Theory and Applications in non-linear Cointegrated VAR Models

Emil Nejstgaard

September 2014
Department of Economics
University of Copenhagen
Øster Farimagsgade 5, building 26
DK-1353 Copenhagen K.
www.econ.ku.dk

Contents

Summary

In this thesis, we consider theory and applications in non-linear cointegration, in particular within the dynamic mixture cointegrated VAR framework called the Autoregressive Conditional Root (ACR) cointegrated model. These and similar models have found applications to many different data series within macroeconomics and finance. The thesis is comprised of four chapters, where the chapters one through three are concerned with the ACR model framework and the fourth discusses a parameter identification problem seen in the ACR model, but also in other similar non-linear autoregressive models.

In chapter one, we present the ACR cointegrated framework, including a number of novel extensions, namely a constant in the cointegration relations, allowing for multiple regimes and letting the covariance matrices of the error terms be regime dependent. We provide a representation theory for the process and give conditions for stationarity. We consider likelihood-based inference on all parameters, when the cointegration relations are estimated and show that, contrary to the case when the cointegration relations are fixed from the outset, asymptotic theory is non-standard. More precisely, the asymptotic distribution of the maximum likelihood estimator is a nuisance parameter depend function of Brownian motions.

In chapter two, we introduce an encompassing framework based on generalized linear restrictions that facilitates estimation of many differently specified ACR cointegrated models. An EM algorithm for estimating the parameters under these restrictions is presented and its performance is evaluated through a small Monte Carlo simulation study. We further discuss testing based on likelihood ratio statistics for two separate cases; a *regular* and an *irregular* case. The irregular case refers to statistics where nuisance parameters are unidentified under the null hypothesis, while the regular case refers to tests where no such problems occur. We show for the regular case that the asymptotic theory developed in chapter one can be applied give convergence in distribution of the likelihood ratio test. In the irregular case, a uniform central limit theory is needed. While such theory is not provided here, we conjecture that it exists as it does other, similar models. Given the non-standard asymptotic distributions, simulation based techniques are required for inference and we propose a bootstrap algorithm to simulate the distributions of the test statistics. The performance of the bootstrap algorithm is investigated through simulations.

Chapter three considers an application of the ACR cointegrated framework to the prices of two of the major crude oil benchmarks, the West Texas Intermediate (WTI) and the Brent. Moreover, the chapter discusses an alteration of the ACR cointegrated model. That is, in chapter one, the asymptotic theory is derived under the assumption that the constant in the cointegration relations is not included in the switching probability functions. Chapter three shows that with a few modifications to the theory of chapter one, such a specification is indeed easily covered. The results from the empirical analysis supports the presence of non-linearities related to a decoupling of the WTI from historical benchmarks observed around 2011. Evidence in favor of non-linearities is less pronounces when this period is excluded from the sample and we use a linear cointegrated VAR to find that the WTI historically and until 2011 has been weakly exogenous.

Chapter four focuses on a parameter identification problem that arises in the ACR framework as well as in other models such as the smooth transition autoregressive models. We discuss the

origin of this problem, which turns out to be linked to the fact that both certain specifications of the ACR model and the logistic smooth transition autoregressive (LSTAR) model, approximates the Threshold Autoregressive (TAR) model as the investigated parameter diverges. This has as a consequence that the likelihood becomes flat in some directions and rippled in others, making numerical optimization tedious. We propose a reparametrization that facilitates numerical analysis. Moreover, we discuss information criteria as selection tools between the LSTAR and TAR models and show by simulations that these consistently select the correct model.

The thesis also points to some areas of interest for future research within these frameworks. In particular, asymptotic theory is needed for the likelihood ratio test in the irregular case, discussed in chapter two. Likewise, asymptotic theory for the bootstrap algorithm proposed in chapter two needs developing and more simulation studies could be considered to better understand small sample performances. Chapter three makes obvious that the ACR cointegrated models lacks adapted misspecification tests. Finally, the model-selection based on information criteria discussed in chapter four was not verified theoretically, since it is unclear how the likelihood ratio statistic behaves when the true data generating process is a TAR model. Investigating this issue further with the goal of developing a proper test statistic for selection between smooth and non-smooth regime switching models is clearly of interest.

Resumé

I denne afhandling behandles teori og anvendelser af modeller for ikke-lineær kointegration, særligt såkaldte *Autoregressive Conditional Root* (ACR) modeller. Disse og lignende modeller er blevet anvendt bredt til analyse af forskellige tidsrækker indenfor makroøkonomi og makofinansiering. Afhandlingen består af fire kapitler, hvoraf tre kapitler beskæftiger sig med ACR-modelrammen og det fjerde diskuterer et parameteridentifikationsproblem som ses i ACR-modellen, men også i andre lignende ikke-lineære modeller.

I kapitel ét præsenteres ACR-kointegrationsmodellen med en række nye udvidelser. Mere præcist bliver teorien for modellen udvidet til at tillade for et konstantled i kointegrationsrelationerne, modellen udvides til at kunne indeholde flere regimer og der tallades for at fejlledskovariansmatricerne kan være regimeafhængige. Vi giver en repræsentationsteori for processen og opsætter betingelser for stationaritet. Ydermere udvikler vi den asymptotiske teori for det tilfælde hvor kointegrationsparametrene er estimeret og viser at, i modsætning til tilfældet hvor disse ikke er estimeret, fås asymptotisk teori der ikke kan baseres på approximationer med χ^2-fordelingen. Mere præcist vises det, at de asymptotisk fordelinger er funktioner af brownske bevælgelser, der afhænger af de estimerede parametre.

I kapitel to introduceres en metode, baseret på generaliserede lineære restriktioner, som letter estimering og specificering ACR cointegrerede modeller. En *EM*-algoritme til at estimere parametrene under disse restriktioner udledes og dennes egenskaber analyseres gennem et mindre Monte Carlo simulationsstudie. Vi diskuterer yderligere kvotienttest for to separate tilfælde: et regulært tilfælde og et irregulært tilfælde. Det irregulære tilfælde refererer til tests, hvor såkaldte nuissanceparametre ikke er identificeret under nulhypotesen, mens det regulære tilfælde refererer til test hvor der ikke opstår sådanne problemer. Vi viser for det regulære tilfældet, at den asymptotiske teori udviklet i kapitel ét, kan anvendes og giver konvergens i fordeling af kvotienttestet. I det irregulære tilfælde er der behov for en uniform central grænseværdisætning som ikke er dækket af teorien fra kapitel ét. Vi formoder, at en sådan teori eksisterer som det er tilfældet for andre, nært beslægtede modeller. Da de asymptotiske fordelinger afhænger af de estimerede parametre, er simulationsbaserede metoder nødvendige. Vi foreslår en bootstrapalgorithme til brug for estimering af de relevante fordelinger og undersøger algoritmens egenskaber gennem simuleringer.

Kapitel tre kigger på en anvendelse af ACR-kointegrationsmodellen på to centrale råoliepriser, mere præcist prisen på West Texas Intermediate (WTI) og prisen på Brent. Desuden diskuterer kapitlet en ændring af ACR-kointegrationsmodellen som den blev præsenteret i kapitel ét, nemlig at regimeskiftsandsynlighederne inddrager konstanten i kointegrationsrelationerne. Vi verificerer at den asymptotiske teori udledt under denne ændring, med små justeringer svarer til den der blev udledt i kapitel ét. Resultaterne fra den empiriske analyse støtter ikke-lineare effekter relateret til en afkobling af WTI fra sit historiske benchmark niveau relativt til andre råoliepriser, som skete i begyndelsen af 2011. Omvendt er indikationerne til fordel for ikke-lineariteter mindre udtalte når modellen estimeres på en stikprøve der udelukker denne periode. Vi finder gennem estimering af en lineær kointegreret VAR for at finde, at WTI historisk har været svagt eksogen op til 2011.

Kapitel fire fokuserer på et parameteridentifcationsproblem, som forkommer i ACR-kointegrationsmodeller

samt i andre modeller såsom STAR modeller. Vi diskuterer oprindelsen af dette problem, som viser sig at være tilknyttet til, at begge modeller approksimerer en såkaldt Threshold Autoregression (TAR), når den undersøgte parameter divergerer. Dette har som konsekvens, at likelihoodfunktionen bliver flad i nogle retninger og rillet i andre, hvilket besværliggør numerisk optimering. Vi foreslår en omparametrisering der letter de numeriske aspekter. Derudover diskuterer vi informationskriterier som udvælgelsesmetode mellem LSTAR og TAR modeller, og vi viser ved simuleringer, at disse vælger den rigtige model når antallet af observations er stort.

Afhandlingen peger også på nogle områder der kunne have interesse for den videre forskning på dette område. Specielt er der behov for asymptotisk teori for kvotienttests i det irregulære tilfælde der behandles i kapitel to. Ligeledes vil det være af interesse at udvikle asymptotisk teori for den bootstrapalgoritme som blev foreslået i kapitel to. Kapitel tre gør klart, at ACR-kointegrationsmodellen mangler misspecifikationstests. Endelig er modeludvælgelsen baseret på informationskriterier diskuteret i kapitel fire, er ikke verificeret teoretisk, da det er uklart, hvordan kvotienttestet (og dermed også informationskriterierne) opfører sig, når den sande, datagenererende proces er en TAR model. At undersøge dette spørgsmål yderligere med det formål at udvikle et test for valg mellem glatte og aprupte regimeskiftmodeller er klart af interesse.

Acknowledgments

I would thank my thesis supervisor, Anders Rahbek, for helping and pushing me to get this far.

Also, discussions with Søren Johansen, Heino Bohn Nielsen and Paolo Paruolo have been very helpful.

I would also like to give credit to my fellow PhD students at university of Copenhagen during these years for great lunch discussions and continual sparing on the day-to-day challenges of writing a PhD thesis. In particular, sharing offices with Andreas Noack Jensen, Line Elvstrøm Ekner, Rasmus Søndergård Pedersen and Andreas Lund Hetland has been a great experience.

During my visit to CREST, Paris, I was very well received by the researchers at the laboratories of macroeconomics and finance. I would like to thank, in particular, Frédérique Bec for helping me arrange the stay at CREST.

Finally, I would like to thank my friends and family for putting up with me and being supportive during the more challenging periods of the PhD program.

1 Likelihood-based inference in dynamic mixture cointegrated VAR models

This chapter is based on joint work with Paolo Paruolo[1] and Anders Rahbek[2].

We consider likelihood-based asymptotic inference in a general class of regime-switching, or dynamic-mixture, cointegrated vector error correction models. This framework allows for epochs of non-stationary behavior, asymmetric error correction and switching error covariance; this extends previously-introduced classes of processes. Unlike previous results on non-linear switching cointegrated models, we discuss asymptotic inference on all parameters, including the cointegrating vectors and switching covariances. To do so, we introduce a new functional central limit theory for non-stationary switching processes, find explicit conditions for existence of moments and derive limiting distributions.

1.1 Introduction

This paper discusses general dynamic-mixture models with cointegration, called Autoregressive Conditional Root (ACR) cointegrated models. We study properties of ACR processes and related likelihood-based inference. Members of this class were introduced in Bec and Rahbek (2004). Non-linear cointegrated models have found many applications in macroeconomics and finance; key examples include studies of purchasing power parities, term structures of interest rates, forward parities, see inter alia Corradi et al. (2000); Lo and Zivot (2001); Hansen and Seo (2002); Seo (2003); Bec and Rahbek (2004); Psaradakis et al. (2004); Bec et al. (2006); Kapetanios et al. (2006); Clarida et al. (2006); Bec et al. (2008); Lof (2012).

This paper introduces several novelties. First, we introduce the general class of ACR processes. This is shown to be a flexible class of non-stationary and regime-switching processes which nests several ones found in the literature. In particular the processes in Bec and Rahbek (2004) and Bec et al. (2008) are special cases. Secondly, we discuss likelihood inference on all parameters in the model, including the cointegration parameters. We derive limit distributions and discuss the accuracy of the asymptotic approximation with a small Monte Carlo simulation study.

The ACR class is a comprehensive one. It allows for multiple regimes, some with, and some without, mean-reverting behavior; the processes have epochs of unstable (even explosive, and hence in particular, bubble-like) behavior, that are brought to an end by epochs of mean-reverting behavior. Similarly to linear cointegrated processes, some linear combinations of the

[1]Research Officer at the European comission, Joint Research Center
[2]Professor in econometrics at University of Copenhagen.

process are stationary, while the whole process is non-stationary. Also in this case, cointegraton is associated with the presence of a linear attractor set.[3]

Moreover, the adjustment towards the attractor set is regime-dependent, with probabilities that govern switching among states which depend on observable stationary variables, including deviations from the attractor set. This gives rise to very general specifications that, similarly to Saikkonen (2005, 2008), allow for asymmetric adjustment. Several papers have argued in favor of asymmetric adjustment, see e.g., Hansen and Seo (2002) and Kılıç (2011), or (possibly asymmetric) polynomial models, see, e.g. Baghli (2005); Escribano (2004). A different class of models with non-linear adjustment has been discussed in Kristensen and Rahbek (2010, 2013), who consider cointegrated non-linear smooth transition models without regime switching.[4]

We also allow for regime-dependent covariances, reflecting time-varying volatility, which is found widely in applications, see the discussion in Cavaliere et al. (2010a).

The new extended ACR class is naturally also related to Markov Switching (MS) models, see Hamilton (1994, Chapter 22) and Lange et al. (2011). However, MS and ACR processes differ in a fundamental way: the ACR class has probabilities of switching that depend on past observables, and it is an 'observation-driven' class of processes; conversely, MS are usually examples of 'parameter-driven' processes, see Cox (1981).

We provide a full theory for inference on all parameters, including the cointegrating parameters, which hitherto have been assumed known in existing literature on regime-switching cointegrated models. A new representation theory is provided, which decomposes the process into a stationary and geometrically ergodic term, a linear deterministic trend and a stochastic trend. Also a new functional central limit theory for switching processes is introduced.

The cointegration parameters are found to be super-consistent, including T and $T^{3/2}$ rates of convergence, while remaining parameters are found to be standard $T^{1/2}$ consistent, where T denotes the number of time periods. Interestingly, and in line with Kristensen and Rahbek (2010, 2013) for smooth transition models which do not include switching, inference is found to be not block-orthogonal between the cointegration parameters and the remaining parameters, and moreover that the asymptotic distributions depend on nuisance parameters.

The rest of the paper is organized as follows. Section 1.2 describes the class of ACR cointegrated class of processes of interest. Section 1.3 provides the representation theory. Section 1.5 illustrates the ACR process class via simulation. Section 1.4 presents leading specifications. Section 1.6 discusses identification, Section 1.7 presents properties of the QMLE and discusses likelihood-based inference. Section 1.8 concludes. Proofs are placed in Appendices.

The following notation is used throughout: "$\overset{w}{\to}$" and "$\overset{p}{\to}$" denote weak convergence and convergence in probability, respectively as $T \to \infty$. The dimension of the system is denoted n and n_a is the number of elements in a vector a. The cointegration rank is denoted by r, and we define β as an $n \times r$ matrix of rank, $r < n$. $\beta_\perp$ indicates any $n \times (n-r)$ matrix of rank $n-r$ for which $\beta'\beta_\perp = 0$. We also let $\bar{\beta} := \beta\left(\beta'\beta\right)^{-1}$, such that the orthogonal projection identity can be written as $\mathcal{I}_n = \bar{\beta}\beta' + \bar{\beta}_\perp\beta'_\perp$, where $\mathcal{I}_n$ is the identity matrix of size n. We use c to denote a generic constant; $\mathrm{vech}\left(A\right)$ and $\mathrm{vec}\left(A\right)$ indicate the column-stacking operators that act

[3] See Gao and Phillips (2011) and Karlsen et al. (2007) and references therein on estimation of non-linear cointegrating relations.

[4] See also the discussion of smooth transition models as an approximation to threshold models in Seo (2011).

on the lower triangular portion of A and the whole of A respectively; $\|\cdot\|$ denotes a norm and $|\cdot|$ for the absolute value of a number. We further use use the fact that $\mathcal{D}_\Omega \text{vech}(\Omega_j) = \text{vec}(\Omega_j)$, $\mathcal{D}_\Lambda \text{vech}(\Lambda) = \text{vec}(\Lambda)$ where $\mathcal{D}_\cdot$ are *duplication* matrices of appropriate dimensions, see e.g. Magnus and Neudecker (1999, Chapter 3). In the same line of though, $\mathcal{I}_n$ is used as the identity matrix of size $(n \times n)$. We write the indicator function of $a > b$ as $\mathbf{1}\{a > b\}$. For partial derivatives of a scalar function f with respect to vectors u, v, w say, we make use of the notation

$$\partial_u f := \frac{\partial f}{\partial u'} \quad , \quad \partial^2_{uv} f := \frac{\partial^2 f}{\partial u \partial v'} \quad \text{and} \quad \partial^3_{uvw} f := \partial_s \text{vec}\left(\partial^2_{uv} f\right),$$

see e.g. Magnus and Neudecker (1999).

For a set of square matrices $\mathcal{M} := \{M_i\}_{i=1}^k$ we indicate by

$$\mathcal{M}^k = \left\{ \prod_{i=1}^k M_i : M_i \in \mathcal{M},\, i = 1, 2, ..., k \right\}$$

and we employ the definition of joint spectral radius

$$\rho(\mathcal{M}) = \limsup_{k \to \infty} \left(\sup_{M \in \mathcal{M}^k} \|M\| \right)^{\frac{1}{k}}, \tag{1.1}$$

see e.g. Liebscher (2005). Note that for a set $\mathcal{M}$ consisting of a single square matrix M, $\mathcal{M} := \{M\}$, the joint spectral radius $\rho(\mathcal{M}) = \rho(M)$ is the spectral radius of the matrix M.

1.2 The ACR cointegrated process

In this section we define the class of processes of interest, the cointegrated ACR processes. This can be seen as a non-linear extension of the linear cointegration model, see Johansen (1996) and Saikkonen (2008); Kristensen and Rahbek (2010, 2013). The non-linearity is introduced through a regime-switching mechanism; the probability of switching among states is taken to depend on past, observable, stationary variables, including cointegrating relations. Importantly, the ACR class allows one to model, inter alia, different speeds of return to equilibrium depending on the distance of the process from equilibrium. Moreover it also models time-changing volatility.

1.2.1 Dynamics given the states

We describe the ACR process in steps: we first describe the dynamics of observables given the states, and subsequently we describe the probability of switching among different states. An n-dimensional ACR cointegrated process is defined as the process, X_t, generated by the equation,

$$\Delta X_t = \sum_{j \in \mathbb{M}} \mathbf{1}\{s_t = j\} \left(\alpha_j \beta^{\star\prime} X^\star_{t-1} + \Gamma_j \Delta \mathbb{X}_{t-1} + V_j \epsilon_t \right) \quad \text{with} \quad \epsilon_t \sim \text{i.i.d.}(0, I_n), \tag{1.2}$$

where, when $k > 1$ we define $\Delta \mathbb{X}_t := \left(\Delta X'_t : \Delta X'_{t-1} : ... : \Delta X'_{t-k+2} \right)'$, $X^\star_t = (X'_{t-1} : 1)'$, Δ is the first difference operator, $\Delta X_t = X_t - X_{t-1}$ and k is the number of lags of X_t included in the model. When $k = 1$ we suppress $\Delta \mathbb{X}_t$. The stochastic, univariate, switching variable s_t

is unobserved, and takes values in the set $\mathbb{M} := \{1, 2, ..., m\}$. The n-dimensional innovation ϵ_t has mean 0 and covariance matrix I_n, and it has some well-defined density with respect to the Lebesque measure, see the following Definition 1.1.

Below we consider the probability p_{jt} of $s_t = j$ conditional on the realization of X_{t-1}, $\Delta\mathbb{X}_{t-1}$ and s_{t-h}, $h \geq 1$. The specification of p_{jt} is given for different regime switching mechanisms, including smooth transition ones.

As will be shown below, it is fundamental for the interpretation of the cointegrating properties of the process X_t, that some non-empty subset $\mathbb{M}_1$, say, of the states $\mathbb{M}$, $\mathbb{M}_1 \subseteq \mathbb{M}$, imply 'mean-reverting' behavior, and that these mean-reverting states are reached with probability tending to one for large deviations from equilibrium. A process specified in this way will be referred to as an ACR cointegrated process.

For the ACR cointegrated process the coefficients appearing in the ΔX_t equation can be interpreted by appealing to linear cointegration terminology. In fact, as shown in the next section, the matrix $\beta^\star := (\beta' : \beta_D')'$ in (1.2) collects the cointegration vectors β of dimension $n \times r$, as well as the intercept term β_D of dimension $1 \times r$. The regime-specific adjustment matrices α_j and short-run dynamics matrices $\Gamma_j := (\Gamma_{j,1} : ... : \Gamma_{j,k-1})$ are of dimensions $n \times r$ and $n \times n\,(k-1)$, respectively. Finally $\Omega_j = V_j V_j'$ indicate regime-specific covariances, where V_j are assumed to be $n \times n$ matrices of full column rank.

The random variables ϵ_t and s_t are assumed to be independent conditionally on their past, and moreover that their distributions depend only on (parts of) Z_{t-1}, where $Z_t := (X_t'\beta : \Delta\mathbb{X}_t')'$. More precisely, we assume that for any Borel-measurable set A and $j \in \mathbb{M}$,

$$\begin{aligned} &s_t, \epsilon_t \mid (X_q, s_q)_{q=-k,...,t-1} \overset{D}{=} s_t, \epsilon_t \mid Z_{t-1}, \\ &\Pr(s_t = j, \epsilon_t \in A \mid Z_{t-1}) = \Pr(\epsilon_t \in A \mid Z_{t-1}) \Pr(s_t = j \mid Z_{t-1}), \end{aligned} \tag{1.3}$$

where $\overset{D}{=}$ indicates equality in distribution. Moreover, the innovations ϵ_t are assumed independent of Z_{t-1},

$$\Pr(\epsilon_t \in A \mid Z_{t-1}) = \Pr(\epsilon_t \in A) \tag{1.4}$$

for any Borel-measurable set A. Finally, the conditional distribution of s_t is allowed to depend on (a subset z_t of) Z_t, that is $\Pr(s_t = j \mid Z_{t-1}) = \Pr(s_t = j \mid z_{t-1})$; in the following we indicate this probability as

$$p_{jt} := \Pr(s_t = j \mid z_{t-1}). \tag{1.5}$$

The vector of variables z_t that enters in the specification of p_{jt} is a function of $Z_t := (X_t'\beta : \Delta\mathbb{X}_t')'$, which we write as $z_t = \psi' Z_t = \psi_\beta' \beta X_t + \psi_\Delta' \Delta\mathbb{X}_t$, with $\psi := \left(\psi_\beta' : \psi_\Delta'\right)'$ conformable with the partition of Z_t. Observe that ψ is a fixed selection matrix and is not considered a parameter to be estimated. Concrete specifications for p_{jt} are discussed in Section 1.4; these probabilities are indexed by a vector of coefficients indicated by γ in the following.

1.2.2 Mean-reverting states and transition probabilities

We next define mean-reverting states and illustrate the condition for the process to reach these states with probability one as deviations from the attractor set become large.

We allow for multiple mean-reverting states, collected in $\mathbb{M}_1$, and possibly non-mean-reverting states, $\mathbb{M}_2$ say, such that $\mathbb{M} = \mathbb{M}_1 \cup \mathbb{M}_2$. The idea is to ensure that the system has a dynamic behavior in these states similar to linear cointegration in $\mathbb{M}_1$, while in the remaining $\mathbb{M}_2$ states there are no requirements for the dynamics. In particular, if the model is specified with $\mathbb{M} = \mathbb{M}_1$ and only one state, the cointegrated ACR process reduces to the classic (non-switching) linear cointegrated VAR process. The further requirement mentioned, that is that $\mathbb{M}_1$ is reached with probability tending to one for large deviations, implies that indeed the 'cointegrating', or mean-reverting, behavior is overall dominating the dynamics of the process.

More precisely, collect the regimes in the two subsets, $\mathbb{M}_1 = \{1, ..., m_1\}$ and $\mathbb{M}_2 = \{m_1 + 1, ..., m\}$, such that $\mathbb{M} = \mathbb{M}_1 \cup \mathbb{M}_2$. Then the regimes in $\mathbb{M}_1$ are called 'mean-reverting' if

$$\rho\left(\mathcal{A}_{\mathbb{M}_1}\right) < 1, \quad \text{where} \quad \mathcal{A}_{\mathbb{M}_1} := \{\mathbb{A}_j, j \in \mathbb{M}_1\} \tag{1.6}$$

and

$$\mathbb{A}_j := \begin{pmatrix} A_{j1} & \cdots & \cdots & A_{jk} \\ I_n & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I_n & 0 \end{pmatrix}, \qquad A_{j1} := \begin{pmatrix} I_r + \beta'\alpha_j + \beta'\Gamma_{j1}\bar{\beta} & \beta'\Gamma_{j1}\bar{\beta}_\perp \\ \beta'_\perp\alpha_j + \beta'_\perp\Gamma_{j1}\bar{\beta} & \beta'_\perp\Gamma_{j1}\bar{\beta}_\perp \end{pmatrix},$$

$$A_{ji} := \begin{pmatrix} \beta'\left(\Gamma_{ji} - \Gamma_{ji-1}\right)\bar{\beta} & \beta'\Gamma_{ji}\bar{\beta}_\perp \\ \beta'_\perp\left(\Gamma_{ji} - \Gamma_{ji-1}\right)\bar{\beta} & \beta'\Gamma_{ji}\bar{\beta}_\perp \end{pmatrix}, \qquad i = 2, ..., k,$$

with $\Gamma_{jk} := 0$. Here $\mathbb{A}_j$ is the companion matrix of regime j in the companion form representation of process (1.2), namely

$$\mathbb{Y}_t = \sum_{j \in \mathbb{M}} \mathbf{1}\left\{s_t = j\right\}\left(\mathbb{A}_j\mathbb{Y}_{t-1} + \mathbb{U}_{jt}\right) = \mathbb{A}_t\mathbb{Y}_{t-1} + \mathbb{U}_t \tag{1.7}$$

where $\mathbb{U}_j := J\left(\alpha_j\beta'_D + V_j\epsilon_t\right)$ and $\mathbb{Y}_t := \left(Y'_t : Y'_{t-1} : \cdots : Y'_{t-k+1}\right)'$, $Y_t := \left(X'_t\beta : \Delta X'_t\beta_\perp\right)'$, $J := \left(\mathcal{I}_n : 0\right)'$.

It is interesting to compare the requirement (1.6) with the usual conditions for cointegration in a linear VAR. In the latter case, the so-called I(1) conditions[5] are stated directly in terms of the cointegrating parameters β, the adjustment coefficients α and the short-run dynamics matrices Γ_h. In the present case, one needs to resort to the companion form representation, and to the notion of generalized spectral radius, because a Moving Average representation of the process involves companion matrices $\mathbb{A}_j$ multiplied in all possible orders.

As formally stated in Theorem 1.4 below, $\beta' X_t$ and ΔX_t are found to be stationary, and hence X_t cointegrated, in the system governed by the coefficient matrices in $\mathcal{A}_{\mathbb{M}_1}$. This is sufficient to ensure that the same applies in general, provided the probability to access $\mathbb{M}_1$ goes to 1 as $\|z_{t-1}\|$ increases. This ensures that the dynamics governed by the coefficient matrices in $\mathcal{A}_{\mathbb{M}_1}$, i.e. the dynamics of the states in $\mathbb{M}_1$, are dominating.

Formally we require that with probability tending to one as $\|z_{t-1}\|$ increases, the states $\mathbb{M}_1$

[5] See Johansen (1996) inter alia.

are reached, i.e.

$$\sum_{j\in\mathbb{M}_1} p_{jt} \to 1 \quad \text{for } \|z_{t-1}\| \to \infty. \tag{1.8}$$

We call this the stability requirement for p_{jt}. Observe that $\mathbb{M}_1$ and $\mathbb{M}_2$ allow for asymmetric adjustment behavior, and hence generalize the formulation of Bec and Rahbek (2004) and Bec et al. (2008), while incorporating the asymmetries introduced in Saikkonen (2005, 2008).

We finally collect the assumptions on the process in the following formal definition.

Definition 1.1. [ACR cointegrated process] A process X_t generated by (1.2), with ϵ_t and s_t conditionally independent, see (1.3), ϵ_t with p.d.f. with respect to Lebesque measure, positive at the origin, and independent of the past (see (1.4)), probability of switching p_{jt} in (1.5) satisfying mean-reverting (or stable) behavior (1.8) for a non-empty set of states $\mathbb{M}_1 \subseteq \mathbb{M}$, see (1.6), is called an '**ACR cointegrated process**'.

From (1.3) we need to specify the conditional probabilities of switching to regime j. The choice of parametrization for these *predicted state probabilities* has to satisfy the stability requirement in equation (1.8); we discuss some of the possible choices below. In the asymptotic results, we consider a generic parametric specification of p_{jt} satisfying regularity conditions as a function of z_{t-1} and of the vector of parameters, γ.

1.3 Representation

In this section, we give a representation theorem for ACR cointegrated processes defined in the previous section. We first introduce a mild assumption on the moments.

Assumption 1.2. *[Moments] The n-dimensional vector sequence $\{\epsilon_t\}$ has moments of order $2q$, $E\left(\|\epsilon_t\|^{2q}\right) < \infty$, for some $q \geq 1$.*

We next introduce more notation and a technical assumption. Because Z_t is a function of the state variables $\mathbb{Y}_t$ in (1.7), and we also write $z_t = \eta'\mathbb{Y}_t$ for an appropriate $n_{\mathbb{Y}} \times n_z$ matrix η of full column rank.

In terms of the companion form representation in eq. (1.7), we next state an assumption concerning the short-term dynamics of the process. In particular this assumption is not needed if $k = 1$ or if the Γ_j parameters are identical across regimes or when $\mathbb{M}_2$ is empty.

Assumption 1.3. *[Control over drift function] There exists some $i \in \mathbb{M}_1$ such that one has $(\mathbb{A}_j - \mathbb{A}_i)\,\eta_\perp = 0$ for any $j \in \mathbb{M}_2$.*

We can then state the following representation results.

Theorem 1.4. *Consider an ACR cointegrated process in definition* 1.1 *and let Assumptions* 1.2 *and* 1.3 *hold with $q \geq 1$. Then the following properties apply:*

1. *The process $\mathbb{Y}_t$ in* (1.7) *is geometrically ergodic, with finite moments of order $2q$. In particular, the initial value $\mathbb{Y}_0$ can be given a distribution such that $\mathbb{Y}_t$ and hence, $\beta' X_t$ and $\beta'_\perp \Delta X_t$ (and hence also ΔX_t) are stationary. For this choice of initial values, define the expectations $\mu_1 := E(\beta' X_t)$, $\mu_2 := E(\beta'_\perp \Delta X_t)$ and the de-meaned stationary processes $v_t := \beta' X_t - \mu_1$, $\xi_t := \beta'_\perp \Delta X_t - \mu_2$.*

2. *The law of large numbers applies to any measurable function* $f(\beta' X_{t-i}, \Delta X_{t-i}; i = 0, 1, ...)$, *with* $E(f(\cdot)) < \infty$.

3. *The process* X_t *has representation*

$$X_t = \tau t + \overline{\beta}_\perp \sum_{i=1}^{t} \xi_i + \bar{\beta}(v_t + \mu_1) + \overline{\beta}_\perp \beta'_\perp X_0, \tag{1.9}$$

where $\tau := \overline{\beta}_\perp \mu_2$, $\kappa := \beta_\perp \mu_{2\perp}$.

Proof. The proof is given in Appendix 1.A. □

Theorem 1.4 establishes that ACR processes have cointegration properties similar to the ones of linear cointegrated processes. Eq. (1.9) in fact shows that X_t can be decomposed into a linear trend, $n - r$ random walks, a stationary component and initial values. This result is a consequence of the fact that $\mathbb{Y}_t$ is a geometrically ergodic Markov chain.

The role of Assumptions 1.2 and 1.3 for the results of the representation Theorem 1.4 is the following. Assumption 1.2 is needed for the existence of moments and for the working of the drift criterion. Assumption 1.3 is also needed in the drift criterion to control the directions of $\mathbb{Y}_t$ not entering p_{jt}, in (1.5), which are controlled via requirement (1.8) on their mean-revertion.

1.4 Specification of p_{jt}

In this section we discuss parametric specifications for p_{jt} in (1.5) that impose restriction (1.8) for the mean-reverting states. As illustrated in the example process in the previous section, specifications of these transition probabilities can be hierarchical, in the sense that probabilities are first assigned to the classes $\mathbb{M}_1$ and $\mathbb{M}_2$, and then to the conditional probability of state h within class $\mathbb{M}_j$. Of course, non-hierarchical alternatives are also possible.

Here we focus on hierarchical specifications, because they allow to incorporate condition (1.8) in a straightforward way. Specifically, we indicate as $p_t = \Pr(s_t \in \mathbb{M}_1 \mid z_{t-1})$ the probability that s_t belongs to the mean-reverting states in $\mathbb{M}_1$; the conditional probabilities for s_t to be in state j given that $j \in \mathbb{M}_i$, $i = \{1, 2\}$ are indicated as $\pi_{jt \cdot i} = \Pr(s_t = j \mid j \in \mathbb{M}_i, z_{t-1})$, where we recall that $z_t = \psi' Z_t = \psi'_\beta \beta X_t + \psi'_\Delta \Delta \mathbb{X}_t$. This gives

$$p_{jt} := \Pr(s_t = j \mid z_{t-1}) = \begin{cases} p_t \pi_{jt \cdot 1} & \text{for } j \in \mathbb{M}_1 \\ (1 - p_t)\, \pi_{jt \cdot 2} & \text{for } j \in \mathbb{M}_2, \end{cases} \tag{1.10}$$

Note that $\sum_{j \in \mathbb{M}_1} p_{jt} = p_t$ by construction. Condition (1.8) can be incorporated in this specification by ensuring that $p_t \to 1$ for increasing $\|z_{t-1}\|$.

Some possible specifications of p_t and $\pi_{jt \cdot i}$ are given in the following subsections; the complete vector of parameters involved in these specifications is indicated by γ. The parameters in the first layer specification are indicated by ϱ, the ones of the second layer specification by ζ, with $\gamma = (\varrho' : \zeta')'$. Before discussing specifications, we first state smoothness conditions needed in the likelihood analysis on the switching probability p_{jt} as a function both of the parameter vector γ and of z.

Assumption 1.5. *[Smoothness of p_{jt}] The first, second and third order derivatives of $\log p_{jt}$ w.r.t. z and γ exist and are bounded as $\|z\| \to \infty$ uniformly over γ, such that*

$$\|\partial_u \log p_{jt}\| \leq c, \quad \left\|\partial^2_{uv} \log p_{jt}\right\| \leq c \quad \textit{and} \quad \left\|\partial^3_{uvw} \log p_{jt}\right\| \leq c,$$

where c are generic constants and u, v, $w \in \{z, \gamma\}$.

The smoothness condition of p_{jt} holds for the following specifications, as verified in Appendix 1.F.

1.4.1 First layer specifications

An exponential specification for p_t is

$$p_t = 1 - \exp\left(-g\left(z_{t-1}; \varrho\right)\right), \quad \text{with} \quad g\left(z_{t-1}; \varrho\right) = \left(z_{t-1} - \mu\right) \Lambda \left(z_{t-1} - \mu\right), \tag{1.11}$$

where Λ is a square, positive definite matrix of order n_z, and $\mu \in \mathbb{R}^{n_z}$; here $\varrho := \left(\text{vech}\left(\Lambda\right)' : \mu'\right)'$. Note that as $\|z_{t-1}\| \to \infty$, one has $g\left(z_{t-1}; \varrho\right) \to \infty$ and $p_t \to 1$ as required in (1.8).

Exponential specifications are popular in the smooth transition literature is (see. e.g. Teräsvirta et al. (2010b)). This specification of p_t is symmetric in z around μ, i.e. deviation $z_{t-1} - \mu$ and $-\left(z_{t-1} - \mu\right)$ give the same value of p_t. The exponential specification can be made asymmetric in several ways; one possibility is to replace $g\left(z_{t-1}; \varrho\right)$ in (1.11) with $g\left(z_{t-1}; \varrho\right) h\left(z_{t-1}; \varrho\right)$ where $h\left(z_{t-1}; \varrho\right) = 0.5 + \left(1 - \exp\left(-\varpi' z_{t-1}\right)\right)^{-1}$ is a logistic function, with $\varpi \in \mathbb{R}^{n_z}$.

A different possibility is to select p_t as a logistic distribution function given by,

$$p_t = \frac{\exp\left(g\left(z_{t-1}; \varrho\right)\right)}{\exp\left(g\left(z_{t-1}; \varrho\right)\right) + 1}, \qquad g\left(z; \varrho\right) = \left(z_{t-1} - \mu\right)' \Lambda \left(z_{t-1} - \mu\right) - \varpi \tag{1.12}$$

where ϖ is a (non-negative) scalar and other parameters are defined as in (1.11), and $\varrho := \left(\text{vech}\left(\Lambda\right)' : \mu' : \varpi\right)'$. Again here, for $\|z_{t-1}\| \to \infty$ one has $g\left(z_{t-1}; \varrho\right) \to \infty$ and $p_t \to 1$ as required in (1.8). Different specifications exists, but the ones reported above are the most popular; in particular the exponential appears to work well in our illustrations below.

1.4.2 Second layer specifications

For the second-layer probabilities $\pi_{jt\cdot i}$, one may select a multinomial logistic specification,

$$\pi_{jt\cdot i} = \pi_{j\cdot i}\left(z_{t-1}\right) = \frac{\exp\left(\zeta_j' \left(z_{t-1} - \mu\right)\right)}{\sum_{\ell \in \mathbb{M}_i} \exp\left(\zeta_\ell' \left(z_{t-1} - \mu\right)\right)} \tag{1.13}$$

with $\sum_{j \in \mathbb{M}_i} \pi_{jt\cdot i} = 1$, $i = \{1, 2\}$. All unrestricted ζ_j parameters are collected in the vector ζ. Again, other specifications are of course possible.

Note that the hierarchical specification nests the one in Bec and Rahbek (2004) and Bec et al. (2008) which corresponds to choosing (1.12), with $m = 2$, $m_1 = 1$, and

$$g\left(z_{t-1}; \varrho\right) = a + b \left\|z_{t-1}\right\| = -\Lambda\mu + \Lambda \left\|z_{t-1}\right\|$$

Table 1.1: The Joint Spectral Radius of regimes in $\mathbb{S}_1$ and the regime specific characteristic roots

$\text{eig}_i\mathbb{A}_1$	$\mathbb{M}_1$	0.8728	0.1000	-0.5728	0.0000
$\text{eig}_i\mathbb{A}_2$	$\mathbb{M}_1$	0.8240	0.1130	-0.5370	0.0000
$\text{eig}_i\mathbb{A}_3$	$\mathbb{M}_2$	1.0000	0.1000	-0.5000	0.0000
JSR Interval			$[0.8728; 0.8828]$		

where $\varrho = (a, b)'$, $\varpi = 0$ and $\|z_t\| = \sqrt{z_t' z_t}$.

1.5 A simulated process

In this section, we report the simulated path of an ACR process that satisfies the assumptions of Theorem 1.4. Let $\mathbb{M}_1 = \{1, 2\}$, and $\mathbb{M}_2 = \{3\}$, $X_t = (x_{1t}, x_{2t})'$ with $k = 2$ in (1.2) with $\epsilon_t \sim i.i.dN(0, I_2)$, $\Omega_j = V_j V_j' = \sigma_j^2 \cdot I_2$, $\sigma_1^2 = 0.01$, $\sigma_2^2 = 0.05$, $\sigma_3^2 = 0.1$, $\alpha_1 = (-0.2, 0)'$, $\alpha_3 = (0.0)'$, $\beta^\star = (1, -1, -1)'$ and $\mathbb{X}_{t-1} = X_{t-1}$. We restrict the short run parameters to be the same across regimes, i.e.

$$\Gamma_1 = \Gamma_2 = \Gamma_3 = \begin{pmatrix} -0.5 & 0.3 \\ 0.0 & 0.1 \end{pmatrix}.$$

Moreover, we select $z_t = \beta' X_t$ with $p_{jt} = p_{jt}\pi_{jt}\mathbf{1}\{j \in \mathbb{M}_1\} + (1 - p_{jt})\mathbf{1}\{j \in \mathbb{M}_2\}$ for $j = 1, 2, 3$ where $p_t = 1 - \exp\left(-\Lambda (z_{t-1} - \mu)' (z_{t-1} - \mu)\right)$ with $\Lambda = 3$ and $\mu = 1$. Finally, we choose

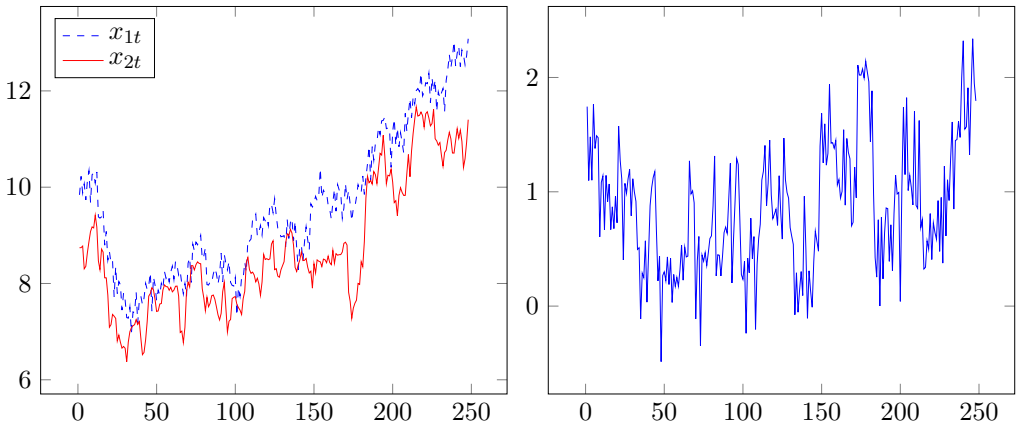
$$\pi_{jt\cdot 1} = \exp\left(\zeta_2' (z_{t-1} - \mu)\right) / \left(1 + \exp\left(\zeta_2' (z_{t-1} - \mu)\right)\right) \tag{1.14}$$

with $\zeta_2 = 3$.

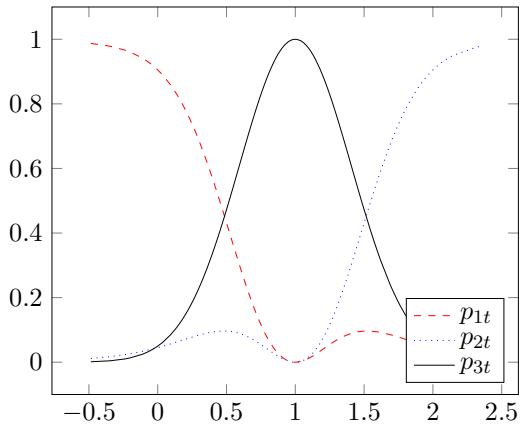
We first verify that this process satisfies the assumptions of Theorem 3: ϵ_t is Gaussian and hence satisfies the definition as well as Assumption 1.2 for any q; the condition in (1.3) is satisfied by p_{jt} due to the choice of the function p_t. Moreover, Assumption 1.3 is satisfied because $\Gamma_1 = \Gamma_2 = \Gamma_3$. In order to verify that the joint spectral radius in (1.1) is less than one, we use the Gripenberg Algorithm, see Gripenberg (1996) and Jungers (2009), as implemented in the JSR louvain toolbox for MATLAB. In Table 1.1, we report the calculated roots of each regime as well as the joint spectral radius with an accuracy of 0.001; the table shows that the system satisfies the mean reverting conditions.

The stylized system incorporates the novel elements, while still being fairly simple. The series are seen to be comoving and contain an upward trend which is a consequence of the asymmetric regime structure resulting in $\tau_0 \neq 0$, where τ_0 was defined in (1.9). The inclusion of the constant is what gives the nonzero mean in the cointegration relations, graph (B). Note also the sustained deviations from equilibrium, which are generated by the fact that the inner regime is a regime with no error correction and cointegration. Finally, (C) depicts the regime predicted state probabilities as functions of the cointegration relations. Here, the choice of the switching probability structure becomes visible; it is clearly seen that for $z_{t-1} \approx 1$, the probability of being in regime three is close to one, while large positive deviations result in a regime switch to regime two and large negative deviations result in a regime switch to regime one. This stylized system incorporates all of the novel elements that we introduce, while still being fairly simple. The

Figure 1.1: Illustration of an ACR cointegrated system



(a) $X_t = (x_{1t}, x_{2t})'$ (b) $\beta' X_{t-1} = x_{1t} - x_{2t}$



(c) $\beta' X_{t-1} \times p_{jt}$ for $j \in \mathbb{M}$

series of graphs given in figure 1.1 illustrates the behavior of the system.

1.6 Parameter Identification and Normalization

With X_t defined in (1.2) the parameters to be estimated are given by: (i) the parameters which enter the equilibrium correction part directly, that is α_j, $\beta^\star$, Γ_j and $\Omega_j = V_j V_j'$ for $j \in \mathbb{M}$; and (ii) the parameters that govern the switching through the probabilities p_{jt} (collected in the vector, γ), also with $j \in \mathbb{M}$. In this section we discuss identification.

1.6.1 Identification of the cointegrating parameters

As in the linear case, the cointegration matrix is only identified up to a normalization; in the following we introduce a coordinate system that imposes identification. Making use of the results from Theorem 1.4, we present a normalized version of β, where the stationary and non-stationary directions are separated in a way that eases the presentation of the asymptotic results.

Define the true value $\beta_0^\star$ of $\beta^\star$ as $(\beta_0' : 0')'$; we wish to normalize $\beta^\star$ as $\widetilde{\beta}^\star$ so as to make $\bar{\beta}_0^{\star\prime}\widetilde{\beta}^\star = \mathcal{I}_r$; this is accomplished by setting $\widetilde{\beta}^\star := \beta^\star(\bar{\beta}_0'\beta)^{-1}$. We next decompose $\widetilde{\beta}^\star$ into relevant components. If $\tau_0 \neq 0$ as defined in Theorem 1.4, we use orthogonal projections on the space spanned by β_0, τ_0 and κ_0, where $\kappa_0 := (\beta_0 : \tau_0)_\perp$. This gives

$$\widetilde{\beta}^\star - \beta_0^\star = (\bar{\kappa}_0^\star : \overline{\tau}_0^\star : i_{n+1}) \begin{pmatrix} \kappa_0'\tilde{\beta} \\ \tau_0'\tilde{\beta} \\ \tilde{\beta}_D \end{pmatrix} =: (\bar{\kappa}_0^\star : \overline{\tau}_0^\star : i_{n+1}) \begin{pmatrix} b \\ b_D \end{pmatrix}$$

where $b := (b_\kappa' : b_\tau')' := \left(\tilde{\beta}'\kappa_0 : \tilde{\beta}'\tau_0\right)'$, $\kappa_0^\star := (\kappa_0' : 0)'$, $\tau_0^\star := (\tau_0' : 0)'$, i_{n+1} is the last column in $\mathcal{I}_{n+1}$ and $b_D := \tilde{\beta}_D := \beta_D(\bar{\beta}_0'\beta)^{-1}$ by definition.

Using this parametrization, we can rewrite the term $\alpha_j\beta^{\star\prime}X_t^\star$ as $\alpha_j(\beta'\bar{\beta}_0)\tilde{\beta}^{\star\prime}X_t^\star =: \widetilde{\alpha}_j\tilde{\beta}^{\star\prime}X_t^\star$, where $\tilde{\beta}^\star$ is identified and $\widetilde{\alpha}_j := \alpha_j(\bar{\beta}_0'\beta)^{-1}$ is the corresponding identified adjustment coefficient in regime j. Observe that

$$\begin{aligned} \alpha_j\beta^{\star\prime}X_{t-1}^\star &= \tilde{\alpha}_j\tilde{\beta}^{\star\prime}X_t^\star = \tilde{\alpha}_j\left(\beta_0^{\star\prime}X_t^\star + (b' : b_D')\,(\bar{\kappa}_0^\star : \overline{\tau}_0^\star : i_{n+1})'\,X_t^\star\right) \\ &= \tilde{\alpha}_j\left(\beta_0'X_t + b'\mathcal{X}_{t-1} + b_D' i_{n+1}'\right). \end{aligned} \tag{1.15}$$

where $(\bar{\kappa}_0^\star : \overline{\tau}_0^\star)'\,X_t^\star$ decomposes the process X_t in the various components given in Theorem 1. In addition, we have used the definition $\mathcal{X}_t := (\bar{\kappa}_0 : \overline{\tau}_0)'\,X_t$. This also identifies α_j as $\widetilde{\alpha}_j$. The same reasoning is applied to the parameters in γ that govern p_{jt}, where $\beta'X_t = \beta_0'X_t + b'\,(\bar{\kappa}_0 : \overline{\tau}_0)'\,X_t$, where $b := (b_\kappa' : b_\tau')'$. For the sake of notational ease, we shall not distinguish notationally between $\widetilde{\alpha}_j$, $\widetilde{\gamma}$ and α_j, γ in the following and we define further .

Remark. While convenient for deriving the results on the asymptotic theory of the QMLE estimator, the identification principles given here are not necessarily the most convenient to implement in practice. Alternatives that might be preferred when taking the model to the data are the linear restrictions discussed in Johansen (1996) or their generalizations given by Boswijk and Doornik (2004). We discuss these further in chapter two.

The regime specific covariance matrices of the error terms, $\Omega_j = V_j V_j'$. To avoid degenerate distributions, we consider the vectors as $\text{vech}(\Omega_j)$.

Identification of the probability parameters entering the regime p_{jt} will be dependent on the choice of the switching structure and can be considered on a case by case basis.

