Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models*

Dennis Kristensen† — UCL and Columbia University

Anders Rahbek‡ — University of Copenhagen

4th November, 2011

Abstract

Estimators and tests are developed and analyzed for a general class of vector error correction models which allow for asymmetric and non-linear error correction. For a given number of cointegration relationships, 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 we believe useful in general for analysis of non-stationary non-linear time series models. We provide 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 and improve results in Kristensen and Rahbek (2010, Journal of Econometrics) where estimation, but not testing, of symmetric non-linear error correction was 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 Juan Carlos Escanciano for very helpful discussions and suggestions. We also 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: d.kristensen@ucl.ac.uk

‡E-mail: anders.rahbek@econ.ku.dk.

1 Introduction

We develop estimators and test statistics for a class of nonlinear vector error correction models with r unknown cointegration vectors, β, with r known. Estimators and test statistics are based on the Gaussian (quasi-)likelihood, and we propose both Lagrange Multiplier (LM), Wald 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 explicit conditions involving the nonlinear transfer function and existence of relevant moments. As part of the theoretical analysis, new functional central limit theorems along with (uniform) weak convergence of stochastic integrals are developed which should be of independent interest in the analysis of nonlinear, non-stationary models.

Allowing for unknown cointegration relations complicates 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 of 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.

We would like to stress that while our framework allows for testing a broad range of different hypotheses, we do not address the issue of testing for the number of cointegration relationships, r. To be more precise, we require throughout that β is identified under null. This rules out testing for the number of cointegrating vectors. In order to develop cointegration rank tests, we need to specify how the error correction mechanism changes as the number of cointegration relation changes. While this is obvious in a linear setting, this is non-trivial in our general, nonlinear framework. For specific parameterizations of the nonlinear transfer function, it should be possible to analyze cointegration rank tests using the techniques developed in this paper. We leave this for future research.

The paper offers a number of novel contributions relative to the existing literature: First, the asymptotic theory for the QMLE's extend the ones of Kristensen and Rahbek (2010) who restrict themselves to a smaller class of nonlinear error correction models that does not include asymmetric adjustments. Our class of models contains their model as a special case, and in addition include, but is not restricted to, asymmetric smooth transition (see e.g. Saikkonen, 2008), as well as the (possibly asymmetric) polynomial models (see e.g. Baghli, 2005 and Escribano, 2004). This is an important extension since asymmetric adjustments have been found in many empirical studies; see e.g. Hansen and Seo (2002) and Kilic (2011). Our results for the QMLE's also complement the ones of Seo (2011) who consider estimation of threshold error correction models using kernel smoothers to handle discontinuities implied by the thresholds.

Second, to the authors' knowledge, this is the first paper to develop a rigorous framework for testing in smooth, multivariate models with non-stationary regressors. There is a large literature on sup-testing in a stationary setting: 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), Kilic (2011) and Seo (2006). Since β is assumed known, all regressors can effectively be treated as stationary; as consequence, their models and results are in line with Hansen (1996).

Taking into account the estimation of β proves to be a non-trivial extension since we have to deal with non-linearities and non-stationary regressors simultaneously. The most related study that also deals with these two features is Caner and Hansen (2001) who test for linearity in univariate threshold autoregressions with unit roots using a sup-test. 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. Hansen and Seo (2002) and Nedeljkovic (2009) also develop sup-tests of linearity in threshold and smooth transition cointegration models respectively. However, they (implicitly) assume that the estimation uncertainty of β has no impact on the asymptotic behaviour of their test statistic, and so effectively are back in the aforementioned framework of Hansen (1996).

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 one will in general expect loss of power against the nonlinear alternative of interest, since a misspecified model is being employed in the testing; see, for example Franq et al (2010), for Monte Carlo evidence in the stationary case.

To establish our theoretical results, it proves necessary to develop a new functional central limit theorems (FCLT's) uniformly over the unidentified parameters, as well as uniform weak convergence to stochastic integrals. 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, 2011). 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 let $C[0,1]$ and $D[0,1]$ denote the space of continuous and *cadlag* functions respectively, and $\mathcal{L}_\infty(\mathcal{A})$ the space of uniformly bounded functions on a given domain $\mathcal{A}$; see van der Vaart and Wellner (1996, Ch. 1.5). We use $\stackrel{P}{\to}$ and $\stackrel{D}{\to}$ to denote convergence in probability and distribution respectively on $\mathbb{R}^k$; We use $\stackrel{W}{\to}$ to denote weak convergence on function spaces, where these will be specified for each case. All convergences take place as $T \to \infty$. Furthermore, $df(x; dx)$ 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 and parameters

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 is a martingale difference sequence. 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:

$$g\left(\beta' X_{t-1}\right) = \alpha\beta' X_{t-1} + \delta\psi\left(\beta' X_{t-1};\xi\right). \tag{2.2}$$

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. (2.2) 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,

$$\delta\psi(z,\xi) = \sum_{s=1}^{S} \delta_s \psi_s(z,\xi) \text{ with } \delta := (\delta_1, ..., \delta_S), \ \psi(z,\xi) := (\psi_1(z,\xi), ..., \psi_S(z,\xi))'. \tag{2.3}$$

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

$$\psi_s(z,\xi) := \left[1 + \exp\left\{(z-\omega_s)' A_s (z-\omega_s)\right\}\right]^{-1} z, \tag{2.4}$$

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 $\psi(z,\xi) = o(1)$ 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. (2.2) 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

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

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 (2011). However, we shall 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 (2011).

Our model does not include deterministic trends and/or exogenous (stationary) regressors. We believe that our analysis could be extended to handle these more general cases, but to avoid overly lengthy assumptions and proofs we leave such extensions for future research.

Regarding identification, then as common in the cointegration literature the $(p \times r)$-dimensional parameter β is only identified up to a normalization. A number of different normalizations of β exist in the literature; most of these can all be expressed in terms of a $(p \times (p-r))$ dimensional matrix κ_0, such that

$$\beta - \beta_0 = \kappa_0 b, \tag{2.6}$$

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. We will here remain flexible regarding the specific choice of κ_0 and merely assume that κ_0 has been chosen such that there exists a matrix sequence K_T with $K_T^{-1}\kappa_0' X_{[Ts]}$ converging weakly, c.f. Assumption 4.5. In practice, the choice of κ_0 will normally be guided by two important issues: First, in many cases it will be useful to choose κ_0 such that $K_T^{-1}\kappa_0' X_{[Ts]}$ has a convenient asymptotic limit. For example, for the symmetric error correction model in Kristensen and Rahbek (2010), κ_0 is chosen such that the limiting distribution is split into a stochastic and deterministic component. We allow for this normalization, but do not restrict our attention to this particular choice since in other (asymmetric) models such a decomposition may not be available. Second, depending on the hypotheses of interest, κ_0 should be chosen accordingly. Consider, for example, the case $p = 2$ and $r = 1$ such that $\beta = [\beta_1, \beta_2]' \in \mathbb{R}^2$; if one is interested in testing hypotheses involving only β_2, a convenient choice is $\kappa_0 = [0, 1]$ such that β_1 is fixed while β_2 is a free parameter; see Luukkonen, Ripatti and Saikkonen (1996) for a further discussion of normalization in cointegrating VAR systems.

Given the chosen normalization, we can rewrite the model in eq. (2.1) as a nonlinear regression model:

$$\Delta X_t = g\left(Z_{0,t-1} + b' Z_{1,t-1}\right) + \Phi Z_{2,t-1} + \varepsilon_t, \tag{2.7}$$

where $Z_{0,t}$, $Z_{1,t}$ and $Z_{2,t}$ are defined as

$$Z_{0,t} := \beta_0' X_t \in \mathbb{R}^r, \quad Z_{1,t} := \kappa_0' X_t \in \mathbb{R}^{p-r}, \quad Z_{2,t} := \left(\Delta X_t', ..., \Delta X_{t-k+1}'\right)' \in \mathbb{R}^{pk}. \tag{2.8}$$

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,

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

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 shall allow that the short-run parameter ξ is not identified under the null of interest, leading to non-standard testing problems. On the other hand, as mentioned, we require throughout that β is identified under the null, which again rules out testing for the number r of cointegrating vectors.

We consider in turn hypotheses involving either short- or long-run parameters. Note that we do not consider testing for joint hypotheses on both short and long-run parameters. Joint testing is essentially straightforward in terms of writing up the test-statistics. However, there are issues regarding identification, see e.g. Johansen (2010) for the linear VAR model, which we wish to address elsewhere.

2.3.1 Testing Short-Run Parameters

We wish to consider general hypotheses involving the short-run parameters $\eta = (\alpha, \delta, \Phi)$ and ξ (cf. eq. (2.9)). To do so, consider restrictions on the form,

$$H_0 : R' vec(\eta,\xi) = 0, \quad (2.12)$$

where R is a known $(m \times d)$-matrix of full rank 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. Note that we require β to be identified under H_0.

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, that is, $\delta = 0$, or $R' vec(\eta,\xi) = vec(\delta) = 0$. 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. (2.2) 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, test for zero rows in α and δ, and, more generally, test for linear constraints involving these: $R'vec\left(\eta,\xi\right)=R'_{\alpha,\delta}vec\left(\alpha,\delta\right)=0$ for some matrix $R_{\alpha,\delta}$. Note that for β to be identified under H_0 this excludes $R_{a,\delta}$ to be a full rank square matrix of dimension $p\left(r+r_{\delta}\right)$.

