Studies in Nonlinear Dynamics and Econometrics

Quarterly Journal
January 1998, Volume 2, Number 4
The MIT Press

Studies in Nonlinear Dynamics and Econometrics (ISSN 1081-1826) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; tel.: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals $40.00, Institutions $130.00. Canadians add 7% GST. Prices subject to change without notice.

Testing the Expectations Theory of the Term Structure of Interest Rates Using Model-Selection Methods

John C. Chao
Department of Economics
University of Maryland
College Park, Maryland, USA
chao@econ.umd.edu

Chaoshin Chiao
Institute of International Economics
National Dong Hwa University
Hualien, Taiwan

Abstract. *In this paper, we propose a model-selection approach to testing the expectations theory of the term structure of interest rates. Our method is based on the posterior information criterion (PIC) developed and analyzed by Phillips and Ploberger (1994, 1996) and extended to provide order estimation of cointegrating rank by Chao and Phillips (1997). This methodology has the advantage that issues of order selection—i.e., the determination of lag length and cointegrating rank in a vector autoregression—and hypothesis testing are treated within the same framework. Applying our procedure to interest-rate data from the* International Financial Statistics*, we find the expectations theory to be inconsistent with the data.*

Acknowledgments. John Chao is grateful for the financial support from the Office of Graduate Studies and Research of the University of Maryland.

Keywords. cointegration, expectations theory, integrated processes, model selection, PIC, term structure of interest rates, vector autoregression

1 Introduction

Traditionally, most empirical studies in economics involving time-series data have been conducted using models that are linear in parameters. Oftentimes, these models belong to the class of autoregressive models or to their multivariate generalization, the vector-autoregressive models (VAR) in levels. Since the pathbreaking empirical analysis of Davidson et al. (1978), however, it has been widely recognized that the error-correction model, which has nonlinear restrictions on some of its coefficients, fits many macroeconomic and financial time series very well, and often better than strictly linear time-series models such as the level VARs.

Parallel to this development is the important work of Granger (1981), Granger and Weiss (1983), and Engle and Granger (1987) on cointegration. In these papers, Granger and his co-authors conceptualize and formalize the idea that several nonstationary I(1) processes (i.e., stochastic processes that are stationary upon first differencing) may move together in such a way so that taking a linear combination of them may result in a stationary I(0) process. Moreover, they show that this phenomenon of cointegration is intimately related to the error-correction model in the sense that every cointegrated time series has an error-correction representation.

The papers of Granger and his co-authors have stimulated much theoretical research in nonstationary regression and cointegration. A main conclusion of this subsequent research is that in general the classical Wald statistic has a nonstandard asymptotic distribution when it is constructed from level regressions with untransformed I(1) variables, but has the usual chi-square distribution when knowledge of cointegrating relationships is used to bring nonstationary regressions into error-correction form [see, for example, Phillips (1991) and Toda and Phillips (1993)].

This theoretical research has also had a profound impact on empirical work. One area which has, in particular, witnessed the application of the new theory of nonstationary regression and cointegration is in the testing of rational expectations present-value models of the asset markets. In two influential papers, Campbell and Shiller (1987, 1989) both estimate and use cointegrating relationships implied by economic theory to transform a nonstationary I(1) system into one that is I(0). They are then able to construct Wald tests that have the conventional chi-square asymptotic distribution in testing the implications of the present-value models for stock prices and for the term structure of interest rates.

Hence, for a system of I(1) processes, both theoretical and practical concerns point to the need for identifying the existence and number of cointegrating relationships, either through estimation or through prior information supplied by economic theory. In addition, if the underlying data-generating mechanism is taken to be well modeled by a VAR process, then there is the additional task of choosing the lag order of the VAR. Recently, Phillips (1996) and Chao and Phillips (1997) have argued that the problem of jointly estimating the cointegrating rank (i.e., the number of linearly independent cointegrating vectors) and the lag order of a VAR system can be best viewed within the framework of model selection. Moreover, they advocate the application of a newly developed model-selection procedure, the *posterior information criterion* (PIC) of Phillips and Ploberger (1994, 1996) for this task of joint order estimation. Chao and Phillips (1997), in particular, have shown that the use of PIC in this context has an advantage over the alternative sequential-likelihood-ratio test procedure proposed by Johansen (1992), in that the former results in consistent estimation of the cointegrating rank, whereas the latter does not.

The purpose of this paper is to apply PIC to a reexamination of the expectations theory of the term structure of interest rates.[1] Our approach is similar to that of Campbell and Shiller (1987), in that we test the expectations theory as restrictions on a VAR. However, our use of the model-selection procedure PIC has a number of advantages over the statistical procedures adopted by Campbell and Shiller (1987). First, whereas Campbell and Shiller (1987) use two distinctively different statistical methods for the two tasks of lag-order selection and hypothesis testing (i.e., they use an information criterion, AIC, to select the lag order of their VAR, and then test the restrictions of the present-value model using the classical Wald test), we show that by employing PIC, we can determine the lag order and the cointegrating rank as well as test restrictions implied by the expectations theory, all within a single coherent framework. Second, in our framework, bar charts and histograms of PIC values can easily be constructed to assess the strength and robustness of our inference with respect to lag and cointegrating-rank selection. Finally, Chao and Phillips (1997) have shown that PIC leads to consistent estimation of the lag order, even in nonstationary VARs, whereas AIC, the lag-selection procedure that Campbell and Shiller (1987) employed, is known to be inconsistent in estimating the lag order for even stationary autoregressive processes and has, in fact, a tendency to overestimate the lag order.[2]

[1]See Chao and Phillips (1997) for an application of a variant of the PIC methodology to a study of the present-value model for stock prices using the data set of Campbell and Shiller (1987).

[2]See Shibata (1976) and Tsay (1984) for theoretical results pointing to the inconsistency of AIC in lag selection. See also Sawa (1978) for Monte

The paper proceeds as follows. In Section 2, we briefly describe the rational expectations present-value model for the term structure of interest rates, and the restrictions that it imposes on a vector error-correction model (VECM). Section 3 gives a summary presentation of our model-selection procedure, PIC. For details on this procedure, the readers are referred to Chao and Phillips (1997). Section 4 gives our empirical results. Finally, we offer our concluding thoughts in Section 5.

