Fractional integration and cointegration in financial time series

Stakėnas, P.

Publication date
2012

Chapter 5

Likelihood-based inference in fractionally cointegrated systems

5.1 Introduction

Various aspects of fractionally (co)integrated systems have been analyzed in the preceding chapters, whereas this chapter studies likelihood-based inference in a parametric model for a fractionally cointegrated time series.

To set up the discussion, suppose we observe an n-dimensional time series X_t which is cointegrated with cointegration rank r: $X_t \sim I(d)$, but $\beta' X_t \sim I(d-b)$ for some $n \times r$ non-zero full column rank matrix β, $b > 0$ and some positive semi-definite long-run variance-covariance matrix of $\Delta_+^{d-b}\beta' X_t$. Methods of inference in fractionally cointegrated systems depend on the goals of an empirical researcher, but typically the parameters β and $\psi = (d, b)$ are of primary interest in empirical analysis, while the dynamics of the short-run component u_t is of secondary importance. If that is the case, one may conduct feasible (optimal) inference for the parameters of interest using semiparametric methods, which, with a few exceptions, are based on exploiting properties of the discrete Fourier transform of the (possibly fractionally filtered and/or tapered) time series X_t. Robinson and Marinucci (2001b) and Robinson and Marinucci (2001a) analyzed narrow-band least squares estimation of β in a certain parameter space achieving optimal inference, while Chen and Hurvich (2003b) proposed a modified version of the NBLS estimator, which uses tapered and differenced data. Nielsen and Frederiksen (2011) suggested fully modified NBLS estimator of β eliminating the asymptotic bias of the NBLS estimator in case of "weak" cointegration. It is also possible to estimate β with a two-step procedure: one could use initial stage, or first-step, estimate of β to estimate ψ (cf. Velasco (2003)) and then use this estimate for a second-step more accurate estimation of β. Chapter 4 reflects this idea in the time domain: a preliminary estimate of β was used to estimate b, which in turn was used in the DOLS regression to estimate β optimally. Another recent and even more general result in the semiparametric frequency domain framework was achieved in Hualde and Robinson (2010).

Another way to estimate parameters of interest is to conduct likelihood-based inference in a parametric model estimating all the parameters of a cofractional system jointly. A model for fractionally cointegrated time series was considered already in Granger (1986) and was analyzed in Łasak (2008), however its solution has not been derived and its properties are unknown. Other notable contributions include Gaussian maximum likelihood inference in a fractional ARIMA model in Dueker and Startz (1998), maximum likelihood estimation of β in a multivariate type I model in Jeganathan (1999) and generalized least-squares inference by Robinson and Hualde (2003). Recently Johansen and Nielsen (2012) have developed inference in the fractional VAR model, suggested by Johansen (2008), which allows for fractional cointegration:

$$\Delta^d X_t = \Delta^{d-b} L_b \alpha\beta' X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{5.1}$$

where ε_t is an i.i.d. zero mean series with a positive definite variance-covariance matrix, $0 < b \leq d$, α, β are $n \times r$ matrices. The solution of the model has the property $X_t \sim I(d)$, but $\beta' X_t \sim I(d-b)$, thus allowing for fractional cointegration. This model is probably the most general of all considered, allowing both for "weak" ($b < 1/2$) and "strong" ($b > 1/2$) cointegration, as well as a deterministic term ρ.

A possible generalization of the above-described fractional cointegration framework is to allow different integration orders between cointegration errors $\beta' X_t$. This is a non-trivial generalization and currently there are no parametric models proposed which would allow for that, although a two-step semiparametric procedure of Hualde and Robinson (2010) allows for inference on β in this case with standard asymptotic properties. This chapter proposes a model allowing for different memories for cointegration errors assuming that the driving $I(0)$ process is a finite-order VAR process: if $\Delta(\delta)$ is an $r \times r$ matrix with fractional filters $\Delta_+^{\delta_i}$ on a diagonal, then $\Delta(\delta)\beta' X_t$ is a finite order VAR process. Inference in this model is based on a Gaussian likelihood which is a nonlinear function of model parameters and hence we call the resulting estimator a non-linear least squares (NLS) estimator. The chapter formulates the model and studies inference in the model, deriving the asymptotic distribution of the NLS estimator along with a Wald test for β.

The rest of the chapter is organized as follows: Section 5.2 presents the framework we are dealing with. Section 5.3 outlines the idea of estimation and presents the main results.

The results of Monte Carlo simulations used to evaluate the finite sample performance of the estimator are presented in Section 5.4, Section 5.5 presents an empirical application, while Section 5.6 concludes. Proofs are given in the appendices.

We use the following notation: $\stackrel{P}{\to}$ means convergence in probability, $\stackrel{d}{\to}$ denotes convergence in distribution, $\Rightarrow$ denotes weak convergence of probability measures on the space of continuous functions induced by the corresponding processes. Also, an informal statement "tightness of a process" is understood formally as tightness of a sequence of probability measures on a space of continuous functions induced by the sequence of processes at hand. The Euclidian norm of a matrix, vector or scalar Z is denoted as $||Z|| = \sqrt{tr(Z'Z)}$. We will also use spectral norm defined on a space of real finite matrices as: $||Z||_1 = \sqrt{\lambda_{\max}(Z'Z)}$. The i-th order derivative of the function $f(x,y)$ with respect to x is denoted as $D_x^i f$. $AsVar(Z)$ denotes the asymptotic variance of an asymptotically stationary random variable Z. We will say that an element A in Euclidian space is $O_p(1)$, if $||A|| = O_p(1)$. By $\otimes$ we denote Kronecker multiplication and by $\circ$ we denote Hadamard multiplication. $Sp(\alpha)$ denotes the Euclidian subspace generated by a collection α of vectors in Euclidian space. $\iota_{ij}^{k\times l}$ denotes a $k \times l$ matrix of zeros, only with its ij-th element equal to 1 and ι_j^n denotes $n \times 1$ vector of zeros with 1 in its j-th coordinate. $int(A)$ denotes the interior of an Euclidian set A.

5.2 Preliminaries: model, assumptions, parameters

We are concerned with a multivariate fractionally cointegrated time series:

Assumption 5.1. *The observed n-dimensional time series X_t satisfies:*

$$\Delta(\psi)RX_t = u_t \sim I(0), \ \ t \geq 0, \tag{5.2}$$

where $\psi = (\delta_1, \ldots, \delta_r, d)$ and $\Delta(\psi) = blockdiag\left(diag(\Delta_+^{\delta_1}, \ldots, \Delta_+^{\delta_r}), \Delta_+^d I_{n-r}\right)$ is a block-diagonal matrix with $0 \leq \delta_1 < \delta_2 < \ldots < \delta_{r-1} < \delta_r < d$. R is an $n \times n$ full-rank matrix. Additionally, it holds: $\delta_i - \delta_j \neq 1/2$, $d - \delta_i \neq 1/2$, $\forall i, j = 1 \ldots r$.

Note, that the time series X_t is integrated of order d, but the first r coordinates of the vector RX_t are integrated of order smaller than d, hence X_t is cointegrated and has cointegration rank r. We do not restrict either the integration order d or the memory of cointegration residuals δ_i[1], hence assuming very general framework for the fractionally

[1]Apart from a technical assumption that no cointegration strength is equal to 1/2. The assumption $\delta_i \neq \delta_j$ may be relaxed as discussed further.

cointegrated system X_t. Note, that it is assumed that the observed time series X_t is cointegrated and the cointegration rank is known, which might seem somewhat restrictive, but it is not an overly restrictive assumption, since it can be easily verified outside the model (see Breitung and Hassler (2002), Chen and Hurvich (2003b), Nielsen and Shimotsu (2007), Łasak (2010) among others).

The following assumption is made on the error term u_t:

Assumption 5.2. *$\Phi(L)u_t = \varepsilon_t$, where $\Phi(L)$ is a matrix lag polynomial of order k: $\Phi(L) = I_n + \sum_{i=1}^{k} \Phi_i L^i$, such that: $\det(\Phi(z)) \neq 0$, $\forall |z| \leq 1$ and ε_t is an i.i.d. $(0, \Sigma)$ process with $\Sigma > 0$, $E||\varepsilon_t||^q < \infty$ with $q = 4$, if $d - \delta_1 < 1/2$ and $q = \max\{6, (\underline{b} - 1/2)^{-1}\} + \epsilon$, if $d - \delta_1 > 1/2$ for some $\epsilon > 0$ with $\underline{b} = \min\{\min_{i,j:\delta_i - \delta_j > 1/2}(\delta_i - \delta_j), \min_{i:d-\delta_i>1/2}(d - \delta_i)\}$.*

Although Assumption 5.2, defining an $I(0)$ process u_t as a finite order VAR process, might seem restrictive and some other parametrizaton, such as VARMA, might be desirable, in practice multivariate processes in the time domain mostly are modeled with finite order VAR models, hence from a practical point of view we do not find this assumption too restrictive. However, we point out, that Hualde and Robinson (2010) allows for a more general class of linear processes u_t, namely of infinite order with some additional smoothness conditions on its spectral density.

If $d - \delta_1 > 1/2$, Assumption 5.2 ensures that a multivariate fractional invariance principle holds[2] for u_t with all $b > \underline{b} - \eta > 1/2$ for some[3] $\eta = \eta(\epsilon) > 0$:

$$T^{1/2-b} \sum_{t=0}^{[Ts]} \Delta_+^{1-b} u_t \Rightarrow \Phi(1)^{-1} W_{b-1}(s). \tag{5.3}$$

Here $W_{b-1}(s)$ is type II fractional Brownian motion with covariance matrix Σ, defined as follows:

$$W_b(0) = 0, \;\; a.s.$$
$$W_b(s) = \frac{1}{\Gamma(b+1)} \int_0^s (s-u)^b dW(u),$$

where $W(u)$ is a Brownian motion with covariance matrix Σ.

An important issue in our analysis is the identification of cointegration subspaces, which implies additional restrictions on the matrix R. If r_i is the i-th row of the matrix

[2]Since $\Phi(L)$ has roots outside unit circle, that implies 1/2-summability of coefficients of $\Psi(L) = \Phi(L)^{-1}$ and Theorem 1 in Marinucci and Robinson (2000) for u_t applies.

[3]η in Assumption 5.2 depends on ϵ: $\eta(\epsilon) = q^{-1} - (\underline{b} - 1/2)^{-1/2}$.

R, then additivity features[4] of integrated processes imply that identifying restrictions on the matrix R have to be such, that $Sp(r_1, \dots, r_s) \cap Sp(r_{s+1}, \dots, r_n) = \{0\}$, $\forall s \leq r$. Hualde (2009) showed[5] that it is possible to achieve identification of the matrix R with the following scheme:

$$R = R(\beta) = \begin{pmatrix} 1 & \beta_1' & & & & & \\ 0 & 1 & \beta_2' & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 1 & \beta_r' & & \\ 0 & \dots & 0 & 0 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.4}$$

where $\beta_i \in \mathbb{R}^{n-i}$ are unrestricted and $\beta = (\beta_1', \dots, \beta_r')'$. Restrictions (5.4) on the matrix R are identifying restrictions for cointegration subspaces generating cointegration errors with different integration orders with respect to Definition 1.2, as shown in Hualde (2009), but in our setting, with $u_t \sim VAR(k)$, restrictions (5.4) are not identifying, but rather over-identifying in a sense that adding restrictions (5.4) to the Assumptions 5.1-5.2 makes the model smaller. Nevertheless, in the following we assume these restrictions suggested by Hualde (2009) and also used in Hualde and Robinson (2010):

Assumption 5.3. *The matrix R in the Assumption 5.1 has the structure (5.4).*

We note that Assumptions 5.1-5.3 imply that the model is not closed under linear transformations, i.e. if X_t belongs to the model, then AX_t may not belong to the model for the full rank $n \times n$ matrix A, what is a certain disadvantage of the model. In particular, the model is not invariant to permutations of multivariate series components, i.e. if the series $X_t = (X_{1t}, \dots, X_{nt})'$ satisfies (5.2) with restrictions (5.4), then the series $\tilde{X}_t = (X_{i_1 t}, \dots, X_{i_n t})'$ does not necessarily satisfies (5.2), (5.4), where $\{i_k\}$ is a permutation of $\{1, \dots, n\}$. Assumption 5.2 is not relevant in this respect: non-invariance to permutations holds if the matrix PRP' does not admit LU factorization for some permutation matrix P, reordering the rows of R, even with a broader class of $I(0)$ processes (such as in Definition 1.1). In particular, if at least one of the upper off-diagonal elements of R is zero, then such a permutation will exist. That might pose some additional considerations in empirical applications, as we discuss in Section 5.5.

