

# LIKELIHOOD-BASED INFERENCE IN NONLINEAR ERROR-CORRECTION MODELS

DENNIS KRISTENSEN\* AND ANDERS RAHBK†

7 OCTOBER 2007

## Abstract

We consider a class of vector nonlinear error correction models where the transfer function (or loadings) of the stationary relationships is nonlinear. This includes in particular the smooth transition models.

A general representation theorem is given which establishes the dynamic properties of the process in terms of stochastic and deterministic trends as well as stationary components. In particular, the behaviour of the cointegrating relations is described in terms of geometric ergodicity. Despite the fact that no deterministic terms are included, the process will have both stochastic trends and a linear trend in general.

Gaussian likelihood-based estimators are considered for the long-run cointegration parameters, and the short-run parameters. Asymptotic theory is provided for these and it is discussed to what extent asymptotic normality and mixed normality can be found. A simulation study reveals that cointegration vectors and the shape of the adjustment are quite accurately estimated by maximum likelihood, while at the same time there is very little information about some of the individual parameters entering the adjustment function.

## 1 Introduction

In this paper we study likelihood-based estimation of the parameters of a class of multivariate, or *vector*, nonlinear error-correction models (ECMs).

---

\*Columbia University and CREATES. Email: dk2313@columbia.edu

†University of Copenhagen and CREATES. Email: rahbek@math.ku.dk

Our main contribution is to give a full asymptotic theory for the likelihood estimators, including the cointegrating relationships and vector error correction (or adjustment) parameters. The much applied smooth transition error correction model (STECM) – originating from Granger and Teräsvirta (1993) and for which the vector version is discussed in van Dijk, Teräsvirta and Franses (2002) – is particularly included in our class of models considered, and a simulation-based study of the properties of the estimators for such models is included.

Nonlinear ECMs have recently been applied to account for non-linear adjustment of key macroeconomic and financial time series to stable, or cointegrated, relationships, which are *linear* combinations of the included variables. See for example the term-structure studies in Anderson (1997), Balke and Fomby (1997), Bec and Rahbek (2004), Corradi, Swanson and White (2000), Hansen and Seo (2002) and Seo (2003) where the (speed of) adjustment is parametrized as a function of interest rate spreads, and similarly, for example, Kapetanios, Shin and Snell (2006) and Psaradakis, Sola and Spagnolo (2004) for studies of asset prices and dividends. See also Escribano (2004) and references therein.

With  $p$ -dimensional observations  $X_t$ ,  $t = 1, \dots, T$ , we consider the maximum likelihood estimator (MLE) of the parameter vector  $\theta = (\eta, \beta)$  based on the Gaussian likelihood function. With  $\beta$  is a  $(p \times r)$ -dimensional matrix,  $\beta$  parametrizes the  $r < p$ , ‘long-run’ cointegration vectors, while  $\eta$  is the ‘short-run’ parameter vector, which parametrizes the nonlinear adjustment of  $\Delta X_t$  to the stable relationships  $\beta' X_{t-1}$  and lagged differences,  $\Delta X_{t-i}$ ,  $i = 1, \dots, k$ .

When deriving the asymptotic properties of the ML estimator of  $\theta = (\eta, \beta)$ ,  $\hat{\theta}$ , we first study the dynamic properties of the model. In particular, we show that similar to the linear case, the process  $X_t$  can be decomposed into (i) stationary, geometrically ergodic components, (ii) a linear trend due to the nonlinearity term in the conditional mean, and (iii) stochastic trend components asymptotically equivalent to random walks. That is, correctly normalized the stochastic trend components satisfy an invariance principle or functional central limit theorem. These findings are closely related to the results in Bec and Rahbek (2004) and Saikkonen (2005, 2006), see also Corradi et al (2000) for the Markov case ( $k = 0$ ).

Next we provide a detailed study of the asymptotic behaviour of the score function, observed information and third order differentials of the likelihood function. Consistency of  $\hat{\theta}$  and the joint asymptotic distributions of the adjustment and cointegration parameter estimators,  $\hat{\eta}$  and  $\hat{\beta}$  can be found from the likelihood theory provided in Lemma 7 and 8 in the appendix.

In general we find that the MLE of the cointegrating vectors  $\beta$  are super-consistent in all directions but one in which it is  $T^{3/2}$ -consistent, while the

short-run parameters are consistent at the usual  $\sqrt{T}$ -rate. The higher rate in one direction reflects the fact that despite no deterministic terms in the model, a linear trend is induced in the process. It corresponds to the case of a linear ECM model with an unrestricted constant, where the constant aggregates to a linear trend, see Johansen (1996). Also we find that short-run and long-run parameters are not asymptotically orthogonal, or uncorrelated, with  $\beta$  as in the linear ECMs. Moreover, the cointegrating relationships turn out not to be asymptotically mixed Gaussian, unless certain quite restrictive regularity conditions hold as discussed below. From our simulation study of the STECM, we conclude that  $\beta$  is quite accurately estimated and likewise the adjustment function itself show little empirical variation, while  $\eta$  has quite a large empirical variation.

While asymptotic likelihood inference for nonlinear *vector* error correction models has not been considered elsewhere, inference was studied in de Jong (2001, 2002) for the *single* equation, or *partial model*, case. In *single* equation nonlinear error-correction models, the  $p$ -dimensional observations  $X_t$  are decomposed as  $X_t = (Y_t, Z_t)'$ , with  $Y_t$  univariate and  $Z_t$   $(p - 1)$ -dimensional. Under the assumption that  $Z_t$  is an I(1) explanatory variable satisfying some invariance principle and the assumption of a single cointegrating relation, de Jong (2001, 2002) studies asymptotic inference for the parameters of the single equation model of  $Y_t$  given  $Z_t$ , as well as lags of these. In de Jong (2001) the cointegration relation is assumed super-consistently estimated from elsewhere, while in de Jong (2002) it is estimated in the single equation nonlinear regression. In accordance with our results on the joint distribution of  $\hat{\theta}$ , de Jong (2001, 2002) find that the short-run parameters are not asymptotically Gaussian. Further discussion of the results are given after Theorem 5.

With the focus of deriving tests for ‘stationarity-ergodicity’, that is cointegration in Markovian nonlinear error correction models, Corradi et al (2000) finds asymptotic properties of the linear OLS estimator in the case of a single cointegration vector, when the data generating process is a nonlinear error correction model. Likewise, Kapetanios and Shin (2006) and Seo (2006) study test statistics for cointegration in single equation nonlinear error correction models. Pitarakis and Gonzalo (2006, 2007) study threshold error correction models and testing.

The remains of the paper are organised as follows: In Section 2, we introduce the model and then proceed to establish the dynamic properties of the process under regularity conditions in Section 3. We propose estimators of the unknown parameters in Section 4, and derive their asymptotic properties under additional conditions. Section 5 contains the results of a simulation study. We conclude in Section 6. All proofs have been relegated to Appendix A, while Appendix B and C contain lemmas, and auxiliary results respec-

tively. Finally Appendix D and E contain Figures 1–12 and Tables 1 and 2 respectively.

Some notation has been used throughout: With  $\beta$  a  $(p \times r)$  dimensional matrix of rank  $r < p$ ,  $\beta_{\perp}$  is the  $(p \times (p - r))$  dimensional matrix of rank  $p - r$  for which  $\beta' \beta_{\perp} = 0$ . Also  $\bar{\beta} = \beta (\beta' \beta)^{-1}$ , such that  $I = \bar{\beta} \beta' + \bar{\beta}_{\perp} \beta'_{\perp}$ . We use  $c$  to denote a generic constant.

## 2 The Model

Consider the class of discrete time vector process  $\{X_t\}$ ,  $X_t \in \mathbb{R}^p$ , solving

$$\Delta X_t = g(Z_{t-1}; \gamma) + \Phi_1 \Delta X_{t-1} + \dots + \Phi_k \Delta X_{t-k} + \varepsilon_t, \quad (1)$$

with  $Z_t = \beta' X_t$ ,  $\{\varepsilon_t\}$  an i.i.d. error process which satisfy,

$$E[\varepsilon_t] = 0, \quad \Omega \equiv E[\varepsilon_t \varepsilon_t'] < \infty, \quad (2)$$

and  $t = 1, \dots, T$ . The parameters to be estimated are given by the ‘long-run’ cointegration parameter matrix  $\beta \in \mathbb{R}^{p \times r}$ , the ‘short-run’ parameters  $\gamma \in \mathcal{G} \subseteq \mathbb{R}^d$  and  $\Phi_i \in \mathbb{R}^{p \times p}$  parametrizing the nonlinear adjustment in  $g(\cdot)$  and the lagged differences respectively. Finally,  $\Omega$  is a  $p$ -dimensional positive definite covariance matrix. The nonlinear error correction function  $g$ , is a possibly nonlinear function specified as  $g : \mathbb{R}^r \times \mathcal{G} \mapsto \mathbb{R}^p$ .

As mentioned a key example is given by the smooth transition error correction model (STECM) in Granger and Teräsvirta (1993) and van Dijk, Teräsvirta and Franses (2002). A general vector version of the STECM which allows for more than one cointegration relation can be represented in terms of the nonlinear error correction, or response, function  $g$ , given by

$$g(z; \gamma) = (\psi(z) \tilde{\alpha} + \alpha) z. \quad (3)$$

Here  $\alpha$  and  $\tilde{\alpha}$  are  $(p \times r)$ -dimensional matrices, while  $\psi(\cdot)$  is a general function for which  $\psi(z) = o(1)$  as  $\|z\| \rightarrow \infty$ . A key example is the logistic specification, where

$$\psi(z) = (1 + \exp\{(z - \omega)' A (z - \omega)\})^{-1}, \quad (4)$$

with  $A$  a positive definite  $(r \times r)$ -dimensional matrix while  $\omega$  is an  $r$ -dimensional vector. The parameter  $\gamma$  is in this case given by

$$\gamma = (\text{vec}(\alpha)', \text{vec}(\tilde{\alpha})', \omega, \text{vec}(A)')' \in \mathbb{R}^d$$

with  $d = r(2p + r + 1)$ .

### 3 Properties of the Process

We make the following assumptions:

**A.1** The sequence  $\{\varepsilon_t\}$  is i.i.d. mean-zero random variables on  $\mathbb{R}^p$  with positive definite covariance matrix  $\Omega$ . The marginal distribution is given by a continuous density  $f_\varepsilon > 0$  satisfying (2) and  $E[\|\varepsilon_t\|^{2s}] < \infty$  for some  $s > 1$ .

**A.2** The function  $g(\cdot; \gamma) : \mathbb{R}^r \mapsto \mathbb{R}^p$  satisfies

$$g(z; \gamma) = \alpha z + o(\|z\|) \quad \text{as } \|z\| \rightarrow \infty,$$

for some  $\alpha \in \mathbb{R}^{p \times r}$ .

**A.3** With the characteristic polynomial  $A(z)$  defined by,

$$A(\lambda) = I_p(1 - \lambda) - \alpha\beta'\lambda - \sum_{i=1}^k \Phi_i(1 - \lambda)\lambda^i, \quad \lambda \in \mathbb{C},$$

assume that  $A(\lambda)$  has exactly  $(p - r)$  roots at  $\lambda = 1$ , while the remaining roots satisfy  $|\lambda| > 1$ .

Assumption A.1 implies that  $X_t$  can be embedded in a Markov chain which is shown below to be geometrically ergodic. It is particularly satisfied if  $\varepsilon_t$  are assumed to be i.i.d. Gaussian, which is used when defining the estimation function below.

Assumption A.2 states that for large values of the cointegrating relations, the nonlinearity is vanishing. This assumption is satisfied for many of existing nonlinear error-correction models, such as in analyses of real exchange rates or yield curve dynamics, see e.g. Dumas (1992) and Sercu et al.(1995) for the former and Anderson (1997) for the latter.

Assumption A.3 is the well-known cointegration regularity condition from linear vector autoregressive models. This assumption is, together with A.2, crucial when establishing Theorem 1 below. Note that upon estimation, it can be verified by computation of the roots of  $A(\lambda)$ .

Assumptions A.2 and A.3, may be reformulated to allow for the case where the linear adjustment coefficient  $\alpha$  in A.2 is allowed to depend on the direction of  $z$  and not the size alone. However, as shown in Saikkonen (2006) this implies that the regularity condition involves the concept of generalized spectral radius which is very difficult to verify for practical purposes.

**Theorem 1** Assume that (A.1)-(A.3) hold.

(i) The process  $\mathbf{Y}_t = (X_t'\beta, \Delta X_t'\beta_\perp, \dots, X_{t-k}'\beta, \Delta X_{t-k}'\beta_\perp)'$  is geometrically ergodic. In particular, the initial value  $\mathbf{Y}_0$  can be given an initial distribution such that  $\mathbf{Y}_t$  is stationary and ergodic, with  $E[\|\mathbf{Y}_t\|^{2s}] < \infty$ .

(ii) The process  $X_t$  has the representation

$$X_t = C \sum_{i=1}^t s_i + C\mu t + \nu_t + D, \quad (5)$$

where  $C = \beta_\perp \left( \alpha'_\perp \left( I - \sum_{i=1}^k \Phi_i \right) \beta_\perp \right)^{-1} \alpha'_\perp$ ,  $\mu = Eg(\beta'X_t; \gamma)$ , and

$$s_t = \varepsilon_t + g_0(\beta'_0 X_{t-1}) - \mu$$

is a stationary and also geometrically mixing sequence with  $E[s_t] = 0$  and  $E[\|s_t\|^{2s}] < \infty$ . Moreover,  $D$  is a constant which satisfies  $\beta'D = 0$  and depends on initial values  $(X_0, \Delta X_0, \dots, \Delta X_{-k+1})$ . Finally  $\nu_t$  is a stationary and geometrically mixing sequence.

**Remark 2** (i) The process contains a linear trend term induced by  $\mu = Eg(\beta'X_t; \gamma)$  which in most cases will be non-zero. An important exception is the linear case, and also the case of STECM with  $\alpha \neq \tilde{\alpha}$  but proportional such that  $\alpha'_\perp \tilde{\alpha} = 0$ ; see also the discussions in Bec and Rahbek (2004).

(ii) The random sequence  $\{s_t\}$  driving the stochastic trend is not necessarily a Martingale Difference. This will have important implications for the asymptotic properties of the MLE in the general case as we shall see in the next Section.

## 4 Likelihood-based Estimation of the Parameters

We here define estimators of the unknown parameters  $\gamma, \beta, \Phi_1, \dots, \Phi_k$  and  $\Omega$  based on the Gaussian likelihood. Let in the following  $\gamma_0$  and  $\beta_0$  denote the true parameter values, and likewise for the remaining parameters.

