Economic Forecasting
Lecture 2: Forecasting the Trend

Richard G. Pierse

1 Introduction

Trends are an important feature of many economic time series. In this lecture we look at some simple models to forecast trends. There are two main classes of model that have been proposed in the literature: *deterministic trend* models and *stochastic trend* models. These two classes of model imply very different implications for the effect of shocks on the future path of a time series. Tests are available to help us to choose between them.

2 Stationarity

Before we start looking at models of trends we need to define the important concept of stationarity. There are weaker and stronger definitions of stationarity but, for our purposes, it is sufficient to consider the definition of what is known as *weak stationarity* or *covariance stationarity.*

A time series $\{y_t\}$ is said to be *covariance stationary* if it satisfies the following three conditions:

$$\mathrm{E}(y_t) = \mu \tag{1}$$

$$\mathrm{var}(y_t) \equiv \mathrm{E}(y_t - \mu)^2 = \sigma^2 \tag{2}$$

and

$$\mathrm{cov}(y_t, y_{t-s}) \equiv \mathrm{E}(y_t - \mu)(y_{t-s} - \mu) = \gamma_s. \tag{3}$$

The first condition (1) states that the mean of the series is the same in every time period t. The second condition states that the *variance* of the series is also the same in every time period t. The third condition states that the *covariance* between the two observations y_t and y_{t-s} depends only on the *distance* between them, s, and is the same in every time period t.

A trended time series is one with a non-constant mean and so *violates* the first condition for covariance stationarity. We will see that a time series with a *stochastic trend* also violates the second and third conditions. Detrending of a trended time series removes the trend to leave a detrended series that will be covariance stationary. The method of detrending will depend on the form of the trend: *deterministic* or *stochastic*.

3 Deterministic Trend Models

3.1 The linear trend model

The simplest deterministic trend model is the *linear trend* model:

$$T_t = \beta_0 + \beta_1 TIME_t$$

where *TIME* is an artificial *time dummy* taking the values $TIME_1 = 1$, $TIME_2 = 2$, $TIME_3 = 3$ *etc.* so that we can say that $TIME_t = t$. The parameter β_0 is the *intercept* and gives the value of the trend when $t = 0$. The parameter β_1 is the *slope*; when β_1 is positive, the trend is increasing and when β_1 is negative, the trend is decreasing. The absolute size of β_1 specifies the steepness of the slope of the trend. In economics and business, linear trends are typically increasing, corresponding to growth, but this need not be the case and there are examples of economic series with decreasing trends such as male labour force participation (Figure 1).

In order to forecast with the linear trend model we must first estimate the unknown parameters β_0 and β_1. Writing

$$y_t = \beta_0 + \beta_1 TIME_t + \varepsilon_t \tag{4}$$

where y_t is the time series to be forecast and ε_t is a disturbance with $\mathrm{E}(\varepsilon_t) = 0$ and $\mathrm{E}(\varepsilon_t^2) = \sigma^2$, the parameters β_0 and β_1 can be estimated by *OLS* regression applied to equation (4). Taking expectations of (4),

$$\mathrm{E}(y_t) = \beta_0 + \beta_1 t$$

which depends on time so that the series y_t is *not stationary* although

$$\mathrm{var}(y_t) = \mathrm{E}(\varepsilon_t^2) = \sigma^2$$

so that it does have constant variance.

