Richard Harris
and Robert Sollis

APPLIED TIME SERIES MODELLING AND FORECASTING

Applied Time Series Modelling and Forecasting

Richard Harris and Robert Sollis

Durham University

WILEY

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

A catalogue record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-84443-4

Project management by Originator, Gt Yarmouth, Norfolk (typeset in 10/12pt Times)
Printed and bound in Great Britain by TJ International Ltd, Padstow, Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Preface

This book is intended for students and others working in the field of economics who want a relatively non-technical introduction to applied time series econometrics and forecasting involving non-stationary data. The emphasis is on the why and how, and as much as possible we confine technical material to boxes or point to the relevant sources that provide more details. It is based on an earlier book by one of the present authors entitled *Using Cointegration Analysis in Econometric Modelling* (see Harris, 1995), but as well as updating the material covered in the earlier book, there are two major additions involving panel tests for unit roots and cointegration, and the modelling and forecasting of financial time series.

We have tried to incorporate into this book as many of the latest techniques in the area as possible and to provide as many examples as necessary to illustrate them. To help the reader, one of the major data sets used is supplied in the Statistical Appendix, which also includes many of the key tables of critical values used for various tests involving unit roots and cointegration. There is also a website for the book (http://www.wiley.co.uk/harris) from which can be retrieved various other data sets we have used, as well as econometric code for implementing some of the more recent procedures covered in the book.

We have no doubt made some mistakes in interpreting the literature, and we would like to thank in advance those readers who might wish to point them out to us. We would also like to acknowledge the help we have received from those who have supplied us with their econometric programming code, data, and guidance on the procedures they have published in articles and books. Particular thanks are due to Peter Pedroni (for his generous offer of time in amending and providing software programmes for Chapter 7), and Robert Shiller for allowing us to use his Standard & Poor's (S&P) Composite data in Chapter 8. We would also like to thank Jean-Phillipe Peters for help with the *G@RCH 2.3* programme, also used in Chapter 8. Others who generously provided software include Jorg Breitung, David Harvey, Robert Kunst and

Johan Lyhagen. Of course, nobody but ourselves take responsibility for the contents of this book.

We also thank Steve Hardman at Wiley, for his willingness to support this project and his patience with seeing it to fruition. Finally, permission from the various authors and copyright holders to reproduce the Statistical Tables is gratefully acknowledged.

Contents

1 Introduction and Overview

Since the mid-1980's applied economists attempting to estimate time series econometric models have been aware of certain difficulties that arise when unit roots are present in the data. To ignore this fact and to proceed to estimate a regression model containing non-stationary variables at best ignores important information about the underlying (statistical and economic) processes generating the data, and at worst leads to nonsensical (or spurious) results. For this reason, it is incumbent on the applied researcher to test for the presence of unit roots and if they are present (and the evidence suggests that they generally are) to use appropriate modelling procedures. De-trending is not appropriate (Chapter 2) and simply differencing the data[1] to remove the non-stationary (stochastic) trend is only part of the answer. While the use of differenced variables will avoid the spurious regression problem, it will also remove any long-run information. In modelling time series data we need to retain this long-run information, but to ensure that it reflects the co-movement of variables due to the underlying equilibrating tendencies of economic forces, rather than those due to common, but unrelated, time trends in the data.

Modelling the long run when the variables are non-stationary is an expanding area of econometrics (both theoretical and applied). It is still fairly new in that while it is possible to find antecedents in the literature dating back to, for example, the seminal work of Sargan (1964) on early forms of the error-correction model, it was really only in 1986 (following the March special issue of the *Oxford Bulletin of Economics and Statistics*) that cointegration became a familiar term in the literature.[2] It is also a continually expanding area, as witnessed by the number of articles that have been published since the mid-1980s. There have been and continue to be major new developments.

[1] That is, converting x_t to Δx_t, where $\Delta x_t = x_t - x_{t-1}$, will remove the non-stationary trend from the variable (and if it does not, because the trend is increasing over time, then x_t will need to be differenced twice, etc.).

[2] Work on testing for unit roots developed a little earlier (e.g., the PhD work of Dickey, 1976 and Fuller, 1976).

