

Hansen, Peter Reinhard

Working Paper

Structural changes in the cointegrated vector autoregressive model

Working Paper, No. 2000-20

Provided in Cooperation with:

Department of Economics, Brown University

Suggested Citation: Hansen, Peter Reinhard (2001) : Structural changes in the cointegrated vector autoregressive model, Working Paper, No. 2000-20, Brown University, Department of Economics, Providence, RI

This Version is available at:

<https://hdl.handle.net/10419/80180>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**Structural Changes
in the
Cointegrated Vector Autoregressive Model**

Peter Reinhard Hansen

Working Paper No. 00-20

First Version: October, 2000

This Version: September, 2001



**Brown University
Department of Economics, Box B
Providence, RI 02912, USA
Phone: (401) 863 9864
Email: Peter_Hansen@brown.edu**

JEL Classification: C3, C32

Keywords: Structural Change, Cointegration, Vector Autoregression, Term Structure, Expectations Hypothesis

Abstract

This paper generalizes the cointegrated vector autoregressive model of Johansen (1988) to allow for structural changes. We take the time of the change points and the number of cointegration relations as given. Estimation under various hypotheses is made possible by a new estimation technique, which makes it simple to derive a number of interesting likelihood ratio tests. For example, one can test for m structural changes against $m + k$ structural changes, or test linear parameter restrictions in the presence of structural changes. The asymptotic distribution of the LR statistic is χ^2 in both cases.

The model is applied to US term structure data, and structural changes in September 1979 and October 1982, which coincide with changes in the Fed's policy, are found to be significant. After accounting for these structural changes, we cannot, contrary to previous studies, reject the long-run implication of the expectations hypothesis.

1 Introduction

The modeling of structural changes in cointegrated processes has been addressed by several authors. In the vector autoregressive framework, Seo (1998) derived the Lagrange multiplier (LM) test for structural changes in cointegration relations and adjustment coefficients, and Inoue (1999) derived a rank test for cointegrated processes with a broken trend. Other approaches to modeling structural changes in cointegrated processes are the recursive estimation to identify structural changes by H. Hansen and Johansen (1999), the combination of cointegration and Markov switching by Krolzig (1996), the co-breaking theory by Hendry (1995), and a test for a cointegration relation with a structural change against an $I(1)$ alternative was given by Gregory and B. E. Hansen (1996).

One of the main contributions of this paper is the development of a flexible framework in which structural changes can be formulated. In this framework it is possible to test for structural changes in any subset of the parameters, referred to as a partial structural change.¹ So it is possible to test for a structural change in a particular cointegration relation or a particular coefficient, leaving other relations and parameters unchanged. Throughout this paper, we take the change points and the number of cointegration relations in each of the subsamples as given. The number of cointegration relations is not required to be constant over the full sample, but the number of cointegration relations at any point in time is taken as given. The more general cases where the change points are unknown, or the number of cointegration relations needs to be estimated, are not treated in this paper.² Because the change points are taken as given, the results are, in this sense, less general than those of Seo (1998). However, the results are more general in other ways, because the framework allows for multiple structural changes, non-constant cointegration rank, and it is possible to test for changes in any subset of the parameters. The test statistic invoked in this paper is the likelihood ratio (LR) test, which is shown to have an asymptotic χ^2 distribution. Another contribution of this paper is that it enables hypotheses testing under the maintained hypothesis that the underlying process exhibits structural changes. The asymptotic χ^2 results remain valid in this situation, as long as the hypotheses do not violate the assumed number of cointegration relations.

Another contribution of this paper is the introduction of a new estimation technique, the *generalized reduced rank regression* (GRRR) technique. This technique has an applicability beyond the estimation problems that arise from structural changes. Estimation of the cointegrated vector autoregressive model (VAR) was solved by Johansen (1988) as an eigenvalue problem, also known as reduced rank regression. This technique is directly applicable to estimation under simple linear restrictions on cointegration relations, β , and adjustment coefficients,

¹Partial structural changes in stationary processes has been analysed by Bai and Perron (1998) and Bai (1999).

²The case with unknown change points leads to non-standard asymptotic distributions. See Seo (1998) who used the framework of Andrews and Ploberger (1994), or B. E. Hansen (1992) and H. Hansen and Johansen (1999) who applied the framework of Nyblom (1989).

α . Johansen and Juselius (1992) proposed a switching algorithm for estimation under slightly more general restrictions. Boswijk (1995) derived a general estimation technique that can handle linear restrictions on $\text{vec}(\alpha)$ and $\text{vec}(\beta)$, where $\text{vec}(\cdot)$ is the vectorization operator. The estimation technique of Boswijk (1995) is applicable to several estimation problems we face with structural changes in the cointegrated VAR. The GRRR technique introduced in this paper is a generalization of his technique in two directions. First of all, the GRRR technique allows for linear restrictions on all parameters apart from the variance parameter. Secondly, the GRRR technique does not require a constant covariance matrix but allows us to model structural changes in the covariance matrix.

The framework developed in this paper is applied to the US term structure of interest rates. The results are that the long-run implication of the expectations hypothesis cannot be rejected once structural changes have been accounted for.

The paper is organized as follows. Section 2 contains the statistical formulation of structural changes in the cointegration VAR. The estimation problems are treated in Section 3, and Section 4 contains the asymptotic analysis. Section 5 contains an empirical analysis of the expectations hypothesis applied to the US term structure of interest rates. Section 6 concludes, and the appendix contains proofs.

2 The Statistical Model

In this section we formulate structural changes in the p -dimensional cointegrated vector autoregressive model. The model with constant parameters is given by

$$\Delta X_t = \alpha \beta' X_{t-1}^* + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where $\varepsilon_t \sim iid N(0, \Omega)$, and where D_t is a q -dimensional vector that contains deterministic terms such as a constant, a linear trend, and seasonal dummies. The variable X_{t-1}^* consists of X_{t-1} and restricted deterministic variables. For example a linear trend, t , may be included in X_{t-1}^* , rather than in D_t , in order to avoid a quadratic trend in the process, X_t . We refer to α as the adjustment coefficients, and to β as the cointegration parameters, see Johansen (1988). The dimensions of the parameters are as follows: α is $p \times r$, β is $p_1 \times r$, where p_1 is the dimension of X_{t-1}^* , Γ_i is $p \times p$, $i = 1, \dots, k-1$, Φ is $p \times q$, and Ω is $p \times p$.

Structural changes are introduced by allowing the parameters to change their values at the change-points: T_1, \dots, T_{m-1} , where $0 < T_1 < \dots < T_{m-1} < T$, and we map these change-points into the unit interval by defining $\rho_i \equiv T_i/T$, $i = 1, \dots, m$. In the most general situation we allow for structural changes in all parameters. However, in most applications it is desirable to keep some parameters constant, such that the number of free parameters is small relative to the sample size, and we can impose constancy of one or more parameters, across two or more subsamples, using simple parameter restrictions.

The generalization of (1) is the following:

$$\Delta X_t = \alpha(t)\beta(t)'X_{t-1}^* + \sum_{i=1}^{k-1} \Gamma_i(t)\Delta X_{t-i} + \Phi(t)D_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

where $\{\varepsilon_t\}$ is a sequence of independent Gaussian variables with mean zero and variance $\Omega(t)$. The time-varying parameters are piecewise constant and given by

$$\begin{aligned} \alpha(t)\beta(t)' &= \alpha_1\beta_1'1_{1t} + \dots + \alpha_m\beta_m'1_{mt}, \\ \Gamma_i(t) &= \Gamma_{1,i}1_{1t} + \dots + \Gamma_{m,i}1_{mt}, \quad i = 1, \dots, k-1, \\ \Phi(t) &= \Phi_11_{1t} + \dots + \Phi_m1_{mt}, \\ \Omega(t) &= \Omega_11_{1t} + \dots + \Omega_m1_{mt}, \end{aligned}$$

where we have defined an indicator function for each of the m subsamples

$$1_{jt} \equiv 1(T_{j-1} + 1 \leq t \leq T_j), \quad j = 1, \dots, m,$$

with the conventions $T_0 \equiv 0$ and $T_m \equiv T$.

The Granger representation we obtain for this process shows that the columns of β_j define the cointegration relations in subsample j . Since β_i and β_j are not required to have the same number of columns, $i \neq j$, this formulation allows for changes in the number of cointegration relations. Thus, the rank of $\alpha(t)\beta(t)'$ may vary across subsamples, in which case α_j and β_j have r_j columns, where r_j denotes the cointegration rank in subsample j , $j = 1, \dots, m$. All other parameters have dimensions that are constant across subsamples.³

We define the variables $Z_{0t} = \Delta X_t$, $Z_{1t} = (1_{1t}X_{t-1}^{*'}, \dots, 1_{mt}X_{t-1}^{*'})'$, $\tilde{Z}_{2t} = (\Delta X_{t-1}', \dots, \Delta X_{t-k+1}', D_t')'$, and $Z_{2t} = (1_{1t}\tilde{Z}_{2t}', \dots, 1_{mt}\tilde{Z}_{2t}')'$ and denote the number of variables \tilde{Z}_{2t} by p_2 , such that Z_{1t} and Z_{2t} consist of mp_1 and mp_2 variables, respectively. Further, we define the parameters $A = (\alpha_1, \dots, \alpha_m)$,

$$B = \begin{pmatrix} \beta_1 & 0 & \dots & 0 & 0 \\ 0 & \beta_2 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \beta_{m-1} & 0 \\ 0 & 0 & \dots & 0 & \beta_m \end{pmatrix},$$

and $C = (\Psi_1, \dots, \Psi_m)$, where $\Psi_j = (\Gamma_{j,1}, \dots, \Gamma_{j,k-1}, \Phi_j)$, $j = 1, \dots, m$. In the following, we use the notation

³In a more general setting one could include different deterministic terms, D_t , in different subsamples in which case the dimension of $\Phi(t)$ could differ across subsamples. If the (number of) variables included in X_t varies across time, it will cause variation in p , which implies additional variation in the dimensions of some of the parameters.

$B = \text{diag}(\beta_1, \dots, \beta_m)$ to save space. These definitions enable us to express (2) as the regression equation

$$Z_{0t} = AB'Z_{1t} + CZ_{2t} + \varepsilon_t, \quad t = 1, \dots, T, \quad (3)$$

which has constant parameters. Apart from the block diagonal structure of B and the non-constancy of $\Omega(t)$, this equation has the form of a reduced rank regression problem, see Anderson (1984) or Johansen (1996).

We can impose the required structure of B using parameter restrictions of the form

$$\text{vec}(B) = H\varphi + h, \quad (4)$$

where $\text{vec}(\cdot)$ is the vectorization operator. The matrix H is a known $mp_1(r_1 + \dots + r_m) \times p_\varphi$ matrix, h is a known $mp_1(r_1 + \dots + r_m)$ dimensional vector, and φ is a vector, which consists of p_φ parameters. The vector h , will serve as a tool to normalize/identify the parameters in A and B . Besides the restrictions on B , we also consider restrictions on (A, C) of the form

$$\text{vec}(A, C) = G\psi, \quad (5)$$

where G is a known $p(r_1 + \dots + r_m + mp_2) \times p_\psi$ matrix, and where ψ is a vector with p_ψ free parameters.

The restrictions (4) and (5) are useful for three aspects of this framework. First of all, they server to impose the required block-diagonal structure of B . Second, the restrictions can be used to impose constancy of some (or all) of the parameters. This can be useful in practice, as the number of free parameters increase linearly with the number of subsamples, unless some parameters are held constant. In applications, where the point of origin is the model with constant parameters, it will often be relevant to test for structural changes in a particular subset of parameters, rather than all parameters. The third use of the restrictions (4) and (5) is to impose other parameter restrictions of interest, for example, that two coefficients equal one another.

2.1 Structural Changes in the Cointegration Parameters

In Section 4, we show that the asymptotic distribution of the LR tests is χ^2 with degrees of freedom that equals the reduction of free parameters. In order to calculate the number of free parameters in a given model, we need the following lemma, taken from Johansen (1996).

Lemma 1 *The function $f(x, y) = xy'$, where x is $p \times r$ ($r \leq p$) and y is $p_1 \times r$ ($r \leq p_1$), is differentiable at all points, with a differential given by*

$$Df(x, y) = x(dy)' + (dx)y'$$

where dy is $p \times r$ and dx is $p_1 \times r$. If x and y have full rank r , then the tangent space at (x, y) , given by $\{x(dy)' + (dx)y' : dx \in \mathbb{R}^{p_1 \times r}, dy \in \mathbb{R}^{p \times r}\}$, has dimension $(p + p_1 - r)r$.

The lemma is useful for calculating the number of free parameters of a product of matrices. For example, in the model with $m - 1$ structural changes we can calculate the number of free parameters in $\alpha(t)\beta(t)'$ by setting $x = \alpha_i$ ($p \times r_i$) and $y = \beta_i$ ($p_1 \times r_i$) and find the number to be $\sum_{i=1}^m (p + p_1 - r_i)r_i$. If the rank is constant over the full sample, then the expression simplifies to $m(p + p_1 - r)r$.

2.1.1 A Structural Change in the Adjustment Coefficients: α

Consider the case where the structural changes do not affect the cointegration relations. In this case, we can express the model as

$$Z_{0t} = (\alpha_1, \dots, \alpha_m) \text{diag}(\beta, \dots, \beta)' Z_{1t} + \Psi Z_{2t} + \varepsilon_t,$$

and the constancy of $\beta(t)$ is straight forward to impose with a suitable choice of H . Since β is constant, so is the cointegration rank r , and the number of free parameters in $\alpha(t)\beta(t)'$ is given by $(mp + p_1 - r)r$.

2.1.2 Structural Changes in the Cointegration Relations: β

When the structural change is solely due to changes in the cointegration relations $\beta(t)$ while $\alpha(t)$ remains constant, the model simplifies to

$$\begin{aligned} \Delta X_t &= \alpha \beta'_1 1_{1t} X_{t-1}^* + \dots + \alpha \beta'_m 1_{mt} X_{t-1}^* + \Psi Z_{2t} + \varepsilon_t \\ &= \alpha (\beta'_1, \dots, \beta'_m) Z_{1t} + \Psi Z_{2t} + \varepsilon_t. \end{aligned}$$

Thus, this model can also be formulated in the form of (3), but without the need of the restrictions (4) and (5). In this situation only a constant cointegration rank, r , is meaningful and the matrices A and B have dimensions $p \times r$ and $mp_1 \times r$ respectively. So the number of free parameters in AB' is given by $(p + mp_1 - r)r$.

