Discussion Papers
Department of Economics
University of Copenhagen

No. 12-02

Unit root vector autoregression with volatility induced stationarity

Anders Rahbek and Heino Bohn Nielsen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

UNIT ROOT VECTOR AUTOREGRESSION WITH VOLATILITY INDUCED STATIONARITY

ANDERS RAHBEK AND HEINO BOHN NIELSEN

Department of Economics, University of Copenhagen

ABSTRACT: We propose a discrete-time multivariate model where lagged levels of the process enter both the conditional mean and the conditional variance. This way we allow for the empirically observed persistence in time series such as interest rates, often implying unit-roots, while at the same time maintain stationarity despite such unit-roots. Specifically, the model bridges vector autoregressions and multivariate ARCH models in which residuals are replaced by levels lagged. An empirical illustration using recent US term structure data is given in which the individual interest rates have unit roots, have no finite first-order moments, but remain strictly stationary and ergodic, while they co-move in the sense that their spread has no unit root. The model thus allows for volatility induced stationarity, and the paper shows conditions under which the multivariate process is strictly stationary and geometrically ergodic. Interestingly, these conditions include the case of unit roots and a reduced rank structure in the conditional mean, known from linear co-integration to imply non-stationarity. Asymptotic theory of the maximum likelihood estimators for a particular structured case (so-called self-exciting) is provided, and it is shown that $\sqrt{T}-$convergence to Gaussian distributions apply despite unit roots as well as absence of finite first and higher order moments. Monte Carlo simulations confirm the usefulness of the asymptotics in finite samples.

KEYWORDS: Vector Autoregression; Unit-Root; Reduced Rank; Volatility Induced Stationarity; Term Structure; Double Autoregression.
JEL CLASSIFICATION: C32.

Rahbek is also affilliated with CREATES. Funding from the The Danish Council for Independent Research, Social Sciences, is gratefully acknowledged (Grant no. 10-07977). We would like to thank Søren Tolver Jensen for numerous discussions. Also we are grateful to comments from Nour Meddahi, Stephane Gregoir and Søren Johansen. The paper was presented at financial econometrics conferences in Copenhagen (HUKU), Heidelberg and Toulouse.

1 Introduction and Summary

This paper presents a new multivariate time series model which captures important stylized facts of the dynamics of term structure data. In particular the model allows for the typically observed persistence in interest rates often detected as unit-roots in the conditional mean in empirical analyses. But contrary to classic autoregressive models, the variables here enter the conditional variance as well. This can induce stationarity, such that despite unit-roots the multivariate process is stationary. In short, the model allows for co-movement of individual series such that for example spreads have no unit-roots, while at the same time the individual time series are allowed to have unit-roots and to be persistent but remain stationary. We present theory for inference as well as discuss properties of the applied model, and moreover demonstrate by our empirical analysis that the model captures surprisingly well dynamic features of US term structure data. Regarding the results on inference we show that standard asymptotic inference applies despite the fact that an immediate implication of the unit-roots present in the model is that the processes will have fat tails and only finite low order moments.

Our insistence on allowing for unit-roots is based on the rich literature in econometrics from which it is by now a stylized empirical fact that term structure data, as well as other financial economic time series, are persistent and appear to have unit-roots when modelled as autoregressive (AR) processes. However, the implied non-stationarity is often questioned from an economic, or finance, point of view, and alternative models which allow unit-root in the conditional mean, but are stationary due to the formulation of the conditional volatility have been proposed in the literature. A large part of such literature deals with continuous time models, where a key example is the celebrated Cox-Ingersoll-Ross model, cf. Cox et al. (1985), where both drift and volatility terms are functions of the level of the continuous time process. Our proposed model also fits well within the rich term structure literature, both in continuous as well as in discrete time, see for example Gourieroux, Monfort and Polimenis (2002), Le, Singleton and Dai (2010) and Carta, Fantazzini and Maggi (2008) for affine term structure models.

With respect to discrete time series models, the proposed model bridges two different time series modeling approaches: *co-integration analysis*, which is multivariate and allows for unit-roots or reduced rank, and the more recent *double autoregressive modeling*, which so far has been univariate. In the co-integrated vector AR models, see Johansen (1996), unit-roots, or equivalently, reduced rank in the conditional mean imply that the stochastic trends are non-stationary and have random walk behavior. This contrasts the univariate double autoregressive (DAR) models where a unit-root does not necessarily imply non-stationarity as lagged values of the process enter the conditional variance, see e.g. Borkovec and Klüppelberg (2001) and Ling (2004). To fix ideas consider initially the univariate DAR of order one as given by,

$$x_t = x_{t-1} + \left[\sqrt{\omega_{xx} + \phi_{xx}^2 x_{t-1}^2}\right] z_{xt}, \tag{1}$$

where z_{xt} is i.i.d. $\mathrm{N}(0,1)$, and $\omega_{xx} > 0$, $\phi_{xx}^2 > 0$. Despite the unit-root in the conditional mean, the process is strictly stationary provided $0 < \phi_{xx}^2 \lesssim 2.42$, which contrasts the case where squared lagged differences enter the conditional variance in which case unit-roots indeed imply non-stationarity, see Lange, Rahbek and Jensen (2011). In terms of our proposed term structure modeling the univariate unit-root DAR can be thought of as a model for the short term interest rate, driving the level and volatility of interest rates with different maturities. Consider thus next y_t defined as the spread between an interest rate with long maturity and the short term rate x_t, with the dynamics of y_t given by,

$$y_t = \rho y_{t-1} + \left[\sqrt{\omega_{yy} + \phi_{yy}^2 y_{t-1}^2 + \phi_{yx}^2 x_{t-1}^2}\right] z_{yt}, \tag{2}$$

with z_{yt} i.i.d. $\mathrm{N}(0,1)$, and ω_{yy}, ϕ_{yy}^2, $\phi_{yx}^2 > 0$. This way, the lagged short term rate x_t, enters as a variance factor of the spread y_t. That is, a stationary factor with a unit-root enters the conditional variance of the spread, where if $|\rho| < 1$ the interest rates are co-moving. Note that in terms of the affine term structure framework of Le, Singleton and Dai (2010) also the joint process $(y_t, x_t)'$ can be used as a multivariate (unobservable) factor.

Within the unit-root and co-integration literature much attention has been devoted to the inclusion and role of constant terms. A key problem in this strand of literature is that a constant term μ may induce a linear trend due to the implied aggregation as caused by the unit-roots. This has lead to various ways of including restricted constants, and other deterministic terms, in multivariate co-integrated vector AR models, cf. Johansen (1996). However, as we show below, we do avoid such issues here and can include an unrestricted constant vector μ in our multivariate model. This is novel in the framework of nonlinear time series. In terms of (1), we show that adding a constant term, μ_x say, on the right hand side does not imply a linear trend, nor that the properties of x_t in terms of stationarity and (geometric) ergodicity are changed, despite the unit-root in x_t.

For an introduction to the recent univariate DAR models, inference and estimation have been explored for univariate DAR models in Ling (2004, 2007) and Ling and Li (2008), while extremal and tail behavior have been analyzed for DAR models of order one in Borkovec (2000), Borkovec and Klüppelberg (2001) and Klüppelberg and Pergamenchtchikov (2004). As also used in the mentioned references DAR processes with Gaussian innovations may be restated as random coefficient autoregressions (RCAR), for which Aue, Horváth and Steinebach (2006) and Berkes, Horváth and Ling (2009) provide results, as well as key references, on estimation and inference. In fact, the RCAR approach is applied in Fong and Li (2004), where co-integration is discussed with random coefficients. The parametrization in Fong and Li (2004) of the conditional variance is quite different when compared to ours, and moreover Fong and Li (2004) apply a local approach where the conditional variance parameters loading the levels vanish at the rate of T, where T denotes the number of observations. Klüppelberg and Pergamenchtchikov (2007) study extremal behavior of a class of multivariate RCAR processes with finite second order moments, which excludes the unit-roots which is a main interest here.

The paper is structured as follows: In the next section we describe in detail our proposed multivariate model. The model is a vector AR model with reduced rank structure in the conditional mean, allowing for unit-roots, while the conditional variance is cast in line with multivariate BEKK ARCH models but in levels. Next, we derive conditions for stationarity, geometric ergodicity and existence of moments, where it is emphasized that the unit-roots imply that the process in general, while being stationary, will only have finite small order moments. Asymptotic theory of the maximum likelihood estimators for the applied US term structure model is given. It is found that despite the fact that the processes lack finite even second and first order moments, maximum likelihood estimators (MLEs) are asymptotically standard $\sqrt{T}-$Gaussian distributed, which is supported by Monte Carlo simulations. Observe that our focus is on the Markovian case of one lag in the conditional mean and variance; in the last sub-section we discuss the general non-Markovian case of further lags as well as other extensions left for future research. It is important though to stress that we found for the US term structure application the Markovian case to be sufficient.

Throughout, the following notation is applied: The symbols $\xrightarrow{D}$ and $\xrightarrow{P}$ are used to indicate weak convergence and convergence in probability, respectively. For any $p \times r$ matrix α of rank r, $r < p$, let $\alpha_{\perp}$ denote a $p \times (p-r)$ matrix whose columns form a basis of the orthogonal complement of span(α). For any square matrix A, $|A|$ denotes the determinant, $\rho(A)$ denotes the spectral radius, while $\|A\|$ denotes the norm of A, where the Euclidean norm is given by$\|A\|^2 = tr\{A'A\}$. We furthermore apply the non-standard notation $A^{\times 2} := AA'$. Finally, for any matrices A and B, $(A \otimes B)$ is the Kronecker product.

2 The Reduced Rank AR Model with Volatility Induced Stationarity

As explained in the introduction we wish to formulate a multivariate AR model which on the one hand allows for unit-roots in the conditional mean, while at the same time, allows levels to appear in the multivariate conditional ARCH part, which – possibly – induce stationarity. The model is notationally a little involved, and we discuss therefore in separate steps the conditional AR mean and the conditional ARCH parametrization to allow for level induced stationarity. As mentioned we study the Markovian case here, and discuss the non-trivial extension of the model to the non-Markovian case in Section 6.

Consider first the conditional mean part:

2.1 Conditional Mean

Consider the p-dimensional vector autoregressive model of order one with a constant term as given by,

$$Y_t = \mu + \Phi Y_{t-1} + \eta_t, \quad \eta_t = \Omega_t^{1/2} z_t$$

where Φ is $(p \times p)$-dimensional, μ is p-dimensional and z_t i.i.d. $\mathrm{N}(0, I_p)$ such that Ω_t is the conditional variance of Y_t. Before specifying the parametrization of the conditional covariance Ω_t, we note that $k \geq 1$ unit-roots in the autoregressive polynomial, $A(z) := I_p - \Phi z$, $z \in \mathbb{C}$, implies that $\Pi := \Phi - I_p$ has reduced rank, $r = p - k$, under the well-known assumption:

Assumption 1 *With $A(z) = I_p - \Phi z$, $z \in \mathbb{C}$, assume that $|A(z)| = 0$ implies that there are k roots at $z = 1$, while the remaining roots are larger than one in absolute value.*

As applied repeatedly in co-integration analysis, the reduced rank r of Π can be parametrized explicitly,

$$\Pi = \alpha\beta' = \sum_{i=1}^{r} \alpha_i \beta_i',$$

where α_i, β_i are p-dimensional vectors, $i = 1, 2, ..., r$, and $\alpha = (\alpha_1, \ldots, \alpha_r)$, $\beta = (\beta_1, \ldots, \beta_r)$. Thus we may rewrite the vector AR model, using Π and first order differences $\Delta Y_t = Y_t - Y_{t-1}$,

$$\Delta Y_t = \mu + \alpha\beta' Y_{t-1} + \varepsilon_t, \quad \varepsilon_t = \Omega_t^{1/2} z_t. \tag{3}$$

Next, observe that under Assumption 1, the following skew-projection identity from Johansen (1996) applies,

$$I_p = \beta_{\perp,\alpha_\perp} \alpha_\perp' + \alpha_\beta \beta', \tag{4}$$

where $\alpha_\perp = (\alpha_{\perp,1}, .., \alpha_{\perp,k})$ and $\beta_\perp = \left(\beta_{\perp,1}, ..., \beta_{\perp,k}\right)$, both of dimension $(p \times k)$, and where $\beta_{\perp,\alpha_\perp} := \beta_\perp \left(\alpha_\perp' \beta_\perp\right)^{-1}$ and $\alpha_\beta := \alpha \left(\beta' \alpha\right)^{-1}$. Use the skew-projection to decompose Y_t as follows,

$$Y_t = \beta_{\perp,\alpha_\perp} \alpha_\perp' Y_t + \alpha_\beta \beta' Y_t. \tag{5}$$

By pre-multiplying in (3) with $\alpha_\perp'$, it follows that the k-dimensional process, $\alpha_\perp' Y_t$, is a unit-root process, and the k linear combinations $\alpha_\perp' Y_t$ in (5) are *common stochastic unit-root factors* which are loaded by the matrix $\beta_{\perp,\alpha_\perp}$ into Y_t. Likewise, pre-multiplying with β' in (3) gives, $\beta' Y_t = \beta'\mu + (I_r + \beta'\alpha)\beta' Y_{t-1} + \beta'\varepsilon_t$. That is, the r linear combinations in the conditional mean, the *conditional mean relationships*, $\beta' Y_t$, are autoregressive with no unit-roots; we say that they co-move.

Observe that as mentioned it is customary in co-integration analysis to consider restrictions on the constant vector μ to avoid the implied aggregation, and hence linear deterministic trend, arising from the reduced rank of Π. Despite the fact that $\alpha'_\perp Y_t$ indeed have unit-roots here these turn out not to imply aggregation of μ, see Theorem 1; this is due to the fact that the lagged levels enter the conditional variance to be defined next.

2.2 Conditional Variance

We propose a version of the BEKK ARCH model of Engle and Kroner (1995) for Ω_t in (3) as it fits well our US data application. Naturally alternative choices for the functional specification of Ω_t could be chosen by replacing lagged residuals (here η_t) with Y_t in the rich class of multivariate ARCH formulations, see e.g. Bauwens, Laurent, and Rombouts (2006) for a survey of multivariate ARCH specifications. While such constructions appear straightforward, we emphasize that each of these choices require separate treatment as the different specifications will have different implications for the properties of the model.

Consider initially the unrestricted BEKK ARCH model of order h from Engle and Kroner (1995) as given by,

$$\Omega_t = \Omega + \sum_{i=1}^{h} (A_i Y_{t-1})^{\times 2}, \tag{6}$$

where A_i are $(p \times p)$ matrices and $\Omega > 0$. Observe that in the BEKK ARCH model, the conditional variance is represented in terms of sums of squared terms and cross-product terms are omitted. The unrestricted BEKK ARCH is discussed in detail in Engle and Kroner (1995), including identification and generality of the parametrization in (6). While one can indeed apply this directly, we propose more structure to the BEKK ARCH such that the impact of the linear combinations $\beta' Y_{t-1}$ and $\alpha'_\perp Y_{t-1}$ in the conditional variance is transparent as in the factor-ARCH models discussed in *inter alia* Bauwens et al. (2006). To do so, we use the skew-projection (4) again, which can be used to rewrite any $(p \times p)$-dimensional matrix ϕ, as follows,

$$\phi = \left[\alpha_\beta \phi_{\beta\beta} + \beta_{\perp,\alpha_\perp} \phi_{\alpha_\perp \beta}\right] \beta' + \left[\beta_{\perp,\alpha_\perp} \phi_{\alpha_\perp \alpha_\perp} + \alpha_\beta \phi_{\beta \alpha_\perp}\right] \alpha'_\perp \tag{7}$$

where α_β $(p \times r)$ and $\beta_{\perp,\alpha_\perp}$ $(p \times k)$ are defined in (4), and $\phi_{\beta\beta} := \beta' \phi \alpha_\beta$ $(r \times r)$, $\phi_{\alpha_\perp \beta} := \alpha'_\perp \phi \alpha_\beta$ $(r \times k)$, $\phi_{\alpha_\perp \alpha_\perp} := \alpha'_\perp \phi \beta_{\perp,\alpha_\perp}$ $(k \times k)$ and finally, $\phi_{\beta \alpha_\perp} := \beta' \phi \beta_{\perp,\alpha_\perp}$ $(r \times k)$.

