Statistical inference for stochastic coefficient regression models

Suhasini Subba Rao
Email: suhasini.subbarao@stat.tamu.edu

May 6, 2009

Abstract

The classical multiple regression model plays a very important role in statistical analysis. The typical assumption is that changes in the response variable, due to a small change in a given regressor, is constant over time. In other words, the rate of change is not influenced by any unforeseen external variables and remains the same over the entire time period of observation. This strong assumption may, sometimes, be unrealistic, for example, in areas like social sciences, environmental sciences etc. In view of this, we propose stochastic coefficient regression (SCR) models with stationary, correlated random errors and consider their statistical inference. We assume that the coefficients are stationary processes, where each admits a linear process representation. We propose a frequency domain method of estimation, the advantage of this method is that no assumptions on the distribution of the coefficients are necessary. We illustrate the methodology with simulations and compare their performance. These models are fitted to two real data sets and their predictive performance are also examined.

Keywords and phrases Gaussian maximum likelihood, frequency domain, locally stationary time series, multiple linear regression, nonstationarity, stochastic coefficients.

1 Introduction

The classical multiple linear regression model is ubiquitous in many fields of research. However, in situations where the response variable $\{Y_t\}$ is observed over time, it is not always possible to assume that the influence the regressors $\{x_{t,j}\}$ exert on the response Y_t is constant over time. A classical example, given in Burnett and Guthrie (1970), is when predicting air quality as a function of pollution emission. The influence the emissions have on air quality on any given day may depend on various factors such as the meterological conditions on the current and previous days. Modelling the variable influence in a deterministic way can be too complex and a simpler method could be to treat the regression coefficients as stochastic. In order to allow for the influence of the previous regression coefficient on the current coefficient, it is often reasonable to assume that the underlying unobservable regression coefficients are stationary processes and each coefficient admits a linear process representation. In other words, a plausible model for modelling the varying influence of regressors on the response variable is

$$Y_t = \sum_{j=1}^{n} (a_{j,0} + \alpha_{t,j}) x_{t,j} + \varepsilon_t = \sum_{j=1}^{n} a_{j,0} x_{t,j} + X_t, \tag{1}$$

where $\{x_{t,j}\}$ are the deterministic regressors, $\{a_{j,0}\}$ are the mean regressor coefficients, $\mathbb{E}(X_t) = 0$ and satisfies $X_t = \sum_{j=1}^{n} \alpha_{t,j} x_{t,j} + \varepsilon_t$, $\{\varepsilon_t\}$ and $\{\alpha_{t,j}\}$ are jointly stationary linear time series with $\mathbb{E}(\alpha_{t,j}) = 0$, $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\alpha_{t,j}^2) < \infty$ and $\mathbb{E}(\varepsilon_t^2) < \infty$. We observe that this model includes the classical multiple regression model as a special case, with $\mathbb{E}(\alpha_{t,j}) = 0$ and $\text{var}(\alpha_{t,j}) = 0$. The above model is often refered to as a stochastic coefficient regression (SCR) model. Such models have a long history in statistics (see Hildreth and Houck (1968) and Swamy (1970, 1971), Burnett and Guthrie (1970), Rosenberg (1972, 1973), Duncan and Horn (1972), Fama (1977), Bruesch and Pagan (1980), Swamy and Tinsley (1980), Synder (1985), Pfeffermann (1984), Newbold and Bos (1985), Stoffer and Wall (1991) and Franke and Gründer (1995). For a review of this model the reader is refered to Newbold and Bos (1985). A comparision of the SCR model with other statistical models is given Section 2 and a test for randomness of the stochastic coefficients against the alternative that they are correlated is considered in Section 4.

In the aforementioned literature, it is usually assumed that $\{\alpha_{t,j}\}$ satisfies a parametric linear time series model and the Gaussian maximum likelihood (GML) is used to estimate the unknown parameters. In the case $\{Y_t\}$ is Gaussian, the estimators are asymptotically normal and the variance of these estimators can be obtained from the inverse of the Information matrix. Even in the situation $\{Y_t\}$ is non-Gaussian, the Gaussian likelihood is usually used as the objective function to be maximised, in this case the objective function is often called the quasi-Gaussian likelhood (quasi-GML). The quasi-GML estimator is a consistent estimate of the parameters (see Ljung and Caines (1979), Caines (1988), Chapter 8.6 and Shumway and Stoffer (2006)), but when $\{Y_t\}$ is non-Gaussian, obtaining an expression for the standard errors of the quasi-GML estimators seems to be almost imposible. Therefore implicitly it is usually assumed that $\{Y_t\}$ is Gaussian, and most statistical inference is based on the assumption of Gaussianity. In several situations the assumption of Gaussianity may not be plausible, and there is a need for estimators which are free of distributional assumptions. In this paper we address this issue.

In Section 3, two methods to estimate the mean regression parameters and the finite number of parameters which are characterising the impulse response sequences of the linear processes are considered. The suggested methods are closely related and are based within the frequency domain, this is because spectral methods usually don't require distributional assumptions, are computationally fast, and can be analysed asymptotically (see Whittle (1962), Walker (1964), Dzhapharidze (1971), Hannan (1971, 1973), Dunsmuir (1979), Taniguchi (1983), Giraitis and Robinson (2001), Dahlhaus (2000) and Shumway and Stoffer (2006)). Both of the proposed methods offer an alternative perspective of the SCR model based within the frequency domain, and are free of any distributional assumptions. In Section 5.2 and 5.3 we consider the asymptotic properties of the estimators proposed. A theoretical comparison of our estimators with the GML estimator, in most cases, is not possible, this is because it is usually not possible to obtain the asymptotic variance of the GML estimator. However, if we consider a subclass of SRC models, where the regressors are smooth, then the asymptotic variance of the GML estimator can be derived. And in Section 5.4 we compare our frequency domain estimator with the GML estimator, for the subclass SCR models with smooth regressors, and show that both estimators have asymptotically equivalent distributions.

In Section 6 we consider some simulations and two real data sets. The two real data are taken from the field of economics and evironmental sciences. In the first case, the SCR model is used to examine the relationship between monthly inflation and nominal T-bills interest rates, where monthly inflation is the response and the T-bills rate is the regressor. We confirm the findings off Newbold and Bos (1985), who observe that the regression coefficient is stochastic. In the second case, we consider the influence man made emissions (the regressors) have on particulate matter (the reponse variable) in Shenandoah National Park, U.S.A. Typically, it is assumed that man-made emissions linearly influence the amount of particulate matter and a multiple linear regression model is fitted to the data. We show that there is clear evidence to suggest that the regression coefficients are random, hence the dependence between man-made emissions and particulate matter is more complicated than previously thought.

The proofs can be found in the appendix.

2 The stochastic coefficient regression model

2.1 The model

Throughout this paper we will assume that the response variable $\{Y_t\}$ satisfies (1), where the regressors $\{x_{t,j}\}$ are observed and the following assumptions are satisfied.

Assumption 2.1 *(i) The stationary time series $\{\alpha_{t,j}\}$ and $\{\varepsilon_t\}$ satisfy the following MA(∞) representations*

$$\alpha_{t,j} = \sum_{i=0}^{\infty} \psi_{i,j} \eta_{t-i,j}, \quad \text{for } j = 1, \ldots, n, \quad \varepsilon_t = \sum_{i=0}^{\infty} \psi_{i,n+1} \eta_{t-i,n+1}, \tag{2}$$

where for all $1 \leq j \leq n+1$, $\sum_{i=0}^{\infty} |\psi_{i,j}| < \infty$, $\sum_{i=0}^{\infty} |\psi_{i,j}|^2 = 1$, $\mathbb{E}(\eta_{t,j}) = 0$, $\mathbb{E}(\eta_{t,j}^2) = \sigma_{j,0}^2 < \infty$, $\{\eta_{t,j}\}$ are independent, identically distributed (iid) variables over t and j.

The parameters $\{\psi_{i,j}\}$ are unknown but have a parametric form, that is there is a known function $\psi_{i,j}(\cdot)$, such that for some vector $\boldsymbol{\theta}_0 = (\boldsymbol{\vartheta}_0, \boldsymbol{\Sigma}_0)$, $\psi_{i,j}(\boldsymbol{\vartheta}_0) = \psi_{i,j}$ and $\boldsymbol{\Sigma}_0 = diag(\sigma_{1,0}^2, \ldots, \sigma_{n+1,0}^2) = var(\underline{\eta}_t)$ where $\underline{\eta}_t = (\eta_{t,1}, \ldots, \eta_{t,n+1})$.

(ii) We define the compact parameter spaces Ω, Θ_1 and Θ_2. Ω contains n-dimensional vectors, Θ_1 contains q-dimensional vectors and Θ_2 contains $(n+1) \times (n+1)$-dimensional diagonal, matrices, moreover for every q-dimensional vector $\boldsymbol{\vartheta} \in \Theta_1$, we have $\sum_{i=0}^{\infty} |\psi_{i,j}(\boldsymbol{\vartheta})|^2 = 1$. We shall assume that $\boldsymbol{a}_0$, $\boldsymbol{\vartheta}_0$ and $\boldsymbol{\Sigma}_0$ lie in the interior of Ω, Θ_1 and Θ_2 respectively.

We define the transfer function $A_j(\boldsymbol{\vartheta}, \omega) = (2\pi)^{-1/2} \sum_{k=0}^{\infty} \psi_{k,j}(\boldsymbol{\vartheta}) \exp(ik\omega)$, and the spectral density $f_j(\boldsymbol{\vartheta}, \omega) = |A_j(\boldsymbol{\vartheta}, \omega)|^2$. Using the above notation the spectrum of the time series $\{\alpha_{t,j}\}$ is $\sigma_{j,0}^2 \cdot f_j(\boldsymbol{\vartheta}_0, \omega)$. Let $c_j(\boldsymbol{\theta}, t-\tau) = \sigma_j^2 \int f_j(\boldsymbol{\vartheta}, \omega) \exp(i(t-\tau)\omega) d\omega$. Hence $cov(\alpha_{t,j}, \alpha_{\tau,j}) = c_j(\boldsymbol{\theta}_0, t-\tau)$.

It should be noted that it is straightforward to generalise (2), such that the vector time series $\{\boldsymbol{\alpha}_t = (\alpha_{t,1}, \ldots, \alpha_{t,n})\}_t$ has a vector MA(∞) representation. However, using this generalisation makes the notation quite cumbersome. For this reason we have considered the simpler case (2).

Example 2.1 *The SCR model where the stochastic coefficients and error $\{\alpha_{t,j}\}$ and ε_t satisfy an autoregressive model are examples of processes which satisfy Assumption 2.1. That is $\{\alpha_{t,j}\}$ and ε_t satisfy*

$$\alpha_{t,j} = \sum_{k=1}^{p_j} \phi_{k,j} \alpha_{t-k,j} + \eta_{t,j} \quad j = 1, \ldots, n \quad \text{and} \quad \varepsilon_t = \sum_{k=1}^{p_{n+1}} \phi_{k,n+1} \varepsilon_{t-k} + \eta_{t,n+1},$$

where $\eta_{t,j}$ are iid random variables with $\mathbb{E}(\eta_{t,j}) = 0$ and $var(\eta_{t,j}) = \rho_j^2$ and the roots of the characteristics polynomial $1 - \sum_{k=1}^{p_j} \phi_{k,j} z^k$ lie outside unit circle. In this case, the true parameters are $\boldsymbol{\vartheta}_0 = (\phi_{1,1}, \ldots, \phi_{p_{n+1},n+1})$ and $\boldsymbol{\Sigma}_0 = diag(\frac{\rho_1^2}{(\int g_1(\omega)d\omega)}, \ldots, \frac{\rho_{n+1}^2}{(\int g_{n+1}(\omega)d\omega)})$, where $g_j(\omega) = \frac{1}{2\pi}|1 - \sum_{k=1}^{p_j} \phi_{k,j}\exp(ik\omega)|^{-2}$.

We mention that identification of $\boldsymbol{\theta}_0$ is only possible if there does not exist another $\boldsymbol{\theta}^* \in \Theta_1 \otimes \Theta_2$ such that for all t and τ, $\sum_{j=1}^{n} x_{t,j} x_{\tau,j} c_j(\boldsymbol{\theta}_0, t-\tau) + c_{n+1}(\boldsymbol{\theta}_0, t-\tau) = \sum_{j=1}^{n} x_{t,j} x_{\tau,j} c_j(\boldsymbol{\theta}^*, t-\tau) + c_{n+1}(\boldsymbol{\theta}^*, t-\tau)$.

2.2 A comparision of the SCR model with other statistical models

In this section we show that the SCR model is closely related to several popular statistical models. Of course, the SCR model includes the multiple linear regression model as a special case, with $\mathrm{var}(\alpha_{t,j}) = 0$ and $\mathbb{E}(\alpha_{t,j}) = 0$.

2.2.1 Varying coefficient models

In several applications, linear regression models with time-dependent parameters are fitted to the data. Examples include the varying-coefficient models considered by Martinussen and Scheike (2000), where $\{Y_t\}$ satisfies

$$Y_t = \sum_{j=1}^{n} \alpha_j(\frac{t}{T}) x_{t,j} + \varepsilon_t, \quad t = 1, \ldots, T \tag{3}$$

and $\{\alpha_j(\cdot)\}$ are smooth, unknown functions and $\{\varepsilon_t\}_t$ are iid random variables with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathrm{var}(\varepsilon_t) < \infty$. Comparing this model with the SCR model, we observe that the difference between the two models lies in the modelling of the time-dependent coefficients. In (3) the coefficient of the regressor is assumed to be deterministic, whereas the SCR model treats the coefficient as a stationary time series.

2.2.2 Locally stationary time series

In this section we show that a subclass of SCR models, where the regressors are slowly varying, and the class of locally stationary processes defined in Dahlhaus (1996) are closely related. To indicate that the regressors are smooth we will suppose that there exists smooth functions $\{x_j(\cdot)\}$ such that the regressors satisfy $x_{t,j} = x_j(\frac{t}{N})$ for some value N (setting $\frac{1}{T}\sum_t x_{t,j}^2 = 1$) and $\{Y_{t,N}\}$ satisfies

$$Y_{t,N} = \sum_{j=1}^{n} a_{j,0} x_j(\frac{t}{N}) + X_{t,N}, \quad \text{where} \quad X_{t,N} = \sum_{j=1}^{n} \alpha_{t,j} x_j(\frac{t}{N}) + \varepsilon_t \quad t = 1, \ldots, T. \tag{4}$$

Roughly speaking, a locally stationary process is a time series which is nonstationary, but in any local neighbourhood can be approximated by a stationary process. We now show that $\{Y_{t,N}\}$, based on the above definition, can be considered as a locally stationary time series.

Proposition 2.1 *Suppose Assumption 2.1(i,ii) is satisfied, and let us suppose the regressors satisfy $\sup_{j,v} |x_j(v)| < \infty$, $Y_{t,N}$ is defined as in (4) and let*

$$Y_t(v) = \sum_{j=1}^{n} a_{j,0} x_j(v) + X_t(v), \quad \textit{where} \quad X_t(v) = \sum_{j=1}^{n} \alpha_{t,j} x_j(v) + \varepsilon_t \quad t = 1, \ldots, T. \tag{5}$$

Then for v kept fixed, $\{Y_t(v)\}$ is a stationary time series. Moreover we have $|Y_{t,N} - Y_t(v)| = O_p(|\frac{t}{N} - v|)$ and $|X_{t,N} - X_t(v)| = O_p(|\frac{t}{N} - v|)$.

PROOF. The proof is straightforward hence we omit the details. □

We now show the converse, that is the class of locally stationary linear processes defined in Dahlhaus (1996), can be approximated by an SCR model with slowly varying regressors. We first define the locally stationary linear time series

$$\tilde{X}_{t,N} = \int A_{t,N}(\omega)\exp(it\omega)dZ(\omega) \tag{6}$$

where $\{Z(\omega)\}$ is a complex valued orthogonal process on $[0, 2\pi]$ with $Z(\lambda+\pi) = Z(\lambda)$, $\mathbb{E}(Z(\lambda)) = 0$, and $\mathbb{E}\{dZ(\lambda)dZ(\mu)\} = \eta(\lambda+\nu)d\lambda d\mu$, $\eta(\lambda) = \sum_{j=-\infty}^{\infty}\delta(\lambda+2\pi j)$ is the periodic extension of the Dirac delta function. Furthermore, there exists a Lipschitz continuous function $A(\cdot)$, such that $\sup_{\omega,t}|A(\frac{t}{N},\omega) - A_{t,N}(\omega)| \leq KN^{-1}$, where K is a finite constant which does not depend on N.

Proposition 2.2 *Let us suppose that $\{\tilde{X}_{t,N}\}$ is a locally stationary process, which satisfies (6) and $\sup_u \int |A(u,\lambda)|^2 du < \infty$. Then for any basis $\{x_j(\cdot)\}$ of $L_2[0,1]$, and for every δ there exists an n_δ, such that*

$$\tilde{X}_{t,N} = \sum_{j=1}^{n_\delta}\alpha_{t,j}x_j(\frac{t}{N}) + O_p(\delta + N^{-1}), \tag{7}$$

where $\{\boldsymbol{\alpha}_t\} = \{(\alpha_1, \ldots, \alpha_{t,n_\delta})\}_t$ is a second order stationary vector time series.

PROOF. See Appendix A.1. □

Proposition 2.2 implies that class of locally stationary linear processes defined in Dahlhaus (1997) can be approximated by an SCR model, where the regressors are basis functions of $L_2[0,1]$. Therefore, if the covariance structure of a time series is varying slowly over time it is possible to fit an SCR model, with smooth regressors, to the time series. However, the choice of n_δ, is an interesting issue, but in view of the mathematical analysis required will not be considered in this paper.

