $\boldsymbol{\alpha}'\mathbf{P} = 0$ where $\mathbf{y}_t$ is cointegrated with the cointegrating matrix $\boldsymbol{\alpha}$.

An interesting specialisation of (10.3.9) arises when sectoral outputs are pairwise cointegrated. Consider first the simple case where $m = 2$, and let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ be the cointegrating vector. Then using the result in Proposition 1 and noting that $\sigma_{22}P_{12} = \sigma_{11}P_{21}$, we have

$$P_{21} = (-\alpha_1/\alpha_2)P_{11},$$
$$P_{11}/P_{22} = (\alpha_2/\alpha_1)^2(\sigma_2 2/\sigma_{11}).$$

Substituting these results in (10.3.9), it is now easily seen that[13]

$$\begin{aligned}(\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w})^{1/2}P_y &= |w_1\sigma_{11}^{1/2}P_1 - w_2\sigma_{22}^{1/2}| \text{ if } \alpha_1/\alpha_2 > 0,\\ &= w_1\sigma_{11}^{1/2}P_1 + w_2\sigma_{22}^{1/2}P_2, \text{ if } \alpha_1/\alpha_2 < 0,\end{aligned}$$

[10] The use of low-order ARIMA processes in the estimation of persistence at the aggregate levels has been crittcized by Cochrane (1988) and Christiano and Eichenbaum (1989).

[11] The counterpart of Cochrane's measure V, based on the disaggregated model, is given by

$$V_d = \{\mathbf{w}'\mathbf{A}(1)\boldsymbol{\Sigma}\mathbf{A}(1)'\mathbf{w}/\,\mathrm{V}(\Delta Y_t)\}.$$

[12] This follows immediately from (10.3.5) and (10.3.8).

[13] Notice that in the present case $\mathbf{w} = (w_1, w_2)'$, and $\boldsymbol{\Sigma}$ is the variance matrix of $(\epsilon_{1t}, \epsilon_{2t})'$.

where $P_i = \sqrt{P_{ii}} > 0$. Therefore, the value of the aggregate persistence measure crucially depends on whether in large samples the sectoral outputs are correlated positively (i.e., $\alpha_1/\alpha_2 < 0$) or negatively (i.e., $\alpha_1/\alpha_2 > 0$). Under the more likely case where $\alpha_1/\alpha_2 < 0$, P_y can be written as the weighted sum of the two sectoral persistence measures, namely,

$$P_y = \lambda_1 P_1 + \lambda_2 P2, \tag{10.3.10}$$

where $\lambda_i = (w_i^2 \sigma_{ii}/\mathbf{w}'\mathbf{\Sigma}\mathbf{w})^{1/2}, i = 1, 2$. This result readily extends to the m-sector case.

Proposition 3. *Let P_i be the sectoral persistence measures, P_y the aggregate persistence measure, and suppose thut the sectoral output levels $y_{it}, i = 1, 2, \ldots, m$, are pairwise cointegrated. Then, assuming that sectoral outputs are correlated positively, we have*

$$P_y = \sum_{i=1}^{m} \lambda_i P_i, \tag{10.3.11}$$

where $\lambda_i = (w_i^2 \sigma_{ii}/\mathbf{w}'\mathbf{\Sigma}\mathbf{w})^{1/2}, i = 1, 2, \ldots, m$. (See appendix **??** *for a proof.)*

The above result can also be written as

$$f_{\Delta y}^{1/2}(0) = \sum_{i=1}^{m} w_i f_{\Delta y_i}^{1/2}(0), \tag{10.3.12}$$

which provides a decomposition of the spectral density of the aggregate growth rate at zero frequency, $f_{\Delta y}(0)$, in terms of the spectral densities of the sectoral growth rates also evaluated at the zero frequency, $f_{\Delta y_i}(0)$.

10.3.3 Measurement of persistence in models with macroeconomic shocks

Consider now the model

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \d{}(L) v_t + \mathbf{A}(L)\boldsymbol{\epsilon}_t, \tag{10.3.13}$$

where v_t represents a *scalar* white-noise process with mean zero and constant variance σ_t^2, and $\d{}(L)$ is an $m \times 1$ vector of lag polynomials,

$$\mathbf{d}(L) = \mathbf{d}_0 + \mathbf{d}_1 L + \mathbf{d}_2 L^2 + \cdots, \quad \sum_{i=1}^{\infty} i|\mathbf{d}_i| < \infty.$$

This model provides a generalisation of (10.3.1) and allows for the effect of a common shock, v_t, on sectoral output growths in addition to the sector-specific shocks, ϵ_{it}. We refer to v_t as the 'macro' shock and in this paper identify it with the unexpected growth of money supply.[14] We use the following specification for the money growth equation:

$$\Delta m_t = \boldsymbol{\beta}' \mathbf{z}_t + v_t, \tag{10.3.14}$$

[14] Other types of macroeconomic shocks, such as unexpected changes in oil prices or exchange rates, may also be considered. Indeed, the inclusion of more than one type of shock can be easily accomodated within this framework, subject to identifying restrictions similar to that discussed below [see Lee et al. (1992)].

where $\mathbf{z}_t$ is a vector of predetermined variables to be specified more fully in the next section. To ensure that the parameters of the equation systems (10.3.13) and (10.3.14) are identified, we assume that v_t and $\boldsymbol{\epsilon}_t$ are uncorrelated.

The identifying nature of the restriction $\mathrm{Cov}(\boldsymbol{\epsilon}_t, v_t) = \boldsymbol{\Sigma}_{\epsilon v} = 0$ can be easily shown in the case of normally distributed shocks. Suppose $\boldsymbol{\Sigma}_{\epsilon v} \neq 0$. Then, under the normality assumption we may write

$$\boldsymbol{\epsilon}_t = (\sigma_v^{-2}\boldsymbol{\Sigma}_{\epsilon v})v_t + \mathbf{u}_t,$$

where v_t and $\mathbf{u}_t$ are now uncorrelated. Using this result in (10.3.13) gives

$$\Delta y_t = \boldsymbol{\mu} + [\mathbf{d}(L) + \sigma_v^{-2}\mathbf{A}(L)\boldsymbol{\Sigma}_{\epsilon v}]v_t + \mathbf{A}(L)\mathbf{u}_t. \tag{10.3.15}$$

The joint maximum likelihood (ML) estimation of (10.3.15) and (10.3.14) now yields consistent estimates of $\mathbf{A}(L)$ and $\mathbf{d}^*(L) = \mathbf{d}(L) + \sigma_v^{-2}\mathbf{A}(L)\boldsymbol{\Sigma}_{\epsilon v}$. However, it is not possible to recover a consistent estimate of $\mathbf{d}(L)$, unless $\Sigma_{\epsilon v} = 0$.[15]

Under the above assumptions, it is now possible to decompose the persistence of output fluctuations into the components due to 'money' and due to 'other' shocks. This can be done both at the level of individual sectors and for the economy as a whole. Here we focus on the decomposition of the aggregate persistence measure, P_y. Using (10.3.13), the spectral density of $\Delta Y_t = \mathbf{w}'\Delta y_t$ at zero frequency is now given by

$$2\pi f_{\Delta y}(0) = \sigma_v^2[\mathbf{w}'\mathbf{d}(1)]^2 + \mathbf{w}'\mathbf{A}(1)\boldsymbol{\Sigma}\mathbf{A}(1)'\mathbf{w}.$$

Scaling this by the conditional variance of ΔY_t,

$$\mathrm{V}(\Delta Y_t|\Omega_{t-1}) = \sigma_v^2(\mathbf{w}'\mathbf{d}_0)^2 + \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w},$$

yields the following decomposition of P_y:

$$P_y^2 = \lambda P_m^2 + (1-\lambda)P_O^2, \tag{10.3.16}$$

where P_m is the component of the persistence measure due to 'monetary' shocks, P_O is the component due to 'other' shocks:

$$P_m = \mathbf{w}\mathbf{d}(1)/\mathbf{w}'\mathbf{d}_O, \tag{10.3.17}$$

$$P_O = \{\mathbf{w}'\mathbf{A}(1)\boldsymbol{\Sigma}\mathbf{A}(1)'\mathbf{w}/\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}\}^{1/2}, \tag{10.3.18}$$

and λ is the mixture coefficient defined by

$$\lambda = \sigma_v^2(\mathbf{w}'\mathbf{d}_O)^2/(\sigma_v^2(\mathbf{w}'\mathbf{d}_O)^2 + \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}). \tag{10.3.19}$$

Similar decompositions can also be obtained at the sectoral levels. The formulae for the decomposition of P_i, the sectoral persistence measures, can be obtained from (10.3.16)–(10.3.19) by replacing $\mathbf{w}$ in these expressions with $\mathbf{e}_i$. Note that care must be taken in interpreting the persistence measure due to 'monetary' shocks since misspecification in (10.3.13) can result in biased estimates for $\mathbf{d}(L)$. In particular, the omission of other important macroeconomic shocks which are negatively corre- lated to 'money' shocks (through, for example, a feedback rule in which monetary policy aims to offset the effects of the omitted shock) will result in downward bias in the measure of persistence due to 'money' shocks.

[15] This results also highlights the general difficulty involved with the decomposition of output innovations into 'supply' and 'demand' shocks, and shows that such a decomposition is only meaningful if one is prepared to assume that the two types of shocks are contemporaneously uncorrelated. [See Blanchard and Quah (1989).]

10.4 Empirical results: Measures of sectoral and aggregate persistence for the U.S. economy

The persistence effect of shocks to the U.S. real GNP has been extensively investigated at the aggregate level. [See, for example, Campbell and Mankiw (1987a,b), Clark (1987a), Cochrane (1988), Watson (1986), Haubrich and Lo (1989), and Durlauf (1989).] Here we apply the methods discussed in the previous section and provide a disaggregated analysis based on a multisectoral model composed of ten sectors[16] The available data set covers the 1947–87 period.

10.4.1 Testing for unit roots at the sectoral levels

Table 10.1: Augmented Dickey-Fuller statistics for tests of a unit root[a] in U.S. sectoral outputs (in logs); 1952–1987.

	Dickey-Fuller statistics[b]			
Sectors	ADF(1)	ADF(2)	ADF(3)	ADF(4)
1. Agriculture	−1.58	−0.35	0.06	0.14
2. Mining	−0.17	−0.10	−0.30	−0.52
3. Construction	−2.14	−2.14	−2.15	−2.17
4. Dur. manuf.	−2.99	−2.96	−2.67	−2.78
5. Nondur. manuf.	−1.82	−0.93	−0.52	−0.38
6. Transport	−2.95	−2.73	−2.65	−2.72
7. Utilities	−0.51	−0.52	−0.50	−0.53
8. Trade	−2.65	−2.31	−2.04	−1.54
9. Services	−1.46	−1.22	−0.78	−0.62
10. Government	−1.99	−1.11	−0.70	−0.60

[a]The underlying augmented Dickey-Fuller regressions contain a simple linear time trend and are based on the same number of observations.

[b]The (asymptotic) 5% and 10% critical values are −3.54 and −3.20 respectively.

The first stage in the analysis is to test for unit roots in sectoral outputs (measured in logarithms).[17] Table 10.1 gives the Augmented Dickey-Fuller (ADF) statistics for four different lag lengths computed over the sample period 1952–87.[18] [See Fuller (1976) and Dickey and Fuller (1981).] All the underlying ADF regressions are estimated on the same data set and include simple linear trends. The relevant 5% and 10% critical values for

[16] The sources of the data and the detals of the sectoral classifications are given m appendix **??**.

[17] Although the aggregate persistence measure, P_Y, proposed in this paper is applicable even if (log) outputs in one or more sectors arc trend-stationary (see section 3.2), nevertheless at the estimation stage where finite-order AR or ARMA processes are fitted to the data, it is more appropriate to exclude sectors with trend-stationary output processes from the analysis. Ambiguities, however, arise when we are not sure whether the output of a particular sector is trend- stationary or not.

[18] The choice of the sample period in the computation of the ADF statistics is governed by the available data and the highest order chosen for the ADF text, namely 4.

the ADF statistics are equal to −3.54 and −3.20. respectively.[19] None of the tests come even close to rejecting the unit root hypothesis, and this is true of all the sectors. With the exception of the durable manufacturing, this finding is in accordance with the results reported in (Durlauf, 1989, table 6). The difference in the two results in the case of the durable manufacturing seems to be primarily due to the different procedure used by Durlauf to correct the simple DF statistic for the residual serial correlation.[20] To check the robustness of the unit root tests to the specification of the trend, following Perron (1988, 1989a), we also computed ADF statistics assuming a different growth path before and after the first oil shock in 1973 for each sector. This is the 'changing growth' model in Perron (1989a). The results are summarized in table 10.2. Only in the case of the 'government' sector is there a clear cut case against the unit root hypothesis. For all the other sectors the hypothesis is either not rejected, or if rejected, the rejection has been confined to one out of the four ADF statistics that were computed for each sector. With the exception of the 'government' sector, Perron's alternative trend specification does not significantly alter the conclusion reached earlier, namely that the hypothesis of a unit root in sectoral outputs cannot be rejected.

Table 10.2: ADF statistics for tests of a unit root in sectoral outputs under Perron's 'changing growth' model;[a] 1952–1987.

	Dickey-Fuller statistics[b]			
Sectors	ADF(1)	ADF(2)	ADF(3)	ADF(4)
1. Agriculture	−4.21[b]	−3.05	−2.73	−2.59
2. Mining	−2.55	−2.39	−3.10	−4.74[b]
3. Construction	−2.10	−1.61	−1.79	−1.75
4. Dur. manuf.	−3.15	−3.24	−3.01	−3.21
5. Nondur. manuf.	−4.39[b]	−3.24	−2.77	−2.61
6. Transport	−2.22	−1.99	−1.94	−2.19
7. Utilities	−1.82	−1.75	−1.97	−1.94
8. Trade	−3.45	−3.21	−3.01	−2.38
9. Services	−3.40	−3.52	−3.06	−2.81
10. Government	−6.62[b]	−5.26[b]	−4.65[b]	−4.80[b]

[a]The Augmented Dickey-Fuller (ADF) statistics are based on the OLS residuals computed over the period 1947–87 from the regressions of sectoral outputs (in logs) on an intercept, a simple linear trend, t, and the broken trend line defined as $DT_t^* = t - T_\beta$ if $t > T_\beta$ and 0 otherwise, where $T_\beta = 27$ is the time break in 1973 and $t = 1$ in 1947. See model B in Perron (1989a).

[b]Statistical significance at the 5% level. For relevant critical values, see table V.B in Perron (1989a), with $\lambda = T_\beta / T = 0.65$.

[19] These critical values are obtained using the simulated response surfaces gven in (MacKinnon, 1990, table I).

[20] Durlauf (1989) employs the Phillips-Perron type correction instead of the augmentation of the simple DF regression by lagged values of Δy_{it}. [See Phillips (1987a) and Phillips and Perron (1988).] But recent Monte Carlo evidence by Schwert (1989) suggests that the ADF procedure tends to have better small sample properties than the Phillips-Perron method.

Finally, before we can utilize the multisectoral model (10.3.1), it is important to show that output growths, Δy_{it}, are in fact stationary. Table 10.3 gives the ADF statistics for the test of a unit root in sectoral growth rates. For most sectors (six out of ten) there is a clear-cut rejection of the unit root hypothesis, and even for the sectors where the evidence is mixed the hypothesis is still rejected on the basis of the simple DF and the ADF(1) statistics. We therefore proceed with the presumption that sectoral growth rates are stationary and that the multisectoral model is a suitable framework for the analysis of persistence in the U.S. post-war economy.

Table 10.3: Augmented Dickey-Fuller statistics for tests of a unit root in U.S. sectoral output growths,[a]
1952–1987.

	Dickey-Fuller statistics[b]			
Sectors	DF	ADF(1)	ADF(2)	ADF(3)
1. Agriculture	-6.73^b	-6.34^b	-4.56^b	-3.25^b
2. Mining	-5.15^b	-3.81^b	-2.11	-1.42
3. Construction	-3.78^b	-4.43^b	-3.05^b	-2.70
4. Dur. manuf.	-6.21^b	-4.95^b	-4.52^b	-3.67^b
5. Nondur. manuf.	-6.23^b	-6.26^b	-4.91^b	-3.76^b
6. Transport	-5.60^b	-4.69^b	-3.79^b	-2.52
7. Utilities	-5.39^b	-3.25^b	-2.28	-2.16
8. Trade	-5.18^b	-4.68^b	-4.22^b	-4.22^b
9. Services	-4.68^b	-3.91^b	-3.70^b	-3.00^b
10. Government	-10.46^b	-11.10^b	-10.43^b	-8.65^b

[a]The ADF regressions contain an intercept term but not a time trend.

[b]Statistical significance at the 5% level.

10.5 Estimates of the persistence measures

To obtain estimates of the persistence measures we need consistent estimates of $\mathbf{A}(L)$ and $\mathbf{\Sigma}$, the parameters of the multisectoral model (10.3.1). For this purpose we initially considered two different versions of a second-order vector autoregressive, VAR(2), version of (10.3.1): a fully unrestricted version.[21]

$$M_1: \quad (\mathbf{I}_m - \mathbf{C}_1 L - \mathbf{C}_2 L^2)\Delta\mathbf{y}_t = \mathbf{C}(L)\Delta\mathbf{y}_t = \mathbf{a} + \boldsymbol{\epsilon}_t \tag{10.5.1}$$

where $\mathbf{C}_1 = \{c_{1,ij}\}$ and $\mathbf{C}_2 = \{c_{2,ij}\}$ are $m \times m$ matrices and $\mathbf{a}$ is a vector of fixed constants: and a version that restricts the coefficients of $\Delta y_{j,t-1}$ (and $\Delta y_{j,t-2}$ in the ith output growth equation to be the same for all $j (j \neq i)$. That is,

$$M_2: \quad \Delta y_{it} = a_i + c_{1,ii}\Delta y_{i,t-1} + c_{2,ii}\Delta y_{i,t-2} + b_{1i}\Delta\mathbf{y}_{-i,t-1} + b_{2i}\Delta_{-i,t-2} + \epsilon_{it}, \quad i = 1, 2, \ldots, m,$$

[21] Notice that given the available sample size, the highest order VAR that we can fit to the data is a second-order one.

where

$$\Delta y_{-i,t} = \sum_{j=1, j \neq i}^{m} \Delta y_{jt}.$$

Model M_2 imposes $2m(m-2)$ parametric restrictions on M_1, and has the interpretation that output growth in sector i is related to a simple aggregation of output growths in the rest of the economy. In terms of the parameters of M_1 these restrictions are

$$c_{1,ij} = b_{1i}, \quad c_{2,ij} = b_{2i}, \quad i, j = 1, 2, \ldots, m, j \neq i. \tag{10.5.2}$$

With $m = 10$, M_2 imposes 160 restrictions on M_1. Despite these restrictions, this model still allows for feedbacks from the rest of the economy to the ith sector. In addition to M_1 and M_2 we also estimated a third model, M_3, which imposed further parameter restrictions on M_2 by dropping regressors in M_2 whose coefficients had a t-ratio (in absolute value) less than unity. All the three models were estimated by the Full Information Maximum Likelihood (FIML) method over the period 1955–87.[22] Table 10.4 gives

Table 10.4: Maximised log-likelihood values[a]

Models	LLF	N
M_1	897.99	210
M_2	825.62	50
M_3	819.87	31

[a] LLF is the maximised log-likelihood values and N is the number of estimated regression coefficients.

the maximised log-likelihood values (LLF) and the number of regression coefficients estimated under each of the three models. The log-likelihood ratio statistic for the test of M_2 against M_1 is equal to 144.74 and for the test of M_3 against M_2 it is equal to 11.50. Both these statistics are well below the 95 percent critical values of the chi-squared distribution with 160 and 19 degrees of freedom, respectively.[23] The nonrejection of M_2 against M_1 is particularly noteworthy, and has important consequences for the precision with which persistence measures are estimated. The use of an overparameterised model such as M_1 will, in general, lead to poorly determined persistence measures, and this can be seen clearly in the estimates of the aggregate persistence measures obtained for the U.S. economy on the basis of models M_1 to M_3 (see the last row of table 10.5).[24] The highly overparameterised model M_1 yields a large but extremely unreliable estimate of the aggregate persistence measure, $\widehat{P}_y$. Under M_1, the (asymptotic) standard error of $\widehat{P}_y$

[22] Clearly, in the case of the unrestricted VAR(2) model, M_1, the FIML and the OLS estimates coincide. In the case of the restricted VAR models, M_2 and M_3, the FIML estimates are computed by iterating on the Seemingly Unrelated Regression Equations (SURE) estimates, described in Zellner (1962).