[4]Namely, that if $X_{1t} \sim I(d_1)$, $X_{2t} \sim I(d_2)$ for $d_1 < d_2$, then $\alpha_1 X_{1t} + \alpha_2 X_{2t} \sim I(d_2)$ with $\alpha_2 \neq 0$.

[5]The paper studied more general setting allowing for restrictions of the type: $\delta_1 = \dots = \delta_{r_s} < \delta_{r_s+1} = \dots = \delta_{r_{s-1}} < \dots < \delta_{r_2+1} = \dots = \delta_{r_1}$.

Denote the parameters of the model as: $\kappa = (vec(\Sigma)', \theta')'$, where $\theta' = (\beta', \tau')$, $\tau' = (\psi, \Phi')$, $\beta' = (\beta_1', \dots, \beta_r')$, $\Phi' = (vec(\Phi_1)', \dots, vec(\Phi_k)')$. Then inference on the parameters κ in the model (5.2) is based on the Gaussian likelihood conditioned on the first k observations. For $T > k$, the conditional Gaussian log-likelihood function for the model (5.2) is:

$$l_T(\kappa) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\det(\Sigma) + tr\left(\Sigma^{-1}\frac{1}{T}\sum_{t=k}^{T}\hat{\varepsilon}_t(\theta)\hat{\varepsilon}_t(\theta)'\right), \tag{5.5}$$

where $\hat{\varepsilon}_t(\theta) = \Phi(L)\Delta(\psi)R(\beta)X_t$. We may concentrate out the parameter Σ and consider the scaled profile likelihood:

$$l_T(\theta) = \log\det(\hat{\Sigma}_T(\theta)), \tag{5.6}$$

$$\hat{\Sigma}_T(\theta) = T^{-1}\sum_{t=k}^{T}\hat{\varepsilon}_t(\theta)\hat{\varepsilon}_t(\theta)'. \tag{5.7}$$

For studying consistency of estimates we need to define an appropriate parameter space for the model parameters. Throughout the chapter we use a subscript “0” for the true values of the parameters: $\kappa_0 = (vec(\Sigma_0)', \theta_0')'$, etc. We define a compact set in the Euclidian space[6] $\Theta = \mathbb{D}_\beta \times \mathbb{D}_\psi \times \mathbb{D}_\Phi \subset \mathbb{R}^{\bar{k}}$ with the following properties:

1. $\theta_0 \in int(\Theta)$,

2. Denote $\delta_{r+1,0} = d_0$ and for every $i = 1, \dots, r+1$ define:

$$\Delta_i^1 = \begin{cases} 1/2, & \text{if } d_0 - \delta_{i0} < 1/2, \\ \min\limits_{i,j:\delta_{j0}-\delta_{i0}>1/2} \delta_{j0} - \delta_{i0} - 1/2, & \text{otherwise.} \end{cases}$$

$$\Delta_i^2 = \min_{j\geq i:\delta_{j0}-\delta_{i0}<1/2} 1/2 - (\delta_{j0} - \delta_{i0}).$$

Then $\mathbb{D}_\psi : -\Delta_i^2 + \eta_1 \leq \delta_i - \delta_{i0} \leq \Delta_i^1 - q^{-1} - \eta_1$, for any arbitrary small $\eta_1 > 0$.

3. $\mathbb{D}_\Phi : \det(\Phi(z)) \neq 0$, $\forall |z| \leq 1$, $\forall \Phi \in \mathbb{D}_\Phi$.

The properties 1. and 3. of the parameter space are obvious, while the somewhat obscure looking property 2. ensures that $l_T(\theta)$ converges uniformly on the parameter space, which is crucial for deriving consistency of the estimators. Further we define the non-linear least

[6]Note the dimension of parameter θ: $dim(\theta) = \bar{k} = kn^2 + (2n - r - 1)r/2 + r + 1$.

squares (NLS) estimator of the model as:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} l_T(\theta). \tag{5.8}$$

The main results of this chapter are presented in the next section and concern the asymptotic properties of estimator $\hat{\theta}$: its consistency, asymptotic normality as well as feasible inference procedures for a linear hypothesis on the parameter θ.

5.3 Main results

Section 5.2 presented the model and laid the framework for inferential analysis. In this Section we present the main results of the chapter. In a general case, when both "weak" and "strong" cointegration relations in the model (5.2) are present, the asymptotic distribution of the estimator $\hat{\theta}$ is a combination of normal and mixed normal distributions and we need additional notation for describing it. Define $L = L_{n,r}$ - an $r(2n-r-1)/2 \times nr$ selection matrix, which could be implicitly defined as: $Lvec(R(\alpha)'(I_r, 0_{r\times n-r})') = \alpha$, for any $\alpha \in \mathbb{R}^{r(2n-r-1)/2}$. We also need the following operator transforming an $r(2n-r-1)/2$ vector into an $n \times r(2n-r-1)/2$ matrix: $C(\alpha) = C_{n,r}(\alpha) = (\iota_1^n \otimes \alpha_1', \ldots, \iota_r^n \otimes \alpha_r')$. Then define the matrices:

$$N_i = diag(T^{b_{i1}-1/2}, \ldots, T^{b_{in}-1/2}), \tag{5.9}$$

$$M_T(\psi) = Lblockdiag(N_1, \ldots, N_r)L', \tag{5.10}$$

$$P = L(I_r \otimes R_0)L', \tag{5.11}$$

with $b_{ij} = \max\{1/2, \delta_{j0} - \delta_i\}, j \leq r$, $b_{ij} = \max\{1/2, d_0 - \delta_i\}, j > r$. In addition, we define $\xi_t(\psi) = (\Delta_+^{\delta_1} X_t' R_0', \ldots, \Delta_+^{\delta_r} X_t' R_0')'$ and S_β^S, S_β^N - full row rank selection matrices selecting stationary and nonstationary coordinates of the vector $L\xi_t(\psi_0)$, respectively. Then the following theorem gives the asymptotic distribution of the NLS estimator $\hat{\theta}$:

Theorem 5.1. *Let the Assumptions 5.1-5.3 hold. Then for the estimator (5.8) it holds:*

$$\sqrt{T}\begin{pmatrix} \begin{pmatrix} S_\beta^N \\ S_\beta^S \end{pmatrix} M_T(\psi_0)P^{-1\prime}(\hat{\beta} - \beta_0) \\ \hat{\tau} - \tau_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \left(\int_0^1 H(u)H(u)'du\right)^{-1} \int_0^1 H(u)\Sigma_0^{-1/2} dW(u) \\ N(0, \Gamma) \end{pmatrix}, \tag{5.12}$$

where $H(u) = S_\beta^N G_\psi(u)\Phi_0(1)'\Sigma_0^{-1/2}$ *and* $G_\psi(u) = (G^1(u)', \ldots, G^r(u)')'$ *with* $G^i(u) = \iota_i^{n\prime} \otimes$

$\zeta_i(u)$, $\zeta_i(u)$ is a $n-i$ vector with its j-th coordinate equal to $\iota_j^{n\prime}\Psi(1)W_{\delta_{j0}-\delta_{i0}-1}(u)\mathbb{1}_{\delta_{j0}-\delta_{i0}>1/2}$, alternatively it can be written as: $G'_\psi(u) = C(L\zeta(u))$, where $\zeta(u)$ is a weak limit of $N^{-1}(\psi_0)O_\xi^N\xi_{[Tu]}(\psi_0)$. Γ is some positive definite matrix.

Denote Γ_{β^S} - a principal minor of the matrix Γ, corresponding to the covariance matrix of $S_\beta^S M(\psi_0)P^{-1\prime}(\hat{\beta}-\beta_0)$. Then we have:

$$\sqrt{T}\begin{pmatrix} S_\beta^N \\ S_\beta^S \end{pmatrix} M_T(\psi_0)P^{-1\prime}(\hat{\beta}-\beta_0) \xrightarrow{d} \begin{pmatrix} \left(\int_0^1 H(u)H(u)'du\right)^{-1}\int_0^1 H(u)\Sigma_0^{-1/2}dW(u) \\ N(0,\Gamma_{\beta^S}) \end{pmatrix}. \tag{5.13}$$

The asymptotic distribution of the estimator $\hat{\beta}$ is rather complicated, since each component of (rotated) $\hat{\beta}$ has a different rate of convergence: $(P^{-1\prime}\hat{\beta})_{ij}$ is $T^{b_{ij}}$-consistent, hence the asymptotic distribution of $\sqrt{T}M_T(\psi_0)P^{-1\prime}(\hat{\beta}-\beta_0)$ is a combination of normally and mixed normally distributed random variables.

Next we turn to the inference problem on the parameter β and consider Wald test for testing linear restrictions on β. Consider the linear hypothesis $H_0: K\beta = m$, where K is $s \times m(2n-r-1)/2$ matrix, then a Wald test statistic for the null based on NLS estimates of $\hat{\vartheta}$ is:

$$W_\beta = \tag{5.14}$$

$$\left(K\hat{\beta}-m\right)'\left(K\left(\sum_{t=k}^{T}\left(\hat{\Phi}(L)C(L\varsigma_t(\hat{\psi}))\right)'\hat{\Sigma}^{-1}\hat{\Phi}(L)C\left(L\varsigma_t(\hat{\psi})\right)\right)^{-1}K'\right)^{-1}\left(K\hat{\beta}-m\right), \tag{5.15}$$

where $\varsigma_t(\psi) = (\Delta_+^{\delta_1}X_t', \ldots, \Delta_+^{\delta_r}X_t')'$ and Σ is the variance-covariance matrix of ε_t. We now state corollary of Theorem 5.1, which shows standard asymptotics of the Wald statistic $W_\beta(\psi,\Sigma)$ under the null hypothesis for consistent estimates of ψ and Σ:

Corollary 5.1. *Under the null it holds: $W_\beta \xrightarrow{d} \chi_s^2$.*

Remark 5.1. If $\delta_{i+j}-\delta_j > 1/2$, $\forall j = 1,\ldots,r, i = 1,\ldots,r-j$ and $d-\delta_r > 1/2$, then it is possible to simplify the Wald statistic as follows:

$$W_\beta(\psi,\Omega) = \left(K\hat{\beta}-r\right)'\left(K\left(\sum_{t=k}^{T}\left(C(L\varsigma_t(\psi))\right)'\Omega^{-1}C\left(L\varsigma_t(\psi)\right)\right)^{-1}K'\right)^{-1}\left(K\hat{\beta}-r\right),$$

where Ω is a long-run variance-covariance matrix of u_t and for $W_\beta(\psi, \Omega)$ Corollary 5.1 holds with Σ replaced by Ω.

Remark 5.2. In case there is equality between integration orders of cointegration errors, $\delta_1 = \ldots = \delta_{r_s} < \delta_{r_s+1} = \ldots = \delta_{r_{s-1}} < \ldots < \delta_{r_2+1} = \ldots = \delta_{r_1}$, we suggest the following restrictions for the matrix R, as proposed by Hualde (2009):

$$R = R(\beta) = \begin{pmatrix} I_{r_s} & B_s & & & \\ 0 & I_{r_{s-1}-r_s} & B_2 & & \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & I_{r_1-r_2} & B_1 \\ 0 & \ldots & 0 & 0 & I_{r_0-r_1} \end{pmatrix}, \tag{5.16}$$

where $\beta = (vec(B_1)', \ldots, vec(B_s)')'$ and $vec(B_i) \in \mathbb{R}^{(r_i - r_{i-1})(r_0 - r_i)}$ are unrestricted. For a particular case, when $\delta_1 = \ldots = \delta_r = 0$ and $d = 1$, the restrictions imply a familiar triangular representation for $I(1)/I(0)$ cointegrated processes, introduced by Phillips (1991).

