Nonstationary Time-Series Modeling versus Structural Equation Modeling: With an Application to Japanese Money Demand

Cheng Hsiao and Hiroshi Fujiki

The issues of identification, estimation, and statistical inferences of nonstationary time series and simultaneous equation models are reviewed. It is shown that prior information matters and the advantage of dichotomization of the traditional autoregressive distributed lag model into the long-run equilibrium relation and the short-run dynamic adjustment process as an empirical modeling device may be exaggerated. A Japanese money demand study is used to illustrate that a direct approach yields a more stable long-run and short-run relationship and has better predictive power than the approach of letting the data determine the long-run relationship and modeling the short-run dynamics as an adjustment of the deviation from its equilibrium position.

Key words: Nonstationary time series; Structural model; Identification; Estimation; Prior information

Cheng Hsiao: Department of Economics, University of Southern California, and National Taiwan University
Hiroshi Fujiki: Financial Markets Department, Bank of Japan

This work began while the first author was visiting the Institute for Monetary and Economic Studies, Bank of Japan, and completed while he was a Kuo-Shu Liang Professor at the Department of Economics, National Taiwan University. He wishes to thank the respective institutions for their kind hospitality and research support. Both authors would like to thank Tatsuya Yonetani and a referee for helpful comments and suggestions. Partial research support by National Science Foundation grant SBR-96-19330 and NSC87-2811-H-002-0005 is also gratefully acknowledged. The views expressed in this paper are solely those of the authors and do not represent the view of the Bank of Japan.

I. Introduction

Many macro or financial time-series data are nonstationary. The major difference between the stationary and nonstationary data is that in the former case the absolute time the data are observed is irrelevant, prob $(\underset{\sim}{y}_{t+1} < \underset{\sim}{a}_1, \ldots, \underset{\sim}{y}_{t+m} < \underset{\sim}{a}_m)$ = prob $(\underset{\sim}{y}_{t+1+s} < \underset{\sim}{a}_1, \ldots, \underset{\sim}{y}_{t+m+s} < \underset{\sim}{a}_m)$ or prob $(\underset{\sim}{y}_{t+1} < \underset{\sim}{a}_1, \ldots, \underset{\sim}{y}_{t+m} < \underset{\sim}{a}_m \mid \underset{\sim}{x}_t)$ = prob $(\underset{\sim}{y}_{t+1+m} < \underset{\sim}{a}_1, \ldots, \underset{\sim}{y}_{t+m+s} < \underset{\sim}{a}_m \mid \underset{\sim}{x}_{t+m})$, and the impact of a shock in the system gradually diminishes over time, while in the latter case the absolute time of an event occurs matters and the impact of a shock does not diminish over time (although its relative impact may become smaller). The stationary process may be modeled by an autoregressive process with roots greater than one. The nonstationary process may be modeled by an autoregressive process with roots equal to one (integrated case) or less than one (explosive case). Anderson (1959) and White (1958) have derived the limiting distributions of the least squares estimators for such processes. In this paper, we shall focus on nonstationarity generated by integrated variables since this is the case for most observable economic data. However, because the limiting distribution of the conventional estimator for data that are integrated of order d, $I(d)$, $d = 1, 2, \ldots$, is nonstandard, econometricians, by and large, have been following the suggestion of Box and Jenkins (1970) to take successive differences of the variables to transform nonstationary time series into stationary time series to construct time series or econometric models. The advantage of such a procedure is that conventional $\sqrt{T}$ convergence of an estimator to its true value holds and the limiting distribution of an estimator is normal so that in formulating the Wald-type test statistic of a null hypothesis, it is asymptotically χ-square distributed under the null. The disadvantage is that differencing the data removes the information about long-run relationships among variables. It is not until Engle and Granger (1987) and Granger and Weiss (1983) propose the concept of cointegration that we see an outburst of econometric models that integrate the long-run equilibrium relations and short-run dynamics (e.g., Hendry [1993]) and statistical analysis of time-series regressions with integrated regressors (e.g., Chan and Wei [1988]; Fuller [1996]; Johansen [1988, 1991]; Park and Phillips [1988, 1989]; Phillips [1986, 1987, 1991, 1995]; Phillips and Durlauf [1986]; Phillips and Hansen [1990]; Sims, Stock, and Watson [1990]; and Tsay and Tiao [1990]).

This paper attempts to summarize the contribution of time-series literature from the Cowles Commission structural equation perspective (e.g., Hood and Koopmans [1953]). In particular, given the prevailing view that many macro or financial time series are nonstationary but certain linear combinations of nonstationary series may be stationary:

> This twofold proposition has important implications for both econometricians and economists. For the econometrician, nonstationarity invalidates many standard inference procedures, whose rationale involved the stationarity assumption. On the other hand, least squares now appears as a super-powerful estimator of cointegrating regressions. Furthermore, the tantalizing prospect is held out of letting the data determine the long-run relationships, rather than having to make strong *a priori* specifications, and dynamic relationships can be

formulated to contain an appealing error-correction term, representing the last period's deviation from the long-term equilibrium (e.g., Hendry [1993] and Hendry and Ericsson [1991]). (Johnston [1991])

While this view appears to be dominant among empirical researchers, we wish to argue that nonstationarity does not necessarily call for a different modeling strategy. In particular, we wish to clarify the following points:

[1] The relationship between the time series and dynamic simultaneous equation modeling with or without cointegration.

[2] Is the concept of identification still relevant or what should be the relevant concept of identification?

[3] Does the separation of long-run and short-run relationships require separate identification conditions (e.g., Johansen [1992], Johansen and Juselius [1994], and Hsiao [1996a, 1997a]) and estimation procedures (e.g., Engle and Granger [1987] and Johansen [1988])?

[4] Does the super-consistency result of Phillips and Durlauf (1986) and Stock (1987) render the issue of "simultaneity bias" irrelevant for models involving integrated regressors?

[5] What is the implication of the dichotomization of a set of variables into endogenous and exogenous variables when variables are integrated?

[6] What are the statistical properties of the conventional system estimators (e.g., two-stage least squares [2SLS] and three-stage least squares [3SLS]) and implications on hypothesis testing in structural equation modeling when variables are integrated and a subset of variables can be treated as strictly exogenous?

[7] What are the statistical properties of the conventional system estimators when there does not exist a subset of variables that are strictly exogenous (i.e., in a vector autoregression format)?

In Section II, we propose a basic framework linking time-series models and a structural equation model and discuss the implication of cointegration. Section III discusses a relevant concept for identification involving nonstationary data and demonstrates that the conventional rank condition is necessary and sufficient to identify an equation in a system. The interdependence between the identification of a short-run dynamic adjustment process and long-run equilibrium relation is also demonstrated. Section IV demonstrates that the least squares estimator is inconsistent if the regressors are cointegrated and are correlated with the errors of the equations. Section V argues that under the strict exogeneity assumption there is no need to devise new estimators. The conventional 2SLS and 3SLS estimators are consistent and the limiting distributions are either normal or mixed normal. Therefore, the Wald-type test statistic will be asymptotically χ-square distributed. Section VI argues that without the strict exogeneity assumption the 2SLS and 3SLS estimators remain consistent, but their limiting distributions may involve matrix unit-root distribution unless there is prior knowledge about the direction of nonstationarity. Section VII uses Japanese money demand data to compare the direct approach versus the approach of dichotomizing the behavioral relation into the long-run equilibrium and the short-run adjustment process as an empirical modeling device. The conclusion is in Section VIII.

II. Vector Autoregression and Dynamic Simultaneous Equation Model with or without Cointegration

Vector autoregression (VAR) was made a popular tool for analyzing the dynamics of economic system by Sims (1980). For instance, let $\underset{\sim}{w}_t$ be an $m \times 1$ vector of random variables. The conventional VAR specification assumes that the current value of a variable—for example, w_{1t}—is a function of its own past values, and the past values of other variables, $\underset{\sim}{w}_{t-j}$. This type of specification makes no use of prior theoretical ideas about how these variables are expected to be related and therefore cannot be used to test economic theories or interpret the data in terms of economic principles. To allow for the possibility of making use of prior theoretical ideas, we assume that $\underset{\sim}{w}_t$ has an autoregressive representation of the form

$$\begin{aligned} A(L)\underset{\sim}{w}_t &= A_0\underset{\sim}{w}_t + A_1\underset{\sim}{w}_{t-1} + A_2\underset{\sim}{w}_{t-2} + \ldots + A_p\underset{\sim}{w}_{t-p} \\ &= \underset{\sim}{\varepsilon}_t,\ t = 1, \ldots, T, \end{aligned} \tag{1}$$

where $A(L)$ is an $m \times m$ matrix polynomial in the lag operator L, $A(L) = \sum A_j L^j$, and $\underset{\sim}{\varepsilon}_t$ is an $m \times 1$ vector of independently, identically distributed random variables with mean $\underset{\sim}{0}$ and nonsingular covariance matrix $\Sigma_{\varepsilon\varepsilon}$. A_0 is assumed to be nonsingular and is not equal to an identity matrix. When $A_0 = I_m$, equation (1) is the conventional VAR used by Sims (1980) and others (e.g., Sims, Stock, and Watson [1990]). When $A_0 \neq I_m$, elements of $\underset{\sim}{w}_t$ are related contemporaneously. We shall call it the structural VAR. We assume that the roots of $|A(L)| = 0$ are either one or outside the unit circle.