[23] Notice that for degrees of freedom $v > 100$, we have

$$\sqrt{2\chi_v^2} - \sqrt{(2v-1)} \overset{a}{\sim} \mathrm{N}(0, 1).$$

[24] The aggregate persistence measure, P_y, relates to $Y_t = \sum_{i=1}^{m} y_{it}$, and is computed according to (10.3.8) with $\mathbf{w}' = (1, 1, \ldots, 1)$, $\mathbf{A}(1) = \mathbf{I}_m - \mathbf{C}_1 - \mathbf{C}_2$, and $m = 10$.

Table 10.5: Sectoral and aggregate persistence measures.[a]

	Models[b]		
Sectors	M_1	M_2	M_3
1. Agriculture	2.75	0.84	0.89
	(3.73)	(0.01)	(0.004)
2. Mining	2.61	1.77	1.38
	(3.53)	(0.51)	(0.11)
3. Construction	4.01	3.45	3.64
	(5.02)	(1.46)	(1.54)
4. Dur. manuf.	1.42	0.91	1.00
	(0.72)	(0.04)	(0.02)
5. Nondur. manuf.	1.82	0.58	0.64
	(2.31)	(0.01)	(0.003)
6. Transport	1.63	1.08	1.17
	(0.93)	(0.04)	(0.01)
7. Utilities	5.22	1.77	1.76
	(7.06)	(0.28)	(0.24)
8. Trade	1.43	0.66	0.73
	(1.37)	(0.01)	(0.01)
9. Services	4.44	2.77	3.75
	(6.00)	(2.73)	(6.93)
10. Government	4.69	2.45	2.65
	(5.07)	(0.20)	(0.18)
Aggregate output	2.09	0.76	0.83
	(2.67)	(0.1473)	(0.0849)

[a]The sectoral persistence measures, $P_i = \sqrt{P_{ii}}$, are estimated using (10.3.4). The aggregate persistence measure P_y, is estimated using (10.3.8) with $\mathbf{w}' = (1, 1, \ldots, 1)$. The figures in parentheses are asymptotic standard errors, computed according to the formulae given in appendix **??**.

[b]M_1 is an unrestricted VAR(2) specification of the multisectoral model, and M_2 and M_3 are restricted versions of M_1. See the text for further details.

is estimated to be equal to 2.67, which is well in excess of the value estimated for P_y itself![25] The situation, however, is very different when we consider the estimates of P_y based on the more parsimonious models M_2 and M_3. The best estimate of P_y (in the sense of having the least variance) is obtained under model M_3.[26] It is equal to 0.83 with an (asymptotic) standard error of 0.0849. This estimate is well below unity and is more in line with the estimates obtained by Watson (1986) and Clark (1987a) than the ones obtained by Campbell and Mankiw (1987a).[27]

[25] A derivation of the asymptotic variance of $\widehat{P}_y$, together with the variance of other persistence measures, can be found in appendix **??**.

[26] Also, recall that M_3 could not be rejected against either M_2 or M_3 (see table 10.4).

[27] Notice that our measure of aggregate output is based on the sum of the logarithms of sectoral outputs, while the measure of aggregate output used in the literature is the logarithm of the sum of

For a more direct comparison of the above estimate of the aggregate persistence measure with an estimate based on a univariate model, we fitted a number of ARMA specifications directly to the aggregate output growth defined by $\Delta Y_t = \sum_{i=1}^{m} \Delta y_{it}$.[28] The results for ARMA processes of orders (i, j), $i, j = 1, 2, 3, 4$, are summarised in tables 10.6a and 10.6b. The maximised values of the log-likelihood function given in table 10.6a are all close to one another and the only model which come close to rejecting the ARMA (1,1) specification at the 5% level is ARMA(1,3). Overall, there is very little to choose between the different ARMA processes. In view of this, in table 10.6b we give the estimates of the aggregate persistence measures for all the ARMA specifications. These estimates are all above unity and fall in the range 1.00–1.85, which are clearly compatible with Campbell and Mankiw's estimates, but not with our estimate of P_y based on the multisectoral model, M_3. Turning now to the persistence of output fluctuations at the sectoral levels, using (10.3.10) we also estimated sector-specific persistence measures, P_i, under models M_1 to M_3. The results are summarised in table 10.5. In view of the highly overparameterised nature of M_1, we shall confine our attention to the estimates obtained under the restricted models M_2 and M_3. These two sets of estimates give a similar pattern of persistence across the sectors, with high values estim- ated for 'construction', 'services', and 'government' sectors, and low values estimated for 'agriculture', 'trade', and 'nondurable manufacturing' sectors.

Table 10.6a: Maximised log-likelihood values for different ARMA models fitted to ΔY_t: 1955–1987

	Order of AR			
Order of MA	1	2	3	4
1	80.69	80.71	82.43	83.10
2	80.69	82.63	82.79	83.11
3	83.59	83.84	83.86	83.89
4	83.85	83.87	83.88	83.90

The relationship between the estimates of sectoral persistence measures, $\widehat{P}_y$, given in table 10.5 is complex and, as shown in section 3.1, depends on the pattern of cointegration among the sectoral outputs. We first tested for the presence of cointegration among the sectoral output series using Johansen (1988, 1989) maximum likelihood procedure[29] and found evidence of between five and eight cointegrating vectors, depending on whether we use the trace or the maximum eigenvalue test criterion. These results indicate that there are substantially fewer than ten independent sources of random variation that affect sectoral outputs, although they should be treated with caution, given the paucity of degrees

sectoral outputs. However, the two measures are very closely related. The correlation between the output growth rates calculated using the two measures is equal to 0.97.

[28] To estimate the ARMA models we used the exact ML algorithm proposed in Pesaran (1988a), which allows for the possibility of obtaining ML estimates on the unit circle. This is an important consideration, especially when the aim is the estimation of persistence measure. On this see also Campbell and Mankiw (1987a).

[29] We computed both of Johansen's proposed test statlstlcs, namely the 'trace' and the 'maximal eigenvalue' statistlcs, over the period 1955–87 on the basis of a VAR(2) model allowing for an intercept term, a linear time trend, and a time trend in the underlyng data generatIon process. The computations were carried out on *Microfit 3.30* [Pesaran and Pesaran (1991a)].

Table 10.6b: Aggregate persistence measures estimated on the basis of ARMA models fitted to ΔY_t: 1955–1987

	Order of AR			
Order of MA	1	2	3	4
1	1.22	1.17	0.93	1.15
	(0.0093)	(0.228)	(0.185)	(0.352)
2	1.20	1.53	1.03	1.16
	(0.393)	(0.197)	(0.238)	(0.376)
3	1.00	1.07	1.06	1.07
	(0.307)	(0.316)	(0.329)	(0.382)
4	1.10	1.09	1.08	1.09
	(0.357)	(0.351)	(0.361)	(0.405)

of freedom. We then tested the hypothesis of pairwise cointegration of sectoral outputs. Surprisingly, we found very little statistically significant evidence of pairwise cointegration. The hypothesis of pairwise cointegration was easily rejected for the 'agriculture', 'mining', and 'utilities' sectors. None of these sectors showed significant evidence (at the 5% level) of cointegration with the other sectors in the economy. We only found significant evidence of pairwise cointeg- ration in the case of sectors (3,4), (3,5), (3,8), (5,10), and (8,9).[30] We therefore do not expect to find a simple relationship between the sectoral and the aggregate persistence measures in the case of the U.S. economy.

10.5.1 Persistent effects of 'monetary' and 'other' shocks

In this section we provide evidence on the relative importance of 'monetary' and 'other' shocks for the long-run evolution of the U.S. output. In view of the highly overparameterised nature of M_1, and since M_2 could not be rejected against M_1, we base our analysis on model M_2 and augment it with the current and the one-period-lagged values of the unanticipated growth of money supply, v_t. This augmented model, which we denote by $\widetilde{M}_2$, may be written as

$$\widetilde{M}_2 : \quad \mathbf{C}(L)\Delta \mathbf{y}_t = \mathbf{a} + (\gamma_0 + \gamma_1 L)v_t + \boldsymbol{\epsilon}_t \tag{10.5.3}$$

where $\gamma_j' = (\gamma_{j1}, \gamma_{j2}, \ldots, \gamma_{jm}), j = 0, 1$. Recall that under M_2 there are $2m(m-2)$ restrictions on the coefficients of $\mathbf{C}(L)$ and these are given by (10.5.2). For the money supply growth equation we adopt the following specification:

$$\Delta m_t = \beta_0 + \beta_1 \Delta m_{t-1} + \beta_2 \Delta m_{t-2} + \beta_3 \Delta G_{t-1} + \beta_4 UN_{t-1} + v_t \tag{10.5.4}$$

where ΔG_t is the rate of change of real federal government expenditure, $UN_t = \log[RU_t/(1-RU_t)]$, and RU_t is the unemployment rate.[31] This specification is a simplified version of the money supply growth equation used in Pesaran (1991b) and avoids the complications

[30] Notice that in this application Johansen's test procedure does not satisfy the transitivity property of pairwise cointegration. Given the evidence of cointegratton between sectors 4 and 3 and between sectors 3 and 5, we would expect, *a priori*, to find evidence of cointegration between sectors 4 and 5, but we do not.

[31] See appendix **??** for data sources.

associated with the use of the Barro (1977) type money growth equation which includes the contemporaneous effect of real federal government expenditure.[32] The above output-money equations can also be viewed as a multisectoral, stochastic trend version of the 'New Classical' model where money shocks can affect output levels in the short run. but not in the long run. Under this interpretation $\widetilde{M}_2$ will also be subject to the following further restrictions:

$$H_{NC}: \quad \gamma_0 + \gamma_1 = 0.$$

These restrictions impose zero persistence for money shocks in all sectors.

The consistent and efficient estimation of the parameters of the multivariate model, (10.5.3) and (10.5.4), is discussed in detail in Pesaran (1991b).[33] In order to obtain efficient estimators and avoid some of the difficulties associated with the use of two-step estimators we estimated the parameters of the output and money equations jointly by the FIML method over the period 1955–87.[34] Here we focus on the estimates of the money supply shock coefficients and report these estimates together with their asymptotic standard errors in table 10.7.[35] In this table we also give Wald statistics for testing the joint hypothesis of zero restrictions on the money supply shocks, $\gamma_{0i} = \gamma_{1i} = 0$, and for testing the new classical hypothesis H_{NC}, i.e., $\gamma_{0i} + \gamma_{1i} = 0$, for $i = 1, 2, \ldots, m$.

The results clearly show that the effect of money supply shocks on output is not uniform across the sectors. Money shocks have no statistically significant long-term, or even short-term, effects on outputs in the 'agricultural', 'durable manufacturing', and the 'government' sectors. In the case of 'mining' and 'utilities', money shocks have significant short-term effects, but these effects tend to die out in the long run. For the five remaining sectors, however, money shocks have statistically significant effects on output levels both in the short run and in the long run. Overall, the evidence on the new classical hypothesis is mixed. It is upheld for half of the sectors studied and rejected for the rest. There is clearly a need for further empirical analysis of the possible short-run and long-run impact of monetary shocks on sectoral outputs.

In view of the above results we base our estimates of the persistence measures and their decomposition on a restricted version of $\widetilde{M}_2$, obtained in the following manner:

(i) In the case of sectors 1, 4, and 10, we imposed the zero restrictions $\gamma_{0i} = \gamma_{1i} = 0$.

(ii) In the case of sectors 2 and 7, we imposed the new classical restriction, $\gamma_{0i} + \gamma_{1i} = 0$. As table 10.7 shows, none of these restrictions can be rejected at the 5% level. We also dropped regressors whose coefficients were less than unity (in absolute value). We refer to this further restricted model as $\widetilde{M}_3$.[36]

[32] On this, see Pesaran (1982) for further details.

[33] The problem of estimatlng univariate 'surprise' models is discussed in Pagan (1984, 1986) and reviewed in (Pesaran, 1987, ch 7).

[34] The use of two-step estimators, whereby r_t is estimated first bv the application of the OLS method to (10.5.4) and then used as regressors in (10.3.12), besides being subject to the familiar 'generated regressor' problem, is further complicated in the present multivariate application due to the contemporaneous correlation across the output dlsturbances, ϵ_{it}.

[35] The computation of the FIML estimators were carried out on GAUSS by iterating on a multivariate generalisation of the double-length regression proposed in Pagan (1986). Also see (Pesaran, 1987, pp. I77–179). The details of the algorithm and the associated computer codes can be obtained from the authors on request.

[36] Model $\widetilde{M}_3$ imposes 32 restrictions on $\widetilde{M}_2$. To ensure that we have not inadvertently imposed an invalid restriction on $\widetilde{M}_2$ we also tested the overall validity of the 32 restrictions by the likelihood ratio procedure. The value of the log-likelihood ratio statistic turned out to be equal to 17.23, which is well below the 95% critlcal value of the chi-squared distribution with 32 degrees of freedom.

Table 10.7: FIML estimates of the coefficients of the money supply shocks.

	Coefficient estimates[a]		Test statistics[b]	
Sectors	$\widehat{\gamma}_{0i}$	$\widehat{\gamma}_{1i}$	$\gamma_{0i} = \gamma_{1i} = 0$	$\gamma_{0i} + \gamma_{1i} = 0$
1. Agriculture	−1.10	−0.50	4.35	3.97
	(−1.77)	(−0.81)		
2. Mining	−0.16	1.22	6.56[c]	2.62
	(−0.30)	(2.56)		
3. Construction	1.59	−0.06	14.05[c]	7.49[c]
	(3.74)	(−0.14)		
4. Dur. manuf.	1.04	−0.00	1.34	0.81
	(1.14)	(−0.00)	1.34	0.81
5. Nondur. manuf.	1.19	0.26	12.11[c]	9.23[c]
	(3.30)	(0.74)		
6. Transport	1.22	0.52	14.63[c]	13.06[c]
7. Utilities	−0.32	0.89	7.01[c]	1.58
	(−0.87)	(2.61)		
8. Trade	0.97	0.09	9.39[c]	6.24[c]
	(2.99)	(0.27)		
9. Services	0.38	0.12	9.23[c]	7.65[c]
	(2.76)	(0.83)		
10. Government	0.16	−0.01	1.39	0.77
	(1.17)	(−0.10)		

[a]The estimates $\widehat{\gamma}_{01}$ and $\widehat{\gamma}_{11}$ respectively refer to the FIML estimates of the coefficients of the current and the one-period-lagged unanticipated money growth variable, v_t, in sector i. (The figures in parenthesees are asymptotic t-ratios.)

[b]The test statistics are the Wald statistics for tests of the hypotheses $\gamma_{0i} = \gamma_{1i} = 0$ and $\gamma_{0i} + \gamma_{1i} = 0$, respectively.

[c]Statistical significance at the 5% level.

Using the parameter estimates obtained under the restricted model we com- puted the estimates reported in table 10.8 for the persistence measures decomposed into 'money' and 'other' shocks.[37] At the aggregate level, the persistence of 'money' shocks, P_m, is estimated to be 1.85 with an (asymptotic) standard error of 0.55. The persistence effects of 'money' shocks on aggregate output are, therefore, statistically significant, but the estimate of P_m is subject to a wide margin of uncertainty. However, it is important to note that despite the statistical significance of the long-term impact of money shocks on output, the contribution of P_m to the total persistence measure is rather small. This is true of all the sectors and can be seen clearly from a comparison of the last two columns

[37] The relevant formulae for the decomposition of persistence measures are given at the end of section 3.3. Also see appendix **??** for the derivation of their asymptotic variances of the persistence measures estimated by the FIML method.

Table 10.8: Decomposition of sectoral and aggregate persistence measures by the type of shocks.[a]
FIML estimates 1955–1987.

Sectors	'Money' shocks	'Other' shocks	Total
1. Agriculture	0.00	0.76	0.76
		(0.05)	(0.05)
2. Mining	0.00	1.31	1.25
		(0.25)	(0.23)
3. Construction	0.50	4.93	4.37
	(0.56)	(1.88)	(1.64)
4. Dur. manuf.	0.00	0.69	0.69
		(0.04)	(0.04)
5. Nondur. manuf.	0.55	0.58	0.58
	(0.22)	(0.04)	(0.04)
6. Transport	1.57	0.87	0.98
	(0.29)	(0.04)	(0.05)
7. Utilities	0.00	4.71	4.41
		(2.79)	(2.61)
8. Trade	0.67	0.58	0.60
	(0.15)	(0.05)	(0.05)
9. Services	3.36	2.23	2.38
	(1.57)	(0.99)	(1.03)
10. Government	0.00	2.00	2.00
		(0.24)	(0.24)
Aggregate output	1.83	0.62	0.67
	(0.55)	(0.0763)	(0.0720)

[a]The decomposition of aggregate and sectoral persistence measures are carried out using the formulae (10.3.16)–(10.3.19) and their counterparts at the sectoral levels. The figures are computed using the FIML estimates of model $\widetilde{M}_3$ defined in the text. The figures in parentheses and asymptotic standard errors, computed according to the formulae given in appendix **??**.

of table 10.8. In the case of aggregate output the difference between the persistence measure due to 'other' shocks and the total persistence measure is only 0.05. The reason for this is primarily due to the fact that the size of money shocks (as measured by σ_v^2) compared to the size of the other shocks (as measured by $\mathbf{w}'\mathbf{\Sigma}\mathbf{w}$) has been very small over the sample period [see relations (10.3.16) and (10.3.19)].

A comparison of the results in tables 10.8 and 10.5 also shows that by including money shocks explicitly in the output growth equations it has in fact been possible to obtain a more precisely determined estimate of the aggregate persis- tence measure, as compared to the 'best' estimate obtained on the basis of model M_3, i.e., the model excluding the money shocks. The estimate of the aggregate persistence measure based on model $\widehat{M}_3$ is 0.67 (0.072) as compared to the estimate of 0.83 (0.085) based on model M_3.

Figures in brackets are asymptotic standard errors.[38] In conclusion, while it is true that estimating long-run effects from finite data sets is, in general, a hazardous undertaking, the results reported in this paper show that by utilizing information on sectoral outputs or other relevant information on variables other than the past history of the variable under investigation, it is possible to reduce the margin of errors involved in the estimation of the aggregate persistence measure. In this paper we have only analyzed the effect of incorporating money shocks in the multisectoral model. Other possibilities would be to try oil price shocks and shocks to the capital and foreign exchange markets.