The relations between the different structural changes are displayed in Figure 1, along with the degrees of freedom of some LR tests.

The following example illustrates how the restriction matrix, H , can be used to impose the required structure on B under various hypotheses.

Example 1 Consider the bivariate system, $p = 2$, with two subsamples, $m = 2$, and a single cointegration relation in both subsamples, $r_1 = r_2 = 1$. Since $m = 2$ the process can have no more than one structural change. In this setting $B = \text{diag}(\beta_1, \beta_2)$, where $\beta_1 = (\beta_{1,1}, \beta_{1,2})'$, and where $\beta_2 = (\beta_{2,1}, \beta_{2,2})'$. The zero elements can

be imposed with

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{since } \text{vec}(B) = \begin{pmatrix} \beta_{1,1} \\ \beta_{1,2} \\ 0 \\ 0 \\ 0 \\ 0 \\ \beta_{2,1} \\ \beta_{2,2} \end{pmatrix}.$$

This choice of H implies that $\varphi = (\beta_{1,1}, \beta_{1,2}, \beta_{2,1}, \beta_{2,2})'$. Imposing that the cointegration parameters are constant can be achieved with the restrictions $\beta_{1,1} = \beta_{2,1}$ and $\beta_{1,2} = \beta_{2,2}$, which can be imposed using

$$H_r = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}'.$$

Constancy of $\alpha(t)$ and $\Phi(t)$, as well as additional restrictions on these parameters, can be imposed in a similarly way, using a suitable choice for G .

In this case we have $h = 0$, and B is not identified. However, identification is not needed for testing the hypothesis of constant cointegration relations, because this is a hypothesis on the subspace spanned by the cointegration vector. If inference is to be made on particular coefficients of $\beta(t)$ or $\alpha(t)$, then a normalization is needed. This can be achieved with the following choices of H and h ,

$$H_n = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad H_{n,r} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad h = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

where H_n is the restriction matrix used in the unrestricted (and identified) model, and $H_{n,r}$ is the restriction matrix used in restricted (and identified) model. Both models apply the same h to impose the normalization $\beta_{1,1} = \beta_{2,1} = 1$.

2.2 Temporary and Permanent Cointegration Relations

The scenario where some cointegration relations are present in the full sample, whereas others are only present in a subsample is also simple to formulate in this framework. Consider the case with a single change point, where β define r_1 permanent cointegration relations, and β_e define an additional $r_2 - r_1$ temporary cointegration relations that are linearly independent of β . Thus, $\beta_1 = \beta$ and $\beta_2 = (\beta_1, \beta_e)$ and we obtain the equation

$$Z_{0t} = (\alpha_1, \alpha_{2,1}, \alpha_{2,e}) \begin{pmatrix} \beta & 0 & 0 \\ 0 & \beta & \beta_e \end{pmatrix}' Z_{1t} + \Psi Z_{2t} + \varepsilon_t.$$

In this case we impose the zero elements and cross restrictions of the parameters in B , with a suitable choice of H . Similarly, we can impose that the adjustment coefficients corresponding to β are constant ($\alpha_1 = \alpha_{2,1}$) using the restriction matrix G .

In this case it is slightly more complicated to derive the degrees of freedom in $\alpha(t)\beta(t)'$. The most general model, where the constancy of r_1 cointegration relations is not imposed, results in $(p+p_1-r_1)r_1$ free parameters in $\alpha_1\beta_1'$ plus $(p+p_1-r_2)r_2$ parameters in $\alpha_2\beta_2'$. The model with r_1 constant cointegration relations has $(p+p_1-r_2)r_2$ free parameters in $\alpha_2\beta_2'$ plus $(p+r_2-r_1)r_1$ free parameters in $\alpha_1\beta_1'$, where we used that $\beta_1 = \beta_2\xi$ for some $r_2 \times r_1$ matrix ξ . The most restrictive model, where we impose $\alpha_1 = \alpha_{2,1}$, has $(p+p_1-r_1)r_1$ free parameters in $\alpha_1\beta_1'$ and an additional $[p+(p_1-r_1)-(r_2-r_1)](r_2-r_1) = (p+p_1-r_2)(r_2-r_1)$ free parameters in $\alpha_e\beta_e'$. When calculating these numbers, we used that β_e can be chosen orthogonal to β_1 , i.e., $\beta_e = \beta_{1,\perp}\delta$, where $\beta_{1,\perp}$ is the orthogonal complement⁴ to β_1 , and where δ is some $p_1 - r_1 \times r_2 - r_1$ matrix. Adding the two terms yields a total of $(p+p_1-r_2)r_2 + (r_2-r_1)r_1$ free parameters. The relations between the three nested models are displayed in Figure 2. Notice that the three models have identical cointegration ranks, r_1 and r_2 . Had this not been the case, the asymptotic distribution would not be χ^2 .

The extension to models with multiple sets of temporary cointegration relations in individual and overlapping subsamples is straightforward, although the calculation of degrees of freedom can be somewhat complicated.

2.2.1 A Useful Simplification

As discussed earlier, it will often be useful to keep some parameters constant. In some cases, one might impose the constraints: $\Psi_1 = \dots = \Psi_m = \Psi$, and concentrate the analysis to structural changes in $\alpha(t)$, $\beta(t)$, and $\Omega(t)$. Rather than imposing the constraints on $\Psi(t)$ with the restriction matrix G , we can substitute $\tilde{Z}_{2t} = (\Delta X'_{t-1}, \dots, \Delta X'_{t-k+1}, D'_t)'$ for Z_{2t} , and redefine $C = \Psi$. This will automatically impose the constancy of $\Psi(t)$, which is useful because it reduced the dimensions of G and other matrices that are used in the estimation.

⁴If a is a $p \times r$ matrix with column rank $r < p$, then the orthogonal complement to a , denoted a_\perp is a $p \times (p-r)$ matrix with full column rank, that satisfies $a'_\perp a = 0_{p-r \times r}$. The matrix a_\perp is not unique, but the subspace spanned by the columns of a_\perp is unique.

3 Estimation

Estimation of the cointegrated VAR, and other models that can be expressed in the form of (3), can be solved as an eigenvalue problem by reduced rank regression techniques, if the parameters, A , B , and C are unrestricted and the covariance is constant. The method of reduced rank regression was developed by Anderson (1951) and Izenman (1975), and applied to the cointegrated VAR model by Johansen (1988).

Reduced rank regression is related to a number of classical estimation problems, such as principle components and canonical correlations, see Reinsel and Velu (1998). The advantage of reduced rank estimation is that a closed-form solution can be obtained without the use of iterative estimation techniques. This method is applicable to estimation under simple linear restrictions on the reduced rank parameters. However, the reduced rank regression technique is not applicable to most models with structural changes, because these models involve more complicated parameter restrictions. So a more general estimation technique is needed to estimate the models with structural changes.

Some estimation problems can be formulated as regression problems that can be solved with the *switching algorithm* of Johansen and Juselius (1992). This algorithm is an iterative procedure that in every iteration simplifies the problem to a reduced rank regression by keeping a subset of the parameters fixed. This method has the nice property that it increases the value of the likelihood function in every iteration, but unfortunately applications have shown that convergence can be very slow, and it is possible to construct examples where the method will not converge to the global optimum.

A more general estimation technique was proposed by Boswijk (1995). This method is similar to the switching algorithm, in the sense that it increases the likelihood function in every iteration. It is more general because it can handle estimation problems with linear restrictions on $\text{vec}(A)$ and $\text{vec}(B)$. This method is therefore sufficient for the estimation of models that only involve structural changes in the adjustment coefficients, α , and the cointegration parameters, β . Applications of the method have shown that convergence is obtained in few iterations, and that it does converge to the global optimum. The fast convergence is not surprising because the information matrix is asymptotically block diagonal, due to the faster rate of convergence of the estimator for B .

A related estimation technique is the minimum-distances approach by Elliott (1997, 2000), which can estimate parameters under the general restriction $g(\theta) = c$, where θ is the vector of parameters, c is a constant and g is a well-behaved function. This method minimizes $\theta' \hat{V}_{\hat{\theta}} \theta$ subject to the constraints $g(\theta) = c$, where $\hat{V}_{\hat{\theta}}$ is an estimate of the asymptotic covariance matrix.

As we shall see below, it is possible to estimate under more general restrictions than those considered by Boswijk (1995) and Elliott (1997, 2000). The restrictions (4) and (5) achieve the same generality as the minimum distance method, and can in addition estimate models with a non-constant variance.

3.1 Generalized Reduced Rank Regression

Consider the regression equations (3), given by $Z_{0t} = AB'Z_{1t} + CZ_{2t} + \varepsilon_t$, $t = 1, \dots, T$, where the parameters are subject to the constraints (4) and (5), and where we denote the free parameters in $\Omega_1, \dots, \Omega_m$ by the vector θ . The parameters in the model are given by the three vectors, ψ , φ , and θ .

To ensure that our choice of G , H , and h in (4) and (5), does not violate the assumed rank of A and B , we make the following assumption, which is taken from Boswijk (1995).

Assumption 1 *The matrices H and G have full column rank and H , G , and h are such that A and B have full column rank for all $(\psi', \varphi')' \in \mathbb{R}^n$, except on a set with Lebesgue measure zero, where n is the total number of columns in H and G .*

Assumption 2 *The matrices H and G and the vector h are such that φ and ψ are identified.*

Assumption 3 *The parameters ψ , φ , and θ are variation free. I.e., the parameter space for (ψ, φ, θ) is given by a product space: $\Theta_\psi \times \Theta_\varphi \times \Theta_\theta$, where Θ_κ is the parameter space for $\kappa = \psi$, φ , or θ .*

Assumption 2 is needed for testing some (but not all) hypotheses, and Assumption 3 allows us to construct a useful iterative algorithm for our estimation problem. We can now state the main result of this section.

Theorem 2 *Let the parameter A , B , and C be restricted by $\text{vec}(A, C) = G\psi$ and $\text{vec}(B) = H\phi + h$ and suppose that Assumptions 1 and 3 hold.*

The maximum likelihood estimates of A , B , C , and $\Omega(t)$ satisfy

$$\begin{aligned} \text{vec}(\hat{A}, \hat{C}) &= G \left[G' \sum_{t=1}^T \left[\begin{pmatrix} \hat{B}' Z_{1t} Z_{1t}' \hat{B} & \hat{B}' Z_{1t} Z_{2t}' \\ Z_{2t} Z_{1t}' \hat{B} & Z_{2t} Z_{2t}' \end{pmatrix} \otimes \hat{\Omega}(t)^{-1} \right] G \right]^{-1} \\ &\quad \times G' \sum_{t=1}^T \text{vec} \left(\hat{\Omega}(t)^{-1} Z_{0t} (Z_{1t}' \hat{B}, Z_{2t}') \right), \end{aligned} \quad (6)$$

$$\begin{aligned} \text{vec}(\hat{B}) &= H \left[H' \sum_{t=1}^T \left[\hat{A}' \hat{\Omega}(t)^{-1} \hat{A} \otimes Z_{1t} Z_{1t}' \right] H \right]^{-1} \\ &\quad \times H' \left[\sum_{t=1}^T \text{vec} \left(Z_{1t} (Z_{0t} - \hat{C} Z_{2t})' \hat{\Omega}(t)^{-1} \hat{A} \right) \right. \\ &\quad \left. - \sum_{t=1}^T \left[\hat{A}' \hat{\Omega}(t)^{-1} \hat{A} \otimes Z_{1t} Z_{1t}' \right] h \right] + h, \end{aligned} \quad (7)$$

and

$$\hat{\Omega}_j = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} \hat{\varepsilon}_t \hat{\varepsilon}_t', \quad j = 1, \dots, m, \quad (8)$$

where $\hat{\varepsilon}_t = Z_{0t} - \hat{A}\hat{B}'Z_{1t} - \hat{C}Z_{2t}$. The maximum value of the likelihood function is given by

$$L_{\max}(\hat{A}, \hat{B}, \hat{C}, \hat{\Omega}(t)) = (2\pi)^{-\frac{Tp}{2}} \prod_{t=1}^T |\hat{\Omega}(t)|^{-\frac{1}{2}} \exp\left(-\frac{Tp}{2}\right).$$

The proof, which is given in the Appendix, exploits that the estimation problem reduces to a GLS problem, when either $(A, C, \Omega_{j=1, \dots, m})$ or $(B, \Omega_{j=1, \dots, m})$ is held constant, and the fact that the covariance estimate is the average sum-of-squared residuals when (A, B, C) is held constant. If $\Omega(t)$ is assumed to be constant over two or more subsamples, then the estimate is given by the average sum-of-squared residuals over these subsamples.

The theorem yields a procedure for parameter estimation, in the sense that the parameter estimates can be obtained by iterating on the three equations until convergence, from some initial values of the parameters. Consider the procedure that iterates on the following three equations:

$$\begin{aligned} \psi^{(n)} &= \arg \max_{\psi \in \Theta_\psi} L(\psi, \varphi^{(n-1)}, \theta^{(n-1)}), \\ \varphi^{(n)} &= \arg \max_{\varphi \in \Theta_\varphi} L(\psi^{(n)}, \varphi, \theta^{(n-1)}), \\ \theta^{(n)} &= \arg \max_{\theta \in \Theta_\theta} L(\psi^{(n)}, \varphi^{(n)}, \theta), \end{aligned}$$

$n \geq 1$ until convergence of the likelihood function, L , starting from some initial values of the parameters $(\psi^{(0)}, \varphi^{(0)}, \theta^{(0)})$. An algorithm of this kind is useful whenever a complicated (possibly high dimensional) problem can be divided into smaller problems that are simpler to solve. For more details, see Oberhofer and Kmenta (1974) who considered the case with two subsets of parameters. The algorithm increases the value of the likelihood function in every iteration, and since the likelihood function is bounded by its global maximum, the procedure will eventually converge. Since finding a stationary point of the three equations is equivalent to solving the normal equations, a convergence point, say $(\hat{\psi}, \hat{\varphi}, \hat{\theta})$, will satisfy the normal equations, which in our case is given by (6), (7), and (8). So this procedure will produce the maximum likelihood estimators whenever the normal equations uniquely define the global maximum of L . In practise local maxima may exist, so one should start the algorithm with different initial values of the parameters, and check that the algorithm converges to the same value of the likelihood function.