The purpose of this book is to present to the reader those techniques that have generally gained most acceptance (including the latest developments surrounding such techniques) and to present them in as non-technical a way as possible while still retaining an understanding of what they are designed to do. Those who want a more rigorous treatment to supplement the current text are referred to Banerjee, Dolado, Galbraith and Hendry (1993) and Johansen (1995a) in the first instance and then of course to the appropriate journals. It is useful to begin by covering some introductory concepts, leaving a full treatment of the standard econometric techniques relating to time series data to other texts (see, for example, Hendry, 1995). This is followed by an overview of the remainder of the book, providing a route map through the topics covered starting with a simple discussion of long-run and short-run models (Chapter 2) and then proceeding through to estimating these models using multivariate techniques (Chapters 5 and 6). We then cover panel data tests for unit roots and cointegration (Chapter 7) before concluding with an in-depth look at modelling and forecasting financial time series (Chapter 8).

SOME INITIAL CONCEPTS

This section will review some of the most important concepts and ideas in time series modelling, providing a reference point for later on in the book. A fuller treatment is available in a standard text such as Harvey (1990). We begin with the idea of a data-generating process (hereafter d.g.p.), in terms of autoregressive and moving-average representations of dynamic processes. This will also necessitate some discussion of the properties of the error term in a regression model and statistical inferences based on the assumption that such residuals are 'white noise'.

Data-generating Processes

As economists, we only have limited knowledge about the economic processes that determine the observed data. Thus, while models involving such data are formulated by economic theory and then tested using econometric techniques, it has to be recognized that theory in itself is not enough. For instance, theory may provide little evidence about the processes of adjustment, which variables are exogenous and indeed which are irrelevant or constant for the particular model under investigation (Hendry, Pagan and Sargan, 1984). A contrasting approach is based on statistical theory, which involves trying to characterize the statistical processes whereby the data were generated.

We begin with a very simple stationary univariate model observed over the sequence of time $t = 1, \dots, T$:

$$\left.\begin{aligned} y_t &= \rho y_{t-1} + u_t \qquad |\rho| < 1 \\ \text{or} \quad (1 - \rho L)y_t &= u_t \end{aligned}\right\} \tag{1.1}$$

where L is the lag operator such that $Ly_t = y_{t-1}$. This statistical model states that the variable y_t is generated by its own past together with a disturbance (or residual) term u_t. The latter represents the influence of all other variables excluded from the model, which are presumed to be random (or unpredictable) such that u_t has the following statistical properties: its expected value (or mean) is zero $[E(u_t) = 0]$ fluctuations around this mean value are not growing or declining over time (i.e., it has constant variance denoted $E(u_t^2) = \sigma^2$); and it is uncorrelated with its own past $[E(u_t u_{t-i}) = 0]$. Having u_t in (1.1) allows y_t to also be treated as a random (stochastic) variable.[3]

This model can be described as a d.g.p., if the observed realization of y_t over time is simply one of an infinite number of possible outcomes, each dependent on drawing a sequence of random numbers u_t from an appropriate (e.g., standard normal) distribution.[4] Despite the fact that in practice only a single sequence of y_t is observed, in theory any number of realizations is possible over the same time period. Statistical inferences with respect to this model are now possible based on its underlying probability distribution.

The model given by equation (1.1) is described as a first-order autoregressive (AR) model or more simply an AR(1) model. It is straightforward to derive the statistical properties of a series generated by this model. First, note that (1.1) can be rearranged as:

$$y_t = [1/(1 - \rho L)]u_t \tag{1.2}$$

It can be shown that $1/(1 - \rho L) = (1 + \rho L + \rho^2 L^2 + \rho^3 L^3 \ldots)$, and therefore the AR(1) model (1.1) can be converted to an infinite order moving average of the lagged disturbance terms:[5]

$$y_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots \tag{1.3}$$

Taking expectations gives $E(y_t) = 0$ (since $E(u_t) = 0$ for all t), thus the mean of y_t, when the d.g.p. is (1.1), is zero. The formula for the variance of y_t is $\text{var}(y_t) = E[\{y_t - E(y_t)\}]^2$. Since in this case the mean of y_t is zero, the formula for the variance simplifies to $E(y_t^2)$. Using this gives:

$$\begin{aligned} E(y_t^2) &= E(\rho y_{t-1} + u_t)^2 \\ &= E(\rho^2 y_{t-1}^2) + E(u_t^2) + 2\rho E(y_{t-1} u_t) \\ &= \rho^2 E(y_{t-1}^2) + \sigma^2 \end{aligned} \tag{1.4}$$

[3] In contrast, y_t would be a deterministic (or fixed) process if it were characterized as $y_t = \rho y_{t-1}$, which, given an initial starting value of y_0, results in y_t being known with complete certainty each time period. Note also that deterministic variables (such as an intercept of time trend) can also be introduced into (1.1).