Using (7) in (6) and omitting cross-product terms, the suggested conditional covariance parametrization is then given by,

$$\Omega_t = \Omega + \left(\alpha_\beta \phi_{\beta\beta} \beta' Y_{t-1}\right)^{\times 2} + \left(\beta_{\perp,\alpha_\perp} \phi_{\alpha_\perp \beta} \beta' Y_{t-1}\right)^{\times 2}$$
$$+ \left(\beta_{\perp,\alpha_\perp} \phi_{\alpha_\perp \alpha_\perp} \alpha_\perp' Y_{t-1}\right)^{\times 2} + \left(\alpha_\beta \phi_{\beta \alpha_\perp} \alpha_\perp' Y_{t-1}\right)^{\times 2}. \tag{8}$$

With Ω_t given in (8), we have imposed a structure which allow the lagged unit-root factors $\alpha_\perp' Y_{t-1}$ as well as lagged conditional mean relations $\beta' Y_{t-1}$ to enter the conditional variance of Y_t. More precisely, the role of the parameters α_β and $\beta_{\perp,\alpha_\perp}$ are to load the factors $\beta' Y_{t-1}$ and $\alpha_\perp' Y_{t-1}$ in the conditional variance of the linear combinations $\beta' Y_t$ and $\alpha_\perp' Y_t$ respectively, as follows from using $\alpha_\perp' \alpha_\beta = 0$ and $\beta' \beta_{\perp,\alpha_\perp} = 0$; and hence the remaining parameters ϕ_{ij}, $i, j = \alpha_\perp, \beta$ provide the "magnitude" or "size" of the loaded factors in the conditional variance.

We add more structure by setting $\phi_{\alpha_\perp \beta} = 0$ in (8),

$$\Omega_t = \Omega + \left(\alpha_\beta \phi_{\beta\beta} \beta' Y_{t-1}\right)^{\times 2} + \left(\beta_{\perp,\alpha_\perp} \phi_{\alpha_\perp \alpha_\perp} \alpha_\perp' Y_{t-1}\right)^{\times 2} + \left(\alpha_\beta \phi_{\beta \alpha_\perp} \alpha_\perp' Y_{t-1}\right)^{\times 2}. \tag{9}$$

This further structure has the immediate implication that the k stochastic common factors $\alpha_\perp' Y_t$ have k unit-roots and a conditional variance driven solely by their own past values since $\alpha_\perp' \alpha_\beta = 0$. That is, in (9) the conditional mean of $\alpha_\perp' \Delta Y_t$ is $\alpha_\perp' \mu$, while the conditional variance is given by,

$$\alpha_\perp' \Omega_t \alpha_\perp = \alpha_\perp' \Omega \alpha_\perp + \left(\phi_{\alpha_\perp \alpha_\perp} \alpha_\perp' Y_{t-1}\right)^{\times 2}.$$

On the other hand the conditional covariance of $\beta' Y_t$ is driven by their own past $\beta' Y_{t-1}$ as well as the common unit-root factors $\alpha_\perp' Y_{t-1}$. We shall say that in this case, the k stochastic factors, $\alpha_\perp' Y_t$, are *self-exciting*, while $\beta' Y_t$ are not. It is this version of the model which we successfully apply to the US term structure, and indeed find that the short rate is a self-exciting stochastic factor driving (parts of) the term structure. It is important to underline that while $\alpha_\perp' Y_t$ marginally is a multivariate unit root DAR process, the model does not exclude covariation in the sense that the conditional covariance, $Cov\left(\beta' Y_t, \alpha_\perp' Y_t \mid Y_{t-1}\right) = \beta' \Omega \alpha_\perp$ is unrestricted since Ω is in (9). That is, it is indeed a multivariate model.

2.3 Parameters of the model

For the statistical analysis the general model given by (3) with Ω_t in (8) or (9) is over-parametrized as β and α $(p \times r)$ as well as $\alpha_\perp$ and $\beta_\perp$ $(p \times k)$ are not identified without imposing just-identifying restrictions. As an example, one can apply the classic identification schemes well-known from co-integration, see Johansen (1996), where identified versions of β and α are given by,

$$\beta_c := \beta \left(c' \beta\right)^{-1} \quad \text{and} \quad \alpha_c := \alpha \left(\beta' c\right), \tag{10}$$

with c a known $(p \times r)$ matrix. Likewise, one may apply identified versions of $\beta_\perp$ and $\alpha_\perp$ as given by,

$$\beta_{\perp,c} := \left(I_p - c\beta_c'\right) c_\perp \quad \text{and} \quad \alpha_{\perp,c} := \left(I_p - \beta_c \left(\alpha_c'\beta_c\right)^{-1} \alpha_c'\right) \beta_{\perp,c}. \tag{11}$$

With this, or some other identification scheme, the parameters of the model are given by α, β, $\alpha_\perp$, $\beta_\perp$ and $\Omega > 0$, in addition to the loadings in the BEKK variance, $\phi_{\beta\beta}$ $(r \times r)$, $\phi_{\alpha_\perp\beta}$ $(k \times r)$, $\phi_{\alpha_\perp\alpha_\perp}$ $(k \times k)$ and $\phi_{\beta\alpha_\perp}$ $(r \times k)$.

3 Stationarity and Geometric Ergodicity

We discuss here under which assumptions of the parameters the conditional mean relations $\beta' Y_t$ are stationary and geometrically ergodic, while also Y_t despite its unit-roots in the characteristic polynomial is geometrically ergodic. As emphasized this is different from co-integration where the unit-roots imply that Y_t is non-stationary, which again leads to asymptotic distributions of parameter estimates characterized by Brownian motions. Here the unit-roots imply ergodicity, but at the same time the processes cease to have finite even small moments and hence imply heavy tails in Y_t.

3.1 AR BEKK model

We start by formulating a general result regarding stationarity and geometric ergodicity in the AR model with unit-roots in the conditional mean part, and the general BEKK ARCH Ω_t in (6) which embeds the restricted parametrizations in (8) and (9),

$$\Delta Y_t = \mu + \sum_{i=1}^{r} \alpha_i \beta_i' Y_{t-1} + \Omega_t^{1/2} z_t, \quad \Omega_t = \Omega + \sum_{i=1}^{h} \left(A_i Y_{t-1}\right)^{\times 2}, \tag{12}$$

with z_t i.i.d. $\mathrm{N}(0, I_p)$. Following this we provide more details for the case applied in the empirical example, where the stochastic factors are self-exciting, see (9).

Our first result states as claimed conditions under which, despite the reduced rank, the process Y_t is stationary and geometrically ergodic:

Theorem 1 *The process Y_t given by (12) is stationary, and geometrically ergodic, if the associated top Lyaponov coefficient is strictly negative, that is,*

$$\gamma := \lim_{q \to \infty} \left[\tfrac{1}{q} E \log \left\| \prod_{t=1}^{q} Q_t \right\| \right] < 0. \tag{13}$$

Here $Q_t := \left(I_p + \alpha\beta' + \epsilon_t\right)$ and ϵ_t is $(p \times p)-$dimensional i.i.d. Gaussian with mean zero and a covariance structure defined by $E\left(\epsilon_t \otimes \epsilon_t\right) = \sum_{i=1}^{h} \left(A_i \otimes A_i\right).$

We emphasize that while the characterization here of stationarity of Y_t for a given set of parameters is implicit, the parameter set for which (13) holds is non-empty, as can be illustrated by simulations described below. We discuss this in more detail for the structured model where the conditional variance is parametrized such that the stochastic factors are self-exicing.

The proof of Theorem 1 is located in Appendix A and is based on rewriting the multivariate process as a random coefficient autoregressive process and application of the drift criterion from Markov chain theory, see also Ling (2007) where a similar approach is used for univariate DAR models of general order. The proof, and hence Theorem 1 is not simple to generalize to other of the existing general covariance structures from multivariate ARCH in Bauwens et al. (2006), as many of the non-BEKK models, including the constant conditional correlation model, can not be written on random coefficient form.

Remark 1 *Computing the Lyaponov coefficient in (13) by simulation, one may use that,*

$$\gamma = \lim_{q\to\infty}\left[\tfrac{1}{q}\log\left\|\prod_{t=1}^{q} Q_t\right\|\right] \quad a.s.,$$

as also noted by several authors, see e.g. Ling (2007) and Francq and Zakoian (2010, Theorem 2.3). There is a rich and general literature on efficient computation of $\log\|\prod_{t=1}^{q} Q_t\|$ *and hence* γ*. In the illustrations of this paper we applied ideas from Dieci and Van Vleck (1995), where QR-decompositions are considered.*

Remark 2 *The classic co-integrated AR model is obtained with* $A_i = 0$*, in which case it is well-known that* Y_t *is non-stationary. This conforms with our result as in this case* $Q_t = Q$ *is non-stochastic, and under Assumption 1,*

$$\gamma = \lim_{q\to\infty}\tfrac{1}{q}\log\left\|\prod_{t=1}^{q} Q_t\right\| = \log\rho(Q) = 0.$$

Note also in this respect, that with $\gamma = 0$ *then, contrary to the case of* $\gamma < 0$*, the constant term* μ *accumulates and generates a linear trend in the* $\alpha_\perp' Y_t$ *process.*

Remark 3 *From Xie and Huang (2009, equation 1.2) it follows that* Y_t *has finite* δ'*th order moment,* $E\|Y_t\|^\delta < \infty$*, if the moment Lyapunov coefficient* $\rho_\delta < 0$ *where* $\rho_\delta := \lim_{q\to\infty}\left[\frac{1}{q}\log E\|\prod_{t=1}^{q} Q_t\|^\delta\right]$.

Remark 4 *As used in Klüppelberg and Pergamenchtchikov (2007), the general Lyapunov condition (13) is implied by the more explicit criterion,*

$$\rho\left(E\left(Q_t \otimes Q_t\right)\right) < 1. \tag{14}$$

However, this is a strong assumption for the multivariate case, in the sense that it implies that not only is Y_t *geometrically ergodic but also that it has finite second order moments,*

$E\,\|Y_t\|^2 < \infty$. *Moreover, by definition this stronger assumption does not allow for unit-roots in* $A(z)$, *or equivalently the reduced rank of* Π, $\Pi = \alpha\beta'$, *and hence cannot be used to verify stationarity of* Y_t *under Assumption 1.*

3.2 Self-exciting Stochastic Factors

When turning to the more structured case where the stochastic factors of the model, $\left(\alpha'_{\perp j} Y_t\right)_{j=1,\ldots,k}$, are self-exciting much more can be said in terms of verification of geometric ergodicity. In addition to results for the self-exciting case, we propose a further reduction of the model, which we label as *separability.* Under separability we may have that the conditional mean relations $(\beta'_i Y_t)_{i=1,\ldots,r}$ are stationary with finite second order moments, while the stochastic factors do not have any finite moments.

Consider the p-dimensional process Y_t as given by (3) and (9) with self-exciting factors which we re-state here,

$$\Delta Y_t = \mu + \alpha\beta' Y_{t-1} + \eta_t, \quad \eta_t = \Omega_t^{1/2} z_t, \tag{15}$$

$$\Omega_t = \Omega + \left[\alpha_\beta \phi_{\beta\beta}\beta' Y_{t-1}\right]^{\times 2} + \left[\alpha_\beta \phi_{\beta\alpha_\perp}\alpha'_\perp Y_{t-1}\right]^{\times 2} + \left[\beta_{\perp,\alpha_\perp}\phi_{\alpha_\perp\alpha_\perp}\alpha'_\perp Y_{t-1}\right]^{\times 2}, \tag{16}$$

using the already applied convention that $a_b = a\,(b'a)^{-1}$ for any $(p \times m)$ matrices a, b such that $b'a$ has full rank. The next theorem states that Y_t in this self-exciting case is stationary as induced by the conditional volatility. Moreover, the condition for stationarity does not depend on the contribution from the term $\left[\alpha_\beta \phi_{\beta\alpha_\perp}\alpha'_\perp Y_{t-1}\right]^{\times 2}$. Stated differently, whatever the value or size of $\phi_{\beta\alpha_\perp}$ is, while the individual realizations of Y_t will vary with $\phi_{\beta\alpha_\perp}$, the conclusions regarding geometric ergodicity and stationarity remain unaffected.

Thus we consider the following regularity conditions:

Assumption 2 *Define* $\varrho_t = I_r + \beta'\alpha + \epsilon_t^\varrho$, *and* $\gamma_t = I_{p-r} + \epsilon_t^\gamma$, *where* ϵ_t^ϱ *and* ϵ_t^γ *are independent and i.i.d. Gaussian distributed, with mean zero and covariance structure given by* $E\left(\epsilon_t^\varrho \otimes \epsilon_t^\varrho\right) = \left(\phi_{\beta\beta} \otimes \phi_{\beta\beta}\right)$ *and* $E\left(\epsilon_t^\gamma \otimes \epsilon_t^\gamma\right) = \left(\phi_{\alpha_\perp\alpha_\perp} \otimes \phi_{\alpha_\perp\alpha_\perp}\right)$. *Assume that,*

$$\lambda_\varrho := \lim_{q\to\infty}\left[\tfrac{1}{q} E \log \left\|\prod_{t=1}^{q} \varrho_t\right\|\right] < 0 \quad \textit{and} \quad \lambda_\gamma := \lim_{q\to\infty}\left[\tfrac{1}{q} E \log \left\|\prod_{t=1}^{q} \gamma_t\right\|\right] < 0. \tag{17}$$

One may observe that ϱ_t corresponds to a RCAR model for $\beta' Y_t$ without the $\alpha'_\perp Y_t$ process, and in this sense the assumption addresses a "skeleton" process for $\beta' Y_t$. The fact that $\alpha'_\perp Y_t$ is self-exciting means that γ_t simply is the RCAR coefficient for this process. Note also that the covariance Ω in Ω_t plays no role in the determination of the stochastic behavior.

Thus the next theorem shows as claimed that $\phi_{\beta\alpha_\perp}$ play no role for the dynamic properties of Y_t. The proof is given in Appendix A and is based on application of a novel drift-function for the drift criterion from Markov chain theory:

Theorem 2 *Under Assumption 1 and Assumption 2, the p-dimensional process Y_t in (15)-(16) is geometrically ergodic, and has a stationary representation with $E\left\|Y_t\right\|^{\delta}<\infty$, for some small $\delta \in (0,1)$.*

Next, we consider the mentioned concept of *separability*, by this we mean that $\phi_{\beta\alpha_\perp}=0$ in (16). That is, $\beta' Y_{t-1}$ enter the variance only in the linear combinations $\beta' Y_t$, and likewise for $\alpha_\perp' Y_t$. Note though that Ω is not restricted to be block-diagonal. In the case of separability we observe:

Corollary 1 *Assume that Assumptions 1-2 hold such that the p-dimensional process Y_t in (15)-(16) is geometrically ergodic. Assume furthermore that we have separability, $\phi_{\beta\alpha_\perp}=0$, and that*

$$\rho\left(E\left(\varrho_t \otimes \varrho_t\right)\right)<1.$$

Then the conditional mean relations $\beta' Y_t$ are geometrically ergodic and strongly mixing with geometric rate. Furthermore, they have finite second order moments, $E\left\|\beta' Y_t\right\|^2<\infty$.

