

CREATES Research Paper 2009-4

First and second order non-linear cointegration models

Theis Lange



School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark



Aarhus School of Business
University of Aarhus

Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF
COPENHAGEN



FIRST AND SECOND ORDER NON-LINEAR COINTEGRATION MODELS

Theis Lange
Department of Economics
University of Copenhagen
&
CREATES

Abstract: This paper studies cointegration in non-linear error correction models characterized by discontinuous and regime-dependent error correction and variance specifications. In addition the models allow for autoregressive conditional heteroscedasticity (ARCH) type specifications of the variance. The regime process is assumed to depend on the lagged disequilibrium, as measured by the norm of linear stable or cointegrating relations. The main contributions of the paper are: *i*) conditions ensuring geometric ergodicity and finite second order moment of linear long run equilibrium relations and differenced observations, *ii*) a representation theorem similar to Granger's representations theorem and a functional central limit theorem for the common trends, *iii*) to establish that the usual reduced rank regression estimator of the cointegrating vector is consistent even in this highly extended model, and *iv*) asymptotic normality of the parameters for fixed cointegration vector and regime parameters. Finally, an application of the model to US term structure data illustrates the empirical relevance of the model.

Keywords: Cointegration, Non-linear adjustment, Regime switching, Multivariate ARCH.

JEL codes: C13; C32; C51.

Theis Lange gratefully acknowledges the research support of grant no. 2114-04-0001 from the "Danish Social Sciences Research Council" and of CREATES (funded by the Danish National Research Foundation).

1 Introduction

Since the 1980's the theory of cointegration has been hugely successful. Especially Granger's representation theorem, see Johansen (1995), which provides conditions under which non-stationary vector autoregressive (VAR) models can exhibit stationary, stable linear combinations. This very intuitive concept of stable relations is probably the main reason why cointegration models have been so widely applied (even outside the world of economics). For an up to date discussion see the survey Johansen (2007).

However, recent empirical studies suggest that the adjustments to the stable relations might not be adequately described by the linear specification employed in the traditional cointegration model. When modeling key macroeconomic variables such as GNP, unemployment, real exchange rates, or interest rate spreads, non-linearities can be attributed to transaction costs, which induces a band of no disequilibrium adjustment. For a more thorough discussion see e.g. Dumas (1992), Sercu, Uppal & Van Hulle (1995), Anderson (1997), Hendry & Ericsson (1991), and Escribano (2004). Furthermore, policy interventions on monetary or foreign exchange markets may also cause non-linear behavior, see Ait-Sahalia (1996) and Forbes & Kofman (2000) among others. Such non-linearities can also explain the problem of seemingly non-constant parameters encountered in many applications of the usual linear models. To address this issue Balke & Fomby (1997) suggested the threshold cointegration model, where the adjustment coefficients may switch between a specific set of values depending on the cointegrating relations. Generalizations of this model has lead to the smooth transition models, see Kapetanios, Shin & Snell (2006) and the references therein and the stochastically switching models, see e.g. Bec & Rahbek (2004), and Dufrenot & Mignon (2002) and the many references therein.

Parallel to this development the whole strain of literature devoted to volatility modeling has documented that non-linearities should also be included in the specification of the variance of the innovations. A large, and ever growing, number of autoregressive conditional heteroscedasticity (ARCH) type models, originally introduced by Engle (1982) and generalized by Bollerslev (1986), has been suggested, see e.g. Bauwens, Laurent & Rombouts (2006) for a recent discussion of

multivariate generalized ARCH models.

Motivated by these findings, this paper proposes a cointegration model, which allows for non-linearities in both the disequilibrium adjustment and the variance specifications. The model will be referred to as the first and second order non-linear cointegration vector autoregressive (FSNL-CVAR) model. The adjustments to the stable relations are assumed to be switching according to a threshold state process, which depends on past observations. Thus, the model extends the concept of threshold cointegration as suggested in Balke & Fomby (1997). The main novelty of the FSNL-CVAR model is to adopt a more general variance specification in which the conditional variance is allowed to depend on both the current regime as well as lagged values of the innovations, hereby including an important feature of financial time series.

Constructing a model which embeds many of the previously suggested models opens up the use of likelihood based tests to assess the relative importance of these models. For instance, does the inclusion of a regime dependent covariance matrix render the traditional ARCH specification obsolete or vice versa? Furthermore, since both the mean- and variance parameters depend on the current regime a test for no regime effect in for example the mean equation can be conducted as a simple χ^2 -test, since the issue of vanishing parameters under the null hypothesis, and resulting non-standard limiting distributions see e.g. Davies (1977), has been resolved by retaining the dependence on the regime process in the variance specification.

The present paper derives easily verifiable conditions ensuring geometric ergodicity, and hence the existence of a stationary initial distribution, of the first differences of the observations and of the linear cointegrating relationships. Stability and geometric ergodicity results form the basis for law of large numbers theorems and are therefore an important step not only towards an understanding of the dynamic properties of the model, but also towards the development of an asymptotic theory. The importance of geometric ergodicity has recently been emphasized by Jensen & Rahbek (2007), where a general law of large numbers is shown to be a direct consequence of geometric ergodicity. It should be noted that the conditions ensuring geometric ergodicity do not involve the parameters of adjustment in the inner regime, corresponding to the band of no action in the

example above. The paper also derives a representation theorem corresponding to the well known Granger representation theorem and establishes a functional central limit theorem (FCLT) for the common trends. Finally, asymptotic normality of the parameter estimates is shown to hold under the assumption of known cointegration vector and threshold parameters. The results are applied to US term structure data. The empirical analysis finds clear evidence indicating that the short-term and long-term rates only adjusts to one another when the spread is above a certain threshold. In order to achieve a satisfactory model fit the inclusion of ARCH effects is paramount. Hence the empirical analysis support the need for cointegration models, which are non-linear in both the mean and the variance. Finally, the empirical study shows that adjustments occurs through the short rate only, which is in accordance with the expectation hypothesis for the term structure.

The rest of the paper is structured as follows. Section 2 presents the model and the necessary regularity conditions. Next Section 3 contains the results regarding stability and order of integration. Estimation and asymptotic theory is discussed in Section 4 and the empirical study presented in Section 5. Conclusions are presented in Section 6 and all proofs can be found in Appendix.

The following notation will be used throughout the paper. For any vector $\|\cdot\|$ denotes the Euclidian vector norm and I_p a p -dimensional unit matrix. For some $p \times r$ matrix β of rank $r \leq p$, define the orthogonal complement β_\perp as the $p \times (p-r)$ -dimensional matrix with the property $\beta'\beta_\perp = 0$. The associated orthogonal projections are given by $I_p = \bar{\beta}\beta' + \bar{\beta}_\perp\beta_\perp'$ with $\bar{\beta} = \beta(\beta'\beta)^{-1}$. Finally $\varepsilon_{i,t}$ denotes the i 'th coordinate of the vector ε_t . In Section 2 and 3 and the associated proofs only the true parameters will be considered and the usual subscript 0 on the true parameters will be omitted to avoid an unnecessary cumbersome notation.

2 The first and second order non-linear cointegration model

In this section the model is defined and conditions for geometric ergodicity of process generated according to the model are stated. As discussed the model

is non-linear in both the mean- and variance specification, which justifies referring to the model as the first and second order non-linear cointegration vector autoregressive (FSNL-CVAR) model.

2.1 Non-linear adjustments

Let X_t be a p -dimensional observable stochastic process. The process is driven by both an unobservable i.i.d. sequence ν_t and a zero-one valued state process s_t . It is assumed that the distribution of the latter depends on lagged values of the observable process and that ν_t is independent of s_t . The evolution of the observable process is governed by the following generalization of the usual CVAR model, see e.g. Johansen (1995).

$$\begin{aligned}\Delta X_t &= s_t \left(a^{(1)} \beta' X_{t-1} + \sum_{j=1}^{q-1} \Gamma_j \Delta X_{t-j} \right) \\ &\quad + (1 - s_t) \left(a^{(0)} \beta' X_{t-1} + \sum_{j=1}^{q-1} G_j \Delta X_{t-j} \right) + \varepsilon_t \\ \varepsilon_t &= H^{1/2}(s_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}) \nu_t = H_t^{1/2} \nu_t,\end{aligned}\tag{1}$$

where $a^{(0)}$, $a^{(1)}$, and β are $p \times r$ matrices, $(\Gamma_j, G_j)_{j=1, \dots, q-1}$ are $p \times p$ matrices, and ν_t an i.i.d. $(0, I_p)$ sequence. By letting the covariance matrix H_t depend on lagged innovations $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ the model allows for a very broad class of ARCH type specifications. The exact specification of the covariance matrix will be addressed in the next section, but by allowing for dependence of the lagged innovations the suggested model permits traditional ARCH type dynamics of the innovations. Saikkonen (2008) has suggested to use lagged values of the observed process X_t in the conditional variance specification, however, this leads to conditions for geometric ergodicity, which cannot be stated independently for the mean- and variance parameters and a less clear cut definition of a unit root.

As indicated in the introduction the proposed model allows for non-linear and discontinuous equilibrium correction. The state process could for instance be specified such that if the deviation from the stable relations, measured by $\|\beta' X_{t-1}\|$,

is below some predefined threshold adjustment to the stable relations occurs through $a^{(0)}$ and as a limiting case no adjustment occurs, which could reflect transaction costs. However, if $\|\beta'X_{t-1}\|$ is large adjustment will take place through $a^{(1)}$. For applications along these lines, see Akram & Nymoen (2006), Chow (1998), and Krolzig, Marcellino & Mizon (2002).

2.2 Switching autoregressive heteroscedasticity

Depending on the value of the state process at time t the covariance matrix is given by

$$H_t = D_t^{1/2} \Lambda^{(l)} D_t^{1/2} \quad (2)$$

$$D_t = \text{diag}(\Pi_t)$$

$$\Pi_t = (\pi_{1,t}, \dots, \pi_{p,t})'$$

$$\pi_{i,t} = 1 + g_i(\varepsilon_{i,t-1}, \dots, \varepsilon_{i,t-q}), \quad i = 1, \dots, p \quad (3)$$

with Λ_l a positive definite covariance matrix, $g_i(\cdot)$ a function onto the non-negative real numbers for all $i = 1, \dots, p$, and $l = 0, 1$ corresponds to the possible values of the state process.

The factorization in (2) isolates the effect of the state process into the matrix $\Lambda^{(l)}$ and the ARCH effect into the diagonal matrix D_t . This factorization implies that all information about correlation is contained in the matrix $\Lambda^{(l)}$, which switches with the regime process. In this respect the variance specification is related to the constant conditional correlation (CCC) model of Bollerslev (1990) and can be viewed as a mixture generalization of this model.