[4] The standard normal distribution is of course appropriate in the sense that it has a zero mean and constant variance and each observation in uncorrelated with any other.

[5] This property is known as *invertibility*.

Repeatedly substituting for $E(y_{t-1}^2)$ on the right-hand side of (1.4) leads to a geometric series that converges to $E(y_t^2) = \sigma^2/(1-\rho^2)$.

The autocovariance of a time series is a measure of dependence between observations. It is straightforward to derive the autocovariance for an AR(1) process. Generally, the autocovariance is $\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)]$ for $k \neq 0$, where μ represents the mean of y_t. When y_t is generated by (1.1), since $E(y_t) = 0$, the autocovariance formula simplifies to $E(y_t y_{t-k})$. Using this formula, it can be shown that the kth autocovariance is given by:

$$\gamma_k = \rho^k \gamma_0 \qquad k = 1, 2 \ldots \tag{1.5}$$

The autocorrelation coefficient for a time series is a standardized measure of the autocovariance restricted to lie between -1 and 1. The kth autocorrelation is given by:

$$\frac{E[(y_t - \mu)(y_{t-k} - \mu)]}{E[(y_t - \mu)^2]} = \frac{\gamma_k}{\gamma_0} \tag{1.6}$$

Thus the kth autocorrelation when y_t is generated by (1.1) is given by ρ^k. Note that the autocovariances and autocorrelation coefficients discussed above are population parameters. In practice, the sample equivalents of these amounts are employed. In particular they are used when specifying time series models for a particular data set and evaluating how appropriate those models are, as in the Box–Jenkins procedure for time series analysis (Box and Jenkins, 1970). These authors were the first to develop a structured approach to time series modelling and forecasting. The Box–Jenkins approach recognizes the importance of using information on the autocovariances and autocorrelations of the series to help identify the correct time series model to estimate, and when evaluating the fitted disturbances from this model.

Another simple model that is popular in time series econometrics is the AR(1) model with a constant:

$$y_t = \delta + \rho y_{t-1} + u_t \qquad |\rho| < 1 \tag{1.7}$$

Adding a constant to (1.1) allows y_t to have a non-zero mean. Specifically, the mean of y_t when (1.7) is the d.g.p. is given by $E(y_t) = \delta/(1-\rho)$. To see this note that (1.7) can be written as:

$$(1 - \rho L)y_t = \delta + u_t \tag{1.8}$$

so that

$$\begin{aligned} y_t &= [1/(1-\rho L)](\delta + u_t) \\ &= (1 + \rho + \rho^2 + \cdots)\delta + (u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots) \end{aligned} \tag{1.9}$$

Since we are assuming that $E(u_t) = 0$, the expected value of (1.9) simplifies to:

$$E(y_t) = (1 + \rho + \rho^2 + \cdots)\delta \tag{1.10}$$

which is a geometric series that converges to $E(y_t) = \delta/(1 - \rho)$. To calculate the variance of y_t when the d.g.p. is (1.7), it is easiest to work with the de-meaned series $x_t = y_t - \mu$. We can then rewrite (1.7) as:

$$x_t = \rho x_{t-1} + u_t \tag{1.11}$$

It follows that $\text{var}(y_t) = E(x_t^2)$, and that $E(x_t^2) = \sigma^2/(1 - \rho^2)$. Therefore y_t generated by the AR(1) model with a constant has a mean of $E(y_t) = \delta/(1 - \rho)$ and a variance of $\text{var}(y_t) = \sigma^2/(1 - \rho^2)$.