Remark 5 *Regarding tail and extremal behavior of the r-dimensional process, $\beta' Y_t$, some further results can be deduced from Klüppelberg and Pergamenchtchikov (2004, 2007) under the assumptions in Corollary 1 of separability. In particular, they find under regularity conditions (Klüppelberg and Pergamenchtchikov, 2007, condition H_0), that the tails may be characterized as Pareto-like, that is, the conditional mean relations $\beta' Y_t$ here have finite second order moments and a tail index $\kappa_\beta>2$.*

Remark 6 *Regarding tail and extremal properties of the k-dimensional stochastic factors, $\alpha_\perp' Y_t$ to our knowledge no results are known for $k>1$, while for $k=1$ these are studied in detail in Borkovec (2000) and Borkovec and Klüppelberg (2001). In particular, Borkovec and Klüppelberg (2001, Proposition 2) implies Pareto-like tails with index $\kappa_{\alpha_\perp}<2$ such that $\kappa_\beta>\kappa_{\alpha_\perp}$ as expected.*

4 Asymptotics

As mentioned in the introduction we derive the asymptotics for the bivariate model, which is the model applied in the empirical illustration. In particular it is shown that classic $\sqrt{T}-$convergence to Gaussian distributions apply to all estimators including the reduced rank parameter β, despite the implied heavy tails of the process Y_t. A small Monte Carlo simulation study in Section 4.1 illustrates that the Gaussian approximation works well even for smaller, or moderate, samples such as $T=200,400$.

The empirically applied bivariate model with a self-exciting factor is given by,

$$\begin{aligned}
\Delta Y_t &= \mu+\alpha\beta' Y_{t-1}+\eta_t, \quad \eta_t=\Omega_t^{1/2} z_t \quad \text{and} &(18)\\
\Omega_t &= \Omega+\left[\alpha_\beta \phi_{\beta\beta}\beta' Y_{t-1}\right]^{\times 2}+\left[\alpha_\beta \phi_{\beta\alpha_\perp}\alpha_\perp' Y_{t-1}\right]^{\times 2}+\left[\beta_{\perp,\alpha_\perp}\phi_{\alpha_\perp\alpha_\perp}\alpha_\perp' Y_{t-1}\right]^{\times 2}, &(19)
\end{aligned}$$

where z_t is i.i.d. $\mathrm{N}(0, I_2)$. With $\alpha = (a, 0)'$ and $\beta = (1, b)'$, the parameters of the model, denoted θ, are given by $\mu = (\mu_1, \mu_2)'$, the scalar parameters a, b, $\phi^2_{\beta\beta}$, $\phi^2_{\beta\alpha_\perp}$, and $\phi^2_{\alpha_\perp\alpha_\perp}$, as well as the positive definite covariance matrix,

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} > 0,$$

Thus with the parameters given by,

$$\theta = \left\{a, b, \mu, \Omega, \phi^2_{\beta\beta}, \phi^2_{\beta\alpha_\perp}, \phi^2_{\alpha_\perp\alpha_\perp}\right\}, \tag{20}$$

the log-likelihood function to be maximized equals,

$$L(\theta) = \sum_{t=1}^{T} l_t(\theta) = -\tfrac{1}{2} \sum_{t=1}^{T} \left[\log|\Omega_{t,\theta}| + tr\left\{\Omega_{t,\theta}^{-1} \eta_{t,\theta} \eta_{t,\theta}'\right\}\right], \tag{21}$$

where the notation $\eta_{t,\theta} = (\Delta Y_t - \mu - \alpha\beta' Y_{t-1})$ and $\Omega_{t,\theta}$ is used to emphasize that η_t and Ω_t are functions of θ.

The result in Theorem 3 states that despite the lack of finite low order moments of Y_t, the ML estimator $\hat{\theta}$ is indeed asymptotically Gaussian distributed. The proof is given in Appendix B, and is based on deriving properties of the score, information and third order derivatives of the log-likelihood function in order to use Jensen and Rahbek (2004, Lemma 1). The findings are in line with known results for univariate models, such as in Jensen and Rahbek (2004) where it is shown that the (G)ARCH parameters are $\sqrt{T}$-consistent despite lack of finite moments of the GARCH process, as well as Ling (2004, 2007) for univariate DAR models. However, univariate results do not necessarily generalize to multivariate cases. Specifically, for the BEKK model, being the multivariate generalization of the univariate ARCH model, the findings in Avarucci, Beutner and Zaffaroni (2012) demonstrate that high order moments of the BEKK processes are needed in general for $\sqrt{T}$-Gaussian inference to apply. Our contrasting results below reflect that Avarucci et al. (2012) study the BEKK model with no conditional mean, and therefore our results show that by combining (a restricted) BEKK structure for the conditional variance and a reduced rank AR structure for the conditional mean, a different model, and hence inference, results despite the apparent similarities.

Theorem 3 *Consider the bivariate model in (18)-(19), with z_t i.i.d. $N(0, I_2)$, and parameters given by θ in (20). Then under Assumptions 1-2 and for θ_0 satisfying $\phi^2_{ij,0} > 0$ for $i, j = \beta, \alpha_\perp$, $\Omega_0 > 0$, and $|\Omega_0| > \psi > 0$ for some constant ψ, there exists a fixed open neighborhood $\mathcal{N}(\theta_0)$ of the true value θ_0 such that with probability tending to one as $T \to \infty$, $L(\theta)$ has a unique maximum at $\hat{\theta}$. Moreover, $\hat{\theta}$ is consistent and asymptotically Gaussian,*

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} N(0, \Sigma_{\theta\theta}),$$

with $\Sigma_{\theta\theta}$ consistently estimated by the observed information evaluated at $\theta = \hat{\theta}$.

Remark 7 *The regularity conditions assumed to hold include Assumptions 1 and 2, in addition to $\phi^2_{ij,0} > 0$ for $i,j = \beta, \alpha_\perp$, $\Omega_0 > 0$ and $|\Omega_0| > \psi > 0$ for some constant ψ. That $\phi^2_{ij,0} > 0$ holds corresponds to the assumption in classic ARCH literature of positive parameters, or equivalently, that ARCH effects are present, see for example Francq and Zakoian (2010). That Ω_0 is positive definite and satisfies that the determinant is bounded away from zero is a regularity condition applied, and also discussed in detail, in Comte and Lieberman (2003) as well as Jeantheau (1992), from where the regularity condition originates in the ARCH literature. It implies for example that $|\Omega_t| > \psi$, cf. Appendix B.*

4.1 Simulation Study

To illustrate the usefulness of the derived asymptotics for the estimators, we perform a small Monte Carlo simulation[1]. The data generating process is given by the model in (18)-(19), with parameters θ_0 given under Figure 1. The observations Y_t, for $t = 1, 2, ..., T$, are generated based on pseudo-random draws for z_t i.i.d. $\mathrm{N}(0, I_2)$, and sample lengths of $T = 200$ and $T = 400$ are considered.

For estimation we parametrize $\Omega > 0$ using a Choleski factorization $\Omega = CC'$, and report results for the parameters $c_{11} > 0$, $c_{12} \in \mathbb{R}$, $c_{22} > 0$ in the lower triangular C.

Figure 1 (A)-(I) report kernel estimates of the densities of the ten estimators based on 10^4 Monte Carlo replications. For a sample length of $T = 400$, the kernel densities are quite close to a Gaussian reference distributions with matching mean and variance, and the derived standard asymptotics seem to apply without problems.

For a small sample of $T = 200$, the Gaussian approximation is still useful for most parameters. A small deviation is visible for $\hat{\phi}_{\beta\beta}$, where there is a minor probability mass at zero. The Gaussian approximation is also less accurate for $\hat{\mu}_2$ and $\hat{c}_{22}$.

5 Empirical Application

To illustrate the use and interpretation of the model, consider a bivariate data set for monthly US interest rates, 1981:1-2006:12, covering the short end of the yield curve. The variables considered are the yield of a one year maturity zero coupon bond, Y_{1t}, and a three month treasury bill rate, Y_{2t}. The time series are illustrated in Figure 2 (A), while the spread, $Y_{1t} - Y_{2t}$, is given in Figure 2 (B). Figure 2 (C)-(D) show the first differences, where a pronounced heteroskedasticity is observed.

We estimate the model in (18)-(19) with z_t i.i.d. $\mathrm{N}(0, I_2)$. As in the simulations, we apply a Choleski factorization for $\Omega > 0$, and we thus estimate the parameters a, b, μ, $\Omega = CC'$ and ϕ_{ij}, $i, j = \beta, \alpha_\perp$. Estimates and standard errors are reported in column (i)

[1] Numerical calculations are done using Ox 6.30, see Doornik (2007).

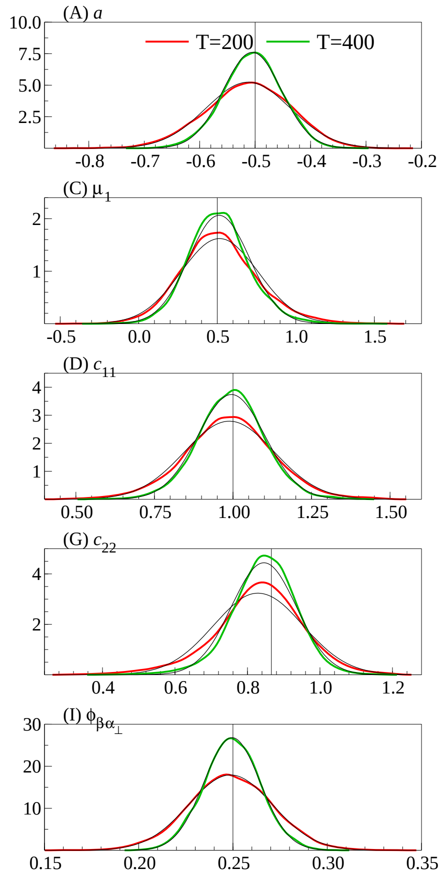

Figure 1: *Kernel density estimates of the simulated distributions of the estimated parameters. Simulations are based on $T = 200$ (red curve) and $T = 400$ (green curve) observations and 10^4 Monte Carlo replications. Thin black curves represent Gaussian distributions with matching mean and variance, and vertical lines indicate the true values in the data generating process. The parameters are given by $a = -0.5$, $b = 0$, $\mu_1 = \mu_2 = 0.2$, $\omega_{11} = 1$, $\omega_{12} = 0.5$, $\omega_{22} = 1$, $\phi_{\beta\beta} = 0.5$, and $\phi_{\beta\alpha_\perp} = \phi_{\alpha_\perp\alpha_\perp} = 0.25$.*

of Table 1.

First, we note that the estimate of b is extremely close to minus unity, suggesting that the conditional mean relationship, $\beta' Y_t$, is actually the interest rate spread. Testing the spread hypothesis, $b = -1$, produces Wald and likelihood ratio statistics of 0.019 and 0.030, that are far from significant in the limiting $\chi^2(1)$ distribution. Imposing the restriction produces the estimates in column (ii) of Table 1, and we note that estimates for the remaining parameters are basically unchanged. Below we continue the interpretation based on the restricted model in column (ii).

An immediate implication of the restricted model is that the interest rate levels, Y_{1t} and Y_{2t}, are unit root processes whereas the spread, $\beta' Y_t = Y_{1t} - Y_{2t}$, is not; the estimated

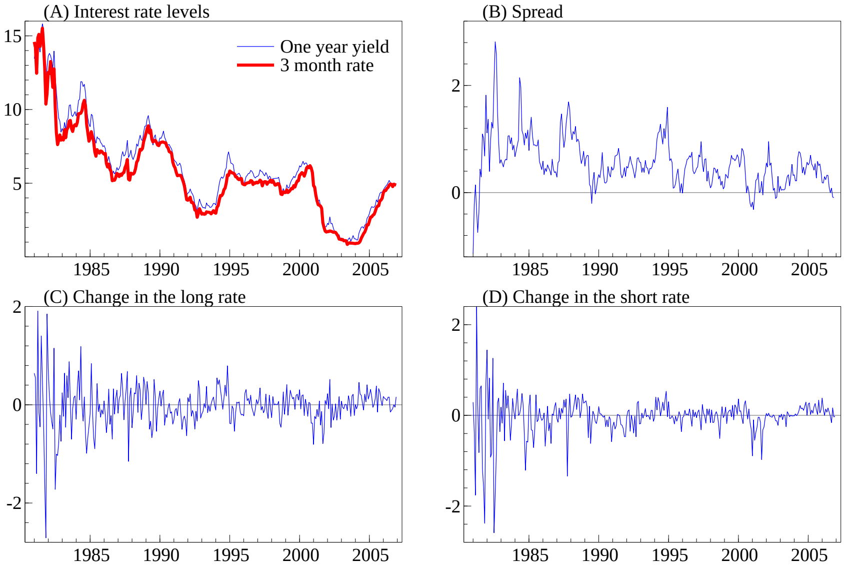

Figure 2: *Monthly US interest rates, 1981-2006.*

adjustment coefficient towards the conditional mean relationship, $\hat{a} = -0.16$, is clearly significant.

Next, note that the parameters in the conditional variance, $\phi_{\beta\beta}$, $\phi_{\beta\alpha_\perp}$, and $\phi_{\alpha_\perp\alpha_\perp}$ are all significantly different from zero. The self-exiting factor is the short rate, $\alpha'_\perp Y_t$, which is a unit root DAR. In contrast, the conditional variance of the spread, $\beta' Y_t$, is driven by the squared self-exciting short rate and the squared spread itself. Also recall that the model allows correlation between the short rate, $\alpha'_\perp Y_t$, and the spread, $\beta' Y_t$.

The estimated conditional variances of Y_{1t} and Y_{2t}, i.e. the diagonal elements in Ω_t, are shown in Figure 3 (A) and (B), and, not surprisingly, the patterns share similarities with the short rate. Figure 3 (C) and (D) show the estimated residuals, $\hat{\eta}_t$, while (E) and (F) present the standardized residuals, $\hat{z}_t$. Note again the pronounced conditional heteroskedasticity, which is to a large extend accounted for by the model.

To assess the stationarity and discuss the dynamic properties of the conditional mean relationship, $\beta' Y_t$ and the self-exiting factor, $\alpha'_\perp Y_t$, we calculate the Lyapunov exponents based on the random coefficient autoregressive representation of the bivariate system, and it is convenient to consider the system multiplied by $(\beta', \alpha'_\perp)'$. A direct application of Theorem 1 would lead to computing the top Lyapunov coefficient of the RCAR model with $Q_t = Q + \epsilon_t$, with $Q = \text{diag}\,(0.840, 1)$ and the variance of ϵ_t given by

$$E\left(\epsilon_t \otimes \epsilon_t\right) = \begin{pmatrix} 0.173^2 & 0 & 0 & 0.024^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.056^2 \end{pmatrix},$$

	(i) Baseline	(ii) Restricted
a	-0.1627 (0.0436)	-0.1600 (0.0348)
b	-1.0042 (0.0305)	-1
μ_1	0.0822 (0.0247)	0.0839 (0.0224)
μ_2	0.0030 (0.0128)	0.0030 (0.0128)
c_{11}	0.1409 (0.0191)	0.1412 (0.0191)
c_{12}	0.0364 (0.0223)	0.0365 (0.0223)
c_{22}	0.0677 (0.0182)	0.0676 (0.0181)
$\phi_{\beta\beta}$	0.1770 (0.0375)	0.1733 (0.0292)
$\phi_{\beta\alpha_\perp}$	0.0243 (0.0035)	0.0242 (0.0032)
$\phi_{\alpha_\perp\alpha_\perp}$	0.0563 (0.0044)	0.0561 (0.0045)
$L(\hat{\theta})$	-16.5021	-16.5169

Table 1: *Estimation results for the model in (18)-(19). Standard errors in parentheses.*

or alternatively, that $E\left(\mathrm{vec}\,(\epsilon_t)\,\mathrm{vec}\,(\epsilon_t)'\right) = \mathrm{diag}\,(0.173^2, 0, 0.024^2, 0.056^2)$. Direct computation of the matrix product in (13) is numerically instable because the Lyapunov stability condition implies that the matrix product converges to zero exponentially fast. Instead, we follow Dieci and Van Vleck (1995) and note that the two Lyapunov exponents, characterizing the dynamics of the system, can be found as the log of the eigenvalues of $\Lambda = \lim_{q\to\infty}((\prod_{t=1}^{q} Q_t)'(\prod_{t=1}^{q} Q_t))^{-2q}$. An efficient and numerically stable algorithm can be implemented using a sequential QR-decomposition of the matrix products, see Dieci and Van Vleck (1995, Section 2.1). For the estimated system, the two simulated Lyapunov exponents based on $q = 10^8$ random matrices, Q_t, $t = 1, 2, ..., q$, are given by

$$\hat{\lambda}_\varrho = \underset{(0.04218)}{-0.19728} \quad \text{and} \quad \hat{\lambda}_\gamma = \underset{(0.000257)}{-0.00158},$$

where numbers in parentheses are standard errors based on numerical delta method. By Theorem 2, the top Lyapunov coefficient $\hat{\gamma} = \max(\hat{\lambda}_\varrho, \hat{\lambda}_\gamma) = \hat{\lambda}_\gamma$ is associated with the self-exiting stochastic factor, $x_t = \alpha_\perp' Y_t$, while the smaller Lyapunov exponent, $\hat{\lambda}_\varrho = -0.19728$, is associated with the conditional mean relationship, $y_t = \beta' Y_t$. Both are significantly negative, implying stationarity and ergodicity of both series despite the presence of a unit root. In addition we note that $\hat{\lambda}_\varrho << \hat{\lambda}_\gamma$, reflecting that the non-unit root spread visually appears to be much more stable than the short rate.