0.A Proof of Proposition 1

We present a proof by induction. Suppose that the proposition holds for $m = s$. That is,

$$P_y(s) = \sum_{i=1}^{s} \lambda_i(s) P_i, \tag{0.A.1}$$

where $P_y(s)$ represents the aggregate persistence measure for the s-sector aggregate, $Y_t(s) = \sum_{i=1}^{s} w_i y_{it}$, $\lambda_i(s) = \{w_i^2 \sigma_{ii} / \mathbf{w}_s' \boldsymbol{\Sigma}_s\}^{1/2}$, $i = 1, 2, \dots, s$, $\mathbf{w}_s' = (w_1, w_2, \dots, w_s)$, and $\boldsymbol{\Sigma}_s$ is the variance matrix of $(\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{st})'$. Consider now adding a further sector to the aggregate, $Y_t(s)$. Namely,

$$Y_t(s+1) = Y_t(s) + w_{s+1} y_{t,s+1}. \tag{0.A.2}$$

Since the cointegration property is transitive it follows from the pairwise cointegration condition that $Y_t(s)$ and $y_{t,s+1}$ are also cointegrated. Moreover, since by assumption $y_{t,s+1}$ and $y_{1t}, y_{2t}, \dots, y_{st}$ are all pairwise positively correlated, it follows that $Y_t(s)$ and $y_{t,s+1}$ will also be positively correlated. Hence, the result (10.3.10) in the text obtained for the case $m = 2$ is also applicable to the right-hand side components of (0.A.2) and we have

$$P_y(s+1) = \mu_1 P_y(s) + \mu_2 P_{s+1}, \tag{0.A.3}$$

where the weights in this case are given by

$$\mu_1 = \{\mathbf{w}' \boldsymbol{\Sigma}_s \mathbf{w}_s / \mathbf{w}_{s+1}' \boldsymbol{\Sigma}_{s+1} \mathbf{w}_{s+1}\}^{1/2},$$

$$\mu_2 = \{\mathbf{w}' \boldsymbol{\Sigma}_s \mathbf{w}_s / \mathbf{w}_{s+1}' \boldsymbol{\Sigma}_{s+1} \mathbf{w}_{s+1}\}^{1/2},$$

Substituting from (0.A.2) in (0.A.3) and after some algebraic simplifications, we have

$$P_y(s+1) = \sum_{i=1}^{s+1} \lambda_i(s+1) P_i,$$

which establishes that if the proposition holds for $m = s$ it will also hold for $m = s + 1$. But we have already established that the proposition holds for $m = 2$ so it should hold for any m. **Q.E.D.**

[38] Given the asymptotic nature of our results, it may be worthwhile to consider the small properties of the persistence measures using Monte Carlo techniques. This is beyond the scope of the present paper however.

0.B Derivation of the variance of the persistence measures

This appendix gives a derivation of the variance of persistence measures estimated on the basis of the multisectoral model (10.5.3) that contains the money supply shocks v_t. Clearly the results are also applicable to the estimates of the persistence measures based on the VAR model (10.5.1) which does not include the money shocks. Here the derivations are given in terms of the aggregate persistence measures, but relevant variance expressions for the sectoral persistence measures can be obtained by replacing $\mathbf{w}$ in the expressions below by $w_i\mathbf{e}_i$, where $\mathbf{e}_i$ is the $m \times 1$ selection vector defined in the text.

Consider the following general version of (10.5.3) and (10.5.4):

$$\mathbf{C}(L)\Delta\mathbf{y}_t = \mathbf{a} + \boldsymbol{\gamma}(L)v_t + \boldsymbol{\epsilon}_t, \tag{0.B.1}$$

$$x_t = \boldsymbol{\beta}'\mathbf{z}_t + v_t, \tag{0.B.2}$$

for $t = 1, 2, \ldots, T$, where $x_t = \Delta m_t$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters

$$\mathbf{C}(L) = \mathbf{I}_m + \mathbf{C}_1 L + \cdots + \mathbf{C}_p L^p,$$

$$\gamma(L) = \gamma_0 + \gamma_1 L + \cdots + \gamma_q L^q \tag{0.B.3}$$

and $\mathbf{I}_m$ is an identity matrix of order m. Stacking all the observations using the notations

$$\Delta\mathbf{Y}' = [\Delta\mathbf{y}_1, \ldots, \Delta\mathbf{y}_T], \quad \mathbf{v}' = (v_1, \ldots, v_t),$$

$$\mathbf{E}' = [\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T], \quad \mathbf{x}' = (x_1, \ldots x_T).$$

$$\mathbf{Z}' = [\mathbf{z}_1, \ldots, \mathbf{z}_T], \quad \boldsymbol{\tau}' = (1, \ldots, 1), \boldsymbol{\tau}' \text{ is } 1 \times T,$$

the model (0.B.1) and (0.B.2) can be written as

$$\mathbf{C}(L)\Delta\mathbf{Y}' = \mathbf{a}\boldsymbol{\tau}' + \gamma(L)\mathbf{v}' + \mathbf{E}', \tag{0.B.4}$$

$$\mathbf{x} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{v}. \tag{0.B.5}$$

Using (0.B.3) and casting the above system in vector forms, (0.B.4) now becomes

$$\text{vec}(\Delta\mathbf{Y}) = (\mathbf{I}_m \otimes \mathbf{W})\mathbf{a} + \text{vec}(\mathbf{E}), \tag{0.B.6}$$

where $\mathbf{W} = [\boldsymbol{\tau}, \Delta\mathbf{Y}_0, \mathbf{V}_0]$, $\Delta\mathbf{Y}_0 = [\Delta\mathbf{Y}_{-1}, \ldots, \Delta\mathbf{Y}_{-p}]$, $\mathbf{V}_0 = [\mathbf{v}, \mathbf{v}_{-1}, \ldots, \mathbf{v}_{-q}]$,
and $\mathbf{a} = \text{vec}\left([\mathbf{a}, -\mathbf{C}_1, \ldots, -\mathbf{C}_p, \gamma_0, \ldots, \gamma_q]'\right).$

The parameters of the model (0.B.5) and (0.B.6) which we denote by $\boldsymbol{\theta}$ are the *unrestricted* elements of

$$\{\boldsymbol{\beta}, \mathbf{a}\} = \{\boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{c}^*, \boldsymbol{\gamma}^*\}. \tag{0.B.7}$$

If no restrictions are imposed on $\{\boldsymbol{\beta}, \mathbf{a}\}$ (as in the unrestricted VAR model), then the dimension of $\boldsymbol{\theta}$ is $k + m + pm2 + (q+1)m$. In general, $\mathbf{a}$ will be a function of $\boldsymbol{\theta}$.

The various persistence measures discussed in the paper are all scalar functions of $\boldsymbol{\theta}$. We represent this functional relation by $P(\boldsymbol{\theta})$, and assume that $P(\boldsymbol{\theta})$ is evaluated at the Maximum Likelihood (ML) estimators of $\boldsymbol{\theta}$ which we denote by $\widehat{\boldsymbol{\theta}}$. The asymptotic variance of $P(\widehat{\boldsymbol{\theta}})$ is given by

$$\operatorname{Avar}[P(\widehat{\boldsymbol{\theta}})] = (\partial P/\partial \boldsymbol{\theta}')\operatorname{Avar}(\widehat{\boldsymbol{\theta}})(\partial P/\partial \boldsymbol{\theta}). \tag{0.B.8}$$

To derive $\operatorname{Avar}(\widehat{\boldsymbol{\theta}})$ we first note that the joint log-likelihood function of (0.B.5) and (0.B.6) is proportional to

$$L(\boldsymbol{\theta}) = \frac{T}{2}\log\sigma_v^2 - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2\sigma_v^2}\mathbf{v}'\mathbf{v} - \frac{1}{2}\operatorname{vec}(\mathbf{E})'\left(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_T\right)\operatorname{vec}(\mathbf{E}), \tag{0.B.9}$$

where $\mathbf{I}_T$ is an identity matrix of order T. Other notations are defined in the text. Using (0.B.9) and after some algebra, we have[39]

$$\operatorname{Avar}[\widehat{\boldsymbol{\theta}}]^{-1} = \left[\begin{array}{c}\left(\frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}'}\right)\\ \left(\frac{\partial\mathbf{a}}{\partial\boldsymbol{\theta}'}\right)\end{array}\right]'\left[\begin{array}{cc}\mathbf{Z}_0'\left(\boldsymbol{\gamma}^*\boldsymbol{\Sigma}\boldsymbol{\gamma}^{*\prime}\otimes\mathbf{I}\right)\mathbf{Z}_0+\frac{1}{\sigma_v^2}\mathbf{Z}'\mathbf{Z} & -\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}^{*\prime}\otimes\mathbf{W}'\right)\mathbf{Z}_0\\ -\mathbf{Z}_0'\left(\boldsymbol{\gamma}^*\boldsymbol{\Sigma}^{-1}\otimes\mathbf{W}\right) & \left(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{W}'\mathbf{W}\right)\end{array}\right]\left[\begin{array}{c}\left(\frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}'}\right)\\ \left(\frac{\partial\mathbf{a}}{\partial\boldsymbol{\theta}'}\right)\end{array}\right], \tag{0.B.10}$$

where $\mathbf{Z}_0' = [\mathbf{Z}', \mathbf{Z}_{-1}{}', \ldots, \mathbf{Z}_{-q}{}']'$.

It only remains to derive the first derivatives $\partial P/\partial\boldsymbol{\theta}'$, for the various persistence measures of interest. First consider the persistence measure P_0 defined by (10.3.18) in the text. It is relatively easy to show that

$$\begin{aligned}\frac{\partial P_0}{\partial\boldsymbol{\theta}'} &= \frac{1}{P_0}\left[\frac{\mathbf{w}'\mathbf{A}(1)\boldsymbol{\Sigma}}{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}\right](\mathbf{w}'\mathbf{A}(1)\otimes\mathbf{A}(1)'\mathbf{S}_p)\left(\frac{\partial\mathbf{c}^*}{\partial\boldsymbol{\theta}'}\right)\\ &\quad - \frac{1}{TP_0}\frac{(\mathbf{w}'\mathbf{A}(1)\otimes\mathbf{w}'\mathbf{A}(1)\mathbf{E}'\mathbf{W})}{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}\left(\frac{\partial\mathbf{a}}{\partial\boldsymbol{\theta}'}\right)\\ &\quad + \frac{P_0}{T}\frac{(\mathbf{w}'\times\mathbf{w}'\mathbf{E}'\mathbf{W})}{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}\left(\frac{\partial\mathbf{a}}{\partial\boldsymbol{\theta}'}\right),\end{aligned} \tag{0.B.11}$$

where $\mathbf{S}_p = [\mathbf{I}_m, \mathbf{I}_m, \ldots, \mathbf{I}(m]$ is an $m\times mp$ matrix. Next for P_m defined by (10.3.17) in the text, we have[40]

$$\begin{aligned}\frac{\partial P_m}{\partial\boldsymbol{\theta}'} &= \boldsymbol{\gamma}(1)'\frac{(\mathbf{w}'\mathbf{A}(1)\otimes\mathbf{A}(1)'\mathbf{S}_p)}{\mathbf{w}'\mathbf{d}(0)}\left(\frac{\partial\mathbf{c}^*}{\partial\boldsymbol{\theta}'}\right)\\ &\quad + \frac{(\mathbf{w}'\mathbf{A}(1)\otimes\mathbf{S}_q)}{\mathbf{w}'\mathbf{d}(0)}\left(\frac{\partial\boldsymbol{\gamma}^*}{\partial\boldsymbol{\theta}'}\right) - \frac{P_m}{\mathbf{w}'\mathbf{d}(0)}\mathbf{w}'\left(\frac{\partial\boldsymbol{\gamma}_0}{\partial\boldsymbol{\theta}'}\right),\end{aligned} \tag{0.B.12}$$

where $S_q = (1, 1, \ldots, 1)$ is an $1\times(q+1)$ vector.

Finally, for the measure P_y, defined by (10.3.16) in the text, we have

$$\frac{\partial P_y}{\partial\boldsymbol{\theta}'} = \lambda\frac{P_m}{P_y}\frac{\partial P_m}{\partial\boldsymbol{\theta}'} + \frac{P_m^2}{2P_y}\frac{\partial\lambda}{\partial\boldsymbol{\theta}'} + \frac{(1-\lambda)P_0}{P_y}\frac{\partial P_0}{\partial\boldsymbol{\theta}'} - \frac{P_0^2}{2P_y}\frac{\partial\lambda}{\partial\boldsymbol{\theta}'}. \tag{0.B.13}$$

λ is defined by (10.3.19) and has the following derivatives with respect to θ:

$$\begin{aligned}\frac{\partial\lambda}{\partial\boldsymbol{\theta}'} &= \frac{2(\lambda-\lambda^2)}{\mathbf{w}'\mathbf{d}(0)}\mathbf{w}'\left(\frac{\partial\boldsymbol{\gamma}_0}{\partial\boldsymbol{\theta}'}\right)\\ &\quad + \left(\frac{2\lambda(\mathbf{w}'\otimes\mathbf{w}'\mathbf{E}'\mathbf{W})}{T[\sigma_v^2(\mathbf{w}'\mathbf{d}(0))^2+\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}]}\right)\left(\frac{\partial\mathbf{a}}{\partial\boldsymbol{\theta}'}\right)\frac{2(\lambda-\lambda^2)}{\sigma_v^2}\left(\frac{\mathbf{v}'\mathbf{Z}}{T}\right)\left(\frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}'}\right).\end{aligned} \tag{0.B.14}$$

[39] Also see Pesaran (1991b), where similar derivations can be found for a related class of multivariate rational expectations models.

[40] Notice that in (10.3.17) $\mathbf{d}(1) = \mathbf{A}(1)\gamma(1) = \mathbf{C}^{-1}\gamma(1)$ and $\mathbf{d}(0) = \mathbf{d}_0 = \gamma_0$.

Consistent estimates of (0.B.9) may now be computed for any of the persistence measures of interest using (0.B.11) and the relevant expressions for the derivatives of $\partial P/\partial \boldsymbol{\theta}'$ given above, all evaluated at the ML estimators.

0.C Data

Industrial output data series for the period 1947–87 were taken from the U.S. Department of Commerce publications *The National Income and Product Accounts of the United States, 1929–1982*, and the July 1986 and July 1988 issues of the *Survey of Current Business*. Figures were taken from table 6.2, which provides annual data on Gross National Product by Industry in constant prices (billions of 1982 dollars). The ten-sector classification used in the empirical work was obtained from the more disaggregated figures provided in these publications as described in table 0.9 below.

For the sample period up to 1985, the data used in the estimation of the money supply growth equation are the same as those employed by Rush and Waldo (1988) and Pesaran (1988c). Data for RU_t (the annual average unemployment rate in the total labor force, including military personnel) and for FED, (real Federal Government expenditure) were extended to 1987 using the Economic Report of the President (1989 edition), while M_t (annual average M1) was extended using the Federal Reserve Bulletin (various issues).

Table 0.9

Survey of Current Business industry titles	Abbreviated industry titles	Line(s)
1. Agriculture, Forestry and Fisheries	Agriculture	4
2. Mining	Mining	$8+9+10+11$
3. Construction	Construction	12
4. Durable Manufacturing	Dur. manuf.	$15-25$
5. Nondurable Manufacturing	Nondur. manuf.	$27-36$
6. Transportation and Communications	Transport	$38+46$
7. Electric, Gas and Sanitary Services	Utilities	49
8. Wholesale and Retail Trade	Trade	$50+51$
9. Finance, Insurance, Real Estate and Services	Services	$52+60$
10. Government and Government Enterprises	Government	74

Chapter 11

Choice Between Disaggregate and Aggregate Specifications Estimated by Instrumental Variables Methods

A choice criterion is proposed for discriminating between disaggregate and aggregate models estimated by the instrumental variables method. The criterion, based on prediction errors, represents a generalisation of criteria developed in the context of classical regression models. The article also derives general tests for aggregation bias in the instrumental variables context. The criterion and the tests are applied in an analysis of UK employment demand. It is shown that a model disaggregated by 40 industries predicts aggregate employment better than an aggregate model and that significant biases exist in estimates of the long-run wage and output elasticities obtained from the aggregate model.

Key Words: Aggregation; Instrumental variables; Labour demand; Model selection.

JEL Classification: C12, C43, C52, J23.

The problem of aggregation over micro units has a long tradition in the econometrics literature stretching back to Theil (1954). Two issues in particular have attracted attention. The first concerns the prediction problem of choice between alternative disaggregate and aggregate specifications to predict aggregate variables. This issue was raised in the literature by Grunfeld and Griliches (1960) and reconsidered in a more general context by Pesaran et al. (1989b) (henceforth PPK). The second issue concerns the problem of aggregation bias defined by the deviation of the macro parameters from the average of the corresponding micro parameters. This was first discussed by Theil (1954) and an indirect test proposed by Zellner (1962). Early empirical studies are reported by Boot

[0] Published in *Journal of Business & Economic Statistics* (1994), Vol. 12, pp. 11–21. Co-authors M. H. Pesaran and K. C. Lee. The authors would like to thank two anonymous referees and the editor of this Journal for helpful and constructive comments. Financial support from the ESRC under grant nos. B01250038 and R000233608 and from The Newton Trust of Trinity College is gratefully acknowledged.

and de Wit (1960), Orcutt et al. (1968), and Gupta (1971), for example, whereas more recent work includes that by Heckman and Sedlacek (1988), Keane et al. (1988), and contributions in Barker and Pesaran (1990). In the work of Lee et al. (1990a,b) some general direct tests were derived for the case where the subset of parameters of interest may be a (possibly nonlinear) function of the full vector of parameters.

This article reconsiders both of these issues in the context of models in which the assumption that model regressors and disturbances are uncorrelated cannot be maintained and, to obtain consistent parameter estimates, instrumental variables (IV) methods are used. This situation arises frequently in applied work either due to simultaneity or because expectations are replaced by their realisations under the rational expectations hypothesis in econometric equations. It also arises in models in which nonlinear relations (such as Euler equations) are derived as first-order conditions to optimisation problems at a microlevel (see for example Hansen (1982) and Hansen and Singleton (1982)). When regressors and disturbances are correlated, the usual criterion for choosing between models, the sum of squared residuals, is no longer an appropriate statistic in the sense that its use does not guarentee that the 'true' model will be chosen, even asymptotically. Since the criteria proposed by Grunfeld and Griliches (1960) and by PPK for choosing between alternative disaggregated and aggregated models are based on the sum of squared residuals, these criteria are also inappropriate in these circumstances. Statistics based on the prediction errors of alternative models that provide a valid model-selection criterion can be derived, however. Moreover, since the equation residuals coincide with the prediction errors in the least squares case, the criterion based on the prediction errors advanced in this article represents a generalisation of the criteria that have been considered previously in the literature.