Let

$$A(L) = A^*(L)\nabla + A(1)L^p, \tag{2}$$

where $A^*(L) = A_0^* + \ldots + A_{p-1}^* L^{p-1}$, $A_j^* = \sum_{l=0}^{j} A_l$, and $\nabla = (1 - L)$. We can express equation (1) in the error-correction form

$$A^*(L)\nabla\underset{\sim}{w}_t + A(1)\underset{\sim}{w}_{t-p} = \underset{\sim}{\varepsilon}_t. \tag{3}$$

Following Engle and Yoo (1989), we can factor $A(L)$ as

$$A(L) = U(L)M(L)V(L), \tag{4}$$

where the roots of $|U(L)| = 0$ and $|V(L)| = 0$ are lying outside the unit circle and $M(L)$ is a diagonal matrix with roots equal to one. We choose the normalization that $M(L)$ is diagonal and $|M(L)| = (1 - L)^d$, where $0 \leq d \leq m$. If $d = 0$, the process is stationary. If $d = m$, the process is $I(1)$ and not cointegrated. If $d = K < m$, then there are K linearly independent $I(1)$ processes and $G = m - K$ linearly independent cointegrating relations. There is no loss of generality to let

$$M = \begin{bmatrix} I_G & \underset{\sim}{0} \\ \underset{\sim}{0} & \nabla I_K \end{bmatrix}. \tag{5}$$

For ease of exposition, we shall assume all the elements of $\underset{\sim}{w}_t$ are $I(1)$. Partition $\underset{\sim}{w}_t$ into $(\underset{\sim}{w}'_{1t}, \underset{\sim}{w}'_{2t})'$; where $\underset{\sim}{w}_{1t}$ is $G \times 1$ and $\underset{\sim}{w}_{2t}$ is $K \times 1$, the conformable partition of equation (4) is

$$\begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} = \begin{bmatrix} U_{11}(L) & U_{12}(L) \\ U_{21}(L) & U_{22}(L) \end{bmatrix} \begin{bmatrix} I_G & \underset{\sim}{0} \\ \underset{\sim}{0} & \nabla I_K \end{bmatrix} \begin{bmatrix} V_{11}(L) & V_{12}(L) \\ V_{21}(L) & V_{22}(L) \end{bmatrix}. \tag{6}$$

Then

$$\begin{bmatrix} A_{11}(1) & A_{12}(1) \\ A_{21}(1) & A_{22}(1) \end{bmatrix} \begin{bmatrix} U_{11}(1) \\ U_{21}(1) \end{bmatrix} [V_{11}(1) \;\; V_{12}(1)], \tag{7}$$

where, by construction, $U_{11}(1)$ and $V_{11}(1)$ are nonsingular. Therefore, we have a structural error-correction representation

$$\begin{bmatrix} A^*_{11}(L) & A^*_{12}(L) \\ A^*_{21}(L) & A^*_{22}(L) \end{bmatrix} \begin{bmatrix} \nabla \underset{\sim}{w}_{1t} \\ \nabla \underset{\sim}{w}_{2t} \end{bmatrix} = - \begin{bmatrix} I_G \\ C \end{bmatrix} [A_{11}(1) \;\; A_{12}(1)] \begin{bmatrix} \underset{\sim}{w}_{1,t-p} \\ \underset{\sim}{w}_{2,t-p} \end{bmatrix} + \begin{bmatrix} \underset{\sim}{\varepsilon}_{1t} \\ \underset{\sim}{\varepsilon}_{2t} \end{bmatrix}, \tag{8}$$

where $C = U_{21}(1)U_{11}(1)^{-1}$. We note that the presence of the deviations from long-run equilibrium relations in the last K equations of equation (8) is a linear combination of the long-run relations implied by the first G equations of equation (1), $[A_{11}(1)\underset{\sim}{w}_{1,t-p} + A_{12}(1)\underset{\sim}{w}_{2,t-p}]$. In other words, each of the first G equations of equation (1) implies one structural (or behavioral) long-run relation, if it exists. Such a structural long-run relation is simply the sum of the corresponding current and lagged coefficients. As for the equation that only describes short-run dynamic behavior of $\nabla\underset{\sim}{w}_{2t}$ (the last K equations of equation [8]), more than one deviation from the long-run equilibrium relations, $A_{11}(1)\underset{\sim}{w}_{1,t-p} + A_{12}(1)\underset{\sim}{w}_{2,t-p}$, can appear. Furthermore, by construction, $A_{11}(1)$ is nonsingular, $[A_{11}(1)\; A_{12}(1)] = A_{11}(1)[I_G \;\; \Pi^*]$, when $\Pi^* = A_{11}(1)^{-1} A_{12}(1)$. The cointegrating relations between $\underset{\sim}{w}_1$ and $\underset{\sim}{w}_2$ can be expressed in reduced form $[I_G \;\; \text{II}^*]$ and define the long-run equilibrium relations as

$$\underset{\sim}{w}^*_1 = -\Pi^* \underset{\sim}{w}^*_2. \tag{9}$$

Equation (9) implies that each element of $\underset{\sim}{w}_1$ can be written as a function of $\underset{\sim}{w}_2$. The nonstationarity of $\underset{\sim}{w}_1$ is caused by the nonstationarity of $\underset{\sim}{w}_2$. Hence, $\underset{\sim}{w}_2$ may be viewed as the common trends of Stock and Watson (1988).

Multiplying equation (8) by A_0^{-1}, we obtain the (reduced form) Granger representation (Engle and Granger [1987])

$$\nabla \underset{\sim}{w}_t = D(L)\nabla \underset{\sim}{w}_{t-1} + \underset{\sim}{\alpha}\,[A_{11}(1)\; A_{12}(1)] \begin{bmatrix} \underset{\sim}{w}_{1,t-p} \\ \underset{\sim}{w}_{2,t-p} \end{bmatrix} + \underset{\sim}{\eta}_t, \tag{10}$$

where $D(L) = -A_0^{-1}(A_1^* + A_2^*L + \ldots + A_{p-1}^*L^{p-2})$, $\underset{\sim}{\alpha} = -A_0^{-1}(I_G \;\; C')'$ and $\underset{\sim}{\eta}_t = A_0^{-1}\underset{\sim}{\varepsilon}_t$. Furthermore, if one defines the long-run equilibrium relations in terms of the reduced form, then the adjustment to the deviations from long-run equilibrium coefficients matrix becomes $\underset{\sim}{\alpha}^* = \underset{\sim}{\alpha}A_{11}(1)$. The reduced-form representation allows each ∇w_{gt} to be a function of a number of deviations from equilibrium relations in contrast to the structural form, which implies one long-run relation per equation if such a relation exists. More than one deviation from long-run equilibrium relations can appear in a behavioral equation only if that equation describes short-run dynamics (i.e., in a nonstationary direction).

If

$$A_{21}(L) \equiv \underset{\sim}{0}, \text{ (or } A_{21}^*(L) \equiv \underset{\sim}{0}), \tag{11}$$

then $\underset{\sim}{w}_1$ does not Granger cause $\underset{\sim}{w}_2$ (Granger [1969]). Equation (6) implies that $U_{21}(L) \equiv \underset{\sim}{0}$ and $\underset{\sim}{V}_{21}(L) \equiv 0$. It follows that $\underset{\sim}{\alpha} = (I_G \;\; \underset{\sim}{0})'$. That is, if $\underset{\sim}{w}_1$ does not Granger cause $\underset{\sim}{w}_2$, there is no so-called adjustment to the deviation from the equilibrium relation in the equations describing $\nabla \underset{\sim}{w}_{2t}$ (the last K equations of equation [1]). In addition to equation (11), if

$$E(\underset{\sim}{\varepsilon}_{1t}\underset{\sim}{\varepsilon}'_{2t}) = \underset{\sim}{0}, \tag{12}$$

then $\underset{\sim}{w}_2$ is exogenous. If elements of $\underset{\sim}{w}$ are not cointegrated, $M(L) = \nabla I_m$, under equations (11) and (12), equation (3) becomes

$$A_{11}^*(L)\nabla \underset{\sim}{w}_{1t} + A_{12}^*(L)\nabla \underset{\sim}{w}_{2t} = \underset{\sim}{\varepsilon}_{1t}, \tag{13}$$

$$A_{22}^*(L)\nabla \underset{\sim}{w}_{2t} = \underset{\sim}{\varepsilon}_{2t}, \tag{14}$$

Suppose there exist common factors in $A_{11}^*(L)$ and $A_{12}^*(L)$ such that

$$[A_{11}^*(L) \; A_{12}^*(L)] = U_{11}(L)[\Gamma(L) \;\; B(L)], \tag{15}$$

then equation (13) becomes

$$\Gamma(L)\nabla \underset{\sim}{w}_{1t} + B(L)\nabla \underset{\sim}{w}_{2t} = \ominus(L)\underset{\sim}{\varepsilon}_{1t}, \tag{16}$$

where $\ominus(L) = U_{11}(L)^{-1}$. Equation (16) is the Zellner and Palm (1974) form of a dynamic simultaneous equation model that is expressed in terms of the first difference of the variables.[1] The vector $\nabla \underset{\sim}{w}_{1t}$ is treated as endogenous. The vector $\nabla \underset{\sim}{w}_{2t}$ is treated as exogenous and is assumed to be generated by equation (14). Similar factorization can represent $\nabla \underset{\sim}{w}_{2t}$ as generated by a multivariate autoregressive integrated moving average process (MARIMA).

1. It should be noted that Zellner and Palm start with an ARMA model with variables in level form. Equation (16) is only a special case of their general form as much as it is a special case of equation (1).