The following two corollaries provide solutions to some simpler estimation problems. Although these problems are special cases of the framework in Theorem 2, these results are quite useful for practical implementations. Some of the estimation problems that arises from our analysis do not require the general formulation of Theorem 2, and the computational burden can be reduced substantially by implementing the results of these corollaries.

To simplify some notation and some of the following expressions, we define the moment matrices $M_{ij} = \frac{1}{T} \sum_{t=1}^T Z_{it}Z'_{jt}$, $i, j = 0, 1, 2$, the residuals $R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}$, $R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}$, and the moment matrices of the residuals $S_{ij} = \frac{1}{T} \sum_{t=1}^T R_{it}R'_{jt}$, $i, j = 0, 1$.

Corollary 3 *Let the parameter A , B , and C be restricted by $\text{vec}(A, C) = G\psi$ and $\text{vec}(B) = H\varphi + h$ and suppose that $\{\varepsilon_t\}$ is iid $N(0, \Omega)$. The maximum likelihood estimates of A , B , C , and Ω satisfy the equations*

$$\begin{aligned}\text{vec}(\hat{A}, \hat{C}) &= G \left[G' \left(\begin{pmatrix} \hat{B}' M_{11} \hat{B} & \hat{B}' M_{12} \\ M_{21} \hat{B} & M_{22} \end{pmatrix} \otimes \hat{\Omega}^{-1} \right) G \right]^{-1} G' \text{vec} \left(\hat{\Omega}^{-1} (M_{01} \hat{B}, M_{02}) \right), \\ \text{vec}(\hat{B}) &= H \left[H' \left(\hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes M_{11} \right) H \right]^{-1} H' \\ &\quad \times \left[\text{vec} \left(M_{10} - M_{02} \hat{C}' \right) \hat{\Omega}^{-1} \hat{A} \right] - \left(\hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes M_{11} \right) h \Big] + h,\end{aligned}$$

and $\hat{\Omega} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$, where $\hat{\varepsilon}_t = Z_{0t} - \hat{A} \hat{B}' Z_{1t} - \hat{C}' Z_{2t}$. The maximum value of the likelihood function is given by

$$L_{\max}^{-2/T}(\hat{A}, \hat{B}, \hat{C}, \hat{\Omega}) = (2\pi e)^p |\hat{\Omega}|.$$

If C is unrestricted we obtain the following result of Boswijk (1995).

Corollary 4 *Let A and B be restricted by $\text{vec}(A) = G\zeta$ and $\text{vec}(B) = H\phi + h$, for known G , H , and h . Then the maximum likelihood estimates satisfy*

$$\begin{aligned}\text{vec}(\hat{B}) &= H \left[H' \left(\hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes S_{11} \right) H \right]^{-1} H' \\ &\quad \times \left[\left(\hat{A}' \otimes S_{10} \right) \text{vec} \left(\hat{\Omega}^{-1} \right) - \left(\hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes S_{11} \right) h \right] + h,\end{aligned}\tag{9}$$

and

$$\hat{A} = G \left[G' \left(\hat{B}' S_{11} \hat{B} \otimes \hat{\Omega}^{-1} \right) G \right]^{-1} G' \left(\hat{B}' \otimes \hat{\Omega}^{-1} \right) \text{vec} (S_{01}),\tag{10}$$

$$\hat{C} = M_{02} M_{22}^{-1} - \hat{A} \hat{B}' M_{12} M_{22}^{-1},$$

$$\hat{\Omega} = S_{00} - S_{01} \hat{B} \hat{A}' + \hat{A} \hat{B}' S_{11} \hat{B} \hat{A}' - \hat{A} \hat{B}' S_{10}.\tag{11}$$

If A is unrestricted, then (10) and (11) simplify to

$$\begin{aligned}\hat{A} &= S_{01} \hat{B} \left(\hat{B}' S_{11} \hat{B} \right)^{-1}, \\ \hat{\Omega} &= S_{00} - S_{01} \hat{B} \left(\hat{B}' S_{11} \hat{B} \right)^{-1} \hat{B}' S_{10}.\end{aligned}$$

The maximum value of the likelihood function is given by

$$L_{\max}^{-2/T}(\hat{A}, \hat{B}, \hat{C}, \hat{\Omega}) = (2\pi e)^p |\hat{\Omega}|.$$

With these results we have the tools available to estimate the parameters in the cointegrated vector autore-

gressive model under all the various structural changes that were considered in the previous section. However, the results presented here have a broader applicability and can be used to estimate parameters in models, which are different from the structural change models considered in this paper.

4 Asymptotic Analysis

We let \xrightarrow{p} , \xrightarrow{d} , and \xrightarrow{w} , denote convergence in probability, convergence in distribution, and weak convergence, respectively. To simplicity the analysis we derive the results in the case where the parameters, $\alpha(t)$, $\beta(t)$, and $\Omega(t)$ may have a single structural change at time T_1 , and all other parameters are kept constant. From the analysis, it is clear that the extensions to multiple structural changes and non-constancy of additional parameters lead to the same results. Also, as is the case for the standard model, the asymptotic results depend on the deterministic part of the process. For simplicity we consider the model without deterministic terms. The analysis for other choices of the deterministic term is quite similar.⁵

So we consider the process described by

$$\Delta X_t = \alpha_1 \beta'_1 X_{t-1} 1_{(t \leq T_1)} + \alpha_2 \beta'_2 X_{t-1} 1_{(t > T_1)} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t, \quad (12)$$

where $\varepsilon_t \sim N(0, \Omega(t))$ and independent, and where $\Omega(t) = \Omega_1$ for $t \leq T_1$ and $\Omega(t) = \Omega_2$ for $t > T_1$, and we assume that the usual $I(1)$ assumptions hold in both subsamples.

Assumption 4 *The roots of $A_j(z) = |I(1-z) - \alpha_i \beta'_i z - \sum_{i=1}^{k-1} \Gamma_i (1-z) z^i| = 0$ are outside the unit circle or equal to one, and the matrix $\alpha'_{j\perp} (I - \Gamma_1 - \dots - \Gamma_{k-1}) \beta_{j\perp}$ has full rank $p - r_j$, $j = 1, 2$.*

Note, that we do not require the cointegration rank, r_j , to be the same in the two subsamples.

Lemma 5 (Granger's Representation Theorem) *Define the matrices $C \equiv \beta_{1\perp} (\alpha'_{1\perp} \Gamma \beta_{1\perp})^{-1} \alpha'_{1\perp}$, $D \equiv \beta_{2\perp} (\alpha'_{2\perp} \Gamma \beta_{2\perp})^{-1} \alpha'_{2\perp}$, and $\Gamma \equiv I - \Gamma_1 - \dots - \Gamma_{k-1}$. The process given by*

$$X_t = C \sum_{i=1}^t \varepsilon_i + C(L) \varepsilon_t + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}), \quad t = 1, \dots, T_1, \quad (13)$$

$$\begin{aligned} X_t = & D \sum_{i=T_1+1}^t \varepsilon_i + D \Gamma C \sum_{i=1}^{T_1} \varepsilon_i + D(L) \varepsilon_t + V_t + D C^*(L) \varepsilon_{T_1} \\ & + D \Gamma C (X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}), \quad t = T_1 + 1, \dots, T, \end{aligned} \quad (14)$$

⁵Different deterministic terms lead to different Granger representations that result in different limits. Different linear combinations of some matrices will have different speed of convergence, which one needs to keep track of. In the end, the differences in the speed of convergence cancel each other out, so the differences are mainly a matter of accounting.

satisfies (12), where the polynomials $C(z)$, $C^*(z)$, and $D(z)$ are convergent for $|z| < 1 + \delta$ and where $(1 + \delta)^t V_t \xrightarrow{p} 0$ for some $\delta > 0$.

Note that the cointegration relations in the second subsample, given by $\beta'_2 X_t = \beta'_2 D(L) \varepsilon_t + \beta'_2 V_t$, are not entirely stationary. However, for large t , $\beta'_2 X_t \approx \beta'_2 D(L) \varepsilon_t$ since $\beta'_2 V_t \xrightarrow{p} 0$ exponentially fast as $t \rightarrow \infty$.

Let $\mathbb{D}^p[0, 1]$ denote the space of CADLAG functions on the unit interval, which take values in \mathbb{R}^p , and let $W(u)$ be a standard p -dimensional Brownian motion, i.e., $\text{var}(W(u)) = u \cdot I_p$. We let $\rho \equiv T_1/T$ denote the proportion of observations in the first subsample and define $\rho^+ = (T_1 + 1)/T$. In the asymptotic analysis, we shall, as T approaches infinity, keep ρ constant, whereby $\lim_{T \rightarrow \infty} \rho^+ = \rho$.

The results of Lemma 5 allow us to define

$$\Sigma^{(j)} \equiv \lim_{T \rightarrow \infty} T^{-1} \sum_{t=T_{j-1}+1}^{T_j} \begin{pmatrix} E(B' Z_{1t} Z'_{1t} B) & E(B' Z_{1t} Z'_{2t}) \\ E(Z_{2t} Z'_{1t} B) & E(Z_{2t} Z'_{2t}) \end{pmatrix}, \quad j = 1, 2,$$

and $\Upsilon = \Sigma^{(1)} \otimes \Omega_1^{-1} + \Sigma^{(2)} \otimes \Omega_2^{-1}$.

Lemma 6 *With the assumptions given above, it holds that.*

$$T^{-1} \sum_{t=T_{j-1}+1}^{T_j} \begin{pmatrix} B' Z_{1t} Z'_{1t} B & B' Z_{1t} Z'_{2t} \\ Z_{2t} Z'_{1t} B & Z_{2t} Z'_{2t} \end{pmatrix} \xrightarrow{p} \Sigma^{(j)}, \quad (15)$$

$$\text{vec} \left(T^{-1/2} \sum_{t=1}^T \Omega(t)^{-1} \varepsilon_t (Z'_{1t} B, Z'_{2t}) \right) \xrightarrow{d} N(0, \Upsilon). \quad (16)$$

For $u \in [0, 1]$ we define $F_T(u) \equiv T^{-1/2} Z_{1[Tu]}$, where $[Tu]$ is the integer part of Tu . We also define

$$F(u) = \begin{pmatrix} F_1(u) \\ F_2(u) \end{pmatrix} \equiv \begin{pmatrix} C \Omega_1^{1/2} W(u) 1_{(u < \rho)} \\ \left[D \Gamma C \Omega_1^{1/2} W(\rho) + D \Omega_2^{1/2} (W(u) - W(\rho)) \right] 1_{(u \geq \rho)} \end{pmatrix},$$

and note that $F_T(u), F(u) \in \mathbb{D}^p[0, 1]$. Finally we define the (stochastic) matrix

$$\begin{aligned} \Xi &\equiv \begin{pmatrix} \alpha'_1 \Omega_1^{-1} \alpha_1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} \int_0^\rho F_1(u) F_1(u)' du & 0 \\ 0 & 0 \end{pmatrix} \\ &+ \begin{pmatrix} 0 & 0 \\ 0 & \alpha'_2 \Omega_2^{-1} \alpha_2 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & \int_\rho^1 F_2(u) F_2(u)' du \end{pmatrix}. \end{aligned}$$

Lemma 7 For a process with the representation (13) and (14) it holds that

$$F_T(u) \xrightarrow{w} F(u), \quad (17)$$

$$T^{-2} \sum_{t=1}^T Z_{1t} Z'_{1t} \xrightarrow{d} \int_0^1 F F' du = \begin{pmatrix} \int_0^\rho F_1 F'_1 du & 0 \\ 0 & \int_\rho^1 F_2 F'_2 du \end{pmatrix}, \quad (18)$$

$$H' T^{-1} \sum_{t=1}^T \text{vec} (Z_{1t} \varepsilon'_t \Omega(t)^{-1} A) \xrightarrow{d} H' \text{vec} \begin{pmatrix} \int_0^\rho F_1 dW' \Omega_1^{-1/2} \alpha_1 & 0 \\ 0 & \int_\rho^1 F_2 dW' \Omega_2^{-1/2} \alpha_2 \end{pmatrix}, \quad (19)$$

where the last expression is a mixed Gaussian distribution. That is, conditional on $F(u)$, the distribution is Gaussian with mean zero and variance $H' \Xi H$.

4.1 Properties of Estimators and Likelihood Ratio Tests

Theorem 8 (Consistency) The maximum likelihood estimators, given by the equations (6)–(8) are consistent for the true parameters.

Theorem 9 (Asymptotic Distribution of Estimators) The asymptotic distribution of \hat{A} and \hat{C} is Gaussian,

$$T^{1/2} [(\hat{A}, \hat{C}) - (A, C)] \xrightarrow{d} N(0, G(G' \Upsilon G)^{-1} G'),$$

and the asymptotic distribution of \hat{B} is mixed Gaussian,

$$T(\hat{B} - B) | F \xrightarrow{d} N\left(0, H (H' \Xi H)^{-1} H'\right),$$

where F is the mixing variable.

Theorem 10 (Asymptotic Distribution of LR Tests) Let M_0 and M_1 be two models defined by restrictions on the form (4) and (5) both satisfying Assumptions 1, 2, 3, and 4, and both having the same cointegration rank in each subsample.

If M_1 is a submodel of M_0 with q fewer parameters, then the asymptotic distribution of the likelihood ratio test of M_1 , tested against M_0 , is χ^2 with q degrees of freedom.

5 Empirical Analysis of the US Term Structure of Interest Rates

In this section we analyze the US term structure using the structural change model we developed in Section 2.

5.1 The Expectations Hypothesis

A version of the term structure of interest rates is that the expected future spot rates equals the future rate plus a time-invariant term premium. We adopt the notation from Campbell, Lo, and Mackinlay (1997) and let $p_{n,t}$ denote the log of the price of a unit-par-value discount bond at date t , with n periods to maturity. The continuously compounded yield to maturity for an n period bond is defined as $y_{n,t} = -\frac{1}{n}p_{n,t}$, and the one-period future rate of return, earned from period $t+n$ to $t+n+1$, (known at time t) is given by $1 + F_{n,t} = P_{n,t}/P_{n+1,t}$, such that $f_{n,t} = \log(1 + F_{n,t}) = p_{n,t} - p_{n+1,t}$.

