Cowles Foundation Paper 943
Reprinted from Jack C. Lee, Wesley O. Johnson and Arnold Zellner (eds.), Modelling and Prediction, 1996

An Empirical Bayesian Approach to Cointegrating Rank Selection and Test of the Present Value Model for Stock Prices

John C. Chao
Department of Economics, University of Maryland
College Park, Maryland 20742

Peter C. B. Phillips
Cowles Foundation for Research in Economics, Yale University
P.O. Box 208281, New Haven, Connecticut 06520-8281

Abstract

This paper provides an empirical Bayesian approach to the problem of jointly estimating the lag order and the cointegrating rank of a partially nonstationary reduced rank regression. The method employed is a variant of the Posterior Information Criterion (PIC) of Phillips and Ploberger (1994, 1995) and is similar to the asymptotic predictive odds version of the PIC criterion given in Phillips (1994). Here, we use a proper (Gaussian) prior whose hyperparameters are estimated from an initial subsample of the data. The form of the prior is suggested by the asymptotic posterior distribution of the parameters of the model, and, hence, the criterion can be interpreted as an approximate predictive odds ratio in the case where the sample size is large. Applying this procedure to the extended Campbell-Shiller data set for stock prices and dividends, we find the present value model for stock prices to be inconsistent with the data.

1 Introduction

Order estimation of the lag dimension of a model is an unavoidable task in applied econometric work involving vector autoregressions (VAR's). In VAR models that are partially nonstationary in the sense of Ahn and Reinsel (1990), it is often desirable to estimate not only the lag order of the model but also the number of linearly independent cointegrating vectors, or the cointegrating rank. In conventional econometric practice, however, the estimation of the lag length of a VAR model is often done with statistical methodologies that are very different from those employed for the determination of the cointegrating rank. While information criteria, such as AIC and BIC, are often favored by researchers for lag estimation (but see Pötscher, 1983, for an alternative method based on LM tests), cointegrating rank determination is most often performed using a sequence of classical tests that come within the Neyman–Pearson tradition (see Johansen, 1992).

Recently, Phillips (1992), Chao and Phillips (1994), and Phillips (1994) have argued that cointegrating rank determination is most naturally a problem of order selection. Applying the Posterior Information Criterion (PIC) of Phillips and Ploberger (1994, 1995), they developed statistical procedures which allow for the joint estimation of the lag order and the cointegration rank in a VAR system. Such procedures are particularly appealing in the light of simulation evidence, presented in Toda and Phillips (1994) and Chao (1995), that shows the performance of classical tests of cointegration, such as those put forth by Johansen (1988, 1992), to be sensitive to autoregressive lag specification. In the present paper, we develop a variant of PIC, which is similar in spirit to the predictive odds ratios of Atkinson (1978), O'Hagan (1991), and Geweke (1994) and to the predictive form of PIC given in Phillips (1994). We derive our criterion using a proper Gaussian prior whose hyperparameters are estimated using an initial subsample of the data. As the form of the prior is suggested by the asymptotic posterior distribution of the parameters of the model, the criterion can be interpreted as an approximate predictive odds ratio in the case where the sample size is large.

The second objective of this paper is to illustrate the use of this model selection criterion through an empirical application that tests the rational expectations present value model for stock prices. An important recent application of the technology of nonstationary time series analysis to empirical economic research has been the work of Campbell and Shiller (1987). In their paper, Campbell and Shiller used classical tests of unit roots and cointegration to address issues, raised by Kleidon (1986) and Marsh and Merton (1986), pertaining to possible nonstationarity in the price and dividend processes. They showed that if both stock prices and dividends are I(1) processes (i.e., stationary in first-order differences), then one implication of the present value model is that these variables are cointegrated. Following the suggestion of Engle and Granger (1987), they use a preliminary estimate of the cointegrating vector to transform their bivariate VAR into a stationary system, where conventional statistical procedures can be applied to test the restrictions implied by the present value model.

An important statistical issue which arises in testing the present value model within the VAR framework of Campbell and Shiller (1987) is the specification of the lag order of the system. Since economic theory offers little guidance in this regard, Campbell and Shiller (1987) used the Akaike Information Criterion (AIC) to pre-select the lag length. However, it is well-known from the results of Shibata (1976) and Sawa (1978) that AIC has a tendency to overestimate the lag order. Moreover, Toda (1991) found classical tests of the present value models to be sensitive to variations in lag selection.

Given the need for pre-selection of the lag order and given the sensitivity of conventional testing procedures to lag specification, the use of our model selection criterion has certain advantages. First, it allows us to jointly select the lag and cointegrating rank order of a VAR and to test the implications of the present value model simultaneously in one coherent framework. Such a joint selection procedure is preferred over a sequential procedure which estimates cointegrating rank conditional on some preliminary lag selection because with sequential procedures, there is always some pre-test bias and associated implications for inference. Moreover, the joint-selection procedure makes comparison across the full array of models and gives consistent order estimates of cointegrating rank and lag order, as shown in Chao and Phillips (1994). A further advantage of our approach is that bar charts and histograms of posterior probabilities for models of different dimensions can be readily constructed so that one may assess the robustness of the inferences with respect to lag length and cointegrating rank.

The paper proceeds as follows. In Section 2, we introduce the time series model to be studied and describe our model selection procedure PIC. Section 3 gives a discussion of the rational expectations present value model for the stock market and the restrictions that it imposes on a VAR in error-correction form. Section 4 presents our empirical results. Finally, we offer some concluding thoughts in the fifth and final section.

2 Order Selection and Hypothesis Testing in VAR Models via Posterior Odds

2.1 The Partially Nonstationary VAR Model

The model framework we consider in this paper is similar to that of Chao and Phillips (1994). The setup is the m-dimensional vector autoregressive model of $(p+1)$-order:

$$y_t = \mu + \sum_{i=1}^{p+1} A_i y_{t-1} + \varepsilon_t \ . \tag{1}$$

It is well-known that equation (1) can be rewritten as an error-correction model (ECM):

$$\Delta y_t = \mu + \sum_{i=1}^{p} A_i^* \Delta y_{t-i} + A_* y_{t-1} + \varepsilon_t \ , \tag{2}$$

where $A_i^* = -\sum_{j=i+1}^{p+1} A_j$ and $A_* = \sum_{i=1}^{p+1} A_i - I_m$. We further assume that the following conditions are applicable to our model:

(i) $\det\left[I_m - \sum_{i=1}^{p+1} A_i L^i\right] = 0$ implies that either $L = 1$ or $|L| > 1$.

(ii) $A_* = \alpha\beta'$, where α and β are $m \times r$ matrices of full column rank r, $0 \leq r \leq m$. (If $r = 0$, we take $\alpha = \beta = 0$, and if $r = m$, we take $\alpha = A_*$ and $\beta = I_m$.)

(iii) $\alpha_\perp' \left(\sum_{i=1}^{p} A_i^* - I_m\right)\beta_\perp$ is nonsingular for $0 \leq r < m$, where $\alpha_\perp$ and $\beta_\perp$ are $m \times (m-r)$ matrices of full column rank $m-r$ such that $\alpha_\perp'\alpha = 0 = \beta_\perp'\beta$. (If $r = 0$, we take $\alpha_\perp = \beta_\perp = I_m$.)

(iv) $\varepsilon \equiv$ i.i.d. $N(0, \Omega)$.

Under these assumptions, $\{y_t\}$ is I(1), but $\beta' y_t$ is I(0) with r linearly independent cointegrating vectors. Thus, in the nomenclature of the literature on cointegration, we say that the multivariate system defined by (2) has a cointegrating rank of r. Note that without further restrictions, α and β are unidentified. To achieve identification, we follow Ahn and Reinsel (1990) in selecting a normalized parameterization in which

$\beta' = [I_r,\ B]$. We shall refer to equation (2) as a reduced rank regression of order (p, r), or RRR(p, r) for short.

2.2 The Posterior Information Criterion in Asymptotic Predictive Form

Our objective in this paper is to jointly estimate the cointegrating rank r and the lag order p of the model (2). As in Chao and Phillips (1994), the criterion which we use for model selection is a posterior odds ratio. However, unlike the earlier paper where a uniform prior was employed, we use here a Gaussian prior whose hyperparameters are estimated from an initial subsample of the data. To be explicit, let us begin by defining a class of competing models $M_{p,r}(p = 0, 1, \ldots, \overline{p};\ r = 0, 1, \ldots, \overline{r})$, given in terms of the parameterized measure $\mathbb{P}_T^{\theta_{p,r}}$ with $\theta_{p,r}$ belonging to some parameter space $\Theta_{p,r}$. Here, $M_{p,r}$ denotes the model with cointegrating rank r and order of lagged differences p. Given the data $\mathbf{y} = \{y_t\}_1^T$, we let $L_T(\theta_{p,r}) = d\mathbb{P}_T^{\theta_{p,r}}/d\nu$ denote the likelihood function of the model $M_{p,r}$ with respect to the Lebesgue measure ν. Suppose we take the prior model probability of $M_{p,r}$ and the prior density of $\theta_{p,r}$ to be $\pi_{p,r}$ and $g(\theta_{p,r}|M_{p,r})$ respectively; then, the posterior odds for the competing models M_{p_0,r_0} and M_{p_1,r_1} is defined as

$$\frac{\Pi_T(M_{p_0,r_0}|\mathbf{y})}{\Pi_T(M_{p_1,r_1}|\mathbf{y})} = \frac{\pi_{p_0,r_0}\overline{L}_T(M_{p_0,r_0})}{\pi_{p_1,r_1}\overline{L}_T(M_{p_1,r_1})}\ , \tag{3}$$

where

$$\overline{L}_T(M_{p,r}) = \int_{\Theta_{p,r}} L_T(\theta_{p,r})g(\theta_{p,r}|M_{p,r})d\theta_{p,r} \tag{4}$$

is the Bayesian data density or the marginalized likelihood. Under the additional assumption of a symmetric loss function, which penalizes Type I and Type II errors equally, we obtain the decision rule:

$$\frac{\Pi_T(M_{p_0,r_0}|\mathbf{y})}{\Pi_T(M_{p_1,r_1}|\mathbf{y})} > 1\ , \quad \text{then decide in favor of} \quad M_{p_0,r_0}\ . \tag{5}$$

