

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

Working Paper No. 00-20

October, 2000

Peter Reinhard Hansen

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

JEL Classification: C3, C32

Keywords: Structural Change, Cointegration, Vector Autoregression

## Abstract

This paper generalizes the cointegrated vector autoregressive model of Johansen (1988) to allow for structural changes. Estimation under various hypotheses is made possible by a new estimation technique, that makes it simple to derive a number of interesting likelihood ratio tests. E.g., the test for  $m$  structural changes against  $m + k$  structural changes (occurring at fixed points in time),  $m \in \mathbb{N}_0$ ,  $k \in \mathbb{N}$ , and test of linear parameter restrictions when the null hypothesis allows for structural changes. The asymptotic distribution is  $\chi^2$  in both cases.

The model is applied to US term structure data, and structural changes in September 1979 and October 1982 – points in time with large changes in the Fed’s policy – are found to be significant. After accounting for these structural changes, I cannot, contrary to previous studies, reject the long-run implications of the expectations hypothesis.

# 1 Introduction

The modelling 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 modelling 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 cointegrating 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. The most related paper is the one of Seo (1998), who considered structural changes in cointegration relations and adjustment coefficients, under i.i.d. assumptions. The framework proposed here can handle a class of changes in integrated processes that are more general than previously treated. Partial structural changes<sup>1</sup> such as a structural change in a particular cointegration relation or a change in its mean can be handled, leaving other relations unchanged. In addition, the framework is applicable under weaker assumptions than the i.i.d. assumption. The test statistic invoked in this paper is the likelihood ratio (LR) test and it is shown that its asymptotic distribution is standard  $\chi^2$  when the change points are taken as given<sup>2</sup>. Another contribution of this paper is that it enables hypotheses testing under the maintained hypothesis that the underlying process exhibits structural changes. The asymptotic  $\chi^2$  results remain valid in this situation.

Another main 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 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,  $\beta$ , and adjustment coefficients,  $\alpha$ . Johansen and Juselius (1992) proposed a switching algorithm

---

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

<sup>2</sup>The case of an unknown change point leads to a non-standard asymptotic distribution. See Andrews and Ploberger (1994), Seo (1998), and Hansen (2000c, Chapter 3).

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, by which it achieves a generality similar to the minimum distance approach by Elliott (1997, 1998a), since the generalization to nonlinear restrictions expressed by functions that are well-behaved<sup>3</sup> is straightforward. Secondly, the GRRR technique allows for a general covariance structure and is therefore applicable to models with heteroskedasticity.

The result of this paper is applied to the US term structure of interest rates. The results are that the long-run implications 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 various structural changes in the cointegration model. 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 give some of the details of the cointegrated vector autoregressive model by Johansen (1988). The model is generalized to allow for various structural changes and it is shown how these changes can be formulated as parameter restrictions in a unified framework.

### 2.1 The Cointegrated Vector Autoregressive Model

I take the  $p$ -dimensional vector autoregressive model  $X_t = \Pi_1 X_{t-1} + \dots + \Pi_k X_{t-k} + \Phi D_t + \varepsilon_t$  as my point of origin, where  $\varepsilon_t$  is assumed to be independent and Gaussian distributed with mean zero and variance  $\Omega$ . The variable  $D_t$  contains deterministic terms such as a constant,

---

<sup>3</sup>Functions that are continuously differentiable.

a linear trend and seasonal dummies. The error correction form for the model is

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t,$$

and it is well known that if the characteristic polynomial, here given by  $A(z) = I(1 - z) - \Pi z - \sum_{i=1}^{k-1} \Gamma_i (1 - z)z^i$ , has all its roots outside the unit-circle, then  $X_t$  is stationary. If the polynomial has one or more unit roots, then  $X_t$  is an integrated process as defined by Johansen (1996). A unit root implies that  $\Pi$  has reduced rank  $r < p$  and if the number of unit roots equals  $p - r$ , then the process  $X_t$  is integrated of order one, denoted  $I(1)$ . When  $\Pi$  has reduced rank, it can be written as a product of two  $p \times r$  matrices  $\Pi = \alpha\beta'$ , such that the model can be expressed in the form

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

This process can be inverted to an infinite moving average representation, also known as the Granger representation, see P. R. Hansen (2000b). The representation shows (i) how the adjustment coefficient,  $\alpha$ , relates to the common stochastic trends in the process and (ii) that  $\beta$  defines the cointegration relations.

It is convenient to rewrite the model as

$$Z_{0t} = \alpha\beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t, \quad (2)$$

where  $Z_{0t} = \Delta X_t$ ,  $Z_{1t} = X_{t-1}$ ,  $Z_{2t} = (\Delta X'_{t-1}, \dots, \Delta X'_{t-k+1}, D'_t)'$  and  $\Psi = (\Gamma_1, \dots, \Gamma_{k-1}, \Phi)$ , so we separate the regressors with reduced rank parameters from the regressors with unrestricted parameters. In some situations we want to add variables to the cointegration space, such as exogenous variables or simply a linear trend or a constant. In such cases we redefine  $Z_{1t}$  to include these variables and denotes its dimension by  $p_1$  rather than  $p$ , which denotes the dimension of  $Z_{0t}$ . The regression problem in (2), with no additional restrictions on the parameters, is referred to as a reduced rank regression (RRR).

We define a generalized reduced rank regression, as the following regression problem:

$$\begin{aligned} Z_{0t} &= AB'Z_{1t} + CZ_{2t} + \varepsilon_t, \\ \text{s.t. } \text{vec}(A, C) &= G\psi, \\ \text{vec}(B) &= H\varphi, \end{aligned} \tag{3}$$

where  $G$  and  $H$  are known matrices with full column rank, and  $\{\varepsilon_t\}$  obeys the following assumption.

**Assumption 1**  $\{\varepsilon_t\}$  is a sequence of independent  $p$ -dimensional Gaussian variables, where  $\varepsilon_t$  is independent of  $Z_{1t}$  and  $Z_{2t}$  and has the marginal distribution  $N(0, \Omega(t))$ .

By this formulation the i.i.d. assumption on  $\{\varepsilon_t\}$  is relaxed, by no longer requiring an identical distribution. In this paper we assume that  $\Omega(t)$  is piecewise constant, i.e.,  $\Omega(t) = \Omega_j$ ,  $t = T_{j-1} + 1, \dots, T_j$ , where  $0 = T_1 \leq \dots \leq T_q = T$ . The assumption still implies independence of  $\{\varepsilon_t\}$ . Estimation and inference under a weaker assumption than Assumption 1, including general forms of heteroskedasticity and autocorrelation, is treated in P. R. Hansen (2000a).

Obviously, the estimation problems that can be solved by a RRR can also be solved by a GRRR, by setting  $G$  and  $H$  as identity matrices, and with  $\Omega(t) = \Omega$ .

As shown by Boswijk (1995), the following assumption is necessary for generic identification of the parameters.

**Assumption 2** The matrices  $H$  and  $G$  in (3) have full column rank and 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, ( $n$  denotes the number of column in  $(H, G)$ ).

Let the covariance parameters be expressed as  $\Omega_j = \Omega_j(\theta)$ ,  $\theta \in \Theta_\theta$ ,  $j = 1, \dots, q$ . This formulation does not necessarily impose any restrictions on the parameters.

**Assumption 3** The parameters  $\psi$ ,  $\varphi$ , and  $\theta$  are variation free, that is

$$(\psi, \varphi, \theta) \in \Theta_\psi \times \Theta_\varphi \times \Theta_\theta.$$

This assumption is convenient for the parameter estimation. Suppose that Assumption

3 holds, and 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  $(\psi^{(0)}, \varphi^{(0)}, \theta^{(0)})$ . This procedure has the nice property that the value of the likelihood function is increased in every iteration; the ordering of the three parameters is irrelevant. 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. So whenever the normal equations uniquely define the global maximum of  $L$ , maximum likelihood estimation is achieved with this procedure. The information matrix is asymptotically block diagonal which eliminates existence of local maxima asymptotically. However, in practise local maxima may exist in finite samples, so one should start the algorithm with different initial values of the parameters, and see if the algorithm converges to the same value of the likelihood function.

We need to calculate the degrees of freedom in the parameter  $\alpha(t)\beta(t)'$ . The following lemma, taken from Johansen (1996), is useful for this purpose.

**Lemma 4** *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  $(p + p_1 - r)r$ .

So, in the case of a reduced rank regression, with  $x = \alpha$  and  $y = \beta$ , the parameter space in which  $\Pi = \alpha\beta'$  can vary has dimension  $(p + p_1 - r)r$ .

## 2.2 Structural Changes in the Cointegrated Vector Autoregressive Model

We now show that structural changes in model (1) can be viewed as a particular form of (3). Without loss of generality, we can focus just on changes in  $\alpha$  and  $\beta$ , because changes in the parameters  $\Gamma_1, \dots, \Gamma_{k-1}$  or  $\Phi$  in (1) are easily handled by redefining  $Z_{2t}$  and  $\Psi$ . For now we keep the covariance matrix,  $\Omega$ , constant, but later we also generalize the model to allow for structural changes in this parameter. Estimation, when all parameters change their value is easily done by estimating each subsample with the RRR technique, however in most applications it is desirable to keep some parameters fixed to avoid that the dimension of the parameter space increases too dramatically.

So, the generalization of model (1) that we consider is

$$Z_{0t} = \alpha(t)\beta(t)'Z_{1t} + \Psi Z_{2t} + \varepsilon_t.$$

We shall consider different choices of the time-dependent parameters  $\alpha(t)$  and  $\beta(t)$ . More specifically, we consider various situations where  $\alpha(t)$  and  $\beta(t)$  are piecewise constant, which can be expressed as

$$\alpha(t)\beta(t)' = \alpha_1\beta_1'1_{1t} + \dots + \alpha_q\beta_q'1_{qt}$$

where  $1_{jt}$ ,  $j = 1, \dots, q$  are indicator functions that determine which  $\alpha_j$  and  $\beta_j$  are active. This formulation does not require  $\alpha_i$  and  $\alpha_j$  to have the same number of columns  $i \neq j$ , as long as  $\alpha_j$  and  $\beta_j$  have the same number of columns. So the formulation allows for changes in the number of cointegration relations as well as scenarios where some relations are constant over several sub-samples while other relations change.

