ECONOMETRIC METHODS

Roselyne Joyeux and **George Milunovich**
Department of Economics, Macquarie University, Australia

Keywords: Least Squares, Maximum Likelihood, Generalized Method of Moments, time series, panel, limited dependent variables

Contents

1. Introduction
2. Least Squares Estimation
3. Maximum Likelihood
3.1. Estimation
3.2. Statistical Inference Using the Maximum Likelihood Approach
4. Generalized Method of Moments
4.1. Method of Moments
4.2 Generalized Method of Moments (GMM)
5. Other Estimation Techniques
6. Time Series Models
6.1 Time Series Models: a Classification
6.2 Univariate Time Series Models
6.3 Multivariate Time Series Models
6.4. Modelling Time-Varying Volatility
6.4.1. Univariate GARCH Models
6.4.2. Multivariate GARCH models
7. Panel Data Models
7.1. Pooled Least Squares Estimation
7.2. Estimation after Differencing
7.3.Fixed Effects Estimation
7.4. Random Effects Estimation
7.5. Non-stationary Panels
8. Discrete and Limited Dependent Variables
8.1. The Linear Probability Model (LPM)
8.2. The Logit and Probit Models
8.3. Modelling Count Data: The Poisson Regression Model
8.4. Modelling Censored Data: Tobit Model
9. Conclusion
Glossary
Bibliography
Biographical Sketches

Summary

The development of econometric methods has proceeded at an unprecedented rate over the last forty years, spurred along by advances in computing, econometric theory and the availability of richer data sets. The aim of this chapter is to provide a survey of econometric methods. We present an overview of those econometric methods and

models that we believe to be most useful to an applied economist. Further, we distinguish between econometric methods, which are statistical estimation techniques and econometric models to which estimation methods are applied.

1. Introduction

Econometric analysis is used to develop, estimate and evaluate models which relate economic or financial variables. An applied economic study usually proceeds in the following way:

- Statement of theory or hypothesis. This step requires economic expertise.
- Specification of the econometric model to test the theory: linear or non-linear, univariate or multivariate, single or multiple equations;
- Estimation of the parameters of the chosen model: parametric or non-parametric, Classical or Bayesian estimation.
- Evaluation: diagnostic tests, ex-post forecasting, simulations.
- The model is the ready for control, forecasting or policy purposes.

Econometric methods guide the applied economist through those steps. The development of econometric methods has proceeded at an unprecedented rate over the last forty years, spurred along by advances in computing, econometric theory and the availability of richer data sets. The aim of this chapter is to provide a survey of econometric methods, although we acknowledge at the outset that it is impossible to overview all available econometric methods in one chapter. Volumes have been written on the subject. We have therefore elected to present an overview of those econometric methods and models that we believe to be most useful to an applied economist. Further, we distinguish between econometric methods, which are statistical estimation techniques and econometric models to which estimation methods are applied.

We start this chapter by presenting in Section 2 an overview of the linear regression model, the least squares estimation technique, properties of the least squares estimators and problems associated with the least squares approach. Section 3 introduces the Maximum Likelihood Estimator, which is still one of the most commonly used estimation methods. The Generalized Method of Moments approach is introduced in section 4. The second part of the chapter focuses on econometric models and applications of these three estimation methods. Section 5 is devoted to time series models. We study both univariate and multivariate models. Simultaneous equation models and time-varying variance (GARCH) models are also considered in this section. Panel data models are the subject of Section 6. Panel Data is a very active area of econometrics with longer longitudinal surveys becoming available in microeconometrics. In this section, we also consider panel data models with non-stationary time series. Section 7 looks at limited dependent variables models. Section 9 concludes.

This overview is necessarily brief and selective. We give further references for the interested reader in the text.

2. Least Squares Estimation

Economic theory usually suggests some relationship between a random variable *y* in terms of some other explanatory random variables $x_1,\ldots,x_k$. Although the joint probability function $f\left(y, x_1,...,x_n \mid \boldsymbol{\theta}\right)$ fully characterises this set of variables, we are often interested in one particular factorisation of the joint density given by:

$$f\left(y, x_1,...,x_n \mid \boldsymbol{\theta}\right) = f\left(x_1,...,x_n \mid \boldsymbol{\alpha}\right) \cdot g\left(y \mid x_1,...,x_n, \boldsymbol{\beta}\right).$$

