Structural Breaks in the Cointegrated Vector Autoregressive Model

Peter Reinhard Hansen[†]

Department of Economics

University of California, San Diego

November, 1999

Abstract

We generalize the cointegrated vector autoregressive model of Johansen (1988) to allow for structural breaks. We derive the likelihood ratio test for structural breaks occurring at fixed points in time, and show that it is asymptotically χ^2. Moreover, we show how inference can be made when the null hypothesis is presence of structural breaks.

The estimation technique derived for this purpose can be applied to several other generalizations of the standard model, beyond the structural breaks treated here. For example, the new technique can be applied to estimate models with heteroskedasticity.

We apply our generalized model to US term structure data, accounting for structural breaks that coincide with the changes in the Fed's policy in September 1979 and October 1982. Contrary to previous findings we cannot reject the long-run implications of the expectations hypothesis.

[†]This paper has benefitted from many valuable comments from my supervisor James D. Hamilton and Søren Johansen, Tom Engsted, Graham Elliott and David Hendry. Any errors are mine alone. Financial support from the Danish Social Science Research Council and the Danish Research Academy is gratefully acknowledged.

1. Introduction

The modelling of structural breaks in cointegrated processes has been addressed by several authors. In the vector autoregressive framework, Seo (1998) derived the $\mathcal{LM}$ test for structural breaks 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 breaks in cointegrated processes are the recursive estimation to identify structural breaks by Hansen and Johansen (1993), the combination of cointegration and Markov switching by Krolzig (1996), the co-breaking theory by Hendry (1995), and test for no cointegration in processes with a structural break by Gregory and Hansen (1996).

One of the main contributions of this paper is the development of a flexible framework in which structural breaks can be formulated. The most related paper is the one of Seo (1998), who considered structural breaks in cointegration relations and adjustment coefficients, under i.i.d. assumptions. The framework proposed here can handle a class of breaks in integrated processes that are more general than previously treated. Partial structural breaks[1] such as, a structural break in a particular cointegration relation or 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 $\mathcal{LR}$ test and it is shown that its asymptotic distribution is standard χ^2 when the break point is taken as given[2]. Another contribution of this paper is that it enables hypotheses testing under the maintained hypothesis that the underlying process exhibits structural breaks. The asymptotic χ^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 arises from structural breaks.

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, β, and adjustment coefficients, α. Johansen and Juselius (1992) proposed a switching algorithm for estimation under slightly more general restrictions. Boswijk (1995) derived a general estimation technique that can handle any linear restriction on $\text{vec}(\alpha)$ and $\text{vec}(\beta)$, where $\text{vec}(\cdot)$ is the vectorization operator.

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

[2] The case of an unknown break point leads to a non-standard asymptotic distribution. See Seo (1998) or Andrews and Ploberger (1994). We treat this aspect in a separate paper.

The estimation technique of Boswijk (1995) is applicable to several estimation problems we face with structural breaks 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" 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 breaks have been accounted for.

The paper is organized as follows. Section 2 contains the statistical formulation of various structural breaks 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 breaks and it is shown how these breaks can be formulated as parameter restrictions in a unified framework.

2.1. The Cointegrated Vector Autoregressive Model

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

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t. \tag{2.1}$$

This process can be inverted to an infinite moving average representation, also known as the Granger representation, (see, for example Hansen and Rahbek (1999)). The representation shows (i) how the adjustment coefficient, α, relates to the common stochastic trends in the process and (ii) that β defines the cointegration relations.

It is convenient to rewrite the model as

$$Z_{0t} = \alpha\beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t, \tag{2.2}$$

where $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}$, $Z_{2t} = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}, D'_t)'$ and $\Psi = (\Gamma_1, \ldots, \Gamma_{k-1}, \Phi)$, 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. We shall therefore denote the dimension of Z_{1t} by p_1 rather than p, which denotes the dimension of Z_{0t}. The regression problem in equation (2.2), 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:

$$\begin{aligned} Z_{0t} &= AB' Z_{1t} + C Z_{2t} + \varepsilon_t, \\ \text{s.t. } \operatorname{vec}(A, C) &= G\psi, \\ \operatorname{vec}(B) &= H\varphi, \end{aligned} \tag{2.3}$$

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

Assumption 2.1. *$\{\varepsilon_t\}$ is a sequence of independent p-dimensional Gaussian variables, where ε_t is independent of Z_{1t} and Z_{2t} and has the marginal distribution $N(0, \ (t))$.*

By this formulation the i.i.d. assumption on $\{\varepsilon_t\}$ is relaxed, by no longer requiring an identical distribution. We leave the exact structure of (t), $t = 1, \ldots, T$ to be determined from model-specific assumptions on heteroskedasticity. The assumption still implies independence of $\{\varepsilon_t\}$.

Estimation and inference under a weaker assumption than Assumption 2.1 is treated in a separate paper (See Hansen (1999)).

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 $\quad(t) = \quad$.

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

Assumption 2.2. *The matrices H and G in (2.3) 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 Lebesque measure zero, (n denotes the number of column in (H, G)).*

Let the covariance parameters be expressed as $\quad(t) = \quad_t(\theta)$, $\theta \in \Theta_\theta$, $t = 1, \ldots, T$. This formulation does not necessarily impose any restrictions on the parameters.

Assumption 2.3. *The parameters ψ, φ and θ are variation free, that is*

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

This assumption is convenient for our parameter estimation. Suppose that Assumption 2.3 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.

All the models we consider in this paper satisfy Assumption 2.3. An example of a model that does not satisfy this assumption is the GARCH model. This model has a dependence between the

parameter space of the covariance matrix, typically denoted by H_t, and the other parameters, due to the dependence of $\hat{H}_t$ on the estimated residuals such as $\hat{\varepsilon}_{t-1}$. The failure of Assumption 2.3 to hold for GARCH models is part of the explanation for why GARCH models can be difficult to estimate.

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 2.4. *The function $f(x, y) = xy'$, where x is $p \times r$ $(r \square p)$ and y is $p_1 \times r$ $(r \square p_1)$, 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 Breaks in the Cointegrated Vector Autoregressive Model

We now show that structural breaks in model (2.1) can be viewed as a particular form of (2.3). Without loss of generality, we can focus just on breaks in α and β, because breaks in the parameters $\Gamma_1, \ldots, \Gamma_{k-1}$ or Φ in (2.1) are easily handled by redefining Z_{2t} and Ψ. For now we keep the covariance matrix, , constant, but later we also generalize the model to allow structural breaks in this parameter. Letting all parameters change their value is easily treated 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 increase too dramatically.

