On the Shape of the Likelihood/Posterior in Cointegration Models
Author(s): Frank Kleibergen and Herman K. van Dijk
Reviewed work(s):
Source: *Econometric Theory*, Vol. 10, No. 3/4, Symposium Double Issue: Bayes Methods and Unit Roots (Aug. - Oct., 1994), pp. 514-551
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3532548
Accessed: 20/07/2012 11:09

Econometric Theory, **10**, 1994, 514–551. Printed in the United States of America.

ON THE SHAPE OF THE LIKELIHOOD/POSTERIOR IN COINTEGRATION MODELS

FRANK KLEIBERGEN AND HERMAN K. VAN DIJK
Econometric Institute and Tinbergen Institüt
Erasmus University Rotterdam

A vector autoregressive (VAR) model is specified with equation system parameters, which directly reflect the possible cointegrating nature of the analyzed time series. By using a flat/diffuse prior, we show that the marginal posteriors of the parameters of interest (multipliers of the cointegrating vectors) may be nonintegrable and favor difference stationary models in an undesired way. To choose between stationary, cointegrated, and difference stationary models in a meaningful way, the Jeffreys prior principle is used. We investigate the sensitivity of the posterior results with respect to the construction of the Jeffreys prior. In this context, we also analyze the effect of fixed and stochastic initial values. The theoretical results are illustrated by using a VAR model for short- and long-term interest rates in the United States.

1. INTRODUCTION

Although some small sample properties of estimators in cointegration models are known (see Phillips [21]), classical statistical analysis of cointegration relies mainly on asymptotic distribution theory; see Johansen [10], Engle and Granger [6], Kleibergen and van Dijk [13], and Phillips [18]. Bayesian statistical analysis tends to analyze the small sample properties, which can be derived when the likelihood function is specified. A problem with Bayesian analysis is that analytical formulas for the posterior moments and densities of the parameters are not known for several classes of econometric models. In the present paper we analyze the marginal likelihood (posterior with uniform prior) of the parameters of a cointegration model. It is shown in Section 3 that these marginal likelihoods are not members of a standard class of probability density functions and are ill-behaved in the sense of having asymptotes in the interior of the parameter region. This behavior of the likelihood is due to the nonidentifiedness of certain parameters, which occurs

We thank John Hartigan, Soren Johansen, Katarina Juselius, Peter Phillips, Peter Schotman, and, in particular, three anonymous referees for helpful comments and suggestions, which led to substantial improvements of a preliminary version. The stimulating discussions with other participants of the ESEM meeting at Cambridge and seminars at CORE, INSEE, GREQE, Yale and the universities of Chicago and Kopenhagen are gratefully acknowledged. Any errors are, of course, our own responsibility.

  0266-4666/94 $5.00 + .00

when the model is a difference stationary one. Thus, the results indicate that flat priors are very informative in cointegration models because difference stationary models are "infinitely" favored. To choose in a meaningful way among stationary, cointegration, and difference stationary models, we make use of the Jeffreys prior principle, i.e., we choose a prior that is proportional to the square root of the determinant of the information matrix. In Section 4, we analyze the sensitivity of the posteriors of the parameters of interest with respect to four different cases of a Jeffreys prior. One of these priors has implicitly been used by DeJong [3] to calculate the posteriors of the roots of vector autoregressive (VAR) models. A second one has been used by Phillips [19] to analyze unit roots in univariate autoregressive models. We also discuss the problem of fixed or stochastic initial observations. Here we introduce a prior that approximates the Jeffreys prior for the case of the exact likelihood. Preliminary to our Bayesian cointegration analysis, we discuss in Section 2 several aspects of specification of cointegration models. We note that a companion paper, Kleibergen and van Dijk [13], deals with a classical statistical analysis of our specification of the cointegration model.

To save on indices, the data series are depicted as row vectors. Also, the common expression for the longrun multiplier $\alpha\beta'$ is replaced by $\beta\alpha$, where β stands for the cointegrating vectors.

2. SPECIFICATION OF COINTEGRATION MODELS

Cointegration describes special features of multiple time series; see Engle and Granger [6]. To analyze cointegration, one needs a model that explains the joint behavior of the analyzed time series. We make use of the VAR model. A pth order VAR model of k elements of the series $x_t = (x_{1t} \cdots x_{kt})$, $t = 1, \ldots, T$, which conditions on the first p observations, reads

$$(x_t - \mu - t\delta)\Pi(L) = \varepsilon_t \qquad t = 1, \ldots, T \tag{1}$$

$$\Pi(z) = I_k - \sum_{i=1}^{p} z^i \Pi_i \tag{2}$$

where x_t, μ, and δ are $1 \times k$ *row vectors*. The parameter vectors μ and δ represent the (nonzero) mean and growth level of the analyzed series x_t. The deterministic parameters are modeled in a multiplicative way to guarantee that their interpretation does not change when unit roots become present. For details on the standard assumptions for VAR models, we refer to Lütkepohl [15].

To define cointegration in the VAR model (1), we rule out any explosive and infinite cyclical behavior. The roots of the characteristic polynomial, $|\Pi(z)| = 0$, are therefore assumed to lie outside the unit circle or to be equal to one; see Johansen [10]. If $k - r$ roots of the characteristic polynomial are equal to one, $0 < r < k$, we say that the series generated by the VAR model (1) are cointegrated. Cointegration implies that the matrix of longrun multipliers, $\Pi = -\Pi(1)$, has a lower rank value. As a consequence, this matrix can be specified as the product of two full rank $k \times r$ matrices, β and α':

$$\Pi = \beta\alpha \qquad \beta, \alpha' : k \times r. \tag{3}$$

The cointegrating vectors β show the r stationary cointegrating (equilibrium) relationships $x_t\beta$.

In case of cointegrated series, the VAR model (1) is defined in terms of the nonstationary variables x_t. Models defined in terms of stationary components are often preferred to models defined in terms of nonstationary components. One may respecify the VAR model (1) such that it only contains stationary components. Two of such specifications, which are observationally equivalent with the VAR model (1), are the error correction model (ECM) and the structural form model. Apart from containing only stationary components, Δx_t and $x_{t-p}\beta$, the ECM has the attractive property that the longrun multiplier is directly estimable. The specification of the ECM reads

$$\Delta(x_t - t\delta)\Gamma(L) = (x_{t-p} - \mu - (t-p)\delta)\beta\alpha + \varepsilon_t \qquad t = 1, \ldots, T, \tag{4}$$

where use is made of a decomposition of the VAR lag-polynomial:

$$\Pi(z) = (1 - z)\Gamma(z) - z^p\beta\alpha. \tag{5}$$

The VAR model (1) and the ECM (4) may be considered as reduced form models because they do not explicitly model the cointegrating (equilibrium) relationships. A possible structural form model reads

$$((x_t - \mu - t\delta)\beta \ \ (\Delta x_t - \delta)A)\Phi(L) = \xi_t \qquad t = 1, \ldots, T, \tag{6}$$

where the invertible VAR polynomial $\Phi(z)$ is specified by

$$\Phi(z) = \begin{pmatrix} \Phi_1(z) \\ \Phi_2(z) \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & (1-z)^{-1}I_{k-r} \end{pmatrix} (\beta \ \ A)^{-1}\Pi(z)(\beta \ \ A) \tag{7}$$

and $\xi_t = \varepsilon_t(\beta \ \ A)$ and $(\beta \ \ A)$ has full rank. Possible choices of A are $A = (0 \ \ I_{k-r})'$ or $A = \beta_\perp$ (the orthogonal complement of β). Through the invertibility of the VAR polynomial $\Phi(z)$, the structural model (6) allows one to construct the implied vector moving average (VMA) representation and/or stochastic trend representation; see Johansen [10] and Kleibergen and van Dijk [13].

Given the specification of the longrun multiplier, $\Pi = \beta\alpha$, one is usually confronted with an identification problem. The number of parameters in $\beta\alpha$, $2kr$, is in most cases not equal to the number of parameters in Π, k^2. Thus, parameters in β and α have to be restricted before estimation. Classical cointegration procedures overcome the identification problem by estimating the cointegrating vector with a "data-parametric technique" such as canonical correlations (Box and Tiao [1] and Johansen [10]) or principal components (Stock and Watson [23]). The procedure suggested by Johansen [10] is rather well known because of the elegant relationship between the canonical correlations and the number of cointegrating relationships or unit roots. In principle, one may perform a Bayesian analysis by using a model like Johansen's but a prior has to be specified on the canonical correlations of the system, which is not trivial. In this paper, we construct a model that contains equation system parameters, which reflect a possible departure from a cointegration model, by using a suitable specification of the longrun multiplier Π; see also Kleibergen and van Dijk [13]. Let β and α be redefined as

$$\beta = \begin{pmatrix} I_r & 0 \\ -\beta_2 & I_{k-r} \end{pmatrix}, \qquad \alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix}, \quad \text{then} \tag{8}$$

$$\Pi = \beta\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ -\beta_2\alpha_{11} & -\beta_2\alpha_{12} + \alpha_{22} \end{pmatrix}, \tag{9}$$

where $\beta_2: (k-r) \times r$, $\alpha_{11}: r \times r$, $\alpha_{12}: r \times (k-r)$, $\alpha_{22}: (k-r) \times (k-r)$ are all unrestricted.

We make use of the same symbols β and α as before. However, the interpretation has changed considerably when $\alpha_{22} \neq 0$. When $\alpha_{22} = 0$, Π has a lower rank value and the specification of $\Pi = \beta\alpha$ corresponds with the specification in equation (3) with $\beta = (I_r \quad -\beta_2')'$ and $\alpha = (\alpha_{11} \quad \alpha_{12})$. If $\alpha_{22} \neq 0$, the interpretation of β does not correspond with a cointegrating vector. Tests for the number of cointegrating vectors or unit roots can be performed by testing whether $\alpha_{22} = 0$ for different values of r. The specification of the possible cointegrating vector $\beta(= (I_r \quad -\beta_2')')$ can be considered as a kind of reduced-form specification of the cointegrating vector. Under cointegration, more general specifications of $\beta(= (I_r \quad -\beta_2')')$ can be constructed, but these specifications do not allow for parameters that measure the departure from a cointegration model like α_{22}.

If α_{11} has full rank, the parameters β_2, α_{11}, α_{12}, and α_{22} are exactly identified and can be obtained from Π. The specification of Π in (8) is by no means unique, however, and $\binom{k}{r}$ $(=k!/(r!\,(k-r)!))$ different parameterizations of Π exist, each of which contains a parameter that reflects a departure from a cointegration model.

Short- and long-term interest rates in the United States are used to illustrate the analysis. The short-term U.S. interest rate is the 3-month U.S. treasury bill rate, and the long-term interest rate series refers to securities that have a maturity of 10 years. Both series are obtained from the "Main Economic Indicators" databank of the OECD, from January 1957 to April 1989 (388 observations) and are shown in Figure 1.

The Dickey–Fuller statistics in Table 1 indicate that the hypothesis that both interest rate series are nonstationary cannot be rejected, but the economic theory of term structures indicates that certain relationships between interest rate series should hold; see Campbell and Shiller [2]. Consequently,

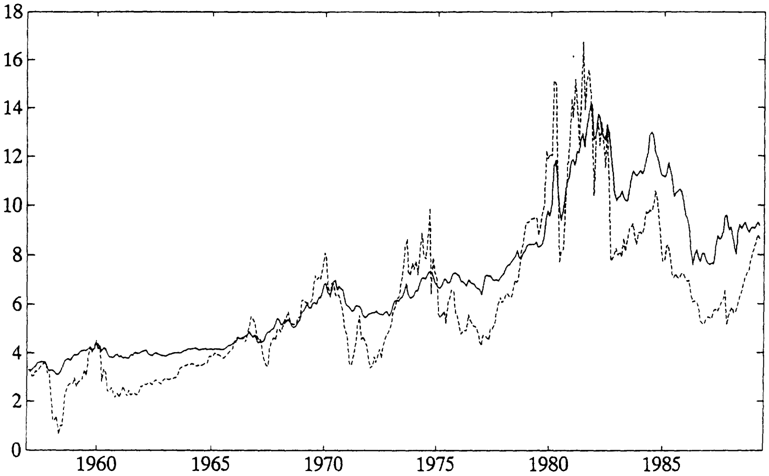

FIGURE 1. U.S. short-term (--) and long-term (—) interest rates, 1957–1989.

the interest rate series may be cointegrated. Table 1 also shows Wald, likelihood ratio, and Lagrange multiplier statistics to test the hypothesis of a certain number of cointegrating relationships. For details on the derivation of these statistics and their asymptotic distributions, we refer to Kleibergen and van Dijk [13]. By using the asymptotic critical values obtained from

TABLE 1. Classical cointegration characteristics of U.S. interest rates ($p = 4$)

	long term	short term	
Dickey–Fuller	−1.26	−2.10	

$$\hat{\Pi}_{OLS} = \begin{bmatrix} \underset{(-1.96)}{-0.056} & \underset{(2.64)}{0.027} \\ \underset{(1.21)}{0.036} & \underset{(-2.85)}{-0.031} \\ \underset{(1.11)}{0.11} & \underset{(1.81)}{0.062} \end{bmatrix} \begin{matrix} \text{long term} \\ \text{short term} \\ \text{constant} \end{matrix}$$

Number of cointegrating relationships	0	1
t_{Wald}	7.17	2.98
t_{LR}	7.10	2.69
t_{LM}	7.04	2.26
Critical (95%)	20.1	9.09

Estimated cointegrating vector	long	short	constant
Maximum likelihood	1.0	0.93 (10.1)	−0.60 (−0.69)

Johansen and Juselius [11], all three statistics indicate that the hypothesis of zero cointegrating vectors, i.e., difference stationary model, cannot be rejected. However, according to the theory of term structures, there should be one cointegrating vector, $(1 \quad -1)'$. We will return to this example in the next sections. To keep matters simple, we do not focus on the heteroskedastic and leptocurtic nature of the disturbances of interest rate models. The hypothesis tested in Table 1 includes both the interest rates and the constant term in the cointegrating relationships such that no linear time trends can become present.

3. PATHOLOGICAL BEHAVIOR OF MARGINAL POSTERIORS

The possibility of pathological behavior of posteriors, i.e., nonintegrability, is a complication for Bayesian analysis. Some examples of models where the posterior shows pathological behavior are discussed in Schotman and van Dijk [22] for univariate autoregressive (AR) models and in Kleibergen and van Dijk [12] for simultaneous equations models (SEM). The cointegration models mentioned in the previous section contain properties of both classes of models discussed in these two papers. One source for pathological behavior is the product of the parameters β and α in model (4). When one of these two parameters equals 0, the other parameter automatically drops out of the model. Thus, the latter parameter is then nonidentified because all possible values of this parameter have the same influence on the model. The specification of the parameter matrices β and α in our analysis is such that only α can be equal to 0 or have a lower rank value; see equations (8)–(9). The product $\beta\alpha$ is such that elements of β_2 or α_{22} become nonidentified when α_{11} has a lower rank value. When, for example, $\alpha_{11} = 0$, the product $\beta\alpha$ becomes

$$\Pi = \begin{pmatrix} 0 & \Pi_{12} \\ 0 & \Pi_{22} \end{pmatrix} = \beta\alpha = \begin{pmatrix} 0 & \alpha_{12} \\ 0 & -\beta_2\alpha_{12} + \alpha_{22} \end{pmatrix}. \tag{10}$$

The $k(k-r)$ different elements of β_2 and α_{22} have to be obtained from the $(k-r)(k-r)$ different elements of Π_{22}. As a consequence, $r(k-r)$ elements of β_2 and α_{22} remain unidentified. If one assigns all these elements to β_2, it follows that β_2 is unidentified when $\alpha_{11} = 0$. Thus, given that α_{11} has full rank, the parameter β_2 is locally identified.