3 Frequency domain estimators

We now consider two methods to estimate the mean regression parameters $\boldsymbol{a}_0 = (a_{1,0}, \ldots, a_{n,0})$ and the parameters $\boldsymbol{\theta}_0$ in the time series model (defined in Assumption 2.1).

3.1 Motivating the objective function

Frequency domain analyis has a number of advantages (see (Brillinger, 2001)), for example it usually does not require any distributional assumptions and can lead to estimators which are easier to evaluate. To motivate the objective function consider the 'localised' finite Fourier transform of $\{Y_t\}_t$ centered at t, that is $J_{Y,t}(\omega) = \frac{1}{\sqrt{2\pi m}}\sum_{k=1}^{m} Y_{t-m/2+k}\exp(ik\omega)$ (where m is even). It is clear we can partition $J_{Y,t}(\omega)$ into the sum of deterministic and stochastic terms, $J_{Y,t}(\omega) = \sum_{j=1}^{n} a_{j,0}J_{t,m}^{(j)}(\omega) + J_{X,t}(\omega)$, where $J_{t,m}^{(j)}(\omega) = \frac{1}{\sqrt{2\pi m}}\sum_{k=1}^{m} x_{t-m/2+k,j}\exp(ik\omega)$, $J_{X,t}(\omega) = \frac{1}{\sqrt{2\pi m}}\sum_{k=1}^{m} X_{t-m/2+k}\exp(ik\omega)$ and (Y_t, X_t) are defined in (1). Let us consider the Fourier transform $J_{Y,t}(\omega)$ at the fundemental frequencies $\omega_k = \frac{2\pi k}{m}$ and define the $m(T-m)$-dimensional vectors $\mathcal{J}_{Y,T} = (J_{Y,m/2}(\omega_1), \ldots, J_{Y,T-m/2}(\omega_m))$ and $\mathcal{J}_{x,T}(\boldsymbol{a}_0) = (\sum_{j=1}^m a_{j,0} J_{m/2,m}^{(j)}(\omega_1), \ldots, \sum_{j=1}^m a_{j,0} J_{T-m/2,m}^{(j)}(\omega_m))$. Using heuristic arguments it can be argued that for large m, $\mathcal{J}_{Y,T}$ has a multivariate complex normal distribution. Therefore, minus the logarithm of the density of $\mathcal{J}_{Y,T}$ is approximately proportional to

$$\ell(\boldsymbol{\theta}_0) = ((\mathcal{J}_{Y,T} - \mathcal{J}_{x,T}(\boldsymbol{a}_0))^H)\Delta(\boldsymbol{\theta}_0)^{-1}((\mathcal{J}_{Y,T} - \mathcal{J}_{x,T}(\boldsymbol{a}_0)) + \log(\det\Gamma(\boldsymbol{\theta}_0)),$$

where $\Delta(\boldsymbol{\theta}_0) = \Delta(\boldsymbol{\theta}_0) = \mathbb{E}((\mathcal{J}_{Y,T} - \mathcal{J}_{x,T}(\boldsymbol{a}_0)(\mathcal{J}_{Y,T} - \mathcal{J}_{x,T}(\boldsymbol{a}_0))^H)$ and H denotes the transpose and complex conjugate (see Picinbono (1996), equation (17)). However, evaluating $\ell(\boldsymbol{\theta}_0)$ involves inverting $\Delta(\boldsymbol{\theta}_0)$, which is difficult due to its large dimension (moreover $\Delta(\boldsymbol{\theta}_0)$ is likely to be singular). Hence it is an unsuitable criterion for estimating the parameters $\boldsymbol{a}_0$ and $\boldsymbol{\theta}_0$. Instead let us consider a related criterion, where we ignore the off-diagonal covariances in $\Delta(\boldsymbol{\theta}_0)$ and replace $\Delta(\boldsymbol{\theta}_0)$ in $\ell(\boldsymbol{\theta}_0)$ with a diagonal matrix which shares the same diagonal as $\Delta(\boldsymbol{\theta}_0)$. Straightforward calculations show that when the $\Delta(\boldsymbol{\theta}_0)$ in $\ell(\boldsymbol{\theta}_0)$ is replaced by its diagonal, what remains is proportional to

$$\tilde{\ell}(\boldsymbol{\theta}_0) = \frac{1}{(T-m)m} \sum_{t=m/2}^{T-m/2} \sum_{k=1}^{m} \left(\frac{|J_{Y,t}(\omega_k) - \sum_{j=1}^n a_{j,0} J_{t,m}^{(j)}(\omega_k)|^2}{\mathcal{F}_{t,m}(\boldsymbol{\theta}_0, \omega_k)} + \log \mathcal{F}_{t,m}(\boldsymbol{\theta}_0, \omega_k) \right), \tag{8}$$

where, under Assumption 2.1,

$$\begin{aligned} \mathcal{F}_{t,m}(\boldsymbol{\theta}_0, \omega) &= \frac{1}{2\pi} \sum_{r=-(m-1)}^{m-1} \exp(ir\omega) \sum_{j=1}^{n+1} c_j(\boldsymbol{\theta}_0, r) \frac{1}{m} \sum_{k=1}^{m-|r|} x_{t-m/2+k,j} x_{t-m/2+k+r,j} \\ &= \sum_{j=1}^{n} \sigma_{j,0}^2 \int_{-\pi}^{\pi} I_{t,m}^{(j)}(\lambda) f_j(\boldsymbol{\vartheta}_0, \omega - \lambda) d\lambda + \sigma_{n+1,0}^2 \int_{-\pi}^{\pi} I_m^{(n+1)}(\lambda) f_{n+1}(\boldsymbol{\vartheta}_0, \omega - \lambda) d\lambda, \end{aligned} \tag{9}$$

letting $x_{t,n+1} = 1$ for all t, $I_{t,m}^{(j)}(\omega) = |J_{t,m}^{(j)}(\omega)|^2$ and $I_m^{(n+1)}(\omega) = \frac{1}{2\pi m} |\sum_{k=1}^m \exp(ik\omega)|^2$.

3.2 Estimator 1

We now use (8) as the basis for the objective function in the first estimator. Replace the true parameters in (8) with $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{n+1}^2)$, $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\Sigma})$ (recall the notation in Assumption 2.1) and replace the summand $\frac{1}{m}\sum_{k=1}^m$ with an integral. This gives the objective function

$$\mathcal{L}_T^{(m)}(\boldsymbol{a}, \boldsymbol{\theta}) = \frac{1}{T_m} \sum_{t=m/2}^{T-m/2} \int_{-\pi}^{\pi} \left\{ \frac{\mathcal{I}_{t,m}(\boldsymbol{a}, \omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta}, \omega)} + \log \mathcal{F}_{t,m}(\boldsymbol{\theta}, \omega) \right\} d\omega, \tag{10}$$

where $T_m = T - m$, m is even and

$$\mathcal{I}_{t,m}(\boldsymbol{a}, \omega) = \frac{1}{2\pi m} \left| \sum_{k=1}^{m} (Y_{t-m/2+k} - \sum_{j=1}^{n} a_j x_{t-m/2+k,j}) \exp(ik\omega) \right|^2. \tag{11}$$

We recall that $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\Sigma})$, hence $\mathcal{L}_T^{(m)}(\boldsymbol{a}, \boldsymbol{\theta}) = \mathcal{L}_T^{(m)}(\boldsymbol{a}, \boldsymbol{\vartheta}, \boldsymbol{\Sigma})$. Let $\boldsymbol{a} \in \Omega \subset \mathbb{R}^n$ and $\boldsymbol{\theta} \in \Theta_1 \otimes \Theta_2 \subset \mathbb{R}^{n+q+1}$. We use $\hat{\boldsymbol{a}}_T$ and $\hat{\boldsymbol{\theta}}_T = (\hat{\boldsymbol{\vartheta}}_T, \hat{\boldsymbol{\Sigma}}_T)$ as an estimator of $\boldsymbol{a}_0$ and $\boldsymbol{\theta}_0 = (\boldsymbol{\vartheta}_0, \boldsymbol{\Sigma}_0)$ where

$$(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\vartheta}}_T, \hat{\boldsymbol{\Sigma}}_T) = \arg \inf_{\boldsymbol{a} \in \Omega, \boldsymbol{\vartheta} \in \Theta_1, \boldsymbol{\Sigma} \in \Theta_2} \mathcal{L}_T^{(m)}(\boldsymbol{a}, \boldsymbol{\vartheta}, \boldsymbol{\Sigma}). \tag{12}$$

We choose m, such that $T_m/T \to 1$ as $T \to \infty$. Hence, m can be fixed, or grow at a rate slower than T.

3.3 Estimator 2

In the case that the number of regressors is relatively large, the minimisation of $\mathcal{L}_T^{(m)}$ can be computationally slow. We recall that the objective function $\mathcal{L}_T^{(m)}$ is evaluated at $\boldsymbol{a}$, $\boldsymbol{\theta} = (\boldsymbol{\Sigma}, \boldsymbol{\vartheta})$. We now suggest a second estimator, which is based on Estimator 1, but estimates the parameters $\boldsymbol{a}$, $\boldsymbol{\theta} = (\boldsymbol{\Sigma}, \boldsymbol{\vartheta})$ in two steps. Empirical studies suggests that it tends to yield a better estimator than Estimator 1. In the first step of the scheme we estimate $\boldsymbol{\Sigma}_0$ and in the second step obtain an estimator of $\boldsymbol{a}_0$ and $\boldsymbol{\vartheta}_0$, thereby reducing the total number parameters to be estimated at each step. An additional advantage of estimating the variance of the coefficients in the first stage is that we can determine whether a coefficient of a regressor is fixed or random.

The two-step parameter estimation scheme

(i) **Step 1** In the first step of the scheme we only estimate the variance $\boldsymbol{\Sigma}_0 = \text{diag}(\sigma_{1,0}, \ldots, \sigma_{n+1,0})$ and the mean coefficients $\boldsymbol{a}_0$. We construct the estimator as if the stochastic coefficients $\{\alpha_{t,j}\}$ and errors $\{\varepsilon_t\}$ were independent identically distributed Gaussian random variables, using this we estimate $\boldsymbol{\Sigma}_0$ (we note that iid and Gaussianity of $\{\alpha_{t,j}\}$ and $\{\varepsilon_t\}$ is not required, it is simply used to construct the objective function). Under these conditions

$$(Y_1, \ldots, Y_T) \sim MVN_T\Big((\sum_{j=1}^{n} a_j x_{1,j}, \ldots, \sum_{j=1}^{n} a_j x_{T,j}), \text{diag}(\sigma_1(\boldsymbol{\Sigma}_0), \ldots, \sigma_T(\boldsymbol{\Sigma}_0)) \Big),$$

where $\text{var}(Y_t) = \sigma_t(\boldsymbol{\Sigma}_0) = \sum_{j=1}^{n} \sigma_{j,0}^2 x_{t,j}^2 + \sigma_{n+1,0}^2$. In this case the log density of $(Y_1, \ldots, Y_T)$ is proportional to

$$\mathcal{L}_T(\boldsymbol{a}_0, \boldsymbol{\Sigma}_0) = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{(Y_t - \sum_{j=1}^{n} a_{j,0} x_{t,j})^2}{\sigma_t(\boldsymbol{\Sigma}_0)} + \log \sigma_t(\boldsymbol{\Sigma}_0) \right). \quad (13)$$

Replacing the true parameters in $\mathcal{L}_T(\boldsymbol{a}_0, \boldsymbol{\Sigma}_0)$ with $(\boldsymbol{a}, \boldsymbol{\Sigma})$ leads to to the objective function $\mathcal{L}_T(\boldsymbol{a}, \boldsymbol{\Sigma})$, where $\sigma_t(\boldsymbol{\Sigma}) = \sum_{j=1}^{n} \sigma_j^2 x_{t,j}^2 + \sigma_{n+1}^2$. Let $(\bar{\boldsymbol{a}}_T, \tilde{\boldsymbol{\Sigma}}_T) = \arg\min_{\boldsymbol{a} \in \Omega, \boldsymbol{\Sigma} \in \Theta_2} \mathcal{L}_T(\boldsymbol{a}, \boldsymbol{\Sigma})$.

(ii) **Step 2** We now use $\tilde{\boldsymbol{\Sigma}}_T$ to estimate $\tilde{\boldsymbol{a}}_T$ and $\boldsymbol{\vartheta}_0$. We substitute $\tilde{\boldsymbol{\Sigma}}_T$ into $\mathcal{L}_T^{(m)}$, keep $\tilde{\boldsymbol{\Sigma}}_T$ fixed and minimise $\mathcal{L}_T^{(m)}$ with respect to $(\boldsymbol{a}, \boldsymbol{\vartheta})$. We use as parameter estimates of $(\boldsymbol{a}_0, \boldsymbol{\vartheta}_0)$ $(\tilde{\boldsymbol{a}}_T, \tilde{\boldsymbol{\vartheta}}_T)$ where

$$(\tilde{\boldsymbol{a}}_T, \tilde{\boldsymbol{\vartheta}}_T) = \arg \min_{\boldsymbol{a} \in \Omega, \boldsymbol{\vartheta} \in \Theta_1} \mathcal{L}_T^{(m)}(\boldsymbol{a}, \boldsymbol{\vartheta}, \tilde{\boldsymbol{\Sigma}}_T). \quad (14)$$

We choose m, such that $T_m/T \to 1$ as $T \to \infty$.

3.4 Practical issues

In Section 5 we obtain the asymptotic sampling properties of the two estimators described above. We first consider some practical issues related to the two estimators.

The Newton Raphson method can be used for the minimisation purposes. In order to facilitate the iterative procedure, in Appendix A.1 we give explicit expressions for the derivatives of the criterion with respect to the parameters. Since the procedure is nonlinear, the choice of initial values is vital. Recalling that the SCR model can be written as a multiple linear regression model with heteroscadastic, dependent errors (see (1)), the mean regression parameters $\boldsymbol{a}_0$ can be estimated using ordinary least square. We found that the ordinary least square estimates serve as best initial values for the mean parameter coefficient here. The choice of initial values for the other set of parameters is not so obvious. But, it seems that final convergent values seem to be robust for initial choice. Choice of 'm' is also an important part of the algorithm. Possible choices for m are discussed in Section 5.4. However, evidence from simulations suggests that there is not a great difference in the estimates for different values of m (see Section 6).

Remark 3.1 (Estimation with missing observations) *There may be some instances when the response variable is not observed. We can still use the above criterion by introducing dummy variables as below. Define the indicator variable* δ_t *such that*

$$\delta_t = \begin{cases} 1 & \text{response } Y_t \text{ observed at time } t \\ 0 & \text{response } Y_t \text{ unobserved at time } t \end{cases}.$$

We now redefine the estimator using the response $\{Y_t\delta_t\}$ *and regressors* $\{(x_{t,1}\delta_t, \ldots, x_{t,n}\delta_t, \delta_t)\}_t$*. More precisely, we define local periodogram*

$$\mathcal{I}_{t,m,\delta}(\boldsymbol{a}, \omega) = \frac{1}{2\pi m}\left|\sum_{k=1}^{m}(Y_{t-m/2+k} - \sum_{j=1}^{n} a_j x_{t-m/2+k,j})\delta_{t-m/2+k}\exp(ik\omega)\right|^2.$$

It is straightfoward to show that the expectation of $\mathcal{I}_{t,m,\delta}(\boldsymbol{a}, \omega)$ *at the true parameter* $\boldsymbol{a}_0$ *is* $\mathcal{I}_{t,m,\delta}(\boldsymbol{a}_0, \omega) = \mathcal{F}_{t,m,\delta}(\boldsymbol{\theta}_0, \omega)$ *where*

$$\mathcal{F}_{t,m,\delta}(\boldsymbol{\theta}_0, \omega) = \sum_{j=1}^{n}\sigma_{j,0}^2\int_{-\pi}^{\pi} I_{t,m,\delta}^{(j)}(\lambda) f_j(\boldsymbol{\vartheta}_0, \omega - \lambda)d\lambda + \sigma_{n+1,0}^2\int_{-\pi}^{\pi} I_{m,t,\delta}^{(n+1)}(\lambda) f_{n+1}(\boldsymbol{\vartheta}_0, \omega - \lambda)d\lambda,$$

the spectral densities $\{f_j\}$ *are defined in Assumption 2.1 and*

$$I_{m,\delta}^{(n+1)}(\omega) = \frac{1}{2\pi m}\Big|\sum_{k=1}^{m}\delta_t\exp(ik\omega)\Big|^2, \quad I_{t,m,\delta}^{(j)}(\omega) = \frac{1}{2\pi m}\Big|\sum_{k=1}^{m} x_{t-m/2+k,j}\delta_{t-m/2+k}\exp(ik\omega)\Big|^2.$$

By replacing $\mathcal{F}_{t,m}$ *and* $I_{t,m}$ *in (10) with* $\mathcal{F}_{t,m,\delta}$ *and* $I_{t,m,\delta}$ *respectively, to obtain a new objective function (based on Estimator 1) that allows for missing response values. We minimise this new criterion with respect to* $\boldsymbol{a}$ *and* $\boldsymbol{\theta}$ *to obtain estimators of* $\boldsymbol{a}_0$ *and* $\boldsymbol{\theta}_0$*. In a similar way, Estimator 2 can be adapted to allow for missing observations.*

4 Testing for randomness of the coefficients in the SCR model

Before fitting a SCR model to the data, it is of interest to check whether there is any evidence to suggest the coefficients are random. Bruesch and Pagan (1980) have proposed a Lagrange Multiplier test, to test the possibility that the parameters of a regression model are fixed against the alternative that they are random. Their test statistic is constructed under the assumption that the errors in the regression model come from a known distribution and are identically distributed. Further, Newbold and Bos (1985), Chapter 3, argue that the test proposed in Bruesch and Pagan (1980) can be viewed as the sample correlation between the squared residuals and the regressors (under the assumption of Gaussianity). In this section, we suggest a distribution free version of the test given in Newbold and Bos (1985), to test the hypothesis that the parameters are fixed against the alternative that they are random. Further, we propose a test to test the hypothesis the parameters are random (iid) against the alternative that they are stochastic (and correlated).