1.6.2 Rate of convergence and T- normalizations

We collect the parameters to be estimated into n_θ-dimensional parameter vector, θ, defined as

$$\theta := \left(\text{vec}\left(b'\right)' : \text{vec}\left(b_D'\right)' : \text{vec}\left(\alpha\right)' : \text{vec}\left(\Gamma\right)' : \text{vech}\left(\Omega\right)' : \gamma'\right)' := \left(\text{vec}\left(b'\right)' : \vartheta'\right)', \tag{1.16}$$

where $\alpha := (\alpha_1 : \alpha_2 : ... : \alpha_m)$, $\Gamma := (\Gamma_1 : \Gamma_2 : ... : \Gamma_m)$, $\vartheta := (\text{vec}\left(b_D'\right)' : \text{vec}\left(\alpha\right)' : \text{vec}\left(\Gamma\right)' : \text{vech}\left(\Omega\right)' : \gamma')'$, where $\text{vech}\left(\Omega\right) = \left(\text{vech}\left(\Omega_1\right)' : \cdots : \text{vech}\left(\Omega_m\right)'\right)'$. Separating θ into the subgroups $\text{vec}\left(b\right)$ and ϑ proves practical for deriving the asymptotic results because of the difference in the speed of convergence.

It turns out that ϑ is $T^{\frac{1}{2}}$- consistent while the different directions of b given by b_κ and b_τ are T- and $T^{\frac{3}{2}}$- consistent, respectively. In particular, in order to define the different normalizations, we define

$$W_{\text{vec}(b)T} := \mathcal{I}_r \otimes W_{bT} := \mathcal{I}_r \otimes \text{diag}\left(T\mathcal{I}_{n-r-1}, T^2\right), \quad W_T := T\text{diag}\left(W_{\text{vec}(b)T}, \mathcal{I}_{n_\vartheta}\right). \tag{1.17}$$

1.7 Likelihood analysis

We state the log-likelihood function as a function of the vector of parameters, θ. The Gaussian log-likelihood is given by $L_T\left(\theta\right) = \sum_{t=1}^T \ell_t\left(\theta\right)$ where

$$\ell_t\left(\theta\right) = \log\left(\sum_{j\in\mathbb{M}} p_{jt}\left(\theta\right)\phi_{jt}\left(\theta\right)\right), \tag{1.18}$$

and the function $\phi_{jt}\left(\theta\right)$ denotes the Gaussian density for a specific regime j and is given by

$$\log\phi_{jt}\left(\theta\right) = -\frac{1}{2}\left(n\log\left(2\pi\right) + \log\left|\Omega_j\right| + \varepsilon_{jt}'\Omega_j^{-1}\varepsilon_{jt}\right), \quad \text{with} \tag{1.19}$$

$$\begin{aligned}\varepsilon_{jt} &:= \Delta X_t - \alpha_j\beta^{\star\prime}X_{t-1}^\star - \Gamma_j\Delta\mathbb{X}_{t-1} \\ &= \Delta X_t - \alpha_j\left(\beta_0'X_{t-1} + b'\mathcal{X}_{t-1} + b_D'i_{n+1}'\right) - \Gamma_j\Delta\mathbb{X}_{t-1}\end{aligned} \tag{1.20}$$

In the following, we often omit to indicate explicitly that $\ell_t\left(\theta\right)$, $\phi_{jt}\left(\theta\right)$ and $p_{jt}\left(\theta\right)$ are functions of θ. When computing derivatives, it is useful to write $\ell_t = \log\left(\sum_{j\in\mathbb{M}}\exp\lambda_{jt}\right)$ where we define $\lambda_{jt} := \log\left(p_{jt}\phi_{jt}\right)$, which are well defined because $p_{jt}, \phi_{jt} > 0$ thanks to the model specification. In fact, it is simple to verify that $\partial_\theta\ell_t = \sum_{j\in\mathbb{M}} p_{jt}^\star\partial_\theta\lambda_{jt}$, where

$$p_{jt}^\star := \frac{p_{jt}\phi_{jt}}{\sum_{i\in\mathbb{M}} p_{it}\phi_{it}} = \Pr\left(s_t = j \mid Z_t, Z_{t-1}\right), \tag{1.21}$$

give the filtered probability of being in regime j at time t given past and current observables. Note the difference between the switching probability p_{jt} from (1.10) and the filtered probability,

$p^{\star}_{jt}$, given by (1.21). Estimation of the parameter can be done by direct numerical optimization of the likelihood or by using EM-algorithms such as those discussed in Bec and Rahbek (2004), Bec et al. (2008) and chapter two of this thesis, where we discuss estimation and bootstrap-based testing in detail.

1.7.1 Properties of the QMLE

The likelihood-based inference is based on the representation results in Theorem 1.4, which has a number of implications. In particular it implies convergence results for the derivatives of the log-likelihood function with respect to the parameters. To formalize them, first recall that $\partial_\theta \ell_t = \sum_{j\in\mathbb{M}} p^{\star}_{jt} \partial_\theta \lambda_{jt}$, and observe that, in particular,

$$\partial_{\mathrm{vec}(b')} \lambda_{jt} = h'_{vjt} \otimes \mathcal{X}_{t-1}, \qquad h_{vjt} := \alpha'_j \Omega_j^{-1} \varepsilon_{jt} + \psi'_\beta \left(\partial_z \log p_{jt}\right)'.$$

Remark that $\partial_{\mathrm{vec}(b')} \lambda_{jt}$ contains the cumulation of $h_{\kappa t} := \bar{\kappa}' \Delta X_{t-1}$, which is shown in Theorem 1.4 to be a stationary, mean zero function of the geometrically ergodic process. It is hence useful to define

$$h_t := \left(h'_{vt} : h'_{\kappa t} : h'_{\vartheta t}\right)', \qquad h_{vt} := \sum_{j\in\mathbb{M}} p^{\star}_{jt} h'_{vjt}, \quad h_{\vartheta t} := \sum_{j\in\mathbb{M}} p^{\star}_{jt} \partial_\vartheta \lambda_{jt}. \tag{1.22}$$

Specific expressions of $\partial_\cdot \lambda_{jt}$ for the various components of the parameter vector ϑ are reported in the Appendices.

We consider the properties of the estimator in a local neighborhood of θ_0, the true values of the parameter vector, θ, from (1.16). The following two theorems give convergence of the score and the asymptotic distribution of the maximum likelihood estimator.

Theorem 1.6. *Define h_t as in* (1.22) *calculated at θ_0, and its long-run variance $\Sigma := \Sigma(0) + \sum_{i=1}^{\infty} \left(\Sigma(i) + \Sigma'(i)\right)$ where $\Sigma(i) := \mathrm{cov}\left(h_t, h_{t+i}\right)$. With Σ positive definite, and under Assumptions 1.2-1.5 with $q = 2$, then as $T \to \infty$*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\cdot]} h_t \xrightarrow{w} \mathcal{B}(\cdot) := \left(\mathcal{B}'_v(\cdot), \mathcal{B}'_\kappa(\cdot), \mathcal{B}'_\vartheta(\cdot)\right)', \tag{1.23}$$

where, $\mathcal{B}(\cdot)$ is a Brownian motion with covariance Σ and the subscripts v, κ and ϑ correspond to the partitioning of h_t in (1.22).

Proof. The proof is given in Appendix 1.B. □

Theorem 1.7. *Let Assumptions 1.2-1.5 hold with $q = 3$; and let θ and W_T be defined as in* (1.16) *and* (1.17). *Then with θ_0 being the true value for θ, there exists a unique maximum point $\hat{\theta}$ of L_T in the neighborhood $\mathcal{N} = \left\{\theta : \left\| T^{-\frac{1}{2}} W_T^{\frac{1}{2}} \left(\hat{\theta} - \theta_0\right)\right\| < \epsilon\right\}$ for some $\epsilon > 0$, which satisfies*

$$W_T^{\frac{1}{2}} \left(\hat{\theta} - \theta_0\right) \xrightarrow{w} \mathbb{H}^{-1}\mathbb{S}, \qquad \textit{where} \quad \mathbb{S} := \begin{pmatrix} \mathrm{vec}\left(\int_0^1 F(s)\, \mathrm{d}\mathcal{B}_v(s)'\right) \\ \mathcal{B}_\vartheta(1) \end{pmatrix},$$

$$F(s) := \begin{pmatrix} \mathcal{B}_\kappa(s) \\ s \end{pmatrix}, \qquad \mathbb{H} := \begin{pmatrix} \int_0^1 F(s) F(s)'\, \mathrm{d}s \otimes \Sigma_{vv} & \int_0^1 F(s)\, \mathrm{d}s \otimes \Sigma_{v\vartheta} \\ \int_0^1 F(s)'\, \mathrm{d}s \otimes \Sigma_{\vartheta v} & \Sigma_{\vartheta\vartheta} \end{pmatrix},$$

and Σ_{su} *are blocks of* Σ *conformable with* (1.23)*.*

Proof. The proof is given in Appendix 1.C. □

Contrary to the case of the linear cointegrated VAR model, the asymptotic distribution is not mixed Gaussian and hence there is no straight forward way to normalize the parameter estimates and obtain nuisance parameter free distributions of standard test statistics. This finding is equivalent to what has been found in other models of cointegration with non-linear adjustment, see inter alia Hansen and Seo (2002), Kristensen and Rahbek (2010, 2013) and Seo (2011). The consequence is that it is cumbersome to evaluate relevant statistics such as confidence intervals for the estimated parameters, since the distributions of these statistics must be simulated on a case by case basis. Note that if the cointegration relations are known from the outset, most asymptotics are standard normal and one avoids having to rely on simulation based techniques as is otherwise necessary, see Bec and Rahbek (2004); Bec et al. (2008). However, testing for a reduction in the number of regimes will be non-standard in general (regardless of whether or not β is considered fixed) due to the fact that some or all of the parameters in the switching probability function will vanish when imposing the null hypothesis. A special case of this test is the test for linearity, which essentially evaluates whether the number of regimes in a given model can be reduced to one. The test for linearity has been discussed extensively in the literature, see among others Davies (1987), Hansen (1996), Caner and Hansen (2001), Hansen and Seo (2002) and Kristensen and Rahbek (2013).

1.8 Conclusion

We have presented a series of novel extensions to the literature on the ACR model. More precisely, we have considered asymmetric regime-switching structures and likelihood-based inference when the cointegration relations are considered unknown and when the error covariance matrix is regime dependent. Our results on the asymptotic theory are in line with recent research on similar models and show that the asymptotic distribution of the QMLE in the cointegrated ACR model is non-standard and nuisance parameter dependent. In chapter two, we look an estimation algorithm designed for this framework and at a bootstrap resampling scheme for simulating the distributions of test statistics of interest.

1.A Proof of Theorem 1.4

The claim in (i) is shown in Lemma 1.8 below. The claim in (ii) follows by the LLN for geometrically ergodic processes, see Jensen and Rahbek (2007). For the claim in (iii), observe that

$$X_t = \left(\overline{\beta}\beta' + \overline{\beta}_\perp\beta'_\perp\right) X_t = \overline{\beta}\beta' X_t + \overline{\beta}_\perp\beta'_\perp \sum_{i=1}^{t} \Delta X_t + \overline{\beta}_\perp\beta'_\perp X_0,$$

Then add and subtract $\overline{\beta}_\perp\mu_2 t$ and $\overline{\beta}\mu_1$ to obtain

$$\begin{aligned} X_t &= \overline{\beta}_\perp\mu_2 t + \overline{\beta}_\perp \sum_{i=1}^{t} \left(\beta'_\perp \Delta X_i - \mu_2\right) + \overline{\beta}\beta' X_t + \overline{\beta}_\perp\beta'_\perp X_0 \\ &=: \tau t + \overline{\beta}_\perp \sum_{i=1}^{t} \xi_i + \overline{\beta}\left(v_t + \mu_1\right) + \overline{\beta}_\perp\beta'_\perp X_0 \end{aligned}$$

where $\tau := \overline{\beta}_\perp\mu_2$, $\xi_i := \beta'_\perp \Delta X_i - \mu_2$ and $v_t := \beta' X_t - \mu_1$. □

Lemma 1.8. *Under Assumptions 1.2-1.3,* $\mathbb{Y}_t = \left(Y'_t, ..., Y'_{t-k-1}\right)'$ *defined in* (1.7) *is a geometrically ergodic Markov Chain, with* $E\left\|\mathbb{Y}_t\right\|^{2q} < \infty$.

Proof. By similar arguments as in Bec and Rahbek (2004) and Assumptions 1.2-1.3, $\mathbb{Y}_t = \left(Y'_t, ..., Y'_{t-1}\right)'$ given in (1.7) is a Markov Chain on $\mathbb{R}^{(k-1)n_y}$ for which the drift criterion in Meyn and Tweedie (1993, Theorem 15.0.1 (iii)) can be applied, as it is irreducible, aperiodic and compact subsets, $\mathbb{K}$, of $\mathbb{R}^{(k-1)n_y}$ are small. Let $\|\cdot\|$ denote the Euclidean distance.

We use the drift function proposed by Saikkonen (2008) and Liebscher (2005),

$$g\left(y\right) = 1 + \sum_{l=0}^{N-1} \rho^{-2ql} \sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}} \left\|\mathbb{A}y\right\|^{2q}, \tag{1.24}$$

where N and q are positive integers; $q = 1$ corresponds to the choice in Saikkonen (2008) and $q = 1/2$ corresponds to the one in Liebscher (2005). Note also that the definition implies $g\left(y\right) - 1 > \left|\left|y\right|\right|^{2q}$.

For notational convenience, we set $E_y\left(\cdot\right) := E\left(\cdot|\mathbb{Y}_{t-1} = y\right)$. Moreover, we denote $\eta' y$ as y_η with η being a selection matrix that picks out those elements of $\mathbb{Y}_{t-1}$ which affect the probability of switching. Remark that for some $i \in \mathbb{M}_1$ as given by the definition of the ACR process, we can rewrite $\mathbb{Y}_t$ as

$$\begin{aligned} \mathbb{Y}_t &= \left(\mathbf{1}\left\{s_t \in \mathbb{M}_1\right\}\left(\mathbb{A}_t - \mathbb{A}_i\right) + \mathbb{A}_i\right)\mathbb{Y}_{t-1} + \mathbf{1}\left\{s_t \in \mathbb{M}_2\right\}\left(\mathbb{A}_t - \mathbb{A}_i\right)\overline{\eta}\eta'\mathbb{Y}_{t-1} + \mathbb{U}_t \\ &=: \mathbb{B}_t\mathbb{Y}_{t-1} + \mathbb{C}_t\eta'\mathbb{Y}_{t-1} + \mathbb{U}_t. \end{aligned}$$

We wish to evaluate $E_y\left[g\left(\mathbb{Y}_t\right)\right]$ and hence

$$E_y\left[\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}} \left\|\mathbb{A}\mathbb{Y}_t\right\|^{2q}\right] = E_y\left[\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}} \left\|\mathbb{A}\mathbb{B}_t y + \mathbb{A}\left(\mathbb{C}_t y_\eta + \mathbb{U}_t\right)\right\|^{2q}\right]. \tag{1.25}$$

First, note that since $i \in \mathbb{M}_1$, it holds by definition of $\mathbb{B}_t$ that $\mathbb{B}_t \in \mathcal{A}_{\mathbb{M}_1}$ and so for any power,

$n > 0$, we have

$$E_y\left[\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left\|\mathbb{A}\mathbb{B}_t y\right\|^{2n}\right] \leq \sup_{\mathbb{A}\in\mathcal{A}^{l+1}_{\mathbb{M}_1}}\left\|\mathbb{A}y\right\|^{2n}. \tag{1.26}$$

Next, recall that $\mathbb{U}_t := \sum_{j\in\mathbb{M}} \mathbf{1}\{s_t = j\} J(\mu + V_j)\epsilon_t$ such that with $E\left[\left\|\epsilon_t\right\|^{2n}\right] < \infty$ one has $E\left[\left\|\mathbb{U}_t\right\|^{2n}\right] < \infty$. One finds

$$\begin{aligned}
E_y\left[\left\|\mathbb{C}_t y_\eta + \mathbb{U}_t\right\|\right]^n &\leq E_y\left[\left\|\mathbb{C}_t y_\eta\right\| + \left\|\mathbb{U}_t\right\|\right]^n \\
&\leq c\sum_{i=0}^{n}\binom{n}{i}\left\|y_\eta\right\|^{n-i} E_y\left[\left\|\mathbb{C}_t\right\|^{n-i}\left\|\mathbb{U}_t\right\|^i\right] \\
&\leq c\sum_{i=0}^{n}\binom{n}{i}\left\|y_\eta\right\|^{n-i}\sqrt{E_y\left\|\mathbb{C}_t\right\|^{2(n-i)} E_y\left\|\mathbb{U}_t\right\|^{2i}} \\
&\leq c\sum_{i=0}^{n}\binom{n}{i}\left\|y_\eta\right\|^{n-i}\sqrt{q(y_\eta)} \\
&\leq c\left\|y_\eta\right\|^n\sqrt{q(y_\eta)}
\end{aligned} \tag{1.27}$$

where we have used $E_y\left[\left\|\mathbb{C}_t y_\eta\right\|^{n-i}\right] \leq \left\|y_\eta\right\|^{n-i} E_y\left[\left\|\mathbb{C}_t\right\|^{n-i}\right]$ and we have applied the notation $q(y_\eta) := 1 - p(y_\eta)$, with $p(y_\eta) = \Pr(s_{t+1} = 1|\eta'\mathbb{Y}_t = y_\eta)$, see below (1.2).

To ease notation in the following evaluation, we set $F_t = \mathbb{A}\mathbb{B}_t y$ and $G_t = \mathbb{A}\left(\mathbb{C}_t y_\eta + \mathbb{U}_t\right)$ and write,

$$\begin{aligned}
E_y\left[\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left\|\mathbb{A}\mathbb{Y}_t\right\|^{2q}\right] &= E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left(\left\|F_t + G_t\right\|^2\right)^q \\
&\leq E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left(\left\|F_t\right\|^2 + \left|\left(G_t + 2G_t\right)' G_t\right|\right)^q \\
&= E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\sum_{i=0}^{q}\binom{q}{i}\left\|F_t\right\|^{2(q-i)}\left|\left(G_t + 2F_t\right)' G_t\right|^i \\
&\leq \underbrace{E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left\|F_t\right\|^{2q}}_{(a)} + \underbrace{\sum_{i=1}^{q}\binom{q}{i} E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left(\left\|F_t\right\|^{2(q-i)}\left|\left(G_t + 2F_t\right)' G_t\right|^i\right)}_{(b)}
\end{aligned}$$

Term (a) is treated in (1.26). For the generic term in the sum in (b), one has, using $\left\|\mathbb{A}\mathbb{B}_t y\right\|^2 \leq \left\|\mathbb{B}_t y\right\|^2\left\|\mathbb{A}'\mathbb{A}\right\|$, that

$$\begin{aligned}
&E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left(\left\|F_t\right\|^{2(q-i)}\left|\left(G_t + 2F_t\right)' G_t\right|^i\right) \\
&= E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left(\left|\left(\mathbb{B}_t y\right)'\mathbb{A}'\mathbb{A}\mathbb{B}_t y\right|^{q-i}\left|\left(\left(\mathbb{C}_t y_\eta + \mathbb{U}_t\right) + 2\mathbb{B}_t y\right)'\mathbb{A}'\mathbb{A}\left(\mathbb{C}_t y + \mathbb{U}_t\right)\right|^i\right) \\
&\leq E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left(\left(\left\|\mathbb{B}_t y\right\|^2\left\|\mathbb{A}'\mathbb{A}\right\|\right)^{q-i}\left(\left\|\left(\mathbb{C}_t y_\eta + \mathbb{U}_t\right) + 2\mathbb{B}_t y\right\|\left\|\mathbb{A}'\mathbb{A}\right\|\left\|\mathbb{C}_t y_\eta + \mathbb{U}_t\right\|\right)^i\right) \\
&\leq \sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left\|\mathbb{A}'\mathbb{A}\right\|^q E_y\left(\left\|\mathbb{B}_t y\right\|^{2(q-i)}\left\|\left(\mathbb{C}_t y_\eta + \mathbb{U}_t\right) + 2\mathbb{B}_t y\right\|^i\left\|\mathbb{C}_t y_\eta + \mathbb{U}_t\right\|^i\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq cE_y\left(\left\|\mathbb{B}_t y\right\|^{2(q-i)}\left(\left\|\mathbb{C}_t y_\eta+\mathbb{U}_t\right\|+2\left\|\mathbb{B}_t y\right\|\right)^i\left\|\mathbb{C}_t y_\eta+\mathbb{U}_t\right\|^i\right)\\
&\leq c\sum_{k=0}^{i}\binom{i}{k}E_y\left(\left\|\mathbb{C}_t y_\eta+\mathbb{U}_t\right\|^{k+i}\left\|\mathbb{B}_t y\right\|^{2q-i-k}\right)\\
&\leq c\sum_{k=0}^{i}\binom{i}{k}\sqrt{E_y\left\|\mathbb{C}_t y_\eta+\mathbb{U}_t\right\|^{2(k+i)}E_y\left\|\mathbb{B}_t y\right\|^{2(2q-i-k)}}\\
&\leq c\sum_{k=0}^{i}\binom{i}{k}\sqrt{\left\|y_\eta\right\|^{2(k+i)}\sqrt{q(y_\eta)}\left\|y\right\|^{2(2q-i-k)}}\\
&= c\sum_{k=0}^{i}\binom{i}{k}\left\|y_\eta\right\|^{(k+i)}\left\|y\right\|^{(2q-i-k)}q(y_\eta)^{\frac{1}{4}}\\
&= c\left(\frac{\left\|y_\eta\right\|}{\left\|y\right\|}\right)^i\left\|y\right\|^{2q}q(y_\eta)^{\frac{1}{4}}\sum_{k=0}^{i}\binom{i}{k}\left(\frac{\left\|y_\eta\right\|}{\left\|y\right\|}\right)^k\\
&= c\left(\frac{\left\|y_\eta\right\|}{\left\|y\right\|}\right)^i\left\|y\right\|^{2q}q(y_\eta)^{\frac{1}{4}}\left(1+\frac{\left\|y_\eta\right\|}{\left\|y\right\|}\right)^i\leq c_i\left\|y\right\|^{2q}\left(\frac{\left\|y_\eta\right\|}{\left\|y\right\|}\right)^i q(y_\eta)^{\frac{1}{4}}
\end{aligned}
$$

Note that $\left\|y_\eta\right\|/\left\|y\right\|\leq 1$, so that $c(1+(\left\|y_\eta\right\|/\left\|y\right\|)^i)\leq 2c$ which gives the new constant. Hence,

$$
\begin{aligned}
(b) \;&=\; \sum_{i=1}^{q}\binom{q}{i}E_y\sup_{\mathbb{A}\in\mathcal{A}^l}\left(\left\|F_t\right\|^{2(q-i)}\left|\left(G_t+2F_t\right)'G_t\right|^i\right)\\
&\leq\; \sum_{i=1}^{q}\binom{q}{i}c_i\left\|y\right\|^{2q}\left(\frac{\left\|y_\eta\right\|}{\left\|y\right\|}\right)^i q(y_\eta)^{\frac{1}{4}}\\
&\leq\; q(y_\eta)^{\frac{1}{4}}\left\|y\right\|^{2q-1}\left\|y_\eta\right\|\sum_{i=1}^{q}\binom{q}{i}c_i\left(\frac{\left\|y_\eta\right\|}{\left\|y\right\|}\right)^{i-1}=cq(y_\eta)^{\frac{1}{4}}\left\|y\right\|^{2q-1}\left\|y_\eta\right\|=:u_l\left(y\right),
\end{aligned}
$$

where $(1+\left\|y_\eta\right\|/\left\|y\right\|)^{q-1}\max_i c_i\leq 2^{q-1}\max_i c_i$ which gives the new constant. Thus,

$$
E_y\sup_{\mathbb{A}\in\mathcal{A}^l_{\mathbb{M}_1}}\left\|\mathbb{A}\mathbb{Y}_t\right\|^{2q}\leq\sup_{\mathbb{A}\in\mathcal{A}^{l+1}_{\mathbb{M}_1}}\left\|\mathbb{A}y\right\|^{2q}+u_l(y).
$$

Next note that, it holds for the ACR cointegrated process that $\rho\left(\mathcal{A}_{\mathbb{M}_1}\right)<1$ and hence, we can find a real number $\rho\in\left(\rho\left(\mathcal{A}_{\mathbb{M}_1}\right),1\right)$ and an integer N such that $\sup_{\mathbb{A}\in\mathcal{A}^N_{\mathbb{M}_1}}\left\|\mathbb{A}\right\|\leq\rho^N$, see Liebscher (2005, eq: (33)). This implies from (1.26) that

$$
\sup_{\mathbb{A}\in\mathcal{A}^N_{\mathbb{M}_1}}\left\|\mathbb{A}y\right\|^{2q}\leq\rho^{2qN}\left\|y\right\|^{2q}. \tag{1.28}
$$

Define

$$
u(y):=\sum_{l=0}^{N-1}\rho^{-2ql}u_l(y)=cq(y_\eta)^{\frac{1}{4}}\left\|y\right\|^{2q-1}\left\|y_\eta\right\|.
$$

Taking conditional expectations in (1.24) and inserting (1.28) one finds

$$
E_y(g\left(\mathbb{Y}_t\right)-1)\leq\sum_{l=0}^{N-1}\rho^{-2ql}\left(\sup_{\mathbb{A}\in\mathcal{A}^{l+1}_{\mathbb{S}_1}}\left\|\mathbb{A}y\right\|^{2q}+u_l(y)\right)
$$

$$
\begin{aligned}
&= \sum_{l=0}^{N-1} \rho^{-2ql} \sup_{\mathbb{A}\in\mathcal{A}_{\mathbb{M}_1}^{l+1}} \|\mathbb{A}y\|^{2q} + u(y) \\
&= \sum_{l=1}^{N-1} \rho^{-2q(l-1)} \sup_{\mathbb{A}\in\mathcal{A}_{\mathbb{M}_1}^{l}} \|\mathbb{A}y\|^{2q} + \rho^{-2q(N-1)} \sup_{\mathbb{A}\in\mathcal{A}_{\mathbb{M}_1}^{N}} \|\mathbb{A}y\|^{2q} + u(y) \\
&\leq \sum_{l=1}^{N-1} \rho^{-2q(l-1)} \sup_{\mathbb{A}\in\mathcal{A}_{\mathbb{M}_1}^{l}} \|\mathbb{A}y\|^{2q} + \rho^{2q} \|y\|^{2q} + u(y) \\
&= \sum_{l=0}^{N-1} \rho^{-2q(l-1)} \sup_{\mathbb{A}\in\mathcal{A}_{\mathbb{M}_1}^{l}} \|\mathbb{A}y\|^{2q} + u(y) = \rho^{2q} (g(y) - 1) + u(y) \qquad (1.29) \\
&\leq \frac{\rho^{2q}+1}{2} (g(y) - 1) - \frac{1-\rho^{2q}}{2} ||y||^{2q} + u(y) \\
&= \frac{\rho^{2q}+1}{2} (g(y) - 1) + ||y||^{2q} \left(\frac{u(y)}{||y||^{2q}} - \frac{1-\rho^{2q}}{2} \right)
\end{aligned}
$$

In the last two lines we have used the fact that

$$
\begin{aligned}
\rho^{2q} (g(y) - 1) &= \frac{\rho^{2q}+1}{2} (g(y) - 1) - \frac{1-\rho^{2q}}{2} (g(y) - 1) \\
&\leq \frac{\rho^{2q}+1}{2} (g(y) - 1) - \frac{1-\rho^{2q}}{2} ||y||^{2q}
\end{aligned}
$$

because $g(y) - 1 \geq ||y||^{2q}$, see the definition of g in (1.24). Consider now

$$
\frac{u(y)}{||y||^{2q}} = c \frac{\|y\|^{2q-1}}{||y||^{2q}} q(y_\eta)^{\frac{1}{4}} \|y_\eta\| = c q(y_\eta)^{\frac{1}{4}} \frac{\|y_\eta\|}{\|y\|}
$$

which we represent as a product of $a(y) := cq(y_\eta)^{\frac{1}{4}}$ and $b(y) := \frac{\|y_\eta\|}{\|y_\eta\| + \|y_{\eta_\perp}\|}$, where $\|y\| = \|y_\eta\| + \|y_{\eta_\perp}\|$ is the orthogonal decomposition of $||y||$ on η and its orthogonal complement. When $\|y\| \to \infty$, either $\|y_\eta\|$ or $\|y_{\eta_\perp}\|$ diverge, or both. When $\|y_\eta\| \to \infty$ (irrespectively of whether $\|y_{\eta_\perp}\|$ diverges or not) one has $a(y) \to 0$ by definition of the ACR cointegrated process and $b(y)$ is bounded by 1; hence the product $a(y)b(y) \to 0$. When $\|y_{\eta_\perp}\| \to \infty$ and $\|y_\eta\|$ is bounded, one has that $b(y) \to 0$ and $a(y)$ is bounded, so that $a(y)b(y) \to 0$. Hence we conclude that $\frac{u(y)}{||y||^{2q}} \to 0$ as $\|y\| \to \infty$. This implies that one can define a compact set $\mathbb{K} = \{y : ||y|| \leq R\}$, with R large enough so for $||y|| > R$ one has that $\frac{u(y)}{||y||^{2q}} \leq \frac{1-\rho^{2q}}{2}$. Hence on $\mathbb{K}^c$ one has

$$
E_y g(\mathbb{Y}_t) < \frac{\rho^{2q}+1}{2} g(y).
$$

It follows that the condition in Meyn and Tweedie (1993, Theorem 15.0.1 (iii)) is satisfied and that $\mathbb{Y}_t$ is a geometrically ergodic Markov Chain with finite moments of order $2q$. This completes the proof. □

1.B Proof of Theorem 1.6

As in Kristensen and Rahbek (2010, Proof of Theorem 2), Theorem 1.6 holds by applying the functional central limit theorem (FCLT) in Meyn and Tweedie (1993, Theorems 17.4.2 and 17.4.4). Below, we first establish that h_t in (1.22) has mean zero, and that h_t is a function of a geometrically ergodic Markov Chain $\mathbb{Z}_t$, where the variance of h_t is bounded by the drift function that applies to $\mathbb{Z}_t$. These conditions are sufficient for Meyn and Tweedie (1993, Theorems 17.4.2 and 17.4.4) to hold, as in Kristensen and Rahbek (2010, Proof of Theorem 2).

We first state a lemma on the first (conditional) moments of quantities involving $p_{jt}^\star$ and p_{jt}, which are used in proving that h_t in (1.22) has mean zero.

Lemma 1.9. *With* $p_{jt}^\star := E\left(\mathbf{1}\left\{s_t = j\right\} \mid Z_t, Z_{t-1}\right)$, $p_{jt} := E\left(\mathbf{1}\left\{s_t = j\right\} \mid Z_{t-1}\right)$ *defined in* (1.21), (1.10) *and* $Z_t = (X_t'\beta, \Delta\mathbb{X}_t')'$, *the following holds:*

$$E\left(p_{jt}^\star \mid Z_{t-1}\right) = p_{jt}, \tag{1.30}$$

$$E\left(p_{jt}^\star \varepsilon_t \mid Z_{t-1}\right) = 0, \tag{1.31}$$

$$E\left(p_{jt}^\star \left(\varepsilon_{jt}\varepsilon_{jt}' - \Omega_j\right) \mid Z_{t-1}\right) = 0, \tag{1.32}$$

$$E\left(\sum_{j\in\mathbb{M}} p_{jt}^\star \left(\partial_\cdot \log p_{jt}\right)' \mid Z_{t-1}\right) = 0. \tag{1.33}$$

Moreover, with M_t *measurable with respect to* $A_t := (Z_t', Z_{t-1}')'$ *but not with respect to* Z_{t-1} *alone, we have*

$$E\left(\left(p_{jt}^\star - p_{jt}\right) M_t \mid Z_{t-1}\right) = \operatorname{cov}\left(\mathbf{1}\left\{s_t = j\right\}, M_t \mid Z_{t-1}\right). \tag{1.34}$$

Proof. Eq. (1.30) holds by applying iterated expectations. Next, because ε_{jt} in (1.20) is a function of A_t, and $\mathbf{1}\left\{s_t = j\right\} \cdot \varepsilon_{jt} = \mathbf{1}\left\{s_t = j\right\} \cdot V_j\epsilon_t$, one has

$$\begin{aligned} E\left(p_{jt}^\star \varepsilon_{jt} \mid Z_{t-1}\right) &= E\left(E\left(\mathbf{1}\left\{s_t = j\right\} \mid A_t\right)\varepsilon_{jt} \mid Z_{t-1}\right) = E\left(E\left(\mathbf{1}\left\{s_t = j\right\} \varepsilon_{jt} \mid A_t\right) \mid Z_{t-1}\right) \\ &= E\left(\mathbf{1}\left\{s_t = j\right\} V_j\epsilon_t \mid Z_{t-1}\right) = E\left(\mathbf{1}\left\{s_t = j\right\} \mid Z_{t-1}\right) V_j E\left(\epsilon_t \mid Z_{t-1}\right) = 0, \end{aligned}$$

where the last equality follows from the fact that ϵ_t is i.i.d. with 0 mean and ϵ_t and s_t are independent conditionally on Z_{t-1}. This proves (1.31). Similarly,

$$\begin{aligned} E\left(p_{jt}^\star \left(\varepsilon_{jt}\varepsilon_{jt}' - \Omega_j\right) \mid Z_{t-1}\right) &= E\left(E\left(\mathbf{1}\left\{s_t = j\right\} \mid A_t\right)\left(\varepsilon_{jt}\varepsilon_{jt}' - \Omega_j\right) \mid Z_{t-1}\right) \\ &= E\left(E\left(\mathbf{1}\left\{s_t = j\right\} \left(\varepsilon_{jt}\varepsilon_{jt}' - \Omega_j\right) \mid A_t\right) \mid Z_{t-1}\right) \\ &= E\left(\mathbf{1}\left\{s_t = j\right\}\left(V_j\epsilon_t\epsilon_t'V_j' - \Omega_j\right) \mid Z_{t-1}\right) \\ &= E\left(\mathbf{1}\left\{s_t = j\right\} \mid Z_{t-1}\right) E\left(V_j\epsilon_t\epsilon_t'V_j' - \Omega_j \mid Z_{t-1}\right) = 0, \end{aligned}$$

where the last equality follows from the fact that $\Omega_j = V_jV_j'$ and $\epsilon_t\epsilon_t'$ is i.i.d. with mean I and ϵ_t and s_t are independent conditionally on Z_{t-1}. This proves (1.32). Next we prove that

$$\sum_{j=1}^{m} p_{jt}^\star \partial_\cdot \log p_{jt} = \sum_{j=1}^{m-1}\left(\frac{1}{p_{jt}}\left(p_{jt}^\star - p_{jt}\right) + \frac{1}{p_{mt}}\sum_{i=1}^{m-1}\left(p_{jt}^\star - p_{jt}\right)\right)\partial_\cdot p_{jt}, \tag{1.35}$$

which implies (1.33) applying (1.30) because p_{jt} and $\partial_{\cdot} p_{jt}$ are measurable with respect to Z_{t-1}. Eq. (1.35) is proved noting that, because $p_{mt} = 1 - \sum_{i=1}^{m-1} p_{it}$, one has $\partial_{\cdot} \log p_{mt} = \partial_{\cdot} \log \left(1 - \sum_{i=1}^{m-1} p_{it}\right) = -\left(1 - \sum_{i=1}^{m-1} p_{it}\right)^{-1} \left(\sum_{i=1}^{m-1} \partial_{\cdot} p_{it}\right)$ and hence

$$\begin{aligned}
\sum_{j=1}^{m} p_{jt}^{\star} \partial_{\cdot} \log p_{jt} &= \sum_{j=1}^{m-1} \frac{p_{jt}^{\star}}{p_{jt}} \partial_{\cdot} p_{jt} - \sum_{j=1}^{m-1} \frac{1 - \sum_{i=1}^{m-1} p_{it}^{\star}}{1 - \sum_{i=1}^{m-1} p_{it}} \partial_{\cdot} p_{jt} \\
&= \sum_{j=1}^{m-1} \left(\frac{p_{jt}^{\star}}{p_{jt}} - \frac{1 - \sum_{i=1}^{m-1} p_{it}^{\star}}{1 - \sum_{i=1}^{m-1} p_{it}} \right) \partial_{\cdot} p_{jt} \\
&= \sum_{j=1}^{m-1} \left(\left(\frac{1}{p_{jt} p_{mt}} \right) \left(1 - \sum_{i=1}^{m-1} p_{it} \right) p_{jt}^{\star} - \left(1 - \sum_{i=1}^{m-1} p_{it}^{\star} \right) p_{jt} \right) \partial_{\cdot} p_{jt} \\
&= \sum_{j=1}^{m-1} \frac{1}{p_{jt} p_{mt}} \left(p_{jt}^{\star} - p_{jt} - \sum_{i=1}^{m-1} p_{it} \left(p_{jt}^{\star} - p_{jt} \right) + \sum_{i=1}^{m-1} \left(p_{it}^{\star} - p_{it} \right) p_{jt} \right) \partial_{\cdot} p_{jt} \\
&= \sum_{j=1}^{m-1} \left(\frac{1}{p_{jt}} \left(p_{jt}^{\star} - p_{jt} \right) + \frac{1}{p_{mt}} \sum_{i=1}^{m-1} \left(p_{jt}^{\star} - p_{jt} \right) \right) \partial_{\cdot} p_{jt}.
\end{aligned}$$

Finally consider

$$\begin{aligned}
\operatorname{cov}\left(\mathbf{1}\left\{s_t = j\right\}, M_t | Z_{t-1}\right) &= E\left(\left(\mathbf{1}\left\{s_t = j\right\} - E\left(\mathbf{1}\left\{s_t = j\right\} | Z_{t-1}\right)\right) M_t | Z_{t-1}\right) \\
&= E\left(E\left(\left(\mathbf{1}\left\{s_t = j\right\} - E\left(\mathbf{1}\left\{s_t = j\right\} | Z_{t-1}\right)\right) M_t | A_t\right) | Z_{t-1}\right) \\
&= E\left(E\left(\left(\mathbf{1}\left\{s_t = j\right\} - E\left(\mathbf{1}\left\{s_t = j\right\} | Z_{t-1}\right)\right) | A_t\right) M_t | Z_{t-1}\right) \\
&= E\left(\left(E\left(\mathbf{1}\left\{s_t = j\right\} | A_t\right) - E\left(\mathbf{1}\left\{s_t = j\right\} | Z_{t-1}\right)\right) M_t | Z_{t-1}\right) \\
&= E\left(\left(p_{jt}^{\star} - p_{jt}\right) M_t | Z_{t-1}\right),
\end{aligned}$$

which gives (1.34). □

Lemma 1.10. *Under Assumptions 1.2-1.5, h_t in (1.22) has 0 mean, $E(h_t) = 0$.*

Proof. Using results in Lemma 1.9,

$$\begin{aligned}
E(\partial_{\operatorname{vec}(b_D')} \lambda_t \mid Z_{t-1}) &= E(\sum_{j \in \mathbb{M}} p_{jt}^{\star} \varepsilon_{jt}' \mid Z_{t-1}) \Omega_j^{-1} \alpha_j = 0 \\
E\left(\partial_{\operatorname{vec}(\alpha_i)} \lambda_t \mid Z_{t-1}\right) &= \operatorname{vec}\left(\Omega_i^{-1} E\left(p_{it}^{\star} \varepsilon_{it} \mid Z_{t-1}\right) X_{t-1}^{\star\prime} \beta^{\star}\right)' = 0 \\
E\left(\partial_{\operatorname{vec}(\Gamma_i)} \lambda_t \mid Z_{t-1}\right) &= \operatorname{vec}\left(\Omega_i^{-1} E\left(p_{it}^{\star} \varepsilon_{it} \mid Z_{t-1}\right) \Delta \mathbb{X}_{t-1}'\right)' = 0 \\
E\left(\partial_{\operatorname{vec}(\Omega_i)} \lambda_t \mid Z_{t-1}\right) &= \operatorname{vec}\left(\frac{1}{2} E\left(p_{it}^{\star} \Omega_i^{-1} \left(\varepsilon_{it} \varepsilon_{it}' - \Omega_i\right) \Omega_i^{-1} \mid Z_{t-1}\right)\right)' = 0 \\
E\left(\partial_{\gamma} \lambda_t \mid Z_{t-1}\right) &= E\left(\sum_{j \in \mathbb{M}} p_{jt}^{\star} \left(\partial_{\gamma} \log p_{jt}\right) \mid Z_{t-1}\right)' = 0 \\
E\left(h_{bt} \mid Z_{t-1}\right) &= \sum_{j \in \mathbb{M}} \left(\alpha_j \Omega_j^{-1} E\left(p_{jt}^{\star} \varepsilon_{jt} \mid Z_{t-1}\right)\right) + E\left(\sum_{j \in \mathbb{M}} p_{jt}^{\star} \left(\partial_z \log p_{jt}\right)' \mid Z_{t-1}\right) = 0.
\end{aligned}$$

Finally, by Theorem 1.4 one has $E\left(\Delta X_t' \kappa\right) = 0$. □

Lemma 1.11. *Under Assumptions 1.2-1.5, h_t is a function of the Markov chain $\mathbb{Z}_t := \left(\mathbb{Y}_{t-1}', \upsilon_t', \epsilon_t'\right)' \in \mathbb{R}^{n(k-1)} \times [0,1] \times \mathbb{R}^n$, where $\mathbb{Y}_{t-1}$ is given in (1.7) and υ_t is i.i.d. and uniformly distributed on $[0,1]$. $\mathbb{Z}_t$ satisfies the drift criterion with drift function*

$$g_q(y, \upsilon, \epsilon) = g_q(y) + \overline{\upsilon}'\overline{\upsilon} + \overline{\epsilon}'\overline{\epsilon}. \tag{1.36}$$

where $\overline{\upsilon} := (\upsilon \otimes \upsilon)$, $\overline{\epsilon} := (\epsilon \otimes \epsilon)$, and $g_q(y)$ is the drift function used in the proof of Theorem 1.4; hence $\mathbb{Z}_t$ is a geometrically ergodic Markov Chain with bounded $2q$ moments. Moreover, for $q \geq 2$, one has $\|h(y, \upsilon, \epsilon)\|^2 \leq c\left(g_q(y, \upsilon, \epsilon)\right)$.

Proof. First we represent s_t as a function of υ_t and $\mathbb{Y}_{t-1}$, where υ_t is i.i.d. and uniformly distributed on $[0,1]$; we write $s_t = s(\mathbb{Y}_{t-1}, \upsilon_t)$ where

$$s(y, \upsilon) = \sum_{j \in \mathbb{M}} \left(j \cdot \mathbf{1}\left\{ \upsilon \in \left(\sum_{i=1}^{j-1} p_i(y) ; \sum_{i=1}^{j} p_i(y) \right] \right\} \right).$$

Next note that p_{jt} and $\partial_{\cdot} \log p_{jt}$ are functions of $\mathbb{Y}_{t-1}$; moreover, see (1.2), ΔX_t is a function of $\mathbb{Z}_t$. This proves that h_t is a function of $\mathbb{Z}_t$.

In order to show that $\mathbb{Z}_t$ is a geometrically ergodic Markov Chain, we apply the same strategy as in Lemma 1.8, using that $\mathbb{Z}_t := \left(\mathbb{Y}_{t-1}', \upsilon_t', \epsilon_t'\right)'$ is a time-homogenous Markov Chain satisfying the criteria given there, using the drift function $g_q(y, \upsilon, \epsilon)$ in (1.36). By similar arguments as in the proof of Theorem 1.4, it can be shown that the Markov Chain $\mathbb{Z}_t$ is geometrically ergodic and has finite moments of order $2q$. In order to prove $\|h(y, \upsilon, \epsilon)\|^2 \leq c\left(g_q(y, \upsilon, \epsilon)\right)$, first note that

$$g_q(y) = 1 + \|y\|^{2q} + \sum_{l=1}^{N-1} \rho^{-2ql} \sup_{\mathbb{A} \in \mathcal{A}_{\mathbb{M}_1}^l} \|\mathbb{A}y\|^{2q} \geq 1 + \|y\|^{2q}.$$

Inspection of $h(y, \upsilon, \epsilon)$ shows that each component of the h vector in $\|h(y, \upsilon, \epsilon)\|^q$ gives a contribution bounded by $c\left(\|y\|^{2q} + \|\epsilon\|^{2q} + 1\right)$, where the $2q$ exponent comes from the derivatives with respect to Ω_i. Hence $\|h(y, \upsilon, \epsilon)\|^q \leq c\left(\|y\|^{2q} + \|\epsilon\|^{2q} + 1\right)$, and, because, $\overline{\epsilon}'\overline{\epsilon} = \|\epsilon\|^4$, $\overline{\upsilon}'\overline{\upsilon} = \|\upsilon\|^4$ one finds

$$\|h(y, \upsilon, \epsilon)\|^2 \leq c\left(\|y\|^4 + \|\epsilon\|^4\right) \leq c\left(g_2(y) + \|\epsilon\|^4 + \|\upsilon\|^4\right) \leq c\left(g_q(y, \upsilon, \epsilon)\right).$$

□

1.C Proof of Theorem 1.7

We give the proof for Theorem 1.7, mimicking Bec and Rahbek (2004, Proof of Theorem 5) and Kristensen and Rahbek (2010, Proof of Theorem 5). For consistency, the conditions (i)-(iii) in Kristensen and Rahbek (2010, Lemma 11) are verified with $U_T = T^{-1}W_T$ and $Q_T(\theta) = T^{-1}L_T(\theta)$, where W_T is the weight matrix given in (1.17) and L_T is the log-likelihood function defined in (1.18). Condition (i) holds by the definition of the likelihood function and by Assumption 1.5. Conditions $(ii) - (iii)$ are verified in Lemmas 1.12, 1.20 and 1.26 below, which discuss weak limits of the score, hessian and third derivatives, respectively. Specifically, denoting

subsequent increments in θ as $\mathrm{d}\theta, \mathrm{d}\theta^{\dagger}, \mathrm{d}\theta^{\ddagger}$ and letting $\mathrm{d}\theta_T := W_T^{-\frac{1}{2}}\mathrm{d}\theta$, $\mathrm{d}b_T := T^{-\frac{1}{2}}W_{bT}^{-\frac{1}{2}}\mathrm{d}b$, one has

$$\mathrm{d}Q_T\left(\theta_0; U_T^{-\frac{1}{2}}\mathrm{d}\theta\right) = T^{-\frac{1}{2}}\mathrm{d}L_T\left(\theta_0; \mathrm{d}\theta_T\right) = o_p\left(1\right)$$
$$\mathrm{d}^2Q_T\left(\theta_0; U_T^{-\frac{1}{2}}\mathrm{d}\theta, U_T^{-\frac{1}{2}}\mathrm{d}\theta\right) = \mathrm{d}^2L_T\left(\theta_0; \mathrm{d}\theta_T, \mathrm{d}\theta_T^{\dagger}\right) \xrightarrow{w} H_{\infty}\left(\mathrm{d}\theta, \mathrm{d}\theta^{\dagger}\right)$$

and

$$\begin{aligned}\mathrm{d}^3Q_T\left(\theta; U_T^{-\frac{1}{2}}\mathrm{d}\theta, U_T^{-\frac{1}{2}}\mathrm{d}\theta, U_T^{-\frac{1}{2}}\mathrm{d}\theta\right) &= T^{\frac{1}{2}}\mathrm{d}^3L_T\left(\theta; \mathrm{d}\theta_T, \mathrm{d}\theta_T^{\dagger}, \mathrm{d}\theta_T^{\ddagger}\right)\\ &= O_p\left(\left\|\mathrm{d}\theta\right\|\left\|\mathrm{d}\theta^{\dagger}\right\|\left\|\mathrm{d}\theta^{\ddagger}\right\|\right).\end{aligned}$$

The verification of these conditions proves consistency.

The form of the asymptotic distribution is proved verifying condition (iv) in Kristensen and Rahbek (2010, Lemma 12), with $\nu_T = T$ such that $\nu_T^{\frac{1}{2}}U_T^{-\frac{1}{2}} = W_T^{-\frac{1}{2}}$. This condition follows from Lemma 1.12 which gives

$$\mathrm{d}Q_T\left(\theta_0; \nu_T^{\frac{1}{2}}U_T^{-\frac{1}{2}}\mathrm{d}\theta\right) = \mathrm{d}\log L_T\left(\theta_0; \mathrm{d}\theta_T\right) \xrightarrow{w} S_{\infty}\left(\mathrm{d}\theta\right).$$

Consequently, we have $W_T^{\frac{1}{2}}\left(\hat{\theta} - \theta_0\right) \xrightarrow{w} \mathrm{d}\theta_{\infty}$, where θ_{∞} satisfies $S_{\infty}\left(\mathrm{d}\theta\right) = H_{\infty}\left(\mathrm{d}\theta, \mathrm{d}\theta_{\infty}\right)$ for all directions $\mathrm{d}\theta$ and with $S_{\infty}\left(\mathrm{d}\theta\right)$ and $H_{\infty}\left(\mathrm{d}\theta, \mathrm{d}\theta_{\infty}\right)$ given in Lemmas 1.12 and 1.20. This completes the proof of Theorem 1.7.

Lemma 1.12. *Given Assumptions 1.2-1.5 apply with $q = 2$. Then*

$$\mathrm{d}L_T\left(\theta_0; \mathrm{d}\theta_T\right) =: S_T\left(\mathrm{d}\theta_T\right) \xrightarrow{w} S_{\infty}\left(\mathrm{d}\theta\right),$$

where $S_{\infty}\left(\mathrm{d}\theta\right)$ is a function of Brownian motions given by

$$S_{\infty}\left(\mathrm{d}\theta\right) := \left(\mathcal{B}_{\vartheta}\left(1\right)', \mathrm{vec}\left(\int_0^1 F\left(s\right)'\mathrm{d}\mathcal{B}_b\left(s\right)\right)'\right)'\mathrm{d}\theta =: \mathbb{S}\mathrm{d}\theta$$

with $F\left(s\right) := \left(\mathcal{B}_{\kappa}\left(s\right)', s, 1\right)'$, where $\mathcal{B}_{\kappa}$, and $\mathcal{B}_{\vartheta}$ are Brownian motions defined in Theorem 1.6.