The specification in (9) has a parameter, α_{22}, which represents a departure from a cointegration model. We note that for the cointegration model, where $\alpha_{22} = 0$, nonidentifiedness of elements of β_2 is possible. Therefore, even for the cointegration model where $\alpha_{22} = 0$, β_2 is only locally identified (when $(\alpha_{11} \quad \alpha_{12})$ has full rank).

The nonidentifiability problem, which results in pathological behavior of marginal posteriors, is not restricted to parameters of the cointegrating vector. Elements of the deterministic component parameters μ and δ can also become nonidentified for certain specific values of other parameters. This is illustrated by using the ECM (4):

$$\mu\beta\alpha = ((\mu_1 - \mu_2\beta_2)\alpha_{11}(\mu_1 - \mu_2\beta_2)\alpha_{12} + \mu_2\alpha_{22}); \tag{11}$$

$$\begin{aligned}\delta(\Gamma(1) - tL^p\beta\alpha) = \Bigg(&\delta_1 - (\delta_1 \quad \delta_2)\sum_{i=1}^{p-1}\Gamma_{1i} - (t-p)(\delta_1 - \delta_2\beta_2)\alpha_{11} \\ &\delta_2 - (\delta_1 \quad \delta_2)\sum_{i=1}^{p-1}\Gamma_{2i} - (t-p) \\ &\times [(\delta_1 - \delta_2\beta_2)\alpha_{12} + \delta_2\alpha_{22}]\Bigg).\end{aligned} \tag{12}$$

When α has full rank, the elements of μ_1 and μ_2 are identified in the product $\mu\beta\alpha$ in (11). When $\alpha_{22} = 0$, the term $\mu_1 - \mu_2\beta_2$ is identified, and it is not possible to determine the distinct elements of μ_1 and μ_2. The same reasoning holds for the growth term parameter δ, equation (12), which contains nonidentified elements when the ECM equals a cointegration model in second differences, i.e., both $\alpha_{22} = 0$ and $\Gamma(1) = I_k - \Sigma\Gamma_i$ has lower rank. For the cointegration model, $\alpha_{22} = 0$, it follows that only $\nu = \mu_1 - \mu_2\beta_2$ and $(\delta_1 \quad \delta_2)$ are locally identified (when $(\alpha_{11} \quad \alpha_{12})$, $\Gamma(1)$ have full rank).

In classical statistical analysis of stationary nonlinear models, the estimators of locally nonidentified parameters may converge to random variables instead of their fixed true values; see Phillips [17]. Bayesian analysis of models with locally nonidentified parameters is also rather difficult; see Kleibergen and van Dijk [12] and Schotman and van Dijk [22]. In Bayesian analysis, the problem originates from the constancy of the likelihood along the axis of the nonidentified parameters, say β_2, in a nonidentified parameter point, say $\alpha_{11} = 0$. In the ECM (4), for example, when $\alpha_{11} = 0$, equation (10) shows that the likelihood will be constant for all values of β_2 and α_{22} for which $\Pi_{22} = -\beta_2\alpha_{12} + \alpha_{22}$, where both Π_{22} and α_{12} are fixed. So, even for infinite values of β_2 and α_{22}, the likelihood will still have a non-zero value when $\alpha_{11} = 0$ and $\Pi_{22} = -\beta_2\alpha_{12} + \alpha_{22}$. As a consequence, the integral of the likelihood with respect to the parameters (β_2, α_{22}) on the region $\mathbb{R}^{k(k-r)}$ will be infinite. Another way of explaining the problem is that the conditional variance of the parameter β_2, given $\alpha_{11}(= 0)$, is infinite, in the nonidentified parameter points. Thus, the information matrix (= inverse covariance matrix) is singular in a nonidentified parameter point.

Although the likelihood may not be integrable, the posteriors of the parameters may be integrable once a suitable prior is chosen. In the following sections, different priors are constructed, and we analyze whether these priors lead to integrable posteriors. The integrability of the posterior is of great importance because, in case of nonintegrable posteriors, inference is difficult. Yet, it is important to investigate the properties of marginal likelihoods. First, because the data information may be such that the nonidentified points (or regions) in the parameter space are relatively far from the region where the data information is important. In other words, the marginal likelihood has a strong local mode far from the nonidentified parameter value. In this case, the use of uniform priors truncated near the nonidentified parameter points lead to proper posteriors, which are not sensitive to the truncation. Second, to construct a class of prior densities that "conforms in some sense" to the likelihood, one has to know the special features of the latter ones. In the remainder of this section, we investigate the properties of the marginal likelihood of a cointegration model.

So far, three different parameters, μ, δ, and β_2, are mentioned, which could contain nonidentified elements. To keep matters simple, we analyze an ECM without deterministic components to focus on the consequences of the identifiedness problems of β_2. The problems concerning the deterministic components will then be discussed briefly. A respecification of the ECM in (4) without deterministic components yields

$$\Delta X = Z\Gamma + X_{-p}\beta\alpha + \varepsilon, \tag{13}$$

where $\Delta X = (\Delta x_1' \cdots \Delta x_T')'$, $Z = (Z_1 \cdots Z_{p-1})$, $Z_i = (\Delta x_{1-i}' \cdots \Delta x_{T-i}')'$, $X_{-p} = (x_{1-p}' \cdots x_{T-p}')'$, $\varepsilon = (\varepsilon_1' \cdots \varepsilon_T')'$, $\Gamma = (\Gamma_1' \cdots \Gamma_{p-1}')'$. Assume that the disturbances ε_t, $t = 1, \ldots, T$ are independently generated by a multivariate normal distribution with mean 0 and covariance matrix Ω. Then the likelihood reads

$$l(\beta,\alpha,\Gamma,\Omega \mid X,Z) \propto |\Omega|^{-(1/2)T}\exp[-\tfrac{1}{2}tr\,\Omega^{-1}(\Delta X - Z\Gamma - X_{-p}\beta\alpha)' \times (\Delta X - Z\Gamma - X_{-p}\beta\alpha)] \tag{14}$$

A Bayesian analysis starts with the specification of a prior density. Because we want to analyze the properties of the likelihood of a cointegration model in detail, we choose a diffuse prior,

$$p(\beta,\alpha,\Gamma,\Omega) \propto |\Omega|^{-(1/2)h}. \tag{15}$$

The posterior is proportional to the product of the prior and the likelihood:

$$p(\beta,\alpha,\Gamma,\Omega \mid X,Z) \propto |\Omega|^{-1/2(T+h)}\exp[-\tfrac{1}{2}tr\,\Omega^{-1}(\Delta X - Z\Gamma - X_{-p}\beta\alpha)' \times (\Delta X - Z\Gamma - X_{-p}\beta\alpha)]. \tag{16}$$

The identification problems especially concern the parameters α and β and only indirectly influence the remaining parameters Γ and Ω. We are primarily interested in the marginal and conditional posteriors of α and β. To derive these posteriors, we have to integrate the parameters Γ and Ω out of the joint posterior (16). Figure 2 contains an integration scheme for the construction of the joint posterior of α and β, where $M_Z = I_T - Z(Z'Z)^{-1}Z'$. For more details on the integration steps in Figure 2, see Drèze and Richard [5] and Zellner [25]. Given the joint posterior of α and β, one can construct the conditional posteriors of α given β and β given α.

THEOREM 1. *Given equations* (13)–(16), *the kernels of the conditional posteriors of* α, *given* β, *and* β, *given* α, *by using the diffuse prior* (15) *read*

$$p(\beta,\alpha,\Gamma,\Omega|X,Z) \propto |\Omega|^{-1/2(T+h)}\exp[-\tfrac{1}{2}tr\,\Omega^{-1}\epsilon'\epsilon)$$

$\downarrow$

inverted-Wishart step on Ω

$\downarrow$

$$p(\beta,\alpha,\Gamma|X,Z) \propto |\epsilon'\epsilon|^{-1/2(T+h-k-1)}$$

$\downarrow$

matric-variate t step on Γ

$\downarrow$

$$p(\beta,\alpha|X,Z) \propto |(\Delta X - X_{-p}\beta\alpha)'M_Z(\Delta X - X_{-p}\beta\alpha)|^{-1/2(T+h-pk-1)}$$

FIGURE 2. Integration scheme for the construction of the joint posterior of α and β, marginal with respect to Ω and Γ.

$$p(\alpha\,|\,\beta,X,Z) \propto |\Delta X'M_{(Z\ \ X_{-p}\beta)}\Delta X + (\alpha - \hat{\alpha})'\beta'X'_{-p}M_ZX_{-p}\beta(\alpha - \hat{\alpha})|^{-1/2(T+h-pk-1)}, \quad \textbf{(17)}$$

$$p(\beta\,|\,\alpha,X,Z) \propto |(\beta - \hat{\beta})'X'_{-p}M_{(Z\ \ \Delta X\alpha'_{\perp})}X_{-p}(\beta - \hat{\beta}) + (\alpha(\Delta X'M_{(Z\ \ X_{-p})}\Delta X)^{-1}\alpha')^{-1}|^{-1/2(T+h-pk-1)}, \quad \textbf{(18)}$$

where $\hat{\beta} = (X'_{-p}M_ZX_{-p})^{-1}X'_{-p}M_Z\Delta X(\Delta X'M_{(Z\ \ X_{-p})}\Delta X)^{-1}\alpha'(\alpha(\Delta X' \times M_{(Z\ \ X_{-p})}\Delta X)^{-1}\alpha')^{-1}$ *and* $\hat{\alpha} = (\beta'X'_{-p}M_ZX_{-p}\beta)^{-1}\beta'X'_{-p}M_Z\Delta X.$

Proof. See Appendix.

We emphasize that the functional forms of the posteriors in (17) and (18) are independent of the functional form of α and β. So, the conditional posterior corresponding to the case of cointegration, $\alpha_{22} = 0$, as well as the conditional posterior corresponding to the case of no cointegration, $\alpha_{22} \neq 0$, are described by Theorem 1. Notice, however, that when $\alpha_{22} \neq 0$, the matrix α is assumed to have full rank such that $\alpha_{\perp} = 0$ (empty matrix).

By using some rules of matrix analysis for decompositions of the determinants in Theorem 1, the conditional posteriors of the individual parameters α_{11}, α_{12}, α_{22}, and β_2 can be constructed. These conditional posteriors are all proportional to matric-variate t densities, regardless of the chosen model. In Theorem 2, the functional forms of the conditional matric-variate t posteriors are stated (for a definition of a matric-variate t density, see the Appendix).

THEOREM 2. *Given the conditions of Theorem* 1, *let* $\alpha = (\alpha_{11} \quad \alpha_{12})$, *and* $\beta = (I_r \ -\beta'_2)'$. *The conditional posteriors of* α *and* β_2 *become*

$$p(\alpha\,|\,\beta,X,Z) = f_{Mt}(\alpha\,|\,\hat{\alpha},\beta'X'_{-p}M_ZX_{-p}\beta,\Delta X'M_{(Z\ \ X_{-p}\beta)}\Delta X, T + h - pk - r - 1) \quad \textbf{(19)}$$

$$p(\beta_2 \mid \alpha, X, Z) = f_{Mt}(\beta_2 \mid \hat{\beta}_{21}, X'_{-p2} M_{(Z\ \Delta X\alpha'_{\perp})} X_{-p2},$$
$$(\alpha(\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha')^{-1}$$
$$+ (I_r - \hat{\beta}_1)' X'_{-p1} M_{(Z\ \Delta X\alpha'_{\perp}\ X_{-p2})} X_{-p1} (I_r - \hat{\beta}_1),$$
$$T + h - pk - (k - r) - 1). \tag{20}$$

Let $\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix} = (\alpha_1 \quad \alpha_2)$,

and $\beta = \begin{pmatrix} I_r & 0 \\ -\beta_2 & I_{k-r} \end{pmatrix}$.

The conditional posteriors of α_{11}, α_2, *and* β_2 *become*

$$p(\alpha_2 \mid \alpha_1, \beta, X, Z) = f_{Mt}(\alpha_2 \mid \hat{\alpha}_2, \Delta X'_2 M_{(Z\ \Delta X_1\ X_{-p}\beta)} \Delta X_2,$$
$$\beta' X'_{-p} M_{(Z\ \Delta X_1 - X_{-p}\beta\alpha_1)} X_{-p}\beta,$$
$$T + h - pk - k - 1), \tag{21}$$

$$p(\alpha_{11} \mid \beta, X, Z) = f_{Mt}(\alpha_{11} \mid \hat{\alpha}_{11}, \Delta X'_1 M_{(Z\ X_{-p1} - X_{-p2}\beta_2)} \Delta X_1,$$
$$(X_{-p1} - X_{-p2}\beta_2)' M_Z (X_{-p1} - X_{-p2}\beta_2),$$
$$T + h - pk - (k - r) - r - 1), \tag{22}$$

$$p(\beta_2 \mid \alpha_{11}, X, Z) = f_{Mt}(\beta_2 \mid \hat{\beta}_{22}, X'_{-p2} M_Z X_{-p2}, \alpha_{11}^{-1\prime}$$
$$\times [\Delta X_1 M'_{(Z\ X_{-p})} \Delta X_1 + (\alpha_{11} - \hat{\pi}_{11})'$$
$$\times X'_{-p1} M_{(Z\ X_{-p2})} X_{-p1} (\alpha_{11} - \hat{\pi}_{11})] \alpha_{11}^{-1},$$
$$T + h - pk - 2(k - r) - 1), \tag{23}$$

where

$$\hat{\beta} = (\hat{\beta}'_1 \quad \hat{\beta}'_2)',$$
$$\hat{\beta}_{21} = \hat{\beta}_2 - (X'_{-p2} M_{(Z\ \Delta X\alpha'_{\perp})} X_{-p2})^{-1} X'_{-p2} M_{(Z\ \Delta X\alpha'_{\perp})} X_{-p1} (I_r - \hat{\beta}_1),$$
$$\hat{\beta}_{22} = -\hat{\pi}_{21}\alpha_{11}^{-1} - (X'_{-p2} M_Z X_{-p2})^{-1} X'_{-p2} M_Z X_{-p1} (\alpha_{11} - \hat{\pi}_{11})\alpha_{11}^{-1},$$
$$\hat{\Pi}_1 = (X'_{-p} M_Z X_{-p})^{-1} X'_{-p} M_Z \Delta X_1 = \begin{pmatrix} \hat{\pi}_{11} \\ \hat{\pi}_{21} \end{pmatrix},$$
$$\hat{\alpha}_2 = (\beta' X'_{-p} M_{(Z\ \Delta X_1 - X_{-p}\beta\alpha_1)} X_{-p}\beta)^{-1} \beta' X'_{-p} M_{(Z\ \Delta X_1 - X_{-p}\beta\alpha_1)} \Delta X_2,$$
$$\hat{\alpha}_{11} = ((X_{-p1} - X_{-p2}\beta_2)' M_Z (X_{-p1} - X_{-p2}\beta_2))^{-1}$$
$$\times (X_{-p1} - X_{-p2}\beta_2)' M_Z \Delta X_1,$$
$$X_{-p1} : T \times r, \qquad X_{-p2} : T \times (k - r), \qquad X_{-p} = (X_{-p1} \quad X_{-p2}).$$

Proof. Use Theorem 1 and the decomposition theorems of matric-variate t densities stated in Zellner [25, p. 397].

