Lecture 7: Common Factor Tests and Stability Tests

R.G. Pierse

1 Common factor Tests

Consider again the regression model with first order autoregressive errors from lecture 4:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \quad , \quad t = 1, \cdots, T \tag{1.1}$$

and

$$u_t = \rho u_{t-1} + \varepsilon_t \quad , \quad t = 1, \cdots, T \tag{1.2}$$

where

$$-1 < \rho < 1 \,.$$

The transformed version of this equation is

$$Y_t - \rho Y_{t-1} = (1-\rho)\beta_1 + \beta_2(X_{2t} - \rho X_{2,t-1}) + \cdots + \beta_k(X_{kt} - \rho X_{k,t-1}) + \varepsilon_t \tag{1.3}$$

where ε_t obeys all the assumptions of the classical model. Note that this model has $k+1$ parameters.

Compare this with the unrestricted dynamic model:

$$Y_t = \gamma_1 + \rho Y_{t-1} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \gamma_2 X_{2,t-1} + \cdots + \gamma_k X_{k,t-1} + \varepsilon_t \,. \tag{1.4}$$

which has $2k$ parameters. Equation (1.3) is a special case of equation (1.4) that satisfies the $k-1$ *nonlinear* restrictions

$$\gamma_j = -\rho\beta_j \quad , \quad j = 2, \cdots, k \,.$$

This suggests that the autoregressive model can be viewed as a restricted form of a dynamic model, satisfying a particular type of parameter restriction known as *common factor restrictions*. Evidence of serial correlation from the residuals from estimating equation (1.1) may indicate omission of *any or all* of the variables Y_{t-1}, $X_{2,t-1}$, $\cdots$, $X_{k,t-1}$ rather than indicating the AR(1) model (1.3).

Sargan (1964) and Hendry and Mizon (1978) suggest a test of common factor restrictions in the model (1.4). This test is based on the statistic

$$T \log\left(\frac{RSS_r}{RSS_u}\right)$$

where RSS_u is the residual sum of squares from the unrestricted model (1.4) and RSS_r is the residual sum of squares from the restricted model (1.3). This test statistic is a form of *Likelihood Ratio* or *LR* test and it can be shown that

$$T \log\left(\frac{RSS_r}{RSS_u}\right) \sim_a \chi^2_{k-1} \tag{1.5}$$

This is a test of the restrictions implicit in the autoregressive errors hypothesis. Sargan and Hendry and Mizon suggest the following testing procedure on finding significant autocorrelation in OLS estimation of (1.1)

(a) Estimate the unrestricted model (1.4)

(b) Test for *common factors* using the test (1.5)

(c) Only if the null hypothesis is *not* rejected then estimate the AR(1) model (1.3) by Cochrane-Orcutt.

2 Stability tests

Consider the classical regression equation

$$y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \quad , \quad t = 1, \cdots, T\,. \tag{2.1}$$

When this equation is estimated, it is assumed that the model parameters β are constant over the entire sample period. This is the assumption of *parameter constancy* or *stability*. Two different tests relating to this hypothesis have been proposed.

2.1 *F*-Test for coefficient stability

Split the data period into two sub-samples: $1, \cdots, T_1$ and $T_1 + 1, \cdots, T$ with T_1 and $T_2 = T - T_1$ observations respectively. Then, as long as $T_1 > k$ and $T_2 > k$ so that the two sub-samples are both long enough to permit estimation, the equation can be estimated separately for each sub-sample giving regression equations

$$y_t = \beta^1_1 + \beta^1_2 X_{2t} + \cdots + \beta^1_k X_{kt} + u_t \quad , \quad t = 1, \cdots, T_1 \tag{2.2}$$

and

$$y_t = \beta^2_1 + \beta^2_2 X_{2t} + \cdots + \beta^2_k X_{kt} + u_t \quad , \quad t = T_1 + 1, \cdots, T \tag{2.3}$$

respectively.

A test for stability is then a test of the hypothesis that

$$H_0 : \beta_i^1 = \beta_i^2 \quad , \quad i = 1, \cdots, k\,.$$

In this case, pooling the two sub-samples to create the regression equation (2.1) is justified.