Proof: The proof is divided into the following Claims 1.13-1.18, where we index the score conformably with the various parameters in θ.

Claim 1.13. $S_T(\mathrm{d}b_T) \xrightarrow{w} \mathrm{vec}\left(\int_0^1 F\left(s\right)\mathrm{d}\mathcal{B}_v\left(s\right)'\right)'\mathrm{vec}\left(\mathrm{d}b'\right)$.

Claim 1.14. $S_T(\mathrm{d}b_{DT}) \xrightarrow{w} \mathcal{B}_{\mathrm{vec}(\beta_D')}\left(1\right)'\mathrm{vec}\left(\mathrm{d}\left(b_D'\right)\right)$.

Claim 1.15. $S_T(\mathrm{d}\alpha_{iT}) \xrightarrow{w} \mathcal{B}_{\mathrm{vec}(\alpha_i)}\left(1\right)'\mathrm{vec}\left(\mathrm{d}\alpha_i\right)$.

Claim 1.16. $S_T\left(\mathrm{d}\Gamma_{iT}\right) \xrightarrow{w} \mathcal{B}_{\mathrm{vec}(\Gamma_i)}\left(1\right)'\mathrm{vec}\left(\mathrm{d}\Gamma_i\right)$.

Claim 1.17. $S_T\left(\mathrm{d}\Omega_{iT}\right) \xrightarrow{w} \mathcal{B}_{\mathrm{vech}(\Omega_i)}\left(1\right)'\mathrm{vech}\left(\mathrm{d}\Omega_i\right)$.

Claim 1.18. $S_T\left(\mathrm{d}\gamma_T\right) \xrightarrow{w} \mathcal{B}_{\gamma}\left(1\right)'\mathrm{d}\gamma$.

***Verification of Claim 1.13*:** Write $\beta = \beta_0 + (\bar{\kappa}_0 : \bar{\tau}_0)\, b$ and recall that $\lambda_{jt} = \log \phi_{jt} + \log p_{jt}$ and $z_{t-1} := \psi' Z_{t-1} =: \psi'_\beta \beta' X_{t-1} + \psi'_\Delta \Delta \mathbb{X}_{t-1}$, where $\psi = (\psi'_\beta : \psi'_\Delta)'$ is partitioned comfortably with $Z_t := \left((\beta' X_{t-1})' : \Delta \mathbb{X}'_{t-1} \right)'$. Differentiation gives

$$\begin{aligned} \mathrm{d}\lambda_{jt}\left(\mathrm{d}b'\right) &= \varepsilon'_{jt}\Omega_j^{-1}\alpha_j \mathrm{d}b' \mathcal{X}_{t-1} + \partial_z \log p_{jt} \psi'_\beta \mathrm{d}b' \mathcal{X}_{t-1} \\ &= \left(\varepsilon'_{jt}\Omega_j^{-1}\alpha_j + \partial_z \log p_{jt} \psi'_\beta \right) \mathrm{d}b' \mathcal{X}_{t-1} \end{aligned}$$

Hence with $h_{vt} := \sum_{j \in \mathbb{M}} p^\star_{jt} h_{vjt}$, one has $\mathrm{d}\ell_t\left(\mathrm{vec}(\mathrm{d}b')\right) = \left(\mathrm{vec}\left(\mathcal{X}_{t-1} h'_{vt}\right)\right)' \mathrm{vec}(\mathrm{d}b')$ so that, by Theorem 3.1 in Hansen (1992), and Theorem 1.6 one has

$$S_T(\mathrm{d}b_T) \xrightarrow{w} \mathrm{vec}\left(\int_0^1 F(s)\, \mathrm{d}\mathcal{B}_v(s)' \right)' \mathrm{vec}\left(\mathrm{d}b'\right).$$

***Verification of Claim 1.14* :** Similarly, consider $\beta^\star = \beta_0^\star + (\overline{\kappa}_0^\star : \overline{\tau}_0 : i_{n+1})\,(b', b'_D)'$ so that $\mathrm{d}\lambda_{jt}\left(\mathrm{d}b'_D\right) = \varepsilon'_{jt}\Omega_j^{-1}\alpha_j \mathrm{d}b'_D$. Hence

$$\mathrm{d}\ell_t\left(\mathrm{vec}(\mathrm{d}b'_D)\right) = \sum_{j \in \mathbb{M}} p^\star_{jt} \varepsilon'_{jt}\Omega_j^{-1}\alpha_j \mathrm{vec}(\mathrm{d}b'_D),$$

which implies $S_T(\mathrm{d}b_{DT}) \xrightarrow{w} \mathcal{B}_{\mathrm{vec}(\beta'_D)}(1)' \mathrm{vec}\left(\mathrm{d}\left(b'_D\right)\right)$ by Theorem 1.6.

Verification of Claims 1.15, 1.16, 1.17 and 1.18: Start by considering Claim 1.15. It holds by Lemma 1.31 that,

$$S_T\left(\mathrm{d}\alpha_{iT}\right) = T^{-\frac{1}{2}} \sum_{t=1}^{T} p^\star_{it} \left(X^{\star\prime}_{t-1}\beta^\star \otimes \varepsilon'_{it}\Omega_i^{-1} \right) \mathrm{vec}\left(\mathrm{d}\alpha_i\right).$$

which, by Theorem 1.6, implies $S_T\left(\mathrm{d}\alpha_T\right) \xrightarrow{w} \mathcal{B}_{\mathrm{vec}\ \alpha_i}(1)' \mathrm{vec}\left(\mathrm{d}\alpha\right)$. By similar arguments,

$$S_T\left(\mathrm{d}\Gamma_{iT}\right) = T^{-\frac{1}{2}} \sum_{t=1}^{T} p^\star_{it} \left(\Delta \mathbb{X}'_{t-1} \otimes \varepsilon'_{it}\Omega_i^{-1} \right) \mathrm{vec}\left(\mathrm{d}\Gamma_i\right) \xrightarrow{w} \mathcal{B}_{\mathrm{vec}\ \Gamma_i}(1)' \mathrm{vec}\left(\mathrm{d}\Gamma\right)$$

Next, consider Claim 1.17. By Lemma 1.31, one obtains

$$\begin{aligned} S_T\left(\mathrm{d}\Omega_{iT}\right) &= -T^{-\frac{1}{2}} \frac{1}{2} \sum_{t=1}^{T} p^\star_{it} \left(\mathrm{vec}\left(\Omega_i^{-1}\left(\varepsilon_{it}\varepsilon'_{it} - \Omega_i\right)\Omega_i^{-1} \right) \right)' \mathcal{D}_\Omega \mathrm{vech}\left(\mathrm{d}\Omega_i\right) \\ &\xrightarrow{w} \mathcal{B}_{\mathrm{vech}\ \Omega_i}(1)' \mathrm{vech}\left(\mathrm{d}\Omega_i\right) \end{aligned}$$

To show validity of Claim 1.18, observe that by Theorem 1.6

$$S_T\left(\mathrm{d}\gamma_T\right) = T^{-\frac{1}{2}} \sum_{t=1}^{T} \sum_{j \in \mathbb{M}} p^\star_{jt} \partial_\gamma \log p_{jt} \mathrm{d}\gamma \xrightarrow{w} \mathcal{B}_\gamma(1)'\, \mathrm{d}\gamma.$$

□

Before stating Lemma 1.20 and 1.26, we provide some auxiliary results.

Lemma 1.19. *Let* $\mathrm{d}\log p_{jt}^{u}$ *(respectively* $\mathrm{d}\log \phi_{jt}^{u}$*) indicate* $\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\theta^{u}\right)$ *(respectively* $\mathrm{d}\log \phi_{jt}^{u}$*) and* $\mathrm{d}^{2}\log p_{jt}^{uv}$ *(respectively* $\mathrm{d}^{2}\log \phi_{jt}^{uv}$*) indicate* $\mathrm{d}^{2}\log p_{jt}\left(\theta;\mathrm{d}\theta^{u};\mathrm{d}\theta^{v}\right)$ *(respectively* $\mathrm{d}^{2}\log \phi_{jt}\left(\theta;\mathrm{d}\theta^{u};\mathrm{d}\theta^{v}\right)$*), where* $\mathrm{d}\theta^{u},\mathrm{d}\theta^{v}$ *indicate any of* $\mathrm{d}\theta,\mathrm{d}\theta^{\dagger}$*; then at* $\theta=\theta_{0}$ *one has*

$$E\left(\sum_{j\in\mathbb{M}}p_{jt}^{\star}\left(\mathrm{d}\log p_{jt}^{u}\ \mathrm{d}\log p_{jt}^{v}+\mathrm{d}^{2}\log p_{jt}^{uv}\right)\middle|Z_{t-1}\right) = 0, \tag{1.37}$$

$$E\left(\sum_{j\in\mathbb{M}}p_{jt}^{\star}\left(\mathrm{d}\log \phi_{jt}^{u}\ \mathrm{d}\log \phi_{jt}^{v}+\mathrm{d}^{2}\log \phi_{jt}^{uv}\right)\middle|Z_{t-1}\right) = 0. \tag{1.38}$$

Proof. First note $\mathrm{d}^{2}\log p_{jt}^{uv}=\frac{1}{p_{jt}}\mathrm{d}^{2}p_{jt}^{uv}-\mathrm{d}\log p_{jt}^{u}\ \mathrm{d}\log p_{jt}^{v}$, such that

$$\sum_{j\in\mathbb{M}}p_{jt}^{\star}\left(\mathrm{d}\log p_{jt}^{u}\ \mathrm{d}\log p_{jt}^{v}+\mathrm{d}^{2}\log p_{jt}^{uv}\right)=\sum_{j\in\mathbb{M}}\left(p_{jt}^{\star}/p_{jt}\right)\mathrm{d}^{2}p_{jt}^{uv}.$$

Proceeding as in proof of Lemma 1.9, one finds that

$$\sum_{j=1}^{m}\frac{p_{jt}^{\star}}{p_{jt}}\mathrm{d}^{2}p_{jt}^{uv}=\sum_{j=1}^{m-1}\left(\frac{1}{p_{jt}}\left(p_{jt}^{\star}-p_{jt}\right)+\frac{1}{p_{mt}}\sum_{i=1}^{m-1}\left(p_{jt}^{\star}-p_{jt}\right)\right)\mathrm{d}^{2}p_{jt}^{uv}.$$

and hence, the result follows by applying conditional expectations and noting that $\mathrm{d}^{2}p_{jt}^{uv}$ is Z_{t-1} measurable. □

Lemma 1.20. *Provided Assumptions 1.2-1.5 apply, the following weak convergence result holds*

$$\mathrm{d}^{2}L_{T}\left(\theta_{0};\mathrm{d}\theta_{T},\mathrm{d}\theta_{T}^{\dagger}\right)=:H_{T}\left(\mathrm{d}\theta_{T},\mathrm{d}\theta_{T}^{\dagger}\right)\xrightarrow{w}H_{\infty}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger}\right),$$

where $H_{\infty}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger}\right)$ *is a non-degenerate distribution that depends on model specific nuisance parameters. With* θ *defined in (1.16),* $H_{\infty}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger}\right)$ *is given by*

$$\begin{aligned}H_{\infty}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger}\right)&=-\mathrm{d}\theta'\begin{pmatrix}\left(\int_{0}^{1}F\left(s\right)F\left(s\right)'\mathrm{d}s\otimes\Sigma_{vv}\right) & \left(\int_{0}^{1}F\left(s\right)\mathrm{d}s\otimes\Sigma_{v\theta}\right)\\ \left(\int_{0}^{1}F\left(s\right)'\mathrm{d}s\otimes\Sigma_{v\vartheta}\right) & \Sigma_{\vartheta\vartheta}\end{pmatrix}\mathrm{d}\theta^{\dagger}\\&=:-\mathrm{d}\theta'\mathbb{H}\mathrm{d}\theta^{\dagger}\end{aligned}$$

where $F\left(s\right)$ *was defined in Lemma 1.12.*

Proof: First note that the second order derivative of the log-likelihood function can be written as

$$\begin{aligned}\mathrm{d}^{2}L_{T}\left(\theta;\mathrm{d}\theta,\mathrm{d}\theta^{\dagger}\right)=&\sum_{t=1}^{T}\sum_{j\in\mathbb{M}}p_{jt}^{\star}\left\{\left(\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta\right)\right)'\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta^{\dagger}\right)+\mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger}\right)\right\}\\&-\sum_{t=1}^{T}\left(\sum_{j\in\mathbb{M}}p_{jt}^{\star}\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta\right)\right)\left(\sum_{j\in\mathbb{M}}p_{jt}^{\star}\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta^{\dagger}\right)\right)'.\end{aligned} \tag{1.39}$$

By Theorem 1.4 and the definition of λ_{jt} given in section 1.7, it holds that for the parameters collected in ϑ (cf. (1.16)), $\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta\right)$ and $\mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}\vartheta\right)$ are functions of the stationary and geometri-

cally ergodic Markov Chain $\mathbb{Z}_t$ given in Lemma 1.11. Hence, the law of large numbers in Jensen and Rahbek (2007) applies to (1.39) such that

$$\sum_{t=1}^{T}\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left\{\left(\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta_T\right)\right)'\left(\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta_T^{\dagger}\right)\right)+\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\vartheta_T,\mathrm{d}\vartheta_T^{\dagger}\right)\right\}\xrightarrow{p}0$$

and

$$-\sum_{t=1}^{T}\left(\sum_{j\in\mathbb{M}} p_{jt}^{\star}\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta_T\right)\right)\left(\sum_{j\in\mathbb{M}} p_{jt}^{\star}\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta_T^{\dagger}\right)\right)'\xrightarrow{p}-\mathrm{d}\vartheta'\Sigma_{\vartheta\vartheta}\mathrm{d}\vartheta,$$

where the first result holds by use of Lemma 1.35 and the second holds by Theorem 1.6 and the Continuous Mapping Theorem. Due to non-stationarity in the directions of $\mathrm{d}b$, the Claims 1.21-1.25 are verified one by one.

Claim 1.21. $H_T\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger}\right)\xrightarrow{w}-\mathrm{tr}\left\{\left(\mathrm{d}b^{\dagger}\right)'\int_0^1\left(F\left(s\right)F\left(s\right)'\mathrm{d}s\right)\mathrm{d}b\Sigma_{bb}\right\}$.

Claim 1.22. $H_T\left(\mathrm{d}b_T,\mathrm{d}\alpha_T^{\dagger}\right)\xrightarrow{w}-\mathrm{tr}\left\{\mathrm{vec}\left(\mathrm{d}\alpha_i^{\dagger}\right)\int_0^1\left(F\left(s\right)'\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\alpha}\right\}$.

Claim 1.23. $H_T\left(\mathrm{d}b_T,\mathrm{d}\Gamma_T^{\dagger}\right)\xrightarrow{w}-\mathrm{tr}\left\{\mathrm{vec}\left(\mathrm{d}\Gamma_i^{\dagger}\right)\int_0^1\left(F\left(s\right)'\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\Gamma}\right\}$.

Claim 1.24. $H_T\left(\mathrm{d}b_T,\mathrm{d}\Omega_{iT}^{\dagger}\right)\xrightarrow{w}\mathrm{tr}\left\{\mathrm{vech}\left(\mathrm{d}\Omega_i^{\dagger}\right)\int_0^1\left(F\left(s\right)'\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\Omega}\right\}$ for all $i\in\mathbb{M}$.

Claim 1.25. $H_T\left(\mathrm{d}b_T,\mathrm{d}\gamma_T^{\dagger}\right)\xrightarrow{w}-\mathrm{tr}\left\{\left(\mathrm{d}\gamma^{\dagger}\right)\int_0^1\left(F\left(s\right)'\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\gamma}\right\}$.

***Verification of Claim 1.21*:** By Lemma 1.32 we have that the second order derivative of the log-likelihood function in direction $\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger}\right)$, and evaluated in the true parameter, θ_0, is given by

$$\begin{aligned}
&H_T\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger}\right)\\
=&-\sum_{t=1}^{T}\mathcal{X}_{t-1}'\mathrm{d}b_T\varphi_{bt}\varphi_{bt}'\mathrm{d}b_T^{\dagger\prime}\mathcal{X}_{t-1}-T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}b\\
&\times\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left\{v_{jt}v_{jt}'-\alpha_{0j}\Omega_{0j}^{-1}\alpha_{0j}+\psi_{\beta}\left(\partial_{zz}^2\log p_{jt}\right)\psi_{\beta}'\right\}\mathrm{d}b^{\dagger\prime}W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}
\end{aligned}\tag{1.40}$$

where $v_{jt}:=\alpha_{0,j}'\Omega_{0,j}^{-1}\varepsilon_{jt}+\psi_{\beta}\left(\partial_z\log p_{jt}\right)'$ and $\varphi_{bt}:=\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left(\alpha_j'\Omega_j^{-1}\varepsilon_{jt}+\psi_1\left(\partial_z\log p_{jt}\right)'\right)$. Moreover, we define

$$f_t:=\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left(v_{jt}v_{jt}'-\alpha_{0j}'\Omega_{0j}^{-1}\alpha_{0j}+\psi_{\beta}\left(\partial_{zz}^2\log p_{jt}\right)\psi_{\beta}'\right).$$

The expectation of f_t conditional on Z_{t-1} is given by:

$$\begin{aligned}
E\left[f_t\mid Z_{t-1}\right] &= E\left[\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left\{\alpha_{0j}'\Omega_{0j}^{-1}\alpha_{0j}+\psi_{\beta}\left(\partial_z\log p_{jt}\right)'\left(\partial_z\log p_{jt}\right)\psi_{\beta}'\right\}\mid Z_{t-1}\right]\\
&\quad+E\left[-\alpha_{0j}'\Omega_{0j}^{-1}\alpha_{0j}+\psi_{\beta}\left(\partial_{zz}^2\log p_{jt}\right)\psi_{\beta}'\mid Z_{t-1}\right]\\
&= \sum_{j\in\mathbb{M}} p_{jt}\left\{\psi_{\beta}\left(\partial_z\log p_{jt}\right)'\left(\partial_z\log p_{jt}\right)\psi_{\beta}'-\psi_{\beta}\left(\partial_{zz}^2\log p_{jt}\right)'\psi_{\beta}'\right\}=0.
\end{aligned}$$

where we have used Lemma 1.9. Next, observe that

$$\mathcal{X}'_{[Ts]}\mathrm{d}b = \begin{pmatrix} X'_{[Ts]}\overline{\kappa}_0 \\ X'_{[Ts]}\overline{\tau}_0 \end{pmatrix}' (\mathrm{d}b)$$

with $\overline{\kappa}_0$, $\overline{\tau}_0$ given in section 1.6.1 and with $s \in [0;1]$. Using the results given in *Verification of Claim 1.13* along with the Continuous Mapping Theorem, we get that

$$W_{bT}^{-\frac{1}{2}}\mathcal{X}_{[Ts]}\mathcal{X}'_{[Ts]}W_{bT}^{-\frac{1}{2}} \xrightarrow{w} \begin{pmatrix} \mathcal{B}_\kappa(s) \\ s \end{pmatrix}' \begin{pmatrix} \mathcal{B}_\kappa(s) \\ s \end{pmatrix}.$$

Since f_t has mean zero and is a function of the stationary and geometrically ergodic $\mathbb{Z}_{t-1}$ (cf. Lemma 1.11), it holds by Kristensen and Rahbek (2010, Lemma 13) that

$$\sup_m E\left|E\left[f_t \mid Z_{t-m}\right]\right| \to 0 \text{ as } m \to \infty.$$

Then by Hansen (1992, Theorem 3.3), we have

$$\sup_{0\le s\le 1}\left|T^{-1}\sum_{t=1}^{\lceil Ts\rceil} W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}f_t\right| \xrightarrow{p} 0.$$

Thus, it is verified that,

$$-T^{-1}\sum_{t=1}^{T}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\left(f_t\right)\mathrm{d}b^{\dagger\prime}W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1} = o_p(1).$$

Rewrite $\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\varphi_{bt}\varphi'_{bt}\mathrm{d}b^{\dagger\prime}W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}$ using the definition $\omega_t = \varphi_{bt}\varphi'_{bt} - E\left[\varphi_{bt}\varphi'_{bt}\right] := \varphi_{bt}\varphi'_{bt} - \Sigma_{bb}$ and the trace operator to obtain:

$$\begin{aligned} & -T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\varphi_{bt}\varphi'_{bt}\mathrm{d}b^{\dagger}W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1} \\ = \; & -\mathrm{tr}\left\{\mathrm{d}b^{\dagger\prime}T^{-1}\sum_{t=1}^{T}W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\Sigma_{bb}\right\} \\ & -\mathrm{tr}\left\{\mathrm{d}b^{\dagger\prime}T^{-1}\sum_{t=1}^{T}W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\omega_t\right\} \end{aligned} \tag{1.41}$$

First, observe that since ω_t is a function of the stationary process, $\mathbb{Z}_{t-1}$ (cf. Lemma 1.11), and has mean zero, it applies by Kristensen and Rahbek (2010, Lemma 13) that,

$$\sup_m E\left|E\left[\omega_t \mid Z_{t-m}\right]\right| \to 0 : \text{as} : m \to \infty.$$

Consequently, by Hansen (1992, Theorem 3.3) and

$$\sup_{0\le s\le 1}\left|\mathrm{tr}\left\{\left(\mathrm{d}b^{\dagger}\right)'T^{-1}\sum_{t=1}^{[Ts]}W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\omega_t\right\}\right| \xrightarrow{p} 0,$$

demonstrating that the first term in (1.41) is $o_p(1)$. Collecting the results, we obtain the weak convergence result,

$$H_T\left(\mathrm{d}b_T, \mathrm{d}b_T^\dagger\right) \xrightarrow{w} -\mathrm{tr}\left\{\left(\mathrm{d}b^\dagger\right)' \int_0^1 \left(F(s) F'(s)\,\mathrm{d}s\right) \mathrm{d}b \Sigma_{bb}\right\}$$

as was desired.

***Verification of Claim 1.22 and 1.23*:** By Lemma 1.32, the second order derivative in direction $\left(\mathrm{d}b_T, \mathrm{d}\alpha_{i,T}^\dagger\right)$ is given by

$$\begin{aligned} H_T\left(\mathrm{d}b_T, \mathrm{d}\alpha_{i,T}^\dagger\right) &= \mathbf{1}\{j=i\} T^{-1} \sum_{t=1}^T \mathcal{X}_{t-1}' W_{bT}^{-\frac{1}{2}} \mathrm{d}b p_{it}^\star \left(\alpha_{0i}' \left(\Omega_{0i}^{-1} \varepsilon_{it} \varepsilon_{it}' - I\right) \Omega_{0i}^{-1} \mathrm{d}\alpha_i^\dagger \beta_0^{\star\prime} X_{t-1}^\star\right) \\ &\quad + \mathbf{1}\{j=i\} T^{-1} \sum_{t=1}^T \mathcal{X}_{t-1}' W_{bT}^{-\frac{1}{2}} \mathrm{d}b p_{it}^\star \left(\partial_z \log p_{it}\right)' \varepsilon_{it} \Omega_{0i}^{-1} \mathrm{d}\alpha_i^\dagger \beta_0^{\star\prime} X_{t-1}^\star \\ &\quad - T^{-1} \sum_{t=1}^T \mathcal{X}_{t-1}' W_{bT}^{-\frac{1}{2}} \mathrm{d}b \varphi_{bt} \varphi_{\alpha_i t}' \mathrm{vec}\left(\mathrm{d}\alpha_i^\dagger\right), \end{aligned} \tag{1.42}$$

with $\varphi_{\alpha_i t} := p_{it}^\star \left(\beta' X_{t-1}^\star \otimes \varepsilon_{it}' \Omega_i^{-1}\right)$. Consider the conditional expectation of the three first terms, for which Lemma 1.9 can be used along with the assumption of conditional independence between s_t and ε_t given in (1.3) to show that

$$E\left[p_{it}^\star \left(\alpha_{0i}' \left(\Omega_{0i}^{-1} \varepsilon_{it} \varepsilon_{it}' - I\right) + \psi_\beta \left(\partial_z \log p_{it}\right)' \varepsilon_{it}\right) \mid Z_{t-1}\right] = 0.$$

Consequently, we may define a mean zero sequence, f_t, as

$$f_t := p_{it}^\star \left(\alpha_{0i}' \left(\Omega_{0i}^{-1} \varepsilon_{it} \varepsilon_{it}' - I\right) + \psi_\beta \left(\partial_z \log p_{it}\right)' \varepsilon_{it}\right) \Omega_{0i}^{-1} \mathrm{d}\alpha_i^\dagger \beta_0^{\star\prime} X_{t-1}^\star.$$

Next, since f_t is a function of the stationary and geometrically ergodic process, $\mathbb{Z}_{t-1}$, and has mean zero; then by Kristensen and Rahbek (2010, Lemma 13) it also holds that

$$\sup_{0\le s\le 1} E\left|E\left[f_t \mid Z_{t-m}\right]\right| \to 0 \text{ as } m \to \infty$$

and hence Hansen (1992, Theorem 3.3) can once again be applied to ensure

$$\sup_{0\le s\le 1} \left| T^{-1} \sum_{t=1}^{\lceil Ts \rceil} \mathcal{X}_{t-1} W_{bT}^{-\frac{1}{2}} f_t \right| \xrightarrow{p} 0,$$

such that the first three terms disappear asymptotically. Now, consider the second term in (1.42) which, using the definition $\omega_t := \varphi_{bt}\varphi_{\alpha_i t}' - E\left[\varphi_{bt}\varphi_{\alpha_i t}'\right] = \varphi_{bt}\varphi_{\alpha_i t}' - \Sigma_{b\alpha_i}$ can be rewritten as

$$-T^{-1} \sum_{t=1}^T \mathcal{X}_{t-1}' \mathrm{d}b \varphi_{bt} \varphi_{\alpha_i t}' \mathrm{vec}\left(\mathrm{d}\alpha_i^\dagger\right) = -\mathrm{tr}\left\{\mathrm{vec}\left(\mathrm{d}\alpha_i^\dagger\right) T^{-1} \sum_{t=1}^T \mathcal{X}_{t-1}' W_{bT}^{-\frac{1}{2}} \mathrm{d}b_T \Sigma_{b\alpha_i}\right\}$$

$$-\mathrm{tr}\left\{\mathrm{vec}\left(\mathrm{d}\alpha_i^{\dagger}\right) T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}b_T\omega_t\right\}.$$

By similar arguments as above we obtain

$$-T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}b\varphi_{bt}\varphi_{\alpha_i t}'\mathrm{vec}\left(\mathrm{d}\alpha_i^{\dagger}\right)\overset{w}{\to}-\mathrm{tr}\left\{\mathrm{vec}\left(\mathrm{d}\alpha_i^{\dagger}\right)\int_0^1\left(F'\left(s\right)\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\alpha_i}\right\}.$$

Using $\alpha:=(\alpha_1:\cdots:\alpha_m)$, we have

$$H_T\left(\mathrm{d}b_T,\mathrm{d}\alpha_T^{\dagger}\right)\overset{w}{\to}-\mathrm{tr}\left\{\mathrm{vec}\left(\mathrm{d}\alpha^{\dagger}\right)\int_0^1\left(F'\left(s\right)\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\alpha}\right\}$$

where

$$\Sigma_{b\alpha}=[\Sigma_{b\alpha_1}:\cdots:\Sigma_{b\alpha_m}]$$

and $\Sigma_{b\alpha}=Cov\left(\varphi_{bt},\varphi_{\alpha t}\right)$. By similar arguments, it can be shown that for $\Gamma:=(\Gamma_1:...:\Gamma_m)$, one has

$$H_T\left(\mathrm{d}b_T,\mathrm{d}\Gamma_T^{\dagger}\right)\overset{w}{\to}-\mathrm{tr}\left\{\mathrm{vec}\left(\mathrm{d}\Gamma^{\dagger}\right)\int_0^1\left(F'\left(s\right)\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\Gamma}\right\}$$

with

$$\Sigma_{b\Gamma}=[\Sigma_{b\Gamma_1}:\cdots:\Sigma_{b\Gamma_m}]$$

and $\Sigma_{b\Gamma}=Cov\left(\varphi_{bt},\varphi_{\Gamma t}\right)$.

***Verification of Claim 1.24*:** From Lemma 1.32, we have that

$$\begin{aligned}H_T\left(\mathrm{d}b_T,\mathrm{d}\Omega_{i,T}^{\dagger}\right) &= -T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}bp_{it}^{\star}\left(\alpha_{0j}'\Omega_{0j}^{-1}\varepsilon_{jt}+\psi_{\beta}\left(\partial_z\log p_{jt}\right)'\right)\\ &\quad\times\mathrm{tr}\left\{\frac{1}{2}\Omega_{0,i}^{-1}\left(\Omega_{0,i}-\varepsilon_{it}\varepsilon_{it}'\right)\Omega_{0i}^{-1}\mathrm{d}\Omega_i^{\dagger}\right\}\\ &\quad+T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}bp_{it}^{\star}\left(\alpha_{0j}'\Omega_{0j}^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_{0j}^{-1}\varepsilon_{jt}\right)\\ &\quad+T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}b\varphi_{bt}\varphi_{\Omega_i t}'\mathcal{D}_{\Omega}\mathrm{vech}\left(\mathrm{d}\Omega_i^{\dagger}\right),\end{aligned}$$

with $\varphi_{\Omega_i t}:=\frac{1}{2}p_{it}^{\star}\left(\mathrm{vec}\left(\Omega_i^{-1}\left(\varepsilon_{it}\varepsilon_{it}'-\Omega_i\right)\Omega_i^{-1}\right)\right)'\mathcal{D}_{\Omega}$. First note that by independence of ε_{jt} and s_t, and and using Lemmas 1.9 along with the Assumption 1.2 (in particular symmetry of ϵ_t), we have

$$E\left[p_{it}^{\star}\alpha_{0i}'\Omega_{0i}^{-1}\varepsilon_{it}\mathrm{tr}\left\{\frac{1}{2}\Omega_{0i}^{-1}\left(\Omega_{0i}-\varepsilon_{it}\varepsilon_{it}'\right)\Omega_{0i}^{-1}\mathrm{d}\Omega_i^{\dagger}\right\}\mid Z_{t-1}\right]=0.$$

By similar arguments, it holds that

$$E\left[\psi_{\beta}\left(\partial_z\log p_{it}\right)'p_{it}^{\star}\mathrm{tr}\left\{\frac{1}{2}\Omega_{0i}^{-1}\left(\Omega_{0i}-\varepsilon_{it}\varepsilon_{it}'\right)\Omega_{0i}^{-1}\mathrm{d}\Omega_i^{\dagger}\right\}\mid Z_{t-1}\right]=0,$$

and

$$E\left[p_{it}^{\star}\left(\alpha_{0i}'\Omega_{0i}^{-1}\mathrm{d}\Omega_i^{\dagger}\Omega_{0i}^{-1}\varepsilon_{it}\right) \mid Z_{t-1}\right] = 0,$$

such that the first two terms has conditional expectation zero. Hence, we can define the stationary sequence

$$f_t := -p_{it}^{\star}\left(\alpha_{0i}'\Omega_{0i}^{-1}\varepsilon_{it} + \psi_{\beta}'\left(\partial_z \log p_{it}\right)'\right)\mathrm{tr}\left\{\frac{1}{2}\Omega_{0i}^{-1}\left(\Omega_{0i} - \varepsilon_{it}\varepsilon_{it}'\right)\Omega_{0i}^{-1}\mathrm{d}\Omega_i^{\dagger}\right\}$$

with mean zero. Again, the mixing property,

$$\sup_{0\leq s\leq 1} E\left|E\left[f_t \mid Z_{t-m}\right]\right| \to 0 \text{ as } m \to \infty$$

holds, since f_t is stationary with mean zero. Hence, Theorem 3.3. from Hansen (1992) can be applied to ensure

$$\sup_{0\leq s\leq 1}\left|T^{-1}\sum_{t=1}^{\lceil Ts\rceil}\mathcal{X}_{t-1}W_{bT}^{-\frac{1}{2}}f_t\right| \xrightarrow{p} 0,$$

showing that the first two terms are zero asymptotically. Again, define $\omega_t = \varphi_{bt}\varphi_{\Omega_i t}' - E\left[\varphi_{bt}\varphi_{\Omega_i t}'\right] = \varphi_{bt}\varphi_{\Omega_i t}' - \Sigma_{b\Omega_i}$and observe that by similar arguments as before, we have,

$$\begin{aligned} &T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}b\varphi_{bt}\varphi_{\Omega_i t}'\mathrm{vech}\left(\mathrm{d}\Omega_i^{\dagger}\right) \\ \xrightarrow{w}&\mathrm{tr}\left\{\mathrm{vech}\left(\mathrm{d}\Omega_i^{\dagger}\right)\int_0^1\left(F'\left(s\right)\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\Omega_i}\right\}\end{aligned}$$

which holds for all $i \in \mathbb{M}$ as desired.

Verification of Claim 1.25: Consider the second order derivative of the log-likelihood function in direction $\left(\mathrm{d}b_T, \mathrm{d}\gamma_T^{\dagger}\right)$,

$$\begin{aligned} H_T\left(\mathrm{d}b_T, \mathrm{d}\gamma_T\right) &= T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}b\left(\varphi_{bt}\left(\partial_{\gamma}\log p_t\right) + \sum_{j\in\mathbb{M}}p_{jt}^{\star}\psi_{\beta}\left(\partial_{z\gamma}^2\log p_t\right)\right)\mathrm{d}\gamma^{\dagger} \\ &\quad -T^{-1}\sum_{t=1}^{T}\mathcal{X}_{t-1}'W_{bT}^{-\frac{1}{2}}\mathrm{d}b\varphi_{bt}\varphi_{\gamma t}'\mathrm{d}\gamma^{\dagger}\end{aligned}$$

Next, observe that by use of Lemmas 1.9 and 1.19,

$$\begin{aligned} &E\left[\varphi_{bt}\left(\partial_{\gamma}\log p_t\right) + \sum_{j\in\mathbb{M}}p_{jt}^{\star}\psi_{\beta}\left(\partial_{z\gamma}^2\log p_t\right) \mid Z_{t-1}\right] \\ =\ &E\left[\sum_{j\in\mathbb{M}}p_{jt}^{\star}\alpha_{0j}'\Omega_{0j}^{-1}\varepsilon_{jt}\left(\partial_{\gamma}\log p_t\right) + \left\{\psi_{\beta}\left(\partial_z\log p_{jt}\right)'\left(\partial_{\gamma}\log p_t\right) + \psi_{\beta}\left(\partial_{z\gamma}^2\log p_t\right)\right\} \mid Z_{t-1}\right] \\ =\ &\sum_{j\in\mathbb{M}}\alpha_{0j}'\Omega_{0j}^{-1}E\left[p_{jt}^{\star}\varepsilon_{jt} \mid Z_{t-1}\right]\left(\partial_{\gamma}\log p_t\right)\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{j\in\mathbb{M}} p_{jt}\left\{\psi_\beta\left(\partial_z \log p_{jt}\right)'\left(\partial_\gamma \log p_t\right) + \psi_\beta\left(\partial^2_{z\gamma}\log p_t\right)\right\} \\
= \;& 0.
\end{aligned}
$$

Hence, we define a stationary sequence we mean zero,

$$
f_t := \left(v_t\left(\partial_\gamma \log p_t\right) + \sum_{j\in\mathbb{S}} p^\star_{jt}\psi_\beta\left(\partial^2_{z\gamma}\log p_t\right)\right).
$$

It holds by similar arguments as above that

$$
\sup_{0\leq s\leq 1} E\left|E\left[f_t \mid Z_{t-m}\right]\right| \to 0 \text{ as } m\to\infty
$$

and thereby

$$
\sup_{0\leq s\leq 1}\left|T^{-1}\sum_{t=1}^{\lceil Ts\rceil}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}f_t\right| \xrightarrow{p} 0,
$$

which shows that the first term is zero asymptotically. For the final term, we make use of the definition $\omega_t = \varphi_{bt}\varphi'_{\gamma t} - \Sigma_{b\gamma}$ and write

$$
\begin{aligned}
-T^{-\frac{1}{2}}\sum_{t=1}^{T}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\varphi_{bt}\varphi'_{\gamma t}\mathrm{d}\gamma^\dagger \;=\; & -\mathrm{tr}\left\{\mathrm{d}\gamma^\dagger T^{-1}\sum_{t=1}^{T}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\Sigma_{vu}\right\} \\
& -\mathrm{tr}\left\{\mathrm{d}\gamma^\dagger T^{-1}\sum_{t=1}^{T}\mathcal{X}'_{t-1}W_{bT}^{-\frac{1}{2}}\mathrm{d}b\omega_t\right\}.
\end{aligned}
$$

By similar arguments as above it holds that

$$
H_T\left(\mathrm{d}b_T, \mathrm{d}\gamma_T^\dagger\right) \xrightarrow{w} -\mathrm{tr}\left\{\mathrm{d}\gamma^\dagger\int_0^1\left(F'\left(s\right)\mathrm{d}s\right)\mathrm{d}b\Sigma_{b\gamma}\right\}
$$

which was desired. □

Lemma 1.26. *Given Assumptions 1.2-1.5, we have that*

$$
\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\mathrm{d}^3L_T\left(\theta;\mathrm{d}\theta_T,\mathrm{d}\theta_T^\dagger,\mathrm{d}\theta_T^\ddagger\right)\right| = O_p\left(\|d\theta\|\left\|d\theta^\dagger\right\|\left\|d\theta^\ddagger\right\|\right) \tag{1.43}
$$

for the neighborhoods of θ_0 given by

$$
\mathcal{N}_T\left(\theta_0\right) = \left\{\theta : \left\|T^{-\frac{1}{2}}W_T^{\frac{1}{2}}\left(\theta-\theta_0\right)\right\| < e\right\},
$$

where W_T is defined in (1.17).

Proof. Observe initially that by Lemma 1.33, the third order derivative of the log-likelihood function is given by

$$
\begin{aligned}
\mathrm{d}^3L_T\left(\theta;\mathrm{d}\theta,\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right) = \sum_{t=1}^{T}\sum_{j\in\mathbb{M}}\Big\{ & \mathrm{d}^2p^\star_{jt}\left(\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta\right) + \mathrm{d}p^\star_{jt}\left(\mathrm{d}\theta^\dagger\right)\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\ddagger\right) \\
& + \mathrm{d}p^\star_{jt}\left(\mathrm{d}\theta^\ddagger\right)\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\dagger\right) + p^\star_{jt}\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)\Big\},
\end{aligned} \tag{1.44}
$$

where $p^{\star}_{jt}$ and λ_{jt} are given in section 1.7. It holds that,

$$\mathrm{d}p^{\star}_{jt}\left(\mathrm{d}\theta^{\dagger}\right) = p^{\star}_{jt}\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta^{\dagger}\right) - p^{\star}_{jt}\sum_{i\in\mathbb{M}} p^{\star}_{it}\mathrm{d}\lambda_{it}\left(\mathrm{d}\theta^{\dagger}\right),$$

$$\begin{aligned}\mathrm{d}^2 p^{\star}_{jt}\left(\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right) = &\sum_{i\in\mathbb{M}_{-j}}\left(\mathrm{d}p^{\star}_{jt}(\mathrm{d}\theta^{\ddagger})p^{\star}_{it} + p^{\star}_{jt}\mathrm{d}p^{\star}_{it}(\mathrm{d}\theta^{\ddagger})\right)\left(\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta^{\dagger}\right) - \mathrm{d}\lambda_{it}\left(\mathrm{d}\theta^{\dagger}\right)\right)\\ &+ p^{\star}_{jt}\sum_{i\in\mathbb{M}_{-j}} p^{\star}_{it}\left(\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right) - \mathrm{d}^2\lambda_{it}\left(\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right)\right),\end{aligned}$$

where $\mathbb{M}_{-j} := \mathbb{M}\backslash\{j\} = \{\, i \in \mathbb{M}, i \neq j\}$ and

$$\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right) = \mathrm{d}^3\log p_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right) + \mathrm{d}^3\log\phi_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right).$$

Now, using that the filtered probabilities $p^{\star}_{jt}$ are trivially bounded since $0 < p^{\star}_{jt} < 1$ and applying the triangle inequality, one has

$$\begin{aligned}\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\mathrm{d}^3 L_T\left(\theta;\mathrm{d}\theta_T,\mathrm{d}\theta_T^{\dagger},\mathrm{d}\theta_T^{\ddagger}\right)\right| \leq &c_1\left(\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\theta_T,\mathrm{d}\theta_T^{\dagger},\mathrm{d}\theta_T^{\ddagger}\right)\right|\right)\\ &+ c_2\left(\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\theta_T,\mathrm{d}\theta_T^{\dagger},\mathrm{d}\theta_T^{\ddagger}\right)\right|\right)\\ &+ c_3\left(\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}\theta_T,\mathrm{d}\theta_T^{\dagger},\mathrm{d}\theta_T^{\ddagger}\right)\right|\right)\end{aligned}$$

where we have introduced the definitions,

$$\begin{aligned}\Xi_1\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right) &:= \sum_{i=1}^{T}\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta\right)\mathrm{d}\lambda_{it}\left(\mathrm{d}\theta^{\dagger}\right)\mathrm{d}\lambda_{ht}\left(\mathrm{d}\theta^{\ddagger}\right),\\ \Xi_2\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right) &:= \sum_{i=1}^{T}\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta\right)\mathrm{d}^2\lambda_{it}\left(\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right),\\ \Xi_3\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right) &:= \sum_{i=1}^{T}\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}\right),\end{aligned}$$

and c_1, c_2 and c_3 are generic constants counting the number of instances of each term. This shows that in the following, we can limit our attention to checking the conditions,

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\theta_T,\mathrm{d}\theta_T^{\dagger},\mathrm{d}\theta_T^{\ddagger}\right)\right| = O_p\left(\|\mathrm{d}\theta\|\left\|\mathrm{d}\theta^{\dagger}\right\|\left\|\mathrm{d}\theta^{\ddagger}\right\|\right), \tag{1.45}$$

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\theta_T,\mathrm{d}\theta_T^{\dagger},\mathrm{d}\theta_T^{\ddagger}\right)\right| = O_p\left(\|\mathrm{d}\theta\|\left\|\mathrm{d}\theta^{\dagger}\right\|\left\|\mathrm{d}\theta^{\ddagger}\right\|\right), \tag{1.46}$$

and

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}\theta_T,\mathrm{d}\theta_T^{\dagger},\mathrm{d}\theta_T^{\ddagger}\right)\right| = O_p\left(\|\mathrm{d}\theta\|\left\|\mathrm{d}\theta^{\dagger}\right\|\left\|\mathrm{d}\theta^{\ddagger}\right\|\right). \tag{1.47}$$

We verify these conditions for a number of key examples collected in Claims 1.28-1.30, which are then subsequently verified. For the remaining cases, the arguments will be the same and the requirements to existence of moments be inferior or similar to those given in the presented claims.

Claim 1.27. $\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\mathrm{d}^3 L_T\left(\theta;\mathrm{d}b_T,\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right)\right| = O_p\left(\|\mathrm{d}b\|\left\|\mathrm{d}b^{\dagger}\right\|\left\|\mathrm{d}b^{\ddagger}\right\|\right)$.

Claim 1.28. $\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\mathrm{d}^3L_T\left(\theta;\mathrm{d}\gamma_T,\mathrm{d}\gamma_T^\dagger,\mathrm{d}\gamma_T^\ddagger\right)\right| = O_p\left(\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\|\right).$

Claim 1.29. $\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\mathrm{d}^3L_T\left(\theta;\mathrm{d}\alpha_{jT},\mathrm{d}\alpha_{iT}^\dagger,\mathrm{d}\alpha_{hT}^\ddagger\right)\right| = O_p\left(\|\mathrm{d}\alpha_j\|\left\|\mathrm{d}\alpha_i^\dagger\right\|\left\|\mathrm{d}\alpha_h^\ddagger\right\|\right).$

Claim 1.30. $\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\mathrm{d}^3L_T\left(\theta;\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^\dagger,\mathrm{d}\Omega_{hT}^\ddagger\right)\right| = O_p\left(\|\mathrm{d}\Omega_j\|\left\|\mathrm{d}\Omega_i^\dagger\right\|\left\|\mathrm{d}\Omega_h^\ddagger\right\|\right).$

Before turning to verifications of the Claims 1.27-1.30, observe that with $\theta\in\mathcal{N}_T(\theta_0)$, we can write

$$b^{\star\prime} = (b' : b_D') = (b_\kappa' : b_\tau' : b_D') = \left(T^{-\frac{1}{2}}\mathfrak{b}_{\kappa,T}' : T^{-1}\mathfrak{b}_{\tau,T}' : \mathfrak{b}_D'\right)$$

where $\|\mathfrak{b}_{\kappa,T}\| < e$, $\|\mathfrak{b}_{\tau,T}\| < e$, and $\mathfrak{b}_D$ is a constant such that

$$(b' : b_D')\,\mathcal{X}_{t-1}^\star = T^{-\frac{1}{2}}\mathfrak{b}_\kappa'\overline{\kappa}_0'X_{t-1} + T^{-1}\mathfrak{b}_\tau'\overline{\tau}_0'X_{t-1} + \mathfrak{b}_D' \tag{1.48}$$

This has in particular as a consequence that with $\theta\in\mathcal{N}_T(\theta_0)$,

$$\begin{aligned}\left\|(b' : b_D')\,\mathcal{X}_{t-1}^\star\right\| &= \left\|b'\mathcal{X}_{t-1} + b_D'i_{n+1}'\right\| \\ &\leq c\left(\left\|T^{-\frac{1}{2}}\mathfrak{b}_\kappa'\overline{\kappa}_0'X_{t-1}\right\| + \left\|T^{-1}\mathfrak{b}_\tau'\overline{\tau}_0'X_{t-1}\right\| + 1\right)\end{aligned}$$

and thus for some integer, a, one has by the triangle inequality

$$\begin{aligned}T^{-1}\sum_{t=1}^T\left\|(b' : b_D')\,\mathcal{X}_{t-1}^\star\right\|^a &\leq cT^{-1}\sum_{t=1}^T\left(T^{-\frac{a}{2}}\left\|\overline{\kappa}_0'X_{t-1}\right\|^a + T^{-a}\left\|\overline{\tau}_0'X_{t-1}\right\|^a + 1\right) \\ &\leq cT^{-1}\sum_{t=1}^T\left(T^{-\frac{a}{2}}\left\|\overline{\kappa}_0'X_{t-1}\right\|^a + T^{-a}\left\|\overline{\tau}_0'X_{t-1}\right\|^a + 1\right)\end{aligned}$$

In addition, by Theorem 1.4, and for some $a\geq 2q$ with q given in assumption 1.2,

$$cT^{-1}\sum_{t=1}^T\left(T^{-\frac{a}{2}}\left\|\overline{\kappa}_0'X_{t-1}\right\|^a + T^{-a}\left\|\overline{\tau}_0'X_{t-1}\right\|^a + 1\right) = O_p(1) \tag{1.49}$$

Next, observe that

$$\left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\| = \left\|T^{-\frac{1}{2}}\overline{\kappa}_0'X_{t-1} + T^{-1}\overline{\tau}_0'X_{t-1}\right\|$$

and thus by the same arguments, and for $a\geq 2q$, one has

$$T^{-1}\sum_{t=1}^T\left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\|^a = O_p(1). \tag{1.50}$$

The inequalities $\|xy\|\leq\|x\|\|y\|$,$\|x+y\|\leq\|x\|+\|y\|$, $\|x-y\|\leq\|x\|+\|y\|$ and $\|x\|\|y\|\leq\|x\|^2+\|y\|^2$ are used repeatedly in the following.