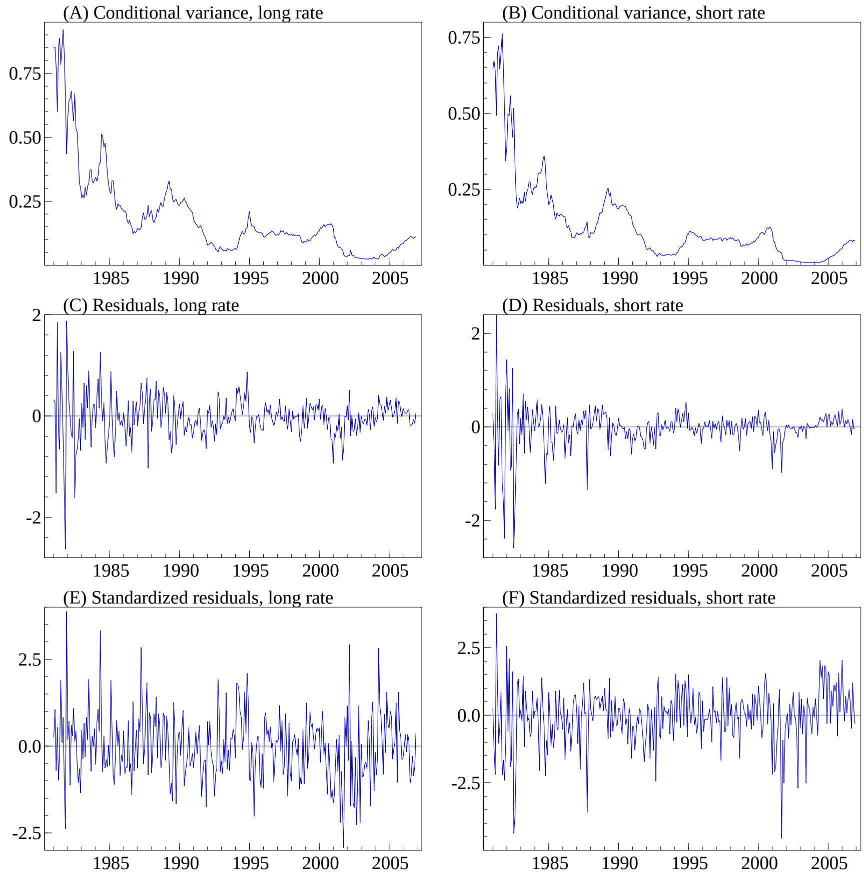

Figure 3: *Conditional variance and residuals.*

6 Extensions and Concluding Remarks

To summarize,we propose to study the multivariate model given by (3), $\Delta Y_t = \alpha\beta' Y_{t-1} + \mu + \Omega_t^{1/2} z_t$, with a structure imposed on Ω_t as in (8) or (9). We derive conditions for geometric ergodicity and establish that for the empirically applied bivariate model classic $\sqrt{T}$-asymptotics hold.

An immediate extension is to include more lags. One may consider the reparametrized p-dimensional vector autoregression of order m say with conditional heteroskedasticity,

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{i=1}^{m-1} \Gamma_i \Delta Y_{t-i} + \eta_t, \quad \eta_t = \Omega_t^{1/2} z_t \quad \text{and} \quad t = 1, ..., T \tag{22}$$

with z_t i.i.d. $\mathrm{N}(0, I_p)$ and α, β as before, while $(\Gamma_i)_{i=1,...,k-1}$ are $(p \times p)$ matrices. The conditional covariance Ω_t in the simplest case of the BEKK ARCH(m) is parametrized as a function of lagged levels Y_t and differences ΔY_t as,

$$\Omega_t = \Omega + \left(\phi_{yy} Y_{t-1}\right)^{\times 2} + \sum_{i=1}^{m-1} \left(\phi_{\delta y, i} \Delta Y_{t-i}\right)^{\times 2}, \tag{23}$$

where $\Omega > 0$, the parameters ϕ_{yy}, $\left(\phi_{\delta y,i}\right)_{1=1,\ldots,m-1}$ are $(p \times p)$ matrices and the initial values $Y_0, \Delta Y_0, \ldots, \Delta Y_{-m+1}$ are fixed in the statistical analysis of the model. The regularity condition in this case replacing Assumption 1 is given by replacing $A(z)$ by $A_m(z) = (1-z) I_p - \alpha\beta' z - \sum_{i=1}^{m-1} \Gamma_i (1-z) z^i$, see Johansen (1996). As to restricting further the model in line with the discussion of the Markovian case, while in principle straightforward in terms of parametrization, we refrain from this here as alone the results on geometric ergodicity at this stage can be generalized. Note finally, that it seems likely that due to the complications of such further added parameters, one should apply bootstrapping rather than classic asymptotic inference, as also used in Cavaliere, Rahbek and Taylor (2010, 2012) for co-integration analysis under heteroscedasticity.

Appendix

Throughout the appendix we use $\mathcal{C}$ and κ to denote generic positive constants.

A Geometric Ergodicity

Proof of Theorem 1: The process Y_t in (15) is a Markov chain, and the drift criterion from Markov chain theory, see Bec and Rahbek (2004), can be used to establish geometric ergodicity and stationarity of Y_t. Observe that the process Y_t in (12) can be rewritten as a random coefficient vector autoregression,

$$Y_t = \mu + Q_t Y_{t-1} + \xi_t,$$

where ξ_t is i.i.d. $\mathrm{N}(0, \Omega)$, with $Q_t := (I_p + \alpha\beta' + \epsilon_t)$ and where ϵ_t is a $(p \times p)$ $-$dimensional Gaussian i.i.d. mean zero and a covariance structure defined by $E(\epsilon_t \otimes \epsilon_t) = \sum_{i=1}^{h} (A_i \otimes A_i)$. Furthermore, ξ_t and ϵ_t are mutually independent. By Tjøstheim (1990) a drift function $d_y(\cdot)$ can be applied to the $k-$step process, Y_{kt} rather than Y_t itself, where

$$Y_{kt} = Q_{(kt,k)} Y_{k(t-1)} + \sum_{j=0}^{k-1} Q_{(kt,j)} (\epsilon_{tk-j} + \mu),$$

using the notation $A_{(t,j)} := A_t A_{t-1} \cdots A_{t-(j-1)}$ for any t-indexed square matrix A_t and $j \geq 1$, while $A_{(t,0)} := I$. An identical recursion is also used in Ling (2007, equation (A.9)), with the exception of the extra term here due to the constant vector μ here. With drift function, $d_y(y) = 1 + \|y\|^\delta$,where $\delta > 0$ is chosen appropriately below, it follows as in Ling (2007, (A.3)) that we can choose $\delta \in (0,1)$ and k such that $E\left\|Q_{(kt,k)}\right\|^\delta < 1$. This we use here to see that μ plays no role in the argument as we get,

$$E\left(d_y(Y_{kt}) \mid Y_{k(t-1)} = y\right) \leq E\left\|Q_{(kt,k)}\right\|^\delta \|y\|^\delta + \mathcal{C}.$$

Hence with M some constant, and for $d_y(y) > M$, we have

$$E\left\|Q_{(kt,k)}\right\|^\delta \|y\|^u + \mathcal{C} = \left[\frac{E\left\|Q_{(kt,k)}\right\|^\delta \|y\|^\delta + \mathcal{C}}{1 + \|y\|^\delta}\right] d_y(y) < \rho d_y(y),$$

where ρ is some constant with $|\rho| < 1$. Thus Y_t is stationary and geometrically ergodic, provided

$$\gamma := \lim_{q \to \infty} \left[\tfrac{1}{q} E \log \left\|\prod_{t=1}^{q} Q_t\right\|\right] < 0.$$

■

Proof of Theorem 2: As in the proof of Theorem 1, observe that the process Y_t in (15) is a Markov chain, and the drift criterion can be used to establish geometric ergodicity and stationarity of Y_t. We do so in three steps: First, we define Z_t as Y_t appropriately rotated, and rewrite Z_t as a random coefficient autoregression. Second, as before using Tjøstheim (1990), the $k-$step process, Z_{kt} rather than Z_t itself is used for the inspection of the drift criterion. Third, for the drift criterion a new drift function is applied to Z_{kt} which exploits the multivariate structure of Z_t, and hence Y_t. We set $\mu = 0$ without loss of generality as it plays no role in the derivations, see the proof of Theorem 1.

It follows immediately that the Markov chain $Z_t = (y_t', x_t')' = (\beta, \alpha_\perp)' Y_t$ has the same transition density as the random coefficient process given by,

$$Z_t = \Phi_t Z_{t-1} + \epsilon_t, \quad \text{where} \quad \Phi_t = \begin{pmatrix} \varrho_t & \xi_t \\ 0 & \gamma_t \end{pmatrix}. \tag{A.1}$$

Here $\epsilon_t = \left(\epsilon_{y,t}', \epsilon_{x,t}'\right)'$ is i.i.d. $\mathrm{N}_p\left(0, (\beta, \alpha_\perp)' \Omega (\beta, \alpha_\perp)\right)$ and independent of the i.i.d. Gaussian $(p \times p)$ matrix sequence Φ_t. The ϱ_t $(r \times r)$, ξ_t $(r \times p - r)$ and γ_t $(p - r \times p - r)$ are all independent and i.i.d. Gaussian, with ϱ_t and γ_t defined in Assumption 2, while $E\gamma_t = 0$ and $E(\gamma_t \otimes \gamma_t) = (I_{p-r} \otimes I_{p-r}) + \left(\phi_{\alpha_\perp \alpha_\perp} \otimes \phi_{\alpha_\perp \alpha_\perp}\right)$.

Now, Z_t satisfies the regularity conditions such that the drift criterion can be applied and we apply a new drift function to the $k-$step process, Z_{kt}. The new drift-function $d(\cdot)$ is given by,

$$d(z) = 1 + \|y\|^\delta + c\|x\|^\delta,$$

where δ and c are constants chosen appropriately below. More precisely, we show that with δ, c and M appropriately chosen constants, then for $d(z) > M$,

$$E\left(d(Z_{kt}) \mid Z_{k(t-1)} = z = (y', x')'\right) < \phi d(z),$$

where $\phi < 1$, while for $d(z) \leq M$ the conditional expectation is bounded.

Now from the definition of Z_t we have as in the proof of Theorem 1,

$$Z_{kt} = \Phi_{(kt,k)} Z_{k(t-1)} + \sum_{j=0}^{k-1} \Phi_{(kt,j)} \epsilon_{tk-j}.$$

Next, by definition of Φ_t in (A.1), for $j \geq 1$,

$$\Phi_{(kt,j)} = \begin{pmatrix} \varrho_{(tk,j)} & \xi^{(tk,j)} \\ 0 & \gamma_{(tk,j)} \end{pmatrix}, \quad \xi^{(t,j)} = \sum_{m=1}^{j} \varrho_{(t,m-1)} \xi_{t-(m-1)} \gamma_{(t-m,j-m)}$$

while $\Phi_{(t,0)} = I_p$, implies $\xi^{(t,0)} = 0$ $(r \times (p-r))$. We thus find, with $z = (y', x')$,

$$E\left(d(Z_{kt}) \mid Z_{k(t-1)} = z\right) = 1 + E\left(\|y_{kt}\|^\delta \mid Z_{k(t-1)} = z\right) + cE\left(\|x_{kt}\|^\delta \mid Z_{k(t-1)} = z\right),$$

where,

$$E\left(\|y_{kt}\|^{\delta} \mid Z_{k(t-1)} = z\right) = E\left\|\varrho_{(tk,k)}y + \xi^{(tk,k)}x + \sum_{j=0}^{k-1}\left(\varrho_{(kt,j)}\epsilon_{y,tk-j} + \xi^{(tk,j)}\epsilon_{x,tk-j}\right)\right\|^{\delta}$$

$$\leq E\left\|\varrho_{(tk,k)}\right\|^{\delta}\|y\|^{\delta} + E\left\|\xi^{(tk,k)}\right\|^{\delta}\|x\|^{\delta} + \mathcal{C}$$

$$E\left(\|x_{kt}\|^{\delta} \mid Z_{k(t-1)} = z\right) = E\left\|\gamma_{(tk,k)}x + \sum_{j=0}^{k-1}\gamma_{(tk,j)}\epsilon_{x,tk-j}\right\|^{\delta} \leq E\left\|\gamma_{(tk,k)}\right\|^{\delta}\|x\|^{\delta} + \mathcal{C}.$$

Collecting terms, we thus find

$$E\left(d\left(Z_{kt}\right) \mid Z_{k(t-1)} = z\right) \leq \left[\frac{\kappa + E\left\|\varrho_{(tk,k)}\right\|^{\delta}\|y\|^{\delta} + \left(E\left\|\xi^{(tk,k)}\right\|^{\delta} + cE\left\|\gamma_{(tk,k)}\right\|^{\delta}\right)\|x\|^{\delta}}{1 + c\|x\|^{\delta} + \|y\|^{\delta}}\right] d\left(z\right).$$

Since (17) are assumed to hold, then as established below, one can choose some small positive δ, $0 < \delta < 1$, and k large enough such that, $E\left\|\varrho_{(tk,k)}\right\|^{\delta} < 1$ and $E\left\|\gamma_{(tk,k)}\right\|^{\delta} < 1$. This again means that, $E\left\|\xi^{(tk,k)}\right\|^{\delta} + cE\left\|\gamma_{(tk,k)}\right\|^{\delta} < c$, provided c is chosen such that,

$$c > \frac{E\left\|\xi^{(tk,k)}\right\|^{\delta}}{1 - E\left\|\gamma_{(tk,k)}\right\|^{\delta}} > 0. \tag{A.2}$$

Hence for some $M > 0$, $d\left(z\right) > M$, then $E\left(d\left(Z_{kt}\right) \mid Z_{k(t-1)} = z\right) < \phi d\left(z\right)$, for some $\phi < 1$. It is simple to see that $E\left(d\left(Z_{kt}\right) | Z_{k(t-1)} = z\right)$ is bounded for $d\left(z\right) \leq M$ and the result hold as desired. That is, Z_{kt} and hence Z_t is geometrically ergodic, see Tjøstheim (1990), and hence Y_t is, since by definition $Y_t = \left(\beta_{\perp,\alpha_{\perp}}, \alpha_{\beta}\right) Z_t$ using (4).