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 of ξ being either identified or unidentified under the null separately:

2.3.2 The parameter ξ identified

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 of the classic LR, LM or Wald-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 as defined in Section 2.4. Finally, the Wald statistic takes the form

$$W_{T}=vec(\hat{\eta},\hat{\xi})'R\left[R'\mathbb{H}_{T,\eta,\xi}(\hat{\theta},\hat{\xi})R\right]^{-1}R'vec(\hat{\eta},\hat{\xi}),\tag{2.15}$$

where $\mathbb{H}_{T,\eta,\xi}(\theta,\xi)$ is the Hessian for (η,ξ). Note that other versions of Lagrange and Wald statistics could be used, see e.g. Newey and McFadden (1994, p. 2222). These will all be asymptotically first-order equivalent under the null, and so we only analyze the versions given above.

2.3.3 The parameter ξ unidentified

Next, consider the case where ξ is unidentified under the null of H_0.[1] As $\eta = (\alpha, \delta, \Phi)$, and lack of identification of ξ can only be achieved from restrictions on δ, the general null in eq. (2.12) can in this case be written as

$$H_0 : R_\eta' vec\,(\eta) = 0,$$

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

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

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

$$\hat{\theta}\left(\xi\right) = \arg \max_{\theta\in\Theta} L_T^* \left(\theta, \xi\right).$$

The sup-LR, sup-LM and sup-Wald test statistics are then obtained by taking supremum of the corresponding standard test statistic over ξ:

$$\sup LR_T := \sup_{\xi\in\Xi} LR_T\left(\xi\right), \quad LR_T\left(\xi\right) = 2\left[L_T^*(\hat{\theta}\left(\xi\right), \xi) - L_T^*(\tilde{\theta}, \xi)\right], \tag{2.16}$$

$$\sup LM_T := \sup_{\xi\in\Xi} LM_T\left(\xi\right), \quad LM_T\left(\xi\right) = \mathbb{S}_T(\tilde{\theta}\left(\xi\right), \xi)'\mathbb{H}_T^{-1}(\tilde{\theta}\left(\xi\right), \xi)\mathbb{S}_T(\tilde{\theta}\left(\xi\right), \xi), \tag{2.17}$$

$$\sup W_T = \sup_{\xi\in\Xi} W_T\left(\xi\right), \quad W_T\left(\xi\right) = vec(\hat{\eta}\left(\xi\right))' R_\eta \left[R_\eta' \mathbb{H}_{T,\eta}(\hat{\theta}\left(\xi\right), \xi) R_\eta\right]^{-1} R_\eta' vec(\hat{\eta}\left(\xi\right)), \tag{2.18}$$

where $\mathbb{H}_{T,\eta}(\theta,\xi)$ is the Hessian w.r.t. η.

2.3.4 Testing Long-Run Parameters

Next, consider hypotheses relating to the long-run parameter β. Recall that β is normalized by eq. (2.6), so we may consider the following hypothesis involving the long-run parameter b,

$$H_{0,b} : R_b' vec\left(b'\right) = 0, \tag{2.19}$$

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

[1] Of course, one may also have that a part of ξ is identified, ξ_1 say, while ξ_2 with $\xi = (\xi_1, \xi_2)$ is not. In that case, redefine η to include also ξ_1 and set $\xi := \xi_2$ in 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 $R_b' vec(b') = b = 0$.

For simplicity consider the case where ξ is identfied. In this case the test statistics are computed in the same way as in section 2.3.3. We first compute the restricted estimators which for ease of notation we still denote $\tilde{\theta}$ 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:

$$LR_{b,T} = 2 \left[L_T^*(\hat{\theta}, \hat{\xi}) - L_T^*(\tilde{\theta}, \tilde{\xi}) \right], \quad LM_{b,T} = \mathbb{S}_T(\tilde{\theta}, \tilde{\xi})' \mathbb{H}_T^{-1}(\tilde{\theta}, \tilde{\xi}) \mathbb{S}_T(\tilde{\theta}, \tilde{\xi}), \quad \text{and} \tag{2.20}$$
$$W_{b,T} = vec(\hat{b}')' R_b \left[R_b' \mathbb{H}_{T,b}(\hat{\theta}, \hat{\xi}) R_b \right]^{-1} R_b' vec(\hat{b}').$$

2.4 Score and Hessian

As is standard, the analysis of likelihood-based estimators and test statistics 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:

$$dL_T^* (\theta, \xi; d\theta, d\xi) = \mathbb{S}_T(\theta, \xi)' vec (d\theta, d\xi) . \tag{2.21}$$

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:

$$d^2 L_T^* (\theta, \xi; d\theta, d\xi; d\theta^*, d\xi^*) = vec (d\theta^*, d\xi^*)' \, \mathbb{H}_T(\theta, \xi) vec (d\theta, d\xi) . \tag{2.22}$$

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,

$$d\psi (z, \xi; dz) = \partial_z \psi (z, \xi) \, dz, \quad \partial_z \psi (z, \xi) = (\partial \psi_i / \partial z_j)_{i,j} \in \mathbb{R}^{r_\delta \times r}, \tag{2.23}$$
$$d\psi (z, \xi; d\xi) = \partial_\xi \psi (z, \xi) \, vec (d\xi) , \quad \partial_\xi \psi (z, \xi) \in \mathbb{R}^{r_\delta \times d_\xi}.$$

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

$$u_t (\xi) := \left(u_{\alpha,t} (\xi)' , u_{\phi,t} (\xi)' , u_{\delta,t} (\xi)' \right)' , \quad v_t (\xi) := [\delta_0 \partial_\xi \psi (Z_{0,t-1}; \xi)]' \, \Omega_0^{-1} \varepsilon_t (\theta_0, \xi) , \quad \text{and} \tag{2.24}$$

$$w_t (\xi) := [\alpha_0 + \delta_0 \partial_z \psi (Z_{0,t-1}; \xi)]' \, \Omega_0^{-1} \varepsilon_t (\theta_0, \xi) ,$$

with

$$u_{\alpha,t}(\xi) := vec\left(\Omega_0^{-1}\varepsilon_t(\theta_0,\xi)\,Z'_{0,t-1}\right), \quad u_{\phi,t}(\xi) := vec\left(\Omega_0^{-1}\varepsilon_t(\theta_0,\xi)\,Z'_{2,t-1}\right) \tag{2.25}$$
$$u_{\delta,t}(\xi) := vec\left(\Omega_0^{-1}\varepsilon_t(\theta_0,\xi)\,\psi(Z_{0,t-1};\xi)'\right).$$

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 $d^2L_T^*\ (\theta_0,\xi;d\theta,d\xi;d\theta^*,d\xi^*)$, 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 explicit expressions.

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 Uniform FCLT and Convergence of Stochastic Integrals

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 results for double indexed random sequences, also referred to as partial sum processes in van der Vaart and Wellner (1996, ch.2.12). 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.

Let $x_{T,t} \in \mathbb{R}^{d_x}$, $t = 1, ..., T$, denote an appropriately normalized triangulary array, which is assumed to converge weakly, see Assumption 3.3 below. Furthermore, let $y_t \in \mathbb{R}^{d_y}$ be a stationary sequence and $e_t \in \mathbb{R}^{d_e}$ a Martingale difference sequence with respect to the natural filtration, $\mathcal{F}_{T,t} = \mathcal{F}(e_t, x_{T,t}, y_t, e_{t-1}, x_{T,t-1}, y_{t-1}...)$. In terms of these, define the following two processes,

$$\phi_T(s,\pi) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} f(y_{t-1},\pi)\,e_t \in \mathbb{R}^{d_x}, \text{ and } \ x_T(s) := x_{T,[Ts]} \in \mathbb{R}^{d_x}, \tag{3.1}$$

where $f : \mathbb{R}^{d_y} \times \Pi \mapsto \mathbb{R}^{d_x \times d_e}$ is a given function, $\pi \in \Pi$ for some compact set $\Pi \subseteq \mathbb{R}^{d_\pi}$ and $s \in [0, 1]$.

We then establish weak convergence results for these processes indexed by $(s, \pi) \in [0, 1] \times \Pi$, and associated stochastic integrals. To do so we impose the following regularity conditions:

Assumption 3.1 *The sequence* (e_t, y_t) *satisfies:*

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

(ii) e_t *is a martingale difference w.r.t.* $\mathcal{F}_{T,t}$ *such that* $E[e_t|\mathcal{F}_{T,t-1}] = 0$ *and* $E[e_t e_t'|\mathcal{F}_{T,t-1}] = \Omega_e$ *for some constant, finite matrix* $\Omega_e \in \mathbb{R}^{d_e \times d_e}$.

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

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

(ii) $\|f(y_{t-1}, \pi) - f(y_{t-1}, \pi')\| \leq B(y_{t-1}) \|\pi - \pi'\|$, *for all* $\pi, \pi' \in \Pi$ *and with* $E[B(y_{t-1})^n] < \infty$.

Assumption 3.3 *The process* $x_T(\cdot) := x_{T,[T\cdot]} \in D[0,1]$ *satisfies:*

(i) *As* $T \to \infty$, $x_T(\cdot) \xrightarrow{W} x(\cdot)$ *on* $D[0,1]$, *where* $x(\cdot)$ *is continuous.*

(ii) *For some integer* $q > 0$, $\sup_{1\leq t\leq T, T\geq 1} E[\|x_{T,t}\|^q] < \infty$.

