UNIVERSITY OF COPENHAGEN

Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models

Kristensen, Dennis; Rahbek, Anders

Publication date:
2010

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Kristensen, D., & Rahbek, A. (2010). *Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models*. Department of Economics, University of Copenhagen.

Discussion Papers
Department of Economics
University of Copenhagen

No. 10-25

Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models

Dennis Kristensen, Anders Rahbek

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)

Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models*

Dennis Kristensen† — Columbia University
Anders Rahbek‡ — University of Copenhagen

October, 2010

Abstract

In this paper, we consider a general class of vector error correction models which allow for asymmetric and non-linear error correction. We provide asymptotic results for (quasi-)maximum likelihood (QML) based estimators and tests. General hypothesis testing is considered, where testing for linearity is of particular interest as parameters of non-linear components vanish under the null. To solve the latter type of testing, we use the so-called sup tests, which here requires development of new (uniform) weak convergence results. These results are potentially useful in general for analysis of non-stationary non-linear time series models. Thus the paper provides a full asymptotic theory for estimators as well as standard and non-standard test statistics. The derived asymptotic results prove to be new compared to results found elsewhere in the literature due to the impact of the estimated cointegration relations. With respect to testing, this makes implementation of testing involved, and bootstrap versions of the tests are proposed in order to facilitate their usage. The asymptotic results regarding the QML estimators extend results in Kristensen and Rahbek (2010, Journal of Econometrics) where symmetric non-linear error correction considered. A simulation study shows that the finite sample properties of the bootstrapped tests are satisfactory with good size and power properties for reasonable sample sizes.

*Both authors are affiliated with CREATES funded by the Danish National Research Foundation. We wish to thank participants at the 18th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, in Novara, Italy, and participants at Montreal Econometrics Seminars and Oxbridge Time Series Group Workshops, Cambridge, for helpful comments and suggestions. Also we are grateful to discussions with M. Seo, LSE, London. The Velux Foundation funded a longer research visit for Kristensen to Copenhagen University, where part of this research was conducted. Part of the research was also conducted while Kristensen visited Princeton University whose hospitality is gratefully acknowledged. Kristensen received research support from the National Science Foundation (grant no. SES-0961596).

†E-mail: dk2313@columbia.edu

‡E-mail: rahbek@econ.ku.dk

1 Introduction

We develop estimators and test statistics for a class of nonlinear vector error correction models with cointegration. Both estimators and test statistics are based on the Gaussian (quasi-)likelihood, and we propose both Lagrange Multiplier (LM) and Likelihood Ratio (LR) test statistics. Our framework allows for testing a wide range of relevant hypotheses. Of particular interest is the hypothesis of nonlinearity, where in general nuisance parameters entering the nonlinear component vanish under the null. We solve this problem by employing sup-tests as advocated in Andrews and Ploberger (1994, 1995), Davies (1987), Hansen (1996) and Hansen and Seo (2002). We derive the asymptotic distributions of both estimators and test statistics under weak restrictions. As part of the theoretical analysis, new functional central limit theorems are developed which are of independent interest in the analysis of nonlinear, non-stationary models.

Allowing for unknown cointegration relations prove to complicate the analysis and the resulting asymptotic distributions of both the quasi-maximum likelihood estimators (QMLE's) and test statistics considerably. In particular, we find non-standard limiting distributions of both estimators and test statistics, when compared to the ones established in linear cointegration models and for nonlinear stationary models, including cointegration models with known long-run parameters. This is due to the fact that the limiting distributions of the estimators of the long-run and short-run parameters are not asymptotically independent. This again spills over to the distribution of the test statistics which are influenced by both the estimated long-run and short-run parameters. This happens even in the case when the null hypothesis only involves restrictions on either of the parameters. If in addition parameters vanish under the null, as is often the case in testing for linearity in the short-run dynamics, the limiting distributions complicate further, and the proposed sup-tests are shown to converge towards a supremum over a squared non-Gaussian process.

As such, our results show that one cannot ignore the estimation of the long-run parameters if these are unknown. This also explains why our findings are different from existing results on testing in nonlinear time series models. In particular, as discussed in further detail below, previous studies investigating sup-tests in cointegration models either assume that the cointegrating relations are known, or that the additional estimation error due to unknown (super consistent) relations does not affect the tests.

Hansen (1996) develops an asymptotic theory for sup-tests in a stationary setting. In this case, the limiting distributions can be written as a supremum over squared Gaussian processes. This theory is extended to threshold and smooth transition cointegration models with known cointegrating relations (β) in Gonzalo and Pitarakis (2006) and Kilic (2009) respectively. Since β is assumed known, their models and results become similar to the ones of Hansen (1996).

Our results regarding sup tests for linearity are related to the ones of Caner and Hansen (2001) who test for linearity in univariate threshold autoregressions with unit roots. We find in the multivariate case, as they do for the univariate case, that the limiting distribution of the sup test statistic consists of two terms: A stationary component due to the short-run parameters and a non-stationary component due to the presence of unknown long-run parameters. On the other hand, our results differ from Hansen and Seo (2002) and Nedeljkovic (2008), where sup-tests of linearity in threshold and smooth transition cointegration models respectively are considered, as they (implicitly) assume that estimators of β have no impact on the asymptotic behaviour of their test statistic. That is, they conclude that limiting distributions can be represented as supremums over squared Gaussian processes. Finally, we note that in a different vein some studies have proposed to test for linearity by approximating the true model using a Taylor expansion of the non-linear component (Choi and Saikkonen, 2004; Kapetanios, Shin and Snell, 2006). This removes the problem of vanishing parameters, but on the other hand introduces asymptotic biases of estimators and test statistics under the alternative, since a misspecified model is being employed.

A number of other studies have developed and analyzed estimators for cointegration models with non-linear error correction. In particular, Kristensen and Rahbek (2010) derive the properties of QMLE for class of smooth nonlinear error-correction models. However, they restrict themselves to the case of symmetric error-correction while we also allow for asymmetric adjustments. Thus, our findings regarding the QMLE generalize and improve upon the results found in that study. Our results also complement the ones of Seo (2010) who consider estimation of threshold error correction models using kernel smoothers to handle discontinuities implied by the thresholds.

To establish our theoretical results, it proves necessary to develop a new functional central limit theorems (FCLT's) uniformly over the unidentified parameters. Such results are useful in the analysis of nonlinear models with non-stationary components, and we therefore establish uniform FCLT's in a general framework that includes, but is not restricted to, the particular class of non-linear error correction models of this study. These results generalize the ones established in Caner and Hansen (2001, Section 2) and will be useful in the analysis of other non-linear time series models; as such, they should be of independent interest.

Due to the highly non-standard limiting distribution of estimators and test statistics, we propose to implement the estimation and testing procedures using bootstrapping based on the ideas developed in Cavaliere, Rahbek and Taylor (2010a,b). In particular, we propose to use the wild bootstrap, which should make the bootstrap tests robust to heteroskedasticity. Seo (2006,2008a) and Gonzalo and Pitarakis (2006) also consider bootstrap methods for testing in non-stationary time series models but in different settings. A simulation study investigates the finite sample performance of the proposed bootstrap version of the sup-LR test. We find that the proposed testing scheme has good size and power properties and so offer a convenient tool for inference in nonlinear error correction models.

The remains of the paper is organized as follows: We present the model and propose estimators and test statistics of the parameters in Section 2. The auxiliary functional central limit theorems (FCLT) are derived in Section 3. These are then in turn used in Section 4 and 5 to derive the limiting distributions of estimators and test statistics respectively. A bootstrap procedure for evaluating the distribution of the test statistic is proposed in Section 6, while Section 7 presents the results of a simulation study. Section 8 concludes. All proofs and lemmas have been relegated to Appendices A-B and C-D respectively.

Throughout, the following notation will be used: We use $\xrightarrow{P}$ and $\xrightarrow{D}$ to denote convergence in probability and distribution respectively; $C(\mathcal{A})$ and $D(\mathcal{A})$ denote the space of continuous and cadlag functions with domain $\mathcal{A}$; $df(x;dx)$ denotes the differential of a mapping $f(x)$ in the direction dx; by $vec(a,b)$, we mean $\left(vec(a)', vec(b)'\right)'$. For any parameter θ, θ_0 will denote its true, data-generating value; for any matrix $m \times n$ matrix A of full column rank $n \leq m$, we define $\bar{A} = A(A'A)^{-1}$, and $A_\perp$ as a $m \times (m-n)$ matrix such that $[A, A_\perp]$ has full rank m and $A'A_\perp = 0$.

2 Framework

2.1 Model

Let $X_t \in \mathbb{R}^p$, $t = 1, ..., T$, be observations from the following error correction model (ECM),

$$\Delta X_t = g\left(\beta' X_{t-1}\right) + \Phi_1 \Delta X_{t-1} + ... + \Phi_k \Delta X_{t-k} + \varepsilon_t, \tag{2.1}$$

where $\Delta X_t = X_t - X_{t-1}$ and the error term ε_t satisfies

$$E\left[\varepsilon_t | \mathcal{F}_{t-1}\right] = 0, \quad \Omega \equiv E\left[\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}\right] < \infty, \tag{2.2}$$

with $\mathcal{F}_{t-1} = \mathcal{F}(X_{t-1}, X_{t-2}, ...)$ denoting the information set based on past values of X_t. The function $g(\cdot)$ describes the (potentially nonlinear) error correction towards the long-run equilibrium. The equilibrium of the process is characterized by the cointegration relations; namely, the $r \geq 1$ linear combinations $\beta' X_t$, with $\beta \in \mathbb{R}^{p \times r}$.

Without loss of generality, we specify $g(\cdot)$ as composed by a linear and nonlinear part:

$$g\left(\beta' X_{t-1}\right) = \alpha \beta' X_{t-1} + \delta \psi\left(\beta' X_{t-1}; \xi\right). \tag{2.3}$$

In this general class of specifications, the deviation from the basic linear ECM is given by the r_δ-dimensional vector function $\psi(\beta' X_{t-1}; \xi)$ multiplied by the $(p \times r_\delta)$-dimensional parameter δ. The parameter ξ in the nonlinear component may contain matrices and we let $d_\xi = \dim(vec(\xi))$ denote the dimension of the vectorized version of ξ. The above specification is sufficiently general to cover most known nonlinear error correction models found in the literature. Note that we here suppress the dependence of $g(\cdot)$ on the parameters, α, δ and ξ.

The form of g in eq. (2.3) embeds various smooth transition error correction models. In general, allowing for S different regimes in $\psi(\cdot)$ indexed by $s = 1, ..., S$, we may write,

$$\delta \psi(z, \xi) = \sum_{s=1}^{S} \delta_s \psi_s(z, \xi) \text{ with } \delta := (\delta_1, ..., \delta_S), \psi(z, \xi) := (\psi_1(z, \xi), ..., \psi_S(z, \xi))'. \tag{2.4}$$

Depending on the functional form of the ψ_s, this formulation allow for both symmetric and asymmetric response functions. A key example of the first type is the logistic STECM in Kristensen and Rahbek (2010), where

$$\psi_s(z, \xi) := \left[1 + \exp\left\{(z - \omega_s)' A_s (z - \omega_s)\right\}\right]^{-1} z, \tag{2.5}$$

with A_s positive definite $(r \times r)$-dimensional matrices, while ω_i are r-dimensional vectors, and $r_\delta = Sr$. The parameter ξ is given by $\xi = (\omega, A)$ with $\omega = (\omega_1, ..., \omega_S)$ and $A = (A_1, ..., A_S)$. With $\psi(z, \xi)$ chosen this way, observe that $\psi(z) = o(1)$ as $\|z\| \to \infty$ and, hence for large deviations as measured by $Z_t = \beta' X_t$, the linear component αz of $g(z; \gamma)$ in eq. (2.3) asymptotically dominates. Also note that the nonlinearity vanishes if indeed $\delta = 0$, in which case the STECM reduces to the linear ECM with $g(z; \gamma) = \alpha z$. To allow for asymmetric responses, Saikkonen (2008) studies alternative general specifications of ψ. An example of Saikkonen (2008) is

$$\psi_s(z, \xi) = \left[1 + \exp\left\{a_s'(z - \omega_s)\right\}\right]^{-1} z, \tag{2.6}$$

with a_s being an $r-$dimensional vector. Depending on whether $(z - \omega_i)$ is orthogonal to a_s as $\|z\| \to \infty$, $\psi_s(z, \xi)$ will also asymptotically be contributing to the linear αz part in the error correction. The above class of models also contains threshold models where $\psi(z, \xi)$ contains indicator functions, see e.g. Hansen and Seo (2003) and Seo (2010). However, we will impose smoothness restrictions on $\psi(z, \xi)$ when analyzing our proposed estimators and test statistics which rule out threshold models. These could potentially however be dealt with by modifying our proposed estimators, replacing indicator functions by kernel smoothers, see e.g. Seo (2010), but will not be considered here.

Regarding identification, then as common in the cointegration literature β is identified up to a normalization and we therefore normalize β conveniently using a $(p \times (p - r))$ dimensional matrix κ_0, such that

$$\beta - \beta_0 = \kappa_0 b, \tag{2.7}$$

and b is the $((p - r) \times r)$ dimensional parameter to be estimated. Thus, $b_0 = 0$ corresponds to the true parameter value β_0. Using this, we can rewrite the model in eq. (2.1) as a nonlinear regression model in terms of $(Z_{0,t}, Z_{1,t}, Z_{2,t})$,

$$\Delta X_t = g\left(Z_{0,t-1} + b' Z_{1,t-1}\right) + \Phi Z_{2,t-1} + \varepsilon_t, \tag{2.8}$$

where

$$Z_{0,t} := \beta_0' X_t \in \mathbb{R}^r, \quad Z_{1,t} := X_t' \kappa_0 \in \mathbb{R}^{p-r}, \quad Z_{2,t} := \left(\Delta X_t', ..., \Delta X_{t-k+1}'\right)' \in \mathbb{R}^{pk}.$$

As argued in Kristensen and Rahbek (2010), the estimator of the error covariance matrix, Ω, will be asymptotically independent of the estimators of the other parameters (appearing in the conditional mean specification). We therefore collect all the conditional mean parameters in ϑ and leave out Ω which is treated separately. Finally, note that under the null of linearity $(\delta = 0)$ the parameter ξ vanishes. To emphasize the role played by the vanishing parameter ξ, we introduce θ which contains all parameter in ϑ *except for* ξ. Furthermore, we differentiate between short-run and long-run parameters and collect the former in η. Thus the parameters of interest are given by:

$$\vartheta := (\theta, \xi) = (b, \eta, \xi), \quad \eta := (\alpha, \delta, \Phi) = (\alpha, \delta, \Phi_1, \Phi_2, ..., \Phi_k). \tag{2.9}$$

We let Θ and Ξ denote the parameter spaces of $\theta = (\beta, \eta)$ and ξ respectively.

2.2 Estimation

Our proposed estimators are based on the Gaussian log-likelihood. In order to write the log-likelihood function, define the residuals,

$$\varepsilon_t(\theta,\xi) = \Delta X_t - \alpha\left(Z_{0,t-1} + b' Z_{1,t-1}\right) - \delta\psi\left(Z_{0,t-1} + b' Z_{1,t-1};\xi\right) - \Phi Z_{2,t-1}. \tag{2.10}$$

Then, given T observations, $X_1, X_2, ..., X_T$, and with the initial values $X_0, \Delta X_0, ..., \Delta X_{-k}$ fixed, the log-likelihood function based on Gaussian errors takes the form,

$$L_T(\theta,\xi,\Omega) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t(\theta,\xi)'\,\Omega^{-1}\varepsilon_t(\theta,\xi). \tag{2.11}$$

We define the corresponding profiled log-likelihood $L_T^*(\theta,\xi) = L_T(\theta,\xi,\Omega^*(\theta,\xi))$ where

$$\Omega^*(\theta,\xi) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\theta,\xi)\,\varepsilon_t(\theta,\xi)',$$

and $\hat{\vartheta}$ is found as,

$$\hat{\vartheta} := (\hat{\theta},\hat{\xi}) = \arg\max_{\theta\in\Theta,\xi\in\Xi} L_T^*(\theta,\xi).$$

As we do not impose any distributional assumptions on the errors, $\hat{\vartheta} = (\hat{\theta},\hat{\xi})$ and $\hat{\Omega} = \Omega^*(\hat{\theta},\hat{\xi})$ are referred to as quasi-maximum likelihood estimators (QMLE's).

2.3 Hypothesis Testing

We are interested in developing inference regarding both short-run (η and ξ) and long-run parameters (β, or b) in the non-linear error correction model. We consider in turn hypotheses involving short- and long-run parameters.

2.3.1 Testing Short-Run Parameters

First, consider the following general hypothesis involving the short-run parameters η and ξ (cf. eq. (2.9)),

$$H_0 : R' vec(\eta,\xi) = 0, \tag{2.12}$$

where R is a known $(m \times d)$-matrix with $d = p(r + d_\delta + pk) + d_\xi$ and $m \leq d$, and we have used the notation $vec(\eta,\xi) = \left[vec(\eta)', vec(\xi)'\right]'$ mentioned in the introduction. Some key examples that are included in the above general formulation include:

Example 1 (Linear error correction) To see if the non-linear components are relevant in explaining the error-correction mechanism, it is of interest to test for their significance. One can do so by testing that there are no nonlinearities in all variables, $R' vec(\eta,\xi) = vec(\delta) = 0$. Alternatively, we may wish to test for presence of non-linear error-correction in individual variables. For example, $R' vec(\eta,\xi) = R_\delta' vec(\delta) = 0$ for some matrix R_δ.

Example 2 (Symmetric response) Suppose that our nonlinear component in eq. (2.3) takes the form

$$\delta\psi\left(z,\xi\right)=\sum_{s=1}^{2}\delta_s\psi_s\left(z,\xi\right),$$

where

$$\psi_s\left(z,\xi\right):=\left[1+\exp\left\{\left(z-\omega_s\right)'A_s\left(z-\omega_s\right)\right\}\right]^{-1}z,\quad s=1,2,$$

such that we have 2 non-linear components in addition to the linear. It is then of interest to test for symmetric responses. That is, $R'vec\left(\eta,\xi\right)=vec\left(\delta_1-\delta_2\right)=0$.

Example 3 (Weak exogeneity) Corresponding to notion of weak exogeneity in linear error correction models with respect to β, we may wish to test for no error correction (neither linear, nor non-linear) in some variables. That is $R'vec\left(\eta,\xi\right)=R'_{\alpha,\delta}\left[\alpha,\delta\right]=0$ for some matrix $R_{\alpha,\delta}$.

Example 4 (# lags) To choose the number of lags included in the model, the following hypothesis is of interest, $R'vec\left(\eta,\xi\right)=vec\left(\Phi_j\right)=0$, for some $j\in\{1,...,k\}$.

Under H_0, some (if not all) parameters in ξ may vanish. One has to check this on a case-by-case basis. One particular case is given in Example 1 where the parameter ξ vanishes under the null of linearity. If this is the case, we face a non-standard testing problem, which is here solved by employing so-called sup-tests. Thus, we treat the two cases (ξ is identified or unidentified under the null) separately:

First, suppose ξ is identified under H_0. In order to test the null, we first obtain the restricted estimator of all parameters, $\vartheta=(\theta,\xi)$, under H_0 which we denote $\tilde{\vartheta}=(\tilde{\theta},\tilde{\xi})$:

$$(\tilde{\theta},\tilde{\xi})=\arg\max_{\substack{\vartheta\\ R'vec(\eta,\xi)=0}}L_T^*\left(\theta,\xi\right).$$

We then propose to test the null by either LR or LM test statistics. The LR statistic compares the log-likelihoods evaluated under the alternative and under the null and is given by

$$LR_T=2\left[L_T^*(\hat{\theta},\hat{\xi})-L_T^*(\tilde{\theta},\tilde{\xi})\right].\tag{2.13}$$

The LM statistic on the other hand, uses the score under the alternative evaluated at the parameter estimates obtained under the null,

$$LM_T=\mathbb{S}_T(\tilde{\theta},\tilde{\xi})'\mathbb{H}_T^{-1}(\tilde{\theta},\tilde{\xi})\mathbb{S}_T(\tilde{\theta},\tilde{\xi}),\tag{2.14}$$

where $\mathbb{S}_T(\theta,\xi)$ and $\mathbb{H}_T(\theta,\xi)$ are the score and Hessian matrices respectively. Here $\mathbb{S}_T(\theta,\xi)$ and $\mathbb{H}_T(\theta,\xi)$ are identified in terms of differentials as introduced in Section 2.4.