In Section 11.1 of this article, new choice criteria are proposed for discriminating between disaggregate and aggregate specifications estimated by IV methods, and their validity in this context is established. In Section 11.2, the issue of aggregation bias is considered in the IV context. Here tests are derived that allow a statistical comparison to be made between different parameters of interest based on aggregate and disaggregate models in which the models are estimated using IV methods. Finally, in Section 11.3 the statistical tools that have been developed are applied in an analysis of employment demand for the UK economy. Labour-demand equations for 40 industrial sectors are estimated using the IV method and compared with their aggregate counterpart. It is established that the disaggregate model outperforms the aggregate model in terms of its ability to predict aggregate employment demand. Furthermore, key long-run elasticities of labour demand estimated by the aggregate and disaggregate models are shown to be significantly different, with elasticities based on the aggregate model overstating the extent of the responsiveness of labour demand to changes in wages and output when compared to estimated elasticities based on the disaggregate model.

11.1 A Choice Criterion under IV Estimation

Suppose we have a disaggregated multisectoral model, denoted H_d, consisting of m sectoral equations, where the dependent variable in the ith equation is $\mathbf{y}_i$, an $n \times 1$ vector of observations for the ith unit $(i = 1, \cdots, m)$. We also have an aggregate model, denoted H_a, given by a single equation, the dependent variable of which is $\mathbf{y}_a = \sum_i \mathbf{y}_i$. Clearly, a disaggregate model can be used to address many questions that the aggregate model cannot. In this section, however, we assume that the primary focus of the analysis is the prediction of the aggregate variable $\mathbf{y}_a$ and consider the derivation of an appropriate selection criterion for choosing between the two models on this basis. This question was first addressed in the literature by Grunfeld and Griliches (1960), and a more general treatment was given by PPK. These works proposed selection criteria for choosing between disaggregate and aggregate models based on sums of squared residuals from the two models. The use of these selection criteria is justified on the grounds that, on average, their use would lead to the choice of the disaggregated model under the assumption that the micro equations are correctly specified. The use of the prediction criteria in the context of choice between models also has implications for model misspecification. When the disaggregate model fits worse than the aggregate model, this would indicate that the disaggregated model is misspecified. This suggests using a Durbin-Hausman type of misspecification test of the disaggregate model, and such a test is developed in the least squares context by Lee et al. (1990b)). A misspecification test of this type, however, serves a quite separate function to that served by the choice criterion. The way to think of the choice criterion is in situations in which an investigator is faced with two models, an aggregate and a disaggregate one, and *must* choose one of them for use in predicting the aggregate variable. The issues of model misspecification and aggregation errors were addressed in more detail by PPK, section 6.

The criteria proposed by Grunfeld and Griliches and by PPK are derived for models in which it could be assumed that regressors and disturbances are uncorrelated. In many instances, however, it is not reasonable to make this assumption, so ordinary least squares (OLS) estimation is no longer appropriate, and the IV estimation method is required to obtain consistent estimates. In these circumstances, the residual vectors obtained from the estimated model depend on the sign and magnitudes of the correlations between the dependent variable and the variables that are determined jointly with it. As a consequence, measures of goodness of fit that are based on the IV residuals cannot be guaranteed to choose a correct model even asymptotically, and the sum of squares of residuals is no longer an appropriate basis for developing model-selection criteria. (See Pesaran and Smith (1994) for further discussion of selection criteria appropriate for choice between models estimated by the IV method.)

In this section we consider alternative statistics, s_d^2 and s_a^2, relating to the disaggregate and aggregate models estimated by the IV method. These statistics are based on prediction errors, which are the appropriate measures for model comparison, and are not subject to the difficulties described previously. Specifically, these statistics are shown to have the property that

$$\operatorname*{plim}_{n\to\infty}(s_d^2|H_d) \leq \operatorname*{plim}_{n\to\infty}(s_a^2|H_d),$$

where probability limits are taken under the hypothesis of the disaggregated model H_d, so that they are valid statistics for use in a choice rule. To this end, consider the general disaggregate model defined by

$$\begin{aligned} H_d: \quad \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i \quad i = 1, \dots, m \\ \mathbf{X}_i &= \mathbf{Z}_i\boldsymbol{\Pi}_i + \mathbf{V}_i \end{aligned} \tag{11.1.1}$$

where $\mathbf{y}_i$ is the $n \times 1$ vector of observations on the dependent variable for the ith sector, $\mathbf{X}_i$ is the $n \times k_i$ matrix of observations on the regressors in (11.1.1) for the ith sector, assumed to have a full column rank, $\boldsymbol{\beta}_i$ is the $k_i \times 1$ vector of the coefficients associated with columns of $\mathbf{X}_i$, and $\mathbf{u}_i$ is the $n \times 1$ vector of disturbances for the ith sector. $\mathbf{Z}_i$

is a $n \times r_i$ matrix of IV's (where $r_i \geq k_i$), $\mathbf{\Pi}_i$ is an $r_i \times k_i$ matrix of parameters, and $\mathbf{V}_i$ is an $n \times k_i$ matrix of disturbances. The disturbances $\mathbf{u}_i$ and $\mathbf{V}_i$ are assumed to be serially uncorrelated within each sector but are contemporaneously correlated across sectors. Formally, the following standard assumptions are made:

Assumption A1: For all $i, j = 1, 2, ..., m$, the probability limits of $\mathbf{u}_i'\mathbf{u}_j/n$, $\mathbf{V}_i'\mathbf{V}_j/n$, and $\mathbf{V}_i'\mathbf{u}_j/n$ exist and are given by σ_{ij}, $\mathbf{\Sigma}_{ij}$, and $\boldsymbol{\delta}_{ij}$, respectively.

Assumption A2: For all $i, j = 1, 2, ..., m$, the instruments, $\mathbf{Z}_i$, are of full-column rank, and are asymptotically uncorrelated with the disturbances $\mathbf{u}_j$ and $\mathbf{V}_j$.

Assumption A3: For all $i, j = 1, 2, ..., m$, the matrices $\mathbf{Z}_i'\mathbf{X}_i/n$ and $\mathbf{Z}_i'\mathbf{Z}_j/n$ have finite probability limits, and the probability limits of $\mathbf{X}_i'\mathbf{X}_i/n$ and $\mathbf{Z}_i'\mathbf{Z}_i/n$ exist and are non-singular.

In general, the matrix $\mathbf{X}_i$ is correlated with $\mathbf{u}_i$ and may include lagged values of the dependent variable, $\mathbf{y}_i$, as well as current and lagged values of other endogenous variables. It is possible, however, that $\mathbf{X}_i$ includes some exogenous variables, in which case we assume that these variables also appear in $\mathbf{Z}_i$, so $\mathbf{V}_i$ and consequently $\mathbf{\Sigma}_{ii}$ will not be of full rank.

The aggregate model is given by

$$H_a: \qquad \mathbf{y}_a = \mathbf{X}_*\boldsymbol{\beta}_* + \mathbf{u}_* \tag{11.1.2}$$

where $\mathbf{y}_a = \sum_{i=1}^m \mathbf{y}_i$ and $\mathbf{X}_*$ is a $n \times k_*$ matrix of regressors, $\boldsymbol{\beta}_*$ is a $k_* \times 1$ vector of the coefficients associated with the columns of $\mathbf{X}_*$ and $\mathbf{u}_*$ is an $n \times 1$ vector of disturbances. It will also be assumed that:

Assumption A4: There exists a set of "aggregate" instruments, $\mathbf{Z}_*$, of full-column rank that are asymptotically uncorrelated with the disturbances $\mathbf{u}_i$ and $\mathbf{V}_i$, and for which the matrices $\mathbf{Z}_*'\mathbf{X}_i/n$ and $\mathbf{Z}_*'\mathbf{Z}_*/n$ have finite probability limits for $i = 1, 2, ..., m$.

No assumption is made in (11.1.2) about the relationship between $\mathbf{X}_*$ and the $\mathbf{X}_i$'s. Model (11.1.2) is to be viewed here as a rival model to (11.1.1) for the purpose of predicting $\mathbf{y}_a$ and has not *necessarily* been derived from (11.1.1) through any formal aggregation procedure. (See Section 11.2, however, on testing for aggregation bias.) Similarly, the instruments of the aggregate model, $\mathbf{Z}_*$, are not *necessarily* related to the disaggregated instrument sets, $\mathbf{Z}_i$, except insofar as by Assumption A4 they would also be valid instruments under H_d. This condition would be satisfied, for example, when the $\mathbf{Z}_*$'s are restricted to include lagged variables only.

Now consider the statistics for the aggregate and disaggregate models, based on the prediction errors, given by

$$s_a^2 = \widehat{\mathbf{e}}_a'\widehat{\mathbf{e}}_a/n \tag{11.1.3}$$

and

$$s_d^2 = \widehat{\mathbf{e}}_d'\widehat{\mathbf{e}}_d/n, \tag{11.1.4}$$

respectively, where

$$\widehat{\mathbf{e}}_d = \sum_{i=1}^m \widehat{\mathbf{e}}_i \tag{11.1.5}$$

and where $\widehat{\mathbf{e}}_a = \mathbf{y}_a - \widehat{\mathbf{X}}_* \widetilde{\boldsymbol{\beta}}_*$, $\widetilde{\boldsymbol{\beta}}_* = (\mathbf{X}_*' \mathbf{P}_* \mathbf{X}_*)^{-1} \mathbf{X}_*' \mathbf{P}_* \mathbf{y}_a$, $\widehat{\mathbf{X}}_* = \mathbf{P}_* \mathbf{X}_*$, $\mathbf{P}_* = \mathbf{Z}_* (\mathbf{Z}_*' \mathbf{Z}_*)^{-1} \mathbf{Z}_*'$, and $\widehat{\mathbf{e}}_i = \mathbf{y}_i - \widehat{\mathbf{X}}_i \widehat{\mathbf{X}} e \boldsymbol{\beta}_i$, $\widetilde{\boldsymbol{\beta}}_i = (\mathbf{X}_i' \mathbf{P}_i \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{P}_i \mathbf{y}_i$, $\widehat{\mathbf{X}}_i = \mathbf{P}_i \mathbf{X}_i$, $\mathbf{P}_i = \mathbf{Z}_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i'$. The estimators $\widetilde{\boldsymbol{\beta}}_*$ and $\widetilde{\boldsymbol{\beta}}_i$ are the generalised IV estimators of the parameters of the aggregate and disaggregate models, respectively. These are consistent IV estimators although for the disaggregate model they are not fully efficient since they do not take into account the contemporaneous covariances between sectors characterised by the nonzero off-diagonal elements in σ_{ij} in Assumption A1. Clearly, the prediction errors of the two models, $\widehat{\mathbf{e}}_a$ and $\widehat{\mathbf{e}}_d$, are different from the usual single-equation residuals, $\mathbf{e}_a = \mathbf{y}_a - \mathbf{X}_* \widetilde{\boldsymbol{\beta}}_*$ and $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i \widetilde{\boldsymbol{\beta}}_i$ because they account for the fact that the regressors $\mathbf{X}_*$ and $\mathbf{X}_i$ are stochastic variables, which, for prediction, must be replaced by their predicted values $\widehat{\mathbf{X}}_*$ and $\widehat{\mathbf{X}}_i$ respectively. The two coincide only in a fixed regressor framework, where $\mathbf{V}_i = 0$ and where OLS is an appropriate estimator. From this perspective, (11.1.3) and (11.1.4) can be viewed as an obvious generalisation of the sum of squared residuals criterion proposed for OLS models by Grunfeld and Griliches (1960) and by PPK.

We now show that the statistics (11.1.3) and 11.1.4) have the desirable property that

$$\operatorname*{plim}_{n\to\infty}(s_d^2|H_d) \leq \operatorname*{plim}_{n\to\infty}(s_a^2|H_d).$$

First note that

$$\widetilde{\boldsymbol{\beta}}_i = (\widehat{\mathbf{X}}_i' \widehat{\mathbf{X}}_i)^{-1} \widehat{\mathbf{X}}_i' \mathbf{y}_i \tag{11.1.6}$$

and

$$\widetilde{\boldsymbol{\beta}}_* = (\widehat{\mathbf{X}}_*' \widehat{\mathbf{X}}_*)^{-1} \widehat{\mathbf{X}}_*' \mathbf{y}_a. \tag{11.1.7}$$

Then we can write $\widehat{\mathbf{e}}_i = (\mathbf{I} - \widehat{\mathbf{Q}}_i)\mathbf{y}_i$, where $\widehat{\mathbf{Q}}_i = \widehat{\mathbf{X}}_i' (\widehat{\mathbf{X}}_i' \widehat{\mathbf{X}}_i)^{-1} \widehat{\mathbf{X}}_i'$. Hence, substituting from (11.1.1), $\widehat{\mathbf{e}}_i = (\mathbf{I} - \widehat{\mathbf{Q}}_i)\mathbf{X}_i \boldsymbol{\beta}_i + (\mathbf{I} - \widehat{\mathbf{Q}}_i)\mathbf{u}_i$, and, since $\widehat{\mathbf{X}}_i' \mathbf{X}_i = \widehat{\mathbf{X}}_i' \widehat{\mathbf{X}}_i$ and $\widehat{\mathbf{Q}}_i \widehat{\mathbf{X}}_i = \widehat{\mathbf{X}}_i$, we have

$$\widehat{\mathbf{e}}_i = (\mathbf{X}_i - \widehat{\mathbf{X}}_i)\boldsymbol{\beta}_i + (\mathbf{I} - \mathbf{Q}_i)\mathbf{u}_i. \tag{11.1.8}$$

However,

$$\begin{aligned}
\mathbf{X}_i - \widehat{\mathbf{X}}_i &= \mathbf{X}_i - \mathbf{Z}_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i' \mathbf{X}_i \\
&= (\mathbf{I} - \mathbf{P}_i)\mathbf{X}_i \\
&= (\mathbf{I} - \mathbf{P}_i)(\mathbf{Z}_i \boldsymbol{\Pi}_i + \mathbf{V}_i) \\
&= (\mathbf{I} - \mathbf{P}_i)\mathbf{V}_i.
\end{aligned}$$

Hence, (11.1.8) can be rewritten as

$$\widehat{\mathbf{e}}_i = (\mathbf{I} - \mathbf{P}_i)\mathbf{V}_i \boldsymbol{\beta}_i + (\mathbf{I} - \widehat{\mathbf{Q}}_i)\mathbf{u}_i \tag{11.1.9}$$

so that

$$\widehat{\mathbf{e}}_d = \sum_i [(\mathbf{I} - \mathbf{P}_i) V_i \boldsymbol{\beta}_i + (\mathbf{I} - \widehat{\mathbf{Q}}_i)\mathbf{u}_i]$$

and

$$\begin{aligned}\widehat{\mathbf{e}}_d'\widehat{\mathbf{e}}_d &= \sum_{i,j} \boldsymbol{\beta}_i'\mathbf{V}_i'(\mathbf{I}-\mathbf{P}_i)(\mathbf{I}-\mathbf{P}_j)\mathbf{V}_j\boldsymbol{\beta}_j \\ &+ \sum_{i,j} \mathbf{u}_i'(\mathbf{I}-\widehat{\mathbf{Q}}_i)(\mathbf{I}-\widehat{\mathbf{Q}}_j)\mathbf{u}_j \\ &+ \sum_{i,j} \boldsymbol{\beta}_i'\mathbf{V}_i'(\mathbf{I}-\mathbf{P}_i)(\mathbf{I}-\widehat{\mathbf{Q}}_j)\mathbf{u}_j \\ &+ \sum_{i,j} \mathbf{u}_i'(\mathbf{I}-\widehat{\mathbf{Q}}_i)(\mathbf{I}-\mathbf{P}_j)\mathbf{V}_j\boldsymbol{\beta}_j. \end{aligned} \tag{11.1.10}$$

But, under Assumptions A1–A3,

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\left(\frac{\widehat{\mathbf{e}}_d'\widehat{\mathbf{e}}_d}{n}|H_d\right) &= \sum_{i,j}\boldsymbol{\beta}_i'\boldsymbol{\Sigma}_{ij}\boldsymbol{\beta}_j + \sum_{i,j}\sigma_{ij} + \sum_{i,j}\boldsymbol{\beta}_i'\boldsymbol{\delta}_{ij} + \sum_{i,j}\boldsymbol{\delta}_{ji}'\boldsymbol{\beta}_j \\ &= \mathrm{E}\left\{\left[\sum_i(\mathbf{u}_i+\mathbf{V}_i\boldsymbol{\beta}_i)\right]'\left[\sum_i \mathbf{u}_i+\mathbf{V}_i\boldsymbol{\beta}_i\right]\right\} \\ &= \mathrm{E}(\boldsymbol{\xi}_a'\boldsymbol{\xi}_a) > 0, \end{aligned} \tag{11.1.11}$$

where $\boldsymbol{\xi}_a = \sum_i \boldsymbol{\xi}_i$ is the vector of aggregate errors of the reduced form equations

$$\begin{aligned}\mathbf{y}_i &= \mathbf{Z}_i\boldsymbol{\Pi}_i\boldsymbol{\beta}_i + \mathbf{V}_i\boldsymbol{\beta}_i + \mathbf{u}_i \\ &= \mathbf{Z}_i\boldsymbol{\Pi}_i\boldsymbol{\beta}_i + \boldsymbol{\xi}_i \end{aligned} \tag{11.1.12}$$

and $\boldsymbol{\xi}_i = \mathbf{V}_i\boldsymbol{\beta}_i + \mathbf{u}_i$.