In terms of the nonlinear error correction model in eq. (2.1), we will choose (in the case of no lagged differences, or $k = 0$), $\pi = \xi$, $y_t = Z_{0,t}$, $f(y_{t-1}; \pi) = \psi(Z_{0,t}; \xi)$, and $x_{T,t} = K_T^{-1} Z_{1,t}$ for some appropriately chosen weighting matrix K_T (see Assumption 4.5) and with $Z_{i,t}$ defined in (2.8). Inparticular for the STECM examples in eq. (2.4) and (2.5), Assumption 3.2 (i) and (ii) hold if $E[\|Z_{0,t}\|^{\max(m,n)}] < \infty$. 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), see next section for details.

Remark 1: In Assumptions 3.1 (ii) one may instead assume that $E[e_t e_t'|\mathcal{F}_{T,t-1}] = \Omega_{e,t}$, with $\Omega_{e,t}$ stationary and $E[\|\Omega_{e,t}\|] < \infty$, thereby allowing for conditional heteroskedasticity. The moment conditions will however in that case be very complicated and we therefore leave this out here. Moreover, the covariance matrix $\Sigma(s_1, \pi_1, s_2, \pi_2)$ defined in the following theorem would have to be changed accordingly.

Remark 2: A general sufficient condition for Assumption 3.2 (ii) to hold 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* $n, m \geq 2$, *the partial sum process* $\phi_T(\cdot, \cdot) \in \mathcal{L}_\infty([0,1] \times \Pi)$ *defined in (3.1), satisfies,*

$$\phi_T(\cdot, \cdot) \xrightarrow{W} \phi(\cdot, \cdot) \quad on\ \mathcal{L}_\infty([0,1] \times \Pi), \tag{3.2}$$

where $\phi(s, \pi)$ *is multi-parameter Gaussian process with covariance kernel,*

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

The theorem is obtained by extending the arguments of Escanciano (2007, Theorem 1) who provide a FCLT result for the stochastic process $\pi \mapsto \phi_T(1,\pi)$. A direct consequence of Theorem 3.4 is the convergence of product moment matrices:

Theorem 3.5 *Under Assumptions 3.1-3.2, with $m, n \geq 2$ and under Assumption 3.3 (i),*

$$\frac{1}{T}\sum_{t=1}^{T} x'_{T,t-1} f(y_{t-1};\cdot) \overset{W}{\to} \int_0^1 x(s)'\, ds E[f(y_{t-1};\cdot)] \quad \text{on } \mathcal{L}_\infty(\Pi).$$

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 3.6 (Convergence to Stochastic Integral) *Assume furthermore that for any fixed $\pi \in \Pi$, $(x_T(\cdot), \phi_T(\cdot,\pi)) \overset{W}{\to} (x(\cdot), \phi(\cdot,\pi))$ on $D[0,1]$.*

(i) Under Assumptions 3.1-3.2 with $m, n \geq 2$ and Assumption 3.3 (i), for any given $\pi \in \Pi$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x'_{T,t-1} f(y_{t-1},\pi) e_t \overset{D}{\to} \int_0^1 x(s)'\, d\phi(s,\pi). \tag{3.3}$$

(ii) Under Assumptions 3.1-3.3 with $n \geq 4$, $m > 3d_\pi$ and $q > \max(3d_\pi, 4)$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x'_{T,t-1} f(y_{t-1},\cdot) e_t \overset{W}{\to} \int_0^1 x(s)'\, d\phi(s,\cdot) \quad \text{on } \mathcal{L}_\infty(\Pi). \tag{3.4}$$

Remark 3: The moment conditions on y_t and e_t in Theorem 3.6(ii) are stronger compared to the other results in this section. In particular, the moment restriction on $f(y_{t-1};\pi)$ on the form $m > 3d_\pi$ and $q > \max(3d_\pi, 4)$ will, unless f is bounded, impose quite strong restrictions on the moments of y_{t-1}. The required number of moments increases linearly in d_π. This "*curse of dimensionality*" stems from the way we establish stochastic equicontinuity or tightness of the stochastic integral, see proof of Theorem 3.6 in the Appendix. We conjecture that the high order moment conditions, while sufficient, are not necessary, and can be avoided by a different proof strategy when establishing weak convergence to stochastic integrals indexed by $\pi \in \Pi$. By comparison, we obtained the weak convergence to the double indexed Gaussian process in Theorem 3.4 with very modest moment restrictions. Similarly Theorem 3.5 is obtained under weak moment restrictions since, contrary to convergence to stochastic integrals, this follows (essentially) by application of the continuous mapping theorem.

Remark 4: 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 pointwise convergence in Theorem 3.6 by verifying the classic conditions of Theorem 2.2 of Kurtz and Protter (1991), or equivalently Theorem 2.1 of Hansen (1992), which do 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.

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, as well as on the parameter space:

Assumption 4.1 *The parameter space* Ξ *for* ξ *is compact and* $\vartheta_0 = (\theta_0, \xi_0)$ *lies in the interior of* $\Theta \times \Xi$.

Assumption 4.2 *The function* $\psi(z, \xi)$ *is three 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 4.3 *The error term* ε_t *is a martingale difference with respect to* $\mathcal{F}_{t-1} = \mathcal{F}(X_{t-1}, X_{t-2}, ...)$*. Furtermore,* $\Omega \equiv E[\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}]$ *and* $E\|\varepsilon_t\|^{q_\varepsilon} < \infty$ *for some* $q_\varepsilon \geq 2$.

Assumption 4.4 *The process* $(Z_{0t}', Z_{2t}')'$ *can be embedded in a stationary and geometrically ergodic Markov chain. Moreover,* $E[\|Z_{0,t-1}\|^{q_0}] < \infty$ *and* $E[\|Z_{2,t-1}\|^{q_2}] < \infty$ *for some* $q_0, q_2 \geq 1$.

Assumption 4.5 *With* κ_0 *defined in (2.6) and for some sequence of diagonal matrices* $K_T \in \mathbb{R}^{(p-r)\times(p-r)}$ *satisfying* $K_T^{-1} \to 0$ *as* $T \to \infty$*, the non-stationary process* $Z_{1,t} = \kappa_0' X_t$ *satisfies: (i)* $K_T^{-1} Z_{1,[T\cdot]} \overset{W}{\to} F(\cdot)$ *on* $D([0,1])$*, for some continuous, stochastic process* $F(s)$ *satisfying* $\int_0^1 F(s)F(s)'ds > 0$ *almost surely; (ii)* $\sup_{T\geq 1}\sup_{t\leq T} E\left\|K_T^{-1}Z_{1,t}\right\|^{q_1} < \infty$ *for some* $q_1 > 0$.

Assumption 4.2 imposes smoothness restrictions and polynomial bounds on $\psi(z, \xi)$. The smoothness restrictions ensures that the first three derivatives of the likelihood w.r.t. the parameters are well-defined, while the polynomial bounds are used in conjunction with Assumption 4.4 to ensure that appropriate moments of these derivatives are well-defined. All proposed specifications of nonlinear error correction found in the literature satisfy this assumption except for threshold models. Thus, we rule out threshold models for which a different proof strategy needs to be used; see e.g. Seo (2011). It is worth pointing out though that threshold models can be approximated up to any degree of precision by a smooth transition model in the sense that as the scale parameter in the smooth transition model converges to zero, the smooth transition model converges towards a threshold model. For example, $[1 + \exp\{(z - \omega_s)/a\}]^{-1} \to \mathbb{I}\{z \geq \omega_s\}$ as $a \to 0^+$. This fundamental feature is the basic building block of the analysis in Linton and Seo (2010) of their smoothed estimators of threshold parameters.

Regarding Assumption 4.4, for a precise definition of geometric ergodicity of a Markov chain, we refer to Meyn and Tweedie (1993). Sufficient conditions for Assumptions 4.4 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. (2.4) and (2.5). Note in this respect that Assumption 4.4 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 $V(y) = 1 + \|y\|^{2q}$, $q > 2$, but not necessarily stationary. 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) and Kristensen and Rahbek (2009). Assumption 4.4 is used to establish stationarity and ergodicity as required by Assumption 3.2 for y_t, with $y_t = \left(Z_{0,t}', Z_{2,t}'\right)'$. Thus the alternative assumption of stationarity and ergodicity could be used instead with no changes in the subsequent results. However, we do not know of any results for stationarity and ergodicity of nonlinear error correction models which are not derived as implied by geometric ergodicity.

In Assumption 4.5**,** 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). Thus, by not specifying $F(\cdot)$, we accomodate for a large class of models, such as the ones included in e.g. Saikkonen (2008).The restriction that $\int_0^1 F(s) F(s)' ds > 0$ almost surely is used to ensure that the information matrix associated with F is non-singular almost surely; see, for example, Theorem 4.7.