Before stating results on the asymptotic behaviour of the likelihood estimators of  $\gamma$  and  $\beta$ , we use the results in Theorem 1 to define a normalized, and hence identified, version of  $\beta$ . With  $C, \mu$  and  $\beta$  defined in Theorem

1 set  $\delta = C\mu \in \mathbb{R}^p$ , and  $\kappa = (\beta, \delta)_\perp \in \mathbb{R}^{p \times (p-r-1)}$ . Then using the coordinate system  $(\beta_0, \kappa_0, \delta_0)$  we have by simple orthogonal projection that  $\beta = \beta_0 \bar{\beta}'_0 \beta + \kappa_0 \bar{\kappa}'_0 \beta + \bar{\delta}_0 \delta'_0 \beta$  and hence, we can define the normalized version as  $\tilde{\beta} \equiv \beta (\bar{\beta}'_0 \beta)^{-1}$ , with

$$\tilde{\beta} - \beta_0 = \kappa_0 b_\kappa + \bar{\delta}_0 b_\delta = (\kappa_0, \bar{\delta}_0) b.$$

This defines the normalized cointegration matrix parameter,

$$b = (b'_\kappa, b'_\delta)' \in \mathbb{R}^{p-r \times r}. \quad (6)$$

With  $\Phi = [\Phi_1 | \dots | \Phi_k] \in \mathbb{R}^{p \times pk}$  define furthermore,

$$\eta = (\gamma, \phi), \quad \phi = \text{vec}(\Phi) \in \mathbb{R}^{p^2 k}. \quad (7)$$

The parameters of interest for our asymptotic results are then given by

$$\theta = (\eta, b) \quad (8)$$

in terms of (6) and (7). Various other normalizations on  $\beta$  exist in the literature; the one chosen here is theoretically appealing as it means that our results can be presented in a straightforward way, see Theorem 5 below.

Note that in the STECM model with  $g(z) = (\alpha + \tilde{\alpha}\psi(z))z$ , where  $\psi$  is given by (4), by a simple reparametrization in terms of  $\mathbf{A}$ ,  $\psi(\cdot)$  is invariant to the proposed normalisation, and so the same holds for the likelihood function defined below. It implies for example that the limiting distribution of  $\beta$  normalized in some other way can be derived from  $\tilde{\beta}$  using simple Taylor expansion arguments. Also the invariance implies that limiting distributions of likelihood based test statistics can be found using the results for  $\tilde{\beta}$ .

Next, in terms of  $\theta$  the model we rewrite the general model as,

$$\Delta X_t = g(Z_{0,t-1} + b'Z_{1,t-1}; \gamma) + \Phi Z_{2,t-1} + \varepsilon_t,$$

where the right hand side  $Z$  variables are defined by,

$$\begin{aligned} Z_{0,t} &= \beta'_0 X_t \in \mathbb{R}^r, \\ Z_{1,t} &= (Z'_{\kappa,t}, Z'_{\delta,t})' = (X'_t \kappa_0, t)' \in \mathbb{R}^{p-r} \quad \text{and} \\ Z_{2,t} &= (\Delta X'_t, \dots, \Delta X'_{t-k+1})' \in \mathbb{R}^{pk}. \end{aligned} \quad (9)$$

By Theorem 1,  $Z_{0,t}$  and  $Z_{2,t}$  are stationary regressors, while  $Z_{1,t}$  is a non-stationary regressor for which all coordinates but one, that is  $p-r-1$ , satisfy an invariance principle, cf. Theorem 3 below. The last coordinate in  $Z_{1,t}$  is a linear deterministic trend.

Assuming that the covariance matrix  $\Omega$  is known, the negative Gaussian log-likelihood function up to a constant and a scale is given by

$$L_T(\theta) \equiv \sum_{t=1}^T \varepsilon_t(\theta)' \Omega^{-1} \varepsilon_t(\theta), \quad (10)$$

where

$$\varepsilon_t(\theta) = \Delta X_t - g(Z_{0,t-1} + b' Z_{1,t-1}; \gamma) - \Phi Z_{2,t-1}.$$

We define the MLE as,

$$\hat{\theta} = \arg \min_{\theta \in \Theta} L_T(\theta),$$

where  $\Theta = \mathcal{G} \times \mathbb{R}^{p^2 k} \times \mathbb{R}^{(p-r) \times r}$ ,  $\mathcal{G} \subseteq \mathbb{R}^d$ . Note that with  $\Omega$  unknown, the MLE is given by the residual sum of squares,

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

and simple iterations maximizing over  $\theta$  and  $\Omega$  (with  $\Omega$  and  $\theta$  respectively fixed) lead to the MLE of  $\Omega$  and  $\theta$ . The asymptotics of  $\hat{\Omega}$  are standard, and will not influence the asymptotics of  $\hat{\theta}$ . We therefore treat it as known here and in the following.

In order to derive the asymptotics of  $\hat{\theta}$ , we need a set of additional regularity conditions. Define  $g_0(z) = g(z; \gamma_0)$ , that is, the  $g(\cdot)$  function evaluated at the true value. In terms of  $g_0$  define the Jacobian matrices w.r.t.  $z$  and  $\gamma$ :

$$\partial_z g_0(z) \equiv \left. \frac{\partial g(z; \gamma)}{\partial z} \right|_{\gamma=\gamma_0} \in \mathbb{R}^{p \times r}, \quad \partial_\gamma g_0(z) \equiv \left. \frac{\partial g(z; \gamma)}{\partial \gamma} \right|_{\gamma=\gamma_0} \in \mathbb{R}^{p \times d} \quad (11)$$

Also, let

$$\partial_{z_i, z_j}^2 g_0(z) \equiv \left. \frac{\partial^2 g(z; \gamma)}{\partial z_i \partial z_j} \right|_{\gamma=\gamma_0} \in \mathbb{R}^p, \quad \partial_{z_i, z_j, z_k}^3 g_0(z) \equiv \left. \frac{\partial^3 g(z; \gamma)}{\partial z_i \partial z_j \partial z_k} \right|_{\gamma=\gamma_0} \in \mathbb{R}^p.$$

Define furthermore the processes  $u_t \in \mathbb{R}^{d+p^2 k}$  and  $v_t \in \mathbb{R}^r$  by

$$\begin{aligned} u_t &= (u'_{1,t}, u'_{2,t})' = (\partial_\gamma g_0(Z_{0,t-1}), Z'_{2,t-1} \otimes I_p)' \Omega^{-1} \varepsilon, \\ v_t &= \partial_z g_0(Z_{0,t-1})' \Omega^{-1} \varepsilon_t. \end{aligned} \quad (12)$$

These will be influential in deriving the asymptotics of the MLE since both the score and the observed information can be expressed in terms of these together with  $Z_{1,t}$ . We shall need the following assumptions:



**A.4** With  $u_t$  and  $v_t$  defined in (12), the covariance matrix,

$$\text{Var} \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{pmatrix}$$

is positive definite.

**A.5** The function  $g : \mathbb{R}^r \times \mathcal{G} \mapsto \mathbb{R}^p$  satisfies that all partial derivatives with respect to  $z$  and  $\gamma$  up to the third order derivatives are  $O(\|z\|)$  uniformly over  $\gamma$ . For example,

$$\begin{aligned} \|\partial_{\gamma_j} g(z; \gamma)\| &\leq c \|z\|, \quad \|\partial_{z, \gamma_k}^2 g(z; \gamma)\| \leq c \|z\| \quad \text{and} \\ \|\partial_{\gamma_i, \gamma_j, z}^3 g(z; \gamma)\| &\leq c \|z\| \end{aligned}$$

where  $c$  does not depend on  $\gamma$ .

Note that in A.4,  $\Sigma_{uu} \in \mathbb{R}^{(d+p^2k) \times (d+p^2k)}$ ,  $\Sigma_{vv} \in \mathbb{R}^{r \times r}$  and  $\Sigma_{uv} \in \mathbb{R}^{(d+p^2k) \times r}$  are given by,

$$\Sigma_{uu} = E \left[ (\partial_\gamma g_0(Z_{0,t-1}), Z'_{2,t-1} \otimes I_p)' \Omega^{-1} (\partial_\gamma g_0(Z_{0,t-1}), Z'_{2,t-1} \otimes I_p) \right]$$

$$\Sigma_{vv} = E \left[ \partial_z g_0(Z_{0,t-1})' \Omega^{-1} \partial_z g_0(Z_{0,t-1}) \right], \quad \text{and}$$

$$\Sigma_{uv} = E \left[ (\partial_\gamma g_0(Z_{0,t-1}), Z'_{2,t-1} \otimes I_p)' \Omega^{-1} \partial_z g_0(Z_{0,t-1}) \right].$$

A.4 then in particular implies that  $\Sigma_{uu}$  and  $\Sigma_{vv}$  are positive definite.

Note also that in the linear case where  $g(z; \gamma) = \alpha z$ ,  $\gamma = \text{vec}(\alpha)$ , then condition A.4 (as well as A.5) trivially holds. Also that as is easily checked, the boundedness assumptions in A.5 hold in particular for the initial STECM example in (3).

Note furthermore that A.5 is implied by the partial derivatives being of order  $\|z\|^s$  for some integer  $s > 1$  together with  $s$  order moments of  $\varepsilon_t$ , c.f. Theorem 1 (i).

**Theorem 3** *With  $u_t$  and  $v_t$  defined in (12), and with  $Z_{\kappa,t}$  defined in (9), then under Assumptions A.1-A.4 with  $r \in [0, 1]$ , the following joint weak convergence holds on  $D[0, 1]^{d+p+p^2k}$ ,*

$$\left( \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u'_t, \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} v'_t, \frac{1}{\sqrt{T}} Z'_{\kappa, [Tr]} \right) \xrightarrow{D} (B'_u(r), B'_v(r), B'_\kappa(r))$$

Here  $B \equiv (B'_u, B'_v, B'_\kappa)'$  is a  $(d + p^2k + p - 1)$ -dimensional Brownian motion with covariance matrix,

$$\Sigma = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{u\kappa} \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{v\kappa} \\ \Sigma_{\kappa u} & \Sigma_{\kappa v} & \Sigma_{\kappa\kappa} \end{bmatrix},$$

see A4. Furthermore,

$$\Sigma = \Sigma(0) + \sum_{h=1}^{\infty} (\Sigma(h) + \Sigma(h)'), \quad \Sigma(h) = \text{Cov}((u_t, v_t, \kappa_t), (u_{t+h}, v_{t+h}, \kappa_{t+h})), \quad (13)$$

where  $\kappa_t = \kappa'_0 C_0 s_t$  with  $s_t$  given in Theorem 1.

**Remark 4** In contrast to the standard FCLT for linear ECM models: The covariance matrices  $\Sigma_{uv}$  and  $\Sigma_{\kappa v}$  are not zero in general. This is a consequence of (i) the nonlinearities in  $g$  and (ii)  $\kappa_t = \kappa'_0 C_0 s_t$  not being a MGD. By definition of  $\kappa_t$  in Theorem 3 and  $v_t$  in (12), then  $\Sigma_{v\kappa} = 0$  is equivalent to

$$E(\partial_z g_0(Z_{0,t-1})') \alpha_{\perp} = 0.$$

To express the asymptotic distribution of the ML estimators, we define the following normalization matrix containing their convergence rates:

$$V_T = T \begin{pmatrix} I_{d+p^2k} & 0 & 0 \\ 0 & T I_{p-r-1} & 0 \\ 0 & 0 & T^2 \end{pmatrix}. \quad (14)$$

**Theorem 5** Assume that Assumptions A.1-A.5 hold. Then there exists a consistent estimator  $\hat{\theta} = (\hat{\eta}, \hat{b})$  of  $\theta = (\eta, b)$  as defined in (6)-(8). As  $T \rightarrow \infty$ ,  $\hat{\theta}$  satisfies:

$$V_T^{1/2} \begin{pmatrix} \hat{\eta} - \eta_0 \\ \text{vec}(\hat{b} - b_0) \end{pmatrix} \xrightarrow{D} H^{-1} S, \quad (15)$$

for a random matrix  $H$  and vector  $S$ .

With  $F(\cdot) = (B'_\kappa(\cdot), \cdot)' \in \mathbb{R}^{p-r}$ , and with  $B_\kappa$ ,  $B_v$  and  $B_u$  defined in Theorem 3, these are given by:

$$H \equiv \begin{pmatrix} \Sigma_{uu} & \int_0^1 F(r)' dr \otimes \Sigma_{uv} \\ \int_0^1 F(r) ds \otimes \Sigma_{vu} & \int_0^1 F(r) F(r)' dr \otimes \Sigma_{vv} \end{pmatrix}, \quad (16)$$

and

$$S \equiv \left( B_u(1), \text{vec} \left( \int_0^1 F dB'_v \right)' \right)'. \quad (17)$$

**Remark 6** (i) The ‘short-run’ parameters are  $\sqrt{T}$ -consistent, while the ‘long-run’ cointegration vectors are super- or  $T$ -consistent in all directions but one, which is  $T^{3/2}$ -consistent. This reflects the presence of the linear trend given in Theorem 1.

Corradi et al (2000, p.47) also notes the general presence of a linear trend induced by nonlinearities in Markov processes, and the thereby implied increased rate of convergence for their OLS based estimator of a single equation cointegration vector.

This differs from the results for the partial model in de Jong (2002), as the linear trend is assumed not to be there by assumption: There,  $X_t = (y_t, Z_t)'$  with  $y_t$  univariate while  $Z_t$  satisfies an invariance principle.

(ii) If  $\mu = Eg_0(Z_{0t}) = 0$ , then the linear trend vanishes, c.f. Theorem 1, and all directions of  $\beta$  will be super-consistent as in the linear cointegrated ECM. This will not necessarily hold in general though.

(iii) From the limit of the observed information as given by  $H$  in (16), it follows that the ‘long-run’ parameters and the ‘short-run’ parameters are not asymptotically orthogonal since  $\Sigma_{uv} \neq 0$  in general. In particular, we observe that  $\hat{\eta}$  is not asymptotically Gaussian, and  $\hat{b}$  not asymptotically mixed Gaussian, implying that usual  $\chi^2$  inference is not possible. This generalizes the findings for the single-equation analysis de Jong (2001, 2002).

The orthogonality condition in de Jong (2002) is equivalent to the sufficient condition here that  $\Sigma_{uv} = 0$ , see Assumption A.4 for the definition.

Note that even of  $\Sigma_{uv} = 0$  then  $\hat{b}$  will still not be asymptotically mixed Gaussian unless  $\Sigma_{kv} = 0$  as well. In this respect, one can observe that in the linear case where  $g(z; \gamma) = \alpha$ , and  $\eta = \text{vec}(\alpha)$ , then the long-run parameters in  $b$  and the short-run parameters in  $\eta$  are orthogonal as  $\Sigma_{uv} = 0$ . But also  $B_\kappa$  and  $B_v$  will be independent, and hence in this case  $\hat{b}$  has a mixed Gaussian distribution. In the STECM case with  $g(\cdot)$  in (3), then

$$E(\partial_z g_0(Z_{0,t-1})') \alpha_\perp = E(\psi(Z_{0,t-1}) + \partial_z \psi(Z_{0,t-1}) Z_{0,t-1}') \tilde{\alpha}' \alpha_\perp,$$

with  $\partial_z \psi(Z_{0,t-1})$  the derivative of  $\psi(\cdot)$  with respect to  $z$  evaluated in  $Z_{0,t-1}$ . Hence  $\Sigma_{v\kappa} = 0$  is implied by proportionality of the adjustments.

