ASECU South-Eastern Europe Journal of Economics 2 (2011) 107-165 SEEJE

# ISSUES AND MODELS IN APPLIED ECONOMETRICS: A PARTIAL SURVEY

ANDREAS A. ANDRIKOPOULOS[a,*]
DIMITRIOS C. GKOUNTANIS[b]
[a]Professor Emeritus, Athens University of Economics and Business,
[b]Department of Economics, University of Southampton, UK

**Abstract**
Econometrics, in its long history, has been and continues to be an important branch not only in general economics (macro and micro), but also in specialized fields in the area of economics, such as financial and spatial economics. This paper surveys some recent developments related to the specification and estimation of econometric models widely used in applied research. Even though we lay emphasis on time series models and their application in financial and spatial econometrics, additional topics, such as limited dependent variable models and simultaneous equation systems, are also reviewed in the paper. However, it should be emphasized that the survey is not unified in the sense that it does not provide an exhaustive review of the development of econometrics through its long history. It simply brings up certain topics(classical and contemporary) which may be of interest to those researchers who are concerned with empirical issues in economics in general and in its specialized fields in particular.

**JEL Classification:** C01, C1, C5
**Keywords:** Classical, Financial, and Spatial Econometric Models

**Corresponding Author:* Prof. Andreas A. Andrikopoulos, 10, Papastratou Street, Gizi, Athens 114 76, Greece. e-mail: aaa@aueb.gr

## 1. Introduction

Since the works of Ciompa [1910] and Frisch [1933], many investigators in the literature of the field have defined econometrics in different but conceptually equivalent ways. Among these investigators we include: (1) Tintner [1953], who defines ***econometrics*** as an important special method for the evaluation of mathematical economic models in numerical terms and for the verification of economic theories; it uses the methods of modern statistics for this purpose. (2) Haavelmo [1944], who defines econometrics as the method of econometric research aiming at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. (3) Samuelson, Koopmans and Stone [1954], who define econometrics as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference. (4) Spanos [1986], stating that econometrics is concerned with the systematic study of economic phenomena using observed data. (5) Geweke, Horowitz and Pesaran [2007], who define econometrics as the method aiming to give empirical content to economic relations for testing economic theories, forecasting, decision making, and for ex post decision policy evaluation.

Regardless of which definition is adopted, econometrics can be thought of as being the application of mathematics and statistical methods in the analysis of economic data; that is data involved in an ***economic model***. The economic models (static or dynamic) contain behavioral relations for the endogenous variables which are considered solutions of optimization problems and they may be planned contingent on either observed outcomes or expectations. The solution of these relations gives the economic equilibrium. The static models focused on the study of the effects that changes in the exogenous variables may have on the endogenous variables, ignoring the process of transition between the involved equilibria, which are taken up by the dynamic economic models, such as growth models. Econometric models, on the other hand, using mathematical and statistical tools, aim to put the economic models in an empirical perspective of economic relations. To this end, a distinction is made between theoretical and applied econometrics. Theoretical econometrics deal with issues concerning the statistical properties, that is properties of the estimators, in an economic model. Applied econometrics, on the other hand, focuses on issues concerning the application of econometric methods, that is methods representing applications of standard statistical models, to evaluate economic theories. The basic difference between econometric and statistical models is that in econometrics the economic data are observational rather than being derived from controlled experiments as assumed in statistical models. This distinction led to the development of methods in dealing, among other things, with identification and estimation of simultaneous equation models. Generally speaking, econometrics is classified into two major categories: Classical and Bayesian Econometrics.

***Classical econometrics***, which reflects the tradition of the ***Cowles Commission,*** makes use of the distinction between endogenous and exogenous variables imposing restrictions to achieve identification, and allowing the investigators to make causal inferences in the absence of controlled experiments. The models treated in the classical econometrics depends on the particular interest of the researchers and the complexity of the relationships they represent. Based on the number of the equations involved the models are described as single-equation models, that is models in which the variable of interest to the researcher is expressed as a function of one or more independent variables; and multiple-equation models, that is models consisting of a set of interrelated variables (simultaneous equation models). A further categorization of the models include: (1) stochastic vs. nonstochastic models; (2) qualitative models vs. quantitative models; (3) time-series vs. cross-section model; and (4) pooled data vs. panel data models. Recently, emphasis was laid on the so-called financial econometric models, usually classified as classical, volatility, and regime-switching models. Special ingredients of classical econometrics include: (1) the correct specification of the model, implying both the selection of the functional form and the choice of the variables which should be included in the model. (2) the choice of the appropriate method of estimation. Depending on the nature of the problem and the available data, methods of estimation include the OLS, the 2SLS, the 3SLS, the method of moments, the generalized method of moments, the SURE and the IV methods. (3) the evaluation of the model in terms of the theoretical, econometric, and statistical criteria.

***Bayesian Econometrics*** differs not only from classical econometrics but also from frequentist econometrics. The basic difference between classical and Bayesian econometrics is that in classical econometrics the researcher works with models, such as regression models, and by using data, estimates, through the application of the appropriate technique, the parameters of the model. Bayesian econometric, on the other hand, uses Bayes's rule to do so. It is based on the subjective view of probability, which argues that uncertainty about anything unknown can be expressed using the rules of probability, and the vector of the coefficients is as a random variable, compared to ***frequentist*** econometrics in which the vector of the coefficients is not a random variable.

The context of this partial survey is organized as follows. Section 2 summarizes the main problems relating to specification, estimation and evaluation of single equation models. Section 3 focuses on time-series models with emphasis on financial econometrics. Classical time-series models (univariate and multivariate), volatility models, regime-switching models, and panel data estimation is the core of the analysis in this Section. In Section 4 the basic Logit, Probit and Tobin models are analyzed and Section 5 discusses basic spatial econometrics. Some issues in simultaneous equation models are discussed in Section 6. The last Section summarizes this review.

## 2. The Linear Regression Model: An Overview

In estimating economic relationships, the most widely used method is the OLS. With this method in applied situations it is usually assumed that the so-called Gauss-Markov assumptions are satisfied. The model and the related assumptions are given below:

$$Y = X\beta + \varepsilon \quad [2.1]$$

where $Y = \{Y_1, Y_2, ..., Y_N\}'$ , $\varepsilon = \{\varepsilon_1, \varepsilon_2, ..., Y_N\}'$ , $\beta = \{\beta_1, \beta_2, ..., \beta_k\}'$ and $X$ is an $Nxk$ matrix of the exogenous variables of the model.

Assumptions:

$$E(\varepsilon) = 0 \longrightarrow E(Y) = X\beta \quad [2.2]$$
$$E(\varepsilon\varepsilon') = \sigma^2 I \quad [2.3]$$
$$Cov(\varepsilon X) = 0 \quad [2.4]$$

The first assumption mean that, on average, the regression line should be correct; that is, if the model includes all the significant exogenous variables, both positive and negative, the error terms will average out to zero. The second assumption states that: (1) each distribution of ε has the same variance, $\sigma^2$, that is the errors are ***homoskedastic;*** and (2) all error terms are pairwise uncorrelated, implying absence of ***autocorrelation***. The third assumption suggests that the matrix X is deterministic and not stochastic. Assumptions [2.2]-[2.4], summarized by the ***Gauss-Markov-Theorem,*** suggest that the OLS estimator, $\hat{\beta} = (X'X)^{-1} X'Y$ ,is the best linear unbiased estimator **(BLUE).** From an empirical point of view some or all the Gauss-Markov assumptions may not be satisfied. In such cases, the issues involved include: (1) the identification of the problem in question; and (2) the derivation of alternative estimators satisfied the Gauss-Markov assumptions. We briefly outline these issues below.

***Heteroskedasticity.*** The problem of heteroskedasticity, usually appearing in cross-section models, refers to the fact that the error terms are mutually uncorrelated but the variance of $\varepsilon_i$ is not constant but varies over the range of observations. That is $Var(\varepsilon_i / X) \neq \sigma^2$ =constant. Various test statistics, each on its own merit, have been developed in the literature for heteroskedasticity testing. Basic test statistics include the Goldfeld-Quandt [1965] test; (2) the Spearman [1904] test; (3) the Glesjer test [1969]; (4) the Breusch-Pagan test [1979,1980]; (5) the White test [1980]; and (6) the Bartlett test [1949]. Alternative methods have been advanced in the literature to cope with the heteroskedasticity problems, such as the weighted least squares(WLS) method, the generalized 2SLS, and the method of the maximum likelihood function (FIML)[1].

---

1. In section 3 we provide an extensive analysis of the heteroskedasticity problem in time-series models, such as the class of ARCH models and their extension.

***Autocorrelation.*** The problem of autocorrelation, common in time-series models, violates the assumption that all error terms are pairwise uncorrelated.
That is $E(\varepsilon\varepsilon') = \sigma^2 (X'X\}^{-1}$ . Omitted important independent variables from the model, models with lagged endogenous variables, and incorrect functional form of the model are some of the causes of autocorrelation. In its general form, the autoregressive model, belonging to the AR(p) category, is written as follows.

$$Y_t = X_t\beta + \varepsilon_t,\quad \varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \ldots + \rho_i\varepsilon_{t-i} + v_t \qquad [2.5]$$

$$-1 \le \rho_i \le +1 \quad v_t \xrightarrow{iid} N(0,\sigma_v^2),\ t=1,2,\ldots,T \qquad [2.6]$$

$\rho_i$=autocorrelation coefficient, p=length of the lagged error.

There are various forms of autocorrelation, each of which leads to a different structure of the autocovariance error matrix. Among these forms, the first-order autocorrelation, AR(1), is the most popular in empirical situations. Focusing on this form, researchers in the field have developed procedures to detect, first, and to cope with, second, the problem of autocorrelation. Formal statistical tests to detect autocorrelation in AR(1) include: (1) The Durbin-Watson [1951] test; (2) the h-Durbin [1970] test; (3) the Von Neumann [1941] ratio; and (4) the Berenblut-Webb [1973] test. The Lagrangian Multiplier test (LM-test), suggested by Breusch [1978] and Godfrey [1978] can be used for detecting higher order autocorrelation. The basic method in estimating models with autocorrelation is the Generalized Least-Squares Method: GLM ή Aitken's Generalized LS)[2].

***Multicollinearity.*** In estimating econometric models it is assumed that cov ($X_i$, $X_j$)=0. In such a case, the matrix X΄X is not invertable. Thus, the estimation of the model with the OLS does not provide unique values of the coefficients of the model. The presence of multicollinearity in a model casts doubts on both the interpretation of the estimates and the correct signs of the coefficients. Various criteria can be used to identify the presence and severity of multicollinearity in a model, such as the: (1) *t*-statistics, $R^2$ and $r^2_{X_iX_j}$ ; (3) criteria of Frisch, Farrar-Glauber [1967], Theil [1965,1968] and Klein [1950a, 1950b]; (4) eigenvalues-condition index, the tolerance and variance inflation factor and the auxiliary regression method. Methods for the solution of the multicollinearity problem include restricted least squares regressions, ridge regressions, transformation of the exogenous variables in an uncorrelated set, combination of cross section and time series data, dropping irelevant variables of the model, the principal component regression, and many others[3].

2. Alternative methods of estimation, particularly applied in low order autoregressive structure include the Cochrane-Orccutt [1949], Hidbreth and Lue [1960], and the Durbin [1960] two-steps procedures.
3. See classical textbooks for complete analysis.

***Specification Errors.*** The violation of the Gauss-Markov assumptions in empirical situations can, in general, be attributed to the misspecification of the model in question. Model misspecification leads to specification errors which are due to: (1) omission of important variables, (2) inclusion of superfluous variables, (3) wrong functional form, (4) wrong specification of the error term, and (5) measurement errors both in the dependent and independent variables in the model[4]. The OLS estimators with: (a) omission of important variables gives biased and inconsistent estimates with large variances and standard errors and (b) inclusion of irrelevant variables, the OLS estimators are unbiased and consistent and the estimated variance is larger than necessary (implying larger confidence intervals than necessary). The OLS estimates are unbiased, consistent and less efficient when the dependent variable is measured with error and biased and inconsistent when the values of the independent variables are measured with errors. The examination of the residuals, the Durbin-Watson statistic, the Lagrange multiplier test and the Ramsey's[1969] RESET test can be used to detect specification errors.

The brief analysis given above is based on the assumption that the investigators reflect the views of the Cowles Commission econometricians in the sense that the analysis focuses on the estimation and evaluation of a particular econometric model. An alternative methodological approach, known as LSE methodology or a general to specific approach, grounded on the works of Leamer [1983] and Hendry [1980], in which the econometric research has shifted from the estimation and evaluation of a given model to the choice among alternative and competing models. More specifically, Leamer[5], for discovering the true model, introduces the extreme bound analysis and Hendry, on the other hand, developed the notion of top-down (general to specific) modelling strategy. In choosing the best among competing models, the most common tests used is the nonnested F-test and the Davidson-MacKinnon[1981] J test[6].

## 3. Time Series Econometrics

Historically, the analysis of economic relationships and their future prediction were based on econometric models, that is models in which the dependent variable(s) is expressed as function of quantitative, qualitative and other random variables. However, the rapid increase of economic relationship, both at national and international level, made it quite difficult to perform and evaluate economic and other predictions. This has led the investigators in the field to develop new techniques of analysis better applied to modern economic theory in general and the financial market in particular;

4. See Cujarati [1988].
5. For criticisms of Leamer's approach and some related issues see Angrist and Pischke [2010].
6. For a historical review of econometrics see Malinvaud[1980], and J.F.Geweke, J.L.Horowitz and M. Hashem Pesaran [2006].

that is markets in which risk and uncertainty constitute an important factor in formulating policies in the proper direction. These new approaches include, among others, models based on time series data and usually classified into four, ***even arbitrary,*** broad categories: (1) classical (univariate and multivariate) models; (2) volatility models; (3) regime-switching models; and (4) panel data models.

### *3.1. Classical Time Series Models*

The basic linear classical time series models include the AR(p),equation [3.1], MA(q), equation [3.2], and ARMA(p,q), equation [3.3], type.

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t,\ \ t = 1,...,T \ \text{ or } \ \phi_p(L)Y_t = \phi_0 + \varepsilon_t \qquad [3.1]$$

$$Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t,\ t = 1,...,T \ \text{ or } \ \theta_0 + \theta_q(L)\varepsilon_t = Y_t \qquad [3.2]$$

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots, \theta_q \varepsilon_{t-q}$$

$$= \phi_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \Leftrightarrow Y_t \sim ARMA(p,q) \qquad [3.3]$$

Where:

$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$, $L$ = lagged operatorς and p = degree of polynomial.

$\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$ and q = degree of polynomial.

The basic issues relating to the above models include: (1) the identification of the specific model fitting better the time series data; (2) the determination of the degree of polynomials; (3) the method of estimation; and (4) the choice of the model that gives the best predictions. In empirical situations, the degree of polynomials, p and q, is determined by the autocorrelation functions (ACF and PACF) and, the Akaike [1973] and Schwartz [1978] information criteria. DF and ADF [1973, 1981] Phillips and Perron [1988] procedures can be used to test for stationarity. The chosen model can be estimated by the Maximum Likelihood Method and the evaluation of the findings is based on the stationarity of the residuals and the goodness of fit of the time series data. The standard criteria are used to test the forecasting ability of the chosen model. Non stationary time series are transformed to stationary using the Box-Jenkins [1970, 1976] methodology. The transformed nonstationary to stationary results in the so-called ARIMA (p,d,q) models, where p and q give the degree of polynomial of AP(p), and MA(q), and d the required differrencing of the series to achieve stationarity.

In the case of models with only one time series, prediction requires that the series must be stationary. Nonstationary time series lead to spurious results in the sense that both the estimators and the relevant statistics are misleading. However, in bivariable (multivariable) time series models the problem of spurious results could be avoided if the series are cointegraded.; cointegration implying that, if two or more variables are I(I) and their linear combination is I(0), then the series are stationary. A classical example of cointegration is the estimation of the Keynesian consumption function $C_t = \beta_0+\beta_1 Y_t+\varepsilon_t$, where C(1) and Y(1). If $\varepsilon_t = C_t - \beta_0 Y_t$ and $\varepsilon_t$(0) then $C_t$ and $Y_t$ are cointegrated, that is stationary. The usual statistical tests in testing Cointegration include the Dickey-Fuller tests in both its version (DF and ADF), the Cointegrated Regression Durbin-Watson test (CRDW), the Sargan and Bhargava [1983] test, and other alternative tests suggested by Maddala and Kim [1998].

The concept of cointegration relates to the issues of the long-run equilibrium, the error correction mechanism (ECM) and the VAR representation of the model. A cointegrated relationship implies long-run equilibrium in the sense that the equilibrium error is stationary, meaning that it fluctuates around zero. However, in the short run there may be disequilibrium. This short run disequilibrium can be corrected and pushed back to the long run equilibrium by utilizing the so-called error correction mechanism (ECM) in which the error term in the cointegrated relationship, $\varepsilon_t = C_t - \beta_0 Y_t$, is considered as an equilibrium error. In a two time series model, the ECM takes the form:

$$\Delta C_t = \gamma_0 + \gamma_1 \Delta Y_t + \gamma_2 \hat{e}_{t-1} + v_t \qquad [3.4]$$

Where Δ= first difference, $\hat{e}_{t-1}$ is the one period lagged value of the residual from regression $C_t = \beta_0 + \beta_1 Y_t + \varepsilon_t$ and $v_t \xrightarrow{iid} N(0,1)$. The coefficient $\hat{\gamma}_2$, if statistically significant, captures the adjustment toward the long run equilibrium. On the other hand, the coefficient $\hat{\gamma}_1$ indicates the short run effect of $Y_t$ on $C_t$.

The vector auroregressive model(VAR), extends the univariate model to multivariate time-series. The VAR model: (1) is considered a powerful tool in describing the dynamic behavior of economic series; (2) provides superior forecast of financial time-series compared to those forecasts obtained in univariate models; (3) avoids the problem of identification in simultaneous equation models; (4) is useful for structural inference and policy analysis; and (5) is suitable for calculating short and long run impacts multipliers. In its simplifying form, the stationary VAR(2) is written as follows:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} Y_{1t}-1 \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} Y_{1t}-2 \\ Y_{2t}-2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \qquad [3.5]$$

$\varepsilon_{1t} \xrightarrow{iid} N(0,1)$, $\varepsilon_{2t} \xrightarrow{iid} N(0,1)$, and $Cov(\varepsilon_{1t}, \varepsilon_{2t}) \neq 0$

In the model stated above, the two endogenous variables, $Y_{it}$ and $Y_{2t}$, have the same regressors, that is the lagged values of $Y_{it}$ and $Y_{2t}$[7]. In essence, the model is a reduced-form model. This implies that the model is just a seemingly unrelated regression (SUR) model. Since in model [3.5] each equation has the same independent variables, each equation may be estimated by the OLS without losing efficiency relative to the GLS. That is, the coefficients of the model are *efficient* and *consistent*. In the case where the number of regressors differs between the equations of the system, the SUR estimates, called near VAR estimates, are efficient. In the bivariable VAR(p) model linear hypotheses can be tested through the Wald procedure and the lag length selection can be obtained through the information criteria such as Akaike (AIC), the Schwartz-Bayesian (BIC) and the Hannan-Quinn (HQC) criteria[8]. Forecasting procedures as those applied in AR(p) models can be also used in the VAR(p) models.

From the economic point of view, it becomes quite difficult to interpret the many estimates involved in a general VAR(p) model. To cope with this problem, researchers in the topic in question use the so-called structural analysis. Three main types of structural analysis appear in the literature: Granger causality tests, impulse response functions, and forecast error variance decompositions.

***Granger Causality Test.*** In the bivariate model [3.5], and in the Granger [1969] framework, $Y_2$ in the first equation of model [3.5] fails to Granger-cause $Y_1$ if all coefficients of $Y_{2t-2}$ are zero. In this case the θ'ς coefficients are lower triangular. That is:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \theta_{11} & 0 \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} Y_{1t} - 1 \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \theta_{11} & 0 \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} Y_{1t} - 2 \\ Y_{2t} - 2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad [3.6]$$

Similarly, the second equation in model [3.5] fails to Granger-cause $Y_2$ if all coefficients of $Y_{1t-1}$ are zero. In this case the θ'ς coefficients are upper triangular. That is:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & 0 \end{bmatrix} \begin{bmatrix} Y_{1t} - 1 \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & 0 \end{bmatrix} \begin{bmatrix} Y_{1t} - 2 \\ Y_{2t} - 2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad [3.7]$$

The same procedure can be used for multivariate VAR(p) models. The Wald statistics can be used to test these restrictions.