Remark 5.3. In practical implementation of model (5.2) estimation, an identification issue for the coefficients Φ arises, i.e. the parameter space for $\mathbb{D}_\Phi$ has to be restricted so that $\Phi(L)$ does not have roots on or inside the unit circle. In practice this may be hard to achieve, especially given that the $\hat{\Phi}$'s are multivariate regression estimates of $\Delta(\hat{\psi})R(\hat{\beta})X_t$ on its k lags. One of the ways to achieve the goal is to consider a penalized conditional likelihood function:

$$l_T^P(\theta) = l_T(\theta) + Pen_T(\theta). \tag{5.17}$$

The idea of the penalty term $Pen_T(\theta)$ is to give an (asymptotically negligible) penalty for the conditional likelihood $l_T(\theta)$, which is inversely proportional to the distance of roots of $\Phi(L)$ to the unit circle. One possible penalty could be simply an unconditional scaled likelihood of the first k observations:

$$Pen_T(\theta) = \frac{1}{T} \log L_X(X_1, \ldots, X_k). \tag{5.18}$$

where $L_X(x_1, \ldots, x_k)$ is the likelihood function of the first k observations with Gaussian error terms ε_t with variance-covariance matrix $\hat{\Sigma}_T(\theta)$. In this case, $\Phi(L)$ is implicitly forced to have roots outside the unit circle. More generally, Theorem 5.1 holds, if the penalty term is a smooth function of the data and parameters satisfying:

- $\sup_{\theta \in \Theta} |Pen_T(\theta)| = o_p(1)$,

- $\sup_{\theta\in\Theta} |D_\theta^2 Pen_T(\theta)| = o_p(1)$,
- $\sqrt{T} blockdiag\{M_T(\psi_0), I_{kn^2+r+1}\} D_\theta Pen_T(\theta_0) = o_p(1)$.

Remark 5.4. The model (5.2) is restrictive since it does not allow for any deterministic components. Consider the following model:

$$\Phi(L)\Delta(\psi)\left(R(\beta)X_t + \mu\right) = \varepsilon_t, \;\; t \geq 0, \tag{5.19}$$

where $\mu \in \mathbb{R}^n$. Divide the parameter μ into subparameters: $\mu = (\mu^{S\prime}, \mu^{N\prime})'$, such that $\mu^{S\prime} = (\mu_1, \ldots, \mu_s)'$ and $\delta_s \leq 1/2 < \delta_{s+1}$ and define a normalization matrix $Q_T(\psi) = diag(T^{-\delta_1}, \ldots, T^{-\delta_s}, T^{-1/2}, \ldots, T^{-1/2})$. Then, if in addition to Assumptions 5.1-5.3 $\delta_s < 1/2$ holds, we have:

$$\sqrt{T}\begin{pmatrix} Q_T(\psi_0)\begin{pmatrix}\hat{\mu}^S - \mu_0^S \\ \hat{\mu}^N - \mu_0^N\end{pmatrix} \\ \begin{pmatrix} S_\beta^N \\ S_\beta^S\end{pmatrix} M_T(\psi_0)P^{-1\prime}(\hat{\beta} - \beta_0) \\ \hat{\tau} - \tau_0 \end{pmatrix} \overset{d}{\to} \mathcal{H}_\theta^{-1}\nabla_\theta, \tag{5.20}$$

where:

$$\mathcal{H}_\theta = \begin{pmatrix} \Gamma_{\mu^S} & 0 & \Gamma_{\mu^S\beta^N} & 0 \\ 0 & \Gamma_{\mu^N} & 0 & 0 \\ \Gamma_{\beta^N\mu^S} & 0 & \Gamma_{\beta^N} & 0 \\ 0 & 0 & 0 & \Gamma \end{pmatrix}, \;\; \nabla_\theta = \begin{pmatrix} \nabla_{\mu^S} \\ \nabla_{\mu^N} \\ \nabla_{\beta^N} \\ \nabla_{\beta^S\tau} \end{pmatrix}. \tag{5.21}$$

∇_{μ^S} is $s \times 1$ vector with i-th element $\iota_i^{n\prime}\Phi_0(1)'\Sigma_0^{-1}\int_0^1 F_{\delta_i} dW$, $\nabla_{\beta^S\tau}$ is a Gaussian vector with a positive definite variance-covariance matrix Γ, $\nabla_{\beta^N} = \int_0^1 H\Sigma_0^{-1/2} dW$ and the i-th element of the $(n-s)\times 1$ vector ∇_{μ^N} is a Gaussian random variable with variance $\iota_{n-s+i}^{n\prime}\Phi_0(1)'\Sigma^{-1}\Phi_0(1)\iota_{n-s+i}^n \sum_{t=1}^\infty F_{\delta_{n-s+i}}^2(t)$, where $F_\delta(u) = u^{-\delta}/\Gamma(1-\delta)$. The vector ∇_θ has a variance-covariance matrix $\mathcal{H}_\theta$, which is the inverse Fisher information under Gaussianity for the parameter θ with the non-zero blocks having the following typical ij-th elements:

$$(\Gamma_{\mu^S})_{ij} = \iota_i^{n\prime}\Phi_0(1)'\Sigma_0^{-1}\Phi_0(1)\iota_j^n \int_0^1 F_{\delta_i}F_{\delta_j},$$

$$(\Gamma_{\mu^N})_{ij} = \iota_{n-s+i}^{n\prime}\Phi_0(1)'\Sigma^{-1}\Phi_0(1)\iota_{n-s+j}^n \sum_{t=1}^\infty F_{\delta_{n-s+i}}(t)F_{\delta_{n-s+j}}(t),$$

$\Gamma_{\beta^N\mu^S} = \int_0^1 H\Sigma_0^{-1/2}\Phi_0(1)F_\delta$, where $F'_\delta = (F_{\delta_1}, \ldots, F_{\delta_s})(I_s, 0_{s\times n})$, $\Gamma_{\beta^N} = \int_0^1 HH'$ and Γ is the same matrix as in Theorem 5.1.

We see that μ_i is not estimated consistently, if $\delta_i > 1/2$, which is in line with results in a fractional regression framework (see also Chapter 4). Nevertheless, inclusion of such μ_i's in the model does not have an effect on the asymptotic distribution of other model parameters, since $\hat{\mu}^N$ is asymptotically uncorrelated with $\hat{\mu}^S, \hat{\beta}, \hat{\tau}$, as can be seen from the Fisher information matrix.

The Wald test statistic for a linear hypothesis $K\beta = m$ in the model (5.19) changes accordingly and is equal to (5.14) with ς_t replaced by ς_t regressed on $(F_{\delta_1}(t), \ldots, F_{\delta_s}(t))'$. It is also possible to consider deterministic terms of higher order in the model (5.2), which would yield a more complicated asymptotic distribution of the NLS estimator, dependent on the orders of stochastic and deterministic terms, but is not pursued here (for a fractional cointegration model with deterministic terms within a fractional regression framework, refer to Robinson and Iacone (2005)).

5.4 Finite sample performance

In this section we present Monte Carlo simulation results illustrating finite-sample performance of the above-described estimation and inference procedures. We simulate the following trivariate fractionally cointegrated system:

$$\Phi(L)\Delta(\psi)R(\beta)X_t = \varepsilon_t, \ \ t = 1\ldots, T, \tag{5.22}$$

where:

$$R(\beta) = \begin{pmatrix} 1 & \beta_{11} & \beta_{12} \\ 0 & 1 & \beta_{21} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \tag{5.23}$$

and $\Phi(L) = (1 - \phi L)I_3$, $\psi = (\delta_1, \delta_2, d)$, ε_t is a Gaussian i.i.d. $(0, \Sigma)$ time series with the ij-th element of the variance-covariance matrix Σ equal to: $\sigma_{ij} = \rho$, $i \neq j$ and $\sigma_{ii} = 1$, $i, j = 1, 2, 3$. The time series is simulated with the following parameter values: $\psi = (0, 0.6, 1.2), (0, 1, 2)$, $\rho = 0, 0.5$, $\phi = 0, 0.3$ and sample size $T = 256, 512, 1024$. In the estimation we minimized the value of the criterion function:

$$l_T(\theta) = \log\det(\hat{\Sigma}_T(\theta)) + Pen_T(\theta), \tag{5.24}$$

where $\hat{\Sigma}_T(\theta)$ is defined as in (5.6), while the penalty used is $Pen_T(\theta) = T^{-1}\sum_{i=1}^{3}\log(1-\lambda_i(\theta))$, where $\lambda_i(\theta)$ are roots of lag polynomial $\hat{\Phi}(L) = I_3 - \hat{\Phi}L$ (we do not restrict $\hat{\Phi}$ to be diagonal). The number of simulations is set to $R = 1000$. We report the bias and root-mean squared error (RMSE) of the NLS estimators $\hat{\beta}$ and $\hat{\psi}$ for different values of the parameters T, ϕ, ρ, ψ. We also simulate the empirical size of the Wald test (5.14) using NLS estimates for $\hat{\psi}$ and an estimate for the long-run covariance matrix Ω, which is constructed as follows: $\hat{\Omega} = \hat{\Phi}(1)^{-1}\hat{\Sigma}\hat{\Phi}(1)^{-1\prime}$, where $\hat{\Sigma}$ is a sample estimate of the variance-covariance matrix of ε_t.

Firstly, we discuss simulation results concerning bias of the estimators, reported in Table 5.1. We note that biases get smaller with increasing T, the bias of $\hat{\psi}$ being quite small and rather robust to the cointegration strength in the system and contemporaneous correlation parameter ρ (however, the bias of $\hat{\psi}$ slightly increases with an increase of ϕ). The bias of $\hat{\beta}$ is more sensitive to the parameter values: a positive (negative) ρ induces a positive (negative) bias of $\hat{\beta}$, whereas an increase in the temporal correlation coefficient ϕ, increases the magnitude of the bias.

Results for the RMSE are reported in the Table 5.2. The RMSE of $\hat{\psi}$ responds to changes in model parameters much in the same manner as the bias: the RMSE is relatively robust to the cointegration strength and parameter ρ, however displaying an increase with an increase of ϕ. More interesting is the RMSE of $\hat{\beta}$. We see a big difference in RMSE with different cointegration strengths in the system: RMSE's of $\hat{\beta}$ with $\psi_0 = (0, 0.6, 1.2)$ are significantly bigger than that with $\psi_0 = (0, 1, 2)$, reflecting strong dependence of the performance of $\hat{\beta}$ on the cointegration strength(s) in the system, as is prescribed by the asymptotic theory. Also, while RMSE with $\psi_0 = (0, 1, 2)$ is not very sensitive to the changes in parameters ρ, ϕ, it is clearly very sensitive to the autoregressive parameter ϕ for the case $\psi_0 = (0, 0.6, 1.2)$ with sample sizes $T = 256, 512$: an increase in ϕ induces increase in the RMSE of $\hat{\beta}$, although the effect becomes rather negligible for $T = 1024$. One might conclude that for moderate sample sizes, persistence in short-run noise plays a big role in the finite sample behaviour of the estimator $\hat{\beta}$, which tends to diminish in large samples.

Next we turn to the simulated empirical sizes of Wald tests, reported in the Table 5.3. Overall, except for the case $T = 1024$, the results are rather disappointing: tests are very much oversized and the empirical sizes are far from the nominal. We briefly discuss a few factors affecting the empirical size: the smaller the cointegration strength, the more distorted empirical sizes are, although the effect diminishes with sample size $T = 1024$. We also see slightly more oversized tests with larger values of ϕ, whereas the coefficient ρ has little impact on the empirical size. Finally, on a positive note, empirical sizes with $T = 1024$ tend not to depend much on the model parameters, although still being too large. We may conclude that while the performance of the test is rather disappointing in small to moderate sample sizes, it displays somewhat better properties in large samples.