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 (2.24)-(2.25), 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 FCLT on $\mathcal{L}_\infty([0,1] \times \Xi)$ where F is defined in Assumption 4.5:

Lemma 4.6 *Suppose that Assumptions 4.1-4.4 hold with* $q_2, q_\varepsilon = 2$ *and* $q_0 = 2\rho,$*, and Assumption 4.5 (i) hold. Then, with* $u_t(\xi)$*,* $v_t(\xi)$ *and* $w_t(\xi)$ *defined in eq. (2.24) and* $Z_{1,t}$

in Assumption 4.5,

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}u_t(\xi)', \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}v_t(\xi)', \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}w_t(\xi)', \left(K_T^{-1}Z_{1,[Ts]}\right)'\right) \tag{4.1}$$
$$\xrightarrow{W}\left(B_u(s,\xi)', B_v(s,\xi)', B_w(s,\xi)', F(s)'\right)$$

on the function space $\mathcal{L}_\infty([0,1]\times\Xi)$*. Here,* 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 FCLT result is used for the asymptotics of the sup statistics where we treat the statistics $LR_T(\xi)$, $LM_T(\xi)$ and $W_T(\xi)$ defined in eqs. (2.16)-(2.18) as sequences of stochastic processes. Note that when the parameter ξ is identified, we only need the above convergence to hold pointwise at $\xi=\xi_0$.

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

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

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: We consider in turn the situations where $\delta_0\neq 0$, and $\delta_0=0$ corresponding to the case where ξ is identified and vanishes under the nul respectively.

Theorem 4.7 *Suppose that Assumptions 4.1-4.4 hold with* $q_2, q_\varepsilon = 2$ *and* $q_0 = 2\rho$*, and Assumption 4.5 (i) hold. Assume furthermore that* $\delta_0\neq 0$*, and* $\Sigma(\xi_0,\xi_0)>0$ *where* $\Sigma(\xi_0,\xi_0)$ *is defined in (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 in the next. 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 4.7 complement the ones of Seo (2011) 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 (2011).

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 4.8 *For $\delta_0 = 0$, suppose that Assumptions 4.1-4.5 hold with $q_0 = \rho \max(4, 3d_\xi)$, $q_1, q_\varepsilon = \max(4, 3d_\xi)$ and $q_2 = 2$. Assume furthermore 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),*

$$T^{1/2} V_{\theta,T}^{1/2} vec\left(\hat{\theta}(\xi) - \theta_0\right) \xrightarrow{W} \mathbb{H}_{\theta\theta}^{-1}(\xi)\, \mathbb{S}_\theta(\xi) \quad on\ \mathcal{L}_\infty(\Xi), \tag{4.7}$$

for a random matrix process $\mathbb{H}_{\theta\theta}(\xi)$ and random vector process $\mathbb{S}_\theta(\xi)$, 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 4.4 and 4.5 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. (2.24). 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-Likelihood Ratio (LR), sup-Lagrange Multiplier (LM) test and sup-Wald (W) tests introduced in eqs. (2.16)-(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 (see Proof of Theorem 4.9 below) 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 $M_\eta = \text{diag}\left(I_{(p-r)r}, (R_\eta)_\perp\right)$, while $\mathbb{S}_\theta(\xi)$ and $\mathbb{H}_{\theta\theta}(\xi)$ are defined in Theorem 4.8. 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

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

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 $M = \text{diag}\left(I_{(p-r)r}, R_\perp\right)$, while $\mathbb{S}$ and $\mathbb{H}$ defined in Theorem 4.7. The following result is then shown in the Appendix:

Theorem 4.9 *Suppose Assumptions 4.1-4.5 hold with q_i specificed below. Assume $H_0 : R'vec(\eta,\xi) = 0$ hold with R having full rank and β identified:*

1. *If ξ_0 is identified under H_0, then with $q_2, q_\varepsilon = 2$ and $q_0 = 2\rho$,*

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

where, with $\mathbb{S}$ and $\mathbb{H}$ given in Theorem 4.7,

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

2. *If ξ is not identified under H_0, then with $q_0 = \rho \max(4, 3d_\xi)$, $q_1, q_\varepsilon = \max(4, 3d_\xi)$ and $q_2 = 2$,*

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

where, with $\mathbb{S}_\theta(\xi)$ and $\mathbb{H}_{\theta\theta}(\xi)$ given in Theorem 4.8,

$$\mathbb{V}_\theta(\xi) := \left[(M_\eta)_\perp \mathbb{H}_{\theta\theta}(\xi)(M_\eta)_\perp\right]^{-1} (M_\eta)_\perp \mathbb{H}_{\theta\theta}^{-1}(\xi)\mathbb{S}_\theta(\xi)$$

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 $E[\psi(Z_{0.t};\xi)] \neq 0$ 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 $(Z_{0,t-1}, Z_{2,t-1}, \psi(Z_{0,t-1};\xi))$ 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 (2011) 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 (2009) 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 $E[\psi(Z_{0.t};\xi)] = 0$.

As already mentioned on p. 9, the problem of vanishing parameters under the null may only involve a subset of the parameters in ξ. For example, suppose that the non-linear component takes the form $\delta\psi(z,\xi) = \sum_{s=1}^{S} \delta_s \psi_s(z,\xi_s)$ 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) The null hypothesis of $H_0 : \delta = 0$ corresponds to choosing $R_\eta = [O_m, I_{d_\delta}]$ where O_m is the $m \times m$-matrix of zeros and $m = \dim(vec(\alpha, \Phi))$. Under the null, the model collapses to a standard linear cointegrating error-correction model with implications discussed after Theorem 4.8. 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 4.8 with $\delta_0 = 0$ (or alternatively, Johansen, 1996), we obtain that

$$\sqrt{T} V_{\theta,T}^{1/2} vec(\tilde{\theta} - \theta_0) \xrightarrow{D} M_\eta \tilde{\mathbb{H}}_{\theta\theta}^{-1} \tilde{\mathbb{S}}_\theta, \quad M_\eta = \operatorname{diag}\left(I_{(p-r)r}, (R_\eta)_\perp\right), \tag{4.14}$$

where, with $B_{\alpha,\phi}(s) = \left(B_\alpha(s)', B_\phi(s)'\right)'$, $F(s)$ and $B_w(s)$ being the Brownian motions described immediately after Theorem 4.8,

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

One can easily check that $M_\eta \tilde{\mathbb{H}}_{\theta\theta}^{-1} \tilde{\mathbb{S}}_\theta$ is the standard asymptotic distribution for the Gaussian QMLE in a linear I(1)-model.

Next, we derive tests for the hypothesis $H_{0,b}$ involving the cointegration relations, $H_{0,b} : R_b' vec(b') = 0$ or, equivalently, $H_{0,b} : vec(b') = (R_b)_\perp \tau$ for some free parameter τ. The proof strategy is identical to the one employed in Theorem 4.9 and so we state the result without proof:

Theorem 4.10 *Suppose Assumptions 4.1–4.5 hold with $q_2, q_\varepsilon = 2$ and $q_0 = 2\rho$, and $H_{0,b} : R_b' vec(b') = 0$ hold with R having full rank. Then the LR and LM test of this hypothesis satisfies*

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

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 4.7 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 4.7-4.10. 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, 2011). 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, 2011), 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, if the alternative is true, the residuals obtained under the null will not be appropriately centered and so the bootstrap procedure would potentially diverge. Since the goal of the bootstrap procedure is to obtain an estimate of the distribution under the null (whether it is true or not), the use of residuals from under the null would be problematic; see Paparoditisa and Politis (2005) for more details.

Given the bootstrap sample X_t^*, $t = 1, ..., T$, we then compute the sup-test statistics with the bootstrap sample replacing the original one. 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 4.6 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 $\tilde{C} = \tilde{\beta}_\perp \left(\tilde{\alpha}_\perp' \left(I - \sum_{i=1}^{k} \tilde{\Phi}_i\right) \tilde{\beta}_\perp\right)^{-1} \tilde{\alpha}_\perp'$, $\sup_{1 \leq t \leq T} R_t^* = o_{P^*}(1)$ 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 4.6 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 tests in a specific example of the smooth transition error correction model (STECM). We focus on the (sup) LR tests as we expect that the LM and Wald tests will perform similarly. The particular model used in the simulation study is given as

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

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. (2.4),

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

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-version of the LR test. The second hypothesis examines whether the spread is stable, $H_R^{(2)} : \beta = (1, -1)'$, such that in this case the parameter ξ does *not* vanish under the null. Given the second null, we choose the normalization $\beta = (1, \beta_2)'$ corresponding to $\kappa_0 = (0, 1)$. c.f. discussion on p. 6.

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)[2]. 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 4.4 and 4.5 hold for the parameters chosen under the nulls and alternative employed.

As part of the LR test statistic, we need to compute the QMLE's under null and alternative; the numerical computation of the QMLE's is discussed below. For the bootstrap we use the set-up in eq. (5.1). In terms 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. (2.16): 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,

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

[2]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

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

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 $\Xi = \{(A, \omega) : 0 \leq A \leq 1 \text{ and } -1 \leq \omega \leq 1\}$, 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 $\sup LR_T$. The choice of Ξ is somewhat ad hoc and it would be of interest to investigate the sensitivity of the test to the choice of Ξ; we leave this for future research.

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. (1994) Empirical Process Methods in Econometrics, *Handbook of Econometrics*, Volume IV, Edited by R.F. Engle and D.L. McFadden, Elsevier.

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.

Baghli, M. (2005) Nonlinear Error-Correction Models for the FF/DM Rate. *Studies in Nonlinear Dynamics & Econometrics* 9, Article 7.