(iv) At the same time, de Jong (2002) finds that the parameters can be orthogonalized by using  $X_t = (Y_t, Z_t)'$ , see (i) above, corrected for their

empirical average, or demeaned, prior to the statistical calculations. Similarly here: if we replace the observations  $X_t$  by  $X_t$  corrected for empirical mean,  $F(\cdot)$  in Theorem 3, would be replaced by  $F(\cdot) - \int_0^1 F(s) ds$  with integral zero asymptotic normality of the short-run parameters. As already noted this does not imply mixed normality of the long-run parameters though. Also observe that one could also detrend as well as demean the observations prior to the analysis, in which case  $\hat{b}$  would be super consistent in all directions as in this case the linear trend would vanish.

- (v) A consistent estimator of the scale or information  $H$  in (16) is given by the observed information, that is the second order derivatives computed in the appendix. Also the covarianc matrix of  $(B_u, B_v, B_\kappa)$  can be estimated consistently based on the ‘HAC’ estimator, cf. de Jong (2002, Theorem 3).

## 5 A Simulation Study

We here investigate some finite-sample properties of the proposed estimator for the smooth transition error correction model (STECM) described in Section 2. For the implementation of the MLE, split the parameters  $\theta = (\theta_1, \theta_2)$  into  $\theta_1 := (\tilde{\alpha}, \alpha, \Phi) \in \mathbb{R}^{2r+pk \times p}$  and  $\theta_2 := (\beta, A, \omega)$ . We can then write the model on a more compact form,

$$\Delta X_t = \theta_1' W_{t-1}(\theta_2) + \varepsilon_t,$$

where

$$W_t(\theta_2) = (\psi(\beta' X_{t-1}; A, c) X_{t-1}' \beta, X_{t-1}' \beta, Z_{2,t-1}')' \in \mathbb{R}^{2r+pk}$$

such that the profile estimator of  $\theta_1$  and  $\Omega$  is given by standard OLS,

$$\hat{\theta}_1(\theta_2) = \left( \sum_{t=1}^T W_t(\theta_2) W_t(\theta_2) \right)^{-1} \left( \sum_{t=1}^T \Delta X_t W_t(\theta_2) \right),$$

$$\hat{\Omega}(\theta_2) = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t(\theta_2) \hat{\varepsilon}_t'(\theta_2), \quad \hat{\varepsilon}_t(\theta_2) = \Delta X_t - \hat{\theta}_1' W_{t-1}(\theta_2).$$

Given these profile estimators, we can in turn estimate  $\theta_2$  by

$$\hat{\theta}_2 = \arg \min_{\theta_2} \log(|\hat{\Omega}(\theta_2)|).$$

In the simulation study, we consider the simplest possible case of a bivariate system ( $p = 2$ ), with one cointegrating relation ( $r = 1$ ) and  $k = 1$  lagged difference entering. In this case  $A \in \mathbb{R}$ ,  $\omega \in \mathbb{R}$ , and  $\beta \in \mathbb{R}^2$ ; Furthermore, we choose the normalization  $\beta_1 = 1$ . We consider three different sample sizes,  $T = 250, 500$  and  $1000$ . For each sample size, we simulate 1000 sample paths for a set of given parameter values and then estimate these using the MLE for both the correctly specified non-linear model, and the incorrectly specified linear model (i.e. with  $\tilde{\alpha} = 0$  but with the inclusion of a constant term  $\mu$ , say). *Empirical* bias, standard deviation (std) and resulting root-mean-square error (RMSE) of the MLE's for the non-linear and linear model are presented in Table 1 and 2 respectively, see Appendix E. We do not report the results for  $\sqrt{\Phi}$  and  $\Omega$  here, and only note that these are estimated with very high precision of the same order as  $\beta$  when using the correctly specified STECM, while they are severely biased when using the misspecified linear ECM.

Regarding  $\beta$ , we observe that the MLE for the STECM is very precise in the sense that it has both a low empirical bias and std. This is consistent with our theoretical results which states that  $\hat{\beta}$  is super consistent. In comparison, the MLE of  $\beta$  based on the linear ECM suffers from additional biases and variances.<sup>1</sup> As also can be seen from Tables 1 and 2, in finite sample the misspecified MLE leads to less precise estimates. And so, there seems to be a considerable loss in using the linear MLE compared to using the correctly specified MLE.

The performance of the MLE's of the individual short-run parameters based on the STECM are highly imprecise with empirical bias and std of an order of magnitude of  $10^5$ . In fact, the MLE based on a linear ECM deliver more precise estimates of the individual parameters despite a high bias. However, when the set of short-run parameter estimates are combined to compute the resulting estimator of  $g(z; \gamma) = (g_1(z; \gamma), g_2(z; \gamma))$ , the MLE based on the STECM give good results as shown in Figure 1, 3, 5, 7, 9 and 11 in Appendix D: Here, we plot the empirical mean and the pointwise 95% empirical confidence intervals of  $g_1(z; \hat{\gamma})$  and  $g_2(z; \hat{\gamma})$  together with  $g_1(z; \gamma_0)$  and  $g_2(z; \gamma_0)$ . From these, we see that the estimates have small empirical biases and their empirical variances are of a much lower order than the individual parameter estimates. So the results reported in Table 1 for the short-run parameter estimates appear to be due to problems of identification of the individual parameters in  $g$ ; the likelihood has no problems identifying the

---

<sup>1</sup>This finding is consistent with the results of Corradi et al (2000) where it is demonstrated that OLS estimation of  $\beta$  based on the (incorrectly specified) linear ECM will remain consistent for our class of models.

function  $g$  itself. Seemingly, the STECM specification is not very attractive from this point of view. The nonlinear component of  $g$  is pinned down by the ‘extreme’ observations of  $\beta' X_{t-1}$  lying out in tails of  $g$ ; since we only have relatively few observations in these regions, the confidence bands tend to grow wider as we move away from the empirical mean of  $\beta' X_{t-1}$ . In particular, for smaller sample sizes (here,  $T = 250$ ), the empirical confidence bands are so wide that it appears likely that a linear specification would be accepted when tested against the correct non-linear specification.

As expected, the estimates of the transfer function  $g(z) = \alpha_1 z$  based on the MLE in the linear ECM are highly biased but exhibit small variance as can be seen from Figures 2, 4, 6, 8, 10 and 12. So while the MLE of  $\beta$  based on the linear ECM yield acceptable estimates, the ones of the short-term parameters are highly unsatisfactory and give very misleading pictures of the shape of the transfer functions.

## 6 Conclusions

The results here contain the estimation theory for the differentiable class of error-correction functions,  $g$ . It is also of interest to extend the results in this paper to the case of regime switching models. In the survey Lange and Rahbek (2007) make a distinction between ‘observation-switching’ (OS) and ‘Markov-switching’ (MS) error correction models. In MS models, such as in Krolzig et al (2002), the switching between regimes is determined by a latent Markov process. In OS models, such as in Bec and Rahbek (2004), the switching process is endogenously modelled with the probability of switching a function of the observed data, or cointegrating relations. The extension to both MS and OS error correction models is of much interest. Also by definition, here threshold error correction models as applied in for example Hansen and Seo (2002), are included if the threshold parameters are assumed known. Non-likelihood based extensions to the case of unknown threshold have been considered by Seo (2007), where a smooth estimating function (‘smooth least squares’) is used to circumvent the non-differentiability of the likelihood function.

The presented asymptotic theory implies as discussed, that even simple hypothesis testing in most cases can not be based on standard  $\chi^2$  distributions, see also de Jong (2002). On top of this is the problem that a key hypothesis of interest is the one of linearity which introduces the additional well-known identification problem as in Hansen and Seo (2002) testing for linearity in a threshold error correction model.

Our set-up assumes that the cointegrating relations are linear combina-

tions of the variables. Also this assumption can be challenged, see for example Bae and de Jong (2006) where non-linear cointegrating relations are studied. In fact, a rich literature exist on estimation and testing in the case of nonlinear cointegration, see e.g. Park and Phillips (2001), Chang and Park (2003), Choi and Saikkonen (2004), Saikkonen and Choi (2004) and Karlsen, Myklebust and Tjøstheim (2007).

## 7 Acknowledgements

Both authors are grateful for the continuing support from the Danish Social Sciences Research Council (DSSRC, Grant nr.2114-04-0001). In addition, D. Kristensen acknowledges financial support from CREATES, funded by the Danish National Research Foundation.

## References

- Anderson, H.M. (1997) Transaction Costs and Nonlinear Adjustment Towards Equilibrium in the US Treasury Bill Market. *Oxford Bulletin of Economics and Statistics* 59, 465-484.
- Bae, Y. and R.M. De Jong (2006) Money Demand Function Estimation by Nonlinear Cointegration. Forthcoming in *Journal of Applied Econometrics*.
- Bec, F. and A. Rahbek (2004) Vector Equilibrium Correction Models with Non-linear Discontinuous Adjustments. *Econometrics Journal* 7, 628-651.
- Boswijk, H.P. (2000) Mixed Normality and Ancillarity in I(2) Systems. *Econometric Theory* 16, 878-904.
- Corradi, V., N.R. Swanson and H. White (2000) Testing for Stationarity-Ergodicity and for Comovements Between Nonlinear Discrete Time Markov Processes. *Journal of Econometrics* 96, 39-73.
- Chang, Y. and J.Y. Park (2003) Index Models with Integrated Time Series. *Journal of Econometrics* 114, 73-106.
- Choi, I. and P. Saikkonen (2004) Testing Linearity in Cointegrating Smooth Transition Regressions. *Econometrics Journal* 7, 341-365.

- Dumas, B. (1992). Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World. *Review of Financial Studies* 5, 153–80.
- Escribano, A. (2004) Nonlinear Error Correction: The case of Money Demand in the UK (1878-2000). *Macroeconomic Dynamics* 8, 76–116.
- Feigin, P.D. and R.L. Tweedie (1985) Random Coefficient Autoregressive Processes: A Markov Chain Analysis of Stationarity and Finiteness of Moments. *Journal of Time Series Analysis* 6, 1-14.
- Granger, C. and T. Teräsvirta (1993) *Modelling Nonlinear Dynamic Relationships*. Oxford: Oxford University Press.
- Hansen, B.E. (1992) Convergence to Stochastic Integrals for Dependent Heterogeneous Processes. *Econometric Theory* 8, 489-500.
- Hansen, B.E. and B. Seo (2002) Testing for Two-regime Threshold Cointegration in Vector Error Correction Models. *Journal of Econometrics* 110, 293-318.
- Jeganathan, P. (1995) Some Aspects of Asymptotic Theory with Applications to Time Series Models. *Econometric Theory* 11, 818-887.
- Jensen, S.T. and A. Rahbek (2004) Asymptotic Inference for Nonstationary GARCH. *Econometric Theory* 20, 1203–1226.
- Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd Ed. Oxford: Oxford University Press.
- 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.
- Kapetanios, G., Y. Shin and A. Snell (2006) Testing for Cointegration in Nonlinear Smooth Transition Error Correction Models. *Econometric Theory* 22, 279-303.
- Karlsen, H.A., Myklebust, T. and D. Tjøstheim (2007) Nonparametric Estimation in a Nonlinear Cointegration Type Model. *Annals of Statistics* 35, 252-299.
- Krolzig, H., M. Marcellino and G. Mizon (2002) A Markov-Switching Vector Equilibrium Correction Model of the UK Labour Market. *Empirical Economics* 27, 233–254.



- Lange, T. and A. Rahbek (2007) An Introduction to Regime Switching Time Series Models. *Handbook of Financial Time Series*, forthcoming.
- 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.
- Park, Y. and P.C.B. Phillips (2001) Nonlinear Regressions with Integrated Time Series. *Econometrica* 69, 117-161.
- Pitarakis, J. Y. and Gonzalo, J., (2006) Threshold Effects in Multivariate Error Correction Models, *Palgrave Handbook of Econometrics*, Vol I, Ch.15.
- Pitarakis, J. Y. and Gonzalo, J. (2007) Threshold Effects in Cointegrating Regressions. *Oxford Bulletin of Economics and Statistics*, forthcoming.
- Psaradakis, Z., M. Sola and F. Spagnolo (2004) On Markov Error Correction Models with an Application to Stock Prices and Dividends. *Journal of Applied Econometrics* 19, 69–88.
- Saikkonen, P. (1995) Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems. *Econometric Theory* 11, 888-911.
- Saikkonen, P. (2005) Stability Results for Nonlinear Error Correction Models. *Journal of Econometrics* 127, 69-81.
- Saikkonen, P. (2006) Stability of Regime Switching Error Correction Models under Linear Cointegration. Forthcoming in *Econometric Theory*.
- Saikkonen, P. and I. Choi (2004) Cointegrating Smooth Transition Regressions. *Econometric Theory* 20, 301-340.
- Seo, B. (2003) Nonlinear Mean Reversion in the Term Structure of Interest Rates. *Journal of Economic Dynamics and Control* 27, 2243-2265
- 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. (2007) Estimation of Nonlinear Error Correction Models, working paper, London School of Economics, UK.

Sercu, P., R. Uppal and C. Van Hulle (1995). The Exchange Rate in the Presence of Transaction Costs: Implications for Tests of Purchasing Power Parity. *Journal of Finance* 50, 1309–19.

Van Dijk, D., T. Teräsvirta and P.H. Franses (2002) Smooth Transition Autoregressive Models – A Survey of Recent Developments. *Econometric Reviews* 21, 1–47.

## A Proofs

**Proof of Theorem 1:** The proof mimics the proof of Theorem 2 in Bec and Rahbek (2004), and is similar in structure to also Saikkonen (2005, proof of Theorem 2).