To implement this posterior odds test for the model in Section 2.1, we must first specify the prior density $g(\theta_{p,r}|M_{p,r})$. Following an approach used in earlier work (e.g., Atkinson (1978), O'Hagan (1991), Geweke (1994), and Phillips (1994)), we divide the sample into two sample periods $[1,\ T_0]$ and $[T_0+1,\ T]$, where $T_0 = [\rho T]$ (the integer part of ρT) for some constant $\rho \in (0,1)$, and use data from the initial sample period $[1,\ T_0]$ to assist in the construction of our prior. Let $\theta_{p,r} = (\phi',\ \omega')'$ $= (\text{vec}(B)',\ \text{vec}(\alpha)',\ \text{vec}(A^*)',\ \omega')'$, where ω is the vector of nonredundant elements of Ω and $A^* = (\mu,\ A_1^*,\ A_2^*,\ \ldots,\ A_p^*)$, and let the prior densities of the model $M_{p,r}$ be

$$g(\Omega|M_{p,r}) = |\Omega|^{-\frac{1}{2}(m+1)} \tag{6}$$

$$g(\phi|M_{p,r}) = (2\pi)^{-\frac{1}{2}(2mr-r^2+m(mp+1))}|V|^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2}(\phi-\overline{\phi})'V(\phi-\overline{\phi})\right\}\ . \tag{7}$$

The hyperparameters $\overline{\phi}$ and V can be estimated using data from the period $[1,\ T_0]$. More specifically, we estimate ϕ by its maximum likelihood estimate $\widehat{\phi}_{T_0}$

$= (\mathrm{vec}(\widehat{B}_{T_0})', \mathrm{vec}(\widehat{\alpha}_{T_0})', \mathrm{vec}(\widehat{A}^*_{T_0})')'$ and V by the asymptotic formula:

$$\widehat{V}_{T_0} = \begin{pmatrix} (\widehat{\alpha}'_{T_0}\widehat{\Omega}^{-1}_{T_0}\widehat{\alpha}_{T_0}\otimes F'Y'_{-1,T_0}Y_{-1,T_0}F) & 0 & 0 \\ 0 & (\widehat{\Omega}^{-1}_{T_0}\otimes\widehat{\beta}'_{T_0}Y'_{-1,T_0}Y_{-1,T_0}\widehat{\beta}_{T_0}) & (\widehat{\Omega}^{-1}_{T_0}\otimes\widehat{\beta}'_{T_0}Y'_{-1,T_0}W_{T_0}) \\ 0 & (\widehat{\Omega}^{-1}_{T_0}\otimes W'_{T_0}Y_{-1,T_0}\widehat{\beta}_{T_0}) & (\widehat{\Omega}^{-1}_{T_0}\otimes W'_{T_0}W_{T_0}) \end{pmatrix} \quad (8)$$

where $\widehat{\beta}_{T_0} = [I_r, \widehat{B}_{T_0}]'$ and $\widehat{\Omega}_{T_0} = (1/(T_0+m+1))\sum_{t=1}^{T_0}(\Delta y_t - \widehat{A}^*_{T_0}W_t - \widehat{\alpha}_{T_0}\widehat{\beta}'_{T_0}y_{t-1})'(\Delta y_t - \widehat{A}^*_{T_0}W_t - \widehat{\alpha}_{T_0}\widehat{\beta}'_{T_0}y_{t-1})$. Here, we have let $Y_{-1,T_0} = (y_0, ..., y_{T_0-1})'$ and $W_{T_0} = (W_1, ..., W_{T_0})'$, with $W_t = (1, \Delta y'_{t-1}, ..., \Delta y'_{t-p})'$, while $F' = [0, I_{m-r}]$ is an $(m-r)\times m$ matrix. Note that $\widehat{V}_{T_0}$ is an estimator of the precision matrix of the asymptotic posterior distribution of ϕ, and, hence, for large T_0, our prior density,

$$g(\phi|M_{p,r},\widehat{\phi}_{T_0},\widehat{V}_{T_0}) = (2\pi)^{-\frac{1}{2}(2mr-r^2+m(mp+1))}|\widehat{V}_{T_0}|^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(\phi-\widehat{\phi}_{T_0})'\widehat{V}_{T_0}(\phi-\widehat{\phi}_{T_0})\right\}, \quad (9)$$

can be interpreted as the approximate (large sample) posterior distribution for the initial sample period $[1, T_0]$ under a uniform prior.

Given our assumption (iv), the likelihood function for the sample period $[T_0+1, T]$ can be written as

$$L_{T_1}(\theta_{p,r}) = (2\pi)^{\frac{-mT_1}{2}}|\Omega|^{\frac{-T_1}{2}}\exp\left\{-\tfrac{1}{2}\sum_{t=T_0+1}^{T}(\Delta y_t - A^*W_t - \alpha\beta' y_{t-1})'\Omega^{-1}(\Delta y_t - A^*W_t - \alpha\beta' y_{t-1})\right\}, \quad (10)$$

where $T_1 = T - T_0 = T - [\rho T]$ and α and β are, as before, $m\times r$ matrices. Combining the likelihood function (10) with the prior densities (6) and (9) and integrating over the parameters $\theta_{p,r} = (\phi,\omega)$ using the Laplace approximation method as in Phillips and Ploberger (1994) and Chao and Phillips (1994), we obtain, up to a multiplicative constant (not involving p and r), the (approximate) Bayesian data density or marginalized likelihood:[1]

$$\overline{L}_{T_1}(M_{p,r}) \sim |\widehat{\Omega}_{T_1}|^{\frac{-T_1}{2}}\left(|\widehat{V}_{T_0}+\widehat{V}_{T_1}|/|\widehat{V}_{T_0}|\right)^{-\frac{1}{2}}\exp\left\{-\tfrac{1}{2}(\widehat{\phi}_{T_1}-\widehat{\phi}_{T_0})'\widehat{V}_{T_0}(\widehat{\phi}_{T_1}-\widehat{\phi}_{T_0})\right\} = \widehat{\overline{L}}_{T_1}(M_{p,r}) \text{ (say)}, \quad (11)$$

where

$$\widehat{\Omega}_{T_1} = \frac{1}{T_1+m+1}\sum_{t=T_0+1}^{T}\left(\Delta y_t - \widehat{A}^*_{T_1}W_t - \widehat{\alpha}_{T_1}\widehat{\beta}'_{T_1}y_{t-1}\right)'(\Delta y_t - \widehat{A}^*_{T_1}W_t - \widehat{\alpha}_{T_1}\widehat{\beta}'_{T_1}y_{t-1}), \quad (12)$$

and

$$\widehat{V}_{T_1} = \begin{pmatrix} (\widehat{\alpha}'_{T_1}\widehat{\Omega}^{-1}_{T_1}\widehat{\alpha}_{T_1}\otimes F'Y'_{-1,T_1}Y_{-1,T_1}F) & 0 & 0 \\ 0 & (\widehat{\Omega}^{-1}_{T_1}\otimes\widehat{\beta}'_{T_1}Y'_{-1,T_1}Y_{-1,T_1}\widehat{\beta}_{T_1}) & (\widehat{\Omega}^{-1}_{T_1}\otimes\widehat{\beta}'_{T_1}Y'_{-1,T_1}W_{T_1}) \\ 0 & (\widehat{\Omega}^{-1}_{T_1}\otimes W'_{T_1}Y_{-1,T_1}\widehat{\beta}_{T_1}) & (\widehat{\Omega}^{-1}_{T_1}\otimes W'_{T_1}W_{T_1}) \end{pmatrix}. \quad (13)$$

[1] See the cited references for details of the Laplace derivation.