The expectations hypothesis states that $f_{n,t} = E_t(y_{1,t+n}) + \Lambda_n$, where Λ_n is the term premium.⁶ The restriction imposed by the expectations hypothesis is that the term premium does not depend on t . From the Fisher-Hicks relation $y_{nt} = n^{-1} \sum_{j=0}^{n-1} f_{j,t}$, $n = 1, 2, \dots$, and the identity $E_t(y_{1,t+j}) = \sum_{i=1}^j E_t(\Delta y_{1,t+i}) + y_{1,t}$, we obtain

$$y_{nt} - y_{1t} = n^{-1} \sum_{j=1}^{n-1} \sum_{i=1}^j E_t(\Delta y_{1,t+i}) + L_n, \quad (20)$$

where $L_n = n^{-1} \sum_{j=0}^{n-1} \Lambda_j$. This shows that if y_{1t} is integrated of order one, $I(1)$, such that the terms $\Delta y_{1,t}$ and $n^{-1} \sum_{j=1}^{n-1} \sum_{i=1}^j E_t(\Delta y_{1,t+i})$ are stationary,⁷ then y_{nt} must be integrated of order one and y_{nt} and y_{1t} are cointegrated with cointegration vector $(1, -1)$ as first analyzed by Campbell and Shiller (1987). Since the relationship will hold for any integer n , any pair of yields to maturity will be cointegrated with cointegration vector $(1, -1)$. We shall call this implication the long-run implication of the expectations hypothesis, which is one of several implications of the expectations hypothesis. Equation (20) is the motivation for modeling interest rates as cointegrated processes and illustrates the convenience of using this framework to test the long-run implication.

The implications of the expectations hypothesis are commonly rejected when tested on US term structure data; this is also the case for the long-run implication as concluded by Hall, Anderson, and Granger (1992), Engsted and Tanggaard (1994), Johnson (1994), and Pagan, Hall, and Martin (1996). Hall, Anderson, and Granger (1992) and Engsted and Tanggaard (1994) attributed their rejection to the unstable period for interest rates between September 1979 and October 1982, when the Fed did not target short interest rates directly. This period is also known as the period with the *nonborrowed reserves operating procedure*. Pagan, Hall, and Martin (1996) gave another possible explanation for the rejection. They extended the cointegration model with a parameter, γ , which denotes the elasticity of volatility with respect to the level of the shortest interest rate. With simulations, they showed that tests on cointegration vectors increasingly over-reject as γ is increase, and found the effect to be substantial as γ increases beyond 0.5.

Whereas the expectations hypothesis has been rejected by most studies of US data, see, e.g., Shiller (1990)

⁶For an overview of the expectations hypothesis theory and empirical studies of interest rates, see Shiller (1990).

⁷The stationarity of $E_t(\Delta y_{1,t+j})$ does not hold in general, but will hold for time-homogeneous processes. In particular it will hold for the vector autoregressive process we consider in this paper.

for an overview, the results from studies of the term structure in other countries are mixed. Hardouvelis (1994) rejected the expectations hypothesis in five of the G7 countries. Cuthbertson (1996) found some evidence in favor of the expectations hypothesis using UK interbank rates and Engsted and Tanggaard (1995) found the long-run implication to hold for Danish data in the period where the central bank targeted interest rates.

5.2 Structural Changes in the US Term Structure of Interest Rates

There are several studies that find evidence of a structural change in the US term structure of interest rates. Hamilton (1988) applied a Markov switching model to 3- and 12-month T-bills, and the model detected a period that precisely coincides with the period with the *nonborrowed reserves operating procedure* as a separate regime. H. Hansen and Johansen (1999) have developed a recursive estimation of the cointegrated vector autoregressive model to detect structural changes. Their application to US data also indicates structural changes around the fall of 1979 and the fall of 1982.

Structural changes of US interest rates have also been analyzed within the framework of continuous time models. Chan, Karolyi, Longstaff, and Sanders (1992) estimated a diffusion process for the short term interest rate and rejected a structural change in October 1979, and then estimated the elasticity of volatility to be 1.5. However Bliss and Smith (1998) found significant structural changes when the possibility of a structural change by the end of 1982 is included in the analysis. They found evidence of structural changes in both 1979 as well as in 1982 when the Fed reversed to target the Fed funds rate. After these changes are accounted for, an elasticity as low as 0.5 is consistent with their data. These studies have shown that the US term structure has had structural changes, and it is not surprising that these changes affect point estimates and inference.

Cavanagh, Elliott, and Stock (1995) showed how standard inference can be misleading when there is a root close to unity. Using this local-to-unity approach, Lanne (1999) rejected the expectation hypothesis for US data in the period 1952:1–1991:2. However, after accounting for a structural change in 1979:10 the hypothesis could not be rejected.

In this paper, interest rates are modeled as $I(1)$ variables.⁸ The fact that nominal interest rates cannot be negative and other considerations are strong arguments against interest rates being $I(1)$. Nevertheless, interest rates may very well be $I(1)$ in a particular sample period. Whenever this is the case, modeling interest rates as $I(1)$ is equivalent to invoking asymptotic results to finite samples. The parallel is that the sample in which interest rates behaved as $I(1)$ need to be long enough for asymptotic results of the $I(1)$ model to be valid, and that any constraint that may prevent interest rates from being $I(1)$, is either irrelevant for the particular sample size (conditional on starting values) or not influential enough to severely distort the asymptotic approximation. See Pagan, Hall, and Martin (1996) for another argument on this matter.

⁸Ait-Sahalia (1996) found the short interest rates to behave as an $I(1)$ process within the band [4%, 18%] and a theoretical model, in which interest rates are like $I(1)$ variables, is given by DenHaan (1995).

5.3 Data

The term structure data were extracted from the Bliss data⁹ that are interpolated by the McCulloch cubic-spline method. This is the same technique as the one used to create the widely used data sets from McCulloch (1990) and McCulloch and Kwon (1993). However the Bliss data differs by not being tax adjusted. The data used in the empirical analysis are monthly US zero-coupon yields with maturities of 1, 3, 6, 9, 12, 60, and 84 months¹⁰ within the sample period 1970:1 – 1995:12. Three of the series are presented in Figure 3.

The yields are stacked in the vector X_t , ordered such that the first element in X_t is the 1-month interest rate at time t . We shall allow for structural changes in the adjustment coefficients, $\alpha(t)$, the cointegration parameters, $\beta(t)$, the covariance, $\Omega(t)$, and the deterministic term, $\Phi D_t = \mu(t) = \alpha(t)\rho(t)$, where $\rho(t)$ is an $r \times 1$ vector. So the deterministic term is a constant that has been restricted into the subspace spanned by $\alpha(t)$, to prevent a deterministic trend in the yields. We allow for two structural changes in 1979:10 and in 1982:11. So our model can be written as

$$\Delta X_t = \alpha(t)\beta^*(t)'X_{t-1}^* + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t,$$

where $X_t^* = (X_t', 1)$ and $\beta^*(t)' = (\beta(t)', \rho(t))$.

We normalize the cointegration relations by

$$\beta^*(t) = \begin{pmatrix} \beta_{11}(t) & \beta_{12}(t) & \cdots & \beta_{16}(t) \\ -1 & 0 & & 0 \\ 0 & -1 & & \\ \vdots & & \ddots & \\ 0 & & & -1 \\ \rho_1(t) & \rho_2(t) & \cdots & \rho_6(t) \end{pmatrix}. \quad (21)$$

Since these relations define the stationary relations, the long-run implication of the expectations hypothesis can be formulated as the parameter restrictions $\beta_{11}(t) = \cdots = \beta_{16}(t) = 1$.

The cointegration relations in (21) can be written as

$$b_n(t)y_{1,t} - y_{n,t} + \rho_n(t), \quad n = 3, 6, 9, 12, 60, \text{ and } 84,$$

where $b_n(t) = \beta_{1i}(t)$, and where the values of n correspond to $i = 1, \dots, 6$. The Granger representation shows that $E(b_n(t)y_{1,t} - y_{n,t} + \rho_n(t)) = 0$, so $\hat{\rho}_i(t)$ can be interpreted as the estimated term premium when $b_n(t)$ is

⁹The data were provided to me by David Marshall, see Bekaert, Hodrick, and Marshall (1997). Interested parties are referred to Robert R. Bliss: rbliss@gsbalum.uchicago.edu.

¹⁰Longer maturities were not selected because precise estimates of these are difficult to obtain by interpolation techniques. See Bliss (1997)

set to unity.

5.4 Estimation Results

The lag length was set to two using Akaike's and Hannan-Quinn's information criteria. The cointegration rank is set at six ($r = 6$) as predicted by the expectations hypothesis and as the existing literature has supported, see Hall, Anderson, and Granger (1992), Engsted and Tanggaard (1994), Johnson (1994), and Pagan, Hall, and Martin (1996). No formal test was applied for this selection, because a rank-test for processes with structural changes (in the general form we consider) is currently unavailable.

Table 1 shows that the covariance matrix clearly differs between the three subsamples. The variance estimates from the three subsamples are given in Table 2.

It is not surprising that the variance of interest rates were much higher in the 1979–1982 subsample when the Fed did not target interest rates directly. One conclusion from Table 1 is that the difference between the variance of interest rates in the first and third subsample is significant. From Table 2 it can be seen that the major difference between the covariance matrix in the first and last subsample is the reduced volatility of the interest rates with shorter maturities. This phenomenon may be explained by the less frequent adjustments of the Fed's target of the Fed's fund rate in the most recent sample, along with fact that the Fed now publicly announces what their target is. The numbers in italic font in Table 2 are the correlation coefficients, which are similar for the three subsamples, although we have not applied a formal test to analyze this question.

Six models with different parameter restrictions were estimated.¹¹ Model 1 in Table 3 with unrestricted parameters is the most general model, which can be represented by the equation

$$\Delta X_t = \alpha(t) [\beta(t)' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

where $\varepsilon_t \sim N(0, \Omega(t))$ and where the parameters are constant within each subsample, i.e., $\alpha(t) = \alpha_1$ for $t \leq 1979:09$, $\alpha(t) = \alpha_2$ for $1979:10 \leq t \leq 1982:10$ and $\alpha(t) = \alpha_3$ for $t \geq 1982:11$, and similarly for $\beta(t)$, $\rho(t)$ and $\Omega(t)$. The long-run implication of the expectations hypothesis requires $b_n = 1$ for $n = 3, 6, 9, 12, 60$, and 84. The point estimates differ from unity by being systematically too small in the two first subsamples and too large in the last subsample.

In Model 2 the long-run implication of the expectations hypothesis is in all subsamples, whereas the term premium, ρ_n , the adjustment coefficients, α_j , and the covariance matrices, Ω_j , may differ across the subsamples, $j = 1, 2, 3$. This model can be written as

$$\Delta X_t = \alpha(t) [\beta' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

¹¹The empirical analysis was performed in Gauss.

where $\varepsilon_t \sim N(0, \Omega(t))$, and where β has the structure required by the long-run implication. The likelihood ratio test of Model 2 against Model 1, has a p -value of 4.95%. This shows that there is not strong evidence against the long-run implication once structural changes in the parameters are accounted for.

Model 3 is a more parsimonious model where in addition to the restrictions in Model 2, the adjustment coefficients are required to span the same subspace, $\alpha(t) = \alpha \cdot \phi(t)$, where $\phi(t)$ is a full rank $r \times r$ matrix. This model can be written as

$$\Delta X_t = \alpha \phi(t) [\beta' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

where $\varepsilon_t \sim N(0, \Omega(t))$. The restriction implies that the orthogonal complement to $\alpha(t)$ is constant, i.e., $\alpha_\perp(t) = \alpha_\perp$. The strength of the adjustments in the three subsamples is expressed in terms of the matrix $\phi(t)$.

Recall the Granger representation from equations (13) and (14), extended with a third subsample,

$$\begin{aligned} X_t &= C \sum_{i=1}^t \varepsilon_i + O_p(1), \quad t = 1, \dots, T_1, \\ X_t &= D \sum_{i=T_1+1}^t \varepsilon_i + D\Gamma C \sum_{i=1}^{T_1} \varepsilon_i + O_p(1), \quad t = T_1 + 1, \dots, T_2, \\ X_t &= E \sum_{i=T_2+1}^t \varepsilon_i + E\Gamma D \sum_{i=T_1+1}^{T_2} \varepsilon_i + E\Gamma D\Gamma C \sum_{i=1}^{T_1} \varepsilon_i + O_p(1), \quad t = T_2 + 1, \dots, T. \end{aligned}$$

An implication of the constancy of α_\perp and β and Γ_1 is that the loading matrix is constant, i.e. $C = D = E = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$. This simplifies the Granger representation to a single equation given by

$$X_t = C \sum_{i=1}^t \varepsilon_i + O_p(1), \quad t = 1, \dots, T,$$

using the fact that $C\Gamma C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp \Gamma \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp = C$.

The term $\alpha'_\perp \sum_{i=1}^t \varepsilon_i$ is called the common stochastic trend in X_t , because it describes the random walk element of X_t , and $C\bar{\alpha}_\perp$ defines how the stochastic trend is loaded into the process X_t , (note that $C\bar{\alpha}_\perp \alpha'_\perp = C$). Thus the non-rejection of Model 3 (a p -value of 10.38% when tested against Model 1) can be interpreted as follows: The long-run implication is consistent with the data and we cannot reject that the common stochastic trend has been a constant linear combination of ε_t , although the variance of the innovations, $\Omega(t)$, has been non-constant. Also, we cannot reject that the loading of the common stochastic trend has been constant.

The last model in Table 3, Model 4, can be expressed as

$$\Delta X_t = \alpha [\beta' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

where $\varepsilon_t \sim N(0, \Omega(t))$. In this model the adjustment coefficients have the same strength in the three subsamples. This is equivalent to Model 3, with the additional restriction that $\phi(t) = \phi$. This model is clearly inconsistent with the term structure data. The fact that the strength of the adjustments are non-constant is not surprising, since changes in volatility and term premium, are likely to affect the strength of the adjustments.

Thus, we find the term structure to have had structural changes in the covariance $\Omega(t)$, the term premium $\rho(t)$, and in the strength of the adjustments. However, the parameters that define fundamentals, such as the common stochastic trend and the cointegration relations between interest rates, have been stable.