If $\underset{\sim}{w}_{1t}$ and $\underset{\sim}{w}_{2t}$ are cointegrated, then $A(1) \neq \underset{\sim}{0}$. Equation (3), under equations (11), (12), and (15), becomes the conventional dynamic simultaneous in level variables,

$$\Gamma(L)\underset{\sim}{w}_{1t} + B(L)\underset{\sim}{w}_{2t} = \underset{\sim}{u}_t, \tag{17}$$

where $\underset{\sim}{u}_t = \ominus(L)\underset{\sim}{\varepsilon}_{1t}$, and the exogenous variables $\underset{\sim}{w}_{2t}$ are generated by equation (14).

Let

$$\Gamma(L) = \Gamma(1)L + (1 - L)\Gamma^*(L), \tag{18}$$

and

$$B(L) = B(1)L + (1 - L)B^*(L). \tag{19}$$

Then equation (17) can be written in the error-correction form

$$\Gamma^*(L)\nabla\underset{\sim}{w}_{1t} + B^*(L)\nabla\underset{\sim}{w}_{2t} + \Gamma(1)\underset{\sim}{w}_{1,t-1} + B(1)\underset{\sim}{w}_{2,t-1} = \underset{\sim}{u}_t, \tag{20}$$

where $\Gamma(1)\underset{\sim}{w}_{1t} + B(1)\underset{\sim}{w}_{2t}$ gives the *implied long-run equilibrium* relation and $\Gamma^*(L)\nabla\underset{\sim}{w}_{1t} + B^*(L)\nabla\underset{\sim}{w}_{2t}$ gives the *implied short-run dynamics*. Equation (20) implies that if $\underset{\sim}{w}_1$ and $\underset{\sim}{w}_2$ are cointegrated, then the dynamic simultaneous equation model should be expressed in levels and the Zellner and Palm (1974) form of equation (16) is a misspecification because it has omitted the term $\Gamma(1)w_{1,t-1} + B(1)\underset{\sim}{w}_{2,t-1}$. On the other hand, if $\underset{\sim}{w}_1$ and $\underset{\sim}{w}_2$ are not cointegrated, then the dynamic simultaneous equation model should be expressed in terms of the first difference of the variables because $A_{11}(1) = \underset{\sim}{0}$ and $A_{12}(1) = \underset{\sim}{0}$ (equation [13]). The transformation of equation (13) into level variables yields

$$\tilde{\Gamma}^*(L)\underset{\sim}{w}_{1t} + \tilde{B}^*(L)\underset{\sim}{w}_{2t} = \underset{\sim}{\varepsilon}_{1t}, \tag{21}$$

but then $\tilde{\Gamma}^*(L)$ and $\tilde{B}^*(L)$ are subject to the common factor restrictions $\tilde{\Gamma}^*(1) \equiv \underset{\sim}{0}$ and $\tilde{B}^*(1) \equiv \underset{\sim}{0}$. In other words, if one takes a structural equation approach, the presence (equation [17]) or absence (equation [16]) of cointegration is preassumed from the way one writes down the simultaneous equation model.

We may summarize the main relations among various formulations of a vector time series $\underset{\sim}{w}'_t = (\underset{\sim}{w}'_{1t}, \underset{\sim}{w}'_{2t})$ as follows: if $A_0 = I_m$, we call the process (equation [1]) a reduced-form VAR. If $A_0 \neq I_m$, we call equation (1) a structural VAR. If $A_{21}(L) \equiv 0$, then $\underset{\sim}{w}_1$ does not Granger cause $\underset{\sim}{w}_2$. If $A_{21}(L) \equiv \underset{\sim}{0}$ and $E\underset{\sim}{\varepsilon}_1\underset{\sim}{\varepsilon}'_2 = \underset{\sim}{0}$, then $\underset{\sim}{w}_2$ can be treated as exogenous variables. If $\underset{\sim}{w}$ are $I(1)$ and are not cointegrated, then $A(1) = \sum A_j = \underset{\sim}{0}$ and $M(L) = \nabla I_m$. Its dynamic simultaneous equation formulation should be expressed in terms of the first difference of the variables (equations [13] or [16]). There is also no error-correction representation of the process. If $\underset{\sim}{w}_1$ and $\underset{\sim}{w}_2$ are cointegrated, then $A(1) \neq \underset{\sim}{0}$ and $M(L) = \begin{bmatrix} I_G & \underset{\sim}{0} \\ 0 & \nabla I_K \end{bmatrix}$. A cointegrated process with linearly independent $I(1)$ variables satisfying exogeneity conditions admits a dynamic simultaneous equation model with variables entering in levels (equation [16]). It also allows an error-correction representation (equation [20]). However, each behavioral equation only implies one long-run relation. More than one long-run relation can appear in an equation only when the error-correction representation is in a reduced-form representation.

III. Identification

The problem of whether it is possible to draw inferences from the probability distributions of the observed variables to an underlying theoretical structure is the concern of econometric literature on identifications (e.g., Hsiao [1983]). Traditionally, identification is approached from the assumption that the density function of random variables $\underset{\sim}{w}$ is known, but the parameter vector $\underset{\sim}{\theta}$ (assumed to belong to a compact set $N \subset R^q$) that characterizes the density function is unknown. Therefore, the problem of identification is reduced to the problem of distinguishing between parameter points (e.g., Koopmans, Rubin, and Leipnik [1950] and Rothenberg [1971]). However, in a nonstationary framework, the timing of events matters and it is hard to define the appropriate simple sufficient statistics. For instance, for a p-th order univariate autoregressive process with independently, identically distributed Gaussian errors, in a finite sample the sufficient statistics are $\sum_{t=1}^{T} w_t w_{t-j}$, $j = 0, 1, \ldots, p$. However, as $T \to \infty$, $\frac{1}{T}\sum w_t w_{t-j} \to \infty$ and $\frac{1}{T^2}\sum w_t w_{t-j}$ all converge to the same random variable if w_t is $I(1)$. Therefore, we find it more convenient to define observationally equivalent structures in terms of conditional density of $\underset{\sim}{w}_t$ given past $\underset{\sim}{w}$'s. Let $\underset{\sim}{w}^{t-} = (\underset{\sim}{w}_{t-1}, \underset{\sim}{w}_{t-2}, \ldots)$ denote the information set before time t.

DEFINITION 3.1. *Two structures, $\underset{\sim}{\theta}$ and $\bar{\underset{\sim}{\theta}}$ in $N \subset R^q$, are said to be observationally equivalent if $f(\underset{\sim}{w}_t \mid \underset{\sim}{w}^{t-}; \underset{\sim}{\theta}) = f(\underset{\sim}{w}_t \mid \underset{\sim}{w}^{t-}; \bar{\underset{\sim}{\theta}})$ for all $\underset{\sim}{w}_t$, $\underset{\sim}{w}^{t-}$, and t.*

DEFINITION 3.2. *The structure $f(\underset{\sim}{w}_t \mid \underset{\sim}{w}^{t-}; \underset{\sim}{\theta})$ is "identified" if there is no other $\bar{\underset{\sim}{\theta}}$ in N that is observationally equivalent.*

Suppose that the maximum order of the autoregressive model (equation [1]) is known to be p and $|A_0| \neq 0$, then the conditional density, $f(\underset{\sim}{w}_t \mid \underset{\sim}{w}^{t-}; \underset{\sim}{\theta})$, is completely specified by the reduced form (Hsiao [1996a]),

$$\underset{\sim}{w}_t = \sum_{j=1}^{p} \Pi_j \underset{\sim}{w}_{t-j} + \underset{\sim}{v}_t, \tag{22}$$

where $\Pi_j = A_0^{-1} A_j$, $j = 1, \ldots, p$, $\underset{\sim}{v}_t = A_0^{-1}\underset{\sim}{\varepsilon}_t$. Equation (22) is the conventional VAR used by Johansen (1988, 1991), Phillips (1995), Sims (1980), and Sims, Stock, and Watson (1990), etc. Since the reduced-form parameters can be consistently estimated by the least squares method (Hsiao [1996a]), identification conditions can then be derived from the algebraic relations between the structural form and reduced form (e.g., Fisher [1967]). Moreover, since equation (1) can be transformed into an error-correction representation,

$$\sum_{j=0}^{p-1} A_j^* \nabla \underset{\sim}{w}_{t-j} + A_p^* \underset{\sim}{w}_{t-p} = \underset{\sim}{\varepsilon}_t, \tag{23}$$

by the nonsingular transformation matrix M,

$$M = \begin{bmatrix} I_m & I_m & \cdots & I_m \\ \underset{\sim}{0} & I_m & \ddots & I_m \\ \cdots & \cdots & \cdots & \\ \underset{\sim}{0} & \underset{\sim}{0} & & I_m \end{bmatrix}, \tag{24}$$

that relates $A = [A_0, A_1, \ldots, A_p]$ to $A^* = [A_0^*, \ldots, A_p^*]$ by

$$A^* = AM. \tag{25}$$

Let $\tilde{A}_1^* = [A_0^*, \ldots, A_{p-1}^*]$; then $\tilde{A}_1^*$ and A_p^* provide the *implied* short-run dynamics and long-run relations of the system (equation [1]). Suppose the g-th equation of equation (1) satisfies the prior restrictions $\underset{\sim}{\alpha}_g' \Phi_g = \underset{\sim}{0}'$, where $\underset{\sim}{\alpha}_g'$ denotes the g-th row of A and Φ_g denotes the $(p + 1)m \times r_g$ matrix with known elements, then $\underset{\sim}{\alpha}_g' \Phi_g = \underset{\sim}{0}'$ is equivalent to $\underset{\sim}{\alpha}_g' M M^{-1} \Phi_g = \underset{\sim}{\alpha}^{*\prime} \Phi_g^* = \underset{\sim}{0}'$ where $\underset{\sim}{\alpha}^*$ denotes the g-th row of A^* and $\Phi_g^* = M^{-1}\Phi_g$ is also a known matrix. Therefore, as shown in Hsiao (1996a),

THEOREM 3.1. *(a) Suppose that the prior restrictions on the g-th equation take the form* $\underset{\sim}{\alpha}_g' \Phi_g = \underset{\sim}{0}'$, *then the g-th equation of equation (1) or equation (23) is identified if and only if*

$$rank\ (A\Phi_g) = m - 1, \tag{26}$$

or

$$rank\ (A^*\Phi_g^*) = m - 1. \tag{27}$$

(b) The structural VAR (equation [1]) is identified if and only if its structural error-correction representation (equation [23]) and its implied short-run dynamics and long-run equilibrium relations are identified.