By defining  $Z_{1jt} = 1_{jt}Z_{1t}$ ,  $j = 1, \dots, q$ , and  $\tilde{Z}_{1t} = (Z'_{11t}, \dots, Z'_{1qt})'$ , we obtain the regression problem

$$Z_{0t} = (\alpha_1, \dots, \alpha_q) \begin{pmatrix} \beta_1 & 0 & \dots & 0 & 0 \\ 0 & \beta_2 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \beta_{q-1} & 0 \\ 0 & 0 & \dots & 0 & \beta_q \end{pmatrix}' \tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t,$$



with a block diagonal structure of the matrix containing the cointegration relations, denoted by  $B$ . This structure can be expressed as a linear restriction on  $\text{vec}(B) = H\varphi$ , and the regression is therefore a special case of (3).

### 2.2.1 Structural Changes in $\alpha$ and $\beta$

Consider a situation with  $q - 1$  structural changes that occur at time  $T_1, \dots, T_{q-1}$ , so that  $\alpha_t$  and  $\beta_t$  can take on  $q$  different values. This can be formulated as

$$\beta_t = \begin{cases} \beta_1 & t = 1, \dots, T_1 \\ \beta_2 & t = T_1 + 1, \dots, T_2 \\ \vdots & \\ \beta_q & t = T_{q-1} + 1, \dots, T, \end{cases}$$

and

$$\alpha_t = \begin{cases} \alpha_1 & t = 1, \dots, T_1 \\ \alpha_2 & t = T_1 + 1, \dots, T_2 \\ \vdots & \\ \alpha_q & t = T_{q-1} + 1, \dots, T. \end{cases}$$

So in this case we define  $Z_{11t} = Z_{1t}1_{(t \leq T_1)}$ ,  $Z_{12t} = Z_{1t}1_{(T_1+1 \leq t \leq T_2)}$ ,  $\dots$ ,  $Z_{1qt} = Z_{1t}1_{(T_{q-1}+1 \leq t \leq T)}$  and  $\tilde{Z}_{1t} = (Z'_{11t}, \dots, Z'_{1qt})'$ , and obtain a model with the form of (3). This formulation allows for a change in the number of cointegration relations. Let  $r_i$  denote the cointegration rank in subsample  $i$ ,  $i = 1, \dots, q$ . Then the dimension of the parameter space of  $\Pi(t) = \alpha(t)\beta(t)'$  is by Lemma 4 found to be  $\sum_{i=1}^q (p + p_1 - r_i)r_i$  where  $r_i$  is the rank of  $\alpha_i\beta'_i$ ,  $i = 1, \dots, q$ . If the rank is constant over the entire sample, the expression for the degrees of freedom simplifies to  $q(p + p_1 - r)r$ .

### 2.2.2 Structural Changes in the Adjustment Coefficients: $\alpha$

If the structural changes only affect the adjustment coefficients,  $\alpha$ , whereas the cointegration relations remain constant, we can express the model as

$$Z_{0t} = (\alpha_1, \dots, \alpha_q) \begin{pmatrix} \beta & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \beta \end{pmatrix}' \tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t,$$

where  $\tilde{Z}_{1t}$  is as defined above. Since  $\beta$  is constant over the sample, so is the cointegration rank  $r$ , and the dimension of the parameter space for  $\Pi(t)$  is simply given by  $(qp + p_1 - r)r$ .

### 2.2.3 Structural Changes in the Cointegrating Relations: $\beta$

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

$$\begin{aligned} Z_{0t} &= \alpha\beta'_1 1_{1t}Z_{1t} + \cdots + \alpha\beta'_q 1_{qt}Z_{1t} + \Psi Z_{2t} + \varepsilon_t \\ &= \alpha(\beta'_1, \dots, \beta'_q) \tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t, \end{aligned}$$

where  $\tilde{Z}_{1t}$  is as defined previously. Here we again obtain an equation of the form of (3), but in this case without additional restrictions on  $A$ ,  $B$ , and  $C$ , i.e.  $G = I$ , and  $H = I_{p_1 r q}$ . In this situation only a constant cointegration rank,  $r$ , is meaningful and the dimension of the parameter space for  $\Pi(t)$  is given by  $(p + qp_1 - r)r$ .

The relations between the different structural changes are displayed in Figure 1, along with the asymptotic distribution (with degrees of freedom) of the LR test of the more restricted model against the more general of the two models. The asymptotic distribution is derived below, and it is not surprisingly found to be  $\chi^2$ .

### 2.2.4 Temporary and Permanent Cointegration Relations

The scenario where some cointegration relations are present in the entire sample, whereas others are only present in a subsample can also be expressed in the form of (3). The simplest situation is where there are  $r_1$  permanent cointegration relations, say  $\beta_1$ , and for  $t \geq T_1 + 1$  there are an additional  $r_2 - r_1$  temporary cointegration relations, say  $\beta_e$ , (linearly independent of  $\beta_1$ ). This situation leads to two different cases – one where the adjustment coefficients corresponding to  $\beta_1$  remain constant, and one where they may differ in the two subsamples. The latter is likely to be the most relevant, since the introduction of an extra adjustment from the added cointegration relations might affect how the process adjusts to the permanent cointegration relations.

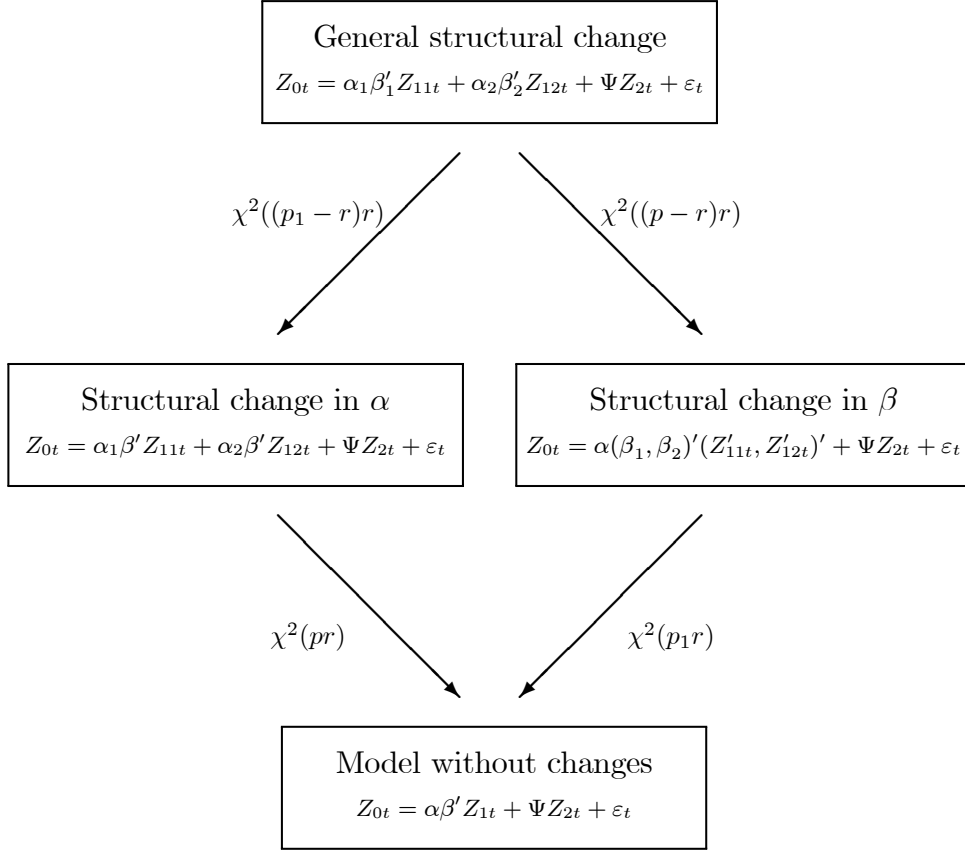


Figure 1: The relations between the different types of structural changes. The asymptotic distribution of the individual LR test is  $\chi^2$  in all cases, with the degrees of freedom reported in the brackets.

First we consider the case where  $\alpha_1$  remains constant. This model is formulated as

$$\begin{aligned}
 Z_{0t} &= \alpha_1 \beta_1' Z_{1t} + \alpha_e \beta_e' Z_{1t} \mathbf{1}_{(t > T_1)} + \Psi Z_{2t} + \varepsilon_t \\
 &= (\alpha_1, \alpha_e) \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_e \end{pmatrix}' \begin{pmatrix} Z_{1t} \\ Z_{1t} \mathbf{1}_{(t > T_1)} \end{pmatrix} + \Psi Z_{2t} + \varepsilon_t,
 \end{aligned}$$

and the dimension of the parameter space for  $\Pi(t)$  is slightly more complicated to derive. The degrees of freedom in  $\Pi_1$  are given by  $(p + p_1 - r_1)r_1$ , and since  $\Pi_2 = \Pi_1 + \alpha_e \beta_e'$  the additional contribution from  $\Pi_2$  is given by  $[p + (p_1 - r_1) - (r_2 - r_1)](r_2 - r_1)$ . Adding the two terms gives the degrees of freedom in  $\Pi(t)$  to be  $(p + p_1 - r_2)r_2 + (r_2 - r_1)r_1$ .

The model where the adjustment coefficients to the permanent cointegration relations

may change, is formulated as

$$\begin{aligned} Z_{0t} &= \alpha_{11}\beta_1'Z_{1t}1_{(t \leq T_1)} + (\alpha_{21}, \alpha_e)(\beta_1, \beta_e)'Z_{1t}1_{(t > T_1)} + \Psi Z_{2t} + \varepsilon_t \\ &= (\alpha_{11}, \alpha_{12}, \alpha_e) \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \beta_1 & \beta_e \end{pmatrix}' \begin{pmatrix} Z_{1t}1_{(t \leq T_1)} \\ Z_{1t}1_{(t > T_1)} \end{pmatrix} + \Psi Z_{2t} + \varepsilon_t, \end{aligned}$$

which is also of the form of (3), but with a more complicated structure of  $H$ , due to the cross restrictions we have on  $B$ . The degrees of freedom are found by adding up the contributions from  $\Pi_1$ ,  $\alpha_{21}\beta_1'$  and  $\alpha_e\beta_e'$ . These are given by  $(p + p_1 - r_1)r_1$ ,  $pr_1$  and  $[p + (p_1 - r_1) - (r_2 - r_1)](r_2 - r_1)$  respectively, where we used that  $\beta_e$  may be chosen orthogonal to  $\beta_1$ . Adding the three terms up, gives the dimension of  $\Pi(t)$  to be  $(p + p_1 - r_2)r_2 + (p + r_2 - r_1)r_1$ .