To simplify notation we will consider simple regression models with just one regressor, the discussion below can be generalised to the multiple regression case. Let us consider the null hypothesis $H_0 : Y_t = a_0 + a_1 x_t + \epsilon_t$ where $\{\epsilon_t\}$ are iid random variables with $\mathbb{E}(\varepsilon_t) = 0$ and $\text{var}(\varepsilon_t) = \sigma_\epsilon^2 < \infty$ against the alternative $H_A : Y_t = a_0 + a_1 x_t + \epsilon_t$, where $\epsilon_t = \alpha_t x_t + \varepsilon_t$ and $\{\alpha_t\}$ and $\{\varepsilon_t\}$ are iid random variables with $\mathbb{E}(\alpha_t) = 0$, $\mathbb{E}(\varepsilon_t) = 0$, $\text{var}(\alpha_t) = \sigma_\alpha^2 < \infty$ and $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$. We observe if the alternative were true, then $\text{var}(\epsilon_t) = x_t^2\sigma_\alpha^2 + \sigma_\varepsilon^2$, hence plotting $\text{var}(\epsilon_t)$ against x_t should give a clear positive slope. The following test is based on this observation. We suppose we observe $\{(Y_t, x_t)\}$ and use OLS to fit the model $a_0 + a_1 x_t$ to Y_t, and let $\hat{\epsilon}_t$ denote the residuals. We use as the test statistic the sample correlation between $\{x_t^2\}$ and $\{\hat{\epsilon}_t^2\}$

$$\mathcal{S}_1 = \frac{1}{T}\sum_{t=1}^{T} x_t^2\hat{\epsilon}_t^2 - \big(\frac{1}{T}\sum_{t=1}^{T} x_t^2\big)\big(\frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2\big). \tag{15}$$

To understand how $\mathcal{S}_1$ behaves under the null and alternative, we rewrite $\mathcal{S}_1$ as

$$\mathcal{S}_1 = \frac{1}{T}\sum_{t=1}^{T}\big(\hat{\epsilon}_t^2 - \mathbb{E}(\epsilon_t^2)\big)\big(x_t^2 - \frac{1}{T}\sum_{s=1}^{T} x_s^2\big) + R_1, \text{ where } R_1 = \frac{1}{T}\sum_{t=1}^{T} x_t^2\Big(\mathbb{E}(\epsilon_t^2) - \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}(\epsilon_s^2)\Big).$$

We observe that in the case that the null is true, then $\mathbb{E}(\epsilon_t^2)$ is constant for all t and $\mathcal{S}_1 = o_p(1)$. On the other hand when the alternative is true we have $\mathbb{E}(\mathcal{S}_1) = R_1$, noting that

$$R_1 = \frac{1}{T}\sum_{t=1}^{T} x_t^2\sigma_\alpha^2\Big(x_t^2 - \frac{1}{T}\sum_{t=1}^{T} x_s^2\Big). \tag{16}$$

We observe that R_1 depends on the amount of variation in the regressors $\{x_t\}$.

Proposition 4.1 *Let $\mathcal{S}_1$ be defined in (15), and suppose the null is true (that is $Y_t = a_0 + a_1 x_t + \epsilon_t$ where $\{\epsilon_t\}$ are iid) and $\mathbb{E}(|\epsilon_t|^{4+\delta}) < \infty$ (for some $\delta > 0$) then we have $\sqrt{T}\mathcal{S}_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma_1)$, where $\Gamma_1 = \frac{var(\epsilon_t^2)}{T}\sum_{t=1}^{T}(x_t^2 - \frac{1}{T}\sum_{s=1}^{T} x_s^2)^2$.*

Suppose the alternative is true (that is $Y_t = a_0 + a_1 x_t + \varepsilon_t$, where $\varepsilon_t = \alpha_t x_t + \epsilon_t$) and let us suppose $\{\varepsilon_t\}$ and $\{\epsilon_t\}$ are iid random variables, $\mathbb{E}(|\varepsilon_t|^{4+\delta}) < \infty$ and $\mathbb{E}(|\alpha_t|^{4+\delta}) < \infty$ (for some $\delta > 0$) then we have

$$\sqrt{T}(\mathcal{S}_1 - R_1) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{1}{T}\sum_{t=1}^{T} var((\alpha_t x_t + \epsilon_t)^2)\big(x_t^2 - \frac{1}{T}\sum_{s=1}^{T} x_s^2\big)^2\Big),$$

where R_1 is defined as in (16).

PROOF. The proof is straightforward using the martingale central limit theorem. □

Therefore, we can use the above proposition to test whether the parameters in a regression model are fixed. We mention that in the case that the parameters in the regression model are fixed, but the variance of the errors vary over time (independent of x_t) using the test statistic $\mathcal{S}_1$ may mistakenly lead to the conclusion that the alternative were true (because in this case R_1 will be non-zero). However, if the variance varies slowly over time, it is possible to modify the test statistic $\mathcal{S}_1$ to allow for a time-dependent variance, we omit the details.

We now adapt the test above to determine whether the parameters in the regressor model are random against the alternative there is correlation. More precisely, consider the null coefficients are random $H_0 : Y_t = a_0 + a_1 x_t + \epsilon_t$, where $\epsilon_t = \alpha_t x_t + \varepsilon_t$ and $\{\alpha_t\}$ are iid random variables and $\{\varepsilon_t\}$ is a stationary time series with $\text{cov}(\varepsilon_0, \varepsilon_k) = c(k)$ against the alternative that the coefficients are stochastic and correlated $H_A : Y_t = a_0 + a_1 x_t + \epsilon_t$, where $\epsilon_t = \alpha_t x_t + \varepsilon_t$ and $\{\alpha_t\}$ and $\{\varepsilon_t\}$ are stationary random variables (that are independent of each other). We observe that if the null were true $\mathbb{E}(\epsilon_t \epsilon_{t-1}) = c(1)$, whereas if the alternative were true then $\mathbb{E}(\epsilon_t \epsilon_{t-1}) = x_t x_{t-1} \rho(1) + c(1)$ (where $\rho(1) = \mathbb{E}(\alpha_t \alpha_{t-1})$), hence plotting $\epsilon_t \epsilon_{t-1}$ against $x_t x_{t-1}$ should give a clear line with a slope. Therefore we define the empirical correlation between $\{\hat{\epsilon}_t \hat{\epsilon}_{t-1}\}$ and $\{x_t x_{t-1}\}$ at lag one as the test statistic

$$\mathcal{S}_2 = \frac{1}{T} \sum_{t=2}^{T} x_t x_{t-1} \hat{\epsilon}_t \hat{\epsilon}_{t-1} - \big(\frac{1}{T} \sum_{t=2}^{T} x_t x_{t-1}\big)\big(\frac{1}{T} \sum_{t=2}^{T} \hat{\epsilon}_t \hat{\epsilon}_{t-1}\big). \tag{17}$$

Now rewriting $\mathcal{S}_2$ we have

$$\mathcal{S}_2 = \frac{1}{T} \sum_{t=2}^{T} \big(\hat{\epsilon}_t \hat{\epsilon}_{t-1} - \mathbb{E}(\epsilon_t \epsilon_{t-1})\big)\big(x_t x_{t-1} - \frac{1}{T} \sum_{s=2}^{T} x_s x_{s-1}\big) + R_2, \tag{18}$$

where $R_2 = \frac{1}{T} \sum_{t=2}^{T} x_t x_{t-1} \bigg(\mathbb{E}(\epsilon_t \epsilon_{t-1}) - \frac{1}{T} \sum_{s=2}^{T} \mathbb{E}(\epsilon_s \epsilon_{s-1}) \bigg)$. Now it is straightforward to see that if the null were true $\mathcal{S}_2 = o_p(1)$, but if the alternative were true then $\mathbb{E}(\mathcal{S}_2) \xrightarrow{\mathcal{P}} R_2$, noting that $R_2 = \frac{1}{T} \sum_{t=2}^{T} x_t x_{t-1} \bigg(x_t x_{t-1} \text{cov}(\alpha_t, \alpha_{t-1}) - \frac{1}{T} \sum_{s=2}^{T} x_s x_{s-1} \text{cov}(\alpha_s, \alpha_{s-1}) \bigg)$.

Proposition 4.2 *Let $\mathcal{S}_2$ be defined in (17), and suppose the null is true, that is $Y_t = a_0 + a_1 x_t + \epsilon_t$, where $\epsilon_t = \alpha_t x_t + \varepsilon_t$, and $\{\alpha_t\}$ are iid random variables with $\mathbb{E}(|\alpha_t|^8) < \infty$ and $\{\varepsilon_t\}$ is a stationary time series which satisfies $\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$ and $\sum_j |\psi_j| < \infty$ and $\mathbb{E}(|\eta_j|^8) < \infty$. Then we have $\sqrt{T}\mathcal{S}_2 \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \Gamma_2\big)$, where $\Gamma_2 = \frac{1}{T^2} \sum_{t_1, t_2 = 1}^{n} cov(\varepsilon_{t_1} \varepsilon_{t_1 - 1}, \varepsilon_{t_2} \varepsilon_{t_2 - 1}) v_{t_1} v_{t_2}$ and $v_t = \big(x_t^2 - \frac{1}{T} \sum_{s=2}^{T} x_s^2\big)^2$.*

On the other hand suppose the alternative were true, that is $Y_t = a_0 + a_1 x_t + \epsilon_t$, where $\epsilon_t = \alpha_t x_t + \varepsilon_t$, and $\{\alpha_t\}$ and $\{\varepsilon_t\}$ are stationary time series which satisfies $\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$, $\alpha_t = \sum_{j=0}^{\infty} \psi_{j,1} \eta_{t-j,1}$, $\sum_j |\psi_j| < \infty$, $\sum_j |\psi_{j,1}| < \infty$, $\mathbb{E}(|\eta_j|^8) < \infty$ and $\mathbb{E}(|\eta_{j,1}|^8) < \infty$. Then we have $\sqrt{T}(\mathcal{S}_2 - R_2) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \Gamma_3\big)$, where $\Gamma_3 = var(\sqrt{T}\mathcal{S}_2)$.

PROOF. The result can be proven by partitioning $\mathcal{S}_3$ into a term which is the sum of martingale differences (and then applying the martingale central limit theorem) and a smaller order term, we omit the details. □

It is worth noting, it is not necessary to limit the test statistic for testing for correlation at lag one (as was done in $\mathcal{S}_2$). It is straightforward to generalise $\mathcal{S}_2$ to test for correlations at larger lags, indeed it may be possible to use a variation of the Portmanteau test.

5 Asymptotic properties of the frequency domain estimators

5.1 Some assumptions

We now consider the asymptotic sampling properties of these estimates. We need the following assumptions on the stochastic coefficients $\{\alpha_{t,j}\}$ and the regressors. We need these to show consistency and also obtaining the sampling distributions. We state two sets of assumptions, one solely on the stochastic coefficients (Assumption 5.1) and the second one on the regressors (Assumption 5.2).

Let $|\cdot|$ denote the Euclidean norm of a vector or matrix respectively.

Assumption 5.1 *(i) The parameter spaces Θ_1 and Θ_2 are such that, there exists a $\delta_1 > 0$, where $\inf_{\boldsymbol{\Sigma}\in\Theta_2} \sigma_{n+1}^2 \geq \sqrt{\delta_1}$ and $\inf_{\boldsymbol{\vartheta}\in\Theta_1} \int_{-\pi}^{\pi} \big(\sum_{r=-(m-1)}^{m-1} (\frac{m-|r|}{m}) \exp(ir\lambda)\big) \cdot f_0(\boldsymbol{\vartheta}, \omega-\lambda) d\lambda \geq \sqrt{\delta_1}$.*

(ii) The parameter spaces Ω, Θ_1 and Θ_2 are compact.

(iii) The coefficients $\psi_{i,j}$ of the MA(∞) representation given in Assumption 2.1, satisfy $\sup_{\boldsymbol{\vartheta}\in\Theta_1} \sum_{i=0}^{\infty} |i| \cdot |\nabla_{\boldsymbol{\vartheta}}^k \psi_{i,j}(\boldsymbol{\vartheta})| < \infty$ (for all $0 \leq k \leq 3$ and $1 \leq j \leq n+1$).

(iv) The innovation sequences $\{\eta_{t,j}\}$ satisfy $\sup_{1\leq j\leq n+1} \mathbb{E}(\eta_{t,j}^8) < \infty$.

Assumption 5.2 *(i) $\sup_{t,j} |x_{t,j}| < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} \frac{\boldsymbol{X}_t \boldsymbol{X}_t'}{var(Y_t)}$ is non-singular, where $\boldsymbol{X}_t' = (x_{t,1}, \ldots, x_{t,n})$, is non-singular for all T.*

(ii) Suppose that $\boldsymbol{\theta}_0$ is the true parameter. There does not exist another $\boldsymbol{\theta}^* \in \Theta_1 \otimes \Theta_2$ such that for all $0 \leq r \leq m-1$ and infinite number of t we have

$$\sum_{j=1}^{n} \big(c_j(\boldsymbol{\theta}_0, r) - c_j(\boldsymbol{\theta}^*, r)\big) \sum_{k=0}^{m-|r|} x_{t-m/2+k,j} x_{t-m/2+k+r,j} = 0.$$

Let $\mathcal{J}_{T,m}(\boldsymbol{\vartheta}, \omega)' = \sum_{t=m/2}^{T-m/2} \mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \omega)^{-1} (J_{T,m}^{(1)}(\omega), \ldots, J_{t,m}^{(n)}(\omega))$. For all T and $\boldsymbol{\vartheta} \in \Theta_1$, $\int \mathcal{J}_{T,m}(\boldsymbol{\vartheta}, \omega) \mathcal{J}_{T,m}(\boldsymbol{\vartheta}, \omega)' d\omega$ is nonsingular and the smallest eigenvalue is bounded away from zero.

(iii) For all T, $\mathbb{E}(\nabla_{\boldsymbol{\theta}}^2 \mathcal{L}_T^{(m)}(\boldsymbol{a}_0, \boldsymbol{\theta}_0))$ and $\mathbb{E}(\nabla_{\boldsymbol{a}}^2 \mathcal{L}_T^{(m)}(\boldsymbol{a}_0, \boldsymbol{\theta}_0))$ are nonsingular matrices and the smallest eigenvalue is bounded away from zero.

Assumption 5.1(i) ensures that $\mathcal{F}_{t,m}$ is bounded away from zero, which implies that $\mathbb{E}|\mathcal{L}_T^{(m)}(\boldsymbol{a}, \boldsymbol{\theta})| < \infty$, similarly Assumption 5.1(iii) implies that $\sum_{r=-\infty}^{\infty} |r \cdot \nabla^k c_{j_1,j_2}(\boldsymbol{\theta}, r)| < \infty$, therefore $\mathbb{E}|\nabla_{\boldsymbol{\theta}}^k \mathcal{L}_T^{(m)}(\boldsymbol{a}, \boldsymbol{\theta})| < \infty$. Assumption 5.2(i,ii) ensures that $\mathbb{E}(\mathcal{L}_T(\cdot))$ and $\mathbb{E}(\mathcal{L}_T^{(m)}(\cdot))$ have a unique minima.

5.2 Sampling properties of Estimator 1: $\hat{\boldsymbol{\alpha}}_T$ and $\hat{\boldsymbol{\theta}}_T$

First we study the asymptotic properties of the $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$. Using these results we obtain the asymptotic properies of the two stage estimator $(\tilde{\boldsymbol{a}}_T, \tilde{\boldsymbol{\vartheta}}_T, \tilde{\boldsymbol{\Sigma}}_T)$. We first show consistency of $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$.