Table 5.1: Monte Carlo RMSE and bias of $\hat{\beta}$, $\hat{\psi}$ for $\psi_0 = (0, 0.6, 1.2)$

			Bias$\times 10^3$						RMSE					
T	ϕ	ρ	β_{11}	β_{12}	β_{21}	δ_1	δ_2	d	β_{11}	β_{12}	β_{21}	δ_1	δ_2	d
256	0	0	-7.3	-8.1	-3.9	12.1	-34.7	-17.9	0.185	0.187	0.105	0.034	0.108	0.085
		0.5	16.2	16.2	5.6	10.0	-27.9	-15.5	0.104	0.103	0.101	0.029	0.093	0.081
	0.3	0	-6.0	-7.4	-5.3	16.5	-67.2	-28.1	0.316	0.309	0.173	0.046	0.179	0.135
		0.5	31.8	32.0	5.7	14.5	-53.3	-26.2	0.254	0.246	0.189	0.040	0.159	0.127
512	0	0	-1.7	-1.7	-0.3	6.1	-9.7	-10.1	0.023	0.023	0.022	0.019	0.048	0.048
		0.5	2.8	3.0	0.1	5.5	-7.7	-4.1	0.022	0.022	0.021	0.017	0.043	0.039
	0.3	0	-5.4	-5.6	-2.0	9.0	-23.2	-21.7	0.221	0.225	0.120	0.028	0.099	0.091
		0.5	11.9	11.9	4.3	7.1	-15.5	-10.8	0.121	0.118	0.086	0.024	0.072	0.071
1024	0	0	0.2	0.2	0.5	2.4	-4.4	-1.9	0.012	0.012	0.011	0.010	0.027	0.026
		0.5	1.2	1.2	-0.1	2.9	-3.6	-0.6	0.011	0.011	0.009	0.010	0.023	0.021
	0.3	0	0.3	0.3	0.3	3.8	-7.9	-2.8	0.014	0.015	0.011	0.014	0.042	0.039
		0.5	1.6	1.6	2.5	3.5	-6.2	-2.6	0.015	0.015	0.065	0.013	0.040	0.042

Table 5.2: Monte Carlo RMSE and bias of $\hat{\beta}$, $\hat{\psi}$ for $\psi_0 = (0, 1, 2)$

			Bias$\times 10^3$						RMSE					
T	ϕ	ρ	β_{11}	β_{12}	β_{21}	δ_1	δ_2	d	β_{11}	β_{12}	β_{21}	δ_1	δ_2	d
256	0	0	2.4	2.4	0.2	9.9	-40.0	-6.4	0.066	0.067	0.009	0.027	0.120	0.051
		0.5	1.8	1.8	-0.3	9.2	-27.5	-4.5	0.015	0.015	0.008	0.025	0.093	0.049
	0.3	0	-0.2	-0.3	0.3	13.1	-77.1	-8.5	0.017	0.017	0.009	0.038	0.195	0.085
		0.5	2.4	2.4	-0.2	13.1	-51.2	-7.7	0.015	0.015	0.009	0.036	0.152	0.076
512	0	0	-0.2	-0.2	-0.3	8.8	-17.9	-3.5	0.007	0.007	0.005	0.023	0.055	0.035
		0.5	0.4	0.4	-0.1	7.8	-10.9	-1.3	0.006	0.006	0.004	0.019	0.053	0.028
	0.3	0	-0.2	-0.2	-0.3	11.5	-34.8	-5.7	0.007	0.007	0.005	0.030	0.109	0.058
		0.5	0.8	0.8	-0.2	11.1	-22.8	-1.5	0.007	0.007	0.004	0.028	0.082	0.037
1024	0	0	0.2	0.2	-0.2	6.7	-8.5	0.7	0.003	0.003	0.002	0.016	0.035	0.023
		0.5	0.2	0.2	-0.1	5.4	-5.1	-0.1	0.003	0.003	0.002	0.014	0.028	0.019
	0.3	0	0.2	0.2	-0.2	8.0	-16.1	0.2	0.003	0.003	0.002	0.021	0.063	0.030
		0.5	0.3	0.3	-0.1	7.3	-7.1	0.1	0.003	0.003	0.002	0.019	0.036	0.025

Table 5.3: Monte Carlo results of empirical sizes of $W(\hat{\psi}, \hat{\Omega})$

			$\psi_0 = (0, 1, 2)$			$\psi_0 = (0, 0.6, 1.2)$		
T	ϕ	ρ	10%	5%	1%	10%	5%	1%
256	0	0	0.19	0.11	0.05	0.31	0.23	0.13
		0.5	0.18	0.13	0.06	0.34	0.27	0.17
	0.3	0	0.23	0.16	0.09	0.36	0.28	0.19
		0.5	0.22	0.17	0.09	0.39	0.34	0.24
512	0	0	0.13	0.08	0.03	0.16	0.10	0.04
		0.5	0.15	0.09	0.03	0.19	0.13	0.07
	0.3	0	0.15	0.10	0.05	0.19	0.14	0.07
		0.5	0.17	0.11	0.04	0.23	0.17	0.09
1024	0	0	0.11	0.06	0.02	0.11	0.07	0.03
		0.5	0.12	0.06	0.02	0.13	0.08	0.03
	0.3	0	0.13	0.07	0.02	0.13	0.08	0.04
		0.5	0.12	0.07	0.02	0.13	0.09	0.04

5.5 Empirical application

In this section we proceed with the empirical analysis of the Treasury yields dataset described in Section 1.3. We analyze a trivariate system $(M3, M6, M1)$ and estimate the model (5.19) based on findings in earlier chapters: since both bivariate systems cointegrate (Section 3.4), the trivariate series also cointegrates and the cointegration rank is $r = 2$; also, since in Section 2.5 we did not reject the null hypothesis that the yields are processes integrated of order 1, we impose $d = 1$ in (5.19). In the estimation of the model we have included 1st and 12th lags in the autoregression[7] and estimation was performed minimizing the value of the objective function (5.24) with the penalty term described in Section 5.4.

As it was mentioned, the ordering of the components of X_t presents an issue in the estimation of the model, since even though $X_t = (X_{1t}, \ldots, X_{nt})'$ belongs to the model, $\tilde{X}_t = (X_{i_1 t}, \ldots, X_{i_n t})'$ might not, where i_k is some permutation of $\{1, \ldots, n\}$. We suggest to solve the problem with the following rule-of-thumb procedure: firstly estimate bivariate systems with one selected component always present, and then order the components correspondingly to the estimates of δ_i's. In our case, the estimates of δ in the systems $(M3, M1)$, $(M6, M1)$ are $0.019, -0.016$ with numerical standard errors 0.09 and 0.08, correspondingly, which does not suggest a natural way of ordering, hence we estimate both trivariate systems: $(M3, M6, M1)$ and $(M6, M3, M1)$.

Likelihood function of both estimated models was very flat displaying many local

[7]Residuals of all considered models in the following did not display any significant autocorrelation after inclusion of these two lags.

maxima (which in some cases were on the boundary of the parameter space) with very close estimates of δ_1, δ_2 for both models. These observations point to the relationship $\{\delta_{10} = \delta_{20}\}$ and thus non-identification of the model with unrestricted β_{11}, explaining the unstable estimation procedure.[8] The restriction $\delta_1 = \delta_2$ can be tested restricting $\beta_{11} = 0$ in the estimation with δ_1, δ_2 being unrestricted. In this case, we get the values $0.071, 0.097$ of (δ_1, δ_2) in the system $(M3, M6, M1)$ with numerical standard errors $0.088, 0.047$, respectively. A Wald test for the null $\delta_1 = \delta_2$ using a numerical Hessian gives a p-value of 0.40 and thus we do not reject the null. Based on these findings, we estimate model (5.19) with restrictions $\beta_{11} = 0$, $\delta = \delta_1 = \delta_2$ and obtain:

$$\hat{\Phi}_1 = \begin{pmatrix} -0.43 & 0.51 & -0.09 \\ -1.09 & 1.20 & -0.08 \\ 1.07 & -0.15 & 0.16 \end{pmatrix}, \quad \hat{\Phi}_{12} = \begin{pmatrix} 0.54 & -0.22 & -0.09 \\ 0.49 & -0.22 & -0.11 \\ 0.05 & -0.06 & 0.19 \end{pmatrix} \tag{5.25}$$

and $(\hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\delta}, \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3) = (-1.04, -1.05, 0.06, 0.01, 0.10, 8.29)$. We note that the numerical standard error for $\hat{\delta}$ is 0.08 meaning that the value 0 falls into the 95% confidence interval for δ thus pointing to a possible non-fractional behaviour of the series.

In Section 4.5 we did not reject the null that the cointegration vector in both bivariate systems is $(1, -1)$ and this implies a specific structure on $R(\beta)$ under restriction $\beta_{11} = 0$: $\beta_{12} = -1$, $\beta_{21} = -1$. We test this hypothesis $(\beta_{12}, \beta_{21}) = (-1, -1)$ with the Wald test (5.14) using NLS estimates for $\hat{\psi}$ and a sample estimate for the variance-covariance matrix Σ. The p-value for the test is 0.074 thus not rejecting the null with conventional 5% significance level.

The estimate of the memory of cointegration errors $\hat{\delta}$ is not significantly different from zero, which stands in contrast to the results in Chapter 4, where the estimates of the memories of the term spreads suggested fractional cointegration (although a different estimator for the integration orders was used). We have also estimated the model (5.19) for the bivariate series $(M3, M1)$, $(M6, M1)$ and obtained estimates 0.24 and 0.21 for the memory of cointegration errors, repsectively, also suggesting fractional cointegration. We conjecture that the less efficient, bivariate system estimation, could be responsible for the discrepant estimation results and conclude the section with the following observations: i) strong support for the fractional behaviour of cointegration errors in the trivariate cointegrated system $(M3, M6, M1)$ was not found; ii) null hypothesis on the structure of cointegration space implied by EH was not rejected.

---

[8] Note, that since the restriction $\delta_2 > \delta_1$ in the estimation procedure was imposed, the model is misspecified and has flat likelihood in the parameter subspace with $\delta_2 \approx \delta_1$ and $\{\beta : \beta_{11} = a, \beta_{12} = a\beta_{21}, a \in \mathbb{R}\}$.

5.6 Conclusions

In this chapter a model for fractionally cointegrated time series was proposed, allowing for different integration orders for cointegration errors and deterministic terms. The model imposes a specific set of identifying restrictions upon cointegration vectors, taken from Hualde (2009). Asymptotic inference in the model parameters based on a conditional Gaussian likelihood has been studied and consistency and asymptotic distribution of the resulting estimator have been derived. Finite sample properties of the estimator were illustrated by means of Monte Carlo simulations. The simulations reveal significant size distortions for the Wald test with moderate sample sizes, but show satisfactory properties of the test for the sample size $T = 1024$. An empirical application for U.S. interest rate data is also provided.

5.7 Appendix

5.7.1 Preliminaries and notation

In this section we introduce notation used in the Appendix. We introduce a local parameterization[9]: $\alpha_i' = (\beta_i' - \beta_{i0}')R_{0i}^{-1}N_{ii}$, for $i = 1, \ldots, r$ where:

$$R_{0i} = (0_{n-i\times i}, I_{n-i})R_0(0_{n-i\times i}, I_{n-i})', \tag{5.26}$$

$$N_{ii} = (0_{n-i\times i}, I_{n-i})N_i(0_{n-i\times i}, I_{n-i})', \tag{5.27}$$

$$N_i = diag(T^{b_{i1}-1/2}, \ldots, T^{b_{in}-1/2}), \tag{5.28}$$

with $b_{ij} = \max\{1/2, \delta_{j0} - \delta_i\}, j \leq r$, $b_{ij} = \max\{1/2, d_0 - \delta_i\}, j > r$ and $R_0 = R(\beta_0)$. In the following, we will derive results for the estimator of $\vartheta = (\alpha', \tau')'$, $\alpha' = (\alpha_1', \ldots, \alpha_r')$, $\tau' = (\psi, \Phi')$, $\psi = (\delta_1, \ldots, \delta_r, d)$, $\Phi' = (vec(\Phi_1)', \ldots, vec(\Phi_k)')$. The parameter space Θ changes accordingly and the new parameter space for ϑ is denoted as Ξ (note that Ξ is expanding in certain directions and depends on T, unless $b_{ij} = 1/2$ for all i, j).

The conditional scaled Gaussian profile log-likelihood[10] of the model with the new

[9] A similar idea of reparameterization is also used in Johansen and Nielsen (2012).