We set without loss of generality  $k = 1$  and  $\Phi = \Phi_1$  :

$$\begin{aligned}\Delta X_t &= g(\beta' X_{t-1}; \gamma) + \Phi \Delta X_{t-1} + \varepsilon_t \\ &= \alpha \beta' X_{t-1} + (g(\beta' X_{t-1}; \gamma) - \alpha \beta' X_{t-1}) + \Phi \Delta X_{t-1} + \varepsilon_t \\ &= \alpha \beta' X_{t-1} + \Phi \Delta X_{t-1} + \tau_t + \varepsilon_t,\end{aligned}$$

where  $\tau_t = g(\beta' X_{t-1}; \gamma) - \alpha \beta' X_{t-1}$ . With  $z_t = \beta' X_t$  and  $w_t = \beta'_\perp \Delta X_t$ , then

$$\mathbf{Y}_t = (z'_t, w'_t, z'_{t-1}, w'_{t-1})'$$

solves

$$\mathbf{Y}_t = \mathbf{A} \mathbf{Y}_{t-1} + \mathbf{B}(z_{t-1}) + \eta_t$$

where using the projection identity  $I = \bar{\beta}_\perp \beta'_\perp + \bar{\beta} \beta'$ ,

$$\mathbf{A} = \begin{pmatrix} I_r + \beta' \alpha + \beta' \Gamma \bar{\beta} & \beta' \Gamma \bar{\beta}_\perp & -\beta' \Gamma \bar{\beta} & 0 \\ \beta'_\perp \alpha + \beta'_\perp \Gamma \bar{\beta} & \beta'_\perp \Gamma \bar{\beta} & -\beta'_\perp \Gamma \bar{\beta} & 0 \\ I_r & 0 & 0 & 0 \\ 0 & I_{p-r} & 0 & 0 \end{pmatrix} \quad (18)$$

$$\mathbf{B}(z_{t-1}) = (\tau'_t \beta, \tau'_t \beta_\perp, 0, 0)' \quad \text{and} \quad \eta_t = (\varepsilon'_t \beta, \varepsilon'_t \beta_\perp, 0, 0)'$$

This is a time-homogenous Markov chain on  $\mathbb{R}^{2p}$ . By Assumption A.1, the 2 (in general  $k + 1$ ) step transition density for  $\mathbf{Y}_t$  is positive and bounded on compact sets in  $\mathbb{R}^{2p} \times \mathbb{R}^{2p}$ . This implies that the Markov chain is irreducible, aperiodic and compact sets are 'small', and the drift criterion as stated in Meyn and Tweedie (1993, Theorem 15.0.1(iii)) can be applied.

Consider a drift function proposed in Feigin and Tweedie (1985) which implies existence of second order moments of  $\mathbf{Y}_t$  – and hence of  $z_t$  and  $w_t$ :

$$V(y) = 1 + y' \mathbf{D} y \geq 1, \quad \mathbf{D} = \sum_{j=0}^{\infty} \mathbf{A}^{j'} \mathbf{A}^j$$

where  $\mathbf{A}$  is defined in (18). This choice of the drift function is well-defined as  $\rho(\mathbf{A} \otimes \mathbf{A}) < 1$ , where  $\rho(\cdot)$  is the spectral radius. That this holds, is implied by Assumption A.3 since  $\rho(\mathbf{A}) < 1$  is equivalent to the condition on the roots of the characteristic polynomial in A.3.

It follows that with  $y = (z'_1, w'_1, z'_2, w'_2)'$ ,

$$\begin{aligned}E(V(\mathbf{Y}_t) | \mathbf{Y}_{t-1} = y) &= \\ &1 + tr\{\mathbf{D}\Sigma\} + y'(\mathbf{D} - \mathbf{I})y + \mathbf{B}(z_1)' \mathbf{D} \mathbf{B}(z_1) + 2\mathbf{B}(z_1)' \mathbf{D} \mathbf{A} y\end{aligned} \quad (19)$$

Next,

$$E(V(\mathbf{Y}_t) | \mathbf{Y}_{t-1} = y) = V(y) \left( 1 - \left[ \frac{y'y - \text{tr}(\Sigma \mathbf{D}) - \mathbf{B}(z_1)' \mathbf{D} \mathbf{B}(z_1) - 2\mathbf{B}(z_1)' \mathbf{D} \mathbf{A} y}{V(y)} \right] \right)$$

Define for some  $\kappa > 1$  the compact set

$$K = \{ y \in \mathbb{R}^{2p} \mid y' \mathbf{D} y \leq \kappa \}$$

On complement of  $K$ ,  $K^c$ , it holds by definition that

$$V(y) = 1 + y' \mathbf{D} y \leq y' \mathbf{D} y \left( 1 + \frac{1}{\kappa} \right) \leq 2y' \mathbf{D} y$$

and therefore

$$\begin{aligned} \left[ \frac{y'y - \text{tr}(\Sigma \mathbf{D}) - \mathbf{B}(z_1)' \mathbf{D} \mathbf{B}(z_1) - 2\mathbf{B}(z_1)' \mathbf{D} \mathbf{A} y}{V(y)} \right] &\geq \frac{1}{2} \inf \left( \frac{y'y}{y' \mathbf{D} y} \right) - \left( \frac{\text{tr}(\Sigma \mathbf{D})}{V(y)} - \frac{\mathbf{B}(z_1)' \mathbf{D} \mathbf{B}(z_1) + 2\mathbf{B}(z_1)' \mathbf{D} \mathbf{A} y}{V(y)} \right) \\ &\geq \frac{1}{2\rho(\mathbf{D})} - \left( \frac{\text{tr}(\Sigma \mathbf{D})}{V(y)} - \frac{\mathbf{B}(z_1)' \mathbf{D} \mathbf{B}(z_1) + 2\mathbf{B}(z_1)' \mathbf{D} \mathbf{A} y}{V(y)} \right). \end{aligned}$$

First, note that

$$\frac{\text{tr}(\Sigma \mathbf{D})}{V(y)} \leq \frac{\text{tr}(\Sigma \mathbf{D})}{1 + \kappa} \rightarrow 0 \text{ as } \kappa \rightarrow \infty.$$

Next, by definition  $V(y) = O(\|y\|^2)$ . Next as  $\mathbf{B}(z_1) = o(\|z_1\|)$  also  $\mathbf{B}(z_1) = o(\|y\|)$  and therefore

$$\frac{\mathbf{B}(z_1)' \mathbf{D} \mathbf{B}(z_1) + 2\mathbf{B}(z_1)' \mathbf{D} \mathbf{A} y}{V(y)} \rightarrow 0 \text{ as } \|y\| \rightarrow \infty.$$

In other words, for  $\lambda_c$  large enough,

$$E(V(\mathbf{Y}_t) | \mathbf{Y}_{t-1} = y) \leq (1 - \gamma)V(y)$$

where  $\frac{1}{2\rho(\mathbf{D})} > \gamma > 0$ .

On  $K$ ,  $E(V(\mathbf{Y}_t) | \mathbf{Y}_{t-1} = y)$  is bounded by a continuous function and hence bounded on the compact set  $K$ . We conclude that  $\mathbf{Y}_t$  is geometrically ergodic with finite second order moment.

To address higher order moments the proof is similar. For the case of fourth order moments, define

$$\bar{\mathbf{Y}}_t = \mathbf{Y}_t \otimes \mathbf{Y}_t, \bar{\mathbf{A}} = \mathbf{A} \otimes \mathbf{A} \text{ and } \bar{\mathbf{B}}(z_1) = \mathbf{B}(z_1) \otimes \mathbf{B}(z_1).$$

As before note that,  $\rho(\bar{\mathbf{A}} \otimes \bar{\mathbf{A}}) < 1$  and therefore the  $p^2 \times p^2$  positive definite matrix  $\bar{\mathbf{D}}$  as well as the drift function  $\bar{V}$  are well-defined, where

$$\bar{\mathbf{D}} = \sum_{i=0}^{\infty} \bar{\mathbf{A}}^i \bar{\mathbf{A}}^i \text{ and } \bar{V}(y) = 1 + \bar{y}' \bar{\mathbf{D}} \bar{y} \text{ with } \bar{y} = (y \otimes y). \quad (20)$$

The rest is then as before, using inequalities in Feigin and Tweedie (1985).

To derive the representation in (5) rewrite the process as,

$$A(L)X_t = \tau_t + \varepsilon_t.$$

Note that by Assumption A.3, by Johansen (1996, Theorem 4.2) the following algebraic identity holds for  $\lambda \neq 1$ ,

$$A(\lambda)^{-1} = C \frac{1}{1 - \lambda} + C(\lambda), \quad (21)$$

where  $C = \beta_{\perp}(\alpha'_{\perp}(I_p - \sum_{i=1}^{k-1} \Gamma_i)\beta_{\perp})^{-1}\alpha'_{\perp}$  and  $C(\lambda) = \sum_{i=0}^{\infty} C_i \lambda^i$  with exponentially decreasing coefficients  $C_i$ . This gives,

$$X_t = C \sum_{i=1}^t (\tau_i + \varepsilon_i) + C(L) (\tau_t + \varepsilon_t) + D = C \sum_{i=1}^t (\tau_i + \varepsilon_i) + \nu_t + D, \quad (22)$$

where  $D$  depends on initial values  $X_0, \Delta X_0, \dots, \Delta X_{-k+1}$  and satisfies  $\beta' D = 0$ . Note that  $\tau_t + \varepsilon_t$  is a (measurable) function of  $\mathbf{Y}_t$ . This implies in particular that  $\tau_t + \varepsilon_t$  and hence  $\nu_t = C(L) (\tau_t + \varepsilon_t)$  are stationary, as  $C(\lambda)$  has exponentially decreasing coefficients. Next, by (22) and the definition of  $C$ , then with  $s_t = \varepsilon_t + g(\beta' X_{t-1}; \gamma) - \mu$ , where  $\mu = E[g(\beta' X_{t-1}; \gamma)]$ ,

$$C \sum_{i=1}^t (\tau_i + \varepsilon_i) = C \sum_{i=1}^t s_i + C \mu t.$$

□

**Proof of Theorem 3:** Note initially by Theorem 1,

$$\begin{aligned} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} u'_t, \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} v'_t, \frac{1}{\sqrt{T}} X'_{[Ts]\kappa} \right) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} (u'_t, v'_t, s'_t C' \kappa) + o_P(1) \\ &\equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} w'_t + o_P(1), \end{aligned}$$

where  $w_t = (u'_t, v'_t, s'_t C' \kappa)' \in \mathbb{R}^{d+p-1}$ . Next, recall from the proof of Theorem 1 that  $s_t = \varepsilon_t + g(\beta' X_{t-1}; \gamma) - \mu$ , where  $\mu = E[g(\beta' X_{t-1}; \gamma)]$ . Hence by definition, see also (12),  $w_t = f(\varepsilon_t, \mathbf{Y}_{t-1})$  with in an obvious notation,

$$f(e, y) = ((\partial_z g_0(y_0), y'_2 \otimes I_p)' \Omega^{-1} e, (e + g_0(y_0) - \mu)' \kappa)'$$

and  $(\varepsilon'_t, \mathbf{Y}'_{t-1})'$  is a geometrically ergodic Markov chain. Geometric ergodicity holds as in the proof of Theorem 1, using that the  $(2+k)$ -step transition density satisfies the regularity conditions there and replacing the drift function  $\bar{V}(y)$  in (20) by  $\tilde{V}(e, y) = \bar{e}'\bar{e} + \bar{V}(y)$ ,  $\bar{e} = (e \otimes e)$ .

With  $(\varepsilon'_t, \mathbf{Y}'_t)'$  geometrically ergodic, the FCLT in Meyn and Tweedie (1993, Theorem 17.4.2 and 17.4.4) can be applied to  $w_t = f(\varepsilon_t, \mathbf{Y}_{t-1})$ . This again holds under the assumptions that  $\|f\|^2 \leq \tilde{V}$  and the long run variance  $\Sigma$  defined in (13) is positive definite. It follows that

$$\begin{aligned} \|f\|^2 &\leq c (\|\partial_z g_0\|^2 \|e\|^2 + \|g_0\|^2 + \|e\|^2 + 1) \\ &\leq c (\|\partial_z g_0\|^4 + \|e\|^4 + \|g_0\|^2 + 1) \\ &\leq c (\|y_0\|^4 + \|e\|^4 + 1) \leq c\tilde{V} \end{aligned}$$

for some generic constant  $c > 0$ . The second last inequality holds by Assumptions A.2 and A.4.  $\square$

**Proof of Theorem 5.** To prove consistency, we verify the conditions of Lemma 10 with  $U_T = \frac{1}{T}V_T$  and  $Q_T(\theta) = \frac{1}{T}L_T(\theta)$ , where  $V_T$  is given in (14). We have that (C.1) holds by (A.1) while (C.2)-(C.3) follow by Lemmas 7, 8 and 9:

$$\begin{aligned} dQ_T(\theta_0; U_T^{-1/2}d\theta) &= dL_T(\theta_0; V_T^{-1/2}d\theta)/\sqrt{T} = o_P(1), \\ d^2Q_T(\theta_0; U_T^{-1/2}d\theta, U_T^{-1/2}d\bar{\theta}) &= d^2L_T(\theta_0; V_T^{-1/2}d\theta, V_T^{-1/2}d\bar{\theta}) \xrightarrow{D} H_\infty(d\theta, d\bar{\theta}), \\ d^3Q_T(\theta; U_T^{-1/2}d\theta, U_T^{-1/2}d\bar{\theta}, U_T^{-1/2}d\tilde{\theta}) &= T^{1/2}d^3L_T(\theta; V_T^{-1/2}d\theta, V_T^{-1/2}d\bar{\theta}, V_T^{-1/2}d\tilde{\theta}) \\ &= O_P(\|d\theta\| \|d\bar{\theta}\| \|d\tilde{\theta}\|). \end{aligned}$$

The asymptotic distribution will follow from Lemma 11 with  $v_T = T$ , such that  $v_T U_T = V_T$ , if we can verify the additional condition (C.4) in Appendix C. But this follows from Lemma 7 since

$$dQ_T(\theta_0; v_T^{-1/2}U_T^{-1/2}d\theta) = dL_T(\theta_0; V_T^{-1/2}d\theta) \xrightarrow{D} S_\infty(d\theta).$$

We conclude that  $V_T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{D} \theta_\infty$ , where  $\theta_\infty$  satisfies  $S_\infty(d\theta) = H_\infty(d\theta, \theta_\infty)$  for all directions  $d\theta$  with  $S_\infty(d\theta)$  and  $H_\infty(d\theta, \theta_\infty)$  given in Lemma 7 and 8 respectively. This implies the result stated in Theorem 5.  $\square$

## B Lemmas

Recall the definitions of  $\partial_z g_0(z)$  and  $\partial_\gamma g_0(z)$  in (11). We then introduce the first order differentials in the directions of  $z$  and  $\gamma$  as given by  $dg(z; \gamma; dz) = \partial_z g(z; \gamma) dz$  and  $dg(z; \gamma; d\gamma) = \partial_\gamma g(z; \gamma) d\gamma$ . The second and third order differentials are written as  $d^2 g(z, \gamma; da, db)$  and  $d^3 g(z, \gamma; da, db, dc)$ , where  $da, db, dc \in \{dz, d\gamma\}$ . When evaluated at the true value  $\gamma_0$  we write  $d^2 g_0(z; \cdot)$  and  $d^3 g_0(z; \cdot)$ . Note that in the case  $r = 1$ , then for example  $d^2 g_0(z; dz_1, dz_2)$  reduces to  $\partial_z^2 g_0(z) dz_1 dz_2$ , where  $\partial_z^2 g_0(z) = \partial^2 g(z) / \partial z^2$  is the second order derivative.