These findings are consistent with many of the suggestions that have been offered to explain the rejection of the expectations hypothesis. The monetary changes in the fall of 1979 and the fall of 1982 had an important impact on the stochastic properties of interest rates. If the structural changes are not accounted for, it distorts inference and can cause a rejection of a true hypothesis, as was suggested by Hall, Anderson, and Granger (1992) and Engsted and Tanggaard (1994). The suggestion by Tzavalis and Wickens (1997) of a time varying term premium is also consistent with our results, since we find $\rho(t)$ to vary as the volatility of interest rates changes. In fact, the spread between long and short interest rates is smaller in the 1979–1982 period where the “short” interest rates were relatively more volatile than “long” interest rates. As pointed out by a referee, this may be explained by agents expecting that the unusually high level of short interest rates would not persist. Finally, we found $\Omega(t)$ to be non-constant. If the variance is modeled to be constant, this is likely to distort inference, similar to the distortion that Pagan, Hall, and Martin (1996) observed in Monte Carlo experiments, where the variance depended on the level of the short interest rate.

The fifth and sixth models in Table 4 replicate previous empirical studies of the US term structure, by having constant parameters. Model 5 is the unrestricted model (with constant parameters) and Model 6 is the submodel in which the long-run implication of the expectations hypothesis is imposed. A test of Model 6 against Model 5 would have lead to a weak rejection of the expectations hypothesis, exactly as previous studies have concluded. Of course, this inference is invalid because Model 5 is inconsistent with the data. The LR test statistic of Model 5 against Model 1 is 1166 and is therefore clearly rejected.

6 Conclusion

We have shown how structural changes in cointegrated processes can be formulated in a unified framework, using the familiar vector autoregressive model. It is possible to formulate and test various structural changes as simple parameter restrictions in this framework, and the parameters can be estimated with the generalized reduced rank regression technique, which is introduced in this paper. This technique is also applicable to estimation problems unrelated to structural changes, see P. R. Hansen (2002) for several examples.

We derived the likelihood ratio test for structural changes occurring at known points in time and showed

how hypotheses can be tested, when the maintained hypothesis is presence of structural changes. Moreover, we derived the asymptotic distributions of the estimators and LR tests that involve nested models, which agree on the cointegration rank. As it is the case for the standard model without structural changes, the estimator of the cointegration parameters is super-consistent and asymptotically mixed Gaussian, and the LR statistic is asymptotically χ^2 .

The combination of cointegration and structural changes may provide a fruitful framework for many economic questions of interest. In this paper, we analyzed the US term structure and found evidence of structural changes that coincide with the Fed's policy changes in September 1979 and October 1982. Contrary to previous studies (see Hall, Anderson, and Granger (1992), Engsted and Tanggaard (1994), or Pagan, Hall, and Martin (1996)), we cannot reject the long-run implication of the expectations hypothesis, once these structural changes are accounted for. In fact, a parsimonious model is consistent with the data. This model has a different covariance, term premium, and strength of adjustments in the three monetary regimes, whereas the cointegration relations and other parameters are stable over the sample period.

In this paper, we took the cointegration rank and the timing of the change points as given. Although this is not an alarming assumption in the empirical analysis of interest rates, this will not be the case in most application. However, it should be possible to use the asymptotic results in the appendix of this paper to: (1) derive a formal test to determine the rank of cointegrated processes with structural changes, (see P. R. Hansen (2000b, chapter 4) for some preliminary results), and (2) to generalized this framework to the case with unknown changes points.

A Appendix of Intermediate Results and Proofs

We introduce the notation, $Z_0 \equiv (Z_{01}, \dots, Z_{0T})$, $Z_1 \equiv (Z_{11}, \dots, Z_{1T})$, $Z_2 \equiv (Z_{21}, \dots, Z_{2T})$, and $E \equiv (\varepsilon_1, \dots, \varepsilon_T)$, so that we can express (3) as

$$Z_0 = AB'Z_1 + CZ_2 + E. \quad (22)$$

We also define $\mathbf{Z}_B \equiv ((Z_1' B, Z_2') \otimes I_p)$ and $\mathbf{Z}_{AC} \equiv (Z_1' \otimes A)K_{p_1, r}$, where $K_{p_1, r}$ is the commutation matrix, uniquely defined by $K_{p_1, r} \text{vec}(B) \equiv \text{vec}(B')$ for any $p_1 \times r$ matrix B . Thus $K_{p_1, r}$ is a $p_1 r \times p_1 r$ matrix consisting of zeros and ones. Finally, we let $\varepsilon \equiv \text{vec}(\varepsilon_1, \dots, \varepsilon_T) = \text{vec}(E)$ and set

$$\Sigma \equiv \text{var}(\varepsilon) = \text{diag}(I_{T_1} \otimes \Omega_1, I_{T_2 - T_1} \otimes \Omega_2, \dots, I_{T - T_{m-1}} \otimes \Omega_m),$$

which is block diagonal. Hence Σ^{-1} is also a block diagonal matrix with Ω_j^{-1} , $j = 1, \dots, m$ as diagonal matrices.

Lemma 11 *With the definitions above, we have the identities*

$$\mathbf{Z}'_{AC}\Sigma^{-1}\mathbf{Z}_{AC} = \sum_{t=1}^T [A'\Omega(t)^{-1}A \otimes Z_{1t}Z'_{1t}], \quad (23)$$

$$\mathbf{Z}'_{AC}\Sigma^{-1}\text{vec}(Z_0 - CZ_2) = \sum_{t=1}^T \text{vec}(Z_{1t}(Z_{0t} - CZ_{2t})'\Omega(t)^{-1}A), \quad (24)$$

$$\mathbf{Z}'_B\Sigma^{-1}\mathbf{Z}_B = \sum_{t=1}^T \left[\begin{pmatrix} B'Z_{1t}Z'_{1t}B & B'Z_{1t}Z'_{2t} \\ Z_{2t}Z'_{1t}B & Z_{2t}Z'_{2t} \end{pmatrix} \otimes \Omega(t)^{-1} \right], \quad (25)$$

$$\mathbf{Z}'_B\Sigma^{-1}\text{vec}(Z_0) = \sum_{t=1}^T \text{vec}(\Omega(t)^{-1}Z_{0t}(Z'_{1t}B, Z'_{2t})), \quad (26)$$

If $\{\varepsilon_t\}$ is iid Gaussian with covariance matrix Ω , the expressions simplify to

$$\begin{aligned} \mathbf{Z}'_{AC}\Sigma^{-1}\mathbf{Z}_{AC} &= T [A'\Omega^{-1}A \otimes M_{11}], \\ \mathbf{Z}'_{AC}\Sigma^{-1}\text{vec}(Z_0 - CZ_2) &= T \text{vec}((M_{10} - M_{02}C')\Omega^{-1}A), \\ \mathbf{Z}'_B\Sigma^{-1}\mathbf{Z}_B &= T \left[\begin{pmatrix} B'M_{11}B & B'M_{12} \\ M_{21}B & M_{22} \end{pmatrix} \otimes \Omega^{-1} \right], \\ \mathbf{Z}'_B\Sigma^{-1}\text{vec}(Z_0) &= T \text{vec}(\Omega^{-1}(M_{01}B, M_{02})). \end{aligned}$$

Proof. The calculations

$$\begin{aligned} \mathbf{Z}'_{AC}\Sigma^{-1}\mathbf{Z}_{AC} &= K'_{p_1,r}(Z_1 \otimes A')\Sigma^{-1}(Z'_1 \otimes A)K_{p_1,r} \\ &= K_{r,p_1} \sum_{t=1}^T (Z_{1t} \otimes A')\Omega(t)^{-1}(Z'_{1t} \otimes A)K_{p_1,r} \\ &= K_{r,p_1} \sum_{t=1}^T (Z_{1t} \otimes A'\Omega(t)^{-1})(Z'_{1t} \otimes A)K_{p_1,r} \\ &= K_{r,p_1} \sum_{t=1}^T (Z_{1t}Z'_{1t} \otimes A'\Omega(t)^{-1}A)K_{p_1,r} \\ &= \sum_{t=1}^T (A'\Omega(t)^{-1}A \otimes Z_{1t}Z'_{1t}). \end{aligned}$$

proves (23), where we used that $Z_{1t} = \text{vec}(Z_{1t})$ and that we can write $\Omega(t)^{-1} = 1 \otimes \Omega(t)^{-1}$. Next,

$$\begin{aligned}
\mathbf{Z}'_{AC} \Sigma^{-1} \text{vec}(Z_0 - CZ_2) &= K_{r,p_1} \sum_{t=1}^T (Z_{1t} \otimes A') \Omega(t)^{-1} (Z_{0t} - CZ_{2t}) \\
&= K_{r,p_1} \sum_{t,\tau=1}^T (Z_{1t} \otimes A' \Omega(t)^{-1}) \text{vec}(Z_{0t} - CZ_{2t}) \\
&= K_{r,p_1} \sum_{t,\tau=1}^T \text{vec}(A' \Omega(t)^{-1} (Z_{0t} - CZ_{2t}) Z'_{1t}) \\
&= \sum_{t=1}^T \text{vec}(Z_{1t} (Z_{0t} - CZ_{2t})' \Omega(t)^{-1} A).
\end{aligned}$$

proves (24), and (25) and (26) are proven similarly.

In the situation where $\{\varepsilon_t\}$ is iid, we have $\Omega(t)^{-1} = \Omega^{-1}$, which proves the last four identities. ■

Proof of Theorem 2. Applying the vec operation to equation (22) yields the equation

$$\begin{aligned}
\text{vec}(Z_0) &= (Z'_1 B \otimes I_p) \text{vec}(A) + (Z'_2 \otimes I_p) \text{vec}(C) + \varepsilon \\
&= [(Z'_1 B, Z'_2) \otimes I_p] \text{vec}(A, C) + \varepsilon = \mathbf{Z}_B G \psi + \varepsilon.
\end{aligned}$$

For fixed values of B and Σ this is a restricted GLS problem with the well-known solution given by

$$\text{vec}(\hat{A}, \hat{C}) = G [G' \mathbf{Z}'_B \Sigma^{-1} \mathbf{Z}_B G]^{-1} G' \mathbf{Z}'_B \Sigma^{-1} \text{vec}(Z_0),$$

which by Lemma 11 simplifies to (6). Similarly, for fixed A , C , and Σ , we have the equation

$$\begin{aligned}
\text{vec}(Z_0 - CZ_2) &= \text{vec}(AB' Z_1) + \varepsilon = (Z'_1 \otimes A) \text{vec}(B') + \varepsilon \\
&= (Z'_1 \otimes A) K_{p_1, r} \text{vec}(B) + \varepsilon = \mathbf{Z}_A \text{vec}(B) + \varepsilon,
\end{aligned}$$

which is also a restricted GLS problem. Its solution is given by

$$\text{vec}(\hat{B}) = H [H' \mathbf{Z}'_{AC} \Sigma^{-1} \mathbf{Z}_{AC} H]^{-1} H' \mathbf{Z}'_{AC} \Sigma^{-1} \text{vec}(Z_0 - CZ_2),$$

which by Lemma 11 reduces to (7). ■

Proof of Corollary 3. Follows from Theorem 2 and Lemma 11. ■

Proof of Corollary 4. From the results for the standard model, see Johansen (1996), we obtain the expressions for \hat{C} and $\hat{\Omega}$. Rather than handling the remaining estimation for A and B as a GLS problem we can obtain the

likelihood equations directly. The concentrated log-likelihood function is (apart from a constant) given by

$$\log L(A, B) = -\frac{T}{2} \text{tr} \{ \Omega^{-1} (S_{00} - AB'S_{10} + AB'S_{11}BA' - S_{01}BA') \}.$$

So the differentials for A and B , in the directions a and b are given by

$$\begin{aligned} D_A \log L(A, B)(a) &= T \text{tr} \{ \Omega^{-1} (S_{01} - AB'S_{11}) Ba' \} \\ &= T [\text{tr} \{ \Omega^{-1} S_{01} Ba' \} - \text{tr} \{ \Omega^{-1} A(B'S_{11}B)a' \}] \\ &= T \text{vec}(a)' [(B' \otimes \Omega^{-1}) \text{vec}(S_{01}) - (B'S_{11}B \otimes \Omega^{-1}) \text{vec}(A)], \end{aligned}$$

and

$$\begin{aligned} D_B \log L(A, B)(b) &= T \text{tr} \{ \Omega^{-1} (S_{01} - AB'S_{11}) bA' \} \\ &= T \text{tr} \{ A' \Omega^{-1} (S_{01} - AB'S_{11}) b \} \\ &= T \text{vec}(b)' [(A' \otimes S_{10}) \text{vec}(\Omega^{-1}) - (A' \Omega^{-1} A \otimes S_{11}) \text{vec}(B)], \end{aligned}$$

where we applied theorem 3 from Magnus and Neudecker (1988, chapter 2). So (9) and (10) are the first order conditions. When A is unrestricted ($G = I$) we see that (10) simplifies to

$$\begin{aligned} \text{vec}(\hat{A}) &= \left[\hat{B}' S_{11} \hat{B} \otimes \hat{\Omega}^{-1} \right]^{-1} \left(\hat{B}' \otimes \hat{\Omega}^{-1} \right) \text{vec}(S_{01}) \\ &= \left(\left[\hat{B}' S_{11} \hat{B} \right]^{-1} \hat{B}' \otimes I_p \right) \text{vec}(S_{01}) = \text{vec}(S_{01} \hat{B} \left[\hat{B}' S_{11} \hat{B} \right]^{-1}) \end{aligned}$$

as claimed, and the simplification of (11) follows directly. ■

Proof of Lemma 5. The representation for the first subsample is given from Granger's representation theorem for the standard model without structural changes, see P. R. Hansen (2000a).¹² The representation of the second subsample is given by

$$X_t = D \sum_{s=T_1+1}^t \varepsilon_s + D(L)\varepsilon_t + V_t + D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i}), \quad t = T_1 + 1, \dots, T,$$

where $D(L)\varepsilon_t + V_t = \sum_{s=0}^{t-1} D_s \varepsilon_{t-s} + D_t X_{T_1}$. In the derivation of the Granger representation, one divides the initial values, (X_0, \dots, X_{-k+1}) , into the “stationary” linear combinations that can be assigned their stationary distribution, $D(L)\varepsilon_t$, and the term $D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i})$. However, as pointed out by a referee, we are not

¹²The original Granger representation for cointegrated VAR processes of Johansen (1991), does not provide a closed-form expression for the initial value (given by $C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i})$) which is important for our analysis.

free to choose a distribution for the “stationary” part of the initial values. The first subsample will result in a value of X_{T_1} , which need not satisfy $D_t X_{T_1} \neq \sum_{s=0}^{\infty} D_{t+s} \varepsilon_{T_1-s}$, which is required for the term to be assigned its stationary distribution stationary. However, the deviation from stationarity is given by $V_t = \sum_{s=0}^{\infty} D_{t+s} \varepsilon_{T_1-s} - D_t X_{T_1}$, which converges to zero in probability exponentially fast, so it play no role in the asymptotic analysis.