[10] In the following, we refer to it as "objective function".

parameterization is defined as follows:

$$l_T(\vartheta) = \frac{1}{2}\log\det(\hat{\Sigma}(\vartheta)), \tag{5.29}$$

$$\hat{\Sigma}(\vartheta) = T^{-1}\sum_{t=k}^{T}\hat{\varepsilon}_t(\vartheta)\hat{\varepsilon}_t'(\vartheta), \tag{5.30}$$

$$\hat{\varepsilon}_t(\vartheta) = \Phi(L)\Delta(\psi)R(\vartheta)X_t. \tag{5.31}$$

For $\vartheta \in \Xi$ we define the gradient vector of $l_T(\vartheta)$ w.r.t. its $\bar{k}$ coordinates:

$$D_\vartheta l_T(\vartheta) = \nabla_T(\vartheta) = \left(\frac{\partial}{\partial vec(\alpha)}l_T(\vartheta), \frac{\partial}{\partial\delta_1}l_T(\vartheta), \ldots, \frac{\partial}{\partial\delta_r}l_T(\vartheta), \frac{\partial}{\partial d}l_T(\vartheta), \frac{\partial}{\partial vec(\Phi)}l_T(\vartheta)\right)'.$$

Then:

$$\nabla_T(\vartheta) = T^{-1}\sum_{t=k}^{T} D_\vartheta\hat{\varepsilon}_t'(\vartheta)\hat{\Sigma}(\vartheta)^{-1}\hat{\varepsilon}_t(\vartheta),$$

where, with some abuse of notation, we denote: $D_\vartheta\hat{\varepsilon}_t'(\vartheta) = (\partial_{\vartheta_1}\hat{\varepsilon}_t(\vartheta), \ldots, \partial_{\vartheta_{\bar{k}}}\hat{\varepsilon}_t(\vartheta))'$. Similarly we define the Hessian matrix of second derivatives:

$$D_\vartheta^2 l_T(\vartheta) = \mathcal{H}_T(\vartheta) = \left(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}l_T(\vartheta)\right)_{i,j=1,\ldots,\bar{k}}.$$

The key to proving asymptotic likelihood properties is to rewrite $l_T(\vartheta)$ as a quadratic form in the parameters α and to this end we rewrite: $R(\vartheta) = \Lambda(\alpha_1' N_{11}^{-1}R_{01}, \ldots, \alpha_r' N_{rr}^{-1}R_{0r})$ $+R(\vartheta_0)$. Here Λ is a matrix function mapping $\mathbb{R}^{(2n-r-1)r/2}$ vector into an upper-diagonal $n \times n$ matrix: $\Lambda(v_1', \ldots, v_r') = R(v_1', \ldots, v_r') - I_n$. Also denote an $nr \times nr$ block diagonal matrix: $N_T(\psi) = blockdiag(N_1, \ldots, N_r)$ and an $n \times nr$ matrix: $D(\alpha) = ((I_r, 0_{r\times n-r})' \otimes \mathbb{1}_n') \circ (\mathbb{1}_r' \otimes \Lambda(\alpha_1', \ldots, \alpha_r'))$, where $\mathbb{1}_j$ is a $j \times 1$ vector of ones. Then we have: $D(\alpha) = D(\beta - \beta_0)(I_r \otimes R_0^{-1})N_T(\psi)$ and hence:

$$\Delta(\psi)\Lambda(\beta-\beta_0)X_t = \Delta(\psi)D(\alpha)N_T^{-1}(\psi)(I_r \otimes R_0)(\mathbb{1}_r \otimes X_t) = D(\alpha)N_T^{-1}(\psi)\xi_t(\psi), \tag{5.32}$$

where $\xi_t(\psi) = (\Delta_+^{\delta_1}X_t'R_0', \ldots, \Delta_+^{\delta_r}X_t'R_0')'$. Then $\Delta(\psi)R(\beta)X_t = \Delta(\psi)R(\beta_0)X_t + \Delta(\psi)\Lambda(\beta - \beta_0)X_t = w_{1t} + D(\alpha)w_{2t}$ and we might rewrite $\hat{\Sigma}(\vartheta)$ as:

$$\hat{\Sigma}(\vartheta) = \Pi_1\Pi_2\left(\frac{1}{T}\sum_{t=k}^{T}W_tW_t'\right)\Pi_2'\Pi_1' = \Pi A_T(\psi)\Pi', \tag{5.33}$$

where: $\Pi_1 = \Pi_1(\Phi) = (I_n, \Phi_1, \ldots, \Phi_k)$, $\Pi_2 = \Pi_2(\alpha) = (I_{(k+1)n}, I_{k+1} \otimes D(\alpha))$, $\Pi = \Pi_1(\Phi)\Pi_2(\alpha)$, $W_t = (W'_{1t}, W'_{2t})'$, $W_{1t} = (w'_{1t}, \ldots, w'_{1t-k})'$, $W_{2t} = (w'_{2t}, \ldots, w'_{2t-k})'$.

In the following, we denote $\Psi(L)$ as the inverse lag polynomial of $\Phi_0(L)$: $\Psi(L) = \Phi_0(L)^{-1}$. α_{ij} is the j-th element of the vector α_i. Then we denote a subparameter α^N of the parameter α, for which it holds: $b_{ij} > 1/2$. Similarly we denote a subparameter α^S for which it holds: $b_{ij} = 1/2$. Then, if S_α^S, S_α^N are full row rank selection matrices selecting stationary and non-stationary coordinates of the vector $L_{n,r}\xi_t(\psi_0)$, respectively, we have: $S_\alpha^S \alpha = \alpha^S$, $S_\alpha^N \alpha = \alpha^N$. The reason behind this division of components of α is that the parameter space Ξ is expanding in directions α^N, while it is not in directions α^S, requiring different technical treatment of these subparameters. Fix $K > 0$ and denote a subspace of parameter space Ξ: $\Xi_K = \{\vartheta \in \Xi : ||\alpha^N|| \leq K\} \subset \Xi$. Then Ξ_K is a compact parameter space.

We sketch the proof of Theorem 5.1: firstly, we prove that the objective function $l_T(\vartheta)$ uniformly converges to its limit $l_\infty(\vartheta)$ with value $\vartheta = \vartheta_0$ being its minimizer. Then we prove existence and consistency of NLS estimator $\hat{\vartheta}$. Finally, expanding the gradient of $l_T(\vartheta)$ around ϑ_0 we get the asymptotic distribution of NLS estimator $\hat{\vartheta}$. The whole proof is divided into lemmas and relies on similar techniques and methods used in Johansen and Nielsen (2012). Although it is not explicit, Assumptions 5.1-5.3 hold throughout the Appendix.

5.7.2 Additional lemmas

Lemma 5.1. *$D_\vartheta^m l_T(\vartheta)$ is tight in $\vartheta \in \Xi_K$ for $m = 0, 1, 2$. In particular, the function $l_T(\vartheta)$ converges weakly on $\vartheta \in \Xi_K$: $l_T(\vartheta) \Rightarrow l_\infty(\vartheta)$ for some (random) function $l_\infty(\vartheta)$.*

Proof. Since $l_T(\vartheta) = \frac{1}{2}\log\det(\hat{\Sigma}(\vartheta))$, due to continuity of the log and $|\cdot|$ functions, it is enough to show tightness of elements of $\hat{\Sigma}(\vartheta)$ and their derivatives in the parameter ϑ. However, $\hat{\Sigma}(\vartheta) = \Pi A_T(\psi)\Pi'$ and thus application of Lemma A.2 in Johansen and Nielsen (2012) implies that it is sufficient to show tightness of all elements and their derivatives of the matrix $A_T(\psi)$ w.r.t. the variable ψ, i.e. tightness of $D_\psi^m(A_T(\psi))_{ij}$ in $\psi \in \mathbb{D}_\psi$ for $m = 0, 1, 2$, $\forall i, j = 1, \ldots, (k+1)(r+1)n$. The matrix $A_T(\psi)$ is composed of blocks of the form $S_{ijT}^{lp}(\psi) = T^{-1}\sum_{t=k}^T V_{lt-i}V'_{pt-j}$, $i, j = 0, \ldots, k$, $l, p = 0, \ldots, r$ where:

$$V_{0t}(\psi) = \Delta(\psi)R_0X_t = \Delta(\psi - \psi_0)\Psi(L)\varepsilon_t,$$
$$V_{it}(\psi) = N_i^{-1}\Delta_+^{\delta_i}R_0X_t = N_i^{-1}\Delta(\delta_i - \psi_0)\Psi(L)\varepsilon_t, i = 1, \ldots, r.$$

Expressions follow from the solution of the model (5.2): $X_t = R_0^{-1}\Delta(-\psi_0)\Psi(L)\varepsilon_t$. Individual elements of $S_{ijT}^{lp}(\psi)$ can be written as: $T^{-1}\sum_{t=k}^T v_{rt-i}v_{st-j}$, $r, s = 0, 1$, where:

$v_{1t}(u_1) = \Delta_+^{-u_1} z_t$ with $|u_1| < 1/2$, and $v_{2t}(u_2) = T^{-u_2+1/2}\Delta_+^{-u_2} z_t$ with $u_2 > 1/2$, where z_t is a component of the vector $\Psi(L)\varepsilon_t$, which is a stationary linear process with absolutely summable coefficients, hence Lemma A.4 in Johansen and Nielsen (2010) applies and tightness of elements of $D_\psi^m A_T(\psi)$ for $m = 0, 1, 2$ follows along with tightness of $D_\psi^m l_T(\vartheta)$.

The second part of the theorem follows easily from results in Johansen and Nielsen (2010): $\forall\psi \in \mathbb{D}_\psi$, elements of $S_{ijT}^{lp}(\psi)$ converge in distribution, hence $A_T(\psi) \xrightarrow{d} A_\infty(\psi)$ and it implies $l_T(\vartheta) \xrightarrow{d} l_\infty(\vartheta) = \Pi A_\infty(\psi)\Pi'$, $\forall\vartheta \in \Xi$. We may similarly prove convergence of finite-dimensional vectors $(l_T(\vartheta_1), \ldots, l_T(\vartheta_s))$ and given that $l_T(\vartheta_0)$ is $O_p(1)$, convergence $l_T(\vartheta) \Rightarrow l_\infty(\vartheta)$ follows from Theorem 7.3 in Billingsley (1968). □

Lemma 5.2. *The NLS estimator exists and is consistent:* $\hat\vartheta \xrightarrow{P} \vartheta_0$.

Proof. Firstly, we prove that $\vartheta = \vartheta_0$ minimizes $l_\infty(\vartheta)$ on Ξ_K. Denote O_ξ^S and O_ξ^N $nr \times nr$ diagonal selection matrices ($O_\xi^S + O_\xi^N = I_{nr}$), selecting variables in the vector $\xi_t(\psi)$ according to their stationarity[11]. Similarly denote O_W^S and O_W^N. Then:

$$\hat\Sigma(\vartheta) = \hat\Sigma_1(\vartheta) + B_T(\vartheta) = \hat\Sigma_1(\vartheta) + o_p(1),$$

where:

$$\hat\Sigma_1(\vartheta) = T^{-1}\Pi\sum_{t=k}^{T} O_W^S W_t W_t' O_W^S \Pi' + T^{-1}\Pi\sum_{t=k}^{T} O_W^N W_t W_t' O_W^N \Pi',$$
$$B_T(\vartheta) = \Pi T^{-1}\sum_{t=k}^{T}(O_W^S W_t W_t' O_W^N + O_W^N W_t W_t' O_W^S)\Pi'.$$

since $T^{-1}\sum_{t=k}^{T}(O_W^S W_t W_t' O_W^N + O_W^N W_t W_t' O_W^S) \Rightarrow 0$ on $\psi \in \mathbb{D}_\psi$ (cf. Johansen and Nielsen (2010), Lemma C.5). Also define: $\hat\Sigma_2(\vartheta) = T^{-1}\Pi\sum_{t=k}^{T} O_W^S W_t W_t' O_W^S \Pi'$ and $l_T^i(\vartheta) = \log\det(\hat\Sigma_i(\vartheta))$, $i = 1, 2$ with corresponding limits $l_\infty^i(\vartheta)$. Further, since the process:

$$\Pi O_W^S W_t = \Phi(L)\Delta(\psi - \psi_0)\Psi(L)\varepsilon_t + \Phi(L)D(\alpha)N(\psi)^{-1}O_\xi^S\xi_t(\psi) = f_\vartheta(L)\varepsilon_t$$

is asymptotically stationary, we have: $\hat\Sigma_2(\vartheta) \xrightarrow{P} \Sigma_2(\vartheta) = Cov(f_\vartheta(L)\varepsilon_t) \geq f_\vartheta(0)\Sigma_0 f_\vartheta(0)'$, with: $f_\vartheta(0) = I_n + D(\alpha)O_\xi^S(\mathbb{1}_r \otimes I_n)$. Since $D(\alpha)O_\xi^S(\mathbb{1}_r \otimes I_n)$ is an upper-triangular matrix with zeros on the diagonal, it holds: $\det(f_\vartheta(0)) = 1, \forall\vartheta \in \Xi_K$ and thus: $l_\infty^2(\vartheta) = \log\det(\Sigma_2(\vartheta)) \geq \log\det(\Sigma_0) = l_\infty^2(\vartheta_0)$. $\det(A + B) \geq \det A$, if A is p.d. and B is p.s.d.,