Although the conditional posteriors stated in Theorem 2 seem to indicate a regular functional form, they contain some peculiarities. For expository purposes, we analyze the mean and variance of the conditional posteriors of β_2, given α.

$$\begin{aligned}
&\mathcal{E}(\beta_2 \,|\, \alpha, \alpha_{22} = 0) \\
&\quad = ((X'_{-p2} M_{(Z\ \Delta X \alpha'_\perp)} X_{-p2})^{-1} X'_{-p2} M_{(Z\ \Delta X \alpha'_\perp)} X_{-p1} \quad I_{k-r})\hat{\beta} \\
&\qquad - (X'_{-p2} M_{(Z\ \Delta X \alpha'_\perp)} X_{-p2})^{-1} X'_{-p2} M_{(Z\ \Delta X \alpha'_\perp)} X_{-p1}.
\end{aligned} \tag{24}$$

$$\begin{aligned}
&\operatorname{var}(\operatorname{vec}(\beta_2) \,|\, \alpha, \alpha_{22} = 0) \\
&\quad = (T + h - pk - (k-r) - r - 3)^{-1} \\
&\qquad \times \big([(\alpha(\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha')^{-1} \\
&\qquad\qquad + (I_r - \hat{\beta}_1)' X'_{-p1} M_{(Z\ \Delta X \alpha'_\perp\ X_{-p2})} X_{-p1} (I_r - \hat{\beta}_1)] \\
&\qquad\qquad \otimes (X'_{-p2} M_{(Z\ \Delta X \alpha'_\perp)} X_{-p2})^{-1} \big).
\end{aligned} \tag{25}$$

$$\begin{aligned}
&\mathcal{E}(\beta_2 \,|\, \alpha_{11}, \alpha_{22} \neq 0) \\
&\quad = ((X'_{-p2} M_Z X_{-p2})^{-1} X'_{-p2} M_Z X_{-p1} \quad I_{k-r}) \hat{\Pi}_1 \alpha_{11}^{-1} \\
&\qquad - (X'_{-p2} M_Z X_{-p2})^{-1} X'_{-p2} M_Z X_{-p1}.
\end{aligned} \tag{26}$$

$$\begin{aligned}
&\operatorname{var}(\operatorname{vec}(\beta_2) \,|\, \alpha_{11}, \alpha_{22} \neq 0) \\
&\quad = (T + h - pk - 2(k-r) - r - 3)^{-1} \\
&\qquad \times (\alpha_{11}^{-1\prime} [\Delta X_1 M'_{(Z\ X_{-p})} \Delta X_1 + (\alpha_{11} - \hat{\pi}_{11})' X'_{-p1} M_{(Z\ X_{-p2})} \\
&\qquad\qquad \times X_{-p1} (\alpha_{11} - \hat{\pi}_{11})] \alpha_{11}^{-1} \otimes (X'_{-p2} M_Z X_{-p2})^{-1}).
\end{aligned} \tag{27}$$

Because $\hat{\beta}$ converges to infinity (see Theorem 1) when α converges to a lower rank value, the mean of the conditional posterior of β_2 with $\alpha_{22} = 0$ also converges to infinity when α converges to a lower rank value. The mean of the conditional posterior of β_2 is, therefore, infinite, when α has a lower rank value. The same argument holds for the variance of the conditional posterior. It is also infinite for lower rank values of α. The finiteness of the mean of the conditional posterior of β_2 when $\alpha_{22} \neq 0$ depends on the rank of α_{11}. When α_{11} has a lower rank value, (26) and (27) show that the mean and variance of the conditional posterior are infinite. The question is how important these infinite means and variances are in practice. If the probability of a lower rank value of α_{11} or α is negligible, one would, in practice, not notice the infinite means and variances of the conditional posteriors. To show the importance of the infinite means and variances of the conditional posterior, we calculate the joint posterior of α_{11} and β_2 for the earlier mentioned fourth-order bivariate ECM describing the joint behavior of U.S. long- and short-term interest rates. Because the model is bivariate, $k = 2$, and we investigate the plausibility of one cointegrating relationship, $r = 1$. The model used allows α_{22} to be different from 0. In Figures 3 and 4, the bivariate posterior and the contourlines of the bivariate posterior, respectively, of α_{11} and β_2 are drawn. To avoid the problems involved with the use of deterministic components, the interest rate series are used in deviation from their means. Both figures show the nonnegligible probability of α_{11} lower rank ($\alpha_{11} = 0$). Thus, the infiniteness of the mean and variance of the conditional posterior of β_2 is really important. This is confirmed by the huge tails of the posterior, which are located at $\alpha_{11} = 0$.

Although it does not hold generally, possible infiniteness of the means and variances of the conditional posteriors affects the marginal posteriors in the models analyzed. In Theorem 3, the marginal posteriors of α and β_2 for the case where $\alpha_{22} = 0$, and the marginal posteriors of α_{11} and β_2 for the case where α_{22} is not restricted to be equal to 0, are given.

THEOREM 3. *Given the conditions stated in Theorem* 1, *let* $\alpha = (\alpha_{11} \quad \alpha_{12})$, *and* $\beta = (I_r - \beta_2')'$. *The kernels of the marginal posteriors of* α *and* β_2 *become*

$$\begin{aligned} p(\beta_2 \mid X, Z) \propto\ & |(\beta_2 - \hat{\beta}_{23})' X_{-p2}' M_Z X_{-p2} (\beta_2 - \hat{\beta}_{23}) \\ & + X_{-p1}' M_{(Z\ X_{-p2})} X_{-p1}|^{1/2(T+h-pk-r-k-1)} |X_{-p1}' \\ & \times M_{(Z\ \Delta X\ X_{-p2})} X_{-p1} + (\beta_2 - \hat{\beta}_{24})' X_{-p2}' M_{(Z\ \Delta X)} \\ & \times X_{-p2} (\beta_2 - \hat{\beta}_{24})|^{-1/2(T+h-pk-r-1)}; \end{aligned} \quad \textbf{(28)}$$

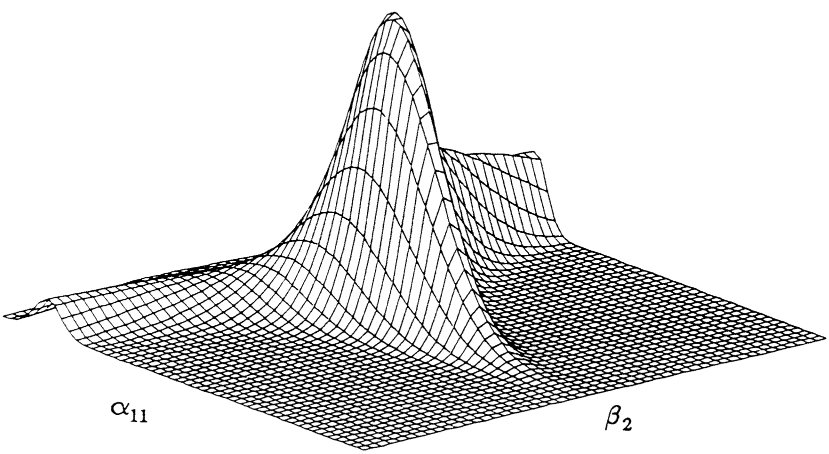

FIGURE 3. Bivariate posterior (α_{11}, β_2) for fourth-order ECM for U.S. interest rates, with α_{22} unrestricted.

$$
\begin{aligned}
p(\alpha \mid X, Z) \propto\ & |X'_{-p1} M_{(Z\ \Delta X\alpha'_{\perp}\ X_{-p2})} X_{-p1}|^{1/2(T+h-pk-1)} \\
& \times |X'_{-p2} M_{(Z\ \Delta X\alpha'_{\perp})} X_{-p2}|^{1/2(T+h-pk-r-1)} \\
& \times |\alpha(\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha'|^{-1/2(T+h-pk-1)} \\
& \times |(I_r - \hat{\beta}_1)' X'_{-p1} M_{(Z\ \Delta X\alpha'_{\perp}\ X_{-p2})} X_{-p1} (I_r - \hat{\beta}_1) \\
& \quad + (\alpha(\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha')^{-1}|^{-1/2(T+h-pk-(k-r)-1)}.
\end{aligned} \tag{29}
$$

Let $\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix} = (\alpha_1 \quad \alpha_2)$,

and $\beta = \begin{pmatrix} I_r & 0 \\ -\beta_2 & I_{k-r} \end{pmatrix}$.

The kernels of the marginal posteriors of α_{11} *and* β_2 *become*

$$
\begin{aligned}
p(\beta_2 \mid X, Z) \propto\ & |(\beta_2 - \hat{\beta}_{23})' X'_{-p2} M_Z X_{-p2} (\beta_2 - \hat{\beta}_{23}) \\
& + X'_{-p1} M_{(Z\ X_{-p2})} X_{-p1}|^{1/2(T+h-pk-(k-r)-2r-1)} \\
& \times |X'_{-p1} M_{(Z\ \Delta X_1\ X_{-p2})} X_{-p1} + (\beta_2 - \hat{\beta}_{25})' X'_{-p2} M_{(Z\ \Delta X_1)} \\
& \quad \times X_{-p2} (\beta_2 - \hat{\beta}_{25})|^{-1/2(T+h-pk-(k-r)-r-1)},
\end{aligned} \tag{30}
$$

$$
\begin{aligned}
p(\alpha_{11} \mid X, Z) \propto\ & |\alpha_{11}|^{-(k-r)} |(\alpha_{11} - \hat{\pi}_{11})' X'_{-p1} M_{(Z\ X_{-p2})} X_{-p1} (\alpha_{11} - \hat{\pi}_{11}) \\
& + \Delta X'_1 M_{(Z\ X_{-p})} \Delta X_1|^{-1/2(T+h-pk-2(k-r)-1)},
\end{aligned} \tag{31}
$$

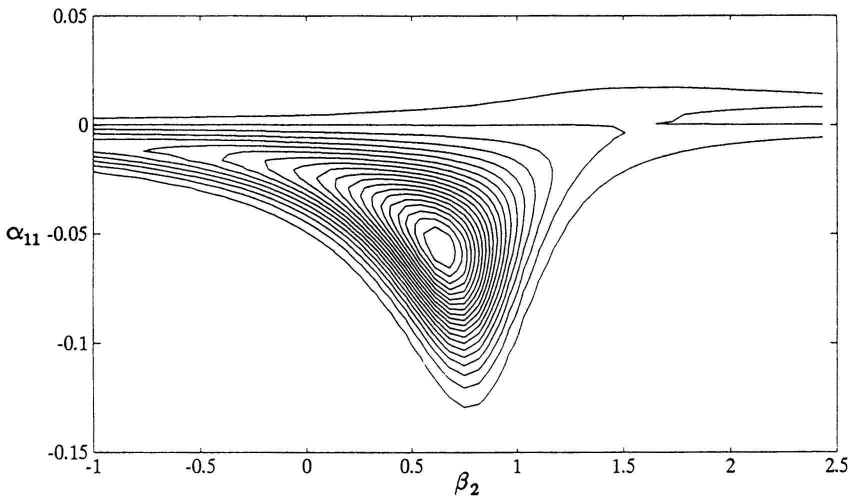

FIGURE 4. Contourlines bivariate posterior (α_{11}, β_2).

where

$$\hat{\beta}_{23} = (X'_{-p2} M_Z X_{-p2})^{-1} X'_{-p2} M_Z X_{-p1},$$

$$\hat{\beta}_{24} = (X'_{-p2} M_{(Z\ \Delta X)} X_{-p2})^{-1} X'_{-p2} M_{(Z\ \Delta X)} X_{-p1},$$

$$\hat{\beta}_{25} = (X'_{-p2} M_{(Z\ \Delta X_1)} X_{-p2})^{-1} X'_{-p2} M_{(Z\ \Delta X_1)} X_{-p1}.$$

Proof. Use the joint posterior of α and β from Figure 2. Integrate out the parameters by using the matric-variate t conditional posteriors stated in Theorem 2.

The marginal posterior of β_2 belongs to the class of 1-1 poly t densities; see Drèze [4]. By using the theory on poly t densities or by using Raleigh quotients (see Kleibergen and van Dijk [12]), one can show that the moments of these marginal posteriors exist up to the degree defined by the difference between the orders of the exponent terms minus r. For the model with $\alpha_{22} = 0$, the moments exist up to the degree $k - r$ (($k - r$)th moment is infinite), whereas for the model where α_{22} is not restricted to 0, even the 0th moment (distribution) is infinite. When the marginal posterior of a certain parameter is integrable, the marginal posteriors of the other parameters are also integrable and vice versa because the order of integration is not important for obtaining finite integrals. The marginal posteriors of the parameters in the cointegration model with $\alpha_{22} = 0$, therefore, are all proper, whereas the marginal posteriors of the parameters in the model where α_{22} is not restricted to 0 are all nonintegrable.

To show the importance of the (non-)integrability of the marginal posteriors, we calculated the marginal posteriors of the different parameters for the ECM for the U.S. short- and long-term interest rates. In Figures 5-8, these marginal posteriors are drawn. Again, $k = 2$ and $r = 1$.

The huge difference in the tails of the marginal posteriors of β_2 with $\alpha_{22} = 0$ and α_{22} not restricted (Figure 5) indicates that the infinite means of the conditional posterior of β_2 are indeed more important for the marginal posterior of β_2 in the model with unrestricted α_{22} than in the model with restricted α_{22}. Figure 5 also shows the nonintegrability of the marginal posterior of β_2 in the model with unrestricted α_{22}. The effect of the infinite means and variances of the conditional posterior of β_2, on the marginal posterior of α for the restricted α_{22} model, is also apparent. In Figures 6 and 7, where the bivariate posterior of $(\alpha_{11}, \alpha_{12})$ is shown, the asymptote at $(\alpha_{11}, \alpha_{12}) = (0,0)$ in the bivariate posterior is clearly visible. When one compares this asymptote with the asymptote in the marginal posterior of α_{11} for the unrestricted α_{22} case (Figure 8), one can conclude that the asymptote of the marginal posterior of α_{11} lies much more within the region with non-negligible probability mass than in the case of the posterior for $(\alpha_{11}, \alpha_{12})$.