---

7. In general terms, the VAR(p) is written as follows:

$$Y_t = \phi + \Theta_1 Y_{t-1} + \Theta_2 Y_{t-2} + \ldots + \Theta_p Y_{t-p} + \varepsilon_i \text{, t=1,…,T} \quad [3.5a]$$

where $Y_i = \left( Y_{1t}, Y_{2t}, \ldots, Y_{nt} \right)'$, $\Theta_i$ are $(nxn)$ coefficient matrices and $\varepsilon_i$ is an $(nx1)$ unobservable zero mean white noise vector process with time invariance matrix $\Sigma$. The VAR(p) model with deterministic terms and exogenous variables is given below:

$$Y_t = \Theta_1 Y_{t-1} + \Theta_2 Y_{t-2} + \ldots + \Theta_p Y_{t-p} + \Phi D_t + G X_t + \varepsilon_i \quad [3.5b]$$

where $D_t$ represents an $(lx1)$ matrix of deterministic components, $X_t$ represents an $(mx1)$ matrix of exogenous variables , and $\Phi$ and $G$ are parameter matrices.

8. For detailed analysis of these cretiria see Lutkepohl [1991] and Hamilton [1994].

***Impulse Response Functions.*** In order to trace out the path effects of structural shocks in the system, Sims [1980] has introduced the Vector Moving Average (VMA) model which, in the most simplified form, is given below:

$$Y_{it} = \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} u_{y_1 t-i} \\ u_{y_2 t-i} \end{bmatrix} \text{ or } Y_t = \mu + \sum_{i=0}^{\infty} \phi_i u_{t-i} \cdot \phi_i = [A_1^i / (1 - b_{12} b_{21})] \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \quad [3.8]$$

As stated in [3.8], the VMA model can be used: (1) as a tool to investigate the interaction between the sequences of the two endogenous variables and, subsequently, to trace out the effects of structural shocks in the entire system; (2) to calculate the short and long run impact multipliers; and (3) to identify the coefficients of the VAR(p) model[9].

***Forecast Error Variance Decomposition.*** The forecast error variance decomposition gives the proportion in the movements in a sequence which is attributable to its own shocks as compared to the shocks of the other variables in the system. Specifically, if the error in variable $Y_{2t}$, $\varepsilon y_{2t}$, does not explain any of the forecast error variance of at all forecast horizons, then the sequence of $Y_{1t}$ is considered exogenous. On the other hand, the $Y_{1t}$ is considered as entirely endogenous if the $\varepsilon y_{2t}$ explains all the forecast error variance in the sequence of $Y_{1t}$ in all forecast horizons.

In their general form, both the VAR(p) and VMA(q) models, contain many variables and long lags. This fact requires, first, a large number of data for their estimation and, second, the estimated coefficients are not very precise. This is due to the fact that the standard error bands do not take into account the parameter uncertainty. The Bayesian VAR(p), BVAR(p), improves the estimates by incorporating in the estimation process: (1) prior information about the structure of the model; (2) possible values of the parameters or functions of these parameters; and (3) parameter uncertainty.

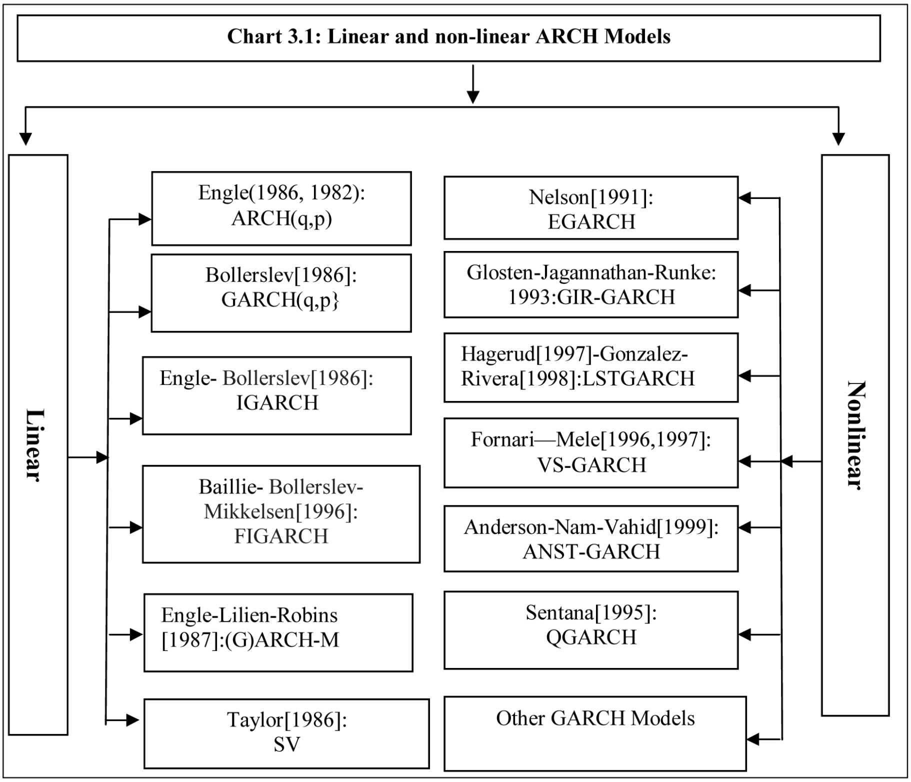
### *3.2. Volatility Models*

Researchers in their attempts to analyze and forecast financial and other time-series data, such as prices of bonds, inflation rates and price of foreign exchange rates, have observed that the prediction of these variables changes considerably from period to period. The prediction error is not stable but varies, particularly in the long run. This volatility may be attributable, among other things, to the volatility of financial markets, governmental policies, and false information about forthcoming events, economic, political and social. The fact that the traditional time-series models (AR, MA and ARMA), particularly models applied in the financial sector, proved to be unsatisfactory both to provide accurate predictions and explain the relationship that exists between returns and risk of financial assets, has resulted in the development of new models, called in the area of financial economics regime-switching volatility econometric models or simply volatility econometric models.

---

9. For a detailed analysis see Sims [1980] and Enders [1995].

In the literature of the field, the most widely used models fall into two major categories: (1) the ARCH-GARCH models and their extensions; and (2) the stochastic volatility models. The first category attempts to capture the most important stylized facts in financial markets, such as volatilities clustering, fat tail and asymmetric distributions of returns, persistence of autocorrelation in absolute or squared returns, and leverage effect. The second category, on the other hand, focuses on models in which the volatility is not directly observable, but is driven by an unobservable random process[10]. In the literature, a distinction is made between linear and nonlinear models (Chart 3.1). This distinction is somewhat arbitrary in the sense that long run time-series may display both linear and nonlinear components.




10. See, for instance, Carroll, Ruppert and Stefanski [1995] and Shephard [1986].

### *3.2.1. Autoreggresive Conditional Heteroskedasticity Models: ARCH(q)*

**The Basic ARCH(q).** The ARCH(q) models, commonly applied in financial economics in which varying volatility clustering are observed, is grounded on the belief that there is the ***suspicion*** that at any point in a time-series its terms will have a characteristic size (or variance). Specifically, the ARCH time series models assume that the variance of the current error term (or its square) is a function of the actual sizes of the error terms (or innovation) of the previous period. The specification of the ARCH model, originally developed by Engle [1982], is based on the following regression model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \ldots + \beta_k X_{kt} + \varepsilon_t = E[Y_t / T_{t-1}] + \varepsilon_t \quad [3.9]$$

This model consists of two parts: (1) The mean of the time-series which can be predicted, given all past information, $T_{t-1}$; and (2) the error term, the variance of which is not constant but a function of the past years'residuals or its square; that is:

$$\begin{aligned} \sigma_t^2 &= E[\varepsilon_t^2 / T_{t-q}] = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2 + \ldots + \gamma_P \varepsilon_{t-q}^2 \\ &= \gamma_0 + \sum_{i=1}^{q} \gamma_i \varepsilon_{t-i}^2 = \gamma_0 + \gamma(L)\varepsilon_t^2 \end{aligned} \quad [3.10]$$

where $q$ is the length of ARCH lags and L=lagged operator. It indicates that the variance of the error includes the constant term, $\gamma_0$, the terms of the ARCH model, $\gamma_1,\ldots\gamma_1$, and the conditional variance, $\sigma_{\varepsilon_t}^2 / \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots, \varepsilon_{t-q}^2$. Positive variance requires that $\gamma_1 \geq 0, \ldots, q \geq 0$ and $\gamma_0 > 0$, and ***stationarity*** implies that $\sum_{i=1}^{q} \gamma_i < 1$.. In a sense, the ARCH(q) model substitutes the sample forecasting variance by a weighted variance, equation [3.10]. In this way, the recent information is given more influence and less importance is attached to the distant past on the conditional variance.

***Estimation***: The ARCH model can be estimated by the OLS or the ML. The procedure is as follows: First, the model [3.9] is estimated with the OLS. Second, from the estimated model we obtain the squares of the residuals. Third, the residuals are regressed on a constant and the q lagged values of these residuals; yielding the conditional variance:

$$\hat{\sigma}_t^2 = \varepsilon_t^2 = \hat{\gamma}_0 + \sum_{i=1}^{q} \hat{\gamma}_i \varepsilon_{t-i}^2 \quad [3.11]$$

***Volatility Hypothesis:*** In testing volatility in a time-series, the null hypothesis and its alternative are as follows: $H_0 : \gamma_1, \ldots, \gamma_n$ **vs.** $H_1$: one or more $\gamma$'s$\neq 0$. In a sample of N residuals under the null hypothesis of no ARCH errors, the test statistic $NR^2$ follows $\chi^2$ distribution with $q$ degrees of freedom. If $TR^2$ is greater than the Chi-square table value, we *reject* the null hypothesis and conclude there is an ARCH effect in the ARCH(q) model. If $TR^2$ is smaller than the Chi-square table value, we *do not reject* the null hypothesis. Acceptance of $H_0$ implies constant variance, that is no volatility in the series.

Forecasting: Based on [3.10], we have:

$$\begin{aligned}\sigma_{h+1}^2 &= \beta_0 + \beta_1\varepsilon_{t-1+1}^2 + \beta_2\varepsilon_{t-2+1}^2 + \beta_3\varepsilon_{t-3+1}^2 + \dots + \beta_p\varepsilon_{t+1-p}^2 \\ &= \beta_0 + \beta_1\varepsilon_t^2 + \beta_2\varepsilon_{t-1}^2 + \beta_3\varepsilon_{t-2}^2 + \dots + \beta_p\varepsilon_{t+1-p}^2\end{aligned} \quad [3.12]$$

$$\sigma_{h+2}^2 = \beta_0 + \beta_1\sigma_{h+1}^2 + \beta_2\varepsilon_h^2 + \beta_3\varepsilon_{h-1} + \dots + \beta_p\varepsilon_{h+2-p}^2 \quad [3.13]$$

$$\sigma_{h+l}^2 = \beta_0 + \sum_{i=1}^{p}\beta_i\sigma_h^2(l-i) \quad , \quad \sigma_h^2(l-i) = \beta_{h+l-i}^2 \quad \text{if} \quad l-i \le 0 \quad [3.14]$$

***Shortcoming:*** Tsay [2005, p.106] summarizes the shortcoming of the ARCH, which include: (1) The ARCH model assumes that positive and negative shocks have the same effects on the volatility of the series. (2) The model does not provide new evidence in understanding the source of variation of a financial time-series. (3) The model is restrictive in the sense that, if the series has a finite fourth moment, $\mathbf{a_2}$ in the ARCH(1) model must be in the range [0,1/3].This constraint becomes complicated for ARCH higher order model. (4) The ARCH models are likely to overpredict the volatility. This is so because these models respond slowly to large isolated shocks to the return series. In addition, the large number of lags in the model complicate the estimation in cases where the series involve the use of daily or weekly data.

**The ARCH-in-Mean ARCH-M.** Engle, Lilien and Robins [1987], extended the ARCH(q) in such a manner that the mean of a series is a function of its own conditional variance. This extension, suitable for application in asset markets, is based on the fact that a risk-averse investor requires compensation for holding a risky asset. In this sense, the ARCH-M: (1) allows the conditional variance to affect the mean of the series; and (2) changes in the conditional variance affect the expected returns of the assets. In its simplest form, the ARCH-M is written as follows:

$$Y_t = \mu + \gamma\sigma_t^2 \quad [3.15]$$

where μ and γ are constant. The parameter γ is called the ***risk premium parameter***. If this parameter is positive it suggests that the expected returns are positively related to the volatility and vs.

### *3.2.2. Extensions of the ARCH(q)*

In order to tackle some of the shortcomings outlined above, the ARCH(q) model was extended and generalized in various ways. Some of these models are discussed below.

**Generalized ARCH Model: GARCH(q,p).** One of the basic shortcomings of the ARCH(q) is that it requires a large number of parameters and high order q to capture the volatility process. In order to tackle this shortcoming, Bollesrslev [1986] introduced the Generalized Autoregressive Conditional Heteroskedasticity model, GARCH(q,p), the basic formulation of which is as follows:

$$\sigma_{tt}^2 = E[\varepsilon_t^2 / T_{t-p}] = \gamma_0 + \gamma_1\varepsilon_{t-1}^2 + ... + \gamma_P\varepsilon_{t-P}^2 + \delta_1\sigma_{t-1}^2 + ... + \delta_q\sigma_{t-q}^2$$

$$= \gamma_0 + \sum_{j=1}^{p}\gamma_j\varepsilon_{t-j}^2 + \sum_{j=1}^{q}\delta_j\sigma_{t-j}^2 = \gamma_0 + \gamma(L)\varepsilon_{t-1}^2 + \delta(L)\sigma_{t-1}^2 \quad [3.16]$$

$$\text{s.t::} \sum_{i=1}^{q}\gamma_i + \sum_{j=1}^{p}\delta_j < 1,\ \gamma_0 \succ 0 \quad \gamma_1 \geq 0,...,p \geq 0 \quad \delta_j \geq 0,...,q \geq 0 \quad [3.17]$$

L=lag operator

The variance of the error in the ARCH(q,p) consists of three parts: (1) the constant term $\gamma_0$; (2) the terms of ARCH, $\gamma_1$,... $\gamma_p$; and (3) the terms of GARCH, $\delta_1$,... $\delta_p$. Conditions [3.17] suggest stationarity and positive conditional variance.

**The GARCH-M.** The GARCH-M model aims to capture the direct relation between returns and the probability of chance in the risk, as this risk is measured by the conditional variance. This is done by augmenting the mean of the variable of our interest by the conditional variance. This model can be expressed as follows:

$$Y_t = \gamma_1 X_t + \gamma_2\sigma_t^2 + \varepsilon_t \quad [3.18]$$

$$\sigma_t^2 = w + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{P}\beta_j\sigma_{t-j}^2 = w + a(L)\varepsilon_t^2 + \beta(L)\sigma_t^2$$

$$\text{s.t.:} \sum_{i=1}^{q}a_i + \sum_{j=1}^{p}\beta_j < 1,\ w > 0,\ a_1 \geq 0,...,p \geq 0 \quad \beta_j \geq 0,...,q \geq 0,\ \gamma_2 = \text{risk premium} \quad [3.19]$$

The model [3.18]-[3.19]: (1) gives the relationship between risk premium and the conditional variance; (2) indicates that the conditional variance is a measure of the risk associated with an asset; (3) allows in the conditional mean to be a function of its variance; and (4) the returns of an asset depends on the degree of the asset in the sense that risk-averter investors require greater compensation for holding this particular asset. Positive variance and stationarity requires that the conditions, given above hold.

Alternative specifications of the linear models include:(1) the Integrated GARCH (IGARCH); (2) the Fractional GARCH (FIGARCH); and (3) the Stochastic Volatility (SV) model. The IGARCH [Engle-Bollerslev,1986] is characterized by the fact that the impacts of past squares of error on the volatility is ***persistent***, implying that the

IGARCH is a unit-root GARCH. That is, the inequality stationarity condition in the GARCH becomes equality restriction in the IGARCH. The FGARCH introduced by Baillie, Bollerslev and Mikkelsen [1993], on the other hand, focuses on the fact that the sum of the coefficients in the GARCH approaches to unity, implying that the effect of any disturbances on the conditional variance is diminishing at slow rates. To put it in another way, the FIGARCH combines the GARCH and the IGARCH model. Finally, the Stochastic Volatility Model (SV)introduced by Taylor [1986] postulates that $\ln(\sigma_t^2)$ is affected by two shocks: (1) the shock which is due to the input of new information in the system, $\ln(\sigma_{t-1}^2)$; and (2) the shock which is attributable to rumors of up-coming events, positive or negative, small or large. It is written as follows:

$$\ln(\sigma_t^2) = \beta_0 + \beta_1 \ln(\sigma_{t-1}^2) + \beta_2 \sigma_t^2 \qquad [3.20]$$
$$\text{s.t.: } \varepsilon_t = z_t \sqrt{\sigma_t^2}\ ,\ , z_t \sim NID(0,1)\ ,\ \sigma_t^2 \sim NID(0,1)\ ,\ Cov(\sigma_t^2, z_t) = 0$$

In general, it has been established in the literature that the linear GARCH models do not perform quite satisfactorily in certain regions of economics and financial time-series data. Structural changes and changes in the behavior of economic units necessitated the construction of non-linear models. Some of these models, which extend the models of ARMA-type[11], are briefly analyzed below.

**The Exponential GARCH (EGARCH):** The EGARCH Model [Nelson, 1991]. As stated above, one of the weakness of the ARCH and GARCH models is that they do not take into account the fact that positive and negative shocks may not have the same effect on the volatility of a series. To put it another way, these models ignore the probability that positive and negative shocks could have different effects on the conditional variance of the errors. The EGARCH model avoids this problem by allowing for asymmetric effects between positive and negative shocks on the conditional variable of a series. That is, the volatility, as measured by the conditional variance, depends both on the magnitude and the signs of the errors. In its simplest form, the EGARCH (1,1) model is written as follows:

$$\log \sigma_t^2 = \beta_0 + \beta_1 \log \sigma_{t-1}^2 + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \delta \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} \qquad [3.21]$$

$$\log \sigma_t^2 = \beta_0 + \beta_1 \log \sigma_{t-1}^2 + (\gamma + \delta) \frac{\varepsilon_{t-1}}{\sigma_{t-1}}\ ,\quad \varepsilon_{t-1} \geq 0 \qquad [3.22]$$

$$\log \sigma_t^2 = \beta_0 + \beta_1 \log \sigma_{t-1}^2 + (\gamma - \delta) \frac{\varepsilon_{t-1}}{\sigma_{t-1}}\ ,\quad \varepsilon_{t-1} \prec 0 \qquad [3.23]$$

11. For alternative GARCH models, see Tsay [1987], Nicholls and Quinn [1982], Melino and Turnbull [1990], and others.

The coefficient γ, which signifies the ***leverage*** effect of $\varepsilon_{t-1}$, indicates ***asymmetry*** effect on the volatility of a series. Specifically γ ≠ 0, indicates assymetry, and γ < 0 indicates that positive disturbances create less volatility than the negative disturbances. Positive (negative) effects on the volatility are measured by γ+δ (γ-δ). For δ=0 the model is non-linear, and in empirical application the γ is expected to be negative. The logarithmic formulation of the model guarantees positive conditional variance, without imposing restrictions on the parameters[12]. Models dealing with the issue of the asymmetry effect include, among others, the Quadratic GARCH [Santana(1995)], equation [3.24], and the GJR-GARCH [Glosten-Jagannathan-Runkle 1993], equation [3.25]. The last term in [3.24] deals with asymmetry effects in the sense that it allows the positive and negative effects on the conditional variance. The GJR-GARCH, on the other hand, postulates that the coefficient of the $\alpha_\iota$ depends on the **sign** of the shock.

$$\sigma_t^2 = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \gamma \varepsilon_{t-1}, \ \gamma = \text{asymmetry coefficient} \quad [3.24]$$

$$\sigma_t^2 = w + a_1 \varepsilon_{t-1}^2 (1 - I[\varepsilon_{t-1} > 0]) + \alpha_2 \varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1 \sigma_{t-1}^2 \quad [3.25]$$

I[.]= indicator function[13].