Finally, we need to establish that $E\left\|\varrho_{(tk,k)}\right\|^{\delta} < 1$ and $E\left\|\gamma_{(tk,k)}\right\|^{\delta} < 1$ for some large k and small δ, provided (17) holds. This follows by standard arguments as in Ling (2007, proof of (A.3)), from which it holds that for some $\delta_{\varrho} < 1$ and k_{ϱ} large enough, $E\left\|\varrho_{(tk_{\varrho},k_{\varrho})}\right\|^{\delta_{\varrho}} < 1$. Likewise for $\gamma_{(tk,k)}$, and we can choose $\delta = \min\left(\delta_{\varrho}, \delta_{\gamma}\right)$ and $k = \max\left(k_{\varrho}, k_{\gamma}\right)$. ■

Proof of Corollary 1: With Y_t in (15)-(16) such that $\phi_{\beta\alpha_{\perp}} = 0$, then $\beta' Y_t$ is a Markov chain, which have a random coefficient representation, $\beta' Y_t = \varrho_t \beta' Y_{t-1} + \xi_{\beta,t}$, $\varrho_t = I_r + \beta'\alpha + \epsilon_t^{\varrho}$ with $\xi_{\beta,t}$ independent of $\varrho_{\beta,t}$, and $\xi_{\beta,t}$ i.i.d. mean zero Gaussian with variance $\beta'\Omega\beta$. Thus by e.g. Feigin and Tweedie (1985, Theorem 3), $\beta' Y_t$ is geometrically ergodic with finite second order moments as claimed. ■

B Asymptotics

Proof of Theorem 3: The result follows by establishing regularity conditions in Lemma 1 from Jensen and Rahbek (2004) for the score, information and third order derivatives of the log-likelihood function respectively. Specifically, Lemma B.1 below establishes condition (A.1) of Jensen and Rahbek (2004, Lemma 1), Lemma B.2 condition (A.2) and finally Lemma B.3 condition (A.3).

The derivations are notationally quite involved, and we start by defining some key variables and expressions. Also to simplify we leave μ out of θ. Thus the parameters in θ, of dimension $p_\theta = 8$, are given by $\theta = (a, b, \tilde{\theta}')'$ where $\tilde{\theta} = (\theta_{12}, \theta_{11}', \theta_{22}')'$, and

$$\theta_{12} = \omega_{12}, \quad \theta_{11} = \left(\omega_{11}, \phi_{\beta\beta}^2, \phi_{\beta\alpha_\perp}^2\right)' \quad \text{and} \quad \theta_{22} = \left(\omega_{22}, \phi_{\alpha_\perp\alpha_\perp}^2\right)'. \tag{B.1}$$

With $y_t(b) := \beta' Y_t$, define also the variables,

$$\lambda_{12,t} = 1, \quad \lambda_{11,t} = \left(1, y_{t-1}^2(b), Y_{2t-1}^2\right)' \quad \text{and} \quad \lambda_{22,t} = \left(1, Y_{2t-1}^2\right)'. \tag{B.2}$$

It will be useful to suppress the dependence on the parameter θ sometimes. We use in particular $y_{t-1} := y_{t-1}(b_0)$ and $\Omega_t := \Omega_{t,\theta_0}$, that is omit θ when quantities are evaluated at $\theta = \theta_0$. When the distinction between θ and θ_0 is important, as for example when providing the uniform bounds for the third derivatives in Lemma B.3, we emphasize this in the arguments.

For any (matrix) function of θ, $f(\theta)$, $df(\theta, d\theta)$ denotes the differential of f in the direction $d\theta$. To save space we let for example $\dot{f}_b$ denote $df(\theta, db)$, that is, the differential in the direction db.

We repeatedly use the definition of $\Omega_{t,\theta}$ and its inverse. Recall that by definition $\Omega_{t,\theta}$ is given by,

$$\Omega_{t,\theta} = \begin{pmatrix} \omega_{t,11} & \omega_{t,12} \\ \omega_{t,21} & \omega_{t,22} \end{pmatrix} = \Omega + S_{11}\left[\phi_{\beta\beta}^2 y_{t-1}^2(b) + \phi_{\beta\alpha_\perp}^2 Y_{2t-1}^2\right] + S_{bb}\phi_{\alpha_\perp\alpha_\perp}^2 Y_{2t-1}^2, \tag{B.3}$$

where we have introduced (two of) the following selection matrices,

$$S_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \; S_{12} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; S_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad S_{bb} = \begin{pmatrix} b^2 & -b \\ -b & 1 \end{pmatrix}. \tag{B.4}$$

Introduce next the notation for the inverse,

$$\Omega_{t,\theta}^{-1} = \begin{pmatrix} \omega_{11}^t & \omega_{12}^t \\ \omega_{21}^t & \omega_{22}^t \end{pmatrix} = \Omega_t^* \left|\Omega_t\right|^{-1} \quad \text{where } \Omega_t^* = \begin{pmatrix} \omega_{t,11} & -\omega_{21} \\ -\omega_{12} & \omega_{t,22} \end{pmatrix}, \tag{B.5}$$

such that $\left(\Omega_{t,\theta}^{-1}\right)_{ij} = \omega_{ij}^t$ and $\left(\Omega_{t,\theta}\right)_{ij} = \omega_{t,ij}$ for $i,j = 1,2$. Moreover, the determinant of $\Omega_{t,\theta}$ we write as,

$$\delta_{(t,\theta)} := \left|\Omega_{t,\theta}\right| = \delta_1 + \delta_2 y_{t-1}^2(b) + \delta_3 Y_{2t-1}^2 + \delta_4 Y_{2t-1}^4 + \delta_5 y_{t-1}^2(b) Y_{2t-1}^2 \tag{B.6}$$

with $\delta_1 = \omega_{11}\omega_{22} - \omega_{12}^2$, $\delta_2 = \omega_{22}\phi_{\beta\beta}^2$, $\delta_3 = \left(\omega_{11}\phi_{\alpha_\perp\alpha_\perp}^2 + \omega_{22}\phi_{\beta\alpha_\perp}^2\right)$, $\delta_4 = \phi_{\beta\alpha_\perp}^2\phi_{\alpha_\perp\alpha_\perp}^2$, $\delta_5 = \phi_{\beta\beta}^2\phi_{\alpha_\perp\alpha_\perp}^2$. Note that by assumption $\delta_{(t)} \geq \delta_1 \geq \psi > 0$.

Score:

For the score we have the following result:

Lemma B.1 *The first order differential of the log-likelihood function is given by,*

$$dl_t(\theta, d\theta) = \tfrac{1}{2} tr\left\{\left(\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}' - I\right)\Omega_{t,\theta}^{-1} d\Omega_{t,\theta}\right\} - tr\left\{\Omega_t^{-1}\eta_{t,\theta} d\eta_{t,\theta}\right\}, \tag{B.7}$$

with $\theta = (a, b, \tilde{\theta}')'$. Under the assumptions of Theorem 3 it holds that the score is asymptotically Gaussian distributed,

$$\frac{1}{\sqrt{T}}\, \partial L(\theta)/\partial\theta\big|_{\theta=\theta_0} \xrightarrow{D} N(0, \Sigma_{\theta\theta}) \quad as \quad T \to \infty. \tag{B.8}$$

Proof of Lemma B.1: The log-likelihood function is given in (21), such that with $\dot{\Omega}_{t,\theta}$ denoting the differential of $\Omega_{t,\theta}$ in the direction $d\theta$, and similarly for $\dot{\eta}_{t,\theta}$, where $\eta_{t,\theta} := \Delta Y_t - \alpha\beta' Y_{t-1} = \Delta Y_t - (a, 0)' y_{t-1}(b)$,

$$\begin{aligned} dL(\theta, d\theta) &= \sum_{t=1}^{T} dl_t(\theta, d\theta) && \text{(B.9)} \\ &= -\tfrac{1}{2}\sum_{t=1}^{T}\left[tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,\theta}\right\} - tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,\theta}\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}'\right\} + 2tr\left\{\Omega_{t,\theta}^{-1}\eta_{t,\theta}\dot{\eta}_{t,\theta}'\right\}\right]. \end{aligned}$$

Part S1: $\partial l_t(\theta)/\partial a$: Standard calculus gives, $\dot{\Omega}_{t,a} = 0$, $\dot{\eta}_{t,a} = -y_{t-1}(b)(1,0)' da$ and hence using (B.9),

$$\partial l_t(\theta)/\partial a\big|_{\theta=\theta_0} = tr\left\{\Omega_{t,\theta}^{-1}\eta_{t,\theta} y_{t-1}(b)(1,0)\right\}\big|_{\theta=\theta_0} = tr\left\{\Omega_t^{-1/2} z_t (1,0)\right\} y_{t-1}, \tag{B.10}$$

which is a martingale difference (MGD) sequence with respect to the filtration $\mathcal{F}_t$ generated by $(\eta_{t-s})_{s=0,\ldots,t-1}$ and Y_0. We find directly,

$$E\left(\left[\partial l_t(\theta)/\partial a|_{\theta=\theta_0}\right]^2 | \mathcal{F}_{t-1}\right) = y_{t-1}^2\omega_{11}^t = y_{t-1}^2\omega_{t,22}/\delta_{(t)}. \tag{B.11}$$

Now using (B.6) we can conclude $E\left(\partial l_t(\theta)/\partial a|_{\theta=\theta_0}\right)^2 < \infty$ since,

$$E\left(\left[\partial l_t(\theta)/\partial a|_{\theta=\theta_0}\right]^2 | \mathcal{F}_{t-1}\right) \leq \frac{y_{t-1}^2\left(\omega_{22} + \phi_{\alpha_\perp\alpha_\perp}^2 Y_{2t-1}^2\right)}{\left[\delta_2 y_{t-1}^2 + \delta_3 Y_{2t-1}^2 + \delta_4 Y_{2t-1}^4 + \delta_5 y_{t-1}^2 Y_{2t-1}^2\right]} \leq \mathcal{C}, \tag{B.12}$$

where we here, and henceforth in the proof, have omitted the subindex "0" on the true parameters.

Part S2: $\partial l_t(\theta)/\partial\tilde{\theta}$: Observe that $\dot{\eta}_{t,\tilde{\theta}} = 0$, and hence by (B.9),

$$dl_t\left(\theta, d\tilde{\theta}\right)\Big|_{\theta=\theta_0} = \tfrac{1}{2}tr\left\{\Omega_t^{-1/2}\left(z_t z_t' - I\right)\Omega_t^{-1/2}\dot{\Omega}_{t,\tilde{\theta}}\right\}. \tag{B.13}$$

As before this is a MGD, and using the identity in Bec, Rahbek and Shephard (2008, equation (48)), we find initially,

$$E\left(\left[dl_t\left(\theta, d\tilde{\theta}\right)\Big|_{\theta=\theta_0}\right]^2 |\mathcal{F}_{t-1}\right) = \tfrac{1}{2}tr\left\{\left[\Omega_t^{-1}\dot{\Omega}_{t,\tilde{\theta}}\right]^2\right\}. \tag{B.14}$$

We consider in turn each term in $\tilde{\theta} = (\theta_{12}, \theta_{11}', \theta_{22}')'$:

Part S2.1: $\partial l_t(\theta)/\partial\theta_{12}$: Using (B.14) and that by definition of $\Omega_{t,\theta}$, $\dot{\Omega}_{t,\theta_{12}} = S_{12}d\theta_{12}$ where S_{12} is defined in (B.4):

$$\begin{aligned} 2E\left(\left[\partial l_t(\theta)/\partial\theta_{12}|_{\theta=\theta_0}\right]^2 |\mathcal{F}_{t-1}\right) &= tr\left\{\left[\Omega_t^{-1}S_{12}\right]^2\right\} = tr\left\{\left[\Omega_t^{*}S_{12}\right]^2\right\}/\delta_{(t)}^2 \\ &= \left(\delta_{(t)} + 2\omega_{12}^2\right)/\delta_{(t)}^2 \leq \left(1 + 2\omega_{12}^2\right)/\psi \leq \mathcal{C}, \end{aligned} \tag{B.15}$$

where we have used the definition of Ω_t^{-1}, see (B.5), and $\delta_{(t)} \geq \delta_1 \geq \psi > 0$.

Part S2.2: $\partial l_t(\theta)/\partial\theta_{11}$: Using (B.13), we find

$$\left(\partial l_t(\theta)/\partial\theta_{11}\right)|_{\theta=\theta_0} = \tfrac{1}{2}tr\left\{\left(z_t z_t' - I\right)\Omega_t^{-1/2}S_{11}\Omega_t^{-1/2}\right\}\lambda_{11,t}, \tag{B.16}$$

and, using (B.14),

$$\begin{aligned} 2E\left(\left[\left(\partial l_t(\theta)/\partial\theta_{yy}\right)\left(\partial l_t(\theta)/\partial\theta_{yy}'\right)\right]|\mathcal{F}_{t-1}\right) &= tr\left\{\left[\Omega_t^{*}S_{11}\right]^2\right\}\lambda_{11,t}\lambda_{11,t}'/\delta_{(t)}^2 \\ &= \omega_{t,22}^2\lambda_{11,t}\lambda_{11,t}'/\delta_{(t)}^2. \end{aligned}$$

Therefore $\left\|E\left(\left[\left(\partial l_t(\theta)/\partial\theta_{yy}\right)\left(\partial l_t(\theta)/\partial\theta_{yy}'\right)\right]|\mathcal{F}_{t-1}\right)\right\| \leq \mathcal{C}$, using in particular that,

$$\omega_{t,22}/\delta_{(t)} \leq \omega_{22}/\delta_1 + 1/\omega_{11}, \quad \omega_{t,22}y_{t-1}^2/\delta_{(t)} \leq 2/\phi_{\beta\beta}^2 \quad \text{and} \quad \omega_{t,22}Y_{2t-1}^2/\delta_{(t)} \leq 2/\phi_{\beta\alpha_\perp}^2. \tag{B.17}$$

Part S2.3: $\partial l_t(\theta)/\partial\theta_{22}$: Using (B.13),

$$\left(\partial l_t(\theta)/\partial\theta_{22}\right)|_{\theta=\theta_0} = \tfrac{1}{2}tr\left\{\left(z_t z_t' - I\right)\Omega_t^{-1/2}S_{22}\Omega_t^{-1/2}\right\}\lambda_{22,t}. \tag{B.18}$$

As before, using (B.14),

$$E\left(\left[\left(\partial l_t(\theta)/\partial\theta_{22}\right)\left(\partial l_t(\theta)/\partial\theta_{22}'\right)\right]^2 |\mathcal{F}_{t-1}\right) = \omega_{t,11}^2\lambda_{22,t}\lambda_{22,t}'/\delta_{(t)}^2$$

and hence,

$$\left\| E\left(\left[\left(\partial l_t\left(\theta\right) / \partial \theta_{22} \right) \left(\partial l_t\left(\theta\right) / \partial \theta_{22}' \right) \right]^2 | \mathcal{F}_{t-1} \right) \right\| \leq \mathcal{C} \tag{B.19}$$

using, similar to before,

$$\omega_{t,11}/\delta_{(t)} \leq \omega_{11}/\delta_1 + 2/\omega_{22} \quad \text{and} \quad \omega_{t,11} Y_{2t-1}^2/\delta_{(t)} \leq 3/\phi_{\alpha_\perp \alpha_\perp}^2. \tag{B.20}$$

Part S3: $\partial l_t\left(\theta\right)/\partial b$ term: By definition $\dot{\eta}_{t,b} = -\left(a, 0\right)' Y_{2t-1} db$ and

$$\dot{\Omega}_{t,b} = \begin{pmatrix} 2\left(b\phi_{\alpha_\perp \alpha_\perp}^2 Y_{2t-1}^2 + \phi_{\beta\beta}^2 y_{t-1}\left(b\right) Y_{2t-1}\right) & -\phi_{\alpha_\perp \alpha_\perp}^2 Y_{2t-1}^2 \\ -\phi_{\alpha_\perp \alpha_\perp}^2 Y_{2t-1}^2 & 0 \end{pmatrix} db. \tag{B.21}$$

Thus,

$$dl_t\left(\theta, db\right)|_{\theta=\theta_0} = \tfrac{1}{2}\left[tr\left\{ \left(z_t z_t' - I\right) \Omega_t^{-1/2} \dot{\Omega}_{t,b} \Omega_t^{-1/2} \right\} - 2tr\left\{ \Omega_t^{-1/2} z_t \dot{\eta}_{t,b}' \right\} \right], \tag{B.22}$$

is a Martingale difference sequence, and we find

$$E\left(\left[dl_t\left(\theta, db\right)|_{\theta=\theta_0} \right]^2 | \mathcal{F}_{t-1} \right) = \tfrac{1}{2} tr\left\{ \left[\Omega_t^{-1} \dot{\Omega}_{t,b} \right]^2 \right\} + \dot{\eta}_{t,b}' \Omega_t^{-1} \dot{\eta}_{t,b}. \tag{B.23}$$

Next, using previously applied bounds, one immediately finds

$$\dot{\eta}_{t,b}' \Omega_t^{-1} \dot{\eta}_{t,b} = a^2 Y_{2t-1}^2 \omega_{t,22}/\delta_{(t)} \leq \mathcal{C}. \tag{B.24}$$

With respect to $tr\left\{ \left[\Omega_t^{-1} \dot{\Omega}_{t,b} \right]^2 \right\} = tr\left\{ \left[\Omega_t^* \dot{\Omega}_{t,b} \right]^2 \right\} / \delta_{(t)}^2$, quite lengthy and tedious calculations show that, with c_i functions of the (true) parameters θ_0,

$$\begin{aligned} tr\left\{ \left[\Omega_t^* \dot{\Omega}_{t,b} \right]^2 \right\} &= \left(c_1 + c_2 Y_{2t-1}^2 + c_3 Y_{2t-1}^4\right) Y_{2t-1}^4 \\ &\quad + \left(Y_{2t-1}^2 y_{t-1}^2\left(b\right)\right)\left(c_4 + c_5 Y_{2t-1}^2 + c_6 Y_{2t-1}^4\right) \\ &\quad + \left(Y_{2t-1}^3 y_{t-1}\left(b\right)\right)\left(c_7 + c_8 Y_{2t-1}^2 + c_9 Y_{2t-1}^4\right). \end{aligned}$$

Using the expression for $\delta_{(t)}$ in (B.6), $tr\left\{ \left[\Omega_t^* \dot{\Omega}_{t,b} \right]^2 \right\} / \delta_{(t)}^2 \leq \mathcal{C}$ at $\theta = \theta_0$. As an example consider a cross-product term with Y_{2t-1} of high power, $c_9 Y_{2t-1}^7 y_{t-1}$, for which,

$$\left| c_9 Y_{2t-1}^7 y_{t-1} \right| / \delta_{(t)}^2 \leq \frac{\left| c_9 Y_{2t-1}^7 y_{t-1} \right|}{\delta_1^2 + \delta_4^2 Y_{2t-1}^8 + 2\delta_4 \delta_5 y_{t-1}^2 Y_{2t-1}^6} \leq \mathcal{C},$$

using $\delta_i > 0$.