The simple time series model (1.1) can be extended to let y_t depend on past values up to a lag length of p:

$$\left.\begin{aligned} & y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + u_t \\ \text{or} \quad & \text{A(L)} y_t = u_t \end{aligned}\right\} \tag{1.12}$$

where A(L) is the polynomial lag operator $1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p$. The d.g.p. in (1.12) is described as a pth-order AR model.[6] The mean, variance and covariance of AR(p) processes when $p > 1$ can also be computed algebraically. For example, for the AR(2) model with a constant:

$$y_t = \delta + \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t \tag{1.13}$$

assuming $\rho_1 + \rho_2 < 1$ and that u_t is defined as before, the mean of y_t is $E(y_t) = \delta/(1 - \rho_1 - \rho_2)$ and the variance of y_t is:[7]

$$\text{var}(y_t) = \frac{(1 - \rho_2)\sigma^2}{(1 + \rho_2)(1 - \rho_1 - \rho_2)(1 + \rho_1 - \rho_2)} \tag{1.14}$$

An alternative to the AR model is to specify the dependence of y_t on its own past as a moving average (MA) process, such as the following first-order MA model:

$$y_t = u_t + \theta u_{t-1} \qquad |\theta| < 1 \tag{1.15}$$

or a model with past values up to a lag length of q:

$$\left.\begin{aligned} & y_t = u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} \\ \text{or} \quad & y_t = B(L) u_t \end{aligned}\right\} \tag{1.16}$$

where $B(L)$ is the polynomial lag operator $1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$. In practice, lower order MA models have been found to be more useful in econometrics than higher order MA models, and it is straightforward to derive the statistical properties of such models. For example, for the first-order MA model (the MA(1) model) given by (1.15), the mean of y_t is simply $E(y_t) = 0$, while the variance of y_t is $\text{var}(y_t) = (1 + \theta^2)\sigma^2$. It turns out that,

[6] Hence, (1.1) was a first-order AR process.

[7] The importance of the assumption $\rho_1 + \rho_2 < 1$ will become clear in the next chapter.

for the MA(1) model, the first autocovariance is $\gamma_1 = \theta\sigma^2$, but that higher autocovariances are all equal to zero. Similarly, the first autocorrelation coefficient is $\rho_1 = \theta/(1+\theta^2)$, but higher autocorrelation coefficients are all equal to zero.

Finally, it is possible to specify a mixed autoregressive moving average (ARMA) model:

$$A(L)y_t = B(L)u_t \tag{1.17}$$

which is the most flexible d.g.p. for a univariate series. Consider, for example, the ARMA(1, 1) model:

$$y_t = \rho_1 y_{t-1} + u_t + \theta_1 u_{t-1} \qquad |\rho_1| < 1,\ |\theta_1| < 1 \tag{1.18}$$

As with the AR(1) model, note that the ARMA(1, 1) model can be rewritten as an infinite order MA process:

$$\begin{aligned} y_t &= (1+\theta_1 L)(1-\rho_1 L)^{-1}u_t \\ &= \sum_{j=0}^{\infty} \omega_j u_{t-j} \end{aligned} \tag{1.19}$$

Since we are assuming that $E(u_t) = 0$, it follows that $E(y_t) = 0$. The variance of y_t is given by:

$$\begin{aligned} E(y_t^2) &= E[(\rho_1 y_{t-1} + u_t + \theta_1 u_{t-1})^2] \\ &= E(\rho_1^2 y_{t-1}^2 + 2\rho_1\theta_1 y_{t-1}u_{t-1} + u_t^2 + \theta_1^2 u_{t-1}^2) \end{aligned} \tag{1.20}$$

Using the autocovariance notation, the variance of y_t can be written:

$$\gamma_0 = \rho_1^2\gamma_0 + 2\rho_1\theta_1\sigma^2 + \sigma^2 + \theta_1^2\sigma^2 \tag{1.21}$$

which can be rearranged as:

$$\gamma_0 = \left(\frac{1+\theta_1^2+2\rho_1\theta_1}{1-\rho_1^2}\right)\sigma^2 \tag{1.22}$$

The higher autocovariances can be obtained in a similar way, and it can be shown that:

$$\begin{aligned} \gamma_1 &= \rho_1\gamma_0 + \theta_1\sigma^2 \\ &= \left(\frac{(1+\rho_1\theta_1)(\rho_1+\theta_1)}{(1-\rho_1^2)}\right)\sigma^2 \end{aligned} \tag{1.23}$$

$$\gamma_2 = \rho_1\gamma_1 \tag{1.24}$$

and $\gamma_k = \rho_1\gamma_{k-1}$ for $k \geq 2$. The autocorrelation coefficients are given by:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1+\rho_1\theta_1)(\rho_1+\theta_1)}{(1+\theta_1^2+2\rho_1\theta_1)} \tag{1.25}$$

and $\rho_k = \rho_1\rho_{k-1}$ for $k \geq 2$.