Here, $Y_{-1,T_1} = [y_{T_0}, ..., y_{T-1}]'$, $W_{T_1} = [W_{T_0+1}, ..., W_T]'$, and the matrices $\widehat{A}^*_{T_1}$, $\widehat{\alpha}_{T_1}$, and $\widehat{\beta}_{T_1}$ are the posterior modes of A^*, α, and β, where the posterior distribution has been updated by the likelihood function (10). A further approximation is possible by taking the transformation

$$-\frac{2}{T_1}\ln\left(\widehat{\widehat{L}}_{T_1}(M_{p,r})\right) \sim \ln|\widehat{\Omega}_{T_1}| + \frac{1}{T_1}\ln\left(|\widehat{V}_{T_0} + \widehat{V}_{T_1}|/|\widehat{V}_{T_0}|\right) , \tag{14}$$

where the term $(1/T_1)(\widehat{\phi}_{T_1} - \widehat{\phi}_{T_0})'\widehat{V}_{T_0}((\widehat{\phi}_{T_1} - \widehat{\phi}_{T_0}) = O_p(T_1^{-1})$ is neglected. If, in addition, we set the prior model probabilities equal across all models so that $\pi_{p,r} = 1/[(\overline{p}+1)(\overline{r}+1)]$, then we can define our information criterion either in terms of expression (11) or in terms of expression (14) as follows:

$$(\widehat{p},\widehat{r}) = \text{argmax PIC}(p,r) , \tag{15}$$

where

$$\text{PIC}(p,r) = \widehat{\widehat{L}}_{T_1}(M_{p,r})/\widehat{\widehat{L}}_{T_1}(M_{\overline{p},\overline{r}}) \tag{16}$$

with $\widehat{\widehat{L}}_{T_1}(M_{p,r})$ as defined in expression (11). Alternatively, we can write this criterion in terms of (14) as

$$(\widehat{p},\widehat{r}) = \text{argmin PIC}'(p,r) , \tag{17}$$

where

$$\text{PIC}'(p,r) = \ln|\widehat{\Omega}_{T_1}| + \frac{1}{T_1}\ln(|\widehat{V}_{T_0} + \widehat{V}_{T_1}|/|\widehat{V}_{T_0}|) . \tag{18}$$

Note that for large values of T_0 and T_1, expression (18) and, especially, expression (16) can be interpreted as criteria that are based on the predictive odds ratio. Note also that expression (18) is in a form analogous to other information criteria (for example — AIC, BIC, and the Fisher information criterion (FIC) of Wei (1992)) in that the right-hand side of expression (18) is comprised of two terms, with the first term being a measure of the goodness of fit and the second term being a penalty function reflecting the complexity of the model. This formulation corresponds closely to the asymptotic predictive odds criterion used in Phillips (1994).

2.3 Hypothesis Testing of Linear Restrictions

The procedure we outline in the last subsection can be readily extended to test linear restrictions of the form:

$$H(M^R) \; : \; \text{vec}(A^*) = Sd + s \;\text{ and }\; \text{vec}(\alpha) = Gc + g , \tag{19}$$

where S, G, s, g, d, and c are respectively an $(m + m^2p) \times q_1$ restriction matrix of rank q_1, an $mr \times q_2$ restriction matrix of rank q_2, an $(m + m^2p) \times 1$ vector of known constants, an $mr \times 1$ vector of known constants, a q_1-vector of basic parameters, and a q_2-vector of basic parameters. Under the hypothesis (19), we have the following (restricted) ECM formulation:

$$\Delta y_t = (I_m \otimes W_t')(Sd + s) + \alpha(c)\beta' y_{t-1} + \varepsilon_t \tag{20}$$

or

$$z_t = X_t d + \alpha(c)\beta' y_{t-1} + \varepsilon_t \ , \tag{21}$$

where $X_t = (I_m \otimes W_t')S$ and $z_t = \Delta y_t - (I_m \otimes W_t')s$. Note that equations (20) and (21) describe a cointegrated system with additional restrictions imposed on the coefficients of its stationary components. Analogous to (9), we take the prior density of $\gamma = (\mathrm{vec}(B)', \, d', \, c')'$ to be

$$g(\gamma | M_{p,r}^R, \, \widetilde{\gamma}_{T_0}, \, \widetilde{P}_{T_0}) = (2\pi)^{-\frac{1}{2}(q_1+q_2)} |\widetilde{P}_{T_0}|^{\frac{1}{2}} \exp\left\{ -\tfrac{1}{2}(\gamma - \widetilde{\gamma}_{T_0})' \widetilde{P}_{T_0} (\gamma - \widetilde{\gamma}_{T_0}) \right\} \ , \tag{22}$$

where $\widetilde{\gamma}_{T_0} = (\mathrm{vec}(\widetilde{B}_{T_0})', \, \widetilde{d}_{T_0}', \, \widetilde{c}_{T_0}')'$ is the posterior mode (or the maximum likelihood estimate) of γ over the sample period $[1, \, T_0]$ and

$$\widetilde{P}_{T_0} = \begin{pmatrix} \begin{matrix} (\alpha(\widetilde{c}_{T_0})'\widetilde{\Omega}_{T_0}^{-1}\alpha(\widetilde{c}_{T_0}) \\ \otimes F'Y_{-1,T_0}'Y_{-1,T_0}F) \end{matrix} & 0 & 0 \\ 0 & G'(\widetilde{\Omega}_{T_0}^{-1} \otimes \widetilde{\beta}_{T_0}' Y_{-1,T_0}' Y_{-1,T_0} \widetilde{\beta}_{T_0})G & G'(\widetilde{\Omega}_{T_0}^{-1} \otimes \widetilde{\beta}_{T_0}' Y_{-1,T_0}' W_{T_0})S \\ 0 & S'(\widetilde{\Omega}_{T_0}^{-1} \otimes W_{T_0}' Y_{-1,T_0} \widetilde{\beta}_{T_0})G & S'(\widetilde{\Omega}_{T_0}^{-1} \otimes W_{T_0}' W_{T_0})S \end{pmatrix} \tag{23}$$

where $\widetilde{\beta}_{T_0} = [I_r, \, \widetilde{B}_{T_0}]'$ and $\widetilde{\Omega}_{T_0} = (1/(T_0 + m + 1)) \sum_{t=1}^{T_0} (z_t - X_t \widetilde{d}_t - \alpha(\widetilde{c}_{T_0})\widetilde{\beta}_{T_0}' y_{t-1})'$ $(z_t - X_t \widetilde{d}_{T_0} - \alpha(\widetilde{c}_{T_0})\widetilde{\beta}_{T_0}' y_{t-1})$. As before, we take our prior density for Ω to be (6), but our likelihood function under the restricted model $M_{p,r}^R$ now becomes

$$\begin{aligned} & L_T^R(B, d, c, \omega) \qquad\qquad (24) \\ = \; & (2\pi)^{\frac{-mT_1}{2}} |\Omega|^{\frac{-T_1}{2}} \exp\left\{ -\tfrac{1}{2} \sum_{t=T_0+1}^{T} (z_t - X_t d - \alpha(c)\beta' y_{t-1})' \Omega^{-1} (z_t - X_t d - \alpha(c)\beta' y_{t-1}) \right\} . \end{aligned}$$

Combining the likelihood function (24) with the prior densities (6) and (22) and integrating over the parameters (B, d, c, ω) using Laplace's method, we obtain, corresponding to (11), the (approximate) Bayesian data density or marginalized likelihood for the restricted model $M_{p,r}^R$:

$$\begin{aligned} \overline{L}_{T_1}(M_{p,r}^R) & \sim |\widetilde{\Omega}_{T_1}|^{\frac{-T_1}{2}} (|\widetilde{P}_{T_0} + \widetilde{P}_{T_1}| / |\widetilde{P}_{T_0}|)^{-\frac{1}{2}} \exp\left\{ -\tfrac{1}{2} (\widetilde{\gamma}_{T_1} - \widetilde{\gamma}_{T_0})' \widetilde{P}_{T_0} (\widetilde{\gamma}_{T_1} - \widetilde{\gamma}_{T_0}) \right\} \qquad (25) \\ & = \widehat{\overline{L}}_{T_1}(M_{p,r}^R) \text{ (say)}, \end{aligned}$$

where

$$\widetilde{\Omega}_{T_1} = \frac{1}{T_1 + m + 1} \sum_{t=T_0+1}^{T} \left(z_t - X_t \widetilde{d}_{T_1} - \alpha(\widetilde{c}_{T_1}) \widetilde{\beta}_{T_1}' y_{t-1} \right)' (z_t - X_t \widetilde{d}_{T_1} - \alpha(\widetilde{c}_{T_1}) \widetilde{\beta}_{T_1}' y_{t-1}) \tag{26}$$

and

$$\widetilde{P}_{T_1} = \begin{pmatrix} \begin{matrix} (\alpha(\widetilde{c}_{T_1})'\widetilde{\Omega}_{T_1}^{-1}\alpha(\widetilde{c}_{T_1}) \\ \otimes F'Y_{-1,T_1}'Y_{-1,T_1}F) \end{matrix} & 0 & 0 \\ 0 & G'(\widetilde{\Omega}_{T_1}^{-1} \otimes \widetilde{\beta}_{T_1}' Y_{-1,T_1}' Y_{-1,T_1} \widetilde{\beta}_{T_1})G & G'(\widetilde{\Omega}_{T_1}^{-1} \otimes \widetilde{\beta}_{T_1}' Y_{-1,T_1}' W_{T_1})S \\ 0 & S'(\widetilde{\Omega}_{T_1}^{-1} \otimes W_{T_1}' Y_{-1,T_1} \widetilde{\beta}_{T_1})G & S'(\widetilde{\Omega}_{T_1}^{-1} \otimes W_{T_1}' W_{T_1})S \end{pmatrix} \tag{27}$$

$\widetilde{\beta}_{T_1}$, $\widetilde{d}_{T_1}$, and $\widetilde{c}_{T_1}$ are the posterior modes of β, d, and c, where the posterior distribution has been updated by the likelihood function (24).