In order to get the representation in the appropriate form we need to modify the second representation such that the initial value depends only on X_t , $t = 0, -1, \dots$, rather than $X_{T_1}, \dots, X_{T_1-k+1}$. This is obtained by substituting the appropriate Granger representations for $X_{T_1}, \dots, X_{T_1-k+1}$, by which we obtain an expression that simplifies to

$$D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i}) = D\Gamma C \sum_{s=1}^{T_1} \varepsilon_s + DC^*(L)\varepsilon_{T_1} + D\Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}),$$

where

$$C^*(L)\varepsilon_{T_1} = (I - (I - \Gamma_1)C)\varepsilon_{T_1} + (C_1 - \Gamma_1 C_0)\varepsilon_{T_1-1} + (C_2 - \Gamma_1 C_1)\varepsilon_{T_1-2} + \dots$$

is a stationary process. ■

Proof of Lemma 6. (15) follows by a law of large numbers, and (16) follows from

$$\text{vec}(\Omega(t)^{-1} \varepsilon_t (Z'_{1t} B, Z'_{2t})) = \left[\begin{pmatrix} B' Z_{1t} \\ Z_{2t} \end{pmatrix} \otimes \Omega(t)^{-1/2} \right] \Omega(t)^{-1/2} \varepsilon_t,$$

$\text{var}(\Omega(t)^{-1/2} \varepsilon_t) = I_p$, and

$$\begin{aligned} T^{-1} \sum_{t=1}^T \left[\begin{pmatrix} B' Z_{1t} \\ Z_{2t} \end{pmatrix} \otimes \Omega(t)^{-1/2} \right] \left[\begin{pmatrix} B' Z_{1t} \\ Z_{2t} \end{pmatrix} \otimes \Omega(t)^{-1/2} \right]' \\ = T^{-1} \sum_{t=1}^T \begin{pmatrix} B' Z_{1t} Z'_{1t} B & B' Z_{1t} Z'_{2t} \\ Z_{2t} Z'_{1t} B & Z_{2t} Z'_{2t} \end{pmatrix} \otimes \Omega(t)^{-1} \xrightarrow{p} \Upsilon, \end{aligned}$$

where we applied (15). ■

Proof of Lemma 7. Define $W_T(u) \equiv T^{-\frac{1}{2}} \sum_{t=1}^{[Tu]} \Omega(t)^{-1/2} \varepsilon_t$. Then by Donsker’s invariance principle we have that $W_T(u) \xrightarrow{w} W(u)$, for $u \in [0, 1]$. The mapping $\mathcal{V}(u) \mapsto \mathcal{V}(u) 1_{\{a \leq u < b\}}$ is Lipschitz and therefore continuous, for any $\mathcal{V} \in \mathbb{D}^p[0, 1]$ and any $a, b \in [0, 1]$.¹³ Consider the random walk terms, of Z_{1t} , in the Granger representation.

¹³This is easy to verify using the sup-norm, since

$$\sup_{0 \leq u \leq 1} |\mathcal{U}(u) - \mathcal{V}(u)| \geq \sup_{a \leq u < b} |\mathcal{U}(u) - \mathcal{V}(u)| = 1 \cdot \sup_{0 \leq u \leq 1} |1_{\{a \leq u < b\}} \mathcal{U}(u) - 1_{\{a \leq u < b\}} \mathcal{V}(u)|,$$

shows that the mapping is Lipschitz, with Lipschitz constant $c = 1$.

For $(u < \rho^+)$ we have that

$$T^{-\frac{1}{2}}C \sum_{t=1}^{[Tu]} \varepsilon_t = C\Omega_1^{1/2}W_T(u)1_{\{u < \rho^+\}} \xrightarrow{w} C\Omega_1^{1/2}W(u)1_{\{u < \rho\}},$$

where we use the continuity of $W(u)$. Similarly, for $u \geq \rho^+$ we have that

$$\begin{aligned} T^{-\frac{1}{2}} \left(D\Gamma C \sum_{t=1}^{T_1} \varepsilon_t + D \sum_{t=T_1+1}^{[Tu]} \varepsilon_t \right) &= D\Gamma C\Omega_1^{1/2}W_T(\rho) + D\Omega_2^{1/2}(W_T(u) - W_T(\rho))1_{\{u \geq \rho^+\}} \\ &\xrightarrow{w} D\Gamma C\Omega_1^{1/2}W(\rho) + D\Omega_2^{1/2}(W(u) - W(\rho))1_{\{u \geq \rho\}}. \end{aligned}$$

Since all other terms in (13) and (14) vanish in probability, we have shown (17), and (18) follows from the continuous mapping theorem.

Next, we see that the Brownian motions F_1 and $\alpha'_1\Omega_1^{-1/2}W$ are independent on $0 \leq u \leq \rho$, since

$$\text{cov}(C\Omega_1^{1/2}W(u_1), \alpha'_1\Omega_1^{-1/2}W(u_2)) = \beta_{1\perp}(\alpha'_{1\perp}\Gamma\beta_{1\perp})^{-1}\alpha'_{1\perp} \min(u_1, u_2)I\alpha_1 = 0,$$

for all $u_1, u_2 \in [0, \rho]$. So

$$T^{-1} \sum_{t=1}^{T_1} X_{t-1} \varepsilon'_t \Omega_1^{-1} \alpha_1 \xrightarrow{d} \int_0^\rho F_1 dW' \Omega_1^{-1/2} \alpha_1,$$

and similarly it follows that

$$T^{-1} \sum_{t=T_1+1}^T X_{t-1} \varepsilon'_t \Omega_2^{-1} \alpha_2 \xrightarrow{d} \int_\rho^1 F_2 dW' \Omega_2^{-1/2} \alpha_2.$$

Since other terms are multiplied by zeroes in H (due to the structure of B), we have shown (19).

To show that the expressions have a mixed Gaussian distribution, we note that $\alpha'_2\Omega_2^{-1/2}dW$ is also independent of F_1 , and $\alpha'_1\Omega_1^{-1/2}dW$ of F_2 , since for $u_1 \in [0, \rho]$ and $u_2 \in [\rho, 1]$ we have that

$$\begin{aligned} E \left[F_1(u_1) (W(u_2) - W(\rho))' \Omega_2^{-1/2} \alpha_2 \right] &= E \left[C\Omega_1^{1/2}W(u_1)(W(u_2) - W(\rho))' \Omega_2^{-1/2} \alpha_2 \right] = 0 \\ E \left[F_2(u_1) W(u_1)' \Omega_1^{-1/2} \alpha_1 \right] &= E \left[(D\Gamma C\Omega_1^{1/2}W(\rho)W(u_1)' \Omega_1^{-1/2} \alpha_1) \right. \\ &\quad \left. + E \left[D\Omega_2^{1/2}(W(u_2) - W(\rho))W(u_1) \Omega_1^{-1/2} \alpha_1 \right] \right] \\ &= D\Gamma C u_1 I \alpha_1 + 0 = 0. \end{aligned}$$

In general, for two independent Brownian motions, W_1 and W_2 , the conditional distribution of $\int_a^b W_1 dW_2'$ given W_1 is Gaussian with mean zero and variance $\int_a^b W_1 W_1' du \otimes \text{var}(W_2(1))$. Hence, given F the stochastic integral is Gaussian with the variance stated in the lemma. ■

Proof of Theorem 8. Rather than proving this directly, we can make use of existing results. Within each of the subsamples one could estimate the unrestricted parameters: α_j , β_j , Ψ_j , and Ω_j , $j = 1, \dots, m$. It follows from Johansen (1988, lemma 13.1, 13.2), that these estimators are consistent for the true parameters (assuming a normalization of β_j). The estimators from the m subsamples can be combined into the estimators \hat{A}_u , \hat{B}_u , \hat{C}_u , and $\hat{\Omega}_u(t)$ that are consistent for the population parameters, A , B , C , and $\Omega_{j=1, \dots, m}$. These estimators need not satisfy all the restrictions imposed by G , H , and h , as is the case for the maximum likelihood estimator. However, the consistency is not affected by imposing valid restrictions, and it follows that the maximum likelihood estimators \hat{A} , \hat{B} , \hat{C} , and $\hat{\Omega}_{j=1, \dots, m}$ are consistent for the population parameters. ■

Proof of Theorem 9. Define $\hat{M}_B = T^{-2} \sum_{t=1}^T [\hat{A}' \hat{\Omega}(t)^{-1} \hat{A} \otimes Z_{1t} Z_{1t}']$, then we have that $\hat{M}_B \xrightarrow{d} M_B$, where $H' M_B H = H' \Xi H$. Consider the expansion of the likelihood equation for \hat{B} ,

$$\begin{aligned} \text{vec}(\hat{B}) &= H [H' \hat{M}_B H]^{-1} H' \left[T^{-2} \sum_{t=1}^T \text{vec} \left(Z_{1t} (Z_{0t} - \hat{C} Z_{2t})' \hat{\Omega}(t)^{-1} \hat{A} \right) - \hat{M}_B h \right] + h \\ &= H [H' \hat{M}_B H]^{-1} H' \left[\hat{M}_B \text{vec}(B) + o_p(T^{-1}) + T^{-2} \sum_{t=1}^T \text{vec} \left(Z_{1t} \varepsilon_t' \hat{\Omega}(t)^{-1} \hat{A} \right) - \hat{M}_B h \right] + h \\ &= \text{vec}(B) + H [H' \Xi H]^{-1} H' T^{-2} \sum_{t=1}^T \text{vec} \left(Z_{1t} \varepsilon_t' \hat{\Omega}(t)^{-1} \hat{A} \right) + o_p(T^{-1}), \end{aligned}$$

where we used the consistency of \hat{A} , \hat{C} , and $\hat{\Omega}_{j=1,2}$. Hence

$$T \text{vec}(\hat{\theta} - \theta) \xrightarrow{w} [H' \Xi H]^{-1} H' \text{vec} \left(\int F(u) dW(u)' \Omega^{-1/2} A \right),$$

and by Lemma 7 the result for \hat{B} follows. Similarly, the expansion

$$\begin{aligned} T^{1/2} \text{vec}(\hat{A} - A, \hat{C} - C) &= G \left[G' \left[T^{-1} \sum_{t=1}^T \begin{pmatrix} \hat{B}' Z_{1t} Z_{1t}' \hat{B} & \hat{B}' Z_{1t} Z_{2t}' \\ Z_{2t} Z_{1t}' \hat{B} & Z_{2t} Z_{2t}' \end{pmatrix} \otimes \hat{\Omega}(t)^{-1} \right] G \right]^{-1} \\ &\quad \times G' \text{vec} \left(T^{-1/2} \sum_{t=1}^T \hat{\Omega}(t)^{-1} \varepsilon_t (Z_{1t}' \hat{B}, Z_{2t}') \right), \end{aligned}$$

shows the result for (A, C) , using Lemma 6, the consistency of $\hat{\Omega}_{j=1,2}$, and the super-consistency of \hat{B} . ■

Lemma 12 Define

$$\begin{aligned} P_{AC} &\equiv \Sigma^{-1/2} \mathbf{Z}_B G [G' \mathbf{Z}_B' \Sigma^{-1} \mathbf{Z}_B G]^{-1} G' \mathbf{Z}_B' \Sigma^{-1/2}, \\ P_B &\equiv \Sigma^{-1/2} \mathbf{Z}_{AC} H [H' \mathbf{Z}_{AC}' \Sigma^{-1} \mathbf{Z}_{AC} H]^{-1} H' \mathbf{Z}_{AC}' \Sigma^{-1/2}. \end{aligned}$$

Then P_{AC} and P_B are orthogonal projection matrices that satisfy $P_{AC} P_B = o_p(1)$.

Proof. That they are orthogonal projection matrices follows from $P = PP$ and $P = P'$, and it holds that $a_T \equiv G' \mathbf{Z}'_B \Sigma^{-1} \mathbf{Z}_B G = o_p(T)$, $b_T \equiv H' \mathbf{Z}'_{AC} \Sigma^{-1} \mathbf{Z}_{AC} H = O_p(T^2)$, and $c_T \equiv G' \mathbf{Z}'_B \Sigma^{-1/2} \mathbf{Z}_{AC} H = O_p(T)$, such that

$$\begin{aligned} \|P_{AC} P_B\| &= \text{tr} \{ (P_{AC} P_B)' (P_{AC} P_B) \} = \text{tr} \{ b_T^{-1} c_T a_T^{-1} a_T a_T^{-1} c_T b_T^{-1} b_T \} \\ &= O_p(T^{-2}) O_p(T) o_p(T^{-1}) O_p(T) = o_p(T^{-1}), \end{aligned}$$

which completes the proof. ■

Proof of Theorem 10. From Theorem 2 we have that the maximum value of the log-likelihood function is given by

$$\log L_{\max} \equiv \log L(\hat{A}, \hat{B}, \hat{C}, \hat{\Omega}_{j=1, \dots, m}) \propto -\frac{T}{2} \left[\sum_{j=1}^m \rho_j \log |\hat{\Omega}_j| \right],$$

where $\hat{\Omega}_j = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} \hat{\varepsilon}_t \hat{\varepsilon}_t'$ and where $\rho_j = (T_j - T_{j-1})/T$ denotes the proportion of observations in subsample j relative to the full sample. We have the identity:

$$\log |\hat{\Omega}_j| = \log |\Omega_j| + \log \left| I + \Omega_j^{-1} (\hat{\Omega}_j - \Omega_j) \right|,$$

which we can Taylor-expand using $\log |I + \delta| = \text{tr}\{\delta\} - \frac{1}{2} \text{tr}\{\delta^2\} + O(\|\delta\|^3)$, where $\delta = \Omega_j^{-1} (\hat{\Omega}_j - \Omega_j)$. Below we show that $(\hat{\Omega}_j - \Omega_j) = O_p(T^{-1/2})$, so the last term is $O_p(T^{-3/2})$ and vanishes in probability. The second term is treated below. Next, we define $\hat{E}_{(j)} \equiv (\hat{\varepsilon}_{T_{j-1}+1}, \dots, \varepsilon_{T_j})$ and $\hat{\varepsilon}_{(j)} = \text{vec}(\hat{E}_{(j)})$, and see that the first term in the expansion (after multiplying with $T\rho_j$ and adding the constant Tp) can be expressed as

$$\begin{aligned} \rho_j T \text{tr} \left\{ \Omega_j^{-1} (\hat{\Omega}_j - \Omega_j) \right\} + Tp &= \rho_j T \text{tr} \left\{ \Omega_j^{-1} \hat{\Omega}_j \right\} = \text{tr} \left\{ \Omega_j^{-1} \hat{E}_{(j)} \hat{E}_{(j)}' \right\} \\ &= \hat{\varepsilon}_{(j)}' [I \otimes \Omega_j^{-1}] \hat{\varepsilon}_{(j)}. \end{aligned}$$