LEMMA 3.1. *For a system involving G cointegrating relations out of* $(G + K)\, I(1)$ *variables, the identification of both the short-run dynamics and long-run relations are interrelated. More specifically, suppose the prior restrictions on the g-th short-run, long-run, and both the short-run and long-run coefficients take the form*

$$\underset{\sim}{\tilde{\alpha}}_{g1}^{*\prime} \tilde{\Phi}_{g1}^* = \underset{\sim}{0},\ \underset{\sim}{\alpha}_{gp}^{*\prime} \tilde{\Phi}_{g2}^* = \underset{\sim}{0},\ and\ (\underset{\sim}{\tilde{\alpha}}_{g1}^{*\prime}, \underset{\sim}{\alpha}_{gp}^{*\prime}) \Phi_{g3}^* = \underset{\sim}{\alpha}_g^{*\prime} \Phi_{g3}^* = \underset{\sim}{0},$$

where $\underset{\sim}{\alpha}_g^{*\prime} = (\underset{\sim}{\tilde{\alpha}}_{g1}^{*\prime}, \underset{\sim}{\alpha}_{gp}^{*\prime})$, *then*

[1] A sufficient condition to identify the g-th long-run relation out of G linearly independent cointegrating relations is that rank $([A_{11}(1), A_{12}(1)]\tilde{\Phi}_{g2}^*) = G - 1$.

(Johansen [1995], Johansen and Juselius [1994], and Pesaran and Shin [1995]).

[2] *Given the identification of the g-th long-run relation, a necessary condition to identify the corresponding short-run dynamics is that there exist at least K restrictions involving the short-run coefficients.*

[3] *A sufficient condition to identify the g-th short-run coefficients is that rank* $(\tilde{A}_1^* \tilde{\Phi}_{g1}^*) = G + K - 1$. *That is, there exist at least* $G + K - 1$ *prior restrictions on the g-th short-run coefficients. However, if the short-run coefficients are identified by* $\underset{\sim}{\alpha}_{g1}^{*\prime}\tilde{\Phi}_{g1}^* = \underset{\sim}{0}$, *so are the corresponding long-run coefficients because rank* $(\tilde{A}_1^* \tilde{\Phi}_{g1}^*) = G + K - 1$ *implies rank* $(A^*\Phi_g^*) = G + K - 1$. *In other words, rank* $(\tilde{A}_1^* \tilde{\Phi}_{g1}^*) = G + K - 1$ *is sufficient to identify both the g-th short-run and long-run coefficients.*

[4] *A necessary and sufficient condition to identify both the g-th short- and long-run relations is that rank* $(A^*\Phi_g^*) = G + K - 1$. *That is, we can have both relations identified without the existence of sufficient information to identify either the short- or the long-run relation.*

In the case of structural equation modeling, the prior restrictions that $A_{21}(L) \equiv \underset{\sim}{0}$ and $E\underset{\sim}{\varepsilon}_{1t}\underset{\sim}{\varepsilon}_{2t}' = \underset{\sim}{0}$ are imposed. Then equation (1) becomes

$$A_{11}(L)\underset{\sim}{w}_{1t} + A_{12}(L)\underset{\sim}{w}_{2t} = \underset{\sim}{\varepsilon}_{1t}, \tag{28}$$

$$A_{22}(L)\underset{\sim}{w}_{2t} = \underset{\sim}{\varepsilon}_{2t}. \tag{29}$$

Then

THEOREM 3.2. *Consider the g-th equation in the system (equation [28]). Suppose there exist prior restrictions of the form* $\underset{\sim}{\alpha}_g'\Phi_g = \underset{\sim}{0}$, *where* Φ_g *denotes a* $(p + 1)(G + K) \times r_g$ *matrix with known elements. A necessary and sufficient condition for the identification of the g-th equation is that*

$$rank\ [(A_{11,0}, \ldots, A_{11,p}, A_{12,0}, \ldots, A_{12,p})\Phi_g] = G - 1. \tag{30}$$

Suppose the k-th equation in the system (equation [29]) is subject to prior restriction of the form $\underset{\sim}{\alpha}_k'\Phi_k = 0$, *where* Φ_k *is a* $(p + 1)K \times r_k$ *matrix with known elements. A necessary and sufficient condition to identify this equation is*

$$rank\ [(A_{22,0}, A_{22,1}, \ldots, A_{22,p})\Phi_k] = K - 1. \tag{31}$$

Equation (28) can be transformed into an error-correction form:

$$\begin{aligned} &A_{11,0}^*\nabla\underset{\sim}{y}_t + \ldots + A_{11,p-1}^*\nabla\underset{\sim}{y}_{t-p+1} + A_{12,0}^*\nabla\underset{\sim}{x}_t + \ldots + A_{12,p-1}^*\nabla\underset{\sim}{x}_{t-p+1} \\ &+ A_{11,p}^*\underset{\sim}{y}_{t-p} + A_{12,p}^*\underset{\sim}{x}_t = \underset{\sim}{\varepsilon}_{1t}. \end{aligned} \tag{32}$$

Since

$$\begin{aligned} (\tilde{A}_1^*, \tilde{A}_2^*) &= [(A_{11,0}^*, A_{12,0}^*), \ldots, (A_{11,p}^*, A_{12,p}^*)] \\ &= [(A_{11,0}, A_{12,0}), \ldots, (A_{11,p}, A_{12,p})]M, \end{aligned} \tag{33}$$

where M is a nonsingular $(p + 1)(G + K) \times (p + 1)(G + K)$ matrix defined in equation (24), it follows that

COROLLARY 3.1. *Suppose the g-th equation of equation (32) satisfies* $\underset{\sim}{\alpha}_g^{*\prime}\Phi_g^* = \underset{\sim}{0}$. *Then a necessary and sufficient condition for the identification of this equation is*

$$rank\ [(\tilde{A}_1^*, \tilde{A}_2^*)\Phi_g^{*\prime}] = G - 1. \tag{34}$$

Moreover, if there exist prior restrictions to identify the g-th long-run relation such that $\underset{\sim}{\alpha}_{g2}^{*\prime}\tilde{\Phi}_{g2}^* = \underset{\sim}{0}$ *has rank* $(\tilde{A}_2^*\tilde{\Phi}_{g2}^*) = G - 1$, *then the g-th short-run relation is also identified. Similarly, if there exist prior restrictions to identify the g-th short-run dynamics, then the g-th long-run relation is also identified.*

IV. The Least Squares Estimator

Phillips and Durlauf (1986) and Stock (1987) have shown that there is no "simultaneity bias" for the least squares estimator when the regressors are $I(1)$. However, this result only holds if the regressors are not cointegrated. Due to behavioral inertia and institutional or technological rigidity, many behavioral equations are better described by an autoregressive distributed lag model (e.g., Jorgenson [1966]). A dynamic structure contains both current and lagged variables. Therefore, even though the regressors are integrated processes, the current and lagged variables are trivially cointegrated. Cointegrated regressors can be transformed into linearly independent $I(0)$ and $I(1)$ components with corresponding transformation of the parameters without changing the structure of the equation. While the coefficients of the linearly independent $I(1)$ process can be consistently estimated by the least squares method, the coefficients of the $I(0)$ component cannot be consistently estimated if the $I(0)$ components are correlated with the error of the equation. Since the parameters of interest involve the transformation of the coefficients of both the $I(0)$ and $I(1)$ components, in general, they cannot be consistently estimated by the least squares method.

To illustrate this, we assume that the g-th equation is identified. (Otherwise, there is no unique least squares estimator.) For ease of exposition, we assume that all variables are $I(1)$ and the prior restrictions are in the form of exclusion restrictions. Then the g-th equation of equation (1) can be written as

$$\underset{\sim}{w}_g = Z_g\underset{\sim}{\delta}_g + \underset{\sim}{\varepsilon}_g, \tag{35}$$

where $\underset{\sim}{w}_g$ and $\underset{\sim}{\varepsilon}_g$ denote the $T \times 1$ vectors of $(w_{g1}, \ldots, w_{gT})'$ and $(\varepsilon_{g1} \ldots, \varepsilon_{gT})$, respectively, and Z_g denotes the included current and lagged values of $\underset{\sim}{w}_t$.

