 Journal of Business & Economic Statistics, January 1995, Vol. 13, No. 1

Estimation of Common Long-Memory Components in Cointegrated Systems

Jesus GONZALO
Department of Economics, Boston University, Boston, MA 02215

Clive GRANGER
Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508

The study of cointegration in large systems requires a reduction of their dimensionality. To achieve this, we propose to obtain the I(1) common factors in every subsystem and then analyze cointegration among them. In this article, a new way of estimating common long-memory components of a cointegrated system is proposed. The identification of these I(1) common factors is achieved by imposing that they be linear combinations of the original variables X_t, and that the error-correction terms do not cause the common factors at low frequencies. Estimation is done from a fully specified error-correction model, which makes it possible to test hypotheses on the common factors using standard chi-squared tests. Several empirical examples illustrate the procedure.

KEY WORDS: Common factors; Cointegration; Error-correction model; Permanent–transitory decomposition.

If x_t and y_t are both integrated of order 1, denoted I(1), so that their changes are stationary, denoted I(0), they are said to be cointegrated if there exists a linear combination $z_t = y_t - Ax_t$, which is I(0). Several useful generalizations can be made of this definition, but this simple form is sufficient for the points proposed in this article. The basic ideas of cointegration were discussed by Granger (1986) and in the book of readings edited by Engle and Granger (1991). A simple constraint that results in cointegration involves an I(1) common factor f_t:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A \\ 1 \end{bmatrix} f_t + \begin{bmatrix} \tilde{y}_t \\ \tilde{x}_t \end{bmatrix}, \tag{1}$$

where $\tilde{y}_t$ and $\tilde{x}_t$ are both I(0). Clearly $z_t = \tilde{y}_t - A\tilde{x}_t$, being a linear combination of I(0) series, will never be I(1) and usually will be I(0). The reverse is also true—if (x_t, y_t) are cointegrated, there must exist a common factor representation of the form (1), as proved by Stock and Watson (1988).

A natural question that arises is how to estimate the common factor f_t, which might be an unobserved factor and is the driving force that results in cointegration. It has been suggested in the literature quoted previously that cointegration can be equated with certain types of equilibrium in that, in the long-run future, the pair of series is expected to lie on the attractor line $x_t = Ay_t$. Although much attention has been given to estimation of the cointegrating vector $(1, -A)$, relatively little attention has been given to estimation of f_t. Notice that when the long-run equilibrium is estimated, the common factor f_t is eliminated. There are several reasons why it is interesting to recover f_t—for example, situations in which the model of the complete set of variables appears very complex, although in fact, if we are interested in the long-run behavior, a simpler representation, using a small set of common long-memory factors could be adequate. This is the case for cointegration in large systems. Economists often conduct research on what might be considered to be natural subdivisions of the macroeconomy. The analysis of the long-run behavior of the whole macrosystem can be conducted by first finding the common factors in every subdivision of the economy and then studying cointegration among them. Another reason for singling out the f_t is that the estimation of this common factor allows one to decompose (y_t, x_t) into two components $(f_t, (\tilde{y}_t, \tilde{x}_t))$ that convey different kinds of information. For example, policymakers may be primarily interested in the trend (permanent component f_t) behavior, but those concerned with business cycles are more interested in the cyclical component (transitory component). Moreover, singling out the common factors allows us to investigate how they are related to other variables. The final goal of any factor model is to be able to identify the common factors with some observable variable. This article proposes a way of achieving this.

The situation studied here has analogies with the decomposition of an I(1) series into *permanent* and *transitory* components, where these components are considered to be I(1) and I(0), respectively. This question was considered by Quah (1992). Because the sum of an I(1) and I(0) series is I(1), it is easily seen that the question, as posed, does not completely identify the I(1) permanent components. To achieve identification, a further condition has to be imposed, such as maintaining that the permanent component is a random walk, or requiring the two components to be orthogonal at all leads and lags. In this article a different condition is used. This is possible because the situation being studied here involves more than one series, and this extra dimension allows a different type of condition to be considered. Basically the conditions imposed are that f_t be a linear combination of (y_t, x_t) and that the part that is left $(\tilde{y}_t, \tilde{x}_t)$, not have any permanent effect on (y_t, x_t). The first condition makes f_t observable; the second one makes f_t a good candidate to summarize the long-run behavior of the original variables. By these two conditions, we identify f_t up to a nonsingular matrix multiplication to the left. The linear combination is easily estimated from a fully specified error-correction model (ECM). This makes the suggested decomposition very convenient, mainly because the ECM takes care of the unit-root problem (see Johansen 1988; Phillips 1991), and therefore hypothesis testing on the linear combination f_t can be conducted using standard chi-squared tests. Another advantage is that any extension (nonlinearities, time-varying parameters, etc.) that could be incorporated in the ECM can be easily taken into account in this decomposition.

This article is organized as follows. Section 1 describes the factor model (1) for p variables and proposes a way to identify the common long-memory factors f_t. Section 2 shows how to estimate the linear combinations that form the common factors and how to test hypotheses on these linear combinations. Section 3 is an application of the method. Section 4 concludes. Proofs of the main results are in the Appendix.

1. FACTOR MODEL

Let X_t be a $(p \times 1)$ vector of I(1) time series with mean 0, for simplicity, and assume that the rank of cointegration is r [there exists a matrix $\alpha_{p\times r}$ of rank r, such that $\alpha' X_t$ is I(0)]. It follows that

1. The vector X_t has an ECM representation

$$\Delta X_t = \underset{p\times r}{\gamma} \underset{r\times p}{\alpha'} X_{t-1} + \sum_{i=1}^{\infty} \Gamma_i \Delta X_{t-i} + \epsilon_t, \tag{2}$$

where $\Delta = I - L$, with L the lag operator.

2. The elements of X_t can be explained in terms of a smaller number $(p-r)$ of I(1) variables, f_t, called (common) factors plus some I(0) components

$$\underset{p\times 1}{X_t} = \underset{p\times k}{A_1} \underset{k\times 1}{f_t} + \underset{p\times 1}{\tilde{X}_t}, \tag{3}$$

where $k = p - r$.

In the standard factor analysis, mostly oriented to cross-section data [for time series, see Peña and Box (1987)], the main objective is to estimate the loading matrix A_1 and the number k of common factors from (3). In our case, these two things are already known once the cointegrating vectors, α, have been estimated: $k = p - r$ and A_1 is any basis of the null space of $\alpha'(\alpha' A_1 = 0)$. The goal of this article is to estimate f_t. In factor analysis this is done from (3), after imposing constraints on f_t and $\tilde{X}_t$ that are not adequate in time series. Even dynamic factor analysis (see Geweke 1977) needs the assumption of stationarity that does not hold here. As will be shown in Section 2, the common factors can be estimated from the ECM (2) instead of from (3).