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

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. *Econometric Theory* 26, 1719-1760.

Cavaliere, G., A. Rahbek and R. Taylor (2011) Bootstrap Determination of the Cointegration Rank in VAR Models. Forthcoming in *Econometrica*.

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.

Escanciano, J.C. (2007) Model Checks Using Residual Marked Empirical Processes. *Statistica Sinica* 17, 115-138.

Franq, C., L. Horvatha and J.-M. Zakoïan (2010) Sup-Tests For Linearity In A General Nonlinear AR(1) Model. *Econometric Theory* 26, 965-993.

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.

Johansen, S. (2010) Some Identification Problems in the Cointegrated Vector Autoregressive Model. *Journal of Econometrics* 158, 262-273.

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. (2011) Testing for Co-integration and Nonlinear Adjustment in a Smooth Transition Error Correction Model. *Journal of Time Series Analysis* 32, 647–660.

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.

Luukkonen, R., A. Ripatti and P. Saikkonen (1996) Testing for a Valid Normalization of Cointegrating Vectors in Vector Autoregressive Processes. *Journal of Business & Economic Statistics* 17, 195-204.

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

Meyn, S.P. and R.L. Tweedie (1993) *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

Meitz, M. and P. Saikkonen (2011) Parameter Estimation in Nonlinear AR-GARCH models. *Econometric Theory* 27, 1236-1278.

Nedeljkovic, M. (2009) Testing for Smooth Transition Nonlinearity in Adjustments of Cointegrating Systems. Department of Economics, Warwick Economic Research Papers, No: 876.

Newey, W.K. and D. McFadden (1986) Large sample estimation and hypothesis testing. *Handbook of Econometrics* volume 4 (eds. R. F. Engle and D. McFadden), 2111-2245. Elsevier

Paparoditis, E. and D.N. Politis (2005) Bootstrap Hypothesis Testing in Regression Models. *Statistics & Probability Letters* 74, 356-365.

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

Pollard, D. (1984) *Convergence of Stochastic Processes.* New York: Springer-Verlag.

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. (2011) Estimation of Nonlinear Error Correction Models. *Econometric Theory* 27, 201–234.

van der Vaart, W. and J.A. Wellner (1996) *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

A Proofs of Section 3

Proof of Theorem 3.4. By standard CLT for stationary martingales, we have that $(\phi_T(s,\pi_1),...,\phi_T(s,\pi_k))$ converges weakly towards $(\phi(s,\pi_1),...,\phi(s,\pi_k))$ for any given finite subset $\{\pi_1,...,\pi_k\}$; see e.g. Brown (1971). The claimed result will now hold if we can show that $\phi_T(s,\pi)$ is asymptotically tight, c.f. Theorem 1.5.4 in Van der Vaart and Wellner (1996) [VW96]. For notational convenience, define

$$G_T(\pi) := \sum_{t=1}^{T} m_t^T(\pi), \quad m_t^T(\pi) := f(y_{t-1};\pi)\, e_t/\sqrt{T},$$

such that $\phi_T(s,\pi) = G_{[Ts]}(\pi)$. The idea is now to combine the arguments in the proof of Theorem 1 of Escanciano (2007) [E07], who show tightness of $\pi \mapsto G_T(\pi)$ in a Martingale setting similar to ours, and Theorem 2.12.1 in VW96, who show tightness of $\phi_T(s,\pi)$ when (y_t, e_t) is i.i.d.

As a first step, we verify that Assumptions W1-W2 of E07 hold so that we can apply the same arguments as in E07's Proof of Theorem 1. E07's Assumption W1 follows straightforwardly by our Assumptions 3.1-3.2 combined with the uniform Law of Large Numbers of Kristensen and Rahbek (2005, Proposition 1)[3]. To verify E07's Assumption W2, we establish the sufficient conditions stated in the discussion on p. 121 in E07: First, the required Lipschitz conditions follows by our Assumption 3.2(ii). The requirement of a uniformly integrable entropy follows by Andrews (1994, Theorem 2) since our function class $\{f(y_{t-1};\pi)\, e_t : \pi \in \Pi\}$ is in Andrews' Class II.

Next, we now proceed as in the proof of Theorem 1 of E07 (see also the proof of Theorem 2.12.1 in VW96) and choose a nested sequence of partitions $\mathcal{P}_q = \{\Pi_{q,k} : k = 1,...,N_q\}$ of Π for $q = 1, 2, ...$ which satisfies $\sum_{q=1}^{\infty} 2^{-q}\sqrt{\log N_q} < \infty$ and Assumption W2 in E07. Furthermore, for each $\Pi_{q,k}$, choose a fixed element $\pi_{q,k} \in \Pi_{q,k}$ and define

$$\text{pr}_q(\pi) := \pi_{q,k}, \quad \Delta_{q,t}^T(\pi) := \sup_{\pi_1,\pi_2 \in \Pi_{q,k}} \left\| m_t^T(\pi_1) - m_t^T(\pi_2) \right\| \text{ if } \pi \in \Pi_{q,k}.$$

Then, according to Theorem 1.5.6 of VW96, it is sufficient to show that for every $\varepsilon, \eta > 0$, there exists $\delta > 0$ and $q_0 \geq 1$ such that

$$\limsup_{T\to\infty} P^* \left(\sup_{|s_1 - s_2| < \delta} \sup_{\pi \in \Pi} \left\| \phi_T(s_1,\pi) - \phi_T\left(s_2, \text{pr}_{q_0}(\pi)\right) \right\| > \varepsilon \right) \leq \eta.$$

By the triangle inequality,

$$\begin{aligned} &\sup_{|s_1 - s_2| < \delta} \sup_{\pi \in \Pi} \left\| \phi_T(s_1,\pi) - \phi_T\left(s_2, \text{pr}_{q_0}(\pi)\right) \right\| \qquad \text{(A.1)} \\ &\leq \sup_{|s_1 - s_2| < \delta} \sup_{\pi \in \Pi} \left\| \phi_T(s_1,\pi) - \phi_T(s_2,\pi) \right\| + \sup_{s_2 \in [0,1]} \sup_{\pi \in \Pi} \left\| \phi_T(s_2,\pi) - \phi_T\left(s_2, \text{pr}_{q_0}(\pi)\right) \right\|. \end{aligned}$$

[3] Kristensen and Rahbek (2005) assume geometric ergodicity, but by inspection of the proof of their Proposition 1, it is easily seen that all arguments are still valid when replacing the assumption of geometric ergodicity with that of stationarity and ergodicity.

First, consider the second term: For a given q_0, introduce the same numbers and indicator functions as in E07, p. 132 for all $q \geq q_0$,

$$
\begin{aligned}
a_q &:= 2^{-q}/\sqrt{\log\left(N_{q+1}\right)}, \\
C_{q-1,t}^T\left(\pi\right) &:= \mathbb{I}\left\{\Delta_{q_0,t}^T\left(\pi\right) \leq a_{q_0}, ..., \Delta_{q-1,t}^T\left(\pi\right) \leq a_{q-1}\right\}, \\
D_{q-1,t}^T\left(\pi\right) &:= \mathbb{I}\left\{\Delta_{q_0,t}^T\left(\pi\right) \leq a_{q_0}, ..., \Delta_{q-1,t}^T\left(\pi\right) \leq a_{q-1}, \Delta_{q,t}^T\left(\pi\right) > a_q\right\}, \\
D_{q_0,t}^T\left(\pi\right) &:= \mathbb{I}\left\{\Delta_{q_0,t}^T\left(\pi\right) > a_{q_0}\right\}.
\end{aligned}
$$

We may now rewrite $m_t^T\left(\pi\right) - m_t^T\left(\mathrm{pr}_{q_0}\left(\pi\right)\right)$ as (c.f. E07, eq. 11)

$$
\begin{aligned}
m_t^T\left(\pi\right) - m_t^T\left(\mathrm{pr}_{q_0}\left(\pi\right)\right) = &\left\{m_t^T\left(\pi\right) - m_t^T\left(\mathrm{pr}_{q_0}\left(\pi\right)\right)\right\} D_{q_0,t}^T\left(\pi\right) \\
&+ \sum_{q=q_0+1}^{\infty} \left\{m_t^T\left(\pi\right) - m_t^T\left(\mathrm{pr}_q\left(\pi\right)\right)\right\} D_{q,t}^T\left(\pi\right) \\
&+ \sum_{q=q_0+1}^{\infty} \left\{m_t^T\left(\mathrm{pr}_q\left(\pi\right)\right) - m_t^T\left(\mathrm{pr}_{q-1}\left(\pi\right)\right)\right\} C_{q,t}^T\left(\pi\right).
\end{aligned}
$$