2 The Present-Value Model and Its Testable Implications

In this section, we briefly describe the present-value model of the term structure of interest rates, and discuss its testable implications. Since a detailed discussion of this model is given in Campbell and Shiller (1987), we focus our attention here only on those features of the model that will be relevant for our subsequent analysis. Following Campbell and Shiller (1987), the present-value model can be stated in its general form as:

$$y_{1t} = \theta (1 - \delta) \sum_{i=0}^{\infty} E \left(y_{2t+i} | I_t \right) + c, \tag{1}$$

where $E(\cdot | I_t)$ denotes the mathematical expectation conditional on the full public information set I_t at time t. Throughout this paper, we shall treat conditional expectations as being equivalent to linear projections on information. While the model given by Equation 1 also includes the present-value model for stock prices and the permanent income theory of consumption as special cases, our discussion here will only focus on its interpretation as a model for the expectations theory of the term structure of interest rates. In this context, the value of the parameter θ is known a priori to equal one; the parameter c has an interpretation as the liquidity premium unrestricted by the model; and δ is the parameter of linearization set to $\frac{1}{(1+R)}$, where R is the discount rate. Here, we let y_{1t} denote the long-term yield and y_{2t} the short-term interest rate.

Let $s_t = y_{1t} - y_{2t}$ be the spread between long- and short-term interest rates. Campbell and Shiller (1987) show that the present-value model denoted by Equation 1 implies two alternative interpretations of the spread:

$$s_t = \sum_{i=0}^{\infty} \delta^i E \left(\triangle y_{2t+i} | I_t \right) + c, \text{ and} \tag{2}$$

$$s_t = \frac{\delta}{1 - \delta} \sum_{i=0}^{\infty} E \left(\triangle y_{1t+i} | I_t \right) + c. \tag{3}$$

If $\triangle y_{2t}$ is stationary, then Equations 2 and 3 further imply that s_t and $\triangle y_{1t}$ are also stationary. Hence, y_{1t} and y_{2t} are cointegrated with cointegrating vector $(1, -1)$.

To test the implications of the expectations theory as expressed in these interpretations of the spread, we first specify the joint dynamics of y_{1t} and y_{2t} as a bivariate VAR of lag order $p + 1$; i.e.,

$$y_t = \mu + \sum_{i=0}^{p+1} B_i y_{t-i} + \varepsilon_t, \tag{4}$$

where $y_t = \left(y_{1t}, y_{2t} \right)'$, $\mu = (\mu_1, \mu_2)'$, and $\varepsilon_t \equiv N(\underset{\sim}{0}, \Omega)$. This formulation is in line with that of Hansen and Sargent (1981), Campbell and Shiller (1987), and Toda (1991), in that it includes lagged values of not only y_{2t} but also y_{1t} in the information set available to the econometrician.

Alternatively, Equation 4 can be written as:

$$\triangle y_t = \mu + \sum_{i=0}^{p} B_i^* \triangle y_{t-i} + B_* y_{t-1} + \varepsilon_t, \tag{5}$$

Carlo evidence on the tendency of AIC to overestimate the lag order in a finite sample.

where $B_i^* = -\sum_{l=i+1}^{p+1} B_l$ and $B_* = \sum_{i=1}^{p+1} B_i - I_2$. Now, if y_{1t} and y_{2t} are cointegrated, then the 2×2 matrix B_* is of reduced rank, and we can write Equation 5 in the form:

$$\triangle y_t = \mu + \sum_{i=0}^{p} B_i^* \triangle y_{t-i} + \gamma\alpha' y_{t-1} + \varepsilon_t, \tag{6}$$

where $\gamma = (\gamma_1, \gamma_2)'$ and α=$(1, \alpha_1)'$. We shall refer to Equation 6 as an unrestricted VECM, and test the expectations theory as restrictions on this model. To do so, it is most convenient for us to work with the spread relationship given by Equation 3. Imposing Equation 3 on the VECM in Equation 6 results in the following set of restrictions:

$$b_{i,11}^* = b_{i,12}^* = 0,\ i = 1, \ldots, p, \tag{7}$$

$$\gamma_1 = \frac{1-\delta}{\delta} = R, \tag{8}$$

$$\mu_1 = -\frac{(1-\delta)}{\delta} c, \text{ and} \tag{9}$$

$$\alpha = (1, \alpha_1)' = (1, -1)', \tag{10}$$

where $(b_{i,11}^*, b_{i,12}^*)$ denotes the first row of the matrix B_i^*.

Before proceeding to discuss our approach for testing the restrictions given by Equations 7–10, we first introduce more notations. In the rest of the paper, we shall let $\Delta Y = (\Delta y_1, \Delta y_2, \ldots, \Delta y_T)'$, $Y_{-1} = (y_0, \ldots, y_{T-1})'$, $W_t(p) = (1, \Delta y_{t-1}', \Delta y_{t-2}', \ldots, \Delta y_{t-p}')'$, $W(p) = (W_1(p), \ldots, W_T(p))'$, $E = (\varepsilon_1, \ldots, \varepsilon_T)'$, and $B^* = (\mu, B_1^*, \ldots, B_p^*)$. With these notations, we can rewrite Equation 6 in a more concise manner, as:

$$\Delta Y = W(p)B^{*\prime} + Y_{-1}\alpha\gamma' + E. \tag{11}$$

In addition, we let $P_X = X\left(X'X\right)^{-1} X'$ denote the matrix that projects onto the range space of X, and let $M_X = I_T - P_X$ denote the matrix that projects onto the orthogonal complement of the range space of X.

3 A PIC Approach to Testing Economic Hypothesis

Our approach to testing the expectations theory for the term structure of interest rates is based on the PIC criterion first developed by Phillips and Ploberger (1994, 1996) and applied to the VAR context by Chao and Phillips (1997). Similar to other information criteria, such as AIC and BIC, hypothesis testing within the PIC framework can be performed by computing the value of the PIC criterion under the various alternative hypotheses (or models) under consideration, and choosing that hypothesis which gives the maximum PIC value. To be more explicit, let us first define the form of the PIC criterion for the (unrestricted) VECM given by Equation 6 (or, alternatively, Equation 11) of the last section. Following Chao and Phillips (1997), we have:

$$\begin{aligned}
&\mathrm{PIC}\left(p, r\right) = \\
&\exp\left\{-\frac{1}{2}\left[\mathrm{vec}\left(\Delta Y - Y_{-1}\widetilde{B}_*\left(p, r\right)'\right)' M_{W(p)}\,\mathrm{vec}\left(\Delta Y - Y_{-1}\widetilde{B}_*\left(p, r\right)'\right)\right]\right\} \\
&\exp\left\{\frac{1}{2}\left[\mathrm{vec}\left(\Delta Y - Y_{-1}\widehat{B}_*\left(\overline{p}\right)'\right)' M_{W(\overline{p})}\,\mathrm{vec}\left(\Delta Y - Y_{-1}\widehat{B}_*\left(\overline{p}\right)'\right)\right]\right\} \\
&\left[\left|\widehat{\Omega}^{-1} \otimes W\left(\overline{p}\right)' W\left(\overline{p}\right)\right|^{\frac{1}{2}} \Big/ \left|\widehat{\Omega}^{-1} \otimes W\left(p\right)' W\left(p\right)\right|^{\frac{1}{2}}\right] \\
&\left[\left|\widehat{\Omega}^{-1} \otimes Y_{-1}' M_{W(\overline{p})} Y_{-1}\right|^{\frac{1}{2}} \Big/ \left|\widetilde{H}\left(p, r\right)\left(\widehat{\Omega}^{-1} \otimes Y_{-1}' M_{W(p)} Y_{-1}\right)\widetilde{H}\left(p, r\right)'\right|^{\frac{1}{2}}\right].
\end{aligned} \tag{12}$$