Next, in the case where ξ is unidentified under the null of H_0, first note that the parameter restrictions in this case cannot involve ξ since we are unable to test for such. So after removing potentially redundant restrictions involving ξ, the general null in eq. (2.12) can be rewritten as

$$H_0: R'_\eta vec\left(\eta\right)=0,$$

for some matrix R_η. The estimator of $\theta = (b, \eta)$ under the null is given by

$$\tilde{\theta} = \arg \max_{\substack{\theta \in \Theta \\ R'_\eta vec(\eta) = 0}} L^*_T(b, \eta, \xi).$$

On the other hand, under the alternative, we compute a profile estimator of θ for any given value of ξ,

$$\hat{\theta}(\xi) = \arg \max_{\theta \in \Theta} L^*_T(\theta, \xi).$$

The sup-LR test is then obtained by taking supremum of the standard LR test over ξ,

$$\sup LR_T := \sup_{\xi \in \Xi} LR_T(\xi), \tag{2.15}$$

where

$$LR_T(\xi) = 2\left[L^*_T(\hat{\theta}(\xi), \xi) - L^*_T(\tilde{\theta}, \xi)\right]. \tag{2.16}$$

The sup-LM test is obtained in a similar manner,

$$\sup LM_T := \sup_{\xi \in \Xi} LM_T(\xi), \tag{2.17}$$

where

$$LM_T(\xi) = \mathbb{S}_T(\tilde{\theta}(\xi), \xi)' \mathbb{H}_T^{-1}(\tilde{\theta}(\xi), \xi) \mathbb{S}_T(\tilde{\theta}(\xi), \xi). \tag{2.18}$$

2.3.2 Testing Long-Run Parameters

Recall that β is identified by eq. (2.7), so consider the following hypothesis involving the long-run parameter b,

$$H_{0,b} : R'_b vec(b') = 0, \tag{2.19}$$

where R_b is a known $(m \times d)$-matrix with $d = (p - r)r$ and $m \leq d$. A key example is the following:

Example 5 (Cointegrating vectors) Economic theory often imposes, or implies, testable restrictions on the cointegrating relations, for example that they are known. One specific example (with $p = 2$ and $r = 1$) is $\beta = (1, -1)'$ corresponding to the spread between the two variables being stable. In terms of $b \in \mathbb{R}$, this can be expressed as $R'_b vec(b') = b = 0$.

The test statistics are computed in the same way as in the previous subsection with identified ξ. We first compute the restricted estimators which for ease of notation we still call $\tilde{\theta}$ and $\tilde{\xi}$:

$$H_{0,b} : (\tilde{\theta}, \tilde{\xi}) = \arg \max_{\substack{\vartheta \\ R'_b vec(b') = 0}} L^*_T(\theta, \xi).$$

The corresponding LR- and LM-test are then given as:

$$LR_{b,T} = 2\left[L^*_T(\hat{\theta}, \hat{\xi}) - L^*_T(\tilde{\theta}, \tilde{\xi})\right],$$

$$LM_{b,T} = \mathbb{S}_T(\tilde{\theta}, \tilde{\xi})' \mathbb{H}_T^{-1}(\tilde{\theta}, \tilde{\xi}) \mathbb{S}_T(\tilde{\theta}, \tilde{\xi}).$$

2.4 Score and Hessian

As is standard, the analysis of likelihood-based estimators and test statistics will focus on the score and Hessian of the log-likelihood. For ease of notation, we here choose to define them in terms of first and second order differentials of the log-likelihood since parameters enter in the form of matrices; see Magnus and Neudecker (1988) for an introduction to the concept of differentials and their use in econometrics. We apply standard notation and let $dL_T^*(\theta, \xi; d\theta, d\xi)$ denote the first-order differential of $L_T^*(\theta, \xi)$ w.r.t. (θ, ξ) in the direction of $d\theta$ and $d\xi$ respectively. The vector score $\mathbb{S}_T(\theta, \xi) = \partial L_T^*(\theta, \xi) / \partial vec(\theta, \xi)$ can then be identified from the differential through the following identity:

$$dL_T^*(\theta, \xi; d\theta, d\xi) = \mathbb{S}_T(\theta, \xi)' vec(d\theta, d\xi).$$

Similarly, with $d^2 L_T^*(\theta, \xi; d\theta, d\xi; d\theta^*, d\xi^*)$ denoting the second order differential, the Hessian $\mathbb{H}_T(\theta, \xi) = \partial L_T^*(\theta, \xi) / (\partial vec(\theta, \xi) \partial vec(\theta, \xi)')$ is given through the following identity:

$$d^2 L_T^*(\theta, \xi; d\theta, d\xi; d\theta^*, d\xi^*) = vec(d\theta^*, d\xi^*)' \, \mathbb{H}_T(\theta, \xi) vec(d\theta, d\xi).$$

To derive expressions of the first and second order differentials of the log-likelihood, some further notation is needed: First, we introduce the differentials of $\psi(z, \xi) \in \mathbb{R}^{r_\delta}$ with respect to $z \in \mathbb{R}^r$ and $vec(\xi) \in \mathbb{R}^{d_\xi}$ in terms of its partial derivatives,

$$\begin{aligned} d\psi(z, \xi; dz) &= \partial_z \psi(z, \xi) \, dz, \quad \partial_z \psi(z, \xi) = (\partial \psi_i / \partial z_j)_{i,j} \in \mathbb{R}^{r_\delta \times r}, \\ d\psi(z, \xi; d\xi) &= \partial_\xi \psi(z, \xi) \, vec(d\xi), \quad \partial_\xi \psi(z, \xi) \in \mathbb{R}^{r_\delta \times d_\xi}. \end{aligned} \tag{2.20}$$

Furthermore, define the processes $u_t(\xi) \in \mathbb{R}^{p(r + r_\delta + pk)}$, $v_t(\xi) \in \mathbb{R}^r$ and $w_t(\xi) \in \mathbb{R}^r$ by

$$\begin{aligned} u_t(\xi) &:= \left(u_{\alpha,t}(\xi)', u_{\phi,t}(\xi)', u_{\delta,t}(\xi)'\right)', \quad v_t(\xi) := [\delta_0 \partial_\xi \psi(Z_{0,t-1}; \xi)]' \, \Omega_0^{-1} \varepsilon_t(\theta_0, \xi), \quad \text{and} \\ w_t(\xi) &:= [\alpha_0 + \delta_0 \partial_z \psi(Z_{0,t-1}; \xi)]' \, \Omega_0^{-1} \varepsilon_t(\theta_0, \xi), \end{aligned} \tag{2.21}$$

with

$$\begin{aligned} u_{\alpha,t}(\xi) &:= vec\left(\Omega_0^{-1} \varepsilon_t(\theta_0, \xi) \, Z_{0,t-1}'\right), \quad u_{\phi,t}(\xi) := vec\left(\Omega_0^{-1} \varepsilon_t(\theta_0, \xi) \, Z_{2,t-1}'\right) \\ u_{\delta,t}(\xi) &:= vec\left(\Omega_0^{-1} \varepsilon_t(\theta_0, \xi) \, \psi(Z_{0,t-1}; \xi)'\right). \end{aligned} \tag{2.22}$$

These processes prove helpful in the analysis of the score and Hessian of log-likelihood. For example, the first-order differential of $L_T^*(\theta, \xi)$ evaluated at θ_0 can be expressed in terms of these (see Appendix C for details),

$$dL_T^*\ (\theta_0, \xi; d\theta, d\xi) = (vec(d\eta))' \sum_{t=1}^T u_t(\xi) + (vec(d\xi))' \sum_{t=1}^T v_t(\xi) + \sum_{t=1}^T Z_{1,t-1}' (db) \, w_t(\xi).$$

Likewise, the second order differential $d^2 L_T^*\ (\theta_0, \xi; d\theta, d\xi; d\theta^*, d\xi^*)$, or equivalently the Hessian $\mathbb{H}_T$, can be expressed in terms of similar processes based on Z_{0t}, Z_{1t}, Z_{2t} and ε_t in addition to first and second order derivatives of ψ; we refer to Appendix C for an explicit expression.

We then wish to analyze the asymptotic properties of the first- and second order differentials; in particular, in the case of ξ vanishing, weak convergence results for averages based on $u_t(\xi)$, $v_t(\xi)$ and $w_t(\xi)$ need to hold uniformly in ξ. To this end, it proves necessary to develop some new functional central limit theorems. The next section is dedicated to this task.

3 FCLT Results for Nonlinear Processes

In order to obtain the asymptotic distributions of the proposed estimators and test statistics when parameters vanish under the null, we first establish novel functional central limits for double indexed random sequences. The results extend Caner and Hansen (2001) to the case of multivariate processes and parameters, and are of general interest for the statistical analysis of non-linear time series models involving non-stationary components. We therefore develop these in a more general setting, not restricted to the class of non-linear error correction models introduced in the previous section.

Consider a sequence of stochastic processes on the form $(x_T(s), \phi_T(s,\pi))$, where $\pi \in \Pi$ for some compact set $\Pi \subseteq \mathbb{R}^{d_\pi}$ and $s \in [0,1]$. The sequence of stochastic processes, $x_T(s) \in \mathbb{R}^{d_x}$, is given by

$$x_T(s) = x_{[Ts]},$$

for some appropriately normalized random sequence x_t which is assumed to converge weakly, see Assumption 3.3 below. The double-indexed sequence $\phi_T(s,\pi)$ is given as

$$\phi_T(s,\pi) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} f(y_{t-1},\pi)\, e_t \tag{3.1}$$

where $f : \mathbb{R}^{d_y} \times \Pi \mapsto \mathbb{R}^{d_x \times d_e}$, and (e_t, y_t) is a sequence of random variables with $e_t \in \mathbb{R}^{d_e}$. We let $\mathcal{F}_t = \mathcal{F}(e_t, x_t, y_t, e_{t-1}, x_{t-1}, y_{t-1}...)$ denote the filtration with respect to current and past values of (e_t, x_t, y_t). We then wish to establish weak convergence results for transformations of this double-indexed process on the space of cadlag functions $D([0,1] \times \Pi)$.

We impose the following conditions:

Assumption 3.1 *The sequence* (e_t, y_t) *with filtration* $\mathcal{F}_t$ *satisfies:*

(i) (e_t, y_t) *is strictly stationary and geometrically ergodic.*

(ii) e_t *is a martingale difference w.r.t.* $\mathcal{F}_{t-1}$ *such that* $E[e_t|\mathcal{F}_{t-1}] = 0$ *and* $E[e_t e_t'|\mathcal{F}_{t-1}] = \Omega_e$.

Assumption 3.2 *The sequences* $f(y_{t-1};\pi)$ *and* e_t *satisfy for some* $m, n, \lambda > 0$*:*

(i) $E[\sup_{\pi\in\Pi} \|f(y_{t-1},\pi)\|^m] < \infty$ *and* $E[\|e_t\|^m] < \infty$.

(ii) $E\|f(y_{t-1},\pi) - f(y_{t-1},\pi')\|^n \leq C\|\pi - \pi'\|^{\lambda n}$, *for all* $\pi, \pi \in \Pi$.

Assumption 3.3 *The process $x_T(s) = x_{[Ts]}$ on $D([0,1])$ satisfies:*

(i) *As $T \to \infty$, $x_T(s) \xrightarrow{D} x(s)$, where $x(\cdot) \in C[0,1]$.*

(ii) $\sup_t E[\|x_t\|^m] < \infty$ *for some $m > 0$.*

Remark 1: In Assumptions 3.1 (ii) one may instead assume that $E[e_t e_t' | \mathcal{F}_{t-1}] = \Omega_{e,t}$, with $\Omega_{e,t}$ stationary and $E[\|\Omega_{e,t}\|] < \infty$, thereby allowing for conditional heteroskedasticity.

Remark 2: A general sufficient condition for Assumption 3.2 (ii) to hold with $\lambda = 1$ is that $f(\cdot;\cdot)$ is continuously differentiable in π with

$$E\left(\sup_{\pi, d\pi} \|df(y_{t-1}, \pi; d\pi)\|^n\right) < \infty.$$

Theorem 3.4 (FCLT) *Under Assumptions 3.1 and 3.2 with $n = 2$ and $m > \max(4, 2d_\pi/\lambda)$, $\phi_T(s, \pi)$ defined in (3.1) satisfies,*

$$\phi_T(\cdot,\cdot) \xrightarrow{D} \phi(\cdot,\cdot), \tag{3.2}$$

where $\phi(s,\pi)$ is multi-parameter Gaussian process on $C([0,1] \times \Pi) \subseteq D([0,1] \times \Pi)$ with covariance kernel,

$$\Sigma(s_1, \pi_1, s_2, \pi_2) = (s_1 \wedge s_2) E\left[f(y_{t-1}; \pi_1) \Omega_e f(y_{t-1}; \pi_2)'\right] > 0.$$

Remark 3: The condition that $m > \max(4, 2d_\pi/\lambda)$ may lead to high moment requirements on y_t. This "*curse of dimensionality*" stems from the way we establish stochastic equicontinuity, see proof of Theorem 3.4. It may suggest that when considering nonlinear alternatives, one should aim for formulations of $f(\cdot)$ which depend on low-dimensional π. This of course is also well-known from estimation of nonlinear models in general. However, we conjecture that the high moment condition, while sufficient, is not necessary. Caner and Hansen (2001) avoid this type of moment conditions as they focus on the case of a univariate nuisance and so their π is of dimension one by definition.

In the next section, we apply Theorem 3.4 on our model in eq. (2.1) where (in the case of no lagged differences, or $k = 0$), $\pi = \xi$, $y_t = \beta' X_t$, $f(y_{t-1}; \pi) = \psi(\beta' X_{t-1}; \xi)$, and $x_t = K_T^{-1} \kappa_0' X_t$ for some appropriately chosen weighting matrix K_T. In particular for the STECM examples in eq. (2.5) and (2.6), Assumption 3.2 (i) and (ii) hold if $E[\|\beta' X_t\|^m] < \infty$ and $E[\|\beta' X_t\|^n] < \infty$ with $\lambda = 1$. Assumption 3.3 holds for the class of nonlinear error correction models introduced in Section 2 under suitable regularity conditions as shown in Kristensen and Rahbek (2010) and Saikkonen (2005).

In addition to the weak convergence in Theorem 3.4, we also need a convergence result for stochastic integrals in terms of the limiting Gaussian process:

Theorem 3.5 (Convergence to Stochastic Integral) *Under Assumptions 3.1-3.3, with $n = 4$, $m > \max(6, 2d_\pi/\lambda)$, and with $(x_T(\cdot), \phi_T(\cdot, \pi)) \xrightarrow{D} (x(\cdot), \phi(\cdot, \pi))$ for any π, then*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x'_{t-1} f(y_{t-1}, \pi) e_t \xrightarrow{D} \int_0^1 x(s)' d\phi(s, \pi),$$

on $C(\Pi)$.

Note that the equivalent Theorem 2 in Caner and Hansen (2001) does not include the condition of joint pointwise convergence of $(x_T(\cdot), \phi_T(\cdot, \pi))$. However, we establish the above result by verifying the conditions of Theorem 2.2 of Kurtz and Protter (1991), or equivalently Theorem 2.1 of Hansen (1992), which require joint convergence of the two processes. The additional requirement is of little concern in our applications though as we have x_t and y_t defined in terms of the same underlying e_t, and the past of this, and so the joint convergence condition will automatically be satisfied.

Finally, we need convergence of product moment matrices:

Theorem 3.6 *Under Assumptions 3.1-3.3, with $m, n > 1$,*

(i)

$$\frac{1}{T} \sum_{t=1}^{T} x'_{t-1} f(y_{t-1}; \pi) \xrightarrow{D} \int_0^1 x(s)' ds E[f(y_{t-1}; \pi)].$$

(ii)

$$\frac{1}{T} \sum_{t=1}^{T} x'_{t-1} f(y_{t-1}; \pi) x_{t-1} \xrightarrow{D} \int_0^1 x(s)' E[f(y_{t-1}; \pi)] x(s) ds.$$

With s fixed, the convergence in eq. (3.2) to a Gaussian process, holds under much less strict conditions. Likewise if π is fixed when the result reduces to an ordinary FCLT result:

Corollary 3.7 (FCLT) *Under Assumption 3.1 and 3.2 with $m > 1$, $n \geq 2$, then, if either s or π are fixed,*

$$\phi_T(s, \cdot) \xrightarrow{D} \phi(s, \cdot), \quad or \quad \phi_T(\cdot, \pi) \xrightarrow{D} \phi(\cdot, \pi),$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x'_{t-1} f(y_{t-1}, \pi) e_t \xrightarrow{D} \int_0^1 x(s)' d\phi(s, \pi),$$

$$\frac{1}{T} \sum_{t=1}^{T} x'_{t-1} f(y_{t-1}; \pi) \xrightarrow{D} \int_0^1 x(s)' ds E[f(y_{t-1}; \pi)].$$

$$\frac{1}{T} \sum_{t=1}^{T} x'_{t-1} f(y_{t-1}; \pi) x_{t-1} \xrightarrow{D} \int_0^1 x(s)' E[f(y_{t-1}; \pi)] x(s) ds$$

on $C(\Pi) \subseteq D(\Pi)$ and $C[0,1] \subseteq D[0,1]$, respectively.

4 Asymptotics of Estimators and Test Statistics

Given the results of the previous section we are now in position to derive the asymptotic distribution of the QMLE of ϑ, both under the null hypothesis of interest and the alternative. The results are used when studying the asymptotics of both the likelihood ratio test statistic and Lagrange multiplier test for general null hypotheses, including the hypothesis of linearity, $\delta_0 = 0$. Furthermore, the results generalize the distributional results of Kristensen and Rahbek (2010) to include the case of asymmetric adjustments in nonlinear error correction models.

4.1 Asymptotics of the QMLE

We start by a list of assumptions on the processes in the score and Hessian:

Assumption 4.1 *The function $\psi(z,\xi)$ is four times differentiable in z and ξ. The function itself and its derivatives are polynomially bounded in z of order $\rho \geq 1$ uniformly over ξ, $\|\psi(z,\xi)\| \leq C(1+|z|^{\rho})$ for some $C > 0$.*

Assumption 4.2 *The processes $(Z_{0t}', Z_{2t}')'$ are stationary and geometrically ergodic with $E[\|Z_{0,t-1}\|^{q_0}] < \infty$ and $E[\|Z_{2,t-1}\|^{q_2}] < \infty$ for some $q_0, q_2 \geq 1$.*

Assumption 4.3 *With κ_0 the $(p \times (p-r))$ dimensional normalization matrix in (2.7), the $(p-r)$ dimensional non-stationary process $\kappa_0' X_t$ satisfies*

$$K_T^{-1}\kappa_0' X_{[Ts]} = K_T^{-1} Z_{1,[Ts]} \xrightarrow{D} F(s),$$

on $s \in [0,1]$, where the process $F(s)$ is a.s. continuous and K_T is a $((p-r)\times(p-r))$ dimensional diagonal matrix for which $K_T^{-1} \to 0$ as $T \to \infty$.

Assumption 4.4 *The parameter space Ξ for ξ is compact.*

Assumption 4.1 rules out threshold models, but these can be approximated up to any degree of precision by a smooth transition model, see also Seo (2010). Otherwise, all proposed specifications of nonlinear error correction satisfy this assumption.

Sufficient conditions for Assumptions 4.2-4.3 for particular specifications of ψ can be found in Bec and Rahbek (2004), Kristensen and Rahbek (2010) and Saikkonen (2005, 2008) amongst others. In particular, they give conditions for the already mentioned STECM, see eqs. (2.5) and (2.6). Note in this respect that Assumptions 4.2 can be replaced by the assumption that,

$$\left(Z_{0t}', ..., Z_{0t-k}', Z_{2t}'\beta_{0\perp}\right) = \left(X_t'\beta_0, ..., X_{t-k}'\beta_0, \Delta X_t'\beta_{0\perp}, ..., , \Delta X_{t-k}'\beta_{0\perp}\right)$$

is a geometrically ergodic Markov chain with drift function $V(y) = 1 + \|y\|^{2q}$, $q > 2$. This way, one is not required to have the initial values of the observations drawn from the invariant distribution, as for example the law of large numbers, and hence the central limit theorem, hold irrespectively of the choice of initial values, see Jensen and Rahbek (2007).

In Assumption 4.3, κ_0 can be used to decompose X_t into trends of different orders. In particular, as demonstrated in Kristensen and Rahbek (2010), when ψ is symmetric the nonlinear error-correction process with $X_t \in \mathbb{R}^p$ has $p-r-1$ common stochastic trends, while there is at most one linear trend. Thus, within their class of models, Assumption 3.3 holds with $F(s)$ being a $(p-r-1)$-dimensional Brownian motion, and a linear trend component. In the general case where symmetry is not imposed, there are at most $p-r$ stochastic trends but the exact number depends on the specific form of ψ; see Saikkonen (2008, p. 308).