Part S4: Cross-terms: Collecting terms, and using the inequalities in (B.20) and (B.17), we conclude that

$$\left\| E\left(\left(\partial l_t\left(\theta\right)/\partial \theta_{22} \right) \left(\partial l_t\left(\theta\right)/\partial \theta_{11}' \right)|_{\theta=\theta_0} | \mathcal{F}_{t-1} \right) \right\| \leq \mathcal{C}. \tag{B.25}$$

We have here used (B.16) and (B.18), as well as the identity Bec et al (2008, equation 48) to see that

$$E\left(\left(\partial l_t(\theta)/\partial\theta_{22}\right)\left(\partial l_t(\theta)/\partial\theta'_{11}\right)|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right)=\tfrac{1}{2}\left(\omega^t_{12}\right)^2\lambda_{22,t}\lambda'_{11,t}=\tfrac{1}{2}\omega^2_{12}\lambda_{22,t}\lambda'_{11,t}/\delta^2_{(t)}.$$

Next, we find

$$\begin{aligned}
&\left\|E\left(\left[\left(\partial l_t(\theta)/\partial\theta_{22}\right)\left(\partial l_t(\theta)/\partial\theta_{12}\right)|_{\theta=\theta_0}\right]|\mathcal{F}_{t-1}\right)\right\|=\left\|\tfrac{1}{2}\left(\omega^t_{12}\right)\left(\omega^t_{22}\right)\lambda_{22,t}\lambda_{12,t}\right\|\\
&=\left\|\tfrac{1}{2}\omega_{12}\omega_{t,11}\lambda_{22,t}\right\|/\delta^2_{(t)}\leq\mathcal{C},\text{ and}\\
&\left\|E\left(\left[\left(\partial l_t(\theta)/\partial\theta_{11}\right)\left(\partial l_t(\theta)/\partial\theta_{12}\right)|_{\theta=\theta_0}\right]|\mathcal{F}_{t-1}\right)\right\|=\left\|\tfrac{1}{2}\left(\omega^t_{12}\right)\left(\omega^t_{11}\right)\lambda_{11,t}\lambda_{12,t}\right\|\\
&=\left\|\tfrac{1}{2}\omega_{12}\omega_{t,22}\lambda_{11,t}\right\|/\delta^2_{(t)}\leq\mathcal{C}.
\end{aligned}$$

Observe furthermore, by definition, $E\left(\left(\partial l_t(\theta)/\partial\tilde{\theta}\right)\left(\partial l_t(\theta)/\partial a\right)\Big|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right)=0$. Next consider the directions da and db, and use (B.10) and (B.22), to see that

$$\left|E\left(\left(\partial l_t(\theta)/\partial a\right)\left(\partial l_t(\theta)/\partial b\right)|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right)\right|\leq\frac{\left(c_1+c_2Y^2_{2t-1}\right)|y_{t-1}||Y_{2t-1}|}{\delta_{(t)}}\leq\mathcal{C},$$

with c_1, c_2 functions of θ_0. Finally, also $\left\|E\left(\left(\partial l_t(\theta)/\partial\tilde{\theta}\right)\left(\partial l_t(\theta)/\partial b\right)\Big|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right)\right\|\leq\mathcal{C}$, as

$$E\left(dl_t\left(\theta,\tilde{\theta}\right)dl_t(\theta,db)\Big|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right)=\tfrac{1}{2}tr\left\{\left[\Omega^*_t\dot{\Omega}_{t,\tilde{\theta}}\right]\left[\Omega^*_t\dot{\Omega}_{t,b}\right]\right\}/\delta^2_{(t)}\leq\mathcal{C},$$

computing the trace of the product, and using the bounds implied by $\delta^2_{(t)}$, as were applied for each case of the $\tilde{\theta}$ parameters (in *Part 2*) and for b (in *Part 3*).

Part S5: Application of CLT: With $\theta=(a,b,\tilde{\theta}')'$, we have shown that $\partial L(\theta)/\partial\theta$ is a martingale difference (MGD) sequence with respect to the filtration $\mathcal{F}_t$ generated by $\left(\eta_{t-s}\right)_{s=0,\ldots,t-1}$ and Y_0. Moreover, as $\left\|E\left(\left(\partial l_t(\theta)/\partial\theta\right)\left(\partial l_t(\theta)/\partial\theta'\right)|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right)\right\|<\mathcal{C}$, and by geometric ergodicity, the regularity conditions of Brown (1971) apply, by the law of large numbers in Jensen and Rahbek (2007). Hence, $\frac{1}{\sqrt{T}}\partial L(\theta)/\partial\theta|_{\theta=\theta_0}\xrightarrow{D}N(0,\Sigma_{\theta\theta})$, as claimed. ■

Information:

Lemma B.2 *For the observed information it holds under the assumptions of Theorem 3 that, as $T\to\infty$,*

$$-\tfrac{1}{T}\partial^2L(\theta)/\partial\theta\partial\theta'\Big|_{\theta=\theta_0}\xrightarrow{P}\Sigma_{\theta\theta}. \tag{B.26}$$

Proof of Lemma B.2: We show below all terms in $d^2l_t(\theta,d\theta,d\theta)$ are bounded and hence that the law of large numbers in Jensen and Rahbek (2007) can be applied.

Part I1: $\partial^2 l_t(\theta)/\partial a^2$: From the proof of Lemma B.1, $\partial l_t(\theta)/\partial a = tr\left\{\Omega_{t,\theta}^{-1}\eta_{t,\theta}\left(y_{t-1}(b),0\right)\right\}$, and hence

$$\partial^2 l_t(\theta)/\partial a^2 = -tr\left\{\Omega_{t,\theta}^{-1}S_{11}\right\}y_{t-1}^2(b), \tag{B.27}$$

with $E\left\|\partial^2 l_t(\theta)/\partial a^2|_{\theta=\theta_0}\right\| = E\left(\omega_{yy}^t y_{t-1}^2\right) \leq \mathcal{C}$ by (B.12). Thus indeed, geometric ergodicity implies that the law of large numbers in Jensen and Rahbek (2007) applies to $-\frac{1}{T}\partial^2 L(\theta)/\partial a^2\big|_{\theta=\theta_0}$.

Part I2: $\partial^2 l_t(\theta)/\partial\tilde{\theta}\partial\tilde{\theta}'$: With $\tilde{\theta}^* \neq \tilde{\theta}$, then as $dl_t\left(\theta, d\tilde{\theta}\right) = \frac{1}{2}tr\left\{\left(\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}' - I\right)\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,\tilde{\theta}}\right\}$, we get,

$$-2d^2 l_t\left(\theta, d\tilde{\theta}, d\tilde{\theta}^*\right) = tr\left\{\left[\Omega_t^{-1}\dot{\Omega}_{t,\tilde{\theta}}\Omega_t^{-1}\dot{\Omega}_{t,\tilde{\theta}^*}\right]\left(2\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}' - I\right)\right\}, \tag{B.28}$$

noting that the second order differential $\ddot{\Omega}_{t,\tilde{\theta},\tilde{\theta}^*} := d\Omega_{t,\theta}\left(d\tilde{\theta}, d\tilde{\theta}^*\right) = 0$. We have

$$\partial^2 l_t(\theta)/\partial\theta_{ij}\partial\theta_{kl}' = -\tfrac{1}{2}tr\left\{\left[\Omega_{t,\theta}^{-1}S_{ij}\right]\left[\Omega_{t,\theta}^{-1}S_{kl}\right]\left(2\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}' - I\right)\right\}\lambda_{ij,t}\lambda_{kl,t}' \tag{B.29}$$

for $i,j,k,l = 1,2$. Using $\eta_t = \eta_{t,\theta_0} = \Omega_t^{1/2}z_t$ and $|tr\{AB\}| \leq \|A\|\|B\|$, the second derivatives at θ_0 are bounded by,

$$\kappa\left[\|z_t\|^2 + 1\right]\left\|\Omega_t^{-1}S_{ij}\right\|\left\|\Omega_t^{-1}S_{kl}\right\|\|\lambda_{ij,t}\|\|\lambda_{kl,t}\|,$$

where κ is some constant and hence the term will have finite expectation provided

$$\left\|\Omega_t^{-1}S_{ij}\right\|\left\|\Omega_t^{-1}S_{kl}\right\|\|\lambda_{ij,t}\|\|\lambda_{kl,t}\| \leq \mathcal{C}, \tag{B.30}$$

as $E\|z_t\|^2 < \infty$. As $\Omega_t^{-1} = |\Omega_t|^{-1}\Omega_t^*$,

$$\left\|\Omega_t^{-1}S_{ij}\right\|\|\lambda_{ij,t}\| = \|\Omega_t^* S_{ij}\|\|\lambda_{ij,t}\|/\delta_{(t)} =: \rho_t(ij).$$

Now if $i = 1, j = 2$,

$$\rho_t^2(12) \leq \left(\omega_{t,22}^2 + \omega_{t,11}^2 + 2\omega_{12}^2\right)/\delta_{(t)}^2 \leq \mathcal{C},$$

using repeatedly the inequalities in (B.20) and (B.17). Likewise,

$$\rho_t^2(22) \leq \kappa\left(\omega_{t,11}^2 + \omega_{12}^2\right)\left(1 + Y_{2t-1}^2\right)/\delta_{(t)}^2 \leq \mathcal{C}, \quad \text{and}$$
$$\rho_t^2(11) \leq \kappa\left(\omega_{t,22}^2 + \omega_{12}^2\right)\left(1 + Y_{2t-1}^2 + y_{t-1}^2\right)/\delta_{(t)}^2 \leq \mathcal{C}.$$

Thus all terms in (B.30) are finite as desired.

Part I3: $\partial^2 l_t(\theta)/\partial b^2$: From the proof of Lemma B.1 (*Part S3*),

$$-2dl_t(\theta, db) = tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\left(I - \Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}'\right)\right\} + 2tr\left\{\Omega_{t,\theta}^{-1}\eta_{t,\theta}\dot{\eta}_{t,b}'\right\}, \tag{B.31}$$

and we find with $\ddot{\Omega}_{t,b,b^*} = d\Omega_{t,\theta}(db, db^*)$, $\ddot{\eta}_{t,b,b^*} := d\eta_{t,\theta}(db, db^*) = 0$,

$$\begin{aligned}
&-2d^2 l_t(\theta, db, db^*) \\
= \; & \underbrace{tr\left\{\Omega_{t,\theta}^{-1}\ddot{\Omega}_{t,b,b^*}\left(I - \Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}'\right)\right\}}_{(a)} + 2\underbrace{tr\left\{\left[\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\right]\left(\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b^*}\right)\left[\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}'\right]\right\}}_{(b1)} \\
& \underbrace{-tr\left\{\left(\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\right)\left[\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b^*}\right]\right\}}_{(b2)} + 2\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\eta}_{t,b^*}\dot{\eta}_{t,b}'\right\}}_{(c)} - 4\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b^*}\Omega_{t,\theta}^{-1}\eta_{t,\theta}\dot{\eta}_{t,b}'\right\}}_{(d)}.
\end{aligned}$$

Consider first (a), which at $\theta = \theta_0$ up to constants equals,

$$tr\left\{\Omega_t^{-1/2}S_{11}\Omega_t^{-1/2}\left(I - z_t z_t'\right)\right\}Y_{2t-1}^2 2\left(\phi_{\alpha_\perp\alpha_\perp}^2 + \phi_{\beta\beta}^2\right)dbdb^*. \tag{B.32}$$

As in *Part I2*, using $|tr\{AB\}| \leq \|A\|\,\|B\|$ and with κ some constant, this is bounded by,

$$\kappa\left(1 + \|z_t\|^2\right)Y_{2t-1}^2\,\|\Omega_t^* S_{11}\|\,/\delta_{(t)}, \tag{B.33}$$

which has finite expectation as $E\,\|z_t\|^2 < \infty$ and

$$Y_{2t-1}^2\,\|\Omega_t^* S_{11}\|\,/\delta_{(t)} \leq \kappa Y_{2t-1}^2\left(1 + Y_{2t-1}^2\right)/\delta_{(t)} \leq \mathcal{C}.$$

Likewise the terms in $(b1)$ and $(b2)$ have finite expectations as $\left\|\Omega_t^*\dot{\Omega}_{t,b}\right\|/\delta_{(t)} \leq \mathcal{C}$, see *Part S3* above in the proof of Lemma B.1. The term in (c) has finite expectation as it is bounded by $\kappa Y_{2t-1}^2\omega_{t,22}/\delta_{(t)} \leq \mathcal{C}$. Finally, the absolute value of (d) is bounded by $\kappa\,\|z_t\|$ as $\left\|\Omega_t^{-1/2}\right\| \leq \mathcal{C}$ and as just applied, $\left\|\Omega_t^*\dot{\Omega}_{t,b}\right\|/\delta_{(t)} \leq \mathcal{C}$.

Part I4: $\partial^2 l_t(\theta)/\partial\tilde{\theta}\partial a$: As $dl_t\left(\theta, d\tilde{\theta}\right) = \frac{1}{2}tr\left\{\left(\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}' - I\right)\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,\tilde{\theta}}\right\}$, then

$$d^2 l_t\left(\theta, d\tilde{\theta}, da\right) = tr\left\{\Omega_{t,\theta}^{-1}\eta_{t,\theta}\left(y_{t-1}^2(b), 0\right)\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,\tilde{\theta}}\right\}da$$

and at $\theta = \theta_0$,

$$\partial^2 l_t(\theta)/\partial\theta_{ij}\partial a\Big|_{\theta=\theta_0} = tr\left\{\Omega_t^{-1/2}z_t\left(y_{t-1}^2, 0\right)\Omega_t^{-1}S_{ij}\right\}\lambda_{ij,t} \tag{B.34}$$

for $i, j = 1, 2$, which is bounded (and moreover all terms have expectation zero).