Here, $\widetilde{B}_*(p, r) = \widehat{\gamma}(p, r)\widehat{\alpha}(p, r)'$, where $\widehat{\gamma}(p, r)$ and $\widehat{\alpha}(p, r)$ are the Gaussian maximum-likelihood estimators of the reduced-rank parameters γ and α in Equation 6, when the cointegrating rank is taken to be r and the order of lagged differences is taken to be p.[3] We let $\widehat{B}_*(\overline{p}) = \Delta Y' M_{W(\overline{p})} Y_{-1}(Y_{-1}' M_{W(\overline{p})} Y_{-1})^{-1}$ denote the least-squares estimator of B_* in a VECM of lag order $\overline{p}$, with $\overline{p}$ being the maximum lag length under consideration. In addition, $\widehat{\Omega} = \Delta Y' M_{(Y_{-1}, W(\overline{p}))} \Delta Y$ is the maximum-likelihood estimator of Ω in the case where Equation 5 has the highest possible order in our setup, i.e., $r = m$ and $p = \overline{p}$. Furthermore, we define

$$\widetilde{H}(p, r) = \begin{bmatrix} (\widehat{\Gamma}(p, r)' \otimes F(r)') \\ (I_m \otimes (I_r, \widehat{\overline{A}}(p, r)')) \end{bmatrix}, \tag{13}$$

where $F(r)$ is the $m \times (m - r)$ matrix such that $F(r)' = [0, I_{m-r}]$. The criterion denoted by Equation 12 can then be used to determine the cointegrating rank and lag order of the (unrestricted) VECM, based on the decision rule:

$$\left(\widehat{p}, \widehat{r}\right) = \arg\max \mathrm{PIC}\left(p, r\right). \tag{14}$$

Although the criterion denoted by Equation 12 may appear complicated, it actually has a simple and intuitive interpretation as a combination of likelihood-ratio statistics, which test the fit of the reduced-rank model of Equation 6, and penalty terms, which reflect the complexity of the model. To see this, we first note that with a bit of straightforward algebra, the criterion denoted by Equation 12 can be rewritten as:

$$\begin{aligned} &\mathrm{PIC}\left(p, r\right) = \\ &\exp\left\{-\frac{1}{2}\,\mathrm{tr}\left[\widehat{\Omega}^{-1}\left(\widetilde{B}_*\left(p, r\right) - \widehat{B}_*\left(p\right)\right) Y_{-1}' M_{W(p)} Y_{-1}\left(\widetilde{B}_*\left(p, r\right) - \widehat{B}_*\left(p\right)\right)'\right]\right\} \\ &\exp\left\{-\frac{1}{2}\,\mathrm{tr}\left[\widehat{\Omega}^{-1}\widehat{B^*}\left(p^*\right) W\left(p^*\right)' M_{(Y_{-1}, W(p))} W\left(p^*\right) \widehat{B^*}\left(p^*\right)'\right]\right\} \\ &\left[\left|\widehat{\Omega}^{-1} \otimes W\left(\overline{p}\right)' W\left(\overline{p}\right)\right|^{\frac{1}{2}}\right] / \left[\left|\widehat{\Omega}^{-1} \otimes W\left(p\right)' W\left(p\right)\right|^{\frac{1}{2}}\right] \\ &\left[\left|\widehat{\Omega}^{-1} \otimes Y_{-1}' M_{W(\overline{p})} Y_{-1}\right|^{\frac{1}{2}}\right] / \left[\left|\widetilde{H}\left(p, r\right)\left(\widehat{\Omega}^{-1} \otimes Y_{-1}' M_{W(p)} Y_{-1}\right) \widetilde{H}\left(p, r\right)'\right|^{\frac{1}{2}}\right], \end{aligned} \tag{15}$$

where $\widehat{B^*}(p^*) = \Delta Y' M_{(Y_{-1}, W(p))} W(p^*)(W(p^*)' M_{(Y_{-1}, W(p))} W(p^*))^{-1}$ denotes the least-squares estimator of the last $m(\overline{p} - p)$ columns of B^* [or the coefficients of the last $(\overline{p} - p)$ lagged differences] in a VECM of lag order $\overline{p}$. Now, note that the trace term in the exponent of the first term, i.e.,

$$\mathrm{tr}\left[\widehat{\Omega}^{-1}\left(\widetilde{B}_*\left(p, r\right) - \widehat{B}_*\left(p\right)\right) Y_{-1}' M_{W(p)} Y_{-1}\left(\widetilde{B}_*\left(p, r\right) - \widehat{B}_*\left(p\right)\right)'\right]$$

is, in fact, the likelihood-ratio statistic for testing that the cointegrating rank is r versus the general alternative that it equals m [cf. Reinsel and Ahn (1992)]. Similarly, the trace expression in the exponent of the second term, i.e.,

$$\mathrm{tr}\left[\widehat{\Omega}^{-1}\widehat{B^*}\left(p^*\right) W\left(p^*\right)' M_{(Y_{-1}, W(p))} W\left(p^*\right) \widehat{B^*}\left(p^*\right)'\right]$$

is simply the likelihood-ratio statistic for testing the null hypothesis that the VECM has p lagged differences

[3] Of course, in our application of PIC here, we work only with a bivariate system, so that strictly speaking, the reduced-rank formulation of Equation 12 is only correct for $r = 1$. When $r = 0$, we take $r = \alpha = 0$, so that correspondingly, $\widehat{\gamma} = \widehat{\alpha} = 0$. On the other hand, when $r = 2$, we take $\gamma = B_*$ and $\alpha = I_2$, so the maximum-likelihood estimators are $\widehat{\gamma} = \widehat{B}_*$ and $\widehat{\alpha} = I_2$.

versus the alternative that the number of lagged differences is $\overline{p} > p$. The third and fourth terms, on the other hand, are terms which, *ceteris paribus*, penalize models for having higher lag order and/or greater cointegrating rank. Note also that the penalty terms of PIC differ from those of AIC, BIC, and the Hannan-Quinn criterion: whereas the penalty functions for these other information criteria depend on a simple parameter count, the penalty terms of PIC compare the Fisher information matrix of the larger model with that of the smaller model and, in effect, use the redundant information introduced by an overparameterized model to penalize excess parameterization. Both Phillips and Ploberger (1994) and Chao and Phillips (1997) have argued that the use of a penalty function based on the Fisher information matrix is a likely explanation for the better finite-sample performance of PIC relative to AIC and BIC, as shown in Monte Carlo experiments reported in these two papers. An information-matrix-based penalty function is advantageous because it takes into account not only the number of regressors included in the alternative models, but also the magnitude of the regressors and the sample information accumulated in the data about the models' parameters. Chao and Phillips (1997) have also shown that choosing p and r based on the decision rule given by Equation 14 leads to consistent estimators of the lag order and cointegrating rank. See Chao and Phillips (1997) for more discussion of the properties of PIC.