As a first step towards establishing the properties of the QMLE's under the null and alternative, we analyze the behaviour of $(u_t(\xi), v_t(\xi), w_t(\xi))$ and X_t where $u_t(\xi)$, $v_t(\xi)$ and $w_t(\xi)$, as defined in (2.21)-(2.22), are the sequences that make up the score and Hessian of the log-likelihood. By applying the general results of Theorem 3.4, we obtain the following uniform FCLT over $(s,\xi) \in [0,1] \times \Xi$:

Lemma 4.5 *Under Assumptions 4.1-4.4 with $q_2 = \max(4, 2d_\xi)$ and $q_0 > q_2 \max(1,\rho)$ given in Assumption 4.2, it holds, as $T \to \infty$,*

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} u_t(\xi)', \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} v_t(\xi)', \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} w_t(\xi)', \left(K_T\kappa_0' X_{[Ts]}\right)'\right) \tag{4.1}$$
$$\stackrel{D}{\to} \left(B_u'(s,\xi), B_v'(s,\xi), B_w'(s,\xi), F'(s)\right)$$

on the function space $[0,1] \times \Xi$. Here F is defined in Assumption 4.3, while B_u, B_v and B_w are Gaussian processes with covariance kernel, $(s_1 \wedge s_2)\Sigma(\xi_1,\xi_2)$ where

$$\Sigma(\xi_1,\xi_2) := Cov\left(\begin{pmatrix} u_t(\xi_1) \\ v_t(\xi_1) \\ w_t(\xi_1)\end{pmatrix}, \begin{pmatrix} u_t(\xi_2) \\ v_t(\xi_2) \\ w_t(\xi_2)\end{pmatrix}\right) := \begin{pmatrix} \Sigma_{(u,v),(u,v)}(\xi_1,\xi_2) & \Sigma_{(u,v),w}(\xi_1,\xi_2) \\ \Sigma_{w,(u,v)}(\xi_1,\xi_2) & \Sigma_{w,w}(\xi_1,\xi_2)\end{pmatrix}. \tag{4.2}$$

The above result will be used to establish (uniform) weak convergence of the score and Hessian of the log-likelihood. The above result is stated uniformly over ξ, which is needed for the asymptotics of the sup statistics when ξ vanishes under the null. In all other situations, we only need the above convergence to hold pointwise at $\xi = \xi_0$. In particular, in the maintained model and under nulls where ξ does not vanish, we can fix ξ at ξ_0 and then apply Corollary 3.7 instead of Theorem 3.4. This in turn allows us to weaken the moment conditions in Lemma 4.5 to $q_0 > 2\max\{\rho, 1\}$ and $q_2 > 2$ when establishing weak convergence of the QMLE and test statistics.

In order to state the asymptotic distribution of the QMLE, define the matrix of convergence rates,

$$V_T^{1/2} = \operatorname{diag}\left(V_{\theta,T}^{1/2}, V_{\xi,T}^{1/2}\right), \quad \text{where } V_{\theta,T}^{1/2} = \operatorname{diag}\begin{pmatrix} I_r \otimes K_T & I_{p(r+r_\delta+pk)}\end{pmatrix} \text{ and } V_{\xi,T}^{1/2} = I_{d_\xi}. \tag{4.3}$$

Here, $V_{\theta,T}$ and $V_{\xi,T}$ contain the rates for the QMLE of θ and ξ respectively. Again, we single out ξ to be able to handle the case of this parameter vanishing.

We now state two separate results for the QMLE: First, we consider the situation where $\delta_0 \neq 0$, and then when $\delta_0 = 0$.

Theorem 4.6 *Suppose that Assumptions 4.1-4.4 hold with $q_0 > 2\max\{1,\rho\}$ and $q_2 > 2$ and $\delta_0 \neq 0$, and that $\Sigma(\xi_0,\xi_0) > 0$ as given in eq. (4.2). Then the following holds: With probability tending to one, there exists a unique minimum point $\hat{\vartheta} = (\hat{\theta},\hat{\xi}) = (\hat{b}',\hat{\eta},\hat{\xi})$ of $L_T^*(\vartheta)$ in the neighbourhood $\{\vartheta : ||\eta-\eta_0|| < \epsilon,\ ||\xi-\xi_0|| < \epsilon$ and $||K_T b|| < \epsilon\}$ for some $\epsilon > 0$. Moreover, with V_T defined in eq. (4.3),*

$$T^{1/2}V_T^{1/2}vec\left(\hat{\vartheta}-\vartheta_0\right) \xrightarrow{D} \mathbb{H}^{-1}\mathbb{S}, \tag{4.4}$$

for a random matrix $\mathbb{H}$ and vector $\mathbb{S}$, given by

$$\mathbb{H} \equiv \begin{pmatrix} \int_0^1 F(s)F(s)'ds \otimes \Sigma_{w,w}(\xi_0,\xi_0) & \int_0^1 F(s)\,ds \otimes \Sigma_{w,(u,v)}(\xi_0,\xi_0) \\ \int_0^1 F(s)'ds \otimes \Sigma_{(u,v),w}(\xi_0,\xi_0) & \Sigma_{(u,v),(u,v)}(\xi_0,\xi_0) \end{pmatrix}, \tag{4.5}$$

and

$$\mathbb{S} \equiv \left(vec\left(\int_0^1 F(s)\,dB_w'(s,\xi_0)\right)',\ B_u'(1,\xi_0),\ B_v'(1,\xi_0)\right)'. \tag{4.6}$$

Finally, note that $\hat{\Omega} \xrightarrow{P} \Omega_0$.

The above result, where $\delta_0 \neq 0$, is an extension of results in Kristensen and Rahbek (2010) as we allow for asymmetry in the error correction as given by the $\psi(\cdot)$ function. Rather than establishing the conditions of Kristensen and Rahbek (2010, Lemmas 11 and 12), we use the more general formulation of Lemmas D.1 and D.2 in Appendix D which allow us also to consider convergence uniformly in ξ. The asymptotic distribution is akin to ones derived in de Jong (2001, 2002) and Kristensen and Rahbek (2010) in the sense that the short- and long-run parameter estimators are not asymptotically independent (as is the case in linear error-correction models). The results in Theorem 4.6 complement the ones of Seo (2010) who derive the asymptotics of estimators based on smoothed likelihood-functions in discontinuous threshold error correction models.

The assumption that $\Sigma(\xi_0,\xi_0) > 0$ is an identification condition that ensures that the limiting distributions of the QMLE is non-degenerate. It proves difficult to give primitive conditions for this to hold. This is a general problem in nonlinear models, where identification has to be verified on a case by case basis, see e.g. Kristensen and Rahbek (2009) and Meitz and Saikkonen (2009).

Next, we examine the behaviour of the QMLE under the null where $\delta_0 = 0$ such that ξ is not identified, or "vanishes". Thus, we state a result that holds uniformly over ξ which we need for the asymptotic analysis of the $\sup LR$-test.

Theorem 4.7 *Suppose that Assumptions 4.1-4.4 hold with* $q_2 = \max(4, 2d_\xi)$ *and* $q_0 > q_2 \max(1, \rho)$ *and* $\delta_0 = 0$*, and that* $\Sigma(\xi_1, \xi_1) > 0$ *for all* $\xi_1, \xi_2 \in \Xi$*, where* $\Sigma(\xi_1, \xi_1)$ *is given in eq. (4.2). Then the following hold uniformly over* ξ*: With probability tending to one, there exists a unique minimum point* $\hat{\theta}(\xi) = (\hat{b}(\xi)', \hat{\eta}(\xi))$ *of* $L_T^*(\theta, \xi)$ *in the neighbourhood* $\{\theta : ||\eta - \eta_0|| < \epsilon$ *and* $||K_T b|| < \epsilon\}$ *for some* $\epsilon > 0$*. Moreover, with* $V_{\theta,T}$ *defined in eq. (4.3),*

$$T^{1/2} V_{\theta,T}^{1/2} vec\left(\hat{\theta}(\xi) - \theta_0\right) \xrightarrow{D} \mathbb{H}_{\theta\theta}^{-1}(\xi)\, \mathbb{S}_\theta(\xi), \tag{4.7}$$

for a random matrix $\mathbb{H}_{\theta\theta}(\xi)$ *and vector* $\mathbb{S}_\theta(\xi)$*, given by*

$$\mathbb{H}_{\theta\theta}(\xi) \equiv \begin{pmatrix} \int_0^1 F(s) F(s)' ds \otimes \Sigma_{w,w} & \int_0^1 F(s)\, ds \otimes \Sigma_{w,u}(\xi,\xi) \\ \int_0^1 F(s)' ds \otimes \Sigma_{u,w}(\xi,\xi) & \Sigma_{u,u}(\xi,\xi) \end{pmatrix}, \tag{4.8}$$

and

$$\mathbb{S}_\theta(\xi) \equiv \left(vec\left(\int_0^1 F(s)\, dB_w'(s) \right)', B_u(1,\xi)' \right)'. \tag{4.9}$$

We note that under the null, the DGP is a standard linear error correction model such that, under the usual I(1) conditions of Johansen (1996), Assumptions 4.2 and 4.3 hold with $F(s)$ being a Brownian motion with covariance matrix $\Sigma_{F,F} = \bar{\beta}_{0,\perp}' C_0 \Omega_0 C' \bar{\beta}_{0,\perp}$, where $C_0 := \beta_{0,\perp}\left(\alpha_{0,\perp}'\left(I - \sum_{i=1}^k \Phi_{0,i}\right)\beta_{0,\perp}\right)^{-1}\alpha_{0,\perp}'$, while $B_u(s,\xi) = \left(B_\alpha(s)', B_\phi(s)', B_\delta(s;\xi)'\right)'$. Also, again due to the model collapsing to a standard I(1) model, the expressions of the variables and parameters entering $\mathbb{S}_\theta(\xi)$ and $\mathbb{H}_{\theta\theta}(\xi)$ above simplify: The process $B_u(s,\xi)$ becomes $B_u(s,\xi) = \left(B_\alpha(s)', B_\phi(s)', B_\delta(s;\xi)'\right)'$ and $B_w(s,\xi) = B_w(s)$ where $B_\alpha(s)$, $B_\phi(s)$ and $B_\delta(s;\xi)$ are the Brownian motions corresponding to the variables $u_{\alpha,t}$, $u_{\phi,t}$ and $u_{\delta,t}$ in eq. (2.21). Here, only $B_\delta(s;\xi)$ depends on ξ since $u_{\alpha,t} = vec\left(\Omega_0^{-1}\varepsilon_t Z_{0,t-1}'\right)$, $u_{\phi,t} = vec\left(\Omega_0^{-1}\varepsilon_t Z_{2,t-1}'\right)$ and $w_t = \alpha_0'\Omega_0^{-1}\varepsilon_t$ under the null. Thus, $F(s)$ is independent of the processes $(B_\alpha(s), B_\phi(s))$ and $B_w(s)$, but is still dependent of $B_\delta(s,\xi)$ and hence of $B_u(s,\xi)$. Finally, the remaining covariances are: $\Sigma_{w,w} = \alpha_0'\Omega_0^{-1}\alpha_0$ and

$$\Sigma_{w,u_\alpha} = E\left[\left(\alpha_0\Omega_0^{-1}\varepsilon_t\right)\left(vec\left(\Omega_0^{-1}\varepsilon_t Z_{0,t-1}'\right)\right)'\right] = E\left[Z_{0,t-1} \otimes I\right]\Omega_0^{-1}\alpha_0 = 0,$$

$$\Sigma_{w,u_\phi} = E\left[\left(\alpha_0\Omega_0^{-1}\varepsilon_t\right) vec\left(\Omega_0^{-1}\varepsilon_t Z_{2,t-1}'\right)\right] = E\left[Z_{2,t-1} \otimes I\right]\Omega_0^{-1}\alpha_0 = 0.$$

4.2 Asymptotics of test statistics

In this section we derive the asymptotic distributions of the tests proposed in Section 2.3. We treat separately the case where ξ is identified and vanishes under the null. We discuss specific examples below.

First, consider the case where ξ is unidentified in which case we employ the sup-Lagrange Multiplier (LM) test and sup-Likelihood Ratio (LR) tests introduced in eqs. (2.15)-(2.16) and (2.17)-(2.18). As noted in Section 2, the null in this case can be written as $H_0 : R_\eta' vec(\eta) = 0$. We then show in the appendix that the restricted estimator satisfies

$$\sqrt{T} V_{\theta,T}^{1/2} vec(\tilde{\theta} - \theta_0) \xrightarrow{D} M_\eta \tilde{\mathbb{H}}_{\theta\theta}^{-1} \tilde{\mathbb{S}}_\theta, \tag{4.10}$$

where

$$\tilde{\mathbb{H}}_{\theta\theta} := M_\eta' \mathbb{H}_{\theta\theta}(\xi) M_\eta\big|_{R_\eta' vec(\eta)=0}, \quad \tilde{\mathbb{S}}_\theta := M_\eta' \mathbb{S}_\theta(\xi)\big|_{R_\eta' vec(\eta)=0}, \tag{4.11}$$

with $M_\eta = \operatorname{diag}\left(I_{(p-r)r}, (R_\eta)_\perp\right)$, while $\mathbb{S}_\theta(\xi)$ and $\mathbb{H}_{\theta\theta}(\xi)$ are defined in Theorem 4.7. Note here, that $\tilde{\mathbb{H}}_{\theta\theta}$ and $\tilde{\mathbb{S}}_\theta$ are independent of ξ as the restriction $R_\eta' vec(\eta_0) = 0$ through M_η removes the components of $\mathbb{S}_\theta(\xi)$ and $\mathbb{H}_{\theta\theta}(\xi)$ that depend on ξ.

The asymptotic distribution of the restricted estimators when ξ is identified is shown to be

$$\sqrt{T} V_{\theta,T}^{1/2} vec(\hat{\vartheta} - \vartheta_0) \xrightarrow{D} M \tilde{\mathbb{H}}^{-1} \tilde{\mathbb{S}}, \tag{4.12}$$

where

$$\tilde{\mathbb{H}} := M' \mathbb{H} M\big|_{R' vec(\eta,\xi)=0}, \quad \tilde{\mathbb{S}} := M' \mathbb{S}\big|_{R' vec(\eta,\xi)=0}, \tag{4.13}$$

and $M = \operatorname{diag}\left(I_{(p-r)r}, R_\perp\right)$, while $\mathbb{S}$ and $\mathbb{H}$ defined in Theorem 4.6. The following result is then shown in the Appendix:

Theorem 4.8 *Suppose Assumptions 4.1-4.4 and $H_0 : R' vec(\eta, \xi) = 0$ hold.*

1. *If ξ_0 is identified under the null, then with $q_0 > 2\max\{1, \rho\}$ and $q_2 > 2$,*

   $$LM_T \xrightarrow{D} \mathbb{V}'\mathbb{V}, \quad LR_T \xrightarrow{D} \mathbb{V}'\mathbb{V},$$

   where

   $$\mathbb{V} := \left(M' \mathbb{H}^{-1} M\right)^{-1/2} M' \mathbb{H}^{-1} \mathbb{S},$$

   with $\mathbb{S}$ and $\mathbb{H}$ given in Theorem 4.6.

2. *If ξ is not identified under the null, then with $q_2 = \max(4, 2d_\xi)$ and $q_0 > q_2 \max(1, \rho)$,*

   $$\sup_{\xi \in \Xi} LM_T(\xi) \xrightarrow{D} \sup_{\xi \in \Xi} \mathbb{V}_\theta(\xi)' \mathbb{V}_\theta(\xi), \quad \sup_{\xi \in \Xi} LR_T(\xi) \xrightarrow{D} \sup_{\xi \in \Xi} \mathbb{V}_\theta(\xi)' \mathbb{V}_\theta(\xi),$$

   where

   $$\mathbb{V}(\xi) := \left(M' \mathbb{H}_{\theta\theta}^{-1}(\xi) M\right)^{-1/2} M' \mathbb{H}_{\theta\theta}^{-1}(\xi) \mathbb{S}_\theta(\xi).$$

   with $\mathbb{S}_\theta(\xi)$ and $\mathbb{H}_{\theta\theta}(\xi)$ given in Theorem 4.7.

Now consider the special case when $E[\psi(Z_{0,t-1}; \xi)] = 0$ which, for example, is satisfied if $\psi(Z_{0,t-1}; \xi)$ is symmetric around zero. In this case, $\Sigma_{w,u}(\xi, \xi) = 0$, such that

$$\mathbb{H}_{\theta\theta}^{-1}(\xi) = \begin{pmatrix} \left[\int_0^1 F(s) F(s)' ds \otimes \Sigma_{w,w}\right]^{-1} & 0 \\ 0 & \Sigma_{u,u}^{-1}(\xi, \xi) \end{pmatrix}.$$

In this case, $\xi \mapsto \mathbb{V}(\xi)$ is a Gaussian process and the limiting distributions of $\sup_{\xi \in \Xi} LM_T(\xi)$ and $\sup_{\xi \in \Xi} LR_T(\xi)$ are as in the stationary case reported in Hansen (1996). In particular, the asymptotic distributions correspond to eq. (C_n^*) in Hansen and Seo (2001, p. 317) who assume $E[\psi(Z_{0.t}; \xi)] = 0$, and hence avoid the contribution from the non-stationary component. Observe however that $E[\psi(Z_{0.t}; \xi)] = 0$ does not necessarily hold, even when the DGP is indeed a linear process. Thus, $E[\psi(Z_{0.t}; \xi)] \neq 0$ in general, and so the limiting distribution reported here is different from the one of Hansen and Seo (2001).

The general result with $E\left[\psi\left(Z_{0.t};\xi\right)\right]\neq 0$ is similar to the results for the sup-Wald test for linearity in threshold unit root models derived in Caner and Hansen (2001) (see also Pitarakis, 2008, Proposition 2). There, the limiting distribution also has two components: One is due to the stationary components of the process (in our case $(Z_{0,t-1}, Z_{2,t-1}, \psi\left(Z_{0,t-1};\xi\right))$ with corresponding score vector $(\mathbb{S}_{\alpha}(\xi), \mathbb{S}_{\Phi}(\xi), \mathbb{S}_{\delta}(\xi)))$ and one due to the non-stationary component (in our case $Z_{1,t-1}$ with corresponding score vector $\mathbb{S}_{b}(\xi)$) The presence of the non-stationary component is due to the fact that b is unknown, and so has to be estimated.

Thus, our result demonstrates that in general one cannot ignore the fact that b is estimated as opposed to known. This is in contrast to, for example, Kilic (2009) who assumes that b is known, and thereby avoid the non-stationary component in the limiting distribution of his sup-Wald test for linearity in error-correction models. Similarly, Nedeljkovic (2008) derives the limiting distribution for a sup-LM test for linearity under the implicit assumption that the estimation error arising from $\tilde{b}$ can be ignored. In both papers, the limiting distribution becomes a supremum over a squared Gaussian process as when $E\left[\psi\left(Z_{0.t};\xi\right)\right]=0$.

The problem of vanishing parameters under the null also appears when the non-linear component takes the form $\delta\psi\left(z,\xi\right)=\sum_{s=1}^{S}\delta_{s}\psi_{s}\left(z,\xi_{s}\right)$ and one wishes to test the hypothesis $\bar{H}_{0}:\delta_{s_0}=0$ for some $s_0\in\{1,...,S\}$. Here, the parameter ξ_{s_0} vanishes under the null. One can easily apply the same arguments as used above to derive the asymptotics of sup-test statistics corresponding to this hypothesis where the supremum is now taken over ξ_{s_0}.

Example 2 (continued) Under the null hypothesis of $H_0:\delta=\delta_0=0$, our model collapses to a standard linear cointegrating error-correction model with implications discussed after Theorem 4.7. In particular, the restricted estimator, $\tilde{\theta}=(\tilde{b}',\tilde{\alpha},\tilde{\Phi},\tilde{\delta})$, where $\tilde{\delta}=0$, is the standard Johansen Gaussian MLE. From Theorem 4.7 with $\delta_0=0$ (or alternatively, Johansen, 1996), we obtain that

$$\sqrt{T}V_{\theta,T}^{1/2}vec(\tilde{\theta}-\theta_0)\stackrel{D}{\rightarrow}R_{\perp}\tilde{\mathbb{H}}_{\theta\theta}^{-1}\tilde{\mathbb{S}}_{\theta}, \tag{4.14}$$

where

$$\tilde{\mathbb{H}}_{\theta\theta}\equiv\begin{pmatrix}\int_0^1 F\left(s\right)F\left(s\right)'ds\otimes\Sigma_{w,w} & 0\\ 0 & \Sigma_{\alpha,\phi}\end{pmatrix}, \tag{4.15}$$

and

$$\tilde{\mathbb{S}}_{\theta}\left(\xi\right)\equiv\left(vec\left(\int_0^1 F\left(s\right)dB_w'\left(s\right)\right)', B_{\alpha,\phi}\left(1\right)'\right)'. \tag{4.16}$$

where $B_{\alpha,\phi}\left(s\right)=\left(B_{\alpha}\left(s\right)',B_{\phi}\left(s\right)'\right)'$ contain the components of $B_u\left(s\right)$ corresponding to $u_{\alpha,t}$ and $u_{\phi,t}$ defined in eq. (2.22). The covariance structure of $F\left(s\right)$, $B_{\alpha,\phi}\left(s\right)$ and $B_w\left(s\right)$ is as discussed after Theorem 4.7.

