Macroeconomic Dynamics, **5**, 2001, 577–597. Printed in the United States of America.
DOI: 10.1017.S1365100500000468

VECTOR AUTOREGRESSIVE PROCESSES WITH NONLINEAR TIME TRENDS IN COINTEGRATING RELATIONS

ANTTI RIPATTI
Bank of Finland

PENTTI SAIKKONEN
University of Helsinki

We extend the conventional cointegrated VAR model to allow for general nonlinear deterministic trends. These nonlinear trends can be used to model gradual structural changes in the intercept term of the cointegrating relations. A general asymptotic theory of estimation and statistical inference is reviewed and a diagnostic test for the correct specification of an employed nonlinear trend is developed. The methods are applied to Finnish interest-rate data. A smooth level shift of the logistic form between the own-yield of broad money and the short-term money market rate is found appropriate for these data. The level shift is motivated by the deregulation of issuing certificates of deposit and its inclusion in the model solves the puzzle of the "missing cointegration vector" found in a previous study.

Keywords: Cointegrated VAR Model, Gradual Structural Change, Nonlinear Deterministic Trend

1. INTRODUCTION

A fairly common finding in empirical analysis of cointegrated time series is that the number of cointegrating relations supported by the data turns out to be smaller than expected on the basis of economic theory. The low power of cointegration tests is often blamed for this but the reason may also be that the data have been affected by structural changes not taken into account in the employed (standard) model. Changes in institutions, policy, and technology are typical examples. If the changes occur as sudden structural breaks at known points of time, they can be modeled by conventional dummy variables. However, as recently discussed by Leybourne et al. (1998) in the context of unit root tests, it may sometimes

The views expressed are those of the authors and not necessarily the views of the Bank of Finland. We thank the participants of the Conference on Nonlinear Modeling of Multivariate Macroeconomic Relations, September 1999, Rotterdam, and two anonymous referees for useful comments and discussions. The usual disclaimer applies.
Address correspondence to: Antti Ripatti, Bank of Finland, P.O. Box 160, FIN-00101 Helsinki, Finland; e-mail: antti.ripatti@bof.fi.

 1365-1005/01 $9.50

be more reasonable to consider structural changes that are gradual and smooth [see also Lin and Teräsvirta (1994) and the references therein]. An example is a structural change due to adopting a new technology, which usually takes time to be completely realized and for which no clear date of change can be pointed out. On the other hand, the realized change may be gradual even if a particular date could be related to it. For instance, regulations can change at a particular date but economic agents may start adjusting their behavior in advance if they know of the forthcoming mandatory change.

Choosing an incorrect number of cointegrating vectors is a rather extreme misspecification that can result from ignoring structural changes. Even if this misspecification is avoided, serious distortions can occur in other inference procedures on cointegrating vectors. In this paper, we therefore consider an extension of a standard cointegrated vector autoregressive (VAR) model in which smooth or continuous deterministic changes are allowed in the constant term of the cointegrating relations. In a typical example, also relevant for the empirical part of the paper, the constant term changes smoothly from one level to another. In the same way as in Lin and Teräsvirta (1994) and Leybourne et al. (1998), we use a logistic function of time to model this. However, in our general model, almost any conceivable smooth function of time can be used to allow for additive structural changes in the cointegrating relations and, although not emphasized in this paper, sudden structural breaks are also possible provided the date(s) of break(s) is (are) a priori known and not estimated from current data.

Recently, Saikkonen (2001a,b) developed a general asymptotic theory of estimation and statistical inference applicable to the model considered in this paper. Since that work is purely theoretical, our purpose here is to discuss related empirical aspects and provide a motivation for the model considered. A further contribution of this paper is that we develop diagnostic tests that can be used to check whether a deterministically changing constant term is really needed in the cointegrating relations of a standard VAR model and, more generally, whether a diagnosed structural change can be adequately described by a chosen function of time. These tests, which are based on the Lagrange multiplier (LM) principle, are similar to the tests developed by Lin and Teräsvirta (1994) and Eitrheim and Teräsvirta (1996) for stationary models. On the other hand, since they are also similar to the variable addition test that Park (1990) proposed for testing the null of cointegration, they have power when the chosen cointegrating rank is too large. This and the lack of a statistical test for cointegration in the presence of a continuous structural change means that our tests are mainly designed for cases in which strong prior information about the cointegrating rank is available.

In the empirical part of the paper, we analyze a data set of four Finnish interest rates. The same data were previously used by Luukkonen et al. (1999), who found two cointegrating vectors instead of the expected three. Here, we show that this unexpected finding can be attributed to a known structural change which appears as a smooth level shift in one of the three cointegrating relations.

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

The rest of the paper is organized as follows. The general model and some particular cases are discussed in Section 2. The main points of Gaussian maximum likelihood (ML) estimation and asymptotic inference are summarized in Section 3. The diagnostic tests of the paper are developed in Section 4 and the empirical example is presented in Section 5. Section 6 concludes.

2. MODEL

Let $\boldsymbol{y}_t$, $t = 1, \ldots, T$, be an s-dimensional time series generated by a VAR process of order p. Using the error correction form of the process, the series is modeled as

$$\Delta \boldsymbol{y}_t = \boldsymbol{d}_t + \Pi \boldsymbol{y}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta \boldsymbol{y}_{t-j} + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T, \tag{1}$$

where Δ is the usual difference operator, Π $(s \times s)$ and Γ_j $(s \times s)$ are unknown parameters, and $\boldsymbol{d}_t$ $(s \times 1)$ is a deterministic sequence to be discussed below. Furthermore, the initial values $\boldsymbol{y}_{-p+1}, \ldots, \boldsymbol{y}_0$ are observable and $\boldsymbol{\epsilon}_t$ is Gaussian white noise, that is, $\boldsymbol{\epsilon}_t \sim \mathrm{NID}(\mathbf{0}, \Omega)$ with Ω positive definite. This last assumption is only made to facilitate the discussion of likelihood-based methods and could be relaxed. It is also assumed that the matrix Π is of rank r $(0 < r < s)$ so that we can write

$$\Pi = \alpha \beta', \tag{2}$$

where α and β are $s \times r$ matrices of full column rank. Since we are interested in time series whose stochastic components are integrated of order 1 and cointegrated, we also assume that the parameters of the model satisfy the conditions of Johansen's (1995, p. 49) version of Granger's representation theorem. Thus, it is assumed that the roots of the characteristic equation

$$\det\left[\left(I_n - \sum_{j=1}^{p-1} \Gamma_j z^j\right)(1-z) - \alpha\beta' z\right] = 0$$

are equal to 1 or lie outside the unit circle and the matrix

$$\alpha_\perp' \left(I_n - \sum_{j=1}^{p-1} \Gamma_j\right) \beta_\perp$$

has full rank $s - r$. Here, $\beta_\perp$, for example, denotes an $s \times (s - r)$ matrix of full column rank and such that $\beta' \beta_\perp = 0$. The above assumptions imply that, with a suitable specification of initial values, both $\Delta \boldsymbol{y}_t$ and $\beta' \boldsymbol{y}_t$ are stationary around deterministic trends [see Johansen (1995, p. 49)].