Now, to test restrictions implied by the expectations theory of the term structure of interest rates, we must also construct the PIC statistic for the VECM restricted by Equations 7–10 in Section 2. To proceed, we note that the restrictions given by Equations 7–10 are linear restrictions that take the form:

$$H\left(M^R\right): \gamma = f_1\phi + g_1, \quad \alpha = g_2, \quad \mathrm{vec}\left(B^*\right) = F_3\left(p\right)\beta, \tag{16}$$

where β and ϕ are respectively, a $(2+2p) \times 1$ vector of basic parameters and a scalar parameter, and where

$$F_3\left(p\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & I_{2p+1} \end{pmatrix}$$

is a $(2+4p) \times (2+2p)$ matrix, and $f_1 = (0, 1)'$, $g_1 = (R, 0)'$, and $g_2 = (1, -1)'$ are 2×1 vectors. Moreover, by imposing the restrictions given by the null hypothesis (Equation 16) on the VECM (Equation 6) of Section 2, we obtain the restricted VECM:

$$\begin{aligned} M_p^R: \quad \Delta y_t &= \left(I_2 \otimes W_t\left(p\right)'\right) F_3\left(p\right)\beta + \left(f_1\phi + g_1\right) g_2' y_{t-1} + \varepsilon_t \\ &= X_t\left(p\right)\beta + \gamma\left(\phi\right) g_2' y_{t-1} + \varepsilon_t, \text{ (say)}, \end{aligned} \tag{17}$$

where $X_t(p) = (I_2 \otimes W_t(p)')F_3(p)$ and $\gamma(\phi) = f_1\phi + g_1$. Following arguments similar to Theorem 3.3 and Corollary 3.4 of Chao and Phillips (1997), we can deduce the form of the PIC statistic for the restricted model (Equation 17) as:

$$\begin{aligned} &\mathrm{PICR}\left(p\right) = \\ &\exp\left\{-\frac{1}{2}\left[\mathrm{vec}\left(\Delta Y - Y_{-1}\overline{B}_*\left(p\right)'\right)' N\left(p\right) \mathrm{vec}\left(\Delta Y - Y_{-1}\overline{B}_*\left(p\right)'\right)\right]\right\} \\ &\exp\left\{\frac{1}{2}\left[\mathrm{vec}\left(\Delta Y - Y_{-1}\widehat{B}_*\left(\overline{p}\right)'\right)' M_{W\left(\overline{p}\right)} \mathrm{vec}\left(\Delta Y - Y_{-1}\widehat{B}_*\left(\overline{p}\right)'\right)\right]\right\} \\ &\left[\left|\widehat{\Omega}^{-1} \otimes W\left(\overline{p}\right)' W\left(\overline{p}\right)\right|^{\frac{1}{2}}\right] / \left[\left|F\left(p\right)'\left(\widehat{\Omega}^{-1} \otimes W\left(p\right)' W\left(p\right)\right) F\left(p\right)\right|^{\frac{1}{2}}\right] \\ &\left[\left|\widehat{\Omega}^{-1} \otimes Y_{-1}' M_{W\left(\overline{p}\right)} Y_{-1}\right|^{\frac{1}{2}}\right] / \left[\left|\overline{H}' N\left(p\right) \overline{H}\right|^{\frac{1}{2}}\right], \end{aligned} \tag{18}$$

where

$$N\left(p\right) = \left(\widehat{\Omega}^{-1} \otimes Y_{-1}' Y_{-1}\right) - \left(\widehat{\Omega}^{-1} \otimes Y_{-1}' W\left(p\right)\right) F_3\left(p\right)$$
$$\left(F_3\left(p\right)' \left(\widehat{\Omega}^{-1} \otimes W\left(p\right)' W\left(p\right)\right) \left(F_3\left(p\right)\right)\right)^{-1} F_3\left(p\right)' \left(\widehat{\Omega}^{-1} \otimes W\left(p\right)' Y_{-1}\right),$$

$$\overline{H} = \left(f_1' \otimes g_2'\right),$$

and

$$\overline{B}_*\left(p\right) = \gamma\left(\widehat{\phi}\left(p\right)\right) g_2',$$

with $\widehat{\phi}(p)$ being the Gaussian maximum-likelihood estimate of ϕ in the restricted model (Equation 17) when the lag order is taken to be p. Given the criterion denoted by Equation 18, the lag choice $\widetilde{p}$ for the restricted model can be made as follows:

$$\widetilde{p} = \arg\max \mathrm{PICR}(p). \tag{19}$$

Finally, let $H(M^R)$ be the null hypothesis defined by Equation 16 (i.e., the set of restrictions implied by the expectations theory), and let $H(M^U)$ be the alternative of an unrestricted VECM as defined by Equation 6 in Section 2. Then we propose to test $H(M^R)$ versus $H(M^U)$ using the decision rule:

$$\text{accept } H\left(M^R\right) \text{ in favor of } H\left(M^U\right) \quad \text{if} \quad \frac{\mathrm{PICR}(\widetilde{p})}{\mathrm{PIC}(\widehat{p}, \widehat{r})} > 1. \tag{20}$$

Note that this decision rule is analogous to one that Bayesians would employ for posterior-odds analysis in the case where the loss function treats type-I and type-II errors symmetrically. Indeed, Chao and Phillips (1997) show that in certain special situations, the PIC criterion can be interpreted as an approximate posterior-odds ratio. The interested reader is referred to that paper for further discussion.

4 Data Description and Empirical Results

In this section, we apply the statistical procedures given in Section 3 to test the expectations theory of the term structure of interest rates described in Section 2. A brief discussion of the data is in order. The data used come from the *International Financial Statistics* published by the IMF, and contain monthly observations extending from 1964:1 to 1997:1. The short-term interest rate y_{2t} in the sample is the one-month U.S. Treasury bill rate, while the long-term interest rate y_{1t} we use is the 10-year U.S. government bond yield. We shall report our empirical results not only for the full sample period, but also for the shorter period from 1964:1 to 1978:8. Results for the latter period are also reported since, as Campbell and Shiller (1987, 1991) have argued, the use of a subsample period ending 1978:8 is likely to avoid any structural change in the data owing to the 1979 monetary regime shift.