One of the conditions that will identify the common factors, f_t, is to impose that f_t be linear combinations of the variables X_t:

$$\underset{k\times 1}{f_t} = \underset{k\times p}{B_1} \underset{p\times 1}{X_t}. \tag{4}$$

This condition not only helps to identify f_t but also to associate the common factors with some observable variables, which is always advisable in factor analysis. The other condition that will identify f_t (up to a nonsingular matrix multiplication to the left) is to impose that $A_1 f_t$ and $\tilde{X}_t$ form the permanent and transitory components X_t, respectively, according to the following definition of a permanent–transitory (P–T) decomposition [part of this definition follows Quah (1992)].

Definition 1. Let X_t be a difference-stationary sequence. A P–T decomposition for X_t is a pair of stochastic processes P_t, T_t such that

1. P_t is difference stationary and T_t is covariance stationary,
2. var (ΔP_t) and var $(T_t) > 0$,
3. $X_t = P_t + T_t$,
4. we let

$$\underset{p\times p}{H^*(L)} \begin{bmatrix} \Delta P_t \\ T_t \end{bmatrix} = \begin{bmatrix} u_{Pt} \\ u_{Tt} \end{bmatrix} \tag{5}$$

be the autoregressive (AR) representation of $(\Delta P_t, T_t)$, with u_{Pt} and u_{Tt} uncorrelated, then

$$\text{(a)} \quad \lim_{h\to\infty} \frac{\partial E_t(X_{t+h})}{\partial u_{Pt}} \neq 0$$

and

$$\text{(b)} \quad \lim_{h\to\infty} \frac{\partial E_t(X_{t+h})}{\partial u_{Tt}} = 0,$$

where E_t is the conditional expectation with respect to the past history.

According to Condition 4, the only shocks that can affect the long-run forecast of X_t are those coming from the innovation term, u_{Pt}, of the permanent component, P_t. Condition (4) is not included in Quah's definition, and it is this that makes P_t and T_t permanent and transitory components, respectively. The next proposition clarifies this condition.

Proposition 1. Let

$$\begin{bmatrix} H_{11}(L) & H_{12}(L) \\ H_{21}(L) & H_{22}(L) \end{bmatrix} \begin{bmatrix} \Delta P_t \\ T_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{6}$$

be the AR representation of $(\Delta P_t, T_t)$. Condition (4) in Definition 1 is satisfied iff the total multiplier of ΔP_t with respect to T_t is 0; equivalently

$$H_{12}(1) = 0. \tag{7}$$

Apart from the instantaneous causality between the innovations (u_{1t}, u_{2t}) of both components that is likely to occur in economics because of temporal aggregation (see Granger 1980), Condition (4) says that T_t does not Granger-cause P_t in the long run or at frequency 0 [see Geweke (1982) and Granger and Lin (1992) for a formal definition of causality at different frequencies]. Let us consider the following example:

$$X_t = P_t + T_t, \tag{8}$$

where

$$\Delta P_t = a_1 T_{t-1} + a_2 \Delta T_{t-1} + u_{1t} \tag{9}$$

and

$$T_t = b_1 \Delta P_{t-1} + u_{2t}. \tag{10}$$

This is a P–T decomposition according to Definition 1 iff $a_1 = 0$. When $a_1 \neq 0$, even though T_t is I(0), this term cannot be called transitory because it will have a permanent effect on X_t (i.e., an effect on the long-run forecast of X_t). Notice that changes in the permanent component can affect the transitory component and also that changes in the transitory component could have an impact on the changes of the permanent component (a transitory impact on the levels of P_t and therefore on X_t).

There are decompositions that do not satisfy Condition (4). For instance, in the decomposition proposed by Aoki (1989), based on a dynamic factor (state-space) model, the I(0) component may have a permanent effect on the levels of the I(1) component and therefore on X_t. Another example is the decomposition of Kasa (1992):

$$X_t = \alpha_\perp(\alpha'_\perp \alpha_\perp)^{-1} f_t + \alpha(\alpha'\alpha)^{-1} z_t, \tag{11}$$

where $f_t = \alpha'_\perp X_t$ and $z_t = \alpha' X_t$. In general (see the proof of the next proposition) $\widetilde{X}_t = \alpha(\alpha'\alpha)^{-1} z_t$ will not be "transitory" according to Condition (4) in Definition 1.

The next proposition shows that the two conditions required for the common factors are enough to identify them up to a nonsingular transformation.

Proposition 2. In the factor model (2) the following conditions are sufficient to identify the common factors f_t:

1. f_t are linear combinations of X_t.
2. $A_1 f_t$ and $\widetilde{X}_t$ form a P–T decomposition.

Substituting (4) in (3), we obtain $\widetilde{X}_t = (I - A_1 B_1) X_t = A_2 \alpha' X_t = A_2 z_t$, where $z_t = \alpha' X_t$. Then, from the ECM (2), it is clear that the only linear combinations of X_t such that $\widetilde{X}_t$ has no long-run impact on X_t are

$$\underset{k\times p}{f_t = \gamma'_\perp}\ \underset{p\times 1}{X_t}, \tag{12}$$

where $\gamma'_\perp \gamma = 0$ and $k = p - r$. These are the linear combinations of ΔX_t that have the "common feature" (see Engle and Kozicki 1990) of not containing the levels of the error correction term z_{t-1} in them.

Once the common factors f_t are identified, inverting the matrix $(\gamma_\perp, \alpha)'$, we obtain the P–T decomposition of X_t proposed in this article:

$$\underset{p\times 1}{X_t} = \underset{p\times k}{A_1}\ \underset{k\times p}{\gamma'_\perp} X_t + \underset{p\times r}{A_2}\ \underset{r\times p}{\alpha'} X_t, \tag{13}$$

where $A_1 = \alpha_\perp(\gamma'_\perp \alpha_\perp)^{-1}$ and $A_2 = \gamma(\alpha'\gamma)^{-1}$.

In the next proposition, it is shown when this common factor decomposition (13) exists.

Proposition 3. If the matrix $\Pi = \gamma_{p\times r} \alpha'_{r\times p}$ has no more than $k = p - r$ eigenvalues equal to 0—that is, if $\det(\alpha'\gamma) \neq 0$—then $(\gamma_\perp, \alpha)'$ is nonsingular and the factor model (13) exists.

Even though f_t is not estimated from the factor model (3), the assumptions made to identify the common factors imply certain constraints on the P–T components that are the counterpart of assumptions imposed in standard factor analysis.

Proposition 4. The factor model

$$X_t = A_1 f_t + A_2 z_t, \tag{14}$$

where $f_t = \gamma_\perp X_t$ and $z_t = \alpha' X_t$, satisfies the following properties:

1. The common factors f_t are not cointegrated.
2. $\mathrm{Cov}(\Delta f^*_{it}, z^*_{j,t-s}) = 0$ ($i = 1, \ldots, k$; $j = 1, \ldots, p - k$; $s \geq 0$), where $\Delta f^*_{it} = \Delta f_{it} - E(\Delta f_{it} \mid \mathrm{lags}(\Delta X_{t-1}))$ and $z^*_{jt} = z_{jt} - E(z_{jt} \mid \mathrm{lags}(\Delta X_{t-1}))$.