In particular cases of the above model, which can now be called standard, the deterministic sequence $\boldsymbol{d}_t$ may contain an intercept term, seasonal dummies, and perhaps also a linear time trend. In this paper, we are mainly interested in nonlinear deterministic trends. The idea is to allow for the possibility that a standard model may fail because the mean of the error correction term $\beta'\boldsymbol{y}_t$ changes in a nonlinear fashion. For simplicity, we consider only the case in which the deterministic term $\boldsymbol{d}_t$ can be absorbed into the cointegrating relations. This means that the sequence $\boldsymbol{d}_t$ is of the form

$$\boldsymbol{d}_t = -\alpha \boldsymbol{g}_t(\boldsymbol{\mu}), \tag{3}$$

where $\boldsymbol{g}_t(\boldsymbol{\mu})$ is a generally nonlinear deterministic function of the time index t and a parameter vector $\boldsymbol{\mu}$. Thus, inserting (2) and (3) into (1) yields

$$\Delta \boldsymbol{y}_t = \alpha(\beta'\boldsymbol{y}_{t-1} - \boldsymbol{g}_t(\boldsymbol{\mu})) + \sum_{j=1}^{p-1} \Gamma_j \Delta \boldsymbol{y}_{t-j} + \boldsymbol{\epsilon}_j, \quad t = 1, \ldots, T. \tag{4}$$

A convenient way to specify the sequence $\boldsymbol{g}_t(\boldsymbol{\mu})$ is to follow the approach used in nonparametric regression and assume that $\boldsymbol{g}_t(\boldsymbol{\mu}) = \boldsymbol{g}(t/T; \boldsymbol{\mu})$ where $\boldsymbol{g}(\cdot; \boldsymbol{\mu})$ is a suitable function defined on the interval [0, 1]. Two examples that illustrate how the function $\boldsymbol{g}(\cdot; \boldsymbol{\mu})$ may be specified are given by

$$\boldsymbol{g}(x; \boldsymbol{\mu}) = \boldsymbol{\nu} + \{1 + \exp[-\gamma(x - \tau)]\}^{-1}\boldsymbol{\delta} \tag{5}$$

and

$$\boldsymbol{g}(x; \boldsymbol{\mu}) = \boldsymbol{\nu} + [1 - \exp\{-\gamma(x - \tau)^2\}]\boldsymbol{\delta}. \tag{6}$$

Here $\boldsymbol{\mu} = [\boldsymbol{\nu}'\boldsymbol{\delta}'\gamma\tau]'$ with $\boldsymbol{\nu}$ and $\boldsymbol{\delta}$ unknown $r \times 1$ parameter vectors while γ and τ are scalar parameters with $\gamma > 0$ and $0 < \tau < 1$. These functions have been used to model a smooth or continuous structural change in the coefficients of a dynamic regression model. A recent paper in this area is by Lin and Teräsvirta (1994), which also contains references to earlier work. In (5) the smooth change is modeled by a logistic function, whereas in (6) the density function of the normal distribution is essentially used. To see the idea of (5), suppose that the components of $\boldsymbol{\delta}$ are positive. Then the parameter vector $\boldsymbol{\nu}$ is the constant term of the cointegrating relations that applies at the beginning of the sample whereas $\boldsymbol{\nu} + \boldsymbol{\delta}$ is a new constant term that applies after a smooth increasing change. The parameter τ determines the average location of the change and γ is a slope parameter that indicates how rapid the change is. The smaller the value of the parameter γ is the longer it takes for the constant term to reach its new level. When the value of γ is "large," we are close to the case where a single structural break occurs. In fact, this case is obtained in the limit by letting $\gamma \to \infty$ because, then, (5) approaches $\boldsymbol{\nu} + \mathrm{I}(x \geq \tau)\boldsymbol{\delta}$, where

$I(\cdot)$ is an indicator function taking the value 1 if the indicated condition is true and 0 elsewhere. When the value of the parameter τ is known, this amounts to using a conventional step dummy. In the case of (6), the change in the constant term is nonmonotonic and symmetric about τ. This specification can be used, for instance, to describe the situation in which the mean of the cointegration relations first decreases smoothly from $\boldsymbol{\nu}$ to $\boldsymbol{\delta}$ and then increases smoothly back to $\boldsymbol{\nu}$. Again, γ is a slope parameter that determines how rapid these changes are.

Obvious extensions of the above specifications are obtained by defining the function $\boldsymbol{g}(x;\boldsymbol{\mu})$ as a linear combination of functions of the type given on the right-hand sides of (5) and (6). Another extension, proposed by Lin and Teräsvirta (1994), is given by

$$\boldsymbol{g}(x;\boldsymbol{\mu}) = \boldsymbol{\nu} + \left\{1 + \exp\left[-\gamma\left(x^k + \tau_1 x^{k-1} + \cdots + \tau_{k-1}x + \tau_k\right)\right]\right\}^{-1}\boldsymbol{\delta}, \qquad \textbf{(7)}$$

where $\tau_1, \ldots, \tau_k$ $(k > 1)$ are scalar parameters and the rest of the notation is as in (5). When $k = 1$, (7) reduces to (5). In addition to this choice, Lin and Teräsvirta (1994) consider the value $k = 3$. With these extensions, one can, for instance, allow for more than a single smooth transition in the constant term.

As the above discussion shows, our approach includes a variety of interesting possibilities to model structural changes in cointegrating relations. In some applications, continuous changes may be more natural than sudden breaks, which have recently received considerable attention. Whether the change is continuous or not, the above examples make clear that in these cases it is quite natural to proceed as in nonparametric regression and use the scaled time index t/T in the trend model instead of t. A similar scaling also has been used by other authors to model trends or varying parameters. For instance, Phillips and Hansen (1990) use a similarly scaled time index in cointegrated systems, although only in models that are linear in parameters. In cointegrating regressions, the same idea is also used by Park and Hahn (1999), who develop nonparametric methods to analyze deterministic changes in cointegrating vectors. Andrews and McDermott (1995), who scale the time index in a slightly different way, consider general nonlinear deterministic trends in a parametric framework but do not allow integrated processes. Dahlhaus (1996a,b) makes extensive use of this approach in his locally stationary models and provides an insightful discussion of its motivation.

The above discussion also implies that we are mainly thinking of deterministic trends that are not "large" like those implied by the linear specification $\boldsymbol{d}_t = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 t$. Assuming a linear trend in our formulation means that $\boldsymbol{d}_t = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2(t/T)$, which is "small" in the sense that it remains bounded as the sample size T tends to infinity. Note however, that in the case of a linear trend, the inference procedures to be discussed in the next section are valid even if the time index is not scaled by the sample size. The scaling is only required in nonlinear functions of time. On the other hand, the test procedures to be developed in

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

Section 3 are affected by the presence of a linear trend in the cointegrating relations so that, for simplicity, this case will be excluded there.

Since $\boldsymbol{g}_t(\boldsymbol{\mu}) = \boldsymbol{g}(t/T; \boldsymbol{\mu})$ is assumed, the process $\boldsymbol{y}_t$ also depends on the sample size, so that, strictly speaking, a notation like $\boldsymbol{y}_{tT}$ should be used. Fortunately, however, this dependence is very simple and therefore does not cause any theoretical complications. Indeed, a version of Granger's representation theorem given by Saikkonen (2001a) shows that the dependence of $\boldsymbol{y}_t$ on T is only due to the deterministic sequence $\boldsymbol{g}(t/T; \boldsymbol{\mu})$. The stochastic part of $\boldsymbol{y}_t$ is independent of T and identical to its counterpart in Johansen's (1995, p. 49) version of Granger's representation theorem. The assumptions imposed on the function $\boldsymbol{g}(x; \boldsymbol{\mu})$ also imply that the processes $\Delta \boldsymbol{y}_t$ and $\beta \boldsymbol{y}_{t-1} - \boldsymbol{g}_t(\boldsymbol{\mu})$, which are stationary in standard cases, can be considered here as nearly stationary or asymptotically stationary [see Saikkonen (2001a)].

3. ML ESTIMATION AND STATISTICAL INFERENCE

The parameters in (4) with a chosen specification of the sequence $\boldsymbol{g}_t(\boldsymbol{\mu})$ or, equivalently, the function $\boldsymbol{g}(x; \boldsymbol{\mu})$ can be estimated by ML. Since the relevant asymptotic estimation theory has been developed by Saikkonen (2001a,b), we only briefly summarize the main points here. First, note that Saikkonen (2001a,b) also proves results when some nuisance parameters are not identified but, unless otherwise stated, identifiability is assumed here. Thus, it is assumed that the cointegrating vectors can be written as $\beta' = [I_r - A(\boldsymbol{\phi})]$, where $A(\boldsymbol{\phi})$ is a continuously differentiable function of the underlying identifiable parameter vector $\boldsymbol{\phi}$. The function $\boldsymbol{g}(x; \boldsymbol{\mu})$ is assumed to be continuously differentiable with respect to the latter argument. A number of technical regularity conditions are also imposed on the function $\boldsymbol{g}(x; \boldsymbol{\mu})$. These conditions are satisfied by the examples given in (5) and (6) if the values of the parameters γ and τ are restricted as $0 < c_1 \leq \gamma \geq c_2 < \infty$ and $0 < c_3 \leq \tau \leq c_4 < 1$. Many other choices of the function $\boldsymbol{g}(x; \boldsymbol{\mu})$ are also allowed. A general sufficient condition is that $\boldsymbol{g}(x; \boldsymbol{\mu})$ is continuously differentiable as a function of $(x, \boldsymbol{\mu})$. Since the function $\boldsymbol{g}(x; \boldsymbol{\mu})$ may be discontinuous with respect to its first argument, conventional dummy variables are also included. The most important case that is excluded is that of structural breaks with unknown brake dates or dummy variables with dates of jump depending on unknown parameters. Finally, note that the results of Saikkonen (2001a,b) also allow for the possibility that the short-run parameters α and Γ are smooth functions of an underlying structural parameter vector, but we shall not discuss that extension here.