The former model is obviously nested in the latter, and both models are nested in the model where there are not necessarily any relations between the cointegration relations in the two samples. This model has a structure as given above with  $r_1$  cointegration relations in the first subsample and  $r_2$  in the second. So the model has  $(p + p_1 - r_1)r_1 + (p + p_1 - r_2)r_2$  free parameters in  $\Pi(t)$ . The relations between these three models are displayed in Figure 2. Below we prove that the likelihood ratio test for this hypothesis is asymptotically  $\chi^2$  with degrees of freedom that correspond to the difference in dimensionality of  $\Pi(t)$ , as one would expect.

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

### 2.2.5 Structural Changes in the Covariance Matrix

Structural changes in the covariance matrix fits naturally into the GRRR framework. A single structural change in the covariance matrix at time  $T_1$ , will cause  $\Omega(t) \equiv \text{var}(\varepsilon_t) = \Omega_1$  for  $t \leq T_1$  and  $\Omega(t) \equiv \text{var}(\varepsilon_t) = \Omega_2$  for  $t > T_1$ . Combining structural changes in the covariance matrix with change in other parameters will also lead to a GRRR estimation problem.

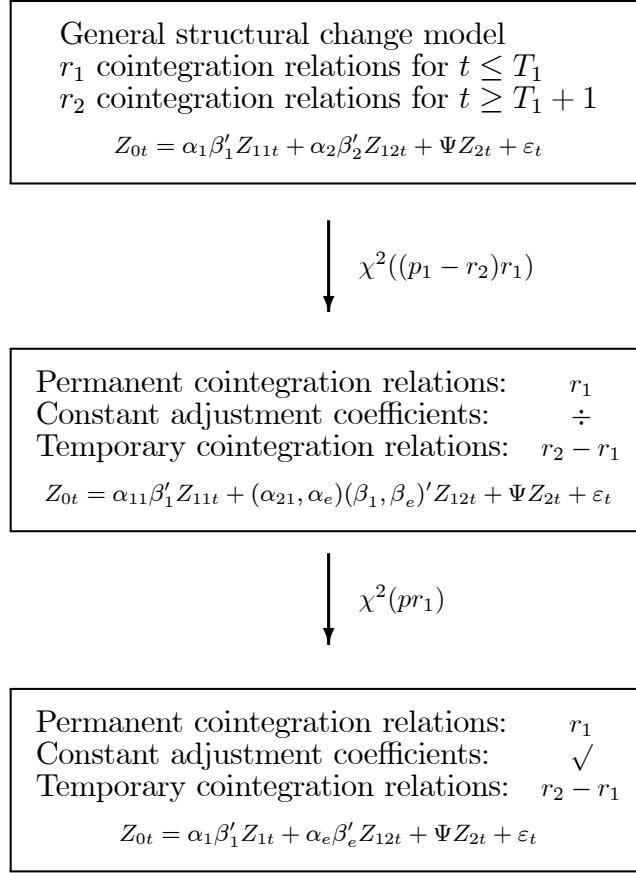


Figure 2: The relations between the different models with structural changes and a change in the number of cointegration relations. The distribution of the LR test statistic between two of the models is asymptotically  $\chi^2$  with the degrees of freedom given in the figure.

### 2.2.6 Linear Restriction on Adjustment Coefficients and Cointegration Relations

Combining hypotheses of structural changes with linear restrictions on the cointegration relations will not complicate the estimation problem, because the two parameter restrictions can jointly be formulated as a linear restriction  $\text{vec}(B) = H\varphi$ , where  $H$  is a known matrix and  $\varphi$  a vector of parameters. Adding linear restrictions to the adjustment coefficients,  $\alpha_1, \dots, \alpha_q$  can be formulated as  $\text{vec}(A) = G\psi$ , and is therefore also a GRRR problem.

## 3 Estimation

Estimation of the cointegrated vector autoregressive model, and other models that have the structure of (2), can be explicitly solved as an eigenvalue problem by reduced rank regression

techniques. The method of reduced rank regression was developed by Anderson (1951) and Izenman (1975) and applied to the  $I(1)$  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 an explicit solution is obtained without iterations, similar to an OLS estimation. This method is applicable to estimation under simple linear restrictions on the reduced rank parameters. However, in most of the structural change models we face restrictions that are beyond what the simple reduced rank regression technique can handle. So a more general estimation technique is needed.

A few of the problems can be formulated as regression problems that can be handled by 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  $\text{vec}(B)$  and  $\text{vec}(A)$ . This method is therefore sufficient for some of the estimation problems that arise from structural change models, that only involves structural changes in the adjustment coefficients,  $\alpha$ , and the cointegrating relations,  $\beta$ . 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.

A related estimation technique is the minimum distances approach by Elliott (1997, 1998a), which can estimate parameters under the general restriction  $g(\theta) = c$ , where  $\theta$  is the vector of parameters,  $c$  is a constant and  $g$  is a well-behaved function. This method minimizes  $\theta' \hat{V}_\theta \theta$  subject to the constraints  $g(\theta) = c$ , where  $\hat{V}_\theta$  is an estimate of the asymptotic covariance matrix. This method is asymptotically equivalent to the maximum likelihood estimation, and with suitable choice of  $\hat{V}_\theta$  and if applied iteratively, (by recursive reestimation of  $\hat{V}_\theta$  as the estimate of  $\theta$  changes), the minimum distance methods leads to the same estimator as

the maximum likelihood method.

As we shall see below, it is possible to estimate under more general restrictions than those considered by Boswijk (1995) and Elliott (1997, 1998a). By handling restrictions as formulated in model (3) we obtain the same generality as the minimum distance method, and can in addition estimate models with a non-constant variance.

In the following we consider the reduced rank regression model  $Z_{0t} = AB'Z_{1t} + CZ_{2t} + \varepsilon_t$ , with various restrictions on the parameters. We denote the dimension of  $Z_{0t}$ ,  $Z_{1t}$ , and  $Z_{2t}$  by  $p$ ,  $p_1$ , and  $p_2$  respectively, and for notional convenience 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  $S_{ij} = \frac{1}{T} \sum_{t=1}^T R_{it}R'_{jt}$ ,  $i, j = 0, 1$ .

### 3.1 Reduced Rank Regression

Estimation of reduced rank regressions is described in the following theorem.

**Theorem 5 (Reduced Rank Regression)** *The unrestricted estimators of (3) are given by*

$$\hat{B} = (\hat{v}_1, \dots, \hat{v}_r)\phi \quad (4)$$

$$\hat{A}(B) = S_{01}B(B'S_{11}B)^{-1} \quad (5)$$

$$\hat{\Omega} = S_{00} - S_{01}BA' + AB'S_{11}BA' - AB'S_{10}, \quad (6)$$

$$\hat{C} = M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}, \quad (7)$$

where  $(\hat{v}_1, \dots, \hat{v}_r)$  are the eigenvectors corresponding to the  $r$  largest eigenvalues  $\hat{\lambda}_1, \dots, \hat{\lambda}_r$  of the eigenvalue problem

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

and where  $\phi$  is any  $r \times r$  full rank matrix, by which  $\hat{B}$  can be normalized. The maximum value of the (conditional) likelihood function is given by

$$L_{\max}^{-2/T}(\hat{A}, \hat{B}, \hat{C}, \hat{\Omega}) = (2\pi e)^P |S_{00}| \prod_{i=1}^r (1 - \lambda_i).$$

An algebraic proof that uncovers the structure of the problem is given in the appendix whereas the original proof can be found in Johansen (1996).

This theorem is directly applicable to the cointegrated vector autoregressive model given by (1). The maximum likelihood estimate is obtained by defining  $Z_{0t} = \Delta X_t$ ,  $Z_{1t} = X_{t-1}$  and  $Z_{2t} = (\Delta X'_{t-1}, \dots, \Delta X'_{t-k+1}, D'_t)'$ .

### 3.2 Generalized Reduced Rank Regression

**Theorem 6** *Let the parameter  $A$ ,  $B$ , and  $C$  be restricted by  $\text{vec}(A, C) = G\psi$  and  $\text{vec}(B) = H\phi$  and suppose that Assumptions 1-3 hold.*

*The maximum likelihood estimates  $\hat{A}$ ,  $\hat{B}$ ,  $\hat{C}$ , and  $\hat{\Omega}(t)$  of  $A$ ,  $B$ ,  $C$ , and  $\Omega(t)$  will satisfy*

$$\text{vec}(\hat{A}, \hat{C}) = G \left[ G' \sum_{t=1}^T \left[ \begin{pmatrix} \hat{B}' Z_{1t} Z'_{1t} \hat{B} & \hat{B}' Z_{1t} Z'_{2t} \\ Z_{2t} Z'_{1t} \hat{B} & Z_{2t} Z'_{2t} \end{pmatrix} \otimes \hat{\Omega}(t)^{-1} \right] G \right]^{-1} \quad (8)$$

$$\times G' \sum_{t=1}^T \text{vec} \left( \hat{\Omega}(t)^{-1} Z_{0t} (Z'_{1t} \hat{B}, Z'_{2t}) \right),$$

$$\text{vec}(\hat{B}) = H \left[ H' \sum_{t=1}^T \left[ \hat{A}' \hat{\Omega}(t)^{-1} \hat{A} \otimes Z_{1t} Z'_{1t} \right] H \right]^{-1} \quad (9)$$

$$\times H' \sum_{t=1}^T \text{vec} \left( Z_{1t} (Z_{0t} - \hat{C} Z_{2t})' \hat{\Omega}(t)^{-1} \hat{A} \right)$$

and  $\hat{\Omega}(t) = \Omega_t(\hat{\theta})$ , where  $\hat{\theta}$  is given from the (model specific) equation

$$\hat{\theta} = \arg \max_{\theta \in \Theta_\theta} L(\hat{A}, \hat{B}, \hat{C}, \theta, Z_0, Z_1, Z_2).$$

*The maximum value of the likelihood function is given by*

$$L_{\max}(\hat{A}, \hat{B}, \hat{C}, \hat{\Omega}) = (2\pi)^{-\frac{Tp}{2}} \prod_{t=1}^T |\hat{\Omega}(t)|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \sum_{t=1}^T \hat{\varepsilon}'_t \hat{\Omega}(t)^{-1} \hat{\varepsilon}_t \right),$$

where  $\hat{\varepsilon}_t = Z_{0t} - \hat{A} \hat{B}' Z_{1t} - \hat{C} Z_{2t}$ .