As mentioned before, the parameter β_2 is not the only parameter that can become nonidentified; the parameters of the deterministic components can

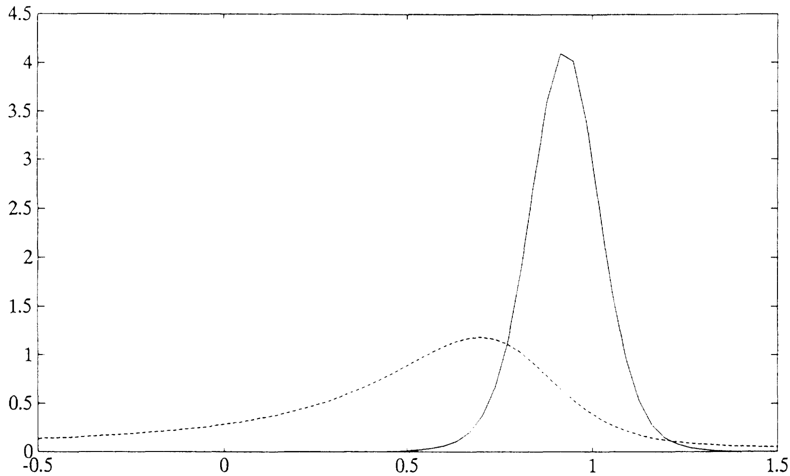

FIGURE 5. Marginal posterior β_2 with $\alpha_{22} = 0$ (–) and α_{22} not restricted (– –).

also become nonidentified. Because of the similarity between the nonidentifiedness of μ and δ, and the nonidentifiedness of β_2, the conditional and marginal posteriors for the parameters μ and δ are not constructed.

4. JEFFREYS PRIORS FOR COINTEGRATION MODELS

As shown in the previous section, diffuse priors can be highly informative in an undesired way in cointegration models because they may lead to nonintegrable posteriors. We proceed with the construction of priors, which lead to balanced posteriors. We propose the class of Jeffreys priors. The reason the Jeffreys priors overcome the problems encountered is that they are proportional to the square root of the determinant of the information matrix,

$$p(\theta) \propto |I(\theta)|^{1/2}, \tag{32}$$

where $I(\theta^*) = -\mathcal{E}((\partial^2 \ln l(\theta^*))/(\partial\theta \quad \partial\theta'))$, the information matrix evaluated in θ^*. For the nonidentified parameter values, the information matrix (= inverse covariance matrix) is singular. This is due to the infinite variance of the conditional posteriors of the nonidentified parameters in these particular parameter points. As a consequence, the Jeffreys prior penalizes nonidentified parameter points and overcomes the problems encountered with the diffuse priors, which where located exactly at the nonidentified parameter values. Another attractive property of the Jeffreys prior is that the result-

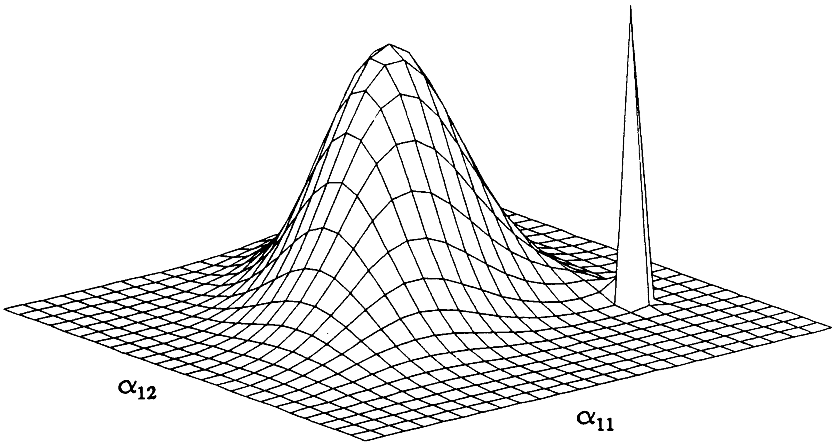

FIGURE 6. Bivariate posterior $(\alpha_{11}, \alpha_{22})$ with $\alpha_{22} = 0$.

ing posteriors are invariant with respect to the parameterization of the model. The information matrix of a parameter, say θ, is equal to a quadratic form of the Jacobian matrix of the transformation of the analyzed parameter specification toward another specification, say $\eta = \eta(\theta)$, with respect to the information matrix of the latter specification. Given $\eta = \eta(\theta)$, where $\eta(\theta)$ is differentiable, it follows that

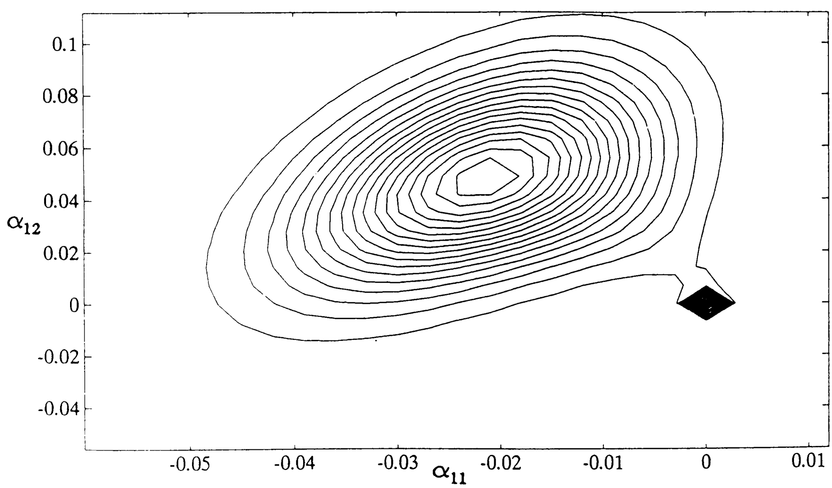

FIGURE 7. Contourlines bivariate posterior $(\alpha_{11}, \alpha_{12})$ with $\alpha_{22} = 0$.

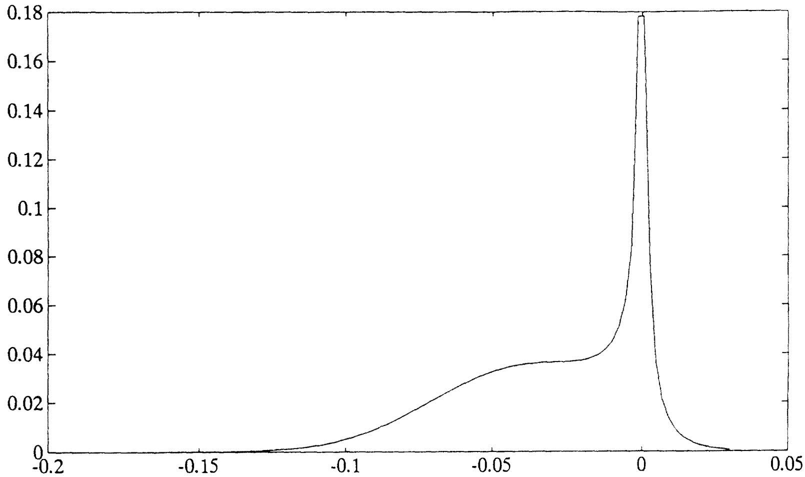

FIGURE 8. Marginal posterior α_{11}, α_{22} unrestricted.

$$
\begin{aligned}
I(\theta) &= -\mathcal{E}\left(\frac{\partial^2 \ln l(\theta)}{\partial\theta\;\partial\theta'}\right) = -\left(\frac{\partial\eta}{\partial\theta'}\right)'\mathcal{E}\left(\frac{\partial^2 \ln l(\eta(\theta))}{\partial\eta\;\partial\eta'}\right)\left(\frac{\partial\eta}{\partial\theta'}\right) \\
&= \left(\frac{\partial\eta}{\partial\theta'}\right)' I(\eta(\theta))\left(\frac{\partial\eta}{\partial\theta'}\right).
\end{aligned}
\tag{33}
$$

To construct the information matrix of a particular specification, it is convenient to construct the information matrix with respect to a specification for which the information matrix can rather straightforwardly be derived. As a next step, we construct the desired information matrix by taking the outlined quadratic form. For the analyzed cointegration models, the VAR specification (1) allows a rather straightforward construction of the information matrix,

$$
I(\theta) = \begin{bmatrix}
W'(\Omega^{-1}\otimes I_T)W & 0 & 0 \\
0 & \begin{matrix} \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1} Y_{1t}'Y_{1t}\right)\right) & \cdots & \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1} Y_{1t}'Y_{pt}\right)\right) \\ \vdots & & \vdots \\ \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1} Y_{pt}'Y_{1t}\right)\right) & \cdots & \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1} Y_{pt}'Y_{pt}\right)\right) \end{matrix} & 0 \\
0 & 0 & \frac{1}{2}TD_k'(\Omega^{-1}\otimes\Omega^{-1})D_k
\end{bmatrix},
\tag{34}
$$

where $\theta = (\mu \quad \delta \quad \text{vec}(\Pi_1)' \cdots \text{vec}(\Pi_p)' \, v(\Omega)')' : k_\theta \times 1$, $W = (I_k \otimes (\iota \quad \tau_0)) - \sum_{i=1}^{p} (\Pi_i' \otimes (\iota \quad \tau_{-i}))$, $Y_t = (Y_{1t} \cdots Y_{pt})$, $Y_{it} = (x_{t-i} - \mu - \delta(t-i))$, $\tau_i = (1+i \cdots T+i)'$, $k_\theta = 2\frac{1}{2}k + (p+\frac{1}{2})k^2$, and D_k is a duplication matrix mapping all $\frac{1}{2}k(k+1)$ different elements of the symmetric matrix, Ω, into the vector $v(\Omega)$. To calculate the information matrix of the ECM, we also need the Jacobian matrix of the transformation of the ECM to the VAR notation, i.e., the Jacobian matrix of the transformation from α, β, and Γ_i to Π_i (see Magnus and Neudecker [16]):

$$J((\Pi),(\Gamma,\beta,\alpha)) = \begin{bmatrix} I_{kk} & 0 & \cdot & \cdot & \cdot & 0 \\ -I_{kk} & I_{kk} & \cdot & & & \cdot \\ 0 & -I_{kk} & \cdot & \cdot & \cdot & \cdot \\ \vdots & & \cdot & \cdot & I_{kk} & 0 \\ 0 & \cdot & \cdot & 0 & -I_{kk} & J(\beta,\alpha) \end{bmatrix}, \tag{35}$$

where

$$J(\beta,\alpha) = \left(\left(I_k \otimes \begin{pmatrix} I_r \\ -\beta_2 \end{pmatrix} \right) ((\alpha_1' \quad J(\alpha_{22})) \otimes \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix} \right) \right), \quad \alpha_1 = (\alpha_{11} \quad \alpha_{12})$$

and $J(\alpha_{22}) = (0 \quad I_{k-r})'$ for the model with the unrestricted α_{22}, and $J(\alpha_{22})$ drops out of the Jacobian matrix when α_{22} is restricted to 0. By using the Jacobian matrix in (35) and the information matrix in (43), the Jeffreys prior can be constructed.

THEOREM 4. *The Jeffreys prior for the unrestricted α_{22} model reads*

$$p(\Gamma,\alpha,\beta,\nu,\Omega)$$
$$\propto |\Omega|^{-1/2(pk+k+1)} |W'(\Omega^{-1} \otimes I_T)W|^{1/2} |\alpha_{11}| \left| \mathcal{E}\left(\sum_{t=1}^{T} Y_t' Y_t \right) \right|^{1/2k}; \tag{36}$$

and for the model with $\alpha_{22} = 0$, the Jeffreys prior reads

$$p(\Gamma,\alpha,\beta,\nu,\Omega) \propto |\Omega|^{-1/2((p-1)k+k+1)} |\alpha\Omega^{-1}\alpha'|^{1/2(k-r)} |W'(\Omega^{-1} \otimes I_T)W|^{1/2}$$
$$\times \left| \mathcal{E}\left(\sum_{t=1}^{T} (\Delta Y_{1t} \cdots \Delta Y_{p-1t})'(\Delta Y_{1t} \cdots \Delta Y_{p-1t}) \right) \right|^{1/2(k)}$$
$$\times \left| \mathcal{E}\left(\sum_{t=1}^{T} Y_{pt}' M_{(\Delta Y_{1t} \cdots \Delta Y_{p-1t})} Y_{pt} \right) \right|^{1/2(r)}$$
$$\times \left| \begin{pmatrix} I_r \\ -\beta_2 \end{pmatrix}' \mathcal{E}\left(\sum_{t=1}^{T} Y_{pt}' M_{(\Delta Y_{1t} \cdots \Delta Y_{p-1t})} Y_{pt} \right) \begin{pmatrix} I_r \\ -\beta_2 \end{pmatrix} \right|^{1/2(k-r)} \tag{37}$$

Proof. See Appendix.

The expectation $\mathcal{E}(\sum Y_t' Y_t)$ can be constructed in several different ways. Therefore, there is not one general expression for the class of Jeffreys prior. We proceed with the construction of the Jeffreys prior for four different, a priori, plausible expressions of the expectation $\mathcal{E}(\sum Y_t' Y_t)$.

Case (i). $\mathcal{E}(\sum Y_t' Y_t)$ is constant over different parameter values. For instance, let $\mathcal{E}(\sum Y_t' Y_t)$ be equal to the realized value of $\sum Y_t' Y_t$. However, by assuming constancy of $\mathcal{E}(\sum Y_t' Y_t)$, one neglects the time series nature of Y_t.

Case (ii). $\mathrm{Vec}(\mathcal{E}(Y_0'Y_0)) = (I_{kp} - (A' \otimes A'))^{-1} \mathrm{vec}((e_1 \otimes I_k)'\Omega(e_1 \otimes I_k))$, i.e., the variance of the initial observations equals the asymptotic variance of the series.

Case (iii). $\mathcal{E}(Y_0'Y_0) = 0$. The series start at their expected value, 0. This case extends the priors used by Phillips [19] to multivariate models.

Case (iv). Construct the Jeffreys prior of the exact likelihood by incorporating the probability density functions of the initial observations.

We will analyze the four cases in more detail.

(i) $\mathcal{E}(\sum Y_t' Y_t) =$ constant (full rank) over all different parameter values. The Jeffreys prior of the ECM then becomes

$$\begin{aligned} p(\Gamma,\alpha,\beta,\nu,\delta,\Omega) \propto\ & |\Omega|^{-1/2((p-1)k+k+1)} |\alpha\Omega^{-1}\alpha'|^{1/2} \\ & \times |(\Gamma(1)\Omega^{-1/2} M_{(\Omega^{-1/2}\alpha')}\Omega^{-1/2}\Gamma(1)' \otimes \iota'\iota) \\ & \quad + (\beta\alpha\Omega^{-1}\alpha'\beta' \otimes \tau_0' M_\iota \tau_0)|^{1/2} \\ & \times |J(\beta,\alpha)'(\Omega^{-1} \otimes \Sigma) J(\beta,\alpha)|^{1/2}, \end{aligned} \tag{38}$$

where $\Sigma = \mathcal{E}(\sum Y_{pt}' Y_{pt}) =$ constant. The prior in (38) only depends on the parameters α, β, Ω, and $\Gamma(1)$. Instead of using the parameter μ, the prior in (38) is specified in terms of $\nu = \mu\beta$. The parameter $\mu\beta$ is also identified when $\alpha_{22} = 0$, which does not hold for μ separately. The assumption of linear time trends in the analyzed interest rates is rather odd. For reasons of simplicity and because of the empirical illustration with interest rates, the ECM does not contain a growth term parameter δ.