Denote $\boldsymbol{\theta} = [\boldsymbol{\theta}_1' \boldsymbol{\theta}_2']'$ where $\boldsymbol{\theta}_1 = [\boldsymbol{\phi}' \boldsymbol{\mu}']'$ and $\boldsymbol{\theta}_2 = \text{vec}[\alpha\Gamma]$. Here vec signifies the usual columnwise vectorization operator. Thus, conditioning on the initial values $\boldsymbol{y}_{-p+1}, \ldots, \boldsymbol{y}_0$, we can write the log-likelihood function of the data as

$$L_T(\boldsymbol{\theta}, \Omega) = -\frac{T}{2}\log\det(\Omega) - \frac{1}{2}\text{tr}\left[\Omega^{-1}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t(\boldsymbol{\theta})\boldsymbol{\epsilon}_t(\boldsymbol{\theta})'\right], \tag{8}$$

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

where

$$\epsilon_t(\boldsymbol{\theta}) = \Delta \boldsymbol{y}_t - \alpha[\boldsymbol{y}_{1,t-1} - A(\phi)\boldsymbol{y}_{2,t-1} - \boldsymbol{g}_t(\boldsymbol{\mu})] - \Gamma \boldsymbol{q}_t,$$

with $\Gamma = [\Gamma_1 \ldots \Gamma_{p-1}]$, $\boldsymbol{q}_t = [\Delta \boldsymbol{y}'_{t-1} \ldots \Delta \boldsymbol{y}'_{t-p+1}]'$ and $\boldsymbol{y}_t = [\boldsymbol{y}'_{1t}\,\boldsymbol{y}'_{2t}]'$ partitioned in an obvious way. ML estimators of $\boldsymbol{\theta}$ and Ω, denoted by $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}'_1 \hat{\boldsymbol{\theta}}'_2]$ and $\hat{\Omega}$, are obtained by maximizing the function $L_T(\boldsymbol{\theta}, \Omega)$. This maximization problem is, of course, highly nonlinear. Saikkonen (2001a) shows that, under suitable regularity conditions, the ML estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\Omega}$ exist with probability approaching 1 and are consistent. The limiting distribution of $\hat{\boldsymbol{\theta}}$ is derived by Saikkonen (2001b). The estimators $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are asymptotically independent of each other and of the estimator $\hat{\Omega}$. The limiting distribution of $\hat{\boldsymbol{\theta}}_1$ is mixed normal and that of $\hat{\boldsymbol{\theta}}_2$ is normal. To be able to describe these results more precisely, we define

$$G_{1t}(\boldsymbol{\theta}) = -\begin{bmatrix} \left[\dfrac{\partial \mathrm{vec} A(\phi)'}{\partial \boldsymbol{\phi}}\right](y_{2,t-1} \otimes \alpha') \\ \left[\dfrac{\partial \boldsymbol{g}_t(\boldsymbol{\mu})'}{\partial \boldsymbol{\mu}}\right] \otimes \alpha' \end{bmatrix}$$

and

$$G_{2t}(\boldsymbol{\theta}_1) = \boldsymbol{z}_t(\boldsymbol{\theta}_1) \otimes I_s,$$

where the symbol $\otimes$ signifies Kronecker's product and $\boldsymbol{z}_t(\boldsymbol{\theta}_1) = [\boldsymbol{q}'_t \boldsymbol{u}_{t-1}(\boldsymbol{\theta}_1)']'$ with $\boldsymbol{u}_{t-1}(\boldsymbol{\theta}_1) = \boldsymbol{y}_{1,t-1} - A(\boldsymbol{\phi})\boldsymbol{y}_{2,t-1} - \boldsymbol{g}_t(\boldsymbol{\mu})$. Note that here, for example, $\partial \boldsymbol{g}_t(\boldsymbol{\mu})'/\partial \boldsymbol{\mu} = (\partial \boldsymbol{g}_t(\boldsymbol{\mu})/\partial \boldsymbol{\mu}')'$ as in Lütkepohl (1996, p. 173). From Saikkonen (2001b) we can now conclude that, under regularity conditions,

$$\hat{M}^{1/2}_{1\cdot 2}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) \xrightarrow{d} N(\mathbf{0}, I), \tag{9}$$

where

$$\hat{M}_{1\cdot 2} = \sum_{t=1}^{T} \hat{G}_{1t}\hat{\Omega}^{-1}\hat{G}'_{1t} - \sum_{t=1}^{T} \hat{G}_{1t}\hat{\Omega}^{-1}\hat{G}'_{2t}\left(\sum_{t=1}^{T} \hat{G}_{2t}\hat{\Omega}^{-1}\hat{G}'_{2t}\right)^{-1} \sum_{t=1}^{T} \hat{G}_{2t}\hat{\Omega}^{-1}\hat{G}'_{1t}$$

with $\hat{G}_{1t} = G_{1t}(\hat{\boldsymbol{\theta}})$ and $\hat{G}_{2t} = G_{2t}(\hat{\boldsymbol{\theta}}_1)$. For $\hat{\boldsymbol{\theta}}_2$ we have

$$\hat{M}^{1/2}_{2\cdot 1}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) \xrightarrow{d} N(0, I), \tag{10}$$

where $\hat{M}_{2\cdot 1}$ is defined in the same way as $\hat{M}_{1\cdot 2}$ except that the roles of the subscripts 1 and 2 are interchanged. Note that (9) would also hold if $\hat{M}_{1\cdot 2}$ were replaced by the first matrix in its defining equation and similarly for (10). $G_{1t}(\boldsymbol{\theta})$ contains integrated processes and $G_{2t}(\boldsymbol{\theta}_1)$ contains asymptotically stationary processes;

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

this fact, together with (9) and (10), explains why the limiting distribution of $\hat{\boldsymbol{\theta}}_1$ is mixed normal and that of $\hat{\boldsymbol{\theta}}_2$ is normal.

Approximate standard errors can be obtained for the components of $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ by taking square roots of the diagonal elements of the matrices $\hat{M}_{1\cdot 2}$ and $\hat{M}_{2\cdot 1}$, respectively. More generally, it is shown by Saikkonen (2001b) that Wald tests with asymptotic chi-squared distributions under the null hypothesis can be constructed in the usual way to test (possibly nonlinear) hypotheses on the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$; similar results also are obtained for corresponding likelihood ratio (LR) and LM tests. In these Wald tests the matrices $\hat{M}_{1\cdot 2}$ and $\hat{M}_{2\cdot 1}$ provide natural estimators of the needed information matrices; their constrained counterparts can be used in corresponding LM tests. Note, however, that these tests assume that the parameters of the model are identified. This particularly means that, if (5) or (6) is specified, testing the hypothesis $\boldsymbol{\delta} = 0$ is not possible because it implies that the parameters γ and τ are not identified. Test procedures for this kind of hypothesis are considered in the next section.

4. TEST PROCEDURES

An interesting question in the model introduced in Section 2 is whether any nonlinear time trend is actually needed or whether a specified nonlinear time trend, such as (5), (6), or (7), can really adequately describe a diagnosed structural change. Tests based on the LM principle appear convenient because they only require estimating parameters of the null model, which in the present context is a considerably simpler task than estimating parameters of the unrestricted model.