The proof exploits that the estimation problem reduces to a GLS problem, when  $(A, C, \theta)$  or  $(B, \theta)$  are hold constant. The proof is given in the Appendix.

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. As described in the paragraph following Assumption 3, this procedure will converge to parameter values that satisfy the normal equations.

We now treat situations with fewer parameter restrictions.



**Corollary 7** Let the parameter  $A$ ,  $B$ , and  $C$  be restricted by  $\text{vec}(A, C) = G\psi$  and  $\text{vec}(B) = H\phi$  and suppose that  $\{\varepsilon_t\}$  is i.i.d. Gaussian  $N(0, \Omega)$ .

The maximum likelihood estimates of  $A$ ,  $B$ ,  $C$ , and  $\Omega$  satisfy the equations

$$\begin{aligned}\text{vec}(\hat{A}, \hat{C}) &= G \left[ G' \left( \begin{array}{cc} \hat{B}' M_{11} B & \hat{B}' M_{12} \\ M_{21} \hat{B} & M_{22} \end{array} \right) \otimes \hat{\Omega}^{-1} \right]^{-1} G' \text{vec} \left( \hat{\Omega}^{-1} (M_{01} \hat{B}, M_{02}) \right), \\ \text{vec}(\hat{B}) &= H \left[ H' \left( \hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes M_{11} \right) H \right]^{-1} H' \text{vec} \left( M_{10} - M_{02} \hat{C}' \hat{\Omega}^{-1} \hat{A} \right), \\ \hat{\Omega} &= T^{-1} (Z_0 - \hat{A} \hat{B}' Z_1 - \hat{C}' Z_2) (Z_0 - \hat{A} \hat{B}' Z_1 - \hat{C}' Z_2)'.\end{aligned}$$

The maximum value of the likelihood function is given by

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

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

**Corollary 8** Let  $A$  and  $B$  be restricted by  $\text{vec}(A) = G\zeta$  and  $\text{vec}(B) = H\phi$ , for known matrices  $G$  and  $H$ . Then the maximum likelihood estimates satisfy the equations

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

and

$$\begin{aligned}\hat{A} &= G \left[ G' \left( \hat{B}' S_{11} \hat{B} \otimes I_p \right) G \right]^{-1} G' \left( \hat{B}' \otimes \hat{\Omega}^{-1} \right) \text{vec} \left( S_{01} \right), \\ \hat{\Omega} &= S_{00} - S_{01} \hat{B} \hat{A}' + \hat{A} \hat{B}' S_{11} \hat{B} \hat{A}' - \hat{A} \hat{B}' S_{10}, \\ \hat{C} &= M_{02} M_{22}^{-1} - \hat{A} \hat{B}' M_{12} M_{22}^{-1}.\end{aligned} \quad (11)$$

The maximum value of the likelihood function is given by

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

**Corollary 9** Let  $B$  be restricted by  $\text{vec}(B) = H\phi$ . Then the maximum likelihood estimates

satisfy the equations

$$\begin{aligned}
\text{vec}(\hat{B}(A, \Omega)) &= H \left[ H' \left( \hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes S_{11} \right) H \right]^{-1} H' \left( \hat{A}' \otimes S_{10} \right) \text{vec} \left( \hat{\Omega}^{-1} \right), \\
\hat{A}(B) &= S_{01} \hat{B} \left( \hat{B}' S_{11} \hat{B} \right)^{-1}, \\
\hat{\Omega}(B) &= S_{00} - S_{01} \hat{B} \left( \hat{B}' S_{11} \hat{B} \right)^{-1} \hat{B}' S_{10}, \\
\hat{C} &= M_{02} M_{22}^{-1} - \hat{A} \hat{B}' M_{12} M_{22}^{-1}.
\end{aligned}$$

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 autoregressive model under all the various structural changes considered in the previous section. However, the theorems presented here have a broader applicability, and can be used to estimate models with parameter restrictions that need not be related to structural changes, for example models with heteroskedasticity.

### 3.3 Applicability

**Example 1 (Structural changes in the covariance matrix)** Consider the cointegrated vector autoregressive model (1), with a structural change at time  $T_1$ , in the sense that  $\alpha(t) = \alpha_1$ ,  $\beta(t) = \beta_1$  and  $\Omega(t) = \Omega_1$  for  $t \leq T_1$  and  $\alpha(t) = \alpha_2$ ,  $\beta(t) = \beta_2$  and  $\Omega(t) = \Omega_2$  for  $t \geq T_1 + 1$ . This estimation problem can be written in the form of (3). The maximum likelihood estimators of  $\Omega_1$  and  $\Omega_2$  are given by

$$\begin{aligned}
\hat{\Omega}_1 &= T_1^{-1} \sum_{t=1}^{T_1} \hat{\varepsilon}_t \hat{\varepsilon}_t' \\
\hat{\Omega}_2 &= (T - T_1)^{-1} \sum_{t=T_1+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'.
\end{aligned}$$

So  $\Omega(t)$ ,  $t = 1, \dots, T$  can be expressed in the functional form required by Theorem 6.

## 4 Asymptotic Analysis

For simplicity, we derive the asymptotic results in the case of a single structural change at time  $T_1$ , and with the number of cointegrating relations being constant,  $r$ . However, it will be clear that the results hold in the general situation with multiple changes, and varying number of cointegrating relations.

The process is described by

$$\Delta X_t = \alpha_1 \beta_1' X_{t-1} 1_{(t \leq T_1)} + \alpha_2 \beta_2' X_{t-1} 1_{(t > T_1)} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t,$$

where  $\varepsilon_t$  is i.i.d.<sup>4</sup>  $N(0, \Omega(t))$ ,  $\Omega(t) = \Omega_1$  for  $t \leq T_1$  and  $\Omega(t) = \Omega_2$  for  $t > T_1$ .

In addition, we assume that the usual  $I(1)$  assumptions hold in both subsamples. Specifically, that the roots of

$$\left| I(1-z) - \alpha_i \beta_i' z - \sum_{i=1}^{k-1} \Gamma_i (1-z) z^i \right| = 0$$

are outside the unit disc or equal to one, and that  $\alpha'_{i\perp} (I - \Gamma_1 - \dots - \Gamma_{k-1}) \beta_{i\perp}$  has full rank  $p - r$ ,  $i = 1, 2$ .

### 4.1 The Granger Representation for Change Processes

In order to study the process's asymptotic properties, we need to derive the Granger representation for this process. The individual Granger representations for each of the sub-samples are given by

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

and

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

where  $C = \beta_{1\perp} (\alpha'_{1\perp} \Gamma \beta_{1\perp})^{-1} \alpha'_{1\perp}$ ,  $D = \beta_{2\perp} (\alpha'_{2\perp} \Gamma \beta_{2\perp})^{-1} \alpha'_{2\perp}$  and  $\Gamma = I - \Gamma_1 - \dots - \Gamma_{k-1}$ , see P. R. Hansen (2000b).

In order to get the representation in the appropriate form we need to express the second

---

<sup>4</sup>The asymptotic results will hold under more general conditions, though not always with the same asymptotic distribution. Both the Gaussian assumption and the i.i.d. assumption can be relaxed to  $\{\varepsilon_t\}$  satisfying a Functional Central Limit Theorem, (see White (2000)).

representation with initial values depending only on  $X_t$ ,  $t = 0, -1, \dots$ , rather than  $D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i})$ . This is obtained by the expression

$$\begin{aligned}
D(X_{T_1} - \sum_{i=1}^{k-1} \Gamma_i X_{T_1-i}) &= D \left[ C \sum_{i=1}^{T_1} \varepsilon_i + C(L)\varepsilon_{T_1} + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right. \\
&\quad \left. - \Gamma_1 \left( C \sum_{i=1}^{T_1-1} \varepsilon_i + C(L)\varepsilon_{T_1-1} + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right) \right. \\
&\quad \vdots \\
&\quad \left. - \Gamma_{k-1} \left( C \sum_{i=1}^{T_1-k+1} \varepsilon_i + C(L)\varepsilon_{T_1-k+1} + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right) \right] \\
&= D \left[ \Gamma C \sum_{i=1}^{T_1} \varepsilon_i + C^*(L)\varepsilon_{T_1} + \Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \right] \\
&= D\Gamma C \sum_{i=1}^{T_1} \varepsilon_i + DC^*(L)\varepsilon_{T_1} + D\Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}),
\end{aligned}$$

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. So altogether we have the Granger representation

$$X_t = C \sum_{i=1}^t \varepsilon_i + C(L)\varepsilon_t + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \quad t = 1, \dots, T_1 \quad (12)$$

$$\begin{aligned}
X_t &= D \sum_{i=T_1+1}^t \varepsilon_i + D\Gamma C \sum_{i=1}^{T_1} \varepsilon_i + D(L)\varepsilon_t + DC^*(L)\varepsilon_{T_1} \\
&\quad + D\Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \quad t = T_1 + 1, \dots, T.
\end{aligned} \quad (13)$$

Note that we have the stationary cointegrating relations in the second sub-sample  $\beta_2' X_t = \beta_2' D(L)\varepsilon_t$ , which is identical to what it would have been in the case of a constant process. For the first sub-sample the results are trivially the same as in the standard case without changes.

## 4.2 The Continuous Time Limits

In an asymptotic study of the process, we shall, as  $T$  approaches infinity, keep the proportion of observations in each sub-sample constant. So we define  $\rho = \frac{T_1}{T}$ , which denotes the fraction

of observations in the first sub-sample.