[11] Note, that the components of $\xi_t(\psi)$ are either (asymptotically) stationarity or nonstationary in the whole parameter space.

implies $l^1_T(\vartheta) \geq l^2_T(\vartheta)$. Thus the following inequality holds:

$$l_\infty(\vartheta) = l^1_\infty(\vartheta) \geq l^2_\infty(\vartheta) \geq l^2_\infty(\vartheta_0) = \log\det(\Sigma_0), \forall \vartheta \in \Xi_K.$$

For $\vartheta = \vartheta_0$: $f_\vartheta(L) = f_\vartheta(0) = I_n$, hence $l_\infty(\vartheta_0) = \log\det(\Sigma_0)$ and thus ϑ_0 minimizes $l_\infty(\vartheta)$. We show uniqueness of the minimizer: $l_\infty(\vartheta) = l^2_\infty(\vartheta)$ implies $D(\alpha)O^N_\xi = 0$, while $l^2_\infty(\vartheta) = l^2_\infty(\vartheta_0)$ implies $f_\vartheta(L) = f_\vartheta(0)$. That implies: $D(\alpha)O^S_\xi = 0$ and $\Phi(L)\Delta(\psi - \psi_0)\Psi(L) = \Phi(0)\Psi(0) = I_n$. Since $\Phi(L)$ and $\Psi(L)$ do not have unit roots, this implies: $\psi = \psi_0$ and $\Phi(L) = \Psi^{-1}(L) = \Phi_0(L)$. On the other hand, $0 = D(\alpha)O^S_\xi + D(\alpha)O^N_\xi = D(\alpha)$, implying $\alpha = \alpha_0 = 0$. Hence, $\vartheta = \vartheta_0$ and the minimizer is unique.

Next we show that the minimum ϑ_0 is well separated in Ξ_K, in a sense that $\forall \eta, \epsilon > 0$, $\exists \delta > 0$:

$$P\left(\inf_{\vartheta\in\Xi_K:||\vartheta-\vartheta_0||>\epsilon} l_\infty(\vartheta) < l_\infty(\vartheta_0) + \delta\right) < \eta.$$

The minimum is well separated in the directions of ψ, Φ and α^S: this follows from well-separation of $l^2_\infty(\vartheta)$ in the above directions. We show that the minimum is also well separated in the directions of α^N.

Since $\det(A+B) \geq \det(A) + \lambda_{\max}(B)\lambda^{n-1}_{\min}(A) + \det(B)$, if A is p.d. and B is p.s.d.[12], to show that $l_\infty(\vartheta)$ is well separated in the directions of α^N, it is enough to show that the limits of $\lambda_{\min}(\hat{\Sigma}_1(\vartheta))$ and $\lambda_{\max}(\hat{\Sigma}_1(\vartheta) - \hat{\Sigma}_2(\vartheta))$ are bounded from zero uniformly on $\forall \vartheta \in \Xi_K : ||\alpha^N|| > \epsilon > 0$. $\hat{\Sigma}_1(\vartheta)$ uniformly converges to variance-covariance matrix of (asymptotically) stationary process $f_\vartheta(L)\varepsilon_t$, hence the limit of $\lambda_{\min}(\hat{\Sigma}_1(\vartheta))$ is bounded from zero. Further we have:

$$\hat{\Sigma}_1(\vartheta) - \hat{\Sigma}_2(\vartheta) = C_T(\alpha,\psi) \xrightarrow{d} C_\infty(\alpha,\psi) = \Phi(1)D(\alpha)O^N_\xi A_\infty(\psi)O^N_\xi D'(\alpha)\Phi'(1).$$

Suppose $\lambda_{\max}(C_\infty(\alpha,\psi))$ is not bounded from zero on the set $\{\vartheta \in \Xi_K : ||\alpha^N|| > \epsilon > 0\}$. Then there is a sequence $\vartheta_k : ||\alpha^N_k|| \to \infty$ and $\lambda_{\max}(C_\infty(\alpha_k,\psi_k)) \to 0$. Then $aD(\alpha) = D(a\alpha), \forall a \in \mathbb{R}$ implies: $\lambda_{\max}(C_\infty(\alpha_k/||\alpha_k||,\psi_k)) \to 0$ and since the set $\{\vartheta : ||\alpha|| = 1\}$ is compact, it implies: $\exists \vartheta_\infty : \lambda_{\max}(C_\infty(\alpha_\infty,\psi_\infty)) = 0$, which is not true, since the p.s.d. matrix $C_\infty(\alpha,\psi)$ is not zero for any ϑ. Hence $\lambda_{\max}(C_\infty(\alpha,\psi))$ is also bounded from zero and thus $l_\infty(\vartheta)$ is well-separated in the directions of α^N.

---

[12]This follows from a more general inequality: $\det(A+B) \geq \prod_{i=1}^n(\lambda_i(A) + \lambda_i(B))$, if A, B are p.s.d. matrices.

Next we show that $\hat{\vartheta} \in \Xi_K$ with arbitrary high probability:

$$\forall \epsilon > 0, \exists T_\epsilon > 0 : P(\hat{\vartheta} \notin \Xi_K) < \epsilon, \forall T > T_\epsilon. \tag{5.34}$$

Let us consider a function on a set $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\} \times \mathbb{R}^+$:

$$f_T(\vartheta, \phi) = \log\det(\hat{\Sigma}(\vartheta, \phi)) = \log\det\left(\frac{1}{T}\sum_{t=k}^{T}(\hat{\varepsilon}_t^S + \phi\hat{\varepsilon}_t^N)(\hat{\varepsilon}_t^S + \phi\hat{\varepsilon}_t^N)'\right), \tag{5.35}$$

where:

$$\hat{\varepsilon}_t^S = \hat{\varepsilon}_t^S(\vartheta) = \Phi(L)w_{1t} + \Phi(L)D(\alpha)O_\xi^S w_{2t}, \tag{5.36}$$
$$\hat{\varepsilon}_t^N = \hat{\varepsilon}_t^N(\vartheta) = \Phi(L)D(\alpha)O_\xi^N w_{2t}. \tag{5.37}$$

This is a reparameterization of the objective function introducing a new parameter ϕ which equals $||\alpha^N||$ in the old parameterization, hence if $\hat{\vartheta}$ is the maximizer of $l_T(\vartheta)$, then it is easy to verify that $(\hat{\alpha}^S, \hat{\alpha}^N/||\hat{\alpha}^N||, \hat{\psi}, \hat{\Phi}, ||\hat{\alpha}^N||)$ is a maximizer of $f_T(\vartheta, \phi)$ and vice versa. Our goal is to prove that for the maximizer of $f_T(\vartheta, \phi)$ on the set $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\} \times \mathbb{R}^+$, it holds $\hat{\phi} < M$ for some $M \in \mathbb{R}$ with arbitrarily high probability, which implies[13] (5.34).

The function $f_T(\vartheta, \phi)$ can be seen as an objective function in the model: $\hat{\varepsilon}_t^S(\vartheta) + \phi\hat{\varepsilon}_t^N(\vartheta) = \varepsilon_t$ with true value $(\vartheta_0, 0)$. Denote $(\hat{\vartheta}, \hat{\phi}) = \arg\min_{(\vartheta,\phi)} f_T(\vartheta, \phi)$, then multivariate regression methods (see Lütkepohl (2005), section 5.2) give an explicit solution for $\hat{\phi}$:

$$\hat{\phi} = tr\left(\hat{\Sigma}^{-1}(\hat{\vartheta}, \hat{\phi})\frac{1}{T}\sum_{t=k}^{T}\hat{\varepsilon}_t^N(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta})\right) tr\left(\hat{\Sigma}^{-1}(\hat{\vartheta}, \hat{\phi})\frac{1}{T}\sum_{t=k}^{T}\hat{\varepsilon}_t^S(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta})\right). \tag{5.38}$$

From the inequality $tr(AB) \le \lambda_{max}(B)tr(A)$ for a p.s.d. matrix A and a symmetric matrix B we obtain:

$$\hat{\phi} \le tr^2\left(\hat{\Sigma}^{-1}(\hat{\vartheta}, \hat{\phi})\right)\lambda_{\max}\left(\frac{1}{T}\sum_{t=k}^{T}\hat{\varepsilon}_t^N(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta})\right)\lambda_{\max}\left(\frac{1}{T}\sum_{t=k}^{T}\hat{\varepsilon}_t^S(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta})\right). \tag{5.39}$$

Note:

$$B_T(\hat{\vartheta}) = \frac{1}{T}\sum_{t=k}^{T}\left(\hat{\varepsilon}_t^S(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta}) + \hat{\varepsilon}_t^S(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta})\right), \quad C_T(\hat{\vartheta}) = \frac{1}{T}\sum_{t=k}^{T}\hat{\varepsilon}_t^N(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta}). \tag{5.40}$$

---

[13]Since the choice of K is arbitrary, in case $K < M$ one might choose K bigger.

Since $B_T(\vartheta)$, $C_T(\vartheta)$ converge uniformly on the compact $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\}$ according to Lemma 5.1 and the limits $||B_\infty(\vartheta)||$, $||C_\infty(\vartheta)||$ are bounded from above almost surely, the maximum eigenvalues of the sample moments of $\hat{\varepsilon}_t^N(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta})$ and $\hat{\varepsilon}_t^S(\hat{\vartheta})\hat{\varepsilon}_t^{N\prime}(\hat{\vartheta})$ are bounded with arbitrarily high probability.

$tr(\hat{\Sigma}^{-1}(\vartheta, \phi))$ is bounded from above, if the maximal element on the diagonal of the matrix is bounded, which is true, if $\widehat{Var}(\hat{\varepsilon}_{it}(\vartheta, \phi)|\hat{\varepsilon}_{jt}(\vartheta, \phi), j = 1, \ldots, n\backslash\{i\}) > \epsilon > 0$, $\forall i = 1, \ldots, n$ with arbitrarily high probability uniformly on $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\} \times \mathbb{R}^+$. Suppose the contrary, i.e. $\widehat{Var}(\hat{\varepsilon}_{it}(\vartheta, \phi)|\hat{\varepsilon}_{jt}(\vartheta, \phi), j = 1, \ldots, n\backslash\{i\})$ is not uniformly bounded from below. Then for some i^* there is a sequence (ϑ_k, ϕ_k), such that it holds:

$$\widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta_k) + \phi_k \hat{\varepsilon}_{i^*t}^N(\vartheta_k)|\hat{\varepsilon}_{jt}(\vartheta_k, \phi_k), j = 1, \ldots, n\backslash\{i^*\}) \xrightarrow{P} 0.$$

Since $\{\vartheta_k\} \subset \Xi_K$, either there exists a subsequence k_n: $\vartheta_{k_n} \to \vartheta_\infty \in \Xi_K$, $\phi_{k_n} \to \infty$:

$$\widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta_\infty) + \phi_{k_n} \hat{\varepsilon}_{i^*t}^N(\vartheta_\infty)|\hat{\varepsilon}_{jt}(\vartheta_\infty, \phi_{k_n}), j = 1, \ldots, n\backslash\{i^*\}) \xrightarrow{P} 0,$$

or there exists a subsequence k_n: $\vartheta_{k_n} \to \vartheta_\infty \in \Xi_K$, $\phi_{k_n} \to \phi_\infty$:

$$\widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta_\infty) + \phi_\infty \hat{\varepsilon}_{i^*t}^N(\vartheta_\infty)|\hat{\varepsilon}_{jt}(\vartheta_\infty, \phi_\infty), j = 1, \ldots, n\backslash\{i^*\}) \xrightarrow{P} 0.$$

In any case this is not possible, if $\widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta)|\hat{\varepsilon}_{jt}(\vartheta, \phi), j = 1, \ldots, n\backslash\{i^*\}) > \epsilon > 0$ uniformly on $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\} \times \mathbb{R}^+$. One can verify that it holds: $\widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta)|\hat{\varepsilon}_{jt}^N(\vartheta), j = 1, \ldots, n\backslash\{i^*\}) = \widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta)) + o_p(1)$ and $\widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta)|\hat{\varepsilon}_{jt}^S(\vartheta), j = 1, \ldots, n\backslash\{i^*\}) > \epsilon$, for some $\epsilon > 0$ with arbitrarily high probability uniformly on $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\} \times \mathbb{R}^+$. The latter holds since $\det(\Sigma_2(\vartheta)) \geq \det(\Sigma_2(\vartheta_0))$. Hence $\widehat{Var}(\hat{\varepsilon}_{i^*t}^S(\vartheta)|\hat{\varepsilon}_{jt}(\vartheta, \phi), j = 1, \ldots, n\backslash\{i^*\}) > \epsilon$ uniformly on $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\} \times \mathbb{R}^+$ and thus boundedness from above of $tr(\hat{\Sigma}^{-1}(\vartheta, \phi))$ with arbitrarily high probability on the set $\{\vartheta \in \Xi_K : ||\alpha^N|| = 1\} \times \mathbb{R}^+$ follows, which implies $\hat{\phi} < M$ for some $M < \infty$ with arbitrarily high probability and hence (5.34) is proved.