**Lemma 7** *Assume that A.1-A.4 hold. Then for the -2log-likelihood function  $L_T(\theta, \Omega)$  defined in (10) and with  $d\theta = (d\eta, db)$ ,  $db = (db'_\kappa, db'_\delta)'$  the following hold:*

$$dL_T(\theta_0; V_T^{-1/2} d\theta) = S_T(\theta_0; V_T^{-1/2} d\theta) \xrightarrow{D} S_\infty(d\theta),$$

where  $V_T$  is defined in (14) and

$$S_\infty(d\theta) = - \left[ \int_0^1 dB'_u(s) ds d\eta + \text{tr}(db' \int_0^1 F dB'_v) \right],$$

with  $B = (B'_u, B'_v, B'_\kappa)'$  defined in Theorem 3 and  $F(s) = (B'_\kappa(s), s)'$ .

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

$$S_T(\theta; V_T^{-1/2} d\theta) = T^{-1/2} S_{\eta,T}(\theta; d\eta) + T^{-1} S_{\kappa,T}(\theta; db_\kappa) + T^{-3/2} S_{\delta,T}(\theta; db_\delta)$$

where

$$\begin{aligned} S_{\eta,T}(\theta; d\gamma) &= - \sum_{t=1}^T [\partial_\gamma g(Z_{0,t-1} + b' Z_{1,t-1}) d\gamma + (Z'_{2,t-1} \otimes I_p) d\phi]' \Omega^{-1} \varepsilon_t(\theta), \\ S_{\kappa,T}(\theta; db_\kappa) &= - \sum_{t=1}^T Z'_{\kappa,t-1} db_\kappa \partial_z g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \varepsilon_t(\theta) \\ S_{\delta,T}(\theta; db_\delta) &= - \sum_{t=1}^T Z_{\delta,t-1} db_\delta \partial_z g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \varepsilon_t(\theta) \end{aligned}$$

Evaluated at the true parameter value  $\theta_0 = (\eta_0, 0)$ , we get

$$\begin{aligned} S_{\eta,T}(\theta_0; d\eta) &= -d\eta' \sum_{t=1}^T u_t, & S_{\kappa,T}(\theta_0; db_\kappa) &= - \sum_{t=1}^T Z'_{\kappa,t-1} db_\kappa v_t, \\ S_{\delta,T}(\theta_0; db_\delta) &= - \sum_{t=1}^T Z_{\delta,t-1} db_\delta v_t, \end{aligned}$$

where  $u_t \in \mathbb{R}^{d+p^2k}$  and  $v_t \in \mathbb{R}^r$  are defined in Theorem 3. From this theorem,

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

By Hansen (1992, Theorem 2.1), as joint convergence holds by Theorem 3,

$$T^{-1}S_{\kappa,T}(\theta_0; db_\kappa) = -T^{-1} \sum_{t=1}^T (Z'_{\kappa,t-1} db_\kappa) v_t \xrightarrow{D} - \int_0^1 (B'_\kappa(s) db_\kappa) dB_v(s).$$

Finally, by yet another application of Theorem 3,

$$T^{-3/2}S_{\delta,T}(\theta_0; db_\delta) = -T^{-2} \sum_{t=1}^T (Z'_{\delta,t-1} db_\delta) v_t \xrightarrow{D} - (db_\delta) \int_0^1 s dB_v(s).$$

The three convergence results above hold simultaneously since the convergence in Theorem 3 does. By collecting terms, the desired result is obtained.  $\square$

**Lemma 8** *Assume that A.1-A.4 hold. With  $d\theta = (d\eta, db)$ ,  $db = (db'_\kappa, db'_\delta)'$ , then for the  $-2\log$ -likelihood function  $L_T(\theta)$  defined in (10) satisfies*

$$d^2 L_T(\theta_0; V_T^{-1/2} d\theta, V_T^{-1/2} d\tilde{\theta}) = H_T(\theta_0; V_T^{-1/2} d\theta, V_T^{-1/2} d\tilde{\theta}) \xrightarrow{w} H_\infty(d\theta, d\tilde{\theta}),$$

where  $H_\infty(d\theta, d\tilde{\theta}) > 0$  a.s. is given by

$$\begin{aligned} H_\infty(d\theta, d\tilde{\theta}) &= d\eta' \Sigma_{uu} d\tilde{\eta} + \text{tr} \left\{ d\tilde{b}' \int_0^1 F(s) F'(s) ds db \Sigma_{vv} \right\} \\ &\quad + \int_0^1 F(s)' ds db \Sigma_{vv} d\tilde{\gamma} + d\gamma' \Sigma_{uv} db' \int_0^1 F(s) ds \end{aligned}$$

with  $\Sigma$  and  $B = (B'_u, B'_v, B'_\kappa)'$  defined in Theorem 3 and  $F(s) = (B'_\kappa(s), s)'$ .

**Proof.** The second order differential is given by,

$$\begin{aligned} H_T(\theta; V_T^{-1/2} d\theta, V_T^{-1/2} d\tilde{\theta}) &= T^{-1} H_{\eta,\eta} + T^{-2} H_{\kappa,\kappa} + T^{-3} H_{\delta,\delta} + T^{-2} (H_{\eta,\delta} + H_{\delta,\eta}) \\ &\quad (23) \\ &\quad + T^{-3/2} (H_{\eta,\kappa} + H_{\kappa,\eta}) + T^{-5/2} (H_{\delta,\kappa} + H_{\kappa,\delta}) \end{aligned}$$



where we have used the notation that  $H_{\eta,\delta} = d^2 L_T(\theta; d\eta, d\tilde{b}_\delta)$  and so forth. If we can prove that the following six claims hold simultaneously, the proof is complete:

$$\text{Claim 1 : } T^{-1} H_{\eta,\eta} \xrightarrow{P} d\eta' \Sigma_{uu} d\eta,$$

$$\text{Claim 2 : } T^{-2} H_{\kappa,\kappa} \xrightarrow{D} \text{tr}\{(db_\kappa)' \int_0^1 B_\kappa(s) B_\kappa(s)' ds (d\tilde{b}_\kappa) \Sigma_{vv}\},$$

$$\text{Claim 3 : } T^{-3} H_{\delta,\delta} \xrightarrow{D} \int_0^1 s^2 ds (db_\delta) \Sigma_{vv} (d\tilde{b}'_\delta)$$

$$\text{Claim 4 : } T^{-2} H_{\delta,\eta} \xrightarrow{D} \left( \int_0^1 s ds \right) db_\delta \Sigma_{vu} d\tilde{\gamma},$$

$$\text{Claim 5 : } T^{-3/2} H_{\kappa,\eta} \xrightarrow{D} \int_0^1 B_\kappa(s)' ds db_\kappa \Sigma_{vu} d\tilde{\gamma}.$$

$$\text{Claim 6 : } T^{-5/2} H_{\delta,\eta} \xrightarrow{D} \text{tr}\{d\tilde{b}_\kappa \int_0^1 B_\kappa(s) s ds db_\delta \Sigma_{vv}\}.$$

**Proof of Claim 1:** We have

$$\begin{aligned} H_{\eta,\eta} &= d\gamma' \sum_{t=1}^T \partial_\gamma g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \partial_\gamma g(Z_{0,t-1} + b' Z_{1,t-1}) d\tilde{\gamma} \\ &\quad - \sum_{t=1}^T d^2 g(Z_{0,t-1} + b' Z_{1,t-1}; d\gamma, d\tilde{\gamma})' \Omega^{-1} \varepsilon_t(\theta) \\ &\quad + d\phi' \sum_{t=1}^T (Z_{2,t-1} \otimes I_p)' \Omega^{-1} (Z_{2,t-1} \otimes I_p) d\tilde{\phi} \\ &\quad + d\phi' \sum_{t=1}^T (Z_{2,t-1} \otimes I_p)' \Omega^{-1} \partial_\gamma g(Z_{0,t-1} + b' Z_{1,t-1}) d\tilde{\gamma} \\ &\quad + d\gamma' \sum_{t=1}^T \partial_\gamma g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} (Z_{2,t-1} \otimes I_p) d\tilde{\phi} \end{aligned}$$

Evaluated at  $\theta = \theta_0 = (\gamma_0, 0)$ , the result holds by the law of large numbers for geometrically ergodic processes,  $H_{\eta,\eta} \xrightarrow{P} d\eta' \Sigma_{uu} d\tilde{\eta}$ .

**Proof of Claim 2:** The differential  $H_{\kappa,\kappa}$  takes the form

$$\begin{aligned} T^{-2}H_{\kappa,\kappa} &= T^{-2}tr\left\{d\tilde{b}'_{\kappa}\sum_{t=1}^TZ_{\kappa,t-1}Z'_{\kappa,t-1}db_{\kappa}\partial_zg(Z_{0,t-1})'\Omega^{-1}\partial_zg(Z_{0,t-1})\right\} \quad (24) \\ &\quad - T^{-2}\sum_{t=1}^T\varepsilon_t'\Omega^{-1}d^2g(Z_{0,t-1};db'_{\kappa}Z_{\kappa,t-1},d\tilde{b}'_{\kappa}Z_{\kappa,t-1}) \end{aligned}$$

Rewrite the first term on the right hand side as

$$T^{-2}d\tilde{b}'_{\kappa}\sum_{t=1}^TZ_{\kappa,t-1}Z'_{\kappa,t-1}d\tilde{b}_{\kappa}\Sigma_{vv} + T^{-2}d\tilde{b}'_{\kappa}\sum_{t=1}^TZ_{\kappa,t-1}Z'_{\kappa,t-1}d\tilde{b}_{\kappa}w_{t-1} \quad (25)$$

where,

$$\begin{aligned} w_{t-1} &= \partial_zg(Z_{0,t-1})'\Omega^{-1}\partial_zg(Z_{0,t-1}) - \Sigma_{vv}, \\ \Sigma_{vv} &= E\left[\partial_zg(Z_{0,t-1})'\Omega^{-1}\partial_zg(Z_{0,t-1})\right]. \end{aligned}$$

Note that by Theorem 3 and the CMT, that the first term in (25) converges weakly to the trace of  $(db_{\kappa})'\int_0^1B_{\kappa}(s)B_{\kappa}(s)'ds(d\tilde{b}_{\kappa})\Sigma_{vv}$ . Next,  $U_{n[n]}\xrightarrow{w}U$ , where  $U(s) = B_{\kappa}(s)B_{\kappa}(s)'$  and with  $n = T$ ,  $v_{b,t} \equiv e_t$ ,  $U_{n,t} \equiv T^{-1}Z_{\kappa,t-1}Z'_{\kappa,t-1}$  and  $\mathcal{F}_t = \sigma(Z_{\kappa,t}, Z_{\kappa,t-1}, \dots, Z_{\kappa,0})$ , Hansen (1992, Theorem 3.3) gives directly that the second term in (25) tends to zero, provided

$$\sup_t E|E(w_{t-1}|\mathcal{F}_{t-m})| \rightarrow 0, \quad m \rightarrow \infty. \quad (26)$$

This holds by Lemma 12. Next turn to the second term in (24) which can be rewritten as

$$\begin{aligned} &- T^{-2}\sum_{t=1}^T\varepsilon_t'\Omega^{-1}(I_p \otimes Z'_{\kappa,t-1}db_{\kappa})(D_z^2g_0(Z_{0,t-1}))d\tilde{b}'_{\kappa}Z_{\kappa,t-1} \\ &= -T^{-2}\sum_{t=1}^Ttr\{vec(I_p \otimes Z'_{\kappa,t-1}db_{\kappa})vec(Z'_{\kappa,t-1}d\tilde{b}_{\kappa})'(D_z^2g_0(Z_{0,t-1})' \otimes \varepsilon_t'\Omega^{-1})\}, \end{aligned}$$

with  $D_z^2g_0(z) = \left(\frac{\partial^2g_i(z;\gamma_0)}{\partial z_j\partial z_k}\right)_{i=1,\dots,p,j,k=1,\dots,r}$ , see Magnus and Neudecker (1988, p.108). As before, the result now holds by Hansen (1992, Theorem 3.3) as the process  $D_z^2g_0(Z_{0,t-1})' \otimes \varepsilon_t'\Omega^{-1}$  is a Martingale difference sequence. We conclude that

$$T^{-2}H_{\kappa,\kappa} \xrightarrow{D} tr\left\{(db_{\kappa})'\int_0^1B_{\kappa}(s)B_{\kappa}(s)'ds(d\tilde{b}_{\kappa})\Sigma_{vv}\right\}. \quad (27)$$

**Proof of Claim 3:** The differential  $H_{\delta,\delta}$  takes the form:

$$\begin{aligned} T^{-3}H_{\delta,\delta} &= T^{-3}db_\delta \sum_{t=1}^T Z_{\delta,t-1}^2 \partial_z g(Z_{0,t-1})' \Omega^{-1} \partial_z g(Z_{0,t-1}) d\tilde{b}'_\delta \quad (28) \\ &\quad - T^{-3} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} d^2 g(Z_{0,t-1}; db'_\delta Z_{\delta,t-1}, d\tilde{b}'_\delta Z_{\delta,t-1}) \end{aligned}$$

Rewrite the first term on the right hand side as

$$T^{-3}db_\delta \sum_{t=1}^T Z_{\delta,t-1}^2 \Sigma_{vv} d\tilde{b}'_\delta + T^{-3}db_\delta \sum_{t=1}^T Z_{\delta,t-1}^2 w_{t-1} d\tilde{b}'_\delta \xrightarrow{D} \int_0^1 s^2 ds (db_\delta) \Sigma_{vv} (d\tilde{b}'_\delta).$$

To see that the limit is given as above, observe that by Theorem 3 and the CMT, the first term converges weakly to  $\int_0^1 s^2 ds (db_\delta) \Sigma_{vv} (d\tilde{b}'_\delta)$ . With the same notation as before, the process  $U_{n,t} \equiv T^{-2}Z_{\delta,t-1}^2$  satisfies  $U_{n[n]} \xrightarrow{D} U$ , where  $U(s) = s^2$ . Hansen (1992, Theorem 3.3) then yields that the second term tends to zero. Next turn to the second term in (28) which can be rewritten as

$$\begin{aligned} &- T^{-3} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} (I_p \otimes Z'_{\delta,t-1} db_\delta) (D^2 g(Z_{0,t-1})) d\tilde{b}'_\delta Z_{\delta,t-1} \\ &= -T^{-3} \sum_{t=1}^T \text{tr} \{ \text{vec} (I_p \otimes Z'_{\delta,t-1} db_\delta) \text{vec} (Z'_{\delta,t-1} d\tilde{b}'_\delta)' (D^2 g(Z_{0,t-1})' \otimes \varepsilon'_t \Omega^{-1}) \}, \end{aligned}$$

and by the same arguments as before, we conclude that this is  $o_P(1)$ .