Next, we derive tests for the hypothesis $H_{0,b}$ involving the cointegration relations, $H_{0,b}$: $R_b'vec\left(b'\right)=0$ or, equivalently, $H_{0,b}:vec\left(b'\right)=\left(R_b\right)_{\perp}\tau$ for some free parameter τ. The proof strategy is identical to the one employed in Theorem 4.8 and so we state the result without proof:

Theorem 4.9 *Suppose Assumptions 4.1-4.4 with $q_0 > 2\max\{1,\rho\}$ and $q_2 > 2$, and $H_{0,b}$: $R_b' vec(b') = 0$ hold. Then the LR and LM test of this hypothesis satisfies*

$$LM_{b,T} \xrightarrow{D} \mathbb{V}_b'\mathbb{V}_b, \quad LR_{b,T} \xrightarrow{D} \mathbb{V}_b'\mathbb{V}_b,$$

where

$$\mathbb{V}_b := \left(M_b'\mathbb{H}^{-1}M_b\right)^{-1/2} M_b'\mathbb{H}^{-1}\mathbb{S},$$

with $\mathbb{S}$ and $\mathbb{H}$ given in Theorem 4.6 and $M_b = \operatorname{diag}\left(I_{(p-r)r}, (R_b)_\perp\right)$

Note that the we here avoid any of the complications normally found in the literature on tests involving cointegration relations such as Johansen (1992, Theorem C.1) and Rahbek, Kongsted and Jørgensen (1999, Appendix B). In these and other studies, one formulates the hypotheses in terms of β; this has as consequence that one has to rotate the coordinate system of the free parameter τ in such a way that $(R_b)_\perp' Z_{1,t}$ has a well-behaved asymptotic distribution. In contrast, since we write the hypothesis $H_{0,b}$ in terms of the normalized parameter b, we avoid this problem here.

5 Bootstrap Procedure

In order to draw inference for the parameters, we need to be able to evaluate the limiting distributions in Theorems 4.6-4.9. These are highly non-standard and so we here propose to use bootstrapping in their implementation.

We here consider a bootstrap procedure similar to the one analyzed in Cavaliere, Rahbek and Taylor (2010a,b). First, consider bootstrapping the distributions of the sup-LR and sup-LM tests. We bootstrap under the null of $\delta_0 = 0$ in which case the model is a standard linear error-correction model. With $\tilde{\theta}$ denoting the restricted estimator, we first compute

$$\Delta X_t^* = \tilde{\alpha}\tilde{\beta}' X_{t-1}^* + \tilde{\Phi}\left(\Delta X_{t-1}^{*\prime}, ..., \Delta X_{t-k}^{*\prime}\right)' + \varepsilon_t^*, \quad t = 1, ..., T, \tag{5.1}$$

where, as in Cavaliere et al (2010a,b), the resampled errors ε_t^* are generated using the so-called Wild bootstrap. That is, $\varepsilon_t^* := \hat{\varepsilon}_t\omega_t$, where ω_t is i.i.d. $N(0,1)$ and $\hat{\varepsilon}_t$, $t = 1, ..., T$, are the residuals obtained under the alternative,

$$\hat{\varepsilon}_t := \Delta X_t - \hat{\alpha}\hat{\beta}' X_{t-1} - \hat{\delta}\psi\left(\hat{\beta}' X_{t-1}; \hat{\xi}\right) - \hat{\Phi}\left(\Delta X_{t-1}', ..., \Delta X_{t-k}'\right)', \quad t = 1, ..., T. \tag{5.2}$$

If $\hat{\delta} = 0$, we fix $\hat{\xi}$ at an arbitrary fixed value, say $\bar{\xi}$, chosen by the econometrician. Instead of using the residuals obtained under the alternative, one could use the ones obtained under the null. However, these would not be consistent under the alternative, in which case the bootstrap procedure would therefore not yield a consistent estimate of the distribution of interest.

Given the bootstrap sample X_t^*, $t = 1, ..., T$, we then compute the sup-LR and the sup-LM test statistics with the bootstrap sample replacing the original one; let $\sup_{\xi\in\Xi} LM_T^*(\xi)$ and $\sup_{\xi\in\Xi} LR_T^*(\xi)$ denote the resulting statistics. Computing, say, N, bootstrap samples, we obtain N realizations of the test statistics, and we use their empirical distributions to compute critical values.

In order to show that the above procedure is consistent under the null, we need to establish that Lemma 4.5 holds for the bootstrap sample. As a first step towards showing this, we note that Cavaliere et al (2010a, Lemma A.4) can be employed to show that X_t^* has the representation,

$$X_t^* = \tilde{C} \sum_{i=0}^{t} \varepsilon_{t-i}^* + \sqrt{T} R_t^*, \tag{5.3}$$

where $\tilde{C} = \tilde{\beta}_\perp \left(\tilde{\alpha}_\perp' \left(I - \sum_{i=1}^k \tilde{\Phi}_i \right) \tilde{\beta}_\perp \right)^{-1} \tilde{\alpha}_\perp'$, $\sup_{1 \le t \le T} R_t^* = o_{P^*}(1)$ and P^* denotes the bootstrap probability measure conditional on data $\{X_t\}$. Moreover, $\sum_{i=0}^t \varepsilon_{t-i}^*$ satisfies an FCLT under P^*, cf. Cavaliere et al (2010a, Lemma A.5). What remains to be shown is that the remaining terms in Lemma 4.5 also satisfies a FCLT under P^*, which in turn then could be utilized to verify that Lemmas C.1-C.3 remain valid weakly in probability for the bootstrap sample. We leave the theoretical proof of this last part for future research, and instead verify the validity of the bootstrap procedure through simulations.

6 A Simulation Study

We here investigate some finite-sample properties of the proposed LR-based tests in a specific example of the smooth transition error correction model (STECM) as given by,

$$\Delta X_t = g\left(\beta' X_{t-1}\right) + \Phi \Delta X_{t-1} + \varepsilon_t, \quad g\left(\beta' X_{t-1}\right) = \alpha \beta' X_{t-1} + \delta \psi\left(\beta' X_{t-1}; \xi\right). \tag{6.1}$$

We consider the bivariate case, $p = 2$, with $r = 1$ cointegrating relations, and with $S = 1$ symmetric nonlinear component on the form given in eq. (2.5),

$$\psi(z, \xi) = \left[1 + \exp\left\{(z - \omega)' A (z - \omega)\right\}\right]^{-1} z, \quad \xi = (A, \omega).$$

We are interested in the following two hypotheses: The first hypothesis of interest is the one of linearity in both components, $H_R^{(1)} : \delta = (\delta_1, \delta_2)' = (0, 0)'$; in this case, ξ vanishes under the null, and we have to employ the sup-LR test. The second hypothesis examines whether the spread is stable, $H_R^{(2)} : \beta = (1, b)' = (1, -1)'$, such that in this case the parameter ξ does *not* vanish under the null.

We wish to analyze the performance of the bootstrapped tests under the null (empirical size) as well as under the alternative (empirical power, or rejection probabilities). Under the respective nulls ($H_R^{(k)}$ for $k = 1, 2$) and the corresponding alternatives, the data-generating parameters were chosen to match estimates obtained by fitting the corresponding linear and non-linear models to the bivariate term structure data considered in Bec and Rahbek (2004)[1]. All parameter values used to simulate under the nulls and alternative are given in Appendix E, and we choose the errors to be i.i.d. normally distributed. Note that Assumption 4.2 and 4.3 hold for the parameters chosen under the nulls and alternative employed.

For the implementation of the (sup) LR tests, we compute the QMLE's under the null and alternative as described below. For the bootstrap we use the set-up in eq. (5.1). In terms

[1] Note that, for this particular data set, Bec and Rahbek (2004), treating β as known, used conventional LR-tests to conclude that $H_R^{(2)}$ was accepted

of notation, as previously defined in eq. (2.9), set $\vartheta = (\theta, \xi) = (\beta, \eta, \xi)$, with $\theta := (b, \eta)$, $\eta := (\alpha, \delta, \Phi) \in \mathbb{R}^{2\times(2+2)}$, $\xi = (A, \omega) \in \mathbb{R}^2$ and $\beta = (1, b)'$.

We first discuss the practical implementation of the $\sup LR_T$ test statistic for linearity as given in eqs. (2.17)-(2.16): Under the null of $H_R^{(1)}$ the QMLE's $\tilde{\theta} = (\tilde{\beta}, \tilde{\eta})$ are standard, see Johansen (1996), and $L_T^*(\tilde{\theta}) = -\frac{T}{2}\log|\hat{\Omega}^*(\tilde{\theta})|$, with

$$\hat{\Omega}^*(\tilde{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\tilde{\theta})\varepsilon_t(\tilde{\theta})'.$$

Under the alternative $H_A^{(1)}$, that is with (6.1) unrestricted, write the model on compact form as,

$$\Delta X_t = \eta' W_{t-1}(\beta, \xi) + \varepsilon_t, \quad W_t(\beta, \xi) = \left(X_{t-1}'\beta, \psi\left(\beta' X_{t-1}; \xi\right), Z_{2,t-1}'\right)' \in \mathbb{R}^{2r+pk}.$$

Observe that profile estimators of η and Ω are given by standard OLS estimation,

$$\hat{\eta}(\beta, \xi) = \left(\textstyle\sum_{t=1}^{T} W_t(\beta, \xi) W_t(\beta, \xi)'\right)^{-1}\left(\textstyle\sum_{t=1}^{T} W_t(\beta, \xi)\Delta X_t'\right), \quad \text{and} \tag{6.2}$$

$$\hat{\Omega}^*(\beta, \xi) = \frac{1}{T}\textstyle\sum_{t=1}^{T}\hat{\varepsilon}_t(\beta, \xi)\hat{\varepsilon}_t(\beta, \xi)', \quad \hat{\varepsilon}_t(\beta, \xi) = \Delta X_t - \hat{\eta}(\beta, \xi)' W_{t-1}(\beta, \xi). \tag{6.3}$$

Given these estimators, we can in turn estimate β for fixed ξ,

$$\hat{\beta}(\xi) = \arg\min_{b\in\mathbb{R}} \log(|\hat{\Omega}^*(\beta, \xi)|),$$

and finally $\sup LR_T$ is computed as,

$$\sup LR_T := T\sup_{\xi\in\Xi}(\log|\hat{\Omega}^*(\tilde{\theta})| - \log|\hat{\Omega}^*(\hat{\theta}(\xi))|).$$

For the particular parameterization, we here choose $\Xi = \{(A, b) : 0 \leq A \leq 1 \text{ and } -1 \leq b \leq 1\}$, and then computed the sup test in practice by evaluating $\log|\Omega^*(\tilde{\theta})| - \log|\Omega^*(\hat{\theta}(\xi))$ on a discrete uniform grid of size 50×50 over Ξ, and then simply choosing the maximum value as an approximation of $\sup LR_T$.

Next, consider the LR_T statistic for testing $H_R^{(2)}$ or stability of the spread: Under both null and alternative, we proceed as before and first use OLS to obtain profile estimates $\hat{\eta}(\beta, \xi)$ and $\hat{\Omega}^*(\beta, \xi)$. Next, under the null $H_A^{(2)}$, $\tilde{\beta} = (1, -1)'$ and $\tilde{\xi} := \arg\min_\xi \log(|\hat{\Omega}^*(\tilde{\beta}, \xi)|)$, while under the alternative, cf. (6.2)-(6.3),

$$(\hat{\beta}, \hat{\xi}) := \arg\min_{(\beta,\xi)} \log(|\hat{\Omega}^*(\beta, \xi)|),$$

and the LR_T statistic readily follows, $LR_T := T(\log|\hat{\Omega}^*(\tilde{\beta}, \tilde{\xi})| - \log|\hat{\Omega}^*(\hat{\beta}, \hat{\xi})|)$, see eq. (2.13).

Three different sample sizes, $T = 250$, 500 and 1000, are considered. For each sample size, 1000 sample paths are simulated for the set of given parameter values (see Appendix E). Next, parameters are estimated as described above using the MLE both under the alternative, and under the null. For the bootstrap, we use $N = 399$ repetitions (see Andrews and Buchinsky, 2001; Cavaliere et al, 2010a,b).

The estimators, test statistics and the bootstrap procedure were implemented in Matlab. In the implementation of the bootstrap procedure, the Matlab numerical maximization routine used to compute the QMLE's under the alternative did not converge for a few of the bootstrap samples; this might be caused by non-identification in the population of the parameters. Moreover, Matlab in those samples reported a negative value of $\sup LR_T$. For these samples, we simply set $\sup LR_T = 0$. Since $\sup LR_T > 0$ this fix means that the estimated distribution of $\sup LR_T$ is pushed to the left and so we will tend to overreject. It's not entirely clear to us how to adjust the bootstrap distribution for this effect. One could potentially leave out the bootstrap samples where non-convergence occurs.

Tables 1 reports the size (i.e. the rejection frequencies under the null) of the bootstrap versions of the LR_T test when we test for $H_R^{(1)}$. From these results, we see that for moderate and large sample sizes ($T = 500$ and 1000) the bootstrap test have very good size properties for both null hypohteses. In smaller sample sizes ($T = 250$), the size begin to deteriorate but is still acceptable.

	1% nominal level	5% nominal level	10% nominal level
$T = 250$	0.4%	4.3%	9.9%
$T = 500$	1.3%	4.8%	10.1%
$T = 1000$	0.9%	5.4%	11.1%

Table 1: Size of bootstrap version of $\sup LR_T$ test for $H_R^{(1)} : \delta = 0$.

The corresponding size performance for the LR_T test of $H_R^{(2)}$ are reported in Table 2. Qualitatively the same picture as for the test of $H_R^{(1)}$ appears: For moderate and large samples, the size is good while in smaller samples it is less precise.

	1% nominal level	5% nominal level	10% nominal level
$T = 250$	0.4%	4.7%	11.8%
$T = 500$	1.0%	5.3%	11.7%
$T = 1000$	1.3%	6.3%	11.7%

Table 2: Size of bootstrap version of LR_T test for $H_R^{(2)} : \beta = (1, -1)$.

Next, we examine the power of the LR_T test for the two hypotheses. The results for $H_R^{(1)}$ are reported in Table 3 The test tends to have low power in small samples, and for example only rejects the incorrect hypothesis of $\delta = 0$ with 16% probability for $T = 250$. However, as the sample size grows, the power quickly improvves and with $T = 500$ observations the bootstrap test exhibit acceptable power properties; for example, it rejects the incorrect null of $\delta = 0$ with 67.6% probability at a 5% level. In large samples ($T = 1000$), the power is very good for the sup-test with rejection probabilities close to 100%.

	1% nominal level	5% nominal level	10% nominal level
$T = 250$	2.7%	16.0%	29.2%
$T = 500$	37.5%	67.6%	78.1%
$T = 1000$	93.5%	97.0%	97.8%

Table 3: Power of bootstrap version of $\sup LR$ test for $H_0^{(1)} : \delta = 0$.

The power of the test of $H_R^{(2)}$ is not quite as impressive as can be seen in Table 4. For example, it rejects at a 5% level with probability 49.5% and 76.4% for sample sizes of $T = 500$ and $T = 1000$ which is significantly lower than the corresponding rejection probabilities reported in Table 3. This is to some extent probably a consequence of the DGP, which under the alternative of $H_R^{(2)}$ is not too far away from the null with β_0 having been chosen as $\beta_0 = (1, -0.9282)'$, cf. Appendix E. Hence it is more difficult to detect the departure from the null in finite samples.

	1% nominal level	5% nominal level	10% nominal level
$T = 250$	3.8%	17.0%	29.7%
$T = 500$	23.2%	49.5%	63.2%
$T = 1000$	63.5%	76.4%	81.4%

Table 4: Power of bootstrap version of LR_T test for $H_0^{(2)} : \beta = (1, -1)$.

7 Conclusion

We have here proposed and analyzed likelihood-based estimators and tests in a class of nonlinear vector error correction models. The properties of estimators and tests prove to be non-standard in two distinct ways: First, due to the dependence between short- and long-run parameter estimators, their asymptotic distributions are not comparable to the standard Dickey-Fuller type asymptotics found in linear models. This in term affects the test statistics. For example, tests only involving short-run parameters will in general not follow χ^2 in contrast to the situation in the linear cointegration model. The distribution of the test statistics get even more involved in the case of testing for linearity of the error correction mechanism due to vanishing parameters under the null.

Due to the complicated nature of the distributions, we proposed to implement the tests using a wild bootstrap procedure, and through simulations we demonstrated that the resulting class of tests perform well both in terms of size and power. It would be of interest to show theoretically that the bootstrap procedure is consistent.

References

Andrews, D.W.K. and M. Buchinsky (2001) A Three-Step Method for Choosing the Number of Bootstrap Repetitions. *Econometrica* 68, 23-51.

Andrews, D.W.K. and W. Ploberger (1994) Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative. *Econometrica* 62, 1383-1414.

Andrews, D.W.K. and W. Ploberger (1995) Admissibility of the Likelihood Ratio Test When A Nuisance Parameter Is Present Only Under the Alternative. *Annals of Statistics* 23, 1609-1629.

Bec, F. and A. Rahbek (2004) Vector Equilibrium Correction Models with Non-linear Discontinuous Adjustments. *Econometrics Journal* 7, 628-651.

Bickel, P.J. and M.J. Wichura (1971) Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *Annals of Mathematical Statistics* 42, 1656-1670.

Brown, B.M. (1971) Martingale Central Limit Theorems. *Annals of Mathematical Statistics* 42, 59-66.

Caner, M. and B. Hansen (2001) Threshold Autoregression with a Unit Root. *Econometrica* 69, 1555–1596.

Cavaliere, G., A. Rahbek and R. Taylor (2010a) Testing for Cointegration in Vector Autoregressions with Non-Stationary Volatility. *Journal of Econometrics* 158, 7-24.

Cavaliere, G., A. Rahbek and R. Taylor (2010b) Co-integration Rank Tests under Conditional Heteroskedasticity. Forthcoming in *Econometric Theory.*

Choi, I. and P. Saikkonen (2004) Testing Linearity in Cointegrating Smooth Transition Regressions. *Econometrics Journal* 7, 341–365.

Davies, R.B. (1987) Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative. *Biometrika* 74, 33-43.

Doukhan, P., P. Massart and E. Rio (1995) Invariance Principles for Absolutely Regular Empirical Processes. Annales de l'Insitute Henri Poincaré 31, 393-427.

Gonzalo, J. and J.-Y. Pitarakis (2006) Threshold Effects in Multivariate Error Correction Models. *In Palgrave Handbook of Econometrics: Econometric Theory* (eds. T.C. Mills, K. Patterson), Vol. 1, 578-609.

de Jong, R.M. (2001) Nonlinear Estimation Using Estimated Cointegrating Relations. *Journal of Econometrics* 101, 109-122.

de Jong, R.M. (2002) Nonlinear Minimization Estimators in the Presence of Cointegrating Relations. *Journal of Econometrics* 110, 241-259.

Hall, P. and C.C. Heyde (1980) *Martingale Limit Theory and its Applications*. New York: Academic Press.

Hansen, B.E. (1992) Convergence to Stochastic Integrals for Dependent Heterogeneous Processes. *Econometric Theory* 8, 489-500.

Hansen, B.E. (1996) Inference when a Nuisance Parameter is not Identified under the Null Hypothesis. *Econometrica* 64, 413-430.

Hansen, B.E. and B. Seo (2002) Testing for Two-regime Threshold Cointegration in Vector Error Correction Models. *Journal of Econometrics* 110, 293-318.

Johansen, S. (1992) Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59, 1551-1580.

Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd Ed. Oxford: Oxford University Press.

Kapetanios, G. Y. Shin and A. Snell (2006) Testing for Cointegration in Nonlinear Smooth Transition Error Correction Models. *Econometric Theory* 22, 279–303.

Kilic, R. (2009) Testing for Co-integration and Nonlinear Adjustment in a Smooth Transition Error Correction Model. Manuscript, Georgia Institute of Technology.

Kristensen, D. and A. Rahbek (2005) Asymptotics of the QMLE for a Class of ARCH(q) Models, *Econometric Theory* 21, 946-961.

Kristensen, D. and A. Rahbek (2009) Asymptotics of the QMLE for Non-linear ARCH Models. *Journal of Time Series Econometrics* 1(1), Article 2.

Kristensen, D. and A. Rahbek (2010) Likelihood-Based Inference for Cointegration with Nonlinear Error-Correction. *Journal of Econometrics* 158, 78-94.

Kurtz, T.G. and P. Protter (1991) Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations. *Annals of Probability* 19, 1035-1070.

Magnus and Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.

Meitz, M. and P. Saikkonen (2009) Parameter estimation in nonlinear AR-GARCH models. Working paper, University of Helsinki.

Nedeljkovic, M. (2008) Testing for Smooth Transition Nonlinearity in Adjustments of Cointegrating Systems, manuscript, Department of Economics, University of Warwick.

Pitarakis, J.-Y. (2008) Comment on: Threshold Autoregression with a Unit Root. *Econometrica* 76, 1207–1217.

Rahbek, A. H.C. Kongsted and C. Jørgensen (1999) Trend Stationarity in the I(2) Cointegration Model. *Journal of Econometrics* 90, 265-289.

Saikkonen, P. (2005) Stability Results for Nonlinear Error Correction Models. *Journal of Econometrics* 127, 69-81.

Saikkonen, P. (2008) Stability of Regime Switching Error Correction Models under Linear Cointegration. *Econometric Theory* 24, 294-318.