By using Theorem 4, it can be shown that the Jeffreys prior for the unrestricted α_{22} model reads

$$p(\Gamma,\alpha,\beta,\nu,\Omega) \propto |\Omega|^{-1/2(pk+k+2)} |\alpha| |\alpha_{11}|, \tag{39}$$

whereas for the model where $\alpha_{22} = 0$, the Jeffreys prior reads

$$
\begin{aligned}
&p(\Gamma,\alpha,\beta,\nu,\Omega)\\
&\quad\propto |\Omega|^{-1/2((p-1)k+k+1)}|\alpha\Omega^{-1}\alpha'|^{1/2}\\
&\quad\times\left|\begin{pmatrix} \left(\Omega^{-1}\otimes\begin{pmatrix} I_r\\ -\beta_2\end{pmatrix}'\Sigma\begin{pmatrix} I_r\\ -\beta_2\end{pmatrix}\right) & \left(\Omega^{-1}\alpha'\otimes\begin{pmatrix} I_r\\ -\beta_2\end{pmatrix}'\Sigma\begin{pmatrix} 0\\ I_{k-r}\end{pmatrix}\right)\\ \left(\alpha\Omega^{-1}\otimes\begin{pmatrix} 0\\ I_{k-r}\end{pmatrix}'\Sigma\begin{pmatrix} I_r\\ -\beta_2\end{pmatrix}\right) & \left(\alpha\Omega^{-1}\alpha'\otimes\begin{pmatrix} 0\\ I_{k-r}\end{pmatrix}'\Sigma\begin{pmatrix} 0\\ I_{k-r}\end{pmatrix}\right)\end{pmatrix}\right|^{1/2}\\
&\quad\propto |\Omega|^{-1/2(pk+r+1)}|\alpha\Omega^{-1}\alpha'|^{1/2(k-r+1)}\left|\begin{pmatrix} I_r\\ -\beta_2\end{pmatrix}'\Sigma\begin{pmatrix} I_r\\ -\beta_2\end{pmatrix}\right|^{1/2(k-r)}, \qquad (40)
\end{aligned}
$$

where $\Sigma = \sum_{t=1}^{T} \bar{Y}_{pt}'\bar{Y}_{pt}$, $\bar{Y}_{pt} = \left(x_{t-p} - \sum_{t=1}^{T} x_{t-p}/T\right)$. The Jeffreys prior of the model with unrestricted α_{22}, equation (39), still allows analytical derivation of some of the marginal and conditional posteriors. However, one has to rely on numerical techniques for the approximation of the marginal posteriors of the model with $\alpha_{22} = 0$ when using the prior (40). In Theorem 6, the marginal posterior of α_1 is stated when using the prior (39).

THEOREM 5. *When using the Jeffreys prior, equation* (39), *the marginal posterior density of* α_1 *in the unrestricted* α_{22} *model becomes*

$$p(\alpha_1\,|\,X,Z) = f_{Mt}(\alpha_1\,|\,\hat{\Pi}_1, X_{-p1}'M_{(Z\ \ X_{-p2}\ \ \iota)}X_{-p1}, \Delta X' M_{(Z\ \ X_{-p}\ \ \iota)}\Delta X, T), \qquad (41)$$

where $\hat{\Pi}_1 = (X_{-p1}'M_{(Z\ \ X_{-p2}\ \ \iota)}X_{-p1})^{-1}X_{-p1}'M_{(Z\ \ X_{-p2}\ \ \iota)}\Delta X.$

Proof. Integrate the joint posterior with respect to the other parameters.

Theorem 5 shows that in the model with unrestricted α_{22}, the Jeffreys prior leads to a proper posterior. Because the marginal posterior of α_{11} is proper, the marginal posterior of β_2 is also proper. It can be proved (see Kleibergen and van Dijk [12]) that the marginal posteriors of β_2 and α_{22} are bounded by Cauchy densities. As a consequence, the marginal posteriors of β_2 and α_{22} will have infinite means and variances.

To derive properties of the posteriors of the parameters of the model with the restricted α_{22}, we note that there is a mathematical equivalence between the ECM and the incomplete SEM. Let

$$
\begin{aligned}
V_1 &= WG + \xi_1\\
V_2 &= V_1 B + \xi_2, \qquad (42)
\end{aligned}
$$

where $V_1 = M_Z\Delta X_1$, $V_2 = M_Z\Delta X_2$, $W = M_Z(X_{-p}\ \ \iota)$, $\xi_1 = M_Z\varepsilon_1$, $\xi_2 = M_Z(\varepsilon_2 - \xi_1 B)$, $B = \alpha_{11}^{-1}\alpha_{12}$, and $G = (\alpha_{11}'\ \ -\alpha_{11}'\beta_2'\ \ -\alpha_{11}'\nu')' = (G_{11}'\ \ G_{21}'\ \ g_{31}')'$. By using bounding functions, Kleibergen and van Dijk [12] showed that the posterior moments of the parameter B exist up to the degree of overidentification plus 1 $(= k - r + 2)$. We are interested in the parameter $(\beta_2' \quad \nu')'$ $(= -(G_{21}' \quad g_{31}')'G_{11}^{-1})$, however. The parameter B is defined as $\alpha_{11}^{-1}\alpha_{12} = G_{11}^{-1}\alpha_{12}$. The inverse element, G_{11}^{-1}, is the same for both β_2 $(= -G_{21}G_{11}^{-1})$ and B. Because the Jeffreys prior leads to invariant posteriors, the posterior moments of both parameters will exist up to the same order $(k - r + 2)$. The $(k - r + 1)$th moment is therefore the finite integer posterior moment of (β_2, ν) of the highest order. The posterior moments of the parameters in which no "inversions" are involved, i.e., the parameters $(\alpha_{11}, \alpha_{12})$, will exist up to approximately the order T.

To illustrate the implications of the Jeffreys prior for the marginal posteriors, we again use the example of U.S. interest rates. In Figure 9, the marginal posteriors of α_{11}, α_{12}, and α_{22} using the Jeffreys prior are presented. All these posteriors behave regularly, and, as proved in Theorem 5, the marginal posteriors of α_{11} and α_{12} are both matric-variate t, whereas the marginal posterior of α_{22} is bounded by Cauchy densities just like the marginal posterior of β_2. The difference between the marginal posterior of α_{11} using a diffuse prior (Figure 8) and the posterior using a Jeffreys prior (Figure 9) is apparent. Figure 10 shows the marginal posteriors of β_2, both using a diffuse prior and a Jeffreys prior. Again, the difference between these posteriors is clear-cut. Figures 9–10 contain the marginal posteriors of the parameters in the model with the unrestricted α_{22}, and Figures 11 and 12

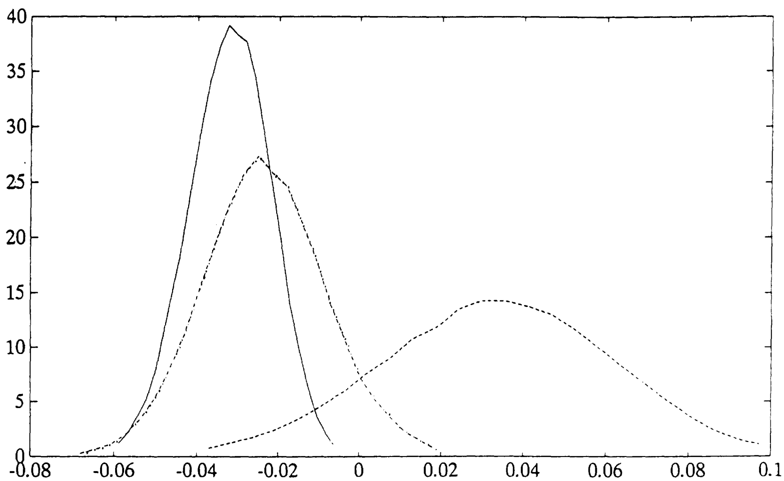

FIGURE 9. Marginal posteriors α_{11} (—), α_{12} (--), and α_{22} (-·-) for U.S. interest rates by using Jeffreys prior.

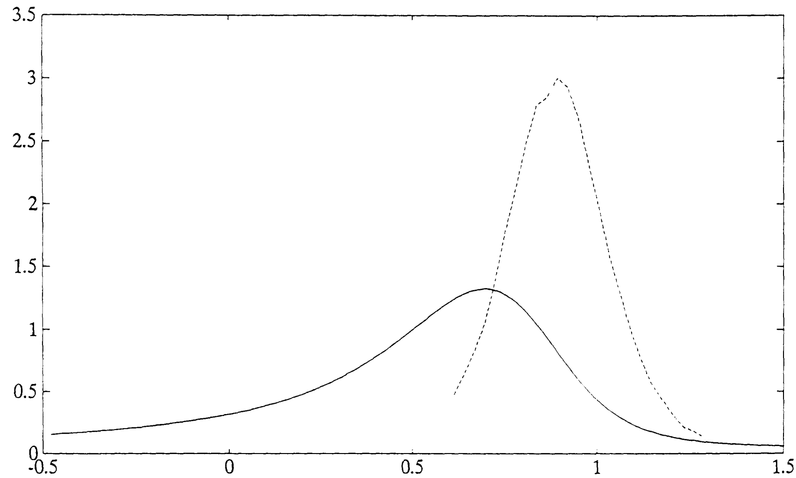

FIGURE 10. Marginal posterior β_2 with diffuse prior (–) and Jeffreys prior (--).

contain the marginal posteriors in the model with $\alpha_{22} = 0$. Figure 11 contains the marginal posteriors of α_{11} and α_{12}, and Figure 12 contains the marginal posterior of β_2, both using a diffuse prior and the Jeffreys prior. Figures 11 and 12 are drawn by using a model, which incorporates a constant

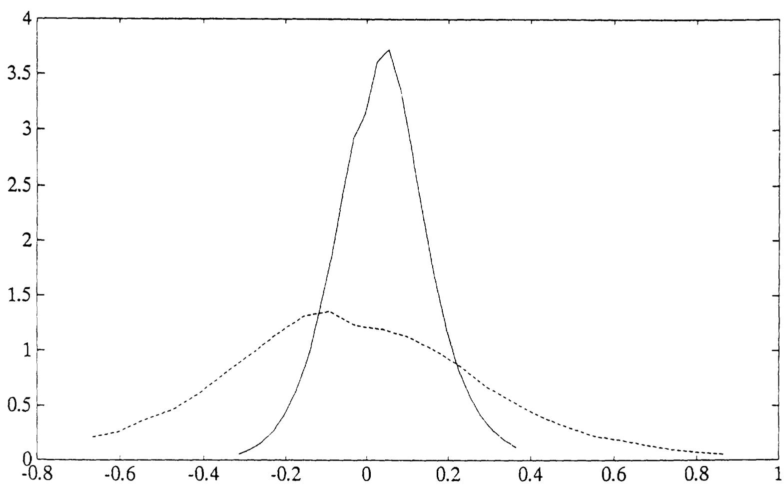

FIGURE 11. Marginal posteriors α_{11} (–) and α_{12} (--) by using Jeffreys prior with $\alpha_{22} = 0$.

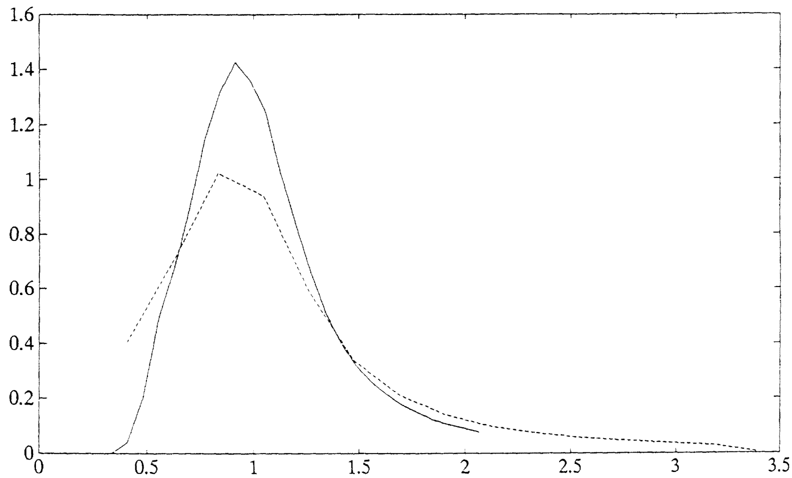

FIGURE 12. Marginal posterior β_2 by using diffuse prior (--) and Jeffreys prior (–) with $\alpha_{22} = 0$.

term in the cointegrating vector. As a consequence, the resulting posteriors cannot be compared with the posteriors drawn in the previous section.

All posteriors in this paper are calculated numerically by using Importance Sampling with a multivariate t importance function (see Kloek and van Dijk [14], van Dijk and Kloek [24], and Geweke [7]), except for the posteriors of the parameters of the unrestricted α_{22} model with constant expectation Jeffreys prior (Table 2). These posteriors are calculated by directly generating the parameters Γ_i, Π from a matric-variate t density and solving for α and β. The computer program SISAM (Hop and van Dijk [9]) was used to perform these calculations. The relative numerical error (which equals the

TABLE 2. Moments posteriors

	α_{22} Unrestricted		$\alpha_{22} = 0$		
	Mean	S.D.	Mean	S.D.	Relative numerical error
α_{11}	−0.032	0.01	0.036	0.11	0.0015
α_{12}	0.03	0.03	−0.026	0.31	0.0041
β_2	0.92	0.13	1.05	0.38	0.0021
α_{22}	−0.020	0.014	0	0	0

numerical error of the posterior mean divided by the posterior standard deviation) shows the accuracy of the calculated posterior means in case Importance Sampling has been used.

Because the posteriors in the previous figures are proper densities, it now becomes possible to perform hypothesis testing by using these posteriors. Two Bayesian testing procedures are highest posterior density (HPD) regions and posterior odds. Posterior odds testing with posteriors calculated by Importance Sampling can be done by using methods suggested by Geweke [8]. Especially for posterior odds testing, properness of the posteriors is crucial. Otherwise, one would always choose the hypothesis with an improper posterior. In the present paper, the testing procedures will be informal, and we only rely on HPD regions.

An interesting hypothesis to test concerns the presence of unit roots in the VAR polynomial. The model with the restricted α_{22} already assumes that a unit root is present. In the unrestricted α_{22} model, the parameter α_{22} measures the departure from a cointegration model. In the unrestricted α_{22} model, one can therefore test for the presence of unit roots by using the marginal posterior of α_{22}. The 95% HPD region of the marginal posterior of α_{22} for the interest-rate series (Figure 9) contains 0. So, according to usual Bayesian procedures, one cannot reject the hypothesis of one unit root in the ECM. The hypothesis of cointegration is, in several cases, not restricted to the parameter α_{22} but also deals with the deterministic components. The interpretation of deterministic components changes when unit roots are present; see Johansen [10] and Kleibergen and van Dijk [13].