where the parameter vector $\boldsymbol{\theta}$ is partitioned into $\left[\boldsymbol{\alpha}, \boldsymbol{\beta}\right]$ and $\boldsymbol{\alpha}$ is a vector of parameters associated with the conditional distribution of *y* while $\boldsymbol{\beta}$ collects parameters of the joint density of the explanatory variables $x_1,...,x_k$. Provided that the process generating the explanatory variables takes place outside the conditional density of *y*, i.e. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ do not share common elements and no restrictions link them, the conditional probability density $f\left(y \mid x_1,...,x_n, \boldsymbol{\beta}\right)$ can be analysed in isolation from the marginal density of the explanatory variables $g\left(x_1,...,x_n \mid \boldsymbol{\alpha}\right)$. This is important because in many economic applications it is the conditional distribution of the dependent variable that is of primary interest, and in particular the conditional mean $E\left(y \mid x_1,...,x_n\right)$ and more recently conditional variance $Var\left(y \mid x_1,...x_n\right)$ equations. In most cases the conditional mean equation is specified as a linear function in (unknown) parameters $\boldsymbol{\beta}$ although various non-linear specifications are becoming common.

For example the dependent variable could be an individual's weekly earnings, and the explanatory variables could be education, work experience, age, etc… The economic theory is often assumed to be well approximated by a linear model written as:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon \qquad (1)$$

where ε is a random error term assumed to pick up all other factors not included in the *x*'s. It is typically different from 0 because:

- there are variables left out;
- unpredictable nature of human behaviour;
- errors of measurement.

The Least Squares (LS) estimation technique chooses the parameters which minimize the sum of the squared error terms $\sum_{t=1}^{T} \varepsilon_t^2$ (assuming we have *T* observations). It does not require any additional assumption beside linearity of the model. It is easy to show that the solution to this minimization problem gives the following *k+1* equations:

$$\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - ... - \hat{\beta}_k x_{ti}\right) = 0 \qquad (2)$$

$$\frac{1}{T}\sum_{t=1}^{T} x_{1t}\left(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - ... - \hat{\beta}_k x_{kt}\right) = 0 \qquad (3)$$

….

$$\frac{1}{T}\sum_{t=1}^{T} x_{kt}\left(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - ... - \hat{\beta}_k x_{kt}\right) = 0 \qquad (4)$$

To show that the least squares estimators have desirable statistical properties it is however necessary to make the following assumptions.

Assumption 1: the model is linear in the parameters and is correctly specified.
Assumption 2: $E(\varepsilon \mid x_1,...,x_k) = E(\varepsilon) = 0$

This means that ε is uncorrelated with the explanatory variables.

Assumption 3: the variance of ε is constant and the ε*'s* are uncorrelated across observations.
Assumption 4: there are no exact linear relationships among the explanatory variables.
The model can be broken into two parts:

- The part of *y* which is explained by the model:

$$E(y \mid x_1,...,x_k) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

- The part of *y* which is left unexplained:

$$\varepsilon = y - E(y \mid x_1,...,x_k) = y - (\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)$$

Under assumptions 1 through 4, it can be shown that the least squares estimators have minimum variance among all linear unbiased estimators. The least squares estimators are said to be best linear unbiased estimators (BLUE).

The sampling distribution of an estimator as *T* tends to infinity is its asymptotic distribution. An estimator is said to be asymptotically unbiased if the mean of its asymptotic distribution is equal to the true value of the parameter. An estimator is consistent if the probability that the estimator takes on a value close to the true value goes to 1 as the number of observations *T* goes to infinity.

Under assumptions 1-4 the least squares estimators are consistent.

If we add the assumption that ε_t is normally distributed then the least squares estimators are also normally distributed. Statistical inference on the parameters can then be conducted.

These assumptions might fail for many reasons:

- Assumption 1: there might be omitted relevant or included irrelevant variables or the model might be non-linear;
- Assumption 2 : the explanatory variables might be endogenous; there might be omitted relevant variables which are correlated with the included ones;
- Assumption 3: heteroskedasticity or autocorrelation might be present.
- Assumption 4: multicollinearity might be a problem.

If assumption 3, $E\left(\varepsilon_t^2\right)=\sigma^2$ **,** is violated the error terms are said to be heteroskedastic. If heteroskedasticity is present, the least squares estimators are unbiased and linear but are not minimum variance among all unbiased estimators. Better estimators can be found by taking into account the heteroskedasticity. The usual least squares standard errors are not consistent estimates of the true variances. White's (1980) heteroskedasticity-consistent covariance matrix estimator has to be used in this case.