If we add up the terms from each of the subsample we get $\hat{\varepsilon}' \Sigma \hat{\varepsilon}$, and we can express $\hat{\varepsilon}_t$ as

$$\begin{aligned} \hat{\varepsilon}_t &= \varepsilon_t + (AB' - \hat{A}\hat{B}') Z_{1t} + (C - \hat{C}) Z_{2t} \\ &= \varepsilon_t + \left[(A, C) - (\hat{A}, \hat{C}) \right] \begin{pmatrix} B' Z_{1t} \\ Z_{2t} \end{pmatrix} + A(B - \hat{B})' Z_{1t} + (\hat{A} - A)(B - \hat{B})' Z_{1t}. \end{aligned}$$

Since the last term, $(\hat{A} - A)(B - \hat{B})' Z_{1t} = O_p(T^{-1/2}) O_p(T^{-1}) O_p(T) = o_p(1)$, we find after some algebra that

$$\hat{\varepsilon} = \varepsilon + \mathbf{Z}_B \text{vec} \left[(A, C) - (\hat{A}, \hat{C}) \right] + \mathbf{Z}_{AC} \text{vec}(B - \hat{B}) + o_p(1),$$

and since

$$\begin{aligned}\text{vec} \left[(A, C) - (\hat{A}, \hat{C}) \right] &= G \left[G' \mathbf{Z}'_B \Sigma^{-1} \mathbf{Z}_B G \right]^{-1} G' \mathbf{Z}'_B \Sigma^{-1} \varepsilon, \\ \text{vec}(B - \hat{B}) &= H \left[H' \mathbf{Z}'_{AC} \Sigma^{-1} \mathbf{Z}_{AC} H \right]^{-1} H' \mathbf{Z}'_{AC} \Sigma^{-1} \varepsilon + o_p(T^{-1}),\end{aligned}$$

we find that

$$\hat{\varepsilon}' \Sigma \hat{\varepsilon} = \eta' (I_T - P_{AC} - P_B) \eta + o_p(1),$$

where $\eta = \Sigma^{-1/2} \varepsilon \sim N(0, I_T)$. Now let G_0 and H_0 be the restrictions matrices for M_0 , and G_1 and H_1 for M_1 , and denote the corresponding projection matrices by $P_{AC,0}$, $P_{AC,1}$, $P_{B,0}$, and $P_{B,1}$. Since M_1 is nested in M_0 , it holds that $\text{sp}(G_1) \subset \text{sp}(G_0)$ and $\text{sp}(H_1) \subset \text{sp}(H_0)$, and the rank of $P_{AC,0} - P_{AC,1} + P_{B,0} - P_{B,1}$ equals $q \equiv \dim(\text{sp}(G_0)) + \dim(\text{sp}(H_0)) - \dim(\text{sp}(G_1)) - \dim(\text{sp}(H_1))$, where $\text{sp}(D)$ denotes the space spanned by the columns of a matrix, D , and $\dim(\text{sp}(D))$ denotes the dimension of this space. Finally, the difference between the two second order terms in the Taylor expansion is proportional to $T \text{tr} \left\{ \left(\Omega_j^{-1} (\hat{\Omega}_1 - \hat{\Omega}_0) \right)^2 \right\} = O_p(T^{-1})$, and by a standard orthogonality argument we find that $-2 \log L_{\max,1} / L_{\max,0} = \eta' Q \eta + o_p(1) \xrightarrow{d} \chi^2_{(q)}$. ■

Corollary 13 *The information matrix is block diagonal with three blocks corresponding to the three sets of parameters, (A, C) , B , and $\Omega_{j=1, \dots, m}$.*

Proof. From the linear model it is well known that the sum-of-squared residuals covariance estimator is asymptotically uncorrelated of $(\hat{A}, \hat{B}, \hat{C})$, and the asymptotic independence of (\hat{A}, \hat{C}) and \hat{B} follows from $P_{AC} P_B = o_p(1)$. ■

Acknowledgements

This paper has benefitted from many valuable comments from James D. Hamilton and Søren Johansen, Tom Engsted, Graham Elliott, Niels Haldrup, David Hendry, Juan Toro, and three anonymous referees. I also thank participants of seminars at the European University Institute, Brown University, UCSD, and the 2000 World Congress of the Econometric Society in Seattle. I am responsible for all remaining errors. Financial support from the Danish Social Science Research Council and the Danish Research Academy is gratefully acknowledged.

References

- Ait-Sahalia, Y., 1996. Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies*, 9, 385–426.
- Anderson, T. W., 1951. Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *Annals of Mathematical Statistics*, 22, 327–351.

- Anderson, T. W., 1984. An Introduction to Multivariate Statistical Analysis, 2nd edn. John Wiley and Sons, New York.
- Andrews, D. W. K., and W. Ploberger, 1994. Optimal Tests When a Nuisance Parameter is Present Only under the Alternative. *Econometrica*, 62, 1383–1414.
- Bai, J., 1999. Likelihood Ratio Test for Multiple Structural Changes. *Journal of Econometrics*, 91, 299–323.
- Bai, J., and P. Perron, 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66, 47–78.
- Bekaert, G., R. J. Hodrick, and D. A. Marshall, 1997. On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates. *Journal of Financial Economics*, 44, 309–348.
- Bliss, R. R., 1997. Testing Term Structure Estimation Methods. *Advances in Futures and Options Research*, 9, 197–231.
- Bliss, R. R., and D. C. Smith, 1998. The Elasticity of Interest Rate Volatility: Chan, Karolyi, Longstaff and Sanders Revisited. *Journal of Risk*, 1, 21–46.
- Boswijk, P. H., 1995. Identifiability of Cointegrated Systems. Working paper, Tinbergen Institute.
- Campbell, J. Y., A. W. Lo, and A. C. Mackinlay, 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, New Jersey.
- Campbell, J. Y., and R. J. Shiller, 1987. Cointegration and Tests of Present Value Models. *Journal of Political Economy*, 95, 1062–1088.
- Cavanagh, C. L., G. Elliott, and J. H. Stock, 1995. Inference in Models with Nearly Nonstationary Regressors. *Econometric Theory*, 11, 1131–1147.
- Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992. An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance*, 47, 1209–1227.
- Cuthbertson, K., 1996. The Expectations Hypothesis of the Term Structure: The UK Interbank Market. *Economic Journal*, 106, 578–92.
- DenHaan, W. J., 1995. The Term Structure of Interest Rates in Real and Monetary Economies. *Journal of Economic Dynamics and Control*, 19, 909–940.
- Elliott, G., 1997. Minimum Distance Methods and Cointegrated Models. UCSD manuscript.
- Elliott, G., 2000. Estimating Restricted Cointegrating Vectors. *Journal of Business and Economic Statistics*, 18, 91–99.
- Engsted, T., and C. Tanggaard, 1994. Cointegration and the US Term Structure. *Journal of Banking and Finance*, 18, 167–181.
- Engsted, T., and C. Tanggaard, 1995. The Predictive Power of Yield Spreads for Future Interest Rates: Evidence from the Danish Term Structure. *Scandinavian Journal of Economics*, 97, 149–59.
- Gregory, A. W., and B. E. Hansen, 1996. Residual-Based Test for Cointegration in Models with Regime Shifts. *Journal of Econometrics*, 70, 99–126.
- Hall, A. D., H. M. Anderson, and C. W. J. Granger, 1992. A Cointegration Analysis of Treasury Bill Yields. *The Review of Economic And Statistics*, 74, 116–126.
- Hamilton, J. D., 1988. Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates. *Journal of Economic Dynamics and Control*, 12, 385–423.
- Hansen, B. E., 1992. Test for Instability with $I(1)$ Processes. *Journal of Business and Economic Statistics*, 10, 321–335.

- Hansen, H., and S. Johansen, 1999. Some Tests for Parameter Constancy in Cointegrated VAR-Models. *Econometrics Journal*, 2, 306–333.
- Hansen, P. R., 2000a. The Granger Representation Theorem: A Closed-Form Expression for $I(1)$ Processes. Brown University Economics Working Paper.
- Hansen, P. R., 2000b. Structural Changes in Cointegrated Processes. Ph.D. thesis, University of California at San Diego.
- Hansen, P. R., 2002. Generalized Reduced Rank Regression. Brown University, Economics Working Paper 2002-02.
- Hardouvelis, G., 1994. The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries. *Journal of Monetary Economics*, 33, 255–83.
- Hendry, D. F., 1995. A Theory of Co-Breaking. Working Paper, Institute of Economics and Statistics, Oxford.
- Inoue, A., 1999. Test for Cointegrating Rank with a Trend-Break. *Journal of Econometrics*, 90, 215–237.
- Izenman, A. J., 1975. Reduced-Rank Regression for the Multivariate Linear Model. *Journal of multivariate analysis*, 5, 248–264.
- Johansen, S., 1988. Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12, 231–254.
- Johansen, S., 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59, 1551–1580.
- Johansen, S., 1996. Likelihood Based Inference in Cointegrated Vector Autoregressive Models, 2nd edn. Oxford University Press, Oxford.
- Johansen, S., and K. Juselius, 1992. Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK. *Journal of Econometrics*, 53, 211–244.
- Johnson, P. A., 1994. On the Number of Common Unit Roots in the Term Structure of Interest Rates. *Applied Economics*, 26, 815–820.
- Krolzig, H.-M., 1996. Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifts. SFB 373 Discussion Paper 25/1996, Humboldt-University Berlin.
- Lanne, M., 1999. Near Unit Roots and the Predictive Power of Yield Spreads for Changes in Long-Term Interest Rates. *The Review of Economics and Statistics*, 81, 393–398.
- Magnus, J. R., and H. Neudecker, 1988. Matrix Differential Calculus with Applications in Statistics and Economics. Wiley, New York.
- McCulloch, J. H., 1990. US Term Structure Data. 1946-1987. In: Friedman, B., and F. Hahn (Eds.), *Handbook of Monetary Economics*. Vol. 1, North-Holland, Amsterdam, pp. 672–715.
- McCulloch, J. H., and H. C. Kwon, 1993. US Term Structure Data. 1947-1991. Ohio State University, Working Paper 93-6.
- Nyblom, J., 1989. Testing for the Constancy of Parameters over Time. *Journal of the American Statistical Association*, 84, 223–230.
- Oberhofer, W., and J. Kmenta, 1974. A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models. *Econometrica*, 42, 579–590.
- Pagan, A. R., A. D. Hall, and V. Martin, 1996. Modelling the Term Structure. *Handbook of Statistics*, 14, 91–118.
- Reinsel, G. C., and R. P. Velu, 1998. Multivariate Reduced-Rank Regression. Springer, New York.

- Seo, B., 1998. Tests for Structural Change in Cointegrated Systems. *Econometric Theory*, 14, 222–259.
- Shiller, R. J., 1990. The Term Structure of Interest Rates. In: Friedman, B., and F. Hahn (Eds.), *Handbook of Monetary Economics*. Vol. 1, North-Holland, Amsterdam, pp. 627–722.
- Tzavalis, E., and M. Wickens, 1997. Explaining the Failures of the Term Spread Models of the Rational Expectations Hypothesis of the Term Structure. *Journal of Money, Credit and Banking*, 29, 364–80.

| $\Delta X_t - \alpha(t)\beta^*(t)'X_{t-1}^* - \Gamma_1\Delta X_{t-1} \sim N(0, \Omega(t))$ | | | | |
|--|---|--------------------|--|--|
| Model | $\max \log L(\alpha(t), \beta^*(t), \Gamma_1, \Omega(t))$ | Degrees of freedom | LR(M _i M ₀) (p-value) | |
| M ₀ : $\Omega(t)$ | 2009.25 | 295 | — | |
| M ₁ : $\Omega_1 = \Omega_3$ | 1824.94 | 270 | 368.61 (0.0000) | |
| M ₂ : $\Omega_1 = \Omega_2 = \Omega_3$ | 1631.77 | 239 | 754.96 (0.0000) | |

Table 1: The maximum value of the likelihood function for different specification of the covariance matrix, $\Omega(t)$.

| The Estimated Covariance Matrices, $\Omega(t)$ | | | | | | | |
|--|--------------------|--|--|--|--|--|--|
| 1970:3–1979:9 | $\hat{\Omega}_1 =$ | $\begin{pmatrix} 0.30 & 0.28 & 0.25 & 0.22 & 0.14 & 0.10 & 0.09 \\ 0.98 & 0.27 & 0.25 & 0.22 & 0.15 & 0.11 & 0.10 \\ 0.91 & 0.96 & 0.25 & 0.23 & 0.17 & 0.12 & 0.11 \\ 0.84 & 0.88 & 0.96 & 0.23 & 0.17 & 0.13 & 0.11 \\ 0.66 & 0.75 & 0.88 & 0.92 & 0.15 & 0.12 & 0.11 \\ 0.58 & 0.67 & 0.76 & 0.86 & 0.98 & 0.10 & 0.09 \\ 0.58 & 0.68 & 0.78 & 0.81 & 0.99 & 0.99 & 0.08 \end{pmatrix}$ | | | | | |
| 1979:10–1982:10 | $\hat{\Omega}_2 =$ | $\begin{pmatrix} 1.75 & 1.68 & 1.51 & 1.28 & 0.92 & 0.63 & 0.54 \\ 0.97 & 1.70 & 1.58 & 1.33 & 0.97 & 0.68 & 0.59 \\ 0.93 & 0.99 & 1.50 & 1.30 & 0.97 & 0.69 & 0.61 \\ 0.89 & 0.94 & 0.98 & 1.18 & 0.90 & 0.65 & 0.57 \\ 0.82 & 0.88 & 0.93 & 0.98 & 0.72 & 0.54 & 0.48 \\ 0.73 & 0.80 & 0.86 & 0.91 & 0.97 & 0.43 & 0.39 \\ 0.69 & 0.76 & 0.84 & 0.89 & 0.96 & 0.99 & 0.35 \end{pmatrix}$ | | | | | |
| 1982:11–1995:12 | $\hat{\Omega}_3 =$ | $\begin{pmatrix} 0.10 & 0.09 & 0.08 & 0.07 & 0.07 & 0.06 & 0.05 \\ 0.95 & 0.09 & 0.09 & 0.09 & 0.09 & 0.08 & 0.07 \\ 0.80 & 0.95 & 0.10 & 0.11 & 0.11 & 0.10 & 0.09 \\ 0.64 & 0.87 & 0.99 & 0.12 & 0.12 & 0.11 & 0.11 \\ 0.61 & 0.83 & 0.96 & 0.96 & 0.13 & 0.13 & 0.12 \\ 0.53 & 0.74 & 0.88 & 0.88 & 0.99 & 0.13 & 0.13 \\ 0.46 & 0.67 & 0.82 & 0.92 & 0.96 & 0.99 & 0.12 \end{pmatrix}$ | | | | | |