Next, we show consistency of $\hat{\vartheta}$. Choose $\eta, \epsilon > 0$. Then $\exists T_1 > 0 : P(\hat{\vartheta} \notin \Xi_K) < \eta/3, \forall T > T_1$. Further, well-separation implies $\exists \delta(\eta, \epsilon) > 0$:

$$P\left(\inf_{\vartheta \in \Xi_K : ||\vartheta - \vartheta_0|| > \epsilon} l_\infty(\vartheta) < l_\infty(\vartheta_0) + \delta\right) < \eta/3.$$

Also, since $l_T(\vartheta)$ is uniformly convergent on Ξ_K, $\exists T_2$:

$$P\left(\sup_{\vartheta\in\Xi_K}|l_T(\vartheta)-l_\infty(\vartheta)|>\delta/3\right)<\eta/3,\ \ \forall T>T_2.$$

Thus $\forall T>\max\{T_1,T_2\}$:

$$P(||\hat\vartheta-\vartheta_0||>\epsilon)<P\left(\inf_{\vartheta\in\Xi_K:||\vartheta-\vartheta_0||>\epsilon}l_\infty(\vartheta)<l_\infty(\vartheta_0)+\delta\right)+P\left(\sup_{\vartheta\in\Xi_K}|l_T(\vartheta)-l_\infty(\vartheta)|>\delta/3\right)$$
$$+P\left(\hat\vartheta\in\Xi_K,||\hat\vartheta-\vartheta_0||>\epsilon,\inf_{\vartheta\in\Xi_K:||\vartheta-\vartheta_0||>\epsilon}l_\infty(\vartheta)\geq l_\infty(\vartheta_0)+\delta,\sup_{\vartheta\in\Xi_K}|l_T(\vartheta)-l_\infty(\vartheta)|\leq\delta/3\right)$$
$$+P(\hat\vartheta\notin\Xi_K)\leq\eta+P\left(\vartheta\in\Xi_K,l_\infty(\hat\vartheta)\geq l_\infty(\vartheta_0)+\delta,\sup_{\vartheta\in\Xi_K}|l_T(\vartheta)-l_\infty(\vartheta)|\leq\delta/3\right)=\eta,$$

because probability of the latter event is zero, since both events imply:

$$l_T(\hat\vartheta)-l_T(\vartheta_0)=l_T(\hat\vartheta)-l_\infty(\hat\vartheta)+l_\infty(\hat\vartheta)-l_\infty(\vartheta_0)+l_\infty(\vartheta_0)-l_T(\vartheta_0)\geq\delta-\frac{2\delta}{3}=\frac{\delta}{3}>0,$$

while $\hat\vartheta$ being the minimizer of $l_T(\vartheta)$. Hence $\hat\vartheta$ is consistent for ϑ_0 and the lemma is proved. □

Lemma 5.3. *Denote S^S_α, S^N_α - full row rank selection matrices selecting stationary and non-stationary coordinates of the vector $L_{n,r}\xi_t(\psi_0)$, respectively. Then the following convergence holds:*

$$T^{-1/2}\sum_{t=k}^{T}S^N_\alpha D_\alpha\hat\varepsilon'_t(\vartheta_0)\Sigma_0^{-1}\varepsilon_t\xrightarrow{d}\int_0^1 S^N_\alpha G_\psi(u)\Phi'_0(1)\Sigma_0^{-1}dW,\tag{5.41}$$

where $G_\psi(u)=(G^1(u)',\ldots,G^r(u)')'$ with $G^i(u)=\iota_i^{n\prime}\otimes\zeta_i(u)$, $\zeta_i(u)$ is a $r-i$ vector with its j-th coordinate equal to $\mathbb{1}_{\delta_{j0}-\delta_{i0}>1/2}\iota_{1j}^{1\times n}\Psi(1)W_{\delta_{j0}-\delta_{i0}-1}(u)$. Alternatively, using definitions of Section 5.3, it can be written as: $G'_\psi(u)=C(L_{n,r}\zeta(u))$, where $\zeta(u)$ is a weak limit of $O^N_\xi N^{-1}(\psi_0)\xi_{[Tu]}(\psi_0)$.

Proof. Let us take a typical row of a matrix $D_\alpha\hat\varepsilon'_t(\vartheta_0)$ representing a *superconsistent* component α_{ij}, which is the j-th coordinate of the vector α_i:

$$D_{\alpha_{ij}}\hat\varepsilon_t(\vartheta_0)=\Phi_0(L)\iota^{n\times nr}_{i(in-n+j+1)}N(\psi_0)^{-1}\xi_t(\psi_0)=T^{\delta_{i0}-\delta_{j0}+1/2}\Phi_0(L)\iota^{n\times n}_{ij}\Delta_+^{\delta_{i0}-\delta_{j0}}u_t.$$

Then applying the fractional invariance principle (5.3):

$$D_{\alpha_{ij}}\hat\varepsilon_{[Tu]}(\vartheta_0)\Rightarrow\Phi_0(1)\iota^{n\times n}_{ij}\Psi(1)W_{\delta_{j0}-\delta_{i0}-1}(u).$$

In addition, we have joint convergence:

$$(D_{\alpha_{ij}}\hat{\varepsilon}_{[Tu]}(\vartheta_0), T^{-1/2}\Delta_+^{-1}\varepsilon_{[Tu]}) \Rightarrow (\Phi_0(1)\iota_{ij}^{n\times n}\Psi(1)W_{\delta_{j0}-\delta_{i0}-1}(u), W(u)),$$

hence Theorem 2.2 in Kurtz and Protter (1991) implies:

$$T^{-1/2}\sum_{t=k}^{T} D_{\alpha_{ij}}\hat{\varepsilon}_t'(\vartheta_0)\Sigma_0^{-1}\varepsilon_t \xrightarrow{d} \int_0^1 W_{\delta_{j0}-\delta_{i0}-1}'\Psi'(1)\iota_{ji}^{n\times n}\Phi_0'(1)\Sigma_0^{-1}dW. \tag{5.42}$$

With some matrix algebra we may express the whole matrix $D_\alpha\hat{\varepsilon}_t(\vartheta_0)$ as follows:

$$D_\alpha\hat{\varepsilon}_t(\vartheta_0) = \Phi_0(L)C(L_{n,r}N^{-1}(\psi_0)\xi_t(\psi_0)).$$

Then weak convergence of $S_\alpha^N N^{-1}(\psi_0)\xi_{[Tu]}(\psi_0)$ implies convergence: $S_\alpha^N D_\alpha\hat{\varepsilon}_{[Tu]}'(\vartheta_0) \Rightarrow S_\alpha^N G_\psi(u)\Phi_0(1)'$. Finally, by similar argument to the above, we have:

$$T^{-1/2}\sum_{t=k}^{T} S_\alpha^N D_\alpha\hat{\varepsilon}_t'(\vartheta_0)\Sigma_0^{-1}\varepsilon_t \xrightarrow{d} \int_0^1 S_\alpha^N G_\psi(u)\Phi_0'(1)\Sigma_0^{-1}dW. \tag{5.43}$$

□

Lemma 5.4. *It holds:*

$$\frac{1}{T}\sum_{t=k}^{T} S_\alpha^N D_\alpha\hat{\varepsilon}_t'(\vartheta_0)\Sigma_0^{-1}D_\tau\hat{\varepsilon}_t(\vartheta_0) \xrightarrow{P} 0,$$
$$\frac{1}{T}\sum_{t=k}^{T} S_\alpha^N D_\alpha\hat{\varepsilon}_t'(\vartheta_0)\Sigma_0^{-1}D_\alpha\hat{\varepsilon}_t(\vartheta_0)S_\alpha^{S\prime} \xrightarrow{P} 0,$$
$$\frac{1}{T}\sum_{t=k}^{T}\begin{pmatrix} S_\alpha^S D_\alpha\hat{\varepsilon}_t'(\vartheta_0) \\ D_\tau\hat{\varepsilon}_t'(\vartheta_0)\end{pmatrix}\Sigma_0^{-1}(D_\alpha\hat{\varepsilon}_t(\vartheta_0)S_\alpha^{S\prime}, D_\tau\hat{\varepsilon}_t(\vartheta_0)) \xrightarrow{P} \Gamma,$$

where Γ is a positive definite matrix.

Proof. $D_\tau\hat{\varepsilon}_t(\vartheta_0)$, $S_\alpha^S D_\alpha\hat{\varepsilon}_t'(\vartheta_0)$ are matrices with components which are asymptotically stationary processes, whereas $S_\alpha^N D_\alpha\hat{\varepsilon}_t'(\vartheta_0)$ has only non-stationary components and hence the first two limits follow from Lemma C.5 in Johansen and Nielsen (2010). The components of $S_\alpha^S D_\alpha\hat{\varepsilon}_t'(\vartheta_0)$ and $D_\tau\hat{\varepsilon}_t'(\vartheta_0)$ are asymptotically stationary processes, hence their product moments converge to their corresponding limits and thus the third convergence follows. We show that the limit matrix Γ is positive definite. Γ is singular, if there exists a non-zero vector $\lambda \in \mathbb{R}^{\bar{k}}$: $\lim_{t\to\infty} Cov(\lambda' D_\vartheta\hat{\varepsilon}_t'(\vartheta_0)) = 0$, or: $\lim_{t\to\infty}\sum_i \lambda_i\partial_{\vartheta_i}\hat{\varepsilon}_t(\vartheta_0) = 0$.

This translates to:

$$\lambda(L)u_t + D(\lambda)N^{-1}(\psi_0)\xi_t(\psi_0) + \ln\Delta\Phi_0(L)diag(\lambda_1,\ldots,\lambda_r,\lambda_{r+1},\ldots,\lambda_{r+1})u_t = f_{\lambda,t}(L)\varepsilon_t \xrightarrow{P} 0.$$

Here $\lambda(L)$ is a $VAR(k)$ matrix lag polynomial with unrestricted λ's as matrix coefficients. That implies $f_{\lambda,t}(0) = 0$, and thus $D(\lambda) = 0$. From here:

$$\lambda(L) + \ln\Delta\Phi_0(L)diag(\lambda_1,\ldots,\lambda_r,\lambda_{r+1},\ldots,\lambda_{r+1}) = 0.$$

However, $\ln\Delta\Phi_0(L)$ is an infinite lag polynomial, while $\lambda(L)$ is finite and hence $\lambda_1 = \ldots = \lambda_{r+1} = 0$. From here: $\lambda(L) = 0$, hence a non-zero linear combination such that $\lambda' D_\vartheta \hat{\varepsilon}'_t(\vartheta_0) \xrightarrow{P} 0$ does not exist and thus the matrix Γ is positive definite. □

Lemma 5.5. *Define $S^N_\vartheta, S^S_\vartheta, S^\tau_\vartheta$ - full row rank selection matrices selecting stationary and non-stationary components of $D_\alpha\hat{\varepsilon}_t(\vartheta_0)$ and components of $D_\tau\hat{\varepsilon}_t(\vartheta_0)$ from the vector $D_\vartheta\hat{\varepsilon}_t(\vartheta_0)$, respectively. Then it holds:*

$$\sqrt{T}\begin{pmatrix} S^N_\vartheta \\ S^S_\vartheta \\ S^\tau_\vartheta \end{pmatrix} \nabla_T(\vartheta_0) \xrightarrow{d} \begin{pmatrix} \int_0^1 S^N_\alpha G_\psi(u)\Phi'_0(1)\Sigma_0^{-1}dW \\ N(0,\Gamma) \end{pmatrix}, \tag{5.44}$$

where $G_\psi(u)$ is defined in Lemma 5.3 and matrix Γ is defined in Lemma 5.4.