**Other Non-Linear GARCH.** Alternative specifications of non-linear GARCH models, each on its own merit, have been suggested and/or applied mainly in the field of financial economics. These models include: (1) the NAGARCH [Engle and Ng, 1993] in which the conditional variance is a function of the shocks and the volatility of the past and positive shocks causes more volatility than negative shocks of the same magnitude; (2) the NGARCH [Engle and Bollerslev, 1986] the basic characteristic of which is that volatility is a non-linear symmetric function of both the shocks and the volatility of the past; (3) the TSGARCH [Taylor, 1986 and Schwert, 1989] which is symmetric with the conditional variance being a function of the moving

12. In its general form, the EGARCH is written as follows:

$$\ln(\sigma_t^2) = w + \sum_{i=1}^{q} a_i \ln(\sigma_{t-1}^2) + \sum_{i=1}^{p} \beta_i [[|\varepsilon_{t-i}| / (\sigma_{t-i}^2)]^2 - (2/\pi)] + \gamma[\varepsilon_{t-1} / (\sigma_{t-1}^2)^{1/2}].$$

13. . ..ternatively, [3.20] is written as follows:

$$\sigma_t^2 = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{P} \beta_i \sigma_{t-i}^2 + \gamma d_{t-1}^{+} \varepsilon_{t-1}^2$$

$$\text{s.t.:} \sum_{i=1}^{q} a_i + \sum_{i=1}^{p} \beta_i < 1 \ , \ w > 0 \ \ a_1 \geq 0, \ldots, p \geq 0 \ \ \beta_i \geq 0, \ldots, q \geq 0$$

where $d_t$ is a dummy variable taking the value 1 if $\varepsilon_{t-1}>0$ and zero otherwise and γ = asymmetry coefficient. Volatility is a function of the nature of the shock. That is, positive shock creates smaller effect on the volatility than that of a negative shock of the same magnitude.

average of the absolute lagged disturbances; and (4) the LSTGARCH [Hagurad, 1987 and Gonzalez-Rivera, 1998], the ANSTGARCH [Anderson, et al., 1999] and the VSGARCH [Formari-Mele, 1997].

### *3.2.3. Regime-Switching Models*

Even though the time series linear models have been and continue to be used in applied econometric studies, they appear to be unsuccessful in identifying nonlinearities in certain time series data, particularly data in financial time series (stock prices, exchange rates, and other financial assets). It has been argued in the literature that the observed non-linearity could be better identified if we take into account the fact that the behavior of the time series involved depends on the regime or the stage prevalent at a specific time period. To put it in another way, there are cases where the structures of the time series are not stable but are characterized by abnormal situations which are not captured by models of the ARMA and ARIMA-type and the GARCH models as well. In dealing with this problem, new methodological approaches have been developed, usually termed ***regime-switching*** approaches. These approaches deal with two main categories of models: (1) models in which the regimes are determined by an observable variable (parametric models); and (2) Markov-switching models in which the regimes are determined by an unobserved variable (probabilistic models).

***Parametric Models.*** The basic parametric models include: (1) the threshold autoregressive models (TAR); (2) the Self-Exciting threshold autoregressive models (SETAR); and (3) the smooth transition autoregressive models (STAR). The basic idea of these models is that the time-series consists of two or more regimes and there is a mechanism through which the model switches between these regimes in accordance with whether a particular time-series reaches or approaches a **threshold** value.

***TAR Models:*** In the TAR models, the regime in period t is determined by an observed variable, say $q_t$, relative to a threshold value, say γ, which is assumed to be weakly exogenous. In this manner, the model becomes a fractional linearization of a nonlinear model in the R domain of $q_t$,

where R= $\{r_0, r_1, \cdots r_{l-1}, r_l\}$, being locally linear[14], and $-\infty = \prec r_0 < r_1 < \cdots < r_{l-1} < r_l = +\infty$. The most popular class of the TAR models are based on autoregressive process. For instance, the specification of a two regime TAR can be written as follows:

$$\begin{aligned} Y_t = &(\phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \cdots + \phi_p^{(1)} Y_{t-p}) I(q_{t-1} \le \gamma) + \varepsilon_t \\ &+ (\phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \cdots + \phi_p^{(2)} Y_{t-p}) I(q_{t-1} > \gamma) + \varepsilon_t \end{aligned} \quad [3.26]$$

14. Tong [1990] classifies the TAR models into three basic categories (1) Piecewise Linear Models); (2) Smooth Threshold Autoregressive Models, and (3) Piecewise Polynomial Models.

Where $\varepsilon_t \xrightarrow{iid} N(0,1)$, conditional upon the history of $\Omega_{t-i} = (y_{t-1}, y_{t-2},..., y_{t-p})$ with $E[\varepsilon_t / \Omega_{t-1}] = 0$ and $E[\varepsilon_t^2 / \Omega_{t-i}] = \sigma^2$, $\phi_i = \{\phi_{0.i}, \phi_{1.i},..., \phi_{.t-p}\}'$ ,i=1,2, and I(A) is an indicator function with I(A)=1 if the event A occurs and I(A)=0 otherwise[15].

***SETAR Models:*** The SETAR models differ from the TAR in the sense that the threshold variable $q_t$ is assumed to be a lagged value of the time series itself; that is $q_t=y_{t-d}$, for a certain integer d>0, and non observed as the specification of the TAR assumes. This implies that the time series may be linear within each regime but not linear for the whole time series. The basic formulation of the SETAR is given below.

$$\begin{aligned} Y_t &= (\phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \cdots + \phi_p^{(1)} Y_{t-p}) I(Y_{t-d} \le \gamma) \\ &\quad + (\phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \cdots + \phi_q^{(2)} Y_{t-q}) I(Y_{t-d} > \gamma) + \varepsilon_t \\ &= \phi^{(1)} x_t^{p}{}' I(Y_{t-d} \le \gamma) + \phi^{(2)} x_t^{q}{}' I(Y_{t-d} > \gamma) + \varepsilon_t = x_t(\gamma)'\theta + \varepsilon_t \end{aligned} \quad [3.27]$$

where :

$$\phi^{(1)} = (\phi_0^{(1)},..., \phi_p^{(1)})', \ \phi^{(2)} = (\phi_0^{(2)},..., \phi_q^{(2)})', x_t^p = (1, Y_{t-1},..., Y_{t-p})', \ x_t^q = (1, Y_{t-1},..., Y_{t-q})'$$

$$'x_t(\gamma) = (x_t^{p}{}' I(Y_{t-d} \le \gamma) \quad x_t^{q}{}' I(Y_{t-d} > \gamma)),' \theta = (\phi^{(1)}{}' \ \phi^{(2)}{}')'$$

***STAR Model.*** One of the basic shortcomings of the SETAR model is the fact that the conditional mean function is non continuous, implying that the border between the regimes is given by the threshold variable, $y_{t-d}$. Tong [1978] and Terasvirta [1994] have shown that a smooth transition between the regimes is obtained if the indicator function, $I(y_{t-1}>\gamma)$, is replaced by a continuous function, say $G(y_{t-d};c, \gamma)$. This function is assumed to be a continuous function bounded between 0 and 1, and the transition variable, $y_{t-d}$ is taken in most cases as being lagged endogenous variable for certain integer d>0. Based on this transition function, the specification of a two regime STAR model takes the following form:

$$\begin{aligned} Y_t &= (\phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \cdots + \phi_p^{(1)} Y_{t-p}) \{1 - G(y_{t-d}; c, \gamma)\} \\ &\quad + (\phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \cdots + \phi_q^{(2)} Y_{t-q}) G(y_{t-d}; c, \gamma) + \varepsilon_t \end{aligned} \quad [3.28]$$

In empirical applications better results are obtained if the transition smoothing function used in the STAR model, $G(y_{t-d}; c, \gamma)$, is replaced either by a logistic function, given by [3.29] or an exponential function, given by [3.30].

$$G(q_t; s, \gamma) = [1 + \exp(-s(q_t - \gamma))]^{-1} \quad [3.29]$$

$$G(q_t; s, \gamma) = 1 - [\exp(-s(q_t - \gamma)^2] \quad [3.30]$$

15. In compact form the model is written as follows:

$$Y_t = x_t^{'} \phi^{(1)} I(q_{t-1} \le \gamma) + x_t^{'} \phi^{(2)} I(q_{t-1} > \gamma) + \varepsilon_t = x_t(\gamma)' \theta + \varepsilon_t$$

With $x_t(\gamma) = (x_t^{'} I(q_{t-1} \le \gamma) \ \ x_t^{'} I(q_{t-1} > \gamma))'$, θ=$(\phi^{(1)}{}', \phi^{(2)}{}')'$

The logistic function leads to LSTAR model, and the exponential to ESTAR model. Extensions and/or modifications of the basic regime-switching models include Multiple Regime STAR Models (MRSTAR), Time Varying STAR Models (TVSTAR), Vector STAR Models (VESTAR) models and others[16].

***Probabilistic Models.*** The basic difference of the probabilistic models from the parametric is that the regime relative to the threshold is determined by probabilistic criteria, such as Markov-Chain procedures. Hamilton [1989] introduced the so-called Markov Switching Model (MSW), in which for an AR(1) model and two regimes can be written as follows:

$$y_t = \begin{cases} \phi_{0.1} + \phi_{1.1} Y_{t-1} + \varepsilon_{1t} & if\ s_t = 1 \\ \phi_{0.2} + \phi_{1.2} Y_{t-1} + \varepsilon_{2t} & if\ s_t = 2 \end{cases} \quad [3.31]$$

Model [3.31] indicates that: (1) the process $s_t$, corresponding to a **first-order Markov-process or Markov-chain procedure**, depends only on the regime of the previous period, $s_{t-1}$; and (2) the transition from the first to the second regime depends on the transition probabilities given below.

$$P_i = \begin{bmatrix} P(s_t = 1/s_{t-1} = 1) & P(s_t = 2/s_{t-1} = 1) \\ P(s_t = 1/s_{t-1} = 2) & P(s_t = 2/s_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad [3.32]$$

where:

$$\begin{cases} p_1(s_t = 1) = [1 - p_{22}] : [2 - p_{11} - p_{22}],\ p_{11}+p_{12}=1,\ p_1+p_2=1 \\ p_1(s_t = 2) = [1 - p_{11}] : [2 - p_{11} - p_{22}],\ p_{21}+p_{22}=1,\ p_1+p_2=1 \end{cases}$$

In [3.32] the probabilities $p_{ij}$ show the probability that regime i in period t-1 is followed by regime j in the second period, t. The value of $p_i$ determines the speed of transition between regimes. In contrast to SETAR model, based on a deterministic process, the MSW follows a stochastic process[17]. Model [3.32] can be extended to include multiple regimes models[18].

Depending on the particular specification of the model, important issues in empirical applications include: (1) the choice of the estimation method; (2) the evaluation of the findings; and (3) the predictability of the variable(s) the model is designed to analyze. Methods of estimation include both parametric approaches, such as the ML and its variants, the GMM, and Bayesian and Monte-Carlo methods. A variety of tests

16. For an analysis of these models see Tong [1983,1990], and van Dijk, Terasvirta and Franses [2002].
17. For variance of the MSW see McCulloch and Tsay [1993] and Hamilton [1994].
18. See Boldin [1996] and the reference cited there.

have been used or suggested to evaluate the performance and the accuracy of the predictions. Tsay [2005] classifies the nonlinearity tests as parametric and nonparametric. The parametric tests include the Ramsey [1969] test, the Tsay [1969] F-test, and the threshold test [Tsay, 1989]. The Ljung-Box [1978], used to check for model inadequacy, the Brock, Dechert, and Scheinkman[1987], used to test the iid assumption of a time series, and the bispectral test, used to test for linearity and Gaussianity[19].

### *3.2.4. Panel Data Econometrics*

Baltagi [2005], Hsiao [2003] and Klevmarken [1989] provide the reasons for which panel data should be used in empirical applications. They point out, among other things, that panel data: (1) control for individual heteroskedasticity; (2) give more informative data, more variability, less collinearity between variables, more degrees of freedom and more efficiency; (3) are able to study better the dynamics of adjusttment, to construct and test more complicated behavioral models, to identify and measure effects that cannot be identified in cross-section and time-series data; and (4) micro panel data are more accurate than that on macro level and panel unit root tests have standard asymptotic distributions. Design and data collection problems, distortion of measurement errors, selectivity problems and short time-series dimension are some of the limitations of panel data[20]. The panel data models widely used and/or proposed in the literature are classified into two major categories: Fixed Effects and Random Effects Models, both static and dynamic, balanced and unbalanced. A brief review of some of these models is given below.

***Static Panel Data Models(SPD).*** In analyzing the SPD models, we consider, first, the following regression model (RM):

$$Y_{it} = \alpha_i + X'_{it}\beta + \varepsilon_{it} = \alpha_i + \sum_{i=1}^{k} \beta_i X_{it} + \varepsilon_{it}\,,\ \varepsilon_{it} \xrightarrow{iid} N(0,\sigma_\varepsilon^2) \qquad [3.33]$$

i = 1,2,…N and t = 1,2,…, T

where i refers to cross-sectional units, such as individuals, firms, countries, households, assets, etc and t to time period. The characteristics of [3.33] are that a change in X's affects all the units and all periods in the same manner but the average level for units i and j may be different. Based on [3.33] and a set of assumptions, we could specify the following Models:

19. For an excellent analysis see Tsay [2005].

20. For a detailed analysis of both advantages and shortcomings of panel data see Baltagy [2005 and Hsiao [2003].

***The Pool Regression(PR) Model.*** The PR or Constant Coefficient (CC) model assumes that intercepts and slopes in [3.33] are constant for all i and t. In this case the PR model takes the following form:

$$Y_{it} = X'_{it}\beta + a + \varepsilon_{it} \qquad [3.34]$$

where $Y_{it}$ is an *Nx*l vector of the endogenous variable, $X_{it}$ is an *Nxk* matrix of the exogenous variables and β and α are parametric constants. Given enough degrees of freedom, an estimate could be made of N time-series regressions, each corresponding to the $i^{th}$ unit. However, in the case where the β and α are constant, both over time and cross-section units, more efficient estimates can be obtained by pooling all cross-section data[21].

***The Fixed Effect(FE) Model.*** The PR model discussed above has been criticized on the ground that both intercept and slopes are assumed to be constant. The FE model is based on the assumption that, given constant slope, the intercept term varies over cross section units and time. In its simplified form, the FE model is written as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ .Y_N \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_N \end{bmatrix} \beta + \begin{bmatrix} i & 0 & \cdots & 0 \\ 0 & i & \cdots & 0 \\ . & . & \cdots & 0 \\ . & . & \cdots & 0 \\ . & . & ... & 0 \\ 0 & 0 & \cdots & i \end{bmatrix} \begin{bmatrix} \alpha_1 \\ a_2 \\ . \\ . \\ . \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_N \end{bmatrix} \; or \; Y_{it} = X'_{it}\beta + Z_i\alpha + \varepsilon_{it} \qquad [3.35]$$

The FE model, as stated above, indicates that the regressand variable, Y, is a function of two groups of regressors: (1) the first group, $X'_{it}$, without constant term, and (2) the second group, $Z'_i\alpha$, reflecting the heterogenous or individual effect. In practice, the FE model is implemented through the Least Squares Dummy (1 if i=j and 0 other-wise) Variable (LSVD) Estimator written as follows[22]:

$$Y_{it} = \sum_{j=1}^{N} \alpha_j d_{ij} + \sum_{i=1}^{k} \beta_i X_{ik} + \varepsilon_{it} \qquad [3.36]$$

21. See Pindyck and Rubinfeld [1998].

22. There are two basic problems associated with the use of dummies in the FE model: (1) the dummy variables do not directly identify what causes the regression model to shift over time and over individuals, and (2) uses up a substantial number of degrees of freedom.

The FE model involves a number of statistical tests, such as individual, group, time, and a group and time effect tests. The individual coefficients are tested with the t-statistic. The group effect test, on the other hand, uses the F-statistic to test the following hypotheses: $H_0$:$\alpha_1$=$\alpha_2$=...=$\alpha_N$=0, or $H_0$:$\alpha_1$–$\alpha_2$=0 and others[23]. Analogous is the procedure of testing the time and group-time hypothesis[24].

In spite of the simplicity, the FE model is not without shortcomings. First, in practice, PE models includes a very large number of cross-sectional units of observations. This fact requires too many dummy variables for their specification. Inclusion of many dummy variables in the PE model reduces the degrees of freedom, leading to adequately powerful statistical tests, increases the standard errors of the coefficients, due to multicolinearity, and the error terms may be correlated with the individual effects.

***Random Effects(RE) Models.*** There are cases where there are factors that affect the dependent variable but are excluded from the estimation procedure. The RE model takes this problem into account by considering the coefficient $\alpha_i$ in [3.33] as a random factor taking the place of the excluded variables in the regression model. That is,

$$Y_{it} = \mu + X'_{it}\beta + \varepsilon_{it} = \mu + \sum_{i=1}^{k}\beta_i X_{it} + (\varepsilon_{it} + \alpha_i)$$
$$\varepsilon_{it} \xrightarrow{iid} N(0,\sigma_\varepsilon^2) \quad \alpha_i \xrightarrow{iid} N(0,\sigma_a^2) \qquad [3.37]$$

The error term in [3.37] consists of two parts. The first part is given by $\alpha_i$ which is assumed to be constant over time and the second part, given by $\varepsilon_{it}$, is assumed to be uncorrelated over time. This structure of the error term has some implication concerning the method of estimation. Variants of the RE models include: (1) The Random Parameter Model and the Error Component Model. The first model assumes that the parameters vary over the cross-sectional units. The second model, on the other hand, called also variance component model, assumes that the error term in the regression model consists of three components: the cross-section specific error, the time specific error, and the error affecting only the particular observation[25].

---

23. In the FE framework, the value of the F-statistic is calculated as follows:

$$F = \frac{\left(R_U^2 - R_R^2\right)/(N-1)}{\left(1 - R_U^2\right)/(NT - N - K}$$, where $R_U^2$ corresponds to unrestricted model, that is to the FE model, $R_R^2$ corresponds to restricted model, Y=Xβ+ε, N=number of groups, NT=total number of observations , K=number of regressors in the model and T=total number of temporal observations.

24. See Greene [2003].

25. Hausman [1978] specification test can be used for choosing between FE and RE models. This test exploits the fact that the parameters in the FE model should not be statistically different from those in the RE model. If the value of the $X^2$ statistic is greater than the critical value we conclude that the parameter estimates in the RE model are statistically significant from those in the FE model. This implies that the RE model is misspecified.

***Dynamic Panel Data Models (DPD).*** The unique characteristic of the DPD model is that it allows us to take into account dynamic elements, such as persistence, partial adjustment, habit formation and others, and understand better the dynamics of adjustment taking place in economic relationships, as characterized by the inclusion of one or more lagged regressors in the specification of the model. The specification of a DPD model with one lagged regressor is as follows:

$$Y_{it} = X'_{it}\beta + \rho Y_{i,t-1} + \alpha_i + \varepsilon_{it}, \ |\rho| \prec 1 \qquad \text{i=1,2,...N, t=1,2,...,T,} \qquad [3.38]$$
$$\varepsilon_{it} \xrightarrow{iid} N(0,\sigma_{\varepsilon}^2) \text{ and } \ \alpha_i \xrightarrow{iid} N(0,\sigma_a^2)$$

The model, as specified above, gives two sources of persistence over time: autocorrelation, due to the presence of lagged variable in the regressors, and the heterogeneity effect among the individual units[26].

The issues of stationarity and cointegration has been an important subject of analysis in DPD models. This is so because, among other things, the observed heterogeneity between the cross-sectional units violates the stationarity conditions required for testing hypotheses in PD models.

In testing for stationarity a number of alternative unit root tests, each on its own merit, have been developed in the literature. Baltagi [2005] classifies the panel unit roots tests into two basic categories: (1) panel unit roots tests based on the cross-sectional independence assumption and includes Levin, Lin and Chu [2002] test, the Im, Pesaran and Sin [2003] test, the Harris and Tsavalis [1999] test, the Breitung-Meyer [1994] test, and the Residual-Based LM test. (2) panel unit roots tests based on the cross-sectional dependence assumption was suggested by Pesaran [2003,2004]. Test for cointegration in PD models can be found in the works of Kao [1999], McCoskey and Kao [1998], Pendroni [2002,2004] and Larsson, et al. [2001].