The second part of assumption 3 is the assumption of no autocorrelation: $cov(\varepsilon_t,\varepsilon_s)=0, t\neq s$. In the presence of autocorrelation the least squares estimators are still linear unbiased estimators, but they are not minimum variance. As was the case for heteroskedasticity the usual least squares standard errors are incorrect in the presence of autocorrelation. Newey and West (1987b) propose a heteroskedasticity and autocorrelation-consistent covariance matrix estimator which generalises White's estimator.

In the following two sections of this chapter we consider two other estimation methods:

- Maximum likelihood;
- Generalized method of moments.

These two estimation techniques have the advantage over the least squares technique that heteroskedasticity and autocorrelation can be dealt with more directly in a more unified framework. They can also be applied when the model is non-linear.

We shall see in what follows that the three estimation methods considered in this chapter all give the same solution for linear models with normally distributed error terms.

3. Maximum Likelihood

3.1. Estimation

Maximum Likelihood Estimation (MLE) approach involves forming an assumption about the underlying probability distribution function (*pdf*) that generates the observed data set, and then estimating parameters of the assumed distribution. Although there are many cases, especially in financial applications, where it may seem inappropriate to assume knowledge of the underlying *pdf*, White (1982) has been shown that for correctly specified moment equations the maximum likelihood estimator, now interpreted as Quasi Maximum Likelihood Estimator (QMLE), is consistent. In other words, as long as conditional moments (e.g. mean and/or variance equations) are

correctly specified QMLE will produce estimates that converge to their true parameter values as the sample size increases, although less efficiently than if the correct likelihood function had been used.

To illustrate the maximum likelihood approach we consider the previously specified linear model now written in the vector notation:

$$y_t = \mathbf{x}_t^{'}\boldsymbol{\beta} + \varepsilon_t \tag{5}$$

where $\boldsymbol{\beta}$ is an $\left[(k+1)\times 1\right]$ vector of unknown parameters, $\mathbf{x}_t$ is an $\left[(k+1)\times 1\right]$ vector of explanatory variables and ε_t is a white noise process with zero mean and variance σ^2. The parameters of interest can then be grouped into a $((k+2)\times 1)$ vector $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix}$.

The maximum likelihood estimation approach typically involves two steps:

- Specification of a probability distribution for ε_t.
- Computation and maximisation of the likelihood function.

The joint *pdf* of the observed sample takes the following form:

$$\begin{aligned} f\left(y_1,.....,y_T \mid \mathbf{x},\boldsymbol{\beta},\sigma^2\right) &= f\left(y_1 \mid \mathbf{x},\boldsymbol{\beta},\sigma^2\right)\cdots f\left(y_T \mid \mathbf{x},\boldsymbol{\beta},\sigma^2\right) \\ &= \prod_{t=1}^{T} f\left(y_t \mid \mathbf{x},\boldsymbol{\beta},\sigma^2\right). \end{aligned} \tag{6}$$

In practice, ε_t is generally assumed to be Gaussian white noise (Hamilton, 1994, p. 117), although other *pdf*s (e.g. *t*-distribution, general error distribution, etc…) are seen in the literature. Assuming normality

$$\varepsilon_t \sim i.i.d.N\left(0,\sigma^2\right) \tag{7}$$

and after a change of variable the conditional likelihood function of y_t can be written as

$$f\left(y_1,.....,y_T \mid \mathbf{x},\boldsymbol{\beta},\sigma^2\right) = \prod_{t=1}^{T}\left(2\pi\sigma^2\right)^{-1/2}\exp\left(-\frac{\left(y_t - \mathbf{x}_t^{'}\boldsymbol{\beta}\right)^2}{2\sigma^2}\right). \tag{8}$$

The maximum likelihood estimate of $\boldsymbol{\theta}$ is found by maximising the above function, which is often interpreted as the probability of observing the realised data sample. In practice we often take logs of the likelihood function in order to simplify algebraic manipulations:

$$L(\boldsymbol{\theta}) = -\frac{T}{2}\ln\left(2\pi\sigma^2\right) - \sum_{t=1}^{T}\left(\frac{\left(y_t - \mathbf{x}_t'\boldsymbol{\beta}\right)^2}{2\sigma^2}\right) \tag{9}$$

where $L(\boldsymbol{\theta})$ is now known as the log-likelihood function.

Maximization of the log-likelihood function involves differentiating $L(\boldsymbol{\theta})$ with respect to the parameters of interest $(\boldsymbol{\theta})$ and setting each of the resulting equations to zero. Although, in some instances, it may be possible to find a closed form solution to the resultant system of equations, e.g. in the case of a linear regression model, in more complicated situations there are no closed form solutions and $L(\boldsymbol{\theta})$ must be maximized numerically. In either case, the outcome of maximizing equation (9) results in a maximum likelihood estimate vector $\hat{\boldsymbol{\theta}}$ based on a Gaussian likelihood function.