***Verification of Claim 1.27*:** The terms of interest are

$$\begin{aligned}T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}b_T,\mathrm{d}b_T^\dagger,\mathrm{d}b_T^\ddagger\right) &= T^{\frac{1}{2}}\sum_{t=1}^T\mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T\right)\mathrm{d}\lambda_{it}\left(\mathrm{d}b_T^\dagger\right)\mathrm{d}\lambda_{ht}\left(\mathrm{d}b_T^\ddagger\right), \\ T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}b_T,\mathrm{d}b_T^\dagger,\mathrm{d}b_T^\ddagger\right) &= T^{\frac{1}{2}}\sum_{t=1}^T\mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T\right)\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b_T^\dagger,\mathrm{d}b_T^\ddagger\right)\end{aligned}$$

and

$$\begin{aligned} T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right) &= T^{\frac{1}{2}}\sum_{t=1}^{T}\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right) \\ &= T^{\frac{1}{2}}\sum_{t=1}^{T}\mathrm{d}^3\log p_{jt}\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right) \end{aligned}$$

We have

$$\begin{aligned} &\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right)\right| \\ \leq & cT^{\frac{1}{2}}\sum_{t=1}^{T}\left\|\mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T\right)\right\|\left\|\mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T\right)\right\|\left\|\mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T\right)\right\| \\ \leq & cT^{-1}\sum_{t=1}^{T}\left(\left(\|\Delta X_t\|+\left\|\beta_0'X_{t-1}\right\|+\left\|b'\mathcal{X}_{t-1}\right\|+\|\Delta\mathbb{X}_{t-1}\|+1\right)\left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\|\right)^3\times \\ & \|\mathrm{d}b\|\left\|\mathrm{d}b^{\dagger}\right\|\left\|\mathrm{d}b^{\ddagger}\right\|. \end{aligned}$$

Using (1.49) and (1.50) with $q\geq 3$, we write

$$cT^{-1}\sum_{t=1}^{T}\left(\|\Delta X_t\|^6+\left\|\beta_0'X_{t-1}\right\|^6+\left\|b'\mathcal{X}_{t-1}\right\|^6+\|\Delta\mathbb{X}_{t-1}\|^6+\left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\|^6+1\right)=O_p\left(1\right)$$

Next, we have

$$\begin{aligned} &\left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right)\right| \\ \leq & T^{\frac{1}{2}}\sum_{t=1}^{T}\left|\mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T\right)\right|\left|\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right)\right| \\ \leq & cT^{-1}\sum_{t=1}^{T}\left(\left(\|\Delta X_t\|+\left\|\beta_0'X_{t-1}\right\|+\left\|b'\mathcal{X}_{t-1}\right\|+\|\Delta\mathbb{X}_{t-1}\|+1\right)\left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\|\|\mathrm{d}b\|\right)\times \\ & \left(\left(\left\|\beta_0'X_{t-1}\right\|+\left\|b'\mathcal{X}_{t-1}\right\|+\|\Delta\mathbb{X}_{t-1}\|+1\right)\left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\|^2\right)\left\|\mathrm{d}b^{\dagger}\right\|\left\|\mathrm{d}b^{\ddagger}\right\| \\ \leq & cT^{-1}\sum_{t=1}^{T}\left(\|\Delta X_t\|^2+\left\|\beta_0'X_{t-1}\right\|^2+\left\|b'\mathcal{X}_{t-1}\right\|^2+\|\Delta\mathbb{X}_{t-1}\|+1\right)^2\times \\ & \left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\|^3\|\mathrm{d}b\|\left\|\mathrm{d}b^{\dagger}\right\|\left\|\mathrm{d}b^{\ddagger}\right\|. \end{aligned}$$

Again, ignoring cross terms and with $q\geq 3$ and using (1.49) and (1.50), one has

$$cT^{-1}\sum_{t=1}^{T}\left(\|\Delta X_t\|^4+\left\|\beta_0'X_{t-1}\right\|^4+\left\|b'\mathcal{X}_{t-1}\right\|^4+\|\Delta\mathbb{X}_{t-1}\|^4+\left\|W_{bT}^{-\frac{1}{2}}\mathcal{X}_{t-1}\right\|^6+1\right)=O_p\left(1\right).$$

Next, by assumption 1.5, one has

$$\left|T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}b_T,\mathrm{d}b_T^{\dagger},\mathrm{d}b_T^{\ddagger}\right)\right|$$

$$
\begin{aligned}
&\leq T^{\frac{1}{2}} \sum_{t=1}^{T} \left| \mathrm{d}^3 \log p_{jt} \left(\mathrm{d}b_T, \mathrm{d}b_T^{\dagger}, \mathrm{d}b_T^{\ddagger} \right) \right| \\
&= T^{-1} \sum_{t=1}^{T} \left| \mathrm{vec} \left(\psi_\beta' \mathrm{d}b^{\dagger\prime} W_{bT}^{-\frac{1}{2}} \mathcal{X}_{t-1} \mathcal{X}_{t-1}' \mathrm{d}b W_{bT}^{-\frac{1}{2}} \psi_\beta \right)' \times \right. \\
&\quad \left. \partial^3_{zzz} \log p_{jt} \psi_\beta' \mathrm{d}b^{\ddagger\prime} W_{bT}^{-\frac{1}{2}} \mathcal{X}_{t-1} \right| \\
&\leq c T^{-1} \sum_{t=1}^{T} \left(\left\| \beta_0' X_{t-1} \right\| + \left\| b' \mathcal{X}_{t-1} \right\| + \left\| \Delta \mathbb{X}_{t-1} \right\| \right) \left\| W_{bT}^{-\frac{1}{2}} \mathcal{X}_{t-1} \right\|^3 \left\| \mathrm{d}b \right\| \left\| \mathrm{d}b^{\dagger} \right\| \left\| \mathrm{d}b^{\ddagger} \right\|
\end{aligned}
$$

where by similar arguments as before

$$
c T^{-1} \sum_{t=1}^{T} \left(\left\| \beta_0' X_{t-1} \right\|^2 + \left\| b' \mathcal{X}_{t-1} \right\|^2 + \left\| \Delta \mathbb{X}_{t-1} \right\|^2 + \left\| W_{bT}^{-\frac{1}{2}} \mathcal{X}_{t-1} \right\|^6 + 1 \right) = O_p(1).
$$

Thus in summary,

$$
\sup_{\theta \in \mathcal{N}_T(\theta_0)} \left| T^{\frac{1}{2}} \Xi_1 \left(\mathrm{d}b_T, \mathrm{d}b_T^{\dagger}, \mathrm{d}b_T^{\ddagger} \right) \right| = O_p \left(\left\| \mathrm{d}b \right\| \left\| \mathrm{d}b^{\dagger} \right\| \left\| \mathrm{d}b^{\ddagger} \right\| \right),
$$

$$
\sup_{\theta \in \mathcal{N}_T(\theta_0)} \left| T^{\frac{1}{2}} \Xi_2 \left(\mathrm{d}b_T, \mathrm{d}b_T^{\dagger}, \mathrm{d}b_T^{\ddagger} \right) \right| = O_p \left(\left\| \mathrm{d}b \right\| \left\| \mathrm{d}b^{\dagger} \right\| \left\| \mathrm{d}b^{\ddagger} \right\| \right)
$$

and

$$
\sup_{\theta \in \mathcal{N}_T(\theta_0)} \left| T^{\frac{1}{2}} \Xi_3 \left(\mathrm{d}b_T, \mathrm{d}b_T^{\dagger}, \mathrm{d}b_T^{\ddagger} \right) \right| = O_p \left(\left\| \mathrm{d}b \right\| \left\| \mathrm{d}b^{\dagger} \right\| \left\| \mathrm{d}b^{\ddagger} \right\| \right)
$$

as desired.

Verification of Claim 1.28: Using Lemmas 1.31, 1.32 and 1.33, it is clear that we need to verify

$$
\begin{aligned}
T^{\frac{1}{2}} \Xi_1 \left(\mathrm{d}\gamma_T, \mathrm{d}\gamma_T^{\dagger}, \mathrm{d}\gamma_T^{\ddagger} \right) &= T^{-1} \sum_{t=1}^{T} \left(\mathrm{d} \log p_{jt} \left(\mathrm{d}\gamma \right) \right) \left(\mathrm{d} \log p_{it} \left(\mathrm{d}\gamma^{\dagger} \right) \right) \left(\mathrm{d} \log p_{ht} \left(\mathrm{d}\gamma^{\ddagger} \right) \right) \\
&= O_p \left(\left\| \mathrm{d}\gamma \right\| \left\| \mathrm{d}\gamma^{\dagger} \right\| \left\| \mathrm{d}\gamma^{\ddagger} \right\| \right),
\end{aligned}
$$

$$
\begin{aligned}
T^{\frac{1}{2}} \Xi_2 \left(\mathrm{d}\gamma_T, \mathrm{d}\gamma_T^{\dagger}, \mathrm{d}\gamma_T^{\ddagger} \right) &= T^{-1} \sum_{t=1}^{T} \left(\mathrm{d} \log p_{jt} \left(\mathrm{d}\gamma \right) \right) \left(\mathrm{d}^2 \log p_{it} \left(\mathrm{d}\gamma, \mathrm{d}\gamma^{\ddagger} \right) \right) \\
&= O_p \left(\left\| \mathrm{d}\gamma \right\| \left\| \mathrm{d}\gamma^{\dagger} \right\| \left\| \mathrm{d}\gamma^{\ddagger} \right\| \right),
\end{aligned}
$$

and

$$
\begin{aligned}
T^{\frac{1}{2}} \Xi_3 \left(\mathrm{d}\gamma_T, \mathrm{d}\gamma_T^{\dagger}, \mathrm{d}\gamma_T^{\ddagger} \right) &= T^{-1} \sum_{t=1}^{T} \left(\mathrm{d}^3 \log p_{jt} \left(\mathrm{d}\gamma, \mathrm{d}\gamma^{\dagger}, \mathrm{d}\gamma^{\ddagger} \right) \right) \\
&= O_p \left(\left\| \mathrm{d}\gamma \right\| \left\| \mathrm{d}\gamma^{\dagger} \right\| \left\| \mathrm{d}\gamma^{\ddagger} \right\| \right).
\end{aligned}
$$

For the first term, observe that by Assumption 1.5, we have

$$\begin{aligned}\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\gamma_T,\mathrm{d}\gamma_T^\dagger,\mathrm{d}\gamma_T^\ddagger\right)\right| \leq & cT^{-1}\sum_{t=1}^{T}\|Z_{t-1}\|^3\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\| \\ \leq & cT^{-1}\sum_{t=1}^{T}\left\|\left(X_{t-1}'\beta:\Delta\mathbb{X}_{t-1}\right)'\right\|^3\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\| \\ \leq & cT^{-1}\sum_{t=1}^{T}\left(\left\|\beta_0'X_{t-1}\right\|^3+\left\|b'\mathcal{X}_{t-1}\right\|^3+\left\|\Delta\mathbb{X}_{t-1}\right\|^3\right)\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\|.\end{aligned}$$

and by similar arguments are in Verification of Claim 1.27, with $q \geq 2$, one has

$$cT^{-1}\sum_{t=1}^{T}\left(\left\|\beta_0'X_{t-1}\right\|^3+\left\|b'\mathcal{X}_{t-1}\right\|^3+\left\|\Delta\mathbb{X}_{t-1}\right\|^3\right)=O_p\left(1\right).$$

For the second and third terms, we have again by Assumption 1.5 that

$$\left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\gamma_T,\mathrm{d}\gamma_T^\dagger,\mathrm{d}\gamma_T^\ddagger\right)\right| \leq cT^{-1}\sum_{t=1}^{T}\|Z_{t-1}\|^2\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\|$$

and

$$\left|T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}\gamma_T,\mathrm{d}\gamma_T^\dagger,\mathrm{d}\gamma_T^\ddagger\right)\right| \leq cT^{-1}\sum_{t=1}^{T}\|Z_{t-1}\|\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\|.$$

which are bounded given the results from before. Hence, it holds that

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\gamma_T,\mathrm{d}\gamma_T^\dagger,\mathrm{d}\gamma_T^\ddagger\right)\right|=O_p\left(\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\|\right),$$

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\gamma_T,\mathrm{d}\gamma_T^\dagger,\mathrm{d}\gamma_T^\ddagger\right)\right|=O_p\left(\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\|\right),$$

and

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}\gamma_T,\mathrm{d}\gamma_T^\dagger,\mathrm{d}\gamma_T^\ddagger\right)\right|=O_p\left(\|\mathrm{d}\gamma\|\left\|\mathrm{d}\gamma^\dagger\right\|\left\|\mathrm{d}\gamma^\ddagger\right\|\right)$$

as desired.

***Verification of Claim 1.29*:** The terms we need to consider are given by

$$T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\alpha_{jT},\mathrm{d}\alpha_{iT}^\dagger,\mathrm{d}\alpha_{hT}^\ddagger\right)=T^{-1}\sum_{t=1}^{T}\left(\mathrm{d}\log\phi_{jt}\left(\mathrm{d}\alpha_j\right)\right)\left(\mathrm{d}\log\phi_{it}\left(\mathrm{d}\alpha_i^\dagger\right)\right)\left(\mathrm{d}\log\phi_{ht}\left(\mathrm{d}\alpha_h^\ddagger\right)\right),$$

$$T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\alpha_{jT},\mathrm{d}\alpha_{iT}^\dagger,\mathrm{d}\alpha_{hT}^\ddagger\right)=T^{-1}\sum_{t=1}^{T}\left(\mathrm{d}\log\phi_{jt}\left(\mathrm{d}\alpha_j\right)\right)\left(\mathrm{d}^2\log\phi_{it}\left(\mathrm{d}\alpha_i^\dagger,\mathrm{d}\alpha_i^\ddagger\right)\right),$$

and

$$T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}\alpha_{jT},\mathrm{d}\alpha_{iT}^\dagger,\mathrm{d}\alpha_{hT}^\ddagger\right)=T^{-1}\sum_{t=1}^{T}\mathrm{d}^3\log\phi_{jt}\left(\mathrm{d}\alpha_j,\mathrm{d}\alpha_i^\dagger,\mathrm{d}\alpha_h^\ddagger\right)=0,$$

which shows $T^{-\frac{1}{2}}\Xi_3(\cdot)$ is bounded by definition. Observe next that using the results from Lemmas 1.31, 1.32 and 1.33, it holds that

$$\begin{aligned}
|\mathrm{d}\log\phi_{it}(\mathrm{d}\alpha_i)| &= \left|\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}'\mathrm{d}\alpha_j\beta^{\star\prime}X_{t-1}^{\star}\right\}\right| \\
&\leq c\left(\|\varepsilon_{jt}\|^2 + \|\beta_0'X_{t-1}\| + \|b'\mathcal{X}_{t-1}\| + 1\right)\|\mathrm{d}\alpha_j\| \\
&\leq c\left(\|\Delta X_t\|^2 + \|\beta_0'X_{t-1}\|^2 + \|b'\mathcal{X}_{t-1}\|^2 + \|\Delta\mathbb{X}_{t-1}\|^2 + 1\right)\|\mathrm{d}\alpha_j\|
\end{aligned}$$

$$\begin{aligned}
\left|\mathrm{d}^2\log\phi_{it}\left(\mathrm{d}\alpha_i^{\dagger},\mathrm{d}\alpha_i^{\ddagger}\right)\right| &= \left|\mathrm{tr}\left\{\Omega_j^{-1}X_{t-1}^{\star\prime}\beta^{\star}\mathrm{d}\alpha_j^{\dagger\prime}\mathrm{d}\alpha_j^{\ddagger}\beta^{\star\prime}X_{t-1}^{\star}\right\}\right| \\
&\leq c\left(\|\beta_0'X_{t-1}\|^2 + \|b'\mathcal{X}_{t-1}\|^2 + 1\right)\|\mathrm{d}\alpha_j\|.
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\alpha_j,\mathrm{d}\alpha_i^{\dagger},\mathrm{d}\alpha_h^{\ddagger}\right)\right| \leq & cT^{-1}\sum_{t=1}^{T}\left(\|\Delta X_t\|^2 + \|\beta_0'X_{t-1}\|^2 + \|b'\mathcal{X}_{t-1}\|^2 + \|\Delta\mathbb{X}_{t-1}\|^2 + 1\right)^3 \times \\
& \|\mathrm{d}\alpha_j\|\left\|\mathrm{d}\alpha_j^{\dagger}\right\|\left\|\mathrm{d}\alpha_j^{\ddagger}\right\| \\
\leq & cT^{-1}\sum_{t=1}^{T}\left(\|\Delta X_t\|^6 + \|\beta_0'X_{t-1}\|^6 + \|b'\mathcal{X}_{t-1}\|^6 + \|\Delta\mathbb{X}_{t-1}\|^6 + 1\right)\times \\
& \|\mathrm{d}\alpha_j\|\left\|\mathrm{d}\alpha_j^{\dagger}\right\|\left\|\mathrm{d}\alpha_j^{\ddagger}\right\|,
\end{aligned}$$

and

$$\begin{aligned}
& \left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\alpha_j,\mathrm{d}\alpha_i^{\dagger},\mathrm{d}\alpha_h^{\ddagger}\right)\right| \\
\leq & cT^{-1}\sum_{t=1}^{T}\left(\|\Delta X_t\|^2 + \|\beta_0'X_{t-1}\|^2 + \|b'\mathcal{X}_{t-1}\|^2 + \|\Delta\mathbb{X}_{t-1}\|^2 + 1\right)\times \\
& \left(\|\beta_0'X_{t-1}\|^2 + \|b'\mathcal{X}_{t-1}\|^2 + 1\right)\|\mathrm{d}\alpha_j\|\left\|\mathrm{d}\alpha_j^{\dagger}\right\|\left\|\mathrm{d}\alpha_j^{\ddagger}\right\| \\
\leq & cT^{-1}\sum_{t=1}^{T}\left(\|\Delta X_t\|^4 + \|\beta_0'X_{t-1}\|^4 + \|b'\mathcal{X}_{t-1}\|^4 + \|\Delta\mathbb{X}_{t-1}\|^4 + 1\right)\times \\
& \|\mathrm{d}\alpha_j\|\left\|\mathrm{d}\alpha_j^{\dagger}\right\|\left\|\mathrm{d}\alpha_j^{\ddagger}\right\|.
\end{aligned}$$

Applying the same arguments as in the proof of claim 1.27, it holds that

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\alpha_{j,T},\mathrm{d}\alpha_{i,T}^{\dagger},\mathrm{d}\alpha_{h,T}^{\ddagger}\right)\right| = O_p\left(\|\mathrm{d}\alpha_j\|\left\|\mathrm{d}\alpha_j^{\dagger}\right\|\left\|\mathrm{d}\alpha_j^{\ddagger}\right\|\right)$$

and

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\alpha_{j,T},\mathrm{d}\alpha_{i,T}^{\dagger},\mathrm{d}\alpha_{h,T}^{\ddagger}\right)\right| = O_p\left(\|\mathrm{d}\alpha_j\|\left\|\mathrm{d}\alpha_j^{\dagger}\right\|\left\|\mathrm{d}\alpha_j^{\ddagger}\right\|\right).$$

***Verification of Claim 1.30*:** The terms that need to be checked are

$$T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^{\dagger},\mathrm{d}\Omega_{hT}^{\ddagger}\right)=T^{-1}\sum_{t=1}^{T}\left(\mathrm{d}\log\phi_{jt}\left(\mathrm{d}\Omega_j\right)\right)\left(\mathrm{d}\log\phi_{jt}\left(\mathrm{d}\Omega_j^{\dagger}\right)\right)\left(\mathrm{d}\log\phi_{jt}\left(\mathrm{d}\Omega_j^{\ddagger}\right)\right),$$

$$T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^{\dagger},\mathrm{d}\Omega_{hT}^{\ddagger}\right)=T^{-1}\sum_{t=1}^{T}\left(\mathrm{d}\log\phi_{jt}\left(\mathrm{d}\Omega_i\right)\right)\left(\mathrm{d}^2\log\phi_{it}\left(\mathrm{d}\Omega_j^{\dagger},\mathrm{d}\Omega_j^{\ddagger}\right)\right),$$

and

$$T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^{\dagger},\mathrm{d}\Omega_{hT}^{\ddagger}\right)=T^{-1}\sum_{t=1}^{T}\mathrm{d}^3\log\phi_{jt}\left(\mathrm{d}\Omega_i,\mathrm{d}\Omega_j^{\dagger},\mathrm{d}\Omega_j^{\ddagger}\right).$$

Observe initially that using the definitions of $\mathrm{d}\lambda_{jt}\left(\mathrm{d}\Omega_j\right)$, $\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Omega_j\mathrm{d}\Omega_i^{\dagger}\right)$ and $\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Omega_j,\mathrm{d}\Omega_i^{\dagger},\mathrm{d}\Omega_h^{\ddagger}\right)$ from Lemmas 1.31, 1.32 and 1.33, it holds that

$$\left|\mathrm{d}\log\phi_{jt}\left(\mathrm{d}\Omega_j\right)\right|=\left|\mathrm{d}\lambda_{jt}\left(\mathrm{d}\Omega_j\right)\right|\leq c\left\|\varepsilon_{jt}\right\|^2\left\|\mathrm{d}\Omega_j\right\|,$$

$$\left|\mathrm{d}^2\log\phi_{it}\left(\mathrm{d}\Omega_j^{\dagger},\mathrm{d}\Omega_j^{\ddagger}\right)\right|=\left|\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Omega_j\mathrm{d}\Omega_i^{\dagger}\right)\right|\leq c\left\|\varepsilon_{jt}\right\|^2\left\|\mathrm{d}\Omega_j\right\|\left\|\mathrm{d}\Omega_i^{\dagger}\right\|$$

and

$$\left|\mathrm{d}^3\log\phi_{jt}\left(\mathrm{d}\Omega_i,\mathrm{d}\Omega_j^{\dagger},\mathrm{d}\Omega_j^{\ddagger}\right)\right|=\left|\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\Omega_j,\mathrm{d}\Omega_i^{\dagger},\mathrm{d}\Omega_h^{\ddagger}\right)\right|\leq c\left\|\varepsilon_{jt}\right\|^2\left\|\mathrm{d}\Omega_j\right\|\left\|\mathrm{d}\Omega_i^{\dagger}\right\|\left\|\mathrm{d}\Omega_h^{\ddagger}\right\|,$$

such that

$$\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^{\dagger},\mathrm{d}\Omega_{hT}^{\ddagger}\right)\right|\leq cT^{-1}\sum_{t=1}^{T}\left\|\varepsilon_{jt}\right\|^6\left\|\mathrm{d}\Omega_j\right\|\left\|\mathrm{d}\Omega_i^{\dagger}\right\|\left\|\mathrm{d}\Omega_h^{\ddagger}\right\|,$$

$$\left|T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^{\dagger},\mathrm{d}\Omega_{hT}^{\ddagger}\right)\right|\leq cT^{-1}\sum_{t=1}^{T}\left\|\varepsilon_{jt}\right\|^4\left\|\mathrm{d}\Omega_j\right\|\left\|\mathrm{d}\Omega_i^{\dagger}\right\|\left\|\mathrm{d}\Omega_h^{\ddagger}\right\|,$$

and

$$\left|T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^{\dagger},\mathrm{d}\Omega_{hT}^{\ddagger}\right)\right|\leq cT^{-1}\sum_{t=1}^{T}\left\|\varepsilon_{jt}\right\|^2\left\|\mathrm{d}\Omega_j\right\|\left\|\mathrm{d}\Omega_i^{\dagger}\right\|\left\|\mathrm{d}\Omega_h^{\ddagger}\right\|.$$

Now, notice that ignoring cross-terms, we obtain

$$cT^{-1}\sum_{t=1}^{T}\left\|\varepsilon_{jt}\right\|^6\leq cT^{-1}\sum_{t=1}^{T}\left(\left\|\Delta X_t\right\|^6+\left\|\beta_0'X_{t-1}\right\|^6+\left\|b'\mathcal{X}_{t-1}\right\|^6+\left\|\Delta\mathbb{X}_{t-1}\right\|^6+1\right),$$

which for $\theta\in\mathcal{N}_T\left(\theta_0\right)$ is bounded using the same arguments as in the proof of claim 1.27, such that

$$\sup_{\theta\in\mathcal{N}_T(\theta_0)}\left|T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}\Omega_{jT},\mathrm{d}\Omega_{iT}^{\dagger},\mathrm{d}\Omega_{hT}^{\ddagger}\right)\right|=O_p\left(\left\|\mathrm{d}\Omega_j\right\|\left\|\mathrm{d}\Omega_i^{\dagger}\right\|\left\|\mathrm{d}\Omega_h^{\ddagger}\right\|\right).$$

Similar results then follow for $T^{\frac{1}{2}}\Xi_2\left(\cdot\right)$ and $T^{\frac{1}{2}}\Xi_3\left(\cdot\right)$ only with lower requirements for the existence of moments. □

1.D Likelihood Derivatives

Let $\mathrm{d}c_{\cdot}^{u}$ indicate $\mathrm{d}c_{\cdot}(\theta;\mathrm{d}\theta^{u})$, $\mathrm{d}^{2}c_{\cdot}^{uv}$ indicate $\mathrm{d}^{2}c_{\cdot}(\theta;\mathrm{d}\theta^{u};\mathrm{d}\theta^{v})$ and $\mathrm{d}^{3}c_{\cdot}^{uvw}$ indicate $\mathrm{d}^{3}c_{\cdot}(\theta;\mathrm{d}\theta^{u};\mathrm{d}\theta^{v};\mathrm{d}\theta^{w})$ where $\mathrm{d}\theta^{u},\mathrm{d}\theta^{v}$ and $\mathrm{d}\theta^{w}$ indicate any of $\mathrm{d}\theta,\mathrm{d}\theta^{\dagger},\mathrm{d}\theta^{\ddagger}$.

Lemma 1.31. *The first order differential of ℓ_t is*

$$\mathrm{d}\ell_{t}^{u}=\sum_{j\in\mathbb{M}}p_{jt}^{\star}\mathrm{d}\lambda_{jt}^{u} \tag{1.51}$$

where $\mathrm{d}\lambda_{jt}$ is defined in section 1.7. We have for any $(j,i)\in\mathbb{M}^{2}$,

$$\mathrm{d}\lambda_{jt}(\mathrm{d}b)=\mathcal{X}_{t-1}'\mathrm{d}b\left(\alpha_{j}'\Omega_{j}^{-1}\varepsilon_{jt}+\psi_{\beta}'\left(\partial_{z}\log p_{jt}\right)'\right) \tag{1.52}$$

$$\mathrm{d}\lambda_{jt}(\mathrm{d}b_{D})=\mathrm{d}b_{D}\alpha_{j}'\Omega_{j}^{-1}\varepsilon_{jt} \tag{1.53}$$

$$\mathrm{d}\lambda_{jt}(\mathrm{d}\alpha_{i})=\mathbf{1}\left\{j=i\right\}X_{t-1}^{\star\prime}\beta^{\star}\mathrm{d}\alpha_{i}'\Omega_{i}^{-1}\varepsilon_{it}. \tag{1.54}$$

$$\mathrm{d}\lambda_{jt}(\mathrm{d}\Gamma_{i})=\mathbf{1}\left\{j=i\right\}\Delta\mathbb{X}_{t-1}'\mathrm{d}\Gamma_{i}'\Omega_{i}^{-1}\varepsilon_{it} \tag{1.55}$$

$$\mathrm{d}\lambda_{jt}(\mathrm{d}\Omega_{i})=\mathbf{1}\left\{j=i\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_{i}^{-1}\mathrm{d}\Omega_{i}\Omega_{i}^{-1}\left(\varepsilon_{it}\varepsilon_{it}'-\Omega_{i}\right)\right\} \tag{1.56}$$

$$\mathrm{d}\lambda_{jt}(\mathrm{d}\gamma)=\partial_{\gamma}\log p_{jt}'\mathrm{d}\gamma. \tag{1.57}$$

Proof. The derivation of the score follows by applying standard differential calculus, see e.g. Magnus and Neudecker (1999). □

Lemma 1.32. *The second order differential of ℓ_t is*

$$\begin{aligned}\mathrm{d}^{2}\ell_{t}^{uv} &= \sum_{j\in\mathbb{M}}\left(\mathrm{d}p_{jt}^{\star v}\mathrm{d}\lambda_{jt}^{u}+p_{jt}^{\star}\mathrm{d}^{2}\lambda_{jt}^{uv}\right)= \\ &= \sum_{j\in\mathbb{M}}p_{jt}^{\star}\mathrm{d}\lambda_{jt}^{v}\mathrm{d}\lambda_{jt}^{u}-\sum_{j\in\mathbb{M}}p_{jt}^{\star}\mathrm{d}\lambda_{jt}^{u}\sum_{i\in\mathbb{M}}p_{it}^{\star}\mathrm{d}\lambda_{it}^{v}+\sum_{j\in\mathbb{M}}p_{jt}^{\star}\mathrm{d}^{2}\lambda_{jt}^{uv}\end{aligned}$$

where

$$\mathrm{d}p_{jt}^{\star v}=p_{jt}^{\star}\mathrm{d}\lambda_{jt}^{v}-p_{jt}^{\star}\sum_{i\in\mathbb{M}}p_{it}^{\star}\mathrm{d}\lambda_{it}^{v} \tag{1.58}$$

and

$$\mathrm{d}^{2}\lambda_{jt}^{uv}=\mathrm{d}^{2}\log p_{jt}^{uv}+\mathrm{d}^{2}\log\phi_{jt}^{uv}. \tag{1.59}$$

Next, we have for any $(j,i,h)\in\mathbb{M}^{3}$,

$$\begin{aligned}\mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b^{\dagger}\right) &= \mathcal{X}_{t-1}'\mathrm{d}b\left(\psi_{\beta}'\partial_{zz}^{2}\log p_{jt}\psi_{\beta}-\alpha_{j}'\Omega_{j}^{-1}\alpha_{j}\right)\mathrm{d}b^{\dagger\prime}\mathcal{X}_{t-1} \\ \mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b_{D}^{\dagger}\right) &= \mathcal{X}_{t-1}'\mathrm{d}b\alpha_{j}'\Omega_{j}^{-1}\alpha_{j}\mathrm{d}b_{D}^{\dagger} \\ \mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\alpha_{i}^{\dagger}\right) &= \mathbf{1}\left\{j=i\right\}\mathcal{X}_{t-1}'\mathrm{d}b\left(\mathrm{d}\alpha_{j}^{\dagger\prime}\Omega_{j}^{-1}\varepsilon_{jt}-\alpha_{j}'\Omega_{j}^{-1}\mathrm{d}\alpha_{j}^{\dagger}\beta^{\star\prime}X_{t-1}^{\star}\right) \\ \mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\Gamma_{i}^{\dagger}\right) &= -\mathbf{1}\left\{j=i\right\}\mathcal{X}_{t-1}'\mathrm{d}b\left(\alpha_{j}'\Omega_{j}^{-1}\mathrm{d}\Gamma_{j}^{\dagger}\Delta\mathbb{X}_{t-1}\right) \\ \mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\Omega_{i}^{\dagger}\right) &= -\mathbf{1}\left\{j=i\right\}\mathcal{X}_{t-1}'\mathrm{d}b\left(\alpha_{j}'\Omega_{j}^{-1}\mathrm{d}\Omega_{j}^{\dagger}\Omega_{j}^{-1}\varepsilon_{jt}\right) \\ \mathrm{d}^{2}\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\gamma^{\dagger}\right) &= \mathcal{X}_{t-1}'\mathrm{d}b\psi_{\beta}\partial_{z\gamma}^{2}\log p_{jt}\mathrm{d}\gamma^{\dagger}\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}b_D^\dagger\right) &= \mathrm{d}b_D\alpha_j'\Omega_j^{-1}\alpha_j\mathrm{d}b_D \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\alpha_i^\dagger\right) &= \mathbf{1}\left\{j=i\right\}\mathrm{d}b_D\left(\mathrm{d}\alpha_j^{\dagger\prime}\Omega_j^{-1}\varepsilon_{jt}-\alpha_j'\Omega_j^{-1}\mathrm{d}\alpha_j^\dagger\beta^{\star\prime}X_{t-1}^\star\right) \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\Gamma_i^\dagger\right) &= -\mathbf{1}\left\{j=i\right\}\mathrm{d}b_D\left(\alpha_j'\Omega_j^{-1}\mathrm{d}\Gamma_j^\dagger\Delta\mathbb{X}_{t-1}\right) \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\Omega_i^\dagger\right) &= -\mathbf{1}\left\{j=i\right\}\mathrm{d}b\left(\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\varepsilon_{jt}\right) \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\gamma^\dagger\right) &= \mathrm{d}b_D\psi_\beta\partial_{z\gamma}^2\log p_{jt}\mathrm{d}\gamma^\dagger
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\alpha_h^\dagger\right) &= -\mathbf{1}\{j=i=h\}X_{t-1}^{\star\prime}\beta^\star\mathrm{d}\alpha_j'\Omega_j^{-1}\mathrm{d}\alpha_j^\dagger\beta^{\star\prime}X_{t-1}^\star \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\Gamma_h^\dagger\right) &= -\mathbf{1}\{j=i=h\}X_{t-1}^{\star\prime}\beta^\star\mathrm{d}\alpha_j'\Omega_j^{-1}\mathrm{d}\Gamma_j^\dagger\Delta\mathbb{X}_{t-1} \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\Omega_h^\dagger\right) &= -\mathbf{1}\{j=i=h\}X_{t-1}^{\star\prime}\beta^\star\mathrm{d}\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\varepsilon_{jt} \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\gamma^\dagger\right) &= 0
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Gamma_i,\mathrm{d}\Gamma_h^\dagger\right) &= -\mathbf{1}\{i=j=h\}\Delta\mathbb{X}_{t-1}'\mathrm{d}\Gamma_j'\Omega_j^{-1}\mathrm{d}\Gamma_j^\dagger\Delta\mathbb{X}_{t-1} \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Gamma_i,\mathrm{d}\Omega_h^\dagger\right) &= -\mathbf{1}\{i=j=h\}\Delta\mathbb{X}_{t-1}'\mathrm{d}\Gamma_j'\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\varepsilon_{jt} \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Gamma_i,\mathrm{d}\gamma^\dagger\right) &= 0
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Omega_i,\mathrm{d}\Omega_h^\dagger\right) =& -\mathbf{1}\{i=j=h\}\frac{1}{2}\left(\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\}\right. \\
&+\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\} \\
&\left.+\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}d\Omega_j^\dagger\right\}\right) \\
\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\Omega_i,\mathrm{d}\gamma^\dagger\right) =& 0
\end{aligned}$$

and

$$\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\gamma,\mathrm{d}\gamma^\dagger\right) = \mathrm{d}\gamma'\left(\partial_{\gamma\gamma}^2\log p_{jt}\right)'\mathrm{d}\gamma^\dagger$$

Proof. The results follow by applying standard matrix differential calculus, see e.g. Magnus and Neudecker (1999). □

Lemma 1.33. *The third order differential of the log-likelihood contribution from (1.18) is given by*

$$\begin{aligned}
\mathrm{d}^3\ell_t\left(\theta;\mathrm{d}\theta,\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right) = \sum_{j\in\mathbb{M}}&\left\{\mathrm{d}^2p_{jt}^\star\left(\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta\right)+\mathrm{d}p_{jt}^\star\left(\mathrm{d}\theta^\dagger\right)\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\ddagger\right)\right. \\
&\left.+\mathrm{d}p_{jt}^\star\left(\mathrm{d}\theta^\ddagger\right)\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\dagger\right)+p_{jt}^\star\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)\right\},
\end{aligned}$$

where

$$\begin{aligned}\mathrm{d}^2 p^\star_{jt}\left(\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right) = &\sum_{i\in\mathbb{M}_{-j}}\left(\mathrm{d}p^\star_{jt}(\mathrm{d}\theta^\ddagger)p^\star_{it}+p^\star_{jt}\mathrm{d}p^\star_{it}(\mathrm{d}\theta^\ddagger)\right)\left(\mathrm{d}\lambda_{jt}\left(\mathrm{d}\theta^\dagger\right)-\mathrm{d}\lambda_{it}\left(\mathrm{d}\theta^\dagger\right)\right)\\ &+p^\star_{jt}\sum_{i\in\mathbb{M}_{-j}}p^\star_{it}\left(\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)-\mathrm{d}^2\lambda_{it}\left(\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)\right)\end{aligned}$$

and

$$\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)=\mathrm{d}^3\log p_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right)+\mathrm{d}^3\log\phi_{jt}\left(\mathrm{d}\theta,\mathrm{d}\theta^\dagger,\mathrm{d}\theta^\ddagger\right).$$

Moreover,

$$\begin{aligned}\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b^\dagger,\mathrm{d}\Gamma_i^\ddagger\right)=&0,\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b^\dagger,\mathrm{d}b_D^\ddagger\right)=&0,\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b^\dagger,\mathrm{d}b^\ddagger\right)=&\mathrm{vec}\left(\psi_\beta'\mathrm{d}b^{\dagger\prime}\mathcal{X}_{t-1}\mathcal{X}_{t-1}'\mathrm{d}b\psi_\beta\right)'\\ &\times\partial^3_{zzz}\log p_{jt}\psi_\beta' db^{\ddagger\prime}\mathcal{X}_{t-1},\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b^\dagger,\mathrm{d}\alpha_i^\ddagger\right)=&-\mathbf{1}\left\{j=i\right\}2\mathcal{X}_{t-1}'\mathrm{d}b\mathrm{d}\alpha_j^{\ddagger\prime}\Omega_j^{-1}\alpha_j\mathrm{d}b^{\dagger\prime}\mathcal{X}_{t-1}\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b^\dagger,\mathrm{d}\Omega_i^\ddagger\right)=&\mathbf{1}\left\{j=i\right\}\mathcal{X}_{t-1}'\mathrm{d}b\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^\ddagger\Omega_j^{-1}\alpha_j\mathrm{d}b^{\dagger\prime}\mathcal{X}_{t-1},\\ \mathrm{d}\lambda^3_{jt}\left(\mathrm{d}b,\mathrm{d}b^\dagger,\mathrm{d}\gamma^\ddagger\right)=&\mathbf{1}\left\{j=i\right\}\mathrm{vec}\left(\psi_\beta'\mathrm{d}b^{\dagger\prime}\mathcal{X}_{t-1}\mathcal{X}_{t-1}'\mathrm{d}b\psi_\beta\right)'\times\\ &\partial^3_{zz\gamma}\log p_{jt}\mathrm{d}\gamma^\ddagger,\end{aligned}$$

and

$$\begin{aligned}\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b;\mathrm{d}b_D^\dagger,\mathrm{d}b_D^\ddagger\right)=&\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b;\mathrm{d}b_D^\dagger,\mathrm{d}\gamma^\ddagger\right)=\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b_D^\dagger,\mathrm{d}\Gamma_h^\ddagger\right)=0,\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b_D^\dagger,\mathrm{d}\alpha_h^\ddagger\right)=&-\mathbf{1}\left\{j=i\right\}2\mathcal{X}_{t-1}'\mathrm{d}b\mathrm{d}\alpha_j^{\ddagger\prime}\Omega_j^{-1}\alpha_j\mathrm{d}b_D^{\dagger\prime},\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}b_D^\dagger,\mathrm{d}\Omega_h^\ddagger\right)=&\mathbf{1}\left\{j=i\right\}\mathcal{X}_{t-1}'\mathrm{d}b\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^\ddagger\Omega_j^{-1}\alpha_j\mathrm{d}b_D^{\dagger\prime},\end{aligned}$$

and

$$\begin{aligned}\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b;\mathrm{d}\alpha_i^\dagger,\mathrm{d}\gamma^\ddagger\right)=&0,\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\alpha_i^\dagger,\mathrm{d}\Gamma_h^\ddagger\right)=&-\mathbf{1}\left\{i=j=h\right\}\mathcal{X}_{t-1}'\mathrm{d}b\mathrm{d}\alpha_j^{\dagger\prime}\Omega_j^{-1}\mathrm{d}\Gamma_j^\ddagger\Delta\mathbb{X}_{t-1},\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\alpha_i^\dagger,\mathrm{d}\alpha_h^\ddagger\right)=&-\mathbf{1}\{i=j=h\}2X_{t-1}'\beta^\star\mathrm{d}\alpha_j^{\dagger\prime}\Omega_j^{-1}\mathrm{d}\alpha_j^\ddagger\mathrm{d}b'\mathcal{X}_{t-1},\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\alpha_i^\dagger,\mathrm{d}\Omega_h^\ddagger\right)=&\mathbf{1}\{i=j=h\}\mathcal{X}_{t-1}'\mathrm{d}b\\ &\times\left(\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^\ddagger\Omega_j^{-1}\mathrm{d}\alpha_j^\dagger\beta^{\star\prime}X_{t-1}^\star-\mathrm{d}\alpha_j^{\dagger\prime}\Omega_j^{-1}\mathrm{d}\Omega_j^\ddagger\Omega_j^{-1}\varepsilon_{jt}\right)\end{aligned}$$

and

$$\begin{aligned}\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\Gamma_i^\dagger,\mathrm{d}\Gamma_h^\ddagger\right)=&\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b;\mathrm{d}\Gamma_i^\dagger,\mathrm{d}\gamma^\ddagger\right)=0,\\ \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\Gamma_i^\dagger,\mathrm{d}\Omega_h^\ddagger\right)=&\mathbf{1}\{j=i=h\}\mathcal{X}_{t-1}'\mathrm{d}b\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^\ddagger\Omega_j^{-1}\mathrm{d}\Gamma_j^\dagger\Delta\mathbb{X}_{t-1},\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\Omega_i^{\dagger},\mathrm{d}\Omega_h^{\ddagger}\right) &= -2\mathcal{X}_{t-1}'\mathrm{d}b\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\varepsilon_{jt}\\
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\Omega_i^{\dagger},\mathrm{d}\gamma^{\ddagger}\right) &= 0,
\end{aligned}$$

and

$$\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b,\mathrm{d}\gamma^{\dagger},\mathrm{d}\gamma^{\ddagger}\right) = \mathrm{vec}\left(\mathrm{d}\gamma^{\dagger}\mathcal{X}_{t-1}'\mathrm{d}b\right)'\partial^3_{z\gamma\gamma}\log p_{jt}\mathrm{d}\gamma^{\ddagger}.$$

Next,

$$\begin{aligned}
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}b_D^{\dagger},\mathrm{d}b^{\ddagger}\right) =& \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}b_D^{\dagger},\mathrm{d}b^{\ddagger}\right) = \mathrm{d}\lambda^3_{jt}\left(\mathrm{d}b_D,\mathrm{d}b_D^{\dagger},\mathrm{d}\gamma^{\ddagger}\right)\\
=& \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}b_D^{\dagger},\mathrm{d}\Gamma_i^{\ddagger}\right) = 0,\\
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}b_D^{\dagger},\mathrm{d}\alpha_i^{\ddagger}\right) =& -\mathbf{1}\left\{j=i\right\}2\mathrm{d}b_D\mathrm{d}\alpha_j^{\ddagger\prime}\Omega_j^{-1}\alpha_j\mathrm{d}b_D^{\dagger\prime}\\
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}b_D^{\dagger},\mathrm{d}\Omega_i^{\ddagger}\right) =& \mathbf{1}\left\{j=i\right\}\mathrm{d}b_D\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\alpha_j\mathrm{d}b_D^{\dagger\prime},
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\alpha_i^{\dagger},\mathrm{d}\alpha_h^{\ddagger}\right) =& \mathrm{d}\lambda^3_{jt}\left(\mathrm{d}b_D,\mathrm{d}\alpha_i^{\dagger},\mathrm{d}\gamma^{\ddagger}\right) = \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\alpha_i^{\dagger},\mathrm{d}\Gamma_i^{\ddagger}\right) = 0,\\
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\alpha_i^{\dagger},\mathrm{d}\Omega_i^{\ddagger}\right) =& \mathbf{1}\{i=j=h\}\mathrm{d}b_D\\
&\times\left(\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\alpha_j^{\dagger}\beta^{\star\prime}X_{t-1}^{\star}-\mathrm{d}\alpha_j^{\dagger\prime}\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\varepsilon_{jt}\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\Gamma_i^{\dagger},\mathrm{d}\Omega_h^{\ddagger}\right) &= 0\\
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_D,\mathrm{d}\Omega_i^{\dagger},\mathrm{d}\Omega_h^{\ddagger}\right) &= -2\mathrm{d}b_D\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\varepsilon_{jt}.
\end{aligned}$$

Next, we have

$$\begin{aligned}
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\alpha_h^{\dagger},\mathrm{d}\alpha_l^{\ddagger}\right) &= \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\alpha_i;\mathrm{d}\alpha_h^{\dagger},\mathrm{d}\Gamma_l^{\ddagger}\right) = \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\alpha_i;\mathrm{d}\alpha_h^{\dagger},\mathrm{d}\gamma^{\ddagger}\right)\\
&= \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\Gamma_h^{\dagger},\mathrm{d}\Gamma_l^{\ddagger}\right) = \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\Gamma_i,\mathrm{d}\Gamma_h^{\dagger},\mathrm{d}\Gamma_l^{\ddagger}\right)\\
&= \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\Gamma_i,\mathrm{d}\Gamma_h^{\dagger},\mathrm{d}\gamma^{\ddagger}\right) = \mathrm{d}\lambda_{jt}\left(\mathrm{d}\Omega_i,\mathrm{d}\Omega_h^{\dagger},\mathrm{d}\gamma^{\ddagger}\right) = 0
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\alpha_h^{\dagger},\mathrm{d}\Omega_l^{\ddagger}\right) =& \mathbf{1}\{j=i=h=l\}X_{t-1}^{\star\prime}\beta^{\star}\mathrm{d}\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\alpha_j^{\dagger}\beta^{\star\prime}X_{t-1}^{\star}\\
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\alpha_i,\mathrm{d}\Gamma_h^{\dagger},\mathrm{d}\Omega_l^{\ddagger}\right) =& \mathbf{1}\left\{i=j=h=l\right\}X_{t-1}^{\star\prime}\beta^{\star}\mathrm{d}\alpha_j'\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Gamma_j^{\dagger}\Delta\mathbb{X}_{t-1}\\
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\Gamma_i,\mathrm{d}\Gamma_h^{\dagger},\mathrm{d}\Omega_l^{\ddagger}\right) =& -\mathbf{1}\{j=i=h=l\}\Delta\mathbb{X}_{t-1}'\mathrm{d}\Gamma_j\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Gamma_j^{\dagger}\Delta\mathbb{X}_{t-1}\\
\mathrm{d}\lambda_{jt}\left(\mathrm{d}\Omega_i,\mathrm{d}\Omega_h^{\dagger},\mathrm{d}\Omega_l^{\ddagger}\right) =& \mathbf{1}\{j=i=h=l\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\}\\
&+\mathbf{1}\{j=i=h=l\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&+ \mathbf{1}\{j=i=h=l\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\}\\
&+ \mathbf{1}\{j=i=h=l\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\}\\
&+ \mathbf{1}\{j=i=h=l\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\}\\
&+ \mathbf{1}\{j=i=h=l\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^{\dagger}\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\left(\varepsilon_{jt}\varepsilon_{jt}'-\Omega_j\right)\right\}\\
&+ \frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega^{\dagger}\right\}\\
&+ \frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^{\ddagger}\Omega_j^{-1}\mathrm{d}\Omega^{\dagger}\right\}
\end{aligned}
$$

and finally,

$$
\mathrm{d}^3\lambda_{jt}\left(\mathrm{d}\gamma,\mathrm{d}\gamma^{\dagger},\mathrm{d}\gamma^{\ddagger}\right)=vec\left(\mathrm{d}\gamma^{\dagger}\mathrm{d}\gamma'\right)'\partial^3_{\gamma\gamma\gamma}\log p_{jt}d\gamma^{\ddagger}.
$$

Proof. The derivations follows by standard matrix calculus applying the notation from Magnus and Neudecker (1999). □

1.E Auxiliary Lemmas

Lemma 1.34. *With ϕ_{jt} defined in (1.19), it holds that*

$$
E\left[\sum_{j\in\mathbb{M}}p^{\star}_{jt}\left(\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\theta\right)\right)'\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\theta^{\dagger}\right)\right)-\mathrm{d}^2\log\phi_{jt}\left(\theta;\mathrm{d}\theta;\mathrm{d}\theta^{\dagger}\right)\right)\mid Z_{t-1}\right]=0.
$$

Proof. Note first that for some parameter, $A\in\{\theta\}\setminus\Omega$, one has

$$
\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}A\right)=-\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}\left(\theta\right)\mathrm{d}\varepsilon_{jt}'\left(\theta;\mathrm{d}A\right)\right\},
$$

$$
\mathrm{d}^2\log\phi_{jt}\left(\theta;\mathrm{d}A,\mathrm{d}A^{\dagger}\right)=-\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\varepsilon_{jt}\left(\theta;\mathrm{d}A^{\dagger}\right)\mathrm{d}\varepsilon_{jt}'\left(\theta;\mathrm{d}A\right)\right\}
$$

and

$$
\begin{aligned}
&\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}A\right)\right)'\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}A^{\dagger}\right)\right)\\
=&\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}\left(\theta\right)\varepsilon_{jt}\left(\theta\right)'\Omega_j^{-1}\mathrm{d}\varepsilon_{jt}\left(\theta;\mathrm{d}A\right)\mathrm{d}\varepsilon_{jt}'\left(\theta;\mathrm{d}A^{\dagger}\right)\right\}.
\end{aligned}
$$

Note further that

$$
\begin{aligned}
E\left[\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}A\right)\right)'\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}A^{\dagger}\right)\right)\mid Z_{t-1}\right]&=\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\varepsilon_{jt}\left(\theta;\mathrm{d}A\right)\mathrm{d}\varepsilon_{jt}'\left(\theta;\mathrm{d}A^{\dagger}\right)\right\}\\
&=\mathrm{d}^2\log\phi_{jt}\left(\theta;\mathrm{d}A,\mathrm{d}A^{\dagger}\right).
\end{aligned}
$$

Next, observe that for some parameter $\Omega_i\in\{\Omega_1,\Omega_2,...,\Omega_m\}$ it holds that

$$
\begin{aligned}
\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_i\right)&=-\mathbf{1}\left\{j=i\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\right\}+\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\\
&=-\mathbf{1}\left\{j=i\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j-\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\\
&=-\mathbf{1}\left\{j=i\right\}\frac{1}{2}\mathrm{tr}\left\{\left(I_n-\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\right)\Omega_j^{-1}\mathrm{d}\Omega_j\right\}
\end{aligned}
$$

and

$$\begin{aligned}
&\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_i\right)\right)'\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_h^\dagger\right)\right)\\
=&\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}\\
&+\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}\\
&-\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}\\
&-\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j\right\}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{d}^2\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_j,\mathrm{d}\Omega_j^\dagger\right)=&\mathbf{1}\left\{j=i=h\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\\
&-\mathbf{1}\left\{j=i=h\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\\
&-\mathbf{1}\left\{j=i=h\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\mathrm{d}\Omega_j\right\}
\end{aligned}$$

Moreover, using $E\left[\varepsilon_{jt}\varepsilon_{jt}'\mid Z_{t-1}\right]=\Omega_j$ it holds that

$$E\left[\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_j\right)\mid Z_{t-1}\right]=-\mathbf{1}\left\{j=i\right\}\frac{1}{2}E\left[\mathrm{tr}\left\{\left(I_n-\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\right)\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\mid Z_{t-1}\right]=0,$$

and

$$\begin{aligned}
&E\left[\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_j\right)\right)'\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_j^\dagger\right)\right)\right]\\
=&\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}E\left[\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\varepsilon_{jt}\varepsilon_{jt}'\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\varepsilon_{jt}\mid Z_{t-1}\right]\\
&-\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}\mathrm{tr}\left\{\Omega_j^{-1}\Omega_j^{-1}\mathrm{d}\Omega_j\right\}.
\end{aligned}$$

Now, assuming ε_{jt} symmetric, we have

$$\begin{aligned}
&E\left[\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_j\right)\right)'\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_j^\dagger\right)\right)\right]\\
=&\mathbf{1}\left\{j=i=h\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}+\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\\
&-\mathbf{1}\left\{j=i=h\right\}\frac{1}{4}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\\
=&\mathbf{1}\left\{j=i=h\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\right\}.
\end{aligned}$$

Finally, it holds that

$$\begin{aligned}
&E\left[\mathrm{d}^2\log\phi_{jt}\left(\theta;\mathrm{d}\Omega_j,\mathrm{d}\Omega_j^\dagger\right)\mid Z_{t-1}\right]\\
=&\mathbf{1}\left\{j=i=h\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\mathrm{d}\Omega_j\right\}-\mathbf{1}\left\{j=i=h\right\}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\mathrm{d}\Omega_j\right\}\\
=&-\mathbf{1}\left\{j=i=h\right\}\frac{1}{2}\mathrm{tr}\left\{\Omega_j^{-1}\mathrm{d}\Omega_j^\dagger\Omega_j^{-1}\mathrm{d}\Omega_j\right\},
\end{aligned}$$

such that

$$E\left[\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left(\left(\mathrm{d}\log\phi_{jt}\right)'\left(\mathrm{d}\log\phi_{jt}\right)-\mathrm{d}^2\log\phi_{jt}\right)\mid Z_{t-1}\right]=0$$

as was desired. □

Lemma 1.35. *With* $\mathrm{d}\lambda_{jt}$ *given in section 1.7.1, we have that*

$$E\left[\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left(\left(\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta\right)\right)'\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta^{\dagger}\right)-\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\theta,\mathrm{d}\vartheta^{\dagger}\right)\right)\mid Z_{t-1}\right]=0,$$

where ϑ *is give by (1.16).*

Proof. First observe that

$$\begin{aligned}
&\left(\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta\right)\right)'\mathrm{d}\lambda_{jt}\left(\mathrm{d}\vartheta^{\dagger}\right)-\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}\vartheta,\mathrm{d}\vartheta^{\dagger}\right)\\
=&\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)-\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta,\mathrm{d}\vartheta^{\dagger}\right)\\
&+\left(\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)-\mathrm{d}^2\log p_{jt}\left(\theta;\mathrm{d}\vartheta;\mathrm{d}\vartheta^{\dagger}\right)\\
&+\left(\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)+\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\left(\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)\right)
\end{aligned}$$

Now, it holds by Lemmas 1.34 and 1.19 that

$$E\left[\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)-\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta,\mathrm{d}\vartheta^{\dagger}\right)\mid Z_{t-1}\right]=0,$$

and

$$E\left[\sum_{j\in\mathbb{M}} p_{jt}^{\star}\left(\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)-\mathrm{d}^2\log p_{jt}\left(\theta;\mathrm{d}\vartheta;\mathrm{d}\vartheta^{\dagger}\right)\mid Z_{t-1}\right]=0.$$

Observe next that all cross products,

$$\left(\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)\right)\quad\text{and}\quad\left(\mathrm{d}\log\phi_{jt}\left(\theta;\mathrm{d}\vartheta\right)\right)'\left(\mathrm{d}\log p_{jt}\left(\theta;\mathrm{d}\vartheta^{\dagger}\right)\right)$$

are zero. This completes the proof. □

1.F Smoothness of probability parametrization

In this Appendix, we verify that Assumption 1.5 is satisfied by the specifications in Section 1.4. Note that $\log p_{jt}=\log p_t+\log\pi_{jt.1}$ for $j\in\mathbb{M}_1$ and $\log p=\log\left(1-p_t\right)+\log\pi_{jt.2}$ for $j\in\mathbb{M}_2$. We show that Assumption 1.5 holds separately for $\log p_t$ and $\log\pi_{jt.i}$; this implies that Assumption 1.5 holds for $\log p_{jt}$.