So, the generalization of model (2.1) that we consider is

$$Z_{0t} = \alpha(t)\beta(t)'Z_{1t} + \Psi Z_{2t} + \varepsilon_t. \tag{2.4}$$

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

$$\alpha(t)\beta(t)' = \alpha_1\beta_1' I_{1t} + \cdots + \alpha_q\beta_q' I_{qt} \tag{2.5}$$

where I_{jt}, $j = 1, \ldots, q$ are indicator functions that determine which α_j and β_j are active. This formulation does not require α_i and α_j to have the same number of columns $i \neq j$, as long as α_j and β_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 subsamples while other relations change.

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

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

with block diagonal structure of the matrix containing the cointegration relations, denoted by B. This structure can be expressed as a linear restriction on $\mathrm{vec}(B) = H\varphi$, and the regression is therefore a special case of equation 2.3.

2.2.1. Structural Breaks in α and β

Consider a situation with $q-1$ structural breaks that occur at time $T_1, \ldots, T_{q-1}$, so that α_t and β_t can take on q different values. This can be formulated as

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

and

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

So in this case we define $Z_{11t} = Z_{1t} I(t \square T_1)$, $Z_{12t} = Z_{1t} I(T_1 + 1 \square t \square T_2)$, $\ldots$, $Z_{1qt} = Z_{1t} I(T_{q-1} + 1 \square t \square T)$ and $\tilde{Z}_{1t} = (Z'_{11t}, \ldots, Z'_{1qt})'$, and obtain a model with the form of equation (2.3). This formulation allows for a change in the number of cointegration relations. Let r_i denote

the cointegration rank in subsample i, $i = 1, \ldots, q$. Then the dimension of the parameter space of $\Pi(t) = \alpha(t)\beta(t)'$ is by Lemma 2.4 found to be $\sum_{i=1}^{q}(p + p_1 - r_i)r_i$ where r_i is the rank of $\alpha_i\beta_i'$, $i = 1, \ldots, q$. 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 Breaks in the Adjustment Coefficients: α

If the structural breaks only affect the adjustment coefficients, α, whereas the cointegration relations remain constant, we can express the model as

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

where $\tilde{Z}_{1t}$ is as defined above. Since β 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 Breaks in the Cointegrating Relations: β

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

$$\begin{aligned} Z_{0t} &= \alpha\beta_1' I_{1t} Z_{1t} + \cdots + \alpha\beta_q' I_{qt} Z_{1t} + \Psi Z_{2t} + \varepsilon_t \\ &= \alpha\left(\beta_1', \ldots, \beta_q'\right) \tilde{Z}_{1t} + \Psi Z_{2t} + \varepsilon_t, \end{aligned}$$

where $\tilde{Z}_{1t}$ is as defined previously. Here we again obtain an equation of the form of (2.3), but in this case without the additional restrictions A and B, i.e. $G = I_{pr}$, and $H = I_{p_1 rq}$. 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 breaks are displayed in Figure 2.1, along with the relevant asymptotic distribution and degrees of freedom. The asymptotic distribution is derived below, and in not surprisingly found to be asymptotically χ^2.

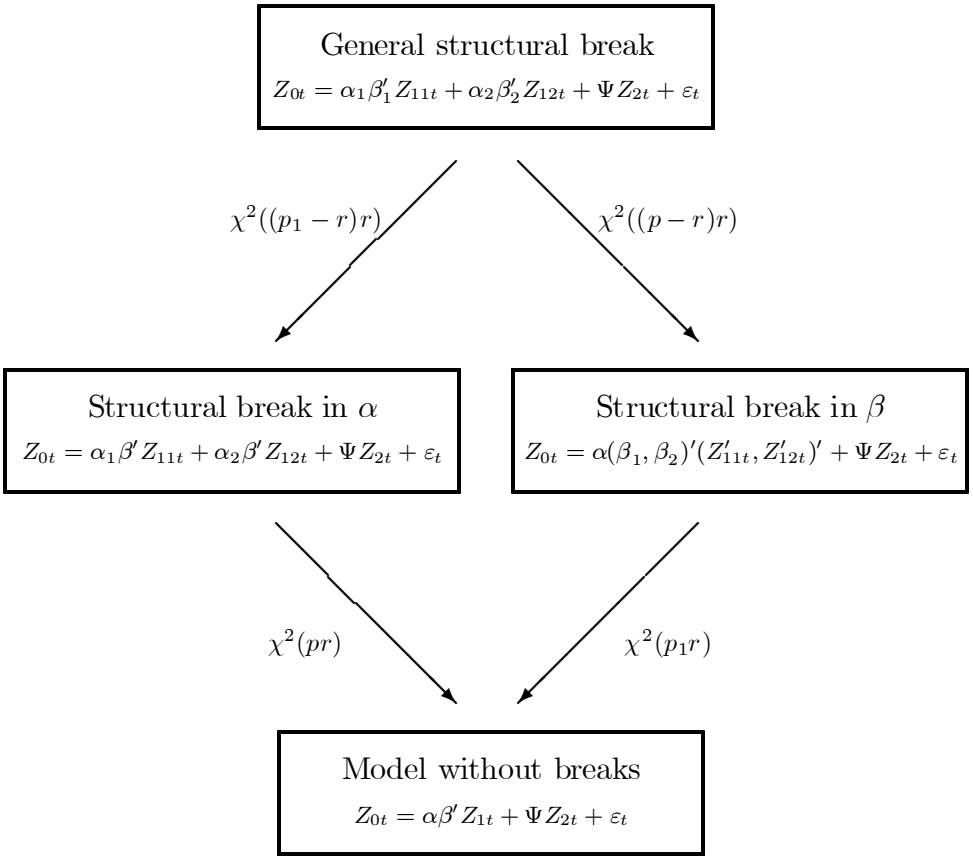

Figure 2.1: The relations between the different types of structural breaks. The asymptotic distribution of the individual $\mathcal{LR}$ test is χ^2 in all cases, with the degrees of freedom reported in the brackets.

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 equation (2.3). The simplest situation is where there are r_1 permanent cointegration relations, say β_1, and for $t \geq T_1 + 1$ there are an additional $r_2 - r_1$ temporary cointegration relations, say β_e, (linearly independent of β_1). This situation leads to two different cases – one where the adjustment coefficients corresponding to β_1 remain constant, and one where they may differ in the two subsamples. The latter is likely 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.