For ease of exposition, we first discuss how to obtain a test for the null hypothesis that states that the sequence $\boldsymbol{g}_t(\boldsymbol{\mu})$ in (4) reduces to a constant. It is obvious that parametric tests cannot be obtained without suitable assumptions of the sequence $\boldsymbol{g}_t(\boldsymbol{\mu})$ or the function $\boldsymbol{g}(x;\boldsymbol{\mu})$. Our assumptions are fairly general and apply in a number of cases, including those discused in Section 2. Partition the parameter vector $\boldsymbol{\mu}$ into three subvectors as $\boldsymbol{\mu} = [\boldsymbol{\nu}'\boldsymbol{\lambda}'\gamma']'$ and assume that for some known value of γ, denoted by γ_0, we have $\boldsymbol{g}(x;\boldsymbol{\nu},\boldsymbol{\lambda},\gamma_0) = \boldsymbol{\nu}$ for all $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$. Thus, under the null hypothesis $\gamma = \gamma_0$, we have a standard model with a constant term in the cointegrating relations. This formulation applies to the examples in (5) and (6), although in the former the notation has to be redefined. In both cases the parameter γ is scalar with $\gamma_0 = 0$ while $\boldsymbol{\lambda} = [\boldsymbol{\delta}'\tau]'$.

In the preceding setup the previously mentioned identification problem appears in the parameter $\boldsymbol{\lambda}$, which is not identified under the null hypothesis. There has recently been a great interest in this problem and significant progress has been made by Andrews (1993), Andrews and Ploberger (1994), and Hansen (1996). However, instead of following the approach of these authors, we proceed as Lin and Teräsvirta (1994) did and obtain a simple LM-type test by replacing the function $\boldsymbol{g}(x;\boldsymbol{\mu})$ with a Taylor approximation. In other words, we assume that the function $\boldsymbol{g}(x;\boldsymbol{\mu})$ is continuously differentiable with respect to $\boldsymbol{\mu}$ (or at least γ) and consider the approximation

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

$$\boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma) \approx \boldsymbol{\nu} + \frac{\partial \boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma_0)}{\partial \gamma'}(\gamma - \gamma_0). \tag{11}$$

Since the parameter $\boldsymbol{\lambda}$ appears on the right-hand side, we still have an identification problem, which cannot be solved unless suitable assumptions are made for the partial derivatives therein. We assume that the vector $\partial \boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma_0)/\partial \gamma$ can be written as a polynomial of x so that, for some integers $1 \leq n_1 < \cdots < n_q$,

$$\frac{\partial \boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma_0)}{\partial \gamma} = \underset{\sim}{\psi}_0 + \underset{\sim}{\psi}_1 x^{n_1} + \cdots + \underset{\sim}{\psi}_q x^{n_q}, \tag{12}$$

where the parameters $\underset{\sim}{\psi}_0, \ldots, \underset{\sim}{\psi}_q$ $(r \times 1)$ are functions of $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$. This assumption is satisfied in (5) and (6). In (5), $q = 1$, $n_1 = 1$, and $\partial \boldsymbol{g}(x; \boldsymbol{\mu})/\partial \gamma$ evaluated at $\gamma = 0$ is $0.25(x - \tau)\boldsymbol{\delta}$, which is of the form in equation (12). In (6), we have $q = 2$, $n_1 = 1$, and $n_2 = 2$. Now $\partial \boldsymbol{g}(x; \boldsymbol{\mu})/\partial \gamma$ evaluated at $\gamma = 0$ is $\boldsymbol{\delta}(x - \tau)^2$, which is a special case of (12). The same conclusion clearly applies to (7) and extensions of (5) and (6), where the function $\boldsymbol{g}(x; \boldsymbol{\mu})$ is a linear combination of functions of types (5) and (6).

Now, using (11) and (12) leads us to replace the original model (4) with the approximation

$$\Delta \boldsymbol{y}_t = \alpha(\beta' \boldsymbol{y}_{t-1} - \boldsymbol{\nu} - \Psi \boldsymbol{w}_t) + \Gamma \boldsymbol{q}_t + \boldsymbol{e}_t, \qquad t = 1, \ldots, T, \tag{13}$$

where $\Psi = [\psi_1 \ldots \psi_q]$, $\boldsymbol{w}_t = [(t/T)^{n_1} \ldots (t/T)^{n_q}]'$, and $\boldsymbol{e}_t$ is an error term that equals the true error term $\boldsymbol{\epsilon}_t$ when the linearity hypothesis holds or when $\Psi = 0$. Note that, here, ψ_i stands for $\underset{\sim}{\psi}_i(\gamma - \gamma_0)$ $(i = 1, \ldots, q)$ and, for convenience, $\underset{\sim}{\psi}_0 = 0$ has been assumed so that $\boldsymbol{\nu} + \underset{\sim}{\psi}_0(\gamma - \gamma_0) = \boldsymbol{\nu}$ holds. It is a simple matter to test the null hypothesis $\Psi = 0$ in (13). One possibility is to apply the reduced rank regression method directly to (13) and test the null hypothesis $\Psi = 0$ by using the LR test based on the assumption that the error term $\boldsymbol{e}_t$ is Gaussian white noise [see Johansen (1995)]. An alternative possibility is to use the corresponding LM test, which requires only the application of reduced rank regression to the null model or to (13) with the constraint $\Psi = 0$. This is the test considered in this paper. However, we do not describe it explicitly here because it can be obtained as a special case of the more general test to be developed shortly.

Before generalizing the preceding test, we note that a test obtained for the null hypothesis $\Psi = 0$ in (13) obviously can be seen as a version of the variable addition test that Park (1990) proposed for testing the null of cointegration. This means that a rejection of the null hypothesis $\Psi = 0$ may also be an indication that the cointegrating rank has been erroneously specified too large. To look at this from another angle, suppose economic theory has suggested the cointegrating rank $r = r_0$ and one starts with the standard model, that is, (13) with $\Psi = 0$. Suppose further that the data clearly support the choice $r = r_0 - 1$ against $r = r_0$ (e.g., the LR test of the null hypothesis $r = r_0 - 2$ is clearly rejected but the null hypothesis

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

$r = r_0 - 1$ cannot be rejected at any reasonable significance level). This might also happen when $r = r_0$ is the correct specification but the data are generated by (4) with $\boldsymbol{g}_t(\boldsymbol{\mu})$ given by (5) or (6). In this situation, testing the null hypothesis $\Psi = 0$ in (13) with $r = r_0$ is also likely to lead to a rejection. Thus, if there is no prior information about the cointegrating rank, it is difficult to say whether one should adopt (4) with $r = r_0 - 1$ and $\boldsymbol{g}_t(\boldsymbol{\mu}) = \boldsymbol{\nu}$ or, alternatively, $r = r_0$ and $\boldsymbol{g}_t(\boldsymbol{\mu})$ modeled by a nonlinear function of time. To be able to discriminate between these two cases, a statistical test for cointegration in the general model (4) would be needed but, to the best of our knowledge, there is no such test available at present. This problem can be avoided, however, when the specification of the cointegrating rank is based on economic theory, which, we believe, is fairly typical in practice. Another point worth emphasizing here is that if a standard model is found inadequate and a nonlinear trend is added to the cointegrating relations, a meaningful interpretation of the specified nonlinear trend and reasons leading to the failure of the standard model would be reasonable to have. Otherwise, including a nonlinear time trend in the cointegrating relations seems rather pointless.