Note that from the decomposition

$$y_t = T_t + C_t + S_t + u_t$$

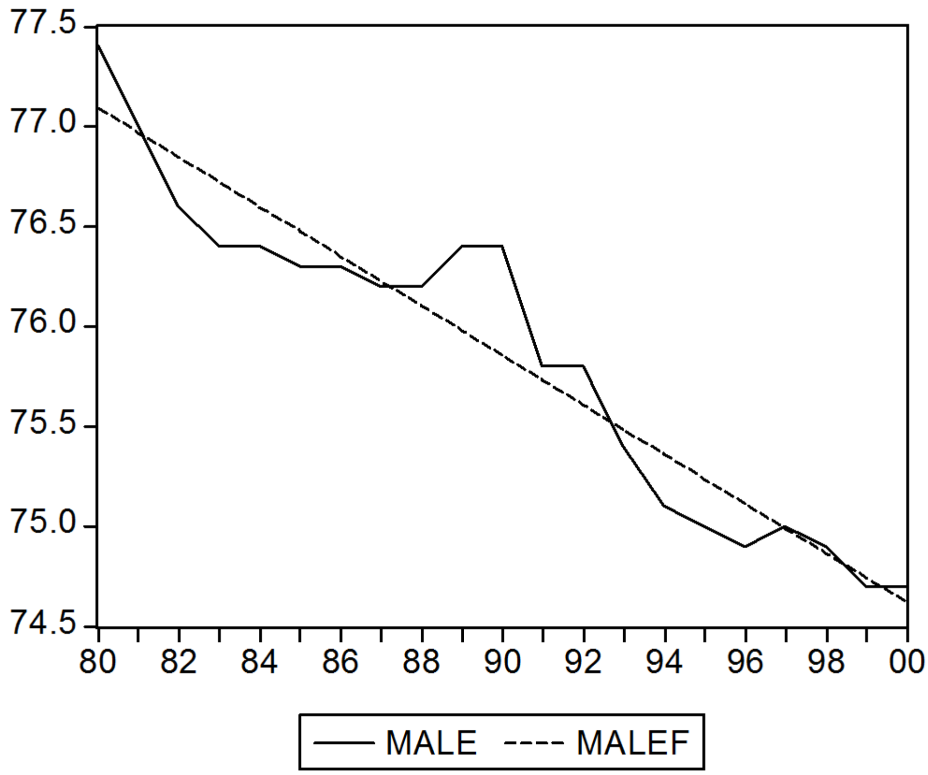

Figure 1: US Male Participation Rates: 1980-2000

discussed in last week's lecture it follows from (4) that the *OLS* disturbance

$$\varepsilon_t = C_t + S_t + u_t$$

is the sum of the cyclical, seasonal and irregular components of the time series. While it is reasonable to assume that each of these components has zero mean with constant variance, it is likely that the cyclical component (and hence ε_t) will be *autocorrelated*. In this case, although the *OLS* parameter estimates will still be unbiased, the standard errors of those parameter estimates will underestimate the true standard errors.

Having estimated the parameters, the forecast of the linear trend model is given by

$$\widehat{T}_t = \widehat{\beta}_0 + \widehat{\beta}_1 \, TIME_t$$

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ denote *OLS* estimates of the parameters β_0 and β_1. This forecast will be a straight line as shown by the dotted lines in Figures 1 and 2.

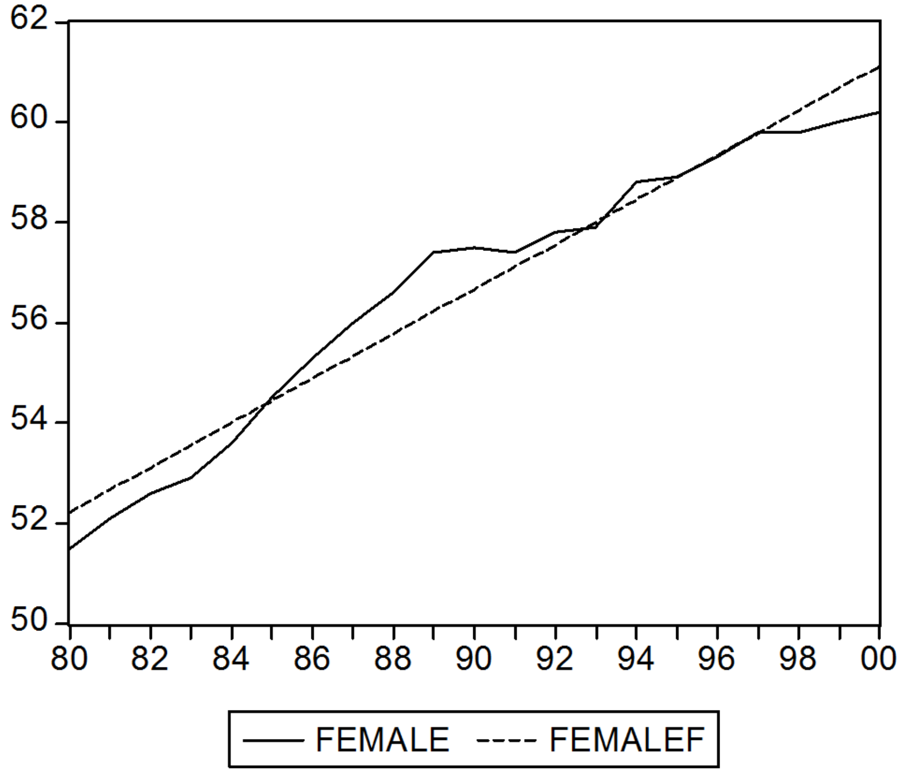

Figure 2: US Female Participation Rates 1980-2000

3.2 The quadratic trend model

In the linear trend model, the predicted trend is a straight line. The quadratic trend model

$$T_t = \beta_0 + \beta_1 \, TIME_t + \beta_2 \, TIME_t^2$$

allows for a non-straight trend line. Various shapes are possible according to the values of the parameters β_1 and β_2. When both parameters are positive, the trend will be monotonically increasing but at a super-linear rate. Conversely, when both parameters are negative, the trend will be monotonically decreasing at a super-linear rate. When $\beta_1 < 0$ and $\beta_2 > 0$, the trend will have a u-shape and when $\beta_1 > 0$ and $\beta_2 < 0$, the trend will have an inverted u-shape.

Figure 3 shows a quadratic trend fitted to the female participation rate time series (the coefficient β_2 was not significant for the case of the male participation rate series, showing that a linear trend is sufficient there).