First we consider the case where α_1 remains constant. This model is formulated as

$$Z_{0t} \quad = \quad \alpha_1\beta_1' Z_{1t} + \alpha_e\beta_e' Z_{1t} I_{(t>T_1)} + \Psi Z_{2t} + \varepsilon_t$$

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

and the dimension of the parameter space for $\Pi(t)$ is slightly more complicated to derive. The degrees of freedom in Π_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 Π_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

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

which is also of the form of equation (2.3), 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 Π_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 β_e may be chosen orthogonal to β_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 most general 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 2.2. Below we prove that the likelihood ratio test for this hypothesis is asymptotically χ^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 Breaks in the Covariance Matrix

Structural breaks in the covariance matrix also leads to a GRRR. The simplest case is a single structural break in the covariance matrix at time T_1. So $\mathrm{var}(\varepsilon_t) =$ $_1$ for $t \square T_1$ and $\mathrm{var}(\varepsilon_t) =$ $_2$

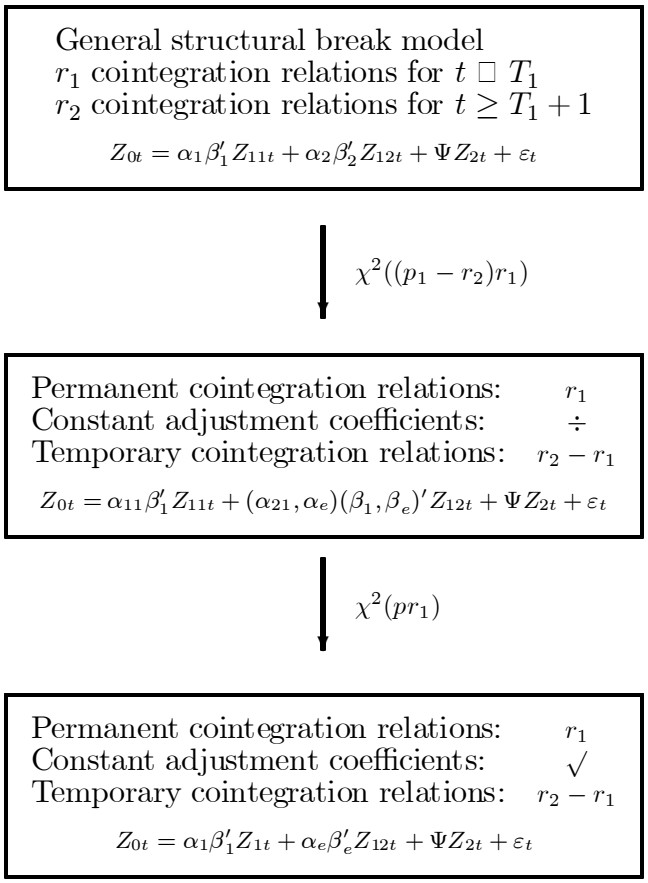

Figure 2.2: The relations between the different models with structural breaks and a shift in the number of cointegration relations. The distribution of the $\mathcal{LR}$ test statistic between two of the models is asymptotically χ^2 with the degrees of freedom given in the figure.

for $t > T_1$, which implies the following structure on the covariance matrix

$$\Sigma = \begin{pmatrix} I_{T_1} & {}_1 & 0 \\ 0 & I_{T-T_1} & {}_2 \end{pmatrix}.$$

The combination of structural breaks in the covariance matrix as well as other parameters, will clearly also lead to a GRRR.

2.2.6. Linear Restriction on Adjustment Coefficients and Cointegration Relations

Combining hypotheses of structural breaks 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 $\text{vec}(B) = H\varphi$ for a known matrix H and some parameters φ. Adding linear restrictions to the adjustment coefficients, $\alpha_1, \ldots, \alpha_q$ can be formulated as

$\text{vec}(A) = G\psi$, and is therefore also a GRRR.

3. Estimation

Estimation of the cointegrated vector autoregressive model, and other models that have the structure of equation (2.2), 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 applied to the $I(1)$ model by Johansen (1988).

The advantage of reduced rank estimation is that an explicit solution is obtained without iterations. Fortunately this method is applicable to estimation under simple linear restrictions on the reduced rank parameters. However, in most of the structural break models we face restrictions that are beyond what the 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. Even more problematic is that general convergence to the global optimum cannot be proven; indeed it is easy to construct examples where the method will not converge.

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 in that it can handle estimation problems with linear restrictions on $\text{vec}(B)$ and $\text{vec}(A)$. This method is therefore sufficient for most of the estimation problems that arise from structural break models. Applications of the method have shown that convergence is obtained in few iterations, and that it does converge to the global optimum.

More general yet is the minimum distances approach by Elliott (1997, 1998a), which can estimate parameters under the general restriction $g(\theta) = c$, where θ is the vector of parameters, c is a constant and g is a well-behaved function. This method minimizes $\theta' \hat{V}_{\hat{\theta}} \theta$ subject to the constraints $g(\theta) = c$, where $\hat{V}_{\hat{\theta}}$ is an estimate of the asymptotic covariance matrix. This method is asymptotically equivalent to the maximum likelihood estimation, and with suitable choice of $\hat{V}_{\hat{\theta}}$ and if applied iteratively, (by recursive reestimation of $\hat{V}_{\hat{\theta}}$ as the estimate of θ 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 (2.3) we obtain the same generality as the minimum distance method by Elliott (1997, 1998a), and can in addition estimate models with heteroskedasticity.

In the following we consider the reduced rank regression model

$$Z_{0t} = AB'Z_{1t} + CZ_{2t} + \varepsilon_t, \tag{3.1}$$

with various restrictions on the parameters, under Assumptions 2.1 and 2.2. 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 $M_{ij} = \frac{1}{T}\sum_{t=1}^{T} Z_{it}Z'_{jt}, \quad i,j = 0,1,2$, 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 $S_{ij} = \frac{1}{T}\sum_{t=1}^{T} R_{it}R'_{jt}, \quad i,j = 0,1$.

3.1. Reduced Rank Regression

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

Theorem 3.1 (Reduced Rank Regression). *The unrestricted estimators of Model (2.3) are given by*

$$\begin{aligned}
\hat{B} &= (\hat{v}_1, \ldots, \hat{v}_r)\phi & (3.2)\\
\hat{A}(B) &= S_{01}B\left(B'S_{11}B\right)^{-1} & (3.3)\\
\hat{\Omega} &= S_{00} - S_{01}BA' + AB'S_{11}BA' - AB'S_{10}, & (3.4)\\
\hat{C} &= M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}, & (3.5)
\end{aligned}$$

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

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

and where ϕ 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

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

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 equation 2.1. The maximum likelihood estimate is obtained by defining $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}$ and $Z_{2t} = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}, D'_t)'$.