Expression (25) enables us to make decisions on the lag order and the cointegrating rank of the system simultaneously with decisions about the validity of the restrictions represented by (19). Let $H(M^R)$ be the null hypothesis defined by (19) and let $H(M^U)$ be the alternative of an unrestricted ECM as defined in Section 2.2, then the decision rule for choosing M^R over M^U can be stated as:

$$\text{Accept } H(M^R) \text{ in favor of } H(M^U) \text{ if } \widehat{\widehat{L}}_{T_1}(M^R_{\tilde{p},\tilde{r}})/\widehat{\widehat{L}}_{T_1}(M^U_{\hat{p},\hat{r}}) > 1 , \tag{28}$$

where

$$\begin{aligned} (\hat{p},\hat{r}) &= \text{argmax PIC}U(p,r) , \\ \text{PIC}U(p,r) &= \widehat{\widehat{L}}_{T_1}(M^U_{p,r})/\widehat{\widehat{L}}_{T_1}(M^U_{\overline{p},\overline{r}}) , \end{aligned} \tag{29}$$

and

$$\begin{aligned} (\tilde{p},\tilde{r}) &= \text{argmax PIC}R(p,r) , \\ \text{PIC}R(p,r) &= \widehat{\widehat{L}}_{T_1}(M^R_{p,r})/\widehat{\widehat{L}}_{T_1}(M^U_{\overline{p},\overline{r}}) . \end{aligned} \tag{30}$$

Note that $\tilde{p}$ and $\tilde{r}$ are the estimated lag and cointegrating rank order of a (possibly) cointegrated system having additional restrictions of the form (19) and, thus, may be different from $\hat{p}$ and $\hat{r}$, which are the order estimates of a (possibly) cointegrated system having no additional restriction.

3 The Present Value Model and its Testable Implications

In this section, we briefly describe the present value model 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 which will be relevant for our subsequent empirical analysis. Formally, the present value model can be written as:

$$y_{2t} = \theta(1-\delta)\sum_{i=0}^{\infty}\delta^i E(y_{1t+i}|I_t) + \text{const}, \tag{31}$$

where $E(\cdot|I_t)$ denotes the mathematical expectation conditional on the full public information I_t at time t and where y_{1t} and y_{2t} are, respectively, the dividend and stock price at time t. Here, as elsewhere in this paper, we treat conditional expectations as being equivalent to linear projections on information. Moreover, in the context of the stock market, $\theta = \delta/(1-\delta)$ and const is restricted to be zero so that equation (31) has the simplified form:

$$y_{2t} = \sum_{i=0}^{\infty}\delta^{i+1}E(y_{1t+i}|I_t) . \tag{32}$$

We follow Campbell and Shiller (1987) in defining the random variable $s_t = y_{2t} - \theta y_{1t}$, which they referred to as the "spread." Subtracting θy_{1t} from both sides of equation (32) and rearranging the terms, it is easy to show, as in Campbell and Shiller (1987), that the present value model implies two alternative interpretations of the spread, *viz.*

$$s_t = \theta \sum_{i=1}^{\infty} \delta^i E(\Delta y_{1t+i} | I_t) \ , \quad \text{and} \tag{33}$$

$$s_t = \frac{\delta}{1-\delta} E(\Delta y_{2t+1} | I_t) \ . \tag{34}$$

For our purposes, it is most convenient to work with the relationship (34). To put equation (34) in a more useful form, we multiply both sides by $-(1-\delta)/\delta$ to obtain

$$s_t^* = -E(\Delta y_{2t+1} | I_t) \ , \tag{35}$$

where the left-hand side of equation (35) has the equivalent forms

$$\begin{aligned} s_t^* &= y_{1t} - \frac{1-\delta}{\delta} y_{2t} \\ &= y_{1t} - (1/\theta) y_{2t} \\ &= y_{1t} + b y_{2t} \ \ \text{(say)}. \end{aligned} \tag{36}$$

If Δy_{2t} is stationary, then equation (35) implies that s_t^* is also stationary, from which we deduce (from equation (36)) the cointegration of y_{1t} and y_{2t} with cointegrating vector $(1, b)$.

The statistical model we use to describe the joint dynamics of y_{1t} and y_{2t} is a bivariate version of the general error-correction model (2), which we will rewrite here with the reduced rank restriction imposed:

$$\Delta y_t = \mu + \sum_{i=1}^{p} A_i^* \Delta y_{t-i} + \alpha \beta' y_{t-1} + \varepsilon_t \ . \tag{37}$$

This representation is in line with that of Hansen and Sargent (1981), Campbell and Shiller (1987), Toda (1991), and DeJong and Whiteman (1994), in that it includes the lagged values of not only y_{1t} but also y_{2t} in the information set that is available to the econometrician. Imposing the relationships (35) and (36) on this error-correction model yields the following set of restrictions:

$$a_{i,21}^* = a_{i,22}^* = 0 \ , \quad i = 1, \ldots, p \ , \tag{38}$$

$$\mu_2 = 0 \ , \tag{39}$$

$$\alpha_2 = -1 \ , \tag{40}$$

$$(1, b) = (1, -(1-\delta)/\delta) = (1, -1/\theta) \ , \tag{41}$$

where $(a_{i,21}^*, a_{i,22}^*)$ is the second row of the matrix A_i^* and α_2 is the second element of the vector $\alpha = (\alpha_1, \alpha_2)'$. The restrictions (38)–(40) are of the form (19) and can therefore be tested using the procedure outlined in the last section. Moreover, for a given value of the discount factor δ, equation (41) is also of the form (19). Following Campbell and Shiller (1987), we shall, in the next section of the paper, test the present value model both for the case where the discount factor δ takes on a value implied by the sample mean return on stocks and for the case where the cointegrating vector $(1, b)$ is left unrestricted.

4 Data Description and Empirical Results

In this section, we apply the test procedures discussed in Section 2 to the Campbell–Shiller data set for stock prices and dividends, updated to 1992. A brief discussion of the data is in order. As explained in Campbell and Shiller (1987), the term y_{2t} is the real stock price computed by dividing the Standard and Poor's stock price index for January by the January producer price index normalized so that the 1976 producer price index equals 100. Real dividend y_{1t} on the other hand, is constructed by dividing the nominal dividend series by the annual average producer price index also normalized so that the 1967 producer price index equals 100. It should be noted that the nominal dividend series before and after 1926 were collected from different sources. Since 1926, the nominal dividend series used in the construction is the dividends per share taken from the Standard and Poor's statistical service. Before 1926, the nominal dividend was taken from Cowles (1939).

Note also that a difficulty arises in pairing y_{1t} and y_{2t} since they are not measured contemporaneously. In our data set, y_{2t} is the beginning-of-period stock price while the dividend y_{1t} is paid sometime within period t. Since y_{1t} is not observable at the start of period t, West (1988) and others have argued that treating y_{1t} and y_{2t} as observations from the same period may lead to spurious rejection of the present value model. To circumvent this problem, we follow Campbell and Shiller (1987) and Toda (1991) in writing the VECM (37) in terms of $y_t = (y_{1t-1}, y_{2t})'$, where both variables are now in the information set at the start of time t. Note that cointegration of y_{1t} and y_{2t} implies that y_{1t-1} and y_{2t} are also cointegrated.

The remainder of this section is divided into three subsections, each discussing the results from a different test procedure. The results of unit root tests and tests of cointegration are given in Subsections 4.1 and 4.2, respectively, while Subsection 4.3 presents the results of testing the full set of restrictions implied by the present value model.

4.1 Unit Root Tests

As our setup depends critically on the assumption that both y_{1t} and y_{2t} are I(1) processes, we begin our empirical analysis by testing both the stock price series and the dividend series for unit roots. In their work, Campbell and Shiller (1987) ran Dickey–Fuller regressions on the two series and found that the unit root null hypothesis cannot be rejected at the 10% level for either series. Here, we take a different approach to unit root testing in an effort to bring additional statistical evidence in support of the hypothesis that both real stock prices and real dividends are well-described by I(1) processes. The method we use is closely related to the model selection criterion PICF detailed in Phillips (1992, 1995) and is, in fact, the univariate version of the multivariate test procedure we outlined in Section 2. To test unit root models against alternatives which may be trend stationary, we compare a general autoregressive model with trend (written in difference form), *viz.*,

$$H(M_{p,\ell}^{REF}) \ : \ \Delta y_t = a_0 y_{t-1} + \sum_{i=1}^{p-1} a_i \Delta y_{t-i} + \sum_{j=0}^{\ell} b_j t^j + \epsilon_t \tag{42}$$

with one which explicitly incorporates a unit root

$$H(M_{p,\ell}^{UR}) \ : \ \Delta y_t = \sum_{i=1}^{p-1} a_i \Delta y_{t-i} + \sum_{j=0}^{\ell} b_j t^j + \varepsilon_t \ . \tag{43}$$

Decisions about unit roots can then be made on the basis of the criterion:

$$\text{Decide in favor of unit root if } \widehat{\widetilde{L}}_{T_1}(M_{\widetilde{p},\widetilde{\ell}}^{UR})/\widehat{\widetilde{L}}_{T_1}(M_{\widehat{p},\widehat{r}}^{REF}) > 1 \ , \tag{44}$$

where

$$\begin{aligned} (\widehat{p},\widehat{\ell}) &= \text{argmax PIC}^{REF}(p,\ell) \ , \\ \text{PIC}^{REF}(p,\ell) &= \widehat{\widetilde{L}}_{T_1}(M_{p,\ell}^{REF})/\widehat{\widetilde{L}}_{T_1}(M_{\bar{p},\bar{\ell}}^{REF}) \ , \end{aligned} \tag{45}$$

and

$$\begin{aligned} (\widetilde{p},\widetilde{\ell}) &= \text{argmax PIC}^{UR}(p,\ell) \ , \\ \text{PIC}^{UR}(p,\ell) &= \widehat{\widetilde{L}}_{T_1}(M_{p,\ell}^{UR})/\widehat{\widetilde{L}}_{T_1}(M_{\bar{p},\bar{\ell}}^{REF}) \ , \end{aligned} \tag{46}$$

The formulae for the (approximate) marginalized likelihoods, $\widehat{\widetilde{L}}_{T_1}(M_{p,\ell}^{UR})$ and $\widehat{\widetilde{L}}_{T_1}(M_{p,\ell}^{REF})$, are analogous to their multivariate counterparts presented in Section 2 (see equations (11) and (25)). Hence, for brevity, we will not state them here.