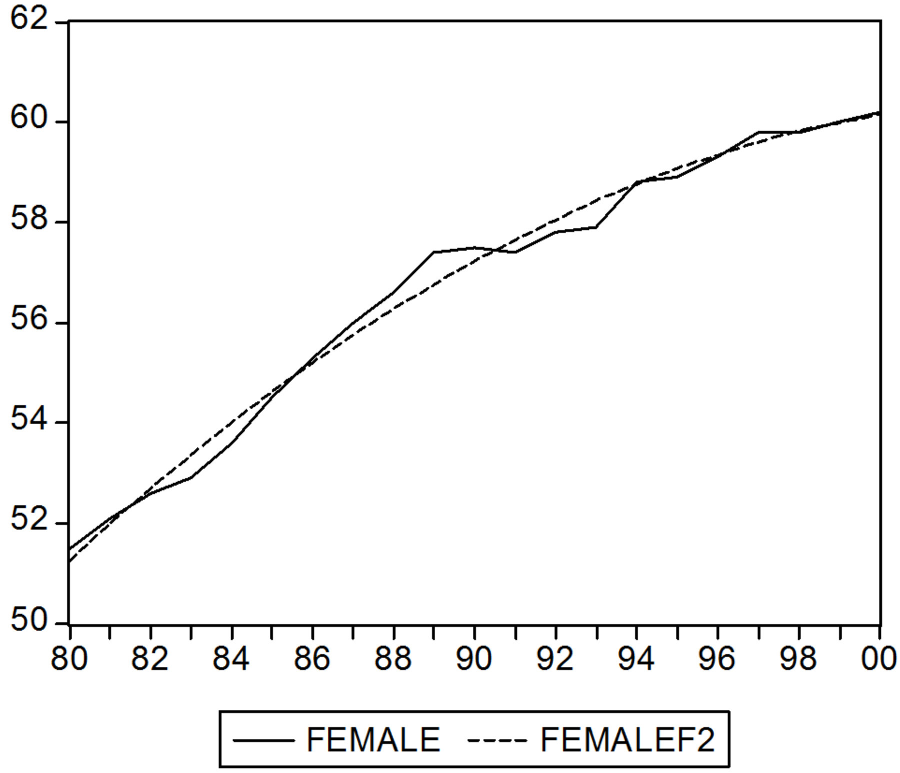

Figure 3: Quadratic trend fitted to US Female Participation Rates

Higher order polynomial trends are possible such as the k-th order polynomial trend

$$T_t = \beta_0 + \beta_1 \, TIME_t + \beta_2 \, TIME_t^2 + \beta_3 \, TIME_t^3 + \cdots + \beta_k \, TIME_t^k.$$

The higher the order of the polynomial, the more flexible is the shape of the trend. However, there is a serious danger of over-fitting when k is allowed to be too large and, in practice, linear or quadratic trends are normally sufficient.

3.3 The exponential trend model

An alternative nonlinear model for the trend is the *exponential trend* model

$$T_t = \beta_0 e^{\beta_1 \, TIME_t}.$$

Taking logarithms we get

$$\log T_t = \log \beta_0 + \beta_1 TIME_t$$

so that the logarithm of the trend is a linear function of time. The parameters of this model can therefore be estimated by fitting a linear trend model to the logged time series.

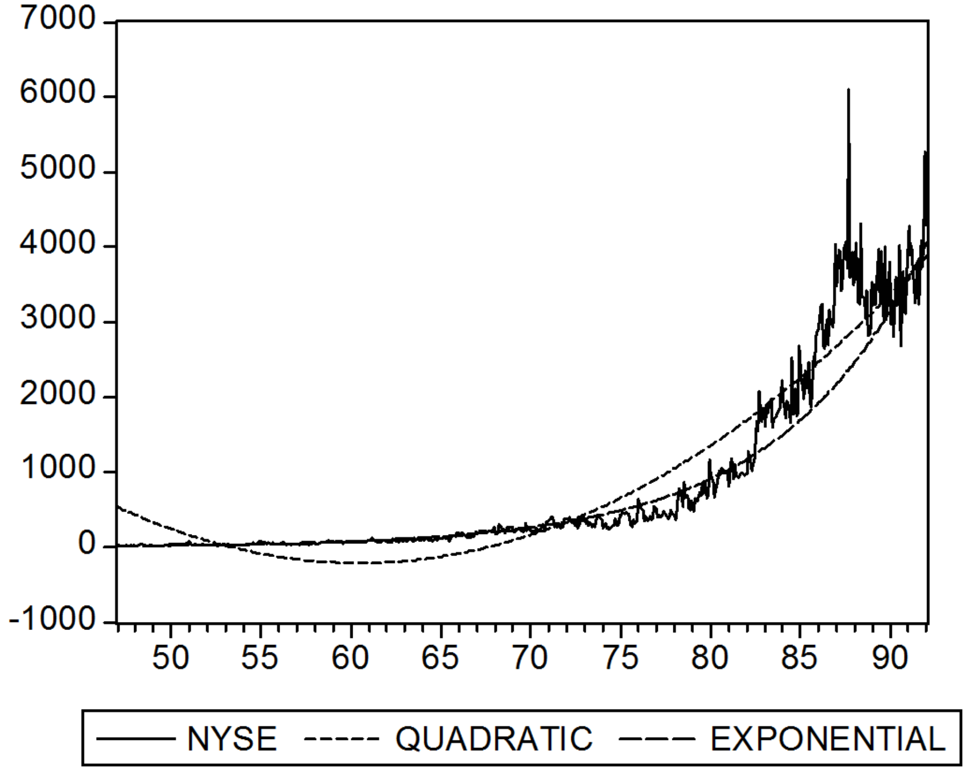

Figure 4: Quadratic v. exponential trend forecasts for NYSE data

Figure 4 shows a comparison between the quadratic trend model and the exponential trend model fitted to the New York Stock Exchange volumes data set. It appears from 4 that the exponential trend model fits the data better despite being based on fewer estimated parameters. This reflects the log-linearity characteristic of many economic time series.