The first property follows from Proposition 3 and the second from the ECM (2). This second property is another way of expressing that z_t does not cause f_t in the long run.

Properties (1) and (2) are equivalent to the assumptions made in standard factor analysis on the uncorrelatedness of the factors and the orthogonality between the factors and the error term ($A_2 z_t$).

As mentioned before, most of the P–T decompositions have been designed and used in a univariate framework. Stock and Watson (1988) proposed a common-trends decomposition that basically extends the univariate decomposition proposed by Beveridge and Nelson (1981) to cointegrated systems. The next proposition shows the connection between the common-trends decomposition of Stock and Watson and decomposition (14).

Proposition 5. The random-walk component (in the Beveridge–Nelson sense) of the I(1) common factor f_t in the decomposition (14) corresponds to the common trend of the Stock–Watson decomposition.

The advantage of our decomposition with respect to the common-trends model of Stock and Watson is that in our case it is easier to estimate the common long-memory components and to test hypotheses on them, as is shown in Section 2.

Notice that alternative definitions of f_t will vary only by I(0) components and therefore will be cointegrated.

In the univariate case, part of the literature has been oriented to obtaining orthogonal P–T decomposition (see Bell 1984; Quah 1992; Watson 1986). To the best of our knowledge, nothing has been written about the multivariate case. From the factor model (14) an orthogonal decomposition can be obtained such that the corresponding Δf_t and z_t are uncorrelated at all leads and lags. First, project z_t on Δf_{t-s} for all s and get the residuals

$$\widetilde{z}_t = z_t - P[z_t \mid \Delta f_{t-s} \forall s]. \tag{15}$$

Then define the new I(1) common factors $\widetilde{f}_t$ as

$$\widetilde{f}_t = (A'_1 A_1)^{-1} A'_1 (X_t - A_2 \widetilde{z}_t). \tag{16}$$

It is clear that $\Delta\widetilde{f}_t$ and $\widetilde{z}_t$ are uncorrelated at all leads and lags, but notice that, unless the $\widetilde{z}_t$ are linear combinations of current X_t, $\widetilde{f}_t$ will not be a linear combination of contemporaneous X_t. This is what is lost if orthogonality is required. To obtain an orthogonal P–T decomposition (according to Definition 1), one has to allow the common factors to be linear combinations of future, present, and past values of X_t.

2. ESTIMATION AND TESTING

In this section it is shown how to estimate and test hypotheses on $\gamma_\perp$. Most of the proofs in this section are based on Johansen and Juselius (1990).

Consider a finite ECM with Gaussian errors,

$$\mathcal{H}_1: \Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{q-1}\Delta X_{t-q+1} + \epsilon_t, \quad t = 1,\ldots,T, \tag{17}$$

where $\epsilon_1,\ldots,\epsilon_T$ are $\mathrm{IIN}_p(0,\Lambda)$, $X_{-q+1},\ldots,X_0$ are fixed, and

$$\underset{p\times p}{\Pi} = \underset{p\times r}{\gamma}\ \underset{r\times p}{\alpha'}. \tag{18}$$

Following Johansen (1988), we can concentrate the model with respect to Π, eliminating the other parameters. This is done by regressing ΔX_t and X_{t-1} on $(\Delta X_{t-1},\ldots,\Delta X_{t-q+1})$. This gives residuals R_{0t} and R_{1t} and residual product matrices

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt}, \qquad i,j = 0,1. \tag{19}$$

The remaining analysis will be performed using the concentrated model

$$R_{0t} = \gamma\alpha' R_{1t} + \epsilon_t. \tag{20}$$

The estimate of α is determined by reduced-rank regression in (20) (see Ahn and Reinsel 1990; Anderson 1951; Johansen 1988) and is found by solving the eigenvalues problem

$$\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0 \tag{21}$$

for eigenvalues $\widehat{\lambda}_1 > \cdots > \widehat{\lambda}_p$ and eigenvectors $\widehat{V} = (\widehat{v}_1,\ldots,\widehat{v}_p)$. The maximum likelihood estimators are given by $\widehat{\alpha} = (\widehat{v}_1,\ldots,\widehat{v}_r)$, $\widehat{\gamma} = S_{01}\widehat{\alpha}$, and $\widehat{\Lambda} = S_{00} - \widehat{\gamma}\widehat{\gamma}'$.

Finally the maximized likelihood function becomes

$$L_{\max}^{-2/T} = |\widehat{\Lambda}| = |S_{00}|\prod_{i=1}^{r}(1-\widehat{\lambda}_i) = |S_{00.1}|\left[\prod_{i=r+1}^{p}(1-\widehat{\lambda}_i)\right]^{-1}, \tag{22}$$

where $S_{00.1} = S_{00} - S_{01}S_{11}^{-1}S_{10}$.

The next theorem shows how to estimate $\gamma_\perp$.

Theorem 1. Under the hypothesis of cointegration H_2: $\Pi = \gamma\alpha'$, the maximum likelihood estimator of $\gamma_\perp$ is found by the following procedure: First solve the equation

$$\left|\lambda S_{00} - S_{01}S_{11}^{-1}S_{10}\right| = 0, \tag{23}$$

giving the eigenvalues $\widehat{\lambda}_1 > \cdots > \widehat{\lambda}_p$ and eigenvectors $\widehat{M} = (\widehat{m}_1,\ldots,\widehat{m}_p)$, normalized such that $\widehat{M}'S_{00}\widehat{M} = I$. The choice of $\widehat{\gamma}_\perp$ is now

$$\widehat{\gamma}_\perp = (\widehat{m}_{r+1},\ldots,\widehat{m}_p), \tag{24}$$

which gives the maximized likelihood function (22).

Notice, as Johansen (1989) pointed out, the duality between $\gamma_\perp$ and α. This is the idea of the proof of the preceding theorem. Both estimates come from the canonical correlation analysis between R_{0t} and R_{1t}. They are the canonical vectors and can be found by solving the following equations:

$$\begin{bmatrix} -\lambda_i S_{00} & S_{01} \\ S_{10} & -\lambda_i S_{11}\end{bmatrix}\begin{bmatrix}\widehat{m}_i \\ \widehat{v}_i\end{bmatrix} = 0, \qquad i = 1,\ldots,p, \tag{25}$$

with the normalizations $\widehat{M}'S_{00}\widehat{M} = I_p$ and $\widehat{V}'S_{11}\widehat{V} = I_p$.