So far the d.g.p. underlying the univariate time series y_t contains no economic information. That is, while it is valid to model y_t as a statistical process (cf. the Box–Jenkins approach), this is of little use if we are looking to establish (causal) linkages between variables. Thus, (1.1) can be generalized to include other variables (both stochastic, such as x_t, and deterministic, such as an intercept), for example:

$$y_t = \alpha_0 + \gamma_0 x_t + \alpha_1 y_{t-1} + u_t \tag{1.26}$$

Since x_t is stochastic, let its underlying d.g.p. be given by:

$$x_t = \xi x_{t-1} + \varepsilon_t \qquad |\xi| < 1 \quad \text{and} \quad \varepsilon_t \sim \text{IN}(0, \sigma_\varepsilon^2) \qquad (1.27)^8$$

If u_t and ε_t are not correlated, we can state that $E(u_t \varepsilon_s) = 0$ for all t and s, and then it is possible to treat x_t *as if* it were fixed for the purposes of estimating (1.26). That is, x_t is independent of u_t (denoted $E(x_t u_t) = 0$) and we can treat it as (strongly) exogenous in terms of (1.26) with x_t being said to Granger-cause y_t. Equation (1.26) is called a conditional model in that y_t is conditional on x_t (with x_t determined by the marginal model given in (1.27)). Therefore, for strong exogeneity to exist x_t must not be Granger-caused by y_t, and this leads on to the concept of weak exogeneity.

Note, if (1.27) is reformulated as:

$$x_t = \xi_1 x_{t-1} + \xi_2 y_{t-1} + \varepsilon_t \tag{1.28}$$

then $E(x_t u_t) = 0$ is retained, but since past values of y_t now determine x_t the latter can only be considered *weakly* exogenous in the conditional model (1.26).[9]

Lastly, weak exogeneity is a necessary condition for super-exogeneity, but the latter also requires that the conditional model is structurally invariant; that is, changes in the distribution of the marginal model for x_t (equation (1.27) or (1.28)) do not affect the parameters in (1.26). In particular, if there are regime shifts in x_t then these must be invariant to (α_i, γ_0) in (1.26).

All three concepts of exogeneity will be tested later, but it is useful at this point to provide a brief example of testing for super-exogeneity in order to make the concept clearer.[10] Assuming that known institutional (e.g., policy)

[8] Note that $\varepsilon_t \sim \text{IN}(0, \sigma_\varepsilon^2)$ states that the residual term is independently and normally distributed with zero mean and constant variance σ_ε^2. The fact that σ_ε^2 is multiplied by a (not shown) value of 1 means that ε_t is not autocorrelated with its own past.

[9] That is, x_t still causes y_t, but not in the Granger sense, because of the lagged values of y_t determining x_t. For a review of these concepts of weak and strong exogeneity, together with their full properties, see Engle, Hendry and Richard (1983).

[10] This example is based on Hendry (1995, p. 537). Further discussion of super-exogeneity can be found in Engle and Hendry (1993), Hendry (1995, p. 172) and Favero (2001, p. 146).

and historical shifts (shocks) can be identified that affected x_t, it should be possible to construct a dummy variable (e.g., POL_t) that augments (1.28):

$$x_t = \xi_1 x_{t-1} + \xi_2 y_{t-1} + \xi_3 \text{POL}_t + \varepsilon_t \tag{1.28*}$$

Assuming that the estimate of $\hat{\xi}_3$ is (highly) significant in determining x_t, then super-exogeneity can be tested by including POL_t in the conditional model (1.26), and if this dummy is significant then super-exogeneity is rejected.[11]

The importance of these three concepts of exogeneity are discussed in Favero (2001, p. 146): (i) if we are primarily interested in inference on the (α_i, γ_0) parameters in (1.26), then if x_t is weakly exogenous we only need to estimate (1.26) and not also (1.28); (ii) if we wish to dynamically simulate y_t and x_t is strongly exogenous, again we only need to estimate (1.26) and not also (1.28); and (iii) if the objective of modelling y_t is for econometric policy evaluation, we only need to estimate the conditional model (1.26) if x_t has the property of being super-exogenous. The latter is a necessary condition to avoid the Lucas Critique (see Lucas, 1976). For example, suppose y_t is a policy variable of government (e.g., the money supply) and x_t is the instrument used to set its outcome (e.g., the interest rate), then x_t must be super-exogenous to avoid the Lucas Critique. Otherwise, setting x_t would change the policy model (the parameters of 1.26), and the policy outcome would not be what the model (1.26) had predicted.[12]