Donsker's invariance principle gives

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tu]} \varepsilon_t \xrightarrow{w} W(u), \quad u \in [0, 1],$$

where  $W(u)$  is a Brownian motion with covariance matrix  $\Omega$ , and where  $\xrightarrow{w}$  denotes weak convergence. We can split this into two independent Brownian motions which gives us

$$T^{-\frac{1}{2}} \left( \sum_{t=1}^{T_1} \varepsilon_t + \sum_{t=T_1+1}^{[Tu]} \varepsilon_t \right) \xrightarrow{w} W(\rho) + W(u) - W(\rho), \quad u > \rho.$$

So the random walk element in  $X_t$  in each of the sub-samples, has the continuous time limits:

$$\begin{aligned} T^{-\frac{1}{2}} C \sum_{t=1}^{[Tu]} \varepsilon_t &\xrightarrow{w} CW(u), & u \leq \rho \\ T^{-\frac{1}{2}} \left( D\Gamma C \sum_{t=1}^{T_1} \varepsilon_t + D \sum_{t=T_1+1}^{[Tu]} \varepsilon_t \right) &\xrightarrow{w} D\Gamma CW(\rho) + D(W(u) - W(\rho)), & u > \rho. \end{aligned} \quad (14)$$

Equation (14) has an important implication for unit root tests, in processes with structural changes. Standard Dickey-Fuller type distributions, such as  $\int (dB) B' [\int BB' du]^{-1} \int B (dB)'$  do not define the asymptotic distribution in this situation, because the Gaussian term  $D\Gamma CW_1(\rho)$ , that comes from the initial values, does not disappear. A unit root test based on observation after a structural change will therefore involve a term such as  $\int (dB) (B + Z)' [\int (B + Z)(B + Z)' du]^{-1} \int (B + Z) (dB)'$ . However, this problem does not occur if a constant (restricted to the cointegration space or unrestricted) is included as regressor, see P.R. Hansen (2000c, Chapter 4).

From the Granger representation we find that

$$T^{-1/2} \begin{pmatrix} X_{[Tu]} 1_{(u \leq \rho)} \\ X_{[Tu]} 1_{(u > \rho)} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} CW(u) 1_{(u \leq \rho)} \\ [D\Gamma CW_1(\rho) + D(W(u) - W(\rho))] 1_{(u > \rho)} \end{pmatrix},$$

since the other terms vanish in probability. Let

$$B = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix},$$

let  $B_\perp$  be the orthogonal complement to  $B$ , i.e.  $B'_\perp B = 0$  and let  $\bar{B}_\perp = B_\perp(B'_\perp B_\perp)^{-1}$ . We define

$$G(u) = \begin{pmatrix} G_1(u) \\ G_2(u) \end{pmatrix} = \bar{B}'_\perp \begin{pmatrix} CW(u)1_{(u \leq \rho)} \\ [D\Gamma CW_1(\rho) + D(W(u) - W(\rho))]1_{(u > \rho)} \end{pmatrix}$$

and by the continuous mapping theorem we have with  $u = t/T$  that

$$\begin{aligned} & T^{-2} \sum_{t=1}^T \bar{B}'_\perp \begin{pmatrix} X_{[Tu]}1_{(u \leq \rho)} \\ X_{[Tu]}1_{(u > \rho)} \end{pmatrix} \begin{pmatrix} X_{[Tu]}1_{(u \leq \rho)} \\ X_{[Tu]}1_{(u > \rho)} \end{pmatrix}' \bar{B}_\perp \xrightarrow{w} \int_0^1 G(u)G(u)du \\ &= \begin{pmatrix} \int_0^\rho G_1(u)G_1(u)du & 0 \\ 0 & \int_\rho^1 G_2(u)G_2(u)du \end{pmatrix}. \end{aligned}$$

With this notation, the asymptotic results for unrestricted parameter estimates ( $G = I$  and  $H = I$ ) of  $A$ ,  $B$ ,  $C$ , and  $\Omega(t)$ , say  $\hat{A}_u$ ,  $\hat{B}_u$ ,  $\hat{C}_u$ , and  $\hat{\Omega}_u(t)$ , follows from Johansen (1988, lemma 13.1, 13.2). The results are that (a normalized)  $\hat{B}_u$  is super consistent, with a mixed Gaussian asymptotic distribution, and that  $\hat{A}_u$  is asymptotically normal. Further it also follows that the LR test of some over identifying restrictions, have a  $\chi^2$  asymptotic distribution.

Consistency is not affected by imposing valid restrictions, and the results for the restricted parameter estimates given by expanding the normal equations. Assume for simplicity that  $\Omega(t)$  is constant, then

$$\begin{aligned} \text{vec}(\hat{B}) &= H \left[ H' \left[ \hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes T^{-2} \sum_{t=1}^T Z_{1t} Z'_{1t} \right] H \right]^{-1} \\ &\quad \times H' \sum_{t=1}^T \text{vec} \left( T^{-1} Z_{1t} (AB' Z_{1t} + (C - \hat{C}) Z_{2t} + \varepsilon_t)' \Omega^{-1} A \right) \\ &= \text{vec}(B) + H \left[ H' \left[ \hat{A}' \hat{\Omega}^{-1} \hat{A} \otimes \sum_{t=1}^T Z_{1t} Z'_{1t} \right] H \right]^{-1} \\ &\quad \times H' \sum_{t=1}^T \text{vec} \left( Z_{1t} \varepsilon'_t \Omega^{-1} A \right) + o_p(1), \end{aligned}$$

which by the consistency of  $\hat{A}$ ,  $\hat{C}$  and  $\hat{\Omega}$  shows that

$$T\text{vec}(\hat{B} - B) \xrightarrow{w} H \left[ H' \left[ A' \Omega^{-1} A \otimes B_{\perp} \int_0^1 G(u) G'(u) du B_{\perp}' \right] H \right]^{-1} \\ \times H' \text{vec} \left( \int G(u) dW \Omega^{-1} A \right),$$

which is a mixed Gaussian distribution. Similarly

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

which asymptotically has a Gaussian distribution. The case with a varying  $\Omega(t)$  leads to the same results, although the expressions have a more complicated structure.

From these results it follows by arguments similar to the ones of Johansen (1996, theorems 13.7, 13.9), that the likelihood ratio test has an asymptotically  $\chi^2$  distribution, for hypotheses that can be formulated as linear restrictions.

## 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<sup>5</sup> states that

$$f_{n,t} = E_t(y_{1,t+n}) + \Lambda_n,$$

where  $\Lambda_n$  is the term premium. The restriction imposed by the expectations hypothesis is that the term premium does not depend on  $t$ . From the Fisher-Hicks relation  $y_{nt} = n^{-1} \sum_{j=0}^{n-1} f_{jt}$ ,  $n = 1, 2, \dots$ , 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

$$y_{nt} - y_{1t} = n^{-1} \sum_{j=1}^{n-1} \sum_{i=1}^j E_t(\Delta y_{1,t+i}) + L_n. \quad (15)$$

where  $L_n = n^{-1} \sum_{j=0}^{n-1} \Lambda_j$ . This equation shows that if  $y_{1t}$  is  $I(1)$ , 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<sup>6</sup>, 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. This is only one of several implications of the expectations hypothesis. Equation (15) is the motivation for modelling 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,  $\gamma$ , for 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  $\gamma$  is increase, and found the effect to be

---

<sup>5</sup>For an overview of the expectations hypothesis theory and empirical studies of interest rates, see Shiller (1990).

<sup>6</sup>The 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.



substantial as  $\gamma$  increases beyond 0.5.

Whereas the expectations hypothesis has been rejected by most studies of US data (see Shiller (1990) 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 implications 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 modelled as  $I(1)$  variables<sup>7</sup>. The fact that nominal interest

---

<sup>7</sup>Ait-Sahalia (1996) found the short interest rates to behave as an  $I(1)$  process within the band [4%, 18%]

rates cannot be negative and other considerations are strong arguments against interest rates being  $I(1)$  forever. Nevertheless, interest rates may very well be  $I(1)$  in a particular sample period. Whenever this is the case, modelling 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  $I(1)$  has had no relevance in the sample period analyzed. See Pagan, Hall, and Martin (1996) for another argument on this matter.

### 5.3 Data

The term structure data were extracted from the Bliss data<sup>8</sup> 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 months<sup>9</sup> within the sample period 1970:1 – 1995:12. 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  $t$ . The most general model can be expressed as

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

where  $\alpha(t)$ ,  $\beta(t)$  and  $\mu(t)$  are piecewise constant with two change points: in 1979:10 and in 1982:10. To avoid a deterministic trend in the yields, the constant is restricted by  $\mu(t) = \alpha(t)\rho(t)$ , so the model can be rewritten 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  $\beta^{*'} = (\beta(t)', \rho(t))$ .

---

and a theoretical model in which interest rates are similar to a random walk is given by DenHaan (1995).

<sup>8</sup>The 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.

<sup>9</sup>Longer maturities were not selected because precise estimate of these are difficult to obtain by interpolation techniques. See Bliss (1997)

We may normalize the cointegration relations by

$$\beta(t)^* = \begin{pmatrix} \beta_{11,t} & \beta_{12,t} & \cdots & \beta_{1r,t} \\ -1 & 0 & & 0 \\ 0 & -1 & & \\ \vdots & & \ddots & \\ 0 & & & -1 \\ \rho_{1,t} & \rho_{2,t} & \cdots & \rho_{r,t} \end{pmatrix}. \quad (16)$$

Since these relations define the stationary relations, the long-run implications of the expectations hypothesis – that the spreads  $y_{n,t} - y_{1,t}$  are stationary – can be formulated as the parameter restrictions  $\beta_{11,t} = \cdots = \beta_{1r,t} = 1$ .

The individual cointegration relations in (16) can be written as

$$b_{n,t}y_{1,t} - y_{n,t} + \rho_{n,t}, \quad n = 3, 6, 9, 12, 60, 84,$$

where the maturities  $n = 3, 6, 9, 12, 60, 84$  and  $b_{n,t}$  correspond to  $i = 1, \dots, r$  and  $\beta_{1i,t}$  in equation (16). The Granger representation shows that  $E(b_{n,t}y_{1,t} - y_{n,t} + \rho_{n,t}) = 0$ , so  $\hat{\rho}_{n,t}$  can be interpreted as the estimated term premia when  $b_{n,t}$  is 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. No formal test was applied for this selection.

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.

$\Delta X_t - \alpha(t)\beta^*(t)'X_{t-1}^* - \Gamma_1\Delta X_{t-1} \sim N(0, \Omega(t))$			
Model	$\max \log L(\alpha(t), \beta^*(t), \Gamma_1, \Omega(t))$	Degrees of freedom	LR( $M_i M_0$ ) ( <i>p</i> -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 the model with changing reduced rank parameters, and changing covariance  $\Omega_t$ .