***Estimation of Panel Data Models.*** The choice of the estimation method in Panel data econometrics depends on the model considered and the assumptions made. Specifically, assuming, first, residual homogeneity and normality and, second, that the errors are independent and homoskedastic, the OLS may be used in estimating both the CC and PE model. The Estimated (EGLS) or Feasible (FGLS) Generalized Least Squares Method can be applied for heteroskedastic models. The Generalized Method of Moments (GMM) can be applied in estimating DPD, that is models with lagged dependent variable(s), and models with heteroskedasticity, autocorrelation and outliers[27].

---

26. For a comprehensive analysis of the DPD models see Baltagi [[2005].
27. The statistical Packages LIMDEP, STATA and SAS, among others, can be used for panel data analysis.

## 4. Limited Dependent Variables Model

Traditionally, in econometric models the dependent variable is a quantitative variable while the independent variables include both quantitative and qualitative variables, such as dummy variables. These models cannot be used to analyze cases where the dependent variable is dichotomous, that is the dependent variable has only two possible values (1 or 0), categorical, or qualitative, called also quantal response. Linear models of this nature are considered inappropriate in predicting the outcome of such binary choices because some of the assumptions of the linear regression models are violated. Specifically, the error terms are heteroskedastic, correlated with explanatory variables, and the predicted value of the dependent variable may fall outside of the range of 1 and zero. These problems are tackled by the ***logit*** model, based on a cumulative logistic probability function, and ***probit*** model, based on a cumulative normal probability function. Models in which the dependent variable is a mix of discrete and continuous outcomes are termed ***limited response*** models. Briefly, we analyze these models below.

### *4.1. The Linear Probability Model [LPM]*

The Simple LPM is given by [4.1]. In this model, it is assumed that the $X_i$ is non stochastic and $E(\varepsilon_i) = 0$.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \ , \quad D = Y_i = \begin{cases} Y = 1, \textit{ if the person is unemployed} \\ Y = 0, \ \textit{otherwise} \end{cases} \qquad [4.1]$$

For values of Y 1 or zero the $\varepsilon_t$ in [4.1] takes the values of $(1 - \beta_0 - \beta_1 X_i)$ and $(-\beta_0 - \beta_1 X_1)$, respectively. Denoting now by $P_1$ and $1\text{-}P_1$ the probabilities for 1 and zero and assuming that $E(\varepsilon_t)$=0, we obtain:

1. $E(\varepsilon_i / X_i) = (1 - \beta_0 - \beta_1 X_i)(P_1) + (-\beta_0 - \beta_1 X_i (1 - P_1) = (1 - \beta_0 - \beta_1 X_i) \neq 0$.

This suggests that the normality assumption of the error is violated.

2. $$\begin{aligned} \sigma_\varepsilon^2 &= E[\varepsilon_i - E(\varepsilon_i)]^2 = E(\varepsilon_i)^2 = \sigma_Y^2 = E(Y_i / X_i)[1 - E(Y_i) \\ &= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) = P_1(1 - P_1) \end{aligned}.$$

This suggests that the errors are heteroskedastic.

The estimation of [4.1] with the OLS will give insufficient estimates and the WLS is recommended. Further, in empirical applications of the model the restriction that the predicted values of the dependent variable must fall within the rage of 1 and zero is violated . Some doubt is also cast on the validity of the $R^2$ as a measure of the goodness of fit of the model.

### *4.2. The Probit Model*

The probit model avoids the problem that in LPM the predicted value of the dependent variable violates the restriction that $0 \leq E(Y_i / X_i \leq 1$ for all the values of the independent variable(s). To this end, the probit model transforms the values of the independent variable(s) to a probability ranging in value between 0 and 1, with the property that the probability increases as the X increases and vice versa. The model: (1) introduces an unobservable utility index, $I_i = \beta_0 + \beta_1 X_i$; (2) assumes that there exists a critical (threshold) level of $I_i$, $I_i^*$, such that if $I_i \succ I_i^*$ the event will occur, otherwise it will not; and (3) for the unobserved $I_i$ and $I_i^*$ the probability that $I_i^* \leq I_i$ can be calculated, given the assumption of normality, from the standard normal distribution, given below:

$$P_i = \Pr ob(Y=1) = \Pr ob(I_i^* \leq I_i) = F(I_i)$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{T_i} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0+\beta_1 X_i} e^{-t^2/2} dt \quad t \to N(0,1) \qquad [4.2]$$

Inversion of [4.2] gives the required information on $I_i$, making, thus, the model ope-rational. That is:

$$I_i = F^{-1}(I_i) = F^{-1}(P_i) = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad [4.3]$$

The probit model is estimated by Maximum Likelihood Estimation, the OLS and the WLS.

### *4.3. The Logit Model*

The logit model deviates from the probit model in the sense that the cumulative normal distribution is replaced by the cumulative logistic distribution function, given below.

$$P_i = E(Y-1/X_i) = F(\beta_0 + \beta_1 X_i) = 1/[1+e^{-(\beta_0+\beta_1 X_i)}] = 1/(1+e^{-Z_i})$$
$$-\infty \lhd Z = \beta_0 + \beta_1 X_i \lhd \infty, \text{ and } \quad 0 \prec P_i \leq 1 \qquad [4.4]$$

Given the fact that the model is nonlinear, the application of linear methods of estimation is not appropriate. The problem of nonlinearity can be circumvented by expressing the ***odds ratio*** in logarithmic form. That is,

$$Z_i = \ln[P_i/(1-P_i)] = \beta_0 + \beta_1 X_i \qquad [4.5]$$

The logit model is estimated by Maximum Likelihood Estimation, the OLS and the WLS.

### *4.4. Censored Regression Models (CRM)*

The CRM due to James Tobin refers to the model in which the dependent variable is limited (censored) from above or below. In mathematical terms, the CRM expresses the observed level of the dependent variable Y in terms of the ***latent*** variable Y*. That is:

$$Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \mathrm{Y_i} = \begin{cases} \beta_0 + \beta_1 X_i + \varepsilon_i & \text{if } Y_i^* \succ 0 \\ 0, \ \text{otherwise} & \text{if } Y_i^* \le 0 \end{cases} \qquad [4.6]$$

The CRM is distinguished from the so-called truncated sample in which the information on the regressors is available only if the regressad is observed. The CRM can be estimated by the method of maxiumum likelihood[28]. Statistical packages, such as the E-views, Stata, Shazam, Rats and many others can be used to estimate the CRM[29].

## 5. Spatial Econometrics: SE

Spatial econometrics, as a branch of econometrics in general, deals with spatial autocorrelation (spatial interaction, or spatial dependence) and structure (spatial heterogeneity) in both cross-sectional and panel data regression models. Spatial ***dependence*** refers to the error term which appears in the regression model and can be due to the fact that there is a spatial correlation of omitted variables, aggregate variables, and errors in variable measurement. Spatial ***heterogeinity***, on the other hand, refers to the fact that economic activity is unevenly distributed over space. Originally, SE was centered in applied research in areas of urban and regional economics, regional science and topology. In recent years it has been applied in a wide range of empirical research, such as labor economics, local public finance, international economics, agricultural and environmental economics, public economics, housing and transportation markets. In the literature, the spatial econometric models are classified in three basic categories: (1) cross-sectional spatial econometric models; (2) panel spatial econometric models; and (3) limited dependent variable spatial econometric models. A brief review of these models is given below[30].

---

28. In the Tobin model the maximum likelihood function is given by:

$$\log L = \sum_{Y_i \succ 0} -\frac{1}{2}\left[\log(2\pi) + \log\sigma^2 + \frac{\left(Y_i - \beta X_i\right)^2}{\sigma^2}\right] + \sum_{Y_i \succ 0} \log\left[1 + F\left(\frac{\beta X_i}{\sigma}\right)\right].$$

29. For limited dependent variables using panel data estimation procedures and related issues see, among others, Baltagi [2005] and Kyriazidou [1997, 2001].
30. For an excellent review of the development of spatial econometrics see Anselin, L [2010].

*5.1. Cross-Sectional Spatial Models (CRSM)*

Both from a theoretical and empirical point of view, the CRSM deals with the specification of two broad classes of models: (1) spatial lag models (SLM); and (2) spatial error models (SEM). An important ingredient of these models and their extensions is the modification of the traditional econometric models to include spatial elements in their formulation. These spatial elements are summarized by the so-called ***weight matrix***, W, which is used to generate spatial exogenous variables in the model in question. This weight matrix: (1) is a squared matrix the dimension of which is equal to the number of observation in the i location; (2) is nonzero if location i and j are neighbors, and zero otherwise; and (3) is used to formalize a notion of locational similarity and important in testing hypotheses. In practice, alternative criteria have been used to define boundaries, such as common boundaries, distance bands and social distances[31].

**The SLM.** The SLM modifies the classical linear cross-sectional model (equation 5.1) by allowing the dependent variable Y to be a function of Y in neighboring locations (equation 5.2).

$$Y = X\beta + \varepsilon\,, \quad \varepsilon \xrightarrow{iid} n(0, I\sigma^2) \qquad [5.1]$$

$$Y = \rho WY + X\beta + \varepsilon \rightarrow Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon\,, \qquad [5.2]$$

where ρ is the spatial autocorrelation coefficient, Y is *nx*l vector of the observations on the dependent variable, *W* is a *nxn* matrix of weights, X is a *nxk* matrix of the regressors, and β and ε are *nx*l vectors of the coefficients and the error terms, respectively. The WY is termed ***spatial lag.***

The model, as stated in [5.2], is nonlinear in terms of ρ and β, follows a spatial AR(1) autoregressive process[32] and is no longer homoskedastic, implying that the estimation of the model with the OLS will give biased and inconsistent estimates. It can, however, be used for the calculation of local and global spatial multipliers[33].

**The SEM.** The SEM postulates that the spatial influence on the dependent variable is coming through the error term. It is formulated as follows:

$$Y = X\beta + \varepsilon \qquad [5.3]$$

$$\varepsilon = \rho W\varepsilon + v \rightarrow (I - \rho W), \quad \varepsilon = (I - \rho W)^{-1} v \qquad [5.4]$$

31. For a detailed analysis see Anselin [2003a,2003b) and the references cited there.
32. For higher order autoregressive process see Blommestein [1983, 1985].
33. For a detailed analysis see Anselin [1988a, 1988b],LeSage and Pace [2009] and references cited there.

Substituting [5.4] into [5.3] yields the SEM. That is:

$$Y = X\beta + \rho W\varepsilon + v = X\beta + (1_n - \rho W)^{-1} v\,, \quad v \xrightarrow{iid} n(0, \sigma^2 I_n) \qquad [5.5]$$

where v is an *nx*l error vector, ρ is a scalar coefficient and Y, X ,β, and W are as before. Variants of the SLM and the SEM include the Durbing Spatial Model (SDM) and the General Spatial Model (GSM).

**The SDM.** The SDM extends the SLM by allowing for explanatory regressors from neighboring observations. In mathematical terms, the SDM is given below.

$$\begin{aligned} &Y = \rho WY + X\beta + WX\gamma + \varepsilon \rightarrow (I - \rho W)Y = X\beta + WX\gamma + \varepsilon \\ &\quad \rightarrow Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} WX\gamma + (I - \rho W)^{-1}\varepsilon \end{aligned} \qquad [5.6]$$

The kx1 vector of γ gives the marginal effect of the explanatory variables from neighboring observations, while the WX product reflects the effect of an average of neighboring units, X-values, on Y.

**The GSM.** The GSM due to Anselin [1988], combines the SLM and the SEM. In simple mathematical terms, the specification of the GSM has as follows:

$$Y = \rho W_1 Y + X\beta + \varepsilon \rightarrow Y = (I - \rho W_1)^{-1} X\beta + (I - \rho W_1)^{-1}\varepsilon \qquad [5.7]$$

$$\varepsilon = \gamma W_2 \varepsilon + v \rightarrow \varepsilon = (I - \gamma W_2)^{-1} v \;, \text{ v} \xrightarrow{iid} n(0, \sigma^2 I) \qquad [5.8]$$

Substitution of [5.8] into [5.7] gives the GSM. That is,

$$Y = (I - \rho W_1)^{-1} X\beta + (I - \rho W_1)^{-1}(I - \gamma W_2)^{-1} v \qquad [5.9]$$

where $W_1$ and $W_2$ are the spatial weight matrices. Depending on the contiguity definition, these matrices: (1) contains first-order contiguity relations or functions of distance and (2) takes zeros -values in positions reflecting non-contiguous observational units, and ones in positions associated with neighboring units. The GSM, as stated above, includes both the LSM and SEM. For instance, if we impose the restriction that X=0 and $w_2$=0, the GSM reduces to SAR, given by [5.1], above. Similarly, for $w_2$=0 the GSM reduces to SEM, given by [5.5].

From an empirical point of view, special attention of the researchers centers around the structure of the weight matrix and its properties, the determination of special boundaries, the identification of spatial dependence and the methods of estimation of the spatial cross-sectional models.

The specification of the weight matrix, W, is obtained by the application of spatial lag operator in the dependent variable, WY, the explanatory variables, WX, and in error term, Wε. The correct specification of the weight matrix depends on the model in question and the criteria used in its construction (common boundaries, distance bands, social distances, etc.). Since, however, the objective of the researcher is to obtain asymptotic properties for the estimators and specification tests, the W is subject to regularity conditions[34].

As was mentioned before, the most important difference between the classical econometric models and the spatial econometric models is the spatial autocorrelation in CRSM. Ignoring spatial autocorrelation, when present, could lead to biased and inconsistent estimates in the SLM and biased t-statistics and inconsistent estimates in the SEM. Various test statistics have been developed in the literature to identify the presence (absence) of spatial autocorrelation. These diagnostic tests are classified by Anselin [2003] in two categories: (1) Tests are based on ML, and include the Wald, the Likelihood Ratio Test, and Lagrange Multiplier tests; and (2) specification tests, which include Moran's [1948, 1950], and the Kelejian-Robinson `[1992] test. In essence, these tests are, more or less, similar to the tests applied in time-series models.

Various methods have been used or proposed in the literature for estimating the spatial SLM and SEM models. Three are the basic methods of estimation: (1) the maximum likelihood (ML); the generalized method of moments (GMM); and (3) the Bayesian Estimation. The log ML for the SLM and SEM are given by [5.10] and [5.11], respectively.

$$\ln(\beta,\rho,\sigma^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln\left|I-\rho W\right| - \frac{1}{2\sigma^2}\left[\{(I-\rho W)Y - X\beta\}'\{(I-\rho W)Y - X\beta\}\right] \qquad [5.10]$$

$$\ln(\beta,\rho,\sigma^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln\left|I-\rho W\right| - \frac{1}{2\sigma^2}\left[\left\{(Y - X\beta)'(I-\rho W)'\right\}\{(I-\rho W)(Y - X\beta)\}\right] \qquad [5.11]$$

An alternative of the LM estimator is the method of moments and its variants (instrumental variables, generalized moments, and the GMM). The basic difference between the LM and the GMM is that the latter does not require the assumption of normality and avoids computational problems inherent in the LM estimator[35].

---

34. For further analysis of this issue see Cliff and Ord [1873, 1981], Anselin [1988b,2002], Upton-Fingleton [1985], Dietz [2002], Case [1991,1992], Lee [2002], and Kelejian and Prucha [2002,1999].
35. For reviews of the methods estimation see Anselin [1988b) and LeSage and Pace [2008, 2009]. It should be noted that the proper method of estimation depends on the structure of the weighted matrix and the objective of the researcher.

### *5.2. Panel Spatial Econometric Models (PSEM)*

The panel data econometric models (homogenous and heterogenous), introduced in section 3, have been extended through the modification of the cross-sectional-n-dimensional weight matrix, $W_n$, to the panel spatial dimensional matrix, $W_{nT}$. Specifically, the specification of the homogeneous panel data models are given by [5.12] and [5.13], respectively:

$$Y_{it} = X'_{it}\beta + \varepsilon_{it}, \text{ i=1,2,...,n and t=1,2,...,T, } n \geq T \qquad [5.12]$$

$$Y_{it} = \alpha_i + X'_{it}\beta_i + \varepsilon_{it} \qquad [5.13]$$

where $Y_{it}$ is an observation on the dependent variable at i and t, $X_{it}$ a $kx1$ vector of the explanatory variables, $\beta$ a $kx1$ vector of the explanatory variables, and $\varepsilon_{it}$ the error term[36]. The basic difference between [5.12] and [5.13] is that in [5.12] the intercept and the slope coefficients are assumed to be constant over the cross-sectional units as compared to [5.13] where the coefficients are specific to each cross-sectional unit.

Denoting the $nTxnT$ spatial weight matrix, defined as $W_{nT} = I_T \otimes W_n$, where $\otimes$ is the Kronecker product, the spatial SLM (in matrix notation) and SEM models are given by [5.14] and [5.15], respectively.[37]

$$\text{SPM: } Y = \rho W_{NT} Y + X\beta + \varepsilon \qquad [5.14]$$

$$\text{SEM: } Y_{it} = \rho \sum_j w_{itj} Y_{ij} + \alpha_i + X_{it}\beta + \varepsilon_{it} \qquad [5.15]$$

where SPM is the spatial pooled model and SFM the spatial fixed-effects model. In the SFM model $w_{itj}$ give the elements of the (i,t) row of W and $X_{it}$ the (i,t) row of X, excluding a constant. The LM estimator can be used for estimating both the SPM and the SEM. Software packages include the R system, the Matlab and the GeoDa.

### *5.3. Limited Dependent Variable Spatial Econometric Models (LDVSEM)*

The limited dependent variable models (Logit, Probit and Tobin) discussed in section 4 have been extended to include spatial elements in their formulation. This extension is achieved through the introduction of the spatial weight matrix in the model which the researcher is interested in estimating. We consider, as an example, the spatial ***lag*** model based on the ***latent*** specification. This model, in matrix notation and in its reduced form, takes the following form:

---

36. In Compact form, model [5.13] is written as $Y=X\beta+\varepsilon$, where Y is a $nTx1$ vector, X is a $nTxk$ matrix and $\varepsilon$ is a $nTx1$ vector.
37. For the complete specification of the panel spatial models, the estimation methods and the related statistics, see Anselin [2005] and Baltagi [2005].

$$Y^* = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon \quad [5.16]$$

[5.16] indicates that $Y^*$ is determined both by the values for X in location i and the values of X at locations j. j=1,2,…,k. The basic methods of estimating the Probit model are the maximum likelihood and the GMM procedures[38].

## 6. Structural Equation Models: SEM

The basic issues related to the SEM center, among other things, on estimation methods, identification, simultaniality and exogeneity. Historically, these issues go back to the works of Fisher [1925], who was concerned with the applicability of sampling theory in analyzing economic data, Frisch [1934], who was concerned mainly with problems of multicollinearity and measurement errors, Koopmans [1937] and Haavelmo [1944], who emphasizes the use of stochastic models in econometrics. In these works attempts were made to develop procedures incorporating economic theory, data, econometric methods, and computing techniques.

**Identification.** The concept of identification of structural parameters, particularly in SEM, was and continues to be an important issue in applied econometrics. It can be found in the works of Working [1927] , Koopmans and Leipnik [1950], who derived rank and order conditions for the identification of a single equation in a complete SEM, ignoring the distinctions between endogenous and exogenous variables, while Wegge [1865] and Fisher [1996] provided a solution of the identification problems by imposing restrictions on the elements of the variance-covariance matrix of the structural disturbances. In SEM, the solution of the problem of identification depends on whether there exists a sufficient number of a priori restrictions for the derivation of the structural equations from the reduced-form parameters.

**Exogeneity.** The procedure of choosing the method of estimating an SEM, called Cowles Foundation Approach, was criticised on various grounds. Among others, in this method: (1) the classification of the variables into endogenous and exogenous is in certain cases arbitrary; (2) the identification requires *a priori* which variables should be included and which should be excluded in the equation[39]; (3) the method ignores the fact that changes in the exogenous variables may change the structural coefficients, that is the coefficients in the SEM are not independent of changes in

38 Remember that in the latent specification, the model specifies the dependent variable $Y_i^*$ is a linear function of an *index function* and a random error term, that is $Y_i^* = X_i'\beta + \varepsilon$. The observed counter-part $Y_i^*$, $Y_i$, equals to one if $Y_i^* \succ 0$ and zero otherwise.