**Proof of Claim 4:** The differential  $H_{\delta,\eta}$  takes the form:

$$\begin{aligned} T^{-2}H_{\delta,\eta}^0 &\quad (29) \\ &= T^{-2}db_\delta \sum_{t=1}^T Z_{\delta,t-1} \partial_z g(Z_{0,t-1})' \Omega^{-1} \left[ \partial_\gamma g(Z_{0,t-1}) d\tilde{\gamma} + (Z'_{\delta,t-1} \otimes I_p) d\tilde{\phi} \right] \end{aligned}$$

$$\begin{aligned}
& + T^{-2} \sum_{t=1}^T [\partial_\gamma g(Z_{0,t-1}) d\gamma + (Z'_{2,t-1} \otimes I_p) d\phi]' \Omega^{-1} \partial_z g(Z_{0,t-1}) Z'_{\delta,t-1} d\tilde{b}'_\delta \\
& - T^{-2} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} d^2 g(Z_{0,t-1}; db'_\delta Z_{\delta,t-1}, d\tilde{\gamma}) \\
& - T^{-2} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} d^2 g(Z_{0,t-1}; d\tilde{b}'_\delta Z_{\delta,t-1}, d\gamma)
\end{aligned}$$

We need here the second order differential in terms of  $w = (z', \gamma')' \in \mathbb{R}^{r+d}$

$$D_{z,\gamma}^2 g_0(z) = \left( \frac{\partial^2 g_i(w)}{\partial w_j \partial w_k} \right)_{i=1,\dots,p,j,k=1,\dots,r+d} \Big|_{w=(z',\gamma_0)'}$$

Using this with  $0_d = (0, 0, \dots, 0)$  of dimension  $d$ , we can write the second term in (29) as

$$T^{-2} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} (I_p \otimes (Z'_{\delta,t-1} db_\delta, 0'_d)) (D_{z,\gamma}^2 g_0(Z_{0,t-1})) (0'_r, d\tilde{\gamma}')' = o_P(1)$$

by similar arguments as before. The first term can be written as

$$\begin{aligned}
& T^{-2} db_\delta \sum_{t=1}^T Z_{\delta,t-1} \partial_z g(Z_{0,t-1})' \Omega^{-1} \left[ \partial_\gamma g(Z_{0,t-1}) d\tilde{\gamma} + (Z'_{2,t-1} \otimes I_p)' d\tilde{\phi} \right] \\
& = T^{-2} db_\delta \sum_{t=1}^T Z_{\delta,t-1} \Sigma_{vu} d\tilde{\eta} + T^{-2} db_\delta \sum_{t=1}^T Z_{\delta,t-1} w_{t-1} d\tilde{\eta},
\end{aligned}$$

where

$$w_{t-1} = \partial_z g(Z_{0,t-1})' \Omega^{-1} \left[ \partial_\gamma g(Z_{0,t-1}), (Z'_{2,t-1} \otimes I_p) \right] - \Sigma_{vu}$$

is a stationary mean-zero sequence. By Theorem 3,

$$T^{-2} db_\delta \sum_{t=1}^T Z_{\delta,t-1} \Sigma_{vu} d\tilde{\eta} \xrightarrow{D} \left( \int_0^1 s ds \right) db_\delta \Sigma_{vu} d\tilde{\eta},$$

while  $T^{-2} db_\delta \sum_{t=1}^T Z_{\delta,t-1} w_{t-1} = o_P(1)$  by the same arguments used before.

**Proof of Claim 5:** The differential  $H_{\kappa,\eta}$  takes the form:

$$\begin{aligned} & T^{-3/2} H_{\kappa,\eta} \\ &= T^{-3/2} \sum_{t=1}^T Z'_{\kappa,t-1} db_{\kappa} \partial_z g (Z_{0,t-1})' \Omega^{-1} \left[ \partial_{\gamma} g (Z_{0,t-1}) d\tilde{\gamma} + (Z'_{2,t-1} \otimes I_p) d\tilde{\phi} \right] \\ &- T^{-3/2} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} d^2 g (Z_{0,t-1}; db'_{\kappa} Z_{\kappa,t-1}, d\tilde{\gamma}). \end{aligned}$$

where the second term is

$$T^{-3/2} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} (I_p \otimes (Z'_{\kappa,t-1} db_{\kappa}, 0'_d)) (D^2 g (Z_{0,t-1})) (0'_r, d\tilde{\gamma})' = o_P(1)$$

by similar arguments as before. The first term can be written as

$$\begin{aligned} & T^{-3/2} \sum_{t=1}^T Z'_{\kappa,t-1} db_{\kappa} \partial_z g (Z_{0,t-1})' \Omega^{-1} \left[ \partial_{\gamma} g (Z_{0,t-1}) d\tilde{\gamma} + (Z_{2,t-1} \otimes I_p)' d\tilde{\phi} \right] \\ &= T^{-3/2} \sum_{t=1}^T Z'_{\kappa,t-1} db_{\kappa} \Sigma_{vu} d\tilde{\eta} + T^{-3/2} \sum_{t=1}^T Z'_{\kappa,t-1} db_{\kappa} w_{t-1} d\tilde{\eta}, \end{aligned}$$

where  $T^{-3/2} \sum_{t=1}^T Z'_{\kappa,t-1} db_{\kappa} w_{t-1} = o_P(1)$  while

$$T^{-3/2} \sum_{t=1}^T Z'_{\kappa,t-1} db_{\kappa} \Sigma_{vu} d\tilde{\eta} \xrightarrow{D} \int_0^1 B_{\kappa}(s)' ds db_{\kappa} \Sigma_{vu} d\tilde{\eta}.$$

**Proof of Claim 6:** The differential  $H_{\kappa,\delta}$  takes the form:

$$\begin{aligned} T^{-5/2} H_{\kappa,\delta} &= T^{-5/2} db_{\delta} \sum_{t=1}^T Z_{\delta,t-1} \partial_z g (Z_{0,t-1})' \Omega^{-1} \partial_z g (Z_{0,t-1}) d\tilde{b}'_{\kappa} Z_{\kappa,t-1} \\ &- T^{-5/2} \sum_{t=1}^T \varepsilon'_t \Omega^{-1} d^2 g (Z_{0,t-1}; db'_{\delta} Z_{\delta,t-1}, d\tilde{b}'_{\kappa} Z_{\kappa,t-1}). \end{aligned}$$

Write

$$\begin{aligned} & T^{-5/2} db_{\delta} \sum_{t=1}^T Z_{\delta,t-1} \partial_z g (Z_{0,t-1})' \Omega^{-1} \partial_z g (Z_{0,t-1}) d\tilde{b}'_{\kappa} Z_{\kappa,t-1} \\ &= T^{-5/2} \text{tr} \left\{ d\tilde{b}'_{\kappa} \sum_{t=1}^T Z_{\kappa,t-1} Z_{\delta,t-1} db_{\delta} \Sigma_{vv} \right\} + T^{-5/2} \text{tr} \left\{ d\tilde{b}'_{\kappa} \sum_{t=1}^T Z_{\kappa,t-1} Z_{\delta,t-1} db_{\delta} w_{t-1} \right\}, \end{aligned}$$

where  $w_{t-1}$  has been redefined as

$$w_{t-1} = \partial_z g(Z_{0,t-1})' \Omega^{-1} \partial_z g(Z_{0,t-1}) - \Sigma_{vv}.$$

The proof of the claim now follows along the same lines as before.  $\square$

**Lemma 9** *Under (A.1)-(A.5),*

$$\sup_{\theta \in \mathcal{N}_T(\theta_0)} \left| T^{1/2} d^3 L_T(\theta, V_T^{-1/2} d\theta, V_T^{-1/2} d\bar{\theta}, V_T^{-1/2} d\tilde{\theta}) \right| = O_P(\|d\theta\| \|d\bar{\theta}\| \|d\tilde{\theta}\|)$$

for a sequence of neighborhoods

$$\mathcal{N}_T(\theta_0) = \left\{ \theta : \|T^{-1/2} V_T^{1/2} (\theta - \theta_0)\| < \epsilon \right\}$$

of  $\theta_0$  and with  $V_T$  defined in (14).

**Proof.** Write the third order differential as,

$$d^3 L_T(\theta, d\theta, d\bar{\theta}, d\tilde{\theta}) = \sum_{i,j,k} G_{\theta_i, \theta_j, \theta_k} d\theta_i d\bar{\theta}_j d\tilde{\theta}_k.$$

Below we consider each of the terms  $G_{\theta_i, \theta_j, \theta_k}$  normalized as indicated in  $\sqrt{T} d^3 L_T(\theta, V_T^{-1/2} d\theta, V_T^{-1/2} d\bar{\theta}, V_T^{-1/2} d\tilde{\theta})$  and argue that they vanish as  $T \rightarrow \infty$  as desired. As the arguments, apart from normalization, are identical for the individual derivatives, we state explicitly  $G_{\gamma_i, \gamma_j, \gamma_k}$  and give the argument in detail for this derivative. The orders of magnitude and expressions for the remaining derivatives are listed below. We here leave out derivatives w.r.t.  $\phi$  since they are similar to the ones of  $\gamma$ .

**Claim 1:**  $T^{-1} G_{\gamma_i, \gamma_j, \gamma_k} = O_P(1)$ .

The derivative is given by

$$\begin{aligned} G_{\gamma_i, \gamma_j, \gamma_k} &= \sum_{t=1}^T \partial_{\gamma_i, \gamma_k}^2 g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \partial_{\gamma_j} g(Z_{0,t-1} + b' Z_{1,t-1}) + \quad (30) \\ &\quad \sum_{t=1}^T \partial_{\gamma_i} g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \partial_{\gamma_j, \gamma_k}^2 g(Z_{0,t-1} + b' Z_{1,t-1}) + \\ &\quad \sum_{t=1}^T \partial_{\gamma_k} g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \partial_{\gamma_i, \gamma_j}^2 g(Z_{0,t-1} + b' Z_{1,t-1}) - \\ &\quad \sum_{t=1}^T \varepsilon_t' \Omega^{-1} \partial_{\gamma_i, \gamma_j, \gamma_k}^3 g(Z_{0,t-1} + b' Z_{1,t-1}) \end{aligned}$$

Next, note that with  $\theta \in \mathcal{N}_T(\theta_0)$ , we can write,

$$b' = (b'_\kappa, b'_\delta) = \left( \frac{1}{\sqrt{T}} h'_{\kappa,T}, \frac{1}{T} h_{\delta,T} \right) \quad (31)$$

where  $\|h_{\kappa,T}\| < \epsilon$  and  $\|h_{\delta,T}\| < \epsilon$ . This way for example,

$$b' Z_{1,t} = \frac{1}{\sqrt{T}} h'_{\kappa,T} Z_{\kappa,t} + \frac{1}{T} h_{\delta,T} Z_{\delta,t}.$$

Hence considering the first term in (30) and using A.5, then with  $c$  denoting a generic constant:

$$\begin{aligned} & \left\| \frac{1}{T} \sum_{t=1}^T \partial_{\gamma_i, \gamma_k}^2 g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \partial_{\gamma_j} g(Z_{0,t-1} + b' Z_{1,t-1}) \right\| \\ & \leq c \frac{1}{T} \sum_{t=1}^T \left( \|Z_{0,t-1}\|^2 + \frac{1}{T} \|Z_{\kappa,t-1}\|^2 + \frac{1}{T^2} \|Z_{\delta,t-1}\|^2 \right) \end{aligned} \quad (32)$$

The first term in (32) is bounded by stationarity of  $Z_{0,t}$  and  $E \|Z_{0,t-1}\|^2 < \infty$  by Theorem 1 (ii). The second term is bounded using that by Theorem 2 the FCLT applies to  $Z_{\kappa,t}$ , and the continuous mapping theorem applied to the mapping  $x \rightarrow \int_0^1 \|x\|^2 du$ , where  $x \in D[0, 1]$ . For the third term note simply that  $\frac{1}{T^2} \|Z_{\delta,t-1}\|^2 = O(1)$ .

The arguments are identical for the other terms in (30) apart from the last term, which is given by:

$$\begin{aligned} & \sum_{t=1}^T \varepsilon'_t \Omega^{-1} \partial_{\gamma_i \gamma_j \gamma_k}^3 g(Z_{0,t-1} + b' Z_{1,t-1}) \\ & = \sum_{t=1}^T (\Delta X_t - g(Z_{0,t-1} + b' Z_{1,t-1}))' \Omega^{-1} \partial_{\gamma_i \gamma_j \gamma_k}^3 g(Z_{0,t-1} + b' Z_{1,t-1}) \end{aligned}$$

Using again A.5, together with the inequality,  $\|x\| \|y\| \leq \|x\|^2 + \|y\|^2$ , we make the following evaluations:

$$\begin{aligned} & \left\| \frac{1}{T} \sum_{t=1}^T (\Delta X_t - g(Z_{0,t-1} + b' Z_{1,t-1}))' \Omega^{-1} \partial_{\gamma_i \gamma_j \gamma_k}^3 g(Z_{0,t-1} + b' Z_{1,t-1}) \right\| \\ & \leq \frac{c}{T} \sum_{t=1}^T \|\Delta X_t\|^2 + \frac{c}{T^2} \sum_{t=1}^T \|Z_{\kappa,t-1}\|^2 + \frac{c}{T^3} \sum_{t=1}^T \|Z_{\delta,t-1}\|^2 + \frac{c}{T} \sum_{t=1}^T \|Z_{0,t-1}\|^2 \end{aligned}$$

That the last expression is  $O_P(1)$  follows by the arguments above.

**Claim 2:**  $T^{-3/2}G_{\gamma_i, \gamma_j, b_{\kappa, k}} = O_P(1)$  and  $T^{-2}G_{\gamma_i, \gamma_j, b_\delta} = O_P(1)$ .

To see this note initially that with  $b_k$  equal to either  $b_{\kappa, k}$  or  $b_\delta$ ,

$$\begin{aligned} G_{\gamma_i, \gamma_j, b_k} &= \sum_{t=1}^T \partial_{\gamma_j, z}^2 g(Z_{0, t-1} + b' Z_{1, t-1})' \Omega^{-1} \partial_{\gamma_i} g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, k, t-1} \\ &\quad + \sum_{t=1}^T \partial_{\gamma_j} g(Z_{0, t-1} + b' Z_{1, t-1})' \Omega^{-1} \partial_{\gamma_j, z}^2 g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, k, t-1} \\ &\quad - \sum_{t=1}^T \varepsilon_t' \Omega^{-1} \partial_{\gamma_i, \gamma_j, z}^3 g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, k, t-1} \\ &\quad + \sum_{t=1}^T \partial_z g(Z_{0, t-1} + b' Z_{1, t-1})' \Omega^{-1} \partial_{\gamma_i, \gamma_j}^2 g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, k, t-1}, \end{aligned}$$

Hence the same arguments as for the proof of Claim 1 gives the desired.