The Estimated Covariance Matrices, $\Omega(t)$	
1970:3–1979:9	$\hat{\Omega}_1 = \begin{pmatrix} 0.30 & 0.28 & 0.25 & 0.22 & 0.14 & 0.10 & 0.09 \\ 0.28 & 0.27 & 0.25 & 0.22 & 0.15 & 0.11 & 0.10 \\ 0.25 & 0.25 & 0.25 & 0.23 & 0.17 & 0.12 & 0.11 \\ 0.22 & 0.22 & 0.23 & 0.23 & 0.17 & 0.13 & 0.11 \\ 0.14 & 0.15 & 0.17 & 0.17 & 0.15 & 0.12 & 0.11 \\ 0.10 & 0.11 & 0.12 & 0.13 & 0.12 & 0.10 & 0.09 \\ 0.09 & 0.10 & 0.11 & 0.11 & 0.11 & 0.09 & 0.08 \end{pmatrix}$
1979:10–1982:10	$\hat{\Omega}_2 = \begin{pmatrix} 1.75 & 1.68 & 1.51 & 1.28 & 0.92 & 0.63 & 0.54 \\ 1.68 & 1.70 & 1.58 & 1.33 & 0.97 & 0.68 & 0.59 \\ 1.51 & 1.58 & 1.50 & 1.30 & 0.97 & 0.69 & 0.61 \\ 1.28 & 1.33 & 1.30 & 1.18 & 0.90 & 0.65 & 0.57 \\ 0.92 & 0.97 & 0.97 & 0.90 & 0.72 & 0.54 & 0.48 \\ 0.63 & 0.68 & 0.69 & 0.65 & 0.54 & 0.43 & 0.39 \\ 0.54 & 0.59 & 0.61 & 0.57 & 0.48 & 0.39 & 0.35 \end{pmatrix}$
1982:11–1995:12	$\hat{\Omega}_3 = \begin{pmatrix} 0.10 & 0.09 & 0.08 & 0.07 & 0.07 & 0.06 & 0.05 \\ 0.09 & 0.09 & 0.09 & 0.09 & 0.09 & 0.08 & 0.07 \\ 0.08 & 0.09 & 0.10 & 0.11 & 0.11 & 0.10 & 0.09 \\ 0.07 & 0.09 & 0.11 & 0.12 & 0.12 & 0.11 & 0.11 \\ 0.07 & 0.09 & 0.11 & 0.12 & 0.13 & 0.13 & 0.12 \\ 0.06 & 0.08 & 0.10 & 0.11 & 0.13 & 0.13 & 0.13 \\ 0.05 & 0.07 & 0.09 & 0.11 & 0.12 & 0.13 & 0.12 \end{pmatrix}$

Table 2: The estimated covariance matrices,  $\hat{\Omega}_j$ ,  $j = 1, 2, 3$ , from the most general change model.

It is not surprising that the variance of interest rates (see Table 2) 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.

Six models with different parameter restrictions were estimated<sup>10</sup>. The estimations results are given in Tables 3 and 4.

Model 1 in Table 3 is the most general model, where the parameters are left unrestricted. This model can be represented by the equation

$$\begin{aligned}\Delta X_t &= \alpha(t) [\beta(t)' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \\ \varepsilon_t &\sim N(0, \Omega(t)),\end{aligned}$$

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 imposed as the parameter restriction  $b_n = 1$  for all  $n$  in all three subsamples, whereas the term premia,  $\rho_n$ , adjustment coefficients,  $\alpha_i$ , as well as the covariance matrices,  $\Omega_i$ , may differ across the subsamples,  $i = 1, 2, 3$ . This model can be written as

$$\begin{aligned}\Delta X_t &= \alpha(t) [\beta' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \\ \varepsilon_t &\sim N(0, \Omega(t)),\end{aligned}$$

where  $\beta$  has the structure required by the long-run implications. 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

---

<sup>10</sup>The empirical analysis was performed in Gauss.

1: Unrestricted Change Model $\alpha(t), \beta(t), \rho(t), \Omega(t)$				$2 \log L$ 4018.49	$\#f$ 295	LR –	$p$ -value –
	n	3	6	9	12	60	84
1970:3–1979:9	$b_n$	0.9831	0.9767	0.9162	0.7473	0.6154	0.5947
	$\rho_n$	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
	$\rho_n$	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
	$\rho_n$	-0.2384	-0.4607	-0.9011	-2.0401	-2.8354	-3.0585
2: Expectations Hypothesis $\alpha(t), \beta(t) = \beta, \rho(t), \Omega(t)$				$2 \log L$ 3989.58	$\#f$ 277	LR 28.91	$p$ -value 0.0495
	n	3	6	9	12	60	84
1970:3–1979:9	$b_n$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\rho_n$	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
	$\rho_n$	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
	$\rho_n$	0.2106	0.3694	0.6307	1.0520	1.3919	1.5010
3: Constant $\alpha_\perp$ & Expectations Hypothesis $\alpha(t) = \alpha\phi(t), \beta(t) = \beta, \rho(t), \Omega(t)$				$2 \log L$ 3978.44	$\#f$ 265	LR 40.05	$p$ -value 0.1038
	n	3	6	9	12	60	84
1970:3–1979:9	$b_n$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\rho_n$	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
	$\rho_n$	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
	$\rho_n$	0.2123	0.3753	0.6523	1.1248	1.5229	1.6487
4: Constant $\alpha$ & $\beta$ & EH. $\rho(t)$ may change. $\alpha(t) = \alpha, \beta(t) = \beta, \rho(t), \Omega(t)$				$2 \log L$ 3784.01	$\#f$ 199	LR 234.48	$p$ -value 0.0000
	n	3	6	9	12	60	84
1970:3–1979:9	$b_n$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\rho_n$	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
	$\rho_n$	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
	$\rho_n$	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 and the test statistic (tested against the most general model) with the correspondings  $p$ -value. The cointegration parameters  $b_n$  and term premia  $\rho_n$  from the cointegration relations  $b_n y_{1,t} - y_{n,t} + \rho_n$  are reported for each model and subsample.

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

$$\begin{aligned}\Delta X_t &= \alpha \phi(t) [\beta' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \\ \varepsilon_t &\sim N(0, \Omega(t)).\end{aligned}$$

The restriction implies that the orthogonal complement to  $\alpha$  is constant, i.e.,  $\alpha_\perp(t) = \alpha_\perp$ . The different strength of the adjustments between the three subsamples are expressed in terms of the matrix  $\phi(t)$ .

Recall the Granger representation from equations (12) and (13), and here extended with a third subsample:

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

An implication of the constancy of  $\alpha_\perp$  and  $\beta$  and  $\Gamma_1$  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  $C\Gamma C = C$ .

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  $C\bar{\alpha}_\perp \alpha'_\perp = C$ ). 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 implications are consistent with the data and we cannot reject that the common stochastic trend has been a constant linear combination of  $\varepsilon_t$ , and we cannot reject that the loading of the common

stochastic trend has been constant. The non-constancy of the common stochastic trend comes from the changing variance of  $\varepsilon_t$ .

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

$$\begin{aligned}\Delta X_t &= \alpha [\beta' X_{t-1} + \rho(t)] + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \\ \varepsilon_t &\sim N(0, \Omega(t)).\end{aligned}$$

In this model the adjustment coefficients have the same strength in the three subsamples. This is equivalent to the additional restriction:  $\phi(t) = \phi$  on Model 3. This model is clearly inconsistent with the term structure data. The fact that the strength of the adjustments are non-constant is not puzzling, since the changes appear along with changes in volatility and term premia.

Thus, we find the term structure to have had structural changes in the covariance  $\Omega(t)$  and the term premia  $\rho(t)$  along with changes in the strength of the adjustments to dis-equilibria in the cointegration relations. However fundamentals such as the common stochastic trend and stable relationships between interest rates have remained relatively unchanged in the sample analyzed.

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, the result can be incorrect inference, and a possible 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 the results, since we find  $\rho(t)$  to vary as the volatility of interest rates changes. Finally, my finding of a changing variance is likely to distort hypothesis testing if not accounted for, which is similar to the volatility effect found by Pagan, Hall, and Martin (1996).

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



5: No Structural Changes $\alpha(t) = \alpha, \beta(t) = \beta, \rho(t) = \rho, \Omega(t) = \Omega$				$2 \log L$	$\#f$	LR	$p$ -value
				2852	131	-	-
	n	3	6	9	12	60	84
1970:3–1995:12	$b_n$	1.0390	1.0417	1.0520	1.0529	1.0239	1.0191
	$\rho_n$	0.0011	0.1680	0.2951	0.6209	1.1478	1.2875
6: No Changes & Expectations Hypothesis $\alpha(t) = \alpha, \beta(t) = H\phi, \rho(t) = \rho, \Omega(t) = \Omega$				$2 \log L$	$\#f$	$\mathcal{LR}$	$p$ -value
				2825	125	26.84	0.0002
	n	3	6	9	12	60	84
1970:3–1995:12	$b_n$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\rho_n$	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 strongly rejected against model 1.

test statistic of Model 5 against Model 1 is 1166. Its distribution is asymptotically  $\chi^2$  with 164 degrees of freedom, and is therefore clearly rejected.

## 6 Conclusion

This paper shows 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. Moreover, the parameters can be estimated under these restrictions with the new generalized reduced rank regression technique. This technique is also applicable to estimation problems unrelated to structural changes, see P. R. Hansen (2000a) for several examples.

I derived the likelihood ratio test for structural changes occurring at known points in time, and showed that it is asymptotically  $\chi^2$ . Moreover, it was shown how hypotheses can be tested, when the maintained hypothesis is presence of structural changes. I derived the asymptotic distributions of the parameter estimates and likelihood ratio tests. Similar to the standard model without structural changes, the estimate of the cointegration relations is super-consistent and asymptotically mixed Gaussian, and the LR statistic is asymptotically  $\chi^2$ .

This combination of cointegration and structural changes may provide a fruitful framework for many economic questions of interest. In this paper I 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)) I cannot reject the long-run implications 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 structure in the three monetary regimes, and along with changes in the covariance matrix, only the term premia and the strength of adjustment coefficients changes, the cointegrating relations are stable over the sample period.

In this paper, the cointegration rank was taken as given. Although this is reasonable when interest rates are analyzed, this need not always be the case. A formal test to determine the rank of cointegrated processes is derived in P. R. Hansen (2000c, Chapter 4).

## A Appendix: Proofs

### A.1 Algebraic Treatment of the Reduced Rank Regression

Before I give the proofs of Theorem 5 I derive some intermediate results. The following lemma is a consequence of Poincaré's theorem, however, a new proof is presented here.