Proposition 5.1 *Suppose Assumptions 2.1, 5.1(i,ii) and 5.2 are satisfied and the estimators $\hat{\boldsymbol{a}}_T$, $\hat{\boldsymbol{\theta}}_T$ are defined as in (12). Then we have $\hat{\boldsymbol{a}}_T \xrightarrow{\mathcal{P}} \boldsymbol{a}_0$ and $\hat{\boldsymbol{\theta}}_T \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$, as $T_m \to \infty$ and $T \to \infty$.*

PROOF. In Appendix A.2. □

We now prove asymptotic normality of $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$, which uses Taylor expansion arguments. Noting that $\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0) = (\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0), \nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$, the Taylor expansion of $\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$ about $\hat{\boldsymbol{a}}_T$ and $\hat{\boldsymbol{\theta}}_T$ is

$$\begin{pmatrix} \hat{\boldsymbol{a}}_T - \boldsymbol{a}_0 \\ \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \end{pmatrix} = \nabla^2\mathcal{L}_T^{(m)}(\bar{\boldsymbol{a}}_T, \bar{\boldsymbol{\theta}}_T)^{-1} \begin{pmatrix} \nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0) \\ \nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0) \end{pmatrix}, \tag{19}$$

where $\bar{\boldsymbol{a}}_T$ and $\bar{\boldsymbol{\theta}}_T$ lie between $\hat{\boldsymbol{a}}_T$ and $\boldsymbol{a}_0$ and $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$ respectively. Expressions for the first and second order derivatives can be found in Appendix A.2. It is worth noting that using Lemma A.2, (19) and under Assumption 5.2(iii) we have

$$\Big((\hat{\boldsymbol{a}}_T - \boldsymbol{a}_0), (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)\Big) = O_p(\frac{1}{\sqrt{T}}). \tag{20}$$

In the following theorem we obtain the asymptotic distribution of $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$. We now define elements of the matrix $\mathbb{E}(\nabla^2\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$ and $\mathrm{var}(\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$. We have

$$\begin{aligned} \mathbb{E}(\nabla_{\boldsymbol{a}}^2\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))\rfloor_{j_1,j_2} &= \frac{1}{T_m}\sum_{t=m/2}^{T-m/2}\int \mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1} J_{t,m}^{(j_1)}(\omega) J_{t,m}^{(j_2)}(-\omega)d\omega \qquad (21) \\ \mathbb{E}(\nabla_{\boldsymbol{\theta}}^2\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) &= \frac{1}{T_m}\sum_{t=m/2}^{T-m/2}\int \frac{\nabla\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)\nabla\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)'}{(\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega))^2}d\omega, \end{aligned}$$

and further

$$\begin{aligned} W_T^{(m)} &= T\mathrm{var}(\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) \qquad (22) \\ &= \begin{pmatrix} W_{1,T} & T\mathrm{cov}(\nabla_{\boldsymbol{\beta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0), \nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) \\ T\mathrm{cov}(\nabla_{\boldsymbol{\beta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0), \nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) & T\mathrm{var}(\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) \end{pmatrix} \end{aligned}$$

with $\boldsymbol{\beta} = (\boldsymbol{a}, \boldsymbol{\vartheta})$ and

$$W_{1,T} = \mathrm{var}\begin{pmatrix} \sqrt{T}\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\sigma}_0) \\ \sqrt{T}\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\sigma}_0) \end{pmatrix}. \tag{23}$$

We partition $W_T^{(m)}$ into matrix blocks, because the partitioning is useful for obtaining the sampling properties of Estimator 2 in Section 5.3.

Theorem 5.1 *Suppose Assumptions 2.1, 5.1 and 5.2 hold, then we have*

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{a}}_T - \boldsymbol{a}_0 \\ \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, (V_T^{(m)})^{-1} W_T^{(m)} (V_T^{(m)})^{-1}\Big), \tag{24}$$

as $T_m/T \to 1$ and $T \to \infty$, where $W_T^{(m)}$ is defined in (22) and

$$V_T^{(m)} = \begin{pmatrix} \mathbb{E}(\nabla_{\boldsymbol{a}}^2\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) & 0 \\ 0 & \mathbb{E}(\nabla_{\boldsymbol{\theta}}^2\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) \end{pmatrix}.$$

PROOF. See Appendix A.2. □

We observe that $V_T^{(m)}$ is a block diagonal matrix, this is because a straightforward calculation gives $\mathbb{E}(\nabla_{\boldsymbol{a}}\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) = 0$.

An expression for $V_T^{(m)}$ in terms of $\mathcal{F}_{t,m}$ is given in (21) we now obtain a similar expression for $W_T^{(m)}$. The expression for $W_T^{(m)}$ is quite cumbersome, to make it more interpretable we state it in terms of operators. Let ℓ_2 denote the space of all square summable (vector) sequences. We define the general operator $\Gamma^{(k_1,k_2),(\boldsymbol{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\cdot)$ which acts on the d_1 and d_2-dimensional column vector sequences $\underline{g} = \{\underline{g}(k)\} \in \ell_2$ and $\underline{h} = \{\underline{h}(k)\} \in \ell_2$ (to reduce notation we do not specify the dimensions of the vectors in ℓ_2)

$$\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\underline{g},\underline{h})(s,u) \;=\; \sum_{r_1=-k_1}^{m-k_1}\sum_{r_2=-k_2}^{m-k_2} \underline{g}(r_1)\underline{h}(r_2)'\mathcal{K}^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_{s+r_1},\mathbf{x}_u,\mathbf{x}_{u+r_2}}(s-u,r_1,r_2), \tag{25}$$

where $\mathbf{x}_s = (x_{s,1},\ldots,x_{s,n},1)$ are the regressors observed at time s and the kernel $\mathcal{K}^{(\mathbf{j})}$ is defined as either

$$\mathcal{K}^{(0)}_{\mathbf{x}_{s_1},\mathbf{x}_{s_2},\mathbf{x}_{s_3},\mathbf{x}_{s_4}}(k,r_1,r_2) \;=\; \sum_{j=1}^{n+1} x_{s_1,j}x_{s_2,j}x_{s_3,j}x_{s_4,j}\mathrm{cov}(\alpha_{0,j}\alpha_{r_1,j},\alpha_{k,j}\alpha_{k+r_2,j})$$

$$\mathcal{K}^{(j_1)}_{\mathbf{x}_{s_1},\mathbf{x}_{s_2},\mathbf{x}_{s_3},\mathbf{x}_{s_4}}(k,r_1,r_2) \;=\; \sum_{j=1}^{n+1} x_{s_1,j}x_{s_2,j}x_{s_3,j}x_{s_4,j_1}\mathrm{cov}(\alpha_{0,j}\alpha_{r_1,j},\alpha_{k,j}) \quad 1\le j_1\le n$$

$$\text{or } \mathcal{K}^{(j_1,j_2)}_{\mathbf{x}_{s_1},\mathbf{x}_{s_2},\mathbf{x}_{s_3},\mathbf{x}_{s_4}}(k,r_1,r_2) \;=\; \sum_{j=1}^{n+1} x_{s_1,j}x_{s_2,j_1}x_{s_3,j}x_{s_4,j_2}\mathrm{cov}(\alpha_{0,j},\alpha_{k,j}) \quad 1\le j_1,j_2\le n \tag{26}$$

where for notational convenience we set $\alpha_{t,n+1} := \varepsilon_t$ and $x_{t,n+1} := 1$. We will show that $\mathrm{var}(\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0),\boldsymbol{\theta}_0)$ can be written in terms of $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$. Suppose $\mathcal{G}_s(\cdot)$ and $\mathcal{H}_u(\cdot)$ are square integrable vector functions, then it can be shown

$$\mathrm{var}\begin{pmatrix} \int_0^{2\pi}\mathcal{H}_u(\omega)X_u\sum_{r_2=-k_2}^{m-k_2}x_{u+r_1,j_1}\exp(ir_2\omega)d\omega \\ \int_0^{2\pi}\mathcal{G}_s(\omega)X_s\sum_{r_1=-k_1}^{m-k_1}X_{s+r_1}\exp(ir_1\omega)d\omega \end{pmatrix}$$

$$= \begin{pmatrix} \Gamma^{(k_2,k_2),(j_1,j_1)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\underline{h}_u,\bar{\underline{h}}_u)(u,u) & \Gamma^{(k_1,k_2),(j_1)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\bar{\underline{g}}_s,\underline{h}_u)(s,u) \\ \Gamma^{(k_1,k_2),(j_1)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\underline{g}_s,\bar{\underline{h}}_u)(s,u) & \Gamma^{(k_1,k_1),(0)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\underline{g}_s,\bar{\underline{g}}_s)(s,s) \end{pmatrix} \tag{27}$$

where $\underline{g}_s = \{\underline{g}_s(r) = \int\mathcal{G}_s(\omega)\exp(ir\omega)d\omega\}_r$, $\underline{h}_s = \{\underline{h}_s(r) = \int\mathcal{H}_s(\omega)\exp(r\omega)d\omega\}_r$, $\bar{\underline{g}}_s = \{\underline{g}_s(-r)\}_r$ and $\bar{\underline{h}}_s = \{h_s(-r)\}_r$.

Using the above notation we can write

$$W_T^{(m)} \;=\; T\mathrm{var}(\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) = \frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=k_2}^{T-m+k_2} \tag{28}$$

$$\begin{pmatrix} \Gamma^{(k_1,k_2),(\mathbf{j},\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(a_{s+m/2-k_1},a_{u+m/2-k_2})(s,u) & \Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\underline{b}_{s+m/2-k_1},a_{u+m/2-k_2})(s,u) \\ \Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\bar{\underline{b}}_{s+m/2-k_1},a_{u+m/2-k_2})(s,u) & \Gamma^{(k_1,k_2),(0)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(\underline{b}_{s+m/2-k_1},\bar{\underline{b}}_{u+m/2-k_2})(s,u) \end{pmatrix},$$

where $\Gamma^{(k_1,k_2),(\mathbf{j},\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T} = \{\Gamma^{(k_1,k_2),(j_1,j_2)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}\}_{j_1,j_2=1,\ldots,n}$ and $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T} = \{\Gamma^{(k_1,k_2),(j)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}\}_{j=1,\ldots,n}$, $\underline{b}_s = \{\underline{b}_s(k)\}$, $a_s = \{a_s(k)\}$ with

$$a_s(k) = \int\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}\exp(ik\omega)d\omega, \quad \underline{b}_s(k) = \int\nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}\exp(ik\omega)d\omega, \tag{29}$$

(noting that $\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}$ is a symmetric function hence $a_s = \bar{a}_s$). We will use (28) in Section 5.4, when we compare Estimator 1 with the Gaussian maximum likelihood estimator.

Remark 5.1 *The expression (28) can be used to estimate $W_T^{(m)}$. Examining $W_T^{(m)}$ we observe that the only parameters in $W_T^{(m)}$ which we need to estimate (in addition to $(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$) are the cumulants $cum(\eta_{t,j},\eta_{t,j},\eta_{t,j})$ and $cum(\eta_{t,j},\eta_{t,j},\eta_{t,j},\eta_{t,j})$. We estimate the joint cumulants from the joint moments. We suggest the following approximate expression to obtain the estimates. We group the observations $\{Y_t\}$ in $(n+1)$ blocks, each of length $M = T/(n+1)$, and evaluate the empirical third moment within each block. If the size of each block $M = T/(n+1)$ is large, we obtain the following approximate equations*

$$\frac{1}{M}\sum_{s=1}^{M} Y_{Mr+s}^3 \approx \frac{1}{M}\sum_{j=1}^{n+1}\mathbb{E}(\eta_{t,j}^3)\sum_{s=1}^{M} x_{Mr+s,j}^3 \sum_{i=0}^{\infty}\psi_{i,j}(\boldsymbol{\vartheta}_0)^3.$$

Obviously this equation is true for $r = 1,\ldots,(n+1)$. Therefore, we have $(n+1)$ linear simultaneous equations in the unknown $\{\mathbb{E}(\eta_{t,j}^3)\}$, which, if we replace $\boldsymbol{\vartheta}_0$ with $\hat{\boldsymbol{\vartheta}}_T$, we can solve for. Thus we have an estimator of $\mathbb{E}(\eta_{t,j}^3)$. Using a similar method, but evaluating the empirical fourth moment, we can obtain an estimator of the fourth order cumulant.

5.3 Sampling properties of Estimator 2: $\tilde{\boldsymbol{\Sigma}}_T$, $\tilde{\boldsymbol{a}}_T$, $\tilde{\boldsymbol{\vartheta}}_T$

We now obtain the sampling properties of Estimator 2. We first consider the properties of the variance estimator $\tilde{\boldsymbol{\Sigma}}_T$.

Proposition 5.2 *Suppose Assumptions 2.1, 5.1 and 5.2 are satisfied, let $\mathcal{L}_T(\boldsymbol{a},\boldsymbol{\Sigma})$ and $\tilde{\boldsymbol{\Sigma}}_T$ be defined as in (13). Then we have*

$$\begin{aligned}\sqrt{T}\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0) &\stackrel{\mathcal{D}}{\to} \mathcal{N}\Big(0, var(\sqrt{T}\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0))\Big) \qquad (30)\\ \sqrt{T}(\tilde{\boldsymbol{\Sigma}}_T-\boldsymbol{\Sigma}_0) &\stackrel{\mathcal{D}}{\to} \mathcal{N}\Big(0, \mathbb{E}\big(\nabla^2_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0)\big)^{-1} var(T^{1/2}\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0))\mathbb{E}\big(\nabla^2_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0)\big)^{-1}\Big),\end{aligned}$$

as $T\to\infty$.

PROOF. See Appendix A.3. □

We now consider the properties of $(\tilde{\boldsymbol{a}}_T,\tilde{\boldsymbol{\vartheta}}_T)$, which are obtained in Step 2 of Estimator 2.

Theorem 5.2 *Suppose Assumptions 2.1, 5.1 and 5.2 hold, then we have*

$$\sqrt{T}\begin{pmatrix}\tilde{\boldsymbol{a}}_T-\boldsymbol{a}_0\\ \tilde{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}_0\end{pmatrix}\stackrel{\mathcal{D}}{\to}\mathcal{N}\Big(0,(\tilde{V}_T^{(m)})^{-1}\tilde{W}_T^{(m)}(\tilde{V}_T^{(m)})^{-1}\Big), \qquad (31)$$

as $T_m/T\to 1$ and $T\to\infty$, where

$$\tilde{W}_T^{(m)} = W_{T,1} + \begin{pmatrix} 0 & \Xi_1\\ \Xi_1' & \Xi_2\end{pmatrix}, \quad \tilde{V}_T^{(m)} = \begin{pmatrix}\mathbb{E}(\nabla^2_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0)) & 0\\ 0 & \mathbb{E}(\nabla^2_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0))\end{pmatrix}$$

with $W_{T,1}$ defined as in (23) and

$$\begin{aligned}
\Xi_1 &= cov\Big(\sqrt{T}\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0),\sqrt{T}\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0)\Big)\mathcal{Q}_T' \\
\Xi_2 &= 2cov\Big(\sqrt{T}\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0),\sqrt{T}\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0)\Big)\mathcal{Q}_T' + \mathcal{Q}_T var(\sqrt{T}\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0))Q_T',
\end{aligned}$$

Q_T is a $q\times(n+1)$-dimensional matrix defined by

$$\begin{aligned}
\mathcal{Q}_T &= \Big(\frac{1}{T_m}\int\sum_{t=m/2}^{T-m/2}(\nabla_{\boldsymbol{\vartheta}}\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0,\omega)^{-1})\otimes\underline{H}_{(t,m)}(\omega)d\omega\Big)\mathbb{E}\big(\nabla^2_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0)\big)^{-1} \\
h_j^{(t,m)}(\omega) &= \int_{-\pi}^{\pi} I_{t,m}^{(j)}(\lambda)f_j(\boldsymbol{\vartheta}_0,\omega-\lambda)d\lambda, \qquad (32)
\end{aligned}$$

$\underline{H}_{(t,m)}(\omega)=(h_1^{(t,m)}(\omega),\ldots,h_{n+1}^{(t,m)}(\omega))$ and noting that $\otimes$ denotes the tensor product.

Remark 5.2 *Comparing the variances of the two estimators in (24) and (31), we observe that they are similar. In particular $\tilde{V}_T^{(m)}$ is a submatrix of $V_T^{(m)}$, and $\tilde{W}_T^{(m)}$ is a submatrix of $W_T^{(m)}$, plus the additional terms Ξ_1 and Ξ_2. The terms Ξ_1 and Ξ_2 are due to the estimation of $\boldsymbol{\Sigma}_0$ in the first stage of the scheme. However, its worth noting that $\sqrt{T}(\tilde{\boldsymbol{a}}_T-\boldsymbol{a}_0)$ (Estimator 2) asymptotically has the same distribution as $\sqrt{T}(\hat{\boldsymbol{a}}_T-\boldsymbol{a}_0)$ (Estimator 1).*

5.4 The Gaussian likelihood and asymptotic efficiency of Estimator 1

In this section we compare the asymptotic properties of the frequency domain estimator $(\hat{\boldsymbol{a}}_T,\hat{\boldsymbol{\theta}}_T)$ with the Gaussian maximum likelihood estimator (GMLE). We recall the GMLE is the estimator, which minimises the objective function which is constructed as if the stochastic coefficients $\{\alpha_{t,j}\}$ and errors $\{\varepsilon_t\}$ were Gaussian. However, unlike the frequency domain estimators, there does not (in general) exist an explicit expression for the asymptotic variance of the Gaussian maximum likelihood. Instead we will consider a subclass of SCR models, where the regressors vary slowly over time and do the comparison for this subclass. We will show that for this subclass an asymptotic expression for the asymptotic distributional variance of the GMLE can be derived. We will assume that the regressors are such that there exists a 'smooth' function, $x_j(\cdot)$, such that $x_{t,j}=x_j(\frac{t}{N})$ and $Y_t:=Y_{t,N}$ satisfies

$$Y_{t,N}=\sum_{j=1}^{n}(a_{j,0}+\alpha_{t,j})x_j(\frac{t}{N})+\varepsilon_t \quad t=1,\ldots,T. \qquad (33)$$

In the following lemma we obtain the asymptotic distribution of the GMLE under the asymptotic framework that both T and $N\to\infty$. In order to succinctly represent the asymptotic variance we define the operator

$$\Gamma^{(\boldsymbol{j})}_{\mathbf{x}_s,\mathbf{x}_u}(\underline{g},\underline{h})(k) = \sum_{r_1=-\infty}^{\infty}\sum_{r_2=-\infty}^{\infty}\underline{g}(r_1)\underline{h}(r_2)\mathcal{K}^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s,\mathbf{x}_u,\mathbf{x}_u}(k,r_1,r_2), \qquad (34)$$

where $\underline{g}=\{\underline{g}(r)\}$ and $\underline{h}=\{\underline{h}(r)\}$ are d_1 and d_2-dimensional column vector sequences which are in ℓ_2 and $\mathcal{K}^{(\mathbf{j})}$ is defined in (26).