A test of the hypothesis of parameter constancy can be based on the residual sum of squares from the two regressions (denoted as RSS_1 and RSS_2 respectively), and the residual sum of squares of the regression on the pooled data (RSS). On the assumption that the two sub-samples are *independent*, then

$$\frac{(RSS - RSS_1 - RSS_2)}{(RSS_1 + RSS_2)} \frac{T - 2k}{k} \sim F_{k,T-2k}\,.$$

This test is known as the *F-test for coefficient stability.* An important maintained assumption in this analysis is that the error variance of u_t in the two-sub-samples is equal. If this is not the case then the pooled regression equation will be heteroscedastic and the test will be invalid. Note that this test is sometimes confusingly referred to as a 'Chow' test although it is quite different from the test in the next section.

2.2 Chow test of predictive failure

Chow (1960) proposes a test for *predictive failure.* The test is based on estimation of the model on the first sub-sample, and over the complete sample. Thus it can be computed even when $T_2 < k$ so that there are two few observations in the second sub-sample to permit separate estimation.

Chow shows that

$$\frac{(RSS - RSS_1)}{RSS_1} \frac{T - k}{T_2} \sim F_{T_2,T-k}$$

This test is equivalent to a joint test on the dummies D_1 to D_{T_2} being equal to zero in the model

$$y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \gamma_1 D_1 + \cdots \gamma_{T_2} D_{T_2} + u_t \quad , \quad t = 1, \cdots, T\,. \tag{2.4}$$

where D_j is a dummy variable taking the value 1 in period $T_1 + j$ and zero elsewhere. Pesaran, Smith and Yeo (1985) provide a useful interpretation of the Chow test.

The two hypotheses tested in the coefficient stability test and the predictive failure test are quite different. The first tests whether the estimated parameters in the two sub-samples are significantly different from each other. The Chow test, on the other hand tests whether the parameters estimated over the first sub-sample continue to predict over the second sub-sample.

Failure of either coefficient stability or predictive failure test may be interpreted as evidence of a structural break in the data. The Great Crash, the oil price shocks of the 1970s and exchange rate revaluations in a fixed exchange rate regime, are all examples of events which might be expected to cause such structural breaks at specific points in time. However, unless there is good *a priori* reason to expect that such a structural break has occurred, then failure of these tests is more likely to be evidence of misspecification of the equation. Conversely, evidence of coefficient stability and predictive power outside the sample period is encouraging evidence in favour of the estimated equation.

3 Diagnostic Tests in *EViews*

The econometrics package *EViews* makes available several standard diagnostic tests after running any *OLS* regression. Some of these tests are presented in two forms: an *LM* form which is asymptotically distributed as chi-squared with appropriate degrees of freedom, and an F version which is derived from the chi-squared form by the relationship:

$$F(p) = \frac{T-k}{p}\frac{\chi^2(p)}{n-\chi^2(p)} \sim_a F_{p,T-k}$$

where $\chi^2(p)$ is the original chi-squared statistic with p degrees of freedom, T is the number of observations, and k is the number of regressors in the *auxiliary* regression from which the chi-squared statistic is calculated. The F version, sometimes known as a *'modified' LM statistic*, will generally have more power is small samples.

3.1 Serial Correlation

EViews provides the *Breusch-Godfrey Lagrange Multiplier* test for autocorrelation, Breusch (1978) and Godfrey (1978), as a serial correlation diagnostic statistic, based on the regression of the OLS residuals e_t on the k regressors and the lagged residuals e_{t-1},e_{t-2},$\cdots$,e_{t-p}. The choice of p is left to the user but, conventionally, we would choose $p = 4$ for quarterly data, and $p = 1$ for annual data. The *LM* form of the statistic, based on the R^2 from this auxiliary regression, is given by

$$(T-p)R^2 \sim_a \chi^2(p)\,.$$

The null hypothesis is that there is no serial correlation, the alternative hypothesis that there is serial correlation of pth order. In *EViews*, this test is selected from the *Residual tests* menu on the *View* button in the equation window.