Consider now the aggregate prediction criterion. We have $\widehat{\mathbf{e}}_a = \mathbf{y}_a - \widehat{\mathbf{X}}_*\widetilde{\boldsymbol{\beta}}_*$, and, under H_d,

$$\widehat{\mathbf{e}}_a = \left(\mathbf{I}-\widehat{\mathbf{Q}}_*\right)\left(\sum_i \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i\right), \tag{11.1.13}$$

where

$$\begin{aligned}\widehat{\mathbf{Q}}_* &= \widehat{\mathbf{X}}_*(\widehat{\mathbf{X}}_*'\widehat{\mathbf{X}}_*)^{-1}\widehat{\mathbf{X}}_*' \\ &= \mathbf{P}_*\mathbf{X}_*(\mathbf{X}_*'\mathbf{P}_*\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{P}_*. \end{aligned}$$

Substituting from (11.1.12),

$$\begin{aligned}\widehat{\mathbf{e}}_a &= \left(\mathbf{I}-\widehat{\mathbf{Q}}_*\right)\left[\sum_i \mathbf{Z}_i\boldsymbol{\Pi}_i\boldsymbol{\beta}_i + \sum_i(\mathbf{u}_i+\mathbf{V}_i\boldsymbol{\beta}_i)\right] \\ &= \left(\mathbf{I}-\widehat{\mathbf{Q}}_*\right)(\mathbf{f}_a+\boldsymbol{\xi}_a) \end{aligned}$$

so that, taking probability limits under H_d,

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\left(\frac{\widehat{\mathbf{e}}_a'\widehat{\mathbf{e}}_a}{n}|H_d\right) &= \operatorname*{plim}_{n\to\infty}\left(\frac{\mathbf{f}_a'[\mathbf{I}-\widehat{\mathbf{Q}}_*]\mathbf{f}_a}{n}\right) + \operatorname*{plim}_{n\to\infty}\left(\frac{\boldsymbol{\xi}_a'[\mathbf{I}-\widehat{\mathbf{Q}}_*]\boldsymbol{\xi}_a}{n}\right) \\ &+ 2\operatorname*{plim}_{n\to\infty}\left(\frac{\mathbf{f}_a'[\mathbf{I}-\widehat{\mathbf{Q}}_*]\boldsymbol{\xi}_a}{n}\right). \end{aligned} \tag{11.1.14}$$

But, since $\mathbf{u}_i$ and $\mathbf{V}_i$ are asymptotically distributed independently of $\mathbf{Z}_*$, by Assumption A4, it follows that

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\left(\frac{\boldsymbol{\xi}_a'[\mathbf{I}-\widehat{bfQ}_*]\boldsymbol{\xi}_a}{n}\right) &= \operatorname*{plim}_{n\to\infty}\left(\frac{\boldsymbol{\xi}_a'\boldsymbol{\xi}_a}{n}\right)\\ &= E(\boldsymbol{\xi}_a'\boldsymbol{\xi}_a)\end{aligned}$$

and

$$\operatorname*{plim}_{n\to\infty}\left(\frac{\mathbf{f}_a'[\mathbf{I}-\widehat{\mathbf{Q}}_*]\boldsymbol{\xi}_a}{n}\right)=0.$$

Hence,

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\left(\frac{\widehat{\mathbf{e}}_a'\widehat{\mathbf{e}}_a}{n}|H_d\right) &= E(\boldsymbol{\xi}_a'\boldsymbol{\xi}_a)+\operatorname*{plim}_{n\to\infty}\left(\frac{\mathbf{f}_a'[\mathbf{I}-\widehat{\mathbf{Q}}_*]\mathbf{f}_a}{n}\right)\\ &\geq E(\boldsymbol{\xi}_a'\boldsymbol{\xi}_a),\end{aligned}\qquad(11.1.15)$$

where the inequality follows because the second term in (11.1.15), namely

$$\operatorname*{plim}_{n\to\infty}\left(\frac{\mathbf{f}_a'[\mathbf{I}-\widehat{\mathbf{Q}}_*]\mathbf{f}_a}{n}\right),$$

is a positive semidefinite quadratic form. Comparing (11.1.15) with (11.1.11) establishes the result that

$$\operatorname*{plim}_{n\to\infty}\left(\frac{\widehat{\mathbf{e}}_d'\widehat{\mathbf{e}}_d}{n}|H_d\right)\leq\operatorname*{plim}_{n\to\infty}\left(\frac{\widehat{\mathbf{e}}_a'\widehat{\mathbf{e}}_a}{n}|H_d\right).\qquad(11.1.16)$$

In general we have not made any assumptions about the relationship between the disaggregate model H_d and the aggregate model H_a. It is interesting, however, to look at the special case where the aggregate model has been derived from a formal aggregation of the disaggregate model so that $\mathbf{X}_*=\mathbf{X}_a=\sum_i\mathbf{X}_i$ and $\mathbf{Z}_*=\mathbf{Z}_a=\sum_i\mathbf{Z}_i$. In this case the best that the aggregate model can do is to predict as well as the disaggregate model so that the two criteria coincide, and the conditions under which this will occur are the conditions for perfect aggregation, discussed for the least squares case by PPK. In the fixed regressor context of PPK, it is well known that sufficient conditions are when either $\boldsymbol{\beta}_i=\boldsymbol{\beta}$, for all i, $i=1,...,m$ (the *microhomogeneity* hypothesis) or when $\mathbf{X}_i=\mathbf{X}_a\boldsymbol{\Lambda}_i$ for all i (the *compositional-stability* hypothesis) where $\boldsymbol{\Lambda}_i$ are square full-rank matrices satisfying $\sum_i\boldsymbol{\Lambda}_i=\mathbf{I}$. (See also Lewbel (1992) for the application of a stochastic version of the compositional-stability hypothesis in the context of aggregating log-linear microequations.) In the present IV framework, however, where there is more than one variable determined simultaneously, these two conditions are no longer sufficient to achieve perfect aggregation, and an additional condition on the $\mathbf{Z}_i$'s is also needed. One such condition is given by

$$\mathbf{Z}_i=\mathbf{Z}_a\boldsymbol{\Gamma}_i\qquad(11.1.17)$$

for all i, $i=1,...,m$ where $\boldsymbol{\Gamma}_i$ are square full-rank matrices of fixed coefficients. This condition ensures that $\widehat{\mathbf{X}}_i=\mathbf{P}_a\mathbf{X}_i$, where $\mathbf{P}_a=\mathbf{Z}_a(\mathbf{Z}_a'\mathbf{Z}_a)^{-1}\mathbf{Z}_a'$, and, together with either the microhomogeneity hypothesis or the compositional-stability hypothesis, it is sufficient to ensure that $\widehat{\mathbf{e}}_d=\widehat{\mathbf{e}}_a$ so that disaggregate and aggregate criteria coincide. Condition (11.1.17) is a compositional stability hypothesis for the IV's of the disaggregate model.

Clearly a special case is where $\mathbf{\Gamma}_i = \mathbf{I}$ for all i which is where a common set of instruments is used across all sectors.

The prediction criteria (11.1.3) and 11.1.4) can be modified to incorporate degrees-of-freedom corrections. Clearly, such corrections will not affect the asymptotic properties of the statistics but we conjecture that they might improve their performance in finite samples. PPK derived corrections that ensured unbiasedness of the criteria in the least squares context. While no formal proof can be given in the present context, by analogy, we suggest using similar correction factors. This has the advantage of ensuring consistency with the criteria of PPK in the limiting case where $\mathbf{X}_i \subset \mathbf{Z}_i$ and $\mathbf{X}_* \subset \mathbf{Z}_*$ and the models (11.1.3) and (11.1.4) collapse to the fixed regressor models considered by PPK. Hence, the following modified criteria are suggested:

$$s_a^2 = \widehat{\mathbf{e}}_a' \widehat{\mathbf{e}}_a/(n - k_*) \tag{11.1.18}$$

and

$$s_d^2 = \sum_i \sum_j \widehat{\mathbf{e}}_i' \widehat{\mathbf{e}}_j / \{n - k_i - k_j + \mathrm{tr}(\widehat{\mathbf{Q}}_i \widehat{\mathbf{Q}}_j)\}, \tag{11.1.19}$$

where, as before, $\widehat{\mathbf{Q}}_i = \widehat{\mathbf{X}}_i(\widehat{\mathbf{X}}_i'\widehat{\mathbf{X}}_i)^{-1}\widehat{\mathbf{X}}_i'$.

11.2 Testing for Aggregation Bias under IV Estimation

Another important aspect in the comparison of aggregate and disaggregate models is the issue of *aggregation bias*. This concept was originally formalised by Theil (1954), who defined aggregation bias as the deviations of the parameters of a *macro* equation from the average of the corresponding parameters of the *micro* equations. Other definitions of aggregation bias are also used in the literature. For example, in his analysis of aggregating log-linear relations with fixed slope coefficients, Lewbel (1992) defined aggregation bias as the percentage difference between the common slope coefficient of the micro relations and the probability limit of the slope coefficient in the analogue aggregate equation and showed that this bias depends on the extent of the dependence between the regressors and the disturbances of the aggregate model. In the context of our application, where the micro equations are linear but have *different* slope coefficients, an adaptation of Lewbel's condition for no aggregation bias requires the micro coefficients, $\boldsymbol{\beta}_i$, to be distributed with a common mean, $\boldsymbol{\beta}_a$, such that $\boldsymbol{\beta}_i - \boldsymbol{\beta}_a$ are distributed independently of the regressors in all the micro equations. This condition yields the familiar random-coefficients model discussed by Zellner (1969). The condition that $\boldsymbol{\beta}_i - \boldsymbol{\beta}_a$ and the regressors of the micro equations are independently distributed is not, however, likely to be satisfied if the micro equations contain lagged dependent variables. (On this, see Pesaran and Smith (1992)). In general, however, where the slope coefficients differ across the micro equations, Theil's definition will still be appropriate for dynamic models, and will therefore be adopted in the rest of the article.

Here we generalise the tests for aggregation bias derived by Lee et al. (1990a,b) to the case where the macro and micro models are estimated by the IV method. In the application of the tests of aggregation bias, it is only meaningful to consider the case where the macro model is defined to be an analogue of the micro relations (11.1.1), given by

$$H_a: \qquad \mathbf{y}_a = \mathbf{X}_a \beta_a + \mathbf{u}, \tag{11.1.2´}$$

where $\mathbf{X}_a = \sum_{i=1}^m \mathbf{X}_i$. Here, the coefficients β_a can be interpreted as the 'average' counterparts of β_i. Such an interpretation of (11.1.2′) arises naturally in the case of random-coefficients models mentioned previously. The familiar method of testing for aggregation bias in the context of the micro relations (11.1.1) is to test directly the micro homogeneity hypothesis—namely, $H_\beta : \beta_1 = \beta_2 = \cdots = \beta_m$. An alternative approach, which is less restrictive, would be to test the equality of β_a from the macro equation with the average of the coefficients of the micro equations—namely,

$$H_0 : \eta_\beta = \beta_a - m^{-1} \sum_{i=1}^{m} \beta_i = 0. \tag{11.2.1}$$

Clearly H_0 implies H_β, but not *vice versa*. In what follows, we focus on tests of H_0 and its generalisation (which was discussed in detail by Lee et al. (1990b)), when the micro and macro equations are estimated by the IV method. The generalisation of H_0 covers situations in which the parameters of interest are (possibly nonlinear) functions of the micro parameters and their macro counterparts. In this general case, the hypothesis of no aggregation bias may be defined as

$$H_0 : \widehat{\mathbf{e}}ta_h = \mathbf{g}(\boldsymbol{\beta}_a) - \mathbf{h}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) = 0, \tag{11.2.2}$$

where $\mathbf{h}$ and $\mathbf{g}$ are assumed to be continuous and differentiable vector functions of dimension s, and where $\mathbf{g}(\boldsymbol{\beta}_a) = \mathbf{h}(\boldsymbol{\beta}_a, \ldots, \boldsymbol{\beta}_a)$. This formulation includes the hypothesis expressed at (11.2.1) as a special case and also allows the possibility of defining bias as the deviation of a function of the macro parameters from an average of the same function of the micro parameters or from a function of the average of the micro parameters or some other general form. In all cases, the null hypothesis that there is no aggregation bias would not be rejected under the micro homogeneity hypothesis H_β. On the other hand, it would be possible that no evidence of aggregation bias is found even when micro homogeneity does not hold, so that testing H_0 provides a less restrictive test for the presence of aggregation bias than the familiar test of the micro homogeneity hypothesis H_β. (Clearly this approach to testing for the presence of aggregation errors is distinct from that based on tests of mispecification in an aggregate model in which measures of distributional effects, calculated across the micro units, are employed (e.g., see Stoker (1986b)).)

Two test statistics are derived corresponding to two different assumptions about the vector of *macro* parameters $\boldsymbol{\beta}_a$. First assume that $\boldsymbol{\beta}_a$ is a vector of known parameters given *a priori* from some 'consensus' view, for example. A test statistic can be constructed based on the vector

$$\widetilde{\boldsymbol{\eta}}_h = \mathbf{g}(\boldsymbol{\beta}_a) - \mathbf{h}(\widetilde{\boldsymbol{\beta}}_1, \ldots, \widetilde{\boldsymbol{\beta}}_m). \tag{11.2.3}$$

On the null hypothesis H_0:

$$\operatorname*{plim}_{n\to\infty} \widetilde{\boldsymbol{\eta}}_h = \boldsymbol{\eta}_h = 0, \tag{11.2.4}$$

and

$$\widehat{\text{Avar}}(\widetilde{\boldsymbol{\eta}}_h) = \sum_{i=1}^{m} \sum_{j=1}^{m} \widetilde{\mathbf{H}}_i \widehat{\text{Avar}}(\widetilde{\boldsymbol{\beta}}_i, \widetilde{\boldsymbol{\beta}}_j) \widetilde{\mathbf{H}}'_j = \widetilde{\boldsymbol{\Omega}}_n, \tag{11.2.5}$$

where $\widetilde{\mathbf{H}}_i = \partial \mathbf{h} / \partial \widetilde{\boldsymbol{\beta}}'_i$, and the variance-covariance matrix of $\boldsymbol{\beta}_i$ in model (11.1.1) is estimated consistently by $\widehat{\text{Avar}}(\widetilde{\boldsymbol{\beta}}_i, \widetilde{\boldsymbol{\beta}}_j) = \widetilde{\sigma}_{ij} (\widehat{\mathbf{X}}'_i \widehat{\mathbf{X}}_i)^{-1} \widehat{\mathbf{X}}'_i \widehat{\mathbf{X}}_j (\widehat{\mathbf{X}}'_j \widehat{\mathbf{X}}_j)^{-1}$, where $\widetilde{\sigma}_{ij}$ is any consistent estimator of σ_{ij}. Then the test statistic for the hypothesis (11.2.2) is given by

$$q_1^* = \widetilde{\boldsymbol{\eta}}_h' \widetilde{\boldsymbol{\Omega}}_n \widetilde{\boldsymbol{\eta}}_h, \tag{11.2.6}$$

and on the null hypothesis, H_0, $q_1^* \overset{a}{\sim} \chi_s^2$.

Second, consider the case in which there is no consensus view on $\boldsymbol{\beta}_a$, so that, instead of being given *a priori*, the parameter vector $\boldsymbol{\beta}_a$ is estimated from the aggregate model (11.1.2′). From Assumption A4,

$$\operatorname*{plim}_{n\to\infty}(\widetilde{\boldsymbol{\beta}}_a | H_d) = \sum_{i=1}^{m} \mathbf{C}_i \boldsymbol{\beta}_i, \tag{11.2.7}$$

where

$$\mathbf{C}_i = \operatorname*{plim}_{n\to\infty}\{(\widehat{\mathbf{X}}_a' \widehat{\mathbf{X}}_a/n)^{-1}(\widehat{\mathbf{X}}_a' \mathbf{X}_i/n)\}.$$

In this case, a test of (11.2.2) can be based on the vector

$$\widetilde{\boldsymbol{\eta}}_h = \mathbf{g}(\widetilde{\boldsymbol{\beta}}_a) - \mathbf{h}(\widetilde{\boldsymbol{\beta}}_1, \ldots, \widetilde{\boldsymbol{\beta}}_m). \tag{11.2.8}$$

On the null hypothesis of no aggregation bias, H_0,

$$\operatorname*{plim}_{n\to\infty}(\widetilde{\boldsymbol{\eta}}_h | H_d) = \mathbf{g}\left(\sum_{i=1}^{m} \mathbf{C}_i \boldsymbol{\beta}_i\right) - \mathbf{h}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) = 0. \tag{11.2.9}$$

The test statistic for this case is given by

$$q_2^* = (\mathbf{g}(\widetilde{\boldsymbol{\beta}}_a) - \mathbf{h}(\widetilde{\boldsymbol{\beta}}_1, \ldots, \widetilde{\boldsymbol{\beta}}_m))' \widetilde{\boldsymbol{\Phi}}_n^{-1} (\mathbf{g}(\widetilde{\boldsymbol{\beta}}_a) - \mathbf{h}(\widetilde{\boldsymbol{\beta}}_1, \ldots, \widetilde{\boldsymbol{\beta}}_m)), \tag{11.2.10}$$

where

$$\widetilde{\boldsymbol{\Phi}}_n = \sum_{i=1}^{m} \sum_{j=1}^{m} \hat{\sigma}_{ij} \widetilde{\boldsymbol{\Psi}}_i \widetilde{\boldsymbol{\Psi}}_j', \tag{11.2.11}$$

and the matrix $\widetilde{\boldsymbol{\Psi}}_i$ (which corresponds to equation (21) of Lee et al. (1990b)) is defined by

$$\widetilde{\boldsymbol{\Psi}}_i = \widetilde{\mathbf{G}}_a (\widehat{\mathbf{X}}_a' \widehat{\mathbf{X}}_a)^{-1} \widehat{\mathbf{X}}_a' - \widetilde{\mathbf{H}}_i (\widehat{\mathbf{X}}_i' \widehat{\mathbf{X}}_i)^{-1} \widehat{\mathbf{X}}_i', \tag{11.2.12}$$

where $\widetilde{\mathbf{G}}_a = \partial \mathbf{g} / \partial \widetilde{\boldsymbol{\beta}}_a'$. On the null hypothesis, H_0, $q_2^* \overset{a}{\sim} \chi_s^2$.

11.3 An Application to Sectoral Labour-Demand Determination

In this section, the statistics that have been developed are applied to aggregate and sectorally disaggregated labour demand functions for the UK economy. This is an area of research that has received considerable attention recently as economists have attempted to understand and explain the causes of the historically high unemployment levels experienced recently in the United Kindom and elsewhere (e.g., see Layard et al. (1991), and the references therein). In particular, much applied research has focused on the responsiveness of labour demand to changes in real wages and in output levels in an effort to evaluate the efficacy of different policies designed to reduce unemployment. Much of this analysis, however, has been carried out using aggregate data, and it is of some interest to consider whether conclusions drawn on the basis of this work are affected by the choice of the level of aggregation used in the analysis.

PPK and Lee et al. (1990b) investigated this question empirically, using annual data for 40 industrial sectors over the period 1956-1984. The data cover the whole of the private sector, excluding the mineral oil and natural gas sector (sector 4) for which data are available only since 1971 when North Sea oil production started to come on line. (Full details of sources and definitions can be found in the data appendix to PPK.) In these works, the following general log-linear dynamic specifications for the sectoral labour demand equations were adopted:

$$\begin{aligned} LE_{it} &= \beta_{i1}/m + \beta_{i2}(T_t/m) + \beta_{i3}LE_{i,t-1} + \beta_{i4}LE_{i,t-2} + \beta_{i5}LY_{it} + \beta_{i6}LY_{i,t-1} + \beta_{i7}LW_{it} \\ &\quad + \beta_{i8}LW_{i,t-1} + \beta_{i9}\overline{LY}_{at} + \beta_{i10}\overline{LY}_{a,t-1} + u_{it}, \\ &\qquad i = 1, 2, 3, 5, 6, \ldots, 41, \quad t = 1956, \ldots, 1984, \end{aligned} \tag{11.3.1}$$

where LE_{it} = log of man-hours employed in sector i at time t, T_t = time trend ($T_{1980} = 0$), LY_{it} = log of sector i output at time t, LW_{it} = log of average product real wage rate per man-hours employed in sector i at time t, $\overline{LY}_{at}$ = average of LY_{it} over the 40 sectors and m = number of sectors, ($m = 40$). This specification can be justified theoretically when employment decisions are made within an industry by cost minimising firms with identical production functions and the same given demand and factor price expectations. The inclusion of lagged employment variables can be justified on the grounds of inertia in revision of expectations, adjustment costs involved in hiring and firing of workers, or aggregation over different labour types (see, for example, Nickell (1984) and Pesaran (1991a)). The variable $\overline{LY}_{at}$, which measures the level of aggregate output (in logs), is a proxy measure intended to capture changes in demand expectations arising from the perceived interdependence of demand in the economy by the firms in the industry. The time trend is included in the specification in order to allow for the effect of neutral technical progress on labour productivity.

OLS estimates of the disaggregated model in (11.3.1), and a restricted version of the model (in which linear parameters restrictions are imposed as a means of avoiding over-parameterisation), were presented in tables I and II of PPK. Using these, evidence is found to suggest that a disaggregate model is superior to its aggregate counterpart in terms of its ability to predict fluctuations in aggregate labour demand and that statistically significant differences exist between estimates of labour-demand elasticities obtained from the estimated aggregate and disaggregated models. In many models of supply-side behaviour, however, it is acknowledged that employment, wage, price, and output levels are determined simultaneously. Furthermore, in these circumstances, it is not clear how aggregate output levels, themselves an aggregation of the outcomes of sectoral output decisions, could be known with certainty prior to the time when sectoral employment decisions are made. Consequently, it might be argued that all of the current-dated explanatory variables in (11.3.1) are potentially correlated with the u_{it}, and that instruments for these variables are required. It is important, therefore, that we establish whether the previous findings are robust to the choice of estimation method, and to this end we have reestimated both the aggregate and disaggregate models using the IV method, and employed the techniques developed in the preceding sections to evaluate their relative performance.