Let M_g be the nonsingular transformation matrix such that $Z_g^* = Z_gM_g = (Z_{g1}^*, Z_{g2}^*)$, where Z_{g1}^* is a $T \times l$ matrix denoting the linearly independent $I(0)$ variables and Z_{g2}^* is a $T \times b$ matrix denoting the linearly independent $I(1)$ variables. Then

$$\begin{aligned} \underset{\sim}{w}_g &= Z_g M_g M_g^{-1} \underset{\sim}{\delta}_g + \underset{\sim}{\varepsilon}_g \\ &= Z_g^* \underset{\sim}{\delta}_g^* + \underset{\sim}{\varepsilon}_g, \end{aligned} \tag{36}$$

where $\underset{\sim}{\delta}_g^* = M_g^{-1} \underset{\sim}{\delta}_g = (\underset{\sim}{\delta}_{g1}^{*\prime}, \underset{\sim}{\delta}_{g2}^{*\prime})'$. The least squares estimator of $\underset{\sim}{\delta}_g$ is equal to

$$\hat{\underset{\sim}{\delta}}_{g,ls} = M_g \hat{\underset{\sim}{\delta}}_{g,ls}^*, \tag{37}$$

where $\hat{\underset{\sim}{\delta}}_{g,ls}^*$ is the least squares estimator of equation (36),

$$\begin{aligned} \hat{\underset{\sim}{\delta}}_{g,ls}^* = \begin{bmatrix} \hat{\underset{\sim}{\delta}}_{g1,ls}^* \\ \hat{\underset{\sim}{\delta}}_{g2,ls}^* \end{bmatrix} = \begin{pmatrix} \underset{\sim}{\delta}_{g1}^* \\ \underset{\sim}{\delta}_{g2}^* \end{pmatrix} + \left[\begin{pmatrix} Z_{g1}^{*\prime} Z_{g1}^* & Z_{g1}^{*\prime} Z_{g2}^* \\ Z_{g2}^{*\prime} Z_{g1}^* & Z_{g2}^{*\prime} Z_{g2}^* \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{T} I_l & \underset{\sim}{0} \\ 0 & \frac{1}{T^2} I_b \end{pmatrix}^{-1} \right] \\ \begin{pmatrix} \frac{1}{T} I_l & 0 \\ 0 & \frac{1}{T^2} I_b \end{pmatrix} \begin{pmatrix} Z_1^{*\prime} \underset{\sim}{\varepsilon}_g \\ Z_2^{*\prime} \underset{\sim}{\varepsilon}_g \end{pmatrix}. \end{aligned} \tag{38}$$

Under fairly general conditions, it can be shown that (e.g., Phillips and Durlauf [1986] and Hsiao [1997a]) $\frac{1}{T} Z_{g1}^{*\prime} Z_{g1}^* \longrightarrow M_{zg1,zg1}^*$, $\frac{1}{T} Z_{g1}^{*\prime} Z_{g2}^* \Rightarrow M_{zg1,zg2}^*$, $\frac{1}{T^2} Z_{g2}^{*\prime} Z_{g1}^* \longrightarrow 0$, $\frac{1}{T^2} Z_{g2}^{*\prime} Z_{g2}^* \Rightarrow M_{zg2,zg2}^*$ and $\frac{1}{T^2} Z_{g2}^{*\prime} \underset{\sim}{\varepsilon}_g \longrightarrow 0$ where $\longrightarrow$ and $\Rightarrow$ denote convergence in probability and in distribution of the associated probability measure, respectively, $M_{zg1,zg1}^*$ is a nonsingular constant matrix of dimension l, $M_{zg2,zg2}^*$ is a nonsingular random matrix almost surely. Therefore, $\hat{\underset{\sim}{\delta}}_{g2}^*$ converges to $\underset{\sim}{\delta}_{g2}^*$, but $\underset{\sim}{\delta}_{g1}^*$ will converge to $\underset{\sim}{\delta}_{g1}^*$ only if $\frac{1}{T} Z_1^{*\prime} \underset{\sim}{\varepsilon}_g \longrightarrow 0$. However, if Z_1^* contains contemporaneous $\nabla \underset{\sim}{w}_t$, then

$$\frac{1}{T} Z_1^{*\prime} \underset{\sim}{\varepsilon}_g \rightarrow \begin{pmatrix} (A_0^{-1} \Sigma_g)_g \\ \underset{\sim}{0} \end{pmatrix}, \tag{39}$$

where Σ_g denotes the g-th column of Σ, and $(A_0^{-1} \Sigma_g)_g$ denotes those elements of $A^{-1} \Sigma_g$ that correspond to included $\underset{\sim}{w}_{gt}$ except for $\underset{\sim}{w}_{gt}$. Therefore $\hat{\underset{\sim}{\delta}}_{g1}^*$ is an inconsistent estimator of $\underset{\sim}{\delta}_{g1}^*$.

The least squares estimator of $\underset{\sim}{\delta}_g$ (equation [35]) is equal to $\hat{\underset{\sim}{\delta}}_{g,ls} = M_g \hat{\underset{\sim}{\delta}}_{g,ls}^*$, which is a linear combination of $\hat{\underset{\sim}{\delta}}_{g1,ls}^*$ and $\hat{\underset{\sim}{\delta}}_{g2,ls}^*$. Since $\hat{\underset{\sim}{\delta}}_{g1}^*$ is inconsistent, the least squares estimator of $\underset{\sim}{\delta}_g$ is inconsistent despite the fact that the variables are $I(1)$. In other words, in a dynamic framework, the issue of "simultaneity bias" raised by the Cowles Commission remains a legitimate concern whether the regressors are $I(0)$ or $I(1)$.

V. Estimation of a Dynamic Simultaneous Equation Model

Interest in the dichotomization of the short-run dynamics and long-run equilibrium relations has led to the development of many new estimators. For instance, Engle and Granger (1987) propose a two-step estimation procedure, Johansen (1988, 1991) has proposed a maximum likelihood estimator that utilizes Anderson (1951) reduced-rank regression techniques, Phillips (1991) and Phillips and Hansen (1990) propose a fully modified least squares estimator, etc. However, because of the nonlinear nature in the dichotomization (e.g., equations [20] and [10]), many of these estimators are difficult to implement and have dubious finite sample properties. Moreover, they tend to estimate the reduced-form specification of equation (10) (i.e., $D_0 = I_m$). Hsiao (1996a,b) argued that from a structural equation perspective, there is no need to devise new estimators. A conventional 2SLS or 3SLS estimator can be implemented and possesses the optimality property in the sense of Phillips (1991) under the exogeneity assumption (equations [11] and [12]).

More specifically: first, the 2SLS estimator of equation (35) and the 3SLS estimator of equation (28) are consistent. The conventional formulae for computing the asymptotic (conditional) covariance matrices of the 2SLS and 3SLS estimators remain valid (Hsiao [1997a,b]) even though the variables may be $I(1)$.

Second, the speed of convergence of the 2SLS and 3SLS estimators to their true values depends on the nature of prior restrictions. For instance, consider the model

$$y_t = \gamma y_{t-1} + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \underset{\sim}{\varepsilon}_t, \tag{40}$$

where y, x_1, and x_2 are all $I(1)$, y, x_1, and x_2 are cointegrated but x_1 and x_2 are not cointegrated. The least squares estimator that is the maximum likelihood estimator of γ, β_1, and β_2 converges to their true values at the speed of $T^{\frac{1}{2}}$. However, if there is prior restriction that $\beta_1 + \beta_2 = c$, then the constrained least squares estimator converges to their true values at the speed of T rather than $T^{\frac{1}{2}}$(Hsiao [1997a]).

Third, the limiting distributions of the 2SLS or 3SLS estimators are either normal or mixed normal. However, the limiting distributions may be singular because the estimation errors typically consist of two components, a $T^{-1/2}$ component and a T^{-1} component. The limiting distribution of an estimator is determined by the component that has a slower rate of convergence (Hsiao [1997a,b]).

Fourth, the possible singularity of the limiting distribution does not create problems of inference because even if a particular linear combination eliminates the $T^{-1/2}$ component, there is still a T^{-1} component. When the estimation error is weighted by its covariance matrix, it will provide the right scale factor to make the limiting distribution non-degenerate. Therefore, consider the null hypothesis $P\underset{\sim}{\delta} = \underset{\sim}{c}$. The Wald test statistic

$$(P\hat{\underset{\sim}{\delta}} - \underset{\sim}{c})'[P \operatorname{cov}(\hat{\underset{\sim}{\delta}})P']^{-1}(P\hat{\underset{\sim}{\delta}} - \underset{\sim}{c}) \tag{41}$$

will be asymptotically χ-square distributed, where $\hat{\underset{\sim}{\delta}}$ is the 2SLS or 3SLS estimator of $\underset{\sim}{\delta}$ and cov $(\hat{\underset{\sim}{\delta}})$ is the conventional formula for approximating the covariance matrix of $\hat{\underset{\sim}{\delta}}$.

Fifth, if one is interested in the dichotomization of the short-run dynamics and long-run equilibrium relations, one can derive the implied short-run and long-run coefficients and their limiting distributions from the relation (equation [25]), which is simply a linear transformation of the conventional estimator of (equation [28]). They are asymptotically efficient.

In short, although the speed of convergence and the limiting distribution are of theoretical interest, for empirical structural model builders, the message is clear—in a structural approach one still needs to worry about the issues of identification and "simultaneity bias," but one need not worry about the issues of nonstationarity and cointegration. All one needs to do in structural model building is to follow the advice of the Cowles Commission.

VI. Estimation of a Structural VAR

When the exogeneity assumptions (equations [11] and [12]) are not imposed, contrary to the stationary case where predeterminedness is sufficient, the conventional structural equation estimators lose much of the appealing features (Hsiao [1996b]).

As argued in Hsiao (1996b): first, the 2SLS and 3SLS estimators remain consistent.

Second, for those parameter estimates that converge to their true values at the speed of $T^{-1/2}$, their limiting distribution is a multivariate normal, possibly singular. For those parameters that converge to their true values at the speed of T^{-1}, their limiting distribution involves (matrix) unit-root distribution.