Proof. Note:

$$\sqrt{T}\begin{pmatrix} S^N_\vartheta \\ S^S_\vartheta \\ S^\tau_\vartheta \end{pmatrix} \nabla_T(\vartheta_0) = \begin{pmatrix} T^{-1/2}\sum_{t=k}^T S^N_\alpha D_\alpha\hat{\varepsilon}'_t(\vartheta_0)\hat{\Sigma}^{-1}\varepsilon_t \\ T^{-1/2}\sum_{t=k}^T S^S_\alpha D_\alpha\hat{\varepsilon}'_t(\vartheta_0)\hat{\Sigma}^{-1}\varepsilon_t \\ T^{-1/2}\sum_{t=k}^T D_\tau\hat{\varepsilon}'_t(\vartheta_0)\hat{\Sigma}^{-1}\varepsilon_t \end{pmatrix}. \tag{5.45}$$

The convergence of the term $T^{-1/2}\sum_{t=k}^T S^N_\alpha D_\alpha\hat{\varepsilon}'_t(\vartheta_0)\Sigma_0^{-1}\varepsilon_t$ follows from Lemma 5.3. We prove convergence for other components. Denote:

$$\begin{pmatrix} T^{-1/2}\sum_{t=k}^T S^S_\alpha D_\alpha\hat{\varepsilon}'_t(\vartheta_0)\Sigma_0^{-1}\varepsilon_t \\ T^{-1/2}\sum_{t=k}^T D_\tau\hat{\varepsilon}'_t(\vartheta_0)\Sigma_0^{-1}\varepsilon_t \end{pmatrix} = T^{-1/2}\sum_{t=k}^T \begin{pmatrix} S^S_\alpha D_\alpha\hat{\varepsilon}'_t(\vartheta_0) \\ D_\tau\hat{\varepsilon}'_t(\vartheta_0) \end{pmatrix} \Sigma_0^{-1}\varepsilon_t = T^{-1/2}\sum_{t=k}^T V_t. \tag{5.46}$$

The random vector series V_t is an asymptotically stationary martingale difference series and the following convergence follows from Lemma 5.4:

$$T^{-1}\sum_{t=k}^{T} E(V_t V_t'|\mathcal{F}_t) = \frac{1}{T}\sum_{t=k}^{T}\begin{pmatrix} S_\alpha^S D_\alpha \hat{\varepsilon}_t'(\vartheta_0) \\ D_\tau \hat{\varepsilon}_t'(\vartheta_0)\end{pmatrix}\Sigma_0^{-1}(D_\alpha\hat{\varepsilon}_t(\vartheta_0)S_\alpha^{S\prime}, D_\tau\hat{\varepsilon}_t(\vartheta_0)) \xrightarrow{P} \Gamma.$$

Here $\mathcal{F}_t$ is the σ-field generated by $\{\varepsilon_j, j \leq t\}$. Now we may apply a central limit theorem for martingale differences series V_t (Corrolary 3.1, Hall and Heyde (1980)) to get the convergence: $T^{-1/2}\sum_{t=k}^{T} V_t \xrightarrow{d} N(0,\Gamma)$. Finally, since the random vector $T^{-1/2}\sum_{t=k}^{T} D_\vartheta \hat{\varepsilon}_t'(\vartheta_0)\Sigma_0^{-1}\varepsilon_t$ is tight in the parameter Σ^{-1} and $\hat{\Sigma}^{-1} \xrightarrow{P} \Sigma_0^{-1}$, we get the convergence:

$$\sqrt{T}\begin{pmatrix} S_\vartheta^N \\ S_\vartheta^S \\ S_\vartheta^\tau \end{pmatrix}\nabla_T(\vartheta_0) \xrightarrow{d} \begin{pmatrix} \int_0^1 S_\vartheta^N G_\psi(u)\Phi_0'(1)\Sigma_0^{-1} dW \\ N(0,\Gamma)\end{pmatrix}. \tag{5.47}$$

□

Lemma 5.6. *It holds:*

$$\begin{pmatrix} S_\vartheta^N \\ S_\vartheta^S \\ S_\vartheta^\tau \end{pmatrix}\mathcal{H}_T(\vartheta_0)\left(S_\vartheta^N, S_\vartheta^S, S_\vartheta^\tau\right) \xrightarrow{d} \begin{pmatrix} \int_0^1 S_\alpha^N G_\psi(u)\Phi_0'(1)\Sigma_0^{-1}\Phi_0(1)G_\psi'(u)S\alpha^{N\prime}du & 0 \\ 0 & \Gamma \end{pmatrix}.$$

Proof. Let us take a typical ij-th element of the Hessian at the point $\vartheta = \vartheta_0$:

$$2(\mathcal{H}_T(\vartheta_0))_{ij} = \partial^2_{\vartheta_i\vartheta_j} l_T(\vartheta_0) = \frac{1}{T}\sum_{t=k}^{T} D_{\vartheta_i}\hat{\varepsilon}_t'(\vartheta_0)\hat{\Sigma}^{-1}D_{\vartheta_j}\hat{\varepsilon}_t(\vartheta_0) + \frac{1}{T}\sum_{t=k}^{T} D_{\vartheta_j}\hat{\varepsilon}_t'(\vartheta_0)\hat{\Sigma}^{-1}D_{\vartheta_i}\hat{\varepsilon}_t(\vartheta_0)$$

$$+\frac{2}{T}\sum_{t=k}^{T} D_{\vartheta_i}\hat{\varepsilon}_t'(\vartheta_0)\hat{\Sigma}^{-1}D_{\vartheta_j}\hat{\Sigma}(\vartheta_0)\hat{\Sigma}^{-1}\varepsilon_t + \frac{2}{T}\sum_{t=k}^{T} D^2_{\vartheta_i\vartheta_j}\hat{\varepsilon}_t'(\vartheta_0)\hat{\Sigma}^{-1}\varepsilon_t.$$

The last two terms are of order $o_p(1)$, as follows from the bounds:

$$\hat{\Sigma}(\vartheta_0) = O_p(1),$$
$$\hat{\Sigma}^{-1}(\vartheta_0) = O_p(1),$$
$$T^{-1}\sum_{t=k}^{T} D_{\vartheta_i}\hat{\varepsilon}_t(\vartheta_0)\varepsilon_t' = o_p(1),$$
$$T^{-1}\sum_{t=k}^{T} D^2_{\vartheta_i\vartheta_j}\hat{\varepsilon}_t(\vartheta_0)\varepsilon_t' = o_p(1),$$
$$D_{\vartheta_i}\hat{\Sigma}(\vartheta_0) = O_p(1),$$

for every i, j. The first two bounds are obvious, third bound follows from Lemma 5.5. Similarly the fourth bound follows, while the fifth bound follows from the third.

As we have seen in Lemma 5.3: $S_\alpha^N D_\alpha \hat{\varepsilon}'_{[Tu]}(\vartheta_0) \Rightarrow S_\alpha^N G_\psi(u)\Phi_0(1)'$. Then the continuous mapping theorem implies:

$$\frac{1}{T}\sum_{t=k}^{T} S_\alpha^N D_\alpha \hat{\varepsilon}_t'(\vartheta_0)\Sigma_0^{-1} D_\alpha \hat{\varepsilon}_t(\vartheta_0) S_\alpha^{N\prime} \xrightarrow{d} \int_0^1 S_\alpha^N G_\psi(u)\Phi_0'(1)\Sigma_0^{-1}\Phi_0(1)G_\psi'(u)S_\alpha^{N\prime} du. \quad (5.48)$$

Lemma 5.4 gives limits of the off-diagonal blocks of the matrix and the proof is completed. □

5.7.3 Proofs of main theorems

Proof of Theorem 5.1. For the NLS estimator $\hat{\vartheta}$ it holds: $\nabla_T(\hat{\vartheta}) = 0$ and a Taylor expansion for the gradient gives: $0 = \nabla_T(\vartheta_0) + \mathcal{H}_T(\tilde{\vartheta})(\hat{\vartheta} - \vartheta_0)$ for some $||\tilde{\vartheta} - \vartheta_0|| \leq ||\hat{\vartheta} - \vartheta_0||$. Due to tightness of $\mathcal{H}_T(\vartheta)$ and consistency of $\hat{\vartheta}$ (Lemmas 5.1, 5.2), we have: $||\mathcal{H}_T(\tilde{\vartheta}) - \mathcal{H}_T(\vartheta_0)|| \xrightarrow{P} 0$ (see Lemma A.3 in Johansen and Nielsen (2012)). The limit of $\left(S_\vartheta^{N\prime}, S_\vartheta^{S\prime}, S_\vartheta^{\tau\prime}\right)' \mathcal{H}_T(\vartheta_0)\left(S_\vartheta^{N\prime}, S_\vartheta^{S\prime}, S_\vartheta^{\tau\prime}\right)$ and its invertibility is given in Lemmas 5.4 and 5.6. Then upon noting that $S^{-1\prime} = S = \left(S_\vartheta^{N\prime}, S_\vartheta^{S\prime}, S_\vartheta^{\tau\prime}\right)'$ for a selection matrix, Lemma 5.5 gives the asymptotic distribution of $\hat{\vartheta}$:

$$\sqrt{T}\begin{pmatrix} S_\vartheta^N \\ S_\vartheta^S \\ S_\vartheta^\tau \end{pmatrix}(\hat{\vartheta} - \vartheta_0) \xrightarrow{d} \begin{pmatrix} \left(\int_0^1 H(u)H(u)'du\right)^{-1}\int_0^1 H(u)\Sigma_0^{-1/2}dW(u) \\ N(0,\Gamma) \end{pmatrix}, \quad (5.49)$$

where $H(u) = S_\alpha^N G_\psi(u)\Phi_0(1)'\Sigma_0^{-1/2}$. Finally, the last step is to recover the asymptotic distribution of the parameters in terms of the initial parameterization. Since $\alpha =$

$M(\psi)P^{-1\prime}(\beta-\beta_0)$, where $M(\psi) = blockdiag(N_{11},\dots,N_{rr}),\ P = blockdiag(R_{01},\dots,R_{0r})$ hence:

$$\hat{\vartheta}-\vartheta_0 = \begin{pmatrix}\hat{\alpha}\\ \hat{\tau}-\tau_0\end{pmatrix} = \begin{pmatrix}M(\hat{\psi})P^{-1\prime} & 0\\ 0 & I_{kn^2+r+1}\end{pmatrix}\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\tau}-\tau_0\end{pmatrix}.$$

Since $\hat{\psi}-\psi_0 = O_p(T^{-1/2})$, we can substitute ψ_0 for $\hat{\psi}$ in $M(\hat{\psi})$ and we have:

$$\sqrt{T}\begin{pmatrix}\begin{pmatrix}S_\alpha^N\\ S_\alpha^S\end{pmatrix}M(\psi_0)P^{-1\prime} & 0\\ 0 & I_{kn^2+r+1}\end{pmatrix}(\hat{\theta}-\theta_0)\xrightarrow{d}\begin{pmatrix}\left(\int_0^1 H(u)H(u)'du\right)^{-1}\int_0^1 H(u)\Sigma_0^{-1/2}dW(u)\\ N(0,\Gamma)\end{pmatrix}. \quad (5.50)$$

□