4.1 Unit Root Tests

We begin our empirical analysis by testing the two interest-rate series for unit roots. To be consistent with the model-selection approach that we take in our subsequent multivariate analysis, we test unit roots using the univariate version of PIC as detailed in Phillips and Ploberger (1994). The Phillips-Ploberger PIC approach to unit-root testing requires comparing a general autoregressive model with trend (and written in difference form), viz.,

$$H\left(M_{p,l}^{\mathrm{REF}}\right): \Delta y_t = a_0 y_{t-1} + \sum_{i=1}^{p-1} a_i \Delta y_{t-i} + \sum_{j=0}^{l} b_j t^j + \varepsilon_t, \tag{21}$$

Table 1
Unit Root Tests

Sample Period	Variable	Lag Selected[a]	Trend Selected	PIC Ratio in Favor of a Unit Root
1964:1–1978:8	y_{1t}	3	-1[b]	198.643
	y_{2t}	2	0	2.570e−25
1964:1–1997:1	y_{1t}	3	-1	498.699
	y_{2t}	7	-1	266.355

[a] The maximum lag order $\overline{p}$ is set equal to 12.
[b] -1 denotes the absence of a constant term.

with one that explicitly incorporates a unit root:

$$H\left(M_{p,l}^{\mathrm{UR}}\right) : \Delta y_t = \sum_{i=1}^{p-1} a_i \Delta y_{t-i} + \sum_{j=0}^{l} b_j t^j + \varepsilon_t. \tag{22}$$

Unit-root tests can then be mounted using the test criterion:

$$\frac{\mathrm{PIC}^{\mathrm{UR}}(\widetilde{p}, \widetilde{r})}{\mathrm{PIC}^{\mathrm{REF}}(\widehat{p}, \widehat{r})} = \left\{\left(\frac{1}{\widehat{\sigma}^2}\right) y_{-1}' M_Z y_{-1}\right\}^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\widehat{\sigma}^2 \widehat{a}_0^{\,2} \left(y_{-1}' M_Z y_{-1}\right)\right\}, \tag{23}$$

where $y_{-1} = (y_0, \ldots, y_{T-1})'$, Z is a $T \times (p + l)$ matrix whose t^{th} row is the $1 \times (p + l)$ vector $(\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}, 1, t, \ldots, t^l)'$ and $\widehat{a}_0$ is the least-squares estimator of a_0 in Equation 21. Here, $\widehat{p}$ and $\widehat{l}$ are the lag order and trend degree selected by PIC for the reference model given by Equation 21, while $\widetilde{p}$ and $\widetilde{l}$ are the lag order and trend degree selected for the unit-root model denoted by Equation 22. Given the criterion denoted by Equation 23, decisions about unit roots can be made based on the rule:

$$\text{decide in favor of unit root if } \quad \frac{\mathrm{PIC}^{\mathrm{UR}}(\widetilde{p}, \widetilde{r})}{\mathrm{PIC}^{\mathrm{REF}}(\widehat{p}, \widehat{r})} > 1. \tag{24}$$

Table 1 documents the results of our unit-root test using the test criterion denoted by Equation 23 and the decision rule given by Equation 24. The last column of Table 1 provides some evidence that both interest-rate series can be characterized as I(1) processes, although the evidential support is not uniform across all sample periods for both series. Of the two, the 10-year government bond yield seems to have the stronger support in favor of its characterization as an I(1) process. In both the full sample period and the short sample period from 1964:1 to 1978:8, the PIC criterion decides in favor of a unit-root characterization for this series. The one-month Treasury bill rate, on the other hand, is found to have a unit root in the full sample period but not in the short sample period. Note also that in all but one case (the one case being the Treasury bill rate in the full sample period), our PIC criterion selects a rather parsimonious dynamic specification with three or less lags. This is not surprising, since model-selection procedures by design will favor the more parsimonious model in the absence of strong evidence for the larger model.

4.2 Lag Order and Cointegrating Rank Estimation

Since some (although not unambiguous) evidence for the presence of unit roots in our interest-rate series is found in the last section, the natural next step in our analysis is to test to see whether the short-term and the long-term interest rates are cointegrated. From Equations 7–10, it is apparent that the cointegration of the short- and long-term interest rates is but one of the restrictions of the expectations theory of the term structure. However, testing for cointegration in the absence of the other restrictions of the expectations theory may nevertheless be of interest, especially if one wishes to obtain an appropriate time-series representation for the long-term and short-term interest rates for forecasting purposes. Hence, in this section, we shall proceed to estimate the lag order and the cointegrating rank of the (unrestricted) VECM (Equation 6) using the

Table 2
Estimation of Cointegrating Rank and the Order of Lagged Differences

Sample Period	$\hat{p}$	$\hat{r}$	p^m	r^m
1964:1–1978:8	3	1	2.490	0.447
1964:1–1997:1	2	1	3.920	0.885

Notes: $\hat{p}$, $\hat{r}$, p^m, and r^m are as defined in Equations 14 and 25.

decision rule given in Equation 14. Note that Equation 14 selects the mode amongst possible PIC values. In addition to this modal criterion, we shall also obtain point estimates for p and r by taking a weighted average using the PIC values as weights, i.e.,

$$\left(p^m, r^m\right) = \text{round}\left\{\left(\sum_{p,r} \text{PIC}\left(p, r\right)\right)^{-1} \sum_{p,r} \left(\text{PIC}\left(p, r\right) \cdot \left(p, r\right)\right)\right\}. \tag{25}$$

The advantage of using a mean criterion such as Equation 25 to supplement a modal criterion like Equation 14 is that the former is affected by and therefore alerts the investigators to cases (i.e., order combinations) where an appreciable mass of PIC values may occur in regions away from the mode.

Table 2 reports order estimates (p, r) under both the modal criterion of Equation 14 and the mean criterion of Equation 25 for the bivariate system comprised of y_{1t} (the 10-year government bond yield) and y_{2t} (the one-month Treasury bill rate). Note that the modal criterion finds one cointegrating vector both for the short sample period extending from 1964:1 to 1978:8 and for the full sample period. The (prior to rounding) cointegrating-rank estimates from the mean criterion are .447 and .885, respectively, for the short and the full sample periods. That these estimates are both closer to .5 than to 0 or 1 is indicative of the considerable uncertainty which surrounds the inference on cointegration in this case. Figures 1 and 2, which give bar charts of PIC values in (p, r) space for the two sample periods, also reveal that the inference on cointegration here is not very sharp, as an appreciable mass of PIC values is observed for every value of the cointegrating rank r under consideration. The weak inference on cointegration and also on unit roots serves to explain our apparently contradictory finding of a cointegrating vector for the short sample period when the results of our unit-root tests suggest that y_{1t} is I(1) and y_{2t} is I(0) during this period, in which case they cannot possibly be cointegrated. The apparently conflicting results are due simply to the fact that the data is not very informative about the existence of unit roots and cointegration. Note also that for the short sample period, the estimate of the cointegrating rank based on the mean criterion rounds to zero and is, thus, more in accord with the unit-root test results.