Table 1: Unit Root Tests for the Sample Period 1871–1992

Variable	Initialization T_0	Lag selected[a] under $H(M^{UR})$	Trend selected[b] under $H(M^{UR})$	Posterior odds in favor of a unit root
y_{1t}	0 (uniform prior)	1	0[c]	42.662
	22	1	0	121.750
	26	1	0	778.628
	30	1	0	170.532
	34	1	0	8.880
	38	3	0	5.179
	42	1	0	1.540
y_{2t}	0 (uniform prior)	1	0	33.357
	22	1	0	357.095
	26	4	0	7691.930
	30	4	0	476.178
	34	1	0	23.647
	38	1	0	46.517
	42	1	0	27.456

[a]The maximum lag length $\bar{p}$ is set equal to 9.
[b]The maximum trend degree $\bar{\ell}$ is set equal to 1.
[c]0 denotes the inclusion of a constant term but not a linear trend.

Table 1 documents the results of unit root tests using the test criterion (44). As our empirical Bayesian approach inevitably involves a subjective choice of the sample split point T_0, the results in Table 1 and beyond will always be reported for several different values of T_0 so as to give an indication of the sensitivity of our results to the choice of the split point. Note also that there is a tradeoff in the choice of T_0: as T_0 increases, the hyperparameters of the prior distribution are estimated more precisely, but T_1 decreases so a smaller portion of the sample is being used for model comparison.

The results presented in Table 1 corroborate those obtained by Campbell and Shiller (1987) in that both dividends and stock prices are found to have a unit root specification, although the strength of the evidence in favor of a unit root (as measured by the posterior odds) varies with different values of T_0. Our criterion, however, does not favor a linear trend specification for either series.

4.2 Estimation of the Lag Order and the Cointegrating Rank

Sometimes, the question of whether dividends and stock prices are cointegrated is of independent interest. In particular, it can be quite independent of any interest in the validity of the present value model. For instance, we may only wish to obtain an appropriate time series representation for the variables y_{1t} and y_{2t} for forecasting purposes. Hence, in this section, we set out to estimate the lag order and the cointegrating rank of the model (37) using the test criterion (15) given in Section 2. Note that the criterion (15) selects the mode amongst possible PIC values. Alternatively, one could also construct point estimates of p and r by taking a weighted average using the PIC values as weights:

$$(p^m, r^m) = \text{round}\left\{\left(\sum_{p,r}\text{PIC}(p,r)\right)^{-1}\sum_{p,r}[\text{PIC}(p,r)\times(p,r)]\right\}. \qquad (47)$$

An advantage of using a mean criterion like (47) in addition to the modal criterion (15) is that it is affected by and therefore alerts the investigator 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) from both the modal criterion (15) and the mean criterion (46) for different values of T_0, the last observation used to construct the prior. From Table 2, we see that a cointegrating rank of zero was selected by both criteria regardless of initialization. These findings are roughly in accord with previous results obtained by Campbell and Shiller (1987), Phillips and Ouliaris (1988), and Toda (1991) for the shorter version of the same data set covering the period 1871–1986. Phillips and Ouliaris (1988) and Toda (1991) found no evidence of cointegration. Campbell and Shiller (1987), on the other hand, rejected the null hypothesis of no cointegration using a Dickey–Fuller regression but failed to reject the same null hypothesis when an Augmented Dickey–Fuller regression was used.

The sharpness of our inference on the lag order and the cointegrating rank is portrayed in Figures 1(a)–(g), where we depict bar charts of PIC values in (p, r) space for different choices of T_0. Note that in each of these figures, our selection of the RRR(1,0) specification is well-determined in the sense that it has far and away the

Table 2: Estimation of Cointegrating Rank and Order of Lagged Differences for the Sample, Period 1871–1992

Initialization T_0	$\widehat{p}$	$\widehat{r}$	p^m	r^m
0 (uniform prior)	1	0	1	0
22	1	0	1	0
26	1	0	2	0
30	1	0	1	0
34	1	0	1	0
38	1	0	1	0
42	1	0	1	0

Notes: $\widehat{p}$, $\widehat{r}$, p^m, and r^m are as defined in expressions (15) and (47).

highest PIC value amongst competing models. That our data strongly favors the RRR(1,0) specification is also reflected in the close agreement between the order estimates from the modal criterion and those from the mean criterion in Table 2, with the only exception being the selection of 2 lags by the mean criterion in the case $T_0 = 26$.

4.3 Tests of the Present Value Model

We proceed now to test the restrictions (38)-(41) of the present value model using the test criterion (28). To help summarize our results, we define the following statistics:

$$\tau_1 = \widehat{\widehat{L}}_{T_1}(M^{PV}_{\widetilde{p},1})/\widehat{\widehat{L}}_{T_1}(M^{U}_{\widetilde{p},\widetilde{r}})$$

and

$$\tau_2 = \widehat{\widehat{L}}_{T_1}(M^{PV}_{\widetilde{p},1})/\widehat{\widehat{L}}_{T_1}(M^{U}_{\widetilde{p},1}) \; .$$

Here, M^{PV} and M^U denote, respectively, the null model which satisfies the present value restrictions and the unrestricted VECM given by equation (2). The statistic τ_1 compares the restricted model of the chosen lag order $\widetilde{p}$ with the model having the highest density amongst those in the class of unrestricted VECM's while τ_2 compares the same null model with an unrestricted model of the same order $(\widetilde{p}, 1)$. We test the present value relations both for the case where $(1-\delta)/\delta = R = .085$ and the case where R is left unrestricted. The number .085 is the sample mean return on stocks for the period 1871–1992 and is used here as a possible discount rate.

Tables 3 and 4 summarize our results which do not seem to be sensitive in any substantive way to whether R is taken to be .085 or left unrestricted. Focusing on the τ_1 statistic, we see that our criterion favors a RRR(1,0) specification over the present value model uniformly over the different choices of T_0. Rejection of the present value model by our procedure is on the whole consistent with the results of Campbell and Shiller (1987) and Toda (1991), who also rejected the full set of present value

Table 3: Model Selection Test of the Present Value Restrictions (R = .085)

Initialization T_0	$\widehat{p}$ under $H(M^{PV})$	p^m under $H(M^{PV})$	τ_1	τ_2
0 (uniform prior)	1	1	0.198	1.194
22	1	2	0.018	0.113
26	1	1	0.017	0.115
30	1	1	0.008	4.552
34	1	1	0.035	3.131
38	1	1	0.057	1.562
42	1	1	0.030	7.066

Table 4: Model Selection Test of the Present Value Restriction (R unrestricted)

Initialization T_0	$\widehat{p}$ under $H(M^{PV})$	p^m under $H(M^{PV})$	τ_1	τ_2
0 (uniform prior)	1	1	0.362	1.823
22	1	1	0.307	1.942
26	1	1	0.085	0.569
30	1	1	0.090	51.434
34	1	1	0.049	4.353
38	1	1	0.055	1.511
42	1	2	0.051	12.070

restrictions using the classical Wald test. Only with the exclusion of the restriction corresponding to our equation (39) did Campbell and Shiller (1987) find favorable evidence for the present value model. In addition, our results are in agreement with those of DeJong and Whiteman (1994) in the case where a relatively tight Minnesota prior was used. We note, however, that those authors found more favorable evidence for the present value model when they allowed their priors to be more diffuse.

Looking at the τ_2 statistics, we see that for a majority of the cases under consideration, the present value model of the chosen lag order $\widetilde{p}$ compares favorably with a reduced rank regression of the same lag order and one cointegrating vector. This suggests that in most cases the rejection of the present value model by our criterion is primarily a rejection of the hypothesis that dividends and stock prices are cointegrated, which we showed in Section 3 to be an implication of the present value model when the data are well-described as integrated processes.

To assess the sensitivity of our results to lag specification, we turn our attention to Figures 2(a)–(g) which, for different choices of the initialization T_0, plot PIC values for four models (VECM with r = 0, VECM with r = 1, present value model with

$R = .085$, and present value model with R unrestricted) against different lag specifications. Note that for the cases where $T_0 \leq 34$, our results are only mildly sensitive to variations in the lag length. In these cases, our choice of an unrestricted model with no cointegration over the present value model can be overturned only if one decides to condition upon lag orders that are extremely small ($p = 0$) or moderately large ($p \geq 5$). On the other hand, in the two cases where the initial sample size is taken to be 38 and 42, the choice between the same two models depends more critically on the lag order selected. Interesting enough, these cases where our results are most sensitive to lag specification are also cases where our inference on the lag order is very sharp, as is evident from Figures 2(f) and 2(g). Hence, even in these cases, there is a clear choice in favor of an unrestricted model with no cointegration.