Part I5: $\partial^2 l_t(\theta)/\partial a\partial b$: Now $dl_t(\theta, da) = tr\left\{\Omega_{t,\theta}^{-1}\eta_{t,\theta}(1, 0)\right\}y_{t-1}(b)\,da$, such that

$$\begin{aligned}
d^2 l_t(\theta, da, db) \;=\; & \underbrace{tr\left\{\Omega_{t,\theta}^{-1}\eta_{t,\theta}(1,0)\right\}Y_{2t-1}dadb}_{(a)} + \underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\eta}_{t,b}(1,0)\right\}y_{t-1}(b)da}_{(b)} \\
& \underbrace{+tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\Omega_{t,\theta}^{-1}\eta_{t,\theta}(1,0)\right\}y_{t-1}(b)da}_{(c)}
\end{aligned}$$

As above, the absolute value of (a) squared is bounded by $\kappa \|z_t\|^2 Y_{2t-1}^2 \left(1+Y_{2t-1}^2\right)/\delta_{(t)}$ and hence, as $Y_{2t-1}^2 \left(1+Y_{2t-1}^2\right)/\delta_{(t)} \leq \mathcal{C}$, has finite expectation. Likewise, (b) is bounded by $\kappa |y_{t-1}| |Y_{2t-1}| \left(1+Y_{2t-1}^2\right)/\delta_{(t)} \leq \mathcal{C}$. And finally, (c) is bounded by $\kappa \|z_t\|$ times,

$$\begin{aligned}
&\left\|\Omega_t^{-1/2}\right\| \left\|(1,0)\,\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\right\| |y_{t-1}| \\
\leq\;& \frac{\left(c_1+c_2 y_{t-1}^2+c_3 Y_{2t-1}^2\right)^{1/2}\left(c_4 Y_{2t-1}^2+c_5 Y_{2t-1}^4+c_6|y_{t-1}Y_{2t-1}|+c_7\left|y_{t-1}Y_{2t-1}^3\right|\right)}{\delta_{(t)}^{3/2}} |y_{t-1}| \leq \mathcal{C},
\end{aligned}$$

with c_i constants, cf. the evaluations applied in *Part S3* in the proof of Lemma B.1 and the definition of $\delta_{(t)}$ in (B.6).

Part I6: $\partial^2 l_t(\theta)/\partial\tilde{\theta}\partial b$: Observe that with $\ddot{\Omega}_{t,\tilde{\theta},b} := d^2\Omega_{t,\theta}\left(d\tilde{\theta}, db\right)$, $-2d^2 l_t\left(\theta, d\tilde{\theta}, db\right)$ is given by,

$$\begin{aligned}
&\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\ddot{\Omega}_{t,\tilde{\theta},b}\left(I-\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}'\right)\right\}}_{(a)} - 2\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,\tilde{\theta}}\Omega_{t,\theta}^{-1}\eta_{t,\theta}\dot{\eta}_{t,b}'\right\}}_{(b)} + \\
&\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,\tilde{\theta}}\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\left(2\Omega_{t,\theta}^{-1}\eta_{t,\theta}\eta_{t,\theta}'-I\right)\right\}}_{(c)}.
\end{aligned}$$

Next, at $\theta=\theta_0$, (a) is bounded by $\kappa\left(1+\|z_t\|^2\right)\left\|\Omega_t^{-1}\ddot{\Omega}_{t,\tilde{\theta},b}\right\|$, where $\ddot{\Omega}_{t,\tilde{\theta},b}=0$ apart from the cases where $\tilde{\theta}=\phi_{11}^2,\phi_{22}^2$. We find

$$\left\|\Omega_t^{-1}\ddot{\Omega}_{t,\phi_{11}^2,b}\right\| \leq \kappa\left\|\Omega_t^{-1}S_{11}\right\| |y_{t-1}Y_{2t-1}| \leq \kappa |y_{t-1}Y_{2t-1}|\left(1+Y_{2t-1}^2\right)/\delta_{(t)} \leq \mathcal{C},$$

and likewise, $\left\|\Omega_t^{-1}\ddot{\Omega}_{t,\phi_{22}^2,b}\right\| \leq \kappa Y_{2t-1}^2\left(1+Y_{2t-1}^2\right)/\delta_{(t)} \leq \mathcal{C}$. Similarly, we get for the term in (c) that its absolute value is bounded by $\kappa\left(1+\|z_t\|^2\right)$ times $\left\|\Omega_t^{-1}\dot{\Omega}_{t,\tilde{\theta}}\right\|$ and $\left\|\Omega_t^{-1}\dot{\Omega}_{t,b}\right\|$. The last two terms have been argued to be bounded by a constant in *Part I2* and *Part I3* above respectively. Finally, the term in (b) is bounded by $\kappa\|z_t\|$ using again $\left\|\Omega_t^{-1}\dot{\Omega}_{t,\tilde{\theta}}\right\| \leq \mathcal{C}$ and that $\left\|\Omega_t^{-1/2}\right\|^2 Y_{2t-1}^2 \leq \mathcal{C}$. ■

Third Derivatives:

The third derivatives are considered with θ varying in a compact neighbourhood of the true value θ_0, $\theta \in \mathcal{K}(\theta_0)$. Thus we let, $\phi_{ij}^2 \in [\phi_{ij}^{L2}, \phi_{ij}^{U2}]$ for $i,j=\alpha_\perp,\beta$ and with $\phi_{ij}^{L2}>0$. Moreover, $\omega_{ij}\in[\omega_{ij}^L,\omega_{ij}^U]$ for $i,j=1,2$, $\omega_{11}^L,\omega_{22}^L>0$ and with $|\Omega|=\omega_{11}\omega_{22}-\omega_{12}^2\geq\psi>0$. Finally, $a\in[a^L,a^U]$ and $b\in[b^L,b^U]$.

Lemma B.3 *With $\mathcal{K}(\cdot)$ just defined, and under the assumptions of Theorem 3, it holds that*

$$\sup_{\theta\in\mathcal{K}(\theta_0)}\left|\tfrac{1}{T}\partial^3 L(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k\right| \leq \mathcal{V}_T$$

where $\mathcal{V}_T \xrightarrow{P} \mathcal{V} < \infty$ *and the indices* i, j *and* k *applied to* θ *refer to individual entries in* θ, *where* $\theta = \left(a, b, \tilde{\theta}'\right)$, $\tilde{\theta} = (\theta_{12}, \theta'_{11}, \theta'_{22})'$.

Proof of Lemma B.3: Observe first that trivially, $\partial^3 l_t(\theta)/\partial a^3 = 0$. Next, consider the parameter entry combinations in turn, observing that by definition of the neighbourhood $\mathcal{K}(\theta_0)$,

$$\delta_{(t,\theta)} \geq \psi + \left[\delta_2^L + \delta_5^L Y_{2t-1}^2\right] y_{t-1}^2(b) + \left[\delta_3^L + \delta_4^L Y_{2t-1}^2\right] Y_{2t-1}^2, \tag{B.35}$$

with $\delta_2^L = \omega_{22}^L \phi_{\beta\beta}^{L2} > 0$, $\delta_3^L = \left(\omega_{11}^L \phi_{\alpha_\perp\alpha_\perp}^{L2} + \omega_{22}^L \phi_{\beta\alpha_\perp}^{L2}\right) > 0$, $\delta_4^L = \phi_{\beta\alpha_\perp}^{L2}\phi_{\alpha_\perp\alpha_\perp}^{L2} > 0$, $\delta_5^L = \phi_{\beta\beta}^{L2}\phi_{\alpha_\perp\alpha_\perp}^{L2} > 0$.

Part TD1: $\partial^3 l_t(\theta)/\partial\theta_{ij}\partial a^2$ for $i, j = 1, 2$. From (B.27), $\partial^2 l_t(\theta)/\partial a^2 = -tr\left\{\Omega_{t,\theta}^{-1} S_{11}\right\} y_{t-1}^2(b)$ such that,

$$\partial^3 l_t(\theta)/\partial a^2 \partial\theta_{ij} = tr\left\{\Omega_{t,\theta}^{-1/2} S_{ij} \Omega_{t,\theta}^{-1} S_{11} \Omega_{t,\theta}^{-1/2}\right\} y_{t-1}^2(b)\, \lambda_{ij,t}.$$

and we find,

$$\left\|\partial^3 l_t(\theta)/\partial\theta_{ij}\partial a^2\right\| \leq \mathcal{C}\left\|\Omega_{t,\theta}^{-1/2} S_{ij} \Omega_{t,\theta}^{-1/2}\right\| \|\lambda_{ij,t}\| \left\|\Omega_{t,\theta}^{-1/2} S_{11} \Omega_{t,\theta}^{-1/2}\right\| y_{t-1}^2(b),$$

which are uniformly bounded for $i, j = 1, 2$. Recall that $\lambda_{11,t} = \left(1, y_{t-1}^2(b), Y_{2t-1}^2\right)'$ such that evaluation of the second term, $\left\|\Omega_{t,\theta}^{-1/2} S_{11} \Omega_{t,\theta}^{-1/2}\right\| y_{t-1}^2(b)$ is included by evaluating for $i, j = 1$,

$$\begin{aligned}
&\left\|\Omega_{t,\theta}^{-1/2} S_{11} \Omega_{t,\theta}^{-1/2}\right\| \|\lambda_{11,t}\| = \sqrt{tr\left\{\left[\Omega_{t,\theta}^{-1} S_{11}\right]^2\right\}}\, \|\lambda_{11,t}(\theta)\| \\
&= |\omega_{t,22}(\theta)|\, \|\lambda_{11,t}\| / \delta_{(t,\theta)} \leq \frac{\left(\omega_{22}^U + \phi_{\alpha_\perp\alpha_\perp}^{U2} Y_{2t-1}^2\right)\left(1 + Y_{2t-1}^2 + y_{1t-1}^2(b)\right)}{\psi + \left[\delta_2^L + \delta_5^L Y_{2t-1}^2\right] y_{t-1}^2(b) + \left[\delta_3^L + \delta_4^L Y_{2t-1}^2\right] Y_{2t-1}^2} \\
&\leq \left[\frac{\left(\omega_{22}^U + \phi_{\alpha_\perp\alpha_\perp}^{U2} Y_{2t-1}^2\right)}{\left(\psi + \delta_3^L Y_{2t-1}^2\right)} + \frac{\left(\omega_{22}^U + \phi_{\alpha_\perp\alpha_\perp}^{U2} Y_{2t-1}^2\right)}{\left(\delta_3^L + \delta_4^L Y_{2t-1}^2\right)} + \frac{\left(\omega_{22}^U + \phi_{\alpha_\perp\alpha_\perp}^{U2} Y_{2t-1}^2\right)}{\left(\delta_2^L + \delta_5^L Y_{2t-1}^2\right)}\right] \leq \mathcal{C}.
\end{aligned} \tag{B.36}$$

Next, for $i, j = 2$,

$$\begin{aligned}
&\left\|\Omega_{t,\theta}^{-1/2} S_{22} \Omega_{t,\theta}^{-1/2}\right\| \|\lambda_{22,t}\| = |\omega_{t,11}(\theta)|\, \|\lambda_{22,t}\| / \delta_{(t,\theta)} \\
&\leq \frac{\left(\omega_{11}^U + \phi_{\beta\beta}^{U2} y_{t-1}^2(b) + \phi_{\beta\alpha_\perp}^{U2} Y_{2t-1}^2\right)\left(1 + Y_{2t-1}^2\right)}{\psi + \left[\delta_2^L + \delta_5^L Y_{2t-1}^2\right] y_{t-1}^2(b) + \left[\delta_3^L + \delta_4^L Y_{2t-1}^2\right] Y_{2t-1}^2} \\
&\leq \frac{\left(\omega_{11}^U + \phi_{\beta\alpha_\perp}^{U2} Y_{2t-1}^2\right)}{\psi + \delta_3^L Y_{2t-1}^2} + \frac{\left(\omega_{11}^U + \phi_{\beta\alpha_\perp}^{U2} Y_{2t-1}^2\right)}{\delta_3^L + \delta_4^L Y_{2t-1}^2} + \frac{\phi_{\beta\beta}^{U2}\left(1 + Y_{2t-1}^2\right)}{\delta_2^L + \delta_5^L Y_{2t-1}^2} \leq \mathcal{C}.
\end{aligned} \tag{B.37}$$

And finally,

$$\left\|\Omega_{t,\theta}^{-1/2} S_{12} \Omega_{t,\theta}^{-1/2}\right\| \leq \kappa \frac{|\omega_{t,11}(\theta)| + |\omega_{t,22}(\theta)| + 2|\omega_{t,12}(\theta)|}{\psi + \left[\delta_2^L + \delta_5^L Y_{2t-1}^2\right] y_{t-1}^2(b) + \left[\delta_3^L + \delta_4^L Y_{2t-1}^2\right] Y_{2t-1}^2} \leq \mathcal{C}.$$

Part TD2: $\partial^3 l_t(\theta)/\partial\theta_{ij}\partial\theta'_{kl}\partial a$ for $i,j,k,l=1,2$: Using (B.29),

$$\partial^3 l_t(\theta)/\partial\theta_{ij}\partial\theta'_{kl}\partial a = 2tr\left\{\Omega_{t,\theta}^{-1}S_{ij}\Omega_{t,\theta}^{-1}S_{kl}\Omega_{t,\theta}^{-1}\eta_{t,\theta}(1,0)\right\}\lambda_{ij,t}\lambda'_{kl,t}y_{t-1}(b),$$

for $i,j,k,l=1,2$. Applying the uniform bounds in (B.36) and (B.37),

$$\begin{aligned}
&\left\|\partial^3 l_t(\theta)/\partial\theta_{ij}\partial\theta'_{kl}\partial a\right\| \\
&\leq \mathcal{C}\left\|\Omega_{t,\theta}^{-1/2}S_{ij}\Omega_{t,\theta}^{-1/2}\right\|\|\lambda_{ij,t}\|\left\|\Omega_{t,\theta}^{-1/2}S_{kl}\Omega_{t,\theta}^{-1/2}\right\|\|\lambda_{kl,t}\|\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\|\left\|(1,0)\,\Omega_{t,\theta}^{-1/2}\right\||y_{t-1}(b)| \\
&\leq \mathcal{C}\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\|\left\|(1,0)\,\Omega_{t,\theta}^{-1/2}\right\||y_{t-1}(b)|.
\end{aligned}$$

Next, rewrite $\eta_{t,\theta}$ as

$$\eta_{t,\theta} = \left(\eta_{t,\theta}-\eta_t\right)+\eta_t = \left[(a-a_0)\,y_{t-1}(b)+a_0\,(b_0-b)\,Y_{2t-1}\right](1,0)'+\Omega_t^{1/2}z_t. \qquad \text{(B.38)}$$

Using this,

$$\begin{aligned}
\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\| &\leq \kappa\left[\left\|(1,0)\,\Omega_{t,\theta}^{-1/2}\right\|(|y_{t-1}(b)|+|Y_{2t-1}|)+\left\|\Omega_{t,\theta}^{-1/2}\Omega_t^{1/2}\right\|\|z_t\|\right] \qquad \text{(B.39)}\\
&\leq \kappa\left[\frac{\left(\omega_{22}^U+\phi_{\alpha_\perp\alpha_\perp}^{U2}Y_{2t-1}^2\right)^{1/2}(|y_{t-1}(b)|+|Y_{2t-1}|)}{\left(\psi+\left[\delta_2^L+\delta_5^L Y_{2t-1}^2\right]y_{t-1}^2(b)+\left[\delta_3^L+\delta_4^L Y_{2t-1}^2\right]Y_{2t-1}^2\right)^{1/2}}+\|z_t\|\right]\leq \mathcal{C}\left[1+\|z_t\|\right],
\end{aligned}$$

where we have used that $\left\|\Omega_{t,\theta}^{-1/2}\Omega_t^{1/2}\right\|^2 = tr\left\{\Omega_{t,\theta}^{-1}\Omega_t\right\}\leq\mathcal{C}$. Moreover,

$$\left\|(1,0)\,\Omega_{t,\theta}^{-1/2}\right\|^2 y_{t-1}^2(b)\leq\frac{\left(\omega_{22}^U+\phi_{\alpha_\perp\alpha_\perp}^{U2}Y_{2t-1}^2\right)}{\delta_2^L+\delta_5^L Y_{2t-1}^2}\leq\mathcal{C}. \qquad \text{(B.40)}$$

Hence we can use $\mathcal{V}_T=\frac{1}{T}\sum_{t=1}^T\mathcal{C}\left[1+\|z_t\|\right]$ since z_t has finite first order moment.