Now consider the case where (4) with a chosen functional form of $\boldsymbol{g}_t(\boldsymbol{\mu})$ has been adopted and the question is whether an additional or alternative nonlinear term should be used in the model. In this context, we proceed in the same way as in Eitrheim and Teräsvirta (1996) and assume that the correct specification of the nonlinear trend can be written as

$$\boldsymbol{g}_t^*(\boldsymbol{\mu}, \boldsymbol{\upsilon}) = \boldsymbol{g}_t(\boldsymbol{\mu}) + \boldsymbol{h}_t(\boldsymbol{\upsilon}), \qquad \textbf{(14)}$$

where $\boldsymbol{h}_t(\boldsymbol{\upsilon})$ depends on a parameter vector $\boldsymbol{\upsilon}$ and $h_t(\boldsymbol{\upsilon}) = 0$ under the null hypothesis. Moreover, $\boldsymbol{h}_t(\boldsymbol{\upsilon})$ is supposed to be such that $\boldsymbol{g}_t^*(\boldsymbol{\mu}, \boldsymbol{\upsilon})$ satisfies the regularity conditions required from $\boldsymbol{g}_t(\boldsymbol{\mu})$ in order for the theoretical results of Section 2 to apply. In particular, we assume that $\boldsymbol{h}_t(\boldsymbol{\upsilon}) = \boldsymbol{h}(t/T; \boldsymbol{\upsilon})$ for a suitable function $\boldsymbol{h}(x; \boldsymbol{\upsilon})$ similar to $\boldsymbol{g}(x; \boldsymbol{\mu})$. The sequence $\boldsymbol{g}_t(\boldsymbol{\mu})$ is supposed to contain an additive constant term, although this would not be necessary. Now, the obvious idea is to linearize the function $\boldsymbol{h}(x; \boldsymbol{\upsilon})$ in the same way as was done for the function $g(x; \boldsymbol{\mu})$ in (11) and assume that the related vector of partial derivatives can be written as a polynomial similar to (12). Thus, instead of (13), we consider its generalization,

$$\Delta \boldsymbol{y}_t = \alpha[\beta' \boldsymbol{y}_{t-1} - \boldsymbol{g}_t(\boldsymbol{\mu}) - \Psi \boldsymbol{w}_t] + \Gamma \boldsymbol{q}_t + \boldsymbol{e}_t, \qquad t = 1, \ldots, T, \qquad \textbf{(15)}$$

where the error term $\boldsymbol{e}_t$ equals the true error term $\boldsymbol{\epsilon}_t$ when the null hypothesis $h_t(\boldsymbol{\upsilon}) = 0$ or, equivalently, $\Psi = 0$ holds. We derive an LM test for this latter form of the null hypothesis by assuming that $\boldsymbol{e}_t \sim N(\boldsymbol{0}, \Omega)$ holds in (15). The test requires a related constrained ML estimation, which can be carried out by specializing the discussion of Section 2 to the present context. Thus, we again assume that $\beta' = [I_r - A(\phi)]$ and that $\boldsymbol{g}_t(\boldsymbol{\mu})$ satisfies the regularity conditions discussed in Section 2. In what follows, we use a hat symbol to indicate constrained ML estimators obtained from (15) with the constraint $\Psi = 0$. We need the score of vec Ψ

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

based on (15) with the assumption $\boldsymbol{e}_t \sim \mathrm{NID}(\mathbf{0}, \Omega)$. It is straightforward to see that this score evaluated at constrained ML estimators is $\sum_{t=1}^{T} \hat{G}_{3t}\hat{\Omega}^{-1}\hat{\epsilon}_t$, where $\hat{G}_{3t} = \boldsymbol{w}_t \otimes \hat{\alpha}'$ and $\hat{\epsilon}_t$ is the obvious constrained residual. Specializing the general LM test of Saikkonen (2001b) to the present context leads to the test statistic

$$S = \left(\sum_{t=1}^{T} \hat{G}_{3t}\hat{\Omega}^{-1}\hat{\epsilon}_t\right)' \hat{M}_{3\cdot 12}^{-1} \left(\sum_{t=1}^{T} \hat{G}_{3t}\hat{\Omega}^{-1}\hat{\epsilon}_t\right), \tag{16}$$

where $\hat{M}_{3.12}$ is defined in the same way as $\hat{M}_{1.2}$ in (12) except that $\hat{G}_{1t}$ is replaced with $\hat{G}_{3t}$ and $\hat{G}_{2t}$ is replaced with $[\hat{G}_{1t}'\hat{G}_{2t}']'$. The limiting null distribution of test statistic S can be derived from the general results of Saikkonen (2001b) by observing that under the null hypothesis we really have $\boldsymbol{e}_t \sim \mathrm{NID}(\mathbf{0}, \Omega)$. Thus, we can conclude that, under the null hypothesis and appropriate regularity conditions, test statistic S has a standard chi-squared limiting distribution; that is,

$$S \xrightarrow{d} \chi^2_{qr}. \tag{17}$$

Of course, large values of the test statistic are critical for the null hypothesis. We do not give formal results about the properties of our test under the alternative hypothesis but we do note that its power should be reasonable when the functions $\boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma)$ and $\boldsymbol{h}(x; \boldsymbol{\upsilon})$ are not orthogonal to the polynomial defined by the components of $\boldsymbol{w}_t$. This is clearly the case in the examples discussed in Section 2. Note, however, that from a significant value of test statistic S, it is not possible to deduce the exact form of the function $\boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma)$ or $\boldsymbol{h}(x; \boldsymbol{\upsilon})$ that one should entertain. This is, of course, clear because choosing $\boldsymbol{w}_t = t/T$ was found reasonable in the case of (5) but this choice is also used when the need for a linear trend is tested. The situation remains the same even if a linear trend can be ruled out a priori. For instance, if (6) is suspected, it is reasonable to choose $\boldsymbol{w}_t = [(t/T)(t/T)^2]'$ so that the resulting test should also have power against (5). Graphical methods may be useful when the specification of the function $\boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma)$ or $\boldsymbol{h}(x; \boldsymbol{\upsilon})$ is considered. This point is exemplified in the next section.

The preceding discussion also makes clear that our test with $\boldsymbol{w}_t = t/T$ breaks down if the null model is augmented by a linear trend. This is simply because then the additional regressor $\boldsymbol{w}_t = t/T$ used in the test already appears in the model. Of course, this difficulty could be circumvented by using higher powers $(t/T)^2$, $(t/T)^3, \ldots,$ but we will not pursue this matter explicitly in this paper [cf. Lin and Teräsvirta (1994)].

The expression of test statistic S reveals that it can be computed as a LM test statistic for the need of the regressor $\hat{G}_{3t}'$ in the auxiliary regression of $\hat{\epsilon}_t$ on $\hat{G}_{1t}'$, $\hat{G}_{2t}'$ and $\hat{G}_{3t}'$ with the error term treated as normal with covariance matrix $\hat{\Omega}$. Although the derivation of test statistic S may appear somewhat ad hoc, it can in some cases, such as those given by (5) and (6), be motivated by the LM principle. To demonstrate this, consider (6) in the case in which the null hypothesis implies that $\boldsymbol{g}_t(\boldsymbol{\mu}) = \boldsymbol{\nu}$. This null hypothesis can be formulated by restricting $\gamma = 0$ in (6). The score of γ evaluated at $\gamma = 0$ is $\boldsymbol{\delta}'\alpha'\Omega^{-1}\sum_{t=1}^{T}[(t/T) - \tau]^2\epsilon_t$. If the values of the parameters $\boldsymbol{\delta}$ and τ were known, then a proper LM test could be based on this score. Then we could use test statistic S with $\hat{G}_{3t} = [(t/T) - \tau]^2\boldsymbol{\delta}'\hat{\alpha}'$. Of course, this test statistic is infeasible because it depends on the values of the nuisance parameters $\boldsymbol{\delta}$ and τ, which are unknown in practice. A standard approach in a case such as this is to take the supremum of the test statistic over the possible values of the unknown nuisance parameters [see Davies (1977, 1987), Andrews (1993), and Andrews and Ploberger (1994)]. However, using the above auxiliary regression interpretation of test statistic S and well-known properties of least-squares theory, it can be seen that our feasible version of test statistic S is obtained precisely in this way. Thus, in this special case, our test can be motivated as a "supLM test." The same result is also obtained in the case of (5) and therefore we may call our test statistic an LM-type or a score-type test statistic.

An alternative motivation for test statistic S is obtained by observing that its derivation is based on the idea of replacing the function $\boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma)$ or $\boldsymbol{h}(x; \boldsymbol{v})$ with a polynomial approximation. This makes the time trend linear in parameters and thereby solves the involved identification problem. This interpretation is to some extent nonparametric. Indeed, such an approximation can be made arbitrarily accurate by taking the degree of the approximating polynomial large enough. However, when one has prior information about the possible form of the function $\boldsymbol{g}(x; \boldsymbol{\nu}, \boldsymbol{\lambda}, \gamma)$ or $\boldsymbol{h}(x; \boldsymbol{v})$, it is worthwhile to make use of this information and choose the polynomial approximation accordingly. For instance, in (5) and (6), first- and second-order polynomials seem quite reasonable and parsimonious choices, even if they could not be motivated by the LM principle.