The support of the hypothesis of a difference stationary ECM for the two separate models for the U.S. interest rates is quite different. For the unrestricted α_{22} model, the assumption of a difference stationary model implies that α_{11}, α_{12}, and α_{22} are all equal to 0. When analyzing the three different marginal posteriors of these parameters (Figure 9), we conclude that, although 0 lies in the 95% HPD region of the marginal posteriors of α_{12} and α_{22}, it does not lie in the 95% HPD region of the marginal posterior of α_{11}. The 95% HPD region of the trivariate posterior of $(\alpha_{11},\alpha_{12},\alpha_{22})$ also does not contain 0. As a consequence, the hypothesis of a difference stationary model is rejected when using this posterior. For the model with the restricted α_{22}, the hypothesis of a difference stationary model corresponds with $(\alpha_{11},\alpha_{12}) = (0,0)$. The 95% HPD regions of the marginal posteriors of both of these parameters contain 0 (Figure 11), and so does the 95% HPD region of the bivariate posterior of $(\alpha_{11},\alpha_{12})$. Thus, the hypothesis of a difference stationary model cannot be rejected. The different conclusions regarding the plausibility of the hypothesis of a difference stationary model are due to the structure imposed on the cointegrating vector and the deterministic components. That is, one imposed a unit root in the restricted α_{22} model and restricts the constant term to lie in the cointegration space.

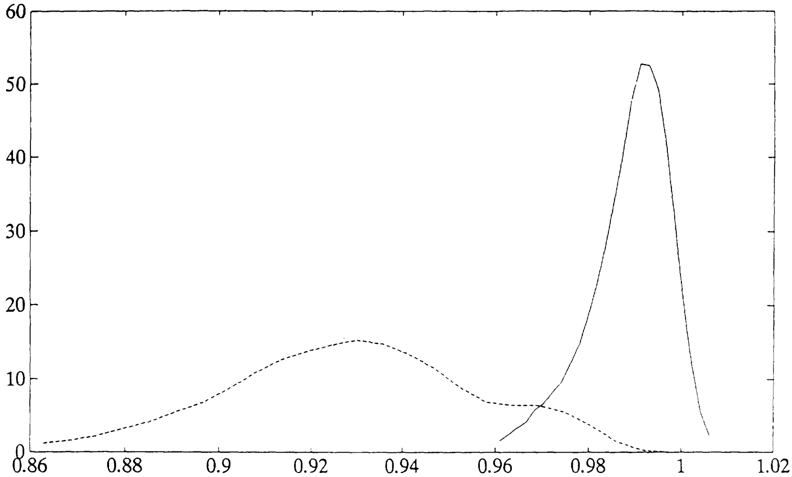

Figure 13. Marginal posterior modulus largest (–) and second largest (– –) roots for VAR model of U.S. interest rates.

For testing for cointegration one can also analyze the posteriors of the roots of the characteristic polynomial $|\Pi(z^{-1})| = 0$; see DeJong [3]. When 1 lies in the 95% HPD regions of the posteriors of the roots, the hypothesis of cointegration cannot be rejected. The problem with the roots is their possible complexity. The comparison of complex and real roots is not straightforward because of their different implications (cycle). To show the possible cointegrating nature of the U.S. interest rates from the root perspective, we calculated the posteriors of the largest roots. The problem of the complex roots is partly overcome by taking the modulus of the roots. In Figure 13 (see also Table 3), the marginal posterior of the largest and second largest roots (in modulus) are drawn. Figures 14 and 15 contain the bivariate posterior and the contourlines of the bivariate posterior of the largest and second largest root. The posteriors are calculated by generating parameters α, β, and Γ from their posteriors (unrestricted α_{22} model, Jeffreys prior) and by calculating the roots of the implied VAR polynomial. The 95% HPD region of the pos-

Table 3. Moments posteriors

	Mean	S.D.
Largest root	0.99	0.0078
Second largest root	0.93	0.025

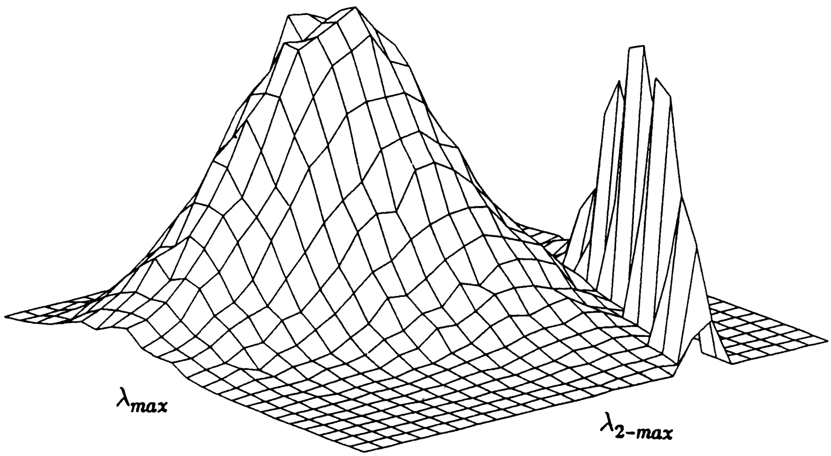

FIGURE 14. Bivariate posterior moduli, largest and second largest roots.

terior of the largest roots contains 1. So, the hypothesis of a unit root cannot be rejected. The second largest root confirms the statements made with the unrestricted α_{22} specification. Again, the hypothesis of a difference stationary model is rejected.

There is, however, also another peculiarity regarding the second largest

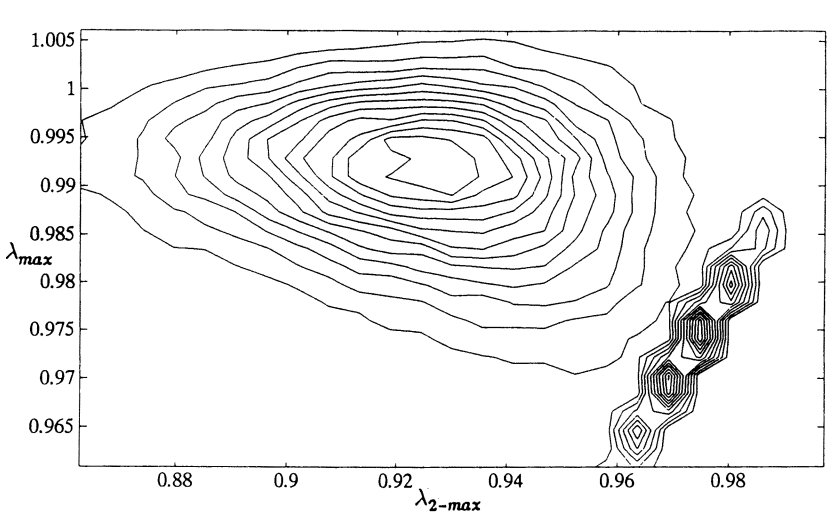

FIGURE 15. Contourlines bivariate posterior, largest and second largest roots.

root. The marginal posterior of the second largest root is almost bimodal. The bivariate posterior of the largest and second largest root (Figures 14 and 15) explain why the bimodality occurs. The second mode is the result of complex roots. These complex roots are always pairwise, which explains why the bivariate posterior consists of two "distinct posteriors." The first posterior only contains real largest roots, whereas the largest roots tend to be complex in the second posterior. As a consequence, the latter posterior exactly lies on the line $x = y$ as indicated by the contourlines. The (fat) left tail of the posterior of the largest root also almost completely consists out of complex roots. When complex, the largest root tends to be less than its real counterpart because of its complex conjugate, which leads to the same nonstationary properties. A complex root with a certain modulus as a consequence induces much more nonstationary kind of behavior than a real root with the same modulus. So the complications of the possible complex nature of the roots are not overcome straightforwardly by taking the modulus of the roots. Another problem arises when the largest roots have a double multiplicity, which happens when the series are $I(2)$. One should evaluate the eigenvectors of the implied companion matrices to check whether the eigenvalues have a higher-order multiplicity. One may conclude, therefore, that unit root testing by using the posteriors of the roots in multivariate models is not straightforward.

In the previous section, we assumed the expectation $\mathcal{E}(\sum Y_t' Y_t)$ to be constant over the different parameter values. As a consequence, the observations are fixed, and a typical feature of time series is the stochastic nature of the observations. In the classical statistical paradigm, this would correspond with the use of normal limiting distributions although these limiting distributions tend to be nonnormal for unit root time series. In the Bayesian analysis of unit roots, several theoreticians have tried to model the observations stochastically; see Phillips [19] and Schotman and van Dijk [22]. To model the observations stochastically for the VAR models, we need to construct the expectation $\mathcal{E}(\sum Y_t' Y_t)$. Theorem 4 shows that the expectation $\mathcal{E}(\sum Y_t' Y_t)$ enters the Jeffreys prior for the unrestricted α_{22} and restricted α_{22} models in different ways.

The expectation $\mathcal{E}(\sum Y_t' Y_t)$ can be constructed by using the VAR(1) specification of Y_t, $Y_t = Y_{t-1}A + v_t$, where $v_t = (\varepsilon_t' \quad 0 \cdots 0)'$ and A is the companion matrix of the VAR(p) model in (1).

$$A = \begin{bmatrix} \Pi_1 & I_k & 0 & \vdots & 0 \\ \cdot & 0 & \cdot & & 0 \\ & & \cdot & & \\ \cdot & \cdot & & \cdot & I_k \\ \Pi_p & 0 & \cdot & \cdot & 0 \end{bmatrix}. \tag{43}$$

By using the companion matrix, the expectation $\mathcal{E}(\sum Y_t' Y_t)$ becomes

$$\mathcal{E}\left(\sum_{t=1}^{T} Y_t' Y_t\right) = \sum_{t=1}^{T-1} A'^t \mathcal{E}(Y_0' Y_0) A^t + \sum_{t=0}^{T-2} (T-t-1) A'^t (e_1 \otimes I_k)' \Omega (e_1 \otimes I_k) A^t, \tag{44}$$

where e_1 is given as the first p-dimensional unity vector, $e_1 = (1 \quad 0 \cdots 0)'$. By using vec operators, the expectation can be calculated exactly:

$$\begin{aligned} \operatorname{vec}\left(\mathcal{E}\left(\sum_{t=1}^{T} Y_t' Y_t\right)\right) &= T(I_{kp} - (A' \otimes A'))^{-1} \operatorname{vec}((e_1 \otimes I_k)' \Omega (e_1 \otimes I_k)) \\ &\quad + (I_{kp} - (A' \otimes A')^T)(I_{kp} - (A' \otimes A'))^{-1} \\ &\quad \times [\operatorname{vec}(\mathcal{E}(Y_0' Y_0)) - (I_{kp} - (A' \otimes A'))^{-1} \\ &\qquad \times \operatorname{vec}((e_1 \otimes I_k)' \Omega (e_1 \otimes I_k))]. \end{aligned} \tag{45}$$

In equation (45), the only unknown term concerns $\operatorname{vec}(\mathcal{E}(Y_0'Y_0))$. In the present paper, we analyze two different expressions of $\operatorname{vec}(\mathcal{E}(Y_0'Y_0))$. The first expression, Case (ii), takes the exact expectation, $\operatorname{vec}(\mathcal{E}(Y_0'Y_0)) = (I_{kp} - (A' \otimes A'))^{-1}\operatorname{vec}((e_1 \otimes I_k)'\Omega(e_1 \otimes I_k))$, whereas the second expression, Case (iii), assumes that the series started in equilibrium, $Y_0 = 0$ such that $\mathcal{E}(Y_0'Y_0) = 0$. Another expression of the Jeffreys prior is discussed in Case (iv), where the initial observations are incorporated in the likelihood (exact likelihood) and the Jeffreys prior of the resulting likelihood is constructed.

(ii) $\operatorname{vec}(\mathcal{E}(Y_0'Y_0)) = (I_{kp} - (A' \otimes A'))^{-1}\operatorname{vec}((e_1 \otimes I_k)'\Omega(e_1 \otimes I_k))$ (see Lütkepohl [15]). The expectation $\operatorname{vec}(\mathcal{E}(\sum Y_t' Y_t))$ then becomes $T(I_{kp} - (A' \otimes A'))^{-1}\operatorname{vec}((e_1 \otimes I_k)'\Omega(e_1 \otimes I_k))$. As a consequence, $\operatorname{vec}(\mathcal{E}(\sum Y_t' Y_t))$ and the Jeffreys prior will be infinite when A has an eigenvalue equal to 1, which corresponds with a unit root in the VAR polynomial. By using this kind of Jeffreys prior, the posterior will again be improper and favor unit root models. The same reasoning holds if we replace $\mathcal{E}(Y_0'Y_0)$ by the observed value of $Y_0'Y_0$.

(iii) $\mathcal{E}(Y_0'Y_0) = 0$. The posterior means of the largest and second largest roots suggest that the model is explosive. However, the series themselves (Figure 1) and the posterior of the roots shown in Figure 13 indicate that the plausibility of an explosive model is really very small. So, not only in the univariate models investigated in Phillips [19], but also in the multivariate model does the Jeffreys prior with $\mathcal{E}(Y_0'Y_0) = 0$ lead to more explosive VAR models than the Jeffreys prior with fixed expectation (Table 4).

(iv) Another interesting posterior can be constructed by using the proba-

TABLE 4. Moments posterior

	Mean	S.D.	Relative numerical error
Largest root	1.18	0.06	0.061
Second largest root	1.08	0.22	0.072

bility density function of the initial observations. If we assume normality, the probability density function (p.d.f.) of the initial observations reads

$$p(Y_0|\theta) \propto |\mathcal{E}(Y_0'Y_0)|^{-1/2}\exp[-\tfrac{1}{2}\mathrm{vec}(Y_0)'(\mathcal{E}(Y_0'Y_0))^{-1}\mathrm{vec}(Y_0)] \tag{46}$$

When constructing the joint posterior of all parameters as proportional to the product of the exact likelihood and the Jeffreys prior of the exact likelihood, it becomes possible to partly offset the term $|\sum \mathcal{E}(Y_t'Y_t)|^{1/2(k)}$ appearing in the Jeffreys prior in (36). To construct the Jeffreys prior of the exact likelihood, we have to construct the information matrix of the exact likelihood. The information matrix of the exact likelihood equals the initial information matrix in (34) minus the expectation of the second-order derivatives of the log of the p.d.f. of the initial observations.