For example, suppose that $p = 2$, $q = 1$, $g_i(\varepsilon_{i,t-1}) = \alpha_i \varepsilon_{i,t-1}^2$, and $s_t = 1$ almost surely for all t . Then the conditional correlation between $X_{1,t}$ and $X_{2,t}$ is given by the off-diagonal element of Λ_1 , which illustrates that the model in this case is reduced to the traditional cointegration model with the conditional variance specified according to the CCC model.

Since the functions g_1, \dots, g_p allow for a feedback from past realizations of the innovations to the present covariance matrix it is necessary to impose some regularity conditions on these functions in order to discuss stability of the cointegrating

relations $\beta' X_t$ and ΔX_t .

Assumption 1. (i) For all $i = 1, \dots, p$ there exists constants, denoted $\alpha_{i,1}, \dots, \alpha_{i,q}$, such that for $\|(\varepsilon'_{t-1}, \dots, \varepsilon'_{t-q})'\|$ sufficiently large it holds that $g_i(\varepsilon_{i,t-1}, \dots, \varepsilon_{i,t-q}) \leq \sum_{j=1}^q \alpha_{i,j} \varepsilon_{i,t-j}^2$.

(ii) For all $i = 1, \dots, p$ the sequence of constants satisfies $\max_{l=0,1} \Lambda_{i,i}^{(l)} \sum_{j=1}^q \alpha_{i,j} < 1$.

The assumption essentially ensures that as the lagged innovations became large the covariance matrix responds no more vigorously than an ARCH(q) process with finite second order moment. However, for smaller shocks the assumption allows for a broad range of non-linear responses.

2.3 The State Process

Initially recall that the state or switching variable s_t is zero-one valued. Next define the $r + p(q - 1)$ -dimensional variable z_t as

$$z_t = (X'_{t-1}\beta, \Delta X'_{t-1}, \dots, \Delta X'_{t-q+1})'. \quad (4)$$

By assumption the dynamics of the state process are given by the conditional probability

$$P(s_t = 1 \mid X_{t-1}, \dots, X_0, s_{t-1}, \dots, s_0) = P(s_t = 1 \mid z_t) \equiv p(z_t). \quad (5)$$

Some theoretical results regarding univariate switching autoregressive models where the regime process is similar to (5) can be found in Gouriéroux & Robert (2006).

The transition function $p(\cdot)$ will be assumed to be an indicator function taking the value one outside a bounded set as suggested by Balke & Fomby (1997). In the transaction cost example it is intuitively clear that as the distance from the stable relation increases so does the probability of adjustment to the stable relation. This leads to:

Assumption 2. The transition probability $p(z)$ defined in (5) is zero-one valued and tends to one as $\|z\| \rightarrow \infty$.

On the basis of Assumption 2 it is natural to refer to the regime where $s_t = 0$ as the inner regime and the other as the outer regime. As in Bec & Rahbek (2004) additional inner regimes can be added without affecting the validity of the results, the only difference being a more cumbersome notation. Extending the model to include additional outer regimes, as in Saikkonen (2008), leads inevitably to regularity conditions expressed in terms of the joint spectral radius of a class of matrices, which are in most cases impossible to verify.

It follows that the FSNL-CVAR model allows for epochs of equilibrium adjustments and epochs without. Furthermore the model allows these epochs to be characterized by different correlation structures and have general ARCH type variance structure. In the next section stationarity and geometric ergodicity of $\beta'X_t$ and $\beta'_\perp\Delta X_t$ are discussed.

3 Stability and order of integration

In the first part of this section conditions which ensure geometric ergodicity of $\beta'X_t$ and ΔX_t will be derived. In the second part of the section we derive a representation theorem corresponding to the well known Granger representation theorem and establish a FCLT for the common trends. The results presented in this section rely on by now classical Markov chain techniques, see Meyn & Tweedie (1993) for an introduction and definitions.

3.1 Geometric Ergodicity

Note initially that the process X_t generated by (1), (2), and (5) is not in itself a Markov chain due to the time varying components of the covariance matrix. Define therefore the process $V_t = (X'_t\beta, \Delta X'_t\beta'_\perp)'$ where the orthogonal projection $I_p = \bar{\beta}\bar{\beta}' + \bar{\beta}_\perp\bar{\beta}'_\perp$, has been used. Furthermore define the stacked processes

$$\bar{V}_t = (V'_t, \dots, V'_{t-q+1})', \quad \bar{\varepsilon}_t = (\varepsilon'_t, \dots, \varepsilon'_{t-q+1})',$$

and the selection matrix

$$\varphi = (I_p, 0_{p \times (q-1)p})'.$$

By construction the process \bar{V}_t is generated according to the VAR(1) model given by

$$\bar{V}_t = s_t \mathbf{A} \bar{V}_{t-1} + (1 - s_t) \mathbf{B} \bar{V}_{t-1} + \eta_t, \quad (6)$$

where $\eta_t = \varphi(\beta', \beta'_\perp)' \varepsilon_t$ and hence is a mean zero random variable with variance $(\beta', \beta'_\perp)' H_{t,s_t} (\beta', \beta'_\perp)$. The matrices \mathbf{A} and \mathbf{B} are implicitly given by (1), see (19) in the appendix for details. Finally the Markov chain to be considered can be defined as $Y_t = (\bar{\varepsilon}'_t, \bar{V}'_{t-q})'$. Note that V_t, \dots, V_{t-q+1} and s_t, \dots, s_{t-q+1} are computable from Y_t , which can be seen by first computing s_{t-q+1} then V_{t-q+1} and repeating.

As hinted earlier it suffices to assume that the usual cointegration assumption is satisfied for the parameters of the outer regime;

Assumption 3. *Assume that the rank of $a^{(1)}$ and β equals r and furthermore that there are exactly $p - r$ roots at $z = 1$ for the characteristic polynomial*

$$A(z) = I_p(1 - z) - a^{(1)}\beta'z - \sum_{i=1}^{q-1} \Gamma_i(1 - z)z^i, \quad z \in \mathbb{C} \quad (7)$$

while the remaining roots are larger than one in absolute value.

Next the main result regarding geometric ergodicity of the Markov chain Y_t will be stated, with the central conditions expressed in terms of the matrices \mathbf{A} and \mathbf{B} . In the subsequent corollaries some important special cases are considered.

Theorem 1. *Assume that:*

- (i) For some $qp \times n$, $qp \geq n \geq 0$ matrix μ of rank n , it holds that $(\mathbf{A} - \mathbf{B})\mu = 0$.
- (ii) The largest in absolute value of the eigenvalues of \mathbf{A} is smaller than one.
- (iii) The stochastic state variable s_t is zero-one valued and the state probability is zero-one valued and satisfies $p(\gamma'\bar{v}) \rightarrow 0$ as $\|\gamma'\bar{v}\| \rightarrow \infty$ with $\gamma = \mu_\perp$ and $\bar{v} \in \mathbb{R}^{pq}$
- (iv) ν_t is i.i.d. $(0, I_p)$ and has a continuous and strictly positive density with respect to the Lebesgue measure on \mathbb{R}^p .
- (v) The generalized ARCH functions g_1, \dots, g_p satisfy the conditions listed in Assumption 1.

Then Y_t is a geometrically ergodic process, which has finite second order moment. Furthermore there exists a distribution for Y_0 such that the sequence $(Y_t)_{t=1}^T$ is stationary.

The formulation of the FSNL-CVAR model is in essence a combination of the non-linear cointegration model of Bec & Rahbek (2004) and the constant conditional correlation model of Bollerslev (1990). It is therefore not surprising that the following corollaries show that the stability conditions for the FSNL-CVAR model are simply a combination of the stability conditions associated with the these two "parent" models.

Corollary 1. *Under Assumption 1, 2, and 3, and (iv) of Theorem 1 it holds that Y_t is a geometrically ergodic process with finite second order moment and that there exists a distribution for Y_0 such that the sequence $(Y_t)_{t=1}^T$ is stationary.*

In the special case where the transition probability in (5) only depends on $\beta'X_t$ Assumption 2 does not hold. However, using Theorem 1 the following result can be established.

Corollary 2. *Consider the case where the transition probability in (5) only depends on $z_t = \beta'X_t$ and that $\Gamma_j = G_j$ for $j = 1, \dots, q - 1$. Under Assumption 1, 2, and 3 and (iv) of Theorem 1 it holds that Y_t is a geometrically ergodic process with finite second order moment and that there exists a distribution for Y_0 such that the sequence $(Y_t)_{t=1}^T$ is stationary.*

3.2 Non-stationarity

When considering linear VAR models the concept of $I(1)$ processes is well defined, see Johansen (1995). This is in contrast to non-linear models, such as the FSNL-CVAR model, where there still exists considerable ambiguity as to how to define $I(1)$ processes. In this paper we follow Corradia, Swanson & White (2000) and Saikkonen (2005) and simply define an $I(1)$ process as a process for which a functional central limit theorem (FCLT) applies. In Theorem 2 below we establish conditions for which the $(p-r)$ common trends of X_t have a non-degenerate long-run variance and a FCLT applies.

Theorem 2. *Under the assumptions of Theorem 1 the process X_t given by (1), (2), and (5) has the representation*

$$X_t = C \sum_{i=1}^t (\varepsilon_i + (\Phi^{(0)} - \Phi^{(1)})u_i) + \tau_t, \quad (8)$$

where the processes τ_t , see 20, and $u_t = (1 - s_t)z_t$ are stationary and z_t is defined in (4). Furthermore the parameters C , $\Phi^{(0)}$, and $\Phi^{(1)}$ are defined by

$$\begin{aligned} C &= \beta_{\perp} (a^{(1)'}_{\perp} (I_p - \sum_{i=1}^{k-1} \Gamma_i) \beta_{\perp})^{-1} a^{(1)'}_{\perp}, \quad \Phi^{(0)} = (a^{(0)}, G_1, \dots, G_{k-1}), \\ \Phi^{(1)} &= (a^{(1)}, \Gamma_1, \dots, \Gamma_{k-1}). \end{aligned}$$

The $(p-r)$ common trends of X_t are given by $\sum_{i=1}^t c_i$, where $c_t = a^{(1)'}_{\perp} (\varepsilon_t + (\Phi^{(0)} - \Phi^{(1)})u_t)$. A FCLT applies to $c_t - Ec_T$, if

$$\Upsilon = \psi' \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{pmatrix} \psi > 0, \quad \text{where } \psi' = a^{(1)'}_{\perp} (I_p, \Phi^{(0)} - \Phi^{(1)}). \quad (9)$$

The Σ matrices are the long run variances and the exact expression can be found in (21).