In time series analysis, a distinction is made between conditional and exact likelihood functions. The difference is based on the treatment of the first *p* observations, where *p* is the number of dependent variable lags specified in the model (e.g. in an autoregressive model of order two, AR(2), *p = 2)*. The conditional likelihood function assumes that the first *p* lags are fixed, that is the likelihood function of the remaining sample is specified conditional on the first *p* observations, which are in turn set to either their realized or expected values. The exact likelihood function (see Hamilton, 1994, Ch 5 for details), on the other hand, is a product of probability densities of all observations, including the first *p* lags. Since it is impossible to model the first *p* observations using a conditional model (i.e. there is no data to condition the first *p* observations on in an AR(*p*) model) an unconditional *pdf* is specified for the initial *p* observations. When the sample size *T* is large, the difference between the two approaches is small.

3.2. Statistical Inference Using the Maximum Likelihood Approach

Results presented here assume that data is strictly stationary and that parameters of interest $(\boldsymbol{\theta})$ do not fall on the boundary of the allowable parameter space. Given these conditions, and in large samples, the MLE estimate vector $\hat{\boldsymbol{\theta}}$ is approximately normally distributed with:

$$\hat{\boldsymbol{\theta}} \approx N\left(\boldsymbol{\theta}, \Im^{-1}\right) \tag{10}$$

where the covariance matrix is given by the inverse of the information matrix. Generally, there are two alternative approaches to calculating the information matrix. The first approach evaluates the second derivative of the likelihood function at the estimated parameter vector $\hat{\boldsymbol{\theta}}$:

$$\Im_{2D} = -\left.\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{11}$$

Alternatively the information matrix can be computed as the outer product of the gradient vectors. If the log-likelihood function takes the form as in equation (9) then the log of each probability density is given by:

$$f\left(y_t \mid \mathbf{x}, \boldsymbol{\beta}, \sigma^2\right) = -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{\left(y_t - \mathbf{x}_t'\boldsymbol{\beta}\right)^2}{2\sigma^2}. \tag{12}$$

Differentiating equation (12) with respect to $\boldsymbol{\theta}$ and evaluating at $\hat{\boldsymbol{\theta}}$ gives:

$$\mathbf{h_t}\left(\hat{\boldsymbol{\theta}}\right) = \left.\frac{\partial f\left(y_t \mid \mathbf{x}, \boldsymbol{\beta}, \sigma^2\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{13}$$

where $\mathbf{h_t}\left(\boldsymbol{\theta}\right)$ is a $\left((k+2)\times 1\right)$ vector of derivatives. The information matrix can then be constructed as:

$$\mathfrak{I}_{OP} = \sum_{t=1}^{T}\left[\mathbf{h}_t\left(\hat{\theta}\right)\right]\left[\mathbf{h}_t\left(\hat{\theta}\right)\right]'. \tag{14}$$

QMLE Covariance Matrix

In instances where the underlying density function is mispecified the MLE covariance matrix is biased even though the estimated parameter vector $\hat{\boldsymbol{\theta}}$ is consistent. In such cases, one can consistently compute the QMLE covariance matrix as:

$$\Sigma_{QMLE} \cong \left[\mathfrak{I}_{2D}\mathfrak{I}_{OP}^{-1}\mathfrak{I}_{2D}\right]^{-1} \tag{15}$$

Hypothesis Testing

In general, there are three approaches to testing a set of restrictions on the parameter vector $\hat{\boldsymbol{\theta}}$ that include: the Wald test, the likelihood ratio test and the Lagrange multiplier test. Under the null hypothesis the three tests are asymptotically equivalent, although their small sample properties are unknown. In the following discussion we will denote the unrestricted $\left((k+2)\times 1\right)$ parameter vector as $\hat{\boldsymbol{\theta}}$ and the restricted $\left[\left(k+2-m\right)\times 1\right]$ parameter vector as $\tilde{\boldsymbol{\theta}}$, implying that there are *m* restrictions.