39. See T.C.Liu [1960].

the exogenous variables[40]. The Huasman [1978] specification test and the Granger [1969] causality test are used to identify and cope with the exogeneity problem.

**Simultaneity.** Relating to the issue of endogeneity (or exogeneity) is the issue of simultaneity inherent in many models, particularly in macroeconomic models. Assume, for instance, the following simple Keynesian model.

$$C_t = \beta_0 + \beta_1 Y_t + \varepsilon_t \text{, where } \varepsilon_t \xrightarrow{i.i.d} N(0, \sigma^2) \quad [6.1]$$

$$Y_t = C_t + I_t \text{, where } E(I_t, \varepsilon_t) = 0 \quad [6.2]$$

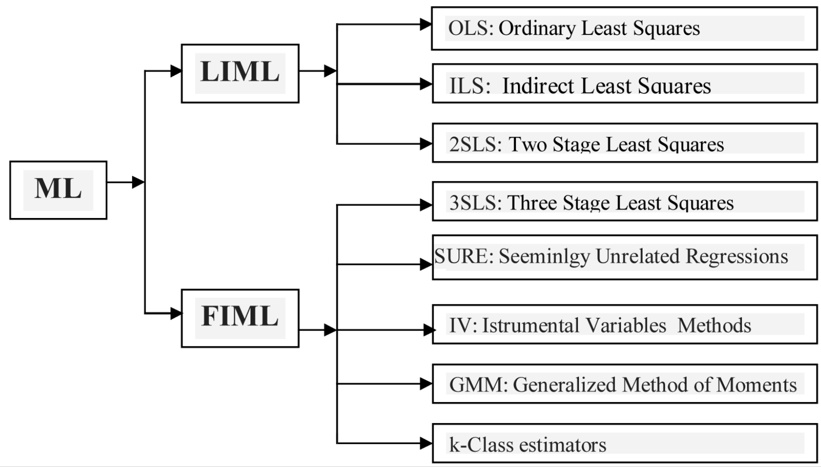
The fact that in this model $Y_t$ appears as exogenous variable in [6.1] and endogenous in [6.2] implies that both $C_t$ and $Y_t$ are endogenous and are jointly determined in the system. To put it in an other way, the fact that $Y_t$ is an exogenous variable in [6.1] and $C_t$ in [6.2] influences $Y_t$, suggests that $E(Y_t, \varepsilon_t)$=0. This further suggests that the OLS estimator of $\beta_1$ in [6.1] will be biased and inconsistent. The ILS or IV methods can be used to tackle the problem of simultaneity.

**Estimation.** In SEM, the methods of estimation are classified into two major cate-gories: Limited Information Maximum Methods (LIML), developed by Anderson and Rubin [1949] and Full Information Maximum Likelihood (FIML), originally proposed by Koopmans and others [1950]. The first category deals with the estimation of a single equation in a SEM and includes the OLS, the 2SLS and the ILS methods. The second category, on the other hand, deals with estimation of a complete SEM and includes the 3SLS, the SURE and the IV approach. Both the LIML and FIML are based on the joined probability distribution of the endogenous variables conditional on the exogenous variables in a given model and yield consistent estimates. The basic difference between the single-equation models and system estimation is that the single equation estimation does not require the full specification of the entire system. Chart 6.1 summarizes these methods of estimation[41].

40. See R.E.Lucas [1976] and Maddala-Sims[1998].

41. For identification and estimation of nonlinear SEM see Fernanzdez-Villaverde and Rubio-Ramirez [2005].

Chart 6.1: SEM-Methods of Estimation




*6.1. Instrumental Variables in SEM*

It has been well established in the literature of the field of econometrics that the presence of **nuisance** parameters in SEM creates a number of problems relating both to the estimation procedures and testing hypotheses in many empirical applications. For instance, Dofour [2007, pp. 786-788] has stated that:

> Weak instruments are notorious for causing serious statistical difficulties on several fronts: {1} parameter estimation;{2} confidence interval consideration; {3} hypothesis testing. In addition, Dofour emphasizes four properties required for a satisfactory solution to the problem of making inference in structural equations: {1} the method should be based on proper pivotal {ideally, a finite-sample pivot}; {2} robustness to the presence of weak instruments; {3} robustness to excluded weak instruments; {4} robustness to the formulation of the model for the explanatory endogenous variables Y {which is desirable in many practical situations}.

*6.1.1. The Specification of the Model*

Alternative IV models have been proposed or applied in the literature to deal with nuisance parameters. In this survey we focus on the most simplified form of SEM, given below.

$$y_1 = y_2\beta + X\gamma_1 + u \qquad [6.3]$$

$$y_2 = Z\pi + X\phi + v_2 \qquad [6.4]$$

where $y_1$ and $y_2$ are respectively nx1 vectors of the two endogenous variables, X is a nx$k_1$ matrix of the exogenous variables, Z is a nx$k_2$ matrix of the IVs, and β, π, γ and φ are unknown parameters. The first equation is a structural and the second a reduced-form equation. By substituting equation [6.4] into equation [6.3] and rearranging, terms, the reduced-form equation for $y_1$ has as follows:

$$y_1 = Z\pi\beta + X\gamma + v_1 \qquad [6.5]$$

In compact form, the reduced-form equation can be written as:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = Z\pi \begin{bmatrix} \beta \\ 1 \end{bmatrix} + X \begin{bmatrix} \gamma \\ \phi \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \text{ or } Y = Z\pi\alpha' + X\xi + V \qquad [6.6]$$

where $\alpha = [\beta \quad 1]$, $\xi = [\gamma \quad \phi]$, $\gamma = \gamma_1 + \phi\beta$, $v_1 = u + v_2\beta$, and $Y = [\,y_1 \;\; y_2\,]$.

In the context of [6.6], the following distributional assumptions are made:

1. The errors in the reduced-form equations, $v_1$ and $v_2$, are assumed to be iid across rows having a zero mean of a bivariate normal distribution with 2x2 nonsingular covariance matrix Ω. That is;

$$\Omega = E\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} x[v_1 \quad v_2] = E\begin{bmatrix} v_1 v_1 & v_2 v_1 \\ v_{12} & v_2 v_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}, \text{ i=1...,n} \qquad [6.7]$$

where the subscript i refers to the i[th] observation of our sample.

2. The variable Z is assumed, entailing no loss of generality, to be orthogonal, that is $Z'X = 0$. Specifically, if we denote with $\vec{Z}$ an original nxk matrix of the IV with $Z'X \neq 0$, then set $Z = M_X \vec{Z} = \left[ I - X(X'X)' X \right] \vec{Z}$.
3. In addition, the IVs included in the model are considered as ***weak*** instruments, meaning that the partial correlation between the instruments and the included endogenous variables is low[42], or, alternatively, the correlation between v and u is zero or close to zero. It is further assumed that $E(u_i / X_i, Z_i) = 0$.

The π, which measures the strength of the IVs, is the most important **nuisance** parameter.[43] This strength is measured by $\lambda = \pi' Z' Z \pi$ .The ratio of $\lambda / w_{22}$ is called concentration parameter. This parameter governs the quality of the standard large-sample normal approximation to the distribution of IV estimators.[44] Small values of $\lambda / w_{22}k$ correspond to weak instruments and the expected value of the first-stage F statistic in testing the hypothesis ***π=0*** in equation [6.5] is asymptotically $1 + \lambda / (w_{22}k)$ under weak instrument asymptotics[45].

42. See, Andrews, et al. [2007].
43. For the properties of the IVs see Andrews, et al. [2006].
44. See Rothenberg [1984].
45. There are three basic reason of using the above simplified model in analyzing the issue in question: (1) It is considered to be the most important in empirical applications; (2) the model can be easily extended for heteroskedastic and autocorrelated errors, and (3) asymptotic results for non-normal errors, both with random or fixed exogenous variables and IVs come in line to the finite sample results for normal errors with fixed exogenous variables and IVs.

*6.1.2. Estimation Methods*

In dealing with the elimination of nuisance parameters in econometric models a large number of methods have been proposed in the literature over the years. These methods are usually classified into two major categories: (1) Frequentist and Fisherian Methods and (2) Bayesian Methods. Below, we briefly analyze some of these methods.

## A. Frequentist and Fisherian Methods

**Profile Likelihood**: The basic idea of the profile likelihood is based on the assumption that we have a probability model for our data which depends on two sets of parameters: (1) the parameters $\pi = (\pi_1,...,\pi_k)$, that is parameters which are in the interest of the researcher and (2) the parameters $\theta = (\theta_1,...,\theta_l)$, that is nuisance parameters which are not of the interest to the researchers but could affect the parameters π. By denoting now the probability density by $f(x/\pi,\theta)$ and further assuming that we have independent observations $X = (X_1,...,X_n)$, the **full** likelihood function has as follows:

$$L(\pi,\theta / X) = \prod_{i=1}^{n} f(X_i / \pi,\theta) \qquad [6.8]$$

In drawing inference on the basis of [6.8], our objective is to eliminate the $\theta = (\theta_1,...,\theta_l)$. To do so, a standard technique for constructing confidence intervals is to find a corresponding hypothesis test, and then to invert that test. The hypothesis and the corresponding **likelihood ratio** statistic are given by [6.9] and [6.10], respectively:

$$H_0 : \pi = \pi_0 \text{ vs } H_0 : \pi \neq \pi_0 \qquad [6.9]$$

$$\lambda(\pi_0 / X) = \frac{\sup\{L(\pi_0,\theta / X;\theta\}}{\sup\{L(\pi,\theta / X);\pi,\theta\}} \qquad [6.10]$$

In [6.9]: (1) the supremum $\{L(\pi_0,\theta / X);\theta\}$ is found over the full space; (2) the supremum $\{L(\pi,\theta / X);\pi;\theta\}$ is found only over the subspace with $\pi = \pi_0$; (3) λ does not depend on **θ** but is a function of $\pi_0$ and X; (4) the -2logλ converges in distribution to a chi-square random variable with k degrees of freedom; and (5) in the context of nuisance parameters, the function λ is called **profile likelihood**.

**Integrated Likelihood.** Perhaps the least studied approach is elimination of nuisance parameters through integration, in the sense that this is viewed as an almost incidental byproduct of Bayesian analysis and is hence not something which is deemed to require separate study. There is, however, considerable value in considering integrated likelihood on its own, especially versions arising from default or noninformative priors. In the context of nuisance parameters, the essence of the integrated likelihood is to achieve some form of likelihood, say $L^A(\theta)$, including the parameters of interest, **π**, only. Elimination of **θ** by simple aggregation results in the so-called ***uniform-integrated likelihood,*** which is of the form:

$$L^U(\pi) = \int L(\pi,\theta)d\theta \quad [6.11]$$

Typically, the integrated likelihoods are written as follows:

$$L(\pi) = \int L(\pi,\theta)\lambda(\theta/\pi)d\theta \quad [6.12]$$

Where $\lambda(\theta / \pi)$ is the weight function for θ. In terms of Bayesian methodologies, discussed below, the term $\lambda(\theta / \pi)$ is the conditional prior density of θ given π.

**Marginal-Conditional Likelihoods.** In certain cases, marginal and conditional likelihoods can be used to eliminate nuisance parameters in an econometric model. This can be achieved when the **full likelihood** is broken down into the product of a marginal likelihood and a conditional likelihood. Marginal and conditional likelihoods are considered special cases of the Cox's partial likelihood[46]. Specifically, if there exists a partition (y,z) of the data x, such that:

$$f(x/\pi,\theta) = h(x)f_1(y/\pi,\theta)f_2(z/y,\pi) \quad \text{or} \quad [6.13]$$
$$f(x/\pi,\theta) = h(x)f_1(y/\pi,\theta)f_2(z/y,\pi,\theta), \quad [6.14]$$

then ignoring the term $hf_1$ (in 6.13) or the terms $hf_2$ (in 6.14), the partial likelihood of π is obtained. Given the fact, however, that the ignored term does not depend on π, there is some loss of information[47].

### B. Bayesian Approaches

Bayesian approaches of tackling nuisance parameters are much more straightforward in nature, compared to their likelihood-based counterparts. Instead of treating them explicitly, we decide for a prior and we subsequently compute the posterior distribution, the center-piece of Bayesian information, with the use of the Bayesian Theorem for up-dating our prior beliefs from the data:

$$p(\theta/y) = \frac{pdf_Y(y/\theta)\pi(\theta)}{\int_{\Omega_\theta} pdf_Y(y/\theta')\pi(\theta')d\theta'} \propto pdf_Y /(y/\theta)\pi(\theta) \quad [6.15]$$

The basic difference between the frequentist approach and the Bayesian approach consists of the fact that the frequentist approach is based on the maximum likelihood ratio tests, providing, thus, significance tests and confidence intervals for the

46. See Cox [1975].
47. For further analysis see Basu [1975,1977], Bernardo, et al.[1999], and Beger, J.O, et al. [1999, p.3].

parameters of interest, which are valid for both moderate and large sample sizes. In contrast, the Bayesian analysis creates random sample from the joint posterior distribution of all parameters. In this manner, the joint distribution of only the parameters of interest can be found through the marginalization over the nuisance parameters. To put it another way, the nuisance parameters are integrated out in the Bayesian method.

### *6.1.3. Detecting IVs in SEM*

In recent years theoretical econometricians have developed various tools for drawing inference in instrumental variables regressions when the instruments are weak. These tools, which are well accepted in empirical works, include the first-stage F-statistic, 2SLS, the $R^2$; the partial $R^2$, the Hahn-Huasman test, and many others, such as the biased and the size methods[48]. The basic hypothesis in these approaches is that $H_0 : \pi = 0$ vs. $H_1 : \pi \neq 0$, where $\boldsymbol{\pi}$ is the coefficient of Z in equation [6.4].

***The First-Stage F-Statistic.*** The simplest of these tools is the first-stage F-statistic of the first-stage regression in the 2SLS. If this statistic is large, that is F>10, we could infer that the instruments are strong, so that the 2SLS output can be used. In contrast, the $H_0$ is rejected if the first-stage F-**statistic** is small, implying that the 2SLS can be biased and the corresponding confidence intervals can be misleading. Moreira [2003], assuming a single included endogenous regressor in the equation of interest and weak instrument large, has shown that: (1) conditional likelihood ratio statistic effectively produces valid and fully efficient confidence intervals and hypothesis testing regardless of whether instruments are weak, strong, or irrelevant; and (2) the LIML method of estimation produces better result than the 2SLS, when the instruments are weak.

***Other Tests.*** In testing $H_0$, alternative approaches have been used, including: (1) The size method, which controls the size of a Wald test of testing the $H_0$: $\beta=\beta_0$ in the equation of interest instead of controlling bias; (2) Hahn-Hausman[2003] test, which tests the null hypothesis of strong instruments under which: (a) the 2SLS estimator; and (b) the inverse of the 2SLS estimator from the ***reverse*** regression should be the same; (3) The first-stage of $R^2$ and Partial $R^2$. The calculation of the first-stage is based on the regression involving both IVs and exogenous regressors, and the partial $R^2$ on the basis of the IVs only. In this context, if the $R^2$ is high and the partial is $R^2$ low, Z is considered as weak instrument.

---

48. See Shea [1997] Stock and Yogo [2005] and Hahn and Hausman [2003].
49. For further analysis of the topic in question and proposed extensions see Staiger and Stock [1997], Hall et al. [1996], and Shea [1997].

### *6.1.4. Testing Hypotheses*

The objective in the model [6.3]-[6.4] is to test the null hypothesis $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$ , treating π, $\gamma_1$ and φ as nuisance parameters. In testing this hypothesis, two major approaches have been used in the literature: (1) fully robust methods, that is methods where the inference is valid for any value of the concentration parameter; and (2) partially robust methods, that is methods that are less sensitive to weak instruments than 2SLS estimator.

Several statistics, each on its own merits, have been developed and empirically performed to test the above hypothesis. These methods and tests statistics are usually classified in two major categories: fully and partially robust statistical tests. The first category includes: (1) A family robust Gaussian Tests; (2) the Gaussian Similar Tests, such as the Anderson- Rubin [1949], the Kleibergen's [2002] and the Moreira [2003] statistics; and (3) Some Conservative Tests, such as those suggested by Staiger and Stock [1997], Wang and Zivot [1998] and Zivot, Startz and Nelson [1998]. On the other hand, partially robust tests includes: (1) the k-Class Estimators, such as LIML, Fuller-k Estimator [1977], and the Bias-Adjusted 2SLS and the Jacknife instrumental variables, suggested by Angrist-Krueger [1999].

### *6.1.5. Applied SEM*

An important area in consumer demand theory is the derivation of complete systems of demand equations and their empirical verification. In classical demand theory any system of demand functions, given the quantity consumed of each commodity as a function of total expenditure (income) and all commodity prices, is derived from any utility function (direct or indirect). The system satisfies certain restrictions connecting income and price slopes (i.e., the Engel aggregation condition, the Cournot aggregation condition, the symmetry condition, and the homogeneity condition). However, consistency of the derived demand functions with utility maximization, in the traditional framework, requires, among other things, certain restrictions on the form of the utility function. Additivity and homotheticity play an important role in formulating tests of the theory of demand. Additivity and homotheticity imply that the elasticities of substitution between pairs of commodities are constant and equal and the expenditure proportions are independent of total expenditures (income). In systems of this nature, the quantities consumed of each of the commodities are usually considered endogenous variables with commodity prices and consumer's income typically treated as exogenous variables. From an empirical point of view, this approach allows the identification of the interdependence among commodities or group of commodities, such as the effects of price changes of certain commodities on consumer demand for other commodities. In practice, the complete demand systems fall into two basic categories: Flexible and Inflexible. The distinction between "flexible" vs. "inflexible" is clearly made by Caves and Christensen[50]. According to them, a functional form

50. See Caves and Christensen [1980].

may be said to be flexible if, given any arbitrary data point (one observation on prices and income), the appropriate choice of parameters can result in any set of price and income elasticities. A functional form is said to be inflexible (nonflexible) if it is not capable of achieving the full set of price and income elasticities. The most popular inflexible functional forms extensively used in the literature include: (1) the Geary-Stone utility function; and (2) the constant elasticity of substitution utility function. The translong utility function (direct or indirect) is used to derive the flexible system of demand equations. A brief review of these models (Chart 6.2) is given below.