Note that sufficient conditions for Υ being positive definite are $\text{sp}(\Phi^{(0)}) = \text{sp}(\Phi^{(1)})$ or $a^{(1)'}_{\perp} \Sigma_{\varepsilon\beta} = 0$.

4 Estimation and asymptotic normality

In this section it is initially established that the usual estimator in the linear cointegrated VAR model of the cointegration vector β , which is based on reduced rank regression (RRR), see Johansen (1995), is consistent even when data is generated by the much more general FSNL-CVAR model. The second part of this section considers estimation and asymptotic theory of the remaining parameters.

Define $\hat{\beta}$ as the usual RRR estimator of the cointegration vector defined in Johansen (1995) Theorem 6.1. Introduce the normalized estimator $\tilde{\beta}$ given by $\tilde{\beta} = \hat{\beta}(\bar{\beta}'\hat{\beta})^{-1}$, where $\bar{\beta} = \beta(\beta'\beta)^{-1}$. Note that this normalization is clearly not feasible in practice as the matrix β is not known, however, the purpose of the normalized version is only to facilitate the formulation of the following consistency result.

Theorem 3. *Under the assumptions of Theorem 1 and the additional assumption that Υ defined in (9) is positive definite, $\tilde{\beta}$ is consistent and $\tilde{\beta} - \beta = o_p(T^{1/2})$.*

Theorem 3 suggests that once the cointegrating vector β has been estimated using RRR the remaining parameters can be estimated by quasi-maximum likelihood using numerical optimization. To further reduce the curse of dimensionality the parameters $\Lambda^{(1)}$ and $\Lambda^{(0)}$ can be concentrated out of the log-likelihood function as discussed in Bollerslev (1990).

In order to discuss asymptotic theory restrict the variance specification in (2) to linear ARCH(q), that is replace (3) by

$$\pi_{i,t} = 1 + \sum_{j=1}^q \alpha_{i,j} \varepsilon_{i,t-j}^2 \quad (10)$$

and define the parameter vectors

$$\begin{aligned} \theta^{(1)} &= \text{vec}(\Phi^{(1)}, \Phi^{(0)}), & \theta^{(2)} &= (\alpha_{1,1}, \dots, \alpha_{p,1}, \alpha_{2,1}, \dots, \alpha_{p,q}), \\ \theta^{(3)} &= (\text{vech}(\Lambda^{(1)})', \text{vech}(\Lambda^{(0)})')', \end{aligned}$$

and $\theta = (\theta^{(1)'}, \theta^{(2)'}, \theta^{(3)'})'$. As is common let θ_0 denote the true parameter value. If the cointegration vector β and the threshold parameters are assumed known the

realization of state process is computable and the quasi log-likelihood function to be optimized is, apart from a constant, given by

$$L_T(\theta) = \frac{1}{T} \sum_{t=1}^T l_t(\theta), \quad l_t(\theta) = -\log(|H_t(\theta)|)/2 - \varepsilon_t(\theta)' H_t(\theta)^{-1} \varepsilon_t(\theta)/2, \quad (11)$$

where $\varepsilon_t(\theta)$ and $H_t(\theta)$ are given by (1) and (2), respectively. The assumption of known β and λ is somewhat unsatisfactory, but at present necessary to establish the result. Furthermore, the assumption can be partly justified by recalling that estimators of both the cointegration vector and the threshold parameter are usually super consistent. The proof of the following asymptotic normality result can be found in the appendix.

Theorem 4. *Under the assumptions of Corollary 1 and the additional assumption that there exists a constant $\delta > 0$ such that $E[\|\varepsilon_t\|^{4+\delta}]$ and $E[\|\nu_t\|^{4+\delta}]$ are both finite and $\theta^{(2)} > 0$ it holds that when β and the parameters of the regime process are kept fixed at true values there exists a fixed open neighborhood around the true parameter $N(\theta_0)$ such that with probability tending to one as T tends to infinity, $L_T(\theta)$ has a unique minimum point $\hat{\theta}_T$ in $N(\theta_0)$. Furthermore, $\hat{\theta}_T$ is consistent and satisfies*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0, \Omega_I^{-1} \Omega_S \Omega_I^{-1}),$$

where $\Omega_S = E[(\partial l_t(\theta_0)/\partial \theta)(\partial l_t(\theta_0)/\partial \theta)']$ and $\Omega_I = E[\partial^2 l_t(\theta_0)/\partial \theta \partial \theta']$.

The proof is given in the appendix, where precise expressions for the asymptotic variance are also stated.

5 An application to the interest rate spread

In this section an analysis of the spread between the long and the short U.S. interest rates using the FSNL-CVAR model is presented. The analysis is similar to the analysis of German interest rate spreads presented in Bec & Rahbek (2004). However, since the FSNL-CVAR model allows for heteroscedasticity the present analysis will employ daily data unlike the analysis in Bec & Rahbek (2004),

which is based on monthly averages. The well-known expectations hypothesis of the term structure implies that, under costless and instantaneous portfolio adjustments and no arbitrage the spread between the long and the short rate can be represented as

$$R(k, t) - R(1, t) = \frac{1}{k} \sum_{i=1}^{k-1} \sum_{j=1}^i E_t[\Delta R(1, t + j)] + L(k, t), \quad (12)$$

where $R(k, t)$ denotes the k -period interest rate at time t , $L(k, t)$ represents the term premium, accounting for risk and liquidity premia, and $E_t[\cdot]$ the expectation conditional on the information at time t , see e.g. Bec & Rahbek (2004) for details. Clearly, the right hand side is stable or stationary provided interest rate changes and the term premium are stationary (see Hall, Anderson & Granger (1992)). In fact portfolio adjustments are neither costless nor instantaneous. It is therefore reasonable to assume that the spread $S(k, 1, t) = R(k, t) - R(1, t)$, will temporarily depart from its equilibrium value given by (12). However, once portfolio adjustments have taken place (12) will again hold. Hence, the long and the short interest rate should be cointegrated with a cointegration vector of $\beta = (1, -1)'$. Testing this implication of the expectations hypothesis of the term structure has been the focus for many empirical papers, however, the results are not clear cut. Indeed the U.S. spread is found to be stationary in e.g. Campbell & Shiller (1987), Stock & Watson (1988), Anderson (1997), and Tzavalis & Wickens (1998), but integrated of order 1 in e.g. Evans & Lewis (1994), Enders & Siklos (2001), and Bec, Guayb & Guerre (2008). Note however, that when allowing for a stationary non-linear alternative the last two papers reject the hypothesis of non-stationarity of the U.S. spread. Indeed Anderson (1997) establishes that if one considers homogeneous transaction costs which reduces the investors yield on a bond by a constant amount, say λ , then one expects that the yield spread is stationary, but non-linear, since portfolio adjustments will only occur when the difference between the actual spread $S(k, 1, t)$ and the value predicted by (12) is larger in absolute value than λ .

According to Anderson's argument the joint dynamics of short-term and long-term interest rates could be described by the non-linear error correction model

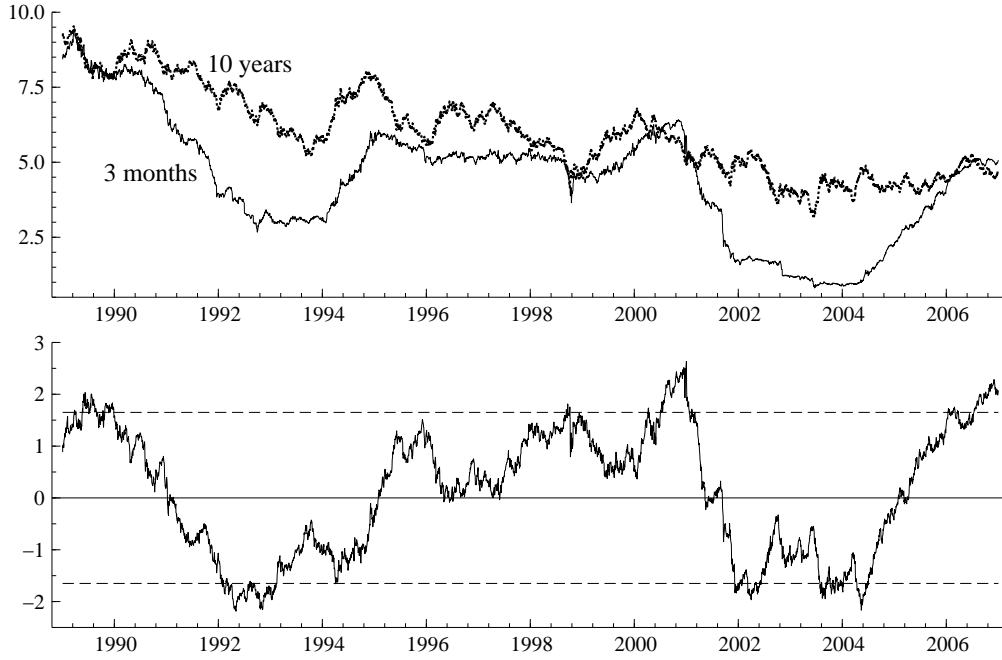


Figure 1: The 3-month and 10-year interest rates (top panel) and the spread between the two series adjusted for their mean (bottom panel). The dashed lines indicate the threshold $\lambda = 1.65$.

given by (1):

$$\Delta X_t = (s_t a^{(1)} + (1 - s_t) a^{(0)}) \beta' X_t + \sum_{j=1}^{k-1} \Delta X_{t-j} + \varepsilon_t, \quad (13)$$

where $X_t = (R_t^S, R_t^L)$, denotes the short and the long rates and the transition function is defined in accordance with Anderson's argument. However, as it is a well established fact that daily interest rates exhibit considerable heteroscedasticity the model must include time dependent variance as in (10).

In the following the proposed FSNL-CVAR model will be applied to daily recordings of the U.S. 3-Month Treasury Constant Maturity Rate and the U.S. 10-Year Treasury Constant Maturity Rate spanning the period from 1/1-1988 to 1/1-2007 yielding a total of 4,500 observations. Data have been downloaded from the web-page of the Federal Reserve Bank of St. Louis. Following Bec & Rahbek (2004) both series are corrected for their average and the state process is therefore given

by $s_t = 1_{\{|S_{t-1}^G| \geq \lambda\}}$, with $S_{t-1}^G = \beta' X_{t-1}$. This amounts to approximate the long-run equilibrium given by (12) by the average of the actual spread, as is common in the literature. Figure 1 depicts the data.