As a first step in the empirical work of this paper, we estimated model (11.3.1) by the generalised IV method using the instruments $Z_{it} = \{1,\ T_t,\ LE_{i,t-1},\ LE_{i,t-2},\ LW_{i,t-1},$

$LW_{i,t-2}$, $LY_{i,t-1}$, $LY_{i,t-2}$, $\overline{LE}_{a,t-1}$, $\overline{LE}_{a,t-2}$, $\overline{LW}_{a,t-1}$, $\overline{LW}_{a,t-2}$, $\overline{LY}_{a,t-1}$, $\overline{LY}_{a,t-2}\}$. This choice of instruments is a natural one given the preceding discussion; the simultaneity of the employment-, price-, output-, and wage-setting decisions in each sector, and the possibility of intersectoral interdependencies, exerted directly through product or labour-market competition, or indirectly through the expectations-formation process, means that lagged sectoral and aggregate variables are likely to provide valid instruments for the current-valued explanatory variables in (11.3.1). Moreover, it is important to be as comprehensive as possible in the choice of instruments for the sectoral regressions; if for any sector i, the included instruments are only a subset of those variables that determine $\mathbf{X}_i$ in model (11.1.1), then the assumed independence of the $\mathbf{Z}_i$ and the $\mathbf{V}_j$ is likely to be violated for $i \neq j$. Similar comments are likely to be true for the assumed independence of the $\mathbf{Z}_*$ and the $\mathbf{V}_i$ if the $\mathbf{Z}_*$ include aggregated values of the $\mathbf{Z}_i$. The second step in the empirical study was to calculate the Wu (1973) T_2 statistic, also known as the Wu-Hausman statistic, for each of the sectoral equations to test for the exogeneity of the current-dated explanatory variables in (11.3.1), and to investigate the relevence of the IV estimation method in this context. In those sectors in which the null of exogeneity was not rejected, we reestimated the labour-demand equations by the OLS method. Finally, for each sector, we undertook a specification search in which variables with t ratios that were less than unity (in absolute terms) were dropped from the list of explanatory variables to obtain a more parsimonious set of employment equations. At each stage of the specification search, a joint test of the parameter restrictions and a test of the exogeneity of the regressors were also carried out. In the case of industries where the exogeneity hypothesis was not rejected, the employment equations were estimated by OLS.

The estimates of the sectoral labour-demand equations obtained through this procedure are given in Table 11.1 and Table 11.2 provides some of the associated summary and diagnostic statistics. Included also in Table 11.2, in the columns headed $\chi^2_{MS}(4)$ and $F_{WH}(3,16)$, are the Sargan (1964) general misspecification test statistics and Wu-Hausman test statistics, respectively, carried out on the (unreported) unrestricted versions of the equations in (11.3.1). The Sargan test statistics serve as a general misspecification test of the joint validity of the model specification and the instruments, and are below their 95% critical values in all sectors. Turning to the Wu-Hausman test results, note that, conservatively working at the 10% level of significance, these statistics suggest the rejection of the exogeneity hypothesis in 6 of the 40 industries—namely, mechanical engineering (15), office goods (16), electrical engineering (17), rubber goods (31), hotels and catering (35), and communications (39). Furthermore, in the course of the specification search procedure, exogeneity of regressors in the restricted version of the labour-demand equation for the office goods sector could not be rejected either. Consequently, in all but five industries the parameter estimates reported in Table 11.1 are obtained by OLS and are equivalent to those in table II of PPK (in which OLS methods were employed throughout). In these five industries, however, exogeneity cannot be assumed to hold, and the IV estimation method has been employed; Wu-Hausman statistics for the test of the exogeneity of regressors in the restricted regressions reported in Table 11.1 for sectors 15, 17, 31, 35, and 39 were 3.14 (2,18), 11.87 (2,22), 3.39 (1,23), 8.83 (2,19), and 18.39 (2,21), respectively, where the relevent degrees of freedom of the F distribution are given in parentheses.

The parameter estimates presented in Table 11.1 are generally of the expected sign and, following the specification search, are generally well determined. In particular, it is

Table 11.1: Disaggregate Labour-Demand Functions (restricted) 1956–1984

Industry	$c/40$	$T/40$	LY_{it}	$LY_{i,t-1}$	$LE_{i,t-1}$	$LE_{i,t-2}$	LW_{it}	$LW_{i,t-1}$	$\overline{LY}_{at}$	$\overline{LY}_{a,t-1}$
1. Agriculture	152.152	—	.2687	.1752	.5312	—	−.4312	—	−.2437	−.1729
	(64.95)		(.14)	(.11)	(.06)		(.08)		(.10)	(.11)
2. Mining	41.200	−.3502	.2734	4 − .4181	1.1604	−.2848	−.2018	—	—	—
	(14.34)	(.07)	(.04)	(.06)	(.09)	(.08)	(.03)			
3. Coke	−351.57	−1.3100	—	.6330	—	—	−.3005	—	1.0448	—
	(44.66)	(.18)		(.15)			(.04)		(16)	
4. Oil	—	—	—	—	—	—	—	—	—	—
5. Petroleum Products	−70.796	−.5087	.3640	—	.5185	—	−.3144	—	—	—
	(71.77)	(.13)	(.13)		(.13)		(.09)			
6. Electric	18.523	—	.1614	—	1.2739	−.5958	−.1732	—	—	—
	(14.70)		(.08)		(.17)	(.16)	(.07)			
7. Gas	−47.110	−.6014	—	.0611	.4191	—	−.1507	—	.5379	—
	(97.22)	(.20)		(.07)	(.15)		(.05)		(.18)	
8. Water	8.168	—	.6536	−.6536	.8112	—	−.4027	.4027	−.6415	.7906
	(18.92)		(.40)	(.40)	(.08)		(.11)	(.11)	(.31)	(.31)
9. Minerals	172.916	—	.2655	—	.6931	—	−.1494	—	−.5337	—
	(79.12)		(.13)		(.08)		(.06)		(.26)	
10. Iron	−349.96	−.9045	.1083	—	.4978	—	−.3873	—	1.1803	—
	(58.87)	(.27)	(.09)		(.08)		(.08)		(.29)	
11. Other metals	−84.826	−.5749	.1817	−.3091	1.2461	−.4796	−.0756	.0756	.5854	—
	(30.72)	(.15)	(.13)	(.13)	(.15)	(.12)	(.05)	(.05)	(.18)	
12. Mineral products	−280.57	−.3729	.3101	—	.6919	—	−.2356	−.2214	.5170	—
	(60.64)	(.21)	(.15)		(.09)		(.11)	(.10)	(.29)	
13. Chemicals	−125.06	—	—	—	.6205	—	−.2810	—	.6049	—
	(23.83)				(.07)		(.03)		(.08)	
14. Metal goods	−32.245	−1231	.4365		.5798	—	−.1671	—	—	—
	(25.53)	(.10)	(.04)		(.05)		(.08)			
15. Mechanical engineering†	−140.40	.2775	.5872	−.2910	.5309	−.1529	−.3090	−.3407	—	.3966
	(63.92)	(.15)	(.10)	(.11)	(.22)	(.16)	(.15)	(.13)		(.21)
16. Office goods	−3.4674	—	.1694	−.1694	1.2748	−.3244	−.3884	.3123	—	—
	(22.75)		(.09)	(.09)	(.20)	(.18)	(.14)	(.13)		
17. Electrical engineering†	−64.345	—	.4199	—	.5345	—	−.9220	.5053	—	—
	(29.12)		(.09)		(.09)		(.18)	(.17)		
18. Motor vehicles	−184.62	−.2365	.4908	−.3811	.9237	−.1783	—	−.1843	.5856	—
	(50.06)	(.11)	(.06)	(.11)	(.16)	(.09)		(.07)	(.18)	
19. Aerospace	200.392	−.6788	.0732	—	.7560	−.4659	—	−.1252	—	—
	(53.12)	(.16)	(.06)		(.17)	(.14)		(.07)		
20. Ships	−.7667	—	.4809	−.4809	1.4717	−.4717	—	—	.5103	−.5103
	(.31)		(.12)	(.12)	(.15)	(.15)			(.20)	(.20)
21. Other vehicles	−132.16	−.4754	.3130	—	.7270	—	−.1432	—	—	.2845
	(54.39)	(.17)	(.07)		(.09)		(.05)			(.11)
22. Instrument engineering	−11.357	−.3580	.3611	—	.5319	—	−.2624	—	—	—
	(47.49)	(.14)	(.10)		(.13)		(.11)			
23. Food	−172.16	−.4510	.6697	—	.3177	.2237	−.1962	—	—	.1157
	(76.05)	(.20)	(.17)		(.17)	(.16)	(.06)			(.12)
24. Drink	−15.180	−.4844	.2933	—	.7283	—	−.0945	.0591	—	—
	(73.49)	(.14)	(.12)		(.12)		(.09)	(.09)		
25. Tobacco	−213.37	−.3959	.7424	—	.7367	.2633	—	—	—	—
	(80.84)	(.12)	(.28)		(.22)	(.22)				
26. Textiles	−68.150	—	.5278	−.1236	.5880	—	−.3428	—	—	—
	(10.02)		(.05)	(.08)	(.06)		(.05)			
27. Clothing	−68.949	—	.4514	—	.5364	—	−.3756	—	—	—
	(11.96)		(.04)		(.04)		(.03)			
28. Timber	60.3106	−.3017	.3769	—	.4312	—	−.2460	.1493	—	—
	(20.95)	(.08)	(.04)		(.06)		(.07)	(.07)		
29. Paper	−44.740	−.3259	.4680	.1585	.3644	—	−.2503	—	—	—
	(13.29)	(.10)	(.07)	(.09)	(.08)	(.04)				
30. Books	58.9249	—	.2973	−.2575	1.4842	−.7029	−.0454	—	—	—
	(20.82)		(.06)	(.06)	(.17)	(.15)	(.05)			

(continued)

Table 11.1: (Continued)

Industry	$c/40$	$T/40$	LY_{it}	$LY_{i,t-1}$	$LE_{i,t-1}$	$LE_{i,t-2}$	LW_{it}	$LW_{i,t-1}$	$\overline{LY}_{at}$	$\overline{LY}_{a,t-1}$
31. Rubber†	−54.009	−.6246	.6726	−.2123	.7116	—	—	—	—	—
	(13.15)	(.12)	(.09)	(.12)	(.09)					
32. Other manufacturing	60.3555	−.3233	.2345	—	.6028	—	—	—	.4274	−.4274
	(20.03)	(.07)	(.04)		(.09)				(.13)	(.13)
33. Construction	7.2409	—	.5490	−.4527	1.0813	−.2453	−.4434	.3376	—	—
	(20.56)		(.08)	(.09)	(.16)	(.11)	(.08)	(.11)		
34. Distribution	109.986	.3892	—	.5034	.5641	—	−.3036	—	—	−.5678
	(43.33)	(.21)		(.20)	(.09)		(.12)			(.17)
35. Hotels†	388.377	.4384	−.8857	1.6050	—	−.4753	−.8499	.3938	—	−.4509
	(205.9)	(.18)	(.52)	(.60)		(.40)	(.31)	(.24)		(.28)
36. Rail	−65.107	—	—	.4077	.8047	—	−.0729	—	—	—
	(26.28)			(.10)	(.05)		(.05)			
37. Land transportation	146.432	−.4542	—	.2451	.9023	−.4855	—	—	—	—
	(37.81)	(.10)		(.07)	(.19)	(.18)				
38. Sea transportation	48.5901	−.1921	.1924	—	1.1919	−.5542	−.0853	—	—	—
	(104.9)	(.11)	(.16)		(.17)	(.22)	(.07)			
39. Communications†	195.189	—	−.2816	—	.8450	−.3766	—	.2841	.5211	—
	(48.50)		(.10)		(.17)	(.17)		(.10)	(.13)	
40. Business	209.651	—	.3108	—	.6781	−.3104	—	—	—	−.1633
	(49.15)		(.08)		(.18)	(.17)				(.05)
41. Services	−39.904	—	.2123	—	.8264	—	−.1408	—	—	—
	(33.31)		(.08)		(.10)		(.07)			

Note: Equations are estimated using the OLS method, *except* in the case of industries denoted † (i.e. industries numbered 15, 17, 31, 35 and 39), in which the IV method was employed. For these five sector, the following variables were included in the instrument set for the ith industry: $c/40$, $T/40$, $LY_{i,t-1}$, $LY_{i,t-2}$, $LE_{i,t-1}$, $LE_{i,t-2}$, $LW_{i,t-1}$, $LW_{i,t-2}$, $\overline{LY}_{a,t-1}$, $\overline{LY}_{a,t-2}$, $\overline{LE}_{a,t-1}$, $\overline{LE}_{a,t-2}$, $\overline{LW}_{a,t-1}$, $\overline{LW}_{a,t-2}$. Variables definitions are provided in the text, and data sources are provided in PPK. Values in parentheses are standard errors.

worth noting that a second lagged dependent variable is included in 17 of the 40 industrial equations, and its coefficient takes a negative sign, as suggested by the theory, in all cases in which the coefficient is statistically different from 0 (see Pesaran (1991a)). The need to include a variable to capture the effects of changes in demand expectations arising from interdependencies in the economy is confirmed by the presence of aggregate output terms in 19 of the 40 sectors. And the signs of the coefficients on the wage and output terms are generally as expected: the sum of the coefficients on current and lagged wage terms is negative in 31 of the sectors (and is not significantly different from 0 in a further 8), but the sum of the coefficients on the sectoral output terms is positive in 33 sectors (and is not significantly different from 0 in a further 4).

Table 11.2 reports the generalised $\bar{R}^2$ as measures of the 'fit' of the IV regressions and also several diagnostic statistics, denoted $\chi^2_{SC}(1)$, $\chi^2_{FF}(1)$, $\chi^2_{N}(2)$, and $\chi^2_{H}(1)$, and distributed approximately as chi-squared variates (with degrees of freedom in parentheses), for tests of residual serial correlation, functional form misspecification, nonnormal errors, and heteroscedasticity, respectively. (For more details of the tests, see Pesaran and Pesaran (1991b)). These statistics indicate that there is evidence of misspecification in only a few cases. For example, there is evidence of residual serial correlation only in the chemicals (13) and construction (33) industries, and this is weak in the former case. The $\chi^2_R(r)$ statistics reported in Table 11.2 for testing the restrictions imposed on

Table 11.2: Summary and Diagnostic Test Statistics for Restricted Labour-Demand Equations of Table 1

Industry	$\overline{GR}^2$	$\hat{\sigma}$	$\chi^2_R(r)$	$\chi^2_{SC}(1)$	$\chi^2_{FF}(1)$	$\chi^2_N(2)$	$\chi^2_H(1)$	$F_{WH}(3,16)$	$\chi^2_{MS}(4)$
1. Agriculture	.9983	.0137	.04(3)	.00	7.40**	.39	2.55	1.03	5.96
2. Mining	.9986	.0158	3.72(3)	.83	.91	.32	.05	.82	5.96
3. Coke	.9771	.0449	5.20(5)	.24	.67	.27	1.87	.02	4.66
4. Oil	—	—	—	—	—	—	—	—	—
5. Petroleum Products	.9178	.0566	4.89(5)	.48	.01	1.83	.85	.78	6.68
6. Electric	.9876	.0190	2.19(5)	.17	1.26	.18	.12	.26	5.62
7. Gas	.9719	.0322	3.97(4)	1.29	.00	4.88*	1.42	1.09	5.91
8. Water	.9279	.0412	.73(4)	1.67	.00	.47	1.05	.36	4.94
9. Minerals	.9760	.0318	2.40(5)	1.36	.16	32.7**	.00	.39	4.21
10. Iron	.9933	.0265	2.49(4)	.08	.19	1.42	.43	1.07	6.88
11. Other metals	.9864	.0250	2.63(2)	.00	3.45*	.20	1.89	1.29	5.63
12. Mineral products	.9935	.0177	3.44(3)	1.11	.23	.76	3.15*	.34	6.38
13. Chemicals	.9795	.0156	6.27(6)	3.51*	1.49	.96	1.14	.44	8.96*
14. Metal goods	.9877	.0192	2.37(5)	.09	.27	.38	1.00	.36	9.02*
15. Mechanical engineering†	.9918	.0148	.49(1)	1.21	6.21*	.77	1.49	2.60*	5.85
16. Office goods	.8922	.0345	10.5(4)**	.05	2.49	7.24**	5.05**	3.00*	3.38
17. Electrical engineering†	.9683	.0224	2.95(5)	.00	7.90**	.69	3.48*	5.40**	1.90
18. Motor vehicles	.9874	.0186	1.29(2)	1.55	8.71**	3.89	.00	.92	5.92
19. Aerospace	.9878	.0268	2.21(4)	.90	.30	1.81	1.30	1.04	3.69
20. Ships	.9818	.0323	9.70(6)	.45	.43	.40	6.26**	.67	2.58
21. Other vehicles	.9973	.0241	1.69(4)	.00	.81	.17	.04	.91	1.99
22. Instrument engineering	.9250	.0257	7.92(5)	.47	3.05*	.00	.84	.29	6.88
23. Food	.9837	.0164	.85(3)	1.69	1.76	1.33	4.38**	.75	8.92
24. Drink	.9232	.0269	2.56(4)	1.32	.02	.94	2.06	.80	6.79
25. Tobacco	.8796	.0497	7.09(6)	.25	.70	.65	6.33**	1.66	4.36
26. Textiles	.9981	.0175	3.18(5)	.05	4.44**	.74	5.09**	.70	5.62
27. Clothing	.9984	.0110	3.76(6)	.36	1.91	.62	.03	.20	3.46
28. Timber	.9864	.0138	4.24(4)	.00	2.56	1.34	.30	.41	2.85
29. Paper	.9927	.0192	2.86(4)	1.09	1.32	1.74	4.41**	1.93	2.30
30. Books	.9306	.0123	4.69(4)	1.70	.00	.14	.44	1.01	2.38
31. Rubber†	.9570	.0193	.67(5)	.11	3.34*	.56	2.48	3.71**	3.43
32. Other manufacturing	.9570	.0137	7.18(5)	.37	.21	1.12	.00	.91	3.45
33. Construction	.9689	.0179	5.54(3)	5.00**	2.13	1.62	1.00	.43	3.83
34. Distribution	.9589	.0143	5.44(4)	.49	.00	.94	2.06	2.12	3.61
35. Hotels†	.9202	.0316	.38(2)	.03	6.48**	2.21	1.77	3.18*	1.01
36. Rail	.9960	.0230	2.36(6)	.28	.00	1.27	1.98	.12	7.10
37. Land transportation	.9747	.0163	4.04(5)	.02	1.75	.64	2.71*	.95	4.00
38. Sea transportation	.9155	.0229	7.87(4)*	.27	6.06**	.39	2.75*	.22	8.64
39. Communications†	.9232	.0203	.33(4)	1.87	1.48	1.01	2.75*	6.39**	1.17
40. Business	.9940	.0128	1.81(5)	.98	2.01	1.98	.17	.49	2.90
41. Services	.9512	.0222	3.21(6)	.06	.53	.39	1.91	.34	6.63

Note: Equations are estimated using the OLS method, *except* in the case of industries denoted † (i.e. industries numbered 15, 17, 31, 35 and 39), in which the IV method was employed. See footnotes to Table 11.1. $\overline{GR}^2$ refers to the generalised R^2 statistic (cf. Pesaran and Smith (1994)). $\hat{\sigma}$ is the estimate of the of the equation's standard error. $\chi^2_R(r)$ is the chi-squared statistic for the Lagrange multiplier test for r linear restrictions imposed on the parameters of the unrestricted equation (where r is given in parentheses). $\chi^2_{SC}(1)$, $\chi^2_{FF}(1)$, $\chi^2_N(2)$ and $\chi^2_H(1)$ are diagnostic statistics, distributed approximately as chi-squared variates (with degrees of freedom in parentheses) for tests of residual serial correlation, functional form misspecification, nonnormal errors, and heteroscedasticity, respectively. (See Pesaran and Pesaran (1991b)). $F_{WH}(3,16)$ is the Wu-Hausman test for the exogeneity of $LYit$, $LWit$ and $\overline{LY}_{at}$ carried out on the unrestricted version of the model (cf. $F(3,16)$). $\chi^2_{MS}(4)$ is Sargan's general misspecification test carried out on the unrestricted version of the model. This latter statistic is the same as the J statistic in the generalised method of moments proposed by Hansen (1982). ** denotes significance at the 5% level and * denotes significance at the 10% level.

the unrestricted labour-demand equations in (11.3.1) to obtain the specifications given in Table 11.1 are below their 95% critical values in all industries other than office goods (16), thus reaffirming the plausibility of our search procedure. In summary, the results of Tables 11.1 and 11.2 indicate that, although there may be room for improving the results—by including industry-specific variables, for example—the specifications considered here provide a reasonable model of labour-demand determination at the sectorally disaggregated level.