Lemma 5.1 *Let us suppose that $\{Y_{t,N}\}$ satisfies (33), where the $\{\alpha_{t,j}\}$ and $\{\varepsilon_t\}$ are Gaussian and satisfy Assumption 2.1. Let*

$$\mathcal{F}(v,\boldsymbol{\theta}_0,\omega) \quad = \quad \sum_{j=1}^{n} x_j(v)^2\sigma_{j,0}^2 f_j(\boldsymbol{\vartheta}_0,\omega) + \sigma_{n+1,0}^2 f_{n+1}(\boldsymbol{\vartheta}_0,\omega). \tag{35}$$

We assume that there does not exist another $\boldsymbol{\theta} \in \Theta_1 \otimes \Theta_2$ such that $\mathcal{F}(v,\boldsymbol{\theta}_0,\omega) = \mathcal{F}(v,\boldsymbol{\theta}_0,\omega)$ for all $v \in [0, T/N]$ and the matrix $\frac{N}{T}\int_0^{T/N} \mathbf{x}(v)\mathbf{x}(v)'dv$, (with $\mathbf{x}(v)' = (x_1(v),\ldots,x_n(v))$) has eigenvalues which are bounded from above and away from zero. Suppose $(\boldsymbol{a}_{mle},\boldsymbol{\theta}_{mle})$ is the Gaussian maximum likelihood estimator of the parameters $(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$. Then we have

$$\sqrt{T}\begin{pmatrix} \boldsymbol{a}_{mle} - \boldsymbol{a}_0 \\ \boldsymbol{\theta}_{mle} - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} \Delta_1^{-1} & 0 \\ 0 & \Delta_2^{-1} \end{pmatrix}\right)$$

with $N \to \infty$ as $T \to \infty$, where

$$(\Delta_1)_{j_1,j_2} = \frac{N}{T}\int_0^{T/N} \sum_{k=-\infty}^{\infty} \Gamma^{(j_1,j_2)}_{\mathbf{x}(v),\mathbf{x}(v)}(a(v),a(v))(k)dv = \frac{N}{T}\int_0^{T/N} x_{j_1}(v)x_{j_2}(v)\mathcal{F}(v,\boldsymbol{\theta}_0,0)^{-1}dv,$$

$$\Delta_2 = 2\frac{N}{T}\int_0^{T/N} \sum_{k=-\infty}^{\infty} \Gamma^{(0)}_{\mathbf{x}(v),\mathbf{x}(v)}(\underline{b}(v),\bar{\underline{b}}(v))(k)dv = 2\frac{N}{T}\int_0^{T/N}\int_0^{2\pi} \frac{\nabla_{\boldsymbol{\theta}}\mathcal{F}(v,\boldsymbol{\theta}_0,\omega)(\nabla_{\boldsymbol{\theta}}\mathcal{F}(v,\boldsymbol{\theta}_0,-\omega))'}{|\mathcal{F}(v,\boldsymbol{\theta}_0,\omega)|^2}d\omega dv,$$

$$a(v,k) \quad = \quad \int \frac{1}{\mathcal{F}(v,\boldsymbol{\theta}_0,\omega)}\exp(ik\omega)d\omega \quad \underline{b}(v,k) = \int \nabla_{\boldsymbol{\theta}}\mathcal{F}(v,\boldsymbol{\theta}_0,\omega)^{-1}\exp(ik\omega)d\omega$$

$a(v) = \{a(v,k)\}$ and $\bar{\underline{b}}(v) = \{\underline{b}(v,-k)\}$.

PROOF. See Appendix A.4.2. □

In practice, for any given set of regressors $\{x_{t,j}\}$, N will not be known, but a lower bound for N can be obtained from $\{x_{t,j}\}$. To ensure the magnitude of the regressors does not influence N, we will assume that the regressors satisfy $\frac{1}{T}\sum_{t=1}^{T} x_{t,j}^2 = 1$ (for all j). To measure the smoothness of the regressors define

$$\hat{N} = \frac{1}{\sup_{t,j}|x_{t,j} - x_{t-1,j}|}. \tag{36}$$

Clearly if $\hat{N}$ is large, this indicates that the regressors are smooth.

We now compare the asymptotic variance of the GMLE and (frequency domain) Estimator 1. We show that the difference primarily depends on how large $\hat{N}$ is. We note that the proof of the proposition below follows from Proposition A.6, in the Appendix A.4.2, and that Proposition A.6(i) gives a useful approximation to $\mathrm{var}(\sqrt{T}\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) = W_T^{(m)}$.

Proposition 5.3 *Suppose Assumptions 2.1, 5.1 and 5.2 hold, $\hat{N}$ is defined as in (36), the operators $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$ and $\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}$ are defined as in (25) and (34) respectively and*

$$\sup_j \int \left|\frac{d^2 f_j(\boldsymbol{\vartheta}_0,\omega)}{d\omega^2}\right|^2 d\omega < \infty \quad \text{and } \sup_j \int \left|\frac{d^2 \nabla_{\boldsymbol{\theta}} f_j(\boldsymbol{\vartheta}_0,\omega)}{d\omega^2}\right|^2 d\omega < \infty. \tag{37}$$

Let $V_T^{(m)}$, $W_T^{(m)}$, $\Delta_{T,N,1}$ and $\Delta_{T,N,2}$ be defined as in (21), (22) and Lemma 5.1 respectively. Then we have

$$\left|W_T^{(m)} - \left(\begin{pmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{pmatrix} + \begin{pmatrix} 0 & \Gamma_{1,2} \\ \Gamma_{1,2}' & \Gamma_2 \end{pmatrix}\right)\right| \leq K\left\{\frac{1}{\hat{N}} + \frac{1}{m} + \frac{1}{T_m} + \frac{m}{\hat{N}}\right\} \tag{38}$$

$$\left|V_T^{(m)} - \begin{pmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{pmatrix}\right| \leq K\frac{m}{\hat{N}} \tag{39}$$

where K is a finite constant,

$$\begin{aligned}
\Gamma_2 &= \frac{N}{T}\int_0^{T/N}\int_0^{2\pi}\int_0^{2\pi} \frac{\nabla_{\boldsymbol{\theta}}\mathcal{F}(v,\boldsymbol{\theta}_0,\omega_1)\nabla_{\boldsymbol{\theta}}\mathcal{F}(v,\boldsymbol{\theta}_0,\omega_1)'}{\mathcal{F}(v,\boldsymbol{\theta}_0,\omega_1)^2\mathcal{F}(v,\boldsymbol{\theta}_0,\omega_2)^2}\mathcal{F}_4(v,\boldsymbol{\vartheta}_0,\omega_1,\omega_2,-\omega_1)d\omega_1 d\omega_2 dv, \\
\Gamma_{1,2} &= \frac{N}{T}\int_0^{T/N}\mathbf{x}(v)\int_0^{2\pi}\int_0^{2\pi} \frac{\nabla_{\boldsymbol{\theta}}\mathcal{F}(v,\boldsymbol{\theta}_0,\omega_2)'}{\mathcal{F}(v,\boldsymbol{\theta}_0,\omega_1)\mathcal{F}(v,\boldsymbol{\theta}_0,\omega_2)^2}\mathcal{F}_3(v,\boldsymbol{\vartheta}_0,\omega_1,\omega_2)\exp(ir\omega_2)d\omega_1 d\omega_2 dv,
\end{aligned} \tag{40}$$

$\mathbf{x}(v)' = (x_1(v),\ldots,x_n(v))$, *$\mathcal{F}(v,\boldsymbol{\theta},\omega)$ is defined in (35),*

$$\begin{aligned}
\mathcal{F}_3(v,\boldsymbol{\vartheta},\omega_1,\omega_2) &= \sum_{j=1}^{n+1}\kappa_{j,3}x_j(v)^3 A_j(\boldsymbol{\vartheta},\omega_1)A_j(\boldsymbol{\vartheta},\omega_2)A_j(\boldsymbol{\vartheta},-\omega_1-\omega_2) \\
\mathcal{F}_4(v,\boldsymbol{\vartheta},\omega_1,\omega_2,\omega_3) &= \sum_{j=1}^{n+1}\kappa_{j,4}x_j(v)^4 A_j(\boldsymbol{\vartheta},\omega_1)A_j(\boldsymbol{\vartheta},\omega_2)A_j(\boldsymbol{\vartheta},\omega_3)A_j(\boldsymbol{\vartheta},-\omega_1-\omega_2-\omega_3),
\end{aligned}$$

$\kappa_{j,3} = cum(\eta_{0,j},\eta_{0,j},\eta_{0,j})$ *and* $\kappa_{j,4} = cum(\eta_{0,j},\eta_{0,j},\eta_{0,j},\eta_{0,j})$.

PROOF. The result immediately follows from Lemma 5.1 and Proposition A.6, in the Appendix. □

Remark 5.3 (Selecting m) *Let us consider the case that $\{\alpha_{t,j}\}$ and $\{\varepsilon_t\}$ are Gaussian, this implies $\Gamma_{1,2} = 0$ and $\Gamma_2 = 0$. Now, comparing the asymptotic variances of the GMLE and $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$ we see that if we let $N \to \infty$, $m \to \infty$ and $m/N \to 0$ (noting that we have replaced $\hat{N}$ with N) as $T \to \infty$, then the GMLE ($(\boldsymbol{a}_{mle},\boldsymbol{\theta}_{mle})$) and $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$ both have the same asymptotic distribution. Hence within this framework, the relative efficiency of the frequency domain estimator compared with the GMLE is one.*

Furthermore, in the case that $\{\alpha_{t,j}\}$ and $\{\varepsilon_t\}$ are Gaussian, (38) suggests a method for selecting m. Since in this case the GMLE is efficient, by using (38) we have

$$\left|(V_T^{(m)})^{-1}W_T^{(m)}(V_T^{(m)})^{-1} - diag(\Delta_1^{-1},\Delta_2^{-1})\right| = O_p\left(\frac{1}{\hat{N}} + \frac{1}{m} + \frac{1}{T_m} + \frac{m}{\hat{N}}\right).$$

Hence the above difference is minimised when $m = \hat{N}^{1/2}$.

6 Simulations and real data analysis

We now consider a simulation study and the results of two real data examples.

6.1 Simulation study

Here we generate three time series, each of length 500 and estimate the parameters, and compare these with the estimates obtained by OLS, GMLE. In each case, we replicate the time series 40 times, and estimates given are based on these 40 independent replications. Along with each average estimate the mean squared error is given.

(i) Model 1: $Y_t^{(1)} = (a_1 + \alpha_t)x_t + \varepsilon_t$, where

$$\alpha_t = \vartheta_1 \alpha_{t-1} + \eta_{t,1}, \quad \varepsilon_t = \eta_{t,2}, \quad x_t = \sin(0.01 \cdot t) + \cos(0.05 \cdot t), \quad t = 1, \dots, 500$$

with $a_1 = 2$, and $\vartheta_1 = 0.7$. In this case the parameters to be estimated are $\boldsymbol{a}_0 = a_1$ and $\boldsymbol{\vartheta}_0 = \vartheta_1$.

(ii) Model 2: $Y_t^{(2)} = (a_1 + \alpha_t)x_t + \varepsilon_t$, where

$$\alpha_t = \vartheta_{1,1}\alpha_{t-1} + \eta_{t,1}, \quad \varepsilon_t = \vartheta_{1,2}\varepsilon_{t-1} + \eta_{t,2}, \quad x_t = \sin(0.01 \cdot t) + \cos(0.05 \cdot t), \quad t = 1, \dots, 500$$

with $a_1 = 2$, $\vartheta_{1,1} = 0.7$ and $\vartheta_{1,2} = -0.6$. In this case $\boldsymbol{a}_0 = a_1$ and $\boldsymbol{\vartheta}_0 = (\vartheta_{1,1}, \vartheta_{1,2})$.

(iii) Model 3: $Y_t^{(3)} = (a_1 + \alpha_t)x_t + \varepsilon_t$, where

$$\alpha_t = \vartheta_{1,1}\alpha_{t-1} + \vartheta_{2,1}\alpha_{t-2} + \eta_{t,1}, \quad \varepsilon_t = \vartheta_{1,2}\varepsilon_{t-1} + \vartheta_{2,2}\varepsilon_{t-2} + \eta_{t,3}, \quad x_t^{(3)} = \sin(0.01 \cdot t),$$

$t = 1, \dots, 500$, with $a_1 = 0$, $\vartheta_{1,1} = 1.5$, $\vartheta_{1,1} = -0.75$, $\vartheta_{1,2} = 1.2$ and $\vartheta_{2,2} = -0.3$. In this case $\boldsymbol{a}_0 = a_1$ and $\boldsymbol{\vartheta}_0 = (\vartheta_{1,1}, \vartheta_{2,1}, \vartheta_{1,2}, \vartheta_{2,2})$.

In all the above simulations the independent random variables $\{\eta_{t,1}\}$, $\{\eta_{t,2}\}$ and $\{\eta_{t,3}\}$ are generated from a Student's t-distribution with 6 degrees of freedom, which makes the time series nonGaussian. We note the time series are nonstationary.

We used ordinary least squares (by using the representation of the SCR as a multiple linear regression model with heteroscadastic correlated errors), the Gaussian maximum likelihood and Estimator 2 (the two-step frequency domain estimator) defined in Section 3.3 to estimate the mean regressor coefficient $\boldsymbol{a}_0$ for all the models. We also estimated the time series model parameters $\boldsymbol{\vartheta}_0$ using the Gaussian maximum likelihood and Estimator 2. For Estimator 2 we used different window lengths $m = 10, 20, 50, 200$ and 400 to do the estimation. The results of the estimation for Models 1, 2 and 3 are given in Tables 1, 2 and 3 respectively. Both the GMLE and two stage scheme require initial values in order to minimise the criterions (eg. $\mathcal{L}_T$ and $\mathcal{L}_T^{(m)}$). We used the ordinary least squares estimator for the initial mean regression coefficient and 0.1 as the initial value for all the other coefficients.

We observe that for all models Estimator 2 tends to give better estimates (with smaller bias and mean square errors (MSE) of the mean regressors $\boldsymbol{a}_0$ than the OLS and the GMLE. The same is true when we compare the GMLE and Estimator 2 of $\boldsymbol{\theta}$, Estimator 2 tends to be better. Overall it seems that all values of m give reasonable estimators, however using $m = 200$ for these models, tends to give the smallest mean squared error.

parameter	OLS	GMLE	m=10	m=20	m=50	m=200	m=400
$a_1 = 2$ (average)	2.01	1.99	1.92	1.99	1.98	1.94	1.96
(mse)	(0.0856)	(0.07)	(0.049)	(0.047)	(0.052)	(0.086)	(0.085)
$\vartheta_{1,1} = 0.7$ (average)		0.65	0.89	0.93	0.917	0.56	0.57
(mse)		(0.004)	(0.034)	(0.020)	(0.0119)	(0.009)	(0.008)

Table 1: Fitting Model 1: $Y_t^{(1)} = (a_1 + \alpha_t)x_t + \varepsilon_t$, where $\alpha_t = \vartheta_1\alpha_{t-1} + \eta_{t,1}$. The estimates for the OLS, GMLE and the Estimator 2 (which is calculated using $m = 10, 20, 50, 200$ and 400) are given in each of the columns. Hence we are comparing three estimators and and also the robustness of m. The average estimated value taken over 40 replications is given, together with the mean squared error, which is given in the brackets below the average.

parameter	OLS	GMLE	m=10	m=20	m=50	m=200	m=400
$a_1 = 2$ (average)	2.01	2.01	1.99	2.00	1.99	1.97	1.98
(mse)	(0.094)	(0.08)	(0.050)	(0.050)	(0.056)	(0.089)	(0.089)
$\vartheta_{1,1} = 0.7$ (average)		0.419	(0.85)	0.87	0.77	0.69	0.69
(mse)		(0.084)	(0.041)	(0.040)	(0.026)	(0.026)	(0.0311)
$\vartheta_{1,2} = -0.6$ (average)		0.097	-0.36	-0.49	-0.59	-0.72	-0.715
(mse)		(0.480)	(0.045)	(0.012)	(0.005)	(0.175)	(0.101)

Table 2: Fitting Model 2:$Y_t^{(2)} = (a_1 + \alpha_t)x_t + \varepsilon_t$, where $\alpha_t = \vartheta_{1,1}\alpha_{t-1} + \eta_{t,1}$ and $\varepsilon_t = \vartheta_{1,2}\varepsilon_{t-1} + \eta_{t,2}$. The average estimate and the mean squared error (which is in brackets below). For details see Table 1.

parameter	OLS	GMLE	m=10	m=20	m=50	m=200	m=400
$a_1 = 0$ (average)	-0.17	0.08	-0.17	-0.179	-0.132	-0.127	-0.118
(mse)	(1.045)	(0.008)	(1.034)	(1.151)	(1.227)	(1.240)	(1.274)
$\vartheta_{1,1} = 1.5$ (average)		0.435	1.762	1.41	1.44	1.32	1.44
(mse)		(1.143)	(0.371)	(0.308)	(0.364)	(0.749)	(0.478)
$\vartheta_{2,1} = -0.75$ (average)		0.33	-1.25	-0.63	-0.533	-0.48	-0.54
(mse)		(1.189)	(0.370)	(0.391)	(0.444)	(0.575)	(0.431)
$\vartheta_{1,2} = 1.2$ (average)		0.84	0.98	1.52	1.46	1.37	1.36
(mse)		(0.165)	(0.772)	(0.129)	(0.090)	(0.048)	(0.152)
$\vartheta_{2,2} = -0.3$ (average)		0.65	-0.006	-0.59	-0.59	-0.51	-0.56
(mse)		(0.945)	(0.866)	(0.098)	(0.092)	(0.062)	(0.103)

Table 3: Fitting Model 3: $Y_t^{(3)} = (a_1 + \alpha_t)x_t + \varepsilon_t$, where $\alpha_t = \vartheta_{1,1}\alpha_{t-1} + \vartheta_{2,1}\alpha_{t-2} + \eta_{t,1}$ and $\varepsilon_t = \vartheta_{1,2}\varepsilon_{t-1} + \vartheta_{2,2}\varepsilon_{t-2} + \eta_{t,2}$. The average estimate and the mean squared error (which is in brackets below). For details see Table 1.