We close this section by discussing test statistic S in the special case in which the function $\boldsymbol{g}(x; \boldsymbol{\mu})$ is of the form

$$\boldsymbol{g}(x; \boldsymbol{\mu}) = \boldsymbol{\nu} + \varphi \boldsymbol{f}(x, \boldsymbol{\mu}_2). \tag{18}$$

Here the parameters $\boldsymbol{\nu}$ and φ are unrestricted and φ may be a matrix. Thus, $\boldsymbol{\mu} = [\boldsymbol{\mu}_1' \; \boldsymbol{\mu}_2']'$ with $\boldsymbol{\mu}_1 = [\boldsymbol{\nu}' \; (\operatorname{vec}\varphi)']'$. We also assume that no overidentifying restrictions are imposed on cointegrating vectors, so that $\beta' = [I_r - A(\boldsymbol{\phi})] = [I_r - A]$ with $\phi = \operatorname{vec} A$. The cointegrating vectors are thus identified by normalizing restrictions. A convenient feature of test statistic S in this case is that it is invariant to a particular normalization. To see this, suppose that $\beta' = [I_r - A]$ is transformed as $\beta' \to \xi\beta'$ and that the parameters $\boldsymbol{\nu}$ and φ are transformed similarly as $\boldsymbol{\nu} \to \xi\boldsymbol{\nu}$ and $\varphi \to \xi\varphi$. Here, ξ is any nonsingular $r \times r$ matrix and, if the parameter matrix α is transformed as $\alpha \to \alpha\xi^{-1}$, we obtain a reparameterization of the original model. The ML residuals $\hat{\epsilon}_t$ are clearly invariant to these transformations. From the definitions, it can further be seen that these transformations do not change the value of $\hat{G}_{3t}$ in test statistic S and the same is also true for $\hat{G}_{2t}$, where the transformation amounts to multiplying $\boldsymbol{u}_{t-1}(\hat{\boldsymbol{\theta}}_1)$ by ξ. Finally, consider the

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

sequence $\hat{G}_{1t}$, which, in the present context, becomes

$$\hat{G}_{1t} = -\begin{bmatrix} \boldsymbol{y}_{2,t-1} \otimes \hat{\alpha}' \\ \hat{\alpha}' \\ \boldsymbol{f}_t(\hat{\boldsymbol{\mu}}_2) \otimes \hat{\alpha}' \\ \left[\dfrac{\partial\, \mathrm{vec}\, \boldsymbol{f}_t(\hat{\boldsymbol{\mu}}_2)'}{\partial \boldsymbol{\mu}_2}\right] \otimes \hat{\varphi}'\hat{\alpha}' \end{bmatrix}.$$

From this and the definition of test statistic S, it can readily be seen that the desired invariance also holds for the sequence $\hat{G}_{1t}$ and hence for test statistic S as a whole.

Although our original formulation of test statistic S applies even when overidentifying restrictions are imposed on cointegrating vectors or the parameters $\boldsymbol{\nu}$ and φ in (18), it seems convenient to consider such restrictions only after broader aspects of cointegrating relations have been specified. This means that, in practice, these overidentifying restrictions are specified only after the cointegrating rank and the form of the deterministic sequence $\boldsymbol{g}_t(\boldsymbol{\mu})$ or the function $\boldsymbol{f}(x; \boldsymbol{\mu}_2)$ in (18) have been specified. If this approach is adopted, test statistic S is applied only in cases in which the above-mentioned invariance property holds, so that one does not need to worry about the possible effect of incorrect normalization on the test. In the special case in which the null model can be estimated by the reduced rank regression method, this particularly means that the resulting ML estimators can be used directly to obtain test statistic S. This happens, for example, when $\boldsymbol{g}_t(\boldsymbol{\mu}) = \boldsymbol{\nu}$.

5. EMPIRICAL APPLICATION

As an empirical application of the ideas put forward in the preceding sections, the relationship between own-yield of broad money and a set of interest rates as an opportunity cost of money is studied.[1] It is often suggested [see, e.g., Rasche (1992)] that, in an economy with perfect capital markets, a system with a set of opportunity costs of money and own-yield of money should contain one common trend. This means that, if interest rates are believed to be $I(1)$ processes, they should be cointegrated with cointegrating rank one less than the number of considered series.[2]

We use monthly Finnish data from the Bank of Finland database covering the period 1980–1995. In addition to own-yield of money (IOWN), our data set contains three interest rates, which are 1-month money market rate (I1M), 3-month money market rate (I3M), and 5-year bond rate (IBOND). Graphs of these series[3] are depicted in Figure 1. Recently Luukkonen et al. (1999) analyzed the same data set by using a standard cointegrated VAR model with an intercept term restricted to the cointegrating space. Their results were rather surprising. The own-yield of money was not cointegrated with any of the three interest rates and only two cointegrating relations were found instead of the expected three.

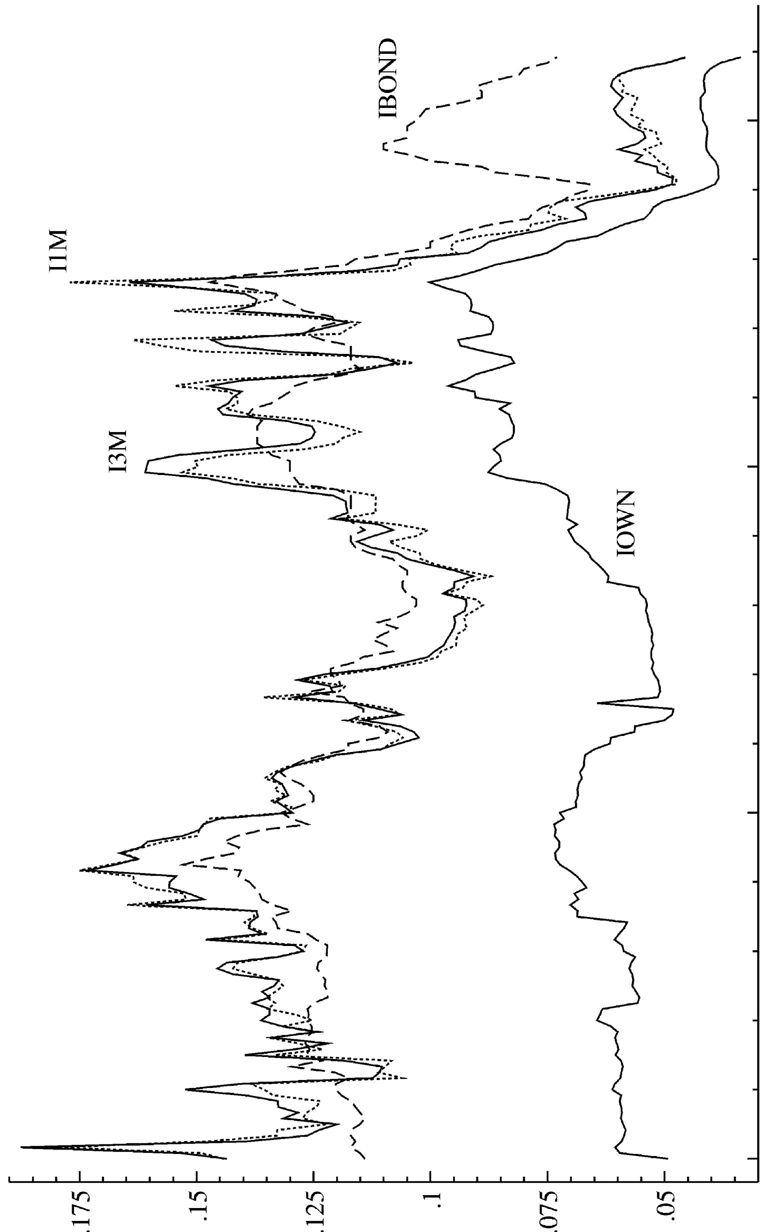

Figure 1. Interest-rate series from Bank of Finland database, 1980–1995: IOWN = own-yield of money, I1M = 1-month money market rate, I3M = 3-month money market rate, and IBOND = 5-year bond rate.