3.2. Generalized Reduced Rank Regression

Theorem 3.2. *Let the parameter A, B and C be restricted by* $\mathrm{vec}(A, C) = G\psi$ *and* $\mathrm{vec}(B) = H\phi$ *and suppose that Assumptions 2.1, 2.2 and 2.3 hold.*

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

$$\begin{aligned}
\mathrm{vec}(\hat{A}, \hat{C}) &= G\left[G'\sum_{t=1}^{T}\left[\begin{pmatrix} \hat{B}'Z_{1t}Z'_{1t}\hat{B} & \hat{B}'Z_{1t}Z'_{2t} \\ Z_{2t}Z'_{1t}\hat{B} & Z_{2t}Z'_{2t}\end{pmatrix} \quad \hat{\ }(t)^{-1}\right] G\right]^{-1} \qquad (3.6)\\
&\quad \times G'\sum_{t=1}^{T}\mathrm{vec}\left(\hat{\ }(t)^{-1}Z_{0t}(Z'_{1t}\hat{B}, Z'_{2t})\right),\\
\mathrm{vec}(\hat{B}) &= H\left[H'\sum_{t=1}^{T}\left[\hat{A}'\hat{\ }(t)^{-1}\hat{A} \quad Z_{1t}Z'_{1t}\right] H\right]^{-1} \qquad (3.7)\\
&\quad \times H'\sum_{t=1}^{T}\mathrm{vec}\left(Z_{1t}(Z_{0t} - \hat{C}Z_{2t})'\hat{\ }(t)^{-1\prime}\hat{A}\right)
\end{aligned}$$

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

$$\hat{\theta} = \arg\max_{\theta\in\Theta_\theta} L(\hat{A}, \hat{B}, \hat{C}, \theta, Z_0, Z_1, Z_2). \qquad (3.8)$$

The maximum value of the likelihood function is given by

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

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 keeping $(A, C, \ (t))$ or $(B, \ (t))$ fixed. 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 2.3, this procedure

will converge to parameter values that satisfy the normal equations.

We now treat situations with fewer parameter restrictions.

Corollary 3.3. *Let the parameter A, B and C be restricted by* $\mathrm{vec}(A, C) = G\psi$ *and* $\mathrm{vec}(B) = H\phi$ *and suppose that* $\{\varepsilon_t\}$ *is i.i.d. Gaussian* $N(0, \quad)$.

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

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

The maximum value of the likelihood function is given by

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

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

Corollary 3.4. *Let A and B be restricted by* $\mathrm{vec}(A) = G\zeta$ *and* $\mathrm{vec}(B) = H\phi$, *for known matrices G and H. Then the maximum likelihood estimates satisfy the equations*

$$\mathrm{vec}(\hat{B}) = H\left[H'\left(\hat{A}'\hat{}^{-1}A \quad S_{11}\right)H\right]^{-1} H'\left(\hat{A}' \quad S_{10}\right)\mathrm{vec}\left(\hat{}^{-1}\right) \tag{3.9}$$

and

$$\begin{aligned}
\hat{A} &= G\left[G'\left(\hat{B}'S_{11}\hat{B} \quad I_p\right)G\right]^{-1} G'\left(\hat{B}' \quad \hat{}^{-1}\right)\mathrm{vec}\left(S_{01}\right) \qquad (3.10) \\
\hat{} &= S_{00} - S_{01}\hat{B}\hat{A}' + \hat{A}\hat{B}'S_{11}\hat{B}\hat{A}' - \hat{A}\hat{B}'S_{10} \\
\hat{C} &= M_{02}M_{22}^{-1} - \hat{A}\hat{B}'M_{12}M_{22}^{-1}.
\end{aligned}$$

The maximum value of the likelihood function is given by

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

Corollary 3.5. *Let B be restricted by* $\mathrm{vec}(B) = H\phi$. *Then the maximum likelihood estimates*

satisfy the equations

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

The maximum value of the likelihood function is given by

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

With these results we have the tools available to estimate the parameters in the cointegrated vector autoregressive model under all the various structural breaks 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 breaks, for example models with heteroskedasticity.

3.3. Applicability

Example 3.6 (Structural breaks in the covariance matrix). *Consider the cointegrated vector autoregressive model (equation (2.1)), with a structural break at time T_1, in the sense that $\alpha(t) = \alpha_1$, $\beta(t) = \beta_1$ and $\quad(t) = \quad_1$ for $t \square T_1$ and $\alpha(t) = \alpha_2$, $\beta(t) = \beta_2$ and $\quad(t) = \quad_2$ for $t \geq T_1 + 1$. This estimation problem can be written in the form of Model 2.3. The maximum likelihood estimators of $\quad_1$ and $\quad_2$ are given by*

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

So $\quad(t)$, $t = 1, \ldots, T$ can be expressed in the functional form required by Theorem 3.2.

Example 3.7 (Heteroskedasticity). *Models with the following type of heteroskedastic errors*

$$\mathrm{var}(\varepsilon_t) = \quad_t = f_\theta(\quad_{t-1}, \quad_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots)$$

can be expressed with the functional form in Theorem 3.2.

4. Asymptotic Analysis

For simplicity, we derive the asymptotic results in the case of a single structural break 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 breaks, and varying number of cointegrating relations.

The process is described by

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

where ε_t is i.i.d.[3] $N(0, \ (t))$, $(t) = \ _1$ for $t \square T_1$ and $(t) = \ _2$ 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 $\alpha_{i\perp}'(I - \Gamma_1 - \cdots - \Gamma_{k-1})\beta_{i\perp}$ has full rank $p-r$, $i = 1, 2$.

4.1. The Granger Representation for Break 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

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

and

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

where $C = \beta_{1\perp} \left(\alpha_{1\perp}' \Gamma \beta_{1\perp}\right)^{-1} \alpha_{1\perp}'$, $D = \beta_{2\perp} \left(\alpha_{2\perp}' \Gamma \beta_{2\perp}\right)^{-1} \alpha_{2\perp}'$ and $\Gamma = I - \Gamma_1 - \cdots - \Gamma_{k-1}$, (see Hansen and Rahbek (1999)).

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

[3] 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 (1999)).

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

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

where

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

is a stationary process. So altogether we have the Granger representation

$$X_t = C\sum_{i=1}^{t} \varepsilon_i + C(L)\varepsilon_t + C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \quad t = 1, \ldots, T_1 \tag{4.1}$$

$$\begin{aligned}
X_t &= D\sum_{i=T_1+1}^{t} \varepsilon_i + D\Gamma C \sum_{i=1}^{T_1} \varepsilon_i + D(L)\varepsilon_t + DC^*(L)\varepsilon_{T_1} \\
&\quad + D\Gamma C(X_0 - \sum_{i=1}^{k-1} \Gamma_i X_{0-i}) \quad t = T_1 + 1, \ldots, T.
\end{aligned} \tag{4.2}$$

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 breaks.