Initially a self-exciting threshold autoregressive (SETAR) model was fitted to the series S_t^G , which indicated a threshold parameter of $\lambda = 1.65$. This value is very close to the threshold parameter value of 1.7 reported in Bec & Rahbek (2004) for a similar study based on monthly German interest rate data. For the remaining part of the analysis the threshold parameter will be kept fixed at 1.65. However, it should be noted that by determining the threshold parameter in such a data dependent way the conditions for the asymptotic results given in Theorem 4 are formally not met. In this respect, recall from the vast literature on univariate threshold models that the threshold parameter is super-consistent and hence can be treated as fixed when making inference on the remaining parameters, we would expect this to hold in this case as well. Furthermore, as can be seen from (13) the short-term parameters Γ_i are assumed to be identical over the two regimes, the estimators and covariances in Theorem 4 should be adjusted accordingly.

Concerning the specification of lag lengths in (13), additional lags were included until there were no evidence of neither autocorrelation nor additional heteroscedasticity in the residuals. This lead us to retain seven lags in the mean equation and six lags in the variance equation. The choice of lag specification was confirmed by both the AIC as well as statistical test indicating that additional lags were not statistically significant at the 5% level.

The parameter estimates of the mean equation are reported in the first two columns of Table 1. Initially it is noted that the estimated parameters seem to confirm our conjecture that when the spread is below the threshold value no adjustment towards the equilibrium occurs. This is confirmed by testing the hypothesis that $a^{(0)} = (0, 0)'$, which is accepted with a p-value of 0.60 using the LR test. In addition the estimates of $a^{(1)}$ indicate that long-term rates do not seem to adjust to disequilibrium. This is confirmed by the LR test of the hypothesis $a_1^{(0)} = a_2^{(0)} = a_2^{(1)} = 0$ which cannot be rejected. The test statistic equals 2.2 corresponding to a p-value of 0.53. The result implies that big spreads significantly affects the short-term rate only, which is in accordance with the expectation hypothesis for the term structure. This conclusion as well as the sign of the estimate

	Unrestricted		Restricted	
	ΔR_t^S	ΔR_t^L	ΔR_t^S	ΔR_t^L
$1_{\{ S_{t-1}^G \geq 1.65\}} S_{t-1}^G$	-0.00116 (-2.26)	0.000996 (1.09)	-0.00136 (-2.86)	-
$1_{\{ S_{t-1}^G < 1.65\}} S_{t-1}^G$	0.000478 (0.92)	-9.81e-005 (-0.11)	-	-
ΔR_{t-1}^S	0.0687 (3.77)	0.00626 (0.30)	0.0687 (3.79)	0.00693 (0.34)
ΔR_{t-1}^L	-0.0163 (-1.82)	0.0636 (3.80)	-0.0163 (-1.83)	0.0629 (3.84)
ΔR_{t-2}^S	-0.0717 (-3.91)	-0.00262 (-0.13)	-0.0713 (-3.88)	-0.00175 (-0.09)
ΔR_{t-2}^L	8.56e-005 (0.01)	-0.0295 (-1.80)	-0.000117 (-0.01)	-0.0301 (-1.86)
ΔR_{t-3}^S	-0.0587 (-2.71)	0.0107 (0.53)	-0.0582 (-2.85)	0.0115 (0.58)
ΔR_{t-3}^L	0.00217 (0.24)	-0.0531 (-3.28)	0.00212 (0.24)	-0.0537 (-3.35)
ΔR_{t-4}^S	0.0583 (3.14)	-0.0177 (-0.89)	0.0589 (3.17)	-0.0165 (-0.82)
ΔR_{t-4}^L	-0.0111 (-1.18)	-0.0224 (-1.37)	-0.0111 (-1.19)	-0.0231 (-1.41)
ΔR_{t-5}^S	0.158 (8.47)	0.0476 (2.46)	0.158 (8.54)	0.0483 (2.50)
ΔR_{t-5}^L	-0.0162 (-1.71)	-0.0116 (-0.66)	-0.0159 (-1.69)	-0.0121 (-0.70)
ΔR_{t-6}^S	-0.0747 (-4.29)	-0.0282 (-1.43)	-0.0749 (-4.30)	-0.0278 (-1.45)
ΔR_{t-6}^L	0.0270 (3.07)	-0.00558 (-0.34)	0.0271 (3.10)	-0.00614 (-0.37)
LM(ARCH 1-8) [p-value]	0.6966 [0.69]	0.6596 [0.73]	0.6963 [0.70]	0.6598 [0.73]
LM(AR 1-8) [p-value]	34.99 [0.32]		34.79 [0.34]	
log-L	14,909.10		14,908.00	

Table 1: Model (13) estimates. t-statistics are reported in parentheses. LM tests of no remaining ARCH and no vector autocorrelation, respectively. Statistically significant parameters are indicated in bold.

of $a_1^{(1)}$ coincides with the findings of Bec & Rahbek (2004). Estimates of this restricted model are reported in the last two columns of Table 1 for the parameters of the mean equation and Table 2 for the parameters of the variance equation.

Table 2 reports the estimates of the variance equations. In order to ease comparison with traditional ARCH models the parametrization has been changed slightly from the one presented in (2) to directly reporting the coefficients of the equation $\Lambda_{1,1}^{(1)}\pi_{1,t} = \Lambda_{1,1}^{(1)} + \sum_{j=1}^6 \Lambda_{1,1}^{(1)}\alpha_{i,j}\varepsilon_{1,t-j}^2$, which gives the conditional variance for the first element of ε_t when $s_t = 1$ and likewise for the other cases. It should be noted that the change of parametrization has been performed after estimating the parameters using the parametrization of (2), which is preferable when performing the numerical optimization of the log-likelihood function as the parameters in $\Lambda^{(1)}$ and $\Lambda^{(1)}$ can be concentrated out. The reported t-statistics have therefore been computed using the delta-method.

The reported parameter estimates clearly demonstrates the presence of heteroscedasticity in the residuals. Examining the covariance matrix of the parameters collected in $\theta^{(3)}$ (covariance matrix not reported) indicates that $\Lambda_{i,i}^{(1)}$ and $\Lambda_{i,i}^{(1)}$ are statistically different for both the short- and the long rate. As expected the correlation is highest when adjustment to disequilibrium occurs ($s_t = 1$), but the hypothesis that the correlations are identical cannot be rejected at the 5% level. Hence the parameter estimates indicates the overall level of variance is highest in the regime where no adjustment occurs, but the correlation might be the same.

Since the parameters of the switching mechanism can be identified based solely on the variance specification the central hypothesis $a^{(1)} = a^{(0)}$, corresponding to no switching in the mean equation, can be tested by the standard LR test statistic, which will be asymptotically χ^2 distributed with two degrees of freedom. Thus avoiding the usual problems of unidentified parameters under the null, see e.g. Davies (1977), Davies (1987), and the survey Lange & Rahbek (2008), often encountered when testing no-switching hypothesis. The test statistic is 18.78 and the hypothesis of no switching in the mean equation is therefore clearly rejected.

It should be noted that the sum of the ARCH coefficients in each column of Table 2 are very close to one, which violates the fourth order moment condition of Theorem 4. However, as argued in Lange, Rahbek & Jensen (2007) based on a univariate model, we expect the asymptotic normality to hold even under a

	$s_t = 1$		$s_t = 0$	
	$\Lambda_{1,1}^{(1)}\pi_{1,t}$	$\Lambda_{2,2}^{(1)}\pi_{2,t}$	$\Lambda_{1,1}^{(0)}\pi_{1,t}$	$\lambda_{2,2}^{(0)}\pi_{2,t}$
Intercept	0.000412 (0.02)	0.00195 (0.12)	0.000475 (0.02)	0.00226 (0.25)
ε_{t-1}^2	0.148 (18.30)	0.0270 (2.80)	0.171 (21.87)	0.0313 (2.79)
ε_{t-2}^2	0.159 (18.14)	0.0430 (3.54)	0.183 (21.51)	0.0499 (3.60)
ε_{t-3}^2	0.235 (24.48)	0.0241 (1.72)	0.271 (35.49)	0.0279 (1.72)
ε_{t-4}^2	0.112 (18.44)	0.0540 (4.24)	0.129 (22.30)	0.0626 (4.34)
ε_{t-5}^2	0.170 (18.63)	0.115 (7.25)	0.196 (22.70)	0.134 (7.65)
ε_{t-6}^2	0.0603 (8.84)	0.0558 (4.70)	0.0695 (9.22)	0.0647 (4.74)
Correlation	0.425 (16.56)		0.396 (29.75)	

Table 2: Estimates of the variance parameters. The parameters in e.g. the first column correspond to the coefficients in the variance equation $\Lambda_{1,1}^{(1)}\pi_{1,t} = \Lambda_{1,1}^{(1)} + \sum_{j=1}^6 \Lambda_{1,1}^{(1)}\alpha_{i,j}\varepsilon_{1,t-j}^2$. t-statistics are reported in parentheses and were computed by the delta-method based on the expressions for the asymptotic variance in Theorem 4. Statistically significant parameters are indicated in bold.

weaker second order moment condition.

6 Conclusion

In this paper we have suggested a cointegrated vector error correction model with a non-linear specification of both adjustments to disequilibrium and variance characterized by regime switches. Since the FSNL-CVAR model embeds many previously suggested models, see the discussion in the introduction, it provides a framework for assessing the relative importance of these models in a likelihood based setup. Furthermore tests of hypothesis such as linearity of the mean equation, which previously led to non-standard limiting distributions, can be conducted as standard χ^2 -tests in the FSNL-CVAR model since the state

process can be identified through the variance specification.

Using Markov chain results we derive easily verifiable conditions under which $\beta'X_t$ and ΔX_t are stable with finite second order moment and can be embedded in a Markov chain, which is geometrically ergodic. The usefulness of this result is enhanced by the recent work of Jensen & Rahbek (2007), which provides a general law of large numbers assuming only geometric ergodicity. Furthermore, a representation theorem corresponding to Granger's representation theorem has been derived and a functional central limit theorem for the common trends established. This is utilized to show that the usual RRR estimator of Johansen (1995) for the cointegrating vector, β , is robust to the model extensions suggested by the FSNL-CVAR model.

Finally, we establish asymptotic normality of the estimated parameters for fixed and known cointegration vector and threshold parameters. Applying the model to daily recordings of the US term structure documents the empirical relevance of the FSNL-CVAR model and the empirical results are in accordance with the expectation hypothesis of the term structure. Specifically it is found that small interest rate spreads are not corrected, while big ones have a significant influence on the short rate only.