Seo, M.H. (2006) Bootstrap Testing for the Null of no Cointegration in a Threshold Vector Error Correction Model. *Journal of Econometrics* 134, 129–150.

Seo, M.H. (2008) Unit Root Test in a Threshold Autoregression: Asymptotic Theory and Residual-Based Block Bootstrap. *Econometric Theory* 24, 1699–1716.

Seo, M.H. (2010) Estimation of Nonlinear Error Correction Models. Forthcoming in *Econometric Theory.*

Straf, M.L. (1972) Weak Convergence of Stochastic Processes with Several Parameters. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, 187-221. University of California Press.

A Proofs of Section 3

Proof of Theorem 3.4. Note initially that finite dimensional convergence follows by standard Martingale CLT results, see e.g. Brown (1971). In particular, $(\phi_T(s,\pi), \phi_T(\tilde{s},\tilde{\pi})) \overset{D}{\to} (\phi(s,\pi), \phi(\tilde{s},\tilde{\pi}))$ for any (s,π) and $(\tilde{s},\tilde{\pi})$ in $[0,1]\times\Pi$. Next, we show stochastic equicontinuity (tightness) by verifying the conditions of Theorem 3 in Bickel and Wichura (1971). Some further notation is needed for this: With $s \in [0,1]$ and $\pi \in \Pi \subseteq \mathbb{R}^{d_\pi}$, let $\gamma = (s,\pi) = (s,\pi_1,...,\pi_{d_\pi})$ and $\tilde{\gamma} = (\tilde{s},\tilde{\pi}) = \left(\tilde{s},\tilde{\pi}_1,...,\tilde{\pi}_{d_\xi}\right)$ be two arbitrary points in $[0,1]\times\Pi$. With $\tilde{\gamma} > \gamma$, define Δy_T on $(\gamma,\tilde{\gamma})$, see Bickel and Wichura (1971),

$$\Delta\phi_T(\gamma,\tilde{\gamma}) :=$$
$$\sum_{\epsilon_0=0,1}\sum_{\epsilon_1=0,1}\cdots\sum_{\epsilon_{d_\pi}=0,1}(-1)^{d_\pi+1-\Sigma_{j=0}^{d_\pi}\epsilon_j}\,\phi_T\left(s-\epsilon_0 ds, \pi_1-\epsilon_1 d\pi_1,...,\pi_{d_\pi}-\epsilon_{d_\pi}d\pi_{d_\pi}\right),$$

where $ds = (s-\tilde{s})$ and $d\pi_i = (\pi_i - \tilde{\pi}_i)$. Direct insertion gives that $\Delta\phi_T(\gamma,\tilde{\gamma})$ can be written as,

$$\Delta\phi_T(\gamma,\tilde{\gamma}) = \frac{1}{\sqrt{T}}\sum_{t=[Ts]}^{[T\tilde{s}]}\Delta f_t(\pi,\tilde{\pi})\,e_t, \tag{A.1}$$

where

$$\Delta f_t(\pi,\tilde{\pi}) = \sum_{\epsilon_1=0,1}\cdots\sum_{\epsilon_{d_\pi}=0,1}(-1)^{d_\pi-\Sigma_{j=1}^{d_\pi}\epsilon_j}\,f\left(y_{t-1};\pi_1-\epsilon_1 d\pi_1,...,\pi_{d_\pi}-\epsilon_{d_\pi}d\pi_{d_\pi}\right). \tag{A.2}$$

Since $(s,\pi)\mapsto\phi(s,\pi)$ is almost surely continuous, then by Straf (1972, Theorem 5.6) in combination with Bickel and Wichura (1971, Theorem 1) it suffices to establish that, for some $q>0$, $\nu>1$,

$$E\left\|\Delta\phi_T(\pi,\tilde{\pi}]\right\|^{2q} \le C\left|\tilde{s}-s\right|^{\nu}\Pi_{i=1,2,..,d_\pi}\left|\tilde{\pi}_i-\pi_i\right|^{\nu}. \tag{A.3}$$

Under Assumptions 3.1-3.2, and using Rosenthal's inequality (see Hall and Heyde, 1980, p.23), Cauchy-Schwarz, and eq. (A.1), it follows that for $q>1$,

$$\begin{aligned}
E\left\|\Delta\phi_T(\pi,\tilde{\pi})\right\|^{2q} &\le \frac{C}{T^q}\left(\sum_{t=[Ts]}^{[T\tilde{s}]}E\left[E\left(\|e_t\|^2\left\|\Delta f_t(\pi,\tilde{\pi})\right\|^2|\mathcal{F}_{t-1}\right)\right]\right)^q \\
&\quad+\frac{C}{T^q}\sum_{t=[Ts]}^{[T\tilde{s}]}E\left[\|e_t\|^{2q}\left\|\Delta f_t(\pi,\tilde{\pi})\right\|^{2q}\right] \\
&\le C\left(\tfrac{[T\tilde{s}]-[Ts]}{T}\right)^q\|\Omega_e\|^q\left(E\left\|\Delta f_t(\pi,\tilde{\pi})\right\|^2\right)^q \\
&\quad+C\left(\tfrac{[T\tilde{s}]-[Ts]}{T^q}\right)\left[E\|e_t\|^{4q}\left(E\left\|\Delta f_t(\pi,\tilde{\pi})\right\|^{4q}\right)\right]^{1/2} \\
&\le C\left(\tfrac{[T\tilde{s}]-[Ts]}{T}\right)^q E\left\|\Delta f_t(\pi,\tilde{\pi})\right\|^{2q}+o(1),
\end{aligned}\tag{A.4}$$

as $T\to\infty$. Observe that for each $i\in D_\pi=\{1,...,d_\pi\}$, one may write $\Delta f_t(\pi,\tilde{\pi}]$ as,

$$\Delta f_t(\pi,\tilde{\pi}) = \sum_{j\in D_\pi,\ ,j\neq i,\ \varepsilon_j=0,1}(-1)^{d_\pi-\Sigma_{j\neq i}\varepsilon_j}\,\partial_i f_t(\pi,\tilde{\pi}), \tag{A.5}$$

where $\partial_i f_t(\pi, \tilde{\pi}]$ denotes the increments of $f(\cdot)$ over the ith coordinate of π,

$$\begin{aligned}\partial_i f_t(\pi, \tilde{\pi}) = & f_t(\pi_1 - \epsilon_1 d\pi_1, ..., \pi_{i-1} - \epsilon_{i-1} d\pi_{i-1}, \tilde{\pi}_i, \pi_{i+1} - \epsilon_{i+1} d\pi_{i+1}, ..., \pi_q - \epsilon_q d\pi_{d_\pi}) \\ & - f_t(\pi_1 - \epsilon_1 d\pi_1, ..., \pi_{i-1} - \epsilon_{i-1} d\pi_{i-1}, \pi_i, \pi_{i+1} - \epsilon_{i+1} d\pi_{i+1}, ..., \pi_q - \epsilon_q d\pi_{d_\pi}),\end{aligned}$$

with $d\pi_i = \tilde{\pi}_i - \pi_i$. Using eq. (A.5), and by Assumption 3.2,

$$E \|\Delta f_t(\pi, \tilde{\pi})\|^2 \leq C 2^{d_\pi - 1} E \|\partial_i f_t(\pi, \tilde{\pi}]\|^2 \leq C |\tilde{\pi}_i - \pi_i|^{2\lambda}$$

As this holds for each i, one gets that for T large enough, eq. (A.4) is bounded by,

$$E \|\Delta \phi_T(\pi, \tilde{\pi})\|^{2q} \leq C |\tilde{s} - s|^q \left(E \|\Delta f_t(\pi, \tilde{\pi})\|^2\right)^q \leq C |\tilde{s} - s|^q \Pi_{i \in D_\pi} |\tilde{\pi}_i - \pi_i|^{2\lambda q / d_\pi},$$

and the desired holds. ■

Proof of Corollary 3.7. For fixed s the result holds by Doukhan et al (1995), see also Hansen (1996), while for fixed π the result holds by Brown (1971). ■

Proof of Theorem 3.5. It follows by standard results that the convergence holds for any given $\pi \in \Pi$, see e.g. Kurtz and Protter (1991, Theorem 2.2). Proceeding as in the proof of Theorem 3.4, define

$$V_T(\pi) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x'_{t-1} f(y_{t-1}; \pi) e_t, \quad \Delta V_T(\pi, \tilde{\pi}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x'_{t-1} \Delta f_t(\pi, \tilde{\pi}) e_t,$$

where Δf_t is defined in eq. (A.2). Again by Rosenthal's inequality (Hall and Heyde, 1980, p.23) and Cauchy-Schwarz inequalities, for $n > 1$,

$$\begin{aligned}& E\left[\|\Delta V_T(\pi, \tilde{\pi})\|^{2q}\right] \\ & \leq \frac{C}{T^q}\left(\sum_{t=1}^{T} E\left[E\left(\|x_{t-1}\|^2 \|e_t\|^2 \|\Delta f_t(\pi, \tilde{\pi})\|^2 |\mathcal{F}_{t-1}\right)\right]\right)^q \\ & + \frac{C}{T^q} \sum_{t=1}^{T} E\left(\|x_{t-1}\|^{2q} \|e_t\|^{2q} \|\Delta f_t(\pi, \tilde{\pi})\|^{2q}\right) \\ & \leq C \|\Omega_e\|^q \left(\left(\sup_t E \|x_{t-1}\|^4\right) E \|\Delta f_t(\pi, \tilde{\pi})\|^4\right)^{q/2} \\ & + C T^{1-q} E\left(\sup_t \left(\|x_{t-1}\|^{2q} \|e_t\|^{2q} \|\Delta f_t(\pi, \tilde{\pi})\|^{2q}\right)\right) \\ & \leq C \|\Omega_e\|^q \left(\sup_t E \|x_{t-1}\|^4 E \|\Delta f_t(\pi, \tilde{\pi})\|^4\right)^{q/2} \\ & + C T^{1-q}\left(E \sup_t \|x_{t-1}\|^{6q} E \|e_t\|^{6q} E \sup_{\pi \in \Pi} \|f(y_{t-1}; \pi)\|^{6q}\right)^{1/3} \\ & \leq C \Pi_{i \in D_\pi} |\tilde{\pi}_i - \pi_i|^{2\lambda q / d_\pi} + o(1).\end{aligned}$$

By the same arguments as employed in the proof of Theorem 3.4, the desired result now follows. ■

Proof of Theorem 3.6. Define the mean-zero sequence $u_t(\pi) = f(y_{t-1};\pi) - E[f(y_{t-1};\pi)]$ and write

$$\frac{1}{T}\textstyle\sum_{t=1}^{T} x_{t-1}' f(y_{t-1};\pi) = \frac{1}{T}\sum_{t=1}^{T} x_{t-1}' E[f(y_{t-1};\pi)] + \frac{1}{T}\sum_{t=1}^{T} x_{t-1}' u_t(\pi).$$

By Assumption 3.3 and the Continuous Mapping Theorem, the first term converges towards the claimed limit. We then need to show that the second term goes to zero in probability uniformly in π. We follow the same arguments as in Caner and Hansen (2001, Proof of Theorem 3): For any given $\delta > 0$, define $N = [1/\delta]$, $t_k = [k\delta T] + 1$ and $t_k^* = t_{k+1} - 1$, and write

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} x_{t-1}' u_t(\pi) &= \frac{1}{T}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*} x_{t-1}{\prime} u_t(\pi) \\ &= \frac{1}{T}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*} (x_{t-1} - x_{t_k-1})' u_t(\pi) + \frac{1}{T}\sum_{k=0}^{N-1} x_{t_k-1}' \sum_{t=t_k}^{t_k^*} u_t(\pi).\end{aligned}$$

The first term is bounded by,

$$\frac{1}{T}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*} \|x_{t-1} - x_{t_k-1}\| \sup_{\pi\in\Xi}\|u_t(\pi)\| \leq \left\{\sup_{|t-t'|\leq T\delta}\|x_t - x_{t'}\|\right\} \times \frac{1}{T}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*}\sup_{\pi\in\Xi}\|u_t(\pi)\|,$$

where, by the law of large numbers,

$$\frac{1}{T}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*}\sup_{\pi\in\Xi}\|u_t(\pi)\| = \frac{1}{T}\sum_{t=1}^{T}\sup_{\pi\in\Xi}\|u_t(\pi)\| \xrightarrow{P} E\left[\sup_{\pi\in\Xi}\|u_t(\pi)\|\right] < \infty,$$

and, by Assumption 3.3,

$$\sup_{|t-t'|\leq T\delta}\|x_t - x_{t'}\| \xrightarrow{D} \sup_{|s-s'|\leq\delta}\left\|x(s) - x(s')\right\|.$$

The limit can be made arbitrarily small due to a.s. continuity of $x(s)$. The second term is bounded by

$$\frac{1}{T}\sum_{k=0}^{N-1}\|x_{t_k-1}\|\left\|\sum_{t=t_k}^{t_k^*} u_t(\pi)\right\| \leq \left\{\sup_{1\leq t\leq T}\|x_t\|\right\} \times \frac{1}{T}\sum_{k=0}^{N-1}\left\|\sum_{t=t_k}^{t_k^*} u_t(\pi)\right\|,$$

where $\sup_{1\leq t\leq T}\|x_t\| = O_P(1)$. Next, $\sup_{\pi\in\Pi}||\sum_{t=1}^{N} u_t(\pi)/N|| \xrightarrow{P} 0$ by Kristensen and Rahbek (2005, Proposition 1) as $N \to \infty$, and hence the arguments following (A.10) in Caner and Hansen (2001, proof of Theorem 3) imply that

$$\sup_{\pi\in\Xi}\frac{1}{T}\sum_{k=0}^{N-1}\left\|\sum_{t=t_k}^{t_k^*} u_t(\pi)\right\| \xrightarrow{P} 0, \quad \text{as } T\delta \to \infty.$$

The proof of the second assertion follows by the same arguments. ■

B Proofs of Section 4

Proof of Lemma 4.5. Choose any $d\alpha$, $d\delta$ and $d\Phi$ and define $\lambda = vec(d\alpha, d\delta, d\Phi)$. We consider the sequence

$$\phi_T(s,\xi) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \{\lambda_u' u_t(\xi) + \lambda_v' v_t(\xi) + \lambda_w' w_t(\xi)\} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} f(y_{t-1};\xi) e_t,$$

with $e_t := \Omega_0^{-1}\varepsilon_t$, $y_{t-1} = \left(Z_{0,t-1}', Z_{2,t-1}'\right)'$, and

$$f(y_{t-1};\xi) := [d\alpha Z_{0,t-1} + d\delta\psi(Z_{0,t-1};\xi) + d\Phi Z_{2,t-1}]'.$$

Also note that $\pi = \xi$. Here, by Assumption 4.1

$$\|f(y_{t-1};\xi)\| \le c(\|Z_{0,t-1}\| + \|\psi(Z_{0,t-1};\xi)\| + \|Z_{2,t-1}\|) \le c(\|Z_{0,t-1}\| + \|Z_{0,t-1}\|^\rho + \|Z_{2,t-1}\|)$$

Thus,

$$\|f(y_{t-1};\xi)\|^m \le c(\|Z_{0,t-1}\|^m + \|Z_{0,t-1}\|^{m\rho} + \|Z_{2,t-1}\|^m).$$

Furthermore, by the differentiability of ψ,

$$\begin{aligned} E\left[\left\|f(y_{t-1};\xi) - f\left(y_{t-1};\xi'\right)\right\|^n\right] &= E\left[\left\|\psi(Z_{0,t-1};\xi) - \psi\left(Z_{0,t-1};\xi'\right)\right\|^n\right] \\ &\le E\left[\left\|\frac{\partial\psi\left(Z_{0,t-1};\bar{\xi}\right)}{\partial\xi}\right\|^n\right]\left\|\xi-\xi'\right\|^n \\ &\le E\left[\|Z_{0,t-1}\|^{\rho n}\right]\left\|\xi-\xi'\right\|^n, \end{aligned}$$

such that $\lambda = 1$. Thus, the requirement $n > 2$ translates into $E\left[\|Z_{0,t-1}\|^{(2+\delta)\rho}\right] < \infty$ for some $\delta > 0$, and the requirement $m > \bar{m} := \max(4, 2d_\xi)$ translates into $E\left[\|Z_{0,t-1}\|^{\bar{m}\bar{\rho}}\right] < \infty$ with $\bar{\rho} = \max(1,\rho)$, and $E\left[\|Z_{2,t-1}\|^{\bar{m}}\right] < \infty$.

This verifies that Assumptions 4.2-4.4 imply that the Assumptions 3.1-3.3 of Theorem 3.4 hold, and hence the result follows for $(u_t'(\cdot), v_t'(\cdot), w_t'(\cdot))$. The joint convergence holds by the marginal convergence in Assumption 4.3, in conjunction with the fact that $(u_t'(\cdot), v_t'(\cdot), w_t'(\cdot))$ and X_t are defined in terms of $(\varepsilon_s)_{s\le t}$. ■

Proof of Theorems 4.6. For ease of notation, we treat $\Omega = \Omega_0$ as known such that $L_T^* = L_T$. The extension to unknown Ω is straigthforward and follows along the lines of Kristensen and Rahbek (2010).

To establish the result, we apply a general formulation in Lemmas D.1 and D.2 in Appendix D below which will allow us to consider convergence uniformly in ξ. To use the results in Section D, set $\gamma = vec(\vartheta)$, $\pi = \xi_0$, $Q_T(\gamma,\pi) = Q_T(\vartheta) = -\frac{1}{T}L_T(\vartheta)$, with $L_T(\vartheta)$ defined in eq. (2.11), $v_T = T$ and $U_T = V_T$, where V_T is defined in eq. (4.3). To prove consistency, we verify the conditions of Lemma D.1: We have that condition (i) holds by Assumption 4.1, while (ii)-(iii) follow by Lemmas C.1, C.2 and C.3:

$$dQ_T(\vartheta_0,\xi_0;U_T^{-1/2}d\gamma) = -\frac{1}{T}dL_T(\vartheta_0;V_T^{-1/2}d\gamma) = o_P(1),$$

$$d^2Q_T\left(\vartheta_0;U_T^{-1/2}d\gamma,U_T^{-1/2}d\bar{\gamma}\right)=-\frac{1}{T}d^2L_T\left(\vartheta_0;V_T^{-1/2}d\gamma,V_T^{-1/2}d\bar{\gamma}\right)\overset{D}{\to}H_\infty\left(d\gamma,d\bar{\gamma}\right),$$

$$\begin{aligned}d^3Q_T\left(\vartheta;U_T^{-1/2}d\gamma,U_T^{-1/2}d\bar{\gamma},U_T^{-1/2}d\check{\gamma}\right)&=-\frac{1}{T}d^3L_T\left(\vartheta;V_T^{-1/2}d\gamma,V_T^{-1/2}d\bar{\gamma},V_T^{-1/2}d\check{\gamma}\right)\\&=O_P(||d\gamma||||d\bar{\gamma}||||d\check{\gamma}||),\end{aligned}$$

with $H_\infty\left(d\gamma,d\bar{\gamma}\right)$ given in C.2. The asymptotic distribution will follow from Lemma D.2 by verifying the additional condition (iv) in Lemma D.2. But this follows from Lemma C.1 since,

$$dQ_T(\vartheta_0;\nu_T^{1/2}U_T^{-1/2}vec\left(d\vartheta\right))=-T^{-1/2}dL_T(\vartheta_0;V_T^{-1/2}vec\left(d\vartheta\right))\overset{D}{\to}S_\infty\left(d\vartheta\right),$$

where $S_\infty\left(d\vartheta\right)$ is given in Lemma C.1. We conclude that $V_T^{1/2}(vec(\hat{\vartheta}_T)-vec(\vartheta_0))\overset{D}{\to}vec(d\vartheta_\infty)$, where ϑ_∞ satisfies $S_\infty\left(d\vartheta\right)=H_\infty\left(d\vartheta,d\vartheta_\infty\right)$ for all directions $d\vartheta$. This together with eq. (D.1) imply the results stated in Theorem 4.6. ■

Proof of Theorems 4.7. We proceed as in the proof of Theorem 4.6: Set $\gamma=vec\left(\theta\right)$, $\pi=vec\left(\xi\right)$, $Q_T\left(\gamma,\pi\right)=-L_T^*\left(\theta,\xi\right)/T$, $v_T=T$ and $U_T=V_{\theta T}$, where $V_{\theta T}$ is defined in (4.3). We can now apply Lemmas D.1 and D.2. The conditions stated there hold by Lemmas C.1, C.2 and C.3. ■