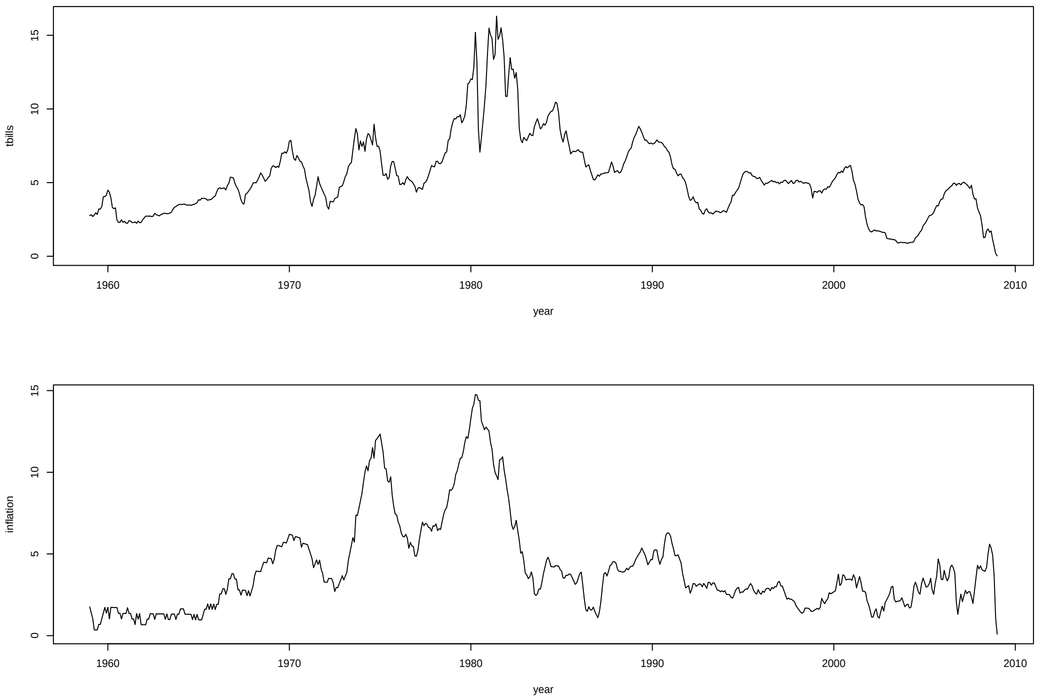

Figure 1: The top plot is 3-month T-bill nominal interest rate taken monthly and lower plot is the monthly inflation rate

6.2 Real data analysis

Example 1: Application to financial time series: Modelling of T-bills and Inflation rates in the US

There are many possible applications of stochastic coefficient regression models in econometrics. One such application is modelling the influence of the nominal interest rate of three month (short term) Treasurey bills (T-bills) on monthly inflation. Fama (1977) argues that the relationship between the T-bills and inflation rate determines whether the market for short term Treasurey bills is efficient or not. In this section we will consider three month T-bills and monthly inflation data observed monthly between January 1959 to December 2008, the data can be obtained from the US Federal reserve, `http://www.federalreserve.gov/releases/h15/data.htm#fn26` and `http://inflationdata.com/inflation/Inflation_Rate/HistoricalInflation.aspx` respectively. A plot of the time series of both sets of observations is given in Figure 1. The estimated correlation coefficient between the three month T-bills and monthly inflation is 0.72. Let Y_t and x_t denote monthly inflation and T-bills interest rate at time t respectively. Fama (1977) and Newbold and Bos (1985) consider the nominal interest rate of three month T-bills and inflation rate data observed every three months between 1953-1980. Fama (1977) fitted the linear regression model $Y_t = a_1 x_t + \varepsilon_t$ ($\{\varepsilon_t\}$ are iid) to the data, and showed that there wasn't a significant departure of a_1 from one, he used this to argue that the T-bills market was efficient. However, Newbold and Bos (1985) argue that the relationship between T-bills and inflation is more complex and suggest that the SCR may be a more appropriate model, where the coefficient of x_t is stochastic and follows an AR(1) model. In other words

$$Y_t = a_0 + (a_1 + \alpha_{t,1})x_t + \varepsilon_t, \quad \alpha_{t,1} = \vartheta_1 \alpha_{t-1,1} + \eta_t \tag{41}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are iid random variables with $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\eta_t) = 0$, $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$ and $\text{var}(\eta_t) = \sigma_\eta^2 < \infty$. Using the GMLE they obtain the parameter estimates $a_0 = -0.97$, $a_1 = 1.09$, $\vartheta_1 = 0.89$, $\sigma_\varepsilon^2 = 1.41$ and $\sigma_\eta^2 = 0.013$. We now fit the same model to the T-bills data observed monthly from January 1959 to December 2008 (600 observations), and use the two-step Estimator 2 to estimate the parameters a_0, a_1, ϑ_1, $\sigma_\alpha^2 = \text{var}(\alpha_{t,1})$ and $\sigma_\varepsilon^2 = \text{var}(\varepsilon_t)$ using Estimator 2. The variances σ_α^2 and σ_ε^2 are estimated in the first step of Estimator 2, the estimates with their standard errors are given Table 4. Note that comparing the estimates with their standard errors we observe that both parameters σ_ε^2 and σ_α^2 appear significant. In the second stage of the scheme we estimate a_0, a_1 and ϑ_1 (we note that because the intercept a_0 appears to be insignificant we also do the estimation excluding the intercept), these estimates are also summarised in Table 4. The estimates for different m are quite close. The model found to be most suitable for the above data when $m = 200$ is

$$Y_t \;=\; (0.73 + \alpha_{t,1})x_t + \varepsilon_t, \quad \alpha_{t,1} = 0.960\alpha_{t-1,1} + \eta_t,$$

where $\sigma_\varepsilon^2 = 1.083^2$ and $\sigma_\alpha^2 = 0.285^2$ (hence $\sigma_\eta^2 = 0.079^2$). We observe that the AR(1) parameter estimate of the stochastic coefficient $\{\alpha_{t,1}\}$ is 0.96. This value is close to one, suggesting that the stochastic coefficients $\{\alpha_{t,1}\}$ could come from a unit root process.

To assess the validity of this model, we obtain one step ahead best linear predictors of Y_t given $\{Y_s\}_{s=1}^{t-1}$ and the current T-bills rate x_t, every month in the year 2008. In order to do the prediction we re-estimate the parameters α_1, θ, σ_ε^2 and σ_α^2 using the observations from 1959-2007. We use the two-step Estimator 2 with $m = 200$ to obtain

$$Y_t \;=\; (0.77 + \alpha_{t,1})x_t + \varepsilon_t, \quad \alpha_{t,1} = 0.965\alpha_{t-1,1} + \eta_t, \tag{42}$$

with $\sigma_\varepsilon^2 = 0.79^2$ and $\sigma_\alpha^2 = 0.30^2$ (hence $\sigma_\eta^2 = 0.18^2$). We also fit the linear regression model $Y_t = a_1 x_t + \varepsilon_t$ to the data and use OLS to obtain the model $Y_t = 0.088 + 0.75x_t + \varepsilon_t$. The predictor using the usual multiple linear regression model and and one-step ahead predictor using the SCR model are given in Figure 2. To do the one-step ahead prediction we use the Kalman filter (using the R package `ss1.R`, see Shumway and Stoffer (2006), Chapter 6 for the details). The mean squared prediction errors over the 12 months using the multiple regression and the SCR model are 8.99 and 0.89 respectively. We observe from the plots in Figure 2 that the multiple regression model always underestimates the true value and the mean square error is substantially larger than the SCR model.

Example 2: Application to environmental time series: Modelling of visibility and air pollution

It is known that air visibility quality depends on the amount of particulate matter (particulate matter negatively effects visibility). Furthermore, air pollution is known to influence the amount of particulate matter (see Hand et al. (2008)). To model the influence of air pollution on particulate matter Hand et al.

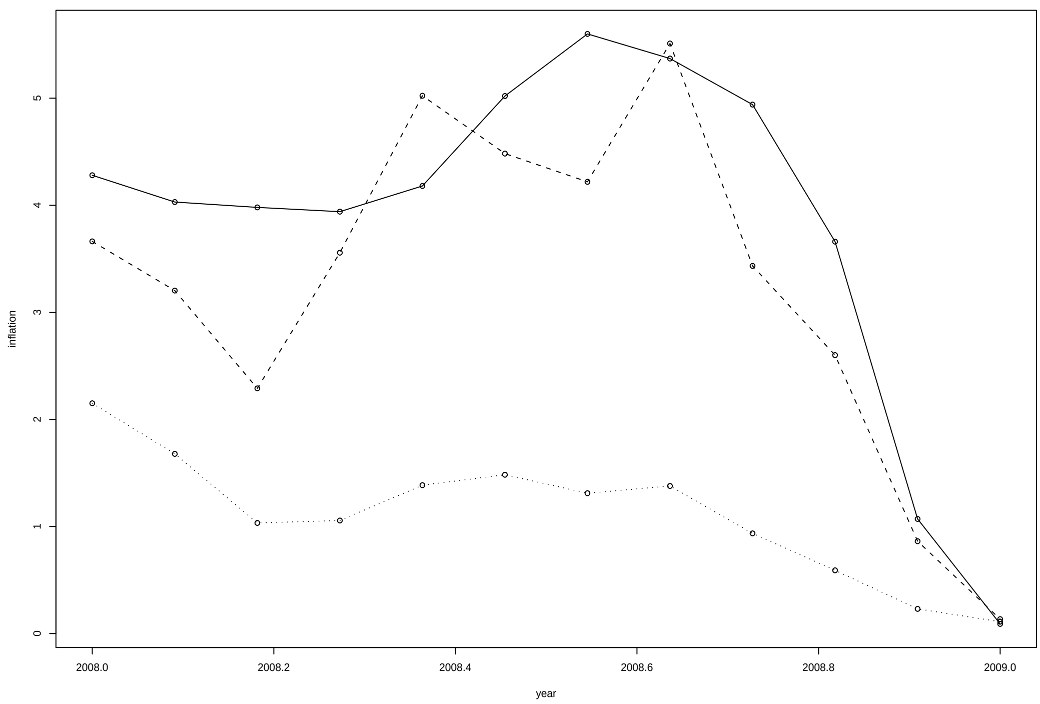

Figure 2: We compare the true inflation rates with their predictions. The continuous thick line $-$ is the true inflation rate. The broad dashed line $--$ is SCR one step ahead predictor given in (42). The fine dashed line $\cdots$ is the linear regression predictor $\hat{Y}_t = 0.088 + 0.750x_t$.

	a_0	a_1	ϑ_1	σ_α	σ_ε
OLS	0.088	0.750			
(s.e.)	(0.18)	(0.029)			
Stage 1	0.088	0.74		0.285	1.083
(s.e.)				(0.011)	(0.059)
$m = 10$ (with intercept)	0.618	0.625	0.981		
(s.e.)	(0.325)	(0.069)	(0.042)		
$m = 10$ (without intercept)		0.741	(0.971)		
(s.e.)		(0.0325)	(0.05)		
$m = 50$ (with intercept)	0.309	0.687	0.969		
(s.e.)	(0.35)	(0.069)	(0.026)		
$m = 50$ (without intercept)		0.743	0.957		
(s.e.)		(0.032)	(0.038)		
$m = 200$ (with intercept)	0.223	0.7327	0.96088		
(s.e.)	(0.44)	(0.022)	(0.024)		
$m = 200$(without intercept)		0.765	0.951		
(s.e.)		(0.029)	(0.030)		
$m = 400$ (with intercept)	0.367	0.725	0.963		
(s.e.)	(0.48)	(0.070)	(0.023)		
$m = 400$(without intercept)		0.773	0.957		
(s.e.)		(0.029)	(0.026)		

Table 4: We fit the model $Y_t = a_0 + (a_1 + \alpha_{t,1})x_t + \varepsilon_t$, where $\alpha_{t,1} = \vartheta_1 \alpha_{t-1,1} + \eta_t$, with and without the intercept a_0. The estimates using least squares and the frequency domain estimator for different m are given. The values in the brackets are the corresponding standard errors.

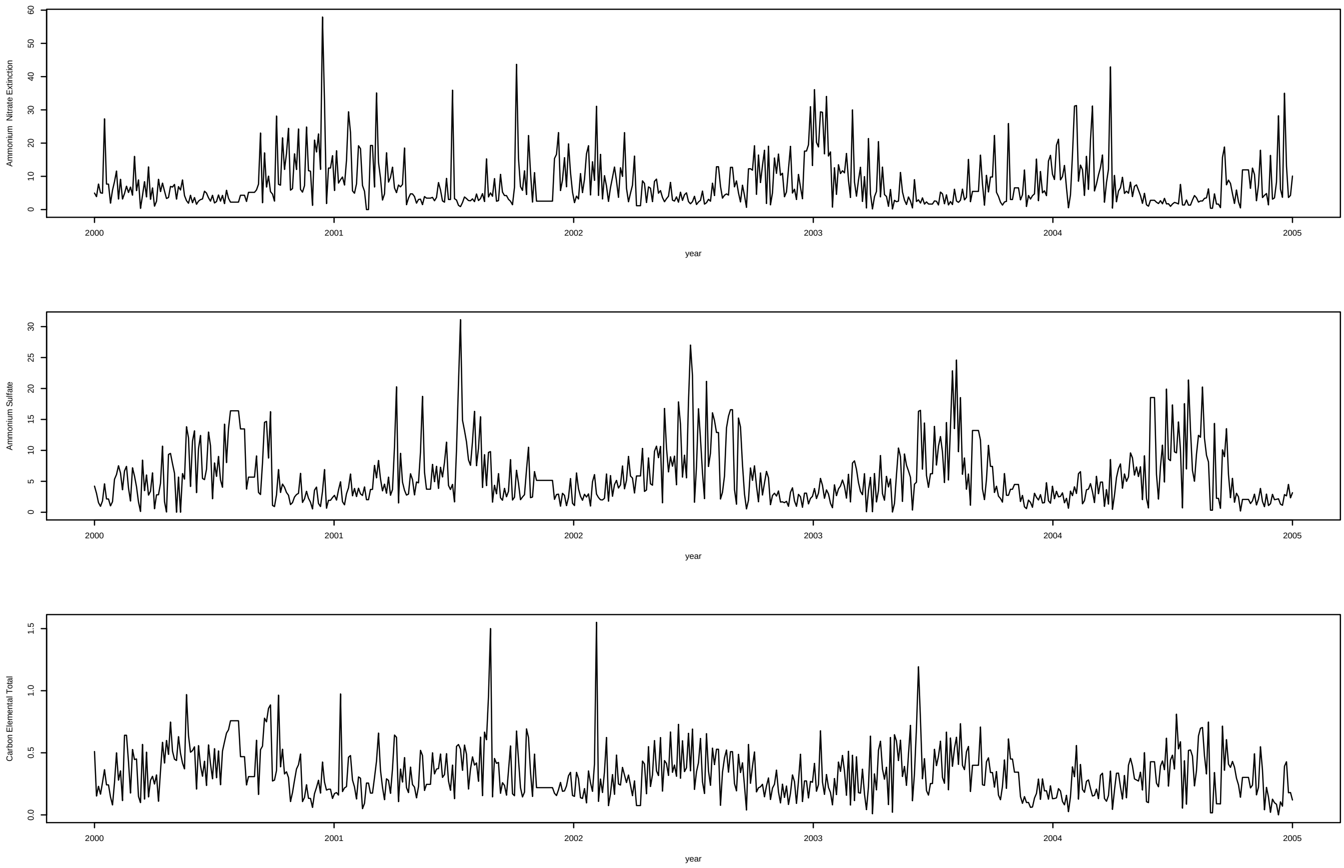

Figure 3: The top plot is 3-day Ammonium Nitrate Extinction (Fine), middle plot is 3-day Ammonium Sulfate (Fine) and lower plot is three day Carbon Elemental Total (Fine)

(2008) (see equation (6)) fit a linear regression model. However, Burnett and Guthrie (1970) argue that the influence of air pollution on air visibility may vary each day, depending on meteorological conditions, and suggest that a SCR model may be more appropriate than a multiple linear regression model. In this section we investigate this possibility. We consider the influence of man made emissions on Particulate Matter (PM2.5-10) in Shenandoah National Park, Virginia, USA. The data we consider is Ammonium Nitrate Extinction (ammNO3f), Ammonium Sulfate (ammSO4f), Carbon Elemental Total (ECF) and Particulate Matter (PM2.5-10) (ammNO3f, ammSO4f and ECF are measured in ug/m3) which has been collected every three days between 2000-2005 (600 observations). We obtain the data from the VIEWS website `http://vista.cira.colostate.edu/views/Web/Data/DataWizard.aspx`. We mention that the influence of man made emissions on air visibility is of particular importance to the US national parks service (NPS), who collected and compiled this data. An explanation of the data and how air pollution influences visibility (light scattering) can be found in Hand et al. (2008).