Third, the conventional formulae of computing the covariance matrices of the 2SLS and 3SLS estimators are no longer valid. Hence, the Wald-type test statistic (equation [41]) is not asymptotically χ-square distributed under the null.

If the direction of nonstationarity is known *a priori*—for example, the last K equations—then as shown in Hsiao (1996b), we can express the last K equations in reduced form and consider the estimation of the system[2]

$$A_{11}(L)\underset{\sim}{w}_{1t} + A_{12}(L)\underset{\sim}{w}_{2t} = \underset{\sim}{\varepsilon}_{1t}, \tag{42}$$

$$\tilde{A}^{*}_{21}(L)\nabla\underset{\sim}{w}_{1t} + \tilde{A}^{*}_{22}(L)\nabla\underset{\sim}{w}_{2t} = \underset{\sim}{\varepsilon}_{2t}, \tag{43}$$

where $\tilde{A}^{*}_{22,0} = I_K$.

The statistical property of the 2SLS estimator of equation (42) will not change with this prior knowledge; however, the 3SLS estimator of equations (42) and (43) will again have the desirable properties for the coefficients of level variables, as in the case when we know certain variables are exogenous. The limiting distribution of the 2SLS and 3SLS estimators will be either normal or mixed normal. The conventional formula of computing the covariance matrix will be valid. The Wald-type test statistic (equation [41]) will again be asymptotically χ-square distributed under the null. Furthermore, the structural form coefficients of the last K equations can be derived from the relationships between the structural form and reduced form.

The difference between applying 2SLS and 3SLS to equation (1) or equations (42) and (43) is that, in the former, the unit roots are either implicitly or explicitly

2. We can eliminate the level variables from the last K equations by adding C times the first G equations of equation (8) to the last K equations, then transforming them into the reduced form (equation [43]).

estimated, while in the latter case it does not involve the estimation of unit roots. In other words, the knowledge of exogeneity or direction of nonstationarity can help eliminate the (matrix) unit-root distribution and make the Wald-type statistics legitimate tools of inference. However, since the identification conditions discussed in Section III do not require such knowledge, it is desirable to pretest for unit roots and explicitly incorporate the pretest results in the specification before implementing the 2SLS and 3SLS estimators, because the use of unit-root information is dramatic, all second-order bias effects will be removed, and the asymptotic distribution becomes symmetric.

VII. Japanese Money Demand Equation: An Illustration

It is often argued that the conventional approach to the money demand function gives unstable income elasticity of money demand. However, it is also argued that the long-run income elasticity of money demand can be recovered from the cointegrating vector of money and income. Instead, we have argued that both the long-run and short-run coefficients can be straightforwardly derived from the estimation of traditional autoregressive distributed lag form and there is no particularly compelling argument in favor of the approach of first letting the data determine the long-run relation, then formulating the dynamic relationships as an adjustment of deviation from the long-term equilibrium. In this section, we estimate the Japanese money demand equation from both approaches to illustrate our point. Seasonally adjusted real M2+CDs, GNP, and the nominal overnight collateralized call rate from 1968/III to 1993/I are used.

Let M denote log real money, Y denote log real income, R denote the interest rate, an autoregressive distributed lag form of a money demand equation is often specified:

$$M_t = C_1 + \alpha_{11}M_{t-1} + \alpha_{12}M_{t-2} + \alpha_{13}Y_t + \alpha_{14}R_t + \varepsilon_{1t}. \tag{44}$$

The augmented Dickey-Fuller (ADF) test (including six lags of differenced term) statistics are –2.68 for Y and –2.87 for M. The critical value of rejecting the null hypothesis of the existence of the unit root at the 5 percent significance level is –3.46. Therefore, we may treat Y and M as $I(1)$ variables. However, as argued in Section II, equation (44) implies that Y and M are cointegrated. The Phillips and Quliaris (1990) test statistic for cointegration between M and Y is –5.346. The critical value of the null of no cointegration at the 5 percent significance level is –3.915. In other words, the statistical test also supports the prior conjecture implied by the specification (equation [44]) that M and Y are cointegrated $I(1)$ variables.

If, as it is sometimes argued (e.g., Goldfeld and Sichel [1990]), the interest rate is set by the monetary authority exogenously and real income is predetermined within the period of analysis, then as discussed in Section V, the least squares method is optimal. The least squares estimate of equation (44) is

$$\hat{M}_t = -0.3255 + 1.6307M_{t-1} - 0.7061M_{t-2} + 0.1079Y_t - 0.0073R_t, \quad (45)$$
$$\begin{array}{lllll}(0.1809) & (0.088) & (0.0887) & (0.042) & (0.0031)\end{array}$$
Durbin-Watson = 2.1174, standard error = 0.0086.

The implied long-run income elasticity of equation (45) equals 0.1079/(1 – 1.6307 + 0.7061) = 1.431.

Given that M and Y are cointegrated, an alternative approach of deriving the estimates of equation (44) is by first estimating the long-run equilibrium relation

$$M_t = a + bY_t + v_t, \quad (46)$$

then substituting equation (46) into equation (44) to estimate the short-run dynamics. The least squares estimate of equation (46) is

$$M_t = -5.7606 + 1.5218Y_t + \hat{v}_t, \quad (47)$$
$$\begin{array}{ll}(0.179) & (0.013)\end{array}$$

where the standard errors are in parentheses. Using (1, –1.5218, 5.7606) as the cointegrating vector, we obtain the short-run dynamic adjustment equation as

$$\nabla\hat{M}_t = 0.0091 + 0.8093\nabla M_{t-1} - 0.0648\hat{v}_{t-1} - 0.0033R_t, \quad (48)$$
$$\begin{array}{llll}(0.0056) & (0.0845) & (0.0288) & (0.0031)\end{array}$$
Durbin-Watson = 2.1436, standard error = 0.0089.

If one believes that M, Y, and R are jointly determined, then one can augment equation (44) by the reduced-form specifications of ∇Y and R,

$$\begin{aligned}\nabla Y_t = c_2 &+ \alpha_{21}\nabla M_{t-1} + \alpha_{22}\nabla M_{t-2} + \alpha_{23}\nabla M_{t-3} + \alpha_{24}\nabla Y_{t-1} + \alpha_{25}\nabla Y_{t-2} \\ &+ \alpha_{26}\nabla Y_{t-3} + \alpha_{27}\nabla R_{t-1} + \alpha_{28}\nabla R_{t-2} + \alpha_{29}\nabla R_{t-3} + \varepsilon_{2t},\end{aligned} \quad (49)$$

$$\begin{aligned}R_t = c_3 &+ \alpha_{31}\nabla M_{t-1} + \alpha_{32}\nabla M_{t-2} + \alpha_{33}\nabla M_{t-3} + \alpha_{34}\nabla Y_{t-1} + \alpha_{35}\nabla Y_{t-2} \\ &+ \alpha_{36}\nabla Y_{t-3} + \alpha_{37}R_{t-1} + \alpha_{38}R_{t-2} + \alpha_{39}R_{t-3} + \varepsilon_{3t}.\end{aligned} \quad (50)$$

We note that, by Theorem 3.1, a necessary and sufficient condition for the identification of equation (44) in a system involving equations (44), (49), and (50) is that the matrix

$$\begin{bmatrix}\alpha_{22} - \alpha_{23}, -(1 + \alpha_{24}), \alpha_{24} - \alpha_{25} & \alpha_{25} - \alpha_{26} & \alpha_{26} & \alpha_{27} & \alpha_{28} & \alpha_{29} \\ \alpha_{32} - \alpha_{33}, -(1 + \alpha_{34}), \alpha_{34} - \alpha_{35} & \alpha_{35} - \alpha_{36} & \alpha_{36} & \alpha_{37} & \alpha_{38} & \alpha_{39}\end{bmatrix} \quad (51)$$

is of rank two. A necessary and sufficient condition to identify equation (49) is that the matrix

$$\begin{bmatrix} 1 & -\alpha_{14} \\ 0 & 1 \end{bmatrix} \tag{52}$$

is of rank two. A necessary and sufficient condition to identify equation (50) is that the matrix

$$\begin{bmatrix} 1 & -\alpha_{13} \\ 0 & 1 \end{bmatrix} \tag{53}$$

is of rank two.

Under the assumption that all three equations are identified, as pointed out in Section VI, equation (44) can be directly estimated by the 3SLS estimator. The limiting distribution of the 3SLS estimator of equation (44) is either asymptotically normal or mixed normal. The conventional Wald-type test statistics will be χ-square distributed. Using M_{t-1}, M_{t-2}, M_{t-3}, M_{t-4}, Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, Y_{t-4}, R_{t-1}, R_{t-2}, R_{t-3} as instruments, the 3SLS estimate of equation (44) is

$$\begin{aligned} \hat{M}_t = {} & -0.3240 + 1.6389M_{t-1} - 0.7126M_{t-2} + 0.1058Y_t - 0.0068R_t, \\ & (0.1677)\ (0.0742) \qquad (0.0732) \qquad (0.0405) \quad (0.0035) \\ & \text{Durbin-Watson} = 2.1327,\ \text{standard error} = 0.0084. \end{aligned} \tag{54}$$

The implied long-run income elasticity of equation (54) is equal to 0.1058/(1 – 1.6389 + 0.7126) = 1.436.