5 Conclusion

This paper argues for and illustrates a Bayesian approach to the joint estimation of the order of lagged differences and the cointegrating rank in a vector error-correction model (VECM). Our method is a variant of the Posterior Information Criterion (PIC), developed and analyzed in Phillips and Ploberger (1994, 1995), and is very similar to the asymptotic predictive odds version of the PIC criterion given in Phillips (1994). In the formulation of the PIC criterion here, we use a proper (Gaussian) prior whose hyperparameters are estimated from an initial subsample (of length T_0) of the data. As the form of the prior is suggested by the asymptotic posterior distribution of the parameters of our model, our criterion can be interpreted, in large samples, as an approximate predictive odds ratio. As in our earlier work (see Chao and Phillips, 1994), this procedure delivers consistent estimates of the lag order and cointegrating rank of a VECM.

Our procedure also has the advantage that it enables us to select the lag and cointegrating rank order of a VECM at the same time as it tests restrictions implied by economic theory, and it does so in the same coherent framework. Hence, the difficulties of accounting for the uncertainty associated with preliminary lag selection that arise with other methods of inference are avoided here. In addition, bar charts of PIC values for models of different dimensions can be readily constructed so that one may assess the sharpness of the inferences with respect to lag length and cointegrating rank.

We applied this method in an empirical analysis of the Campbell-Shiller stock market data, extended through to 1992. Using our procedure to compare models of different lag lengths and cointegrating rank, we consistently select a model with no cointegration for different choices of the initial value T_0. Further examination of the distribution of PIC values finds the lag and rank estimates to be sharply determined in every case.

Finally, we test the full set of present value restrictions using our criterion. We find models which satisfy these restrictions to be less plausible than time series models with no cointegration. These results are by and large consistent with the results of Campbell and Shiller (1987) and Toda (1991) using classical methodologies and with the Bayesian results of DeJong and Whiteman (1994) in the case where a relatively tight Minnesota prior was used.

6 Acknowledgements

We thank Don Andrews, Chris Sims, participants of the Conference on Forecasting, Prediction and Modeling in Statistics and Econometrics, particularly John Geweke and Arnold Zellner, and two anonymous referees for their many helpful comments and suggestions during the evolution of this paper to its present form. An expression of gratitude is also due to Robert Shiller for supplying the data used here. In addition, PCBP thanks the NSF for research support under Grant No. SES 9122142. Last but not least, we are grateful to Glena Ames for expertly keyboarding the manuscript. All the usual disclaimers apply.

References

Ahn, S. K. and G. C. Reinsel (1990). "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

Atkinson, A. C. (1978). "Posterior Probabilities for Choosing a Regression Model," *Biometrika*, 65, 39–48.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer–Verlag.

Campbell, J. Y. and R. J. Shiller (1987). "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

Chao, J. (1995). "Some Simulation Results of Likelihood Ratio Tests for Cointegrating Ranks in Misspecified Vector Autoregressions," in preparation.

Chao, J. C. and P. C. B. Phillips (1994). "Bayesian Model Selection in Partially Nonstationary Vector Autoregressive Processes with Reduced Rank Structure," mimeographed, Yale University.

Cowles, A. (1939). *Common Stock Indexes.* Bloomington, IN: Principia.

DeJong, D. N. and C. H. Whiteman (1992). "More Unsettling Evidence on the Perfect Markets Hypothesis: Trend-Stationarity Revisited," *Federal Reserve Bank of Atlanta Economic Review*, 77, 1–13.

DeJong, D. N. and C. H. Whiteman (1994). "Modeling Stock Prices without Knowing How to Induce Stationarity," *Econometric Theory*, 10, 701–719.

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

Geweke, J. (1994). "Bayesian Comparison of Econometric Models," Federal Reserve Bank of Minneapolis Research Department working paper 532, July.

Hansen, L. P. and T. Sargent (1981). "Exact Linear Rational Expectations Models: Specification and Estimation," Staff Report No. 71, Minneapolis: Federal Reserve Bank.

Johansen, S. (1988). "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamic and Control, 12*, 231–254.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Johansen, S. (1992). "Determination of Cointegrating Rank in the Presence of a Linear Trend," *Oxford Bulletin of Economics and Statistics*, 54, 383–397.

Kleidon, A. W. (1986). "Variance Bounds Test and Stock Price Valuation Models," *Journal of Political Economy*, 94, 953–1001.

Marsh, T. A. and R. C. Merton (1986). "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Prices," *American Economic Review*, 46, 483–498.

O'Hagan, A. (1991). "Discussion on Posterior Bayes Factor (by M. Aitkin)," *Journal of the Royal Statistical Society Series B*, 53, 136.

Phillips, P. C. B. (1992). "Bayes Methods for Trending Multiple Time Series with an Empirical Application to the U.S. Economy," Cowles Foundation Discussion Paper No. 1025, Yale University.

Phillips, P. C. B. (1994). "Model Determination and Macroeconomic Activity," mimeographed, Yale University.

Phillips, P. C. B. (1995). "Bayesian Model Selection and Prediction with Empirical Applications," *Journal of Econometrics* (in press).

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 (1995). "An Asymptotic Theory of Bayesian Inference for Time Series," *Econometrica* (forthcoming).

Phillips, P. C. B. and S. Ouliaris (1988). "Testing for Cointegration Using Principal Components Methods," *Journal of Economic Dynamics and Control*, 12, 205–230.

Pötscher, B. M. (1983), "Order Estimation in ARMA Models by Lagrange Multiplier Tests," *Annals of Statistics*, 11, 872–885.

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," mimeographed, Yale University.

Toda, H. Y. and P. C. B. Phillips (1994). "Vector Autoregression and Causality: A Theoretical Overview and Simulation Study," *Econometric Reviews*, 13, 259–285.

Tsay, R. S. (1984). "Order Selection in Nonstationary Autoregressive Models," *Annals of Statistics*, 12, 1425–1433.

Wei, C. Z. (1992). "On Predictive Least Squares Principles," *Annals of Statistics*, 20, 1–42.

West, K. D. (1988). "Dividend Innovations and Stock Market Volatility," *Econometrica*, 56, 37–62.

Zellner, A. (1971). *An Introduction to Bayesian Inference to Econometrics.* New York: John Wiley and Sons.

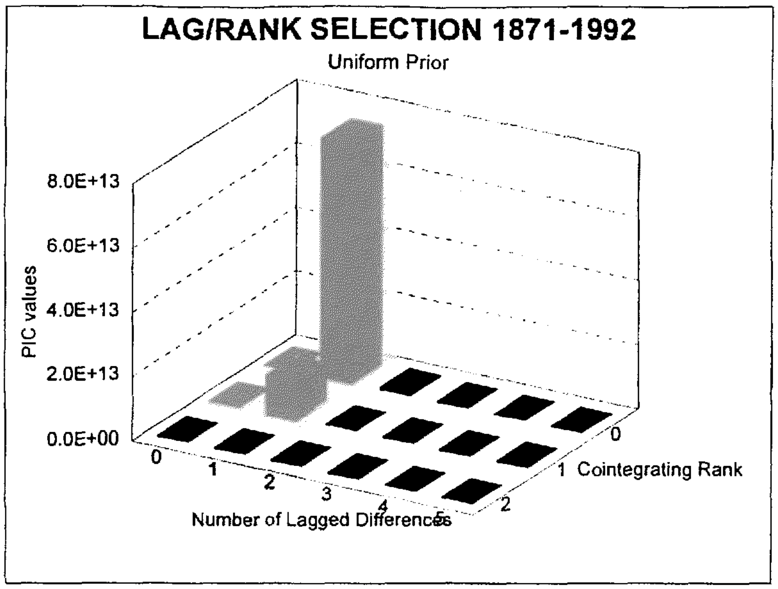

Figure 1(a)

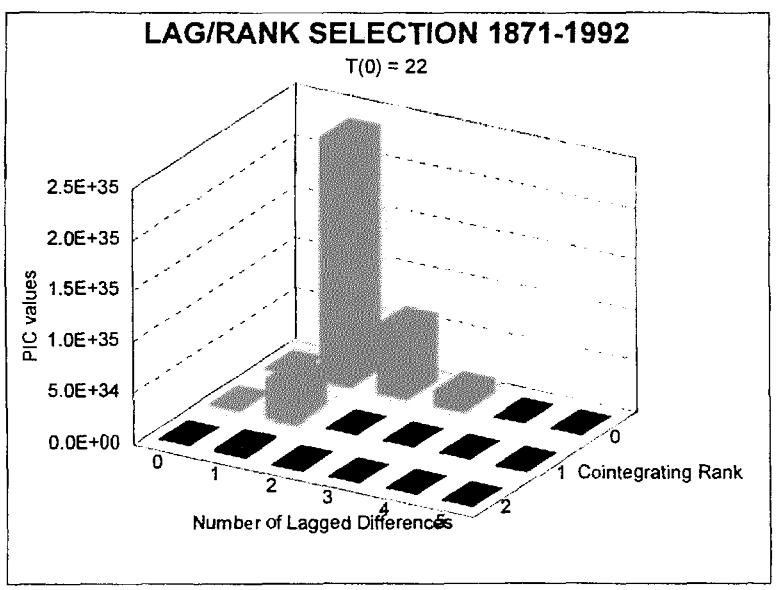

Figure 1(b)