From (25) and the preceding normalizations, it is clear that

$$\widehat{m}_j' S_{01}\widehat{v}_i = 0, \qquad i \neq j. \tag{26}$$

Because $\widehat{\alpha} = (\widehat{v}_1,\ldots,\widehat{v}_r)$ and $\widehat{\gamma} = S_{01}\widehat{\alpha}$, then $\widehat{\gamma}_\perp = (\widehat{m}_{r+1},\ldots,\widehat{m}_p)$.

If for any reason α is not estimated by maximum likelihood or simultaneous reduced-rank least squares [see Gonzalo (1994) for different methods of estimation], the way to estimate $\gamma_\perp$ is the following: Insert the estimate of α, $\widetilde{\alpha}$, into the ECM (17), use this to estimate $\widetilde{\gamma}$, and then solve

$$\left|\lambda S_{00} - \underset{p\times r}{\widetilde{\gamma}}\ \underset{r\times p}{\widetilde{\gamma}'}\right| = 0, \tag{27}$$

giving the eigenvalues $\widetilde{\lambda}_1 > \cdots > \widetilde{\lambda}_p$ ($\widetilde{\lambda}_{r+j} = 0, j = 1,\ldots,p-r$) and eigenvectors $\widetilde{M} = (\widetilde{m}_1,\ldots,\widetilde{m}_p)$ normalized such that $\widetilde{M}'S_{00}\widetilde{M} = I$. The choice of $\widetilde{\gamma}_\perp$ is now $\widetilde{\gamma}_\perp = (\widetilde{m}_{r+1},\ldots,\widetilde{m}_p)$; the eigenvectors corresponding to the eigenvalues equal 0.

To find the asymptotic distribution of $\widehat{\gamma}_\perp$, it is convenient to decompose $\gamma'_\perp$ as follows: $\widehat{\gamma}_\perp = \gamma_\perp\widehat{d} + \gamma\widehat{a}$, where $\widehat{d} = (\gamma'_\perp\gamma_\perp)^{-1}\gamma'_\perp\widehat{\gamma}_\perp$, and $\widehat{a} = (\gamma'\gamma)^{-1}\gamma'\widehat{\gamma}_\perp$.

Theorem 2. When $T \to \infty$,

$$T^{1/2}(\widehat{\gamma}_\perp\widehat{d}^{-1} - \gamma_\perp) \Rightarrow N(0,V), \tag{28}$$

where $\Rightarrow$ means convergence in distribution, $V = \gamma(\gamma'(\Sigma_{00} - \Lambda)\gamma)^{-1}\gamma' \otimes \gamma'_\perp\Lambda\gamma_\perp$, and Σ_{00} var $(\Delta X_t \mid \Delta X_{t-1},\ldots,\Delta X_{t-q+1})$.

As mentioned earlier, one of the advantages of our decomposition is that one can test whether or not certain linear combinations of X_t can be common factor. Johansen (1991) showed how to test the hypotheses on α and γ:

$$\mathcal{H}_3: \underset{p\times r}{\alpha} = \underset{p\times s}{J}\ \underset{s\times r}{\varrho}, \qquad r \le s \le p$$

and

$$\mathcal{H}_{4a}: \underset{p\times r}{\gamma} = \underset{p\times n}{Q}\ \underset{n\times r}{\psi}, \qquad r \le n \le p.$$

In the next theorem it is shown how to test the hypotheses on $\gamma_\perp$:

$$\mathcal{H}_{4b}: \underset{p\times k}{\gamma_\perp} = \underset{p\times m}{G}\ \underset{m\times k}{\theta} \quad \text{with} \quad k = p - r \text{ and } k \le m \le p.$$

Theorem 3. Under the hypotheses $\mathcal{H}_{4b}$: $\gamma_\perp = G\theta$, one can find the maximum likelihood estimator of $\gamma_\perp$ as follows:

First solve

$$\left|\lambda G'S_{00}G - G'S_{01}S_{11}^{-1}S_{10}G\right| = 0 \tag{29}$$

Table 1. Consumption and GNP Regressions (Cochrane 1991)

Left variable		Right variable: const.	$c_{t-1}-y_{t-1}$	Δc_{t-1}	Δc_{t-2}	Δy_{t-1}	Δy_{t-2}	R^2
			1. Vector autoregression					
Δc_t	coeff.	−.43	−.02	.07	−.02	.09	−.02	.06
	t stat.	−.49	−1.23	.90	−.19	1.91	−.40	
Δy_t	coeff.	5.19	.08	.52	.16	.22	.14	.27
	t stat.	3.49	3.45	3.81	1.12	2.74	1.89	

2. P–T decomposition

$$\gamma'_\perp = (1,0); \alpha' = (1,-1).$$

$$\begin{bmatrix} c_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} f_t + \begin{bmatrix} 0 \\ -1 \end{bmatrix} z_t,$$

where $f_t = \gamma'_\perp(c_t, y_t)' = c_t$ and $z_t = \alpha'(c_t, y_t)' = c_t - y_t$.

NOTE: y_t denotes real GNP and c_t denotes log (nondurable + services consumption). Δ denotes first differences, $\Delta y_t = y_t - y_{t-1}$. Data sample: 1947:1–1989:3.

for $\widehat{\lambda}_{4b.1} > \cdots > \widehat{\lambda}_{4b.m}$ and $\widehat{M}_{4b} = (\widehat{m}_{4b.1}, \ldots, \widehat{m}_{4b.m})$ normzlized by $\widehat{M}'_{4b}\,(G'S_{00}G)\,\widehat{M}_{4b} = I$. Choose

$$\widehat{\theta}_{mx(p-r)} = (\widehat{m}_{4b.(m+1)-(p-r)}, \ldots, \widehat{m}_{4b.m}) \quad \text{and} \quad \widehat{\gamma}_\perp = G\widehat{\theta}. \tag{30}$$

The maximized likelihood function becomes

$$L_{\max}^{-2/T}(\mathcal{H}_{4b}) = |S_{00.1}|\left(\prod_{i=r+1}^{p}(1-\widehat{\lambda}_{4b.i+(m-p)})\right)^{-1}, \tag{31}$$

which gives the likelihood ratio test of the hypothesis H_{4b} in H_2 as

$$-2\ln(\cdot\,;\mathcal{H}_{4b} \text{ in } \mathcal{H}_2) = -T\sum_{r+1}^{p}\ln\{(1-\widehat{\lambda}_{4b.i+(m-p)})/(1-\widehat{\lambda}_i)\} \sim \chi^2_{(p-r)\times(p-m)}. \tag{32}$$

Finally one may be interested in estimating α and $\gamma_\perp$ under $\mathcal{H}_3$ and $\mathcal{H}_{4b}$. The way to proceed is to convert $\mathcal{H}_{4b}$ into $\mathcal{H}_{4a}$. Notice that Q (the matrix in $\mathcal{H}_{4a}$) is formed by the $p-m$ eigenvectors of GG' corresponding to the eigenvalues equal to 0. Following theorem 3.1 of Johansen (1991), α and γ can be estimated under $\mathcal{H}_3$ and $\mathcal{H}_{4a}$. Once γ is estimated, we are in the situation described in (27).