Proof of Theorem 4.8. We give a proof of the most complicated case where ξ is not identified under the null; the proof of the other case is analogous. We rewrite the restriction on η as $vec\left(\eta\right)=\left(R_\eta\right)_\perp\tau$ where τ is an unrestricted parameter vector. We first analyze the restricted estimator $\tilde{\theta}$: Under the null ξ vanishes so the restricted log-likelihood does not depend on this parameter. Thus, $L_T^*\left(b,\eta\right)=\tilde{L}_T^*\left(b,\tau\right)$ and $\tilde{L}_T^*\left(b,\tau\right):=L_T^*\left(b,\left(R_\eta\right)_\perp\tau\right)$. Taking differentials w.r.t. (b,τ),

$$d\tilde{L}_T^*\left(b,\tau\right)=\mathbb{S}_{b,T}(\theta)vec\left(db'\right)+\mathbb{S}_{\eta,T}(\theta)'\left(R_\eta\right)_\perp d\tau,$$

$$\begin{aligned}d^2\tilde{L}_T^*\left(b,\tau\right)=&\ vec\left(db'\right)'\mathbb{H}_{bb,T}(\theta)vec\left(db'\right)+d\tau'\left(R_\eta\right)_\perp'\mathbb{H}_{\eta\eta,T}(\theta)\left(R_\eta\right)_\perp d\tau\\&+2vec\left(db'\right)'\mathbb{H}_{b\eta,T}(\theta)\left(R_\eta\right)_\perp d\tau,\end{aligned}$$

where we suppress dependence on db and $d\tau$ in the differentials. Here, $\mathbb{S}_{b,T}(\theta)$ and $\mathbb{H}_{bb,T}(\theta)$ are the score vector and Hessian matrix w.r.t. b defined as the solutions to $dL_T^*\left(\theta;db\right)=\mathbb{S}_{b,T}(\theta)'vec\left(db'\right)$ and $d^2L_T^*\left(\theta;db,db\right)=vec\left(db'\right)'\mathbb{H}_{bb,T}(\theta)vec\left(db'\right)$; similarly with $\mathbb{S}_{\eta,T}(b,\eta)$, $\mathbb{H}_{\eta\eta,T}(\theta)$ and $\mathbb{H}_{b\eta,T}(\theta)$. By the same arguments as used in the proof of Theorem 4.7, we now obtain that $(\tilde{b},\tilde{\tau})$ satisfies

$$\begin{aligned}0&=d\tilde{L}_T^*\left(b_0,\tau_0;d\tau\right)+d^2\tilde{L}_T^*\left(b_0,\tau_0;d\tau,\tilde{\tau}-\tau_0\right)\\&=\mathbb{S}_{b,T}(\theta_0)'vec\left(db'\right)+\mathbb{S}_{\eta,T}(\theta_0)'\left(R_\eta\right)_\perp d\tau\\&+vec(\tilde{b}')'\mathbb{H}_{bb,T}(\theta_0)vec\left(db'\right)+(\tilde{\tau}-\tau_0)'\left(R_\eta\right)_\perp'\mathbb{H}_{\eta\eta,T}(\theta_0)\left(R_\eta\right)_\perp d\tau\\&+vec(\tilde{b}')'\mathbb{H}_{b\eta,T}(\theta_0)\left(R_\eta\right)_\perp d\tau+(\tilde{\tau}-\tau_0)'\left(R_\eta\right)_\perp'\mathbb{H}_{\eta b,T}(\theta_0)vec\left(db'\right)\end{aligned}$$

for any directions $(db, d\tau)$, where we ignore the higher-order remainder term. With $vec\,(db') = K_T^{-1}d\bar{h}$ and $d\tau = 1/\sqrt{T}d\bar{\tau}$, Lemmas C.1-C.2 yield

$$\begin{aligned}
\mathbb{S}_{b,T}(\theta_0)'vec\left(db'\right) &= \mathbb{S}_{b,T}(\theta_0)'K_T^{-1}dh \xrightarrow{D} \mathbb{S}_{b,\infty}(\theta_0)'dh,\\
\mathbb{S}_{\eta,T}(\theta_0)'\left(R_\eta\right)_\perp d\tau &= T^{-1/2}\mathbb{S}_{\eta,T}(\theta_0)'\left(R_\eta\right)_\perp d\bar{\tau} \xrightarrow{D} \mathbb{S}_{\eta,\infty}(\theta_0)'\left(R_\eta\right)_\perp d\bar{\tau},\\
K_T^{-1}\mathbb{H}_{bb,T}(\theta_0)vec\left(db'\right) &= K_T^{-1}\mathbb{H}_{bb,T}(\theta_0,\xi)K_T^{-1}dh \xrightarrow{D} \mathbb{H}_{bb,\infty}(\theta_0)dh\\
T^{-1/2}R_\perp'\mathbb{H}_{\eta\eta,T}(\theta_0)\left(R_\eta\right)_\perp d\tau &= T^{-1}R_\perp'\mathbb{H}_{\eta\eta,T}(\theta_0)\left(R_\eta\right)_\perp d\bar{\tau} \xrightarrow{D} R_\perp'\mathbb{H}_{\eta\eta,\infty}(\theta_0)\left(R_\eta\right)_\perp d\bar{\tau},
\end{aligned}$$

and similar for the cross terms. We conclude that

$$\sqrt{T}\begin{pmatrix} \left(I_r\otimes K_T\right)vec\left(\tilde{b}'\right) \\ \tilde{\tau}-\tau_0 \end{pmatrix} = -\tilde{\mathbb{H}}_{\theta,T}^{-1}\tilde{\mathbb{S}}_{\theta,T} + o_P\left(1\right),$$

where $\tilde{\mathbb{H}}_T \xrightarrow{D} \tilde{\mathbb{H}}$ and $\tilde{\mathbb{S}}_T \xrightarrow{D} \tilde{\mathbb{S}}$ with $\tilde{\mathbb{H}}_{\theta\theta}$ and $\tilde{\mathbb{S}}_\theta$ defined in eq. (4.11). Thus,

$$\sqrt{T}V_{\theta,T}^{1/2}vec\left(\tilde{\theta}-\theta_0\right) = \sqrt{T}V_{\theta,T}^{1/2}\begin{pmatrix} vec\left(\tilde{b}'\right) \\ vec\left(\tilde{\eta}\right)-vec\left(\eta_0\right) \end{pmatrix} = -M_\eta\tilde{\mathbb{H}}_{\theta\theta,T}^{-1}\tilde{\mathbb{S}}_{\theta,T} + o_P\left(1\right),$$

Next, from the proof of Theorem 4.7, for any ξ,

$$\sqrt{T}V_{\theta,T}^{1/2}vec(\hat{\theta}\left(\xi\right)-\theta_0) = \sqrt{T}V_{\theta,T}^{1/2}\begin{pmatrix} vec(\hat{b}\left(\xi\right)') \\ vec\left(\hat{\eta}\left(\xi\right)\right)-vec\left(\eta_0\right) \end{pmatrix} = -\mathbb{H}_{\theta\theta,T}^{-1}\left(\xi\right)\mathbb{S}_{\theta,T}\left(\xi\right) + o_P\left(1\right).$$

Given these results, we derive the asymptotic distributions of the sup-LR and sup-LM test. Regarding the sup-LR test, use a second-order Taylor expansion to obtain

$$\begin{aligned}
LR_T\left(\xi\right) &= 2\left[L_T^*(\hat{\theta}\left(\xi\right),\xi) - L_T^*(\tilde{\theta})\right]\\
&= \frac{1}{2}\mathbb{S}_{\theta,T}(\hat{\theta}\left(\xi\right),\xi)(\hat{\theta}\left(\xi\right)-\tilde{\theta}) + (\hat{\theta}\left(\xi\right)-\tilde{\theta})'\mathbb{H}_{\theta\theta,T}(\bar{\theta}\left(\xi\right),\xi)(\hat{\theta}\left(\xi\right)-\tilde{\theta}),
\end{aligned}$$

where $\bar{\theta}\left(\xi\right)$ lies between $\hat{\theta}\left(\xi\right)$ and $\tilde{\theta}$. Since $\hat{\theta}\left(\xi\right)$ maximizes $L_T^*(\theta,\xi)$, $\mathbb{S}_{\theta,T}(\hat{\theta}\left(\xi\right),\xi)=0$, while

$$\begin{aligned}
-\sqrt{T}V_{\theta,T}^{1/2}vec\left(\hat{\theta}\left(\xi\right)-\tilde{\theta}\right) &= -\sqrt{T}V_{\theta,T}^{1/2}vec\left(\hat{\theta}\left(\xi\right)-\theta_0\right) + \sqrt{T}V_{\theta,T}^{1/2}vec\left(\tilde{\theta}-\theta_0\right)\\
&= \mathbb{H}_{\theta\theta,T}^{-1}(\theta_0,\xi)\mathbb{S}_{\theta,T}(\theta_0,\xi) - M_\eta\tilde{\mathbb{H}}_{\theta\theta,T}^{-1}\tilde{\mathbb{S}}_{\theta,T} + o_P\left(1\right)\\
&\xrightarrow{D} \mathbb{H}_{\theta\theta}^{-1}(\xi)\mathbb{S}_\theta(\xi) - M_\eta\left[M_\eta'\mathbb{H}_{\theta\theta}(\xi)M_\eta\right]^{-1}M_\eta'\mathbb{S}_\theta(\xi)\\
&= \mathbb{P}\left(\xi\right)\mathbb{S}_\theta(\xi),
\end{aligned}$$

where we have employed Lemmas C.1-C.3, and

$$\mathbb{P}\left(\xi\right) := \mathbb{H}_{\theta\theta}^{-1}(\xi) - M_\eta\left[M_\eta'\mathbb{H}_{\theta\theta}(\xi)M_\eta\right]^{-1}M_\eta'.$$

Thus,

$$\lim_{T\to\infty} LR_T\left(\xi\right) \stackrel{d}{=} \mathbb{S}_\theta(\xi)'\mathbb{P}\left(\xi\right)'\mathbb{H}_{\theta\theta}(\xi)\mathbb{P}\left(\xi\right)\mathbb{S}_\theta(\xi) = \mathbb{V}\left(\xi\right)'\mathbb{V}\left(\xi\right),$$

where $\mathbb{V}\left(\xi\right)$ is given in the theorem. For the LM test, use a first order Taylor expansion to write the unrestricted score evaluated at the restricted estimators as

$$\begin{aligned}
\mathbb{S}_{\theta,T}(\tilde{\theta},\xi) &= \mathbb{S}_{\theta,T}(\theta_0,\xi) + \mathbb{H}_{\theta\theta,T}(\theta_0,\xi)\sqrt{T}V_{\theta,T}^{1/2}vec(\tilde{\theta}-\theta_0) + o_P\left(1\right),\\
&\xrightarrow{D} \mathbb{S}_\theta(\xi) - \mathbb{H}_{\theta\theta}(\theta_0,\xi)M_\eta\tilde{\mathbb{H}}_{\theta\theta}^{-1}\tilde{\mathbb{S}}_\theta\\
&= \mathbb{H}_{\theta\theta}(\xi)\mathbb{P}\left(\xi\right)\mathbb{S}_\theta(\xi),
\end{aligned}$$

In conclusion, uniformly in ξ,

$$\lim_{T\to\infty} LM_T(\xi) \overset{D}{=} \mathbb{S}_\theta(\xi)'\mathbb{P}(\xi)'\,\mathbb{H}_{\theta\theta}(\xi)\mathbb{P}(\theta_0,\xi)\,\mathbb{S}_\theta(\theta_0,\xi) = \mathbb{V}(\xi)'\,\mathbb{V}(\xi).$$

■

C Asymptotics of derivatives of likelihood function

In the following, we use the notation $V_T^{-1/2}d\vartheta = unvec(V_T^{-1/2}vec(d\vartheta))$ to save space, and similar for other parameters.

Lemma C.1 *Under Assumptions 4.1-4.4 with $q_0 > 2\max\{1,\rho\}$ and $q_2 > 2$, the log-likelihood function $L_T(\vartheta)$ defined in (2.11) with $d\vartheta = (d\theta, d\xi)$ and $d\theta = (db', d\eta)$ satisfies:*

1. *If $\delta_0 \neq 0$, then as $T \to \infty$,*

$$T^{-1/2}dL_T(\vartheta_0; V_T^{-1/2}d\vartheta) \xrightarrow{D} S_{\theta,\infty}(\xi_0; d\theta) + S_{\xi,\infty}(\xi_0; d\xi),$$

where

$$S_{\theta,\infty}(\xi; d\theta) = \left[tr(db'\int_0^1 F(s)\,dB_w'(s,\xi))\right] + vec(d\eta)'B_u(1,\xi),$$
$$S_{\xi,\infty}(\xi; d\xi) = (vecd\xi)'B_v(1,\xi),$$

and $(B_u', B_v', B_w', F')'$ are defined in (4.1).

2. *If $\delta_0 = 0$, then uniformly over $\xi \in \Xi$, as $T \to \infty$,*

$$T^{-1/2}dL_T(\theta_0,\xi; V_{\theta\theta,T}^{-1/2}d\theta) = S_{\theta,T}(\theta_0,\xi; V_{\theta\theta,T}^{-1/2}d\theta) \xrightarrow{D} S_{\theta,\infty}(\xi; d\theta). \tag{C.1}$$

Proof. The first order differential of $L_T(\theta,\xi)$ is given by

$$T^{-1/2}dL_T(\vartheta; V_T^{-1/2}d\vartheta) = S_{b,T}\left(\theta,\xi; K_T^{-1}db\right) + S_{\eta,T}(\theta,\xi; d\eta) + S_{\xi,T}(\theta,\xi; d\xi)$$

where, with

$$Z_t(b) := Z_{0,t-1} + b'Z_{1,t-1} \tag{C.2}$$

$$\sqrt{T}S_{\eta,T}(\theta,\xi; d\eta) = \sum_{t=1}^T [d\alpha Z_t(b) + d\delta\psi(Z_t(b);\xi) + d\Phi Z_{2,t-1}]'\,\Omega_0^{-1}\varepsilon_t(\theta), \tag{C.3}$$

$$\sqrt{T}S_{b,T}(\theta,\xi; db) = \sum_{t=1}^T Z_{1,t-1}'db\,(\alpha + \delta\partial_z\psi(Z_t(b);\xi))'\,\Omega_0^{-1}\varepsilon_t(\theta). \tag{C.4}$$

$$\sqrt{T}S_{\xi,T}(\theta,\xi; d\xi) = (vec(d\xi))'\sum_{t=1}^T \partial_\xi\psi(Z_t(b);\xi)'\,\delta'\Omega_0^{-1}\varepsilon_t(\theta). \tag{C.5}$$

Proof of part 2 ($\delta_0 = 0$): Evaluated at the parameter value $\vartheta_0(\xi) = (0, \eta_0, \xi)$, with $\delta_0 = 0$, we get

$$S_{\eta,T}(\theta_0,\xi; d\eta) = \tfrac{1}{\sqrt{T}}(vec(d\eta))'\sum_{t=1}^T u_t(\xi), \quad S_{b,T}(\theta_0,\xi; db) = \tfrac{1}{\sqrt{T}}\sum_{t=1}^T Z_{1,t-1}'dbw_t,$$

where $u_t(\xi) \in \mathbb{R}^{p(r+r_\delta+pk)}$ and $w_t \in \mathbb{R}^r$ are defined in eq. (2.21), where we note that $w_t(\xi)$ does not depend on ξ when $\delta_0 = 0$ and we therefore simply write w_t. By the same arguments as in the Proof of Lemma 4.5, we now apply Corollary 3.7 with $x_{t-1} = K_T^{-1} Z_{1,t-1}$ to obtain

$$S_{\eta,T}(\theta_0, \xi; d\eta) = T^{-1/2} (vec(d\eta))' \sum_{t=1}^{T} u_t(\xi) \xrightarrow{D} (vec(d\eta))' \int_0^1 dB_u(s, \xi),$$

on the space $C(\Xi)$, and

$$S_{b,T}\left(\theta_0, \xi; K_T^{-1} db\right) = T^{-1/2} \sum_{t=1}^{T} \left(\left\{Z_{1,t-1}' K_T^{-1}\right\} db\right) w_t \xrightarrow{D} \int_0^1 \left(F(s)' db\right) dB_w(s).$$

The two convergence results above hold simultaneously. This proves the second part of the theorem.

Proof of Part 1 ($\delta_0 \neq 0$)*:* By the same arguments as in the proof of Part 2 the claimed results holds for the derivatives considered there. Simultaneously with those convergence results, it also holds that, by Corollary 3.7,

$$S_{\xi,T}(\theta_0, \xi_0; d\xi) = T^{-1/2} (vec(d\xi))' \sum_{t=1}^{T} v_t(\xi_0) \xrightarrow{D} (vec(d\xi))' \int_0^1 dB_v(s, \xi_0).$$

■

Lemma C.2 *Under Assumptions 4.1-4.4, with $d\vartheta = (d\theta, d\xi)$, $d\theta = (d\eta, db)$ and the log-likelihood function $L_T(\theta, \xi)$ defined in (2.11), the following hold:*

1. *If $\delta_0 \neq 0$, then with $q_0 > 2\max\{1, \rho\}$ and $q_2 > 2$:*

$$-\frac{1}{T} d^2 L_T(\vartheta_0; V_T^{-1/2} d\vartheta, V_T^{-1/2} d\bar{\vartheta})$$
$$\xrightarrow{D} H_{\theta\theta,\infty}(\vartheta_0; d\theta, d\bar{\theta}) + H_{\theta\xi,\infty}(\vartheta_0; d\theta, d\bar{\xi}) + H_{\xi\theta,,\infty}(\vartheta_0; d\xi, d\bar{\theta}) + H_{\xi\xi,,\infty}(\vartheta_0; d\xi, d\bar{\xi})$$

   where

$$H_{\theta\theta,\infty}(\vartheta, d\theta, d\bar{\theta}) = vec(d\eta)' \Sigma_{u,u}(\xi, \xi) vec(d\bar{\eta}) + tr\{db' \int_0^1 F(s) F'(s) ds d\bar{b} \Sigma_{w,w}(\xi, \xi)\} \tag{C.6}$$
$$+ \int_0^1 F(s)' ds db \Sigma_{w,u}(\xi, \xi) vec(d\bar{\eta}) + vec(d\eta)' \Sigma_{u,w}(\xi, \xi) db' \int_0^1 F(s) ds,$$
$$H_{\theta\xi,\infty}(\vartheta, d\theta, d\xi) = vec(d\eta)' \Sigma_{u,v}(\xi, \xi) vec(d\xi) + vec\left(db'\right)' \left(\int F ds \otimes \Sigma_{u,v}(\xi, \xi)\right) vec(d\xi)$$
$$H_{\xi\xi,\infty}(\vartheta, d\xi, d\xi) = vec(d\xi)' \Sigma_{v,v}(\xi, \xi) vec(d\xi).$$

   Here Σ is defined in (4.2) and $(B_u', B_v', B_w', F')'$ in (4.1).

2. *If $\delta_0 = 0$, then with $q_2 = \max(4, 2d_\xi)$ and $q_0 > q_2 \max(1, \rho)$, it holds uniformly over $\xi \in \Xi$,*

$$-\frac{1}{T} d^2 L_T(\theta_0, \xi; V_{\theta,T}^{-1/2} d\theta, V_{\theta,T}^{-1/2} d\bar{\theta}) \xrightarrow{D} H_{\theta\theta,\infty}(\xi; d\theta, d\bar{\theta}),$$

   where $H_{\theta\theta,\infty}(\xi; d\theta, d\bar{\theta})$ given by (C.6) is evaluated at η_0 (with $\delta_0 = 0$).

Proof. Note that,

$$-\frac{1}{T}d^2L_T(\vartheta_0;V_T^{-1/2}d\vartheta,V_T^{-1/2}d\bar{\vartheta}) = H_{\theta\theta,T}(\vartheta_0;V_{\theta,T}^{-1/2}d\theta,V_{\theta,T}^{-1/2}d\bar{\theta}) + H_{\theta\xi,T}(\vartheta_0;V_{\theta,T}^{-1/2}d\theta,V_T^{-1/2}d\bar{\xi})$$
$$+ H_{\xi\theta,T}(\vartheta_0;V_{\xi,T}^{-1/2}d\xi,V_{\theta,T}^{-1/2}d\bar{\theta}) + H_{\xi\xi,T}(\vartheta_0;V_{\xi,T}^{-1/2}d\xi,V_{\xi,T}^{-1/2}d\bar{\xi}),$$

with

$$\begin{aligned}
H_{\theta\theta,T}(\vartheta_0;V_{\theta,T}^{-1/2}d\theta,V_{\theta,T}^{-1/2}d\bar{\theta}) &= H_{\eta,\eta}(\xi) + H_{b,b}\left(\xi,K_T^{-1}db,K_T^{-1}d\bar{b}\right) + H_{\eta,b}\left(\xi,K_T^{-1}d\bar{b}\right)\\
&\quad + H_{b,\eta}\left(\xi,K_T^{-1}db\right),\\
H_{\xi\theta,T}(\vartheta_0;V_{\xi,T}^{-1/2}d\xi,V_{\theta,T}^{-1/2}d\bar{\theta}) &= H_{\xi,\eta}(\xi) + H_{\xi,b}\left(\xi,K_T^{-1}d\bar{b}\right),\\
H_{\xi\xi,T}(\vartheta_0;V_{\xi,T}^{-1/2}d\xi,V_{\xi,T}^{-1/2}d\bar{\xi}) &= H_{\xi,\xi}(\xi),
\end{aligned}$$

where we suppress dependence on all directions except db and have used the notation that $H_{\eta,b}(\xi,db) = -\frac{1}{T}d^2L_T(\theta_0,\xi;d\eta,d\bar{b})$ and so forth.