3.4 Choosing between models: information criteria

We have looked at several different deterministic models of the trend. In each case, the *forecast errors* from each model are given by the *OLS* regression residuals

$$e_t \equiv y_t - \widehat{T}_t.$$

Since the *OLS* estimator by definition *minimises* the sum of squared residuals

$$\sum_{t=1}^{T} e_t^2$$

or, equivalently, the *Mean Square Error (MSE)*

$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2 = \widehat{E}(e_t^2),$$

it follows that the forecasts from each model will be *optimal* if the forecaster has the *quadratic loss function*

$$L(e_t) = e_t^2.$$

How does the forecaster choose between competing models? It might seem tempting to choose the model with the smallest *MSE*, which is equivalent to choosing the model with the largest R^2. However, the *MSE* is not a good criterion for selecting a model because it will *never* favour a parsimonious model with fewer parameters. The reason is that the *MSE cannot increase* when an extra coefficient is added to a regression and can only stay the same or fall. This means that, in choosing between a linear trend and a quadratic trend, the *MSE* criterion will always favour the quadratic trend, simply because it has an extra parameter.

In last week's lecture, we discussed the importance of *parsimony* in selecting a forecasting model. This suggests that, when comparing forecasts from competing models, we should attach a penalty to a model with more parameters. Two important information criteria have been proposed in the literature. Both involve adjusting the *MSE* by a multiplicative penalty related to the number of model parameters, k. In choosing between models, the model with the smallest value of the criterion is selected.

The *Akaike Information Criterion (AIC)* is defined by

$$AIC = \exp(\frac{2k}{T})\frac{1}{T}\sum_{t=1}^{T} e_t^2$$

and the *Schwarz Information Criterion (SIC)* (sometimes known as the *Bayesian Information Criterion (BIC)*), is defined by

$$SIC = T^{\frac{k}{T}} \frac{1}{T} \sum_{t=1}^{T} e_t^2.$$

Both information criteria penalise a model with more parameters (larger k) but the *SIC* involves a higher penalty than the *AIC*. This means that the *SIC* is more likely to choose the more parsimonious model than the *AIC*. Which criterion should we use? It can be shown that the *SIC* is *consistent* in the sense that, when the 'true' model is among the models considered, then the probability of selecting it approaches 1 as the sample size increases. On the other hand, the *AIC* is *asymptotically efficient* in the sense that, as the sample size increases, it will select a sequence of models aproaching the 'true' model at least as fast as any other criterion. In practice, both criteria are used and they often lead to selection of the same model. Where there is a conflict, the *parsimony principle* would suggest using the *SIC* criterion.

To illustrate the use of information criteria to select a model, the table below shows the *MSE* and the *AIC* and *SIC* criteria for the linear, quadratic and exponential trend models applied to the male and female participation rates data graphed in Figures 1–3. (In order to make the exponential model comparable, it was estimated in an anti-log form so that the dependent variable is the same as for the other two specifications.)

	MSE	*AIC*	*SIC*	*MSE*	*AIC*	*SIC*
Linear	0.0496	0.025	0.125	0.3239	1.901	2.000
Quadratic	0.0490	0.107	0.256	0.0768	0.557	0.706
Exponential	0.0498	0.029	0.128	0.3832	2.069	2.169

For the male participation rate (columns 1–3) both *AIC* and *SIC* select the linear model even though, as expected, the quadratic model has the smaller *MSE*. For the female participation rate (columns 4–6) both *AIC* and *SIC* select the quadratic model graphed in Figure 3. Thus, in this example, the two criteria agree on the choice of model although this will not always be the case.

4 Stochastic Trend Models

4.1 The Random Walk with Drift

The simplest stochastic trend model is the random walk with drift

$$T_t = \delta + T_{t-1} + \varepsilon_t \tag{5}$$

where δ is the *drift* parameter and ε_t is a random disturbance that is assumed to be mean zero, constant variance and non-autocorrelated, or, formally,

$$\mathrm{E}(\varepsilon_t) = 0,\ \mathrm{E}(\varepsilon_t^2) = \sigma^2,\ \mathrm{E}(\varepsilon_t \varepsilon_s) = 0,\ s \neq t.$$

Such a disturbance is known as a '*white noise*' process.

The random walk with drift is a *non-stationary* process as can be seen by repeated back-substitution in (5) to get

$$T_t = T_0 + t\delta + \sum_{i=1}^{t} \varepsilon_i$$

so that

$$\mathrm{E}(T_t) = T_0 + t\delta \tag{6}$$

and

$$\mathrm{var}(T_t) \equiv \mathrm{E}(T_t - \mathrm{E}(T_t))^2 = \mathrm{E}(\sum_{i=1}^{t} \varepsilon_i)^2 = t\sigma^2. \tag{7}$$

From (6) it can be seen that the mean is non-constant (except when the drift parameter δ is zero) and depends on t and from (7) it follows that the variance is also non-constant and increases with t. In the limit, as $t \to \infty$, the variance becomes infinite. When the drift parameter δ is positive, the mean is increasing in t and the process has a general upward trend. Conversely, when the drift parameter δ is negative, the mean is decreasing in t and the process has a general downward trend. In the special case when $\delta = 0$, the process has no trend.