3. APPLICATIONS

In the first two examples (consumption and gross national product (GNP), dividends and stock prices), it is shown how to obtain the common factors directly from an ECM. The third application (interest rates in Canada and the United States) shows, step by step, how to estimate the common factors and how to decompose these variables into permanent and transitory components.

3.1 Consumption and GNP, Dividends and Stock Prices

The (vector autoregression) VAR (ECM) models of Tables 1 and 2 are reproduced from Cochrane (1991). Focusing our attention in the consumption–GNP example, it can be seen from the VAR of Table 1 that the error-correction term $(c_{t-1} - y_{t-1})$ does not appear to be significant in the consumption equation; therefore, $\gamma' = (0,1)$ and $\gamma'_\perp = (1,0)$. In other words, the I(1) common factor (permanent component) in our decomposition is

$$f_t = (1,0)\begin{bmatrix} c_t \\ y_t \end{bmatrix},$$

Table 2. Dividend and Price Regressions (Cochrane 1991)

Left variable		Right variable: const.	$d_{t-1}-p_{t-1}$	Δd_{t-1}	Δd_{t-2}	Δp_{t-1}	Δp_{t-2}	R^2
			1. Vector autoregression					
Δd_t	coeff.	20.01	.038	.046	−.06	−.08	−.04	.038
	t stat.	.78	.47	.25	.34	−.65	.32	
Δp_t	coeff.	78.65	.225	.06	−.08	.114	−.09	.14
	t stat.	2.34	2.11	.25	−.36	.68	−.55	

2. P–T decomposition

$$\gamma_\perp = (1,0); \alpha = (1,-1).$$

$$\begin{bmatrix} d_t \\ p_t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} f_t + \begin{bmatrix} 0 \\ -1 \end{bmatrix} z_t$$

where $f_t = d_t$ and $z_t = d_t - p_t$.

NOTE: d_t denotes log dividends and p_t denotes log price (cumulated returns) on the value-weighted New York Stock Exchange portfolio. Δ denotes first differences; Δp_t is the log return. Data sample: 1927–1988.

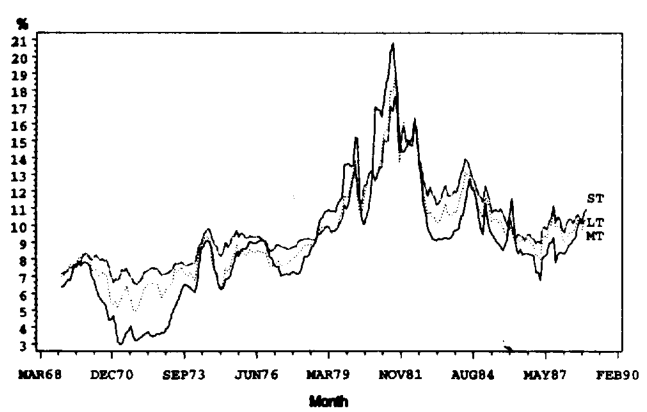

Figure 1. Canada Interest Rates (1969:1–1988:12): ———, Short-Term (ST); ....., Medium-Term (MT); — · — · , Long-Term (LT).

a multiple of the consumption variable. This means that, if consumption is kept fixed, any change in the income is going to affect (c_t, y_t) only through z_t (the transitory component) and therefore will only have transitory effects (see the factor model in Table 1). This is exactly the conclusion reached by Cochrane (1991) through the impulse-response functions:

> GNP's response to a consumption shock is partly permanent but also partly temporary. More importantly, GNP's response to a GNP shock holding consumption constant is almost entirely transitory. This finding has a natural interpretation: If consumption does not change, permanent income must not have changed, so any change in GNP must be entirely transitory. (p. 2)

The same kind of conclusion is obtained in the second example with dividends and stock prices in Table 2. From the factor model it can be seen that a shock in dividends has a permanent (long-run) effect in prices and dividends, but a shock in prices, with no movements in dividends, is completely transitory.

3.2 Interest Rates in Canada and the United States

The main purpose of this application is to find the permanent component that is driving the interest rates of Canada and the United States in the long run. To do that, three interest rates with different maturities have been considered in each country—short-term, medium-term, and long-term interest rates. In Canada (see Fig. 1), the short-term rate is the weighted average of the yields on successful bids for three-month treasury bills (x_{1c}), the medium-term rate refers to government bonds with original maturity of 3 to 5 years (x_{2c}), and the long-term rate refers to bonds with original maturity of 10 years and over (x_{3c}). In the United States (see Fig. 2), the short-term rate is an annual average of the discount rate on new issues of three-month treasury bills (x_{1u}), the medium-term rate refers to 3-year constant maturity government bonds (x_{2u}), and the long-term rate refers to 10-year constant-maturity bonds (x_{3u}). The data consist of 240 monthly observations from 1969:1 to 1988:12 and were obtained from the IMF data base.

To show the potential of our decomposition as a dimension-reduction method, two different approaches have been followed to obtain the common permanent component of the whole set of interest rates. In the first approach, the interest rates are considered within countries, and in each country the I(1) common factor is estimated. The common permanent component between these two I(1) country factors will be the factor that is driving the whole system of interest rates in the long run. In this process the number of variables involved at every step is at most 3. This is what makes this first approach very convenient for analyzing cointegration in big systems.

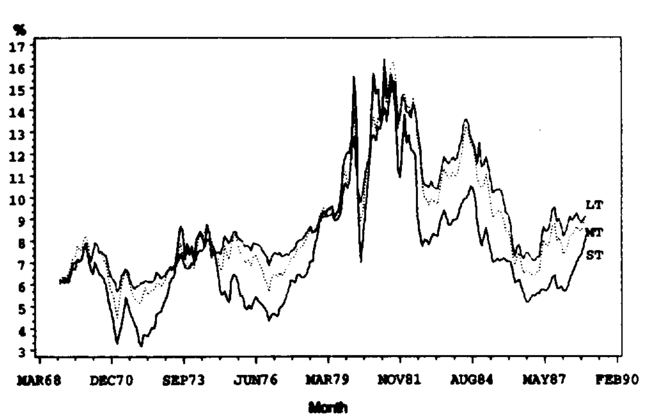

Figure 2. U.S. Interest Rates (1969:1–1988:12): ———, Short-Term (ST); ....., Medium-Term (MT); — · — · , Long-Term (LT).