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

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

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

$$T^{-\frac{1}{2}} \left(\sum_{t=1}^{T_1} \varepsilon_t + \sum_{t=T_1+1}^{[Tu]} \varepsilon_t \right) \xrightarrow{w} W_1(\rho) + W_2(u) - W_2(\rho), \quad u > \rho$$

where W_1 and W_2 are stochastically independent.

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

$$\begin{aligned} T^{-\frac{1}{2}} C \sum_{t=1}^{[Tu]} \varepsilon_t &\xrightarrow{w} CW_1(u), && u \square \rho \\ T^{-\frac{1}{2}} \left(D\Gamma C \sum_{t=1}^{T_1} \varepsilon_t + D \sum_{t=T_1+1}^{[Tu]} \varepsilon_t \right) &\xrightarrow{w} D\Gamma C W_1(\rho) + D\left(W_2(u) - W_2(\rho)\right), && u > \rho. \end{aligned} \quad (4.3)$$

Equation (4.3) has an important implication for unit root tests, in processes with structural breaks. Standard Dickey-Fuller type distributions, such as $\int (dB)\, B' \left[\int BB' du\right]^{-1} \int B\, (dB)'$ do not define the asymptotic distribution in this situation, because the Gaussian term $D\Gamma C W_1(\rho)$, that comes from the initial values, does not disappear. A unit root test based on observation after a structural break will therefore involve a term such as $\int (dB)\, (B+Z)' \left[\int (B+Z)(B+Z)' du\right]^{-1} \int (B+Z)\, (dB)'$. However, this problem does not occur if an unrestricted constant is used as regressor. This aspect of structural breaks in cointegrated processes is treated in Hansen and Johansen (1999).

From the Granger representation we find that the non-vanishing term is given by

$$T^{-1/2} \begin{pmatrix} X_{[Tu]} 1_{(u \square \rho)} \\ X_{[Tu]} 1_{(u > \rho)} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} CW_1(u) 1_{(u \square \rho)} \\ \left[D\Gamma C W_1(\rho) + D\left(W_2(u) - W_2(\rho)\right)\right] 1_{(u > \rho)} \end{pmatrix}.$$

Let

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

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

$$G(u) = \begin{pmatrix} G_1(u) \\ G_2(u) \end{pmatrix} = \bar{B}'_\perp \begin{pmatrix} CW_1(u) 1_{(u \square \rho)} \\ [D\Gamma CW_1(\rho) + D\left(W_2(u) - W_2(\rho)\right)] 1_{(u>\rho)} \end{pmatrix}$$

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

$$T^{-2} \sum_{t=1}^{T} \bar{B}'_\perp \begin{pmatrix} X_{[Tu]} 1_{(u \square \rho)} \\ X_{[Tu]} 1_{(u>\rho)} \end{pmatrix} \begin{pmatrix} X_{[Tu]} 1_{(u \square \rho)} \\ X_{[Tu]} 1_{(u>\rho)} \end{pmatrix}' \bar{B}_\perp \xrightarrow{w} \int_0^1 G(u) G(u) du$$

$$= \begin{pmatrix} \int_0^\rho G_1(u) G_1(u) du & 0 \\ 0 & \int_\rho^1 G_2(u) G_2(u) du \end{pmatrix}.$$

With this notation, the asymptotic results for unrestricted parameter estimates ($G = I$ and $H = I$) of A, B, C and $\quad(t)$, say $\hat{A}_u$, $\hat{B}_u$, $\hat{C}_u$ and $\hat{}_u(t)$, 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 $\mathcal{LR}$ test of some over identifying restrictions, have a χ^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 $\quad(t)$ is constant, then

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

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

$$T\mathrm{vec}(\hat{B} - B) \xrightarrow{w} H \quad \square \quad \square \quad H' \quad A' \quad {}^{-1} A \quad B_\perp \int_0^1 G(u) G'(u) du B'_\perp \Big] H \Big]^{-1}$$
$$\times H' \mathrm{vec} \left(\int G(u) dW \quad {}^{-1\prime} A \right),$$

which is a mixed Gaussian distribution. Similarly

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

which has an Gaussian asymptotic distribution. The case with a varying (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 (1988, Theorem 13.7, 13.9), that the likelihood ratio test has an asymptotically χ^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 break 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 (at time t) earning a return from period $t+n$ to $t+n+1$, 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[4] states that

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

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

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

$n = 1, 2, \ldots$, 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

$$y_{nt} - y_{1t} = n^{-1} \sum_{j=1}^{n-1} \sum_{i=1}^{j} E_t(\Delta y_{1,t+i}) + L_n. \tag{5.1}$$

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[5], 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 (5.1) 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, γ, for the elasticity of volatility with respect to the level of the shortest interest rate. With simulations, they showed that hypothesis tests on cointegration vectors over-reject as γ increases, and found the effect to be substantial as γ increases beyond 0.5.

Whereas the expectations hypothesis has been rejected by most studies of US data (see 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 5 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] 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.

5.2. Structural Breaks in the US Term Structure of Interest Rates

There are several studies that find evidence of a structural break 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. Hansen and Johansen (1993) have developed a recursive estimation of the cointegrated vector autoregressive model to detect structural changes. Their application to US data also indicates structural breaks around the fall of 1979 and the fall of 1982.

Structural breaks 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 shift in October 1979, and then estimated the elasticity of volatility to be 1.5. However Bliss and Smith (1998) found significant structural breaks when the possibility of a structural shift by the end of 1982 is included in the analysis. They found evidence of structural breaks in both 1979 as well as in 1982 when the Fed reversed to target the Fed funds rate. After these breaks 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 breaks, and it is not surprising that these breaks affect point estimates and inference.

Elliott (1998b) 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 break in 1979:10 the hypothesis could not be rejected.

In this paper, interest rates are modelled as $I(1)$ variables. The fact that nominal interest 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[6]. 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.

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

5.3. Data

The term structure data were extracted from the Bliss data[7] 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[8] 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 break 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

$$\Delta X_t = \alpha(t)\beta(t)^{*\prime} X^*_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t,$$

where $X^{*\prime}_t = (X'_t, 1)$ and $\beta^{*\prime} = (\beta(t)', \rho(t))$.