However, despite the effect of the drift, the presence of the stochastic disturbance u_t means that a random walk with drift can wander arbitrarily far from its mean trend path as is illustrated in Figure 5, which simulates a random walk with $\delta = 0.01$ and $\sigma^2 = 0.01$. Despite the positive drift, this series has significant periods of trending downwards.

Another way of writing (5) is

$$\Delta T_t = \delta + \varepsilon_t$$

where Δ is the first-order difference operator defined by $\Delta x_t \equiv x_t - x_{t-1}$. Note that

$$\mathrm{E}(\Delta T_t) = \delta$$
$$\mathrm{var}(\Delta T_t) = \mathrm{E}(\varepsilon_t^2) = \sigma^2$$

and

$$\mathrm{cov}(\Delta T_t, \Delta T_{t-s}) = \mathrm{E}(\varepsilon_t \varepsilon_{t-s}) = 0,\ s \neq t$$

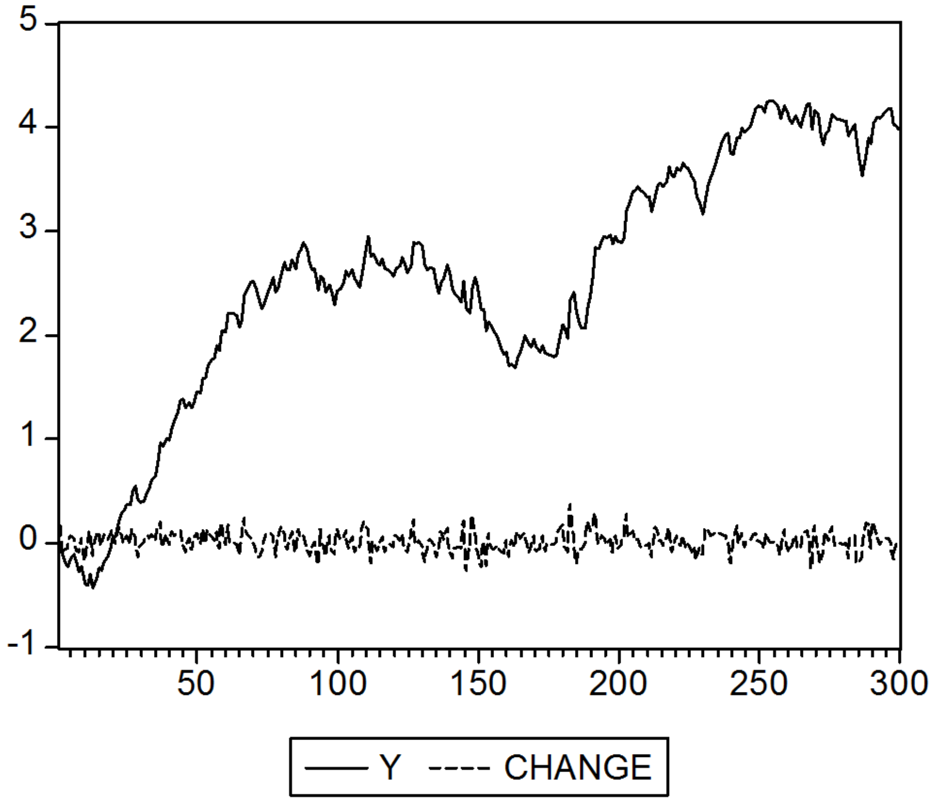

Figure 5: Random walk with drift ($\delta = 0.01$ and $\sigma^2 = 0.01$). Level and Change

so that ΔT_t is *covariance stationary.*

A variable that can be made stationary by differencing is known as an *integrated* variable and denoted as $I(d)$ where d indicates the number of times that the variable needs to be differenced to make it stationary. In this case, d is 1 so we say that T_t is $I(1)$ or '*integrated of order 1*'.

4.2 The log random walk model with drift

A variant of the random walk model with drift is the log random walk model with drift defined by

$$T_t = \delta T_{t-1} e^{\varepsilon_t}.$$

Taking logarithms of this equation we obtain

$$\log(T_t) = \log(\delta) + \log(T_{t-1}) + \varepsilon_t$$

which is a random walk with drift in the logarithm of T_t. In this model the stationary variable is the trend *growth rate*

$$\Delta \log(T_t) = \log(\delta) + \varepsilon_t$$

rather than the difference ΔT_t. This model is probably the most widely used stochastic trend model in economic forecasting.

4.3 Forecasting with random walk models

There is one unknown parameter to be estimated in the random walk model with drift, the drift parameter δ. This can be estimated by the regression of the first difference of the data series Δy_t on an intercept:

$$\Delta y_t = \delta + u_t.$$

Then, *one-step ahead forecasts* can be computed by

$$\Delta \widehat{y}_{t+1,t} = \widehat{\delta}$$

or

$$\widehat{y}_{t+1,t} = y_t + \widehat{\delta}.$$

For the log random walk model with drift we use the regression

$$\Delta \log(y_t) = \log(\delta) + u_t$$

and compute *one-step ahead forecasts* by

$$\Delta \log(\widehat{y}_{t+1,t}) = \log(\widehat{\delta})$$

or

$$\log(\widehat{y}_{t+1,t}) = \log(y_t) + \log(\widehat{\delta})$$

or

$$\widehat{y}_{t+1,t} = y_t \widehat{\delta}.$$

For *multi-step ahead forecasts* we have

$$\begin{aligned} \widehat{y}_{t+h,t} &= \widehat{y}_{t+h-1,t} + \widehat{\delta} \\ &= y_t + h\widehat{\delta} \end{aligned}$$

and

$$\begin{aligned} \widehat{y}_{t+h,t} &= \widehat{y}_{t+h-1,t}\widehat{\delta} \\ &= y_t \widehat{\delta}^h \end{aligned}$$