By the same arguments as in E07, p. 131-132, it therefore follows that

$$
\sup_{s\in[0,1]} \sup_{\pi\in\Pi} \left\|\phi_T\left(s,\pi\right) - \phi_T\left(s,\mathrm{pr}_{q_0}\left(\pi\right)\right)\right\| \leq I_1 + I_2 + 2II_2 + II_3 + III,
$$

where, with $\sum_{t,q}^{[Ts]} = \sum_{t=1}^{[Ts]}\sum_{q=q_0+1}^{\infty}$ and $\Delta m_{q,t}^T\left(\pi\right) = m_t^T\left(\mathrm{pr}_q\left(\pi\right)\right) - m_t^T\left(\mathrm{pr}_{q-1}\left(\pi\right)\right)$,

$$
\begin{aligned}
I_1 &= \sup_{\substack{s\in[0,1] \\ \pi\in\Pi}} \left\|\sum_{t=1}^{[Ts]} \Delta_{q_0,t}^T\left(\pi\right) D_{q_0,t}^T\left(\pi\right)\right\|, \quad I_2 = \sup_{\substack{s\in[0,1] \\ \pi\in\Pi}} \left\|\sum_{t=1}^{[Ts]} E\left[\Delta_{q_0,t}^T\left(\pi\right) D_{q_0,t}^T\left(\pi\right)|\mathcal{F}_{T,t-1}\right]\right\|, \\
II_2 &= \sup_{\substack{s\in[0,1] \\ \pi\in\Pi}} \left\|\sum_{t,q}^{[Ts]} E\left[\Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right)|\mathcal{F}_{T,t-1}\right]\right\|, \\
II_3 &= \sup_{\substack{s\in[0,1] \\ \pi\in\Pi}} \left\|\sum_{t,q}^{[Ts]} \left\{\Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) - E\left[\Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right)|\mathcal{F}_{T,t-1}\right]\right\}\right\| \\
III &= \sup_{\substack{s\in[0,1] \\ \pi\in\Pi}} \left\|\sum_{t,q}^{[Ts]} \Delta m_{q,t}^T\left(\pi\right) C_{q,t}^T\left(\pi\right) - E\left[\Delta m_{q,t}^T\left(\pi\right) C_{q,t}^T\left(\pi\right)|\mathcal{F}_{T,t-1}\right]||\right\|.
\end{aligned}
$$

We show that each of the above terms converges in probability towards zero. Since the arguments are more or less identical to the ones in E07, we only sketch them. To show that I_1 and I_2 tend to zero, note that, using the definition of $m_t^T\left(\pi\right)$, $\Delta_{q_0,t}^T\left(\pi\right) D_{q_0,t}^T\left(\pi\right) \leq 2M_t\mathbb{I}\left\{2M_t > \sqrt{T}a_{q_0}\right\}/\sqrt{T}$ where $M_t = F\left(y_{t-1}\right)\left\|e_t\right\|$ and F is the envelope defined in Assumption 3.2. Thus, for any fixed q_0,

$$
\begin{aligned}
I_1 &\leq \sup_{s\in[0,1],\pi\in\Pi} \sum_{t=1}^{[Ts]} \left\|\Delta_{q_0,t}^T\left(\pi\right) D_{q_0,t}^T\left(\pi\right)\right\| \leq \frac{2}{\sqrt{T}} \sum_{t=1}^{T} M_t\mathbb{I}\left\{2M_t > \sqrt{T}a_{q_0}\right\} \\
&\leq \frac{2}{a_{q_0}T} \sum_{t=1}^{T} M_t^2\mathbb{I}\left\{2M_t > \sqrt{T}a_{q_0}\right\} \xrightarrow{P} 0,
\end{aligned}
$$

since $E\left[M_t^2\right] = E\left[F^2\left(y_{t-1}\right)\right] E\left[\|e_t\|\right] < \infty$. Similarly, $I_2 \overset{P}{\to} 0$. By the same arguments as in E07, p. 133,

$$
\begin{aligned}
II_2 &\leq \left\{ \sup_{q \geq q_0+1} \sup_{\pi \in \Pi} \sum_{t=1}^{T} \left\| E\left[\Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) | \mathcal{F}_{T,t-1}\right] \right\| \right\} \times \left\{ \sum_{q=q_0+1}^{\infty} \frac{2^{-2q}}{a_q} \right\} \\
&\leq K \sum_{q=q_0+1}^{\infty} \frac{2^{-2q}}{a_q}
\end{aligned}
$$

on the set $\Omega_K^T = \left\{\sup_{q \geq 1} \alpha_T\left(\mathcal{P}_q\right)/2^{-2q} \leq K\right\}$, where $\alpha_T\left(\mathcal{P}_q\right)$ is defined in E07, p. 120 and $K > 0$ is a given constant. Thus, II_2 can be made arbitrarily small by choosing K and q_0 large enough. As for II_3, since on the set Ω_K^T the following bounds hold

$$
\begin{aligned}
\left\| \Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) - E\left[\Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) | \mathcal{F}_{T,t-1}\right] \right\| &\leq 2a_{q-1}, \\
\sum_{t=1}^{T} E\left[\Delta_{q,t}^T\left(\pi\right)^2 D_{q,t}^T\left(\pi\right) | \mathcal{F}_{T,t-1}\right] &\leq K2^{-2q},
\end{aligned}
$$

we can apply Lemma 9 in Pollard (1984, p. 177) to obtain that

$$
\begin{aligned}
& P\left(\sup_{s \in [0,1]} \sup_{\pi \in \Pi} \left\| \sum_{t=1}^{[Ts]} \left\{ \Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) - E\left[\Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) | \mathcal{F}_{T,t-1}\right] \right\} \right\| > \sqrt{12K2^{-2q}} | \Omega_K^T \right) \\
& \leq 3P\left(\sup_{\pi \in \Pi} \left\| \sum_{t=1}^{T} \left\{ \Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) - E\left[\Delta_{q,t}^T\left(\pi\right) D_{q,t}^T\left(\pi\right) | \mathcal{F}_{T,t-1}\right] \right\} \right\| > \sqrt{3K2^{-2q}} | \Omega_K^T \right).
\end{aligned}
$$

It now follows by the same arguments as in E07, p. 133-134 that II_3 can be made arbitrarily small in probability by choosing K and q_0 large enough. Similarly, III can be controlled by first applying Lemma 9 in Pollard (1984, p. 177) and then proceeding as in Escanciano (2007).

Next, consider the first term in eq. (A.1): First, note that due to stationarity of (e_t, y_t), for any $s_1 < s_2$,

$$
\phi_T\left(s_1, \pi\right) - \phi_T\left(s_2, \pi\right) = \frac{1}{\sqrt{T}} \sum_{t=[Ts_1]}^{[Ts_2]} f\left(y_{t-1}, \pi\right) e_t \overset{d}{=} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} f\left(y_{t-1}, \pi\right) e_t = \phi_T\left(s, \pi\right),
$$

where $s = s_2 - s_1$. We may therefore proceed as with the second term: With the same definitions as before,

$$
\sup_{|s_1 - s_2| < \delta} \sup_{\pi \in \Pi} \left\| \phi_T\left(s_1, \pi\right) - \phi_T\left(s_2, \pi\right) \right\| \overset{d}{=} \sup_{s < \delta} \sup_{\pi \in \Pi} \left\| \phi_T\left(s, \pi\right) \right\| \leq I_1 + I_2 + 2II_2 + II_3 + III + IV,
$$

where I_1, I_2, II_2, II_3 and III are handled as before while

$$
IV = \sup_{\substack{s \in [0,\delta] \\ \pi \in \Pi}} \left\| \sum_{t=1}^{[Ts]} m_t^T\left(\mathrm{pr}_{q_0}\left(\pi\right)\right) \right\|.
$$

Introducing the following functions,

$$
\bar{m}_{q,t}^T\left(\pi\right) := \sup_{\pi \in \Pi_{q,k}} \left\| m_t^T\left(\pi\right) \right\|, \quad C_{q,t}^T\left(\pi\right) := \mathbb{I}\left\{\bar{m}_{q,t}^T\left(\pi\right) \leq a_q\right\}, \quad D_{q,t}^T\left(\pi\right) := \mathbb{I}\left\{\bar{m}_{q,t}^T\left(\pi\right) > a_q\right\},
$$

we have $IV \leq 2IV_1 + IV_2 + IV_3$ where

$$IV_1 = \sup_{\substack{s\in[0,\delta] \\ \pi\in\Pi}} \left\| \sum_{t=1}^{[Ts]} E\left[m_t^T\left(\mathrm{pr}_{q_0}(\pi)\right) D_{q_0,t}^T(\pi) \,|\mathcal{F}_{T,t-1}\right] \right\|,$$

$$IV_2 = \sup_{\substack{s\in[0,\delta] \\ \pi\in\Pi}} \left\| \sum_{t=1}^{[Ts]} \left\{m_t^T\left(\mathrm{pr}_{q_0}(\pi)\right) D_{q_0,t}^T(\pi) - E\left[m_t^T\left(\mathrm{pr}_{q_0}(\pi)\right) D_{q_0,t}^T(\pi) \,|\mathcal{F}_{T,t-1}\right]\right\} \right\|$$

$$IV_3 = \sup_{\substack{s\in[0,\delta] \\ \pi\in\Pi}} \left\| \sum_{t=1}^{[Ts]} m_t^T\left(\mathrm{pr}_{q_0}(\pi)\right) C_{q_0,t}^T(\pi) - E\left[m_t^T\left(\mathrm{pr}_{q_0}(\pi)\right) C_{q_0,t}^T(\pi)\right] \right\|.$$

We can now employ the same arguments as before to show that each of the terms can be made arbitrarily small in probability by choosing $\delta > 0$ small enough. ■

Proof of Theorem 3.5. Define the mean-zero sequence $u_t(\pi) = f(y_{t-1};\pi) - E[f(y_{t-1};\pi)]$ and write