Wald Test

The Wald test involves estimating only the unrestricted model. Because the MLE estimates are consistent the estimated parameters will converge to their true values, so that when the restrictions tested are valid the difference between the estimated parameters and the restrictions will be close to zero. Consider a set of linear restrictions given by $\mathbf{R}\boldsymbol{\theta} = \mathbf{q}$ where $\mathbf{R}$ is of dimension$\left(m\times\left(k+2\right)\right)$, i.e. the number of rows is

given by the number of restrictions, and $\mathbf{q}$ is an $(m \times 1)$ vector of restrictions. The Wald test can be expressed as:

$$Wald_{statistic} = \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{q}\right]' \left[\mathbf{R}\boldsymbol{\Sigma}_{MLE(\hat{\theta})}\mathbf{R}'\right]^{-1} \left[\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{q}\right] \sim \chi^2_{(m)} \qquad (16)$$

where $\boldsymbol{\Sigma}_{MLE(\hat{\theta})}$ is the asymptotic covariance matrix of the estimated parameter vector $\hat{\boldsymbol{\theta}}_{n\times 1}$ as given in equation (10). The null hypothesis is rejected if the $Wald_{statistic} > \chi^2_{(m)}$ critical value.

There are two shortcomings of the Wald test. First, it is a test against the null hypothesis in which there is no specific formulation of the alternative hypothesis. Second, the Wald test is not invariant to the formulation of the restrictions (see Greene, 2003, p.488 for details).

Likelihood Ratio Test

Unlike the Wald test, the likelihood ratio test requires MLE estimates of both the restricted $\tilde{\boldsymbol{\theta}}$ and the unrestricted $\hat{\boldsymbol{\theta}}$ parameter vectors. The restricted vector can be obtained by keying in restrictions directly into the likelihood function and maximizing it with respect to the restricted parameters. The likelihood ratio test takes the following form:

$$LR_{statistic} = 2\left[L\left(\hat{\boldsymbol{\theta}}\right) - L\left(\tilde{\boldsymbol{\theta}}\right)\right] \sim \chi^2_{(m)} \qquad (17)$$

The null hypothesis of *m* restrictions is rejected when $LR_{statistic} > \chi^2_{(m)}$ critical value.

Lagrange Multiplier Test

In order to conduct a Lagrange multiplier test we need only to estimate the restricted model. The rationale behind this test is that if the restrictions are valid, then the slope of the maximised log-likelihood function should be close to zero. Thus, what is required is the gradient of the conditional (unrestricted) *pdf* evaluated at the restricted parameter vector $\tilde{\boldsymbol{\theta}}$:

$$\mathbf{h}_t\left(\tilde{\boldsymbol{\theta}}\right) = \left.\frac{\partial f\left(y_t \mid \mathbf{x}, \boldsymbol{\beta}, \sigma^2\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}. \qquad (18)$$

To make this operational we need to estimate the restricted model and then use the restricted estimates to evaluate the derivative of the likelihood function. The Lagrange Multiplier statistic takes the form:

$$LM_{statistic} = \left[\sum_{t=1}^{T} \mathbf{h}_t\left(\tilde{\boldsymbol{\theta}}\right)\right]' \tilde{\mathfrak{I}}^{-1} \left[\sum_{t=1}^{T} \mathbf{h}_t\left(\tilde{\boldsymbol{\theta}}\right)\right] \sim \chi^2_{(m)} \tag{19}$$

where the information matrix $\tilde{\mathfrak{I}}$ is also evaluated at the restricted values:

$$\tilde{\mathfrak{I}} = -\left.\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}. \tag{20}$$

If the $LM_{statistic} > \chi^2_{(m)}$ the null hypothesis of *m* restrictions is rejected.

-
-
-

Bibliography

Abelson P and. R. Joyeux (eds), *Economic Forecasting*, Allen and Unwin, Sydney, Australia. [chapter 2 gives an over view of econometric forecasting methods]

Andersen, T. G., Bollerslev, T. and Diebold, F.X. (2002), "Parametric and Nonparametric Volatility Measurement", NBER Working Paper No. T0279. [seminal work on realised volatility models]

Andersen, T.G., and Bollerslev, T. (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance* **52**, 975-1005. [realised volatility reference]

Arellano, M. and O. Bover (1995), "Another Look at the Instrumental Variables Estimation of Error-Components Models," *Journal of Econometrics* **58**, 277-297. [an advanced paper on GMM estimation and panel data]

Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, **58**, 277-297. [an advanced paper on GMM estimation and panel data]

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* **74**, 3-30. [seminal work on integrated GARCH models]

Baltagi, B. (2005), *Econometric Analysis of Panel Data*, third edition, John Wiley and Sons, New York. [excellent overview of econometric methods used for panel data]