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

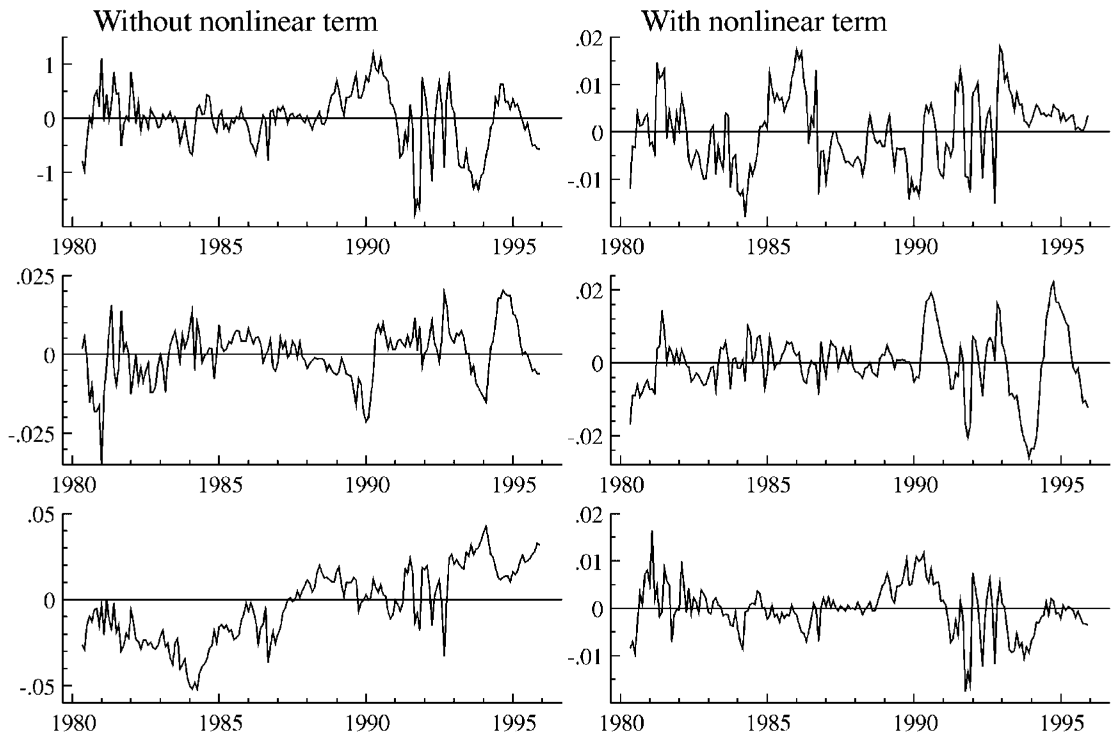

Figure 2. Cointegrating relations of the model with and without nonlinear terms. The first column presents cointegrating relations based on unrestricted reduced rank estimation of the VAR(4) model of Luukkonen et al. (1999). The cointegrating relations are ordered according to the canonical correlations so that the first graph corresponds to the largest canonical correlation. The second column presents cointegrating relations based on the estimates of model (19).

https://doi.org/10.1017/S1365100501023069 Published online by Cambridge University Press

There is a potential explanation for their result. The own-yield of money is a weighted average of the yields of various components of broad money. The Ministry of Finance restricted banks' certificates of deposit (CD) issue until the beginning of 1987 but gradually lifted the quota due to the plans of the Bank of Finland to start open-market operations in March 1997. The CD's were chosen as the material for open-market operations. The open-market operations were started because the Bank of Finland had a long-run objective to improve the functioning of the money markets and to provide central bank financing in a deregulated economic environment. After that period the CD stock grew rapidly and, consequently, the gap between the own-yield of money and the opportunity cost of money diminished substantially. Because of the long-run objectives of the Bank of Finland, the change in the gap can be considered as exogenous with respect to the level of interest rates. The standard model of Luukkonen et al. (1999) assumed that this gap is constant, not changing gradually over time. Thus, the incapability of the employed model to allow for the effects of changing regulations might be the reason for the previous unexpected results.

The left panel of Figure 2 depicts the first three cointegrating relations of the standard reduced-rank VAR(4) model of Luukkonen et al. (1999). The cointegrating relations are ordered according to the canonical correlations. The third cointegrating relation exhibits nonstationary features, which is in line with the cointegrating rank tests of Luukkonen et al. (1999). However, as our previous discussion suggests, this nonstationarity might be modeled by a smooth function of time, which could pick up the "gradualism" in the deregulation and banks' adjustment to the new procedures of the Bank of Finland. To formally test for this idea, we use the test procedure developed in the preceding section. Thus, we continue the analysis by assuming that the cointegrating rank is 3 and the lag length of the VAR model is 4. On the basis of the graph of the third cointegrating relation in the left panel of Figure 2, a function of the logistic form might be appropriate. This would mean choosing $w_t = t/T$ in test statistic S. However, in addition to this choice, we also apply the test with $\boldsymbol{w}_t = [t/T\,(t/T)^2]'$. The outcome of these two tests is reported in the two middle columns of Table 1. When the test is applied with $w_t = t/T$, the null hypothesis of no structural change is rejected even at the 1% significance level but with the other choice of $\boldsymbol{w}_t$ a rejection at the 7% significance level is only possible. The difference between the outcomes of these two tests also suggests that a function of the logistic form is more appropriate than, for instance, a bell-shaped function such as the one in (6).

The first column presents cointegrating relations based on unrestricted reduced rank estimation of the VAR(4) model of Luukkonen et al. (1999). The cointegrating relations are ordered according to the canonical correlations so that the first graph corresponds to the largest canonical correlation. The second column presents cointegrating relations based on the estimates of model (19).

Thus, we augment the model of Luukkonen et al. (1999) by including a logistic function of time in the cointegrating relations. This leads to the model

TABLE 1. Testing continuous structural change

Degree of polynomial	Degrees of freedom	Standard model		Model (19)	
		Test statistic	p-value	Test statistic	p-value
1	3	10.55	0.01	4.41	0.22
2	6	11.49	0.07	5.43	0.49

$$\Delta \boldsymbol{y}_t = \alpha\{\boldsymbol{y}_{1,t-1} - \boldsymbol{A}y_{2,t-1} - \boldsymbol{\nu} - \varphi \boldsymbol{f}[\gamma(t/T - \tau)]\} + \Gamma \boldsymbol{q}_t + \epsilon_t, \qquad \textbf{(19)}$$

where $\boldsymbol{q}_t$ contains lagged differences (lag length is 4), $\boldsymbol{y}_{1,t} = [\text{IOWN}_t\ \text{IBOND}_t\ \text{I3M}_t]'$, and $y_{2,t} = \text{I1M}_t$. Moreover, $\boldsymbol{A}$, $\boldsymbol{\nu}$, and φ are 3×1 parameter vectors and $f[\gamma(t/T - \tau)] = \{1 + \exp[-\gamma(t/T - \tau)]\}^{-1}$. Note that we assume that the form of nonlinearity, that is, the function $f(\cdot)$, is the same in each cointegrating relation.

The right column of Figure 2 shows the cointegrating relations based on the estimates of model (19). According to visual inspection, all three cointegrating relations look stationary, which supports the prior belief of three cointegrating relations. Unfortunately, there seems to be no formal test for cointegration available in the case of smooth deterministic trends. However, since the estimated logistic function turned out to be fairly steep,[4] we approximated it with a step dummy and applied the LR test for cointegration with critical values computed by the Disco program of Nielsen (1993). Because of this approximation and also because the date when the value of this step dummy is changed is determined by the ML estimate of the parameter τ in (19), these critical values should be treated only as approximations of their correct asymptotic counterparts. Using this procedure, the p-value obtained for the null hypothesis that the cointegrating rank is, at most, 2 was approximately 0.07. Thus, even this approximate test was not strongly contrary to the prior belief of three cointegrating relations.

Before discussing ML estimates obtained for the preceding model, we test the adequacy of the specified nonlinearity. Thus, we use the general form of test statistic S and, in the same way as earlier, apply it with both $w_t = t/T$ and $\boldsymbol{w}_t = [t/T\,(t/T)^2]'$. The results, reported in the last two columns of Table 1, imply that the deterministic nonlinearity can be captured by a single logistic function so that the specified model is satisfactory in this respect.