We hence let p indicate either p_t or $\pi_{jt.i}$, and write $p=f\left(\exp\left(g\right)\right)$ where $f\left(x\right)=1-1/x$ and $g\left(z;\varrho\right)=\left(z-\mu\right)\Lambda\left(z-\mu\right)$ for the exponential specification of p_t, $f\left(x\right)=x/\left(x+1\right)$ and $g\left(z;\varrho\right)=\left(z-\mu\right)\Lambda\left(z-\mu\right)-\varpi$ for the logistic specification of p_t, $f\left(x\right)=x/\left(x+c\right)$ and $g\left(z;\zeta\right)=\zeta_j'z$ for the logistic specification of $\pi_{jt.i}$.

We note that for all specifications $p > 0$ and that

$$\begin{aligned}
\partial_u \log p &= p^{-1}\partial_u p, \\
\partial_{uv} \log p &= -p^{-2}\left(\partial_u p\right)'\left(\partial_v p\right) + p^{-1}\partial_{uv} p, \\
\partial_{uvw} \log p &= 2p^{-3}\left(\left(\partial_u p\right)' \otimes \left(\partial_v p\right)'\right)\partial_w p - p^{-2}\left(\left(\partial_v p' \otimes I\right)\partial_{uw} p + \left(I \otimes \left(\partial_u p\right)'\right)\partial_{vw} p\right) + \\
&\quad -p^{-2}\mathrm{vec}\left(\partial_{uv} p\right)\partial_w p + p^{-1}\partial_{uvw} p,
\end{aligned}$$

where, indicating the j-th derivative of f as $f^{(j)}(x)$ and letting $f^{(j)} = f^{(j)}(\exp(g))$,

$$\begin{aligned}
\partial_u p &= \exp(g) f^{(1)} \partial_u g, \\
\partial_{uv} p &= \exp(2g) f^{(2)} \left(\partial_u g\right)'\left(\partial_v g\right) \\
&\quad + \exp(g) f^{(1)} \left(\partial_u g\right)'\left(\partial_v g\right) + \exp(g) f^{(1)} \partial_{uv} g, \qquad (1.60) \\
\partial_{uvw} p &= 2\exp(2g) f^{(2)} \mathrm{vec}\left(\left(\partial_u g\right)'\left(\partial_v g\right)\right)\partial_w g \\
&\quad + \exp(2g) f^{(3)} \mathrm{vec}\left(\left(\partial_u g\right)'\left(\partial_v g\right)\right)\partial_w g \\
&\quad + \exp(2g) f^{(2)} \left(\left(\partial_v g' \otimes I\right)\partial_{uw} g + \left(I \otimes \left(\partial_u g\right)'\right)\partial_{vw} g\right) \\
&\quad + \exp(2g) f^{(2)} \mathrm{vec}\left(\left(\partial_u g\right)'\left(\partial_v g\right)\right)\partial_w g \\
&\quad + \exp(g) f^{(1)} \mathrm{vec}\left(\left(\partial_u g\right)'\left(\partial_v g\right)\right)\partial_w g \\
&\quad + \exp(g) f^{(1)} \left(\left(\partial_v g' \otimes I\right)\partial_{uw} g + \left(I \otimes \left(\partial_u g\right)'\right)\partial_{vw} g\right) \\
&\quad + \exp(2g) f^{(2)} \mathrm{vec}\left(\partial_{uv} g\right)\partial_w g \\
&\quad + \exp(g) f^{(1)} \mathrm{vec}\left(\partial_{uv} g\right)\partial_w g + \exp(g) f^{(1)} \partial_{uvw} g. \qquad (1.61)
\end{aligned}$$

In the following, we show that (i) $\exp(jg) f^{(j)} = O(\exp(-\|g\|))$ with $j = 1, 2$ and (ii) $\partial_u g$, $\partial_{uv} g$, $\partial_{uvw} g$ are at most quadratic in $\|z\|$ for large $\|z\|$. Conditions (i) and (ii) are sufficient to ensure that Assumption 1.5 holds for $\log p_{jt}$ because all terms in (1.60) contain one term $\exp(jg) f^{(j)}$ multiplied by products of $\partial_u g$, $\partial_{uv} g$, $\partial_{uvw} g$. We next discuss conditions (i) and (ii) in turn.

Condition (i). For the case $f(x) = 1 - 1/x$ in the exponential specification of p_t, one has $f^{(j)}(x) = (-1)^{j+1} j! x^{-(j+1)}$ so that $\exp(jg) f^{(j)} = \exp(jg)(-1)^{j+1} j! \exp(-(j+1)g)) = O(\exp(-\|g\|))$. For the case $f(x) = x/(x+c)$ in the logistic specification of p_t (with $c = 1$) and in the logistic specification of $\pi_{jt.i}$, one finds $f^{(j)}(x) = (-1)^{j+1} j! (x-c)^{-(j+1)}$ so that $\exp(jg) f^{(j)} = \exp(jg)(-1)^{j+1} j! (\exp(-g) - c)^{-(j+1)} = O(\exp(-\|g\|))$. Hence in both cases one has $\exp(jg) f^{(j)} = O(\exp(-\|g\|))$ with any $j = 1, 2, ...$

Condition (ii). It is simple to verify that $\partial_u g$, $\partial_{uv} g$, $\partial_{uvw} g$ are at most quadratic in $\|z\|$. In fact, in the case $g(z; \varrho) = (z-\mu)\Lambda(z-\mu)$, the first derivatives are $\partial_\mu g = -\partial_z g = -2(z-\mu)'\Lambda$, $\partial_{\mathrm{vec}\Lambda} g = (z-\mu) \otimes (z-\mu)$, the second derivatives are

$$\begin{aligned}
\partial_{\mu z} g &= \partial_{z\mu} g = -\partial_{\mu\mu} g = -\partial_{zz} g = 2\Lambda, \\
\partial_{\mathrm{vec}\Lambda \mu} g &= -\partial_{\mathrm{vec}\Lambda z} g = -\left(I \otimes (z-\mu)\right) - \left((z-\mu) \otimes I\right), \qquad \partial_{\mathrm{vec}\Lambda \mathrm{vec}\Lambda} g = 0,
\end{aligned}$$

and third order derivatives (where here (uvw) means uvw in any order)

$$\begin{aligned}
\partial_{\mu z \mathrm{vec}\Lambda} g &= \partial_{z\mu \mathrm{vec}\Lambda} g = -\partial_{\mu\mu \mathrm{vec}\Lambda} g = -\partial_{zz \mathrm{vec}\Lambda} g = 2I, \\
\partial_{(z\mu\mu)} g &= \partial_{(\mu zz)} g = \partial_{zzz} g = \partial_{\mu\mu\mu} g = 0, \\
\partial_{\mathrm{vec}\Lambda \mathrm{vec}\Lambda z} g &= \partial_{\mathrm{vec}\Lambda \mathrm{vec}\Lambda \mu} g = 0, \qquad \partial_{\mathrm{vec}\Lambda \mu z} g = -\partial_{\mathrm{vec}\Lambda zz} g = c.
\end{aligned}$$

Similar derivations apply for $g(z;\varrho) = (z-\mu)\Lambda(z-\mu) - \varpi$, which is a translation of the above case, and to the linear case $g(z;\zeta) = \zeta_j' z$.

2 Estimation and testing in dynamic mixture cointegrated VAR models

We discuss model selection, estimation and testing within the ACR cointegrated model discussed in chapter one. A framework based on generalized linear restrictions is used to consider estimation of restricted models. We discuss construction of likelihood ratio tests, in two separate cases. First, a *regular* case, where all parameters are identified under the null hypothesis; and second, an *irregular* case, where some parameters are unidentified under the null hypothesis. We apply the asymptotic theory given in chapter one to derive the asymptotic distribution of the likelihood ratio test in the regular case. In both cases asymptotic inference is non-standard and the distributions are nuisance parameter dependent. We propose a bootstrap algorithm to simulate these distributions and investigate its performance through simulations.

2.1 Introduction

In chapter one, the dynamic mixture cointegrated VAR model called the Autoregressive Conditional Root (ACR) cointegrated model was introduced and the asymptotic theory for the maximum likelihood estimator was derived. This chapter takes a more practical approach and considers estimation and testing by building an encompassing framework that allows for analysis of a wide range of differently specified ACR cointegrated models. The framework is based on the principles of generalized linear restrictions of the type discussed in Boswijk and Doornik (2004). These principles turn out to be practical not only for identifying the cointegration vector and imposing testable restrictions, but also provides a convenient way of selecting which of the model parameters are chosen to be switching. Moreover, we show that it is straight forward to modify the estimators of the EM algorithm from Bec and Rahbek (2004) and Bec et al. (2008) to allow for generalized linear restrictions and to accommodate for estimation of the cointegration relations. The properties of the extended EM algorithm in small samples is evaluated through a simple Monte Carlo simulation study, using the example system discussed in chapter one, section 1.5 as the data generating process.

Next, we discuss implementation and asymptotic theory of the likelihood ratio test. In particular, testing in two different cases is considered. First, a *regular* case where all parameters are identified under the null and under the alternative; and second an *irregular* case where some parameters are unidentified under the null. The irregular case is well known from the literature on non-linear autoregressive models since it arises in particular when one wishes to test for linearity, see inter alia Davies (1987); Andrews and Ploberger (1994); Hansen (1996); Caner and Hansen (2001) and Kristensen and Rahbek (2013). The asymptotic theory developed in chapter

one is applied to show convergence in distribution of the likelihood ratio test in the regular case, while the theory for the irregular case still needs to be developed, since the results of Kristensen and Rahbek (2013) are not directly transferable. However, we do believe that such a theorem can be worked out and we state a convergence result for the irregular case as a conjecture. For both the regular and the irregular cases, the distributions of the test statistics are nuisance parameter depend when the cointegration vectors are estimated. Consequently, inference based on χ^2 approximations is invalid and instead, these distributions must be simulated on a case by case basis.

We introduce a semi-parametric, model-based bootstrap algorithm that relies on resampling of the estimated residuals. This algorithm applies principles from the bootstrap literature on linear dynamic models discussed in, inter alia Bose (1988); Horowitz (2001); Lahiri (2003); Kreiss and Paparoditis (2011); Cavaliere and Taylor (2008); Cavaliere et al. (2010a,b, 2012). The performance of the proposed bootstrap algorithm is investigated through simulation.

This chapter is structured as follows. Section 2.2 presents the model and discusses specification and estimation subject to generalized linear restrictions. Section 2.2.4 contains a Monte Carlo simulation experiment illustrating the small sample properties of the maximum likelihood estimator based on the EM algorithm. Section 2.3 discusses hypothesis testing based on likelihood ratio statistics. Section 2.4 presents the bootstrap algorithm and investigates its validity through two simulation experiments. Finally, section 2.5 concludes

We use the same definitions for standard matrices and operators as in chapter one.

2.2 Model specification and estimation

In this section, we briefly present the model, discuss model specification based on generalized linear restrictions and introduce an iterative algorithm for estimation under these restrictions.

2.2.1 The model

The n-dimensional ACR cointegrated process, X_t, is generated by the equations

$$\begin{aligned}\Delta X_t &= \sum_{j\in\mathbb{M}} \mathbf{1}\{s_t = j\}\left(\alpha_j\beta^{\star\prime}X^{\star}_{t-1} + \Gamma_j\Delta\mathbb{X}_{t-1} + V_j\epsilon_t\right)\\ &= \sum_{j\in\mathbb{M}} \mathbf{1}\{s_t = j\}\left(\Phi_j U_{t-1} + V_j\epsilon_t\right) \quad \text{with} \quad \epsilon_t \sim \text{i.i.d.}\left(0, I_n\right), \end{aligned} \tag{2.1}$$

where the same definitions as in chapter one, section 1.2 have been used. In addition, we introduce the collecting variable, U_t, is defined as $U_t = (X_t^{\star\prime}\beta^{\star}, \Delta\mathbb{X}_t')'$ and the collecting parameters $\Phi_j = (\alpha_j, \Gamma_j)$, $\Phi := (\Phi_1 : \cdots : \Phi_m)$ and $\Omega := (\Omega_1 : \cdots : \Omega_m)$. The parameters are collected into the vector

$$\begin{aligned}\Theta &= \left(\text{vec}\left(\beta^{\star\prime}\right)' : \text{vec}\left(\Phi\right)' : \text{vec}\left(\Omega\right)' : \gamma'\right)'\\ &= \left(\text{vec}\left(\beta'\right)' : \beta_D' : \text{vec}\left(\Phi\right)' : \text{vec}\left(\Omega\right)' : \gamma'\right)'.\end{aligned}$$

We denote the regime switching probability as $p_{jt} := P\left(s_t = j \mid z_{t-1}; \gamma\right)$, where $z_t := \psi' Z_t$,

$Z_t = (X_t'\beta, \Delta\mathbb{X}_t')'$ and ψ is some selection matrix. The vector, γ, is the parameter indexing the probability function. Note that $U_t \neq Z_t$ since only U_t contains the constant term in the cointegration relations. For a full description of the model, the necessary assumptions for stationarity and ergodicity of the process as well as examples of predicted state probabilities, see chapter one.

2.2.2 Model selection and identification of parameters

When applying the ACR cointegrated model to data, it will often be of interest to fix some of the parameters across regimes from the outset and to impose identifying restrictions on other parameters. We will denote the vector of freely varying elements of Θ as θ, where

$$\Theta = H_\Theta \theta + h_\Theta$$

with H_Θ is a selection matrix and h_Θ a normalizing vector such as discussed by Boswijk and Doornik (2004). We further partition θ such that

$$\theta := (\varphi' : \varpi' : \omega' : \delta')' \tag{2.2}$$

where φ, ϖ, ω and δ are the freely varying elements of $\mathrm{vec}\,(\beta^{\star\prime})$, $\mathrm{vec}\,(\Phi)$, $\mathrm{vec}\,(\Omega)$ and δ, respectively.

Example 2.1. As an example, consider the simulated process given in chapter one, section 1.5. In this system, the short run parameters, Γ_j, where set to be equal across regimes, i.e. $\Gamma_1 = \Gamma_2 = \Gamma_3$. This condition ensures that assumption 1.3 is satisfied despite having non-stationary regimes and having that the switching only depends on $\beta' X_{t-1}$. Hence that model is a special case of the model given in (2.1). We can enforce this property by introducing the selection matrix H_Φ and a normalizing vector h_Φ, such that

$$\begin{aligned}
\mathrm{vec}\,(\Phi) &= \left(\mathrm{vec}\,(\alpha_1)' : \mathrm{vec}\,(\Gamma)' : \mathrm{vec}\,(\alpha_2)' : \mathrm{vec}\,(\Gamma)' : \mathrm{vec}\,(\alpha_3)' : \mathrm{vec}\,(\Gamma)'\right)' \\
&= H_\Phi \left(\mathrm{vec}\,(\alpha_1)' : \mathrm{vec}\,(\Gamma)' : \mathrm{vec}\,(\alpha_2)' : \mathrm{vec}\,(\alpha_3)'\right)' + h_\Phi \\
&= H_\Phi \varpi + h_\Phi
\end{aligned}$$

where

$$H_\Phi = \begin{pmatrix}
\mathcal{I}_2 & 0 & 0 & 0 \\
0 & \mathcal{I}_4 & 0 & 0 \\
0 & 0 & \mathcal{I}_2 & 0 \\
0 & \mathcal{I}_4 & 0 & 0 \\
0 & 0 & 0 & \mathcal{I}_2 \\
0 & \mathcal{I}_4 & 0 & 0
\end{pmatrix} \quad \text{and} \quad h_\Phi = \mathbf{0}_{12\times 1}.$$

The vector of freely varying parameters of Φ is then given by ϖ. Observe further that since Ω_j for all $j \in \mathbb{M}$ are symmetric matrices we set

$$\mathrm{vec}\,(\Omega) = \left(\mathrm{vech}\,(\Omega_1)'\,\mathcal{D}_n' : \mathrm{vech}\,(\Omega_2)'\,\mathcal{D}_n' : \mathrm{vech}\,(\Omega_3)'\,\mathcal{D}_n'\right)' = H_\Omega \omega + h_\Omega$$

where $H_\Omega = \text{diag}(\mathcal{D}_n, \mathcal{D}_n, \mathcal{D}_n)$, $h_\Omega = \mathbf{0}_{9\times 1}$ and $\omega = (\omega_1' : \omega_2' : \omega_3')'$, with $\omega_j = \text{vech}(\Omega_j)$ where

$$\Omega_j = \begin{pmatrix} \omega_{11,j} & \omega_{12,j} \\ \omega_{12,j} & \omega_{22,j} \end{pmatrix}$$

for all $j \in \{1,2,3\}$. For identification of the cointegration relations we introduce H_β and h_β such that

$$\text{vec}(\beta^{\star\prime}) = \left(H_\beta' : H_{\beta_D}'\right)' \left(\varphi_\beta' : \varphi_{\beta_D}'\right)' + h_{\beta^\star} =: H_{\beta^\star}\varphi + h_{\beta^\star}$$

and $H = (\mathbf{0}_{2\times 1} : \mathcal{I}_2)'$ and $h_{\beta^\star} = (1 : \mathbf{0}_{1\times 2})'$. Finally, recall that the probability parameters for this example are given by $\gamma = (\Lambda : \mu : \zeta_1 : \zeta_2 : \zeta_3)'$ and that the regime structure was such that $\mathbb{M}_1 = \{1,2\}$, and $\mathbb{M}_2 = 3$. The parameters entering the second layer regime switching specification were $\zeta = (\zeta_1 : \zeta_2 : \zeta_3)'$, one needs to set $\zeta_1 = \zeta_3 = 0$ to identify ζ. We can once again impose that identification through the matrix H_γ and the vector h_γ such that

$$\gamma = (\Lambda : \mu : 0 : \zeta_2 : 0)' = H_\gamma (\Lambda : \mu : \zeta_2)' + h_\gamma =: H_\gamma \delta + h_\gamma.$$

In all the we obtain the unrestricted parameters of the selected, identified model as given by (2.2).

2.2.3 Maximum likelihood estimation

The Gaussian (log-)likelihood as a function of θ can be written as

$$L_T(\theta) = \sum_{t=1}^{T} \ell_t(\theta) = \sum_{t=1}^{T} \sum_{j\in\mathbb{M}} \log(p_{jt}\phi_{jt}) \tag{2.3}$$

with p_{jt} defined in the previous section and ϕ_{jt} given by

$$\log \phi_{jt} \quad = \quad -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\Omega_j) - \frac{1}{2}\varepsilon_{jt}'\Omega_j^{-1}\varepsilon_{jt}$$

and

$$\varepsilon_{jt} := \Delta X_t - \alpha_j \beta^{\star\prime} X_{t-1}^\star - \Gamma_j \Delta \mathbb{X}_{t-1} = \Delta X_t - \Phi_j Z_{t-1}.$$

Define further the filtered probabilities,

$$p_{jt}^\star := P(s_t = 1 \mid Z_t, Z_{t-1}) = \frac{p_{jt}\phi_{jt}}{\sum_{j\in\mathbb{M}} p_{jt}\phi_{jt}}$$

which play a central role in the following estimation algorithm.

The algorithm considered here is an extension of the ones given in Bec et al. (2004) and Bec et al. (2008). As noted in Bec et al. (2008), maximizing the likelihood using the updating recursions discussed in the following is equivalent to applying an EM algorithm. It turns out that the EM-loglikelihood to be maximized in the so-called M-step is given by

$$L_T^{EM}(\theta) = \sum_{t=1}^{T} \sum_{j\in\mathbb{M}} p_{jt}^\star (\log p_{jt} + \log \phi_{jt}). \tag{2.4}$$

Closed form estimators for ϖ and ω given $p^{\star}_{jt}$ are provided in Lemmas 2.12 and 2.13 found in Appendix 2.A. The cointegration parameters, φ_{β}, and the probability parameters, γ, do not have closed form estimators and numerical methods are used for those.

Algorithm 2.2. *Algorithm for maximizing the likelihood*

1. *Set initial values for the parameters, denoted* $\hat{\varphi}_1$, $\hat{\varpi}_1$, $\hat{\omega}_1$ *and* $\hat{\delta}_1$ *and calculate the value of the likelihood function at the initial step,* $L_T(\theta_1)$.
2. *Set the iteration number to* $i = 2$.
3. *Update the filtered probabilities, to obtain* $\hat{p}^{\star}_{jt,i}$ *using* $\hat{\varphi}_i$, $\hat{\varpi}_i$, $\hat{\omega}_i$ *and* $\hat{\delta}_i$.
4. *Find* $\hat{\varphi}_i$ *by numerically maximizing (2.4) with respect to* φ *holding* $\hat{\varpi}_{i-1}$, $\hat{\omega}_{i-1}$, $\hat{\delta}_{i-1}$ *and* $p^{\star}_{jt,i}$ *fixed.*
5. *Calculate* $\hat{\varpi}_i$ *using (2.12) from Lemma 2.12 holding* $\hat{p}^{\star}_{jt,i}$ *fixed.*
6. *Calculate* $\hat{\omega}_i$ *using (2.13) from Lemma 2.13 holding* $\hat{p}^{\star}_{jt,i}$ *fixed.*
7. *Find* $\hat{\delta}_i$ *by numerical maximization of (2.4) with respect to* δ *holding* $\hat{\varpi}_i$, $\hat{\varphi}_i$, $\hat{\omega}_i$ *and* $p^{\star}_{jt,i}$, *fixed.*
8. *Calculate* $L_T(\theta_i) - L_T(\theta_{i-1})$ *and update the iteration number,* $i = i + 1$.
9. *Repeat steps 3-8 until* $L_T(\theta_i) - L_T(\theta_{i-1}) < c$, *for some small value* c.

Remark 2.3. Good candidates for initial values of the parameters of interest can be an important part of estimating ACR cointegrated models, in particular when the dimension of the system grows. Unfortunately there is no sure method to easily identify good initial values. One can use that ϖ and ω are less sensitive to initial values and that given $p^{\star}_{jt}$, closed form estimators are available. That is, a fairly robust procedure for finding good initial values is to do a grid search over the parameters entering the switching probability, namely φ_b and δ, while estimating φ_{β_D}, ϖ and ω using the EM algorithm. This procedure is the generalization of the profile likelihoods proposed in Bec et al. (2008). The problem is that, when many parameters enter the switching probability, the curse of dimensionality will result in an unreasonably severe computational burden. Thus, this approach is in particular infeasible when estimations are done inside the bootstrap algorithms.

When choosing grids for initial value searches, it will often be necessary to do a little trial an error to find the relevant regions of the parameters space. However, one can often take some useful considerations on the shape of the likelihood function and the properties of z_t into account. An example is discussed in chapter four, where the switching probability is logistic. In that case, the likelihood function can be *rippled* in the direction of some probability parameters and of the cointegration vectors, meaning that the grid should be dense in these directions. In contrast, the likelihood function will in direction only have few local maxima but large flat areas. Both observations are related to the fact that a limiting model for the logistic cointegrated ACR model is a threshold error correction model (see e.g. Balke and Fomby (1997); Hansen and Seo (2002); Seo (2011)) which is known to have a non-smooth likelihood function. Indeed, if

one finds that the likelihood is very sensitive to initial values for some parameters, it might be prudent to simplify the switching probability function, fix some of the troublesome parameter or change framework to e.g. the threshold error correction framework.

Remark 2.4. Observe that the restriction matrices can be chosen such that $\Phi_1 = \Phi_2 = ... = \Phi_m$ and $\Omega_1 = \Omega_2 = ... = \Omega_m$, which has the consequence that the model becomes linear. Hence, defining estimation in terms of the restriction principles given above allows for modeling linear cointegration models within the more general ACR cointegrated framework.

Remark 2.5. Observe that in fact, an estimator conditional on the predicted state probabilities for φ_{β_D} can be found in this case. However, in chapter three we show that the theory given in chapter one can be modified such that β_D is allowed to enter the probability of switching and in that case, no closed form estimator will be available for φ_D. The EM-algorithm as it is written in algorithm 2.2 works for both cases and is thus retained here.

Remark 2.6. This algorithm clarifies the extensions to earlier algorithms since excluding the step for calculating $\hat{\varphi}_i$ reduces the algorithm to multiregime versions of the ones discussed in Bec and Rahbek (2004) and Bec et al. (2008). However, other alternatives could be considered for maximizing (2.3). For example, the numerical optimization step for calculating $\hat{\varphi}_i$ and $\hat{\delta}_i$ could be collected into a single step where the EM-logliklihood is maximized with respect to $\left(\hat{\varphi}_i', \hat{\delta}_i'\right)'$.

2.2.4 Performance of the MLE

In this section, we evaluate the performance of the presented, extended EM-algorithm through a simple Monte Carlo simulation study that uses the example from chapter one, section 1.5 as a data generating process.

Monte Carlo simulations of the estimator

With a total of $M = 10^4$ replications, Table 2.1 reports the first summary statistics, including estimated bias (Bias), root mean square error (RMSE), the 5% and 95% percentiles as well as the median (Median). Note that even for a moderate sample of $T = 500$ all parameters are seen to have close to zero bias and small root mean square error, except for the transition parameters Λ and ζ_2. In particular, note that the distribution of ζ_2 is skewed with a think right tail. This is an artifact of the identification difficulties for logistic shape parameters, an effect that is discussion in detail in chapter four. Moreover, ζ_2 only enters the likelihood when the process is in regimes belonging to $\mathbb{M}_1$, which will tend to enhance this problem.

To investigate these effects for different true values of Λ and ζ_2, we consider simulations for combinations of true values for Λ and ζ_2. These are given in table 2.2. Observe that for larger values of Λ, ζ_2 is better identified since more observations will enter regimes in $\mathbb{M}_1$. It it also seen that the bias and root mean square errors grow a lot when the true value of the parameter grows. This is in particular the case for ζ_2 and is a consequence of the logistic structure of this switching probability, which makes the likelihood flatten out in the direction of ζ_2.

In cases with identification difficulties for the probability parameters, one may consider the option to fix one (or even both) of these parameters at some value(s). Doing so would not affect the actual shape of the estimated transition function and these numerical problems would cease

Table 2.1: Monte Carlo simulations of small sample properties of the Maximum Likelihood estimator.

Parameters	True	Bias	RMSE	5 % Percentile	Median	95 % Percentile
$\hat{\varphi}_{\beta}$	-1.000	-0.000	0.008	-1.014	-1.000	-0.986
$\hat{\varphi}_{\beta_D}$	-1.000	-0.012	0.171	-1.329	-0.997	-0.758
$\hat{\alpha}_{1,1}$	-0.200	-0.004	0.037	-0.267	-0.203	-0.145
$\hat{\alpha}_{2,1}$	0.000	0.001	0.016	-0.024	0.001	0.028
$\hat{\alpha}_{2,1}$	0.000	-0.003	0.039	-0.071	-0.001	0.056
$\hat{\alpha}_{2,2}$	0.200	0.015	0.057	0.138	0.207	0.320
$\hat{\alpha}_{3,1}$	0.000	0.003	0.070	-0.110	0.004	0.115
$\hat{\alpha}_{3,2}$	0.000	-0.006	0.076	-0.135	-0.003	0.112
$\hat{\Gamma}_{11}$	-0.500	0.003	0.026	-0.539	-0.497	-0.454
$\hat{\Gamma}_{21}$	0.000	-0.003	0.026	-0.046	-0.002	0.041
$\hat{\Gamma}_{12}$	0.300	-0.003	0.037	0.235	0.297	0.358
$\hat{\Gamma}_{22}$	0.010	0.002	0.037	0.042	0.102	0.163
$\hat{\omega}_{11,1}$	0.010	-0.000	0.002	0.007	0.010	0.014
$\hat{\omega}_{12,1}$	0.000	-0.000	0.001	-0.002	-0.000	0.002
$\hat{\omega}_{22,1}$	0.010	-0.000	0.002	0.007	0.010	0.013
$\hat{\omega}_{11,2}$	0.050	-0.001	0.011	0.031	0.048	0.069
$\hat{\omega}_{12,2}$	0.000	-0.000	0.008	-0.013	-0.000	0.012
$\hat{\omega}_{22,2}$	0.050	-0.001	0.011	0.031	0.048	0.069
$\hat{\omega}_{11,3}$	0.100	-0.000	0.010	0.084	0.100	0.117
$\hat{\omega}_{12,3}$	0.000	-0.000	0.07	-0.012	-0.000	0.011
$\hat{\omega}_{22,3}$	0.100	-0.000	0.010	0.084	0.100	0.118
$\hat{\Lambda}$	3.000	0.305	0.960	2.168	3.158	4.884
$\hat{\mu}$	1.000	-0.010	0.095	0.835	0.990	1.142
$\hat{\zeta}_2$	3.000	10.484	54.619	1.934	3.023	33.291

The sample size is 500 and the number of replications is 10000.

Table 2.2: Monte Carlo Simulations for different values of the probability parameters, Λ and ζ_2.

Parameters	True	Bias	RMSE	5 % Percentile	Median	95 % Percentile
$\hat{\Lambda}$	1.000	0.115	0.364	0.640	1.064	1.736
$\hat{\zeta}_2$	1.000	1.761	21.790	0.513	0.998	2.135
$\hat{\Lambda}$	1.000	0.137	0.351	0.697	1.082	1.735
$\hat{\zeta}_2$	2.000	11.077	54.272	1.153	1.968	35.621
$\hat{\Lambda}$	1.000	0.152	0.346	0.730	1.094	1.745
$\hat{\zeta}_2$	3.000	22.984	75.094	1.535	3.013	173.48
$\hat{\Lambda}$	2.000	0.168	0.674	1.306	2.080	3.303
$\hat{\zeta}_2$	1.000	0.197	5.283	0.547	0.997	1.819
$\hat{\Lambda}$	2.000	0.196	0.645	1.394	2.098	3.258
$\hat{\zeta}_2$	2.000	4.255	33.34	1.273	2.008	7.318
$\hat{\Lambda}$	2.000	0.236	0.641	1.464	2.134	3.334
$\hat{\zeta}_2$	3.000	14.218	62.255	1.780	3.031	62.131
$\hat{\Lambda}$	3.000	0.230	0.984	1.963	3.090	4.949
$\hat{\zeta}_2$	1.000	0.197	7.080	0.544	1.007	1.698
$\hat{\Lambda}$	3.000	0.238	0.913	2.051	3.104	4.855
$\hat{\zeta}_2$	2.000	2.908	29.225	1.333	2.024	4.139
$\hat{\Lambda}$	3.000	0.305	0.960	2.168	3.158	4.884
$\hat{\zeta}_2$	3.000	10.484	54.619	1.934	3.023	33.291

The sample size is 500 and the number of replications is 10000 for each set of parameters.

to exist. Moreover, the limit theory provided in chapter one would still apply for the remaining parameters.

2.3 Testing hypotheses

Testing hypothesis in this framework can be done using the same type of linear restrictions as was considered for model selection above. For testing purposes, we introduce the additional over-identifying restrictions. That is, we can write

$$\theta \quad = \quad H_\theta \upsilon + h_\theta$$

where H_θ and h_θ specifies the additional restrictions on θ that we wish to test and υ is the vector of freely varying parameters under those restrictions. We denote $\tilde{\theta}$ as the *restricted* parameter vector, i.e. the parameter vector where the freely varying elements are given by υ; and $\hat{\theta}$ as the *unrestricted* parameter vector given by (2.2) .

2.3.1 The regular case

Using the definitions given in the previous section, the likelihood ratio test statistic has the form

$$LR_T\left(\hat{\theta}, \tilde{\theta}\right) \quad = \quad -2\left[L_T\left(\tilde{\theta}\right) - L_T\left(\hat{\theta}\right)\right] \tag{2.5}$$

The asymptotic properties follow directly from the asymptotic theory for the estimators given in chapter one, Theorem 1.7, and are stated in the following Theorem 2.7.

Theorem 2.7. *With the ACR cointegrated process defined in chapter one, definition 1.1, using a suitably defined normalizing matrix W_T, and provided assumptions 1.2-1.5 are satisfied with $q \geq 3$, then the following weak convergence result applies to the likelihood ratio test defined in (2.5),*

$$LR_T\left(\hat{\theta}, \tilde{\theta}\right) \quad \xrightarrow{w} \quad \mathbb{V}\left(\theta_0\right)' \mathbb{V}\left(\theta_0\right),$$

where

$$\mathbb{V}\left(\theta\right) = \left(H_{\theta,\perp}' \mathbb{H}\left(\theta_0\right)^{-1} H_{\theta,\perp}\right)^{-1/2} H_{\theta,\perp}' \mathbb{H}\left(\theta_0\right)^{-1} \mathbb{S}\left(\theta_0\right),$$

and where $\mathbb{S}\left(\theta_0\right)$ and $\mathbb{H}\left(\theta_0\right)$ are the limits of $W_T^{-\frac{1}{2}} \mathbb{S}_T\left(\theta_0\right)$ and $W_T^{-\frac{1}{2}} \mathbb{H}_T\left(\theta_0\right) W_T^{-\frac{1}{2}}$ as $T \to \infty$ with the definitions,

$$\mathbb{S}_T\left(\theta_0\right) := \frac{\partial L_T\left(\theta\right)}{\partial \theta} \mid_{\theta=\theta_0} \text{ and } \quad \mathbb{H}_T\left(\theta_0\right) := \frac{\partial^2 L_T\left(\theta\right)}{\partial \theta \partial \theta} \mid_{\theta=\theta_0}.$$

Proof. The proof is given in Appendix 2.B. □

2.3.2 The irregular case

The result in the previous section applies to tests where all freely varying parameters are identified under the null. However, in the ACR cointegrated model (as in many other non-linear

models) a number of typical hypothesis of interest will result in unidentified parameters under the null. This is for example the case when testing for linearity where the switching variable s_t disappears under the null, and the parameters in the predicted state probabilities no longer enter the likelihood. In that case, one may apply supremum statistics similar to those considered in e.g. Davies (1987); Andrews and Ploberger (1994); Hansen (1996); Caner and Hansen (2001) and Kristensen and Rahbek (2013).

In the following, denote the subset of the probability parameters that vanish under the null as $\varrho \subseteq \gamma$ and specify the parameter vector as $\theta = (\eta', \varrho')'$. The vanishing parameters are defined such that $\varrho \in \Xi$ where Ξ is compact and $\Xi \subset \mathbb{R}^{n_\varrho}$. We write the supremum statistic in terms of these parameters, i.e.

$$\sup_{\varrho \in \Xi} LR_T(\hat{\eta}, \tilde{\eta}, \varrho) = \sup_{\varrho \in \Xi} \left(-2\left[L_T(\tilde{\eta}) - L_T(\hat{\eta}, \varrho)\right]\right) \tag{2.6}$$

where $\hat{\eta}$ is estimated in the unrestricted model and $\tilde{\eta}$ is estimated under the null.

Kristensen and Rahbek (2013) provide asymptotic theory for this case in a similar framework. They show that the score of their likelihood function is an asymptotically tight partial sum process for which a uniform Donsker theorem applies, see also van der Vaart and Wellner (1996, Chapter 2.12). Using our notation, the score of the likelihood function in Kristensen and Rahbek (2013) can be written as

$$\mathbb{S}_T(\eta, \varrho) = \sum_{t=1}^{T} f(U_{t-1} : \eta, \varrho)\, \varepsilon_t \tag{2.7}$$

where ε_t is the error term of their model. The function, $f(\cdot)$, is some sufficiently smooth function of U_{t-1}, η and in particular of the vanishing parameters, ϱ[1]. The form given in (2.7) is a also known as a so-called residual-market empirical process, see e.g. Stute (1997); Stute et al. (1998) and Escanciano (2007). Observe that the score of the ACR likelihood function takes on a different form. From chapter one, section 1.7.1 we have that

$$\mathbb{S}_T(\eta, \varrho) = \sum_{t=1}^{T} \sum_{j \in \mathbb{M}} p^{\star}_{jt} \partial_\theta \lambda_{jt} \tag{2.8}$$

where both $p^{\star}_{jt}$ and λ_{jt} are functions of η, ϱ and of U_t and U_{t-1} (not only of U_{t-1}). Hence, the separation between the errors and some function of parameters and lagged values of the process, $f(\cdot)$, seen in (2.7) does not apply to the ACR cointegrated framework and so, the results from Kristensen and Rahbek (2013) do not carry over easily.

To show that a similar result holds for the ACR cointegrated framework, one would thus need to develop an independent theorem of uniform weak convergence of (2.8). This is not done here and we state instead the following conjecture.

Conjecture 2.8. *With the ACR cointegrated process defined in chapter one, definition 1.1, using a suitably defined normalizing matrix W_T and provided assumptions 1.2-1.5 are satisfied with $q \geq 3$, then the following weak convergence result applies to the likelihood ratio test defined*

[1] In fact, here U_t contains β_D which is not included in the framework of Kristensen and Rahbek (2013). However, to avoid unnecessarily complicated notation we ignore that fact here.

in (2.5),

$$\sup_{\varrho\in\Xi} LR_T\left(\hat{\eta},\tilde{\eta},\varrho\right) \quad \xrightarrow{w} \quad \sup_{\varrho\in\Xi} \mathbb{V}\left(\eta_0,\varrho\right)' \mathbb{V}\left(\eta_0,\varrho\right),$$

where

$$\mathbb{V}\left(\eta_0,\varrho\right) = \left(H_{\theta,\perp}'\mathbb{H}\left(\eta_0,\rho\right)^{-1} H_{\theta,\perp}\right)^{-1/2} H_{\theta,\perp}'\mathbb{H}\left(\eta_0,\rho\right)^{-1}\mathbb{S}\left(\eta_0,\varrho\right)$$

where $\mathbb{S}\left(\eta_0,\varrho\right)$ *and* $\mathbb{H}\left(\eta_0,\varrho\right)$ *are the limits of* $W_T^{-\frac{1}{2}}\mathbb{S}_T\left(\eta_0,\varrho\right)$ *and* $W_T^{-\frac{1}{2}}\mathbb{H}_T\left(\eta_0,\varrho\right)W_T^{-\frac{1}{2}}$ *as* $T\to\infty$ *where*

$$\mathbb{S}_{T,W}\left(\eta_0,\varrho\right) := \frac{\partial L_T\left(\theta\right)}{\partial\theta}\mid_{\eta=\eta_0} \text{ and } \quad \mathbb{H}_{T,W}\left(\eta_0,\varrho\right) := \frac{\partial^2 L_T\left(\theta\right)}{\partial\theta\partial\theta}\mid_{\eta=\eta_0}$$

and the subscript W *indicates that the terms are normalized correctly according to the individual speeds of convergence of the parameter estimates.*

Note that here, we have only treated the likelihood ratio test since this is most easily implemented in the bootstrap algorithms. However, similar results will hold for Lagrange multiplier and Wald tests for the same reasons as those given in Kristensen and Rahbek (2013).

2.4 Bootstrapping

We propose a model-based bootstrap algorithm for simulating the distributions of the test statistic which resamples the estimated residuals. However, the innovations of the ACR model, ϵ_t, are unobserved and not straight forwardly calculated. As an estimator, we use

$$\hat{\epsilon}_t = E\left[\sum_{j\in\mathbb{M}} \mathbf{1}\left\{s_t=j\right\} V_j^{-1}\hat{\varepsilon}_{jt} \mid Z_t, Z_{t-1}\right] = \sum_{j\in\mathbb{M}} p_{jt}^{\star} V_j^{-1}\hat{\varepsilon}_{jt} \tag{2.9}$$

where $\hat{\varepsilon}_{jt}$ are the estimated, regime specific residuals. A different estimator in form of so-called prediction error distributions, was given in Bec et al. (2008) could be also be used. However, the estimator proposed here is somewhat more straight forward.

Algorithm 2.9. *A bootstrap algorithm for estimating distributions of the likelihood ratio test statistics.*

1. *Estimate the model under the null to obtain* $\tilde{\theta}$.
2. *Control that the estimated model under the null satisfies assumptions 1.2-1.5, from chapter one.*
3. *Obtain the estimated residuals,* $\tilde{\epsilon}_t$*, using (2.9) and the regime specific residuals,*

$$\tilde{\varepsilon}_{jt} = \Delta X_t - \tilde{\alpha}_j\tilde{\beta}^{\star\prime}X_{t-1} - \tilde{\Gamma}_j\Delta\mathbb{X}_{t-1}.$$

4. *To ensure a zero mean for the empirical distribution of the bootstrap innovations, recenter the estimated ACR residuals,*

$$\tilde{\epsilon}_t^c = \tilde{\epsilon}_t - \frac{1}{T}\sum_{t=1}^{T}\tilde{\epsilon}_t,$$

and denote their empirical distribution as $\tilde{\mathbb{E}}_T = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{\tilde{\epsilon}_t^c \preceq x\}$.

5. *Set the iteration number to* $i = 1$.

6. *Obtain the bootstrap innovations for iteration* i, $\left\{\hat{\epsilon}_{t,i}^b\right\}_{t=1,...,T}$, *by resampling* $\hat{\epsilon}_t^c$ *independently and with replacement from* $\tilde{\mathbb{E}}_T$.

7. *Generate a bootstrap sample,* $\left\{X_{t,i}^b\right\}_{t=1,...,T}$, *using*

$$X_{t,i}^b = X_{t-1,i}^b + \sum_{j\in\mathbb{M}} \mathbf{1}\{s_t = j\}\left(\tilde{\alpha}_j\tilde{\beta}^{\star\prime}X_{t-1,i}^{\star b} + \tilde{\Gamma}_j\Delta\mathbb{X}_{t-1,i}^b + \tilde{V}_j\hat{\epsilon}_{t,i}^b\right)$$

where $\left(X_{0,i}^b, X_{-1,i}^b, ..., X_{-k,i}^b\right) = (X_0, X_{-1}, ..., X_{-k})$ *and where* s_t *is drawn from a multinomial distribution with probability* $p_{jt}^b = \left(s_t = j \mid z_{t-1}^b, \hat{\gamma}_i^b\right)$ *of drawing regime* j.

8. *Estimate the model under null to obtain* $\tilde{\theta}_i^b$ *and estimate the model under the alternative using* $\tilde{\theta}_i^b$ *as initial values to obtain* $\hat{\theta}_i^b$.

9. *Calculate an instance of the statistic of interest,*

$$LR_{T,i}^b\left(\tilde{\theta}_i^b, \hat{\theta}_i^b\right) = -2\left(L_T\left(\tilde{\theta}_i^b\right) - L_T\left(\hat{\theta}_i^b\right)\right) \quad or$$

$$\sup_{\varrho\in\Xi} LR_{T,i}^b\left(\tilde{\eta}_i^b, \hat{\eta}^b, \varrho\right) = \sup_{\varrho\in\Xi}\left(-2\left[L_T\left(\tilde{\eta}_i^b\right) - L_T\left(\hat{\eta}_i{}^b, \varrho\right)\right]\right).$$

10. *Increment the iteration number,* $i = i + 1$.

11. *Repeat steps 5-10 many times and generate bootstrap empirical distributions of the statics of interest,*

$$\mathbb{T}_T^b\left(\tilde{\theta}_i^b, \hat{\theta}_i^b\right) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}\left\{LR_{T,i}^b\left(\tilde{\theta}_i^b, \hat{\theta}_i^b\right) \preceq x\right\} \quad or$$

and

$$\mathbb{U}_T^b\left(\tilde{\eta}_i^b, \hat{\eta}_i^b, \varrho\right) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}\left\{\sup_{\varrho\in\Xi} LR_{T,i}^b\left(\tilde{\eta}_i^b, \hat{\eta}^b, \varrho\right) \preceq x\right\}.$$

Remark 2.10. The i.i.d resampling method used to obtain the bootstrap innovations given in the presented algorithm will eliminate any left over structure conditional variance of the error terms. If one wishes the bootstrap to replicate the presence of such a structure, a so-called wild bootstrap resampling scheme can be considered. Here, the bootstrap innovations are generated such that

$$\tilde{\epsilon}_{t,i}^b = \mathfrak{n}_{t,i}\tilde{\epsilon}_t$$

and $\mathfrak{n}_{t,i}$ is an $i.i.d\,(0,1)$ random variable, see e.g. Wu (1986), Liu (1988), Mammen (1993), Davidson and Flachaire (2008), Cavaliere et al. (2010a), and Kristensen and Rahbek (2013). Observe that Kristensen and Rahbek (2013) suggests a wild bootstrap based on the normal distribution as a candidate for $\mathfrak{n}_{t,i}$. However, in this framework it seems to work better when applying the so-called Rademacher distribution. The reason is that using the Rademacher distribution will make the kurtosis of $\tilde{\epsilon}_{t,i}^b$ match the kutosis of $\tilde{\epsilon}_t$, while the normal distribution

will give a multiplication of 3 and thus heavier tails in the bootstrap innovations. Of course many other distributions could be considered, see Davidson and Flachaire (2008) for a comprehensive discussion.

Remark 2.11. In Algorithm 2.9, step 8, we use the estimated values under the null as initial values for the unrestricted model. This procedure avoids potential negative values of the likelihood ratio test as was observed in Kristensen and Rahbek (2013). It does not ensure that one obtains the best possible maximum at each replication, but neither would using any other set of initial values. This procedure thus seems to be the better option. Of course, with infinite computational power, dense grid searches for optimization would be a useful method for avoiding the problem with local maxima all together.

Algorithm 2.9 provides asymptotically valid estimates of the distributions of the likelihood ratio tests in both the regular and irregular cases if

$$\mathbb{T}_{T,i}^{b}\left(\hat{\theta},\tilde{\theta}\right) \overset{w}{\rightarrow}_{p} \mathbb{V}\left(\theta_0\right)' \mathbb{V}\left(\theta_0\right) \quad \text{and} \quad \mathbb{U}_{T}^{b}\left(\hat{\eta}_i^b,\tilde{\eta}_i^b,\varrho\right) \overset{w}{\rightarrow}_{p} \sup_{\varrho\in\Xi} \mathbb{V}\left(\eta_0,\varrho\right)' \mathbb{V}\left(\eta_0,\varrho\right)$$

where $\overset{w}{\rightarrow}_{p}$ denotes weak convergence in probability as defined by Gine and Zinn (1990) and one would therefore prefer to verify this formally. However, even in the simplest possible scenario of a stationary univariate ACR process, the necessary limit theory is not available and hence the performance of the proposed algorithm will be evaluated through simulations. Note finally that if the cointegration vectors are considered fixed from the outset, then the asymptotic theory for the likelihood ratio statistic in (2.5) is standard χ^2, while the asymptotic distribution of (2.6) will still be nuisance parameter depend and will require simulation, see Bec and Rahbek (2004) and Bec et al. (2008) for more details on the regular case with fixed cointegration relations.