### A. The Inflexible Systems

The LES: The LES, involving n demand equations, is derived from the consumer's utility- maximization problem of the Geary-Stone type subject to total expenditure constraint. The derived demand and expenditure functions and the relevant restrictions are given by [6.16] and [6.17], respectively.

$$X_i = \gamma_i + \frac{\beta_i}{P_i}(E - \sum_{j=1}^{n} \gamma_j P_j) . \ E_i = \gamma_i P_i + \beta_i (E - \sum_{j=1}^{n} \gamma_j P_j) . \text{ i. j=1,..., n .} \qquad [6.16]$$

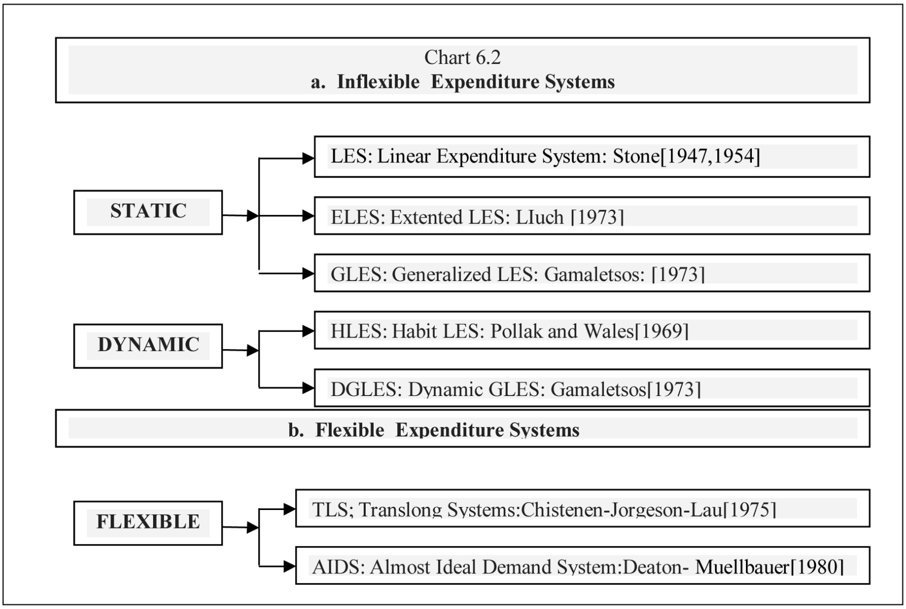
$$\text{s.t: } 0 \prec \beta_i \prec 1 . \sum_{i=1}^{n} \beta_i = 1 . \ X_i \succ \gamma_i . \text{ and } E = \sum_{i=1}^{n} E_i \succ \sum_{j=1}^{n} \gamma_j P_j \qquad [6.17]$$

where E= total expenditures allocated among n commodities. $E_i$ = expenditures on the ith commodity, $X_i$ quantity demanded for commodity i, $P_i(P_1,...,P_n)$, the commodity prices, and $\gamma_i(\gamma_1,...,\gamma_n)$ are the so-called committed(subsistence) quantities. The discretionary expenditures, $E - \sum_{j=1}^{n} \gamma_j P_j$, are allocated among the commodities on the basis of the allocation coefficients, $\partial E_i / \partial E = \beta_i(\beta_1,..,\beta_n)$. The model [6.16] allows the calculation of both uncompensated (Marshialian), $\varepsilon_{ii}$, and compensated (Hicksian), $H_{ii}$, price elasticities of demand and the expenditure elasticities, $\varepsilon_{iy}$. These elasticities are given below.

$$\varepsilon_{ij} = \begin{cases} -1 + (1-\beta_i)(\gamma_i / X_i) . & i = j \\ -(\beta_i \gamma_j P_j)/(X_i P_i) . & i \neq j \text{ . i. j = 1,..., n} \end{cases} \qquad [6.18]$$

$$\text{H}_{\text{ii}} = \begin{cases} -1 + (1-\beta_i)(1 - \gamma_i / X_i) . \text{ i=j} \\ -\beta_i \beta_j \left( E - \sum_{j=1}^{n} P_j \gamma_j \right) / (X_i P_i) \quad i \neq j \end{cases} \qquad [6.19]$$

$$\varepsilon_{iy} = \beta_i / S_i \ , S_i = E_i / E \qquad [6.20]$$




Given [6.16], the own-price elasticities are negative $\varepsilon_{ii}$<0 and, in absolute terms, less than one, the expenditure elasticities, $\varepsilon_{iy}=\beta_i / S_i$, are positive implied that the LES is restricted to normal commodities, and the uncompensated (Marshallian) cross-price elasticities, are negative ($\varepsilon_{ij}$<0), implying complementarity. The fact, on the other hand, that $H_{ij}$ are postulated to be positive implies that the pairs of commodities are net substitutes.

The model [6.16] has been estimated for the Canadian economy using quarterly data covering the period 1963 I-1976 IV. For computational convenience two major concessions have been made. First, only four basic categories (i.e., durables. semi-durables. non-durables. and services) are used in this estimation. Second, by assuming absence of autocorrelation across the expenditure equations, the error structure and its effect on the estimates have been ignored. The method used to estimate the LES is the non-linear least squares, using the Newton-type subroutine in the SHAZAM computing program. This method estimates non-linear regressions in terms of the coefficients but linear in terms of prices and total expenditures by a maximum likelihood procedure[51]. For this application, the initial values are the $\beta_i$'s, equal to one half of the mean values of the expenditure shares and the initial values of the $\gamma_i$´s are set equal to one half of the minimum quantities. Experimentation suggested that the estimates are not overly sensitive to starting values in the plausible range. Table 6.1 reports the findings of this application. This table indicates that: (1)

51. For further information see White [1978].

all coefficients are consistent with the a priori theoretical criteria, that is positive, and statistically significant at the one per cent probability level; (2) The $\beta_i$´s which sum to unity suggest the manner by which discretionary expenditure is allocated among the group of commodities: that is 21.6% of the discretionary expenditures goes to durables, 27.6% to non-durables, 16.3% to semi-durables, and the remaining 34.5% to services; and (3) As indicated by the values of the $\gamma_i$´s coefficients, of the $10277.7 spent on the base quantities, $1316.1 (12.8%) are for durables, $3309.9 (30.3%) for non-durables, $1243.8 (12.2%) for non-durables, and $4408.1 (44.7%) for services.

Table 6.1: The LES Estimates for Canada (1963I-1976IV)

| | Estimates | | | | $R^2$ | Forecasting Criteria | | |
|---|---|---|---|---|---|---|---|---|
| | $\hat{\beta}_i$ | t-value | $\hat{\gamma}_i$ | t- Value | | DW | Theil-U | $R^2$ |
| 1. Durables | 0.216 | 21.98 | 1316.1 | 3.85 | 0.915 | 2.261 | 0.407 | 0.841 |
| 2. Semi-Durables | 0.276 | 59.10 | 3309.9 | 7.33 | 0.991 | 1.716 | 0.256 | 0.950 |
| 3. NonDurables | 0.163 | 6.91 | 1243.6 | 4.50 | 0.994 | 2.206 | 0.439 | 0.912 |
| 4. Services | 0.345 | 21.08 | 4408.1 | 7.05 | 0.982 | 2.502 | 2.506 | 0.712 |
| Total | 1.000 | … | 10277.7 | … | … | … | … | .. |
| | Marshalian Price Elasticities: $\varepsilon_{ij}$ | | | | Hicksian Price Elasticities:$H_{ij}$ | | | |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1. Durables | **-0.450** | -0.155 | -0.141 | -0.155 | **-0.234** | 0.082 | 0.049 | 0.103 |
| 2. Semi-Durables | -0.255 | **-0.410** | -0.249 | -0.299 | 0.040 | **-0.134** | 0.030 | 0.064 |
| 3. NonDurables | -0.136 | -0.158 | **-0.382** | -0.159 | 0.057 | 0.072 | **-0.219** | 0.090 |
| 4. Services | -0.326 | 0.050 | 0.030 | **-0.464** | 0.139 | 0.050 | 0.030 | **-0.119** |
| $\varepsilon_{im}$ | 1.460 | 1.450 | 0.858 | 0.873 | σ= Elasticity of Substitution = 1. | | | |

Table 6.1 reports also the various elasticities implied by the LES. It is indicated in this Table that: (1) All the expenditure elasticities are positive, as the LES postulates, and they ranged from the highest for durable commodities to the lowest for non-durables; (2) The own-price elasticities of demand (uncompensated and compensated) are, as expected, all negative and less than one and vary, in absolute terms, among the groups of commodities; and (3) As expected, the uncompensated cross prices elasticities are negative, suggesting complementarity, and the compensated cross price elasticities are positive, suggesting net substitutability.

The interpretation of these price elasticities of demand is quite straightforward. For example, the effect of a change (increase) in the price of durables will reduce their own demand and, subsequently, will influence the demand for the other commodities in a manner suggested by the ***compensated*** cross-price elasticities. Thus, a ten per cent increase in the price for durable commodities will reduce their own demand by 2.34% ($H_{11}$=-0.234). This price change, given total expenditures, will increase the demand for semi-durables by 0.4% ($H_{21}$=0.040), the demand for non-durables by 0.57% ($H_{31}$=0.057) and the demand for services by 1.39% ($H_{41}$=0.139). A similar interpretation could be provided for the remaining compensated cross-price elasticities.

Three criteria have been used to evaluate the forecasting accuracy of the LES outside the sample period (1977 I-1978 IV): (1) The $R^2$ between actual and predicted consumption expenditures; (2) Theil's inequality, U, statistic; and (3) Durbin-Watson, d, statistic. On the basis of the first criterion. the LES performs fairly well ($0.712 \leq R^2 \leq 0.950$). On the basis of the second criterion which is based on a quadratic loss function, the model, with the exception for services, gives a value below the critical value of unity. Since it is well known that the optimal forecasting process will generate a white noise error process, we evaluate the forecasting accuracy of the LES by using the third criterion, the in-sample Durbin-Watson statistic. This statistic indicates that we cannot reject the hypothesis of white noise error for any of the commodities under the LES regime. It is therefore apparent that when the LES is tested with quarterly data for the Canadian economy, the model performs fairly well in all cases, in terms of its theoretical plausibility, and in three out of four cases, in terms of its forecasting accuracy[52].

**The ELES:** One of the basic shortcomings of the LES is the fact that it considers that total expenditure, E, equals to consumer's income, Y, and E is exogenously determined. In doing so, the LES ignores the consumption-saving decisions and inadequately treats consumption expenditures on durable commodities, which involves consumption-saving decisions. In order to take into account this shortcoming, Lluch [1973] introduced the ELES, given by [6.21].

$$X_i = \gamma_i + \frac{\mu\beta_i}{P_i}(Y - \sum_{j=1}^{n}\gamma_j P_j) \,.\; E_i = \gamma_i P_i + \mu\beta_i (Y - \sum_{j=1}^{n}\gamma_j P_j) \qquad [6.21]$$

$$E = C = (1-\mu)\sum_{j=1}^{n}\gamma_j P_j + \mu Y \,.\; 0 \prec \mu \prec 1 \,.\; S = (1-\mu)(Y - \sum_{j=1}^{n}\gamma_j P_j) \,.\; 0 \leq (1-\mu) \prec 1 \qquad [6.22]$$

where $\mu(1-\mu)$ = marginal propensity to consume (saving). The ELES deviates from the LES in the sense that: (1) The E in [6.16] is replaced by Y in [6.21]; (2) the allocation coefficients in [6.16] is substituted by $\mu\beta_i$, in [6.21]; and (3) the ELES includes both the LES and the aggregate consumption (saving) function of the Keynes type, given by [6.22].

52. For an application of the LES in the financial sector, see Saito [1977] and Andrikopoulos and-Brox [1986].

**The HLES.** The HLES, being dynamic in nature, extends the LES by allowing the subsistence minimum quantities and/or the allocation coefficients to vary in a linear way with either a time trend or the last year's consumption of the ith commodity. In practice, the HLES has been formulated under two alternative specifications: the so-called ***habit formation*** hypothesis and the ***inflation rate*** hypothesis. The habit formation hypothesis decomposes the subsistence quantities into two parts: the physiologically necessary component and psychologically necessary component. The inflation rate hypothesis, on the other hand, assumes that the subsistence minimum quantities in year t are a multiplicative function of the inverse price change of the ith commodity and the previous year's consumption level of the same commodity. Further modification and/or extensions of the LES include the GLES and DGLES, the derivation of which is based on the CES-utility function[53].

## B. Flexible Systems

The models discussed so far are restrictive in the sense that the functional forms of the utility functions used in the derivation of the demand systems are restricted, on a priori basis, to being homothetic and/or additive. This fact imposes serious limitations on applied research. To cope with this problem, attempts have been made, both theoretically and empirically, to derive and to test systems of demand equations from utility and cost functions which do not employ additivity or homotheticity as a part of the maintained hypothesis. These models, usually called ***flexible*** demand models, could be classified into three major classes: The Translong Model (TLM), the Almost Ideal Demand System (AIDS), and the Generalized Leontief (GL) Model.

**TLM.** The first class of flexible systems include those systems the derivation of which is based on translog (direct or indirect) utility or cost functions. The TLM have been extensively applied both in production and consumption theory. We briefly review some of these models below.

**The TL Cost Function.** In the general translog framework, we explicitly assume a well-behaved neoclassical production function, which implies that the sector in question is characterized in terms of a factor minimal cost. In the generalized translog framework, production cost is a function of input prices, output and time, contains all economically relevant information on the underlying technology and is written as:

53. For the derivation of these models and some alternative model specification see Andrikopoulos and Brox [1997].

$$\ln C = \alpha_0 + \gamma_y \ln Y + 0.5\gamma_{yy}(\ln Y)^2 + \sum_{i=1}^{n} \beta_i \ln P_i + 0.5\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} \ln P_i \ln P_j \quad [6.23]$$

$$+ \sum_{i=1}^{n} \gamma_{yi} \ln Y \ln P_i + \sum_{i=1}^{n} \gamma_{i\tau} \ln P_i.T + \gamma_{y\tau} \ln Y.T + \gamma_\tau.T + 0.5\gamma_{\tau\tau}T^2$$

$$\text{s.t.:} \quad \sum_{i=1}^{n} \beta_i = 0 \,.\, \sum_{i=1}^{n} \gamma_{ij} = \sum_{j=1}^{n} \gamma_{ji} = \sum_{j=1}^{n} \gamma_y = \sum_{j=1}^{n} \gamma_{i\tau} = 0 \quad [6.24]$$

Where C is total cost, Y is output, $p_i$´s are the prices of inputs, and T is a time trend representing non-neutral (technological-bias) and scale-augmenting technology. To conform with the theory, restrictions [6.24] are imposed on the translog cost function. No a priori restrictions are imposed with respect to the substitution possibilities among the inputs of production, the extent of economies of scale, the degree of homotheticity, or the particular form of technical change. By Shephard's lemma, the translog cost function yields, in share form, the total cost-minimizing input demands. That is,

$$S_i = \frac{\partial \ln C}{\partial \ln P_i} = \beta_i + \sum_{j=1}^{n} \gamma_{ij} \ln P_j + \gamma_{yi} \ln Y + \gamma_{\tau i}.T \text{ . i=1.....n} \quad [6.25]$$

Given the fact that in the TL framework a direct economic interpretation of the coefficients of the model is not possible, we resort to various elasticities and productivity indexes that are functions of the parameter estimates and yet have standard interpretation. This procedure is facilitated as the model imposes no prior constraints on these measures. In our work, we focus on the following sets of elasticities and productivity indexes.

First, we consider the Allen-Uzawa partial elasticities of substitution, $\sigma_{ij}$, between two inputs i and j and the price elasticities of input demand, $\varepsilon_{ij}$, which are computed from the coefficients of the TL cost function(equation 6.23) via the formulae:

$$\sigma_{ij} = \begin{cases} [\gamma_{ij} + S_i^2 - S_i]/S_iS_j, i = j \\ \cdot S_iS_j]/S_iS_j, \quad i \neq j \end{cases} \qquad \varepsilon_{ij} = \begin{cases} \sigma_{ii}S_i \quad i = j \\ \sigma_{ij}S_j \quad i = j \end{cases} \quad [6.26]$$

Global convexity of the cost function requires that all own-partial elasticities of substitution, $\sigma_{ii}$, are negative at all points. No restrictions are imposed on the cross-partial elasticities of substitution. They can be either positive, suggesting input substitutability, or negative, suggesting input complementarity.

Second, we consider the cost elasticity, total (equation 6.27) and average (equation 6.28). These cost elasticities can be used, among other things, to identify long-run economies of scale, defined as the reduction in total cost as all inputs are changed and the input prices remain constant. For instance, a value of $\varepsilon_{cy}(\varepsilon_{c/y})$ equal to unity (zero), greater (greater) than unity (zero) or less than unity (zero) implies constant, decreasing or increasing scale economies, respectively.

$$\varepsilon_{CY} = \frac{\partial \ln C}{\partial \ln Y} = \gamma_y + \gamma_{yy} \ln Y + \sum_{i=1}^{n} \gamma_{yi} \ln P_i + \gamma_{y\tau}.T \quad [6.27]$$

$$\varepsilon_{C/y} = \frac{\partial \ln(C/Y)}{\partial \ln Y} = \varepsilon_{cy} - 1 \quad [6.28]$$

Third, we consider the rate of growth of technical change (equation 6.29), the rate of growth in total factor productivity (equation 6.30) and the rate of growth of the elasticities of the average input productivities with respect to input prices (equation 6.31), output (equation 6.32) and technology (equation 6.33).

$$\varepsilon_{ct} = \frac{\partial \ln C}{\partial T} = \gamma_\tau + \gamma_{\tau\tau} T + \sum_{i=1}^{n} \gamma_{i\tau} \ln P_i + \gamma_{y\tau} \ln Y \quad [6.29]$$

$$\varepsilon_{y\tau} = -\varepsilon_{ct}.\varepsilon_{cy}^{-1} \quad [6.30]$$

$$\Pi_{ij} = \frac{\partial (Y/X_i)}{\partial \ln P_j} = -(\varepsilon_{ij}) \quad [6.31]$$

$$\Pi_{iy} = \frac{\partial (Y/X_i)}{\partial \ln Y} = 1 - c_{cy} - \gamma_{yi} S_i^{-1} \quad [6.32]$$

$$\Pi_{i\tau} = \frac{\partial (Y/X_i)}{\partial T} = 1 - \varepsilon_{ct} - \gamma_{it} S_i^{-1} \quad [6.33]$$

The time trend, T, inserted in the TLM represents: (1) a non-neutral and scale-augmenting technological change; (2) serves as a proxy of disembodied technical change; and (3) is a 'catchall' variable that captures the effects of technological factors, such as learning by doing and organizational changes[54].

The model outline above was estimated by Andrikopoulos and Vlachou [1995], via Zellners's app- roach, for the vertically integrated system of the Greek Public Power Corporation (GPPC) using time-series annual data (1970-1989) on three inputs (capital, labor and energy).The cost of capital was estimated as user cost and its price of capital services is given by the ratio of the user cost of capital services and capital stock. Three sources of energy were used as inputs: fuel oil, diesel oil, and solid fuels, mainly lignite, all expressed in terms of equivalent thermal units consumed. Labor cost has been taken as being the sum of total wages and salaries paid, including pension and benefits. In the absence of an hourly wage rate, the average annual labor payments per employee was taken as the price of labor. Finally,

54. For a further analysis of this issue see Nelson [1984] and Hulten [1992].

assuming weak separability, the output delivered in the form of high-,medium-, or low-voltage electricity, was transformed to high-voltage gigawatt-hour (GWh) equivalents, using as weights the relative marginal cost of producing and distributing medium-and low-voltage electricity to the marginal cost of high-voltage electricity. Zellner's approach was used to estimate the model. Based on the estimates of the model, not reported here, we focus our analysis on the elasticities and productivity indexes (Table 6.2), which are functions of these estimates, and discuss their implications relating to the operation of the GPPC.

Table 6.2. The TL Cost Function: Elasticities and Productivity Indexes

| | Estimates | t-Value | | Estimates | t-Value | | |
|---|---|---|---|---|---|---|---|
| $\sigma_{KK}$ | -0.441 | 7.44 | $\Pi_{KK}$ | 0.222 | 7.44 | $\varepsilon_{ct}$: | **-0.006** |
| $\sigma_{LL}$ | -0.663 | 1.81 | $\Pi_{LL}$ | 0.108 | 1.81 | SCEF | -4.270 |
| $\sigma_{EE}$ | -0.984 | 4.46 | $\Pi_{EE}$ | 0.302 | 4.46 | NTEF | 4.250 |
| $\sigma_{KL}$ | 0.154 | 1.35 | $\Pi_{KL}$ | -0.026 | 1.35 | NNTE | 0.014 |
| $\sigma_{KE}$ | 0.603 | 5.51 | $\Pi_{LE}$ | -0.030 | 0.72 | $\varepsilon_{yt}$: | **0.017** |
| $\sigma_{LE}$ | 0.093 | 0.72 | $\Pi_{KE}$ | -0.096 | 5.51 | SCEF | 12.784 |
| $\varepsilon_{KK}$ | -0.222 | 7.44 | $\Pi_{Ky}$ | 1.435 | 4.83 | NTEF | -12.725 |
| $\varepsilon_{LL}$ | -0.108 | 1.81 | $\Pi_{Ly}$ | 0.972 | 1.28 | NNTE | -0.042 |
| $\varepsilon_{EE}$ | -0.302 | 4.46 | $\Pi_{Ey}$ | -0.686 | 0.79 | $\varepsilon_{cy}$ | 0.334 |
| $\varepsilon_{KL}$ | 0.026 | 1.35 | $\Pi_{L\tau}$ | 0.002 | 0.03 | $\varepsilon_{c/y}$ | -0.666 |
| $\varepsilon_{KE}$ | 0.096 | 5.51 | $\Pi_{K\tau}$ | -0.051 | 2.99 | $\varepsilon_{cy}^{-1}$ | 2.994 |
| $\varepsilon_{LE}$ | 0.030 | 0.72 | $\Pi_{E\tau}$ | 0.064 | 1.29 | ... | ... |

SCEF=scale effect. NTEF=neutral technological effect. and NNTE=nonneutral technological effect.