Table 2: The estimated covariance matrices, $\hat{\Omega}_j$, $j = 1, 2, 3$, from the most general change model. The variances are in the diagonal, the covariances in the upper triangle, and the correlations are displayed with italic font in the lower triangle.

| Model 1. $\alpha(t), \beta(t), \rho(t), \Omega(t)$ | | | | $2 \log L$ 4018.49 | df 295 | LR – | p -value – |
|---|----------|---------|---------|-----------------------|-------------|--------------|----------------------|
| | | $n = 3$ | $n = 6$ | $n = 9$ | $n = 12$ | $n = 60$ | $n = 84$ |
| 1970:3–1979:9 | b_n | 0.9831 | 0.9767 | 0.9162 | 0.7473 | 0.6154 | 0.5947 |
| | ρ_n | 0.3634 | 0.6356 | 1.1666 | 2.4113 | 3.4517 | 3.6640 |
| 1979:10–1982:10 | b_n | 0.9234 | 0.8455 | 0.7716 | 0.7378 | 0.7179 | 0.6765 |
| | ρ_n | 1.4726 | 2.5655 | 3.5156 | 3.8391 | 3.9931 | 4.4702 |
| 1982:11–1995:12 | b_n | 1.0746 | 1.1391 | 1.2596 | 1.5328 | 1.7390 | 1.7989 |
| | ρ_n | -0.2384 | -0.4607 | -0.9011 | -2.0401 | -2.8354 | -3.0585 |
| Model 2. $\alpha(t), \beta(t) = \beta_0, \rho(t), \Omega(t)$ | | | | $2 \log L$ 3989.58 | df 277 | LR 28.91 | p -value 0.0495 |
| | | $n = 3$ | $n = 6$ | $n = 9$ | $n = 12$ | $n = 60$ | $n = 84$ |
| 1970:3–1979:9 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.2620 | 0.4935 | 0.6592 | 0.8935 | 1.1475 | 1.2357 |
| 1979:10–1982:10 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.6309 | 0.8628 | 0.9917 | 0.9370 | 0.8637 | 0.8800 |
| 1982:11–1995:12 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.2106 | 0.3694 | 0.6307 | 1.0520 | 1.3919 | 1.5010 |
| Model 3. $\alpha(t) = \alpha\phi(t), \beta(t) = \beta_0, \rho(t), \Omega(t)$ | | | | $2 \log L$ 3978.44 | df 265 | LR 40.05 | p -value 0.1038 |
| | | $n = 3$ | $n = 6$ | $n = 9$ | $n = 12$ | $n = 60$ | $n = 84$ |
| 1970:3–1979:9 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.2644 | 0.4999 | 0.6748 | 0.9221 | 1.1861 | 1.2753 |
| 1979:10–1982:10 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.6529 | 0.9089 | 1.0495 | 0.9896 | 0.9065 | 0.9281 |
| 1982:11–1995:12 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.2123 | 0.3753 | 0.6523 | 1.1248 | 1.5229 | 1.6487 |
| Model 4. $\alpha(t) = \alpha, \beta(t) = \beta_0, \rho(t), \Omega(t)$ | | | | $2 \log L$ 3784.01 | df 199 | LR 234.48 | p -value 0.0000 |
| | | $n = 3$ | $n = 6$ | $n = 9$ | $n = 12$ | $n = 60$ | $n = 84$ |
| 1970:3–1979:9 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.2701 | 0.5061 | 0.6798 | 0.9381 | 1.2343 | 1.3332 |
| 1979:10–1982:10 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.5850 | 0.8015 | 0.9598 | 1.2261 | 1.4309 | 1.5107 |
| 1982:11–1995:12 | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.2182 | 0.3826 | 0.6621 | 1.1599 | 1.5995 | 1.7405 |

Table 3: Estimation results: For each model we report the maximum value of the likelihood function, the model's degrees of freedom (df) and the LR statistic (tested against the most general model) with the correspondings p -value. $\beta(t) = \beta$ denotes constant cointegration relations and $\beta(t) = \beta_0$ denotes constant cointegration relations that satisfy the expectations hypothesis. The cointegration parameters b_n and term premia ρ_n from the cointegration relations $b_n y_{1,t} - y_{n,t} + \rho_n$ are reported for each model and subsample.

| | | | | | | | |
|--|----------|---------|---------|--------------------|-------------|-------------|----------------------|
| Model 5. $\alpha(t) = \alpha, \beta(t) = \beta, \rho(t) = \rho, \Omega(t) = \Omega$ | | | | $2 \log L$ 2852 | df 131 | LR – | p -value – |
| 1970:3–1995:12 | | $n = 3$ | $n = 6$ | $n = 9$ | $n = 12$ | $n = 60$ | $n = 84$ |
| | b_n | 1.0390 | 1.0417 | 1.0520 | 1.0529 | 1.0239 | 1.0191 |
| | ρ_n | 0.0011 | 0.1680 | 0.2951 | 0.6209 | 1.1478 | 1.2875 |
| Model 6. $\alpha(t) = \alpha, \beta(t) = \beta_0, \rho(t) = \rho, \Omega(t) = \Omega$ | | | | $2 \log L$ 2825 | df 125 | LR 26.84 | p -value 0.0002 |
| 1970:3–1995:12 | | $n = 3$ | $n = 6$ | $n = 9$ | $n = 12$ | $n = 60$ | $n = 84$ |
| | b_n | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | ρ_n | 0.2719 | 0.4570 | 0.6561 | 0.9888 | 1.3148 | 1.4215 |

Table 4: Estimation results. Testing the expectations hypothesis in the cointegrated VAR without structural changes. Note that the p -value is invalid because Model 5 is clearly rejected when tested against Model 1.

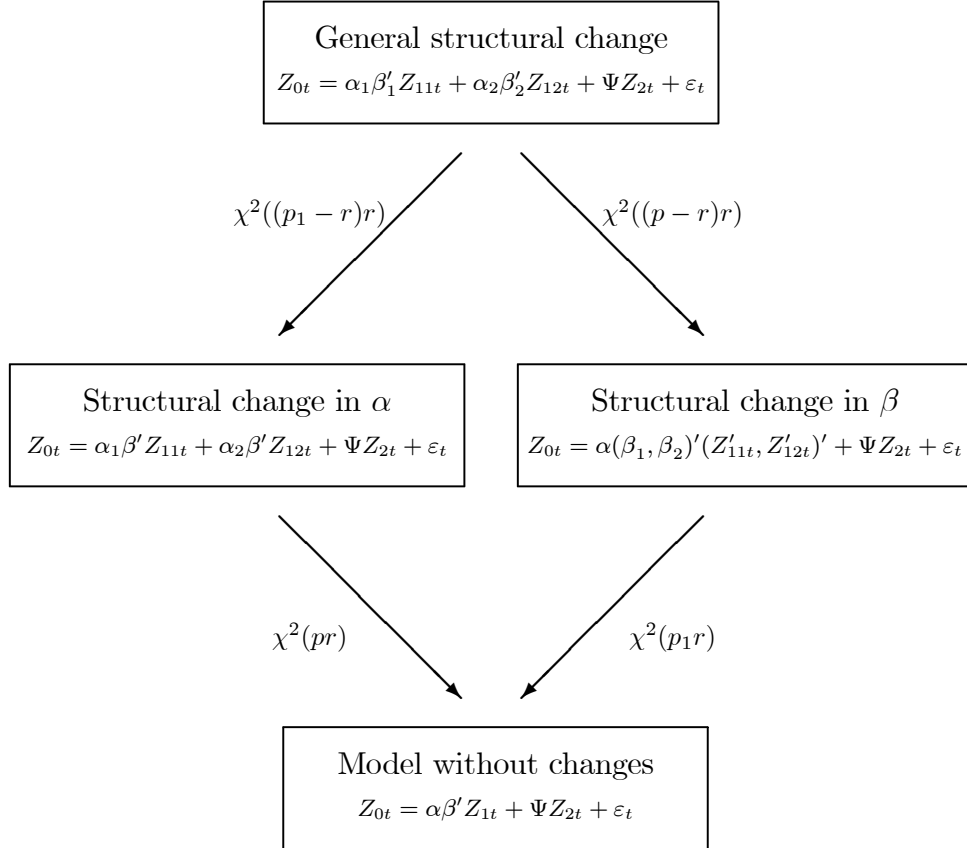


Figure 1: The relations between the different types of structural changes in the cointegration parameters, in the situation with a single structural change. The asymptotic distribution of the LR tests are χ^2 in all cases, with the degrees of freedom reported in the brackets. The LR test of constant parameters against changes in both α and β is asymptotically χ^2 with $(p + p_1 - r)r$ degrees of freedom.

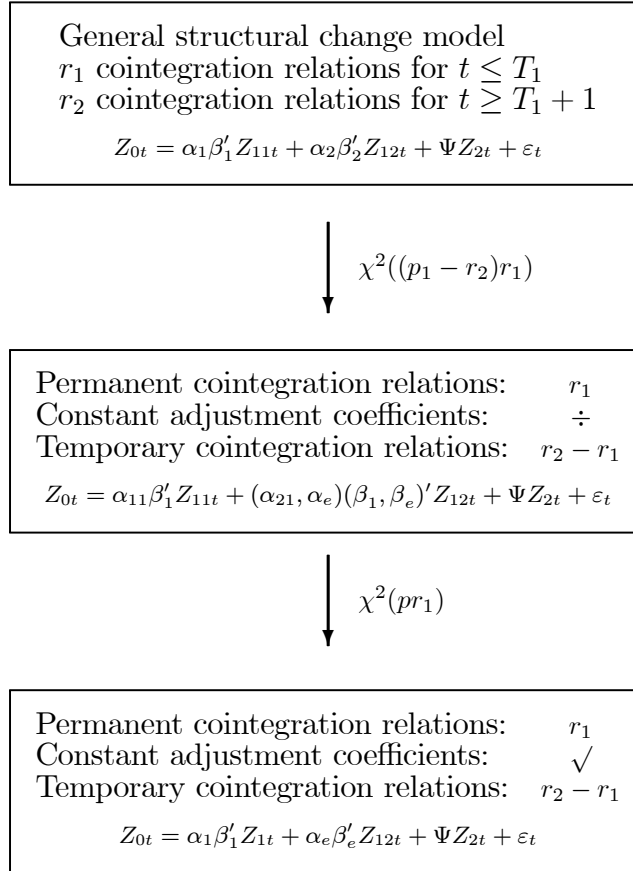


Figure 2: The relations between the different models with structural changes and a change in the number of cointegration relations. The asymptotic distribution of the LR test statistic between two of these models is χ^2 with the $(p_1 - r_2)r_1$, pr_1 , or $(p + r_1 - r_2)r_2$ degrees of freedom.

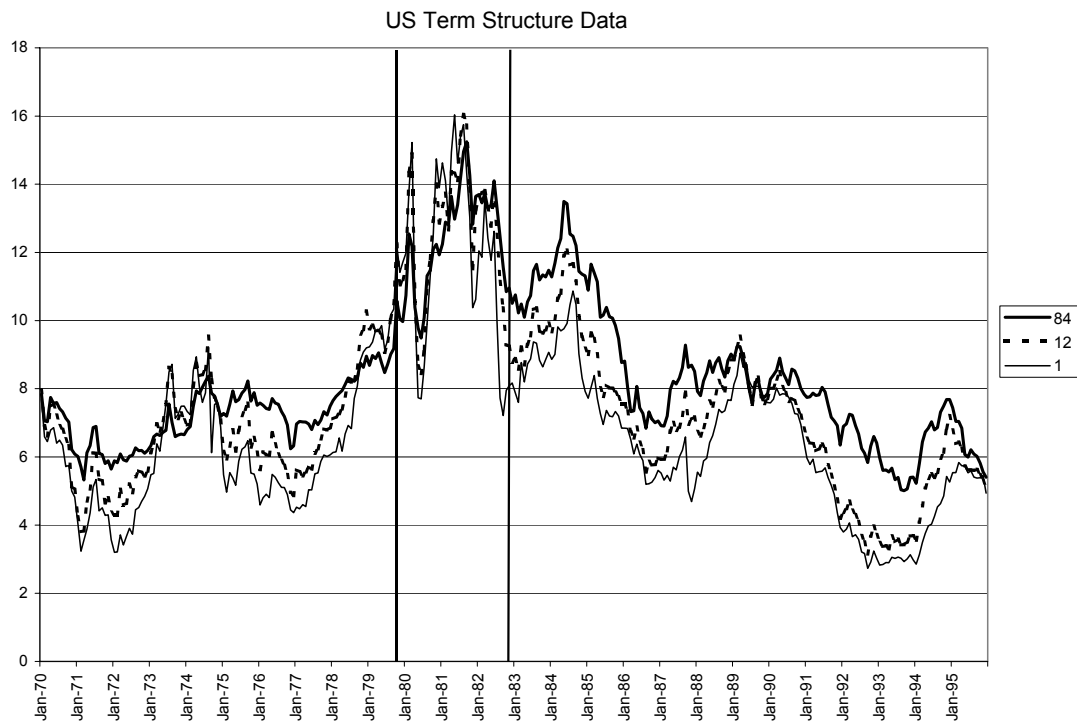


Figure 3: Monthly data of US zero-coupon yields with maturities 1, 12, and 84 month. The vertical lines show the location of the (possible) change points.