The plots of both the air pollution and PM2.5-10 data is given in Figure 3 and 4 respectively. There is a clear seasonal component in all the data sets as seen from their plots. Therefore to prevent spurious correlation between the PM2.5-10 and air pollution we detrended and deseasonalised the PM2.5-10 and

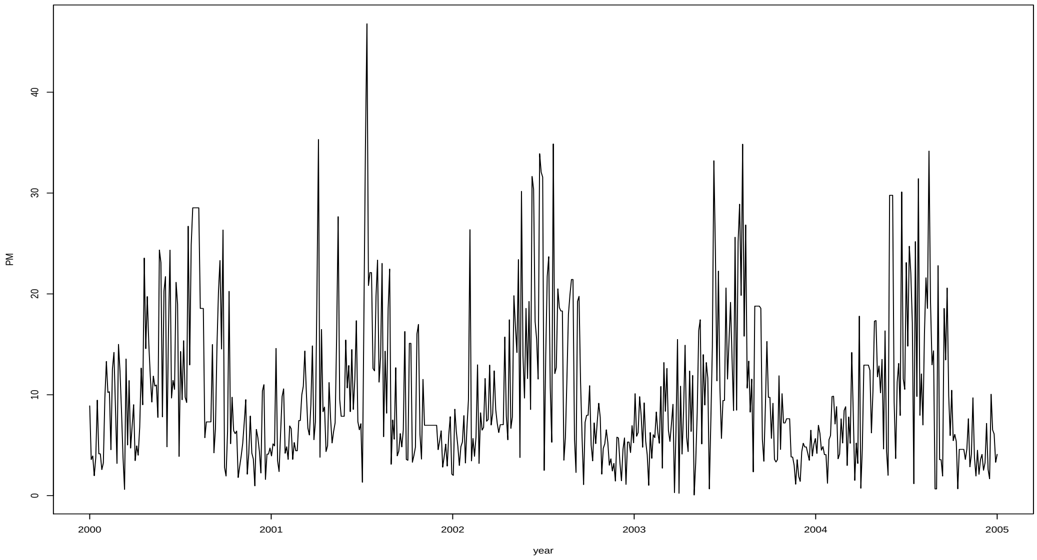

Figure 4: The plot is 3-day Particulate Matter (PM2.5 - PM10)

emissions data. To identify the dominating harmomics we used the maximum periodogram methods suggested in Quinn and Fernandez (1991) and Kavalieris and Hannan (1994). To the detrended and deseasonalise PM2.5-10 and air pollution data we fitted the following model

$$Y_t = (a_1 + \alpha_{t,1})x_{t,1} + (a_2 + \alpha_{t,2})x_{t,2} + (a_3 + \alpha_{t,3})x_{t,3} + \varepsilon_t,$$

where $\{x_{t,1}\}$, $\{x_{t,2}\}$, $\{x_{t,3}\}$ and $\{Y_t\}$ are the detrended and deseasonalised ammNO3f, ammSO4f, ECF and Particulate Matter (PM2.5-10), and $\{\alpha_{t,j}\}$ and ε_t satisfy

$$\alpha_{t,j} = \vartheta_j \alpha_{t-1,j} + \eta_{t,j}, \quad \text{for and } j = 1, 2, 3, \qquad \varepsilon_t = \vartheta_4 \varepsilon_{t-1} + \eta_{t,4},$$

$\{\eta_{t,i}\}$ iid random variables. Let $\sigma_\varepsilon^2 = \text{var}(\varepsilon_t)$, $\sigma_{\alpha,1}^2 = \text{var}(\alpha_{t,1})$, $\sigma_{\alpha,2}^2 = \text{var}(\alpha_{t,2})$ and $\sigma_{\alpha,3}^2 = \text{var}(\alpha_{t,3})$.

We used Estimator 2 to estimate the parameters a_1, a_2, a_3, $\alpha_{t,1}, \alpha_{t,2}, \alpha_{t,3}, \alpha_{t,4}$, σ_ε^2, $\sigma_{\alpha,1}^2$, $\sigma_{\alpha,2}^2$ and $\sigma_{\alpha,3}^2$. At the start off the minimisations of the objective functions $\mathcal{L}_T$ and $\mathcal{L}_T^{(m)}$, as initial values we gave the least squares estimates of $\boldsymbol{a}_0$ for the mean regression coefficients and 0.1 for all the other unknown parameters. In the first stage of the scheme we estimated a_1, a_2, a_3 and $\sigma_\varepsilon^2, \sigma_{\alpha,1}^2, \sigma_{\alpha,2}^2$ and $\sigma_{\alpha,3}^2$. We also fitted parsimonious models where some of the coefficients were kept fixed rather stochastic. The results are summarised in Table 5 (step 1). We observe that the estimate of $\sigma_{\alpha,1}$ is extremely small and insignificant. We observe that the minimum value of the objective function $\mathcal{L}_T$ is about the same when the coefficient of ammNO3f $\{x_{t,1}\}$ is fixed and random (it is 2.012). This suggests that the coefficient of ammNO3f $\{x_{t,1}\}$ is deterministic. This may indicate that the relative contribution of NO3f ($\{x_{t,1}\}$) to the response is constant throughout the period of time and is not influenced by any other extraneous factors. We re-did the minimisation systematically removing $\sigma_{\alpha,2}$ and $\sigma_{\alpha,3}$, but the minimum value of $\mathcal{L}_T$, changed quite substantially (compare the minimum of the objective functions 2.012 with $4.09, 3.99$ and 4.15). Hence the most appropriate model appears to be

$$Y_t = a_1 x_{t,1} + (a_2 + \alpha_{t,2})x_{t,2} + (a_3 + \alpha_{t,3})x_{t,3} + \varepsilon_t,$$

where $\{\alpha_{t,1}\}$ and $\{\alpha_{t,2}\}$ are stochastic coefficients. It is of interest to investigate whether the coefficients of ammSO4f and ECF are purely random or correlated, and we investigate this in the second stage of the frequency domain scheme, where we modelled $\{\alpha_{t,2}\}$ and $\{\alpha_{t,3}\}$ both as iid random variables and as the AR(1) model $\alpha_{t,j} = \vartheta_j \alpha_{t-1,j} + \eta_{t,j}$, for $j = 2, 3$. The estimates for various different models and different values of m are given in Table 5. If we compare the minimum of the objective function where $\{\alpha_{t,2}\}$ and $\{\alpha_{t,3}\}$ are modelled as both iid and satisfying an AR(1) model, we see that there is very little difference between them. Moreover the standard errors for the estimates of ϑ_2 and ϑ_3, are large. Altogether, this suggests that ϑ_2 and ϑ_3 are not significant and $\{\alpha_{t,2}\}$ and $\{\alpha_{t,3}\}$ are uncorrelated over time. Hence it seems plausible that the coefficients of ammSO4f and ECF are random, but independent. To check the possibility that the errors $\{\varepsilon_t\}$ are correlated, we fitted an AR(1) model to the errors. However we observe from Table 5, that the AR(1) parameter does not appear to be significant. Moreover, comparing the minimum of the objective function $\mathcal{L}_{600}^{(m)}$ (for different values of m) fitting iid $\{\varepsilon_t\}$ and an AR(1) to $\{\varepsilon_t\}$ gives almost the same value. This suggests that the errors are independent. In summary, our analysis suggests that the influence of ammNO3f on PM2.5-10 is fixed over time, whereas the influence of ammSO4f and ECF varies purely randomly over time. Using the estimator obtained when $m = 200$ this suggests the model

$$Y_t = 0.255 x_{t,1} + (4.58 + \alpha_{t,2}) x_{t,2} + (1.79 + \alpha_{t,3}) x_{t,3} + \varepsilon_t,$$

where $\{\alpha_{t,2}\}$ and $\{\alpha_{t,3}\}$ are iid random variables, with $\sigma_{\alpha,2} = 1.157$, $\sigma_{\alpha,3} = 0.84$, $\sigma_\varepsilon = 1.296$. Based on our analysis it would appear that the coefficients of pollutants are random, but there is no linear dependence between the current coefficient and the previous coefficient. On possible explanation for the lack of dependence is that the data is taken every three days and not daily. This could mean that the meterological conditions from three days ago has little influence on today's particulate matter. On the other hand if we were to analyse the daily pollutants and daily PM2.5-10 the conclusions could have been different. But this daily data is not available. It is likely that since the data is aggregated (smoothed) over a three day period any possible dependence in the data was removed.

Acknowledgements

The author wishes to thank Dr. Bret Schichtel, at CIRA, Colorado State University for his invaluable advice and explanations of the VIEWs data. This work has been partially supported by the NSF grant under DMS-0806096 and the Deutsche Forschungsgemeinschaft under DA 187/15-1.

	a_1	a_2	a_3	ϑ_2	ϑ_3	ϑ_4	$\sqrt{\text{var}(\alpha_{t,1})}$	$\sqrt{\text{var}(\alpha_{t,2})}$	$\sqrt{\text{var}(\alpha_{t,3})}$	$\sqrt{\text{var}(\varepsilon_t)}$	minL
OLS	0.29	4.57	1.76								
	(0.078)	(0.088)	(0.0908)								
Stage 1	0.38	4.53	1.58				7.10^{-7}	1.25	0.84	1.29	2.102
(s.e.)	(0.048)	(0.079)	(0.076)				(0.07)	(0.097)	(0.098)	(0.046)	
Stage 1	0.56	3.28	-3.09				0.75		3.315	4.496	4.09
(s.e.)	(0.170)	(0.181)	(0.29)				(0.042)		(0.159)	(0.049)	
Stage 1	0.387	4.53	1.58					1.157	0.84	1.296	2.102
(s.e.)	(0.048)	(0.080)	(0.076)					(0.009)	(0.009)	(0.002)	
Stage 1	0.287	3.52	-3.11						3.00	3.83	3.99
(s.e.)	(0.139)	(0.171)	(0.27)						(0.164)	(0.07)	
Stage 1	0.521	5.09	-2.93							4.42	4.15
(s.e.)	(0.131)	(0.145)	(0.65)							(0.011)	
m=10	0.393	4.47	1.630	0.883	-0.144						2.066
(s.e.)	(0.048)	(0.079)	(0.076)	(0.095)	(0.351)						
m=10	0.39	4.47	1.63		-0.13						2.066
(s.e.)	(0.05)	(0.077)	(0.071)		(0.34)						
m=10	0.388	4.467	1.661			0.94					2.084
(s.e.)	(0.039)	(0.061)	(0.058)			(0.029)					
m=50	0.327	4.55	1.75	0.617	-0.235						2.199
(s.e.)	(0.056)	(0.070)	(0.071)	(0.30)	(0.659)						
m=50	0.324	4.54	1.746		-0.32						2.22
(s.e.)	(0.055)	(0.073)	(0.070)		(0.54)						
m=50	0.308	4.55	1.764			0.244					2.21
(s.e.)	(0.049)	(0.062)	(0.062)			(0.127)					
m=200	0.261	4.595	1.770	0.538	0.9722						2.111
(s.e.)	(0.06)	(0.067)	(0.070)	(0.322)	(0.042)						
m=200	0.261	4.60	1.77		-0.087						2.124
(s.e.)	(0.06)	(0.068)	(0.068)		(1.2)						
m=200	0.255	4.58	1.797			0.458					2.1339
(s.e.)	(0.051)	(0.058)	(0.058)			(0.145)					
m=400	0.2531	4.597	1.793	0.979	0.932						2.116
(s.e.)	(0.061)	(0.068)	(0.070)	(0.032)	(0.143)						
m=400	0.25	4.59	1.79		0.92						2.128
(s.e.)	(0.06)	(0.068)	(0.068)		(0.167)						
m=400	0.250	4.580	1.807			0.478					2.139
(s.e.)	(0.051)	(0.0583)	(0.059)			(0.148)					

Table 5: In stage 1 we fitted the model $Y_t = (a_1 + \alpha_{t,1})x_{t,1} + (a_2 + \alpha_{t,2})x_{t,2} + (a_3 + \alpha_{t,3})x_{t,3} + \varepsilon_t$, and various subsets (here we did not model any dependence in the stochastic coefficients). In the second stage (for $m = 10, 50, 200$ and 400) we fitted AR(1) models to $\{\alpha_{t,2}\}$, $\{\alpha_{t,3}\}$ and $\{\varepsilon_t\}$, that is $\alpha_{t,2} = \vartheta_2 \alpha_{t-1,2} + \eta_{t,2}$, $\alpha_{t,3} = \vartheta_3 \alpha_{t-1,3} + \eta_{t,3}$ and $\varepsilon_t = \vartheta_4 \varepsilon_{t-1} + \eta_{t,4}$. The value of the frequency domain likelihood at the minimal value is also given in the column minL. The standard errors of the estimates are given below each estimate in brackets.

Reference

Brillinger, D. R. (2001). *Time series data analysis and theory.* Philadelphia: SIAM.

Bruesch, T. S., & Pagan, A. R. (1980). A simple test for for heteroscedasicity and random coefficient variation. *Econmetrica*, *47*, 1287-1294.

Burnett, T. D., & Guthrie, D. (1970). Estimation of stationary stochastic regression parameters. *J. American Statistical Association*, *65*, 1547-1553.

Caines, P. (1988). *Linear stochastic systems.* New York: Wiley.

Dahlhaus, R. (1996). Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonparametric Statist.*, *6*, 171-191.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Stat.*, *16*, 1-37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Stat.*, *28*, 1762-1794.

Duncan, D. B., & Horn, S. (1972). Linear dynamic recursion estimation from the viewpoint of regression analysis. *J. American Statistical Association*, *67*, 815-821.

Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Ann. Statist.*, *7*, 490-506.

Dzhapharidze, K. (1971). On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary Gaussian processes with rational spectral density. *Theory Probab. Appl.*, *16*, 550-554.

Fama, E. F. (1977). Interest rates and inflation: the message in the entrails. *American Economic Review*, *67*, 487-496.

Franke, J., & Gründer, B. (1995). Athens conference in applied probability and time series analysis: Vol ii. In P. R. und M. Rosenblatt (Ed.), (p. 177-189). New York: Springer.

Giraitis, L., & Robinson, P. (2001). Whittle estimation of ARCH models. *Econometric Theory*, *17*, 608-631.

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

Hand, J. L., , & Malm, W. C. (2008). *Review of the improve equation for estimating ambient light extinction coefficients - final report* (Tech. Rep.). `http://vista.cira.colostate.edu/improve/Publications/GrayLit/016_IMPROVEeqReview/IMPROVEeqReview.htm.`

Hannan, E. J. (1971). Non-linear time series regression. *J. Appl. Probab.*, *8*, 767-780.

Hannan, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.*, *10*, 130-145.

Hildreth, c., & Houck, C. (1968). Some estimates for a linear model with random coefficients. *Journal of the American Statistical Association*, *63*, 584-595.

Kavalieris, L., & Hannan, E. J. (1994). Determining the number of terms in a periodic regression. *Journal of Time Series Analysis*, *15*.

Ljung, L., & Caines, P. (1979). Asymptotic normality of prediction error estimators for approximate system models. *Stochastics*, *3*, 29-46.

Martinussen, T., & Scheike, T. H. (2000). A nonparametric dynamic additive regression model for longitutal data. *Ann. Statist.*, 1000-1025.

Newbold, P., & Bos, T. (1985). *Stochastic parameter regression models.* Beverly Hills: Sage Publications.

Pfeffermann. (1984). On extensions of the gauss-markov theorem to the case of stochastic-regression coefficient. *Journal of the Royal Statistical Society, B*, *46*, 139-148.

Picinbono, B. (1996). Second-order complex random vectors and normal distributions. *IEEE trans. Signal Processing*, *44*, 2637-2640.

Quinn, B. G., & Fernandez, J. M. (1991). A fast and efficient technique for the estimation of frequency. *Biometrika*, *78*, 489-498.

Rosenberg, B. (1972). The estimation of stationary stochastic regression parameters reexamined. *Journal of the American Statistical Association*, *67*, 650-654.

Rosenberg, B. (1973). Linear regression with randomly dispersed parameters reexamined. *Biometrika*, *60*, 65-72.

Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications (with R examples).* New York: Springer.

Stoffer, D. S., & Wall, K. D. (1991). Bootstrappping state space models: Gaussian maximum likelihood estimation. *Journal of the American Statistical Association*, *86*, 1024-1033.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, *38*, 311-323.

Swamy, P. A. V. B. (1971). *Statistical inference in random coefficient models.* Berlin: Springer.

Swamy, P. A. V. B., & Tinsley, P. A. (1980). Linear prediction and estimation methods for regression model with stationary stochastic coefficients. *Journal of Econometrics*, *12*, 103-142.

Synder, R. D. (1985). Recursive estimation of dynamic linear models. *J. Royal Statistical Society, B*, *47*, 272-276.

Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.*, *11*, 157-169.

Walker, A. M. (1964). Asymptotic properties of least squares estimates of parameters of the spectrum of nonstationary non-deterministic time series. *J. Australian mathematical society*, *4*, 363-384.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, *39*, 105-129.

A Appendix

A.1 Proofs for Section 2.2.2

We first prove Proposition 2.2.