3.2 Heteroscedasticity

EViews provides the White (1980) statistic as a test for general heteroscedasticity of unknown form. It is based on a regression of the squared residuals e_i^2 on all (non-redundant) cross-products of the regressors $X_{ji}X_{li}$, for all $j = 1, \cdots, k$, and $l = 1, \cdots, k$. The auxiliary regression takes the form:

$$\begin{aligned} e_i^2 &= \gamma_1 X_{1i}^2 + \gamma_2 X_{1i} X_{2i} + \cdots + \gamma_k X_{1i} X_{ki} \\ &\quad + \gamma_{k+1} X_{2i}^2 + \gamma_{k+2} X_{2i} X_{3i} + \cdots + \gamma_{2k-1} X_{2i} X_{ki} + \cdots + \gamma_{k(k+1)/2} X_{ki}^2 + u_i. \end{aligned}$$

The test statistic is based on the R^2 from this auxiliary regression and is given by

$$nR^2 \sim_a \chi^2_{k(k+1)/2-1}.$$

An alternative version of the test, dropping the cross-products $X_{ji}X_{li}$, $l \neq j$, is based on the regression

$$e_i^2 = \gamma_1 X_{1i}^2 + \gamma_2 X_{2i}^2 + \cdots + \gamma_k X_{ki}^2 + u_i.$$

and is given by

$$nR^2 \sim_a \chi^2_{k-1}.$$

The null hypothesis of these tests is that there is no heteroscedasticity so that errors are homoscedastic. In *EViews*, this test is selected from the *Residual tests* menu on the *View* button in the equation window. Both forms of the test (either including or excluding cross–product terms) are available.

3.3 Functional Form

EViews provides *Ramsey's RESET test*, Ramsey (1969), which is a statistic for testing for functional form misspecification. The test is is based on the R^2 from an an auxiliary regression of e_t on the k regressors and $\widehat{y}_t^2$, the squared fitted values from the original regression. The null hypothesis of this test is that the equation has the correct functional form, and the alternative hypothesis that there is functional misspecification. The statistic is is given by

$$(T-1)R^2 \sim_a \chi^2(1)\,.$$

In *EViews*, this test is selected from the *Stability tests* menu on the *View* button in the equation window.

3.4 Normality

EViews provides the *Jarque-Bera test*, Jarque and Bera (1980), as a test for normality of the residuals. This test is based on the estimated moments of the residuals, given by

$$\mu_j = \frac{1}{T}\sum_{t=1}^{T} e_t^j \quad , \quad j = 1, 2, \cdots .$$

Note that μ_1 is the estimated mean of the residuals and μ_2 their estimated variance. The third moment μ_3 measures *skewness* and the fourth moment μ_4 measures *kurtosis*. The test is based on the fact that, for a normal distribution, these moments have a particular form. The test is given by

$$T\left[\frac{\mu_3^2}{6\mu_2^3} + \frac{1}{24}\left(\frac{\mu_4}{\mu_2^2} - 3\right)^2 + \frac{3\mu_1^2}{2\mu_2} - \frac{\mu_3\mu_1}{\mu_2^2}\right] \sim_a \chi^2(2)\,.$$

The null hypothesis of the test is that the residuals are normally distributed, against the alternative that the distribution is non-normal. In *EViews*, this test is selected from the *Residual tests* menu on the *View* button in the equation window.

References

[1] Breusch, T.S. (1978),'Testing for autocorrelation in dynamic linear models', *Australian Economic Papers*, 17, 334–355.

[2] Chow, G.C. (1960), 'Tests of equality between subsets of coefficients in two linear regression models', *Econometrica*, 28, 591–605.

[3] Godfrey, L.G. (1978), 'Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables', *Econometrica*, 46, 1293–1302.

[4] Hendry, D.F. and G.E. Mizon (1978), 'Serial correlation as a convenioent simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England', *The Economic Journal*, 88, 549–563.

[5] Jarque, C.M. and A.K. Bera (1980), 'Efficient tests for normality, homoscedasticity and serial independence of regression residuals', *Economic Letters*, 6, 255–9.

[6] Pesaran, M.H., R.P. Smith and J.S. Yeo (1985), 'Testing for structural stability and predictive failure: a review', *Manchester School*, 53, 280–295.

[7] Ramsey, J.B. (1969), 'Tests for specification errors in classical least squares regression analysis', *Journal of the Royal Statistical Society, Series B*, 31, 350–71

[8] 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. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, Colston Papers 16, Butterworth, London. Reprinted in D.F. Hendry and K.F. Wallis (eds.) *Econometrics and Quantitative Economics*, (1984), Basil Blackwell, Oxford.

[9] White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica*, 48, 817–838.