Banerjee, A., Cockerill, L., Russell, B. (2001), "An I(2) Analysis of Inflation and the Markup," *Journal of Applied Econometrics* **16**, 221--240. [provides a comparison of panel unit root tests]

Black, F. (1976), 'Studies of Stock Price Volatility Changes', Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section, 177-181. [one of the first studies to document asymmetric effects in volatility]

Blanchard, O. (1989), "A Traditional Interpretation of Macroeconomic Fluctuations," *American Economic Review* **79**, 5, 1146-64. [this paper provides an illustration of VAR models]

Bollerslev, T. (1986), 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics* **31**(3), 307-327. [GARCH extension to Engle's ARCH model]

Bollerslev, T. (1990), 'Modelling the Coherence in Short-run Nominal Exchange Rates: A Generalised ARCH Model', *Review of Economic Statistics* **72**, 498-505. [a multivariate GARCH model]

Bollerslev, T., Engle, R. F. and Wooldridge, J. (1988), 'A capital asset pricing model with time varying covariances', *Journal of Political Economy* **96**, 116-131. [a multivariate GARCH model]

Bomhoff, E.J. (1994), *Financial Forecasting for Business and Economics*, The Dryden Press, London. [provides details on the Kalman filter]

Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, San Francisco, Holden-Day. [the first comprehensive overview of ARIMA models]

Breunig, R. (2006), Lecture Notes, http://econrsss.anu.au.edu.au/Staff/breunig/contact_bb.htm [excellent introductory lecture notes for panel data models]

Choi, I. (2006), "Combination unit root tests for cross-sectionally correlated panels," *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honour of Peter C.B. Phillips*, Cambridge University Press. [an advanced paper on panel unit root tests for cross-sectionally correlated panels]

Christie, A. A. (1982), 'The Stochastic Behavior of Common Stock Variances -Value, Leverage and Interest Rate Effects', *Journal of Financial Economics* **10(4)**, 407-432. [one of the first papers to document asymmetric effect in volatility]

Clements, M.P. and D.F. Hendry (1995), "Forecasting in Cointegrated Systems," *Journal of Applied Econometrics* **10**, 127-146. [a good introduction to forecasting in cointegrated systems]

Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge, Cambridge University Press. The Marshall Lectures on Economic Forecasting. excellent intermediate level book on forecasting]

Cuthbertson, K., Hall, S. and M. Taylor (1992), *Applied Econometric Techniques*, Harvester Wheatsheaf, Hemel Hampstead. [excellent intermediate level book on econometric techniques]

Day, R. and W. Shafer (1985), "Keynesian Chaos," *Journal of Macroeconomics* **7**, 277-95. [an application of chaos theory]

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica* **30**, 193-220. [advanced paper on unit root tests]

Dungey, M., Fry, R., Martin, M, and González-Hermosillo, B. (2005), "Empirical Modelling of Contagion: A Review of Methodologies" *Quantitative Finance* **5(1)**, 9-24. [an application of time-varying volatility models to measuring financial contagion]

Engle, R. and C.W.J. Granger (1987), "Cointegration and Error-Correction: Representation, Estimation and Testing", *Econometrica* **55**, 251-76. [one of the first papers on cointegration]

Engle, R. F. and Kroner, K. (1995), 'Multivariate simultaneous GARCH', *Econometric Theory* **11**, 122-150. [a description of BEKK multivariate GARCH model]

Engle, R.F. (1982), 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica* **50**, 987-1007. [the first ARCH paper]

Engle, R.F. (2002), 'Dynamic conditional correlation - A simple class of multivariate GARCH models', *Journal of Business and Economic Statistics* **20**(3), 339-350. [a multivariate GARCH model]

Engle, R.F. and Bollerslev T. (1986), "Modelling the persistence of conditional variances", *Econometric Reviews* **5***:*1–50. [modelling long memory time-varying variance processes]

Engle, R.F., and Colacito, R. (2004), 'Testing and valuing dynamic correlations for asset allocations', unpublished manuscript, New York University, NY. [an application of MGARCH models to portfolio allocation]

Engle, R.F., Hendry, D.F. and J.-F. Richard (1983), "Exogeneity," *Econometrica* **51**, 2, 277-304. [this paper sets out carefully definitions of exogeneity]

Engle, R.F., Lilien, D.M., and Robins, R.P. (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica* **55**, 391-407. [GARCH in mean model]

Ericsson, N. (1998), "Empirical Modelling of Money Demand," International Finance Discussion Paper No. 61*, Board of Governors of the Federal Reserve System, Washington, D.C., April, available on the World Wide Web at www.bog.frb.fed.us/pubs/ifdp/1998/61*/default.htm. [An excellent reference on the methodology of single-equation modelling]