We may normalize the cointegration relations by

$$\beta(t)^* = \begin{pmatrix} \beta_{11,t} & \beta_{12,t} & \cdots & \beta_{1r,t} \\ -1 & 0 & & 0 \\ 0 & -1 & & \\ \vdots & & \ddots & \\ 0 & & & -1 \\ \rho_{1,t} & \rho_{2,t} & \cdots & \rho_{r,t} \end{pmatrix}. \tag{5.2}$$

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 $\beta_{11,t} = \cdots = \beta_{1r,t} = 1$.

[7] 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.

[8] Longer maturities were not selected because precise estimate of these are difficult to obtain by interpolation techniques. See Bliss (1997)

The individual cointegration relations in equation (5.2) can be written as

$$b_{n,t}y_{1,t} - y_{n,t} + \rho_{n,t}, \quad n = 3, 6, 9, 12, 60, 84, \tag{5.3}$$

where the maturities $n = 3$, 6, 9, 12, 60, 84 and $b_{n,t}$ correspond to $i = 1, \ldots, r$ and $\beta_{1i,t}$ in equation (5.2). 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 5.1 shows that the covariance matrix clearly differs between the three subsamples. The variance estimates from the three subsamples are given in Table 5.2.

$\Delta X_t - \alpha(t)\beta^*(t)'X^*_{t-1} - \Gamma_1\Delta X_{t-1} \sim N(0, \quad (t))$				
	Model	$\max \log L(\alpha(t), \beta^*(t), \Gamma_1, \quad (t))$	Degrees of freedom	$\mathcal{LR}(\mathrm{M}_i\|\mathrm{M}_0)$ (p-value)
M_0:	(t)	2009.25	295	–
M_1:	$_1 = \ _3$	1824.94	270	368.61 (0.0000)
M_2:	$_1 = \ _2 = \ _3$	1631.77	239	754.96 (0.0000)

Table 5.1: The maximum value of the likelihood function for the model with changing reduced rank parameters, and changing covariance $_t$.

It is not surprising that the variance of interest rates (see Table 5.2) were much higher in the 1979–1982 subsample when the Fed did not target interest rates directly. One conclusion from Table 5.1 is that the difference between the variance of interest rates in the first and third subsample is significant. From Table 5.2 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 phenomena 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[9]. The estimations results are given in Tables 5.3 and 5.4.

[9] The empirical analysis was performed in Gauss. Code and documentation can be obtained from http://weber.ucsd.edu/~phansen/.

The Estimated Covariance Matrices, (t)							
1970:3–1979:9	$_1 =$	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
1979:10–1982:10	$_2 =$	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
1982:11-1995:12	$_3 =$	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

Table 5.2: The estimated covariance matrices $_t$ from the most general break model.

1: Unrestricted Break Model $\alpha(t)$, $\beta(t)$, $\rho(t)$, (t)					$2\log L$ 4018.49	$\#f$ 295	$\mathcal{LR}$ –	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	
	ρ_n	0.3634	0.6356	1.1666	2.4113	3.4517	3.6640	
1979:10–1982:10	b_n	0.9234	0.8455	0.7716	0.7378	0.7179	0.6765	
	ρ_n	1.4726	2.5655	3.5156	3.8391	3.9931	4.4702	
1982:11-1995:12	b_n	1.0746	1.1391	1.2596	1.5328	1.7390	1.7989	
	ρ_n	-0.2384	-0.4607	-0.9011	-2.0401	-2.8354	-3.0585	
2: Expectations Hypothesis $\alpha(t)$, $\beta(t)=\beta, \rho(t)$, (t)					$2\log L$ 3989.58	$\#f$ 277	$\mathcal{LR}$ 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	
	ρ_n	0.2620	0.4935	0.6592	0.8935	1.1475	1.2357	
1979:10–1982:10	b_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	ρ_n	0.6309	0.8628	0.9917	0.9370	0.8637	0.8800	
1982:11-1995:12	b_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	ρ_n	0.2106	0.3694	0.6307	1.0520	1.3919	1.5010	
3: Constant $\alpha_\perp$ & Expectations Hypothesis $\alpha(t)=\alpha\phi(t)$, $\beta(t)=\beta$, $\rho(t)$, (t)					$2\log L$ 3978.44	$\#f$ 265	$\mathcal{LR}$ 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	
	ρ_n	0.2644	0.4999	0.6748	0.9221	1.1861	1.2753	
1979:10–1982:10	b_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	ρ_n	0.6529	0.9089	1.0495	0.9896	0.9065	0.9281	
1982:11-1995:12	b_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	ρ_n	0.2123	0.3753	0.6523	1.1248	1.5229	1.6487	
4: Constant α & β & EH. $\rho(t)$ may change. $\alpha(t)=\alpha$, $\beta(t)=\beta$, $\rho(t)$, (t)					$2\log L$ 3784.01	$\#f$ 199	$\mathcal{LR}$ 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	
	ρ_n	0.2701	0.5061	0.6798	0.9381	1.2343	1.3332	
1979:10–1982:10	b_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	ρ_n	0.5850	0.8015	0.9598	1.2261	1.4309	1.5107	
1982:11-1995:12	b_n	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
	ρ_n	0.2182	0.3826	0.6621	1.1599	1.5995	1.7405	

Table 5.3: 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 ρ_n from the cointegration relations $b_n y_{1,t} - y_{n,t} + \rho_n$ are reported for each model and subsample.

Model 1 in Table 5.3 is the most general model, where the parameters are left unrestricted. This model can be represented by the equation

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

where the parameters are constant within each subsample, i.e. $\alpha(t) = \alpha_1$ for $t \square$ 1979:09, $\alpha(t) = \alpha_2$ for 1979:10 $\square$ t $\square$ 1982:10 and $\alpha(t) = \alpha_3$ for $t \geq$ 1982:11, and similarly for $\beta(t)$, $\rho(t)$ and (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 subsamples, whereas term premia (ρ_n) adjustment coefficients (α_i, $i = 1, 2, 3$) as well as the covariance may differ across subsamples. This model can be written as

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

where β 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 against the long-run implication once structural breaks 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

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

The restriction implies that the orthogonal compliment to α 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 (4.1) and (4.2), and here extended with a

third subsample:

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

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

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

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 ε_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 ε_t.

The last model in Table 5.3, Model 4, can be expressed as

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

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 breaks in the covariance (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 breaks 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 our results, since we find $\rho(t)$ to vary as the volatility of interest rates changes. Finally, our 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).