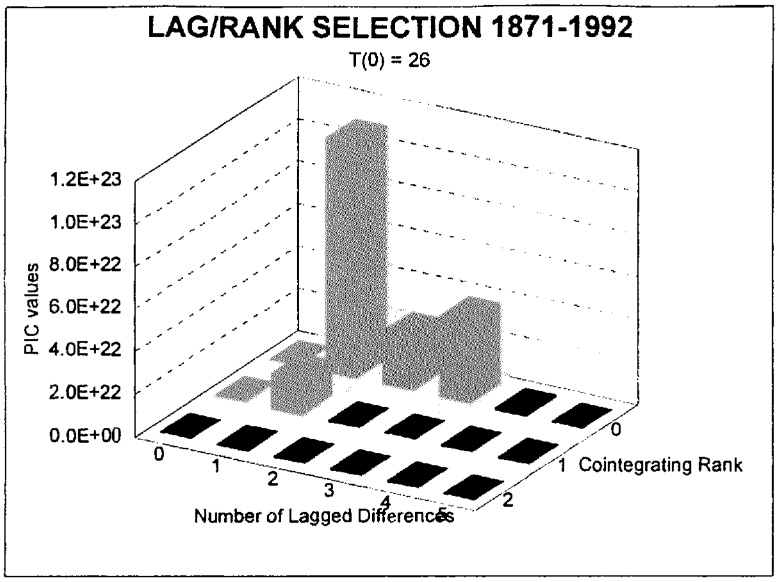

Figure 1(c)

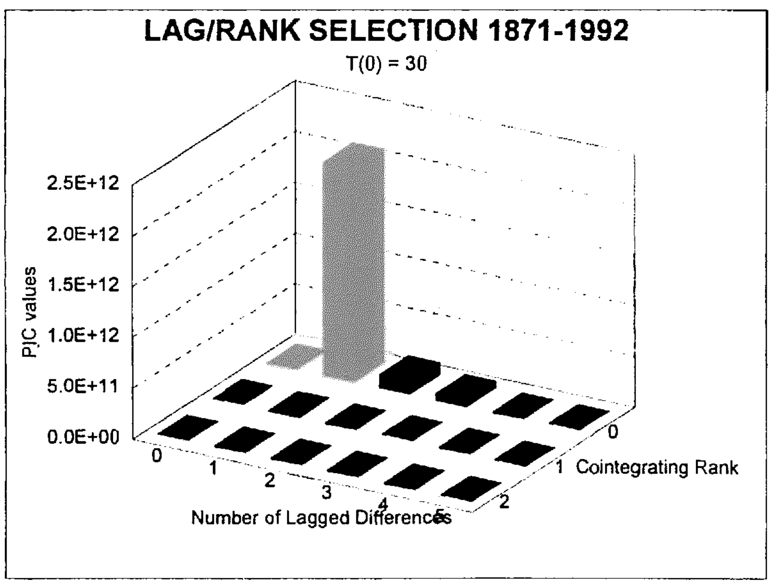

Figure 1(d)

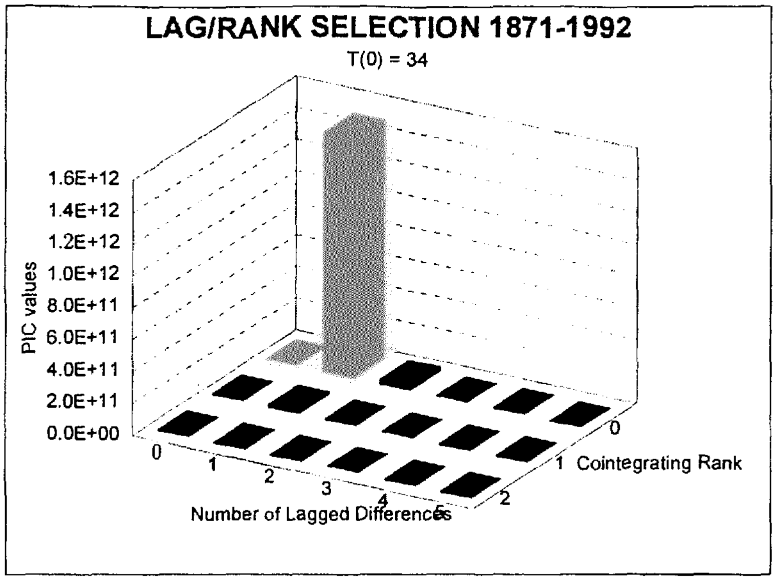

Figure 1(e)

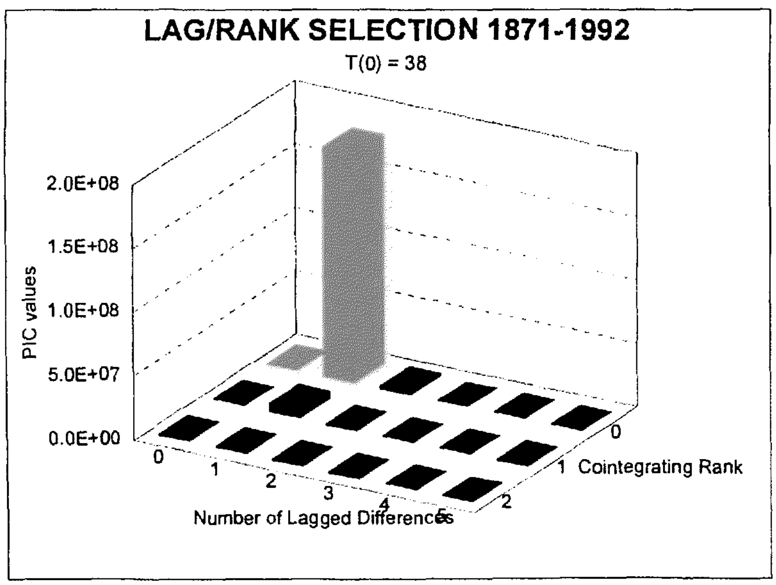

Figure 1(f)

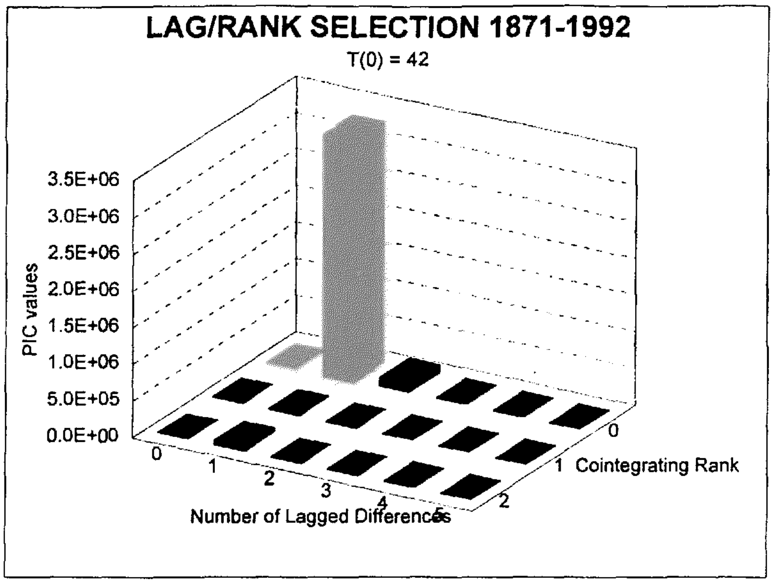
LAG/RANK SELECTION 1871-1992
T(0) = 42
PIC values
3.5E+06
3.0E+06
2.5E+06
2.0E+06
1.5E+06
1.0E+06
5.0E+05
0.0E+00
0
1
2
3
4
5
Number of Lagged Differences
0
1
2
Cointegrating Rank

Figure 1(g)

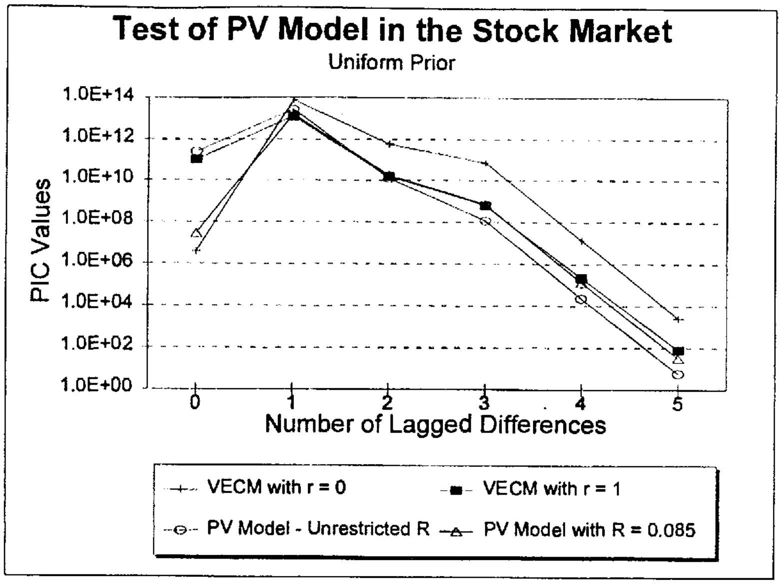

Figure 2(a)