As with the univariate case, the d.g.p. denoted by (1.26) can be generalized to obtain what is known as an autoregressive distributed lag (ADL) model:

$$A(L)y_t = B(L)x_t + u_t \tag{1.29}$$

where the polynomial lag operators $A(L)$ and $B(L)$ have already been defined.[13] Extending to the multivariate case is straightforward, replacing y_t and x_t by vectors of variables, $\mathbf{y}_t$ and $\mathbf{x}_t$.

The great strength of using an equation like (1.29) as the basis for econometric modelling is that it provides a good first approximation to the (unknown) d.g.p. Recall the above arguments that theory usually has little

[11] That is, its exclusion from (1.26) would alter the estimates of (α_i, γ_0). Note also that the residuals $\hat{\varepsilon}_t$ from (1.28*) should not be a significant determinant of y_t in equation (1.26).

[12] For example, suppose the government uses the immediate history of y_t to determine what it wishes current y_t to be; hence, it alters x_t to achieve this policy outcome. However, economic agents also 'know' the model (the policy rule) underlying (1.26) and (1.28*). Thus when POL_t changes, agents alter their behaviour (the parameters of 1.26 change) since they have anticipated the intended impact of government policy. Econometric models that fail to separate out the expectations formulation by economic agents from the behavioural relationships in the model itself will be subject to Lucas's critique.

[13] While we could further extend this to allow for an MA error process, it can be shown that a relatively simple form of the MA error process can be approximated by sufficiently large values of p and q in (1.29).

to say about the form of the (dynamic) adjustment process (which (1.29) is flexible enough to capture), nor about which variables are exogenous (this model can also be used as a basis for testing for exogeneity). In fact, Hendry et al. (1984) argue that the process of econometric modelling is an attempt to match the unknown d.g.p. with a validly specified econometric model, and thus '... economic theory restrictions on the analysis are essential; and while the data are the result of economic behaviour, the actual statistical properties of the observables corresponding to y and z are also obviously relevant to correctly analysing their empirical relationship. In a nutshell, measurement without theory is as valueless as the converse is non-operational.' In practical terms, and according to the Hendry-type approach, the test of model adequacy is whether the model is *congruent* with the data evidence, which in a single equation model is defined in terms of the statistical properties of the model (e.g., a 'white noise' error term and parameters that are constant over time) and whether the model is consistent with the theory from which it is derived and with the data it admits. Finally, congruency requires the model to encompass rival models.[14]

Role of the Error Term u_t and Statistical Inference

As stated above, the error term u_t represents the influence of all other variables excluded from the model that are presumed to be random (or unpredictable) such that u_t has the following statistical properties: its mean is zero $[E(u_t) = 0]$; it has constant variance $[E(u_t^2) = \sigma^2]$; and it is uncorrelated with its own past $[E(u_t u_{t-i}) = 0]$. To this we can add that the determining variable(s) in the model, assuming they are stochastic, must be independent of the error term $[E(x_t u_t) = 0]$.[15] If these assumptions hold, then it is shown in standard texts like Johnston (1984) that estimators like the ordinary least squares (OLS) estimator will lead to unbiased estimates of the parameter coefficients of the model (indeed, OLS is the best linear unbiased estimator). If it is further assumed that u_t is drawn from the (multivariate) normal distribution, then this sufficies to establish inference procedures for testing hypotheses involving the parameters of the model, based on χ^2, t- and F-tests and their associated probability distributions.

Thus, testing to ensure that $u_t \sim \text{IN}(0, \sigma_u^2)$ (i.e., an independently distributed random 'white noise' process drawn from the normal distribution) is an essential part of the modelling process. Its failure leads to invalid inference

[14] A good discussion of congruency and modelling procedures is given in Doornik and Hendry (2001).

[15] Although not considered above, clearly this condition is not met in (1.1) and similar dynamic models, where y_{t-1} is a predetermined explanatory variable, since $E(y_t u_{t-i}) \neq 0$ for $i \geq 1$. However, it is possible to show by applying the Mann–Wald theorem (Johnston, 1984, p. 362) that with a sufficiently large sample size this will not lead to bias when estimating the parameter coefficients of the regression model.