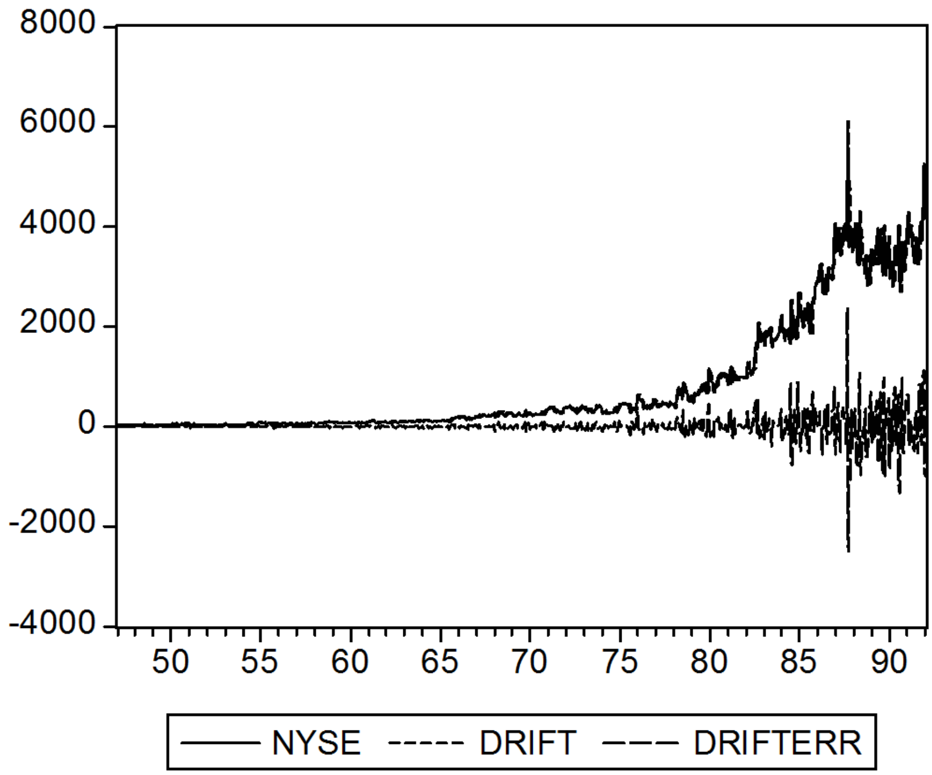

Figure 6: Random walk model with drift fitted to NYSE data

respectively. It is clear that these multi-step ahead forecasts simply project constant changes (growth rates) into the future and will be expected quickly to go off track.

For one-step ahead forecasting however, the random walk models can do very well as is illustrated in Figures 6 and 7 where the random walk model and the log random walk model are both fitted to the NYSE data series. Both models fit notably better than the deterministic trend forecasts in Figure 4, and there is very little to choose between them. (The *AIC* and *SIC* criteria marginally favour the log random walk model). However, the figures also graph the *one-step ahead forecast errors* which clearly show an increasing variance over time, evidence of *heteroscedasticity*.