5: No Breaks $\alpha(t)=\alpha,\ \beta(t)=\beta,\ \rho(t)=\rho,\quad (t)=$				$2\log L$ 2852	$\#f$ 131	$\mathcal{LR}$ –	p-value –
	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
	ρ_n	0.0011	0.1680	0.2951	0.6209	1.1478	1.2875
6: No Breaks & Expectations Hypothesis $\alpha(t)=\alpha,\ \beta(t)=H\phi,\ \rho(t)=\rho,\quad (t)=$				$2\log L$ 2825	$\#f$ 125	$\mathcal{LR}$ 26.84	p-value 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
	ρ_n	0.2719	0.4570	0.6561	0.9888	1.3148	1.4215

Table 5.4: Estimation results. Testing the expectations hypothesis in the cointegrated VAR without structural breaks. Note that the p-value is invalid because model 5 is strongly rejected against model 1.

The fifth and sixth models in Table 5.4 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 $\mathcal{LR}$ test statistic of Model 5 against Model 1 is 1166. Its distribution is asymptotically χ^2 with 164 degrees of freedom, and is therefore clearly rejected.

6. Conclusion

This paper shows how structural breaks 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 breaks as simple parameter restrictions in this framework. Moreover, the parameters can be estimated under these restrictions with the generalized reduced rank regression technique we developed. This technique is also applicable to estimation problems unrelated to structural breaks.

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

This combination of cointegration and structural breaks may provide a fruitful framework for many economic questions of interest. In this paper we analyzed the US term structure and found evidence of structural breaks that coincide with the Fed's policy changes in September 1979 and October 1982. Contrary to previous studies (see Hall, Anderson, and Granger (1992), Engsted and Tanggaard (1994) or Pagan, Hall, and Martin (1996)) we cannot reject the long-run implications of the expectations hypothesis, once these structural breaks are accounted for. In fact, we find a parsimonious model to be consistent with our 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.

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 currently being developed in Hansen and Johansen (1999).

A. Proofs

Before we give the proofs of Theorem 3.1 we derive some intermediate results. The following lemma is a consequence of Poincaré's theorem, however, a direct proof is presented here.

Lemma A.1. *The function* $g(y) = |y'\Lambda y| \,/\, |y'y|$ *where* y *is a* $p \times r$ *matrix,* $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ *and* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ *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 $J \subset \{1, \ldots, p\}$ of cardinality r, and define the $r \times r$ matrices y_J and Λ_J by $y_J = \{y_{ij}\}_{i \in J,\, j=1,\ldots,r}$ 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 $\{1, \ldots, p\}$ containing exactly r different elements (cardinality r). Below, we prove that

$$|y'\Lambda y| = \sum_{J \in \mathbb{D}_p^r} |y_J' \Lambda_J y_J| = \sum_{J \in \mathbb{D}_p^r} |y_J' y_J| \, \Pi_{i \in J} \lambda_i = \sum_{J \in \mathbb{D}_p^r} |y_J|^2 \, \Pi_{i \in J} \lambda_i. \tag{A.1}$$

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

The identity (A.1) 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

$p \backslash r$	1	2	3	4	$\cdots$
1	✓	–	–	–	
2	✓	✓	–	–	
3	✓	?	✓	–	
4	✓	?	?	✓	
$\vdots$	$\vdots$				$\ddots$

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

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

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

which proves the lemma in this case.

Next assume that $(y_{p1}, \ldots, y_{pr}) \neq 0$. We can then choose a full rank $r \times r$-matrix Q, so that $(y_{p1}, \ldots, y_{pr})Q = (0, \ldots, 0, 1)$, and we define the $p-1 \times r-1$ matrix $\tilde{z}$ as the first $r-1$ columns of $\tilde{y}Q$. We then have

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

Applying assumption $(A2)$ on the first term of (A.2) we find

$$\left| Q'\tilde{y}'\tilde{\Lambda}\tilde{y}Q \right| = |Q|^2 \sum_{J \in \mathbb{D}^r_{p-1}} |\tilde{y}'_J \tilde{\Lambda}_J \tilde{y}_J| = |Q|^2 \sum_{J \in \mathbb{D}^r_p, p \notin J} |y'_J \Lambda_J y_J|. \tag{A.3}$$

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

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

where $\tilde{J} = \{J \cup \{p\}\} \in \mathbb{D}^r_p$. So applying assumption $(A1)$ to the second term of (A.2) we have

$$\left| \tilde{z}'\tilde{\Lambda}\tilde{z} \right| \lambda_p = \lambda_p = |Q|^2 \sum_{J \in \mathbb{D}^{r-1}_p, p \in J} |y'_J \Lambda_J y_J| \,. \tag{A.4}$$

Combining the identities (A.2), (A.3) and (A.4) we have shown

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

which completes the proof. ■

In the proof for Lemma A.1 we obtained a representation for $|y'\Lambda y|$ which we formulate as a separate corollary.

Corollary A.2. *Let Λ be a real $p \times p$ diagonal matrix, and y a real $p \times r$ matrix, where $r \square p$. Then with the definitions above, we have that*

$$|y'\Lambda y| = \sum_{J \in \mathbb{D}^r_p} |y'_J \Lambda_J y_J| = \sum_{J \in \mathbb{D}^r_p} |y'_J y_J| \, \Pi_{i \in J} \lambda_i = \sum_{J \in \mathbb{D}^r_p} |y_J|^2 \, \Pi_{i \in J} \lambda_i.$$

Lemma A.3. *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 $f(x) = |x'Mx| / |x'Nx|$ has $\prod_{i=1}^{r} \lambda_i$ as its maximum with is obtained for $x = (v_1, \ldots, v_r)$ where $v_1, \ldots, v_r$ are eigenvectors corresponding to the r largest eigenvalues, $\lambda_1, \ldots, \lambda_r$ 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 $QQ' = I$, $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$ and $\lambda_1 > \lambda_2 > \cdots > \lambda_p \geq 0$. 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 A.1 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 3.1.

The likelihood function is given by

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

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

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

The maximization is done in three steps. First, we maximize with respect to Ψ taking α and β as given, then with respect to α and taking β as given, and finally with respect to β.

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

$$\hat{C}(\alpha, \beta) = M_{02} M_{22}^{-1} - \alpha\beta' M_{12} M_{22}^{-1}, \tag{A.5}$$

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

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

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 β as given, the estimates of α and are given by

$$\hat{\alpha}(\beta) = S_{01}\beta(\beta' S_{11}\beta)^{-1} \tag{A.6}$$

$$\hat{}(\beta) = S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}, \tag{A.7}$$

again using that the errors are Gaussian.