Appendix

Proof of Theorem 1

Consider initially the homogenous Markov chain $Y_t = (\bar{\varepsilon}_t', \bar{V}_{t-q}')'$. Before applying the drift criterion, see e.g. Tjøstheim (1990) or Meyn & Tweedie (1993) it must be verified that the Markov chain Y_t is irreducible, aperiodic, and that compact sets are small. In order to do so it will be verified that the $2q$ -step transition kernel has a density with respect to the Lebesgue measure, which is positive and bounded away from zero on compact sets.

Note that V_t, \dots, V_{t-q+1} and s_t, \dots, s_{t-q+1} are computable from Y_t , which can be seen by first computing s_{t-q+1} then V_{t-q+1} and repeating this procedure. With $h(\cdot \mid \cdot)$ denoting a generic conditional density with respect to an appropriate

measure the $2q$ -step transition kernel can be rewritten as follows (for exposition only the derivations for $q = 2$ are presented).

$$\begin{aligned}
h(Y_t \mid Y_{t-4}) &= h(\varepsilon_t \mid \varepsilon_{t-1}, V_{t-2}, V_{t-3}, Y_{t-4}) h(\varepsilon_{t-1} \mid V_{t-2}, V_{t-3}, Y_{t-4}) \\
&\quad h(V_{t-2} \mid V_{t-3}, Y_{t-4}) h(V_{t-3} \mid Y_{t-4}) \\
&= h(\varepsilon_t \mid s_t, \varepsilon_{t-1}, \varepsilon_{t-2}) h(\varepsilon_{t-1} \mid s_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}) \\
&\quad h(V_{t-2} \mid s_{t-2}, V_{t-3}, V_{t-4}, \varepsilon_{t-3}, \varepsilon_{t-4})) \\
&\quad h(V_{t-3} \mid s_{t-3}, V_{t-4}, V_{t-5}, \varepsilon_{t-4}, \varepsilon_{t-5})).
\end{aligned}$$

By (iv) and since the conditional covariance matrices are always positive definite the last four densities are Lebesgue densities, strictly positive, and bounded away from zero on compact sets. Hence the $2q$ -step transition kernel for Y_t will also have a strictly positive density, which is bounded away from zero on compact sets. This establishes that the Markov chain Y_t is irreducible, aperiodic, and that compact sets are small and we proceed by applying the drift criterion of Tjøstheim (1990).

Denote a generic element of the Markov chain Y_t by $y = (\bar{v}', \bar{e}')'$ and define the drift function f by

$$f(Y_t) = 1 + k \bar{V}'_{t-q} D \bar{V}_{t-q} + \sum_{i=1}^p \sum_{j=0}^{q-1} k_{i,j} \varepsilon_{i,t-j}^2, \quad D \equiv \sum_{i=0}^{\infty} \mathbf{A}^{j'} \mathbf{A}^j, \quad (14)$$

where D is well defined as $\rho(\mathbf{A} \otimes \mathbf{A}) < 1$ by assumption and the positive constants $k, k_{i,j}$ will be specified later. The choice of the matrix D follows Feigin & Tweedie (1985) and it implies the existence of second-order moments of Y_t . Using (6) and

following Bec & Rahbek (2004) it holds that

$$\begin{aligned}
& E[\bar{V}'_{t-q} D \bar{V}_{t-q} \mid Y_{t-1} = y] \\
&= \bar{v}' \mathbf{A}' D \mathbf{A} \bar{v} + (1 - p(\gamma' \bar{v})) \{ \bar{v}' \mathbf{B}' D \mathbf{B} \bar{v} - \bar{v}' \mathbf{A}' D \mathbf{A} \bar{v} \} + E[\eta'_{t-q} D \eta_{t-q} \mid Y_{t-1} = y] \\
&= \bar{v}' \mathbf{A}' D \mathbf{A} \bar{v} + (1 - p(\gamma' \bar{v})) \{ \bar{v}' (\mathbf{A} - \mathbf{B})' D (\mathbf{A} - \mathbf{B}) \bar{v} - 2 \bar{v}' \mathbf{A}' D (\mathbf{A} - \mathbf{B}) \bar{v} \} \\
&\quad + E[\eta'_{t-q} D \eta_{t-q} \mid Y_{t-1} = y] \\
&= \bar{v}' D \bar{v} - \bar{v}' \bar{v} + (1 - p(\gamma' \bar{v})) \{ (\bar{v}' \gamma) \bar{\gamma}' (\mathbf{A} - \mathbf{B})' D (\mathbf{A} - \mathbf{B}) \bar{\gamma} (\gamma' \bar{v}) \\
&\quad - 2 \bar{v}' \mathbf{A}' D (\mathbf{A} - \mathbf{B}) \bar{\gamma} (\gamma' \bar{v}) \} + E[\eta'_{t-q} D \eta_{t-q} \mid Y_{t-1} = y]. \tag{15}
\end{aligned}$$

In the last equality the projection $I_{pq} = \bar{\gamma} \gamma' + \bar{\mu} \mu'$ and (i) have been used. Define for some $\lambda_c > 1$ the compact set

$$C_v = \{ \bar{v} \in \mathbb{R}^{pq} \mid \bar{v}' D \bar{v} \leq \lambda_c \}.$$

On the complement of C_v it holds that

$$\frac{\bar{v}' D \bar{v} - \bar{v}' \bar{v}}{\bar{v}' D \bar{v}} = 1 - \frac{\bar{v}' \bar{v}}{\bar{v}' D \bar{v}} \leq 1 - \inf_{\bar{v} \neq 0} \frac{\bar{v}' \bar{v}}{\bar{v}' D \bar{v}} \leq 1 - \frac{1}{\rho(D)},$$

where $\rho(\cdot)$ denotes the spectral radius of a square matrix. Furthermore note that

$$(\bar{v}' \gamma) \bar{\gamma}' (\mathbf{A} - \mathbf{B})' D (\mathbf{A} - \mathbf{B}) \bar{\gamma} (\gamma' \bar{v}) \asymp \|\gamma' \bar{v}\|^2$$

and that

$$2 \bar{v}' \mathbf{A}' D (\mathbf{A} - \mathbf{B}) \bar{\gamma} (\gamma' \bar{v}) \asymp \|\bar{v}\| \|\gamma' \bar{v}\|,$$

where $h_1(x) \asymp h_2(x)$ denotes that $h_1(x)/h_2(x)$ tends to a non-zero constant as $\|x\|$ tends to infinity. However, by assumption the drift function satisfies $f(y) \asymp \|y\|^2 = \|\bar{\gamma} \gamma' \bar{v} + \bar{\mu} \mu' \bar{v}\|^2 + \|\bar{e}\|^2$ and since $(1 - p(\gamma' \bar{v})) \rightarrow 0$ as $\|\gamma' \bar{v}\| \rightarrow \infty$ it can be concluded that

$$\frac{k(1 - \tilde{p}(\gamma' \bar{v})) \{ (\bar{v}' \gamma) \bar{\gamma}' (\mathbf{A} - \mathbf{B})' D (\mathbf{A} - \mathbf{B}) \bar{\gamma} (\gamma' \bar{v}) - 2 \bar{v}' \mathbf{A}' D (\mathbf{A} - \mathbf{B}) \bar{\gamma} (\gamma' \bar{v}) \}}{f(y)} \rightarrow 0,$$

as $\|\bar{v}\| \rightarrow \infty$. Or in other words, for λ_c adequately large it holds that on the

complement of C_v will

$$kE[\bar{V}_{t-q}D\bar{V}_{t-q} \mid Y_{t-1} = y] \leq \frac{(1 - \delta^*)k\bar{v}'D\bar{v}}{f(y)}f(y) + kE[\eta'_{t-q}D\eta_{t-q} \mid Y_{t-1} = y]. \quad (16)$$

The constant δ^* should be chosen such that $\delta^* \in]0, 1[$ and $\frac{1}{\rho(D)} > \delta^* > 0$. Next consider the final term of (16). By construction it will be positive and there exists positive constants c, c_i where $i = 1, \dots, p$ such that

$$\begin{aligned} kE[\eta'_{t-q}D\eta_{t-q} \mid Y_{t-1} = y] &= kE[\varepsilon'_{t-q}(\beta', \beta'_\perp)\varphi'D\varphi(\beta', \beta'_\perp)'\varepsilon_{t-q} \mid Y_{t-1} = y] \\ &\leq ck + k \sum_{i=1}^p c_i e_{i,q}^2. \end{aligned} \quad (17)$$

As previously define the compact set $C_e = \{\bar{e} \in \mathbb{R}^{pq} \mid \|\bar{e}\|^2 \leq \lambda_c\}$. Furthermore if λ_c is chosen large enough (v) yields that on the complement of C_e it holds that

$$\begin{aligned} &c_i k e_{i,q}^2 + \sum_{j=0}^{q-1} k_{i,j} E[\varepsilon_{i,t-j}^2 \mid Y_{t-1} = y] \\ &= c_i k e_{i,q}^2 + k_{i,0} E[\varepsilon_{i,t}^2 \mid Y_{t-1} = y] + \sum_{j=1}^{q-1} k_{i,j} e_{i,j}^2 \\ &\leq c_i k e_{i,q}^2 + k_{i,0} \bar{\sigma}_i \left(1 + \sum_{j=1}^q \alpha_{i,j} e_{i,j}^2 \right) + \sum_{j=1}^{q-1} k_{i,j} e_{i,j}^2 \\ &= K + (k_{i,0} \bar{\sigma}_i \alpha_{i,1} + k_{i,1}) e_{i,1}^2 + \sum_{j=2}^{q-1} (k_{i,0} \bar{\sigma}_i \alpha_{i,j} + k_{i,j}) e_{i,j}^2 \\ &\quad + (c_i k + k_{i,0} \bar{\sigma}_i \alpha_{i,q}) e_{i,q}^2 \end{aligned} \quad (18)$$

for all $i = 1, \dots, p$ where K is some positive constant and $\bar{\sigma}_i = \max_{l=0,1} \Lambda_{i,i}^{(l)}$. When $\bar{\sigma}_i \sum_{j=1}^q \alpha_{i,j} < 1$ the positive constants k and $k_{i,j}$ can be chosen such that the

inequalities

$$\begin{aligned}
k_{i,0}\bar{\sigma}_i\alpha_{i,1} + k_{i,1} &< k_{i,0} \\
k_{i,0}\bar{\sigma}_i\alpha_{i,j} + k_{i,j} &< k_{i,j-1}, \quad j = 2, \dots, q-1 \\
c_i k + k_{i,0}\bar{\sigma}_i\alpha_{i,q} &< k_{i,q-1}
\end{aligned}$$

are all satisfied, which can be seen by setting $k_{i,0} = 1$ and choosing k very small, see Lu (1996) for details. Hence there exists a constant $\delta_i^{**} \in]0, 1[$ such that the coefficient of $e_{i,j+1}^2$ in (18) is smaller than $(1 - \delta_i^{**})k_{i,j}$ for all $j = 0, \dots, q-1$.