Table 3. Augmented Dickey–Fuller Statistics for Tests of a Unit Root

	ADF(0)	ADF(1)	ADF(2)	ADF(3)	ADF(4)
$x1c_t$	−1.46	−2.18	−2.10	−1.97	−1.93
$x2c_t$	−1.67	−1.93	−1.80	−1.86	−1.63
$x3c_t$	−1.64	−1.75	−1.67	−1.73	−1.6
$x1u_t$	−2.05	−2.7	−2.17	−2.15	−2.07
$x2u_t$	−1.72	−2.35	−1.78	−1.83	−1.79
$x3u_t$	−1.55	−1.88	−1.60	−1.66	−1.62

NOTE: $ADF(q)$ is the t statistic of $\hat{\delta}$ in the regression $\Delta x_t = c + \delta x_{t-1} + \sum_{i=1}^{q} \phi_i \Delta x_{t-i} + e_t$. The critical values (from Mackinnon 1991) for $n = 240$ are 1% (−3.46), 5% (−2.87), and 10% (−2.57). xij_t denotes the i term interest rate in country j at time t, for $i = 1$ (short), $i = 2$ (medium), $i = 3$ (long), $j = c$ (Canada), and $j = u$ (U.S.). Data are from the IMF. Sample period: 1969:1–1988:12.

Table 4. Testing for Cointegration

H_2	Trace	Trace(.90)	λmax	λmax(.90)
		Canada		
$r \leq 2$	3.52	6.50	3.52	6.50
$r \leq 1$	25.22	15.66	21.70	12.91
$r = 0$	56.63	28.71	31.40	18.90
		United States		
$r \leq 2$	3.98	6.50	3.95	6.50
$r \leq 1$	29.18	15.66	25.23	12.91
$r = 0$	61.98	28.71	32.79	18.90
		Canada and United States		
$r \leq 5$	3.79	6.50	3.79	6.50
$r \leq 4$	16.49	15.66	12.70	12.91
$r \leq 3$	36.59	28.71	20.10	18.90
$r \leq 2$	68.89	45.23	32.29	24.78
$r \leq 1$	104.11	66.49	35.23	30.84
$r = 0$	153.87	90.39	49.75	36.35

NOTE: The critical values have been obtained from Osterwald-Lenum (1992). Test statistics for the hypothesis H_2 are for several values of r versus $r + 1$ (λ max) and versus general alternative H_1 (trace) for Canadian and U.S. interest rates data (1969:1–1988:12).

Table 5. Estimation of the Cointegration Structure

Canada				United States				Canada and United States						
							Eigenvalues $\widehat{\lambda}$							
(.123, .086, .014)				(.128, .10, .016)				(.187, .136, .126, .080, .051, .016)						
							Eigenvectors $\widehat{V}^a$							
x1c	−.066	.009	.008	x1u	.091	−.045	.016	x1c	.008	−.007	.068	−.019	−.041	.001
x2c	.109	−.149	−.031	x2u	−.275	−.004	−.062	x2c	.037	.046	−.100	.181	−.022	.015
x3c	−.030	.148	.051	x3u	.191	.046	.076	x3c	−.081	−.074	−.075	−.239	.007	−.008
								x1u	−.083	.073	−.053	−.011	.023	.012
								x2u	.075	−.275	.101	−.014	.019	−.063
								x3u	.032	.242	.039	.089	.030	.070
							Eigenvectors $\widehat{M}^b$							
x1c	−.016	.018	.079	x1u	−.100	−.079	.107	x1c	.034	.015	.035	−.040	−.143	−.006
x2c	.058	−.380	.004	x2u	.324	−.123	−.189	x2c	−.160	.076	−.202	.296	−.099	.034
x3c	.095	−.466	.059	x3u	−.179	.273	.243	x3c	.225	−.023	.031	−.417	.189	−.003
								x1u	−.110	−.053	−.045	−.047	.029	.112
								x2u	.033	.270	.221	−.026	.128	−.220
								x3u	.095	−.276	.004	.190	−.103	.260

NOTE: The Eigenvalues $\hat{\lambda}$ and Eigenvectors $\hat{V}$, $\hat{M}$ based on the normalizations $\hat{V}S_{11}\hat{V} = I$ and $\hat{M}S_{00}M = I$ for Canada and U.S. interest rate data (1969: 1-1988:12).
[a]The first r columns form $\hat{\alpha}$.
[b]The last $p-r$ columns form $\hat{\gamma}_{\perp}$.

The second approach consists of analyzing the cointegration of the whole system (6 variables) without any a priori partition. This second way becomes unfeasible when the number of variables is large (greater than 10). The conclusion obtained by these two different approaches matches perfectly. There is only one common long-memory factor in the whole system formed by the six interest rates, and that factor is the U.S. common permanent component.

To reach the preceding conclusion these steps have been followed:

1. Unit-root tests (Table 3): Using the augmented Dickey–Fuller test, the null of the unit root is not rejected for any of the six interest rates.

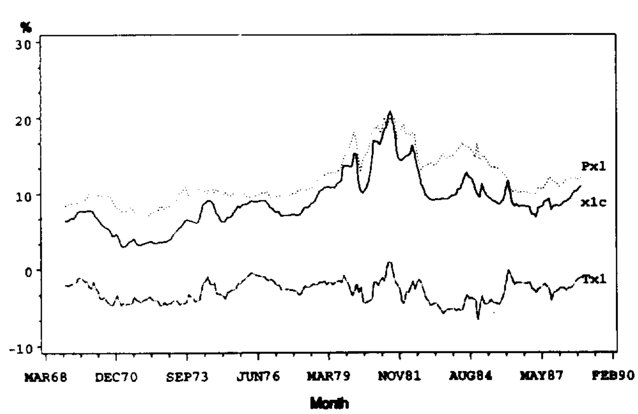

Figure 3. Canada: P–T Decomposition of Short-Term Interest Rates (x_{1c}); $f_1 = -.006x_{1c} + .034x_{2c} - .003x_{3c} + .112x_{1u} - .22x_{2u} + .26x_{3u}$; $Z_1 = .008x_{1c} + .037x_{2c} - .081x_{3c} - .083x_{1u} + .075x_{2u} + .032x_{3u}$; $Z_2 = -.007x_{1c} + .046x_{2c} - .074x_{3c} + .073x_{1u} - .275x_{2u} + .242x_{3u}$; $Z_3 = .068x_{1c} - .100x_{2c} - .075x_{3c} - .053x_{1u} + .101x_{2u} + .039x_{3u}$; $Z_4 = -.019x_{1c} + .181x_{2c} - .239x_{3c} - .011x_{1u} - .014x_{2u} + .089x_{3u}$; $Z_5 = -.041x_{1c} - .022x_{2c} + .007x_{3c} + .023x_{1u} + .019x_{2u} + .030x_{3u}$; $Px_{1c} = 7.86f_1$; $Tx_{1c} = -3.58Z_1 - 6.92Z_2 + 4.52Z_3 + 1.31Z_4 - 18.16Z_5$: ———, x_{1c}; ......, Px_{1c}; −·−·−, Tx_{1c}.