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

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

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

The first term is bounded by,

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

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,

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

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

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

where $\sup_{1\leq t\leq T}\|x_{T,t}\| = O_P(1)$. 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. ■

Proof of Theorem 3.6. Define,

$$V_T(\pi) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x'_{T,t-1} f(y_{t-1};\pi)\, e_t = \sum_{t=1}^{T} x'_{T,t-1}\Delta\phi_T(t/T,\pi) \in \mathbb{R}.$$

It follows by standard results that the convergence for any $\pi\in\Pi$ in (3.3) holds under the listed assumptions, see e.g. Kurtz and Protter (1991, Theorem 2.2). Next, the uniform convergence on $\mathcal{L}_\infty(\Pi)$ in (3.4) holds by Kallenberg (2002, Corollary 16.9) by establishing the tightness condition $E\|V_T(\pi)-V_T(\pi')\|^{2a} \leq c\|\pi-\pi'\|^{d_\pi+b}$, where $d_\pi = \dim(\Pi)$ and $a,b>0$. By Rosenthal's inequality (Hall and Heyde, 1980, p.23) and Cauchy-Schwarz inequalities, for $a>1$, with $f_t(\pi) = f(y_{t-1};\pi)$,

$$\begin{aligned}
&E\left[\|V_T(\pi)-V_T(\pi')\|^{2a}\right] \\
&\leq \frac{C}{T^a}\left(\sum_{t=1}^{T} E\left[E\left(\|x_{T,t-1}\|^2\|e_t\|^2\|f_t(\pi)-f_t(\pi')\|^2|\mathcal{F}_{t-1}\right)\right]\right)^a \\
&+\frac{C}{T^a}\sum_{t=1}^{T} E\left(\|x_{T,t-1}\|^{2a}\|e_t\|^{2a}\|f_t(\pi)-f_t(\pi')\|^{2a}\right) \\
&\leq C\|\Omega_e\|^a\left(\left(\sup_t E\|x_{T,t-1}\|^4\right)E\|f_t(\pi)-f_t(\pi')\|^4\right)^{a/2} \\
&+CT^{1-a}E\left(\sup_t\left(\|x_{T,t-1}\|^{2a}\|e_t\|^{2a}\|f_t(\pi)-f_t(\pi')\|^{2a}\right)\right) \\
&\leq C\|\Omega_e\|^a\left(\sup_t E\|x_{T,t-1}\|^4 E\|f_t(\pi)-f_t(\pi')\|^4\right)^{a/2} \\
&+CT^{1-a}\left(E\sup_t\|x_{T,t-1}\|^{6a}E\|e_t\|^{6a}E\sup_{\pi\in\Pi}\|f_t(\pi)\|^{6a}\right)^{1/3} \\
&\leq C\|\Omega_e\|^a\left(\sup_t E\|x_{T,t-1}\|^4\right)^{a/2}\left(EB(y_{t-1})^4\right)^{a/2}\|\pi-\pi'\|^{2a}+o(1).
\end{aligned}$$

Then with $2a>d_\pi$ the result follows. ■

B Proofs of Section 4

Proof of Lemma 4.6. We show the result by verifying the conditions in Theorem 3.4. Choose any $d\alpha$, $d\delta$ and $d\Phi$ and define $\lambda = vec(d\alpha, d\delta, d\Phi)$. We consider the sequence

$$\phi_T(s,\xi) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\left\{\lambda'_u u_t(\xi)+\lambda'_v v_t(\xi)+\lambda'_w w_t(\xi)\right\} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} f(y_{t-1};\xi)\, e_t,$$

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

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

Also note that $\pi = \xi$. Here, by Assumption 4.2

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

Thus,

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

Hence the requirement that $E\left\|f\left(y_{t-1}\right)\right\|^2 < \infty$, translates into $E\left\|Z_{0,t-1}\right\|^{2\rho}, E\left\|Z_{2,t-1}\right\|^2 < \infty$. Furthermore, by the differentiability of ψ and the polynomial boundedness, a.s.

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

Thus, Assumption 3.3 (ii), holds with $B\left(y_{t-1}\right) = \left\|Z_{0,t-1}\right\|^{\rho}$. Thus the requirement that $EB\left(y_{t-1}\right)^2 < \infty$, translates into $E\left\|Z_{0,t-1}\right\|^{2\rho} < \infty$. This verifies that Assumptions 4.4-4.1 imply that the Assumptions 3.1-3.3 of Theorem 3.4 hold, and hence the result follows for $\left(u'_t(\cdot), v'_t(\cdot), w'_t(\cdot)\right)$. The joint convergence holds by the marginal convergence in Assumption 4.5, in conjunction with the fact that $\left(u'_t(\cdot), v'_t(\cdot), w'_t(\cdot)\right)$ and X_t are defined in terms of $\left(\varepsilon_s\right)_{s\leq t}$. ■

Proof of Theorems 4.7. 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 4.2, while (ii)-(iii) follow by Lemmas C.1, C.2 and C.3:

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

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

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

with $H_\infty(d\gamma, d\bar{\gamma})$ 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,

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

where $S_\infty(d\vartheta)$ is given in Lemma C.1. We conclude that $V_T^{1/2}(vec(\hat{\vartheta}_T) - vec(\vartheta_0)) \xrightarrow{D} vec(d\vartheta_\infty)$, where ϑ_∞ satisfies $S_\infty(d\vartheta) = H_\infty(d\vartheta, d\vartheta_\infty)$ for all directions $d\vartheta$. This together with eq. (D.1) imply the results stated in Theorem 4.7. ■

Proof of Theorems 4.8. We proceed as in the proof of Theorem 4.7: Set $\gamma = vec(\theta)$, $\pi = vec(\xi)$, $Q_T(\gamma, \pi) = -L_T^*(\theta, \xi)/T$, $v_T = T$ 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 4.9. 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 $vec(\eta) = (R_\eta)_\perp \tau$ 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, $L_T^*(b, \eta) = \tilde{L}_T^*(b, \tau)$ and $\tilde{L}_T^*(b, \tau) := L_T^*\left(b, (R_\eta)_\perp \tau\right)$. Taking differentials w.r.t. (b, τ),

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

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

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 $dL_T^*(\theta; db) = \mathbb{S}_{b,T}(\theta)' vec(db')$ and $d^2L_T^*(\theta; db, db) = vec(db')' \mathbb{H}_{bb,T}(\theta) vec(db')$; 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 4.8, we now obtain that $(\tilde{b}, \tilde{\tau})$ satisfies

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

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 4.8, 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 on $\mathcal{L}_\infty\left(\Xi\right)$,

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

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

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

Thus,

$$LR_T\left(\xi\right) \xrightarrow{W} \mathbb{S}_\theta(\xi)'\mathbb{P}\left(\xi\right)'\mathbb{H}_{\theta\theta}(\xi)\mathbb{P}\left(\xi\right)\mathbb{S}_\theta(\xi) = \mathbb{V}_\theta\left(\xi\right)'\mathbb{V}_\theta\left(\xi\right) \quad \text{on } \mathcal{L}_\infty\left(\Xi\right),$$

where $\mathbb{V}_\theta\left(\xi\right)$ 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

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

In conclusion, on $\mathcal{L}_{\infty}(\Xi)$,

$$LM_T(\xi) \xrightarrow{W} \mathbb{S}_\theta(\xi)'\mathbb{P}(\xi)'\,\mathbb{H}_{\theta\theta}(\xi)\mathbb{P}(\theta_0,\xi)\,\mathbb{S}_\theta(\theta_0,\xi) = \mathbb{V}_\theta(\xi)'\,\mathbb{V}_\theta(\xi).$$

■

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 $q_0 = 2\rho$, $q_2 = 2$ and Assumption 4.5 (i), 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 as $T \to \infty$,*

$$T^{-1/2}dL_T(\theta_0,\xi; V_{\theta\theta,T}^{-1/2}d\theta) = S_{\theta,T}(\theta_0,\xi; V_{\theta\theta,T}^{-1/2}d\theta) \xrightarrow{W} S_{\theta,\infty}(\xi; d\theta) \text{ on } \mathcal{L}_{\infty}(\Xi) \quad \text{(C.1)}$$

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

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

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

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

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

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. (2.24), where we note that $w_t(\xi)$ does not depend on ξ when $\delta_0 = 0$ and we therefore simply write w_t. By Lemma 4.6,

$$S_{\eta,T}(\theta_0, \xi; d\eta) = T^{-1/2}(vec(d\eta))' \sum_{t=1}^{T} u_t(\xi) \xrightarrow{W} (vec(d\eta))' \int_0^1 dB_u(s,\xi) \text{ on } \mathcal{L}_\infty(\Xi),$$

and next, as $S_{b,T}$ does not dend on ξ, apply Theorem 3.6 (i) with $x_{T,t-1} := db' K_T^{-1} Z_{1,t-1}$, $\pi = \xi$ and $f(y_{t-1}, \pi)$, e_t as in the proof of Lemma 4.6 to obtain,

$$S_{b,T}\left(\theta_0, \xi; K_T^{-1} db\right) = T^{-1/2} \sum_{t=1}^{T} \left(\left\{Z_{1,t-1}' K_T^{-1}\right\} db\right) w_t \xrightarrow{W} \int_0^1 \left(F(s)' db\right) dB_w(s) \quad \text{on } \mathcal{L}_\infty(\Xi).$$