By combining (16)-(18) it can be concluded that for y outside the compact set $C = C_v \times C_\varepsilon$ will

$$\begin{aligned}
E[f(Y_t) \mid Y_{t-1} = y] &\leq \frac{(1 - \delta)\bar{v}'D\bar{v} + (1 - \delta) \sum_{i=1}^p \sum_{j=0}^{q-1} k_{i,j} e_{i,j+1}^2}{1 + \bar{v}'D\bar{v} + \sum_{i=1}^p \sum_{j=0}^{q-1} k_{i,j} e_{i,j+1}^2} f(y) \\
&\leq (1 - \delta)f(y),
\end{aligned}$$

where $\delta = \min(\delta^*, \delta_1^{**}, \dots, \delta_p^{**}) > 0$. Inside the compact set C the function $E[f(Y_t) \mid Y_{t-1} = y]$ is continuous and hence bounded. This completes the verification of the drift criterion.

Proof of Corollary 1

Assume without loss of generality that $q = 2$. Under the assumptions listed in the corollary the coefficient matrices of (6) are given by

$$\mathbf{A} = \begin{pmatrix} \beta' a^{(1)} - I_r + \beta' \Gamma_1 \bar{\beta} & \beta' \Gamma_1 \bar{\beta}_\perp & -\beta' \Gamma_1 \bar{\beta} & 0 \\ \beta'_\perp a^{(1)} + \beta'_\perp \Gamma_1 \bar{\beta} & \beta'_\perp \Gamma_1 \bar{\beta}_\perp & -\beta'_\perp \Gamma_1 \bar{\beta} & 0 \\ 0 & 0 & I_r & 0 \\ 0 & 0 & 0 & I_{p-r} \end{pmatrix}, \quad (19)$$

and likewise for \mathbf{B} . Hence

$$(\mathbf{A}-\mathbf{B}) = \begin{pmatrix} \beta'(a^{(1)} - a^{(0)}) + \beta'(\Gamma_1 - G_1)\bar{\beta} & \beta'(\Gamma_1 - G_1)\bar{\beta}_\perp & -\beta'(\Gamma_1 - G_1)\bar{\beta} & 0 \\ \beta'_\perp(a^{(1)} - a^{(0)}) + \beta'_\perp(\Gamma_1 - G_1)\bar{\beta} & \beta'_\perp(\Gamma_1 - G_1)\bar{\beta}_\perp & -\beta'_\perp(\Gamma_1 - G_1)\bar{\beta} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

So the matrices γ and μ can be chosen as

$$\gamma = \begin{pmatrix} I_r & 0 & 0 \\ 0 & I_{p-r} & 0 \\ 0 & 0 & I_r \\ 0 & 0 & 0 \end{pmatrix}, \quad \mu = \begin{pmatrix} 0 \\ 0 \\ 0 \\ I_{p-r} \end{pmatrix},$$

which satisfies $\mu = \gamma_\perp$ and the remaining assumptions of Theorem 1. Finally Theorem 1 yields the desired result.

Proof of Corollary 2

Assume again without loss of generality that $q = 2$. In this case the matrices γ and μ can be chosen as

$$\gamma = \begin{pmatrix} I_r \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mu = \begin{pmatrix} 0 & 0 & 0 \\ I_{p-r} & 0 & 0 \\ 0 & I_r & 0 \\ 0 & 0 & I_{p-r} \end{pmatrix},$$

and an application of Theorem 1 completes the proof.

Proof of Theorem 2

The formulation of the theorem as well as the proof owes much to Theorem 4 of Bec & Rahbek (2004). Initially note that the process X_t given by (1), (2), and (5) can be written as

$$A(L)X_t = (\Phi^{(0)} - \Phi^{(1)})u_t + \varepsilon_t,$$

where L denotes the lag-operator and the polynomial $A(\cdot)$ is defined in (7). By the algebraic identity

$$A(z)^{-1} = C \frac{1}{1-z} + C(z),$$

where $C(z) = \sum_{i=0}^{\infty} C_i z^i$ with exponentially decreasing coefficients C_i it holds that

$$\begin{aligned} X_t &= C \sum_{i=1}^t ((\Phi^{(0)} - \Phi^{(1)})u_i + \varepsilon_i) + C(L)((\Phi^{(0)} - \Phi^{(1)})u_t + \varepsilon_t) \\ &= C \sum_{i=1}^t ((\Phi^{(0)} - \Phi^{(1)})u_i + \varepsilon_i) + \tau_t. \end{aligned} \quad (20)$$

Next Theorem 1 yields that $\beta'X_t$ and ΔX_{t-i} are stationary and in turn that $u_t = (1 - s_t)z_t$ is stationary. Since $C(L)$ has exponentially decreasing coefficients it therefore holds that τ_t is stationary. Hence the common trends of X_t are given by

$$\sum_{i=1}^t c_i = a^{(1)'} \sum_{i=1}^t ((\Phi^{(0)} - \Phi^{(1)})u_i + \varepsilon_i).$$

Since $\|u_t\| \leq \|z_t\|$ it holds by Theorem 17.4.2 and Theorem 17.4.4 of Meyn & Tweedie (1993) that a FCLT applies to c_t provided the long run variance Υ ,

$$\Upsilon = \gamma_{cc}(0) + \sum_{h=1}^{\infty} (\gamma_{cc}(h) + \gamma_{cc}(h)'), \quad \gamma_{cc}(h) = \text{Cov}(c_t, c_{t+h}),$$

is positive definite. Note that the long run variance can be written as

$$\Upsilon = \psi' \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon u} \\ \Sigma_{u\varepsilon} & \Sigma_{uu} \end{pmatrix} \psi, \quad \Sigma_{\varepsilon u} = \gamma_{\varepsilon u}(0) + \sum_{h=1}^{\infty} (\gamma_{\varepsilon u}(h) + \gamma_{\varepsilon u}(h)'). \quad (21)$$

With similar expressions for the remaining Σ matrices.

Proof of Theorem 3

By combining Theorem 1 and Theorem 2 one can mimicking the proof of Lemma 13.1 of Johansen (1995) in order to establish the result.

Proof of Theorem 4

Before proving Theorem 4 we initially state and prove some auxiliary lemmas. For notational ease we adopt the convention $\varepsilon_t(\theta_0) = \varepsilon_t$ and likewise for other functions of the parameter vector evaluated in the true parameters. Furthermore, define $\varepsilon_t^d = \text{diag}(\varepsilon_t)$ and let $\mathbf{1}_{(d_1 \times d_2)}$ denote a d_1 times d_2 matrix of ones.

Lemma 1. *Under the assumptions of Theorem 4 it holds that*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial l_t(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, \Omega_S),$$

with $\Omega_S = E[(\partial l_t(\theta_0)/\partial \theta)(\partial l_t(\theta_0)/\partial \theta')] > 0$ as T tends to infinity.

Proof. Initially note that by utilizing the diagonal structure of D_t the first derivatives evaluated at the true parameters can be written as

$$\frac{\partial l_t}{\partial \theta^{(1)}} = -\frac{\partial \varepsilon'_t}{\partial \theta^{(1)}} H_t^{-1} \varepsilon_t - \frac{1}{2} \frac{\partial \Pi'_t}{\partial \theta^{(1)}} D_t^{-1} \xi_t \mathbf{1}_{(p \times 1)} \quad (22)$$

$$\frac{\partial l_t}{\partial \theta^{(2)}} = -\frac{1}{2} \frac{\partial \Pi'_t}{\partial \theta^{(2)}} D_t^{-1} \xi_t \mathbf{1}_{(p \times 1)} \quad (23)$$

$$\begin{aligned} \frac{\partial l_t}{\partial \theta^{(3)}} &= s_t \frac{\partial \text{vec}(\Lambda_1)'}{\partial \theta^{(3)}} \text{vec}(\Lambda_1^{-1} - \Lambda_1^{-1} D_t^{-1/2} \varepsilon_t \varepsilon'_t D_t^{-1/2} \Lambda_1^{-1}) \\ &\quad + (1 - s_t) \frac{\partial \text{vec}(\Lambda_0)'}{\partial \theta^{(3)}} \text{vec}(\Lambda_0^{-1} - \Lambda_0^{-1} D_t^{-1/2} \varepsilon_t \varepsilon'_t D_t^{-1/2} \Lambda_0^{-1}) \end{aligned} \quad (24)$$

$$\frac{\partial \Pi'_t}{\partial \theta^{(1)}} = 2 \sum_{j=1}^q \frac{\partial \varepsilon'_{t-j}}{\partial \theta^{(1)}} A_j \varepsilon_{t-j}^d, \quad A_j = \text{diag}(\alpha_{1,j}, \dots, \alpha_{p,j})$$

$$\frac{\partial \varepsilon'_t}{\partial \theta^{(1)}} = s_t J_1 z_t + (1 - s_t) J_2 z_t$$

$$\xi_t = I_p - \varepsilon_t^d H_t^{-1} \varepsilon_t^d,$$

where J_1 is a $2p(r + p(q - 1))$ times $r + p(q - 1)$ matrix with all ones on the first $p(r + p(q - 1))$ rows and zeros on the remaining rows and the matrix J_2 the

opposite. Finally, the derivative of Π_t' with respect to $\theta^{(2)}$ is a block diagonal matrix with the vectors $(\varepsilon_{1,t-1}, \dots, \varepsilon_{1,t-q})'$ to $(\varepsilon_{p,t-1}, \dots, \varepsilon_{p,t-q})'$ on the diagonal blocks.

Next, note that since the ARCH parameters are all bounded away from zero there exists a constant $k_1 > 0$ such that $\varepsilon_{i,t-j}^2/\pi_{i,t} \leq 1/\alpha_{i,j} < k_1$ for all $j = 1, \dots, q$. By repeating this argument one can conclude that there exists a constant k_2 such that $\|(\partial \Pi_t'/\partial \theta D_t^{-1})\| < k_2$. Combining this with the observations that $E[\|\xi_t\|^2] < \infty$ and $E[\|\nu_t\|^4] < \infty$ yields that $E[\|\partial l_t/\partial \theta\|^2] < \infty$ and $\Omega < \infty$.