Giot, P. and Laurent, S. (2003), "Value-at-risk for long and short trading positions," *Journal of Applied Econometrics* **18(6)**, 641-663. [an application of time-varying volatility models to measuring risk]

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), 'On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks', *Journal of Finance* **48(5)**, 1779-1801. [a GARCH model that takes into account asymmetric effect in volatility]

Green, W.H. (2003), *Econometric Analysis*, fifth edition, Pearson Education International. [An excellent reference on the methodology of econometrics]

Hall, R. (1978), "Stochastic Implications of the Life Cycle – Permanent Income Hypothesis Theory and Evidence," *Journal of Political Economy* **86**, 6, 971-987. [a paper on the life consumption model]

Hamilton, J. (1994), *Time Series Analysis*. Princeton: Princeton University Press. [An excellent reference on the methodology of time series analysis]

Harris, R.D.I., Sollis, R. (2003), *Applied Time Series Modelling and Forecasting*. John Wiley. [An excellent reference on applied time series modelling]

Harvey, A.C. (1990), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge. [provides details on the Kalman filter]

Harvey, A.C. and P.H.J. Todd (1983), "Forecasting Economic Time Series with Structural and Box-Jenkins models: a case study," *Journal of Business and Economic Statistics*, **1**, 299-307. [paper on state space models]

Hendry, D.F. and J.F. Richard (1982), "On the Formulation of Empirical Models in Dynamic Econometrics," *Journal of Econometrics* **20**, 3-33. [provides an introduction to the 'general to specific' approach to modelling]

Hey, J. (1983), *Data in Doubt*, Martin Robertson, Oxford, UK. [presents an introductory exposition to the Bayesian approach]

Hsiao, C. (2003), *Analysis of Panel Data*, Cambridge University Press, New York. [advanced reference on panel data models]

Im, K. S., Pesaran, M. H., Shin, Y. (2003), Testing for unit roots in heterogeneous panels. *Journal of Econometrics* **115**, 53--74. [an advanced paper presenting a widely used test for panel unit roots]

Johansen, S.J. (1988), "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control*, **12**, 231-54. [an advanced paper presenting a widely used test for cointegration]

Johansen, S.J. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregression Models," *Econometrica* **59**, 1551-80. [an advanced paper presenting estimation and testing for cointegrating vectors]

Johansen, S.J. (1992), "Determination of the Cointegrating Rank in the Presence of a Linear Trend," *Oxford Bulletin of Economic and Statistics* **54**, 383-97. [an advanced paper presenting a widely used test for cointegration in the presence of a trend]

Johansen, S.J. and K. Juselius (1990), "The Full Information Maximum Likelihood Procedure for Inference on Cointegration - with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics* **52**, 169-210. [an advanced paper presenting a widely used estimation method for cointegrating vectors]

Johnston, J. and J. Dinardo (1997), *Econometric Methods*, 4th edition, McGraw-Hill, New York. [An excellent reference on the methodology of econometrics]

Joyeux, R. and P. Abelson (2000), "Forecasting Methods: An Overview,", in Abelson P and. R. Joyeux (eds), *Economic Forecasting*, Allen and Unwin, Sydney, Australia. [gives an over view of econometric forecasting methods]

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T.C. Lee (1985), *The Theory and Practise of Econometrics*, 2nd edition, John Wiley and Sons, New York. [An excellent reference on the methodology of econometrics]

Kennedy, P. (1998), *A Guide to Econometrics*, fourth edition, Blackwell Publishers Inc., Oxford, UK. [An excellent introductory reference on the methodology of econometrics]

Longin, F. and Solnik, B. (1995), 'Is the correlation in international equity returns constant: 1960-1990?', *Journal of International Money and Finance* **14(1)**, 3−26. [an empirical study of time-varying correlation in international stock markets]

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Model*, Axel Springer Verlag, New York. [An excellent reference on the methodology of time series analysis and VAR models]

McDonald, J., and Mofitt, R. (1980), "The Uses of Tobit Analysis", *Review of Economics and Statistics*, 62, 318-321. [exposition of the Tobit model]

Maddala, G.S. and S. Wu (1999), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economic and Statistics* **61**, 631-52. [an advanced paper presenting a widely used test for panel unit roots]

Meade, N. (1995), "Neural Network Time Series Forecasting of Financial Markets," *International Journal of Forecasting* **11**, 4, 601-2. [an application of neural networks]