In spite of the less than unambiguous results which our unit-root and cointegration analyses have yielded, it is worth keeping in mind that the goal of model selection is often not so much to test whether a model is true, but rather to choose from a class of models under consideration that which best fits the data while making the appropriate trade-offs with respect to model size. Seen from this perspective, our results suggest that a VECM with $r = 1$ gives a better time-series representation for our data than alternative specification of the cointegrating rank. This is in agreement with the results of Campbell and Shiller (1987) who, using a different data set, find evidence of cointegration between the one-month Treasury bill rate and a long-term interest rate.

With respect to lag selection, Table 2 reports that the modal criterion chooses lag orders of 3 and 2, respectively, for the short and the full sample periods. As with cointegrating-rank selection, both the lag estimates from the mean criterion (which gives $\widetilde{p} = 2.49$ for the short sample period and $\widetilde{p} = 3.92$ for the full sample period) and the bar charts of PIC values depicted in Figures 1 and 2 show considerable uncertainty in regards to lag specification. This is particularly true for the full sample period, where the lag selection of the mean criterion rounds to 4: this is substantially different from the lag order of 2 selected by the modal criterion.

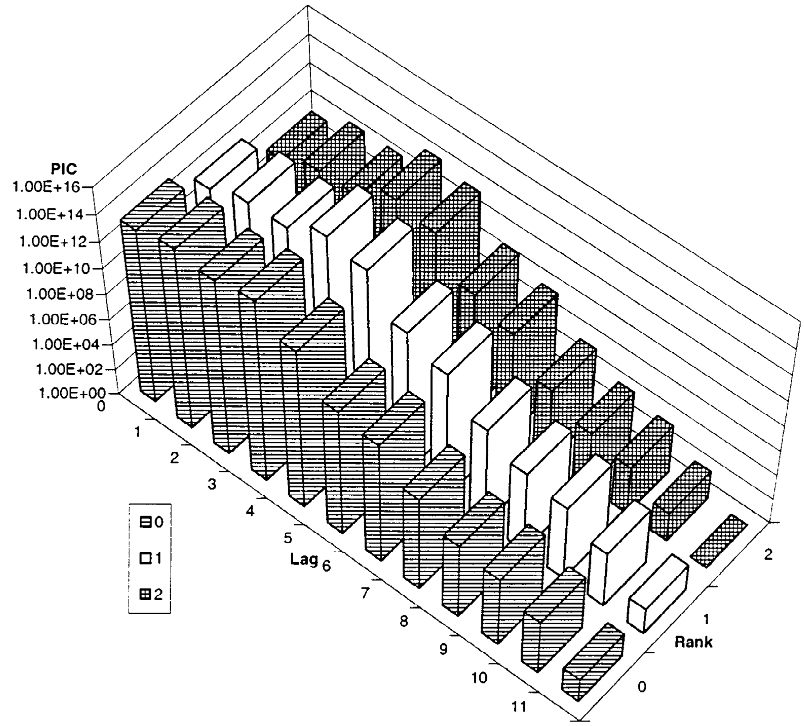

Figure 1
Lag/rank selection 1964:1–1978:8 for the spread between y_2 and y_1.

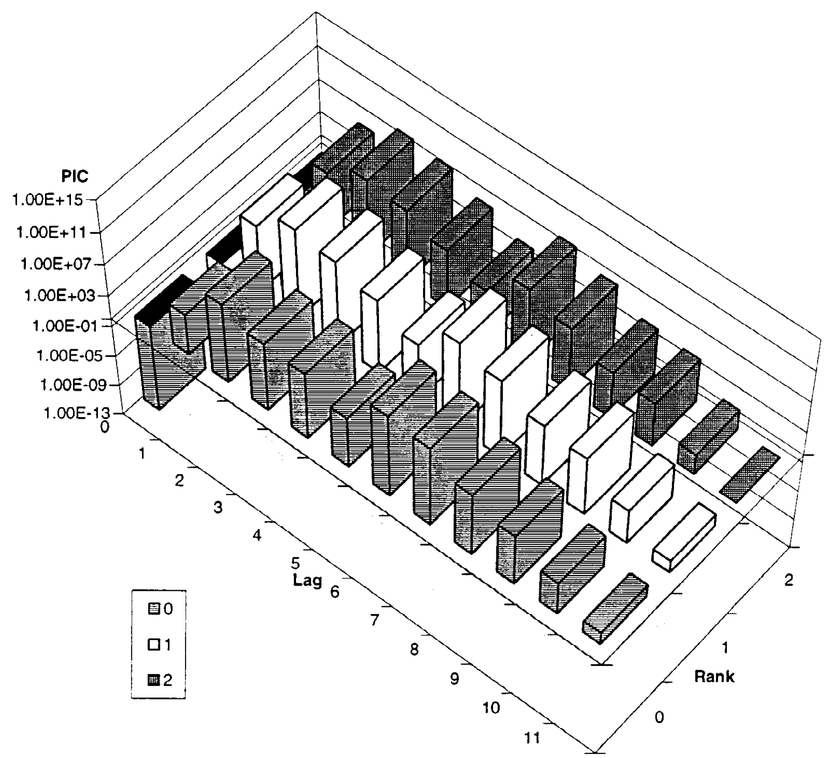

Figure 2
Lag/rank selection 1964:1–1997:1 for the spread between y_2 and y_1.

Table 3
Model Selection Test of the Expectations Theory[c]

Sample Period	$\hat{p}$ under $H(M^R)$	p^m under $H(M^R)$	τ_1	τ_2
1964:1–1978:8 ($R = .0637/12$)	3	2.862	0.278	0.278
1964:1–1997:1 ($R = .07911/12$)	3	3.200	1.017e−18	3.755e−18

[c] R is given by the mean 10-year bond rate.

Table 4
Model Selection Test of the Expectations Theory[d]

Sample Period	$\hat{p}$ under $H(M^R)$	p^m under $H(M^R)$	τ_1	τ_2
1964:1–1978:8	3	2.862	4.639e−02	4.639e−02
1964:1–1997:1	3	3.200	3.286e−19	1.406e−18

[d] R is unrestricted.