$$\begin{aligned} I_{EL}(\theta) &= I(\theta) - \mathcal{E}\left(\frac{\partial^2 \ln p(y_0|\theta)}{\partial\theta\ \ \delta\theta'}\right) \\ &= I(\theta) + \mathcal{E}\Bigg[(I_{k_\theta} \otimes [\mathrm{vec}(\Sigma^{-1}) - (\mathrm{vec}(Y_0)' \otimes \mathrm{vec}(Y_0)')])\frac{\partial^2 \mathrm{vec}(\Sigma)}{\partial\theta\ \ \partial\theta'} \\ &\qquad + \left(\frac{\partial\,\mathrm{vec}(\Sigma)}{\partial\theta'}\right)'(\Sigma^{-1}\otimes\Sigma^{-1})\left(\frac{\partial\,\mathrm{vec}(\Sigma)}{\partial\theta'}\right) \\ &\qquad + \left(\frac{\partial\,\mathrm{vec}(Y_0)}{\partial\theta'}\right)'\Sigma\left(\frac{\partial\,\mathrm{vec}(Y_0)}{\partial\theta'}\right)\Bigg] \\ &= I(\theta) + \left(\frac{\partial\,\mathrm{vec}(\Sigma)}{\partial\theta'}\right)'(\Sigma^{-1}\otimes\Sigma^{-1})\left(\frac{\partial\,\mathrm{vec}(\Sigma)}{\partial\theta'}\right) \\ &\quad + \left(\frac{\partial\,\mathrm{vec}(Y_0)}{\partial\theta'}\right)'\Sigma\left(\frac{\partial\,\mathrm{vec}(Y_0)}{\partial\theta'}\right), \end{aligned} \tag{47}$$

where $\Sigma = (\mathcal{E}(Y_0'Y_0))^{-1}$ and $\mathcal{E}(Y_0) = 0$. In our approach, we neglect the second term in the information matrix (47), which can be justified from an asymptotic argument. The only influence of the initial observations then arises from the last term of the information matrix (47), which is only different from 0 for the deterministic component parameters, μ and δ. For the unrestricted α_{22} specification, the Jeffreys prior then becomes

$$p(\Gamma,\alpha,\beta,\nu,\Omega) \propto |\Omega|^{-1/2(pk+k+2)}|T\alpha'\alpha + (\iota \otimes I_k)'\Sigma^{-1}(\iota \otimes I_k)|^{1/2} \times |\mathcal{E}(Y_0'Y_0)|^{1/2(k)}, \tag{48}$$

where $\iota = (1 \cdots 1)'$. To get sensible posterior results, we assume that the term $|\mathcal{E}(Y_0'Y_0)|^{1/2(k)}$ cancels out with the term $|\mathcal{E}(Y_0'Y_0)|^{-1/2}$ from the exact likelihood. As a consequence, the term $|\mathcal{E}(Y_0'Y_0)|$ does not appear in the resulting posterior. In univariate models, where $k = 1$, the posterior is proportional to the product of the exact likelihood and the Jeffreys prior of the exact likelihood, whereas for the multivariate models we have made a heuristic assumption to get plausible results, which would produce otherwise improper posteriors. By using the approximate Jeffreys prior from (48), the joint posterior then becomes

$$\begin{aligned} p(\Gamma,\alpha,\beta,\nu,\Omega \mid X,Z) &\propto |\Omega|^{-1/2(T+pk+k+2)}|T\alpha'\alpha + (\iota \otimes I_k)'\Sigma^{-1}(\iota \otimes I_k)|^{1/2}|\alpha_{11}| \\ &\times \exp[-\tfrac{1}{2}[\mathrm{vec}(Y_0)'(\mathcal{E}(Y_0'Y_0))^{-1}\mathrm{vec}(Y_0) + \mathrm{tr}\quad \Omega^{-1}\varepsilon'\varepsilon]]. \end{aligned} \tag{49}$$

Before (numerically) integrating the different parameters out of the posterior in (49) to calculate the marginal posteriors, examine the resulting posterior further. We start with the joint posterior of the AR(1) model with mean μ. For details on the model, see Schotman and van Dijk [22].

$$\begin{aligned} p(\mu,\rho,\sigma \mid X,Z) &\propto |\sigma|^{-(T+4)}|T(1-\rho)^2 + (1-\rho^2)|^{1/2} \\ &\times \exp[-\tfrac{1}{2}\sigma^{-2}[(1-\rho^2)(y_0-\mu)^2 + (y - \iota\mu(1-\rho) - \rho y_{-1})' \\ &\times (y - \iota\mu(1-\rho) - \rho y_{-1})]]. \end{aligned} \tag{50}$$

When $\rho = 1$, the model corresponds with a random walk, and the initial observations are not important. As a result, the initial observation, y_0, is deleted from the posterior. Because of the slower convergence of $(1-\rho^2)$ to 0, when ρ converges to 1, compared with $(1-\rho)^2$, μ still has the interpretation of a mean parameter when $\rho = 1$. Contrary to the model without the initial observations where $c = \mu(1-\rho)$ has the interpretation of a growth term when $\rho = 1$. This phenomenon may also be analyzed by using the mean of the conditional posterior of μ on ρ.

$$\begin{aligned} \mu(\rho) &= [\iota'(y - \rho y_{-1}) + (1+\rho)y_0]/[T(1-\rho) + (1+\rho)] \quad \text{initial present;} \\ &= \iota'(y - \rho y_{-1})/[T(1-\rho)] \quad \text{no initial.} \end{aligned} \tag{51}$$

When $\rho = 1$, $\mu(\rho) = (y_T + y_0)/2$ for the posterior with the initial observations, and $(1-\rho)\mu(\rho) = (y_T - y_0)/T$ for the posterior without the initial observations. In the first case, $\mu(\rho)$ still corresponds with the mean of the series, whereas in the second case, $(1-\rho)\mu(\rho)$ measures the average growth of the series. Although the difference is subtle, it may be important because of the crucial role the deterministic components have in unit root testing. For instance, it is now possible to test for $\rho = 1$ without the need to restrict the constant term because its interpretation for $\rho < 1$ is the same as for $\rho = 1$.

The posterior in (50) corresponds partly with the posterior used in Schotman and van Dijk [22]. The only different element in the posterior (50) concerns the term $|T(1-\rho)^2 + (1-\rho^2)|^{1/2}$, which needs to be replaced by $|1-\rho^2|^{1/2}$ to get a similar posterior as used by Schotman and van Dijk [22]. Although the difference between the two posteriors is subtle, it has quite important consequences for the marginal posterior for ρ. For the two different priors, this posterior becomes

$$p(\rho \mid X,Z) \propto f(\rho)^{-1/2(T+1)} \quad \text{Jeffreys prior exact likelihood} \tag{52}$$

$$\propto (1 + T(1-\rho)/(1+\rho))^{1/2} f(\rho)^{-1/2(T)} \quad \text{Schotman and van Dijk [22],} \tag{53}$$

where $f(\rho) = |(1-\rho^2)(y_0 - \mu(\rho))^2 + (y - \iota\mu(\rho)(1-\rho) - \rho y_{-1})'(y - \iota\mu(\rho) \times (1-\rho) - \rho y_{-1})|$ and $\mu(\rho)$ corresponds to the first expression of $\mu(\rho)$ stated in (51). Although the difference between the two marginal posteriors of ρ seems small, it can be quite important. First, the posteriors are in essence only intended for stationary processes. Second, the posterior (52) does allow explosive values for ρ, whereas the posterior (53) has an asymptote at $(T+1)/(T-1)$ (>1). That is, the posterior (53) is not defined for $\rho > 1$.

The analysis for the AR(2) models can be shown to give similar results as for the AR(1) model, i.e., the deterministic components keep the same interpretation even in the presence of unit roots.

Similar results for the exact likelihood of the VAR model with unit roots can be shown by using the structural form model in (6)–(7). The quadratic form $\text{vec}(Y_0)'(\mathcal{E}(Y_0'Y_0))^{-1}\text{vec}(Y_0)$ can be specified as

$$\begin{aligned}
&\text{vec}(Y_0)'(\mathcal{E}(Y_0'Y_0))^{-1}\,\text{vec}(Y_0) \\
&\quad = \text{vec}(Y_0)'(I_k \otimes (\beta \quad A))(\mathcal{E}[(I_k \otimes (\beta \quad A))'Y_0'Y_0(I_k \otimes (\beta \quad A))])^{-1} \\
&\qquad \times (I_k \otimes (\beta \quad A)')\text{vec}(Y_0).
\end{aligned} \tag{54}$$

Because the VAR polynomial $\Phi(z)$ is invertable, the expectation $\mathcal{E}[(I_k \otimes (\beta \quad A))'Y_0'Y_0(I_k \otimes (\beta \quad A))]$ will be finite. When unit roots are present, similar phenomena arise in equation (54) as discussed for the AR(1) model.

Although the specification (54) is theoretically interesting, it is difficult to operationalize to VAR models. The expectation will be hard to calculate because the exact expression, $\text{vec}(\mathcal{E}(Y_0'Y_0)) = (I_{kp} - (A' \otimes A'))^{-1}\text{vec}((e_1 \otimes I_k)'\Omega(e_1 \otimes I_k))$, involves the inversion of a large matrix for VAR models. In the example for the interest rates, we chose to approximate the variance of the initial observations by using the recurrence formulas,

$$\begin{aligned}
F_0 &= (e_1 \otimes I_k)'\Omega(e_1 \otimes I_k) \\
F_i &= A'F_{i-1}A + (e_1 \otimes I_k)'\Omega(e_1 \otimes I_k) \qquad i = 1, \ldots, q
\end{aligned} \tag{55}$$

The recurrence formulas in (55) converge quite fast and also have the attractive property that they allow for explosive values of the parameters, which lead to large variances of the initial observations.

With the recurrence formulas in (55) with 50 iterations, we calculated the marginal posteriors of the parameters of the unrestricted α_{22} model by using the approximate Jeffreys prior of the exact likelihood. Just as for the Jeffreys prior with $\mathcal{E}(Y_0'Y_0) = 0$, the calculated posteriors are less accurate than the other calculated posteriors. For the other posteriors, one can considerably reduce the number of parameters to be integrated out numerically, whereas for the Jeffreys prior with $\mathcal{E}(Y_0'Y_0) = 0$ and the approximate exact likelihood Jeffreys prior, all parameters have to be integrated out numerically. Also, the highly nonlinear functional form of the priors decreases the convergence of the Importance Sampling procedures.

In Figure 16, the marginal posteriors of α_{11}, α_{12}, and α_{22} are drawn by using the approximate exact likelihood Jeffreys prior. Figure 17 contains the posterior of β_2, and the posteriors of the largest and second largest roots of $|\Pi(z^{-1})| = 0$ are drawn in Figure 18. When using the 95% HPD regions of the posteriors in Figures 16–18, one cannot reject the hypothesis of a one unit root nor can one reject the hypothesis of a two unit roots difference stationary model. The marginal posteriors of α_{11} and α_{22}, as well as the marginal posteriors of the largest and second largest roots, confirm this statement. The resulting posteriors confirm the conclusions drawn from the posteriors of the parameters of the restricted α_{22} model with the fixed expectation Jeffreys prior. This indicates the fragility of the inference because different (plausible) priors give different support to the hypothesis of cointegration/differ-

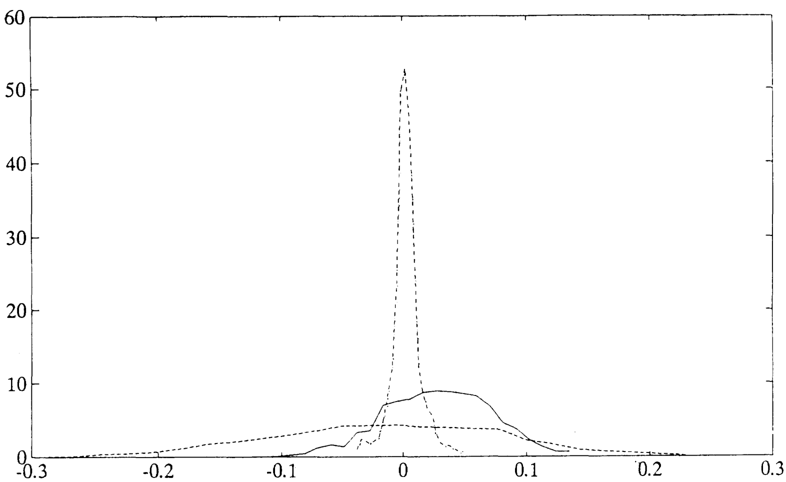

FIGURE 16. Marginal posterior, α_{11} (–), α_{12} (--), and α_{22} (-··-), for U.S. interest rates by using approximate exact likelihood Jeffreys prior.

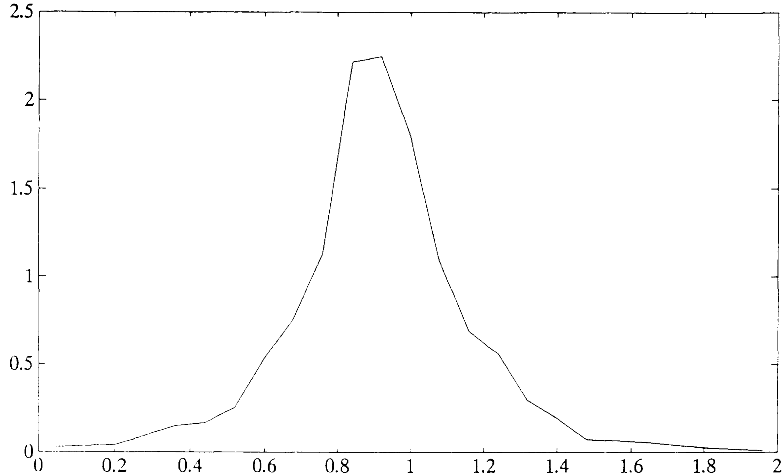

FIGURE 17. Marginal posterior β_2 for U.S. interest rates by using approximate exact likelihood Jeffreys prior.

ence stationarity. Notice, however, that again the role of the constant term is crucial. For the unrestricted α_{22} model with constant expectation Jeffreys prior, it is unrestricted such that it can represent a growth term once $\alpha_{22} = 0$. For the restricted α_{22} model and the unrestricted α_{22} model with exact likelihood Jeffreys prior, the constant term represents the mean of a certain set

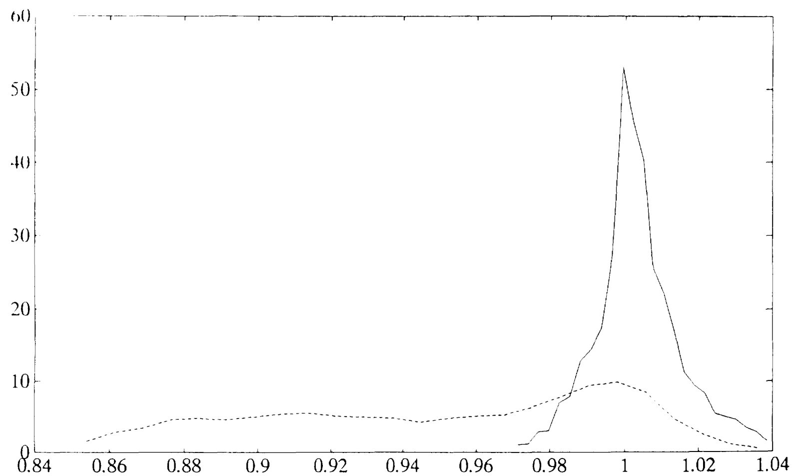

FIGURE 18. Marginal posterior, largest (—) and second largest (--) roots.

of variables. For the models where the constant terms represent means, we could not reject the hypothesis of difference stationarity (two unit roots, two stochastic trends). For the model with unrestricted α_{22} constant expectation Jeffreys prior, however, we could reject this hypothesis but still we have two trends, one stochastic and one deterministic (Table 5).

5. CONCLUSIONS

We have shown that the use of diffuse priors can be quite informative in cointegration models, in the sense that the marginal posterior densities of certain parameters of interest have an asymptote in the interior of the parameter region. Thus, the posteriors may be nonintegrable. This result extends to all models in which parameters can become nonidentified. By using the Jeffreys prior principle, we show that one may overcome this identification problem. The Jeffreys prior is not unique, however, because the expectation of the endogenous variables can be evaluated in several different ways. It is shown that for four different expressions of this expectation, only three imply Jeffreys priors, which lead to proper posteriors. From these three priors, only two priors give plausible posterior outcomes for the analyzed U.S. interest-rate series. The two plausible priors are the Jeffreys prior, which leaves the expectation constant over different parameter values, and the other prior is the so-called approximate exact likelihood Jeffreys prior. The exact likelihood prior is especially interesting because it treats the initial observations in a stochastic manner such that only in the unit root case the model is conditional on the initial observations.