Consider now the aggregate employment equation obtained as an analogue of (11.3.1):

$$\begin{aligned} LE_{at} = b_1 + b_2 T_t + b_3 LE_{a,t-1} + b_4 LE_{a,t-2} + b_5 LY_{at} \\ + b_6 LY_{a,t-1} + b_7 LW_{at} + b_8 LW_{a,t-1} + u_{at}, \end{aligned} \tag{11.3.2}$$

where

$$LE_{at} = \sum_{i=1,\ i\neq 4}^{41} LE_{it} \quad , \quad LY_{at} = \sum_{i=1,\ i\neq 4}^{41} LY_{it} \quad \text{and} \quad LW_{at} = \sum_{i=1,\ i\neq 4}^{41} LW_{it}.$$

Here the dependent variable of interest is assumed to be LE_{at}—that is, the sum of the logarithms of industry employment (in man-hours). Clearly, this is not the dependent variable usually considered in aggregate labour-demand equations (which tend to consider the logarithm of the sum of industry employment). The issue of consistent aggregation in the context of log-linear models has been discussed in the literature (e.g., Lovell (1973); van Daal and Merkies (1981)), and here we simply note that the aggregates employed in (11.3.2) may have some theoretical advantages over standard aggregate measures (i.e., the logarithm of the sum of sectoral employment, wages, or output) when the issue of interest is the analysis of sectoral employment growths. Of course, for our purposes, the specification (11.3.2) also has the advantage of fitting directly within the linear framework of the article.

A restricted version of (11.3.2) was estimated by the IV method using the instrument set $Z_{at} = \{1,\ LE_{a,t-1},\ LE_{a,t-2},\ LY_{a,t-1},\ LY_{a,t-2},\ LW_{a,t-1},\ LW_{a,t-2}\}$, and the following results were obtained:

$$\begin{array}{lllll} LE_{at} = & -137.01 & +0.6840 LE_{a,t-1} & +0.4745 LY_{at} & -0.3830 LW_{at} + \hat{u}_{at} \\ & (20.70) & (0.0569) & (0.0708) & (0.0540) \end{array}$$

$$\hat{\sigma} = 0.3487, \quad s(LE_{at}) = 5.75, \quad \text{Sample} = 1956 - 1984\ (n = 29)$$

$$\chi^2_{SC}(1) = 0.56, \quad \chi^2_{FF}(1) = 0.07, \quad \chi^2_{N}(2) = 4.15, \quad \chi^2_{H}(1) = 2.18, \quad \chi^2_{MS}(3) = 2.68. \tag{11.3.3}$$

Here, standard errors of the estimated parameters are given in brackets, $\hat{\sigma}$ is the estimate of the equation's standard error, $s(LE_{at})$ is the standard deviation of the dependent variable, and the remaining diagnostic statistics are as described in relation to Table 11.1. These IV estimates differ only marginally from those previously obtained using the OLS procedure and reported by PPK, and indeed the Wu-Hausman test fails to reject the exogeneity of the regressors LW_{at} and LY_{at} in this equation. While this finding might appear to suggest that the use of the OLS estimation method would be acceptable, it is in fact most important that the presence of simultaneity is taken into account here. If there is simultaneity in any of the sectoral equations, then it is clear that the aggregate model will be affected by such simultaneity so long as the matrix of regressors in the aggregate model includes aggregated values of the $\mathbf{X}_i$'s. The IV method will be the appropriate estimation procedure for the aggregate model in these circumstances, even if tests of exogeneity of regressors in the aggregate model fail to reflect this, possibly because of lack of power. Given the presence of simultaneity in the determination of employment, wages, and output in 5 of the 40 sectors, it is not appropriate to estimate the aggregate equation using OLS methods, and the IV results reported here are the relevent ones for use in comparison of the aggregate and disaggregate models.

For the two models (11.3.1) and (11.3.2), the statistics s_d^2 and s_a^2 of Section 11.1 were computed in both the uncorrected form and the modified form making an adjustment for degrees of freedom. (These calculations, as with all those described in this section, were carried out using the GAUSS programming language. Copies of the procedures as well as the data used for this analysis, are available from the authors on request.) For the disaggregated model, the uncorrected and modified values of s_d^2 were found to be 0.0742 and 0.1000, respectively, whereas for the aggregate model, the uncorrected and corrected values of s_a^2 were 0.3584 and 0.4158, respectively. It is clear that the criteria favour the disaggregate model, both in the uncorrected forms of (11.1.3) and (11.1.4) and in the corrected forms of (11.1.18) and (11.1.19), which include the degrees-of-freedom adjustments. These results are consistent with the findings of PPK based on the OLS estimates.

Models (11.3.1) and (11.3.2) were also used to test for aggregation bias in the estimates of the long-run elasticities of UK labour demand with respect to wages and output. (For this analysis sectors 20 and 25 had to be excluded because the restricted specifications estimated for these sectors do not seem to possess long-run solutions. The two sectors were consequently also removed from the definition of the aggregate variables entering equation (11.3.2).) For the ith sector, the long-run elasticities of interest are defined by

$$\epsilon_{iw} = \frac{\beta_{i7} + \beta_{i8}}{1 - \beta_{i3} - \beta_{i4}}, \quad \epsilon_{iy} = \frac{\beta_{i5} + \beta_{i6} + \beta_{i9} + \beta_{i10}}{1 - \beta_{i3} - \beta_{i4}},$$

and in considering aggregation bias, we aim to compare the average of each of these sectoral elasticities with the corresponding estimates based on the aggregate specification. As noted previously, these elasticities have been the subject of considerable interest because of their implications for macroeconomic policy. Various aggregate studies (many of which were reviewed by Treasury (1985)), have found a significant effect for real wages on employment demand, although the estimated size of the effect has varied considerably across studies, depending on the coverage of the data and on the specification of the employment equation that is considered. A consensus view has emerged on the basis of these aggregate studies, however, that the elasticity is close to -1, and hence this is the figure employed in the test of aggregation bias when comparison is made with an aggregate measure that is assumed known *a priori*. Similarly, a unit elasticity is used as the consensus figure for the output elasticity. Note that the wage and output elasticities obtained based on the aggregate model of (11.3.3) are -1.2792 (0.2121) and 1.5189 (0.2676), respectively. (Asymptotically valid standard errors are in parentheses.) These estimates are consistent with the hypothesis of wage and output elasticities of -1 and $+1$, respectively.

To examine whether these estimates for the aggregate wage and output elasticities are subject to aggregation bias, we use the restricted versions of IV models (11.3.1) and (11.3.2) and apply the tests defined by the statistics (11.2.6) and (11.2.10) of Section 11.2. A value of -0.5112 was obtained for the average of the sectoral wage elasticities estimated in Model (11.3.1), but the average of the sectoral output elasticities is found to be 0.9094. The q_1^* statistics of (11.2.6) that correspond to these figures, testing the null hypotheses that the wage and output elasticities are equal to their ‘consensus’ values of -1 and $+1$ respectively, are 29.25 and 0.48. Since both statistics are compared to the χ_1^2 distribution, these results provide strong evidence with which to reject the null hypothesis of no aggregation bias in the case of the wage elasticity but no evidence to reject the null in the case of the output elasticity. In contrast, when the q_2^* test statistics of (11.2.10) are calculated in which aggregation bias is defined with respect to the aggregate elasticities obtained from the estimated version of (11.3.2), the test statistic takes the value of 6.82 in the case of the wage elasticity, and 3.36 in the case of the output elasticity. Again each statistic is to be compared to the χ_1^2 distribution, so that there remains strong evidence with which to reject the null hypothesis of no aggregation bias in the case of the wage elasticity, and there is now some marginal evidence with which to reject this hypothesis for the output elasticity. These findings are also in line with those reported by Lee et al. (1990b), using the OLS method.

The results just described, obtained using the statistics appropriate for models estimated using the IV method derived in the previous sections of the article, confirm the findings of PPK and Lee et al. (1990b) that a disaggregate model of employment demand in the United Kingdom outperforms an aggregate model in terms of its predictive power and that there is significant aggregation bias in the estimation of the key wage elasticity of employment demand. The results substantiate the conclusions drawn previously by demonstrating that they cannot be attributed simply to some neglected simultaneity bias. The implications of those findings may be important for policy formulation, and certainly the results indicate that further work at the disaggregate level may be worthwhile.

Bibliography

Aigner, D. J. and S. M. Goldfeld (1974). Estimation and prediction from aggregate data when aggregates are measured more accurately than their components. *Econometrica 42*, 113–134.

Amemiya, T. and R. Y. Wu (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association 67*, 628–632.

Anderson, B. D. A. and J. B. Moore (1979). *Optimal Filtering.* Englewood Cliffs, N. J.: Prentice Hall.

Ando, A. (1971). On a problem of aggregation. *International Economic Review 12*, 306–311.

Arrow, K. J. (1962). The economic implications of learning by doing. *Review of Economic Studies 29*, 155–173.

Artis, M. J. and M. K. Lewis (1974). The demand for money: stable or unstable. *The Banker 124*, 239–247.

Artis, M. J. and M. K. Lewis (1976). The demand for money in the United Kingdom, 1963–1973. *Manchester School 44*, 147–181.

Baily, M. N. and A. K. Chakrabarti (1985). Innovation and productivity in US industry. *Brookings Papers on Economic Activity 2*, 609–632.

Ball, R. J. and E. B. A. S. Cyr (1966). Short term employment functions in British manufacturing industry. *Review of Economic Studies 33*, 179–207.

Barker, T. S. (1970). Aggregation error and estimates of the U.K. import demand function. In K. Hilton and D. F. Heathfield (Eds.), *The Econometric Study of the United Kingdom.* London: Macmillan.

Barker, T. S. and M. H. Pesaran (1990). *Disaggregation in Econometric Modelling.* London: Routledge.

Barker, T. S. and A. W. A. Peterson (1988). *The Cambridge Multisectoral Dynamic Model.* Cambridge: Cambridge University Press.

Barro, R. J. (1977). Unanticipated money growth and unemployment in the United States. *American Economic Review 67*, 101–115.

Beenstock, M. and P. Warburton (1984). *An econometric model of the UK labour market.* London. mimeo.

Beggs, J. J. (1984). Long-run trends in patenting. In Z. Griliches (Ed.), *R&D, Patents and Productivity*. Chicago, Il: University of Chicago Press.

Bergstrom, A. R. (1984). Continuous time stochastic models and issues of aggregation over time. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, Chapter 20. Amsterdam: North-Holland.

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components wth particular attentlon to measurement of the 'business cycle'. *Journal of Monetary Economics 7*, 151–I74.

Blanchard, O. J. and D. Quah (1989). The dynamics of aggregate demand and supply disturbances. *American Economic Review 79*, 655–673.

Boot, J. C. G. and G. M. de Wit (1960). Investment demand: an empirical contribution to the aggregation problem. *International Economic Review 1*, 3–30.

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day. Revised edition.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*, 1509–1526.

Brechling, F. (1965). Relationship between output and employment in British manufacturing industries. *Review of Economic Studies 32*, 187–216.

Breusch, T. and P. Schmidt (1985). *Alternative forms of the Wald test: how long is a piece of string?* Unpublished manuscript.

Brewer, K. R. W. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics 1*, 133–154.

Burgess, S. M. (1988). Employment adjustment in UK manufacturing. *Economic Journal 98*, 81–103.

Campbell, J. Y. and N. G. Mankiw (1987a). Are output fluctuations transitory? *Quarterly Journal of Economics 102*, 875–880.

Campbell, J. Y. and N. G. Mankiw (1987b). Permanent and transitory components in macroeconomic fluctuations. *American Economic Review 77*, 111–117.

Campbell, J. Y. and N. G. Mankiw (1989). International evidence on the persistence of economic fluctuations. *Journal of Monetary Economics 23*, 319–333.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*, 591–605.

Chow, G. C. and A. Lin (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *Review of Economics and Statistics 53*, 372–375.

Chow, G. C. and A. Lin (1976). Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association 71*, 719–721.

Christiano, L. J. and M. Eichenbaum (1989). *Unit roots in real GNP: do we know, and do we care?* NBER: Working Paper Series. No. 3130.

Clark, P. K. (1987a). The cyclical component of U.S. economic activity. *Quarterly Journal of Economics 102*, 797–814.

Clark, P. K. (1987b). *Trend reversion in real output and unemployment.* Stanford University, Stanford, CA.: Graduate School of Business. Discussion paper.

Cochrane, J. H. (1988). How big is the random walk component in GNP? *Journal of Political Economy 96*, 893–920.

Demery, D. and N. Duck (1990). *Are economic fluctuations really persistent?* Bristol, UK: University of Bristol. Mimeo.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual, totals; an approach based on quadratic minimization. *Journal of the American Statistical Association 66*, 99–102.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 427–431.

Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica 49*, 1057–1073.

Diebold., F. X. and M. Nerlove (1989). Unit roots in economic time series: a selective survey. In T. B. Formby and G. F. Rhodes (Eds.), *Advances in Econometrics: Cointegration, Spurious Regression and Unit Roots.* Greenwich, CT: JAI Press.

Dréze, J. H. and J.-F. Richard (1983). Bayesian analysis of simultaneous equations systems. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics.* Amsterdam: North Holland.

Durbin, J. (1954). Errors in variables. *Review of International Statistical Institute 22*, 23–32.

Durbin, J. (1970). Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica 38*, 410–421.

Durlauf, S. N. (1989). Output persistence, economic structure and the choice of stabilization policy. *Brookings Papers on Economic Activity 2*, 69–116.

Durlauf, S. N. (1990). *Spectral based testing of the martingale hypothesis.* Stanford, CA and NBER: Stanford University.

Durlauf, S. N. and P. C. B. Phillips (1986). *Trends versus random walks in time series analysis.* New Haven, CT. Unpublished manuscript.

Edwards., J. B. and G. H. Orcutt (1969). Should aggregation prior to estimation be the rule? *Review of Economics and Statistics 51*, 409–420.

Engle, R. F. (1982a). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. *Econometrica 50*, 987–1008.

Engle, R. F. (1982b). A general approach to Lagrange Multiplier diagnostics. *Journal of Econometrics 20*, 83–104.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica 55*, 251–276.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Evans, G. (1989). Output and unemployment dynamics in the United States 1950-85. *Journal of Applied Econometrics 4*, 213–237.

Fernandez, R. (1981). A methodological note on the estimation of time series. *Review of Economics and Statistics 63*, 471–475.

Fisher, D. (1968). The demand for money in britain: quarterly results 1951 to 1967. *Manchester School 36*, 329–344.

Fisher, D. (1970). The instruments of monetary policy and the generalised trade-off function for Britain, 1955–1968. *Manchester School 38*, 209–222.

Florens, J. P., M. Mouchart, and J.-F. Richard (1974). Bayesian inference in errors-in-variables models. *Journal of Multivariate Analysis 4*, 419–452.

Florens, J. P., M. Mouchart, and J.-F. Richard (1979). *Specification and inference in linear models.* CORE Discussion paper 7984.

Fuller, W. (1976). *lntroductlon to Statlstical Time Series.* New York, NY: Wiley.

Gardner, G., A. C. Harvey, and G. D. A. Phillips (1980). An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Applied Statistics 29*, 311–322.

Geweke, J. (1981). The approximate slopes of econometric tests. *Econometrica 49*, 1427–1442.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica 46*, 1293–1302.

Godley, W. A. H. and J. R. Shepherd (1964). Long-term growth and short-term policy. *National Institute Economic Review 29*, 26–38.

Goodhart, C. A. E. (1980a). Monetary policy in the United Kingdom. In K. Holbik (Ed.), *Monetary Policy in Twelve Industrial Countries*, Chapter 12. Boston, MA: Federal Reserve Bank of Boston.

Goodhart, C. A. E. (1980b). Problems of monetary management: the U.K. experience. In S. C. A (Ed.), *Inflation, Depression and Economic Policy in the West, Lessons from the 1970's.* Oxford: Basil Blackwell.

Goodhart, C. A. E. and A. D. Crockett (1970). The importance of money. *Bank of England Quarterly Bulletin 10*, 159–198.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*, 424–438.

Granger, C. W. J. (1987). Implications of aggregation with common factors. *Econometric Theory 3*, 208–222.

Granger, C. W. J. (1990). Aggregation of time-series variables: a survey. In T. S. Barker and M. H. Pesaran (Eds.), *Disaggregation in Econometric Modelling*. London: Routledge.

Granger, C. W. J. and R. F. Engle (1985). *Dynamic model specification with equilibium constraints: cointegration and error-correction*. Discussion paper 85-18.

Granger, C. W. J. and M. J. Morris (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society A 139*, 246–257.

Gregory, A. W. and M. R. Veall (1985). Formulating Wald tests of nonlinear restrictions. *Econometrica 53*, 1465–1468.

Gregory, A. W. and M. R. Veall (1987). Formulating Wald tests of the restrictions implied by the rational expectations hypothesis. *Journal of Applied Econometrics 2*, 61–68.

Grunfeld, Y. and Z. Griliches (1960). Is aggregation necessarily bad? *Review of Economics and Statistics 42*, 1–13.

Gupta, K. L. (1971). Aggregate bias in linear economic models. *International Economic Review 12*, 293–305.

Hacche, G. (1974). The demand for money in the United Kingdom: experience since 1971. *Bank of England Quarterly Bulletin 14*, 284–305.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of Political Economy 86*, 971–987.

Hansen, L. P. (1982). Large sample properties of generalised method of moments estimators. *Econometrica 50*, 1029–1054.

Hansen, L. P. and K. J. Singleton (1982). Generalised instrumental variables estimation of nonlinear rational expectations models. *Econometrica 50*, 1269–1286.