2. Cointegrations tests (Table 4): Using the Johansen likelihood ratio (LR) test, for a VAR of order 3 (order suggested by the Akaike information criterion), it is found that Canada, as well as the United States, has two cointegrating vectors, and the whole system has five cointegrating vectors. Therefore there is one common I(1) factor in each country, and they are cointegrated, implying that there is only one common permanent component in the whole system.

3. Estimation of the cointegration structure: In Table 5 we provide the estimates of the cointegrating vectors and of the linear combinations that define our common permanent components. From these estimates, following Section 1, all interest rates can be decomposed into permanent and transitory components. Some examples are shown in Figures 3 and 4.

4. Testing hypotheses on the long-memory common factors: From Table 5, the I(1) common factor of the whole system is $f_1 = -.006x_{1c} + .034x_{2c} - .003x_{3c} + .112x_{1u} - .22x_{2u} + .26x_{3u}$. Following Theorem 3, we tested that the U.S. interest

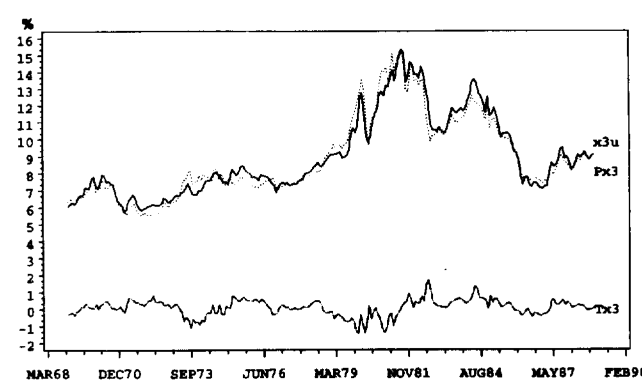

Figure 4. United States P–T Decomposition of Long-Term Interest Rates (x_{1u}); See Definitions of Variables in Figure 3. $Px_{1u} = 5.91f_1$; $Tx_{1u} = 6.01Z_1 - 2,81Z_2 + .252Z_3 - 1.27Z_4 + 1.79Z_5$. ———, x_{3u}; ......, Px_{3u}; −·−·−, Tx_{3u}.

rates are the only variables driving the whole system in the long run; that is,

$$\mathcal{H}_{4b}: \gamma_\perp = G\theta \quad \text{with} \quad G = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Under $\mathcal{H}_{4b}, \widehat{\theta} = (.2, -.25, .27)$. This hypothesis is not rejected with a p value of .86. The same conclusion was obtained when the analysis was done by countries. The common long-memory factor in Canada is $f_{1c} = .08x_{1c} + .004x_{2c} + .06x_{3c}$, and in the United States $f_{1u} = .11x_{1u} - .19x_{2u} + .24x_{3u}$. These two common factors are cointegrated, and the hypothesis tests that in the long-run the driving force of these two common factors is f_{1u} has a p value of .45.

Results in more detail can be found in the paper by Gonzalo and Granger (1992).

4. CONCLUSION

The results of this article have implications on three fronts. In the first place, they provide a new form of estimating the I(1) common factors that ensure that a set of variables are cointegrated, thus allowing us to gain more understanding of the nature of economic time series. Second, they show a new method for estimating the permanent component ("trend") of a time series using multivariate information, and third, they provide a new way of studying cointegration in large systems by using the common long-memory factors of every "natural" subsystem.

Further research needs to be done on the small-sample properties of $\widehat{\gamma}_\perp$ and on how to incorporate different characteristics of the ECM (nonlinearities, time-varying parameters, etc.) in the estimation of the common factors and therefore in the estimation of P–T decompositions.

ACKNOWLEDGMENTS

This research was partially supported by the Sloan Foundation and U.S. National Science Foundation Grant SES-9023037. We thank Chor-Yiu Sin and two anonymous referees for helpful comments.

APPENDIX: PROOFS OF THE MAIN RESULTS

Proof of Proposition 1. Inverting (6),

$$\begin{bmatrix} \Delta P_t \\ T_t \end{bmatrix} = \begin{bmatrix} H^{11}(L) & H^{12}(L) \\ H^{21}(L) & H^{22}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \qquad (A.1)$$

we obtain the moving average representation of ΔP_t

$$\Delta P_t = H^{11}(1)u_{1t} + H^{12}(1)u_{2t} + (I - L)\{\widetilde{H}^{11}(L)u_{1t} + \widetilde{H}^{12}(L)u_{2t}\}, \qquad (A.2)$$

where

$$H^{1j}(L) = H^{1j}(1) + (I - L)\widetilde{H}^{1j}(L), \qquad j = 1, 2, \qquad (A.3)$$

and $u_t = (u_{1t}, u_{2t})$ is a vector white noise with covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Assuming that u_{1t} and u_{2t} are not perfectly correlated, they can be decomposed as

$$u_{1t} = u_{Pt} \quad \text{and} \quad u_{2t} = \Sigma_{11}^{-1}\Sigma_{21}u_{1t} + u_{Tt}. \qquad (A.4)$$

From (A.2) and (A.4),

$$\lim_{h\to\infty} \frac{\partial E_t(p_{t+h})}{\partial u_{Pt}} = H^{11}(1) + H^{12}(1)\Sigma_{11}^{-1}\Sigma_{21} \qquad (A.5)$$

and

$$\lim_{h\to\infty} \frac{\partial E_t(p_{t+h})}{\partial u_{Tt}} = H^{12}(1). \qquad (A.6)$$

Noticing that

$$\lim_{h\to\infty} E_t(X_{t+h}) = \lim_{h\to\infty} E_t(P_{t+h}), \qquad (A.7)$$

(P_t, T_t) will be P–T decomposition according to Definition 1 iff $H^{12}(1) = H_{11}^{-1}(1)H_{12}(1)\,[H_{21}(1)H_{11}(1)^{-1}H_{12}(1) - H_{22}(1)]^{-1} = 0$. In other words, iff the total multiplier of ΔP_t with respect to T_t is 0,

$$H_{11}^{-1}(1)H_{12}(1) = 0. \qquad (A.8)$$

Proof of Proposition 3. If $\gamma\alpha'$ has only $p - r$ eigenvalues equal to 0, then $\text{rank}(\alpha'\gamma) = r$. Taking determinants on the right side of the matrix multiplication

$$\begin{bmatrix} \alpha' \\ \gamma_\perp' \end{bmatrix} \begin{bmatrix} \gamma & \gamma_\perp \end{bmatrix} = \begin{bmatrix} \alpha'\gamma & \alpha'\gamma_\perp \\ 0 & \gamma_\perp'\gamma_\perp \end{bmatrix}, \qquad (A.9)$$

it follows that this matrix has full rank and therefore

$$\text{rank of} \begin{bmatrix} \alpha' \\ \gamma_\perp' \end{bmatrix} = p. \qquad (A.10)$$

Proof of Proposition 5. In this proof, for simplicity it is assumed that X_t follows an AR(q) as in (17).