For any vector c with same dimension as θ define the sequence $l_t^{(1)} = c' \partial l_t / \partial \theta c$, which is a martingale difference sequence with respect to the natural filtration $\mathbb{F}_t = \sigma(X_t, X_{t-1}, \dots)$ since $E[\xi_t | \mathbb{F}_{t-1}] = 0$ and s_t is \mathbb{F}_{t-1} measurable. Under the stated conditions Theorem 1, the law of large number for geometrically ergodic time series, and the central limit of Brown (1971) yield that $T^{-1/2} \sum_{t=1}^T l_t^{(1)} \xrightarrow{D} N(0, c' \Omega_S c)$ and the Cramér-Wold device establishes the lemma.

The positive definiteness of Ω_S can be established by noting that Theorem 1 guarantees that $P(s_t = 1) > 0$, $P(s_t = 0) > 0$, and that all elements of Y_t have strictly positive densities. Hence it holds that the $c' \Omega c = 0$ if and only if $c = 0$ and Ω_S is therefore positive definite, see Lange et al. (2007) for details. \square

Lemma 2. *Under the conditions of Theorem 4 there exists an open neighborhood around the true parameter value $N(\theta_0)$ and a positive constant k_3 , such that*

$$\sup_{\theta \in N(\theta_0)} \left\| \frac{\Pi_t(\theta)'}{\partial \theta^{(1)}} D_t^{-1} \right\|_{\max} < k_3, \quad \sup_{\theta \in N(\theta_0)} \left\| \frac{\Pi_t(\theta)'}{\partial \theta^{(2)}} D_t^{-1} \right\|_{\max} < k_3,$$

and $\|\Lambda^{(l)}\|_{\max}^{-1} < k_3$ for $l = 0, 1$, where $\|\cdot\|_{\max}$ denotes the max norm.

Proof. Let $N(\theta_0) = \{\theta \in \mathbb{R}^{\dim(\theta_0)} \mid \|\theta_0 - \theta\|_{\max} < \delta\}$. Next, note that by construction any term in $\frac{\partial l_t(\theta)}{\partial \theta^{(1)}}$ will also be in the relevant part of D_t , hence it can be concluded that if δ is sufficiently small there exists a positive constant k_3 such that

$$\sup_{\theta \in N(\theta_0)} \left\| \frac{\Pi_t(\theta)'}{\partial \theta^{(1)}} D_t^{-1} \right\|_{\max} \leq \frac{1}{\min_{\theta \in N(\theta_0)} \theta^{(2)}} < k_3$$

and likewise for the derivative with respect to $\theta^{(2)}$. Finally, note that since the

true value of both $\Lambda^{(1)}$ and $\Lambda^{(0)}$ are positive definite and the eigenvalues of a matrix is a continuous function of the matrix itself it holds that δ can be chosen such that $\|\Lambda^{(l)}\|_{\max}^{-1} < k_3$ for $l = 0, 1$.

□

Proof of Theorem 4. The proof is based on a Taylor expansion of the log-likelihood function. To avoid the need for third derivatives we will verify conditions (A.1)-(A.4) of Lemma A.1 in Lange et al. (2007). The asymptotic normality of the score evaluated at the true parameter values has been established in Lemma 1, hence condition (A.1) is satisfied. By directly differentiating (22), (23), and (24) and adopting the notation of Lemma 1 one obtains the following expressions for the second derivatives.

$$\begin{aligned} & \frac{\partial^2 l_t(\theta)}{\partial \theta^{(1)} \partial \theta^{(1)'}} \\ = & -\frac{\partial \varepsilon_t(\theta)'}{\partial \theta^{(1)}} H_t(\theta)^{-1} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^{(1)'}} \\ & -\frac{1}{4} \frac{\partial \Pi_t(\theta)'}{\partial \theta^{(1)}} D_t(\theta)^{-1} (\text{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1} \varepsilon_t(\theta)^d \mathbf{1}_{(p \times 1)}\} - \xi_t + I_p) D_t(\theta)^{-1} \frac{\partial \Pi_t(\theta)}{\partial \theta^{(1)'}} \\ & + \frac{\partial \varepsilon_t(\theta)'}{\partial \theta^{(1)}} H_t(\theta)^{-1} \varepsilon_t(\theta)^d D_t(\theta)^{-1} \frac{\partial \Pi_t(\theta)}{\partial \theta^{(1)'}} - \frac{1}{2} \frac{\partial \Pi_t(\theta)'}{\partial \theta^{(1)}} D_t(\theta)^{-2} \text{diag}\{\xi_t(\theta) \mathbf{1}_{(p \times 1)}\} \frac{\partial \Pi_t(\theta)}{\partial \theta^{(1)'}} \\ & - \sum_{j=1}^q \frac{\partial \varepsilon_{t-j}(\theta)'}{\partial \theta^{(1)}} A_j D_t(\theta)^{-1} \text{diag}\{\xi_t(\theta) \mathbf{1}_{(p \times 1)}\} \frac{\partial \varepsilon_{t-j}(\theta)}{\partial \theta^{(1)'}} \\ & + \frac{1}{2} \frac{\partial \Pi_t(\theta)'}{\partial \theta^{(1)}} D_t(\theta)^{-1} (\text{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1} \mathbf{1}_{(p \times 1)}\} + \varepsilon_t^d H_t^{-1}) \frac{\partial \varepsilon_t(\theta)}{\partial \theta^{(1)'}} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \theta^{(1)} \partial \theta^{(2)'}} = & \frac{\partial \varepsilon_t(\theta)'}{\partial \theta^{(1)}} H_t(\theta)^{-1} \varepsilon_t(\theta)^d D_t(\theta)^{-1} \frac{\partial \Pi_t(\theta)}{\partial \theta^{(2)'}} \\ & + \frac{1}{4} \frac{\partial \Pi_t(\theta)'}{\partial \theta^{(1)}} D_t^{-1} [2 \text{diag}\{\xi_t(\theta) \mathbf{1}_{(p \times 1)}\} - \text{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1} \varepsilon_t(\theta)^d \mathbf{1}_{(p \times 1)}\} \\ & \quad - \varepsilon_t(\theta)^d H_t(\theta)^{-1} \varepsilon_t(\theta)^d] D_t^{-1} \frac{\partial \Pi_t(\theta)}{\partial \theta^{(2)'}} \\ & - \sum_{j=1}^q \frac{\partial \varepsilon_{t-j}(\theta)'}{\partial \theta^{(1)}} \varepsilon_{t-j}^d D_t(\theta)^{-1} \xi_t(\theta)^d \frac{\partial (A_j \mathbf{1}_{(p \times 1)})}{\partial \theta^{(2)'}} \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial \theta^{(1)} \partial \theta^{(3)'}} &= s_t \frac{\partial \varepsilon_t(\theta)'}{\partial \theta^{(1)}} ((\varepsilon_t(\theta)' D_t^{-1/2} \Lambda_1^{-1}) \otimes (\Lambda_1^{-1} D_t^{-1/2})) \frac{\partial \text{vec}(\Lambda_1)}{\partial \theta^{(3)'}} \\
&\quad + (1 - s_t) \frac{\partial \varepsilon_t(\theta)'}{\partial \theta^{(1)}} ((\varepsilon_t(\theta)' D_t^{-1/2} \Lambda_0^{-1}) \otimes (\Lambda_0^{-1} D_t^{-1/2})) \frac{\partial \text{vec}(\Lambda_0)}{\partial \theta^{(3)'}} \\
&\quad - \frac{1}{2} \frac{\partial \Pi_t(\theta)'}{\partial \theta^{(1)}} D_t^{-1} \frac{\partial (\xi_t(\theta) \mathbf{1}_{(p \times 1)})}{\partial \theta^{(3)'}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial \theta^{(2)} \partial \theta^{(2)'}} &= \frac{1}{4} \frac{\partial \Pi_t(\theta)'}{\partial \theta^{(2)}} D_t(\theta)^{-1} [2 \text{diag}\{\xi_t(\theta) \mathbf{1}_{(p \times 1)}\} \\
&\quad - \text{diag}\{\varepsilon_t(\theta)^d H_t(\theta)^{-1} \varepsilon_t(\theta)^d \mathbf{1}_{(p \times 1)}\} + \xi_t - I_p] D_t(\theta)^{-1} \frac{\partial \Pi_t(\theta)}{\partial \theta^{(2)'}}
\end{aligned}$$

$$\frac{\partial^2 l_t(\theta)}{\partial \theta^{(2)} \partial \theta^{(3)'}} = -\frac{1}{2} \frac{\partial \Pi_t(\theta)'}{\partial \theta^{(2)}} D_t(\theta)^{-1} \frac{\partial (\xi_t(\theta) \mathbf{1}_{(p \times 1)})}{\partial \theta^{(3)'}}$$

$$\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial \theta^{(3)} \partial \theta^{(3)'}} &= s_t \frac{1}{2} \frac{\partial \text{vec}(\Lambda_1)'}{\partial \theta^{(3)}} (\Lambda_1^{-1} \otimes I_p) [- (\Lambda_1^{-1} D_t(\theta)^{-1/2} \varepsilon_t(\theta) \varepsilon_t(\theta)' D_t(\theta)^{-1/2}) \otimes I_p \\
&\quad - I_p \otimes (\Lambda_1^{-1} D_t(\theta)^{-1/2} \varepsilon_t(\theta) \varepsilon_t(\theta)' D_t(\theta)^{-1/2}) - I_p] (I_p \otimes \Lambda_1^{-1}) \frac{\partial \text{vec}(\Lambda_1)}{\partial \theta^{(3)'}} \\
&\quad + (1 - s_t) \frac{1}{2} \frac{\partial \text{vec}(\Lambda_0)'}{\partial \theta^{(3)}} (\Lambda_0^{-1} \otimes I_p) [- (\Lambda_0^{-1} D_t(\theta)^{-1/2} \varepsilon_t(\theta) \varepsilon_t(\theta)' D_t(\theta)^{-1/2}) \otimes I_p \\
&\quad - I_p \otimes (\Lambda_0^{-1} D_t(\theta)^{-1/2} \varepsilon_t(\theta) \varepsilon_t(\theta)' D_t(\theta)^{-1/2}) + I_p] (I_p \otimes \Lambda_0^{-1}) \frac{\partial \text{vec}(\Lambda_0)}{\partial \theta^{(3)'}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial (\xi_t(\theta) \mathbf{1}_{(p \times 1)})}{\partial \theta^{(3)'}} &= s_t \varepsilon_t(\theta)^d ((\varepsilon_t(\theta) D_t(\theta)^{-1/2} \Lambda_1^{-1}) \otimes (\Lambda_1^{-1} D_t^{-1/2})) \frac{\partial \text{vec}(\Lambda_1)}{\partial \theta^{(3)'}} \\
&\quad + (1 - s_t) \varepsilon_t(\theta)^d ((\varepsilon_t(\theta) D_t(\theta)^{-1/2} \Lambda_0^{-1}) \otimes (\Lambda_0^{-1} D_t^{-1/2})) \frac{\partial \text{vec}(\Lambda_0)}{\partial \theta^{(3)'}}
\end{aligned}$$

By combining Theorem 1 with the law of large numbers for geometrically ergodic time series and Lemma 2 it can be concluded that the

$$\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{P} \Omega_I,$$

as T tends to infinity. By the same arguments as in the proof of Lemma 1 Ω_I is

positive definite. Hence condition (A.2) of Lemma A.1 in Lange et al. (2007) is satisfied.