Harvey, A. C. (1981a). Finite sample prediction and overdifferencing. *Journal of Time Series Analysis 2*, 221–232.

Harvey, A. C. (1981b). *Time Series Models*. New York: John Wiley.

Harvey, A. C. (1985). Trends and cycles in macroeconomics time series. *Journal of Business and Economic Statistics 3*, 216–27.

Harvey, A. C., S. G. B. Henry, S. Peters, and S. Wren-Lewis (1986). Stochastic trends in dynamic regression models: an application to the employment-output equation. *Economic Journal 96*, 975–985.

Harvey, A. C. and C. R. McKenzie (1983). Missing observations in dynamic econometric models. In E. Parzen (Ed.), *Proceedings of Symposium on Time Series Analysis of Irregularly Observed Data, Texas A&M University, February 1983*. New York: Springer-Verlag. (forthcoming).

Harvey, A. C. and G. D. A. Phillips (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika 66*, 49–58.

Harvey, A. C. and R. G. Pierse (1982). *Estimating missing observations in economic time series.* London: LSE. Econometrics Programme Discussion Paper A33.

Harvey, A. C. and R. G. Pierse (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association 79*, 125–131.

Haubrich, J. G. and A. W. Lo (1989). *The sources and nature of long-term memory in the business cycle.* NBER: Working Paper Series. No. 2951.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*, 1251–1271.

Heckman, J. J. and G. Sedlacek (1988). Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labour market. *Journal of Political Economy 93*, 1077–1125.

Hendry, D. F. (1978a). *GENRAM. A generalised maximum likelihood estimation program for econometric systems.* London School of Economics Computer Unit.

Hendry, D. F. (1978b). *Technical manual for GIVE.* London School of Economics Computer Unit.

Hendry, D. F. (1979). The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors. *Journal of Econometrics 9*, 295–314.

Hendry, D. F. (1980). Predictive failure and econometric modelling in macroeconomics: the transactions demand for money. In P. Ormerod (Ed.), *Economic Modelling*, pp. 217–242. London: Heinemann.

Hendry, D. F. and G. E. Mizon (1979). Serial correlation as a convenient simplification, not a nuisance. *Economic Journal 88*, 549–563.

Hendry, D. F. and J.-F. Richard (1982). On the formulation of empirical models on dynamic econometrics. *Journal of Econometrics 20*, 3–33.

Hendry, D. F. and J.-F. Richard (1983). The econometric analysis of economic time series. *International Statistical Review 51*, 111–163.

Hotta, L. K., R. G. Pierse, and P. L. V. Pereira (1993). *The effect of overlapping aggregation in time series models.* London: London Business School. Unpublished paper.

Johansen, S. (1988). Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control 12*, 231–54.

Johansen, S. (1989). *Likelihood based inference on cointegration: theory and applications.* Bagni di Lucca, Italy: Centro Interuniversitario di Econometria. Lecture notes.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics 22*, 389–395.

Kaldor, N. (1957). A model of economic growth. *Economic Journal 67*, 591–624.

Kaldor, N. (1961). Capital accumulation and economic growth. In F. A. Lutz and D. C. Hague (Eds.), *The Theory of Capital.* London: Macmillan.

Kaldor, N. and J. A. Mirlees (1962). A new model of economic growth. *Review of Economic Studies 29*, 174–192.

Keane, M., R. Moffitt, and D. Runkle (1988). Real wages over the business cycle: estimating the impact of heterogeneity with micro data. *Journal of Political Economy 96*, 1232–1266.

Kelejian, H. H. (1980). Aggregation and disaggregation of non-linear equations. In J. Kmenta and J. B. Ramsey (Eds.), *Evaluation of Econometric Models*. New York: Academic Press.

Kennan, J. (1979). The estimation of partial adjustment models with rational expectations. *Econometrica 47*, 1441–1455.

Kiefer, N. and J.-F. Richard (1979). *A Bayesian approach to hypothesis testing and evaluating estimation strategies.* CORE Discussion paper 7927.

King, R. G., C. I. Plosser, J. H. Stock, and M. Watson (1987). *Stochastic trends and economic fluctuations.* NBER: Discussion Paper Series. No. 2229.

Kiviet, J. (1985). Model selection test procedures in a single linear equation of a dynamic simultaneous system and their defects in small samples. *Journal of Econometrics 28*, 327–362.

Klein, L. R. (1953). *A Textbook of Econometrics.* Evanston, IL: Row Peterson and Company.

Kloek, T. (1961). Note on convenient matrix notations in multivariate statistical analysis and in the theory of linear aggregation. *International Economic Review 2*, 351–360.

Lafontaine, F. and K. J. White (1986). Obtaining any Wald statistic you want. *Economics Letters 21*, 35–40.

Laidler, D. and M. Parkin (1970). The demand for money in the United Kindom 1956-1967: preliminary estimates. *Manchester School 38*, 187–208.

Lancaster, K. J. (1966). Economic aggregation and additivity. In S. R. Krupp (Ed.), *The Structure of Economic Science*. Englewood Cliffs, NJ: Prentice-Hall.

Layard, P. R. G. and S. J. Nickell (1985a). The causes of British unemployment. *National Institute Economic Review 111*, 62–85.

Layard, P. R. G. and S. J. Nickell (1985b). Unemployment, real wages, and aggregate demand in Europe, Japan, and the United States. *Carnegie-Rochester Conference Series on Public Policy 23*, 143–202.

Layard, P. R. G., S. J. Nickell, and R. Jackson (1991). *Unemployment: Macroeconomic Performance and the Labour Market.* Oxford: Oxford University Press.

Leamer, E. E. (1978). *Specification Searches: Ad hoc Inference with Non-experimental Data.* New York: Wiley.

Lee, K. C., M. H. Pesaran, and R. G. Pierse (1989). Aggregation bias in labour demand equations for the U.K. economy. In T. S. Barker and M. H. Pesaran (Eds.), *Disaggregation in Economic Modelling.* London: Routledge. (forthcoming).

Lee, K. C., M. H. Pesaran, and R. G. Pierse (1990a). Aggregation bias in labour demand equations for the U.K. economy. In T. S. Barker and M. H. Pesaran (Eds.), *Disaggregation in Economic Modelling*, Chapter 6, pp. 113–49. London: Routledge.

Lee, K. C., M. H. Pesaran, and R. G. Pierse (1990b). Testing for aggregation bias in linear models. *Economic Journal (Supplement) 100*, 137–150.

Lee, K. C., M. H. Pesaran, and R. G. Pierse (1991). *Persistence of shocks and their sources in a multisectoral model of UK output growth.* Cambridge: DAE Discussion Paper. No. 9114.

Lee, K. C., M. H. Pesaran, and R. G. Pierse (1992). Persistence of shocks and its sources in a multisectoral model of U.K. output growth. *Economic Journal 102*, 342–356.

Lewbel, A. (1992). Aggregation with log-linear models. *Review of Economic Studies 59*, 635–642.

Lippi, M. and L. Reichlin (1991). Trend-cycle decompositions and measures of persistence; does time aggregation matter? *Economic Journal 101*, 314–323.

Long, J. and C. I. Plosser (1987). Sectoral vs. aggregate shocks in business cycles. *American Economic Review 77*, 333–337.

Lovell, C. A. K. (1973). A note on aggregation bias and loss. *Journal of Econometrics 1*, 301–311.

Lubrano, M. and J.-F. Richard (1981). *Specification of the prior density in single equation errors-in-variables models: an application to a U.K. money demand equation.* CORE Discussion paper 8101.

Lucas, R. F. (1976). Econometric policy evaluation: a critique. In K. Brunner and A. M. Meltzer (Eds.), *The Phillips Curve and Labor Markets*. Amsterdam: North Holland. (Carnegie-Rochester Conference Series on Public Policy), Vol. 1.

MacKinnon, J. G. (1990). *Critical values for cointegration tests.* Kingston. Ont. and University of California, San Diego. CA: Queen's University. Unpublished manuscript.

Maddala, G. S. and F. D. Nelson (1974). Maximum likelihood methods for models of markets in disequilibrium. *Econometrica 42*, 1013–1030.

Mankiw, N. G. and M. D. Shapiro (1985). Trends, random walks, and tests of the permanent income hypothesis. *Journal of Monetary Economics 16*, 165–174.

Mankiw, N. G. and M. D. Shapiro (1986). Do we reject too often? small sample properties of tests of rational expectations models. *Economics Letters 20*, 139–145.

McFadden, D. and F. Reid (1975). Aggregate travel demand forecasting from disaggregate demand models. *Transportation Research Board Records 534*, 24–37.

Mills, T. C. (1991). Are fluctuations in U.K. output transitory or permanent? *Manchester School 59*, 1–11.

Molana, H. (1991). The time series consumption function: error correction, random walk and the steady-state. *Economic Journal 101*, 382–403.

Nelson, C. R. and H. Kang (1983). *Pitfalls in the use of time as an explanatory variable in regression.* NBER: Technical Working Paper. No. 30.

Nelson, C. R. and G. J. Plosser (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics 10*, 139–162.

Nickell, S. J. (1984). An investigation of the determinants of manufacturing employment in the United Kingdom. *Review of Economic Studies 51*, 529–558.

Nijman, T. E. and F. C. Palm (1990). Predictive accuracy gain from disaggregate sampling in ARIMA models. *Journal of Business and Economic Statistics 8*, 405–415.

Orcutt, G. H., H. W. Watts, and J. B. Edwards (1968). Data aggregation and information loss. *American Economic Review 58*, 773–787.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review 25*, 221–247.

Pagan, A. (1986). Two stage and related estimators and their applications. *Review of Economic Studies 53*, 517–538.

Palm, F. C. and T. E. Nijman (1994). Missing observations in the dynamic regression model. *Econometrica 52*, 1415–1435.

Pearlman, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika 67*, 232–233.

Perron, P. (1988). Trends and random walks in macroeconomic time series: further evidence from a new approach. *Journal of Economic Dynamics and Control 12*, 297–332.

Perron, P. (1989a). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica 57*, 1361–1401.

Perron, P. (1989b). Testing for a random walk: a simulation experiment of power when the sampling interval is varied. In B. Raj (Ed.), *Advances in Econometrics and Modelling.* Dordrecht: Kluwer Academic Publishers.

Perron, P. (1991). Test consistency with varying sampling frequency. *Econometric Theory 7*, 341–368.

Pesaran, B. (1988a). *Exact maximum likelihood estimation of regression models with invertible general order moving average disturbances.* London: National Institute of Economic and Social Research. Discussion paper no. 36.

Pesaran, B. and M. H. Pesaran (1991a). *Microfit 3.0: An Interactive Econometric Software Package.* Oxford: Oxford University Press. (forthcoming).

Pesaran, M. H. (1973). The small sample problem of truncation remainders in the estimation of distributed lag models with autocorrelated errors. *International Economic Review 14*, 120–131.

Pesaran, M. H. (1982). A critique of the proposed tests of the natural rate-rational expectations hypothesis. *Economic Journal 92*, 529–554.

Pesaran, M. H. (1987). *The Limits to Rational Expectations.* Oxford: Basil Blackwell.

Pesaran, M. H. (1988b). *Costly adjustment under rational expectations: a generalisation.* Los Angeles,. CA: UCLA. Working paper No. 480.

Pesaran, M. H. (1988c). On the policy ineffectiveness proposltion and a Keynesian alternative: a rejoinder. *Economic Journal 98*, 505–508.

Pesaran, M. H. (1991a). Costly adjustment under rational expectations: a generalisation. *Review of Economics and Statistics 73*, 353–358.

Pesaran, M. H. (1991b). Estimation of a simple class of multivariate rational expectations models: a test of the new classical model at a sectoral level. *Empirical Economics 16*, 211–232.

Pesaran, M. H. and B. Pesaran (1987a). *Data-FIT: An Interactive Econometric Software Package.* Oxford: Oxford University Press.

Pesaran, M. H. and B. Pesaran (1987b). *Microfit: An Interactive Econometric Software Package.* Oxford: Oxford University Press.

Pesaran, M. H. and B. Pesaran (1991b). *Microfit 3.0: An Interactive Econometric Software Package.* Oxford: Oxford University Press.

Pesaran, M. H. and R. G. Pierse (1989). A proof of the asymptotic validity of a test for perfect aggregation. *Economic Letters*. (forthcoming).

Pesaran, M. H., R. G. Pierse, and M. S. Kumar (1989a). Econometric analysis of aggregation in the context of linear prediction models. *Econometrica.* (forthcoming).

Pesaran, M. H., R. G. Pierse, and M. S. Kumar (1989b). Econometric analysis of aggregation in the context of linear prediction models. *Econometrica 57*, 861–888.

Pesaran, M. H., R. G. Pierse, and K. C. Lee (1991). Persistence, cointegration and aggregation: a disaggregated analysis of output fluctuations in the U.S. economy. *Journal of Econometrics*. (forthcoming).

Pesaran, M. H. and R. J. Smith (1989). A unified approach to estimation and orthogonality tests in linear single equation econometric models. *Journal of Econometrics*. (forthcoming).

Pesaran, M. H. and R. J. Smith (1994). A generalised R^2 and non-nested tests for regression models estimated by the instrumental variable method. *Econometrica.* (in press).

Pesaran, M. H. and R. P. Smith (1992). *Esimating long run relationships from dynamic heterogenous panels.* Cambridge: University of Cambridge. DAE DP Series no. 9215.

Pesaran, M. H. and A. G. Timmerman (1990). *The statistical significance of the predictability of excess returns on common stocks.* CA: UCLA. Programme in Applied Econometrics Discussion Paper no. 26.

Peterson, A. W. A. (1988). Employment. In T. S. Barker and A. W. A. Peterson (Eds.), *The Cambridge Multisectoral Dynamic Model*, Chapter 10. Cambridge: Cambridge University Press.

Phillips, P. C. B. (1987a). Time series regression with unit roots. *Econometrica 55*, 277–302.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika 74*, 535–547.

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*, 335–346.

Pierce, D. A. (1971). Distribution of residual autocorrelations in the regression model with autoregressive-moving average errors. *Journal of the Royal Statistical Society, Series B 33*, 140–146.

Pierse, R. G. (1982). *PERSEUS: a user's guide.* Program manual available from the author on request.

Powell, J. L. and T. M. Stoker (1985). The estimation of complete aggregation structures. *Journal of Econometrics 30*, 317–344.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* New York: Wiley. (2nd Ed.).

Report, R. (1959). *The Report of the Radcliffe Committee on the Working of the Monetary System.* London: Her Majesty's Stationary Office. Cmnd. 827.

Reynolds, R. A. (1982). Posterior odds for the hypothesis of independence between stochastic regressors and disturbances. *International Economic Review 23*, 479–490.

Richard, J.-F. (1979). Models with several regimes and changes in exogeneity. *Review of Economic Studies*. (forthcoming).

Richard, J.-F. (1980). Models with several regimes and changes in exogeneity. *Review of Economic Studies 47*, 1–20.

Richard, J.-F. (1984). Classical and Bayesian inference in incomplete simultaneous equation models. In D. F. Hendry and K. F. Wallis (Eds.), *Econometrics and Quantitative Economics.* Oxford: Basil Blackwell.

Richard, J.-F. and H. Tompa (1980). On the evaluation of poly-t density functions. *Journal of Econometrics 12*, 335–351.

Rosenberg, B. (1973). Random coefficient models: the analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement 2*, 399–428.

Rush, M. and D. Waldo (1988). On the policy ineffectiveness proposition and a Keynesian alternative. *Economic Journal 98*, 498–503.

Ruud, P. A. (1984). Tests of specification in econometrics. *Econometric Reviews 3*, 211–242.

Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika 71*, 599–607.

Sargan, J. D. (1964). Wages and prices in the United Kingdom; a study in econometric methodology. In P. E. Hart, G. Mills, and J. K. Whitake (Eds.), *Econometrics Analysis for National Economic Planning*, pp. 25–54. London: Butterworths.

Sargent, T. J. (1978). Estimation of dynamic labour demand schedules under rational expectations. *Journal of Political Economy 86*, 1009–1044.

Sasaki, K. (1978). An empirical analysis of linear aggregation problems. the case of investment behavior in Japanese firms. *Journal of Econometrics 7*, 313–331.

Schankerman, M. (1984). Comment on beggs' paper. In Z. Griliches (Ed.), *R&D, Patents and Productivity*. Chicago, Il: University of Chicago Press.

Schmidt, P. (1976). *Econometrics*. New York: Marcel Dekker Inc. Statistics: textbooks and monographs Volume 18.

Schmookler, T. (1966). *Invention and Economic Growth*. Cambridge, MA: Harvard University Press.

Schwert, G. W. (1989). Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics 7*, 147–159.

Shapiro, M. D. and M. W. Watson (1988). Sources of business cycle fluctuations. *NBER Macroeconomics Annual*.

Shiller, R. J. and P. Perron (1985). Testing the random walk hypothesis: power versus frequency of observation. *Economics Letters 18*, 381–386.

Smyth, D. (1984). Short-term employment functions when the speed of the adjustment depends on the unemployment rate. *Review of Economics and Statistics 63*, 138–142.

Stock, J. M. and M. W. Watson (1988). Journal of economic perspectives. *Econometrica 2*, 147–174.

Stoker, T. M. (1984). Completeness, distribution restrictions, and the form of aggregate functions. *Econometrica 52*, 887–907.

Stoker, T. M. (1986a). Aggregation, efficiency and cross-section regression. *Econometrica 54*, 171–188.

Stoker, T. M. (1986b). Simple tests of distributional effects on macroeconomic equations. *Journal of Political Economy 94*, 763–795.

Symons, J. S. V. (1985). Relative prices and the demand for labour in British manufacturing. *Economica 52*, 37–49.

Theil, H. (1954). *Linear Aggregation of Economic Relations*. Amsterdam: North-Holland.

Tompa, H. (1973). *The iterative Simpson method of numerical integration*. CORE Discussion paper 7336.

Treasury, H. M. (1985). *The Relationship Between Wages and Employment; A Review by Treasury Officials*. London: Her Majesty's Stationary Office.

van Daal, J. and A. H. Q. M. Merkies (1981). A simple proof of the Nataf theorem on consistent aggregation. *Economics Letters 7*, 145–150.

Wallis, K. F. (1974). Seasonal adjustment and relations between variables. *Journal of the American Statistical Association 69*, 18–31.

Watson, M. W. (1986). Univariate trending methods with stochastic trends. *Journal of Monetary Economics 18*, 49–75.

White, H. (1984). *Asymptotic Theory for Econometricians.* Orlando, FL: Academic Press.

Winters, L. A. (1980). Aggregation in logarithmic models: some experiments with UK exports. *Oxford Bulletin of Economics and Statistics 42*, 36–50.

Wold, H. and L. Jurlen (1953). *Demand Analysis.* New York: Wiley.

Wren-Lewis, S. (1986). An econometric model of U.K. manufacturing employment using survey data on expected output. *Journal of Applied Econometrics 1*, 297–316.

Wu, D.-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica 41*, 733–750.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association 57*, 348–368.

Zellner, A. (1969). On the aggregation problem: a new approach to a troublesome problem. In K. A. Fox, J. K. Sengupta, and G. V. L. Narasimham (Eds.), *Economic Models, Estimation, and Risk Programming: Essays in Honour of Gerhard Tinter*, pp. 365–378. London: Springer-Verlag.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley & Sons.