2.4.1 Numerical analysis of the bootstrap

Let $X_t = (x_{1t}, x_{2t})'$ be a two-dimensional vector generated by the equation

$$\Delta X_t \quad = \quad s_t\left(\alpha_1\beta^{\star\prime}X_{t-1}^{\star} + V_1\epsilon_t\right) + \left(1-s_t\right)\left(\alpha_2\beta^{\star\prime}X_{t-1}^{\star} + V_2\epsilon_t\right) \tag{2.10}$$

where $\epsilon_t \sim i.i.dN\left(0,1\right)$ and the remaining parameters are defined as in section 2.2.1. The switching probability is given by,

$$p_{1t} = P\left(s_t = 1 \mid Z_{t-1}\right) = 1 - \exp\left(-\gamma z_{t-1}^2\right) = 1 - P\left(s_1 = 0 | Z_{t-1}\right) = 1 - p_{2t}$$

with $z_t := \psi' Z_t = \beta' X_t$. We impose the following identifying and model selective restrictions. The selection matrices are given by $H_{\beta^\star} = \left(\mathbf{0}_{1\times 2} : \mathcal{I}_2\right)'$ $H_\Phi := \mathcal{I}_4$, $H_\Omega = \mathrm{diag}\left(\mathcal{D}_n, \mathcal{D}_n\right)$ and $H_\gamma = 1$; and the corresponding vectors are $h_\beta = \left(1 : \mathbf{0}_{1\times 2}\right)'$, $h_\Phi = \mathbf{0}_{2\times 1}$, $h_\Omega = \mathbf{0}_{6\times 1}$ and $h_\gamma = 0$. The freely varying parameters are defined accordingly.

As a data generating process, we use the following parameter values, $\alpha_1 = \left(-0.5 : 0.2\right)'$, $\alpha_2 = \left(0 : 0\right)'$, $\Omega_1 = 0.01 \cdot \mathcal{I}_n$, $\Omega_2 = 0.05 \cdot \mathcal{I}_n$ $\beta^\star = \left(1 : -1 : 0\right)'$ and $\gamma = 1$.

We investigate the properties of two example test statistics using the bootstrap algorithm to simulation the distributions. The first is a regular case statistic, where all parameters are

identified under the null and under the alternative. The second in an irregular case statistic, where a number of parameters are unidentified under the null.

Case I: No error correction of x_{2t}

Consider a frequently tested hypothesis in empirical applications, namely that of no error correction of a certain variable. In the cointegrated VAR models, this translates into testing for a zero row in the error correction parameter, α. In the linear setup, this amounts to test for weak exogeneity, but that is not generally the case in this setup. The hypothesis that x_{2t} does not error correct can be stated as,

$$\mathcal{H}_0 : a_{2,1} = a_{2,2} = 0$$

where $a_{i,j}$ is the i'th element in vector α_j for $j \in \{1, 2\}$. This hypothesis represents additional restrictions on the parameter vector, θ, which can be introduced using properly defined H_θ, h_θ and υ. In this case, we set $H_\theta = \text{diag}\left(H_{\beta^\star}, G_\Phi, H_\Omega, H_\gamma\right)$ with

$$G_\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

and $h_\theta = \mathbf{0}_{n_\theta}$. The statistic of interest is the likelihood ratio given in (2.5).

Case II: Linearity

Let $X_t = (x_{1,t}, x_{2,t})'$ be generated from

$$\Delta X_t = \alpha \beta^{\star\prime} X^\star_{t-1} + \varepsilon_t, \tag{2.11}$$

where $\varepsilon_t \sim i.i.N\left(0, \Omega\right)$ and α and $\beta^\star$ where $\alpha = (-0.5, 0.2)$, $\beta^\star = (1, -1, 0)'$ and $\Omega = 0.01 \cdot \mathcal{I}_n$. As discussed previously, this model is a sub-model of the ACR cointegrated model model since (2.11) emerges when imposing the following null on the parameters of (2.10);

$$\mathcal{H}_0 : \alpha_1 = \alpha_2 \quad \text{and} \quad \Omega_1 = \Omega_2.$$

Imposing this condition on the system will eliminate s_t from (2.10) making γ abundant. Hence, to test this hypothesis, one must apply the supremum statistics discussed in section 2.3 and the statistic we wish to analyze is given by

$$\max_{\varrho \in \dot{\Xi}} LR_T\left(\hat{\eta}, \tilde{\eta}, \varrho\right) = \max_{\varrho \in \dot{\Xi}} 2\left[L_T\left(\hat{\eta}\right) - L_T\left(\tilde{\eta}, \varrho\right)\right],$$

and here $\varrho = \gamma$.

Observe that we seek to obtain the maximum of the likelihood over the vanishing parameter to get the maximal likelihood ratio statistic. Hence, computationally there is no difference between the two example cases. In both, the problem with obtaining local maxima under the null or the

Table 2.3: Empirical rejection frequencies

	Simulating under the null			Simulating under the alternative		
T	10 %	5 %	1%	10 %	5 %	1%
250	0.095	0.047	0.009	1.000	0.999	0.994
500	0.095	0.045	0.005	1.000	1.000	1.000
1000	0.108	0.053	0.009	1.000	1.000	1.000

(a) Case I

	Simulating under the null			Simulating under the alternative		
T	10 %	5 %	1%	10 %	5 %	1%
250	0.059	0.027	0.003	0.998	0.992	0.961
500	0.072	0.027	0.005	1.000	1.000	1.000
1000	0.067	0.033	0.006	1.000	1.000	1.000

(b) Case II

alternative inside the bootstrap replications is present, however perhaps more so in this case where γ does not enter the likelihood under the null and it is unclear how this problem affects the of the bootstrap. Similar considerations are done in Kristensen and Rahbek (2013), where negative likelihood ratios were observed.

Empirical rejection frequencies

We evaluate the performance of the bootstrap algorithms by estimating the empirical rejection frequencies when generating data under the null and under the alternative respectively. To produce the empirical rejection frequencies, the likelihood ratio statistics are estimated and evaluated against a bootstrap distributions $M = 1000$ times. The bootstrap distributions are generated with $B = 399$ bootstraps. The resulting estimated rejection frequencies are provided in table 2.3.

The bootstrap statistics in the regular case seems very well behaved, displaying perfect empirical power and close to correct empirical size. The test statistic in the irregular case is not as well behaved since the estimated empirical size seem to undershoot the correct values. This result might be linked to numerical problems with estimation when a parameter is unidentified under the null. Interestingly, when conducting a similar experiment for the estimated systems in chapter three, we do not observe these size distortions and they might hence also be related to a poor choice of parameters in this example. The results do, however, point out some of the difficulties that can arise when applying this methodology, underlining the need for more analysis of the small sample behavior of the proposed bootstrap and in particular, for a theoretical verification of the method in the irregular case.

2.5 Conclusion

We have presented a modified EM algorithm for estimation of ACR cointegrated models with general linear restrictions. Some small sample properties of the EM algorithm are discussed through simulations and it is found that estimators for all parameters, safe certain parameter in the probability of switching, are well behaved. In practical applications, it can be considered to fix the parameters that cause problems, in which case the inference on the remaining parameters remains the same.

Numerical analysis of the methods indicate that the regular test have close to correct empirical size while the test in the irregular case is undersized. A better understanding of why this occurs for the irregular case would be useful, in particular one could investigate whether it is related to the numerical difficulties of estimating the non-linear model on data generated under the linear null.

For future research, it would be of interest to develop a theorem to replace conjecture 2.8 and to verify analytically that the proposed bootstrap algorithm is asymptotically valid. As an alternative to the bootstrap, one could look at direct simulations of the asymptotic distributions, using a similar method to that discussed in Kristensen and Rahbek (2010, Theorem 7). Analyzing the validity of such a scheme and investigating its numerical properties, in particular in the irregular case. Note however that such a procedure is inherently asymptotic and as an alternative one could use a fully parametric bootstrap, simulating the behavior of the likelihood ratio statistics for some $\overline{T} \gg T$ and with, say, normally distribution innovations instead of the resampled error used in algorithm 2.9. The two approaches would approximate the same asymptotic statistic and hence which is more useful becomes a question of numerically stability, computational speed and ease of implementation.

2.A Estimators for the EM-algorithm

Lemma 2.12. *Define ϖ as the vector that satisfies*

$$\text{vec}(\Phi) = H_\Phi \varpi + h_\Phi$$

for a selection matrix H_Φ and a normalizing vector h_Φ. An estimator for $\hat{\varpi}$ conditional on the filtered probabilities, $p^\star_{jt}$, for $j \in \mathbb{M}$, is given by

$$\hat{\varpi} = \left(H_\Phi' \mathbf{M} H_\Phi\right)^{-1} \left(\mathbf{Y} H_\Phi - h_\Phi' \mathbf{M} H_\Phi\right) \tag{2.12}$$

where

$$\mathbf{Y} := \left(\text{vec}\left(\Delta \mathbf{X}' \mathbf{U}_1^\star \Omega_1^{-1}\right)', \ldots, \text{vec}\left(\Delta \mathbf{X}' \mathbf{U}_m^\star \Omega_m^{-1}\right)' \right)',$$

$$\mathbf{M} := \begin{pmatrix} \left(\Omega_1^{-1} \otimes \mathbf{U}'\mathbf{U}_1^\star\right) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \left(\Omega_m^{-1} \otimes \mathbf{U}'\mathbf{U}_m^\star\right) \end{pmatrix}$$

and

$$\Delta \mathbf{X} := \left(\Delta X_1', ..., \Delta X_T'\right)', \quad \mathbf{U} := \left(U_1', ..., U_T'\right)'$$

and

$$\mathbf{U}_i^\star := \left(p_{j1}^\star U_1', ..., p_{jT}^\star U_T'\right)' \quad \textit{for} \quad j \in \mathbb{M}.$$

Proof. Observe initially that the first order derivative of the likelihood contribution with respect to Φ can be written as,

$$\mathrm{d}L_T\left(\theta; \text{vec}\left(\mathrm{d}\Phi\right)\right) = \left(\mathbb{S}_{\Phi_1,T}, ..., \mathbb{S}_{\Phi_m,T}\right) \text{vec}\left(\mathrm{d}\Phi\right) = \mathbb{S}_{\Phi,T} \text{vec}\left(\mathrm{d}\Phi\right).$$

and by Lemma 1.31 in chapter one. Using matrix notation, we get

$$\begin{aligned} \mathrm{d}L_T\left(\theta; \text{vec}\left(\mathrm{d}\Phi\right)\right) &= \text{vec}\left(\mathbf{U}_1^{\star\prime} \Delta \mathbf{X} \Omega_1^{-1}\right)' \text{vec}\left(\mathrm{d}\Phi_1\right) \\ &\quad + \cdots + \text{vec}\left(\mathbf{U}_1^{\star\prime} \Delta \mathbf{X} \Omega_m^{-1}\right)' \text{vec}\left(\mathrm{d}\Phi_m\right) \\ &\quad - \text{vec}\left(\Phi_1\right)' \left(\Omega_1^{-1} \otimes \mathbf{U}_1^{\star\prime}\mathbf{U}\right) \text{vec}\left(\mathrm{d}\Phi_1\right) \\ &\quad - \cdots - \text{vec}\left(\Phi_m\right)' \left(\Omega_m^{-1} \otimes \mathbf{U}_m^{\star\prime}\mathbf{U}\right) \text{vec}\left(\mathrm{d}\Phi_m\right) \\ &= \mathbf{Y}' \text{vec}\left(\mathrm{d}\Phi\right) - \text{vec}\left(\Phi\right)' \mathbf{M}' \text{vec}\left(\mathrm{d}\Phi\right) \\ &= \left(\mathbf{Y}' - \text{vec}\left(\Phi\right)' \mathbf{M}'\right) \text{vec}\left(\mathrm{d}\Phi\right) \end{aligned}$$

In terms of the restricted estimator, we get

$$\mathrm{d}L_t\left(\theta; \text{vec}\left(\mathrm{d}\varpi\right)\right) = \left(\mathbf{Y}' H_\Phi - \left(H_\Phi \varpi + h_\Phi\right)' \mathbf{M}' H_\Phi\right) \mathrm{d}\varpi.$$

An estimator can thus be found as by solving for ϖ in

$$\mathbf{Y}' H_\Phi - h_\Phi' \mathbf{M}' H_\Phi - \hat{\varpi}' H_\Phi' \mathbf{M}' H_\Phi = 0$$

which gives

$$\hat{\varpi} = \left(H_\Phi' \mathbf{M} H_\Phi\right)^{-1} \left(\mathbf{Y} H_\Phi - h_\Phi' \mathbf{M} H_\Phi\right)$$

□

Lemma 2.13. *Define ω as the vector containing the freely varying elements of Ω such that*

$$\mathrm{vec}\left(\Omega\right) = H_\Omega \omega + h_\Omega.$$

Then for fixed filtered probabilities, $p_{it}^\star$, an estimator for ω is given by

$$\hat{\omega} = \left(H_\Omega' \mathbf{P} H_\Omega\right)^{-1} H_\Omega' \left(\mathrm{vec}\left(\mathbf{E}\right) - \mathbf{P} h_\Omega\right) \tag{2.13}$$

where

$$\mathbf{P} = \sum_{t=1}^{T} \mathrm{diag}\left(p_{1t}^\star \mathcal{I}_{n^2}, ..., p_{mt}^\star \mathcal{I}_{n^2}\right) \quad \textit{and} \quad \hat{\mathbf{E}} = \sum_{t=1}^{T} \left(p_{1t}^\star \hat{\varepsilon}_{1t} \hat{\varepsilon}_{1t}', \ldots, p_{mt}^\star \hat{\varepsilon}_{mt} \hat{\varepsilon}_{mt}'\right).$$

Proof. Consider the first order derivative of the log-likelihood function in direction Ω,

$$\mathrm{d}L_T\left(\theta; \mathrm{d}\Omega\right) = \left(\mathbb{S}_{\Omega_1}, \ldots \mathbb{S}_{\Omega_m}\right) \mathrm{vec}\left(\mathrm{d}\Omega\right) = \mathbb{S}_\Omega \mathrm{vec}\left(\mathrm{d}\Omega\right).$$

By the results of Lemma 1.31 from chapter one, it holds that

$$\begin{aligned}
\mathbb{S}_\Omega \mathrm{vec}\left(\mathrm{d}\Omega\right) &= -\sum_{t=1} \frac{1}{2} tr\left\{p_{1t}^\star \Omega_1^{-1} \mathrm{d}\Omega_1\right\} - ... - \sum_{t=1} \frac{1}{2} tr\left\{p_{mt}^\star \Omega_m^{-1} \mathrm{d}\Omega_m\right\} \\
&\quad + \sum_{t=1} \frac{1}{2} tr\left\{\Omega_1^{-1} p_{1,t}^\star \varepsilon_{1t} \varepsilon_{1t}' \Omega_1^{-1} \mathrm{d}\Omega_1\right\} \\
&\quad + \cdots + \sum_{t=1} \frac{1}{2} tr\left\{\Omega_m^{-1} p_{mt}^\star \varepsilon_{mt} \varepsilon_{mt}' \Omega_m^{-1} \mathrm{d}\Omega_m\right\} \\
&= \mathrm{vec}\left(\mathrm{d}\Omega\right)' S_2 \mathrm{vec}\left(\mathbf{E}\right) - \mathrm{vec}\left(\mathrm{d}\Omega\right)' S_1 \mathrm{vec}\left(\Omega\right)
\end{aligned}$$

where

$$S_1 = \mathbf{P} S_2, \text{ and } S_2 = \mathrm{diag}\left(\left(\Omega_1^{-1} \otimes \Omega_1^{-1}\right), ..., \left(\Omega_m^{-1} \otimes \Omega_m^{-1}\right)\right).$$

Now, an estimator for ω can be found by solving

$$-H_\Omega' S_1 \left(H_\Omega \hat{\omega} + h_\Omega\right) + H_\Omega' S_2 \mathrm{vec}\left(\mathbf{E}\right) = 0$$

which yields

$$\begin{aligned}
\hat{\omega} &= \left(H_\Omega' S_1 H_\Omega\right)^{-1} H_\Omega' \left(S_2 \mathrm{vec}\left(\mathbf{E}\right) - S_1 h_\Omega\right) \\
&= \left(H_\Omega' S_2 \mathbf{P} H_\Omega\right)^{-1} H_\Omega' S_2 \left(\mathrm{vec}\left(\mathbf{E}\right) - \mathbf{P} h_\Omega\right) \\
&= \left(H_\Omega' \mathbf{P} H_\Omega\right)^{-1} H_\Omega' \left(\mathrm{vec}\left(\mathbf{E}\right) - \mathbf{P} h_\Omega\right).
\end{aligned}$$

□

2.B Proof of Lemma 2.7

We mimic the proof given in proof of Theorem 4.9 (part i) in Kristensen and Rahbek (2013), adapting their arguments to the present framework and notation. Observe initially that

$$\tilde{\theta} - \theta_0 = H_\theta \tilde{\upsilon} + g - H_\theta \upsilon_0 - g = H_\theta \left(\tilde{\upsilon} - \upsilon_0 \right)$$

and that the derivative in the direction of the freely varying parameters is given by

$$\mathrm{d}\tilde{\theta} = \mathrm{d}\theta \left(\upsilon; \mathrm{d}\upsilon \right) = H_\theta \mathrm{d}\upsilon.$$

Consider the a first order Taylor series expansion around the true value, θ_0, of the first derivative of the log-likelihood function evaluated under the imposed restrictions,

$$\begin{aligned} 0 &= \mathrm{d}L_T \left(\theta_0; \mathrm{d}\tilde{\theta} \right) + \mathrm{d}^2 L_T \left(\theta_0; \mathrm{d}\tilde{\theta}, \tilde{\theta} - \theta_0 \right) + r \left(\theta_0 \right) \\ &= \mathbb{S}_T \left(\theta_0 \right)' \mathrm{d}\tilde{\theta} + \left(\tilde{\theta} - \theta_0 \right)' \mathbb{H}_T \left(\theta_0 \right) \mathrm{d}\theta + r \left(\theta_0 \right) \end{aligned}$$

where $r \left(\theta_0 \right)$ is a remainder with higher order terms from the Taylor series expansion. We then look at

$$\begin{aligned} 0 &= \mathrm{d}L_T \left(\theta_0; W_T^{-\frac{1}{2}} \mathrm{d}\tilde{\theta} \right) + \mathrm{d}^2 L_T \left(\theta_0; W_T^{-\frac{1}{2}} \mathrm{d}\tilde{\theta}, \left(\tilde{\theta} - \theta_0 \right) \right) + r \left(\theta_0 \right) \\ &= \mathbb{S}_T \left(\theta_0 \right)' W_T^{-\frac{1}{2}} \mathrm{d}\tilde{\theta} + \left(\tilde{\theta} - \theta_0 \right)' \mathbb{H}_T \left(\theta_0 \right) W_T^{-\frac{1}{2}} \mathrm{d}\tilde{\theta} + r \left(\theta_0 \right) \\ &= \mathbb{S}_T \left(\theta_0 \right)' W_T^{-\frac{1}{2}} \mathrm{d}\tilde{\theta} + \left(\tilde{\theta} - \theta_0 \right)' W_T^{\frac{1}{2}} W_T^{-\frac{1}{2}} \mathbb{H}_T \left(\theta_0 \right) W_T^{-\frac{1}{2}} \mathrm{d}\tilde{\theta} + r \left(\theta_0 \right) \\ &:= \mathbb{S}_{T,W} \left(\theta_0 \right)' H_\theta \mathrm{d}\upsilon + \left(\tilde{\upsilon} - \upsilon_0 \right)' H_\theta' W_T^{\frac{1}{2}} \mathbb{H}_{T,W} \left(\theta_0 \right) H_\theta \mathrm{d}\upsilon + r \left(\theta_0 \right) \\ &= \mathbb{S}_{T,W} \left(\theta_0 \right)' H_\theta \mathrm{d}\upsilon + \left(\tilde{\upsilon} - \upsilon_0 \right)' U_T^{1/2} H_\theta' \mathbb{H}_{T,W} \left(\theta_0 \right) H_\theta \mathrm{d}\upsilon + r \left(\theta_0 \right) \end{aligned}$$

where we have used that one can define a matrix U_T, such that $W_T^{\frac{1}{2}} \left(\tilde{\theta} - \theta_0 \right) = H_\theta U_T^{1/2} \left(\tilde{\upsilon} - \upsilon_0 \right)$ and where, to simplify notation in the following, we have used the definitions, $\mathbb{S}_{T,W} \left(\cdot \right) := W_T^{-\frac{1}{2}} \mathbb{S}_T \left(\cdot \right)$ and $\mathbb{H}_{T,W} \left(\cdot \right) := W_T^{-\frac{1}{2}} \mathbb{H}_T \left(\cdot \right) W_T^{-\frac{1}{2}}$. By rearranging, we obtain

$$W_T^{\frac{1}{2}} \left(\tilde{\theta} - \theta_0 \right) = H_\theta U_T^{1/2} \left(\tilde{\upsilon} - \upsilon_0 \right) = -H_\theta \left(H_\theta' \mathbb{H}_{T,W} \left(\theta_0 \right) H_\theta \right)^{-1} H_\theta' \mathbb{S}_{T,W} \left(\theta_0 \right) + o_p \left(1 \right),$$

where it has been used that $r \left(\theta_0 \right) = o_p \left(1 \right)$ by Lemmas (1.12)-(1.26) from chapter one. Hence, it holds that

$$W_T^{\frac{1}{2}} \left(\tilde{\theta} - \theta_0 \right) = H_\theta U_T^{1/2} \left(\tilde{\upsilon} - \upsilon_0 \right) \xrightarrow{w} -H_\theta \left(H_\theta' \mathbb{H} \left(\theta_0 \right) H_\theta \right)^{-1} H_\theta' \mathbb{S} \left(\theta_0 \right)',$$

where $\mathbb{H} \left(\theta_0 \right) := \mathbb{H}_{\infty,W} \left(\theta_0 \right)$ and $\mathbb{S} \left(\theta_0 \right) := \mathbb{S}_{\infty,W} \left(\theta_0 \right)$ are the limits of the T-normalized Hessian and the score evaluated at the true values. Using a second order Taylor's expansion on the likelihood ratio test in (2.5), we get

$$LR_T \left(\hat{\theta}, \tilde{\theta} \right) = -2 \left(L_T \left(\hat{\theta} \right) - L_T \left(\hat{\theta} \right) - \mathbb{S}_T \left(\hat{\theta} \right) \left(\hat{\theta} - \tilde{\theta} \right) - \frac{1}{2} \left(\hat{\theta} - \tilde{\theta} \right)' \mathbb{H}_T \left(\theta^\star \right) \left(\hat{\theta} - \tilde{\theta} \right) \right)$$

$$
\begin{aligned}
&= 2\mathbb{S}_T\left(\hat{\theta}\right)\left(\hat{\theta}-\tilde{\theta}\right)+\left(\hat{\theta}-\tilde{\theta}\right)'\mathbb{H}_T\left(\theta^\star\right)\left(\hat{\theta}-\tilde{\theta}\right)\\
&= \left(\hat{\theta}-\tilde{\theta}\right)'\mathbb{H}_T\left(\theta^\star\right)\left(\hat{\theta}-\tilde{\theta}\right)
\end{aligned}
$$

where it holds for $\theta^\star$, that $\left|\hat{\theta}-\theta^\star\right| \leq \left|\hat{\theta}-\tilde{\theta}\right|$, and since $\hat{\theta}$ maximizes $L_T\left(\hat{\theta}\right)$, we have $\mathbb{S}_T\left(\hat{\theta}\right)=0$. Next, observe that

$$
\begin{aligned}
-W_T^{\frac{1}{2}}\left(\hat{\theta}-\tilde{\theta}\right) &= -W_T^{\frac{1}{2}}\left(\hat{\theta}-\theta_0\right)+W_T^{\frac{1}{2}}\left(\tilde{\theta}-\theta_0\right)\\
&= \left(\mathbb{H}_{T,W}\left(\theta_0\right)\right)^{-1}\mathbb{S}_{T,W}\left(\theta_0\right)\\
&\quad -H_\theta\left(H_\theta'\mathbb{H}_{T,W}\left(\theta_0\right)H_\theta\right)^{-1}H_\theta'S_{T,W}\left(\theta_0\right)+o_p\left(1\right)\\
&\xrightarrow{w} \mathbb{H}\left(\theta_0\right)^{-1}\mathbb{S}\left(\theta_0\right)-H_\theta\left(H_\theta'\mathbb{H}\left(\theta_0\right)H_\theta\right)^{-1}H_\theta'\mathbb{S}\left(\theta_0\right)\\
&:= \mathbb{P}\left(\theta_0\right)\mathbb{S}\left(\theta_0\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{P}\left(\theta_0\right) &= \mathbb{H}\left(\theta_0\right)^{-1}-H_\theta\left(H_\theta'\mathbb{H}\left(\theta_0\right)H_\theta\right)^{-1}H_\theta'\\
&= \mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\left(H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\right)^{-1}H_{\theta,\perp}\mathbb{H}\left(\theta_0\right)^{-1}
\end{aligned}
$$

such that

$$
\begin{aligned}
LR_T\left(\hat{\theta},\tilde{\theta}\right) &\xrightarrow{w} \mathbb{S}\left(\theta_0\right)'\mathbb{P}\left(\theta_0\right)'\mathbb{H}\left(\theta_0\right)\mathbb{P}\left(\theta_0\right)\mathbb{S}\left(\theta_0\right)\\
&= \mathbb{S}\left(\theta_0\right)'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\left(H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\right)^{-1}H_{\theta,\perp}'\\
&\quad \times\mathbb{H}\left(\theta_0\right)^{-1}\mathbb{H}\left(\theta_0\right)\mathbb{H}\left(\theta_0\right)^{-1}\\
&\quad \times H_{\theta,\perp}\left(H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\right)^{-1}H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}\mathbb{S}\left(\theta_0\right)\\
&= \mathbb{S}\left(\theta_0\right)'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\left(H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\right)^{-1}\\
&\quad \times H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\left(H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\right)^{-1}H_{\theta,\perp}'\\
&\quad \times\left(\mathbb{H}\left(\theta_0\right)\right)^{-1}\mathbb{S}\left(\theta_0\right)\\
&= \mathbb{S}\left(\theta_0\right)'\left(\mathbb{H}\left(\theta_0\right)\right)^{-1}H_{\theta,\perp}\left(H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\right)^{-1}H_{\theta,\perp}'\times\\
&\quad \left(\mathbb{H}\left(\theta_0\right)\right)^{-1}\mathbb{S}\left(\theta_0\right)\\
&= \mathbb{V}\left(\theta_0\right)'\mathbb{V}\left(\theta_0\right)
\end{aligned}
$$

where

$$
\mathbb{V}\left(\theta_0\right)=\left(H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}H_{\theta,\perp}\right)^{-1/2}H_{\theta,\perp}'\mathbb{H}\left(\theta_0\right)^{-1}\mathbb{S}\left(\theta_0\right).
$$

This completes the proof. □

3 Non-linear cointegration analysis of crude oil benchmarks

We analyze the dynamic behavior between two main crude oil benchmarks, the American West Texas Intermediate (WTI) and the European Brent. The series are analyzed using the ACR cointegrated model discussed in chapters one and two. The cointegration relations are considered unknown and are estimated jointly with the remaining parameters. Non-linearities are allowed on the error-correction parameters, short-run parameters and the parameters in the residual covariances. We moreover allow the constant in the cointegration relations to enter the regime switching probability and show that the asymptotic theory given in chapter one goes through with a few modifications. Having jointly estimated all parameters including the cointegration vectors, the asymptotic inference is non-standard and the bootstrap procedure from chapter two is invoked to test hypothesis. We observe non-linearities linked to a recent decoupling of the WTI crude oil price from the international market and find that historically, the Brent crude price has been the only series reacting to disequilibria.

3.1 Introduction

A vast number of crude oils are produced across the world and their corresponding individual prices on the international commodity markets will normally depends quality, geographical placement, logistics and prices on other crudes. In particular the American West Texas Intermediate (WTI) and the European Brent play central roles as benchmarks on the international markets. Despite the geographical distance and the slight difference in quality, these commodities are very close substitutes and arbitrage behavior of economic agents should ensure a stable price relationships, motivating the existence of cointegration. However, in addition to the factors mentioned above, the price of crude oil in different regions may also depend on many political instabilities in regions with high crude oil production, changes in the consumption patterns following introductions of new technologies, financial or economic crises and so on, see e.g. Fattouh (2010). Therefore, the arbitrage among the spreads can be disrupted by numerous factors resulting in dynamics that are not necessarily well captured by linear cointegration models.

In this chapter, we analyze monthly observations of the WTI and Brent price series using the ACR model. The analysis is related to studies such as Hammoudeh et al. (2008), Fattouh (2010), Mann (2012), Ghoshray and Trifonova (2014) and Liao et al. (2014), who apply different types of threshold autoregressive analysis to these same and other series of crude oil prices, though for different periods and data frequencies.

This analysis differs from earlier studies on a number of important points. First, Fattouh (2010) models the spreads directly as univariate self-exiting threshold autoregressions and thus imposes the cointegration relation prior to the analysis and is unable to assess more elaborate dynamics such as potential weak exogeneity of some of the analyzed crudes. Second, Hammoudeh et al. (2008) and Ghoshray and Trifonova (2014) uses a procedure advocated by Enders and Siklos (2001) where an Engle-Granger type two-step approach is applied to extract cointegration relations, which are then modeled within a so-called momentum threshold autoregressive framework, where the regime switching depends on the size and direction of change in the deviations from long-run equilibrium. Such an approach is invalid in the ACR framework, since one cannot ignore estimation of the cointegration vectors when conducting inference on the other parameters. We thus consider joint maximum likelihood estimation of the cointegration parameters along with the remaining parameters. This means that simulation-based methods are required for conducting inference and we apply the bootstrap principle discussed in chapter two. Finally, the method used in this chapter allows for changing short run parameters and changing error covariance, both features that none of the previous papers include. Many other approaches have been considered in the literature studying the dynamic behavior of crude oil prices and a comprehensive literature review is given in Ghoshray and Trifonova (2014).

The data displays a curious decoupling of the series that happened in 2011 and has persisted until now. This change in the behavior is picked up by the ACR model as a regime with weak error correction mechanisms. Excluding this part of the data, we do not find strong support for non-linearities since a linear model cannot be rejected in favor of the ACR model. We also find that the non-linearities have played more important roles in periods of financial turmoil. Finally, we observe that historically, the Brent crude has been reacting strongly to disequilibria, while the WTI has been weakly exogenous.

This paper is structured as follows. Section 3.2 discusses a simple economic motivation for the existence of (potentially non-linear) cointegration between the crude oil benchmarks. Section 3.3 presents the econometric model and discusses consequences for the asymptotic theory when letting the constant in the cointegration vectors enter the regime switching probabilities. Section 3.4 presents the data and gives some discussions of key periods of interest. Section 3.5 gives the results from estimation of the ACR and linear cointegrated VAR models. Finally, section 5 concludes.

3.2 Motivation

Since these crudes are close substitutes, arbitrage behavior should ensure that the spread corrected for differences in quality and transaction costs is stable. A simple encompassing framework for this pricing mechanism is the cost of carry model discussed by Alizadeh and Nomikos (2004) and Fattouh (2010). The long run relationship between the crude oil prices is given by the equation,

$$P_{Brent,t} = P_{WTI,t} - C_{Brent} - D \tag{3.1}$$

where $P_{Brent,t}$ and $P_{WTI,t}$ are specific log transformed crude oil prices and D is the premium arising from differences in the quality of the crude oils. The term, C_{Brent}, is the cost of carry

including transportation, insurance, custom duty, pipeline tariffs etc. Econometrically, one cannot distinguish between C_{Brent} and D in the framework that is applied in the following, where the constant in the cointegration relations will capture both C_{Brent} and D.

The investigated crude oil streams differ on geographical location and slightly on quality. The WTI is based at Cushing Oklahoma, US and the Brent has its base in the North Sea, see Fattouh (2011). The WTI and Brent crudes are so-called light/sweet crudes. The abbreviations "light" or "heavy" refers to the density of the crude oil. Generally, crude oils with lower density will yield higher proportions of final petroleum products than heavier crudes. The "sweet" and "sour" abbreviations refer to the amount of sulfur contained in the crude, where sweet crudes contain less sulfur than sour crudes. The WTI is slightly lighter and sweeter than the Brent and the Brent crude is subject to larger transportation costs, meaning that the WTI should be traded at a premium over the Brent, see Fattouh (2010).

3.3 Econometric model

Suppose that the crude oil price series are given by $X_t = (P_{Brent,t} : P_{WTI,t})'$, a process generated by the equations,

$$\begin{aligned}\Delta X_t &= \sum_{j=\{1,2\}} \mathbf{1}\{s_t = j\}\left(\alpha_j \beta^{\star\prime} X^{\star}_{t-1} + \Gamma_j \Delta X_{t-1} + V_j \epsilon_t\right) \\ &= \sum_{j=\{1,2\}} \mathbf{1}\{s_t = j\}\left(\Phi_j U_{t-1} + V_j \epsilon_t\right)\end{aligned}$$

using the definitions from chapter two. Identification is introduced by setting $\beta_1 = 1$ such that $\beta^{\star} = (1 : \varphi_{\beta} : \varphi_{\beta_D})'$ and the freely varying parameter is given by $\varphi = (\varphi_{\beta} : \varphi_{\beta_D})'$. The probability of switching as a function of the probability parameters, $\gamma = (\Lambda : \kappa)'$, and the cointegration relations, $u_t = \psi' U_t$, is given by

$$p_{1t} = P\left(s_t = 1 \mid u_{t-1}; \gamma\right) = 1 - \exp\left(-u'_{t-1} \Lambda u_{t-1} g\left(u_{t-1}; \kappa\right)\right) = 1 - p_{2t} \tag{3.2}$$

with

$$g\left(u_{t-1}; \kappa\right) = 0.5 + \left(1 + \exp\left(-\kappa' u_{t-1}\right)\right)$$

This specification allows for the parameters to change when large deviations from the equilibrium occur, with deviations measured by the squared cointegration relations. In addition, asymmetry is allowed through $g\left(u_{t-1}; \kappa\right)$ such that the impact of a deviation from equilibrium on the switching probabilities can differ depending on the sign of the deviation.

Observe that we have allowed u_{t-1} to enter the switching probabilities and that this has some consequences for the theory developed in chapter one. First, consider the stability results provided in section 1.4 and recall that the process written on companion form was given by

$$\begin{aligned}\mathbb{Y}_t &= \sum_{j \in \mathbb{M}} \mathbf{1}\{s_t = j\}\left(\mathbb{A}_j \mathbb{Y}_{t-1} + \mathbb{U}_{jt}\right) &(3.3)\\ &= \sum_{j \in \mathbb{M}} \mathbf{1}\{s_t = j\}\left(\mathbb{A}_j \mathbb{Y}^{\star}_{t-1} + \mathbb{E}_{jt}\right) = \mathbb{A}_t \mathbb{Y}^{\star}_{t-1} + \mathbb{E}_t &(3.4)\end{aligned}$$

where $\mathbb{E}_{jt} := JV_j\epsilon_t$, $\mathbb{Y}_t^\star := \left(Y_t^{\star\prime} : Y_{t-1}' : \cdots : Y_{t-k+1}'\right)'$, $Y_t^\star := (X_t^{\star\prime}\beta^\star : \Delta X_t'\beta_\perp)'$, $J := (\mathcal{I}_n : 0)'$, while $\mathbb{Y}_t$, $\mathbb{U}_j$ and $\mathbb{A}_j$ were given in chapter one. Previously, the switching was chosen to depend on $z_t = \psi' Z_t = \eta'\mathbb{Y}_t$, whereas now, the switching will depend on $u_t = \psi' U_t = \eta'\mathbb{Y}_t^\star$. Observe that the selection matrix, η, is the same in the two specifications and hence, it is clear that the stability results that applied to (3.4) will also apply to (3.3). The more important difference comes from added contributions to the likelihood function. When including β_D in the probability of switching, the score, hessian and third order derivatives presented chapter one, no longer applies in the direction of β_D. In appendix 3.A we present the needed modifications of the score, hessian and third order derivatives, and verify convergence results similar to those of chapter one for these modified terms.

The parameters are estimated by maximum likelihood and the Gaussian log-likelihood as a function of the parameter vector,

$$\theta := \left(\operatorname{vec}\left(\beta^{\star\prime}\right)' : \operatorname{vec}\left(\Phi_1\right)' : \operatorname{vec}\left(\Phi_1\right)' : \operatorname{vech}\left(\Omega_1\right)' : \operatorname{vech}\left(\Omega_2\right)' : \gamma'\right)',$$

is given by

$$L_T\left(\theta\right) = \sum_{t=1}^{T} \ell_t\left(\theta\right) = \sum_{t=1}^{T}\sum_{j\in\mathbb{M}} \log\left(p_{jt}\phi_{jt}\right) \tag{3.5}$$

where p_{jt} given in (3.2) which depends on past values of the process and with ϕ_{jt} given by

$$\log\phi_{jt} \quad = \quad -\frac{1}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left(\Omega_j\right) - \frac{1}{2}\varepsilon_{jt}'\Omega_j^{-1}\varepsilon_{jt}$$

where

$$\varepsilon_{jt} := \Delta X_t - \alpha_j\beta^{\star\prime}X_{t-1}^\star - \Gamma_j\Delta\mathbb{X}_{t-1} = \Delta X_t - \Phi_j Z_{t-1}.$$

The so-called filtered probabilities are given by

$$P\left(s_t = j \mid Z_t\right) = p_{jt}^\star = \frac{p_{jt}\phi_{jt}}{\sum_{j\in\mathbb{M}} p_{jt}\phi_{jt}}. \tag{3.6}$$

In the discussions of the model fit, these are used as estimators for the unobserved regime switching variable, s_t.

The parameters of the different models are estimated using the EM algorithm discussed in chapter two. To find initial values for the algorithm, a grid search over the parameters that enter the switching probability, i.e. φ_β, φ_{β_D}, Λ and κ is done. The remaining parameters, Φ_1, Φ_2, $\operatorname{vech}\left(\Omega_1\right)$ and $\operatorname{vech}\left(\Omega_2\right)$ are estimated for each set of φ_β, φ_{β_D}, Λ and κ, using the EM algorithm. This method is the four parameter equivalent of the profile likelihood approach suggested by Bec et al. (2008) for the case of a specification with two probability parameters and fixed cointegration relations.

3.4 Data

The data is obtained from Bloomberg at a monthly frequency from January 1987 to Februrary 2014, where each observation corresponds to the closing price observed the last trading day of

the month. The period spans the era of the marked based pricing system, see Fattouh (2011). We model the logarithmically transformed prices. These series, the first differences and the spread are displayed in figure 3.1.

As mentioned in the introduction, the WTI crude should be trading at a premium over the Brent. This has been true for most of the observed period, except for very short periods of reversal and from about 2011, where a clear departure from this rule is observed. This decoupling has arisen as a result of a drop in the price of WTI relative to other crudes. This is not seen from graphs in figure 3.1, but is evident when comparing with other crude streams such as the Dubai-Fateh (see figure 3.7 in Appendix 3.C). A common explanation for this anomaly is a combination of increased crude oil and gas production in North America along with logistical limitations in Cushing, Oklahoma, the delivery point for the WTI crudes, see e.g. Mann (2012) and Fattouh (2011)[1].

Other that the post 2010 decoupling, there are a couple of other key periods that need to be mentioned. First, a sharp increase in both prices and their first differences is seen around 1990, corresponding to the commencement of the (first) Iraqui war. Around 1997-1999 the oil prices are seen to plunge and the first differences are again large in absolute terms compared to the average. This period corresponds to the Asian financial crisis. Finally, in 2008, the breakout of the global financial crisis spurs similar behavior in the series.

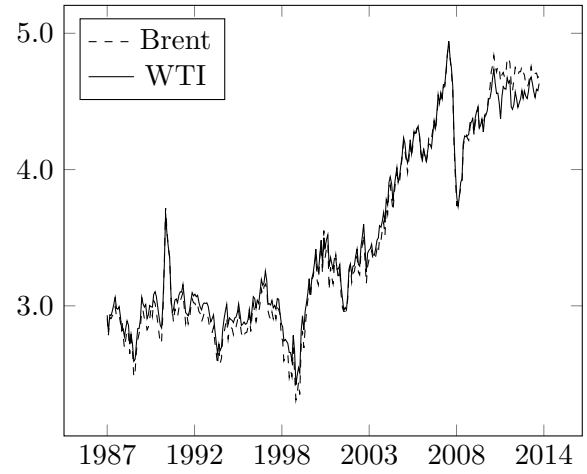
3.5 Results

We estimate the model both on the full sample and on a smaller sample that ends in December 31st 2010. It becomes evident that whether or not this period is included has important consequences for the results. In particular, the non-linearities seem more important in the full sample since they allow the model more flexibility to accommodate this change in the behavior of the series.

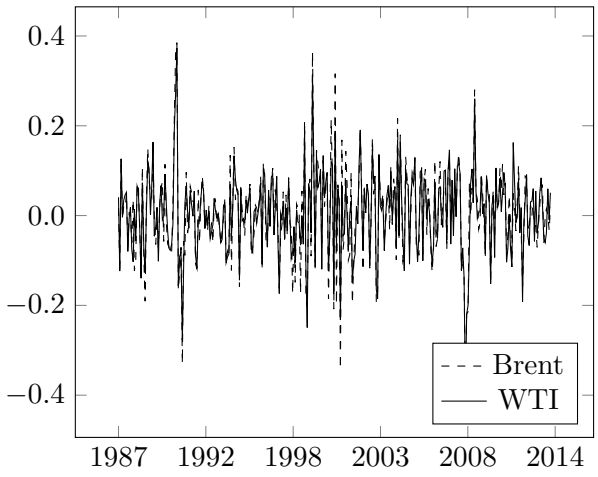
The maximum likelihood estimates of table 3.1 show that the estimated cointegration relations resembles the spread of the series corrected for a premium as discussed in section 3.2. Second, one observes that the error correction coefficients in the outer regime are much smaller than in the inner regime. This is also seen as the spectral radius in this regime is around 0.95, while that of the inner regime is 0.65. The system is stationary since the joint spectral radius is less than unity. Observe that this result is to a large extend driven by the period post 2010, where the discussed decoupling has taken place and the outer regime is primarily identified by observations in this period. That is clear from figure 3.2(c) where the filtered probabilities are depicted. Figure 3.2(e) shows that the estimated level of asymmetry in the regime switching probabilities has the effect that positive changes will result in a regime change faster than negative changes would. Finally, observe that the variance of the error terms seem to be of an order two larger in regime one that in regime two. The estimates are quite different when one considers the period excluding observations at the end of the sample. Then the structure of the regimes is reversed and the outer regime is more stable that the inner regime, though not very much so. The estimated asymmetry of the switching probabilities is also less pronounced.

[1]See also U.S. Energy Information Administration, Today in Energy, June 28 2013, "Price difference between Brent and WTI crude oil narrowing"

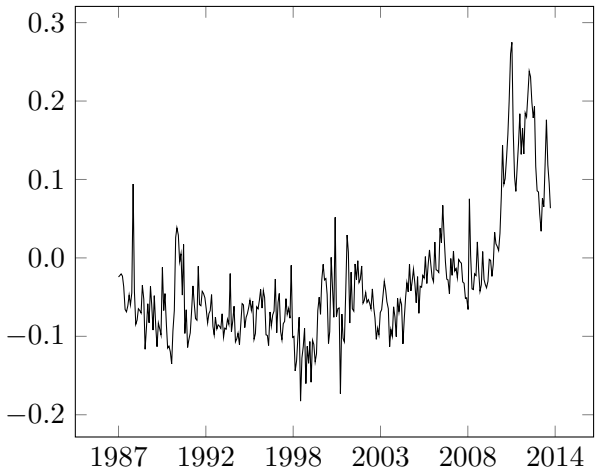
Figure 3.1: Graphical analysis of the series



(a) Levels, log-prices



(b) First Differences of log prices



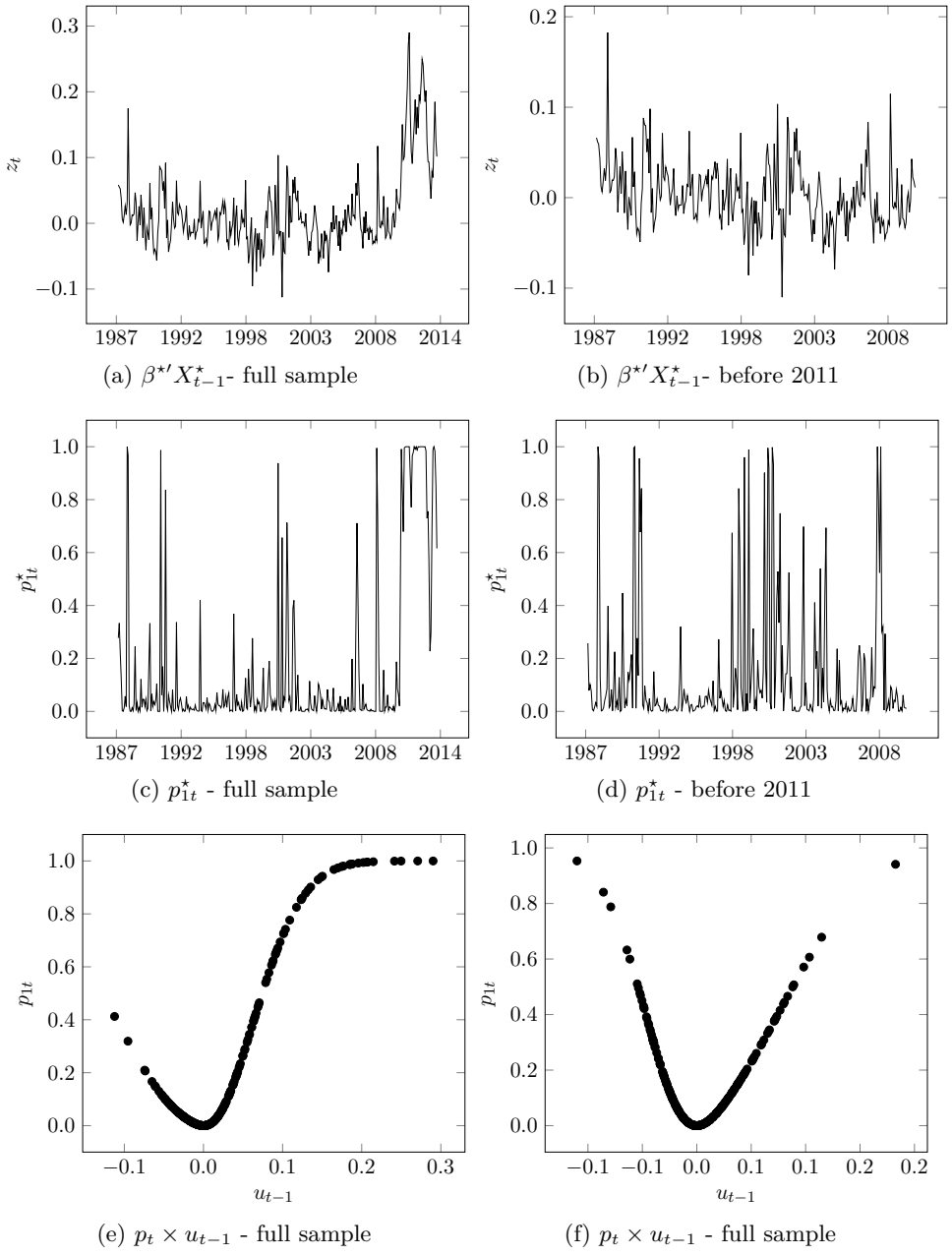
(c) The spread between Brent and WTI

Table 3.1: Maximum likelihood estimates

	Full sample		Before 2011	
	ACR	Linear	ACR	Linear
$\hat{\varphi}_1$	-1.042	-1.090	-1.054	-1.043
$\hat{\varphi}_2$	0.197	0.347	0.239	0.201
$\alpha_{1,1}$	-0.044	-0.109	-0.143	-0.449
$\alpha_{2,1}$	0.011	0.078	0.393	0.033
$\Gamma_{11,1}$	-0.424	-0.117	-0.730	-0.028
$\Gamma_{21,1}$	-0.055	0.092	-0.549	0.112
$\Gamma_{12,1}$	0.172	0.228	1.264	0.157
$\Gamma_{22,1}$	-0.096	0.048	1.102	0.029
$\Omega_{11,1}$	0.005	0.010	0.032	0.010
$\Omega_{12,1}$	0.003	0.009	0.025	0.009
$\Omega_{21,1}$	0.005	0.008	0.021	0.009
$\alpha_{1,2}$	-0.242	-	-0.351	-
$\alpha_{2,2}$	0.169	-	0.037	-
$\Gamma_{11,2}$	0.025	-	0.007	-
$\Gamma_{21,2}$	0.104	-	0.142	-
$\Gamma_{12,2}$	0.228	-	-0.012	-
$\Gamma_{22,2}$	0.097	-	-0.139	-
$\Omega_{11,2}$	0.011	-	0.007	-
$\Omega_{12,2}$	0.009	-	0.006	-
$\Omega_{21,2}$	0.008	-	0.006	-
Λ	84.160	-	169.54	-
κ	67.871	-	-41.51	-
Spectral radius regime one	0.9595	0.8412	0.5926	0.6107
Spectral radius regime two	0.6500	-	0.6785	-
Joint spectral radius	[0.9595 : 0.9695]	-	[0.6785 : 0.689]	-
Log-likelihood value	986.238		890.698	
T	326	326	288	288

The joint spectral radius is calculated using the jsr_louvain toolbox by Raphael Jungers, see Jungers (2009).