4.3 Tests of the Expectations Theory of the Term Structure

We proceed now to test the restrictions (Equations 7–10) of the expectations theory using the test criterion given in Equation 20. To summarize our results, we define the following statistics:

$$\tau_1 = \frac{\text{PICR}\left(\widetilde{p}\right)}{\text{PIC}\left(\widehat{p}, \widehat{r}\right)}, \quad \text{and}$$

$$\tau_2 = \frac{\text{PICR}\left(\widetilde{p}\right)}{\text{PIC}\left(\widetilde{p}, 1\right)}.$$

Here, PICR $(\cdot)$ denotes the form of the PIC criterion under the null model $H(M^R)$, when restrictions implied by the expectations theory are imposed, while PIC$(\cdot,\cdot)$ denotes the form of the criterion when the underlying model is the unrestricted VECM (Equation 6). The statistic τ_1 compares the null model of the chosen lag order $\widetilde{p}$ with the model having the highest PIC value amongst those in the class of unrestricted (bivariate) VECMs. On the other hand, τ_2 compares the same null model with an unrestricted VECM of the same order $(\widetilde{p}, 1)$.

We test the expectations theory for different specifications of the discount rate R. For the results in Table 3, we take $R = .0637/12$ for the short sample period and $R = .07911/12$ for the full sample period, where these discount rates are calculated as the mean monthly rate for the 10-year government bond for the respective sample periods. Similar methods for specifying R were employed in Campbell and Shiller (1987). For the results in Table 4, we leave R unrestricted and estimate it from the data jointly with the other parameters.

Tables 3 and 4 and Figures 3 and 4 report our findings. The results are unambiguous; for both the short sample period and the full sample period, and regardless of the way R is specified, our tests reject the restrictions imposed by the present-value model of the term structure of interest rates. Moreover, inspection of the τ_2 statistic shows that the present-value model of the chosen lag order $\widetilde{p}$ also does not compare favorably with an unrestricted VECM with the same lag order and one cointegrating vector.[4] This observation suggests that rejection of the expectations theory here is not a rejection of the cointegration part of the set of restrictions of the present-value model, but rather it is caused by violation of the other restrictions of the model, i.e., Equations 7–9 in Section 2. This result is also consistent with our cointegration analysis reported earlier which, on the whole, finds cointegration between y_{1t} and y_{2t} while also noting the considerable uncertainty surrounding that inference.

[4]Indeed, for the short sample period in both Tables 3 and 4, $\tau_1 = \tau_2$ since, in this case, the unrestricted model with the highest PIC value happens to have the same order $(p, r) = (3, 1)$ as the restricted model of the chosen lag order $\widetilde{p}$.

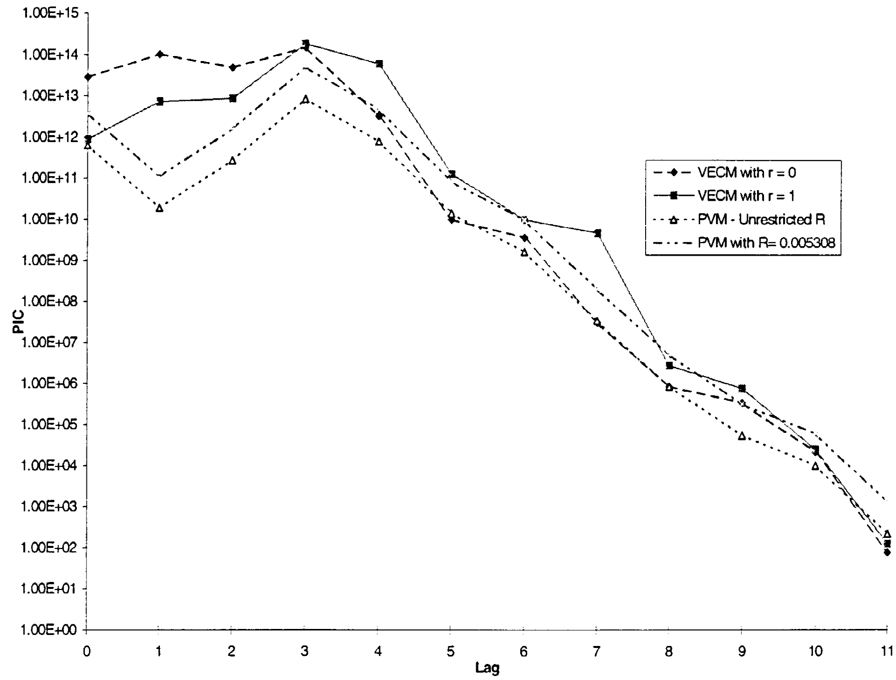

Figure 3
PIC values for the spread between y_2 and y_1: 1964:1–1978:8.

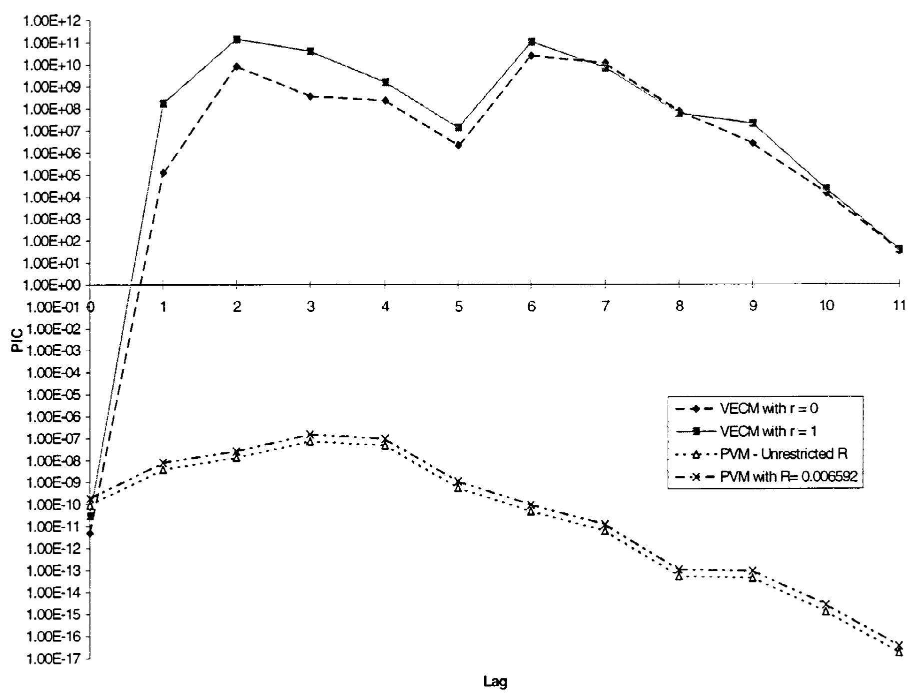

Figure 4
PIC values for the spread between y_2 and y_1: 1964:1–1997:1.