**Claim 3:**  $T^{-2}G_{\gamma_i, b_{\kappa, j}, b_{\kappa, k}}, T^{-5/2}G_{\gamma_i, b_{\kappa, j}, b_\delta}, T^{-3}G_{\gamma_i, b_\delta, b_\delta}$  are  $O_P(1)$ .

Similar to the previous with  $b_k$  equal to either  $b_{\kappa, k}$  or  $b_\delta$ ,

$$\begin{aligned} G_{\gamma_i, b_j, b_k} &= \sum_{t=1}^T \partial_{\gamma_i, z}^2 g(Z_{0, t-1} + b' Z_{1, t-1}) \Omega^{-1} \partial_z g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, j, t-1} Z_{1, k, t-1} \\ &\quad + \sum_{t=1}^T \partial_z g(Z_{0, t-1} + b' Z_{1, t-1}) \Omega^{-1} \partial_{\gamma_i, z}^2 g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, j, t-1} Z_{1, k, t-1} \\ &\quad - \sum_{t=1}^T \partial_{\gamma_i} g(Z_{0, t-1} + b' Z_{1, t-1})' \Omega^{-1} \partial_z^2 g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, j, t-1} Z_{1, k, t-1} \\ &\quad - \sum_{t=1}^T \varepsilon_t' \Omega^{-1} \partial_{z, z, \gamma_i}^3 g(Z_{0, t-1} + b' Z_{1, t-1}) Z_{1, j, t-1} Z_{1, k, t-1} \end{aligned}$$

Hence the same arguments as for the proof of Claim 1 gives the desired.

**Claim 4:**  $T^{-5/2}G_{b_{\kappa, i}, b_{\kappa, j}, b_{\kappa, k}}, T^{-3}G_{b_{\kappa, i}, b_{\kappa, j}, b_\delta}, T^{-7/2}G_{b_{\kappa, i}, b_\delta, b_\delta}$  and  $T^{-4}G_{b_\delta, b_\delta, b_\delta}$  are  $O_P(1)$ .



With  $b_k$  equal to either  $b_{\kappa,k}$  or  $b_\delta$ ,

$$\begin{aligned}
G_{b_i, b_j, b_k, T}(\theta) &= \sum_{t=1}^T \partial_{z,z}^2 g(Z_{0,t-1} + b' Z_{1,t-1}) \Omega^{-1} \partial_z g(Z_{0,t-1} + b' Z_{1,t-1}) Z_{1,i,t-1} Z_{1,j,t-1} Z_{1,k,t-1} \\
&\quad + \sum_{t=1}^T \partial_z g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \partial_{z,z}^2 g(Z_{0,t-1} + b' Z_{1,t-1}) Z_{1,i,t-1} Z_{1,j,t-1} Z_{1,k,t-1} \\
&\quad - \sum_{t=1}^T \partial_z g(Z_{0,t-1} + b' Z_{1,t-1})' \Omega^{-1} \partial_{z,z}^2 g(Z_{0,t-1} + b' Z_{1,t-1}) Z_{1,i,t-1} Z_{1,j,t-1} Z_{1,k,t-1} \\
&\quad - \sum_{t=1}^T \varepsilon_t' \Omega^{-1} \partial_{z,z,z}^3 g(Z_{0,t-1} + b' Z_{1,t-1}) Z_{1,i,t-1} Z_{1,j,t-1} Z_{1,k,t-1}
\end{aligned}$$

As in the previous, arguments as for Claim 1 ends the proof.  $\square$

## C Auxiliary Lemmas

Consider  $Q_T(\theta)$  which is a function of observations  $X_1, \dots, X_T$  and the parameter  $\theta \in \Theta \subseteq \mathbb{R}^d$ . Introduce furthermore  $\theta_0$ , which is an interior point of  $\Theta$ . The proof is based on classic expansions of the likelihood function similar to Jensen and Rahbek (2004). However, here the information is stochastic in the limit and the arguments need be modified as done below in Lemma 10 and 11. This is well-known from the study of regression with non-stationary variables in Saikkonen (1995), and also similarly to the use of local (likelihood) expansions as in Boswijk (2002) and de Jong (2002).

**Lemma 10** *Assume that:*

**C.1**  $Q_T(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$  is three times continuously differentiable in  $\theta$ .

**C.2** There exists a sequence of nonsingular diagonal matrices  $U_T \in \mathbb{R}^{d \times d}$  such that  $U_T^{-1} \rightarrow 0$  and

$$\left( dQ_T(\theta_0; U_T^{-1/2} d\theta), d^2 Q_T(\theta_0; U_T^{-1/2} d\theta, U_T^{-1/2} d\bar{\theta}) \right)_{\theta=\theta_0} \xrightarrow{D} (0, H_\infty(d\theta, d\bar{\theta})).$$

where  $H_\infty(d\theta, d\bar{\theta}) > 0$  a.s.

**C.3**  $\sup_{\theta \in \mathcal{N}_T(\theta_0)} \left| d^3 Q_T(\theta; U_T^{-1/2} d\theta, U_T^{-1/2} d\bar{\theta}, U_T^{-1/2} d\tilde{\theta}) \right| = O_P(\|d\theta\| \|d\bar{\theta}\| \|d\tilde{\theta}\|)$   
over the sequence of local neighbourhoods

$$\mathcal{N}_T(\theta_0) = \left\{ \theta : \|U_T^{1/2}(\theta - \theta_0)\| < \epsilon \right\}.$$

Then with probability tending to one, there exists a unique minimum point  $\hat{\theta}$  of  $Q_T(\theta)$  in  $\mathcal{N}_T(\theta_0)$  which solves  $\partial Q_T(\hat{\theta})/\partial\theta = 0$ . It satisfies  $U_T^{1/2}(\hat{\theta} - \theta_0) = o_P(1)$ .

**Proof of Lemma 10.** Use a second order Taylor expansion to obtain for any bounded sequence  $h_T \in \mathbb{R}^d$  such that  $\theta_0 + U_T^{-1/2}h_T \in N_T(\theta_0)$ ,

$$Q_T(\theta_0 + U_T^{-1/2}h_T) - Q_T(\theta_0) = dQ_T(\theta_0; U_T^{-1/2}h_T) + \frac{1}{2}d^2Q_T(\bar{\theta}; U_T^{-1/2}h_T, U_T^{-1/2}h_T),$$

for some  $\bar{\theta} \in [\theta_0, \theta_0 + U_T^{-1/2}h_T] \in \mathcal{N}_T(\theta_0)$ . Define the bounded sequence  $\bar{h}_T = U_T^{1/2}(\bar{\theta} - \theta_0)$ . Then, by another application of Taylor's Theorem, there exists  $\tilde{\theta} \in [\theta_0, \bar{\theta}] \in \mathcal{N}_T(\theta_0)$  such that

$$\begin{aligned} & \left| d^2Q_T(\bar{\theta}; U_T^{-1/2}h_T, U_T^{-1/2}h_T) - d^2Q_T(\theta_0; U_T^{-1/2}h_T, U_T^{-1/2}h_T) \right| \\ &= \left| d^3Q_T(\tilde{\theta}; U_T^{-1/2}h_T, U_T^{-1/2}h_T, U_T^{-1/2}\bar{h}_T) \right| = O_P(\|h_T\|^2 \|\bar{h}_T\|) = O_P(\epsilon^3), \end{aligned}$$

where we have used (C.3). Thus,

$$\begin{aligned} & Q_T(\theta_0 + U_T^{-1/2}h_T) - Q_T(\theta_0) \\ &= dQ_T(\theta_0; U_T^{-1/2}h_T) + \frac{1}{2}H_\infty(h_T, h_T) \\ &+ \frac{1}{2} \left[ d^2Q_T(\theta_0; U_T^{-1/2}h_T, U_T^{-1/2}h_T) - H_\infty(h_T, h_T) \right] + O_P(\epsilon^3) \\ &= \frac{1}{2}H_\infty(h_T, h_T) + O_P(\epsilon^3), \end{aligned}$$

where the second equality follows by (C.2). By choosing  $\epsilon$  sufficiently small, and as  $H_\infty(h_T, h_T) > 0$  a.s.,  $Q_T(\theta)$  is convex with probability tending to one in the neighbourhood  $\mathcal{N}_T(\theta_0)$ . In particular, there exists a unique minimizer  $\hat{\theta} = \theta_0 + U_T^{-1/2}\hat{h}_T$  which solves the first-order condition,  $dQ_T(\hat{\theta}, d\theta) = 0$  for all  $d\theta$ . Since we can choose  $\epsilon$  arbitrarily small,  $\hat{h}_T = o_P(1)$ , and hence  $U_T^{1/2}(\hat{\theta} - \theta_0) = o_P(1)$  as desired.  $\square$

**Lemma 11** Assume that (C.1)-(C.3) holds and that

**C.4** There exists a sequence of numbers  $v_T \in \mathbb{R}_+$  such that  $v_T^{-1} \rightarrow 0$  and

$$\begin{aligned} & \left( dQ_T(\theta_0; v_T^{1/2}U_T^{-1/2}d\theta), d^2Q_T(\theta_0; U_T^{-1/2}d\theta, U_T^{-1/2}d\bar{\theta}) \right) \\ & \xrightarrow{D} (S_\infty(d\theta), H_\infty(d\theta, d\bar{\theta})). \end{aligned}$$

Then  $v_T^{1/2}U_T^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathbb{H}^{-1}\mathbb{S}$ , where  $\mathbb{H} \in \mathbb{R}^{d \times d}$  and  $\mathbb{S} \in \mathbb{R}^d$  are given through the following identities:

$$S_\infty(d\theta) = \mathbb{S}d\theta, \quad d\theta' \mathbb{H} d\bar{\theta} = H_\infty(d\theta, d\bar{\theta}).$$

**Proof of Lemma 11.** By Lemma 10, we know that  $\hat{\theta}_T$  is consistent and solves the first order condition. A first order Taylor expansion of the score and using (C.3) together with the same arguments as in the proof of Lemma 10 yields

$$\begin{aligned} dQ_T(\theta_0; v_T^{1/2}U_T^{-1/2}d\theta) &= d^2Q_T(\bar{\theta}; U_T^{-1/2}d\theta, U_T^{-1/2}v_T^{1/2}U_T^{1/2}(\hat{\theta} - \theta_0)) \\ &= d^2Q_T(\theta_0; U_T^{-1/2}d\theta, U_T^{-1/2}v_T^{1/2}U_T^{1/2}(\hat{\theta} - \theta_0)) + o_P(1) \end{aligned}$$

such that, by (C.4),

$$S_\infty(d\theta) = H_\infty\left(d\theta, v_T^{1/2}U_T^{1/2}(\hat{\theta} - \theta_0)\right) + o_P(1).$$

This completes the proof.  $\square$

**Lemma 12** *Asssume that  $Z_t$  is a stationary  $V$ -geometrically ergodic time-homogenous Markov chain. Then for any function  $g \leq V$  such that  $E[g(Z_t)] = 0$ ,*

$$\sup_t E |E(g(Z_t) | Z_{t-m}, Z_{t-m-1}, \dots, Z_0)| \rightarrow 0 \quad \text{as } m \rightarrow \infty.$$

**Proof.** From the Markov property,

$$E[g(Z_t) | Z_{t-m} = z, Z_{t-m-1}, \dots, Z_0] = E[g(Z_t) | Z_{t-m} = z].$$

Next, by definition of  $V$ -geometric ergodicity (Meyn and Tweedie, 1992, pp.382),

$$\begin{aligned} \sup_z \frac{|E[g(Z_t) | Z_{t-m} = z]|}{V(z)} &\leq \sup_z \sup_{|f| \leq V} \frac{|E[f(Z_t) | Z_{t-m} = z] - E[f(Z_t)]|}{V(z)} \\ &= \sup_z \frac{\|P^m(\cdot | z) - \pi\|_V}{V(z)} \\ &\leq c\rho^m, \end{aligned} \tag{33}$$

where  $0 < c < \infty$  and  $0 < \rho < 1$ . In total,

$$|E(g(Z_t) | Z_{t-m} = z, Z_{t-m-1}, \dots, Z_0)| \leq c\rho^m V(z).$$

Taking expectations and using that  $\{Z_t\}$  is stationary,

$$\sup_t E |E(g(Z_t) | Z_{t-m}, Z_{t-m-1}, \dots, Z_0)| \leq c\rho^m EV(Z_0).$$

□

## D Figures

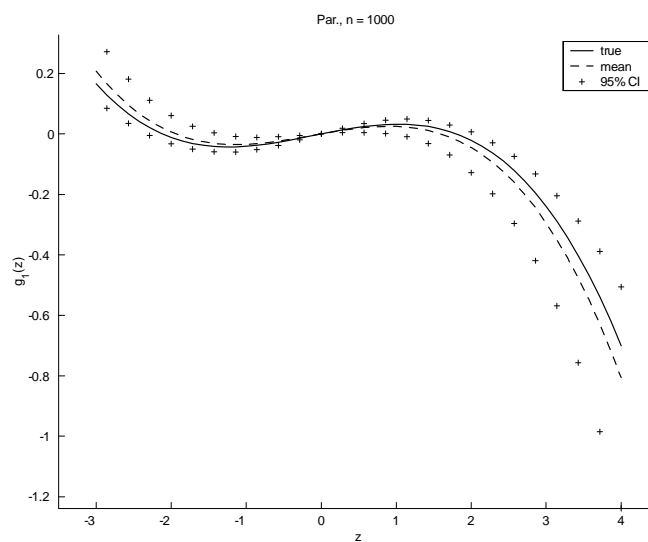


Figure 1: MLE of  $g_1(z)$  in ST ECM,  $T = 1000$  observations

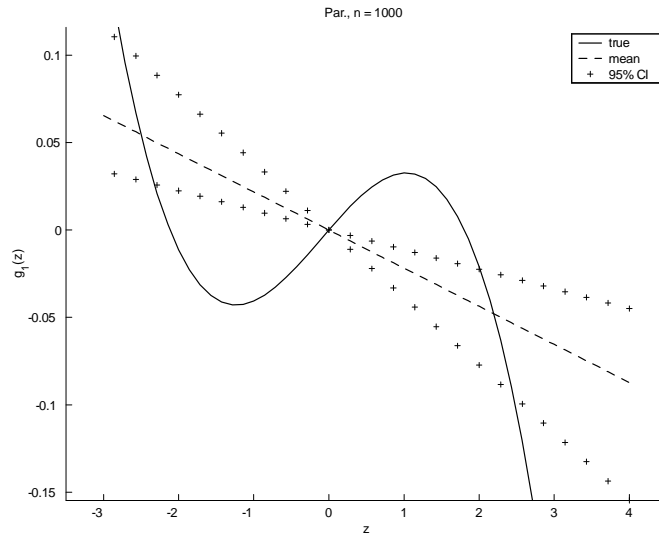


Figure 2: MLE of  $g_1(z)$  in linear ECM,  $T = 1000$ .