Next, let $N(\theta_0) = \{\theta \in \mathbb{R}^{\dim(\theta_0)} \mid \|\theta_0 - \theta\|_{\max} < \delta\}$ denote an open neighborhood around the true parameter value. By inspecting the second derivatives and utilizing Lemma 2 it is evident that $\delta > 0$ can be chosen such that there exists a positive constant k_4 and vector k of positive constants such that

$$E\left[\sup_{\theta \in N(\theta_0)} \left\| \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right\| \right] \leq E[k_4(1 + \|\nu_t \nu_t' \nu_t \nu_t'\|_{\max}) k'(z_t', \dots, z_{t-q}')'], \quad (25)$$

which is finite by assumption. One can therefore define a function F as the point-by-point (in θ) limit of $T^{-1} \sum_{t=1}^T \partial^2 l_t(\theta) / \partial \theta \partial \theta'$ and by dominated convergence the function is continuous. Hence condition (A.3) is also satisfied.

Finally, by Theorem 4.2.1 of Amemiya (1985) the required uniform convergence follows from (25) and condition (A.4) is therefore satisfied. Note that Theorem 4.2.1 is applicable in our setup by Amemiya (1985) p. 117 as the law of large numbers applies due to geometric ergodicity of the Markov chain z_t . This completes the proof. \square

References

- Ait-Sahalia, Y. (1996), ‘Testing continuous-time models of the spot interest rate’, *Review of Financial Studies* **2**, 385–426.
- Akram, Q.F. & R. Nymoen (2006), ‘Econometric modelling of slack and tight labour markets’, *Economic Modelling*, **23**, 579–596.
- Amemiya, T. (1985), *Advanced Econometrics*, Basil Blackwell, UK.
- Anderson, H.M. (1997), ‘Transaction costs and nonlinear adjustment towards equilibrium in the US treasury bill market’, *Oxford Bulletin of Economics and Statistics* **59**, 465–484.
- Balke, N.S. & T.B. Fomby (1997), ‘Threshold cointegration’, *International Economic Review* **38**, 627–645.

- Bauwens, L., S. Laurent & J.V.K. Rombouts (2006), ‘Multivariate GARCH models: A survey’, *Journal of Applied Econometrics* **21**, 79–109.
- Bec, F., A. Guayb & E. Guerre (2008), ‘Adaptive consistent unit-root tests based on autoregressive threshold model’, *Journal of Econometrics* **142**, 94–133.
- Bec, F. & A. Rahbek (2004), ‘Vector equilibrium correction models with non-linear discontinuous adjustments’, *The Econometrics Journal* **7**, 628–651.
- Bollerslev, T. (1986), ‘Generalized autoregressive conditional heteroskedasticity’, *Journal of Econometrics* **31**, 307–327.
- Bollerslev, T. (1990), ‘Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model’, *The Review of Economics and Statistics* **72**, 498–505.
- Brown, B. M. (1971), ‘Martingale central limit theorems’, *Annals of Mathematical Statistics* **42**, 59–66.
- Campbell, J. & R. Shiller (1987), ‘Cointegration and tests of present value models’, *Journal of Political Economy* **95**, 1062–1088.
- Chow, Y. (1998), ‘Regime switching and cointegration tests of the efficiency of futures markets’, *Journal of Futures Markets* **18**, 871–901.
- Corradia, V., N. R. Swanson & H. White (2000), ‘Testing for stationarity-ergodicity and for comovements between nonlinear discrete time Markov processes’, *Journal of Econometrics* **96**, 39–73.
- Davies, R.B. (1977), ‘Hypothesis testing when a nuisance parameter is present only under the alternative’, *Biometrika* **64**, 247–254.
- Davies, R.B. (1987), ‘Hypothesis testing when a nuisance parameter is present only under the alternative’, *Biometrika* **74**, 33–43.
- Dufrenot, G. & V. Mignon (2002), *Recent Developments in Nonlinear Cointegration with Applications to Macroeconomics and Finance*, Kluwer Academic Publishers.

- Dumas, B. (1992), ‘Dynamic equilibrium and the real exchange rate in a spatially separated world’, *Review of Financial Studies* **5**, 153–180.
- Enders, W. & P.L. Siklos (2001), ‘Cointegration and threshold adjustment’, *Journal of Business and Economic Statistics* **19**, 166–176.
- Engle, R.F. (1982), ‘Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation’, *Econometrica* **50**, 987–1007.
- Escribano, A. (2004), ‘Nonlinear error correction: The case of money demand in the United Kingdom’, *Macroeconomic Dynamics* **8**, 76–116.
- Evans, M.D. & K.K. Lewis (1994), ‘Do stationary risk premia explain it all? Evidence from the term structure’, *Journal of Monetary Economics* **33**, 285–318.
- Feigin, P.D. & R.L. Tweedie (1985), ‘Random coefficient autoregressive processes. a Markov chain analysis of stationarity and finiteness of moments’, *Journal of Time Series Analysis* **6**, 1–14.
- Forbes, C. S. & P. Kofman (2000), Bayesian soft target zones. Working Paper 4/2000, Monash University, Australia.
- Gourieroux, C. & C.Y. Robert (2006), ‘Stochastic unit root models’, *Econometric Theory* **22**, 1052–1090.
- Hall, A.D., H.M. Anderson & C.W.J. Granger (1992), ‘A cointegration analysis of treasury bill yields’, *Review of Economics & Statistics* **74**, 116–126.
- Hendry, D.F. & N.R. Ericsson (1991), ‘An econometric analysis of UK money demand in monetary trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz’, *The American Economic Review* **81**, 8–38.
- Jensen, S.T. & A. Rahbek (2007), ‘On the law of large numbers for (geometrically) ergodic Markov chains’, *Econometric Theory* **23**, 761–766.
- Johansen, Soren (1995), *Likelihood-Based inference in cointegrated vector autoregressive models*, Oxford University Press.

- Johansen, Søren (2007), Cointegration. Overview and development, in ‘Handbook of Financial Time Series’, Springer.
- Kapetanios, G., Y. Shin & A. Snell (2006), ‘Testing for cointegration in nonlinear smooth transition error correction models’, *Econometric Theory* **22**, 279–303.
- Krolzig, H.M.L., M.L. Marcellino & G.E.L. Mizon (2002), ‘A Markov-switching vector equilibrium correction model of the UK labour market’, *Empirical Economics* **27**, 233–254.
- Lange, T. & A. Rahbek (2008), *Handbook of Financial Time Series*, Springer, chapter Regime Switching Time Series Models: A survey.
- Lange, T., A. Rahbek & S.T. Jensen (2007), Estimation and asymptotic inference in AR-ARCH model. Working paper Department of Mathematical Sciences, University of Copenhagen.
- Lu, Z. (1996), ‘A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model’, *Statistics & Probability Letters* **30**, 305–311.
- Meyn, S.P. & R.L. Tweedie (1993), *Markov chains and stochastic stability*, Communications and control engineering series, Springer-Verlag, London ltd., London.
- Saikkonen, P. (2005), ‘Stability results for nonlinear error correction models’, *Journal of Econometrics* **127**, 69–81.
- Saikkonen, P. (2008), ‘Stability of regime switching error correction models under linear cointegration’, *Econometric Theory* **24**, 294–318.
- Sercu, P., R. Uppal & C. Van Hulle (1995), ‘The exchange rate in the presence of transactions costs: Implications for tests of purchasing power parity’, *Journal of Finance* **50**, 1309–1319.
- Stock, J. & M. Watson (1988), ‘Testing for common trends’, *Journal of the American Statistical Association* **83**, 1097–1107.
- Tjøstheim, D. (1990), ‘Non-linear time series and Markov chains’, *Advances in Applied Probability* **22**, 587–611.

Tzavalis, E. & M. Wickens (1998), ‘A re-examination of the rational expectations hypothesis of the term structure: Reconciling the evidence from long-run and short-run tests’, *International Journal of Finance & Economics* **3**, 229–239.

Research Papers 2008



- 2008-52: Katarzyna Lasak: Likelihood based testing for no fractional cointegration
- 2008-53: Katarzyna Lasak: Maximum likelihood estimation of fractionally cointegrated systems
- 2008-54: Andrew J. Patton and Allan Timmermann: The Resolution of Macroeconomic Uncertainty: Evidence from Survey Forecast
- 2008-55: Carlos Capistrán and Allan Timmermann: Forecast Combination With Entry and Exit of Experts
- 2008-56: Carlos Capistrán and Allan Timmermann: Disagreement and Biases in Inflation Expectations
- 2008-57: Almut E. D. Veraart: Impact of time-inhomogeneous jumps and leverage type effects on returns and realised variances
- 2008-58: Dennis Kristensen and Yongseok Shin: Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood
- 2008-59: Per Frederiksen and Frank S. Nielsen: Testing for long memory in potentially nonstationary perturbed fractional processes
- 2008-60: Thomas Q. Pedersen: Intertemporal Asset Allocation with Habit Formation in Preferences: An Approximate Analytical Solution
- 2008-61: Jean Jacod, Mark Podolskij and Mathias Vetter: Limit theorems for moving averages of discretized processes plus noise
- 2008-62: Giuseppe Cavaliere, David I. Harvey, Stephen J. Leybourne and A.M. Robert Taylor: Testing for Unit Roots in the Presence of a Possible Break in Trend and Non-Stationary Volatility
- 2008-63: Ole E. Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde and Neil Shephard: Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading
- 2009-01: Roman Frydman, Michael D. Goldberg, Søren Johansen and Katarina Juselius: A Resolution of the Purchasing Power Parity Puzzle: Imperfect Knowledge and Long Swings
- 2009-02: Morten Ørregaard Nielsen: Nonparametric Cointegration Analysis of Fractional Systems With Unknown Integration Orders
- 2009-03: Andrés González, Kirstin Hubrich and Timo Teräsvirta: Forecasting inflation with gradual regime shifts and exogenous information
- 2009-4: Theis Lange: First and second order non-linear cointegration models