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

Proof of Part 1 ($\delta_0 \neq 0$)*:* The convergence results of Part 2 holds, and in addition by Lemma 4.6,

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

■

Lemma C.2 *Under Assumptions 4.1-4.5 for q_i specified below, 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 $q_0, q_2 = 2$ and q_1 unconstrained,*

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

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 $q_0 = \rho \max(4, 3d_\xi)$, $q_1 = \max(4, 3d_\xi)$ and $q_2 = 2$ it holds,*

$$-\frac{1}{T} d^2 L_T(\theta_0, \xi; V_{\theta,T}^{-1/2} d\theta, V_{\theta,T}^{-1/2} d\bar{\theta}) \xrightarrow{W} H_{\theta\theta,\infty}(\xi; d\theta, d\bar{\theta}) \text{ on } \mathcal{L}_\infty(\Xi),$$

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 $\xi\in\Xi$:

$$\begin{aligned}&\text{Claim 2.1}: H_{\eta,\eta}(\xi)\overset{P}{\to}vec(d\eta)'\Sigma_{uu}(\xi)\,vec(d\bar{\eta}),\\&\text{Claim 2.2}: H_{b,b}\left(\xi,K_T^{-1}db,K_T^{-1}d\bar{b}\right)\overset{D}{\to}tr\{(db)'\int_0^1F(s)F(s)'ds(d\bar{b})\Sigma_{w,w}\},\\&\text{Claim 2.3}: H_{b,\eta}\left(\xi,K_T^{-1}db\right)\overset{W}{\to}\int_0^1F(s)'dsdb\Sigma_{w,u}(\xi,\xi)\,vec(d\bar{\eta})\quad\text{on }\mathcal{L}_\infty(\Xi).\end{aligned}$$

Proof of Claim 2.1: We have

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

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

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

with

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

and

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

At θ_0 with $\delta_0 = 0, H^{(1)}_{b,b}\left(\xi, db, d\bar{b}\right) = 0$, applying Theorem 3.5, with $x_{T,t-1} = d\bar{b}'K_T^{-1}Z_{1,t-1}$,

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

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

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

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

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

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

$$H^{(1)}_{b,\eta}(\xi, db) = \frac{1}{T}\sum_{t=1}^{T} Z'_{1,t-1}dbf^{(1)}_{t-1}(\xi)\, d\bar{\eta}'e_t. \tag{C.12}$$

By the same arguments as in the proof of Lemma 4.6, we see that $f^{(1)}_{t-1}(\xi)$ satisfies Assumption 3.2 (i) with

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

Furthermore, by the differentiability of ψ,

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

Thus, the requirement in Theorem 3.6 (ii) translates into $q_0 = \rho\max(4, 3d_\xi)$, and $q_1 = \max(4, 3d_\xi)$. Theorem 3.6 now implies that $\sqrt{T}H^{(1)}_{b,\eta}\left(\xi, K_T^{-1}db\right) = O_P(1)$ and hence, $H^{(1)}_{b,\eta}\left(\xi, K_T^{-1}db\right) = o_P(1)$, uniformly in ξ as desired.

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

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

Applying Theorem 3.5 gives at θ_0,

$$H_{b,\eta}^{(2)}\left(\xi,K_T^{-1}db\right)\overset{W}{\to}\int_0^1 F(s)'\,dsdb\Sigma_{w,u}(\xi,\xi)\,vec(d\bar{\eta})\text{ on }\mathcal{L}_\infty(\Xi)$$

This finishes the proof of part 2.

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

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

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

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

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

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

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

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 $H_{\eta,\xi}(\xi)=H_{\eta,\xi}^{(1)}(\xi)+H_{\eta,\xi}^{(2)}(\xi)$

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

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

By Theorem 3.4, for $\vartheta=\vartheta_0$, $T^{1/2}H_{\eta,\xi}^{(1)}(\xi_0)=O_P(1)$, while by the LLN, $H_{\eta,\xi}^{(2)}(\xi_0)\overset{P}{\to}d\eta'\Sigma_{u,v}(\xi_0,\xi_0)\,d\xi$.

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

$$\partial_{z,\xi}^2\psi(z,\xi)=\frac{\partial vec(\partial_z\psi(z,\xi))}{\partial vec(\xi)'},\tag{C.13}$$

we find,

$$H_{b,\xi}^{(1)}(\xi,db)=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\theta)'\,\Omega_0^{-1}\delta\left(Z_{1,t-1}'db\otimes I_{r_\delta}\right)\partial_{z,\xi}^2\psi(Z_t(b);\xi)\,vec\left(d\bar{\xi}\right),\tag{C.14}$$

$$H_{b,\xi}^{(2)}(\xi,db)=\frac{1}{T}\sum_{t=1}^{T}\left[\{\alpha+\delta\partial_z\psi(Z_t(b);\xi)\}\,db'Z_{1,t-1}\right]'\Omega_0^{-1}\delta\partial_\xi\psi(Z_t(b);\xi)\,vec\left(d\bar{\xi}\right),\tag{C.15}$$

By Theorem 3.6 (i), at ϑ_0, $H_{b,\xi}^{(1)}\left(\xi_0,K_T^{-1}db\right)=o_P(1)$ and by Theorem 3.5, $H_{b,\xi}^{(2)}\left(\xi_0,K_T^{-1}db\right)$ converges towards the claimed limit.

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

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

with

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

and

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

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

Lemma C.3 *Suppose Assumptions 4.1-4.4 hold with $q_0 = 2\rho$, $q_2 = 2$ and Assumption 4.5 (i) holds. 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 $\delta_0 \neq 0$,*

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

   for a sequence of neighborhoods

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

2. *If $\delta_0 = 0$,*

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

   for a sequence of neighborhoods

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

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

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

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

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

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

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

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

where

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

Likewise,

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

Hence, by Assumption 4.2,

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

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

$$\|Z_t(b)\|\leq\|Z_{0,t-1}\|+\epsilon\left\|K_T^{-1}Z_{1,t-1}\right\|\leq\|Z_{0,t-1}\|+\epsilon\sup_{u\in[0,1]}\left\|K_T^{-1}Z_{1,[Tu]}\right\|. \tag{C.16}$$

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

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

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 $H_{\xi,\xi}(\xi)$,

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

and

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

where $\partial^2_{\xi,z}\psi\left(Z_t(b);\xi\right)$ is defined in eq. (C.13), and

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

Thus,

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

and so $\left|d\left(H^{(1)}_{\xi,\xi}(\xi),K_T^{-1}d\tilde{b}\right)\right| = O_P(1)$ by eq. (C.16).

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

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

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

with

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

and

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

Thus, multiplying all directions with K_T^{-1} and using eq. (C.16),

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

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 $\pi\in\Pi$. Let $dQ_T(\gamma_0,\pi;d\gamma)$ and $d^2Q_T(\gamma_0,\pi;d\gamma,d\bar{\gamma})$ denote the first and second order differential of $Q_T(\gamma,\pi)$ w.r.t. γ.

Lemma D.1 *Assume that:*

(i) $Q_T(\gamma,\pi)$ *is three times continuously differentiable in* γ *for all* π.

(ii) *There exists* γ_0 *in the interior of* Γ *and a sequence of nonsingular matrices* $U_T \in \mathbb{R}^{d\times d}$ *such that* $U_T^{-1} = O(1)$ *and*

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

where the convergence takes place on $\mathcal{L}_\infty(\Pi)$*, and where the stochastic process* $\inf_\pi H_\infty(\pi,d\gamma,d\bar{\gamma}) > 0$ *a.s.*

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

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

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 $\gamma_0+U_T^{-1/2}d_T(\pi)\in\mathcal{N}_T(\gamma_0)$, such that in particular $\|d_T(\pi)\| < \epsilon$

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

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 $\bar{d}_T(\pi) = U_T^{1/2}\left(\bar{\gamma}(\pi)-\gamma_0\right)$. 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, using (iii) and $\|d_T(\pi)\|, \|\bar{d}_T(\pi)\| < \epsilon$,

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

Thus,

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

where the second equality follows by (ii). Note here that since $dQ_T(\gamma_0,\pi;d\gamma)$ and $d^2Q_T(\gamma_0,\pi;d\gamma,d\gamma)$ are linear and quadratic in $d\gamma$ respectively, then the pointwise convergence in (ii) implies uniform convergence in $d\gamma$. As $H_\infty(\pi,d_T(\pi),d_T(\pi)) > 0$ 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:*

$$\left(dQ_T\left(\gamma_0,\pi;\nu_T^{1/2}U_T^{-1/2}d\gamma\right),\ d^2Q_T\left(\gamma_0,\pi;U_T^{-1/2}d\gamma,U_T^{-1/2}d\bar{\gamma}\right)\right)$$
$$\xrightarrow{W}\left(S_\infty(\pi,d\gamma),H_\infty(\pi,d\gamma,d\bar{\gamma})\right)\ \text{on}\ \mathcal{L}_\infty(\Pi).$$

Then $\nu_T^{1/2}U_T^{1/2}(\hat{\gamma}(\pi)-\gamma_0)\xrightarrow{W}-\mathbb{H}^{-1}(\pi)\,\mathbb{S}(\pi)$ on $\mathcal{L}_\infty(\Pi)$, 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

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

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 $H_A^{(1)}:\delta_0\neq0$ 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

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