What remains is to maximize the concentrated likelihood function with respect to β. Since

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

the concentrated likelihood is given by

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

So maximizing the likelihood function is equivalent to minimizing

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

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 $\max \frac{|\beta'(S_{10}S_{00}^{-1}S_{01})\beta|}{|\beta' S_{11}\beta|}$. By Lemma A.3 the estimator is given by

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

where λ_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 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, and we find

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

Since the eigenvectors are normalized by $(\hat{v}_1, \ldots, \hat{v}_p)' S_{11} (\hat{v}_1, \ldots, \hat{v}_p) = I$, we have $\hat{\beta}' S_{11} \hat{\beta} = I$, such that (A.6) and (A.7) reduces to (3.3) and (3.4). By inserting these estimates into (A.5) we find (3.5). ■

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

Before we can formulate the general estimation result we need some additional notation. Define $Z_0 = (Z_{01}, \ldots, Z_{0T})$, $Z_1 = (Z_{11}, \ldots, Z_{1T})$, $Z_2 = (Z_{21}, \ldots, Z_{2T})$, and $E = (\varepsilon_1, \ldots, \varepsilon_T)$, so that Model 2.3 can be expressed as

$$Z_0 = AB'Z_1 + CZ_2 + E. \tag{A.8}$$

Next define

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

where $K_{p_1,r}$ is the commutation matrix, uniquely defined by $K_{p_1,r}\mathrm{vec}(B) \equiv \mathrm{vec}(B')$ 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

$$\Sigma = \mathrm{var}(\mathrm{vec}(\varepsilon_1, \ldots, \varepsilon_T)),$$

which is block diagonal under Assumption 2.1. The block diagonal matrices of Σ are given by (t), $t = 1, \ldots, T$, i.e. $\Sigma_{p(t-1)+i,p(t-1)+j} = \quad_{i,j}(t)$ for $i, j = 1, \ldots, p$ and $t = 1, \ldots, T$. Hence Σ^{-1} is a block diagonal matrix with $\quad(t)^{-1}$ as diagonal matrices, $t = 1, \ldots, T$.

Lemma A.4. *With the definitions above, we have the relations:*

$$\mathbf{Z}_{1A}'\Sigma^{-1}\mathbf{Z}_{1A} = \sum_{t=1}^{T} \left[A' \quad (t)^{-1}A \quad Z_{1t}Z_{1t}'\right], \tag{A.9}$$

$$\mathbf{Z}_{1A}'\Sigma^{-1}\mathrm{vec}(Z_0 - CZ_2) = \sum_{t=1}^{T} \mathrm{vec}\left(Z_{1t}(Z_{0t} - CZ_{2t})' \quad (t)^{-1\prime}A\right), \tag{A.10}$$

$$\mathbf{Z}_{1B2}'\Sigma^{-1}\mathbf{Z}_{1B2} = \sum_{t=1}^{T} \left[\begin{pmatrix} B'Z_{1t}Z_{1t}'B & B'Z_{1t}Z_{2t}' \\ Z_{2t}Z_{1t}'B & Z_{2t}Z_{2t}' \end{pmatrix} \quad (t)^{-1}\right], \tag{A.11}$$

$$\mathbf{Z}_{1B2}'\Sigma^{-1}\mathrm{vec}\,(Z_0) = \sum_{t=1}^{T} \mathrm{vec}\left(\quad (t)^{-1}Z_{0\tau}(Z_{1t}'B, Z_{2t}')\right), \tag{A.12}$$

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

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

Proof. The identity

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

which proves (A.9).

Next consider

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

which proves (A.10). Equations (A.11) and (A.12) are proven similarly.

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

Proof of Theorem 3.2. Applying the vec operation to equation (A.8) yields the equation

$$\begin{aligned}\mathrm{vec}(Z_0) &= (Z_1'B \quad I_p)\mathrm{vec}(A) + (Z_2' \quad I_p)\mathrm{vec}(C) + \varepsilon \\ &= [(Z_1'B, Z_2') \quad I_p)]\,\mathrm{vec}(A, C) + \varepsilon \\ &= \mathbf{Z}_{1B2}G\psi + \varepsilon.\end{aligned}$$

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

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

which by Lemma A.4 simplifies to (3.6).

Similarly for fixed A, C, and Σ, we have the equation

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

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

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

which by Lemma A.4 reduces to (3.7). ■

Proof of Corollary 3.3. Follows from Theorem 3.2 and Lemma A.4. ■

Proof of Corollary 3.4. From Theorem 3.1, we obtain the equations for $\hat{C}$ and $\hat{\ }$. Rather that 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

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

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

$$\begin{aligned}D_A \log L(A, B)(a) &= T\mathrm{tr}\left\{\ ^{-1}\left(S_{01} - AB'S_{11}\right)Ba'\right\} \\ &= T\left[\mathrm{tr}\left\{\ ^{-1}S_{01}Ba'\right\} - \mathrm{tr}\left\{I_pA(B'S_{11}B)a'\right\}\right]\end{aligned}$$

$$= T\text{vec}(a)' \left[\left(B' \quad {}^{-1}\right) \text{vec}\left(S_{01}\right) - \left(B'S_{11}B \quad I_p\right) \text{vec}(A) \right],$$

and

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

using Theorem 3 from Magnus and Neudecker (1988, Chapter 2). So equations (3.9) and (3.10) are the first order conditions. ■

Proof of Corollary 3.5. The result follows directly from Theorem 3.1 and Corollary 3.4. ■

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 Hypohesis 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*. Princeson 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.

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*.

——— (1998a): "Estimating Restricted Cointegrating Vectors," *UCSD manuscript*.

——— (1998b): "The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots," *Econometrica*, 66, 149–158.

Engsted, T., and C. Tanggaard (1994): "Cointegration and the US Term Structure," *Journal of Banking and Finance*, 18, 167–181.

——— (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 Treasure 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 (1993): "Recursive Analysis of Eigenvalues in Cointegrated VAR-Models," *Institute of Mathematical Statistics Working Paper 1, University of Copenhagen, Copenhagen, Denmark*.

Hansen, P. R. (1999): "Estimation and Inference in Cointegration Models with a General Covariance Structure," *Unpublished Research Paper*.

Hansen, P. R., and S. Johansen (1999): "Determining the Cointegration Rank in Processes with Structural Breaks," *Unpublished Research Paper*.

Hansen, P. R., and A. Rahbek (1999): "Granger's Representation Theorem: An Explicit Expression," *Unpublished note*.

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.

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

——— (1996): *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 2nd edn.

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 *Handbook of Monetary Economics*, ed. by B. Friedman, and F. Hahn. North-Holland, Amsterdam, vol. 1, 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.

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 *Handbook of Monetary Economics*, ed. by B. Friedman, and F. Hahn. North-Holland, Amsterdam, vol. 1, 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. (1999): *Asymptotic Theory for Econometricians*. Forthcoming at Academic Press, Orlando, 2nd edn.