Figure 3.2: Graphical analysis of the estimated ACR models



(a) $\beta^{\star\prime} X^\star_{t-1}$- full sample

(b) $\beta^{\star\prime} X^\star_{t-1}$- before 2011

(c) $p^\star_{1t}$ - full sample

(d) $p^\star_{1t}$ - before 2011

(e) $p_t \times u_{t-1}$ - full sample

(f) $p_t \times u_{t-1}$ - full sample

The cointegration relation is again seen to be close to the spread corrected for the quality and cost related price premium. For both modeled periods, and prior to the 2011 decoupling, regime one is predominant mostly in the crisis periods indicating that the non-linearities in the cut-off sample are most prominent when the markets are less stable. We also estimate linear cointegrated VAR models for comparisons. They display clear stationarity and in particular indicate that Brent crude seems to be error-correcting while the WTI does not. Standard errors are not reported in table 3.1 since the distributions of t-statistics are non-standard and nuisance parameter dependent in the ACR framework. Instead, table 3.3 tests a number of hypothesis of interest using bootstrap methods to simulate the corresponding distributions.

Misspecification tests are not yet developed for the ACR cointegrated model and instead we look at some properties of the estimated residuals to get indications of possible misspecification. The residuals in the ACR model are unobserved due to the unobserved regime switching, and as a substitute, we use the estimator proposed in chapter two, section 2.4. We display some graphics of these residuals illustrating univariate properties and we calculate portmanteau statistics for left-over autocorrelation, see e.g. Lütkepohl (2005) and Juselius (2006). The graphs are given in appendix 3.B and display very few signs of autocorrelation in the residuals. When looking at the ACF functions for the squared residuals, some indications of left-over heteroskedasticity are seen. We observe from table 3.2 that the portmanteau statistics reject the null of no left-over autocorrelation in for the models estimated on the full sample and thus picks up some autocorrelation that is not easily seen in the ACF plots. Hence, the conclusions from the tests below should be drawn with some caution for the full sample. They do not reject for the model estimated on the small sample, indicating that the left-over autocorrelation reflect the models difficulties in fitting the post 2011 decoupling. The portmanteau statistics for no left-over autocorrelation in the squared errors also reject, giving indications of ARCH effects as was also seen in the ACF functions. In this respect, observe in particular that the ACF for the first couple of lags of squared residuals are smaller in the ACR models than in the linear models. Likewise, the graphs of the errors display much less volatility clustering than in the linear case. This illustrates the models ability to accommodate some conditional heteroskedasticity through the switching error covariance. We handle the remaining left over heteroskedasticity by supplementing with wild bootstrap p-values when evaluating test statistics in the next (see chapter two, remark 2.10, for a discussion on the wild bootstrap modification of the proposed algorithm).

The results from the full sample show that the bootstrap test rejects the null hypothesis that the spread is the cointegration relation, though the estimate of φ_b seems close to minus one. That illustrates the super consistency of the cointegration relations discussed in chapter one, which has as a consequence that tests on the cointegration parameters are very powerful. The hypothesis that premium is zero is likewise very clearly rejected. Next it is seen that tests for asymmetry in the regime switching and for weak exogeneity are not significant. The latter indicates that the structure of the error correction mechanisms in this model is poorly identified. Observe that when testing for linearity of the error covariance, the bootstrap test rejects on a five-percent significance level, providing evidence in favor of non-linearities, at least in Ω_j when modeling the full sample. Linearity is rejected for the full sample. It thus seems that the

Table 3.2: P-values for portmanteau statistics for left over autocorrelation in the residuals and in the squared residuals

	Full sample		before 2011	
	ACR	Linear model	ACR	Linear model
ϵ_t	0.01	0.01	0.14	0.10
ϵ_t^2	0.00	0.00	0.00	0.00

The portmanteau statistic is calculated as described in Lütkepohl (2005) and with the number of lags equal to $T/4$.

Table 3.3: Testing hypothesis

	Full sample			Sample until December 2011		
	ACR		Linear	ACR		Linear
Null Hypothesis	iid-bs	wild-bs	χ^2	iid-bs	wild-bs	χ^2
$\varphi_\beta = -1$	0.00	0.00	0.00	0.00	0.00	0.04
$\varphi_{\beta_D} = 0$	0.00	0.00	0.00	0.00	0.00	0.00
$\alpha_{1,1} = \alpha_{2,1} = 0$	0.75	0.73	0.33	0.35	0.44	0.00
$\alpha_{1,2} = \alpha_{2,2} = 0$	0.53	0.52	0.43	0.84	0.83	0.86
$\Gamma_1 = \Gamma_2$	0.37	0.55	-	0.66	0.72	-
$\Omega_1 = \Omega_2$	0.02	0.03	-	0.21	0.13	-
$\alpha_1 = \alpha_2$,$\Gamma_1 = \Gamma_2$	0.02	0.07	-	0.73	0.78	-
$\kappa = 0$	0.20	0.23	-	0.35	0.37	-
Linearity	0.00	0.02	-	0.27	0.27	-

p-values in the ACR model are based on the bootstrap procedure presented in algorithm 2.9 in chapter two. The number of bootstrap replications is set to 399. The p-value in the linear models are based on χ^2 approximations.

Table 3.4: Empirical rejection frequencies for the i.i.d bootstrap at the 5 % level for linearity tests

	Full sample	Before 2011
Estimated size	0.047	0.042
Estimated power	0.821	0.973

enhanced flexibility of the non-linearities gives a significant improvement over the linear model.

When estimating the model on the data that ends in 2011, the non-linearities are no longer rejected, indicating that it is indeed the period post 2011 that drives the regime switching. The tests of the remaining parameters have similar conclusions as in the previous setup.

Finally, we look at the linear alternative for the small sample and conduct tests based on χ^2 inference. It is seen that the hypothesis that the cointegration relation corresponds to the spread is rejected in the full sample and only marginally at a five percent significance level in the small sample. Likewise the hypothesis that the premium is zero is clearly rejected. Finally, when looking at the full sample, the error correction parameters are insignificant, while in the cut-off sample the Brent crude price has been error correcting very clearly, while the WTI has not. This illustrates the historical oddity of the recent period and indeed indicates that the dynamics of these prices have dramatically changed since 2011.

Weak exogeneity of the WTI versus the Brent is also found in similar studies, such as Alizadeh and Nomikos (2004), where a linear cointegrated VAR model is used to analyze a cost of carry relationship and where including data on freight rates are included as a means of modeling C_{Brent} in (3.1).

Remark 3.1. Observe that in chapter two, the test for linearity displayed small sample size distortions and to investigate whether similar distortions are present in this setup, we estimate empirical rejection frequencies using the estimated parameter values. The results are given in table 3.4, where fortunately, it seems that the test is well behaved, though with some lack of power.

3.6 Conclusion

Having analyzed the dynamic behavior of two crude oil benchmarks, the WTI and Brent using the ACR cointegrated model, we find that a remarkable departure from historical benchmarks has taken place since 2011 and that whether or not this period is included in the model has strong effects on the results. This departure is largely driving the non-linearities, which cannot be reject in favor of a linear cointegrated VAR model for the full sample. Analyzing the period up until December 2011, we do not find as strong support for ACR-type non-linearities and we observe through analysis of a linear cointegrated VAR that the WTI has been weakly exogenous. The estimated cointegration relations resemble the spread corrected for differences in quality and transportation costs, though the tests for the spreads as cointegration relations are rejected. On the theoretical side, we have shown that the asymptotic theory for the ACR cointegrated model developed in chapter one is intact, apart for minor adjustments, when the switching probabilities are taken to depend on the cointegration relations corrected for a non-zero mean rather than

the just the cointegration relations.

The estimated ACR cointegrated model pointed toward non-linearities spawning as a result of the 2011 decoupling. However, the non-linear structure considered in this model is not specifically designed to handle that kind of regime shift and using for example a model with an estimated breakpoint around 2011 could present an alternative. In such a setup one could have two different regimes, one before 2011 and one after. A third alternative could be to model the change in a linear framework, introducing a shift dummy in the cointegration relations.

3.A Appendix to section 3.3

The general likelihood derivatives are given in Lemmas (1.31), (1.32) and (1.33) from chapter one and are functions of p_{jt} and λ_{jt} (defined in (3.6) and section 1.7 of chapter one respectively) and of derivatives thereof. A way to show what happens when β_D is included in the switching probability is the following. Recall that we can identify $\beta^\star$ as in chapter one, section 1.6. We introduce further the notation $b^{\star\prime} = (b' : b'_D)$ along with the corresponding manipulated data vector $\mathcal{X}_t^\star = (\overline{\kappa}_0^\star : \overline{\tau}_0^\star : i_{n+1})' X_{t-1}^\star$. Using this notation, all the derivatives of λ_{jt} with respect to $\mathrm{d}b$, are now altered to be with respect to $\mathrm{d}b^\star$. That leaves the same expressions except that $\mathrm{d}b$ and $\mathcal{X}_{t-1}$ which are replaced by $\mathrm{d}b^\star$ and $\mathcal{X}_t^\star$. Moreover, the expressions derived for $\mathrm{d}b_D$ are redundant and removed.

To check that the asymptotic theory goes through when using this specification, we verify a few central claims modified to this setup. The remaining claims involving $\mathrm{d}b^\star$ will hold for the same reasons as those presented here. We use the notation from chapter one, Lemmas 1.12, 1.20 and 1.26 such that $S_T(\mathrm{d}b^\star)$ is the score evaluated at the true value in direction $\mathrm{d}b^\star$, $H_T\left(\mathrm{d}b^\star, \mathrm{d}b^{\star\dagger}\right)$ is the hessian evaluated at the true value in direction $\left(\mathrm{d}b^\star, \mathrm{d}b^{\star\dagger}\right)$ and $\mathrm{d}b_T^\star = T^{-\frac{1}{2}} W_{b^\star T}^{\frac{1}{2}} \mathrm{d}b^\star$, where $W_{b^\star T} := \mathrm{diag}\left(T I_{n-r-1}, T^2, 1\right)$.

Claim 3.2. $S_T(\mathrm{d}b_T^\star) \xrightarrow{w} \mathrm{vec}\left(\int_0^1 F^\star(s)\, \mathrm{d}\mathcal{B}_v(s)'\right)' \mathrm{vec}\left(\mathrm{d}b^{\star\prime}\right)$

Claim 3.3. $H_T\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}\right) \xrightarrow{w} -\mathrm{tr}\left\{\left(\mathrm{d}b^{\star\dagger}\right)' \int_0^1 \left(F^\star(s) F(s)^{\star\prime}\, \mathrm{d}s\right) \mathrm{d}b^\star \Sigma_{vv}\right\}$

Claim 3.4. $\sup_{\theta \in \mathcal{N}_T(\theta_0)} \left| T^{\frac{1}{2}} \mathrm{d}^3 L_T\left(\theta; \mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}, \mathrm{d}b_T^{\star\ddagger}\right)\right| = O_p\left(\|\mathrm{d}b^\star\| \left\|\mathrm{d}b^{\star\dagger}\right\| \left\|\mathrm{d}b^{\star\ddagger}\right\|\right)$.

Verification of Claim 3.2 Observe initially that $\beta^\star = \beta_0^\star + (\overline{\kappa}_0^\star : \overline{\tau}_0^\star : i_{n+1})\, b^\star$. As in chapter one, we observe that differentiation of λ_{jt} with respect to $b^\star$ gives

$$\begin{aligned} \mathrm{d}\lambda_{jt}\left(\mathrm{d}b^{\star\prime}\right) &= \varepsilon'_{jt}\Omega_j^{-1}\alpha_j \mathrm{d}b^{\star\prime}\mathcal{X}^\star{}_{t-1} + \partial_z \log p_{jt}\psi'_\beta \mathrm{d}b^{\star\prime}\mathcal{X}^\star_{t-1} \\ &= \left(\varepsilon'_{jt}\Omega_j^{-1}\alpha_j + \partial_u \log p_{jt}\psi'_\beta\right)\mathrm{d}b^{\star\prime}\mathcal{X}^\star_{t-1} \end{aligned}$$

such that we can write $h_{vt} := \sum_{j\in\mathbb{M}} p_{jt}^\star h_{vjt}$ and $\mathrm{d}\ell_t(\mathrm{vec}(\mathrm{d}b')) = \left(\mathrm{vec}\left(\mathcal{X}_{t-1}^\star h'_{vt}\right)\right)' \mathrm{vec}(\mathrm{d}b^{\star\prime})$ and hence by the same arguments as in chapter one, we get

$$S_T(\mathrm{d}b_T^\star) \xrightarrow{w} \mathrm{vec}\left(\int_0^1 F^\star(s)\, \mathrm{d}\mathcal{B}_v(s)'\right)' \mathrm{vec}\left(\mathrm{d}\left(b^{\star\prime}\right)\right)$$

where $F^\star(s) = \left(\mathcal{B}_\kappa(s)' : s : 1\right)'$.

Verification of Claim 3.3 As in chapter one, and with $\left(\mathrm{d}b_T, \mathrm{d}b_T^\dagger\right)$ replaced by $\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}\right)$, we have

$$\begin{aligned} H_T\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}\right) = &-\sum_{t=1}^T \mathcal{X}_{t-1}^{\star\prime} \mathrm{d}b_T^\star \varphi_{bt}\varphi'_{bt} \mathrm{d}b_T^{\dagger\prime\star} \mathcal{X}_{t-1}^\star - \sum_{t=1}^T \mathcal{X}^{\star\prime}{}_{t-1} \mathrm{d}b_T^\star \\ &\times \sum_{j\in\mathbb{M}} p_{jt}^\star \left\{v_{jt}v'_{jt} - \alpha_{0j}\Omega_{0j}^{-1}\alpha_{0j} + \psi_\beta\left(\partial_{uu}^2 \log p_{jt}\right)\psi'_\beta\right\} \mathrm{d}b_T^{\star\dagger\prime} \mathcal{X}_{t-1}^\star \end{aligned} \tag{3.7}$$

where $v_{jt} := \alpha_{0,j}'\Omega_{0,j}^{-1}\varepsilon_{jt} + \psi_\beta\left(\partial_u \log p_{jt}\right)'$ and $\varphi_{bt} := \sum_{j\in\mathbb{M}} p_{jt}^\star\left(\alpha_j'\Omega_j^{-1}\varepsilon_{jt} + \psi_1\left(\partial_z \log p_{jt}\right)'\right)$. The inner part of the second term,

$$f_t := \left\{v_{jt}v_{jt}' - \alpha_{0j}\Omega_{0j}^{-1}\alpha_{0j} + \psi_\beta\left(\partial_{zz}^2 \log p_{jt}\right)\psi_\beta'\right\}$$

is a mean zero, martingale difference sequence by the same arguments as in chapter one. Next, observe that

$$\mathcal{X}_{[Ts]}^{\star\prime}\mathrm{d}b^\star = \begin{pmatrix} X_{[Ts]}'\overline{\kappa}_0^\star \\ X_{[Ts]}'\overline{\tau}_0^\star \\ i_{n+1}\end{pmatrix}'\mathrm{d}b^\star$$

with $\overline{\kappa}_0^\star$, $\overline{\tau}_0^\star$ given in section 1.6.1 and with $s\in[0;1]$, and applying the arguments from chapter one, we have

$$W_{b^\star T}\mathcal{X}_{[Ts]}^\star\mathcal{X}_{[Ts]}^{\star\prime}W_{b^\star T} \xrightarrow{w} \begin{pmatrix}\mathcal{B}_\kappa(s)\\ s\\ 1\end{pmatrix}'\begin{pmatrix}\mathcal{B}_\kappa(s)\\ s\\ 1\end{pmatrix}$$

and

$$H_T\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}\right) \xrightarrow{w} -\mathrm{tr}\left\{\left(\mathrm{d}b^{\star\dagger}\right)'\int_0^1\left(F^\star(s)F^\star(s)'\mathrm{d}s\right)\mathrm{d}b^\star\Sigma_{vv}\right\}$$

as was desired.

Verification of Claim 3.4 For verification of this claim, observe that

$$\begin{aligned}\left\|W_{b^\star T}\mathcal{X}_{t-1}^\star\right\| &= \left\|T^{-\frac{1}{2}}\overline{\kappa}_0^{\star\prime}X_{t-1}^\star + T^{-1}\overline{\tau}_0^{\star\prime}X_{t-1}^\star + i_{n+1}'\right\| \\ &= \left\|T^{-\frac{1}{2}}\overline{\kappa}_0'X_{t-1} + T^{-1}\overline{\tau}_0'X_{t-1} + i_{n+1}'\right\|\end{aligned}$$

and thus the same arguments used in 1.49, one has that for some $a\geq 2q$ and with q given in Assumption 1.2 of chapter one, we have

$$T^{-1}\sum_{t=1}^T\left\|W_{b^\star T}\mathcal{X}_{t-1}^\star\right\|^a = O_p(1). \tag{3.8}$$

Next, the terms of interest are given by

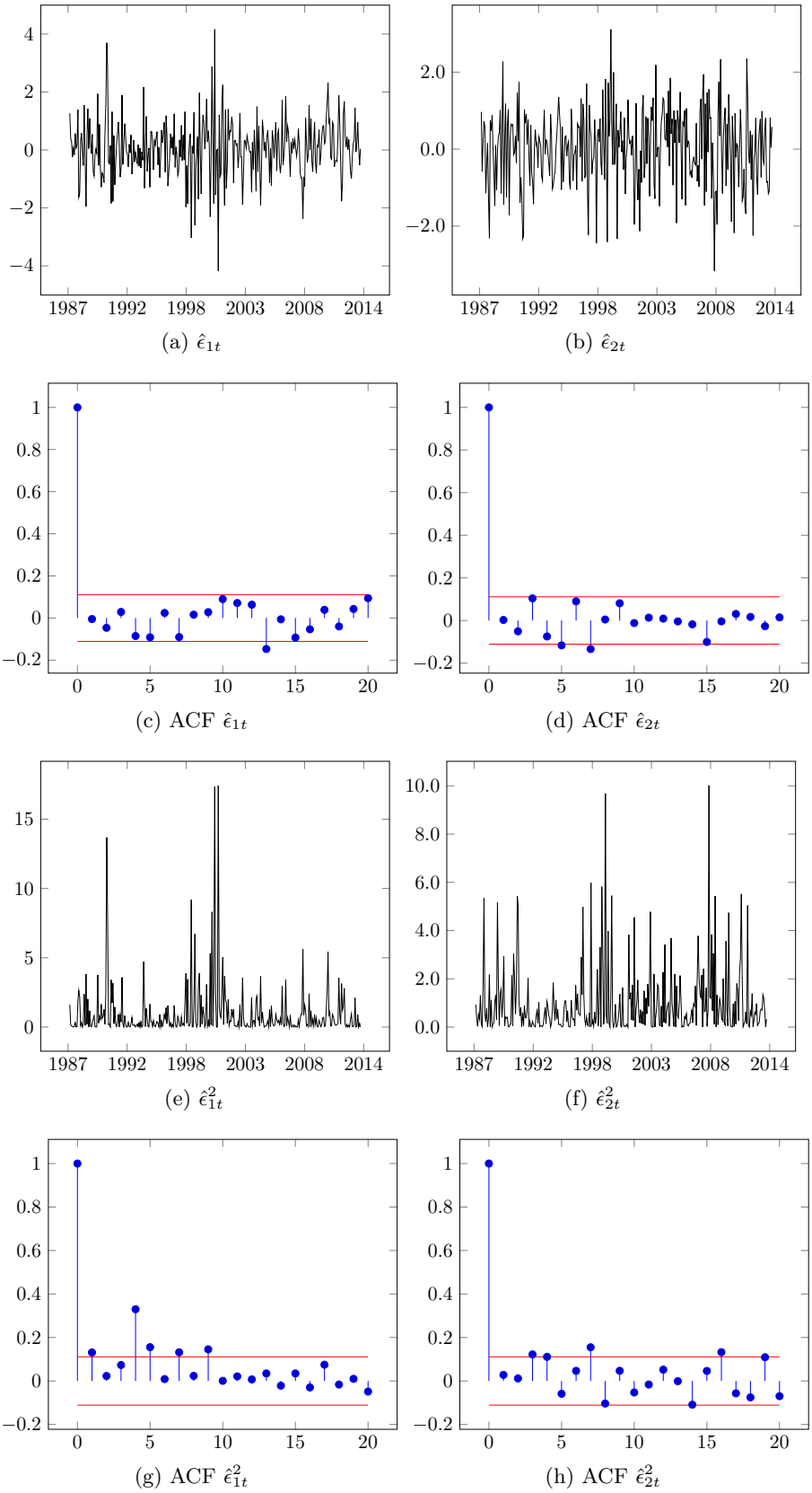
$$\begin{aligned}T^{\frac{1}{2}}\Xi_1\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}, \mathrm{d}b_T^{\star\ddagger}\right) &= T^{\frac{1}{2}}\sum_{t=1}^T \mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T^\star\right)\mathrm{d}\lambda_{it}\left(\mathrm{d}b_T^{\star\dagger}\right)\mathrm{d}\lambda_{ht}\left(\mathrm{d}b_T^{\star\ddagger}\right), \\ T^{\frac{1}{2}}\Xi_2\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}, \mathrm{d}b_T^{\star\ddagger}\right) &= T^{\frac{1}{2}}\sum_{t=1}^T \mathrm{d}\lambda_{jt}\left(\mathrm{d}b_T^\star\right)\mathrm{d}^2\lambda_{jt}\left(\mathrm{d}b_T^{\star\dagger}, \mathrm{d}b_T^{\star\ddagger}\right), \text{ and} \\ T^{\frac{1}{2}}\Xi_3\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}, \mathrm{d}b_T^{\star\ddagger}\right) &= T^{\frac{1}{2}}\sum_{t=1}^T \mathrm{d}^3\lambda_{jt}\left(\mathrm{d}b_T^\star, \mathrm{d}b_T^{\star\dagger}, \mathrm{d}b_T^{\star\ddagger}\right).\end{aligned}$$

Using (3.8) and (1.49) from chapter one, these terms are bounded by the applying the exact same arguments as in chapter one.

3.B Graphical analysis of residuals

In this appendix, we provide a number graphs for the estimated residuals. We give the residuals them selves, ϵ_{it} as well as the residuals squared, ϵ_{it}^2, for $i = \{1, 2\}$.

Figure 3.3: Graphical analysis of the residuals, Brent-WTI, ACR - full sample



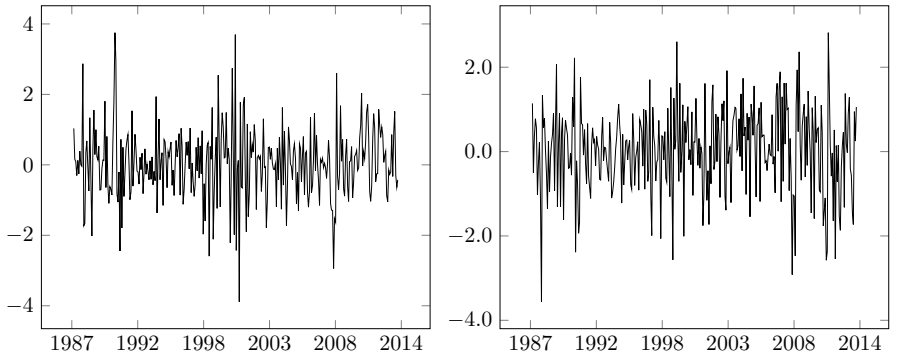
(a) $\hat{\epsilon}_{1t}$ (b) $\hat{\epsilon}_{2t}$

(c) ACF $\hat{\epsilon}_{1t}$ (d) ACF $\hat{\epsilon}_{2t}$

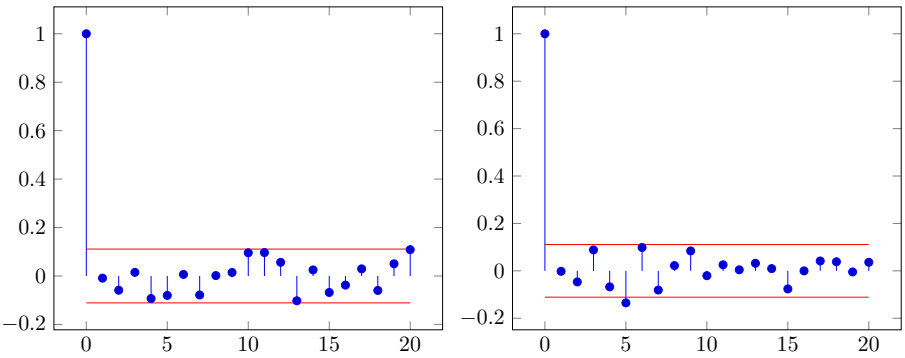
(e) $\hat{\epsilon}^2_{1t}$ (f) $\hat{\epsilon}^2_{2t}$

(g) ACF $\hat{\epsilon}^2_{1t}$ (h) ACF $\hat{\epsilon}^2_{2t}$

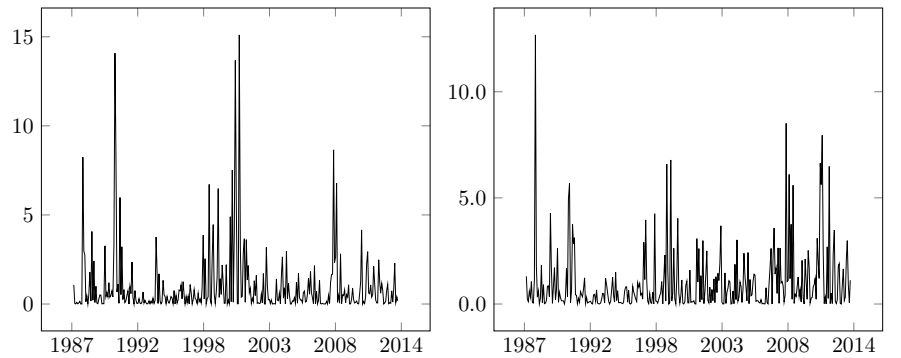
Figure 3.4: Graphical analysis of the residuals, Brent-WTI, Linear - full Sample



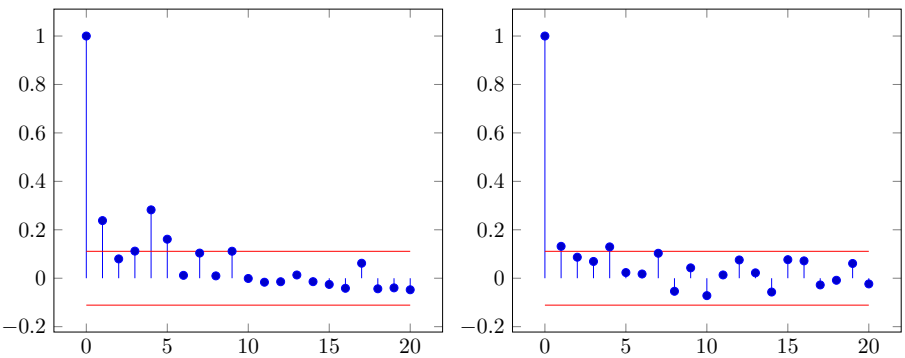
(a) $\hat{\epsilon}_{1t}$ (b) $\hat{\epsilon}_{2t}$



(c) ACF $\hat{\epsilon}_{1t}$ (d) ACF $\hat{\epsilon}_{2t}$

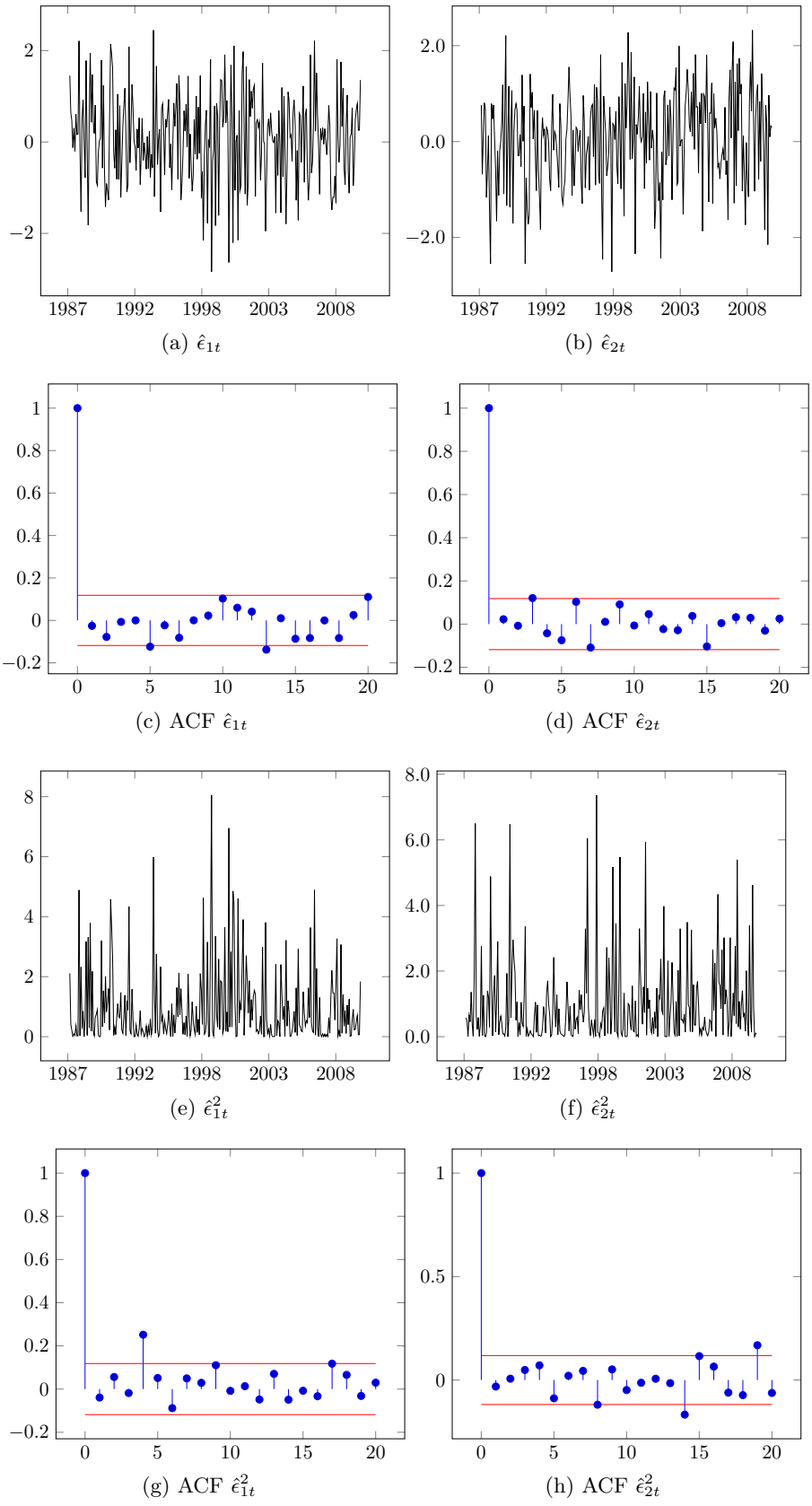


(e) $\hat{\epsilon}^2_{1t}$ (f) $\hat{\epsilon}^2_{2t}$



(g) ACF $\hat{\epsilon}^2_{1t}$ (h) ACF $\hat{\epsilon}^2_{2t}$

Figure 3.5: Graphical analysis of the residuals, Brent-WTI, ACR - before 2011



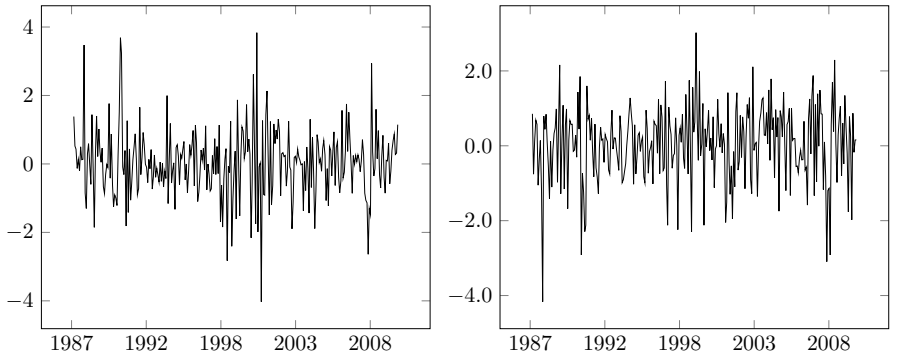
(a) $\hat{\epsilon}_{1t}$ (b) $\hat{\epsilon}_{2t}$

(c) ACF $\hat{\epsilon}_{1t}$ (d) ACF $\hat{\epsilon}_{2t}$

(e) $\hat{\epsilon}^2_{1t}$ (f) $\hat{\epsilon}^2_{2t}$

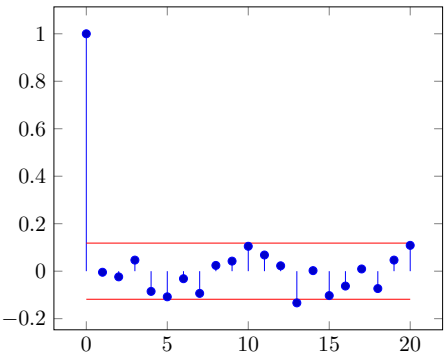
(g) ACF $\hat{\epsilon}^2_{1t}$ (h) ACF $\hat{\epsilon}^2_{2t}$

Figure 3.6: Graphical analysis of the residuals, Brent-WTI, Linear - before 2011

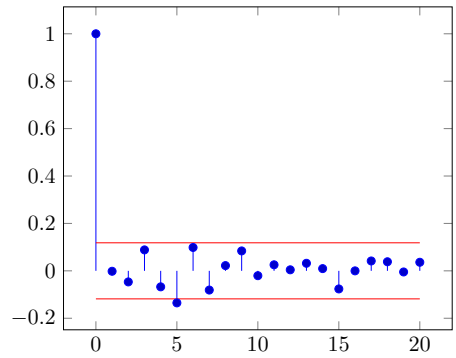


(a) $\hat{\epsilon}_{1t}$

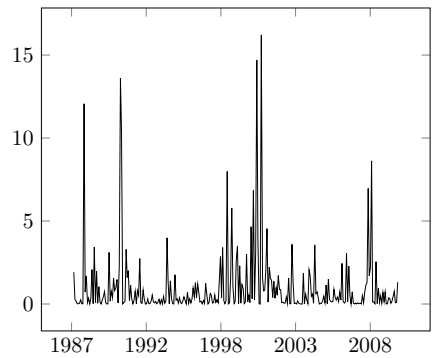
(b) $\hat{\epsilon}_{2t}$



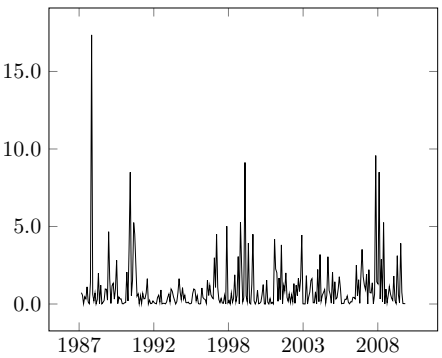
(c) ACF $\hat{\epsilon}_{1t}$



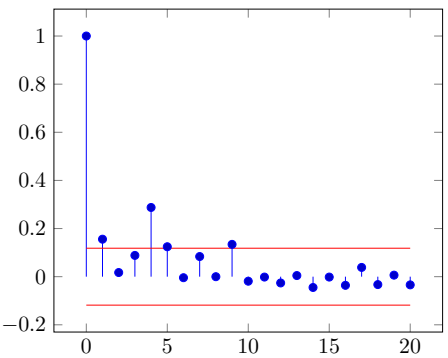
(d) ACF $\hat{\epsilon}_{2t}$



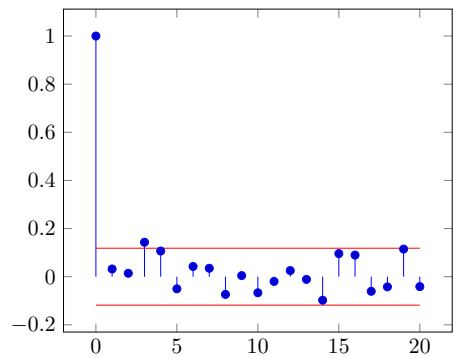
(e) $\hat{\epsilon}^2_{1t}$



(f) $\hat{\epsilon}^2_{2t}$



(g) ACF $\hat{\epsilon}^2_{1t}$

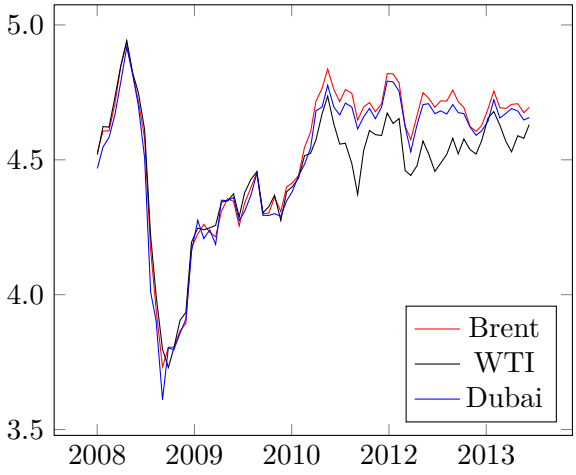


(h) ACF $\hat{\epsilon}^2_{2t}$

3.C Comparing WTI, Brent and Dubai-Fateh

To illustrate that the decoupling of the WTI and the Brent crude oil prices originate primarily from a general decoupling of the WTI from the remaining international crude oil streams, we display a graph where the Dubai-Fateh crude stream is also depicted. It is seen that the spread between the Dubai-Fateh and Brent crude oils is of a similar order before and after 2011, and that the WTI crude price has dropped relatively to the two others.

Figure 3.7: Brent, WTI and Dubai-Fateh, from 2008-2014



4 Smooth vs. non-smooth regime switching

This chapter is based on joint work with Line Elvstrøm Ekner.[1]

As observed in chapter two, section 2.2.4 using logistic switching probabilities can result in identification difficulties for the parameters in the switching probability. This problem is not isolated to the ACR framework, but also accors frequently in other non-linear models such as the logistic smooth transition autoregression (LSTAR), where the problematic parameter is referred to as the speed of transition parameter. We propose a reparametrization of the logistic smooth transition autoregressive model which facilitates identification and estimation of this parameter. Moreover, we show that all derivatives of the likelihood function are approaching zero as the parameter measuring the speed of transition increases, and, hence, the threshold autoregressive (TAR) model always represents at least a local stable point of the LSTAR likelihood function. We propose to use information criteria for the choice between the two models and illustrate the validity of this procedure by means of simulatons. Two empirical applications illustrate the usefulness of our findings.

4.1 Introduction

Regime switching models have become increasingly popular in the time series literature over the last decades and applied to data from potential regime switching processes such as, e.g., the business cycle, the unemployment rate, exchange rates, prices, interest rates, etc. The majority of the models initiate from the threshold autoregressive (TAR) model first presented by Tong and Lim (1980). Nevertheless, the idea of smooth regime switching was discussed by Bacon and Watts (1971), but not formalized in terms of a time series model until Chan and Tong (1986) proposed what they called a smoothed threshold autoregressive model as an extension to the TAR model of Tong and Lim (1980). Heavily cited contributions by Luukkonen et al. (1988) and Teräsvirta (1994) changed the label from "smoothed threshold" to "smooth transition" resulting in the label smooth transition autoregression (STAR) used today. For an overview of the TAR and STAR literature, see Tong (2011), Teräsvirta et al. (2010a), and van Dijk et al. (2002).

The logistic STAR (LSTAR) model differs from the TAR model by having smooth regime switches over time parametrized by the *speed of transition* parameter. The switches in the TAR model are in contrast discontinuous. The primary economic motivation for the LSTAR model is that economic time series are often results of decisions made by a large number of economic agents. Even if agents are assumed to make only dichotomous decisions or change their behavior

[1] We would like to thank Søren Johansen, Myung Hwan Seo and Timo Teräsvirta for very helpful comments and discussions.

discretely, it is unlikely that they do so simultaneously. Hence, any regime switching in economic time series may be more accurately described as taking place smoothly over time. Moreover, the speed of the regime switching can be of separate interest to an economist, e.g., to analyze how fast the economy adapts to another regime or state of the economy. In empirical applications it is, however, often very difficult to identify the speed of transition parameter, and, thus, it is important to test whether this additional parameter of the LSTAR model is at all relevant compared to a TAR model.

The first contribution of this paper is a reparametrization of the LSTAR model. In terms of the proposed parametrization we can explicitly illustrate the problem of distinguishing LSTAR and TAR alternatives.

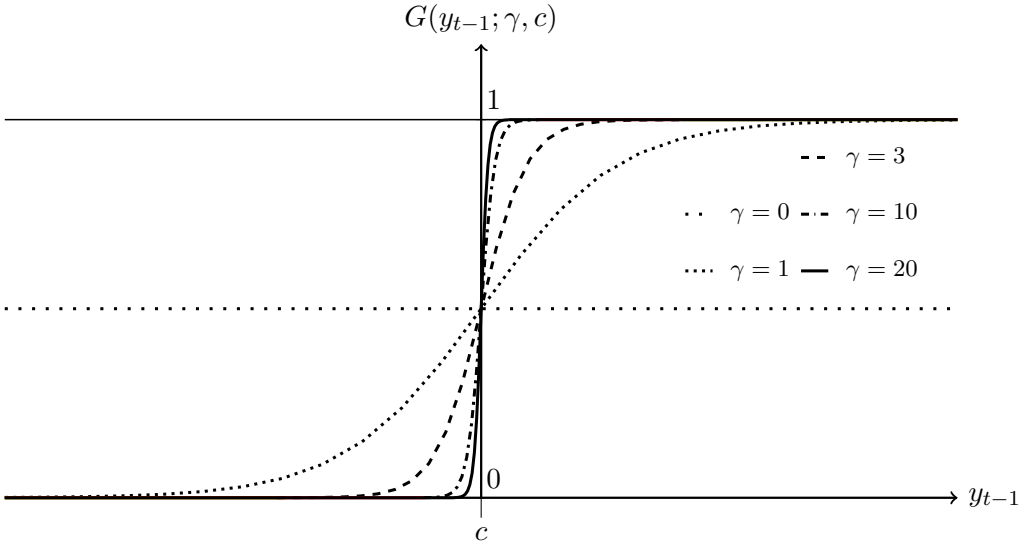
The second contribution is to show that this distinction is complicated both theoretically and in terms of numerical optimization by the fact that all derivatives of the LSTAR likelihood function are approaching zero with faster speed of transition, i.e., when the LSTAR model approaches the TAR model. Using likelihood analysis, we study the consequences of this identification problem for estimation and inference in the LSTAR model. The new parametrization avoids some of the numerical difficulties that arise when applying the original LSTAR parametrization, and, moreover, clarifies that the LSTAR likelihood function can have a maximum corresponding to a TAR model. In the literature of LSTAR models economic theory is used as the only motivation for modeling an LSTAR model instead of a TAR model, see, e.g., Granger and Teräsvirta (1993) and Teräsvirta (1998). However, our new parametrization facilitates a decision based upon the data, possibly in conjunction with economic theory. We show how information criteria provide a neat, but conservative, tool to select an LSTAR model over a TAR model that can be applied if one wishes to comment on the speed of transition.

The related issue of selecting between an LSTAR model and an autoregressive (AR) model is not treated in this paper. Although testing such hypothesis of linearity is non-standard, procedures are available and well-described in the literature of both the (L)STAR and TAR models, see Davies (1987), Luukkonen et al. (1988), Hansen (1996), and Kristensen and Rahbek (2013).

Furthermore, we discuss numerical optimization of the LSTAR likelihood function. In particular, we consider the origin of multiple maxima on the likelihood function and the use of grid search methods. Finally, we illustrate the benefits of the new parametrization and the model selection procedure with data from two published applications. In the first application, the likelihood function for the reparametrized speed of transition parameter reveals that the published result is only a local maximum on the likelihood function, and that the global maximum is a TAR model. In the second example, data contains insufficient information about the speed of transition parameter which then becomes irrelevant, and, as a result, information criteria prefer a TAR model over the published LSTAR model.

The new parametrization can be applied to all kinds of regime switching models where the regime switching is governed by one or more logistic type transition functions. Identification of the speed of transition parameter in the related exponential STAR (ESTAR) model with an exponential transition function has recently been studied by Heinen et al. (2012). However, the problem is different in the ESTAR model since this model approaches an AR model when the

Figure 4.1: The logistic transition function $G_t = \{1 + \exp(-\gamma(y_{t-1} - c))\}^{-1}$ for different values of γ.



speed of transition approaches infinity and not a TAR model. Hence, their results do no carry over to the LSTAR model. Nevertheless, the new parametrization can also be beneficial for estimation of the ESTAR model by facilitating numerical optimization as well as identification of the global maximum of the likelihood function.

4.2 The model and the identification problem

To fix ideas, consider a simple LSTAR model of order one for $y_t \in \mathbb{R}$, cf., Teräsvirta (1994),

$$y_t = \alpha y_{t-1} G_t + \varepsilon_t,\ t = 1, 2, .., T \tag{4.1}$$

with $\varepsilon_t \sim i.i.N(0, \sigma^2)$ and where G_t is the logistic transition function given by

$$G_t := G\left(y_{t-1}; \gamma, c\right) = \left(1 + \exp\left\{-\gamma(y_{t-1} - c)\right\}\right)^{-1}. \tag{4.2}$$

The AR parameter is α, γ is the speed of transition parameter and c is the threshold parameter. While $|\alpha| < 1$ and $c \in \mathbb{R}$, we assume that $\gamma \in \bar{\mathbb{R}}_+$ where $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \infty$. Thus, we extend the original definition of the parameter space for γ to include infinity, hereby making it feasible to discuss both the LSTAR model and the TAR model within the same framework. Figure 4.1 shows how the functional form of G_t changes with γ. In particular, note that as $\gamma \to 0$, $G_t \to \frac{1}{2}$ and as $\gamma \to \infty$, $G_t \to \mathbb{I}_{\{y_{t-1} - c > 0\}}$ where $\mathbb{I}_{\{A\}}$ is the indicator function equal to one when A is true and zero otherwise. Hence, the TAR model is a limiting case of the LSTAR model prevailing when $\gamma = \infty$. This feature of the model is central to the identification problem discussed in this paper.

The related ESTAR model is given by (4.1) and $G\left(y_{t-1}; \gamma, c\right) = 1 - \exp\left\{-\gamma(y_{t-1} - c)^2\right\}$. When $\gamma \to \infty$, $G_t \to 0$ (with a single blip at $y_{t-1} = c$) and the ESTAR model approaches a white noise process or, in a more general case, an AR model. Hence, poor identification of the speed of

transition parameter is, in contrast to the LSTAR model, often anticipated when testing against a linear model, which is standard in the STAR literature.

4.3 Likelihood analysis of the speed of transition parameter

LSTAR models are traditionally estimated by maximum likelihood (ML) or non-linear least squares (NLS). The two approaches are equivalent when the errors are assumed *i.i.d.* Gaussian, and thus the essential insights from the following ML analysis carry over to NLS. Before introducing the new parametrization, we illustrate some of the less attractive consequences of the original parametrization for estimation and inference in the LSTAR model. We are interested in analyzing only the properties of the ML estimator of γ, and, hence, we fix σ^2, α and c. The (log-)likelihood function is, apart from a constant, given by

$$L_T(\gamma) = \sum_{t=1}^{T} \ell_t(\gamma) = -\frac{1}{2}\sum_{t=1}^{T} \varepsilon_t(\gamma)^2\,,\ \varepsilon_t(\gamma) = y_t - \alpha y_{t-1} G_t \tag{4.3}$$

where G_t is the logistic transition function given by (4.2). Lemma 4.1 below provides results on the behavior of the derivatives of the likelihood function as the speed of transition parameter, γ, tends to infinity. Observe, in particular, that both the score and Hessian tend to zero as $\gamma \to \infty$, meaning that the likelihood function becomes flat as the LSTAR model approximates the TAR model. Hence, the TAR model always represents at least a local maximum of the likelihood function.

Lemma 4.1. *With the likelihood function given in (4.3), it holds for $n \geq 1$ that*

$$\lim_{\gamma\to\infty} \frac{\partial^n \ell_t(\gamma)}{(\partial\gamma)^n} = 0. \tag{4.4}$$

The proof is given in the appendix.

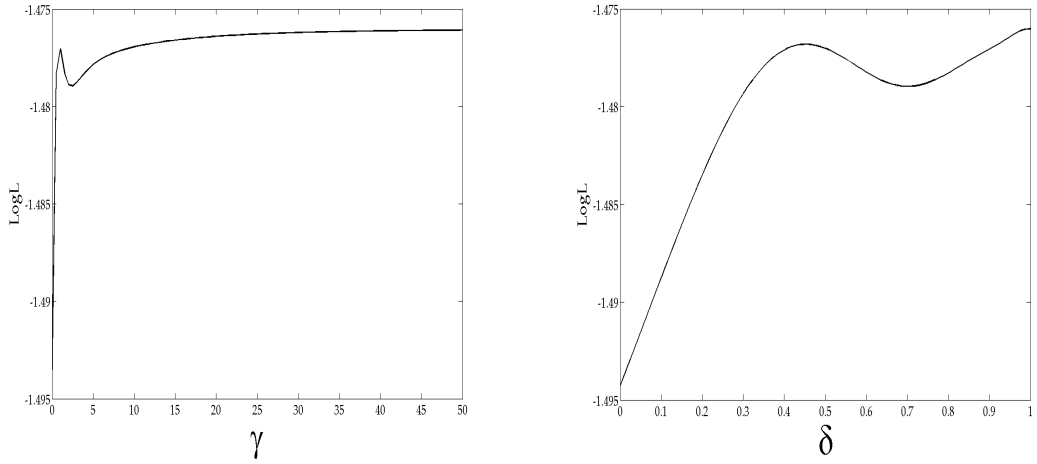
To illustrate the consequences for estimation, we simulate a data set from an LSTAR model with $T = 150$, $\gamma_0 = 2$, $\sigma^2 = 1$, $c = 0$ and $\alpha = 0.5$. The data series and G_t is graphed in figure 4.6 in appendix 4.A while the corresponding likelihood function as a function of γ is depicted in figure 4.2(a) below. Observe that the likelihood function gets flatter as the value of γ grows and the maximum is found, roughly, somewhere in the interval $\gamma \in [35; \infty]$ which does not contain $\gamma_0 = 2$. In empirical applications, one often estimate a large value of γ with a large standard error. As our parametrization below clarifies, this is in effect identical to estimating a TAR model.