As to the ML estimates of the parameters of model (19), we first note that the t values obtained for the last two components of the parameter vector φ were only 0.19 and 0.05. This suggests that the nonlinearity is only present in the first cointegrating relation, which describes the relationship between own-yield of money and 1-month money market rate. The p-value of the LR test for the joint hypothesis $\varphi_2 = \varphi_3 = 0$ was 0.98, supporting this view. Thus, it is reasonable to consider the corresponding restricted version of model (19). The ML estimates obtained for the long-run parameters of this model are as follows (standard errors in parenthesis):

$$\Delta\begin{bmatrix} \text{IOWN} \\ \text{IBOND} \\ \text{I3M} \\ \text{I1M} \end{bmatrix}_t = \begin{bmatrix} -0.137 & 0.045 & 0.077 \\ (0.038) & (0.049) & (0.103) \\ 0.056 & -0.170 & 0.176 \\ (0.125) & (0.030) & (0.040) \\ 0.212 & 0.046 & 0.00145 \\ (0.084) & (0.100) & (0.056) \\ 0.330 & 0.036 & 0.390 \\ (0.074) & (0.156) & (0.186) \end{bmatrix}$$

$$\times \left\{ \begin{bmatrix} \text{IOWN} \\ \text{IBOND} \\ \text{I3M} \end{bmatrix}_{t-1} - \begin{bmatrix} 0.633 \\ (0.038) \\ 0.464 \\ (0.049) \\ 0.963 \\ (0.024) \end{bmatrix} \text{I1M}_{t-1} - \begin{bmatrix} -0.0224 \\ (0.005) \\ 0.064 \\ (0.0058) \\ 0.0056 \\ (0.0029) \end{bmatrix} \right.$$

$$\left. - \begin{bmatrix} 0.0268 \\ (0.0021) \\ 0 \\ \\ 0 \\ \\ \end{bmatrix} \frac{1}{1 + \exp\left[\underset{(51.1)}{-80.1}(t/T - \underset{(0.010)}{0.448}) \right]} \right\} + \hat{\Gamma}\boldsymbol{q}_t + \hat{\epsilon}_t .$$

The estimation results are interesting. The relationship between 1-month money market rate (I1M) and bond rate (IBOND) differs significantly from a one-to-one relationship.[5] This one-to-one relationship is preserved between 1-month and 3-month money market rates. The constant term in both relationships significantly differs from zero, suggesting, for example, that the 3-month money market rate is, on the average, a half percentage point above the 1-month money market rate. This is consistent with the common observation that the yield curve is, on the average, upward sloping.

The relationship between own-yield of money and 1-month money market rate exhibits structural change in the intercept term. On the average, the own-yield of money has moved along with the 1-month money market rate, but the responses have been smaller than unity (0.633). The nonlinear term suggests that the "adjusted gap" between these two variables has narrowed by 2.7 percentage points. The estimate of τ is 0.448, which suggests that the midpoint of this transition period was in April 1987. That is close to the date when the Bank of Finland started its market operations (March 1987). The shift to the new level has been quite rapid ($\gamma = -80$)—graphical investigation suggests that the major part of the shift occurred within 12 months surrounding the midpoint (April 1987). It is quite natural that the transition started half a year before the midpoint, because the banks' CD quotas were extended already before the Bank of Finland started its open-market operations in order to enlarge the money markets. It is also natural to believe that it took some time after the start of the open-market operations to saturate banks' needs to issue CDs. The uncertainty in the estimation of the parameter γ is likely to arise from this rapid shift, which means that we do not have very many observations from the "transition period." Note that this also means that it is not reasonable to interpret this result as an indication of insignificance.[6] The estimate of the parameter α is reasonable. The residuals of the model exhibit no autocorrelation, but they suffer from nonnormality and, in some cases, heteroskedasticity. Trying to improve the model in this respect is outside the scope of this paper, however.

6. CONCLUSION

This paper has argued that the puzzle of "missing cointegrating vector," which is not uncommon in empirical analysis of cointegrated systems, can in some cases be explained by structural changes and solved by including nonlinear deterministic trends in cointegrating relations. This idea was implemented in the paper by extending the conventional cointegrated VAR model to a fairly general form. The case of continuous deterministic trends was emphasized because gradual structural changes are often conceivable and may not be well modeled by conventional dummy variables, which have so far been used mainly in these contexts.

The usefulness of the proposed approach was demonstrated in the paper by an empirical example on interest-rate data. In this example, a previous model was augmented by including a logistic trend term in the cointegrating relations and the puzzle of the "missing cointegrating vector" could thereby be solved. According to our general idea, we could also pinpoint reasons for this augmentation and interpret the parameter estimates in the logistic trend in a reasonable way. However, even in our model, no one-to-one relationship between interest rates was found.

Finally, in the same way as in Lin and Teräsvirta (1994), in the case of a nonlinear time trend the ML estimation required in our approach is demanding and can cause problems. The estimation algorithm that we used may converge slowly and find a local optimum. Good starting values are therefore an important prerequisite to successful parameter estimation.

NOTES

1. The computations were done with Gauss 3.2 with CML library. We also thank Bent Nielsen for letting us use his DisCo program.

2. Here it is implicitly assumed that the expectation hypothesis of the term structure holds and that a possible risk premium is stationary.

3. Because of the devaluation speculations in August–September 1986, the short-term interest rates rose temporarily two percentage points. We adjust the data for this spike, which was particularly acute in the 1-month and 3-month money market rates.

4. See the discussion later in this section.

5. The one-to-one relationship was suggested by, among others, Campbell and Shiller (1987) and Stock and Watson (1988) and recently studied, e.g., by Lanne (1997) and is based on the cointegration implications of the expectation hypothesis of the term structure of interest rates. The one-to-one relationship between the own-yield and the opportunity cost of money is suggested by, e.g., Ripatti (1998).

6. See Lin and Teräsvirta (1994) for a similar result and further discussions.

REFERENCES

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Andrews, D.W.K. & C.J. McDermott (1995) Nonlinear econometric models with deterministically trending variables. *Review of Economic Studies* 62, 343–360.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Campbell, J.Y. & R.J. Shiller (1987) Cointegration test of present value models. *Journal of Political Economy* 95, 1052–1088.

Dahlhaus, R. (1996a) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1–37.

Dahlhaus, R. (1996b) Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In P.M. Robinson & M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series Analysis 2*, pp. 145–159. New York: Springer.

Davies, R.B. (1977) Hypothesis testing when the nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

Davies, R.B. (1987) Hypothesis testing when the nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.

Eitrheim, Ø. & T. Teräsvirta (1996) Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics* 74, 59–75.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

Johansen, S. (1995) *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Lanne, M. (1997) *Essays on Inference in Time Series Models with Near Unit Roots*. Helsinki: University of Helsinki.

Leybourne, S., P. Newbold, & D. Vougas (1998) Unit roots and smooth transitions. *Journal of Time Series Analysis* 19, 83–97.

Lin, C.-F. & T. Teräsvirta (1994) Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics* 62, 211–228.

Lütkepohl, H. (1996) *Handbook of Matrices*. Chichester, England: John Wiley.

Luukkonen, R., A. Ripatti & P. Saikkonen (1999) Testing for a valid normalization of cointegrating vectors in vector autoregressive processes. *Journal of Business & Economic Statistics* 17, 195–205.

Nielsen, B. (1993) *DisCo, Version 1.1, A Program for Simulating Asymptotic Tables of the Cointegration Rank Test*. Institute of Mathematical Statistics, University of Copenhagen.

Park, J.Y. (1990) Testing for unit roots and cointegration by variable addition. In G.F. Rhides & T.B. Fomby (eds.), *Advances in Econometrics*. Greenwich: JAI Press.

Park, J. & S. Hahn (1999) Cointegrating regressions with time varying coefficients. *Econometric Theory* 15, 664–703.

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.

Rasche, R.H. (1992) Money Demand and the Term Structure; Some New Ideas on an Old Problem. Econometrics and economic theory paper 9211, Michigan State University.

Ripatti, A. (1998) Stability of the demand for M1 and harmonized M3 in Finland. *Empirical Economics* 23, 317–337.

Saikkonen, P. (2001a) Consistent estimation in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations. *Econometric Theory* 17, 296–326.

Saikkonen, P. (2001b) Statistical inference in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations. *Econometric Theory* 17, 327–356.

Stock, J.H. & M.W. Watson (1988) Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.