Proof of part 2 ($\delta_0 = 0$)**:** First, consider $H_{\theta\theta,T}$ at $\theta_0 = (0,\eta_0)$, with $\delta_0 = 0$ and ξ freely varying. The following claims are shown to hold uniformly over $\xi \in \Xi$:

$$\begin{aligned}
&\text{Claim 2.1}: H_{\eta,\eta}(\xi) \xrightarrow{P} vec(d\eta)'\,\Sigma_{uu}(\xi)\,vec(d\bar{\eta}),\\
&\text{Claim 2.2}: H_{b,b}\left(\xi,K_T^{-1}db,K_T^{-1}d\bar{b}\right) \xrightarrow{D} tr\{(db)'\int_0^1 F(s)\,F(s)'\,ds(d\bar{b})\Sigma_{w,w}\},\\
&\text{Claim 2.3}: H_{b,\eta}\left(\xi,K_T^{-1}db\right) \xrightarrow{D} \int_0^1 F(s)'\,dsdb\Sigma_{w,u}(\xi,\xi)\,vec(d\bar{\eta})
\end{aligned}$$

Proof of Claim 2.1: We have

$$H_{\eta,\eta}(\xi) = \frac{1}{T}\sum_{t=1}^{T}\left[d\alpha(Z_t(b)) + d\delta\psi(Z_t(b);\xi) + d\Phi Z_{2,t-1}\right]'\Omega_0^{-1} \tag{C.7}$$
$$\times\left[d\bar{\alpha}(Z_t(b)) + d\bar{\delta}\psi(Z_t(b)\ \xi) + d\bar{\Phi}Z_{2,t-1}\right],$$

Evaluated at θ_0,

$$H_{\eta,\eta}(\xi) = \frac{1}{T}vec(d\eta)'\sum_{t=1}^{T}\left[\left(Z_{0,t-1}',\psi(Z_{0,t-1};\xi)',Z_{2,t-1}'\right)'\left(Z_{0,t-1}',\psi(Z_{0,t-1};\xi)',Z_{2,t-1}'\right)\otimes\Omega_0^{-1}\right]$$
$$\times\, vec(d\bar{\eta}),$$

and the result follows by the uniform law of large numbers in Kristensen and Rahbek (2005).

*Proof of Claim 2.2***:** Next, $H_{b,b}\left(\xi,db,d\bar{b}\right) = H_{b,b}^{(1)}\left(\xi,db,d\bar{b}\right) + H_{b,b}^{(2)}\left(\xi,db,d\bar{b}\right)$, where

$$\begin{aligned}
H_{b,b}^{(1)}\left(\xi,db,d\bar{b}\right) &= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\theta)'\,\Omega_0^{-1}\delta\left(Z_{1,t-1}'db\otimes I_{r_\delta}\right)\partial_{zz}^2\psi(Z_t(b);\xi)\,d\bar{b}'Z_{1,t-1} &\text{(C.8)}\\
&= \frac{1}{T}\sum_{t=1}^{T}tr\{vec\left(Z_{1,t-1}'db\otimes I_{r_\delta}\right)vec(Z_{1,t-1}'d\bar{b})'\left(\partial_z^2\psi(Z_t(b);\xi)\,'\otimes\varepsilon_t(\theta)'\,\Omega_0^{-1}\delta\right)\},
\end{aligned}$$

with

$$\partial_{zz}^2 \psi (z, \xi) = \partial vec (\partial_z \psi (z, \xi)) / \partial z',$$

and

$$H_{b,b}^{(2)} \left(\xi, db, d\bar{b}\right) = \frac{1}{T} \sum_{t=1}^{T} \left[\{\alpha + \delta \partial_z \psi (Z_t (b) ; \xi)\} Z_{1,t-1}' d\bar{b} \right]' \Omega_0^{-1} \left[\{\alpha + \delta \partial_z \psi (Z_t (b) ; \xi)\} Z_{1,t-1}' db \right]. \tag{C.9}$$

At θ_0 with $\delta_0 = 0, H_{b,b}^{(1)} \left(\xi, db, d\bar{b}\right) = 0$, such that by Corollary 3.7,

$$\begin{aligned} H_{b,b} \left(\xi, K_T^{-1} db, K_T^{-1} d\bar{b}\right) &= T^{-1} \sum_{t=1}^{T} \left[\alpha_0 d\bar{b}' K_T^{-1} Z_{1,t-1} \right]' \Omega_0^{-1} \alpha_0 db' K_T^{-1} Z_{1,t-1} \\ &\xrightarrow{D} tr \left\{ d\bar{b}' \int_0^1 FF' ds db \Sigma_{w,w} \right\} \end{aligned}$$

with $\Sigma_{w,w} = Var (w_t) = \alpha_0' \Omega_0^{-1} \alpha_0$.

Proof of Claim 2.3: We write $H_{b,\eta} (\xi, db) = H_{b,\eta}^{(1)} (\xi, db) + H_{b,\eta}^{(2)} (\xi, db)$, where

$$H_{b,\eta}^{(1)} (\xi, db) = \frac{1}{T} \sum_{t=1}^{T} \left[\left\{ d\bar{\alpha} + d\bar{\delta} \partial_z \psi (Z_t (b) ; \xi) \right\} db' Z_{1,t-1} \right]' \Omega_0^{-1} \varepsilon_t (\theta) \tag{C.10}$$

$$H_{b,\eta}^{(2)} (\xi, db) = \frac{1}{T} \sum_{t=1}^{T} \left[\{\alpha + \delta \partial_z \psi (Z_t (b) ; \xi)\} db' Z_{1,t-1} \right]' \Omega_0^{-1} \left[d\bar{\alpha} Z_{0,t-1} + d\bar{\delta} \psi (Z_t (b) ; \xi) + d\bar{\Phi} Z_{2,t-1} \right] \tag{C.11}$$

With $\theta = \theta_0$ (such that in particular $b = 0$), set $f_{t-1}^{(1)} (\xi) = (I_r, \partial_z \psi (Z_{0,t-1}; \xi), 0)$ and $e_t = \Omega_0^{-1} \varepsilon_t$, then

$$H_{b,\eta}^{(1)} (\xi, db) = \frac{1}{T} \sum_{t=1}^{T} Z_{1,t-1}' db f_{t-1}^{(1)} (\xi) d\bar{\eta}' e_t.$$

By the same arguments as in the proof of Lemma 4.5, we see that $f_{t-1}^{(1)} (\xi)$ satisfies the conditions of Theorem 3.5 with

$$\left\| f^{(1)} (y_{t-1}; \xi) \right\| \leq c \left(1 + \| \partial_z \psi (Z_{0,t-1}; \xi) \| \right) \leq c \left(1 + \| Z_{0,t-1} \|^{\rho} \right)$$

Thus, $\left\| f^{(1)} (y_{t-1}; \xi) \right\|^m \leq c \left(1 + \| Z_{0,t-1} \|^{m\rho} \right)$. Furthermore, by the differentiability of ψ,

$$\begin{aligned} E \left[\left\| f^{(1)} (y_{t-1}; \xi) - f^{(1)} \left(y_{t-1}; \xi'\right) \right\|^n \right] &= E \left[\left\| \partial_z \psi (Z_{0,t-1}; \xi) - \partial_z \psi \left(Z_{0,t-1}; \xi'\right) \right\|^n \right] \\ &\leq E \left[\left\| \frac{\partial \left[\partial_z \psi \left(Z_{0,t-1}; \bar{\xi}\right) \right]}{\partial \xi} \right\|^n \right] \left\| \xi - \xi' \right\|^n \\ &\leq E \left[\| Z_{0,t-1} \|^{\rho n} \right] \left\| \xi - \xi' \right\|^n, \end{aligned}$$

such that $\lambda = 1$. Thus, the requirement $n > 2$ translates into $E \left[\| Z_{0,t-1} \|^{(2+\delta)\rho} \right] < \infty$ for some $\delta > 0$, and the requirement $m > \bar{m} := \max (4, 2d_\xi)$ translates into $E \left[\| Z_{0,t-1} \|^{\bar{m}\bar{\rho}} \right] < \infty$ with $\bar{\rho} = \max (1, \rho)$. Theorem 3.5 now implies that $H_{b,\eta}^{(1)} \left(\xi, K_T^{-1} db\right) = O_P (1)$ and hence,

$H^{(1)}_{b,\eta}\left(\xi, K_T^{-1}db\right) = o_P(1)$, uniformly in ξ.
Consider $H^{(2)}_{b,\eta}(\xi, db)$ and observe that,

$$H^{(2)}_{b,\eta}(\xi, db) = \frac{1}{T}\sum_{t=1}^{T}\left[\alpha db' Z_{1,t-1}\right]' f^{(2)}_{t-1}(\xi)$$

where

$$f^{(2)}_{t-1}(\xi) = \Omega_0^{-1}\left[d\bar{\alpha}Z_{0,t-1} + d\bar{\delta}\psi\left(Z_t(b);\xi\right) + d\bar{\Phi}Z_{2,t-1}\right]$$

Applying Theorem 3.6 gives at θ_0,

$$H^{(2)}_{b,\eta}\left(\xi, K_T^{-1}db\right) \xrightarrow{D} \int_0^1 F(s)' ds db \Sigma_{w,u}(\xi,\xi) vec(d\bar{\eta}).$$

This finishes the proof of part 2.

Proof of part 1 ($\delta_0 \neq 0$)**:** We state the needed as claims again:

$$\text{Claim 1.1}: H_{\eta,\eta}(\xi_0) \xrightarrow{P} vec(d\eta)' \Sigma_{u,u}(\xi_0,\xi_0) vec(d\bar{\eta}),$$

$$\text{Claim 1.2}: H_{b,b}\left(\xi_0, K_T^{-1}db, K_T^{-1}d\bar{b}\right) \xrightarrow{D} tr\{(db)' \int_0^1 F(s) F(s)' ds (d\bar{b}) \Sigma_{w,w}(\xi_0,\xi_0)\},$$

$$\text{Claim 1.3}: H_{b,\eta}\left(\xi_0, K_T^{-1}db\right) \xrightarrow{D} \int_0^1 F(s)' ds db \Sigma_{w,u}(\xi_0,\xi_0) vec(d\bar{\eta})$$

$$\text{Claim 1.4}: H_{\eta,\xi}(\xi_0) \xrightarrow{P} vec(d\eta)' \Sigma_{u,v}(\xi_0,\xi_0) vec\left(d\bar{\xi}\right)$$

$$\text{Claim 1.5}: H_{b,\xi}\left(\xi_0, K_T^{-1}db\right) \xrightarrow{D} \int_0^1 F(s)' ds db \Sigma_{w,v}(\xi_0,\xi_0) vec\left(d\bar{\xi}\right).$$

$$\text{Claim 1.6}: H_{\xi,\xi}(\xi_0) \xrightarrow{P} vec(d\xi)' \Sigma_{v,v}(\xi_0,\xi_0) vec\left(d\bar{\xi}\right).$$

Proof of Claims 1.1-1.3: They follow as before for claims 2.1-2.3.
Proof of Claim 1.4: The differential $H_{\eta,\xi}(\xi)$ takes the form $H_{\eta,\xi}(\xi) = H^{(1)}_{\eta,\xi}(\xi) + H^{(2)}_{\eta,\xi}(\xi)$

$$H^{(1)}_{\eta,\xi}(\xi) = -\frac{1}{T}\sum_{t=1}^{T}\left[d\delta\partial_\xi\psi\left(Z_t(b);\xi\right) vec\left(d\bar{\xi}\right)\right]' \Omega_0^{-1}\varepsilon_t(\theta),$$

$$H^{(2)}_{\eta,\xi}(\xi) = \frac{1}{T}\sum_{t=1}^{T}\left[d\alpha Z_{0,t-1} + d\delta\psi\left(Z_t(b);\xi\right) + d\Phi Z_{2,t-1}\right]' \Omega_0^{-1}\delta\partial_\xi\psi\left(Z_t(b);\xi\right) vec\left(d\bar{\xi}\right).$$

By Corollary 3.7, at ϑ_0, $T^{1/2}H^{(1)}_{\eta,\xi}(\xi) \xrightarrow{D} Y(1,\xi)$ for an appropriately defined Gaussian process $Y(s,\xi)$, while by Corollary 3.7 with $x_{T,t-1} = 1$, $H^{(2)}_{\eta,\xi}(\xi_0) \xrightarrow{D} d\eta'\Sigma_{u,v}(\xi_0,\xi_0) d\xi$.

Proof of Claim 1.5: The differential $H_{b,\xi}(\xi, db) = H^{(1)}_{b,\xi}(\xi, db) + H^{(2)}_{b,\xi}(\xi, db)$ where, similar to the proof of Claim 1.2, with

$$\partial^2_{z,\xi}\psi(z,\xi) = \frac{\partial vec\left(\partial_z\psi(z,\xi)\right)}{\partial vec(\xi)'}, \tag{C.12}$$

we find,

$$H^{(1)}_{b,\xi}(\xi, db) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\theta)' \Omega_0^{-1}\delta\left(Z'_{1,t-1}db \otimes I_{r_\delta}\right)\partial^2_{z,\xi}\psi\left(Z_t(b);\xi\right) vec\left(d\bar{\xi}\right), \tag{C.13}$$

$$H_{b,\xi}^{(2)}(\xi, db) = \frac{1}{T}\sum_{t=1}^{T}\left[\{\alpha + \delta\partial_z\psi(Z_t(b);\xi)\} db' Z_{1,t-1}\right]' \Omega_0^{-1}\delta\partial_\xi\psi(Z_t(b);\xi)\, vec\left(d\bar{\xi}\right), \quad \text{(C.14)}$$

By Corollary 3.7, at ϑ_0, $H_{b,\xi}^{(1)}\left(\xi_0, K_T^{-1}db\right) \xrightarrow{P} 0$ and $H_{b,\xi}^{(2)}\left(\xi_0, K_T^{-1}db\right)$ converges towards the claimed limit.

Proof of Claim 1.6: The differential $H_{\xi,\xi}(\xi) = H_{\xi,\xi}^{(1)}(\xi) + H_{\xi,\xi}^{(2)}(\xi)$, where

$$H_{\xi,\xi}^{(1)}(\xi) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\theta)' \Omega_0^{-1}\delta\left(vec(d\xi)' \otimes I_{r_\delta}\right)\partial_{\xi,\xi}^2\psi(Z_t(b);\xi)\, vec\left(d\bar{\xi}\right),$$

with

$$\partial_{\xi,\xi}^2\psi(z,\xi) = \frac{\partial vec(\partial_\xi\psi(z,\xi))}{\partial vec(\xi)'},$$

and

$$H_{\xi,\xi}^{(2)}(\xi) = (vec(d\xi))' \frac{1}{T}\sum_{t=1}^{T}\partial_\xi\psi(Z_t(b);\xi)'\delta'\Omega_0^{-1}\delta\partial_\xi\psi(Z_t(b);\xi)\, vec\left(d\bar{\xi}\right).$$

It follows by Corollary 3.7 that at ϑ_0, $H_{\xi,\xi}^{(1)}(\xi_0) \xrightarrow{P} 0$ and $H_{\xi,\xi}^{(2)}(\xi_0) \xrightarrow{P} d\xi'\Sigma_{v,v}(\xi_0,\xi_0)\, d\bar{\xi}$. ■

Lemma C.3 *Assume that Assumptions 4.1-4.4 hold. With $d\vartheta = (d\theta, d\xi)$ and $d\theta = (d\eta, db)$ and the log-likelihood function $L_T(\theta,\xi)$ defined in (2.11), the following hold:*

1. *If $\delta_0 \neq 0$, then with $q_0 > 2\max\{1,\rho\}$ and $q_2 > 2$:*

$$\sup_{\vartheta\in N_T(\vartheta_0)}\left|\frac{1}{T}d^3L_T(\vartheta, V_T^{-1/2}d\vartheta, V_T^{-1/2}d\bar{\vartheta}, V_T^{-1/2}d\check{\vartheta})\right| = O_P(||d\vartheta||||d\bar{\vartheta}||||d\check{\vartheta}||)$$

   for a sequence of neighborhoods

$$\mathcal{N}_T(\vartheta_0) = \{\vartheta : ||\eta - \eta_0|| < \epsilon,\ ||\xi - \xi_0|| < \epsilon \ \text{ and } \ ||K_T b|| < \epsilon\}.$$

2. *If $\delta_0 = 0$, then with $q_0 > 2\max\{1,\rho\}$ and $q_2 > 2$:*

$$\sup_{\substack{\theta\in N_T(\theta_0)\\ \xi\in\Xi}}\left|\frac{1}{T}d^3L_T(\theta,\xi, V_{\theta,T}^{-1/2}d\theta, V_{\theta,T}^{-1/2}d\bar{\theta}, V_{\theta,T}^{-1/2}d\check{\theta})\right| = O_P(||d\theta||||d\bar{\theta}||||d\check{\theta}||)$$

   for a sequence of neighborhoods

$$\mathcal{N}_T(\theta_0) = \{\theta : ||\eta - \eta_0|| < \epsilon, \quad \text{and} \quad ||K_T b|| < \epsilon\}.$$

Proof of Lemma C.3. Write the third order differential as,

$$\frac{1}{T}d^3L_T(\theta,\xi,d\theta,d\bar{\theta},d\tilde{\theta}) = \sum_{i,j} d\left(H_{\theta_i,\bar{\theta}_j}(\xi), d\tilde{\theta}\right).$$

Below we consider each of the terms normalized as indicated in the lemma and argue that they are $O_P(1)$ as $T \to \infty$ as desired. We focus on the most difficult cases when $\delta_0 = 0$, and third order derivatives are considered w.r.t. b and ξ. The remaining cases ($\delta_0 \neq 0$ and derivatives in other directions) proceeds in a completely analogous manner, and only differ in terms of notation.

Claim 1: $\sup_{\xi}\left\|d\left(H_{\xi,\xi}\left(\xi\right),d\tilde{\xi}\right)\right\| = O_P\left(1\right)$. From the proof of Lemma C.2, recall that, the differential $H_{\xi,\xi}\left(\xi\right) = H^{(1)}_{\xi,\xi}\left(\xi\right) + H^{(2)}_{\xi,\xi}\left(\xi\right)$, where

$$H^{(1)}_{\xi,\xi}\left(\xi\right) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\left(\theta,\xi\right)'\Omega_0^{-1}\delta\left(vec\left(d\xi\right)'\otimes I_{r_\delta}\right)\partial^2_{\xi,\xi}\psi\left(;\xi\right)vec\left(d\bar{\xi}\right),$$

$$H^{(2)}_{\xi,\xi}\left(\xi\right) = \frac{1}{T}\left(vec\left(d\xi\right)\right)'\sum_{t=1}^{T}\partial_{\xi}\psi\left(Z_t\left(b\right);\xi\right)'\delta'\Omega_0^{-1}\delta\partial_{\xi}\psi\left(Z_t\left(b\right);\xi\right)vec\left(d\bar{\xi}\right),$$

with $Z_t\left(b\right)$ defined in (C.2). Thus,

$$\begin{aligned}
&d\left(H^{(1)}_{\xi,\xi}\left(\xi\right),d\tilde{\xi}\right)\\
&= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\left(\theta,\xi\right)'\Omega_0^{-1}\delta\left(vec\left(d\xi\right)'\otimes I_{r_\delta}\right)\left(vec\left(d\bar{\xi}\right)'\otimes I\right)\partial^3_{\xi\xi\xi}\psi\left(Z_t\left(b\right);\xi\right)vec\left(d\tilde{\xi}\right)\\
&- \frac{1}{T}\sum_{t=1}^{T}\left(\delta\partial_{\xi}\psi\left(Z_t\left(b\right)\right)vec\left(d\xi\right)\right)'\Omega_0^{-1}\delta\left(vec\left(d\bar{\xi}\right)'\otimes I_{r_\delta}\right)\partial^2_{\xi,\xi}\psi\left(Z_t\left(b\right);\xi\right)vec\left(d\tilde{\xi}\right),
\end{aligned}$$

where

$$\partial^3_{\xi\xi\xi}\psi\left(z;\xi\right) = \frac{\partial vec\left(\partial^2_{\xi\xi}\psi\left(z;\xi\right)\right)}{\partial vec\left(\xi\right)'}.$$

Likewise,

$$\begin{aligned}
&Td\left(H^{(2)}_{\xi,\xi}\left(\xi\right),d\tilde{\xi}\right)\\
&= \left(vec\left(d\xi\right)\right)'\sum_{t=1}^{T}\partial_{\xi}\psi\left(Z_t\left(b\right);\xi\right)'\delta'\Omega_0^{-1}\delta\left(vec\left(d\bar{\xi}\right)'\otimes I\right)\partial^2_{\xi\xi}\psi\left(Z_t\left(b\right);\xi\right)vec\left(d\tilde{\xi}\right)\\
&+ \left(vec\left(d\bar{\xi}\right)\right)'\sum_{t=1}^{T}\partial_{\xi}\psi\left(Z_t\left(b\right);\xi\right)'\delta'\Omega_0^{-1}\delta\left(vec\left(d\xi\right)'\otimes I\right)\partial^2_{\xi\xi}\psi\left(Z_t\left(b\right);\xi\right)vec\left(d\tilde{\xi}\right).
\end{aligned}$$