4.3.1 The δ-parametrization

To minimize the harmful flat areas of the likelihood function, we propose the following reparametrization. We define a new parameter $\delta \in (0; 1]$, such that

$$\delta = \frac{\gamma}{1+\gamma} \tag{4.5}$$

Figure 4.2: The profiled likelihood function as a function of γ(a) and δ(b), respectively. The data set is simulated for $T = 150$, $\gamma_0 = 2$ ($\delta_0 = \frac{2}{3}$), $\sigma^2 = 1$,$c = 0$, $\alpha = 0.5$.



(a) (b)

with $\delta \to 0$ as $\gamma \to 0$ and $\delta \to 1$ as $\gamma \to \infty$. Hence, the transition function in (4.2) is replaced by

$$G\left(y_{t-1};\delta,c\right) = \left(1 + \exp\left\{-\frac{\delta}{1-\delta}(y_{t-1} - c)\right\}\right)^{-1} \tag{4.6}$$

Although lemma 1 also applies to an LSTAR likelihood function with (4.6), the reparametrization has advantages compared to the γ-parametrization. The main advantage is that it emphasizes the part of the likelihood function that is of principal interest in an LSTAR model. Essentially, the new parametrization maps $\gamma \in \bar{\mathbb{R}}_+$ into $\delta \in (0;1]$, where $\delta \in (0;1)$ is an LSTAR model, $\delta = 1$ is a TAR model, and $\delta = 0$ is an AR model. Of particular importance is the mapping of $\gamma \in [U;\infty]$ into $\delta \in [u;1]$, where U is some large value potentially tending to infinity and u is the corresponding value in the δ parametrization. For example, in figure 4.2 set $U \approx 9$ and, hence, $\gamma \in [9;\infty]$ is mapped into $\delta \in [0.9;1]$. This feature can facilitate numerical optimization of the likelihood function because the large flat part of the likelihood function appearing in figure 4.2(a) is now mapped into a much smaller interval as evident from figure 4.2(b). An example hereof is discussed in the next subsection.

The reparametrization highlights two important aspects that were less clear with the original γ-parametrization. First, the likelihood function is bimodal with a well defined local maximum around $\delta = 0.45$, corresponding to an LSTAR model with $\gamma \approx 0.8$ and, thus, not equal to the true value of $\gamma_0 = 2$ ($\delta_0 = \frac{2}{3}$). Apparently, for this particular realization the local maximum undershoots the true value of the speed of transition. Second, the δ-parametrization stresses that the global maximum of the likelihood function is found close to or at the boundary of the parameter space corresponding to a TAR model. The fact that the likelihood function actually continues to increase until $\delta \approx 1$ is less likely to be seen from the γ-parametrization.

Consequences for numerical optimization

Granger and Teräsvirta (1993, p. 123) note that γ tend to be overestimated, and this is also

Table 4.1: Estimated bias in $\hat{\gamma}$ and $\hat{\delta}$ as a function of the stopping criterion for the numerical optimizer.

Number of observations	$T = 150$			$T = 300$		
$S_T(\hat{x}) = \frac{\partial \ell_T(\hat{x})}{\partial \hat{x}}$	$\leq 10^{-2}$	$\leq 10^{-6}$	$\leq 10^{-16}$	$\leq 10^{-2}$	$\leq 10^{-6}$	$\leq 10^{-16}$
$\widehat{BIAS}(\hat{\gamma}) = \sum_{m=1}^{M}(\hat{\gamma}_m - \gamma)$	0.9063	8.2439	49.509	0.6594	5.4652	21.781
$\widehat{BIAS}(\hat{\delta}) = \sum_{m=1}^{M}(\hat{\delta}_m - \delta)$	0.0533	0.0545	0.0545	0.0146	0.0148	0.0148

Note: The DGP is $\gamma = 1$ ($\delta = \frac{1}{2}$), $\sigma^2 = 1$,$c = 0$ and $\alpha = 0.5$. $M = 10,000$ and c and α are fixed in estimation.

observed in the literature on LSTAR models where $\widehat{\gamma}$ has been reported to have a positive sample bias, see e.g. Areosa et al. (2011). This bias may be caused by the estimation of γ without recognizing the behavior of the numerical optimizer when the threshold alternative is the global maximum of a model with a logistic transition function. Table 4.1 shows results from a Monto Carlo study in which the estimated bias in $\hat{\gamma}$ is computed for different values of the stopping criterion of the numerical optimizer used to estimate γ. We focus on the stopping criterion related to the score of the likelihood function. The positive bias in $\hat{\gamma}$ depends heavily on the value of this criterion. This illustrates that the often arbitrary choice of stopping criterion for the numerical optimizer affects the bias in $\hat{\gamma}$. However, for the δ-parametrization, the bias appears (almost) unaffected by the size of the stopping criterion.[2]

4.4 Estimating LSTAR models

The properties of the likelihood function for LSTAR models discussed so far introduce difficulties for numerical optimization. We observe two separate problems that have to be taken into account. First, the likelihood function might have a multiple maxima in the direction of δ, as described in the previous sections. To handle this, it is useful to estimate δ with a derivative based optimizer and using different initial values from the parameter space $\delta \in (0; 1]$. To ensure that the reached maximum is global, it is important to always calculate the additional likelihood value at the limit, $\delta = 1$.

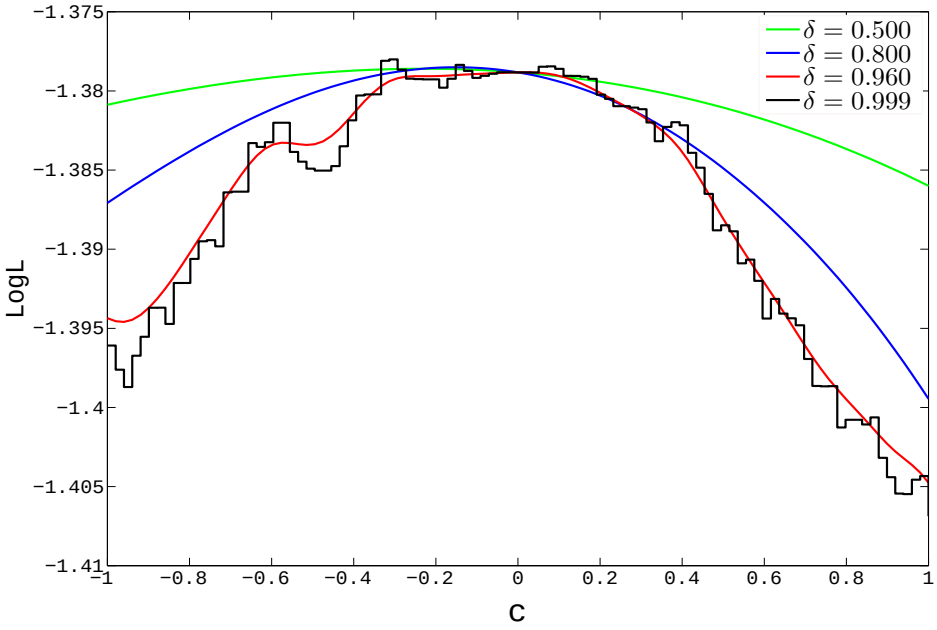
The second difficulty is that the likelihood function approaches the step-wise likelihood function of a TAR model in the direction of c as $\delta \to 1$. Consequently, many local maxima exist in the direction of c and derivative based optimizers will not work well. For an illustration see figure 4.3, which shows the likelihood as a function of c for different values of δ. To circumvent the problem of a step-wise likelihood function, a grid search algorithm over c can be performed with an interval that covers observed values of y_{t-1} spanning from, e.g., the 10th to the 90th percentile of the distribution of y_{t-1}. This grid search technique for c is standard in the TAR literature and ensures that all relevant points for threshold locations are examined. The rest of the parameters are estimated using least squares conditional on the transition function parameters.

When estimating simple models, as the one analyzed in this paper, performing a two di-

[2]Note that the size of the bias is not comparable across parametrizations due to the different scaling of the parameters and that since the ML estimator is consistent, the bias diminishes as T grows.

mensional grid search over δ and c and drawing the profiled likelihood function is generally informative. This approach allows one to take into account both problems. Note that this proposal is by no means new and is in fact standard practice in the literature for finding candidates for initial values, see inter alia Bec et al. (2008) and Teräsvirta et al. (2010a, ch. 12). Our contribution is that the δ-parametrization clarifies the reason for doing the grid search, and we emphasize that the main problem of multiple equilibria of the LSTAR model is related to the fact that the likelihood function approaches a step-wise likelihood function as $\delta \to 1$.

Figure 4.3: Profiled likelihood functions in the direction of c for different values of δ. Data is one realization from an LSTAR model with $T = 300$, $\sigma^2 = 1$, $c = 0$ and $\alpha = 0.5$.



4.5 Selecting between LSTAR and TAR with information criteria

The bimodality of the likelihood function seen in figure 4.2(b) is a common small sample property of LSTAR models. Typically, there exists one inner maximum corresponding to an LSTAR model and a maximum on the boundary of the parameter space ($\delta = 1$) corresponding to a TAR model. The simulation considered above in figure 4.2 is an extreme example of this, where the global maximum of the likelihood function is at $\delta \approx 1$. A more typical case is one with the inner maximum being the global maximum and a local, smaller maximum is found at the TAR solution, see for example figure 4.5(b). A relevant question is therefore whether the likelihood value of the inner maximum is large enough compared to the likelihood value of the boundary TAR maximum to justify estimation of a speed of transition parameter. One way to investigate this question would be to derive a test for the null-hypothesis of $\delta = 1$. However, such a test is highly non-standard since, as Lemma 4.1 shows, all derivatives are zero and, hence, it is not obvious how to obtain critical values. We propose instead to conduct model selection based on information criteria, where no critical values are needed. Information criteria combine a measure of goodness-of-fit with a penalty for model complexity. Comparing information criteria would

therefore indicate whether the additional speed of transition parameter of an estimated LSTAR model leads to a notable improvement of fit compared to a corresponding TAR model. Note that the theoretical foundation for validity of the information criteria when the true model is the TAR model suffers from the same difficulties as a formal test. As a result, we are unable to prove the asymptotic validity of this selection procedure analytically. Rather, we rely on simulation studies which give clear indications that the model selection proceduce is indeed consistent. We conjecture that these simulation results are not specific to the selected models, such that information criteria can be used more generally to select between TAR and STAR models.

Psaradakis et al. (2009) pursue a comparable idea and consider selecting between several non-linear autoregressive models by means of information criteria. In the following, we conduct a similar simulation study for the choice between a TAR model and an LSTAR model using the proposed reparametrization. With $L_T(\delta)$ being the likelihood function given in (4.3) evaluated with respect to δ, the information criteria are of the form

$$IC_T(\delta, k) = -2L_T(\delta) + kc_T, \tag{4.7}$$

where k is the number of estimated parameters which equals 1 for the LSTAR and 0 for the TAR. The term c_T is a function of T that satisfies $\lim_{T\to\infty} c_T = \infty$ and $\lim_{T\to\infty}(T^{-1}c_T) = 0$. We focus on the Bayesian Information Criterion (BIC), Schwarz (1978), with $c_T = \log(T)$, and the Hannan-Quinn Information Criterion (HQIC), Hannan and Quinn (1979), with $c_T = 2\log(\log(T))$. Both criteria fulfill the requirements for c_T. We use the information criteria to estimate the number of parameters, k, and denote this estimate $\hat{k}$. The selection procedure is consistent if $\hat{k} \to k_0$ as $T \to \infty$. We illustrate this selection method using four different models, three LSTAR models with $\delta = \{0.2, 0.5, 0.9\}$, respectively, and a TAR model. We simulate $M = 10,000$ data sets and estimate only the speed of transition parameter δ. The remaining parameters are fixed at the true values: $\sigma^2 = 1$, $\alpha = 0.5$ and $c = 0$. For each replication, we calculate the percentage selected LSTAR models of the two information criteria. The experiment is done for different sample lengths and the selection percentages are given in table 4.2.

Table 4.2: Percentage selected LSTAR models using information criteria.

DGP	LSTAR, $\delta = 0.2$		LSTAR, $\delta = 0.5$		LSTAR, $\delta = 0.9$		TAR, $\delta = 1$	
T	HQIC	BIC	HQIC	BIC	HQIC	BIC	HQIC	BIC
100	63	47	25	13	3	1	3	1
250	92	82	45	25	3	1	3	1
500	99	98	68	46	3	1	3	1
10^3	100	100	90	76	3	1	3	1
10^4	100	100	100	100	6	1	2	0
10^5	100	100	100	100	42	6	0	0
10^6	100	100	100	100	100	100	0	0

Note: Only δ is estimated. Remaining parameters are fixed at true values of $\sigma^2 = 1$, $\alpha = 0.5$ and $c = 0$.

The results show that the slower the speed of transition, the better the performance of the

information criteria. Nevertheless, even with a relatively slow transition speed of $\delta = 0.5$ and $T = 1,000$, BIC and HQIC still select a rather large number of incorrect TAR models, 24% and 10%, respectively. For the LSTAR model with $\delta = 0.9$ the information criteria are apparently punishing too severely for the additional parameter and do not choose the LSTAR model in 100% of the cases until $T = 1,000,000$. Again, this shows that while the identification problem for δ is a small sample problem, sometimes T needs to be extremely large to get a clear distinction between an LSTAR and a TAR model. This finding is also supported by the results of Castle and Hendry (2013, table 3) which shows that data generated from the two models are highly correlated. Observe that when the TAR model is the DGP, the information criteria only choose the LSTAR rarely for small samples and not at all for (very) large samples.

Overall, if model selection based on information criteria prefer an LSTAR, it is a clear indication that the speed of transition is slow enough to make a difference compared to the TAR model. On the other hand, if the TAR is chosen, there is a risk that one has incorrectly fixed $\delta = 1$. However, this only means that the value of δ is irrelevant for the model. Hence, information criteria provide a conservative tool to select LSTAR models over TAR models.

4.6 Empirical applications

We illustrate by two empirical applications from the LSTAR literature the advantages of the δ-parametrization over the γ-parametrization and model selection based on information criteria. The first application illustrates a situation where the δ-parametrization reveals that the reported maximum of the likelihood function is not the global maximum. In the second application, the δ-parametrization confirms that the global maximum is the reported one, but information criteria prefer the TAR model over the LSTAR model because the regime switching is so fast that estimating the additional speed of transition parameter is superfluous.

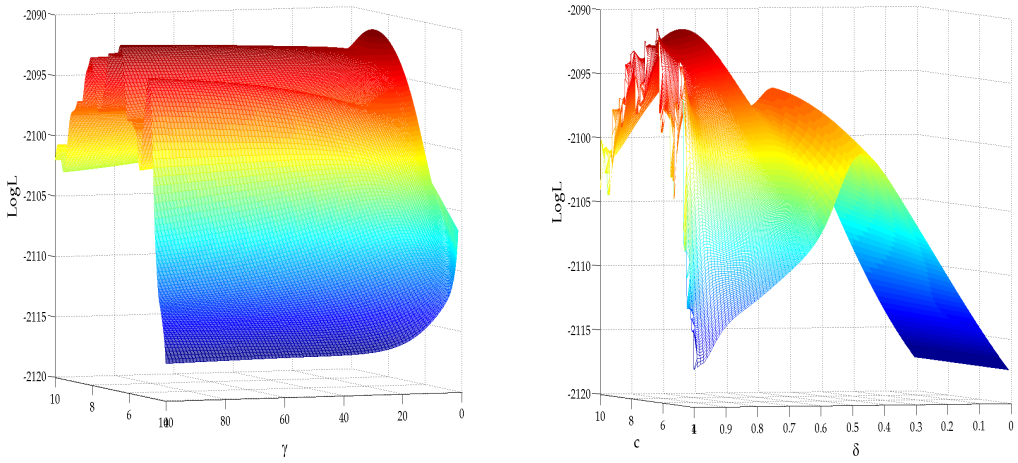
4.6.1 Wolf's annual sunspot numbers

Teräsvirta et al. (2010a, p. 390), illustrate a suggested STAR modeling procedure by analyzing Wolf's annual sunspot numbers dating from 1700 to 1979. The data is published at the Belgian web page of Solar Influences Data Analysis Center.[3] Following Teräsvirta et al. (2010a) the series is transformed as: $y_t = 2\left\{(1+z_t)^{1/2} - 1\right\}$ where z_t is the original series. The motivation for transformation is that the transformed series is easier to model than the untransformed one. The original estimated LSTAR model is reproduced with both parametrizations and given by (standard errors in parenthesis)

$$\begin{aligned} y_t &= \underset{(0.08)}{1.46} y_{t-1} - \underset{(0.13)}{0.76} y_{t-2} + \underset{(0.05)}{0.17} y_{t-7} + \underset{(0.04)}{0.11} y_{t-9} \\ &\quad + (\underset{(0.85)}{2.65} - \underset{(0.13)}{0.54} y_{t-1} + \underset{(0.18)}{0.75} y_{t-2} - \underset{(0.11)}{0.47} y_{t-3} \\ &\quad + \underset{(0.11)}{0.32} y_{t-4} - \underset{(0.07)}{0.26} y_{t-5} - \underset{(0.05)}{0.24} y_{t-8} + \underset{(0.06)}{0.17} y_{t-10}) \times \widehat{G}_t^x \end{aligned} \tag{4.8}$$

[3] http://www.sidc.oma.be/sunspot-data/

Figure 4.4: Profiled likelihood functions of the LSTAR model for Wolf's sunspot numbers, 1710-1979. (a) is for the γ-parametrization and (b) is the for the δ-parametrization.



(a) (b)

$$x = \gamma: \quad \widehat{G}^{\gamma} = 1 + \exp\{\underset{(1.11)}{-5.46}(y_{t-2} - \underset{(0.36)}{7.88})/\widehat{\sigma}_{y_{t-2}}\}^{-1}$$

$$x = \delta: \quad \widehat{G}^{\delta} = 1 + \exp\left\{-\frac{\underset{(0.03)}{0.85}}{1 - \underset{(0.03)}{0.85}}(y_{t-2} - \underset{(0.36)}{7.88})/\widehat{\sigma}_{y_{t-2}}\right\}^{-1}$$

$$T = 270, \qquad RSS = 921.84, \qquad LogL = -2,091.2$$
$$BIC = 4,260.8, \qquad HQIC = 4,230.7$$

The normalization by $\widehat{\sigma}_{y_{t-2}}$ in the transition function is standard in the literature of applied STAR models because it facilitates the choice of grid or initial values for γ, see van Dijk et al. (2002). The profiled likelihood function in direction of c and γ for each parametrization is showed in figure 4.4. The characteristically flatness in the direction of γ is pronounced in figure 4.4(a), and the reported maximum for $(\widehat{c}, \widehat{\gamma}) = (5.46, 7.88)$ appears relatively well-defined. However, figure 4.4(b) reveals that the global maximum is actually the TAR model at the boundary $\delta = 1$, whereas the LSTAR model is only a local maximum. The γ-parametrization has effectively blurred the shape of the likelihood function. At the boundary, the TAR likelihood function is characterized by discrete jumps over the range of c. This implies that performing a careful grid search over potential values of c is crucial for the estimation of c, as discussed in section 4.4 and, more importantly, that inference on c is non-standard, cf., Chan (1993) and Hansen (1997). Estimating the TAR model yields[4]

[4]The grid search of c is performed over values of y_{t-2}, disregarding values in the lower 10% percentile and upper 90% percentile of the distribution of y_{t-2}. No standard error of $\widehat{c}$ is reported due to the non-standard inference on the threshold parameter in a TAR model.

$$
\begin{aligned}
y_t &= \underset{(0.08)}{1.43} y_{t-1} - \underset{(0.14)}{0.77} y_{t-2} + \underset{(0.05)}{0.17} y_{t-7} + \underset{(0.05)}{0.12} y_{t-9} \\
&\quad + (\underset{(0.70)}{2.69} - \underset{(0.11)}{0.45} y_{t-1} + \underset{(0.18)}{0.69} y_{t-2} - \underset{(0.11)}{0.48} y_{t-3} \\
&\quad + \underset{(0.11)}{0.36} y_{t-4} - \underset{(0.07)}{0.27} y_{t-5} - \underset{(0.05)}{0.21} y_{t-8} + \underset{(0.05)}{0.14} y_{t-10}) \times \mathbb{I}(y_{t-2} > 6.39) \,. \qquad (4.9)
\end{aligned}
$$

$$
T = 270, \quad RSS = 920.66, \quad LogL = -2,090.9
$$
$$
BIC = 4,254.6, \quad HQIC = 4,226.6
$$

While the AR parameters are almost identical to those of the LSTAR model in (4.8), the threshold parameter differs between the models. This TAR maximum is preferred by the information criteria to the reported LSTAR model in (4.8) because the TAR model achieves a higher (lower) value of LogL (RSS) in addition to be one parameter short of the LSTAR model.[5] The TAR maximum (4.9) can easily be reproduced with the δ-parametrization by performing a two-dimensional grid search over c and $\delta \in (0; 1]$. A similar exercise for the γ-parametrization produces, depending on the choice of grid for γ as well as the choice of stopping criterion, either the local LSTAR maximum of (4.8) or an invalid maximum with all observations in one regime. Hence, the model that truly maximizes the likelihood function is impossible to estimate with the γ-parametrization because γ is infinity.

Nevertheless, given that an LSTAR process has a TAR maximum as a small sample property, as found in section 4.5, and the relatively small sample size of 270, the LSTAR model cannot be discarded as being the DGP of this sunspot data. In addition, the likelihood function in the region of the local LSTAR maximum in (4.8) and appearing in figure 4.4, seems closely approximated by a quadratic form, and is thus a well defined maximum. Based on these considerations, one could also argue that the LSTAR model may be the DGP of the process.

4.6.2 U.S. unemployment rate

The paper by van Dijk et al. (2002) illustrates a suggested STAR modeling cycle which includes, among others, impulse response and forecasting analysis. The data series is the monthly seasonally unadjusted unemployment rate for U.S. males aged 20 and over for the period 1968:6-1989:12.[6]

The LSTAR model is reproduced with both parametrizations and given by (standard errors in parenthesis)

$$
\Delta y_t = \underset{(0.07)}{0.479} + \underset{(0.07)}{0.645} D_{1,t} - \underset{(0.10)}{0.342} D_{2,t} - \underset{(0.09)}{0.680} D_{3,t} - \underset{(0.11)}{0.725} D_{4,t} - \underset{(0.10)}{0.649} D_{5,t}
$$

[5] Teräsvirta et al. (2010a) reach similar conclusion when estimating a TAR model for the same data later in the book, though, without specifying a measurement. Their TAR model is, however, specified differently and non-nested with (4.9) and (4.8) which makes direct comparisons infeasible.

[6] The series is constructed from data on the unemployment level and labor force for the particular sub population. These two series are published together with Gauss programs used to estimate their model at http://swopec.hhs.se/hastef/abs/hastef0380.htm.

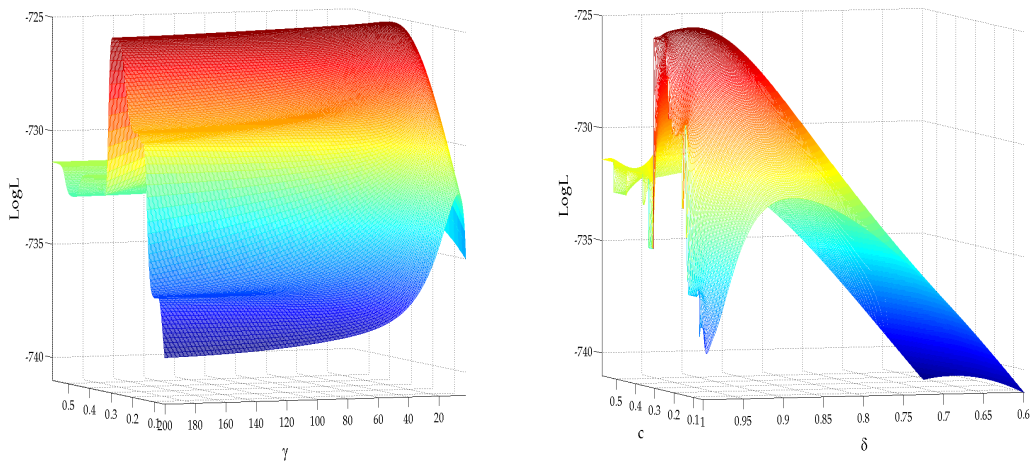
$$
\begin{aligned}
&\underset{(0.09)}{-0.317}D_{6,t} - \underset{(0.09)}{0.410}D_{6,t} - \underset{(0.09)}{0.501}D_{8,t} - \underset{(0.09)}{0.554}D_{9,t} - \underset{(0.07)}{0.306}D_{10,t} \\
&+[\underset{(0.01)}{-0.040}y_{t-1} - \underset{(0.08)}{0.1460}\Delta y_{t-1} - \underset{(0.06)}{0.101}\Delta y_{t-6} + \underset{(0.06)}{0.097}\Delta y_{t-8} - \underset{(0.06)}{0.123}\Delta y_{t-10} \\
&+\underset{(0.07)}{0.129}\Delta y_{t-13} - \underset{(0.06)}{0.103}\Delta y_{t-15}] \times [1 - \widehat{G}_t^x] \\
&+[\underset{(0.01)}{-0.011}y_{t-1} + \underset{(0.08)}{0.225}\Delta y_{t-1} + \underset{(0.08)}{0.307}\Delta y_{t-2} - \underset{(0.07)}{0.119}\Delta y_{t-7} - \underset{(0.09)}{0.155}\Delta y_{t-13} \\
&-\underset{(0.09)}{0.215}\Delta y_{t-14} - \underset{(0.09)}{0.235}\Delta y_{t-15}] \times \widehat{G}_t^x
\end{aligned} \tag{4.10}
$$

$$
\begin{aligned}
x = \gamma: \quad & \widehat{G}^{\gamma} = 1 + \exp\{\underset{(21.75)}{-23.15}(\Delta_{12}y_{t-1} - \underset{(0.04)}{0.274})/\widehat{\sigma}_{\Delta_{12}y_{t-1}}\}^{-1} \\
x = \delta: \quad & \widehat{G}^{\delta} = 1 + \exp\left\{-\frac{\underset{(0.04)}{0.96}}{1 - \underset{(0.04)}{0.96}}(\Delta_{12}y_{t-1} - \underset{(0.04)}{0.274})/\widehat{\sigma}_{\Delta_{12}y_{t-1}}\right\}^{-1}
\end{aligned}
$$

$$
T = 240, \quad RSS = 8.178, \quad LogL = -725.0
$$
$$
BIC = 1,597.9, \quad HQIC = 1,541.8
$$

$D_{s,t}$ is monthly dummy variables where $D_{s,t} = 1$ if observation t corresponds to month s and $D_{s,t} = 0$ otherwise. van Dijk et al. (2002) have sequentially removed all variables with a t-statistic lower than 1 in absolute value. Observe that γ is rather large and imprecisely estimated indicating that data contains little information about the size of this parameter. The profiled likelihood functions for the two parametrizations are displayed in figure 4.5. Because $\widehat{\gamma}$ is so large, the

Figure 4.5: Profiled likelihood functions of the LSTAR model for U.S. male unemployment rate, 1968:6-1989:12. (a) is for the γ-parametrization and (b) is the for the δ-parametrization.



(a) (b)

maximum is visually absorbed by the flatness of the γ-likelihood function in figure 4.5(a). In contrast, the δ-likelihood function in figure 4.5(b) confirms that the reported maximum is in fact the global maximum of the likelihood function. Interestingly, the δ-likelihood function shows

that the local TAR maximum at the boundary leads to only a minor drop in likelihood value compared to the LSTAR model. To check whether this TAR model is preferred by information criteria, the TAR model is estimated and given by[7]

$$\begin{aligned}\Delta y_t &= \underset{(0.07)}{0.473} + \underset{(0.07)}{0.644} D_{1,t} - \underset{(0.10)}{0.343} D_{2,t} - \underset{(0.09)}{0.675} D_{3,t} - \underset{(0.11)}{0.721} D_{4,t} - \underset{(0.10)}{0.641} D_{5,t} \\ &\quad \underset{(0.09)}{-0.308} D_{6,t} - \underset{(0.09)}{0.410} D_{6,t} - \underset{(0.08)}{0.505} D_{8,t} - \underset{(0.09)}{0.546} D_{9,t} \underset{(0.07)}{-0.295}\, D_{10,t} \\ &\quad + [\underset{(0.01)}{-0.040} y_{t-1} - \underset{(0.08)}{0.140} \Delta y_{t-1} - \underset{(0.06)}{0.094} \Delta y_{t-6} + \underset{(0.06)}{0.092} \Delta y_{t-8} - \underset{(0.06)}{0.116} \Delta y_{t-10} \\ &\quad \underset{(0.07)}{+0.136} \Delta y_{t-13} - \underset{(0.06)}{0.106} \Delta y_{t-15}] \times \mathbb{I}(\Delta_{12} y_{t-1} \leq 0.268) \\ &\quad [\underset{(0.01)}{-0.012} y_{t-1} + \underset{(0.08)}{0.227} \Delta y_{t-1} + \underset{(0.08)}{0.307} \Delta y_{t-2} - \underset{(0.07)}{0.094} \Delta y_{t-7} - \underset{(0.09)}{0.146} \Delta y_{t-13} \\ &\quad \underset{(0.09)}{-0.211} \Delta y_{t-14} - \underset{(0.09)}{0.216} \Delta y_{t-15}] \times \mathbb{I}(\Delta_{12} y_{t-1} > 0.268) \end{aligned} \tag{4.11}$$

$$T = 240, \quad RSS = 8.191, \quad LogL = -725.3$$
$$BIC = 1,593.2, \quad HQIC = 1,539.1$$

The information criteria prefer this TAR model implying that the speed of transition is too poorly estimated to make a difference.

This application highlights one of the key points of the present paper, namely that a large and imprecisely estimated γ implies that the LSTAR model is effectively a TAR model. Moreover, we observe the consequences of the flat likelihood function for inference on $\hat{\gamma}$. The estimated standard error of $\hat{\gamma}$ (s.e.($\hat{\gamma}$)) is large due to the flatness of the likelihood function in direction of γ towards infinity, see figure 4.5(a). However, the large s.e.($\hat{\gamma}$) seems less justified towards zero where one observes a large drop in the likelihood. This illustrates how the flatness contaminates the estimation of the variance of $\hat{\gamma}$ for which zero is well within a two s.e.($\hat{\gamma}$). Consequently, one might conclude that γ could be zero, which from a look at the function in figure 4.5(a) seems unlikely. For this reason (and because a test of $\gamma = 0$ results in vanishing parameters and, thus, is a non-standard test), it is common practice in the LSTAR literature not to comment on the s.e.($\hat{\gamma}$). It is seen from the s.e.($\hat{\delta}$) that the δ-parametrization does not suffer from this problem in the present application.

4.7 Conclusion

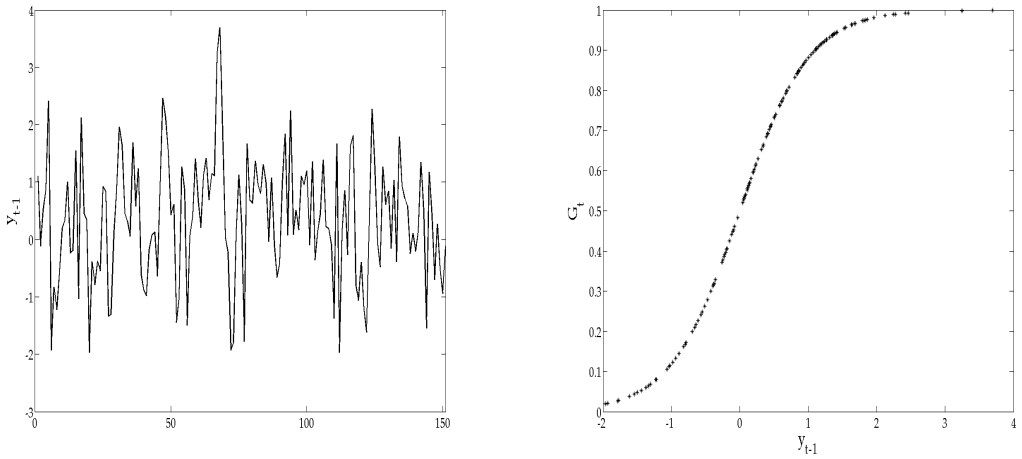
Regime switching models characterized by smooth transitions only differ from discrete regime switching models by the speed of transition parameter. Thus, estimation and identification of this parameter is essential not only for economic interpretation but also for model selection. Nevertheless, the identification problem and its consequences for estimation have received little attention in the literature. We show that the original parametrization of the speed of transition

[7] Similar to the previous TAR estimation, the grid search of c is performed over values of $\Delta_{12} y_{t-1}$, disregarding values in the lower 10% percentile and upper 90% percentile of the distribution of $\Delta_{12} y_{t-1}$. No standard error of $\widehat{c}$ is reported due to the non-standard inference on the threshold parameter in a TAR model.

parameter is problematic as the likelihood function is characterized by large flat areas caused by all derivatives approaching zero with faster speed of transition. This implies that the magnitude of the estimator may depend on the arbitrarily chosen stopping criteria of the numerical optimizer. To circumvent this problem, we propose a reparametrization of the LSTAR model. The reparametrization maps the parameter space of the original speed of transition parameter into a much smaller interval which facilitates identifying the global maximum of the likelihood function as well as numerical optimization. We then show that the TAR model can be the global maximum of a LSTAR likelihood function, while it, by construction, is always at least a local stable point and possibly a maximum. Instead of relying solely on economic theory when justifying the additional parameter of the LSTAR model, we show that information criteria provide a model selection tool. Acknowledging that the LSTAR model considered in this paper is simple and the presented simulation results only apply to this particular framework, the new parametrization provides general insights on the shape of the likelihood function in directions of the two parameters of the transition function that can be generalized to a broad range of other models within the smooth switching literature. For example, the double-logistic smooth transition (D-LSTAR), the Multi-Regime Smooth Transition Autoregression (MR-STAR) and the logistic autoregressive conditional root (LACR) model, see, e.g. , Bec et al. (2010) and Bec et al. (2008).

4.A Simulated LSTAR process and logistic transition function

Figure 4.6: Simulated data series (a) and transition function (b) for the LSTAR model (4.1) with $\gamma = 2$, $c = 0$, $\alpha = 0.5$ and $T = 150$.



(a) (b)

4.B Proof of Lemma 4.1

Observe initially that with G_t defined in (4.2), it holds that

$$\frac{\partial G_t}{\partial \gamma} = G_t \left(1 - G_t\right) \left(y_{t-1} - c\right) =: \psi_t \left(y_{t-1} - c\right)$$

and

$$\frac{\partial^2 G_t}{\left(\partial \gamma\right)^2} = \frac{\partial \psi_t}{\partial \gamma} \left(y_{t-1} - c\right) = \psi_t \left(1 - 2G_t\right) \left(y_{t-1} - c\right)^2 .$$

Moreover, as $\gamma \to \infty$, one has $\psi_t \to 0$ and hence ${}^{\partial G_t}/_{\partial \gamma} \to 0$ and ${}^{\partial^2 G_t}/_{(\partial \gamma)^2} \to 0$. In fact, note that all higher order derivatives will have the form

$$\frac{\partial^n G_t}{\left(\partial \gamma\right)^n} = \psi_t g\left(G_t\right) \left(y_{t-1} - c\right)^n$$

where $g\left(G_t\right)$ is a function consisting of an integer and of sums and products of G_t. In particuler, observe that since $0 < G_t < 1$, it holds for any $n < \infty$ that $g\left(G_t\right) = K$ for a constant $K < \infty$. Thus, we have that

$$\frac{\partial^n G_t}{\left(\partial \gamma\right)^n} \to 0 \quad \text{as} \quad \gamma \to \infty. \tag{4.12}$$

Next, consider the likelihood contribution given by $\ell_t\left(\gamma\right)$ in (4.3). Standard calculus gives

$$\frac{\partial \ell_t\left(\gamma\right)}{\partial \gamma} = \varepsilon_t\left(\gamma\right) \alpha y_{t-1} \frac{\partial G_t}{\partial \gamma} = \alpha y_{t-1} y_t \frac{\partial G_t}{\partial \gamma} - \alpha^2 y_{t-1}^2 G_t \frac{\partial G_t}{\partial \gamma} =: a_t\left(\gamma\right) + b_t\left(\gamma\right).$$

Observe that the higher order derivatives of the terms $a_t\left(\gamma\right)$ and $b_t\left(\gamma\right)$ with respect to γ will be of the respective forms

$$\frac{\partial^n a_t\left(\gamma\right)}{\left(\partial \gamma\right)^n} = \alpha y_{t-1} y_t \frac{\partial^n G_t}{\left(\partial \gamma\right)^n} \quad \text{and} \quad \frac{\partial^n b_t\left(\gamma\right)}{\left(\partial \gamma\right)^n} = -\alpha^2 y_{t-1}^2 \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \frac{\partial^k G_t}{\left(\partial \gamma\right)^k} \frac{\partial^{n-k} G_t}{\left(\partial \gamma\right)^{n-k}}.$$

Consequently, it holds by (4.12) that ${}^{\partial^n \ell_t(\gamma)}/_{(\partial \gamma)^n} \to 0$ as $\gamma \to \infty$. Observe that the same result holds for the parametrization with δ since

$$\frac{\partial^n G_t}{\left(\partial \delta\right)^n} = \frac{\partial^n G_t}{\left(\partial \gamma\right)^n} \frac{\partial^n \gamma}{\left(\partial \delta\right)^n} \quad \text{and} \quad \frac{\partial^n \gamma}{\left(\partial \delta\right)^n} = \frac{n!}{\left(\delta - 1\right)^{(n+1)}},$$

where ${}^{\partial^n \gamma}/_{(\partial \delta)^n}$ is function that grows as $\delta \to 1$. However, the grows rate is slower than the exponential decay of ${}^{\partial^n G_t}/_{(\partial \gamma)^n}$ and, hence, we still have,

$$\lim_{\delta \to 1} \frac{\partial^n G_t}{\left(\partial \delta\right)^n} = 0.$$

□

Bibliography

Alizadeh, A. H. and N. K. Nomikos (2004): "Cost of carry, causality and arbitrage between oil futures and tanker freight markets," *Transportation Research Part E: Logistics and Transportation Review*, 40, 297–316.

Andrews, D. W. K. and W. Ploberger (1994): "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

Areosa, W. D., M. McAleer, and M. C. Medeiros (2011): "Moment-based estimation of smooth transition regression models with endogenous variables," *Journal of Econometrics*, 165, 100–111.

Bacon, D. W. and D. G. Watts (1971): "Estimating the Transition Between Two Intersecting Straight Lines," *Biometrika*, 58, 525–534.

Baghli, M. (2005): "Nonlinear Error-Correction Models for the FF/DM Rate," *Studies in Nonlinear Dynamics & Econometrics*, 9, 1–41.

Balke, N. S. and T. B. Fomby (1997): "Threshold Cointegration," *International Economic Review*, 38, 627–645.

Bec, F., M. Ben Salem, and R. MacDonald (2006): "Real exchange rates and real interest rates : a nonlinear perspective," *Recherches économiques de Louvain*, 72, 177.

Bec, F. and A. Rahbek (2004): "Vector equilibrium correction models with non-linear discontinuous adjustments," *Econometrics Journal*, 7, 628–651.

Bec, F., A. Rahbek, and N. Shephard (2008): "The ACR Model: A Multivariate Dynamic Mixture Autoregression," *Oxford Bulletin of Economics and Statistics*, 70, 583–618.

Bec, F., M. B. Salem, and M. Carrasco (2010): "Detecting Mean Reversion in Real Exchange Rates from a Multiple Regime STAR Model," *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, 395–427.

Bec, F., M. B. Salem, and A. Rahbek (2004): "Nonlinear adjustment towards the purchasing power parity relation: a multivariate approach," *Manuscript. CREST, Paris.*

Bose, A. (1988): "Edgeworth Correction by Bootstrap in Autoregressions," *The Annals of Statistics*, 16, 1709–1722.

Boswijk, H. P. and J. A. Doornik (2004): "Identifying, estimating and testing restricted cointegrated systems: An overview," *Statistica Neerlandica*, 58, 440–465.

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

Castle, J. and D. Hendry (2013): "Semi-automatic Non-linear Model Selection," *Essays in Nonlinear Time Series Econometrics, Oxford University Press, forthcoming. Edited by Haldrup, N., Meitz M. and Saikkonen, P.*

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2012): "Bootstrap Determination of the Co-Integration Rank in Vector Autoregressive Models," *Econometrica*, 80, 1721–1740.

Cavaliere, G., A. Rahbek, and A. R. Taylor (2010a): "Cointegration Rank Testing Under Conditional Heteroskedasticity," *Econometric Theory*, 26, 1719–1760.

——— (2010b): "Testing for co-integration in vector autoregressions with non-stationary volatility," *Journal of Econometrics*, 158, 7–24.

Cavaliere, G. and A. R. Taylor (2008): "Bootstrap Unit root Tests for Time Series with Nonstationary Volatility," *Econometric Theory*, 24, 43–71.

Chan, K. S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *The Annals of Statistics*, 21, 520–533.

Chan, K. S. and H. Tong (1986): "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–190.

Clarida, R. H., L. Sarno, M. P. Taylor, and G. Valente (2006): "The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates," *The Journal of Business*, 79, 1193–1224.

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.

Cox, D. R. (1981): "Statistical Analysis of Time Series: Some Recent Developments [with Discussion and Reply]," *Scandinavian Journal of Statistics*, 8, 93–115.

Davidson, R. and E. Flachaire (2008): "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146, 162–169.

Davies, R. B. (1987): "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternatives," *Biometrika*, 74, 33–43.

Enders, W. and P. L. Siklos (2001): "Cointegration and Threshold Adjustment," *Journal of Business & Economic Statistics*, 19, 166–176.

Escanciano, J. C. (2007): "Model checks using residual marked empirical processes." *Statistica Sinica*, 115–138.

Escribano, A. (2004): "Nonlinear error correction: The case of money demand in the United Kingdom (1878-2000)," *Macroeconomic Dynamics*, 8, 76–116.

Fattouh, B. (2010): "The dynamics of crude oil price differentials," *Energy Economics*, 32, 334–342.

——— (2011): "An Anatomy of the Crude Oil Pricing System," Tech. rep., Oxford Institute for Energy Studies.

Gao, J. and P. C. B. Phillips (2011): "Semiparametric Estimation in Multivariate Nonstationary Time Series Models," Monash Econometrics and Business Statistics Working Paper 17/11, Monash University, Department of Econometrics and Business Statistics.

Ghoshray, A. and T. Trifonova (2014): "Dynamic Adjustment of Crude Oil Price Spreads," *Energy Journal*, 35, 119–136.

Gine, E. and J. Zinn (1990): "Bootstrapping General Empirical Measures," *The Annals of Probability*, 18, 851–869.

Granger, C. and T. Teräsvirta (1993): *Modelling nonlinear economic relationships*, Oxford University Press, USA.

Gripenberg, G. (1996): "Computing the joint spectral radius," *Linear Algebra and its Applications*, 234, 43–60.

Hamilton, J. D. (1994): *Time series analysis*, Princeton Univ Pr.

Hammoudeh, S. M., B. T. Ewing, and M. A. Thompson (2008): "Threshold Cointegration Analysis of Crude Oil Benchmarks," *The Energy Journal*, 29, 79–95.

Hannan, E. J. and B. G. Quinn (1979): "The determination of the order of an autoregression," *Journal of the Royal Statistical Society. Series B (Methodological)*, 190–195.

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

——— (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

——— (1997): "Inference in TAR Models," *Studies in Nonlinear Dynamics & Econometrics*, 2, 1.

Hansen, B. E. and B. Seo (2002): "Testing for two-regime threshold cointegration in vector error-correction models," *Journal of Econometrics*, 110, 293–318.

Heinen, F., S. Michael, and P. Sibbertsen (2012): "Weak identification in the ESTAR model and a new model," *Journal of Time Series Analysis*.

Horowitz, J. L. (2001): "Chapter 52 The Bootstrap," in *Handbook of Econometrics*, ed. by J.J. Heckman and E. Leamer, Elsevier, vol. Volume 5, 3159–3228.

Jensen, S. T. and A. Rahbek (2007): "On the Law of Large Numbers for (geometrically) Ergodic Markov Chains," *Econometric Theory*, 23, 761–766.

Johansen, S. (1996): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models (Advanced Texts in Econometrics)*, Oxford University Press, USA.

Jungers, R. (2009): *The Joint Spectral Radius: Theory and Applications*, Springer.

Juselius, K. (2006): "The Cointegrated VAR Model: Methodology and Applications," *Oxford University Press*.

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

Karlsen, H. A., T. Myklebust, and D. Tjøstheim (2007): "Nonparametric estimation in a nonlinear cointegration type model," *The Annals of Statistics*, 35, 252–299.

Kiliç, R. (2011): "Testing for co-integration and nonlinear adjustment in a smooth transition error correction model," *Journal of Time Series Analysis*, 32, 647–660.

Kreiss, J.-P. and E. Paparoditis (2011): "Bootstrap methods for dependent data: A review," *Journal of the Korean Statistical Society*, 40, 357–378.

Kristensen, D. and A. Rahbek (2010): "Likelihood-based inference for cointegration with nonlinear error-correction," *Journal of Econometrics*, 158, 78–94.

——— (2013): "Testing and inference in nonlinear cointegrating vector error correction models," *Econometric Theory*, 29, 1238–1288.

Lahiri, S. N. (2003): *Resampling Methods for Dependent Data*, Springer.

Lange, T., A. Rahbek, and S. T. Jensen (2011): "Estimation and Asymptotic Inference in the AR-ARCH Model," *Econometric Reviews*, 30, 129–153.

Liao, H.-C., S.-C. Lin, and H.-C. Huang (2014): "Are crude oil markets globalized or regionalized? Evidence from WTI and Brent," *Applied Economics Letters*, 21, 235–241.

Liebscher, E. (2005): "Towards a Unified Approach for Proving Geometric Ergodicity and Mixing Properties of Nonlinear Autoregressive Processes," *Journal of Time Series Analysis*, 26, 669–689.

Liu, R. Y. (1988): "Bootstrap Procedures under some Non-I.I.D. Models," *The Annals of Statistics*, 16, 1696–1708.

Lo, M. C. and E. Zivot (2001): "Theshold Cointegration and Nonlinear Adjustment to the law of one price," *Macroeconomic Dynamics*, 5, 533–576.

Lof, M. (2012): "Heterogeneity in stock prices: A STAR model with multivariate transition function," *Journal of Economic Dynamics and Control*, 36, 1845–1854.

Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*, Springer-Verlag.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

Magnus, J. R. and H. Neudecker (1999): *Matrix differential calculus with applications in statistics and econometrics*, New York: John Wiley.

Mammen, E. (1993): "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *The Annals of Statistics*, 21, 255–285.

Mann, J. (2012): "Threshold Cointegration with Applications to the Oil and Gasoline Industry," Ph.D. thesis, Queen's University.

Meyn, S. and R. Tweedie (1993): *Markov chains and stochastic stability*, Springer-Verlag, London, available at: probability.ca/MT.

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.

Psaradakis, Z., M. Sola, F. Spagnolo, and N. Spagnolo (2009): "Selecting nonlinear time series models using information criteria," *Journal of Time Series Analysis*, 30, 369–394.

Saikkonen, P. (2005): "Stability results for nonlinear error correction models," *Journal of Econometrics*, 127, 69–81.

——— (2008): "Stability of Regime Switching Error Correction Models Under Linear Cointegration," *Econometric Theory*, 24, 294–318.

Schwarz, G. (1978): "Estimating the dimension of a model," *The annals of statistics*, 6, 461–464.

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

Stute, W. (1997): "Nonparametric Model Checks for Regression," *The Annals of Statistics*, 25, 613–641.

Stute, W., W. G. Manteiga, and M. P. Quindimil (1998): "Bootstrap Approximations in Model Checks for Regression," *Journal of the American Statistical Association*, 93, 141–149.

Teräsvirta, T. (1994): "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of American Statistical Association*, 89, 208–218.

——— (1998): "Modelling economic relationships with smooth transition regressions," *Handbook of Applied Economic Statistics, Marcel Dekker: New York*, 507–552.

Teräsvirta, T., D. Tjøstheim, and C. Granger (2010a): *Modelling Nonlinear Economic Time Series*, Advanced Texts in Econometrics Series, Oxford University Press.

Teräsvirta, T., D. Tjøstheim, and C. W. J. Granger (2010b): *Modelling Nonlinear Economic Time Series*, Oxford University Press.

Tong, H. (2011): "Threshold models in time series analysis–30 years on," *Statistics and its Interface*, 4, 107–118.

Tong, H. and K. S. Lim (1980): "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 245–292.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes - With Applications to Statistics*, Springer-Verlag.

van Dijk, D., T. Teräsvirta, and P. H. Franses (2002): "Smooth transition autoregressive models - a survey of recent developments," *Economic Reviews*, 21, 1–47.

Wu, C. F. J. (1986): "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *The Annals of Statistics*, 14, 1261–1295.