**Technological and Productivity Indexes**, The total factor productivity, $c_{yt}$, calculated by using the mean values of the variables involved, grew at an average annual rate of 0.017 percent over the period 1970-1989.The effect of neutral technological change is negative and statistically significant (NTEF= -12.725%, t=3.88), while the effect of non-neutral technological change is negative but insignificant (NNTE= -0.042%, t=0.09). On the other hand, the scale effect is positive and strong enough (SCEF=12.784, t=3.85) to offset the negative effects of technological change. Thus, both technological change and scale economies play a significant role in determining the overall growth rate of total factor productivity, scale economies being, on the average, the dominant one. The rate of technical change has been found negative ($\varepsilon_{ct}$,=-0.006) which is due to the fact that the SCEF over offsets the positive NTEF and NNTE. The rates of growth in the average input productivities, reported also in Table 6.2, indicate that: (1) An increase in the price of an input i will increase the rate of change of its own productivity and an increase in the price of the ith input will increase the productivity of the jth input; (2) an increase in output will increase the productivity of capital and labor and reduce the productivity of capital; and (3) technology affects positively the productivity of labor and energy and negatively the productivity of capital.

**Elasticities-Economies of Scale**, Table 6.2 reports the partial elasticities of substitution, the price elasticities of input demands, and the cost elasticities, total and average. These elasticities indicate that: First, the own price elasticities of substitution, $\sigma_{ii}$, are all negative and statistically significant implying that: (a) the postulate of cost-minimizing factor demand theory are satisfied; and (2) provides evidence that the CPPC, a public firm, does in fact behave as a cost minimizer. Second, pairwise substitutability prevails and the price demands are price inelastic. Third, scale economics are present in the GPPC. This is indicated by the fact that: (a) the total cost elasticity is positive, implying negative cost elasticity, and a dual return to scale, $\varepsilon_{cy}^{-1}$ , greater than one.

Based on the above basic findings, we can conclude that the GPPC, being publicly owned and operated, is relatively efficient and exhibits economies of scale that contribute the most to the rate of growth in total factor productivity. This suggests that policy makers who propose privatization and vertical divestiture of the GPPC or other similar public enterprises have to provide convincing arguments that policies of this nature are really going to improve the efficiency of the industry[55].

The **TL Utility Function**. The TLM derived from the translog utility function and mainly applied in the theory of consumer demand is summarized below: Application of Roy's identity on [6.34] or [6.35] gives the demand functions, equation [6.36] and [6.37]. The elasticities are given in equation [6.38][56].

$$\ln U(X) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln X_i + 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \ln X_i \ln X_j \qquad [6.34]$$

$$-\ln V\left(\frac{P}{M}\right) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln\left(\frac{P_i}{M}\right) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \ln\left(\frac{P_i}{M}\right) \ln\left(\frac{P_j}{M}\right) \qquad [6.35]$$

55. For an application of the TL cost function in the transportation sector, see Andrikopoulos and Loizides [1998].

56. Variants of the TLM include: (1) the homogeneous translog demand system; (2) the linear homogeneous translog system; (3) the Geary-Stone demand system; and (4) the linear homogeneous demand system. These variants are obtained either by imposing certain a priori restrictions on the basic translog utility function as a part of the maintained hypothesis or by empirically testing the theoretical properties of the classical demand theory. For an analytical treatment of these properties together with the statistical tests and methods of estimation, see Andrikopoulos and Brox [1997].

$$S_i = (X_i P_i / M) = \left[ \alpha_i + \sum_{i=1}^{n} \beta_{ij} \ln\left(\frac{P_i}{M}\right) \right] \div \left[ \sum_{i=1}^{n} \alpha_i + \sum_{i-1}^{n} \sum_{j=1}^{n} \beta_{ij} \left(\frac{P_i}{M}\right) \right] \quad [6.36]$$

$$X_i = \left(\frac{M}{P_i}\right) \left[ \alpha_i + \sum_{i=1}^{n} \beta_{ij} \ln\left(\frac{P_i}{M}\right) \right] \div \left[ \sum_{i=1}^{n} \alpha_i + \sum_{i-1}^{n} \sum_{j=1}^{n} \beta_{ij} \left(\frac{P_i}{M}\right) \right] \quad [6.37]$$

$$\varepsilon_{ij} = \begin{cases} -1 + (\partial \ln S_i)/\partial \ln M, \quad \varepsilon_{iy} = 1 + (\partial \ln S_i / \partial \ln M) \\ \partial \ln S_i / \partial \ln M \text{ . i=1..... n} \end{cases} \quad [6.38]$$

**The AIDS.** The AIDS, developed by Deaton and Muellbauer [1980] assumes a specific class of preferences, known as price independent generalized logarithmic linearity (PIGLOG), which: (1) permits exact aggregation over consumers and commodities, and (2) is presented via the cost or expenditure function which defines the minimum expenditure necessary to attain a specific level of utility at given prices. This cost function is given by [6.39].

$$\ln C(U,P) = (1-U)\ln\{A(P)\} + U \ln\{B(P)\} . \quad 0 \le U \le 1 \quad [6.39]$$

In order to make [6.39] operational, the specification of the A{P} and B{P} functions becomes necessary. The criteria used for the choice of the functional forms of A{P} and B{P}include the requirements that: (1) they must lead to a system of demand functions with the desirable properties; (2) the resulting cost function(equation [6.38]) possess enough parameters to be regarded as a flexible functional form; and (3) they be valid presentation of consumers' preferences. Deaton and Muellbauer specify these functions as:

$$\ln\{A(P)\} = \alpha_0 + \sum_{j=1}^{n} \alpha_i \ln P_i + 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln P_i \ \ln P_j \quad [6.40]$$

$$\ln\{B(P)\} = \ln\{A(P)\} + \beta_0 \prod_{i=1}^{n} P_i^{\beta_1} \quad [6.41]$$

By substituting [6.40] and [6.41] into [6.39], rearranging terms, assuming equilibrium, lnC(U.P)=M, and applying Shephard lemma, the demand functions and the implying elasticities are given by expressions [6.42], [6.43], and [6.44], respectively.

$$S_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln P_j + \beta_i \ln\left(\frac{M}{P}\right), \quad \sum_{i=1}^{n} \beta_i = 0 \text{ . i=1,..., n} \qquad [6.42]$$

$$\sigma_{ij} = \begin{cases} \gamma_{ij} + S_i^2 - S_i]/S_i S_j, i = j \\ \gamma_{ij} + S_i S_j]/S_i S_j, \quad i \neq j \end{cases} \qquad [6.43]$$

$$\varepsilon_{ij} = \begin{cases} \sigma_{ii} S_i \quad i = j, \ \varepsilon_{im} = 1 + (\beta_i)/S_i \\ \sigma_{ij} S_j \quad i \neq j, \text{ i. j=1,....n} \end{cases} \qquad [6.44]$$

where $\sigma_{ij}$= partial elasticities of substitution, $\varepsilon_{ij}$=price elasticities and $\varepsilon_{im}$=income elasticities. The system of the demand functions satisfy the adding-up, the homogeneity, and the Slutsky symmetry and negativity conditions. On the assumption that P in [6.42] is known or can be calculated with some accuracy, Zellner's approach can be used to estimate the complete system of demand equations.

**The GL.** The GL model is derived from the generalized Leontief Function. given by [6.45]. Assuming constant returns to scale and applied Shephard's Lemma in [6.44], the demand functions and the implying elasticities are given by equations [6.46]-[6.49][57].

$$C(P_i, P_j, Y) = Y \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \left(P_i P_j\right)^{1/2} \quad \text{I,.., n} \qquad [6.45]$$

$$X_i = \partial C / \partial P_i = Y\{\sum_{i=1}^{n} \gamma_{ji} \left(P_j P_i\right)^{1/2}\} \qquad [6.46]$$

$$\alpha_i = X_i / Y = \sum_{i=1}^{n} \gamma_{ji} \left(P_j P_i\right)^{1/2}\} \qquad [6.47]$$

$$\sigma_{ij} = \begin{cases} (1/2)\{C\gamma_{ij} \left(P_i P_j\right)^{-1/2}\} \div Y\beta_i \beta_j \\ -(1/2)\{C\gamma_{ij} P_j^{1/2} P_i^{-3/2}\} \div Y\beta_i^2 \end{cases} \qquad [6.48]$$

$$\varepsilon_{ij} = \begin{cases} \sigma_i S_i = (1/2)\{\gamma_{ij} \left(P_i / P_j\right)^{-1/2}\} \quad i \neq j \\ \sigma_i S_i = (1/2)\{\sum_{j=1}^{n} \gamma_{ij} \left(P_i / P_j\right)^{-1/2}\} \quad i = j \end{cases} \qquad [6.49]$$

where $P_i$=factor prices. Y=output. The SURE method is used to estimate the demand functions of the inputs concerned.

57. See Berndt [1991].

## 7. Concluding Remarks

The purpose of this paper was not to introduce new materials in econometrics. It was rather directed to bringing up a number of issues, problems, and methods of estimations of econometric models extensively applied in empirical research. Particular attention was paid to time-series models, with the emphasis on financial econometrics. Specification of spatial econometric models, simultaneous equation systems with empirical applications, and limited dependent variable models were also reviewed in this partial survey.

**References**

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Principle". In: B.N. Petrov and F. Cszaki, eds., Second International Symposium on Information Theory, Akademiai Kiado, Budapest, 267-281.

Allen, R.G.D. (1938), Mathematical Analysis for Economist, (London: Macmillan).

Anderson, T.W. and Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics,* 20: 46-63.

Anderson, H.M., K. Nam and F. Vahid (1999), "Asymmetric Nonlinear Smooth Transition GARCH Models". In: Nonlinear Time Series Analysis of Economics and Financial Data, Boston: Kluwer, 191-207.

Andrews, D.W.K., Moreira M.J, et.al. (2007), "Performance of Conditional Wald Tests in IV Regression with Weak Instruments", *Journal of Econometrics,* 139: 116-132.

Andrews, D.W.K., Moreira. M.J., and Stock. J.H. (2006), "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression", *Econometrica,* 74: 715-752.

Andrews, D.W.K. and Stock J.H. (2005), "Inference with Weak Instruments", Paper Prepared for an Invited Session at the 2005 Econometric Society World Congress in London.

Andrews, D.W.K. and Stock J.H. (2005), "Inference with Weak Instruments," Cowles Foundation Discussion Paper.

Andrikopoulos, A.A. and J.A Brox (1986), "The Demand for Deposits and Risk Sensitivity: The Case for Greece, 1955-1980", *Empirical Economics,* 17: 309-328.

Andrikopoulos, A.A., and J.A Brox (1997), Demand Analysis: Theory and Practice (Waterloo, Ontario, Canada: North Waterloo Academic Press).

Andrikopoulos, A.A. and J.A Brox (1998), Demand Analysis: Theory and Practice (Waterloo, Ontario, Canada: North Waterloo Academic Press).

Andrikopoulos, A.A. and J. Loisides (1998), "Cost Structure and Productivity Growth in European Railways Systems", *Applied Economics,* 30(12): 1625-1639.

Andrikopoulos, A.A. and A. Vlachou (1995), "The Structure and Efficiency of the Publicly Owned Electric Power Industry", *Journal of Energy and Development,* 19(1): 57-79.

Angrist, I. and Krueger, A.B. (1999), "Jackknife Instrumental Variables Estimation", *Journal of Applied Econometrics,* 14: 57-67.

Angrist, J.D. and Pischke J.S. (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics", *Journal of Economic Perspectives,* 24(2): 3-30.

Anselin, L. (1988a), "Model Validation in Spatial Econometrics: A Review and Evaluation of Alternative Approaches", *International Regional Science Review,* 11: 279-316.

Anselin, L. (1988b), Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, Dordrecht, Netherlands

Anselin, L. (2002), "Under the Hood. Issues in the Specification and Interpretation of Spatial Regression Models", *Agricultural Economics,* 27(3): 247-267.

Anselin, L. (2003a), "Spatial Externalities", *International Regional Science Review,* 26(2): 147-152.

Anselin, L. (2003b), "Spatial Externalities, Spatial Multipliers and Spatial Economics", *International Regional Science Review,* 26: 153-166.

Anselin, L. (2005), "Special Statistical Modeling in a GIS Environment". In: Maguire DJ, Batty M, Goodchild MF (eds.), GIS, Special Analysis and Modeling, ESRI Press, Redlands, CA.

Anselin, L. (2010), "Thirty years of spatial econometrics" *Papers in Regional Sciences,* 89: 3-25.

Anselin, L. and Bera A. (1998), "Special Dependence in Linear Regression Models with an Introduction to Spatial Economics". In: Ullah A, Giles D.E. (eds.), Handbook of Applied Economic Statistics, Marcel Dekker, New York.

Baillie, R.T. and T. Bollerslev (1996), "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates", *Review of Economic Studies,* 58, 565-585.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics,* 74: 3-30.

Baltagi, B.H. (2005), Econometric Analysis of Panel Data (Third Edition), John Wiley & Sons.

Bartlett, M.S. (1949), "Fitting of Straight Line when Both Variable are Subject to Error", *Biometrics,* pp. 207-212.

Basu, S. (1975), "The Information Content of Price-Earnings, Ratios", *Financial Management,* 4(2): 53-64.

Basu, S. (1977), "The Investment Performance of Common Stocks in relation to their Price-Earnings Ratios", *The Journal of Finance,* 32(3): 663-82

Bekker, P.A. (1994), "Alternative Approximations to the Distribution of Instrumental Variable Estimators", *Econometrica,* 62: 657-681.

Bell, K.P. and Bocksteael, N.E. (2002), "Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data", *The Review of Economic and Statistics,* 82(2): 72-82.

Belsley, D.A., E. Kuh and R.E. Welsh (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, (New York: John Wiley & Sons).

Berenblutt, I.I. and G.I Webb (1973), "A New Test for Autocorrelated Errors in the Linear Regression Model", *Journal of the Royal Statistical Society,* Series B, 35(1): 35-50.

Berdnt, E.R. (1991), The Practice of Econometrics: Classic and Contemporary, (Addison-Wesley).

Berger, J.O, Liseo B. and Wolpert R.L. (1999), "Integrated Likelihood Methods for Eliminating Nuisance Parameters", *Statistical Science,* 14(1): 1-28.

Bernardo, J.M., J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.)' (1999), Bayesian statistics Oxford Science Publications, Oxford.

Black, F. (1976), "The Pricing of Commodity Contracts", *Journal of Financial Economics,* 3, 167-179.

Blommestein, H. (1983), "Specification and Estimation of Spatial Econometric Models: A Discussion of Alternative Strategies for Spatial Economic Modeling", *Regional Science and Urban Economics,* 13: 250-271.

Blommestein, H. (1985), "Specification of Circular Routs in Spatial Dynamic Regression Equations", *Regional Science and Urban Economics,* 15:121-130.

Breitung, J. and W. Meyer (1994), "Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?", *Applied Economics,* 26: 353-361.

Boldin, M.D. (1996), "A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle", *Studies in Non linear Dynamics and Econometrics,* 1: 35-46.

Bollerslev, T. (1986),"Generalized Autoregressive Conditional Heteroscedasticity" *Journal of Econometrics,* 31, 307-27.

Bollerslev, T. and H.O. Vikkelsen (1996), "Modeling and Pricing Long Memory in Stock Market Volatility", *Journal of Econometrics,* 73: 151-84.

Box, G.E.P. and G.M. Jenkins (1970), Time Series Analysis: Forecasting and Control, San Francisco, Holden-Day.

Box, G.E.P. and G.M. Jenkins (1976), Time Series Analysis: Forecasting and Control, revised edition, Holden-Day.

Breusch, T. and A. Pagan (1980), "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica,* 47: 1287-1294

Breusch, T. (1978), "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economic Papers,* 17: 334-355.

Brock, W., Dechert, W.D. and Scheinkman, I. (1987), "A Test for Independence based on the Correlation Dimension", Working Paper, Department of Economics, University of Wisconsin, Madison.

Buse, A. (1992), "The Bias of Instrumental Variable Estimators", *Econometrica,* 60: 173-180.

Campbell, E.P. (2004), "Bayesian Selection of Threshold Autoregressive Models", *Journal of Time Series Analysis,* 25(4): 467-482.

Case, A.C. (1991), "Spatial Patterns in Housing Demand", *Econometrica,* 59: 953-965.

Case, A.C. (1992), "Neighborhood Influence and Technical Change", *Regional Science and Urban Economics,* 22: 491-508.

Caves, D.W. and L.R. Christensen, (1980), "Global Properties of Flexible Functional Forms", *American Economic Review,* 70(3): 422-432.

Chan, K.S. and H.Tong (1986), "On Estimating Thresholds in Autoregressive Models", *Journal of Time Series Analysis,* 7: 179-190.

Christensen, L.R., D.W. Jorgenson and L.J. Lau (1973), "Transcendental Logarithmic Production Functions Frontiers", *Review of Economics and Statistics,* 55(1): 28-45.

Christensen, L.R., D.W. Jorgenson and L.J. Lau (1975), "Transcendental Logarithmic Utility Function", *American Economic Review,* 65(3): 367-383.

Choi I. and P.C.B. Phillips (1992), "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Test in Partially Identified Structural Equations", *Journal of Econometrics,* 51: 113-150.

Cliff, A. and Ord, J.K. (1973), Spatial Autocorrelation, Pion, London, UK.

Cliff, A. and Ord, J.K. (1972), "Testing for Spatial Autocorrelation among Regression Residuals", *Geographical Analysis,* 4: 267-284.

Cliff, A. and Ord, J.K. (1981), Spatial Processes: Models and Applications, Pion, London, UK.

Cochrane, D. and G.H. Orccutt, (1949), "Application of Least Squares Regressions to Relationships Containing Autocorrelated Error Terms", *Journal of the American Statistical Association,* pp.32-61.

Conley, T.G., Topa G. (2002), "Socio-Economic Distance and Spatial Patterns in Unemployment", *Journal of Applied Econometrics,* 7:303-327.

Cox, D.R. (1975). "Partial likelihood", *Biometrika* 62: 269-76.

Deaton, A.S. (1972), "The Estimation and Testing of System of Demand Equations", *European Economic Review,* 3: 390-411.

Deaton, A.S. (1974), "The Analysis of Consumer Demand in the United Kingdom", *Econometrica,* 42: 351-367.

Deaton, A.S., and J. Muellbauer (1980), "An Almost Ideal Demand System", *American Economic Review,* 70: 524-536.

Devidson, R. and Mackinnon, J.G. (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica,* 49:781-793.

Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregresive Time Series with a Unit Root", *Journal of the American Statistical Association,* 74: 140-6.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root", *Econometrica,* 49, 1057 -72.

Diewert, W.E. (1971), "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function", *Journal of Political Economy,* 79(3): 481-507.

Diewert, W.E. (1974), "Intertemporal Consumer Theory and the Demand for Durables", *Econometrica,* 42: 497-516.

Dietz, R.D. (2002), "The Estimation of Neighborhood Effects in the Social Sciences: An Interdisciplinary Approach", *Social Science Research,* 31: 539-575.

Dijk, D., T. Terasvirta and Phillip H. Frances (2002), "Smooth Transition Autoregressive Models-A Survey of Recent Developments", *Econometric Reviews,* 21(1): 1-47.

Dufour, J.M. (2007), "Identification, Weak Instruments, and Statistical Inference in Econometrics", *Canadian Journal of Economics,* 36(4): 769-808.

Durbin, J. (1954), "Errors in Variables", *Review of International Statistical Institute,* pp. 23-32.

Durbin, J. (1960), "Estimation of Parameters in Time Series Regression Models", *Journal of the Royal Statistical Society,* Series B, pp.139-153.

Durbin, J. (1970), "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors Are Lagged Dependent Variables", *Econometrica,* pp. 410-421.

Durbin, J. and G.S. Watson (1950), "Testing for Serial Correlation in Least-Squares Regression", *Biometrika,* 37: 409-428.

Durbin J. and G.S. Watson (1951), "Testing for Serial Correlation in Least-Squares Regression", *Biometrika,* 38: 159-171.

Elhorst, J.P. (2009), "Spatial Panel Data Methods".in: Mandfred M. Fischer and Arthur Getis (eds.), Handbook of Applied Spatial Analysis, Berlin: Springer.

Enders, W. (1995), Applied Econometric Time Series, John Wiley & Sons, Inc.

Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica,* 50: 987-1007.

Engle, R.F, and Bollerslev, T. (1986), "Modelling the Persistence of Conditional Variances", *Econometric Reviews,* 5: 1-50.