**Lemma 10** *The function  $g(y) = |y' \Lambda y| / |y' y|$  where  $y$  is a  $p \times r$  matrix,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  and  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  has maximum value  $\prod_{i=1}^r \lambda_i$  which is attained with  $y$  equal to the first  $r$  unit vectors, that is  $y = (I_r, 0_{r \times p-r})'$ .*

**Proof.** Let  $J$  be an index set  $J \subset \{1, \dots, p\}$  of cardinality  $r$ , and define the  $r \times r$  matrices  $y_J$  and  $\Lambda_J$  by  $y_J = \{y_{ij}\}_{i \in J, j=1, \dots, r}$  and  $\Lambda_J = \{\Lambda_{ij}\}_{i, j \in J}$ . So if  $p = 3$ ,  $r = 2$  and  $J = \{1, 2\}$  we would have  $y_J = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$  and  $\Lambda_J = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ .

Next, let  $\mathbb{D}_p^r$  denote the set of all subsets of  $\{1, \dots, p\}$  containing exactly  $r$  different elements (cardinality  $r$ ). Below, I prove that

$$|y' \Lambda y| = \sum_{J \in \mathbb{D}_p^r} |y'_J \Lambda_J y_J| = \sum_{J \in \mathbb{D}_p^r} |y'_J y_J| \prod_{i \in J} \lambda_i = \sum_{J \in \mathbb{D}_p^r} |y_J|^2 \prod_{i \in J} \lambda_i. \quad (17)$$

So  $g(y) = |y' \Lambda y| / |y' y| = \sum_{J \in \mathbb{D}_p^r} |y_J|^2 \prod_{i \in J} \lambda_i / \sum_{J \in \mathbb{D}_p^r} |y_J|^2$  is a convex combination over the elements in  $\mathbb{D}_p^r$  with values given by  $\prod_{i \in J} \lambda_i$ , with the largest element being  $\prod_{i=1}^r \lambda_i$  corresponding to  $J = \{1, \dots, r\}$ . This value can be obtained with  $\hat{y} = (I_r, 0_{r \times p-r})'$  which therefore maximizes the function  $g(y)$ .

The identity (17) is proved as follows. The second and third equality follows trivially from  $|AB| = |A||B|$  for matrices of proper dimensions, whereas the first equality is showed by induction below. The equality trivially holds for  $r = 1$  or  $p = r$ . So the scheme

$p \setminus r$	1	2	3	4	$\dots$
1	✓	–	–	–	
2	✓	✓	–	–	
3	✓	?	✓	–	
4	✓	?	?	✓	
$\vdots$	$\vdots$				$\ddots$

shows that the equality can be proven by showing it holds for cell  $(p, r)$  when it is assume that it holds for cell  $(p-1, r-1)$ , say assumption **(A1)**, and for cell  $(p-1, r)$ , say assumption **(A2)**.

Define  $\tilde{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{p-1})$  and consider first the case where the last row of  $y$  is a zero-row  $(y_{p1}, \dots, y_{pr}) = (0, \dots, 0)$ . Define in this case  $\tilde{y} = \{y_{ij}\}_{i=1, \dots, p-1}$ , that is  $y$  without the zero-row. By applying assumption **(A2)** we have the relation

$$\begin{aligned}
|y' \Lambda y| &= |\tilde{y}' \tilde{\Lambda} \tilde{y}| = \sum_{J \in \mathbb{D}_{p-1}^r} |y'_J y_J| \cdot \prod_{i \in J} \lambda_i \\
&= \sum_{J \in \mathbb{D}_p^r, p \notin J} |y'_J y_J| \cdot \prod_{i \in J} \lambda_i + \underbrace{\sum_{J \in \mathbb{D}_p^r, p \in J} |y'_J y_J| \cdot \prod_{i \in J} \lambda_i}_{=0} = \sum_{J \in \mathbb{D}_p^r} |y'_J y_J| \cdot \prod_{i \in J} \lambda_i
\end{aligned}$$

which proves the lemma in this case.

Next assume that  $(y_{p1}, \dots, y_{pr}) \neq 0$ , and choose a full rank  $r \times r$ -matrix  $Q$ , so that  $(y_{p1}, \dots, y_{pr})Q = (0, \dots, 0, 1)$  Then define the  $(p-1) \times (r-1)$  matrix  $\tilde{z}$  as the first  $r-1$  columns of  $\tilde{y}Q$ . Then it holds that

$$\begin{aligned}
|Q|^2 |y' \Lambda y| &= \left| Q' \tilde{y}' \tilde{\Lambda} \tilde{y} Q + \begin{pmatrix} 0_{r-1 \times r-1} & 0 \\ 0 & \lambda_p \end{pmatrix} \right| \\
&= |Q' \tilde{y}' \tilde{\Lambda} \tilde{y} Q| + |\tilde{z}' \tilde{\Lambda} \tilde{z}| \lambda_p.
\end{aligned} \tag{18}$$

Applying assumption **(A2)** on the first term of (18) we get

$$\left| Q' \tilde{y}' \tilde{\Lambda} \tilde{y} Q \right| = |Q|^2 \sum_{J \in \mathbb{D}_{p-1}^r} |\tilde{y}'_J \tilde{\Lambda}_J \tilde{y}_J| = |Q|^2 \sum_{J \in \mathbb{D}_p^r, p \notin J} |y'_J \Lambda_J y_J|. \quad (19)$$

Note that for  $J \in \mathbb{D}_{p-1}^{r-1}$  we have that

$$|\tilde{z}_J| = \left| \begin{pmatrix} \tilde{z}_J & 0 \\ 0 & 1 \end{pmatrix} \right| = |y_{\tilde{J}} Q|, \text{ and } \lambda_p |\tilde{\Lambda}_J| = |\Lambda_{\tilde{J}}|$$

where  $\tilde{J} = \{J \cup \{p\}\} \in \mathbb{D}_p^r$ . So applying assumption **(A1)** to the second term of (18) we have

$$\left| \tilde{z}' \tilde{\Lambda} \tilde{z} \right| \lambda_p = \lambda_p = |Q|^2 \sum_{J \in \mathbb{D}_p^{r-1}, p \in J} |y'_J \Lambda_J y_J|. \quad (20)$$

Combining the identities (18), (19), and (20) we have shown

$$|Q|^2 |y' \Lambda y| = |Q|^2 \sum_{J \in \mathbb{D}_p^r, p \notin J} |y'_J \Lambda_J y_J| + |Q|^2 \sum_{J \in \mathbb{D}_p^r, p \in J} |y'_J \Lambda_J y_J| = |Q|^2 \sum_{J \in \mathbb{D}_p^r} |y'_J \Lambda_J y_J|$$

which completes the proof. ■

In the proof for Lemma 10 we obtained a representation for  $|y' \Lambda y|$  which we formulate as a separate corollary.

**Corollary 11** *Let  $\Lambda$  be a real  $p \times p$  diagonal matrix, and  $y$  a real  $p \times r$  matrix, where  $r \leq p$ . Then with the definitions above, we have that*

$$|y' \Lambda y| = \sum_{J \in \mathbb{D}_p^r} |y'_J \Lambda_J y_J| = \sum_{J \in \mathbb{D}_p^r} |y'_J y_J| \prod_{i \in J} \lambda_i = \sum_{J \in \mathbb{D}_p^r} |y_J|^2 \prod_{i \in J} \lambda_i.$$

**Lemma 12** *Let  $x$  be a  $p \times r$  matrix,  $M$  and  $N$  be  $p \times p$  symmetric matrices,  $M$  positive semi-definite and  $N$  positive definite.*

*The function  $f(x) = |x' M x| / |x' N x|$  has  $\prod_{i=1}^r \lambda_i$  as its maximum with is obtained for  $x = (v_1, \dots, v_r)$  where  $v_1, \dots, v_r$  are eigenvectors corresponding to the  $r$  largest eigenvalues,  $\lambda_1, \dots, \lambda_r$  from the eigenvalue problem  $|\lambda N - M| = 0$ .*

**Proof.** The matrix  $\left( N^{-\frac{1}{2}} M N^{-\frac{1}{2}} \right)$  is symmetric positive semi-definite, hence we can diagonalize it as  $N^{-\frac{1}{2}} M N^{-\frac{1}{2}} = Q \Lambda Q'$  where  $Q Q' = I$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  and  $\lambda_1 > \lambda_2 > \dots >$

$\lambda_p \geq 0$ . By defining  $V = N^{-\frac{1}{2}}Q$  and  $y = V^{-1}x$ , we have that  $|x'Mx| / |x'Nx| = |y'\Lambda y| / |y'y|$ . According to Lemma 10 this is maximized by  $\hat{y} = (I_r, 0)'$ , so  $f(x)$  is maximized by  $\hat{x} = V\hat{y} = N^{-\frac{1}{2}}Q\hat{y}$ . ■

**Proof of Theorem 5.**

The likelihood function is given by

$$\begin{aligned} L(\alpha, \beta, \Psi, \Omega) &= \prod_{t=1}^T ((2\pi)^p |\Omega|)^{-\frac{1}{2}} \\ &\times \exp\left(-\frac{1}{2}(Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t})'\Omega^{-1}(Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t})\right). \end{aligned}$$

The estimate of the parameters are found by maximization of the likelihood function, or equivalently by maximization of the logarithm of the likelihood function

$$\begin{aligned} \log L(\alpha, \beta, \Psi, \Omega) &= -\frac{T}{2}|\Omega| - \frac{T}{2}\log(2\pi)^p \\ &- \frac{1}{2}\sum_{t=1}^T (Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t})'\Omega^{-1}(Z_{0t} - \alpha\beta'Z_{1t} - \Psi Z_{2t}). \end{aligned}$$

The maximization is done in three steps. First, we maximize with respect to  $\Psi$  taking  $\alpha$  and  $\beta$  as given, then with respect to  $\alpha$  and  $\Omega$  taking  $\beta$  as given, and finally with respect to  $\beta$ .

The estimate of  $\Psi$ , given  $\alpha$  and  $\beta$ , is found by regressing  $(Z_{0t} - \alpha\beta'Z_{1t})$  on  $Z_{2t}$ , with the Gaussian error term, the estimate is found by OLS

$$\hat{C}(\alpha, \beta) = M_{02}M_{22}^{-1} - \alpha\beta'M_{12}M_{22}^{-1}, \quad (21)$$

where  $M_{ij} = T^{-1}\sum_{t=1}^T Z_{it}Z'_{jt}$ . The concentrated likelihood function is given by

$$\log L(\alpha, \beta, \Omega) = -\frac{T}{2}|\Omega| - \frac{T}{2}\log(2\pi)^p - \frac{1}{2}\sum_{t=1}^T (R_{0t} - \alpha\beta'R_{1t})'\Omega^{-1}(R_{0t} - \alpha\beta'R_{1t}),$$

where the auxiliary residuals ( $Z_{0t}$  and  $Z_{1t}$  corrected for  $Z_{2t}$ ) are given by  $R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}$  and  $R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t}$ .