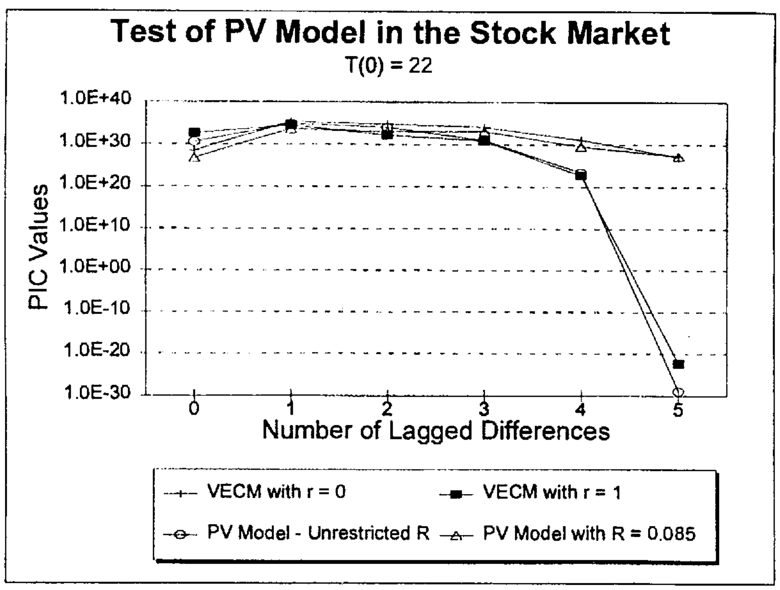

Figure 2(b)

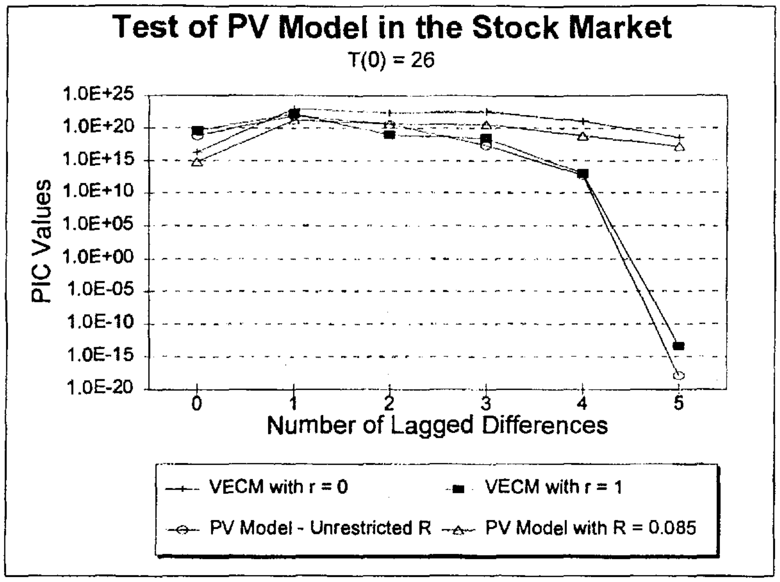

Figure 2(c)

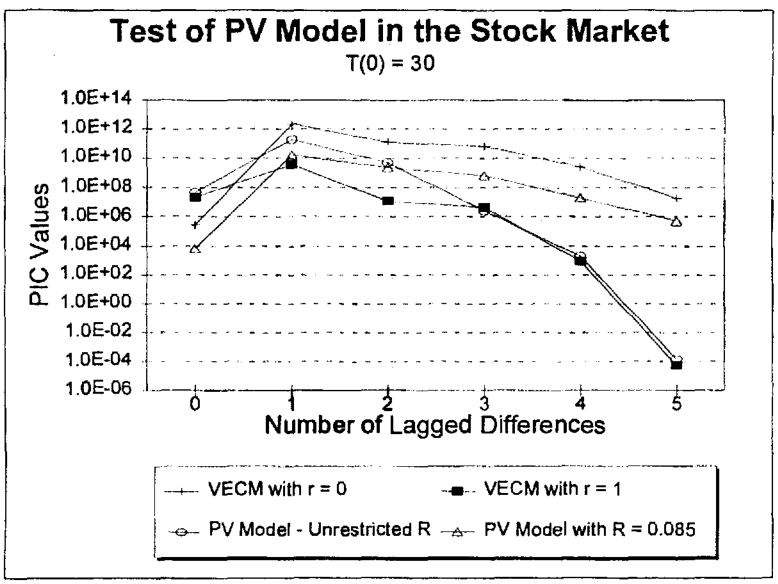

Figure 2(d)

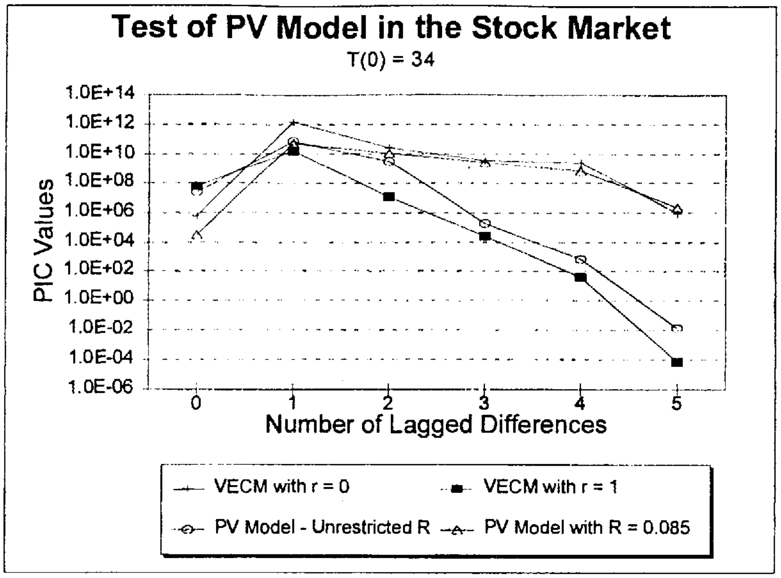

Figure 2(e)

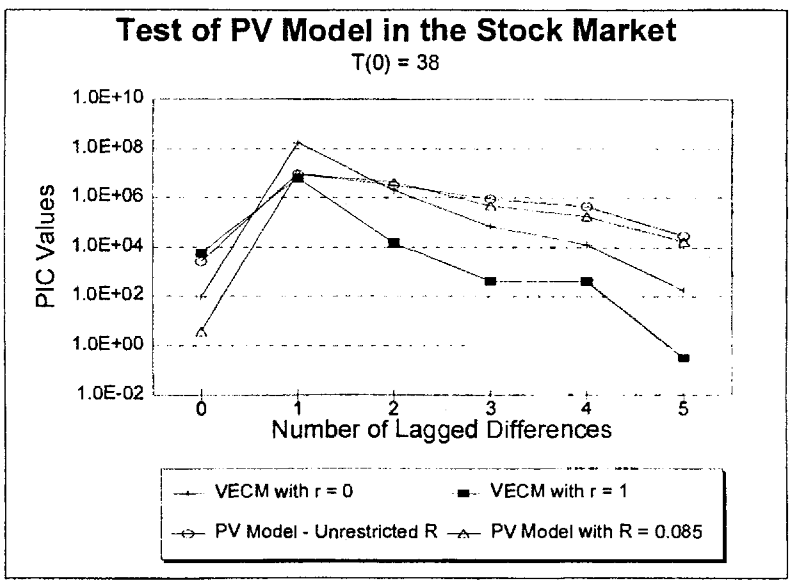

Figure 2(f)

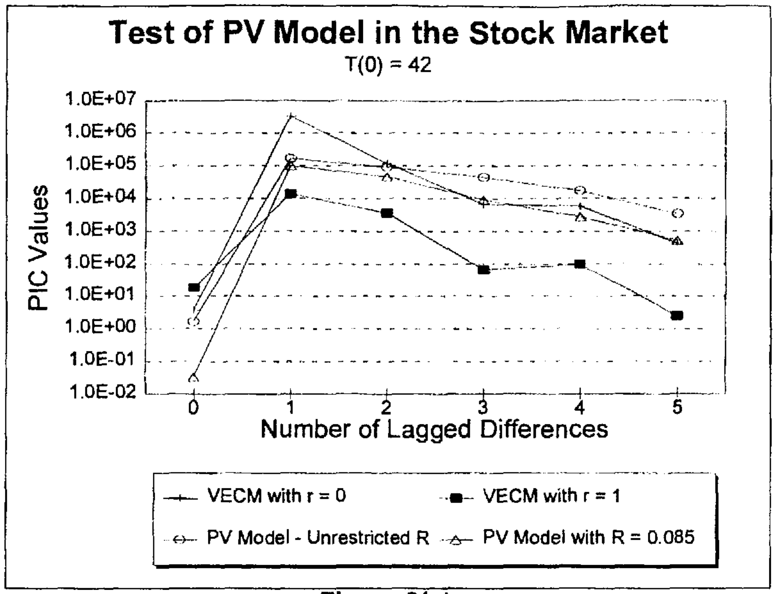
Test of PV Model in the Stock Market
T(0) = 42
1.0E+07
1.0E+06
1.0E+05
1.0E+04
1.0E+03
1.0E+02
1.0E+01
1.0E+00
1.0E-01
1.0E-02
PIC Values
0
1
2
3
4
5
Number of Lagged Differences
VECM with r = 0
VECM with r = 1
PV Model - Unrestricted R
PV Model with R = 0.085

Figure 2(g)