Mittelhammer,R., G. Judge, and D. Muller (2000), *Econometric Foundations*, Cambridge University Press, Cambridge, UK. [provides an extensive overview of econometric methods]

Newey, W. and K. West (1987a), "Hypothesis Testing with Efficient Method of Moments Estmation," *International Economic Review* **28**, 777-787. [an advanced paper which derives Wald, LM and LR tests for the GMM estimators]

Newey, W. and K. West (1987b), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* **55**, 703-08. [an advanced paper on the properties of estimators in the presence of heteroskedasticity or autocorrelation]

Novick, M.R. and P.H. Jackson (1974), *Statistical Methods for Educational and Psychological Research*, McGraw-Hill, New York. [presents an intermediate exposition to the Bayesian approach]

O'Connell, P.G.J. (1998), "The overvaluation of purchasing power parity," *Journal of International Economics* **44**, 1-19. [provides a comparison of panel unit root tests]

Pedroni, P. (1997), "Cross sectional dependence in cointegration tests of purchasing power parity in panels," Working Paper in Economics, Indiana University. [an advanced paper on panel cointegration tests]

Pedroni, P. (1999), "Critical values for cointegration tests in heterogeneous panels with multiple regressors.," *Oxford Bulletin of Economic and Statistics* **61**, 653-678. [an advanced paper on panel cointegration tests]

Pedroni, P. (2004), "Panel Cointegration, Asymptotic and Finite Sample Properties of Pooled Time Series tests with an Application to the PPP hypothesis." *Econometric Theory* **20**, 3, 597--625. [an advanced paper on panel cointegration tests]

Phillips, P.C.B. and H. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* **67**, 1057-1111. [an advanced paper on panel estimators]

Phillips, P.C.B. and H. Moon (2000), "Nonstationary Panel Data Analysis:An overview of Some Recent Developments," *Econometric Reviews* **19**, 263-86. [an advanced paper on panel estimators]

Phillips, P.C.B. and P. Perron (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika* **75**, 335-46. [an advanced paper on a widely used unit root test]

Robertson, D. and M. Wickens (1994), "VAR Modelling," in S. Hall (ed.), *Applied Economic Forecasting Techniques*, Harvester Wheatsheaf, Hemel Hampstead, 29-47. [an introduction to VAR modelling]

Sims, C.A. (1980), "Macroeconomics and Reality," *Econometrica* **48**, 1-48. [this paper provides a rationalization for VAR models]

Van der Ploeg, F. (1986), "Rational Expectations, Risk and Chaos in Financial Markets," *Economy Journal (Supplement)* **96**, 151-62. [an advanced paper on chaos theory]

White, H. (1980), "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* **48**, 817-38. [an advanced paper on the properties of estimators in the presence of heteroskedasticity]

Wold, H.O. (1954), *A Study in the Analysis of Stationary Time Series*, 2nd ed., Almquist and Wicksell, Uppsala, Sweden. [an advanced book on the analysis of time series]

Wooldridge, J. (2002*), Econometric Analysis of Cross-Section and Panel Data*, MIT Press: Cambridge, MA. [An excellent reference on the methodology of panel data models]

Wooldridge, J. (2006*), Introductory Econometrics A Modern Approach*, third edition, Thomson, South Western: Mason, OH. [An excellent introductory reference on the methodology of econometrics]

Biographical Sketches

Roselyne Joyeux received a PhD in Economics and a Master degree in Mathematics from the University of California at San Diego. She has held positions at Cornell University, the University of Auckland, the Centre for Operations Research and Econometrics at the University of Louvain-la-Neuve and the GREQAM at the University of Provence. She is presently a Senior Lecturer at Macquarie University. She has published widely in both econometric theory and economic modelling. Her interests include time series analysis and its applications to macroeconomic modelling, energy economics, carbon emission trading and finance.

George Milunovich is a Senior Lecturer in economics at Macquarie University, Sydney, Australia. He joined Macquarie University in 2004, with previous appointments at the University of New South Wales (UNSW) and consultancies in the private sector. George received an undergraduate degree in Economics from the University of Auckland, a Masters degree in Finance from the University of Sydney and a PhD from UNSW. George's primary areas of research are international stock markets linkages, real estate markets, carbon-dioxide trading and econometrics. He has published in international academic journals such as: Journal of Financial Research, Journal of Multinational Financial Management, Global Finance Journal, Economic Record et cetera. His current interests include pricing efficiency of carbon-dioxide derivative contracts and the role of financial contagion on portfolio risk.