Multiplying the ECM (17) by $\gamma_\perp'$ and substituting $X_t = A_1 f_t + A_2 z_t$ into (17), we get the AR representation of the common factors f_t

$$\Delta f_t = \sum_{i=1}^{q-1} \gamma_\perp'\Gamma_i A_1 \Delta f_{t-i} + \sum_{i=1}^{q-1} \gamma_\perp'\Gamma_i A_2 \Delta z_{t-i} + \gamma_\perp'\epsilon_t. \qquad (A.11)$$

From (A.11), the random-walk part [in the Beveridge–Nelson (1981) sense] of f_t is

$$\left(I - \sum_{i=1}^{q-1} \gamma_\perp'\Gamma_i A_1\right)^{-1} \gamma_\perp'(I - L)^{-1}\epsilon_t. \qquad (A.12)$$

The common trend decomposition of Stock and Watson (1988) is obtained from the Wold representation of ΔX_t,

$$\Delta X_t = C(L)\epsilon_t = C(1)\epsilon_t + \Delta\widetilde{C}(L)\epsilon_t, \qquad (A.13)$$

where

$$C(1) = \alpha_\perp(\gamma_\perp'\Psi\alpha_\perp)^{-1}\gamma_\perp' \qquad (A.14)$$

with

$$\Psi = \text{mean lag matrix in } H_1 = I - \cdots - \Gamma_{q-1} + \Pi. \qquad (A.15)$$

Therefore,

$$C(1) = \alpha_\perp(\gamma'_\perp\alpha_\perp)^{-1} \times \left[I - \left(\gamma'_\perp \sum_{i=1}^{q-1} \Gamma_i \alpha_\perp \right) \alpha_\perp (\gamma'_\perp \alpha_\perp)^{-1} \right]^{-1} \gamma'_\perp. \quad \text{(A.16)}$$

The result follows from noticing that $A_1 = \alpha_\perp(\gamma'_\perp\alpha_\perp)^{-1}$.

Proof of Theorem 1. Johansen (1989) showed that the likelihood function of Model (20) can be expressed as

$$L_{\max}^{-2/T} = |S_{00.1}||\gamma'_\perp S_{00}\gamma_\perp|/|\gamma'_\perp(S_{00} - S_{01}S_{11}^{-1}S_{10})\gamma_\perp|. \quad \text{(A.17)}$$

Therefore L is maximized by maximizing

$$|\gamma'_\perp(S_{00} - S_{01}S_{11}^{-1}S_{10})\gamma_\perp|/|\gamma'_\perp S_{00}\gamma_\perp|. \quad \text{(A.18)}$$

This is accomplished by choosing $\gamma_\perp$ to be the eigenvectors corresponding to the $p - r$ smallest eigenvalues of $S_{01}S_{11}^{-1}S_{10}$ with respect to S_{00} and the maximal value is

$$\prod_{i=r+1}^{p}(1 - \hat{\lambda}_i). \quad \text{(A.19)}$$

The result follows from substituting (A.19) in (A.17).

Proof of Theorem 2. The proof follows from proposition 3.11 of Johansen and Juselius (1990).

Proof of Theorem 3. Substituting $\gamma_\perp$ by $G\theta$ in (A.17), it is clear that θ can be estimated as the eigenvectors corresponding to the $(p - r)$ smallest eigenvalues of $G'S_{01}S_{11}^{-1}S_{10}G$ with respect to $G'S_{00}G$.

The distribution of the LR test follows from proposition (3.13) of Johansen and Juselius (1990).

[*Received February 1992. Revised December 1993.*]

REFERENCES

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

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

Aoki, M. (1989), "A Two-Step Space Time Series Modeling Method," *Computer Mathematical Applications*, 17, 1165–1176.

Bell, W. R. (1984), "Signal Extraction for Nonstationary Time Series," *The Annals of Statistics*, 12, 646–664.

Beveridge, S., and Nelson, C. R. (1981), "A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components With Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151–174.

Cochrane, J. (1991), "Univariate vs. Multivariate Forecasts of GNP Growth and Stock Returns: Evidence and Implications for the Persistence of Shocks, Detrending Methods, and Tests of the Permanent Income Hypothesis," Working Paper 3427, National Bureau of Economic Research, Cambridge, MA.

Engle, R. F., and Granger, C. W. J. (eds.) (1991), *Long-Run Economic Relationships: Readings in Cointegration* (Advance Texts in Econometrics), Oxford, U.K.: Oxford University Press.

Engle, R. F., and Kozicki, S. (1990), "Testing for Common Feature," Discussion Paper 9023, University of California, San Diego, Dept. of Economics.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series Models," in *Latent Variables in Socioeconomic Models*, eds. D. Aigner and A. Goldberger, Amsterdam: North-Holland, pp. 365–383.

——— (1982), "Measurement of Linear Dependence and Feedback Between Multiple Time Series," *Journal of the American Statistical Association*, 378, 304–324.

Gonzalo, J. (1994), "Five Alternative Methods of Estimating Long-Run Equilibrium Relationships," *Journal of Econometrics*, 60, 203–233.

Gonzalo, J., and Granger, C. W. J. (1992), "Estimation of Common Long-Memory Components in Cointegrated Systems," Discussion Paper 4, Boston University, Dept. of Economics.

Granger, C. W. J. (1980), "Testing for Causality: A Personal Viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.

——— (1986), "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics*, 48, 213–228.

Granger, C. W. J., and Lin, J. (1992), "Causality in the Long-Run," Discussion Paper 9215, Academica Sinica.

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

——— (1989), "Likelihood Based Inference on Cointegration. Theory and Applications," unpublished lecture notes, University of Copenhagen, Institute of Mathematical Statistics.

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

Johansen, S., and Juselius, K. (1990), "Maximum Likelihood Estimation and Inference on Cointegration—With Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169–210.

Kasa, K. (1992), "Common Stochastic Trends in International Stock Markets," *Journal of Monetary Economics*, 29, 95–124.

Mackinnon, J. G. (1991), "Critical Values for Cointegration Tests," in *Long-Run Economic Relationships: Readings in Cointegration*, eds. R. Engle and C. Granger, Oxford, U.K.: Oxford University Press, pp. 267–276.

Osterwald-Lenum, M. (1992), "A Note With Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54, 461–472.

Peña, D., and Box, G. E. P. (1987), "Identifying a Simplifying Structure in Time Series," *Journal of the American Statistical Association*, 82, 836–843.

Phillips, P. C. B. (1991), "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

Quah, D. (1992), "The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds," *Econometrica*, 60, 107–118.

Stock, J. H., and Watson, M. W. (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

Watson, M. W. (1986), "Univariate Detrending Methods With Stochastic Trends," *Journal of Monetary Economics*, 18, 1–27.