Figures 3 and 4 plot PIC values for four models (VECM with $r = 0$, VECM with $r = 1$, present-value model with R given by the mean 10-year bond rate, and the present-value model with R unrestricted) against different lag specifications. These figures show that our results are fairly robust to lag selection. For the short sample period, one finds some evidence for the present-value model only if one is willing to condition on lag selection greater than or equal to eight.[5] The results for the full sample period are even more robust, since only with a lag selection of $p = 0$ will one find evidence in favor of the present-value model; but this lag selection is quite unlikely, as PIC values (for all four models) at this lag are much smaller than those at other lags.

Our results are, on the whole, in agreement with the results of Campbell and Shiller (1987) and Toda (1991), both of whom using a different data set reject the expectations theory using classical statistical procedures. However, Campbell and Shiller (1987) have argued that classical tests of the present-value model may be so sensitive to departures from the expectations theory that they obscure some of the theory's merits. That is, the expectations hypothesis may be rejected even when it explains much of the variations in the long-term interest rate.

Hence, in addition to classical tests of hypothesis, they also propose the use of informal procedures to evaluate the "fit" of the present-value model, and find the informal procedures they use to yield encouraging evidence for the expectations theory. It is interesting to note that Chao (1994) applies the model-selection procedure used here to the Campbell-Shiller data set, and obtains results favorable to the expectations theory. Qualitatively, the inference given in Chao (1994) is very similar to that of Campbell and Shiller (1987), based on their informal procedures. This is not surprising, since model-selection methods are, in fact, procedures that evaluate the "goodness of fit" of models while making the appropriate trade-offs on model size. Hence, in our case, the mere fact that even model-selection methods reject the present-value model suggests that the evidence against the expectations theory, at least for the particular data set used here, is very strong.

5 Conclusion

This paper puts forth a model-selection approach to testing the expectations theory of the term structure of interest rates. Our method is based on the information criterion, PIC, as developed and analyzed in Phillips and Ploberger (1994, 1996) and extended to provide order estimation of cointegrating rank by Chao and Phillips (1997). This procedure has the advantage that it allows joint estimation of the lag order and cointegrating rank of a VAR system, and the order estimators obtained from this procedure are completely consistent in the sense that the probabilities of type-I and type-II errors both go to zero asymptotically. In addition, bar charts of the PIC statistics for models of different order combinations can easily be constructed so that one may assess the robustness of the inference with respect to lag order and cointegrating rank.

Applying PIC to monthly time-series data on the Treasury bill rate and a 10-year government bond yield supplied by the publication *International Financial Statistics*, we find evidence for cointegration of these interest rates, although bar charts of PIC values reveal considerable uncertainty about this inference. Testing the full set of restrictions implied by the present-value model of the term structure results in the rejection of this model by our model-selection procedure. Moreover, the rejection of the expectations theory is found to be fairly robust to lag selection, as it can be overturned only if one conditions upon lag orders that are extremely small ($p = 0$) or reasonably large ($p \geq 8$). These lag specifications are considered unlikely, given the small PIC values associated with them. These results are, on the whole, in agreement with those obtained by Campbell and Shiller (1987) and Toda (1991) using a different data set, as these studies also reject the expectations theory using classical tests of hypothesis.

[5] More precisely, we find evidence in favor of the present-value model for $p = 8, 10$, and 11, but not for $p = 9$, since in the latter case the unrestricted VECM with $r = 1$ has a greater PIC value than any of the other models.

References

Ahn, S. K., and G. C. Reinsel (1980). "Estimation of partially nonstationary multivariate autoregressive models." *Journal of the American Statistical Association*, 85:813–823.

Campbell, J. Y., and R. J. Shiller (1987). "Cointegration and tests of present value models." *Journal of Political Economy*, 95:1062–1088.

Campbell, J. Y., and R. J. Shiller (1989). "The dividend-price ratio and expectations of future dividends and discount factors." *Review of Financial Studies*, 1:195–228.

Campbell, J. Y., and R. J. Shiller (1991). "Yield spreads and interest rate movements: A bird's eye view." *Review of Economic Studies*, 58:495–514.

Chao, J. C. (1994). "Essays in Bayesian econometrics." PhD Thesis, Department of Economics, Yale University.

Chao, J. C., and P. C. B. Phillips (1997). "Model selection in partially nonstationary vector autoregressive processes with reduced rank structure." Discussion Paper 1155, Cowles Foundation.

Davidson, J., D. Hendry, F. Srba, and S. Yeo (1978). "Econometric modeling of the aggregate time series relationship between consumers' expenditure and income in the United Kingdom." *Economic Journal*, 88:661–692.

Engle, R. F., and C. W. J. Granger (1987). "Cointegration and error-correction: Representation, estimation, and testing." *Econometrica*, 55:251–276.

Granger, C. W. J. (1981). "Some properties of time series data and their use in economic model specification." *Journal of Econometrics*, 16:121–130.

Granger, C. W. J., and A. A. Weiss (1983). "Time series analysis of error-correcting models." In *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press, pp. 255–278.

Hansen, L. P., and T. Sargent (1981). "Exact linear rational expectations models: Specification and estimation." Staff Report 71, Federal Reserve Bank, Minneapolis, MN.

Johansen, S. (1992). "Determination of cointegrating rank in the presence of a linear trend." *Oxford Bulletin of Economics and Statistics*, 54:383–397.

Phillips, P. C. B. (1996). "Econometric model determination." *Econometrica*, 64:763–812.

Phillips, P. C. B. (1991). "Optimal inference in cointegrated systems." *Econometrica*, 59:283–306.

Phillips, P. C. B., and W. Ploberger (1994). "Posterior odds testing for a unit root with data-based model selection." *Econometric Theory*, 10:774–808.

Phillips, P. C. B., and W. Ploberger (1996). "An asymptotic theory of Bayesian inference for time series." *Econometrica*, 64:381–412.

Reinsel, G. C. and S. K. Ahn (1992). "Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting." *Journal of Time Series Analysis*, 13:353–375.

Sawa, T. (1978). "Information criteria for discriminating among alternative regression models." *Econometrica*, 46:1273–1291.

Shibata, R. (1976). "Selection of the order of an autoregressive model by Akaike's information criterion." *Biometrika*, 63:117–126.

Toda, H. Y. (1991). "An ECM approach to tests of present value models." Mimeograph, Yale University.

Toda, H. Y., and P. C. B. Phillips (1993). "Vector autoregressions and causality." *Econometrica*, 61:1367–1393.

Tsay, R. S. (1984). "Order selection in nonstationary autoregressive models." *Annals of Statistics*, 12:1425–1433.

Editorial Policy

The *SNDE* is formed in recognition that advances in statistics and dynamical systems theory may increase our understanding of economic and financial markets. The journal will seek both theoretical and applied papers that characterize and motivate nonlinear phenomena. Researchers will be encouraged to assist replication of empirical results by providing copies of data and programs online. Algorithms and rapid communications will also be published.

ISSN 1081-1826