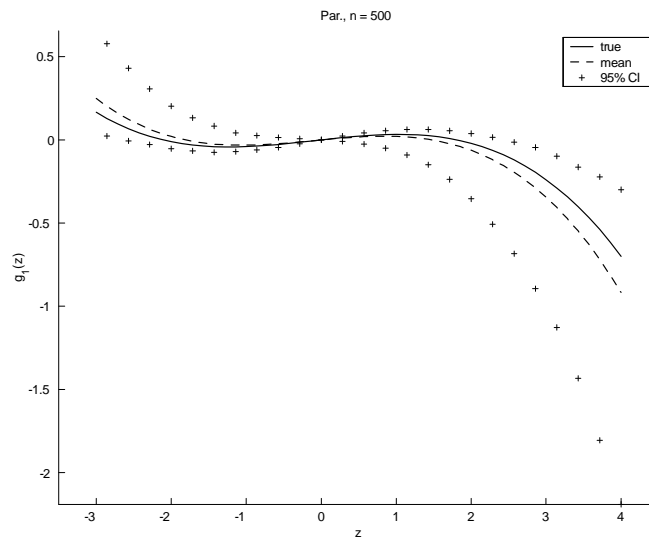


Figure 3: MLE of  $g_1(z)$  in STECM,  $T = 500$  observations

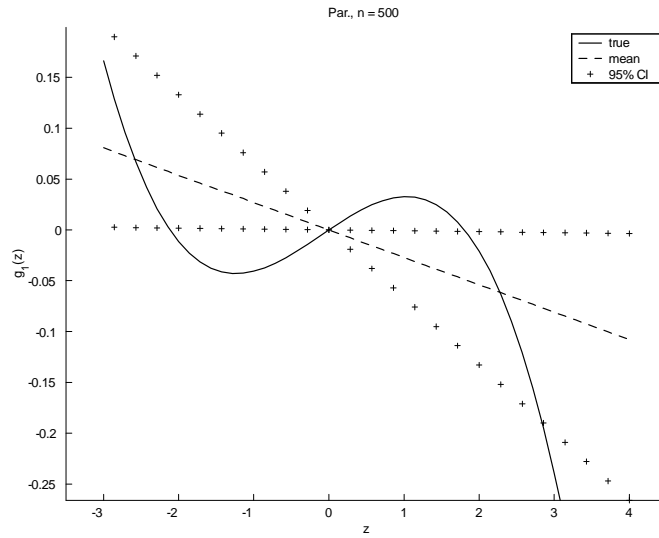


Figure 4: MLE of  $g_1(z)$  in linear ECM,  $T = 500$  observations

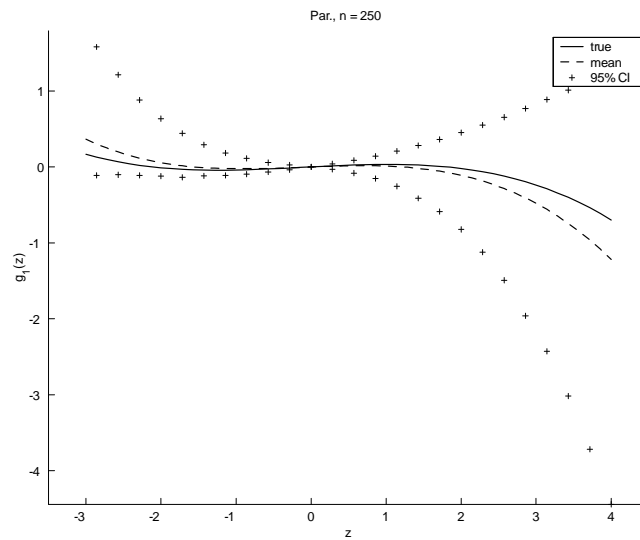


Figure 5: MLE of  $g_1(z)$  in ST ECM,  $T = 250$  observations

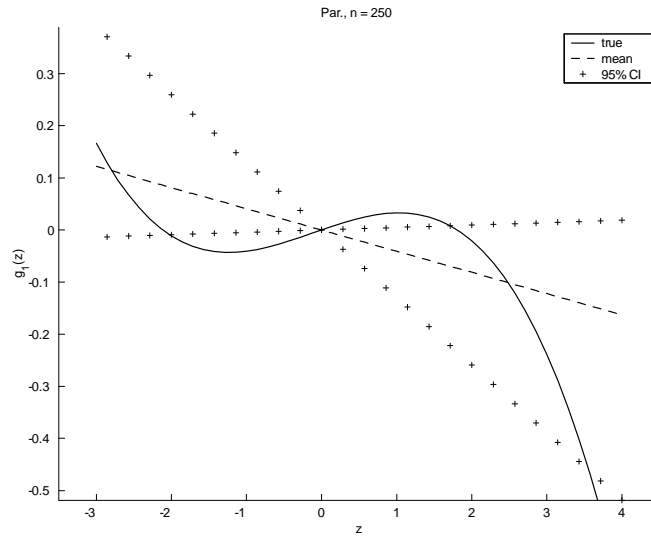


Figure 6: MLE of  $g_1(z)$  in linear ECM,  $T = 250$  observations

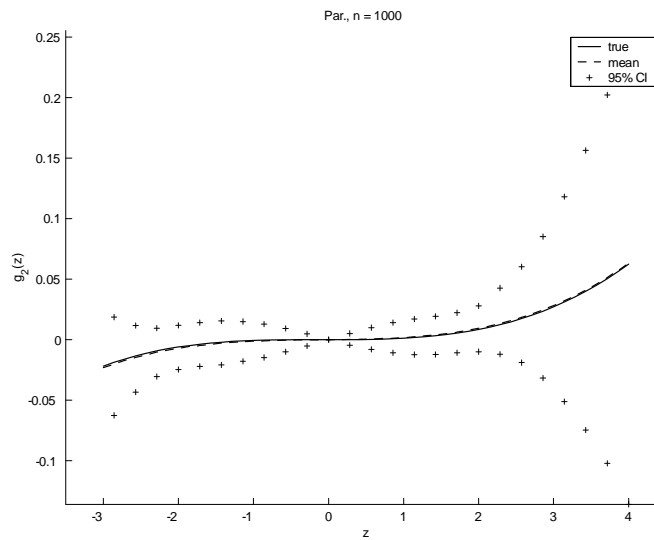


Figure 7: MLE of  $g_2(z)$  in ST ECM,  $T = 1000$  observations



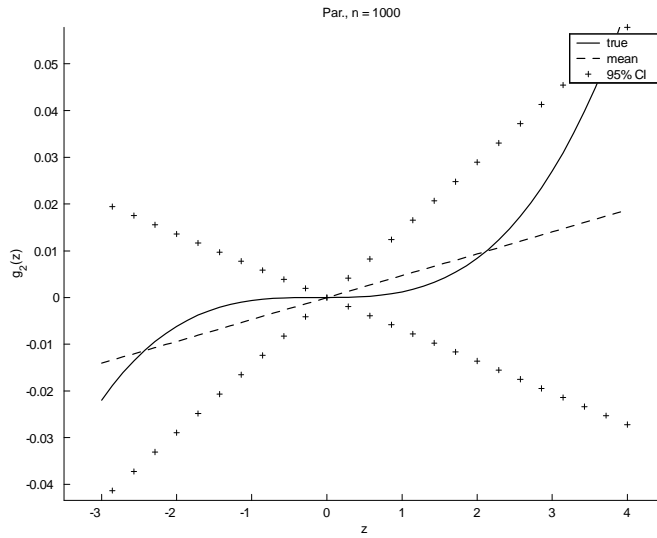


Figure 8: MLE of  $g_2(z)$  in linear ECM,  $T = 1000$  observations

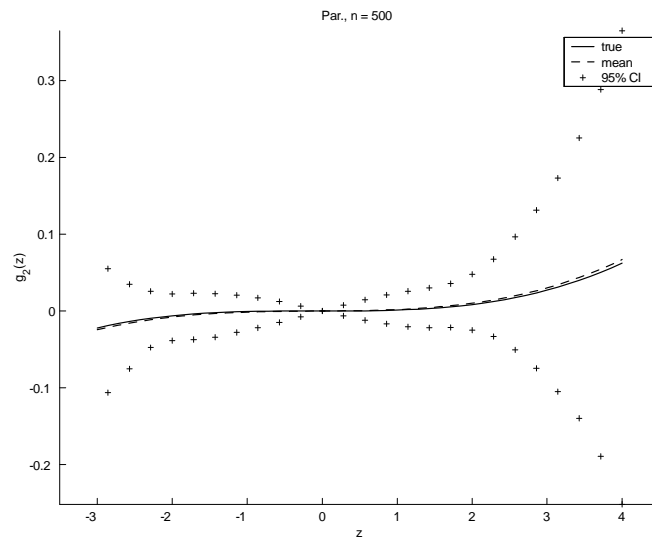


Figure 9: MLE of  $g_2(z)$  in ST ECM,  $T = 500$  observations

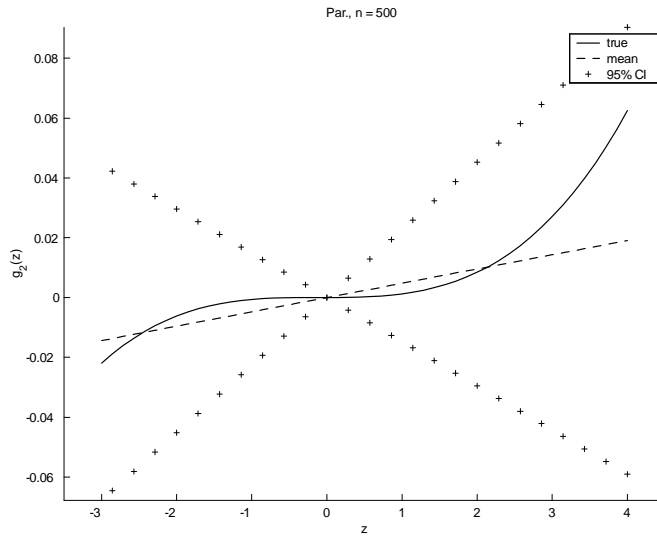


Figure 10: MLE of  $g_2(z)$  in linear ECM,  $T = 500$  observations

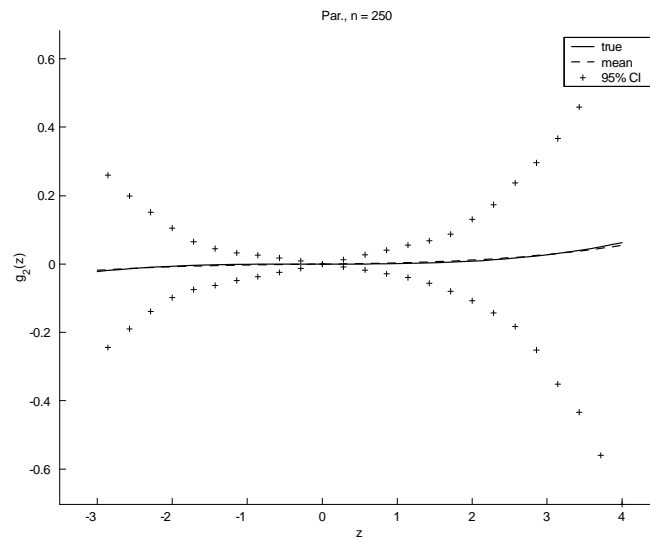


Figure 11: MLE of  $g_2(z)$  in STECM,  $T = 250$  observations

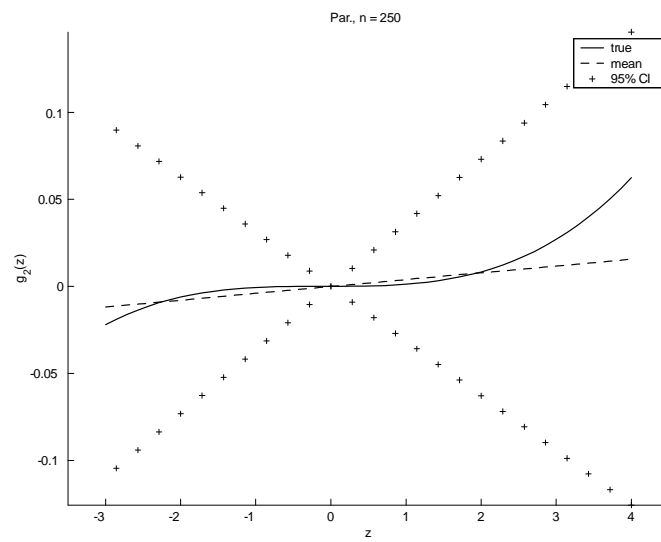


Figure 12: MLE of  $g_2(z)$  in linear ECM,  $T = 250$  observations

	$T = 250$			$T = 500$			$T = 1000$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$A (10^{-5} \times)$	-0.7000	1.7971	1.9284	-0.2383	0.9214	0.9518	-0.1070	0.5559	0.5661
$\omega$	0.6052	19.5887	19.5981	0.0858	4.3513	4.3512	-0.0127	0.2273	0.2276
$\alpha_1 (10^5 \times)$	-2.1154	7.4716	7.7653	-1.1666	4.1769	4.3368	-0.2356	1.2099	1.2326
$\alpha_2 (10^5 \times)$	0.8364	2.907	3.0381	0.4671	1.7008	1.7638	0.0906	0.4600	0.4688
$\tilde{\alpha}_1 (10^5 \times)$	1.0577	3.7358	3.8826	0.5833	2.0884	2.1684	0.1178	0.6049	0.6163
$\tilde{\alpha}_2 (10^5 \times)$	-0.4182	1.4603	1.5190	-0.2336	0.8504	0.8819	-0.0453	0.2300	0.2344
$\beta_2$	0.0117	0.2309	0.2312	-0.0005	0.0703	0.0703	0.0016	0.0255	0.0256

Table 1: Bias, Std and root-MSE (RMSE) of MLE of STECM.

	$T = 250$			$T = 500$			$T = 1000$		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
$\alpha_1 (10^3 \times)$	5.4177	0.0000	5.4177	5.4177	0.0000	5.4177	5.4177	0.0000	5.4177
$\alpha_2 (10^3 \times)$	-0.3764	0.0000	-0.3764	-0.3764	0.0000	-0.3764	-0.3764	0.0000	-0.3764
$\beta_2$	0.0787	4.2421	4.2428	0.0206	2.5716	2.5717	-0.0034	0.2249	0.2249

Table 2: Bias, Std and root-MSE of MLE of Linear ECM ( $\alpha_2 = 0$ ).

True values:  $A = -9.6178 \times 10^{-6}$ ,  $\omega = -0.1554$ ,  $\alpha = 100 \times (-5.41177, 0.3764)$ ,  
 $\tilde{\alpha} = 100 \times (2.7089, -0.1882)$ , and  $\beta_2 = -0.9519$ .