Taking  $\beta$  as given, the estimates of  $\alpha$  and  $\Omega$  are given by

$$\hat{\alpha}(\beta) = S_{01}\beta(\beta'S_{11}\beta)^{-1} \quad (22)$$

$$\hat{\Omega}(\beta) = S_{00} - S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta'S_{10}, \quad (23)$$

again using that the errors are Gaussian.

What remains is to maximize the concentrated likelihood function with respect to  $\beta$ .

Since

$$T^{-1} \sum_{t=1}^T (R_{0t} - \hat{\alpha}(\beta)\beta' R_{1t})' (\Omega(\beta))^{-1} (R_{0t} - \hat{\alpha}(\beta)\beta' R_{1t}) = I,$$

the concentrated likelihood is given by

$$L(\beta) = \left( (2\pi)^p |\hat{\Omega}(\beta)| \right)^{-\frac{T}{2}} \exp \left( -\frac{1}{2} T p \right) = \left( (2\pi e)^p |\hat{\Omega}(\beta)| \right)^{-\frac{T}{2}}.$$

So maximizing the likelihood function is equivalent to minimizing

$$|\hat{\Omega}(\beta)| = |S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}| = |S_{00}| \frac{|\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta' S_{11}\beta|},$$

which is solved by choosing the  $r$  smallest eigenvalues of  $|S_{11}\rho - (S_{11} - S_{10}S_{00}^{-1}S_{01})|$ , or be defining  $\lambda = 1 - \rho$ , choosing the  $r$  largest eigenvalues of  $|S_{11}\rho - S_{10}S_{00}^{-1}S_{01}|$ , which is identical to solve  $\max_{\beta} \frac{|\beta'(S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta' S_{11}\beta|}$ . By Lemma 12 the estimator is given by

$$\hat{\beta} = (\hat{v}_1, \dots, \hat{v}_r),$$

where  $\hat{\lambda}_i$  and  $\hat{v}_i$  are the eigenvalues and eigenvectors to the problem

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

ordered such that  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$ , and we find

$$|\hat{\Omega}(\hat{\beta})| = |S_{00}| \prod_{i=1}^r (1 - \hat{\lambda}_i).$$

Since the eigenvectors are normalized by  $(\hat{v}_1, \dots, \hat{v}_p)' S_{11} (\hat{v}_1, \dots, \hat{v}_p) = I$ , we have  $\hat{\beta}' S_{11} \hat{\beta} = I$ , such that (22) and (23) reduces to (5) and (6). By inserting these estimates into (21) we find (7). ■

## A.2 Algebraic Treatment of the Generalized Reduced Rank Regression

Before we can formulate the general estimation result we need some additional notation. Define  $Z_0 = (Z_{01}, \dots, Z_{0T})$ ,  $Z_1 = (Z_{11}, \dots, Z_{1T})$ ,  $Z_2 = (Z_{21}, \dots, Z_{2T})$ , and  $E = (\varepsilon_1, \dots, \varepsilon_T)$ , so that Model 3 can be expressed as

$$Z_0 = AB'Z_1 + CZ_2 + E. \quad (24)$$

Next define

$$\begin{aligned} \mathbf{Z}_{1B2} &= ((Z_1' B, Z_2') \otimes I_p), \\ \mathbf{Z}_{1A} &= (Z_1' \otimes A) K_{p_1, r}, \end{aligned}$$

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 let  $\varepsilon = \text{vec}(\varepsilon_1, \dots, \varepsilon_T)$  and set

$$\Sigma = \text{var}(\varepsilon),$$

which is block diagonal under Assumption 1. The  $p \times p$  matrices in the diagonal of  $\Sigma$  are given by  $\Omega(t)$ ,  $t = 1, \dots, T$ , formally  $\Sigma_{p(t-1)+i, p(t-1)+j} = \Omega_{i, j}(t)$  for  $i, j = 1, \dots, p$  and  $t = 1, \dots, T$ . Hence  $\Sigma^{-1}$  is a block diagonal matrix with  $\Omega(t)^{-1}$  as diagonal matrices,  $t = 1, \dots, T$ .

**Lemma 13** *With the definitions above, we have the relations:*

$$\mathbf{Z}'_{1A} \Sigma^{-1} \mathbf{Z}_{1A} = \sum_{t=1}^T [A' \Omega(t)^{-1} A \otimes Z_{1t} Z'_{1t}], \quad (25)$$

$$\mathbf{Z}'_{1A} \Sigma^{-1} \text{vec}(Z_0 - CZ_2) = \sum_{t=1}^T \text{vec}(Z_{1t} (Z_{0t} - CZ_{2t})' \Omega(t)^{-1} A), \quad (26)$$

$$\mathbf{Z}'_{1B2} \Sigma^{-1} \mathbf{Z}_{1B2} = \sum_{t=1}^T \left[ \begin{pmatrix} B' Z_{1t} Z'_{1t} B & B' Z_{1t} Z'_{2t} \\ Z_{2t} Z'_{1t} B & Z_{2t} Z'_{2t} \end{pmatrix} \otimes \Omega(t)^{-1} \right], \quad (27)$$

$$\mathbf{Z}'_{1B2} \Sigma^{-1} \text{vec}(Z_0) = \sum_{t=1}^T \text{vec}(\Omega(t)^{-1} Z_{0t} (Z'_{1t} B, Z'_{2t})), \quad (28)$$

If  $\{\varepsilon_t\}$  is i.i.d. Gaussian with covariance matrix  $\Omega$ , the expressions simplify to:

$$\begin{aligned}
\mathbf{Z}'_{1A}\Sigma^{-1}\mathbf{Z}_{1A} &= T [A'\Omega^{-1}A \otimes M_{11}], \\
\mathbf{Z}'_{1A}\Sigma^{-1}\text{vec}(Z_0 - CZ_2) &= T\text{vec}((M_{10} - M_{02}C')\Omega^{-1}A), \\
\mathbf{Z}'_{1B2}\Sigma^{-1}\mathbf{Z}_{1B2} &= T \left[ \begin{pmatrix} B'M_{11}B & B'M_{12} \\ M_{21}B & M_{22} \end{pmatrix} \otimes \Omega^{-1} \right], \\
\mathbf{Z}'_{1B2}\Sigma^{-1}\text{vec}(Z_0) &= T\text{vec}(\Omega^{-1}(M_{01}B, M_{02})).
\end{aligned}$$

**Proof.** The identity

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

which proves (25). We used that  $Z_{1t}$  is a column vector, and that we can write  $\Omega(t)^{-1} = 1 \otimes \Omega(t)^{-1}$  and the formula  $(M_1 \otimes M_2)(M_3 \otimes M_4) = (M_1M_3 \otimes M_2M_4)$  for matrices where the product  $M_1M_3$  and  $M_2M_4$  are well defined.

Next consider

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

which proves (26). Equations (27) and (28) are proven similarly.

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



equations. ■

**Proof of Theorem 6.** Applying the vec operation to equation (24) yields the equation

$$\begin{aligned}
\text{vec}(Z_0) &= (Z'_1 B \otimes I_p) \text{vec}(A) + (Z'_2 \otimes I_p) \text{vec}(C) + \varepsilon \\
&= [(Z'_1 B, Z'_2) \otimes I_p] \text{vec}(A, C) + \varepsilon \\
&= \mathbf{Z}_{1B2} G' \psi + \varepsilon.
\end{aligned}$$

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

$$\text{vec}(\hat{A}, \hat{C}) = G [G' \mathbf{Z}'_{1B2} \Sigma^{-1} \mathbf{Z}_{1B2} G]^{-1} G' \mathbf{Z}'_{1B2} \Sigma^{-1} \text{vec}(Z_0),$$

which by Lemma 13 simplifies to (8).

Similarly for fixed  $A$ ,  $C$ , and  $\Sigma$ , we have the equation

$$\begin{aligned}
\text{vec}(Z_0 - CZ_2) &= \text{vec}(AB'Z_1) + \varepsilon \\
&= (Z'_1 \otimes A) \text{vec}(B') + \varepsilon \\
&= (Z'_1 \otimes A) K_{p_1, r} \text{vec}(B) + \varepsilon \\
&= \mathbf{Z}_{1A} \text{vec}(B) + \varepsilon.
\end{aligned}$$

This is also a restricted GLS problem, with the solution given by

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

which by Lemma 13 reduces to (9). ■

**Proof of Corollary 7.** Follows from Theorem 6 and Lemma 13. ■

**Proof of Corollary 8.** From Theorem 5, we obtain the equations 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

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

holding  $\Omega$  fixed. So the derivatives of  $A$  and  $B$  in the directions  $a$  and  $b$  are given by

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

and

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

using Theorem 3 from Magnus and Neudecker (1988, Chapter 2). So equations (10) and (11) are the first order conditions. ■

**Proof of Corollary 9.** The result follows directly from Theorem 5 and Corollary 8. ■

## Acknowledgements

This paper has benefitted from many valuable comments from James D. Hamilton and Søren Johansen, Tom Engsted, Graham Elliott, Niels Haldrup, David Hendry, and Juan Toro. All errors remain my responsibility. 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.
- 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. Federal Reserve Bank of Atlanta Working Paper. 1997-13a. (March 1998).
- 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., 1998. Estimating Restricted Cointegrating Vectors. UCSD manuscript.
- 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, H., and S. Johansen, 1999. Some Tests for Parameter Constancy in Cointegrated VAR-Models. *Econometrics Journal*, 2, 306–333.
- Hansen, P. R., 2000a. Generalized Reduced Rank Regression. Brown University Economics Working Paper.
- Hansen, P. R., 2000b. The Johansen-Granger Representation Theorem: An Explicit Expression for  $I(1)$  Processes. Brown University Economics Working Paper.

- Hansen, P. R., 2000c. Structural Changes in Cointegrated Processes. Ph.D. thesis, University of California at San Diego.
- 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., 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.
- 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.
- White, H., 2000. Asymptotic Theory for Econometricians, revised edn. Academic Press, San Diego.