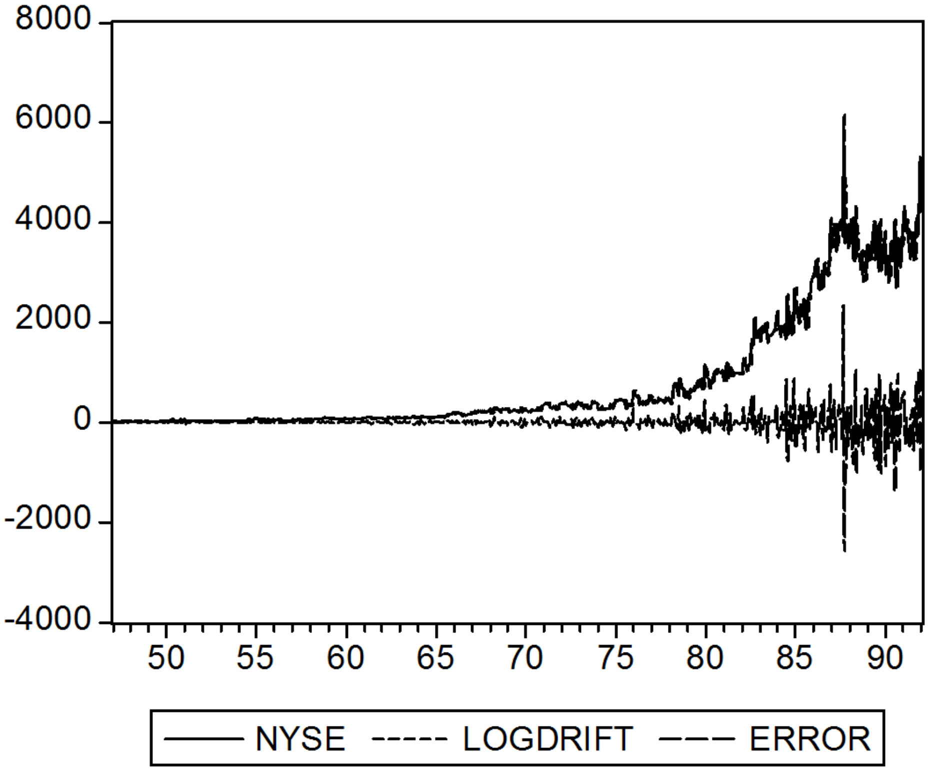

Figure 7: Log random walk model with drift fitted to NYSE data

4.4 The local linear trend model

A more general stochastic trend model than the random walk with drift is the *local linear trend* model defined by the two equations:

$$
\begin{aligned}
T_t &= T_{t-1} + \delta_{t-1} + \varepsilon_t \\
\delta_t &= \delta_{t-1} + \eta_t
\end{aligned}
$$

where ε_t and η_t are both independent, normally distributed, white noise processes with

$$\mathrm{E}(\varepsilon_t) = 0,\ \mathrm{E}(\varepsilon_t^2) = \sigma^2,\ \mathrm{E}(\eta_t) = 0,\ \mathrm{E}(\eta_t^2) = \omega^2.$$

In this model the drift, δ_t, is a stochastic process that follows a random walk. In the special case where $\mathrm{var}(\eta_t) = \omega^2 = 0$, the model reduces to the random walk model with drift

$$\Delta T_t = \delta + \varepsilon_t.$$

Alternatively, in the special case $\text{var}(\varepsilon_t) = \sigma^2 = 0$, the model reduces to

$$\Delta^2 T_t = \eta_{t-1}$$

which is an $I(2)$ stochastic process and, in the general case, the model can be rewritten as

$$\Delta^2 T_t = \eta_{t-1} + \Delta\varepsilon_t$$

which is also an $I(2)$ stochastic process since η_{t-1} and $\Delta\varepsilon_t$ are both stationary.

In the general case, the effect of δ_t is to allow the slope of the trend to change over time whereas in the random walk model with drift the slope is constant at δ. Estimation of the unknown parameters in the *local linear trend* model, the two variances σ^2 and ω^2, can be achieved by using the *Kalman filter* (see Harvey (1989) for details) but this is beyond the scope of these lectures.

As we will see, empirical investigation suggests that many economic time series are *I(1)* (generally in logarithms) but few are *I(2)*. This suggests that the *local linear trend* model may be more flexible than necessary to model the trend in economic variables. However, one area in which a special case of the local linear trend model has proved very popular is in the *Hodrick-Prescott (H-P) filter* used to detrend economic series. It can be shown that applying the *H-P* filter is equivalent to estimating the trend T_t in the model

$$\begin{aligned} y_t &= T_t + u_t \\ T_t &= T_{t-1} + \delta_{t-1} \\ \delta_t &= \delta_{t-1} + \eta_t \end{aligned}$$

where u_t is an irregular stationary component with variance $\mathrm{E}(u_t^2)$, and the model is estimated subject to the restriction that $\eta_t^2) = q\,\mathrm{E}(u_t^2)$ where q is a smoothing parameter (set equal to 1/1600 for quarterly data). Several authors have pointed out that the flexibility of the trend in the *H-P* filter means that it can remove more than just the trend and, paradoxically, can introduce spurious cycles into the residual irregular component.

5 Choosing Between Deterministic and Stochastic Trends

Consider again the linear deterministic trend model

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \tag{8}$$

and the random walk model with drift

$$\Delta y_t = \delta + \varepsilon_t. \tag{9}$$

Without any disturbances ε_t, the two models are indistinguishable. To see this, set $\varepsilon_t = 0$, lag (8) by one period and subtract it from itself to get

$$y_t - y_{t-1} = (\beta_0 + \beta_1 t)\text{-}(\beta_0 + \beta_1(t\text{-}1))$$

or

$$\Delta y_t = \beta_1$$

which is the same as (9) where $\beta_1 = \delta$ is the drift. Initialising the random walk model by setting $y_0 = \beta_0$, both models will generate identical forecasts in the absence of shocks.