The Bayesian cointegration analysis conducted in this paper analyzes some of the possible problems one encounters in this approach. A different approach of analyzing cointegration models by using Bayesian statistical analysis is conducted by Phillips [20] where a restricted version of the model in (4) and (8)–(9) is used. The approach by Phillips [20] differs in the sense that the posteriors are connected with the theory of stochastic martingale

Table 5. Moments posteriors

	Mean	S.D.	Relative numerical error
α_{11}	0.031	0.046	0.027
α_{12}	−0.027	0.11	0.028
α_{22}	0.002	0.037	0.015
β_2	0.93	0.26	0.025
Largest root	1.005	0.013	0.023
Second largest root	0.95	0.05	0.025

processes, which allows one to construct the posteriors recursively. Also, the model used is linear such that no identification problem occurs, but, as a consequence, certain Granger longrun noncausality relationships are implicitly assumed to hold; see also Kleibergen and van Dijk [13].

REFERENCES

1. Box, G.E.P. & G.C. Tiao. A canonical analysis of multiple time series. *Biometrika* 64 (1977): 355-365.
2. Campbell, J. & R.J. Shiller. Cointegration and tests of present value models. *Journal of Political Economy* 95 (1987): 1062-1088.
3. DeJong, D.N. Co-integration and trend-stationarity in macro-economic time series: Evidence from the likelihood function. *Journal of Econometrics*, forthcoming.
4. Drèze, J.H. Bayesian regression using poly t densities. *Journal of Econometrics* 6 (1977): 329-354.
5. Drèze, J.H. & J.F. Richard. Bayesian analysis of simultaneous equations systems. In Z. Griliches & M.D. Intrilligator (eds.), *Handbook of Econometrics, Vol. 1.* Amsterdam: North-Holland, 1983.
6. Engle, R.F. & C.W.J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55 (1987): 251-276.
7. Geweke, J. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57 (1989): 1317-1339.
8. Geweke, J. Exact predictive densities for linear models with ARCH disturbances. *Journal of Econometrics* 40 (1989): 63-86.
9. Hop, J.P. & H.K. van Dijk. SISAM and MIXIN: Two algorithms for the computation of posterior moments and densities using Monte-Carlo integration. *Computer Science in Economics and Management* 5 (1992): 183-220.
10. Johansen, S. Estimation and hypothesis testing of cointegrating vectors in gaussian vector autoregressive models. *Econometrica* 59 (1991): 1551-1581.
11. Johansen, J. & K. Juselius. Maximum likelihood estimation and inference on cointegration—With applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52 (1990): 169-210.
12. Kleibergen, F. & H.K. van Dijk. *Bayesian Simultaneous Equation Model Analysis: On the Existence of Structural Posterior Moments.* Econometric Institute report 9269/A, Erasmus University, Rotterdam, 1992.
13. Kleibergen, F. & H.K. van Dijk. Direct cointegration testing in error correction models. *Journal of Econometrics*, forthcoming.
14. Kloek, T. & H.K. van Dijk. Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* 46 (1978): 1-20.
15. Lütkepohl, H. *Introduction to Multiple Time Series Analysis.* Berlin: Springer-Verlag, 1991.
16. Magnus, J. & H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Economics.* Chichester: Wiley, 1988.
17. Phillips, P.C.B. Partially identified econometric models. *Econometric Theory* 5 (1989): 181-240.
18. Phillips, P.C.B. Optimal inference in cointegrated systems. *Econometrica* 59 (1991): 283-307.
19. Phillips, P.C.B. To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6 (1991): 333-364.
20. Phillips, P.C.B. Bayes Methods for Trending Multiple Time Series with an Empirical Application to the U.S. Economy. Cowles Foundation discussion paper #1025, 1992.
21. Phillips, P.C.B. Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models. *Econometrica* 62 (1994): 73-93.

22. Schotman, P. & H.K. van Dijk. A Bayesian analysis of the unit root hypothesis. *Journal of Econometrics* 49 (1991): 195-238.
23. Stock, J.H. & M.W. Watson. Testing for common trends. *Journal of the American Statistical Association* 83 (1988): 1097-1107.
24. van Dijk, H.K. & T. Kloek. Further experience in Bayesian analysis using Monte-Carlo integration. *Journal of Econometrics* 14 (1980): 307-328.
25. Zellner, A. *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley, 1971.

APPENDIX

The $k \times m$ random matrix T is said to have a matric-variate t (Mt) distribution with λ degrees of freedom if its probability density function is

$$\begin{aligned} p(T) &= f_{Mt}(T \mid \bar{T}, \Sigma, P, \lambda) \\ &= c^{-1} |\Sigma|^{1/2(\lambda)} |P|^{1/2(m)} |\Sigma + (T - \bar{T})' P (T - \bar{T})|^{-1/2(\lambda+k)} \\ &= c^{-1} |\Sigma|^{-1/2(k)} |P|^{-1/2(\lambda+k-m)} |P^{-1} + (T - \bar{T}) \Sigma^{-1} (T - \bar{T})'|^{-1/2(\lambda+k)}, \end{aligned}$$

$$c = \pi^{1/2(km)} \prod_{i=1}^{m} \left[\Gamma\left(\frac{1}{2}(\lambda + 1 - i)\right) \Big/ \Gamma\left(\frac{1}{2}(\lambda + k + 1 - i)\right) \right]$$

where T is a $k \times m$ constant matrix, Σ is $m \times m$ PDS constant matrix, P is a $k \times k$ PDS constant matrix, and $\lambda > m - 1$. The moments of this probability density function exist up to the order $\lambda - m + 1$; see, e.g., Zellner [25], Appendix B5.

Proof of Theorem 1.

$$\begin{aligned} &p(\beta, \alpha \mid X, Z) \\ &\quad \propto |(\Delta X - X_{-p}\beta\alpha)' M_Z (\Delta X - X_{-p}\beta\alpha)|^{-1/2(T+h-pk-1)} \\ &\quad \propto |\Delta X' M_{(Z\ X_{-p}\beta)} \Delta X + (\alpha - \hat{\alpha})' \beta' X'_{-p} M_Z X_{-p} \beta (\alpha - \hat{\alpha})|^{-1/2(T+h-pk-1)} \\ &\quad \propto |\Delta X' M_{(Z\ X_{-p})} \Delta X + (\beta\alpha - \hat{\Pi})' X'_{-p} M_Z X_{-p} (\beta\alpha - \hat{\Pi})|^{-1/2(T+h-pk-1)} \\ &\quad \propto |(X'_{-p} M_Z X_{-p})^{-1} + (X'_{-p} M_Z X_{-p})^{-1} X'_{-p} M_Z \Delta X [(\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \\ &\qquad - (\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha' (\alpha (\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha')^{-1} \\ &\qquad \times \alpha (\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1}] \Delta X' M_Z X_{-p} (X'_{-p} M_Z X_{-p})^{-1} \\ &\qquad + (\beta - \hat{\beta}) \alpha (\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha' (\beta - \hat{\beta})'|^{-1/2(T+h-pk-1)} \\ &\quad \propto |(X'_{-p} M_Z X_{-p})^{-1} + (X'_{-p} M_Z X_{-p})^{-1} X'_{-p} M_Z \Delta X \alpha'_\perp (\alpha_\perp \Delta X' M_{(Z\ X_{-p})} \Delta X \alpha'_\perp)^{-1} \\ &\qquad \times \alpha_\perp \Delta X' M_Z X_{-p} (X'_{-p} M_Z X_{-p})^{-1} \\ &\qquad + (\beta - \hat{\beta}) \alpha (\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha' (\beta - \hat{\beta})'|^{-1/2(T+h-pk-1)} \\ &\quad \propto |(X'_{-p} M_{(Z\ \Delta X \alpha'_\perp)} X_{-p})^{-1} \\ &\qquad + (\beta - \hat{\beta}) \alpha (\Delta X' M_{(Z\ X_{-p})} \Delta X)^{-1} \alpha' (\beta - \hat{\beta})'|^{-1/2(T+h-pk-1)} \end{aligned}$$

$$\propto [|(\alpha(\Delta X' M_{(Z\ X_{-p})}\Delta X)^{-1}\alpha')^{-1} + (\beta-\hat{\beta})' X'_{-p} M_{(Z\ \Delta X\alpha'_\perp)} X_{-p}(\beta-\hat{\beta})| \times |\alpha(\Delta X' M_{(Z\ X_{-p})}\Delta X)^{-1}\alpha'||X'_{-p} M_{(Z\ \Delta X\alpha'_\perp)} X_{-p}|^{-1}]^{-1/2(T+h-pk-1)}.$$

$$\hat{\Pi} = (X'_{-p} M_Z X_{-p})^{-1} X'_{-p} M_Z \Delta X$$

Proof of Theorem 4.

$$\eta = (\mu\delta\ \mathrm{vec}(\Gamma_1)' \cdots \mathrm{vec}(\Gamma_{p-1})'\ \mathrm{vec}(\alpha_{11})'\ \mathrm{vec}(\alpha_{12})'\ \mathrm{vec}(\beta_2)'\ \mathrm{vec}(\alpha_{22})' v(\Omega)')'$$

$$I(\eta) = \left(\frac{\partial\eta}{\partial\theta'}\right)' I(\eta(\theta)) \left(\frac{\partial\eta}{\partial\theta'}\right)$$

$$= \begin{pmatrix} I_{2k} & 0 & 0 \\ 0 & \begin{pmatrix} I_{kk} & 0 & \cdots & & 0 \\ -I_{kk} & I_{kk} & & & \vdots \\ 0 & -I_{kk} & \ddots & & \vdots \\ \vdots & & \ddots & I_{kk} & 0 \\ 0 & \cdots & 0 & -I_{kk} & J(\beta,\alpha) \end{pmatrix} & 0 \\ 0 & 0 & I_{1/2k(k+1)} \end{pmatrix}$$

$$\begin{pmatrix} W'(\Omega^{-1}\otimes I_T)W & 0 & 0 \\ 0 & \begin{matrix} \left(\Omega^{-1}\otimes \mathcal{E}\left(\sum_{t=0}^{T-1} Y'_{1t}Y_{1t}\right)\right) & \cdots & \left(\Omega^{-1}\otimes \mathcal{E}\left(\sum_{t=0}^{T-1} Y'_{1t}Y_{pt}\right)\right) \\ \vdots & & \vdots \\ \left(\Omega^{-1}\otimes \mathcal{E}\left(\sum_{t=0}^{T-1} Y'_{pt}Y_{1t}\right)\right) & \cdots & \left(\Omega^{-1}\otimes \mathcal{E}\left(\sum_{t=0}^{T-1} Y'_{pt}Y_{pt}\right)\right) \end{matrix} & 0 \\ 0 & 0 & \frac{1}{2}TD'_k(\Omega^{-1}\otimes\Omega^{-1})D_k \end{pmatrix}$$

$$\begin{pmatrix} I_{2k} & 0 & 0 \\ 0 & \begin{pmatrix} I_{kk} & 0 & \cdots & & 0 \\ -I_{kk} & I_{kk} & & & \vdots \\ 0 & -I_{kk} & \ddots & & \vdots \\ \vdots & & \ddots & I_{kk} & 0 \\ 0 & \cdots & 0 & -I_{kk} & J(\beta,\alpha) \end{pmatrix} & 0 \\ 0 & 0 & I_{1/2k(k+1)} \end{pmatrix}.$$

For the unrestricted α_{22} model, the Jacobian matrix is square such that

$$
\begin{aligned}
|I(\eta)| &= \left|\left(\frac{\partial\eta}{\partial\theta'}\right)\right|^2 |I(\eta(\theta))| \\
&= |J(\beta,\alpha)|^2 |W'(\Omega^{-1}\otimes I_T)W| \left|\Omega^{-1}\otimes \mathcal{E}\left(\sum_{t=0}^{T-1} Y_t' Y_t\right)\right| |\Omega|^{(k+1)} \\
&= |\alpha_{11}|^2 |W'(\Omega^{-1}\otimes I_T)W| \left|\mathcal{E}\left(\sum_{t=0}^{T-1} Y_t' Y_t\right)\right|^k |\Omega|^{(pk+k+1)},
\end{aligned}
$$

whereas for the restricted α_{22} model,

$$
|I(\eta)| = \left|\left(\frac{\partial\eta}{\partial\theta'}\right)' I(\eta(\theta))\left(\frac{\partial\eta}{\partial\theta'}\right)\right|
$$

$$
= \left|\left\{\begin{matrix}
\left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1}\Delta Y_{1t}'\Delta Y_{1t}\right)\right) & \cdots & \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1}\Delta Y_{1t}'\Delta Y_{p-1t}\right)\right) & \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1}\Delta Y_{1t}' Y_{pt}\right)J(\beta,\alpha)\right. \\
\vdots & & \vdots & \vdots \\
\left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1}\Delta Y_{p-1t}'\Delta Y_{1t}\right)\right) & \cdots & \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1}\Delta Y_{p-1t}'\Delta Y_{p-1t}\right)\right) & \left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1}\Delta Y_{p-1t}' Y_{pt}\right)J(\beta,\alpha)\right. \\
J(\beta,\alpha)'\left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1} Y_{pt}'\Delta Y_{1t}\right)\right) & \cdots & J(\beta,\alpha)'\left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1} Y_{pt}'\Delta Y_{p-1t}\right)\right) & J(\beta,\alpha)'\left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=0}^{T-1} Y_{pt}' Y_{pt}\right)J(\beta,\alpha)\right.
\end{matrix}\right]\right| |W'(\Omega^{-1}\otimes I_T)W||\Omega|^{(k+1)}
$$

$$
\begin{aligned}
&= \left|\mathcal{E}\left(\sum_{t=0}^{T-1}(\Delta Y_{1t}\cdots\Delta Y_{p-1t})'(\Delta Y_{1t}\cdots\Delta Y_{p-1t})\right)\right|^k |W'(\Omega^{-1}\otimes I_T)W| \\
&\quad\times |\Omega|^{((p-1)k+k+1)} \left|J(\beta,\alpha)'\left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=1}^{T} Y_{pt}' M_{(\Delta Y_{1t}\cdots\Delta Y_{p-1t})} Y_{pt}\right)\right)J(\beta,\alpha)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\left|J(\beta,\alpha)'\left(\Omega^{-1}\otimes\mathcal{E}\left(\sum_{t=1}^{T} Y_{pt}' M_{(\Delta Y_{1t}\cdots\Delta Y_{p-1t})} Y_{pt}\right)\right)J(\beta,\alpha)\right| \\
&\quad= |\alpha\Omega^{-1}\alpha'|^{(k-r)} \left|\begin{pmatrix} I_r \\ -\beta_2\end{pmatrix}'\mathcal{E}\left(\sum_{t=1}^{T} Y_{pt}' M_{(\Delta Y_{1t}\cdots\Delta Y_{p-1t})} Y_{pt}\right)\begin{pmatrix} I_r \\ -\beta_2\end{pmatrix}\right|^{(k-r)} \\
&\quad\times \left|\mathcal{E}\left(\sum_{t=1}^{T} Y_{pt}' M_{(\Delta Y_{1t}\cdots\Delta Y_{p-1t})} Y_{pt}\right)\right|^r.
\end{aligned}
$$