If M and Y are jointly dependent, although the least squares estimator of equation (46) remains consistent, it no longer possesses good statistical properties (Phillips and Durlauf [1986]). Phillips and Hansen (1990) (PH) have suggested using a fully modified procedure to remove the impact of endogenous regressors and serial correlation in the residuals.[3] The PH fully modified regression estimates of equation (46) are

$$\begin{aligned} M_t = {} & -7.1147 + 1.6177Y_t + \hat{v}_t. \\ & (0.5418)\ (0.0403) \end{aligned} \tag{55}$$

Under the assumption that Y_t and R_t are exogenous, the least squares estimate of the short-run dynamic adjustment equation using (1, –1.6177, 7.1147) as the cointegrating vector yields

3. The Phillips-Hansen (1990) fully modified least squares estimator of the model $y_t = \underset{\sim}{x}_t \underset{\sim}{\beta} + u_t$ is

$$\hat{\underset{\sim}{\beta}} = \left(\sum_{t=1}^{T} \underset{\sim}{x}_t \underset{\sim}{x}'_t\right)^{-1} \left(\sum_{t=1}^{T} \underset{\sim}{x}_t \underset{\sim}{y}^+_t - T\Delta\Omega_{u\nabla x}\Omega^{-1}_{\nabla x \nabla x}\right)$$

where $y^+_t = y_t - \Omega_{u,\nabla x}\Omega^{-1}_{\nabla x,\nabla x}\nabla\underset{\sim}{x}_t$, $\Omega_{\nabla x,\nabla x} = \sum_{v=-\infty}^{\infty} E(\nabla\underset{\sim}{x}_t \nabla\underset{\sim}{x}'_{t+v})$, $\Omega_{u\nabla x} = \sum_{v=-\infty}^{\infty} E(u_t \nabla\underset{\sim}{x}'_{t+v})$ and $\nabla = \sum_{v=0}^{\infty} E(\underset{\sim}{u}_t \underset{\sim}{u}'_{t-v})$.

$$\nabla\hat{M}_t = \underset{(0.0056)}{0.0081} + \underset{(0.0984)}{0.7981}\nabla M_{t-1} - \underset{(0.0291)}{0.0195}\hat{v}_{t-1} - \underset{(0.0031)}{0.0019}R_t, \tag{56}$$

Durbin-Watson = 2.0931, standard error = 0.0092.

Under the assumption that M_t, Y_t, and R_t are jointly dependent, we should use the 2SLS or 3SLS estimator to estimate the short-run dynamic adjustment equation of ∇M_t. However, since in our case equations (49) and (50) are exactly identified, the 2SLS and 3SLS estimators are identical. The 2SLS estimator using ∇M_{t-1}, ∇M_{t-2}, ∇M_{t-3}, ∇Y_{t-1}, ∇Y_{t-2}, ∇Y_{t-3}, R_{t-1}, R_{t-2}, R_{t-3}, as instruments yields the short-run dynamic adjustment equation as

$$\nabla\hat{M}_t = \underset{(0.0063)}{0.0049} + \underset{(0.1080)}{0.9230}\nabla M_{t-1} - \underset{(0.0594)}{0.0953}\hat{v}_{t-1} + \underset{(0.0043)}{0.0015}R_t, \tag{57}$$

Durbin-Watson = 1.9833, standard error = 0.0097.

As one can see from equations (48), (56), and (57), first, the estimates of the short-run adjustment behavior are sensitive to the estimates of the long-run relation. The adjustment coefficient of the deviation from the long-run equilibrium based on the least squares estimates is more than three times larger in magnitude (–0.0648) than the one based on the PH fully modified regression (–0.0195). Second, the interest rate coefficient, although equation (56) has the correct sign, is no longer statistically significant and is less than half the magnitude of the one derived from direct estimation (equations [45] or [54]). Moreover, equation (57) has the wrong sign.

Transforming equation (48) into the autoregressive distributed form yields

$$\hat{M}_t = -0.3641 + 1.7445M_{t-1} - 0.8093M_{t-2} + 0.0986Y_{t-1} - 0.0033R_t. \tag{58}$$

Transforming equation (56) into the autoregressive distributed lag form yields

$$\hat{M}_t = -0.1306 + 1.7786M_{t-1} - 0.7981M_{t-2} + 0.0315Y_{t-1} - 0.0019R_t. \tag{59}$$

Transforming equation (57) into the autoregressive distributed lag form yields

$$\hat{M}_t = -0.6730 + 1.8277M_{t-1} - 0.9230M_{t-2} + 0.1541Y_{t-1} + 0.0015R_t. \tag{60}$$

Comparing equations (45), (54), (58), (59), and (60), several results are worth noting. First, different methods of estimating the long-run income elasticity yield similar results. The least squares and 3SLS estimation of the autoregressive model yield long-run income elasticity of 1.431 and 1.436, respectively. The least squares and PH method of estimating long-run income elasticity directly yield 1.5218 and 1.6177, respectively. Second, different methods yield very different short-run income elasticities. The least squares and 3SLS of the autoregressive model yield 0.1079 and 0.1058, respectively. The least squares estimate of the error-correction model using 1.5218 as the long-run coefficient yields 0.0986. The least squares and 2SLS estimates using 1.6177 as the long-run coefficient yield short-run income elasticity of

0.0315 and 0.1541, respectively. Third, the implied short-run and long-run behavioral dichotomization are sensitive to the estimation methods used. The estimated short-run income elasticity and interest rate coefficient derived from the least squares estimates of the cointegrating vector are almost twice as large as those derived from the PH method of estimating the cointegrating vector [(0.0986, –0.0033) versus (0.0315, –0.0019)]. This in turn implies very different short- and long-run elasticity of the interest rate. For instance, if the interest rate is at 3.5 percent, then the implied short-run elasticity is –0.094 for equation (58) and –0.054 for equation (59) and the implied long-run elasticity is –1.4542 and –2.7828, respectively. If the interest rate is at 5 percent, then the implied short-run elasticity is –0.066 for equation (58) and –0.038 for equation (59) and the long-run elasticity is –1.018 and –1.948, respectively. Fourth, the direct method yields statistically significant short-run and long-run coefficients of the interest rate of –0.0068 and –0.092 (equation [54]), which in turn implies short-term interest rate elasticities of –0.194 and –0.136 and long-term interest rate elasticities of –2.628 and –1.84 when interest rates are at 3.5 percent and 5 percent, respectively. On the other hand, the two-step method yields statistically insignificant short-term interest rate coefficients of –0.0033 for equation (58) and –0.0019 for equation (59) and long-term interest rate coefficients of –0.0509 and –0.0974, respectively. Fifth, in terms of goodness of fit, the direct method has a standard error of 0.0084, while the indirect methods yield standard errors of 0.0089 for equation (48) and 0.0092 for equation (56).

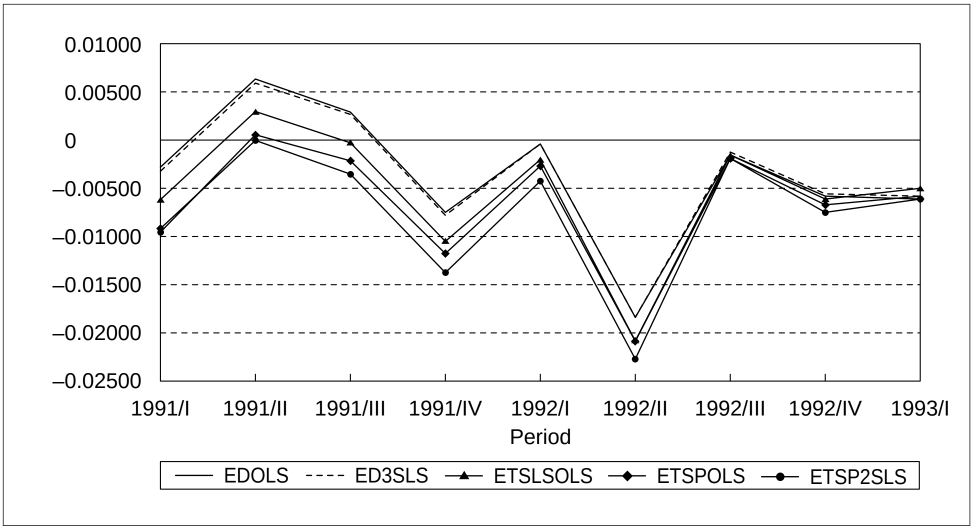
Finally, we compare the predictive performance of the direct method versus the two-step method. We split up the sample into two periods. The first period consists of observations from 1968/III to 1990/IV. The second period consists of observations from 1991/I to 1993/I. The first-period data are used to reestimate model equation (44) either directly or by the two-step method. The estimated models are then used to generate predicted values for the second period. Table 1 provides prediction errors by equation (45), hereafter EDOLS; equation (48), hereafter ETSLOLS;

Table 1 Prediction Comparison between the Direct Method and Two-Step Method

Period	Direct		Two-step		
	EDOLS	ED3SLS	ETSLSOLS	ETSPOLS	ETSP2SLS
1991/I	–0.00274	–0.00321	–0.00612	–0.00949	–0.00905
1991/II	0.00645	0.00607	0.00306	0.00066	0.00005
1991/III	0.00304	0.00270	–0.00018	–0.00202	–0.00355
1991/IV	–0.00744	–0.00771	–0.01058	–0.01187	–0.01386
1992/I	–0.00023	–0.00027	–0.00208	–0.00274	–0.00414
1992/II	–0.01862	–0.01857	–0.02109	–0.02111	–0.02298
1992/III	–0.00136	–0.00106	–0.00181	–0.00154	–0.00179
1992/IV	–0.00570	–0.00543	–0.00668	–0.00601	–0.00737
1993/I	–0.00613	–0.00576	–0.00586	–0.00496	–0.00606
Mean	–0.00364	–0.00369	–0.00570	–0.00657	–0.00764
RMPE	0.00769	0.00762	0.00876	0.00914	0.01014

equation (54), hereafter ED3SLS; equation (56), hereafter ETSPOLS; and equation (57), hereafter ETSP2SL. Figure 1 plots the prediction errors. As one can see, apart from 1991/II and III, the direct method dominates the two-step method in all other time periods. The mean prediction errors as shown in Table 1 are –0.00364 and –0.00369 for the direct method and –0.0057, –0.00657, and –0.00764 for the two-step method. The root mean square prediction errors (RMPE in Table 1) are 0.00769 and 0.00762 for the direct method and 0.00876, 0.00914, and 0.01014 for the two-step method.