Part TD3: $d^3 l_t(\theta,d\theta_{ij},d\theta_{kl},d\theta_{mn})$ for $i,j,k,l,m,n=1,2$: By (B.28), we find that the third order differential is bounded by,

$$\begin{aligned}
&\left\|d^3 l_t(\theta,d\theta_{ij},d\theta_{kl},d\theta_{mn})\right\|\leq\mathcal{C}\left\|\Omega_{t,\theta}^{-1/2}S_{ij}\Omega_{t,\theta}^{-1/2}\right\|\|\lambda_{ij,t}\|\left\|\Omega_{t,\theta}^{-1/2}S_{kl}\Omega_{t,\theta}^{-1/2}\right\|\|\lambda_{kl,t}\|\times\\
&\left\|\Omega_{t,\theta}^{-1/2}S_{ij}\Omega_{t,\theta}^{-1/2}\right\|\|\lambda_{mn,t}\|\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\|^2\leq\mathcal{C}\left[1+\|z_t\|^2\right],
\end{aligned}$$

using (B.36), (B.37) and (B.39).

Part TD4: $\partial^3 l_t(\theta)/\partial a^2\partial b$: Recall that by (B.27), $\partial^2 l_t(\theta)/\partial a^2=-tr\left\{\Omega_{t,\theta}^{-1}S_{11}\right\}y_{t-1}^2(b)$ such that,

$$d^3 l_t(\theta,da,da,db)=tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\Omega_{t,\theta}^{-1}S_{11}\right\}y_{t-1}^2(b)+2tr\left\{\Omega_{t,\theta}^{-1}S_{11}\right\}y_{t-1}(b)\,Y_{2t-1}.$$

Now, $\left\|\Omega_{t,\theta}^{-1/2}S_{11}\Omega_{t,\theta}^{-1/2}\right\| y_{t-1}^2(b) \leq \frac{\left(\omega_{22}^U+\phi_{\alpha_\perp\alpha_\perp}^{U2}Y_{2t-1}^2\right)y_{t-1}^2(b)}{\psi+\left[\delta_2^L+\delta_5^L Y_{2t-1}^2\right]y_{t-1}^2(b)+\left[\delta_3^L+\delta_4^L Y_{2t-1}^2\right]Y_{2t-1}^2} \leq \mathcal{C}$, and likewise, uniformly in $\mathcal{K}(\theta_0)$,

$$\left\|\Omega_{t,\theta}^{-1/2}\dot{\Omega}_{t,b}\Omega_{t,\theta}^{-1/2}\right\| \leq \mathcal{C}, \tag{B.41}$$

using arguments similar to *Part S3*. Finally, for the last term we have,

$$\left|tr\left\{\Omega_{t,\theta}^{-1}S_{11}\right\}y_{t-1}(b)Y_{2t-1}\right| \leq \frac{\left(\omega_{22}^U+\phi_{\alpha_\perp\alpha_\perp}^{U2}Y_{2t-1}^2\right)|y_{t-1}(b)||Y_{2t-1}|}{\psi+\left[\delta_2^L+\delta_5^L Y_{2t-1}^2\right]y_{t-1}^2(b)+\left[\delta_3^L+\delta_4^L Y_{2t-1}^2\right]Y_{2t-1}^2} \leq \mathcal{C}.$$

Part TD5: $\partial^3 l_t(\theta)/\partial a\partial b^2$: From *Part I3,* proof of Lemma B.2, and using that by definition of $\Omega_{t,\theta}$, we have $\dot{\Omega}_{t,a}=0$,

$$\begin{aligned}
-2d^3 l_t(\theta,db,db,da) &= -2\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\ddot{\Omega}_{t,b,b}\Omega_{t,\theta}^{-1}\eta_{t,\theta}\dot{\eta}_{t,a}'\right\}}_{(a)} - 4\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\Omega_{t,\theta}^{-1}\dot{\eta}_{t,a}\dot{\eta}_{t,b}'\right\}}_{(b)} \\
&+4\underbrace{tr\left\{\left[\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\right]\left(\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b}\right)\left[\Omega_{t,\theta}^{-1}\eta_{t,\theta}\dot{\eta}_{t,a}'\right]\right\}}_{(c)} + 4\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\eta}_{t,b}\ddot{\eta}_{t,b,a}'\right\}}_{(d)} \\
&-8\underbrace{tr\left\{\Omega_{t,\theta}^{-1}\dot{\Omega}_{t,b^*}\Omega_{t,\theta}^{-1}\eta_{t,\theta}\ddot{\eta}_{t,b,a}'\right\}}_{(e)}.
\end{aligned}$$

Using $|tr\{AB\}| \leq \|A\|\|B\|$ repeatedly, $\partial^3 l_t(\theta)/\partial a\partial b^2$ is uniformly bounded if $\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\|$, $\left\|\Omega_{t,\theta}^{-1/2}\dot{\Omega}_{t,b}\Omega_{t,\theta}^{-1/2}\right\|$, $\left\|\Omega_{t,\theta}^{-1/2}\ddot{\Omega}_{t,b,b}\Omega_{t,\theta}^{-1/2}\right\|$, $\left\|\Omega_{t,\theta}^{-1/2}\dot{\eta}_{t,b}\right\|$ and $\left\|\Omega_{t,\theta}^{-1/2}\ddot{\eta}_{t,a,b}\right\|$ are. For the first two see (B.41) and (B.39) respectively, and the remaining ones are bounded using similar arguments.

Part TD6: $\partial^3 l_t(\theta)/\partial b^3$: From *Part I3,* proof of Lemma B.2, we find as in *Part TD5,* that $\left|\partial^3 l_t(\theta)/\partial b^3\right|$ is uniformly bounded if, as used in *Part TD5,* $\left\|\Omega_{t,\theta}^{-1/2}\dot{\Omega}_{t,b}\Omega_{t,\theta}^{-1/2}\right\|$, $\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\|$, $\left\|\Omega_{t,\theta}^{-1/2}\ddot{\Omega}_{t,b,b}\Omega_{t,\theta}^{-1/2}\right\|$ and $\left\|\Omega_{t,\theta}^{-1/2}\dot{\eta}_{t,b}\right\|$ are.

Part TD7: $\partial^3 l_t(\theta)/\partial\theta_{ij}\partial b^2$: In addition to the quantities in *Part TD5* and *Part TD6,* $\partial^3 l_t(\theta)/\partial\theta_{ij}\partial b^2$ is uniformly bounded if also $\left\|\Omega_{t,\theta}^{-1/2}\dot{\Omega}_{t,\theta_{ij}}\Omega_{t,\theta}^{-1/2}\right\|$, $\left\|\Omega_{t,\theta}^{-1/2}\dddot{\Omega}_{t,b,b,\theta_{ij}}\Omega_{t,\theta}^{-1/2}\right\|$ and $\left\|\Omega_{t,\theta}^{-1/2}\ddot{\Omega}_{t,b,\theta_{ij}}\Omega_{t,\theta}^{-1/2}\right\|$ are. The first term for $\theta_{ij}=\theta_{12},\theta_{11}$ and θ_{22} respectively, is dealt with in *Part TD1,* while for the last two similar arguments can be used. For example, with $\dddot{\Omega}_{t,b,b,\theta_{ij}}=0$ for $\theta_{ij}=\theta_{12},\theta_{22}$, while

$$\left\|\Omega_{t,\theta}^{-1/2}\dddot{\Omega}_{t,b,b,\theta_{11}}\Omega_{t,\theta}^{-1/2}\right\|^2 \leq \kappa\left(\phi_{\beta\beta}^{2L}+\phi_{\beta\alpha_\perp}^2\right)Y_{2t-1}^2 tr\left\{\left[\Omega_{t,\theta}^{-1}S_{11}\right]^2\right\} \leq \mathcal{C}.$$

Part TD8: $\partial^3 l_t(\theta)/\partial\theta_{ij}\partial\theta'_{kl}\partial b$: Using *Part I2*, simple calculations give that this is uniformly bounded as $\left\|\Omega_{t,\theta}^{-1/2}\dot{\Omega}_{t,\theta_{ij}}\Omega_{t,\theta}^{-1/2}\right\|$ (see *Part TD1)*, $\left\|\Omega_{t,\theta}^{-1/2}\ddot{\Omega}_{t,\theta_{ij},b}\Omega_{t,\theta}^{-1/2}\right\|$ (see *Part TD7)* as well as $\left\|\Omega_{t,\theta}^{-1/2}\dot{\Omega}_{t,b}\Omega_{t,\theta}^{-1/2}\right\|$, $\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\|$ and $\left\|\Omega_{t,\theta}^{-1/2}\dot{\eta}_{t,b}\right\|$ are (see *Part TD5)*.

Part TD9: $\partial^3 l_t(\theta)/\partial a\partial b\partial\theta_{ij}$: Using *Part I5*, we again find this is uniformly bounded as $\left\|\Omega_{t,\theta}^{-1/2}\dot{\Omega}_{t,\theta_{ij}}\Omega_{t,\theta}^{-1/2}\right\|$ (see *Part TD1)*, $\left\|\Omega_{t,\theta}^{-1/2}\dot{\eta}_{t,b}\right\|$ and $\left\|\Omega_{t,\theta}^{-1/2}\eta_{t,\theta}\right\|$ (see *Part TD5)* are, together with

$$\left\|(1,0)\,\Omega_{t,\theta}^{-1/2}\right\|^2\left(Y_{2t-1}^2+y_{t-1}^2(b)\right)\leq\frac{\left(\omega_{22}^U+\phi_{\alpha_\perp\alpha_\perp}^2Y_{2t-1}^2\right)}{\delta_{(t,\theta)}}\left(Y_{2t-1}^2+y_{t-1}^2(b)\right)\leq\mathcal{C},$$

using (B.35), (B.39) and (B.40). ■

References

[1] Aue, A., L. Horváth and J. Steinebach, 2006, *Estimation in Random Coefficient Autoregressive Models*, Journal of Time Series Analysis, 27:61-76.

[2] Avarucci, M., E. Beutner and P. Zaffaroni, 2012, *On Moment Conditions for Quasi-Maximum Likelihood Estimation of Multivariate ARCH Models*, Econometric Theory, forthcoming.

[3] Bauwens, L., S. Laurent and J. V. Rombouts, 2006, *Multivariate GARCH Models: A Survey,* Journal of Applied Econometrics, 21:79-109.

[4] Bec, F. and A. Rahbek, 2004, *Vector Equilibrium Correction Models with Non-Linear Discontinuous Adjustments*, Econometrics Journal, 7:628–651.

[5] Bec, F. A. Rahbek and N. Shephard, 2008, *The ACR Model: A Multivariate Dynamic Mixture Autoregression*, Oxford Bulletin of Economics and Statistics, 583-618.

[6] Berkes, I., L. Horváth, S. Ling, 2009, *Estimation in Nonstationary Random Coefficient Autoregressive Models*, Journal of Time Series Analysis, 30:395–416.

[7] Borkovec, M., 2000, *Extremal Behaviour of the Autoregressive Process with ARCH(1) Errors*, Stochastic Processes and their Applications, 85:189-207.

[8] Borkovec, M. and C. Klüppelberg, 2001, *The Tail of the Stationary Distribution of an Autoregressive Process with ARCH(1) Errors*, Annals of Applied Probability, 11:1220-1241.

[9] Brown, B.M., *Martingale Central Limit Theorems*, Annals of Mathematical Statistics, 42:49-56.

[10] Cavaliere, G., A. Rahbek and R. M. Taylor, 2010, *Cointegration Rank Testing under Conditional Heteroscedasticity,* Econometric Theory, 26:1719–1760.

[11] Cavaliere, G., A. Rahbek and R. M. Taylor, 2012, *Bootstrap Determination of the Co-integration Rank in VAR Models ,* Econometrica, forthcoming.

[12] Carta, A., D. Fantazzini and M. A. Maggi, 2008, *Discrete-time Affine Term Structure Models: An ARCH Formulation*, International Journal of Risk Assessment and Management, 11:164-179.

[13] Comte, F. and O. Lieberman, *Asymptotic Theory for Multivariate GARCH Processes*, 2003, Journal of Multivariate Analysis, 84:61–84.

[14] Cox, J.C., J.E. Ingersoll and S.A. Ross, 1985, *A Theory of Term Structure of Interest Rates,* Econometrica, 53:385–407.

[15] Dieci, L., and E.S. Van Vleck, 1995, *Computation of a few Lyapunov exponents for continuous and discrete dynamical systems*, Applied Numerical Mathematics, 17:275-291.

[16] Doornik, J. A., 2007, *Object-Oriented Matrix Programming Using Ox* (3rd ed.). London: Timberlake Consultants Press.

[17] Engle, R. and K. Kroner, 1995, *Multivariate Simultaneous Generalized ARCH*, Econometric Theory, 11:122-150.

[18] Feigin, P.D. and 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.

[19] Fong, P. W. and W. K. Li, 2004, *Some Results on Cointegration with Random Coefficients in the Error Correction Form: Estimation and Testing*, Journal of Time Series Analysis, 25:419-441.

[20] Francq, C. and J.M. Zakoian, 2010, *GARCH Models: Structure, Statistical Inference and Financial Applications*, Wiley.

[21] Gourieroux, C., A. Monfort and V. Polimenis, 2002, *Affine Term Structure Models*, Working Papers 2002-49, Centre de Recherche en Economie et Statistique (CREST), Paris.

[22] Jeantheau, T., 1998, *Strong Consistency of Estimators for Multivariate ARCH Models*, Econometric Theory, 14:70–86.

[23] Jensen, S. T. and A. Rahbek, 2004, *Asymptotic Inference for Nonstationary GARCH*, Econometric Theory, 20:1203-1226.

[24] Jensen, S. T. and A. Rahbek, 2007, *A Note on the Law of Large Numbers for Functions of Geometrically Ergodic Time Series*, Econometric Theory*, 23:761-766.*

[25] Johansen, S., 1996, *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

[26] Klüppelberg, C. and S. Pergamenchtchikov, 2004, *The Tail of the Stationary Distribution of a Random Coefficient AR(q) Model*, Annals of Applied Probability, 14:971-1005.

[27] Klüppelberg, C. and S. Pergamenchtchikov, 2007, *Extremal Behaviour of Models with Multivariate Random Recurrence Representation*, Stochastic Processes and their Applications, 117:432-456

[28] Lange, T., A. Rahbek and S.T. Jensen, 2011, *Estimation and Asymptotic Inference in the AR-ARCH Model*, Econometric Reviews, 30:129-153.

[29] Le, A., K.J. Singleton and Q. Dai, 2010, *Discrete-Time Affine^Q Term Structure Models with Generalized Market Prices of Risk*, The Review of Financial Studies, 23(5):2184-2227.

[30] Ling, S., 2004, *Estimation and Testing of Stationarity for Double-Autoregressive Models,* Journal of the Royal Statistical Society, Series B, 66:63-78.

[31] Ling, S., 2007, *A Double AR(p) Model: Structure and Estimation*, Statistica Sinica, 17:161-175.

[32] Ling, S. and D. Li, 2008, *Asymptotic Inference for a Nonstationary Double AR(1) Model*, Biometrika, 95:257-263.

[33] Tjøstheim, D., 1990, *Non-linear Time series and Markov Chains*, Advances in Applied Probability, 22:587–611.

[34] Xie, W. C. and Q. Huang, 2009, *Simulation of moment Lyapunov exponents for linear homogeneous stochastic systems*, Journal of Applied Mechanics, 76:031001-1.