Hence, by Assumption 4.1,

$$\begin{aligned}
\left\|d\left(H^{(1)}_{\xi,\xi}\left(\xi\right),d\tilde{\xi}\right)\right\| &\leq c\left\|d\xi\right\|\left\|d\bar{\xi}\right\|\left\|d\tilde{\xi}\right\|\frac{1}{T}\sum_{t=1}^{T}\left\|\varepsilon_t\left(\theta,\xi\right)\right\|\left(1+\left|\left|Z_t\left(b\right)\right|\right|^{\rho}\right)\\
&\leq c\left\|d\xi\right\|\left\|d\bar{\xi}\right\|\left\|d\tilde{\xi}\right\|\\
&\times\frac{1}{T}\sum_{t=1}^{T}\left(\left\|\varepsilon_t\right\| + \left|\left|Z_{0,t-1}\right|\right| + \left|\left|Z_{2,t-1}\right|\right| + \left|\left|Z_t\left(b\right)\right|\right|\right)\left(1+\left|\left|Z_t\left(b\right)\right|\right|^{\rho}\right).
\end{aligned}$$

Next, note that with $\theta\in\mathcal{N}_T(\theta_0)$, we can write, $b = K_T^{-1}h$,where $||h|| < \epsilon$, and hence,

$$\left|\left|Z_t\left(b\right)\right|\right| \leq \left|\left|Z_{0,t-1}\right|\right| + \epsilon\left\|K_T^{-1}Z_{1,t-1}\right\| \leq \left|\left|Z_{0,t-1}\right|\right| + \epsilon\sup_{u\in[0,1]}\left\|K_T^{-1}Z_{1,[Tu]}\right\|. \tag{C.15}$$

As $\sup_{u\in[0,1]}\left\|K_T^{-1}Z_{1,[Tu]}\right\| = O_P\left(1\right)$, we get by the LLN (Kristensen and Rahbek, 2005),

$$\left\|d\left(H^{(1)}_{\xi,\xi}\left(\xi\right),d\tilde{\xi}\right)\right\| = O_P\left(\left\|d\xi\right\|\left\|d\bar{\xi}\right\|\left\|d\tilde{\xi}\right\|\right).$$

Claim 2: $\sup_\xi \left\| d\left(H_{\xi,\xi}(\xi), K_T^{-1}d\tilde{b}\right)\right\| = O_P(1)$. As in Claim 1, given the expression of $H_{\xi,\xi}(\xi)$,

$$
\begin{aligned}
& d\left(H_{\xi,\xi}^{(1)}(\xi), d\tilde{b}\right) \\
& = -\frac{1}{T}\sum_{t=1}^{T} Z_{1,t-1}' d\tilde{b}\left[\alpha + \delta\partial_z\psi(Z_t(b);\xi)\right]' \Omega_0^{-1}\delta\left(vec(d\xi)'\otimes I_{r_\delta}\right)\partial_{\xi,\xi}^2\psi(Z_t(b);\xi)\, vec\left(d\bar{\xi}\right), \\
& + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\theta,\xi)'\Omega_0^{-1}\delta\left(vec(d\xi)'\otimes I_{r_\delta}\right)\left(vec\left(d\bar{\xi}\right)'\otimes I\right)\partial_{\xi\xi,z}^3\psi(Z_t(b);\xi)\, d\tilde{b}' Z_{1,t-1},
\end{aligned}
$$

and

$$
\begin{aligned}
& d\left(H_{\xi,\xi}^{(2)}(\xi), d\tilde{b}\right) \\
& = \frac{1}{T}(vec(d\xi))'\sum_{t=1}^{T}\partial_\xi\psi(Z_t(b);\xi)'\delta'\Omega_0^{-1}\delta\left(vec\left(d\bar{\xi}\right)'\otimes I\right)\partial_{\xi,z}^2\psi(Z_t(b);\xi)\, d\tilde{b}' Z_{1,t-1} \\
& + \frac{1}{T}\sum_{t=1}^{T}\left[\left(vec(d\xi)'\otimes I\right)\partial_{\xi,z}^2\psi(Z_t(b);\xi)\, d\tilde{b}' Z_{1,t-1}\right]'\delta'\Omega_0^{-1}\delta\partial_\xi\psi(Z_t(b);\xi)\, vec\left(d\bar{\xi}\right),
\end{aligned}
$$

where $\partial_{\xi,z}^2\psi(Z_t(b);\xi)$ is defined in eq. (C.12), and

$$
\partial_{\xi\xi z}^3\psi(z;\xi) = \frac{\partial vec\left(\partial_{\xi\xi}^2\psi(z;\xi)\right)}{\partial z'}.
$$

Thus,

$$
\begin{aligned}
& \left| d\left(H_{\xi,\xi}^{(1)}(\xi), K_T^{-1}d\tilde{b}\right)\right| \\
& \leq c\|d\xi\|\left\|d\bar{\xi}\right\|\frac{1}{T}\sum_{t=1}^{T}\left\|Z_{1,t-1}'K_T^{-1}d\tilde{b}\right\|\left(1+\|\partial_z\psi(Z_t(b);\xi)\|\right)\left\|\partial_{\xi,\xi}^2\psi(Z_t(b);\xi)\right\| \\
& + c\|d\xi\|\left\|d\bar{\xi}\right\|\frac{1}{T} \\
& \sum_{t=1}^{T}\|\varepsilon_t(\theta,\xi)\|\left\|\partial_{\xi\xi,z}^3\psi(Z_t(b);\xi)\right\|\left\|d\tilde{b}'K_T^{-1}Z_{1,t-1}\right\| \\
& \leq c\|d\xi\|\left\|d\bar{\xi}\right\|\frac{1}{T}\sum_{t=1}^{T}\left\|Z_{1,t-1}'K_T^{-1}\right\|\left(1+\|Z_t(b)\|^\rho\right)^2 \\
& + c\|d\xi\|\left\|d\bar{\xi}\right\|\frac{1}{T}\sum_{t=1}^{T}\left(\|\varepsilon_t\|+||Z_{0,t-1}||+||Z_{2,t-1}||\right)\left(1+\|Z_t(b)\|^\rho\right)\left\|K_T^{-1}Z_{1,t-1}\right\|,
\end{aligned}
$$

and so $\left| d\left(H_{\xi,\xi}^{(1)}(\xi), K_T^{-1}d\tilde{b}\right)\right| = O_P(1)$ by eq. (C.15).

By identical arguments, $\left| d\left(H_{\xi,\xi}^{(2)}(\xi), K_T^{-1}d\tilde{b}\right)\right| = O_P(1)$.

Claim 3: $\sup_\xi \left\| d\left(H_{b,b}(\xi), K_T^{-1} d\tilde{b}\right)\right\| = O_P(1)$. Given the expression of $H_{b,b}(\xi)$ in the Proof of Lemma C.2, $d\left(H_{b,b}(\xi), d\tilde{b}\right) = d\left(H_{b,b}^{(1)}(\xi), d\tilde{b}\right) + d\left(H_{b,b}^{(2)}(\xi), d\tilde{b}\right)$, where

$$
\begin{aligned}
& d\left(H_{b,b}^{(1)}(\xi), d\tilde{b}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} Z_{1,t-1}' d\tilde{b} \left[\alpha + \delta \partial_z \psi(Z_t(b);\xi)\right]' \Omega_0^{-1} \delta \left(Z_{1,t-1}' db \otimes I_{r_\delta}\right) \partial_{zz}^2 \psi(Z_t(b);\xi) d\bar{b}' Z_{1,t-1} \\
&+ \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t(\theta,\xi)' \Omega_0^{-1} \delta \left(Z_{1,t-1}' db \otimes I_{r_\delta}\right)\left(Z_{1,t-1}' d\bar{b} \otimes I_{r_\delta}\right) \partial_{zzz}^3 \psi(Z_t(b);\xi) d\tilde{b}' Z_{1,t-1}
\end{aligned}
$$

with

$$
\partial_{zzz}^3 \psi(z,\xi) = \partial vec\left(\partial_{zz}^2 \psi(z,\xi)\right)/\partial z',
$$

and

$$
\begin{aligned}
& d\left(H_{b,b}^{(2)}(\xi), d\tilde{b}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} \left[(I \otimes \delta) \partial_{zz}^2 \psi(Z_t(b);\xi) d\tilde{b}' Z_{1,t-1} Z_{1,t-1}' d\bar{b}\right]' \Omega_0^{-1} \left[\{\alpha + \delta \partial_z \psi(Z_t(b);\xi)\} Z_{1,t-1}' db\right] \\
&+ \frac{1}{T}\sum_{t=1}^{T} \left[\{\alpha + \delta \partial_z \psi(Z_t(b);\xi)\} Z_{1,t-1}' d\bar{b}\right]' \Omega_0^{-1} \left[(I \otimes \delta) \partial_{zz}^2 \psi(Z_t(b);\xi) d\tilde{b}' Z_{1,t-1} Z_{1,t-1}' db\right].
\end{aligned}
$$

Thus, multiplying all directions with K_T^{-1} and using eq. (C.15),

$$
\begin{aligned}
\left| d\left(H_{b,b}^{(1)}(\xi), K_T^{-1} d\tilde{b}\right)\right| &\leq \frac{c}{T} \|db\| \|d\bar{b}\| \left\|d\tilde{b}\right\| \sum_{t=1}^{T} \left\|K_T^{-1} Z_{1,t-1}\right\|^3 \left[1 + \|Z_t(b)\|^\rho\right] \|Z_t(b)\|^\rho \\
&= O_P(1),
\end{aligned}
$$

and, by identical arguments, $\left| d\left(H_{b,b}^{(2)}(\xi), K_T^{-1} d\tilde{b}\right)\right| = O_P(1)$. ■

D Auxiliary Lemmas

Consider $Q_T(\gamma,\pi)$ which is a function of observations $X_1, ..., X_T$ and parameters $\gamma \in \Gamma \subseteq \mathbb{R}^d$ and $\pi \in \Pi \subseteq \mathbb{R}^k$. Introduce furthermore γ_0, which is an interior point of Γ. We then state conditions under which $\hat{\gamma}(\pi) = \arg\min_{\gamma\in\Gamma} Q_T(\gamma,\pi)$ is consistent and has a well-defined asymptotic distribution. The proof is based on standard expansions of the likelihood function similar to Kristensen and Rahbek (2010). However, the objective function, and thereby the estimator, depends on a nuisance parameter π, and we state results that hold uniformly over $\pi \in \Pi$. In the following, for any constant c, let $W_T \xrightarrow{D} c$ denote convergence in distribution towards the degenerate distribution at c.

Lemma D.1 *Assume that:*

(i) $Q_T(\cdot,\pi) : \mathbb{R}^d \to \mathbb{R}$ *is three times continuously differentiable in* γ *for all* π.

(ii) *There exists a sequence of nonsingular matrices $U_T \in \mathbb{R}^{d\times d}$ such that $U_T^{-1} = O(1)$ and*

$$\left(dQ_T\ (\gamma_0, \pi; U_T^{-1/2} d\gamma), d^2 Q_T\ (\gamma_0, \pi; U_T^{-1/2} d\gamma, U_T^{-1/2} d\gamma)\right)_{\gamma=\gamma_0} \xrightarrow{D} (0, H_\infty(\pi, d\gamma, d\bar{\gamma})),$$

where the convergence takes place on Π, and where the stochastic process $H_\infty(\pi, d\gamma, d\bar{\gamma}) >$ 0 a.s.

(iii) $\sup_{\pi\in\Pi} \sup_{\gamma\in N_T(\gamma_0(\pi))} \left| d^3 Q_T\ (\gamma, \pi; U_T^{-1/2} d\gamma, U_T^{-1/2} d\bar{\gamma}, U_T^{-1/2} d\tilde{\gamma})\right| = O_P(||d\gamma|| ||d\bar{\gamma}|| ||d\tilde{\gamma}||)$ *over the sequence of local neighborhoods*

$$\mathcal{N}_T(\gamma_0) = \left\{\gamma : ||U_T^{1/2}(\gamma - \gamma_0)|| < \epsilon\right\}.$$

Then with probability tending to one, for any $\pi \in \Pi$, there exists a unique minimum point $\hat{\gamma}(\pi)$ of $Q_T(\gamma, \pi)$ in $\mathcal{N}_T(\gamma_0)$ which solves $\partial Q_T(\hat{\gamma}(\pi), \pi)/\partial\gamma = 0$.

It satisfies $\sup_{\pi\in\Pi} \left\| U_T^{1/2}(\hat{\gamma}(\pi) - \gamma_0)\right\| = o_P(1)$.

Proof of Lemma D.1. Use a second order Taylor expansion to obtain for any bounded sequence $d_T(\pi) \in \mathbb{R}^d$ such that $\gamma_0 + U_T^{-1/2} d_T(\pi) \in N_T(\gamma_0)$,

$$\begin{aligned} Q_T\ (\gamma_0 + U_T^{-1/2} d_T(\pi), \pi) - Q_T\ (\gamma_0, \pi) &= dQ_T\ (\gamma_0, \pi; U_T^{-1/2} d_T(\pi)) \\ &+ \frac{1}{2} d^2 Q_T\ (\bar{\gamma}(\pi), \pi; U_T^{-1/2} d_T(\pi), U_T^{-1/2} d_T(\pi)), \end{aligned}$$

for some $\bar{\gamma}(\pi) \in [\gamma_0, \gamma_0 + U_T^{-1/2} d_T(\pi)] \in \mathcal{N}_T(\gamma_0)$. Define the bounded sequence $\bar{d}_T(\pi) = U_T^{1/2}(\bar{\gamma}(\pi) - \gamma_0)$. Then, by another application of Taylor's Theorem, there exists $\tilde{\gamma}(\pi) \in [\gamma_0, \bar{\gamma}(\pi)] \in \mathcal{N}_T(\gamma_0)$ such that

$$\begin{aligned} &\sup_{\pi\in\Pi} \left| d^2 Q_T\ (\bar{\gamma}(\pi), \pi; U_T^{-1/2} d_T(\pi), U_T^{-1/2} d_T(\pi)) - d^2 Q_T\ (\gamma_0, \pi; U_T^{-1/2} d_T(\pi), U_T^{-1/2} d_T(\pi))\right| \\ &= \sup_{\pi\in\Pi} \left| d^3 Q_T\ (\tilde{\gamma}(\pi), \pi; U_T^{-1/2} d_T(\pi), U_T^{-1/2} d_T(\pi), U_T^{-1/2} \bar{d}_T(\pi))\right| \\ &= O_P\left(\|d_T(\pi)\|^2 \|\bar{d}_T(\pi)\|\right) = O_P(\epsilon^3), \end{aligned}$$

where we have used (iii). Thus,

$$\begin{aligned} &Q_T\ (\gamma_0 + U_T^{-1/2} d_T(\pi), \pi) - Q_T\ (\gamma_0, \pi) \\ &= dQ_T\ (\gamma_0, \pi; U_T^{-1/2} d_T(\pi)) + \frac{1}{2} H_\infty(d_T(\pi), d_T(\pi)) \\ &+ \frac{1}{2}\left[d^2 Q_T\ (\gamma_0, \pi; U_T^{-1/2} d_T(\pi), U_T^{-1/2} d_T(\pi)) - H_\infty(d_T(\pi), d_T(\pi))\right] + O_P(\epsilon^3) \\ &= \frac{1}{2} H_\infty(d_T(\pi), d_T(\pi)) + O_P(\epsilon^3), \end{aligned}$$

where the second equality follows by (ii). As $H_\infty(d_T(\pi), d_T(\pi)) > 0$ a.s., ϵ can be chosen sufficiently small such that $Q_T\ (\gamma, \pi)$ is convex with probability tending to one in the neighbourhood $\mathcal{N}_T(\gamma_0)$. In particular, there exists a unique minimizer $\hat{\gamma}(\pi) = \gamma_0 + U_T^{-1/2} \hat{d}_T(\pi)$ which solves the first-order condition, $dQ_T\ (\hat{\gamma}, \pi, d\gamma) = 0$ for all $d\gamma$. Since we can choose ϵ arbitrarily small, $\sup_\pi \left\|\hat{d}_T(\pi)\right\| = o_P(1)$, and hence $\sup_\pi \left\| U_T^{1/2}(\hat{\gamma}(\pi) - \gamma_0)\right\| = o_P(1)$ as desired. ■

Lemma D.2 *Assume that assumptions (i)-(iii) of Lemma D.1 hold together with:*

(iv) *There exists a sequence of numbers $\nu_T \in \mathbb{R}_+$ such that $\nu_T^{-1} \to 0$ and uniformly over $\pi \in \Pi$:*

$$\left(dQ_T\left(\gamma_0, \pi; \nu_T^{1/2} U_T^{-1/2} d\gamma\right),\ d^2 Q_T\left(\gamma_0, \pi; U_T^{-1/2} d\gamma, U_T^{-1/2} d\bar{\gamma}\right)\right)$$
$$\xrightarrow{D} \left(S_\infty(\pi, d\gamma), H_\infty(\pi, d\gamma, d\bar{\gamma})\right).$$

Then $\nu_T^{1/2} U_T^{1/2}(\hat{\gamma}(\pi) - \gamma_0) \xrightarrow{D} -\mathbb{H}^{-1}(\pi)\,\mathbb{S}(\pi)$, where $\mathbb{H}(\pi) \in \mathbb{R}^{d\times d}$ and $\mathbb{S}(\pi) \in \mathbb{R}^d$ are stochastic process given through the following identities:

$$S_\infty(\pi, d\gamma) = \mathbb{S}(\pi)' d\gamma, \quad d\gamma' \mathbb{H}(\pi)\, d\bar{\gamma} = H_\infty(\pi, d\gamma, d\bar{\gamma}). \tag{D.1}$$

Proof of Lemma D.2. By Lemma D.1, we know that $\hat{\gamma}_T$ is consistent and solves the first order condition. A first order Taylor expansion of the score and using (iii) together with the same arguments as in the proof of Lemma D.1 yields

$$\begin{aligned} 0 &= dQ_T\left(\gamma_0; \nu_T^{1/2} U_T^{-1/2} d\gamma\right) + d^2 Q_T\left(\bar{\gamma}(\pi), \pi; U_T^{-1/2} d\gamma, U_T^{-1/2} \nu_T^{1/2} U_T^{1/2}(\hat{\gamma}(\pi) - \gamma_0)\right) \\ &= dQ_T\left(\gamma_0; \nu_T^{1/2} U_T^{-1/2} d\gamma\right) + d^2 Q_T\left(\gamma_0, \pi; U_T^{-1/2} d\gamma, U_T^{-1/2} \nu_T^{1/2} U_T^{1/2}(\hat{\gamma}(\pi) - \gamma_0)\right) + o_P(1) \end{aligned}$$

such that, by (iv),

$$-S_\infty(\pi, d\gamma) = H_\infty\left(\pi, d\gamma, \nu_T^{1/2} U_T^{1/2}(\hat{\gamma}(\pi) - \gamma_0)\right) + o_P(1).$$

This completes the proof. ■

E Model Specifications in Simulation Study

DGP under $H_R^{(1)}: \delta_0 = 0$: $\beta_0 = (1, -0.8724)'$, $\alpha_0 = (-0.0211, 0.0015)'$, $\delta_0 = (0, 0)'$ and

$$\Phi_0 = \begin{bmatrix} 0.2097 & -0.0907 \\ 0.4468 & 0.4295 \end{bmatrix}, \quad \Omega_0 = \begin{bmatrix} 0.0916 & 0.0242 \\ 0.0242 & 0.0415 \end{bmatrix}.$$

DGP under $H_R^{(2)}: \beta_0 = (1, -1)'$: $\beta_0 = (1, -1)'$, $\alpha_0 = (14.3870, -0.2793)'$, $\delta_0 = (-7.4947, 0.2975)'$, $\omega_0 = 0.1079$, $A_0 = 0.0041$, and

$$\Phi_0 = \begin{bmatrix} 0.2395 & -0.0899 \\ 0.4201 & 0.4034 \end{bmatrix}, \quad \Omega_0 = \begin{bmatrix} 0.0861 & 0.0251 \\ 0.0251 & 0.0417 \end{bmatrix}.$$

DGP under $H_A^{(1)}: \delta_0 \neq 0$ and $H_A^{(2)}: \beta_0 \neq (1, -1)$: $\beta_0 = (1, -0.9282)'$, $\alpha_0 = (14.7819, -0.2765)'$, $\delta_0 = (-7.3486, 0.1382)'$, $\omega_0 = 0.1009$, $A_0 = 0.0037$, and

$$\Phi_0 = \begin{bmatrix} 0.2339 & -0.0970 \\ 0.4193 & 0.4338 \end{bmatrix}, \quad \Omega_0 = \begin{bmatrix} 0.0874 & 0.0247 \\ 0.0247 & 0.0415 \end{bmatrix}.$$