Figure 1 Prediction Errors



In short, both in terms of the stability of the parameter estimates and the prediction comparison, the direct estimation of an autoregressive distributed lag form appears to dominate the approach of dichotomizing economic relations into long-run equilibrium and short-run dynamic adjustment behavior. Neither does the dichotomization of nonstationary data offer any statistical advantages over the direct method in terms of inference and estimation. In fact, not only is economic behavior typically formulated in an autoregressive distributed lag form, it is also much simpler to analyze.

VIII. Conclusion

In this paper, we presented statistical issues of VAR and Cowles Commission structural equation modeling when the data were nonstationary. A basic framework linking time-series models and structural equation models were provided, and the implications of cointegration were discussed. We demonstrated that the same rank condition is necessary and sufficient to make inferences of an underlying theoretical structure from the observed data whether the data are stationary or nonstationary. Moreover, because a dynamic structure introduces trivial cointegration between the current and lagged variables, the "simultaneity bias" of the least squares estimator remains a legitimate concern even when the regressors are $I(1)$. Conventional structural equation estimators like 2SLS and 3SLS still possess desirable statistical properties provided that certain variables in the system satisfy the strong exogeneity assumptions. The Wald-type test statistic will be asymptotically χ-square distributed. However, if the strong exogeneity assumption is relaxed as in the case of fitting a VAR model, although the 2SLS and 3SLS estimators remain consistent, their limiting distributions will involve (matrix) unit-root distributions, which are nonstandard. Thus, the conventional formulae for computing the covariances of the 2SLS and 3SLS estimators are no longer valid and the Wald-type test statistic may not be asymptotically χ-square distributed. It is only in the case that the directions of nonstationarity of a system are known and incorporated in the specification that a 3SLS estimator will again possess the desirable statistical properties in the sense of Phillips (1991).

Two implications appear to follow from the analysis of this paper. First, as argued in sections IV, V, and VI, the dichotomization of the long-run equilibrium and short-run adjustment process of a dynamic relationship is a very useful tool in understanding the statistical properties of the estimators of a conventional autoregressive distributed lag model when variables are $I(1)$, but it probably unnecessarily complicates the empirical model building process. As the analysis of Japanese money demand in Section VII has demonstrated, the estimation of the short-run adjustment process of the deviation from the last period's long-run equilibrium can be very sensitive to the way the long-run relation is estimated. Neither the error-correction formulations can generate more accurate predictions than the traditional autoregressive distributed lag model. In fact, it is easier to understand and much simpler to estimate a traditional autoregressive distributed lag model. Inferences about economic agents' long-run and short-run behavior can be drawn from such a model straightforwardly.

Second, contrary to the common belief that prior information does not matter in a nonstationary framework, in fact, it is more important than in a stationary framework. Not only does the identification of a theoretical structure from the probability distributions of observed variables depend on it, but also whether an estimator will have desirable statistical properties depends on it. Prior information plays a critical role in providing a valid inference. Although the idea of letting the data speak for themselves is admirable, the shortage of degree of freedom and multicollinearity can often hamper proper statistical inference. In fact, because so many possibilities exist, any model could arise by chance through a sequence of tests applied to a heavily parameterized model with a finite number of observations. Without a structure, "We may be asking too much of our data. We want them to test our theories, provide us with estimates of important parameters, and disclose to us the exact form of the interrelationships between the variables (Griliches [1967])." A purely data-based approach is likely to start an investigation from scratch. It is against the principle of division of labor. There are good theories. Econometricians constructing econometric models could make use of the insights of other good economists.

References

Anderson, T. W., "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 22, 1951, pp. 327–351 (Correction, *Annals of Statistics*, 8, 1980, p. 1400).

———, "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations," *Annals of Mathematical Statistics*, 30, 1959, pp. 676–687.

Box, G. E. P., and G. N. Jenkins, *Time Series Analysis, Forecasting and Control*, San Francisco: Holden Day, 1970.

Chan, N. H., and C. Z. Wei, "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 1988, pp. 367–401.

Engle, R. F., and C. W. J. Granger, "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 1987, pp. 251–276.

———, and B. Sam Yoo, "Cointegrated Economic Time Series: A Survey with New Results," Discussion Paper 87-26R, Department of Economics, University of California, San Diego, 1989.

Fisher, F. M., *The Identification Problem in Econometrics*, New York: McGraw-Hill, 1967.

Fuller, W. A., *Introduction to Statistical Time Series*, 2nd ed., New York: Wiley, 1996.

Goldfeld, S. M., and D. E. Sichel, "The Demand for Money," in B. M. Friedman and F. H. Hahn, eds. *Handbook of Monetary Economics*, Vol. 1, Amsterdam: Elsevier Science, 1990, pp. 299–356.

Granger, C. W. J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Models," *Econometrica*, 37, 1969, pp. 424–438.

———, and A. A. Weiss, "Time Series Analysis of Error-Correction Models," in S. Karlin, T. Amemiya, and L. Goodman, eds. *Studies in Econometrics, Time Series and Multivariate Statistics*, New York: Academic Press, 1983, pp. 255–278.

Griliches, Z., "Distributed Lags: A Survey," *Econometrica*, 35, 1967, pp. 16–49.

Hendry, D. F., *Econometrics: Alchemy or Science*, Oxford: Blackwell, 1993.

———, and N. R. Ericsson, "'An Econometric Analysis of U.K. Money Demand in Monetary Trends in the United States and the United Kingdom,' by Milton Friedman and Anna J. Schwartz," *American Economic Review*, 81 (1), 1991, pp. 8–38.

Hood, W. C., and T. C. Koopmans, eds. *Studies in Econometric Method*, Cowles Commission Monograph No. 14, New York: John Wiley & Sons, 1953.

Hsiao, C., "Identification," in Z. Griliches and M. Intriligator, eds. *Handbook of Econometrics*, Vol. I, North-Holland: Amsterdam, 1983, pp. 223–283.

———, "Identification of Cointegrated Vector Autoregressive Process," mimeo, University of Southern California, 1996a.

———, "Estimation of Nonstationary Vector Autoregressive Process," mimeo, University of Southern California, 1996b.

———, "Cointegration and Dynamic Simultaneous Equations Models," *Econometrica*, 65, 1997a, pp. 647–670.

———, "Statistical Properties of the Two Stage Least Squares Estimator under Cointegration," *Review of Economic Studies*, 64, 1997b, pp. 385–398.

Johansen, S., "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 1988, pp. 231–254.

———, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1991, pp. 1551–1580.

———, "Identifying Restrictions of Linear Equations," Preprint, 4, Institute of Mathematical Statistics, University of Copenhagen, 1992.

———, "Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration," *Journal of Econometrics*, 69, 1995, pp. 111–132.

———, and K. Juselius, "Identification of the Long-Run and Short-Run Structure—An Application to the ISLM Model," *Journal of Econometrics*, 69, 1994, pp. 111–132.

Johnston, J., "Cointegrating Vectors—How to Find Them, and What Use Are They When You've Got Them?" mimeo, University of California, Irvine, 1991.

Jorgenson, D., "Rational Distributed Lag Functions," *Econometrica*, 34, 1966, pp. 135–150.

Koopmans, T. C., H. Rubin, and R. B. Leipnik, "Measuring the Equation Systems of Dynamic Economics," in T. C. Koopmans, ed. *Statistical Inference in Dynamic Economic Models*, New York: Wiley, 1950, pp. 53–237.

Park, J. Y., and P. C. B. Phillips, "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 1988, pp. 468–497.

———, and ———, "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5, 1989, pp. 95–131.

Pesaran, M. H., and Y. Shin, "Long-Run Structural Modelling," mimeo, Cambridge University, 1995.

Phillips, P. C. B., "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 1986, pp. 311–340.

———, "Time Series Regression with a Unit Root," *Econometrica*, 55, 1987, pp. 277–301.

———, "Optimal Inference in Cointegrating Systems," *Econometrica*, 59, 1991, pp. 283–306.

———, "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1995, pp. 1023–1078.

———, and S. N. Durlauf, "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 1986, pp. 473–495.

———, and B. E. Hansen, "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 1990, pp. 99–125.

———, and S. Quliaris, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 1990, pp. 165–193.

Rothenberg, T .J., "Identification in Parametric Models," *Econometrica*, 39, 1971, pp. 577–592.

Sims, C. A., "Macroeconomics and Reality," *Econometrica*, 48, 1980, pp. 1–48.

———, J. H. Stock, and M. W. Watson, "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 1990, pp. 113–144.

Stock, J. H., "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1987, pp. 1035–1056.

Tsay, R. S., and G. C. Tiao, "Asymptotic Properties of Multivariate Nonstationary Processes with Applications to Autoregressions," *Annals of Statistics*, 18, 1990, pp. 220–250.

White, J. S., "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics*, 29, 1958, pp. 1188–1197.

Zellner, A., and F. Palm, "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics*, 2, 1974, pp. 17–54.