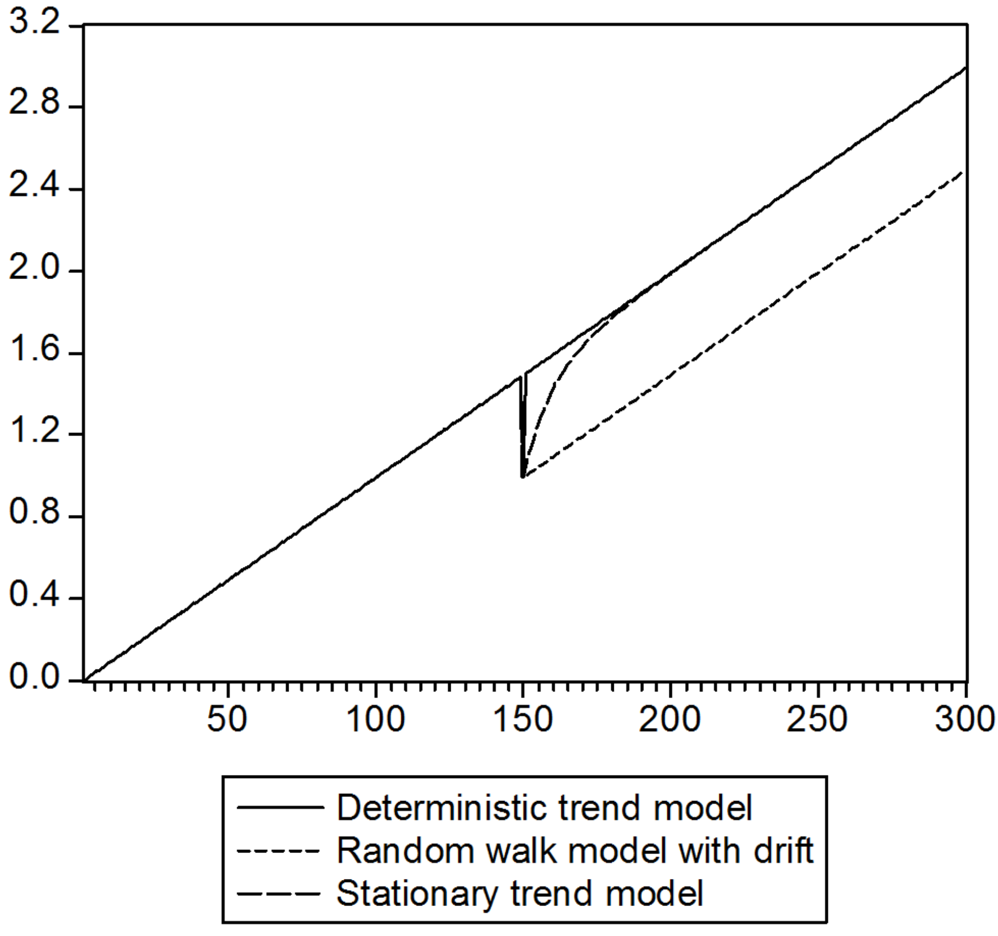

Figure 8: Linear trend v. random walk with drift: response to a single shock

The key difference between the deterministic and stochastic trend models lies in the way that they respond to shocks. In the deterministic model, shocks have only a temporary impact whereas in the random walk with drift, shocks have a permanent effect on the path of the time series. This is illustrated in Figure 8 which shows the response of the two models to a single negative shock in period 150. In the deterministic trend model, the series returns to its trend immediately. In the random walk model with drift, the series never returns to its former trend but continues after the shock from where it is. The shock thus has a permanent effect on the time path.

A third intermediate case is given by the *stationary trend model*

$$y_t = \beta_0 + \beta_1 t + \beta_2 y_{t-1} + \varepsilon_t \tag{10}$$

where the parameter β_2 satisfies $|\beta_2| < 1$. (The deterministic trend model (8) is just a special case of this model in which $\beta_2 = 0$). In this model, the effect of a shock dies away gradually as illustrated by the broken line in Figure 8 and eventually the series resumes its former trend. The time that it takes for the effect of the shock to completely die away depends on the magnitude of the parameter β_2, the closer $|\beta_2|$ is to 1, the longer it takes. In the limiting case where $\beta_2 = 1$, the shock has a permanent effect. This case corresponds with a *unit root* in the dynamics of y_t. Note that the *random walk model with drift* is a special case of a unit root model where $\beta_2 = 1$, $\beta_1 = 0$ and $\beta_0 = \delta$.

5.1 Testing for unit roots

Consider the equation

$$\Delta y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + \varepsilon_t. \tag{11}$$

This is just equation (10) with y_{t-1} subtracted from both sides and with $\alpha \equiv \beta_2 - 1$. This equation nests all three of the models (8), (9) and (10). When $\alpha = 0$ (and $\beta_1 = 0$) we have the random walk model with drift. When $\alpha < 0$, the model corresponds to the stationary trend model with stationary $\beta_2 < 1$ and in the special case when $\alpha = -1$ this is the deterministic trend model.

A test of the hypothesis

$$H_0 : \alpha = 0$$

against the alternative

$$H_1 : \alpha < 0$$

is a test for a *unit root* in the equation (11). It is a test of the random walk model with drift against the deterministic trend models (10) and (8). A test can be based on the *t-statistic*

$$\frac{\widehat{\alpha}}{\sqrt{\widehat{\text{var}}(\widehat{\alpha})}}$$

from *OLS* regression of equation (11). However, it is important to stress that this statistic can not be tested using the standard Student t-distribution. In fact, on the null hypothesis of a unit root, this statistic will be distributed with a distribution known as the *Dickey-Fuller distribution* first tabulated by Fuller (1976).

The *Dickey-Fuller unit root test* provides a way of choosing between deterministic and stochastic trend models. In empirical work, it has been found that the null hypothesis of a unit root is rarely rejected for economic variables. This suggests that stochastic trend models may be more appropriate than deterministic trend models for forecasting economic variables and this in turn implies that it may be difficult to forecast trends over a long horizon, when the best multi-step forecast will be a constant change or constant growth rate forecast.

References

[1] Fuller, W.A. (1976), *Introduction to Statistical Time Series*, Wiley, New York.

[2] Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.