Engle, R.F. and Granger, C.W.J. (1987), "Co-integration and error-correction: Representation, estimation and testing", *Econometrica,* Vol 55, No 2, 251-276.

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

Engle, R.F. and V.K.Ng. (1993), "Measuring and Testing the Impact of News on Volatility", *Journal of Finance,* 48: 1749-1778.

Fernandez-Villaverde, J. and J. Rubio-Ramirez (2005), "Estimating Dynamic Equilibrium Economics: Linear Versus Nonlinear Likelihood", *Journal of Applied Econometrics,* 20: 891-910.

Fisher, R.A. (1925), Statistical Methods for Research Workers, Edinburg: Oliver and Boyed.

Fisher, F.M. (1966) The Identification Problem in Econometrics, New York: McGraw-Hill.

Formari, F. and A. Mele (1996), "Modeling the Changing Asymmetry of Conditional Variances", *Economics Letters,* 50: 197-203.

Formari, F. and A. Mele (1997), "Sign- and Volatility-Switching ARCH Models: Theory and Application to International Stock Markets", *Journal of Applied Econometrics,* 12: 49-65.

Frances, P.H., and D. van Dijk (2004), Non-Linear Time Series Models in Empirical Finance, Cambridge University Press.

Frisch, R. (1933), Pitfalls in the Statistical Construction of Demand and Supply Curves, Leipzig: Hans Buske Verlag.

Frisch, R. (1934), Statistical Confluence Analysis by Means of Complete Regression Systems, Oslo: University Institute of Economics.

Fuller, W.A. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica,* 45: 939-953

Gamaletsos, T. (1973), "Further Analysis of Cross-Country Comparison of Consumer Patterns", *European Economic Review,* 4: 1-20.

Geweke, Horowitz and Pesaran (2007), "Econometrics: A Bird's Eye View, The New Palgrave Dictionary, Second Edition.

Glejser, H. (1969), "A New Test for Heteroscedasticity", *Journal of the American Statistical Association,* 64: 316-323.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance,* 48: 1779-801.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica,* 37: 424-438.

Granger, C.W.J. and N.R. Swanson (1997), "An Introduction to Stochastic Unit-Root Processes", *Journal of Econometrics,* 80: 35-62.

Greene, W.H. (2003), Econometric Analysis, Prentice Hall.

Goldfeld, E. and Quandt, R. (1965), "Some Tests for Homoskedasticity", *Journal of the American Statistical Association,* 60: 539-547.

Goldfrey, L.G. (1978), "Testing Against General Autoregressive and Moving Average Models When the Regression Includes Lagged Dependent Variables", *Econometrica,* 46: 1293-1302.

Goldfrey, L.G. (1978), "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables", *Econometrica,* 46: 1303-1310.

Gonzalez-Riversa, G. (1998), Smooth Transition GARCH Models, *Studies in Nonlinear Dynamics and Econometrics,* 3: 61-78.

Gujarati, D.N. (1988), Basic Econometrics (McGraw-Hill).

Haavelmo, T. (1944), "The probability Approach in Econometrics", *Econometrica,* 12: 1-118.

Hagerud, G.E. (1997), A New Non-linear GARCH Model, PhD Thesis, IFE, Stockholm School of Economics.

Hahn, J. and J. Hausman (2002), "A New Specification Test for the Validity of Instrumental Variables", *Econometrica,* 70: 163-189.

Hahn, J. and J. Hausman (2003), "Weak Instruments: Diagnosis and Cures in Empirical Economics", *American Economic Review, Papers and Proceedings,* 93: 118-125.

Hall, A. G. Rudebusch, and D. Wilcx (1996), "Judging. Instrument Relevance in Instrumental Variables Estimation", *International Economic Review,* 37: 283-298.

Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series Subject to Changes in Regime", *Econometrica,* 57: 116-126.

Hamilton, J.D. (1994), Time Series Analysis: Princeton University Press.

Hansen, B.E. (1997), "Inference in TAR Models", *Studies in Non-linear Dynamics and Econometrics,* 2: 1-14.

Hansen, B.E. (1999), "Testing for Linearity", *Journal of Economic Surveys,* 13(5): 551-576.

Harris, R.D.F. and E. Tzavalis (1999), "Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed", *Journal of Econometrics,* 91: 201-226.

Hausman, J.A. (1978), "Specification Tests in Econometrics", *Econometrica,* 46(6): 1251-1271.

Hendry, D.F. (1979), "Predictive Failure and Econometric Modelling in Macroeconomics: The Transactions Demand for Money", in Paul Ormerod (ed.), Economic Modelling (London: Heinemann).

Hendry, D.F. (1980), "Econometrics-Alchemy or Science?", *Econometrica,* 47: 387-406.

Hendry, D.F. (1983), "Econometric Modeling: The Consumption Function in Retrospect", *Scottish Journal of Political Economy,* 30(3): 193-220.

Hendry, D.F. and J.F. Richard (1983), "The Econometric Analysis of Economic Time Series", *International Statistical Review,* 51: 3-33.

Hildreth, G. and J.Y. Lu (1960), "Demand Relations with Autocorrelated Disturbances", Michigan State University, Agricultural Experiment Station, Tech. Bull, 276.

Hillier, G.H. (1990), "On the Normalization of Structural Equations: Properties of Direction Estimators", *Econometrica,* 58, 1181-1194.

Hulten, C.H. (1992), "Growth Accounting when Technical Change is Embodied in Capital", *American Economic Review,* Review 82: 964-980.

Hsio, C. (2003), Analysis of Panel Data, Cambridge University Press.

Im, K.S. M.H. Pesaran and Y. Shin (2003), "Testing for Unit Roots in Heterogeneous Panels", *Journal of Econometrics,* 115:53-74.

Julie Le Gallo and Hubert Jayet (2008), "Spatial Panel Econometrics". In: Laslo Matryas and Patrick Sevestre (eds.), The Econometrics of Panel Data, New York, N.Y., Springer, 625-659.

Judge, *et al.* (1985), Introduction to the Theory and Practice of Econometrics (John Wiley & Sons).

Johnston, J. (1984), Econometric Methods (New York: McGraw-Hill).

Intrligator, M.D. (1978), Econometric Models, Techniques and Applications (Engelewood Cliffs: Prentice Hall).

Kao, C. (1999) "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data", *Journal of Econometrics,* 90: 1-44.

Kapetanios, G. (2000), "Small Sample Properties of the Conditional Least Squares in SETAR Models", *Economic Letters,* 69: 267-276.

Kapoor, M., Harry H. Kelejian, and Ingmar R. Prucha (2007) "Panel Data Models with Spatially Autocorrelated Error Components", *Journal of Econometrics,* 1: 97-130.

Kelejian, H.H. and Robinson, D.P. (1992) "Spatial Autocorrelation: A New Computationally Simple Test with an Application to per Capita Country Police Expenditures", *Regional Science and Urban Economics,* 22: 317-333.

Kelejian, H.H. and Prucha, I. (1999), "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model", *International Economic Review,* 40: 509-533.

Kelejian, H.H. and Prucha, I. (2002), "2SLS and OLS in a Spatial Autoregressive Model with equal Spatial Weights", *Regional Science and Urban Economics,* 32(6): 691-707.

Kleibergen, F. (2002), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica,* 70:1781-1803.

Kleibergen, F. and Mavorides (2009), "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics,* 27(3), 293-311.

Klein, L.R. (1950a), Economic Fluctuations in the United States, 1921-1941, Clowles Commission Monograph 11, New York: Wiley.

Klein, L.R. (1950b) The Efficiency of Estimation in Econometric Models. In: Essays in Economics and Econometrics, R.W. Pfouts, pp. 216.32. Chapel Hill: University.

Klein, LR. (1962), An Introduction to Econometrics, Prentice-Hall.

Klein, L.R. and H. Rubin (1947-48), "A Constant-Utility Index of the Cost of Living", *Review of Economic Studies,* 15: 84-87.

Klevmarken, N.A. (1989), "Panel Studies: What Can we Learn from Them?, Introduction", *European Economic Review,* 33:523-529.

Koopmans, T.C. (1937), Linear Regression Analysis of Economic Time Series, Haarlem: De Erven F.Bohn for the Netherlands Economic Institute.

Koopmans, T.C., H. Rubin and R.B. Leipnik (1950), Measuring the Equation Systems of Dynamic Economics. In: Statistical Inference in Dynamic Economic Models, (ed.) T.C. Koopmans, Cowles Commission Monograph, No.10, New York: John Wiley.

Krager, H. and P. Kugler (1993). "Nonlinear in Foreign Exchange Markets: A Different Prospective", *Journal of International Money and Finance,* 12, 195-208.

Kyriazidou, E. (1997), "Estimation of Panel Sample Selection Model", *Econometrica,* 65: 1335-1364.

Kyriazidou, E. (2001), "Estimation of Dynamic Panel Sample Selection Models", *Review of Economic Studies,* 68: 543-572.

Larson, R., J. Lyhagen and Lohgren (2001), "Likelihood-Based Cointegration Tests in Heterogeneous Panels", *Econometrics Journals,* 4: 109-142.

Leamer, E.E. (1983). "Lets Take the Con out of Econometrics", *American Economic Review,* vol. 73, pp. 31-43.

Leamer, E.E. (2010), "Tantalus on the Road to Asymptopia", *Journal of Economic Perspectives,* 24(2): 31-46.

Lee, L.F. (2002), "Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models", *Econometric Theory,* 18(2): 252-277.

LeSage, J.P., Pace R.K. (2008) "Spatial Econometric Modeling of Origin-Destination Flows", *Journal of Regional Science,* 48: 941-967.

LeSage, J.P., Pace R.K. (2009) Introduction to Spatial Econometrics, CRC Press, Boca Raton, FL.

Levin, A., C.F. Lin and C. Chu (2002) "Unit Root Test in Panel Data: Asymptotic and Finite Sample Properies", *Journal of Econometrics,* 108: 1-25.

Liu, T.C. (1960), "Underidentification, Structural Estimation and Forecasting", *Econometrica,* 28: 855 865.

Ljung, G.M, and G.E.P. Box (1978), "On a Measure of Lack of a Fit in Time Series Models", *Biometrica,* 65: 297-303.

Lluch, C. (1973), "The Extended Linear Expenditure System", *European Economic Review,* 2: 1-31.

Lluch, C and R. Williams (1975), "Consumer Demand Systems and Aggregate Consumption in the U.S.A.: An Application of the Extended Linear Expenditure System", *Canadian Journal of Economics,* 8: 275-303.

Lucas, R.E. (1976), "Econometric Policy Evaluation: A Critique". In: The Phillips Curve and Labor Markets, (ed.) K. Brunner and A.M. Meltzer, Carnegie-Rochester Conference on Public Policy, Vol.1, Amsterdam: North-Holland, 19-46.

Lutkepohl, H. (1991), Introduction to Multiple Time Series, New York: Springer.

Maddala, G.S. (1992), Introductions to Econometrics (Prentice Hall).

Maddala, G.S. and Kim, L.M. (1998), Unit Roots Cointegration and Structural Change, Cambridge University Press, New York.

Magnalinos, M.A. (1990), "The Classical Principles of Testing Using Instrumental Variable Estimates" *Journal of Econometrics,* 44: 241-279.

Magnalinos, M.A. (1994): "Testing Instrumental Admissibility: Some Refined Asymptotic Results", *Econometrica,* 373-404.

Malinvaud, E. (1980), Statistical Methods in Econometrics, Amsterdam: North-Holland Publishing Co.

Marlano, R.S. (1982), "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous Equation Case", *International Economic Review,* 23: 503-534.

McCoskey, S. and C. Kao (1988), “A residual-Based Test of the Null of Cointegration in Panel Data”, *Econometric Reviews,* 17: 57-84.

McCullock, R.E. and Tsay, R.S. (1993), “Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series”, *Journal of American Statistical Association,* 88: 968-978.

Melino, A and S.M. Turnbull (1990), “Pricing Foreign Currency Options with Stochastic Volatility”, *Journal of Econometrics,* 45: 239-265.

Moran, P. (1948), “The Interpretation of Statistical Maps”, *Biometrica,* 35: 255-260.

Moran, P. (1950), A Test for the Serial Dependence of Residuals”, *Biometrica,* 37: 178-181.

Moreira. M.J. (2003): “A Conditional Likelihood Ratio Test for Structural Models”, *Econometrica,* 71, No.4, 1027-1048.

Nelson, D.B. (1991),“Conditional Heteroscedasticity in Asset Returns: A New Approach”, *Econometrica,* 59: 347-370.

Nelson, R.A. (1984), “Capital Vintage, Time Trends, and the Technical Change in the Electric Power Industry”, *Southern Economic Journal,* pp. 315-332.

Nicholls, D.F. and B.G. Quinn. (1982), Random Coefficient Autoregressive Models: An Introduction, Lecture Notes in Statistics, 11, Springer-Verlag, New York.

Pedroni, F. (2002), “The European Community Household Panels: A Review”, *Empirical Economics,* 27: 63-90.

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: 597-625.

Pesaran, M.H. (2003), “A Simple Panel Unit Test in the Presence of Cross Section Dependence”, Working Paper, Trinity College, Cambridge.

Pesaran, M.H. (2004), “General Diagnostic Tests for Cross-Section Dependence in Panels”, Working Paper, Trinity College, Cambridge.

Petrov and F. Cszaki, (eds.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest, 267-281.

Phillips P.C.B. (1983): “Exact Small Sample Theory in the Simultaneous Equation Model”, in Handbook of Econometrics, Vol.I, ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland, 449-516.

Phillips P.C.B. (1989): “Partially Identified Econometric Models”, *Econometric Theory,* 5: 181-240.

Phillips, P.C.B. and P. Perron (1988), “Testing for a Unit Root in Time series Regression”, *Biometrika,* 75: 335-346.

Pindyck, R.S. and D. Rubinfeld. (1988), Econometric Models and Economic Forecasts, New York: Mcgraw-Hill.

Pollak, R.A. and T.J. Wales (1969), “Estimation of the Linear Expenditure Systems”, *Econometrica,* 37: 611-628.

Pollak, R.A. and T.J. Wales (1979), “Welfare Comparisons and Equivalent Scales”, *American Economic Review,* 69(2): 216-221.

Pollak, R.A. and T.J. Wales (1980), “Comparison of the Quadratic Expenditure Systems and the Translog Demand Systems with Alternative Specifications of Demographic Effects”.

Ramsey, J.B. (1969), “Tests for Specification Errors in Classical Linear Least Squares Regression Analysis”, *Journal of the Royal Statistical Society,* Series B, 31: 350-371.

Rotherberg, T.J. (1984) “Approximating the Distribution of Econometric Estimators and Test Statistics”, Chapter 15 in Hand book of Econometrics, Vol. II (eds. Z. Griliches and M.D. Intriligator), Amsterdam: North Holland, 881-935.

Ruppert C.D, and L.A. Stefanski. (1995), “Measurement Error in Nonlinear Models”, Monographs on Statistics and Applied Probability 63, Chapmans & Hall, NY, NY.

Saito, M. (1977), "Household Flow-of-Funds Equations: Specification and Estimation", *Journal of Money, Credit and Banking,* 90(1): 1-20.

Samuelson, P.A., T.C. Koopmans and J.R.N. Stone (1954), "Report of the Evaluative Committee for Econometrica," *Econometrica,* 22: 141-146.

Santana, E. (1995). "Quadratic ARCH Models", *Review of Economic Studies,* 62: 639-661.

Sargan, J.D. and Bhargava, A.S. (1983) "Testing Residuals from Least Squares Regression for Being Generated by the Caussian Random Walk", *Econometrica,* 51: 213-248.

Schwert, G.W. (1989) "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business of Economics and Statistics,* 7: 147-160.

Schwartz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics,* 6, 461-464.

Shea, J. (1997), "Instrumental Relevance in Multivariant Linear Models: A Simple Measure", *Review of Economics and Statistics,* 79: 348-352.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility", in O.E. Barndorff-Nielsen, D.R. Cox and D.V. Hinkley (eds), Statistical Models in Econometrics, Finance and Other Fields, London: Chapman & Hall, 1-67.

Sims., C. (1980) "Macroeconomics and Reality", *Econometrica,* 48: 1-48.

Spanos, A. (1986), Statistical Foundations of Econometric Modelling, Cambridge University Press, Cambridge,MA.

Spearman, C. (1904), "The Proof and Measumement of Association between two Things", *American Journal of Psychology,* 15: 72-101.

Staiger, D. and Stock J.H. (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica,* 65, No.3, 557-586.

Stock, H.H. and M. Yogo (2001), "Testing for Weak Instruments in Linear IV Regression", Prepared for the festschrift in honor of Thomas Rothenberg.

Stock, H.H, and M. Yogo (2005), "Testing for Weak Instruments in Linear IV Regressions", in Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.

Stock, J. H., J. Wright and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments", *Journal of Business and Economic Statistics.*

Stock, J. H. (2010), "The Other Transformation in Econometric Practice: Robust Tools for Inference", *Journal of Economic Perspectives,* 24(2): 83-94.

Stone, J.R.N. (1947), "On the Interdependence of Blocks of Transactions", Supplement of Journal of the Royal Statistical Society, Chapter VIII, Part 1, pp. 1-32.

Stone, J.R.N. (1954), "Linear Expenditure Systems and Demand Analysis: An Application to the Patterns of British Demand", *Economic Journal,* 64: 511-527.

Taylor, S.J. (1986), Modelling Financial Time Series, New York: John Wiley.

Terasvirta, T. (1994), "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models", *Journal of the American Statistical Association,* 89: 208-218.

Theil, H. (1961), Economic Forecasts and Policy, Amsterdam: North-Holland.

Theil, H. (1961), Principles of Econometrics, New York: Wiley.

Theil, H. (1965), "The Analysis of Disturbances in Regression Analysis", *Journal of the American Statistical Association,* 60: 1067-1079.

Theil, H. (1965), "The Information Approach to Demand Analysis", *Econometrica,* 33: 67-87.

Theil, H. (1968), "A Simplification of the BLUS Procedure for Analysing Regression Disturbances", *Journal of the American Statistical Association,* 63: 242-251.

Theil, H. and A.S. Goldberger (1961), "Pure and Mixed Statistical Estimation in Economics", *International Economic Review,* 2(1): 65-78.

Tintner, G. (1953), "The Definition of Econometrics," *Econometrica,* 21: 31-40.

Tobin, I. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica,* 26: 24-36.

Tong, H. (1978), "On a Threshold Model". In: C.H. Chen (ed.), Pattern Recognition and Signal Processing, Amsterdam: Sijthoff & Noordhoff, 101-141.

Tong, H. (1983), Threshold Models in Nonlinear Time Series Analysis, Lecture Notes In Statistics, Springer-Verlag, New York.

Tong, H. (1990), Non-Linear Time Series: A Dynamic System Approach, Oxford University Press.

Tong, H. and K.S. Lim (1980), "Threshold Autoregressions, Limit Cycles, and Data", *Journal of the Royal Statistical Society,* B-42, 245-292.

Tsay, R.S. (1987). "Conditional Heteroscedastic Time Series Models", *Journal of American Statistical Association,* 82, 590-604.

Tsay, R.S. (1989). "Testing and Modeling Threshold Autoregressive Processes", *Journal of the American Statistical Association,* 84, 231-240.

Tsay, R.S. (2005). Analysis of Financial Time Series, Second Edition, Wiley Series in Probability and Statistics: John Wiley & Sons.

Tzavalis, I, and R.D.F. Harris (1999), "Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed", *Journal of Econometrics* (1999), 91: 201-226.

Upton, G.J. and Fingleton, B. (1995), Spatial Data Analysis by Example. Volume I. Point Pattern and Quantitative Data, Wiley, New York.

Verbeek, M. (2005), A Guide to Modern Econometrics, John Wiley & Sons, Ltd.

Von Neumann, J. (1941), "Distribution of the Ratio of the Mean Square Successive Difference to the Variance", *Annals of Mathematical Statistics,* pp.367-395.

Wang, J. and E. Zivot (1998), "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica,* 66(6): 1389-1404.

Wegge, L. (1965), "Identifiability Criteria for Systems of Equations as a Whole", *Australian Journal of Statistics,* 7: 67-77.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica,* 48: 817-38.

Working, E.J. (1927), "What do Statistical 'Demand Curves' Show?" *Quarterly Journal of Economics,* 41: 212-235.

Zivot, E., R. Startz and C.R. Nelson (1998), "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review,* 39: 1119–1144.