PROOF of Proposition 2.2 We first observe that because $\tilde{X}_{t,N}$ is locally stationary there exists a function $A(\cdot)$, such that

$$\tilde{X}_{t,N} = \int A(\frac{t}{N},\omega)\exp(it\omega)dZ(\omega) + O_p(N^{-1}). \tag{43}$$

Now since $A(\cdot) \in L_2([0,1]\otimes[-\pi,\pi])$, for any δ there exists a n_δ such that

$$\big|A(v,\omega) - \sum_{j=1}^{n_\delta} A_j(\omega)x_j(u)\big| \leq K\delta,$$

where $A_j(\omega) = \int A(v,\omega)x_j(v)dv$, and K is a finite constant independent of δ. Substituting this into (44) gives

$$\tilde{X}_{t,N} = \sum_{j=1}^{n_\delta}\left\{\int A_j(\omega)\exp(it\omega)dZ(\omega)\right\}x_j(\frac{t}{N}) + O_p(\delta + N^{-1}). \tag{44}$$

Let $\alpha_{t,j} = \int A_j(\omega)\exp(it\omega)dZ(\omega)$, since $Z(\omega)$ is a orthogonal increment process with $\mathbb{E}|dZ(\omega)|^2 = d\omega$ it is clear that $\{\boldsymbol{\alpha}_t = (\alpha_{t,1},\ldots,\alpha_{t,n})\}$ is a second order stationary vector time series. $\square$

A.2 Proofs for Section 5.2

We first study $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$. Using (11), we note that $\mathcal{I}_{t,m}(\boldsymbol{a},\omega)$ can be written as

$$\mathcal{I}_{t,m}(\boldsymbol{a},\omega) = \left(J_{t,m}^{(X)}(\omega) + \sum_{j=1}^{n}(a_{j,0}-a_j)J_{t,m}^{(j)}(\omega)\right)\left(J_{t,m}^{(X)}(-\omega) + \sum_{j=1}^{n}(a_{j,0}-a_j)J_{t,m}^{(j)}(-\omega)\right), \tag{45}$$

where

$$J_{t,m}^{(X)}(\omega) = \frac{1}{\sqrt{2\pi m}}\sum_{k=1}^{m} X_{t-m/2+k}\exp(ik\omega). \tag{46}$$

Substituting (45) into $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ gives

$$\begin{aligned}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) &= \frac{1}{T_m}\Big\{V_T(\mathcal{F}_{\boldsymbol{\theta}}^{-1}) + 2\sum_{j=1}^{n}(a_{j,0}-a_j)\Re(D_T^{(j)}(\mathcal{F}_{\boldsymbol{\theta}}^{-1})) + \\ &\sum_{j_1,j_2=1}^{n}(a_{j_1,0}-a_{j_1})(a_{j_2,0}-a_{j_2})H_T^{(j_1,j_2)}(\mathcal{F}_{\boldsymbol{\theta}}^{-1}) + \sum_{t=m/2}^{T-m/2}\int\log\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)d\omega\Big\},\end{aligned}$$

where $\Re(z)$ denotes the real part of the complex variable z,

$$V_T(\mathcal{G}) = \sum_{t=m/2}^{T-m/2}\int\mathcal{G}_t(\omega)|J_{t,m}^{(X)}(\omega)|^2 d\omega = \frac{1}{2\pi m}\sum_{k=1}^{m}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k}X_s X_{s+r}g_{s+m/2-k}(r), \tag{47}$$

$$
\begin{aligned}
H_T^{(j_1,j_2)}(\mathcal{G}) &= \sum_{t=m/2}^{T-m/2}\int \mathcal{G}_t(\omega)J_{t,m}^{(j_1)}(\omega)J_{t,m}^{(j_2)}(-\omega)d\omega = \frac{1}{2\pi m}\sum_{k=1}^{m}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} x_{s,j_1}x_{s+r,j_2}g_{s+m/2-k}(r)\\
\Re\{D_T^{(j)}(\mathcal{G})\} &= \sum_{t=m/2}^{T-m/2}\int \mathcal{G}_t(\omega)\Re\{J_{t,m}^{(X)}(\omega)J_{t,m}^{(j)}(-\omega)\}d\omega = \frac{1}{m}\sum_{k=1}^{m}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} X_s x_{s+r,j}\tilde{g}_{s+m/2-k}(r)
\end{aligned}
$$

where $g_s(r) = \int G_s(\omega)\exp(ir\omega)d\omega$ and $\tilde{g}_s(r) = \int G_s(\omega)\cos(r\omega)d\omega$. In order to numerically minimise $\mathcal{L}_T^{(m)}$ and obtain the sampling properties of $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$ it's necessary to consider the first and second derivatives of $\mathcal{L}_T^{(m)}$. Let $\nabla = (\frac{\partial}{\partial a_1},\ldots,\frac{\partial}{\partial \boldsymbol{\theta}_q})$ and $\nabla\mathcal{L}_T^{(m)} = (\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}, \nabla_\theta\mathcal{L}_T^{(m)})$, where

$$
\begin{aligned}
\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) &= \frac{1}{T_m}\Big\{\big(V_T(\nabla_{\boldsymbol{\theta}}\mathcal{F}^{-1}) + 2\sum_{j=1}^{n}(a_{j,0}-a_j)\Re[D_T^{(j)}(\nabla_{\boldsymbol{\theta}}\mathcal{F}^{-1})]\big)\\
&\quad + \sum_{j_1,j_2}(a_{j_1,0}-a_{j_1})(a_{j_2,0}-a_{j_2})H_T^{(j_1,j_2)}(\nabla_{\boldsymbol{\theta}}\mathcal{F}^{-1}) + \sum_{t=m/2}^{T-m/2}\int\frac{\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}d\omega\Big\}\\
\nabla_{a_j}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) &= \frac{-2}{T_m}\big\{\Re[D_T^{(j)}(\mathcal{F}^{-1})] + \sum_{j_1=1}^{n}(a_{j_1,0}-a_{j_1})H_T^{(j,j_1)}(\mathcal{F}^{-1})\big\} \qquad (48)
\end{aligned}
$$

and the second derivatives are

$$
\begin{aligned}
\nabla_{a_j}\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) &= \frac{-2}{T_m}\big\{\Re[D_T^{(j)}(\nabla_{\boldsymbol{\theta}}\mathcal{F}^{-1})] + \sum_{j_1=1}^{n}(a_{j_1,0}-a_{j_1})H_T^{(j,j_1)}(\nabla_{\boldsymbol{\theta}}\mathcal{F}^{-1})\big\}\\
\nabla_{a_{j_1}}\nabla_{a_{j_2}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) &= \frac{2}{T_m}H_T^{(j_1,j_2)}(\nabla_{\boldsymbol{\theta}}\mathcal{F}^{-1}) \qquad (49)\\
\nabla^2_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) &= \frac{1}{T_m}\Big[V_T(\nabla^2_{\boldsymbol{\theta}}\mathcal{F}^{-1}) + 2\sum_{j=1}^{n}(a_{j,0}-a_j)\Re\{D_T^{(j)}(\nabla^2_{\boldsymbol{\theta}}\mathcal{F}^{-1})\} +\\
&\quad \sum_{j_1,j_2=1}^{n}(a_{j_1,0}-a_{j_1})(a_{j_2,0}-a_{j_2})H_T^{(j_1,j_2)}(\nabla^2_{\boldsymbol{\theta}}\mathcal{F}^{-1})\\
&\quad +\frac{1}{T_m}\sum_{t=m/2}^{T-m/2}\Big\{\int\Big[\frac{\nabla^2_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)} - \frac{\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)'}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)^2}\Big]d\omega\Big\}\Big]. \qquad (50)
\end{aligned}
$$

It is worth noting that since $\nabla\mathcal{F}_{s,m}$ is real, then $\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}$ and $\nabla_{a_j}\mathcal{L}_T^{(m)}$ are also real.

Lemma A.1 *Suppose Assumptions 2.1, 5.1(i-iii) (and $\sup_j \mathbb{E}(\eta_{t,j}^4) < \infty$) and 5.2(i) are satisfied. Let $V_T(\mathcal{G})$ and $D_T^{(j)}(\mathcal{H})$ be defined as in (47), $g_t(r) = \int \mathcal{G}_t(\omega)\exp(ir\omega)d\omega$, $\tilde{g}_t(r) = \int \mathcal{G}_t(\omega)\cos(r\omega)d\omega$ and suppose $\sup_t \sum_r |g_t(r)| < \infty$ and $\sup_t \sum_r |\tilde{g}_t(r)| < \infty$. Then we have $\mathbb{E}(D_T^{(j)}(\mathcal{G})) = 0$,*

$$
\mathbb{E}(V_T(\mathcal{G})) = \sum_{t=m/2}^{T-m/2}\int \mathcal{G}_t(\omega)\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)d\omega, \qquad (51)
$$

$$
var(V_T(\mathcal{G})) \leq T(n+1)\sup_s\sum_r|g_s(r)|\left\{2(\sum_r\rho_2(r))^2 + \sum_{k_1,k_2,k_3}\rho_4(k_1,k_2,k_3)\right\} \qquad (52)
$$

and

$$var(D_T^{(j)}(\mathcal{G})) \leq T(n+1)\sup_s(\sum_r |g_s(r)|)(\sum_r \rho_2(r)), \quad (53)$$

where

$$\rho_2(k) = \kappa_2^2 \sup_j \sum_i |\psi_{i,j}| \cdot |\psi_{i+k,j}|, \quad \rho_4(k_1,k_2,k_3) = \kappa_4 \sup_j \sum_i |\psi_{i,j}| \cdot |\psi_{i+k_1,j}| \cdot |\psi_{i+k_2,j}| \cdot |\psi_{i+k_3,j}|, \quad (54)$$

$\kappa_2 = \sup_j var(\eta_{0,j})$ *and* $\kappa_4 = \sup_j cum(\eta_{0,j},\eta_{0,j},\eta_{0,j},\eta_{0,j})$.

PROOF. The proof of (51) is straightforward, hence we omit the details. To prove (52) we first note that $|\mathrm{cov}(X_t, X_s)| \leq (n+1)\sup_{t,j}|x_{t,j}|^2\rho_2(t-s)$ and $|\mathrm{cum}(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})| \leq (n+1)\sup_{t,j}|x_{t,j}|^4\rho_4(t_2 - t_1, t_3 - t_1, t_4 - t_1)$. Expanding $\mathrm{var}(V_T(\mathcal{H}))$ gives

$$\mathrm{var}(V_T(\mathcal{H})) = \frac{1}{(2\pi m)^2}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m/2+k_1}\sum_{u=k_2}^{T-m/2+k_2}\sum_{r_1=-k_1}^{m-k_1}\sum_{r_2=-k_2}^{m-k_2}\mathrm{cov}(X_sX_{s+r_1}, X_u, X_{u+r_2})h_s(r_1)h_u(r_2). (55)$$

From the definition of $\rho_2(\cdot)$ and $\rho_4(\cdot)$ it is clear that

$$\begin{aligned} & |\mathrm{cov}(X_sX_{s+r_1}, X_u, X_{u+r_2})| \\ \leq\ & \rho_2(s-u)\rho_2(s+r_1-u-r_2) + \rho_2(s+r_1-u)\rho_2(s-u-r_2) + \rho_4(r_1, u-s-r_1, u-s+r_1-r_2). \end{aligned}$$

Substituting the above into (55) gives us (53). Using the same methods we have (53). □

We use the following lemma to prove consistency of the estimator and obtain the rate in (20).

Lemma A.2 *Suppose Assumptions 2.1, 5.1(i-v) and 5.2(i) are satisfied. Then for all* $\boldsymbol{a} \in \Omega$ *and* $\boldsymbol{\theta} \in \Theta$ *we have*

$$var\Big(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})\Big) = O(\frac{1}{T_m}), \quad var\Big(\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})\Big) = O(\frac{1}{T_m}), \quad var\Big(\nabla^2\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})\Big) = O(\frac{1}{T_m}). \quad (56)$$

PROOF. To prove (56), we use Lemma A.1, which means we have to verify the condition $\sum_r |\int \nabla^k_{\boldsymbol{\theta}}\mathcal{F}_{t,m}^{-1}(\boldsymbol{\theta},\omega)\exp(ir\omega)d\omega| < \infty$, for $k = 0,1,2$. This follows from (111) (in Lemma A.9), hence we obtain the result. □

To show convergence in probability of the estimator we need to show equicontinuity in probability of the sequence of random functions $\{\mathcal{L}_T^{(m)}\}_T$. We recall that a sequence of random functions $\{\mathcal{H}_T(\boldsymbol{\theta})\}$ is equicontinuous in probability if for every $\epsilon > 0$ and $\eta > 0$ there exists a $\delta > 0$ such that $\lim_{T\to\infty}\sup P(\sup_{\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_1\leq\delta}|\mathcal{H}_T(\boldsymbol{\theta}_1) - \mathcal{H}_T(\boldsymbol{\theta}_2)| > \eta) < \varepsilon$. In our case $\mathcal{H}_T(\boldsymbol{\theta}_1) := \mathcal{L}_T^{(m)}$ and pointwise it converges to its expectation

$$\begin{aligned} & \mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})) \quad (57) \\ =\ & \frac{1}{T_m}\sum_{t=m/2}^{T-m/2}\int\Big\{\frac{\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)} + \sum_{j_1,j_2=1}^{n}(a_{j_1,0}-a_{j_1})(a_{j_2,0}-a_{j_2})\frac{J_{t,m}^{(j_1)}(\omega)J_{t,m}^{(j_2)}(-\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)} + \log\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)\Big\}d\omega, \end{aligned}$$

Lemma A.3 *Suppose Assumptions 2.1, 5.1(i-v) and 5.2(i) are satisfied, then* *(i)* $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ *is equicontinuous is probability and* $|\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) - \mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))| \stackrel{\mathcal{P}}{\to} 0$

(ii) $\nabla^2\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ *is equicontinuous in probability and* $|\nabla^2\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) - \mathbb{E}(\nabla^2\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))| \stackrel{\mathcal{P}}{\to} 0,$

as $T \to \infty$.

PROOF. We first prove (i). By the mean value theorem

$$|\mathcal{L}_T^{(m)}(\boldsymbol{a}_1,\boldsymbol{\theta}^{(1)}) - \mathcal{L}_T^{(m)}(\boldsymbol{a}_2,\boldsymbol{\theta}^{(2)})| \leq \sup_{\boldsymbol{a}\in\Omega,\boldsymbol{\theta}\in\Theta_1\otimes\Theta_2} |\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})| \cdot |(\boldsymbol{a}_1-\boldsymbol{a}_2,\boldsymbol{\theta}^{(1)}-\boldsymbol{\theta}^{(2)})|, \tag{58}$$

where $\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) = (\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}), \nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))$. From above it is clear if we can show $\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ is bounded in probability, then equicontinuity of $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ immediately follows. Therefore we now bound $\sup_{\boldsymbol{a}\in\Omega\boldsymbol{\theta}\in\Theta_1\otimes\Theta_2}|\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})|$, with a random variable which is bounded in probability. Inspecting (48) and under Assumption 5.1(i) we note that

$$\begin{aligned}
\sup_{\boldsymbol{a},\boldsymbol{\theta}} |\nabla_{a_j}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})| &\leq \frac{K}{T_m}\Big\{\sup_{t,j}|x_{t,j}|\cdot m\cdot S_T + \sum_{j_2=1}^{n}\sup_{\boldsymbol{\theta}}|H_T^{(j,j_2)}(\mathcal{F}_{\boldsymbol{\theta}}^{-1})|\Big\} &(59)\\
\sup_{\boldsymbol{a},\boldsymbol{\theta}} |\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})| &\leq \frac{K}{T_m}\Bigg\{\sup_{\boldsymbol{\theta}} V_T(\nabla_{\boldsymbol{\theta}}\mathcal{F}_{\boldsymbol{\theta}}^{-1}) + \sup_{t,j}|x_{t,j}|\cdot m\cdot S_T \\
&\quad + \sum_{j_1,j_2}\sup_{\boldsymbol{\theta},\boldsymbol{a}} H_T^{(j_1,j_2)}(\nabla_{\boldsymbol{\theta}}\mathcal{F}_{\boldsymbol{\theta}}^{-1}) + \sum_{t=m/2}^{T-m/2}\sup_{\boldsymbol{\theta}}\int\frac{|\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)|}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}d\omega\Bigg\}, &(60)
\end{aligned}$$

noting that to obtain the above we bound the $D_T^{(j)}(\mathcal{F}_{\boldsymbol{\theta}}^{-1})$ in $\nabla_{a_j}\mathcal{L}^{(m)}$ (see (48)) with $|D_T^{(j)}(\mathcal{F}_{\boldsymbol{\theta}}^{-1})| \leq \sup_{t,\omega}|\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1}|\sup_{t,j}|x_{t,j}|\cdot S_T$, where $S_T = \sum_{s=1}^{T}|X_s|$. With the exception of $T_m^{-1}\sup_{\boldsymbol{\theta}} V_T(\nabla_{\boldsymbol{\theta}}\mathcal{F}_{\boldsymbol{\theta}}^{-1})$ and $T_m^{-1}S_T$, all the terms on the right hand side of $\sup_{\beta,\boldsymbol{\theta}}|\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})|$ and $\sup_{\boldsymbol{a},\boldsymbol{\theta}}|\nabla_{a_j}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})|$, above, are deterministic and, under Assumptions 5.1(i,ii,iii) (with $k=1$) and 5.2(i), are bounded. It is straightforward to show that $\sup_T\mathbb{E}(T_m^{-1}S_T) < \infty$ and $\mathrm{var}(T_m^{-1}S_T) = O(T^{-1})$, therefore $T_m^{-1}S_T$ is bounded in probability. By using Lemma A.1 we have that $T_m^{-1}\mathbb{E}(\sup_{\boldsymbol{\theta}} V_T(\nabla_{\boldsymbol{\theta}}\mathcal{F}_{\boldsymbol{\theta}}^{-1})) < \infty$ and $\mathrm{var}\big[\frac{1}{T_m}V_T(\nabla_{\boldsymbol{\theta}}\mathcal{F}_{\boldsymbol{\theta}}^{-1})\big] = O(\frac{1}{T_m})$, hence $\frac{1}{T_m}V_T(\nabla_{\boldsymbol{\theta}}\mathcal{F}_{\boldsymbol{\theta}}^{-1})$ is bounded in probability. Altogether this gives that $\sup_{\boldsymbol{a},\boldsymbol{\theta}}|\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})|$ and $\sup_{\boldsymbol{a},\boldsymbol{\theta}}|\nabla_{a_j}\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})|$ are both bounded in probability. Finally, since $\nabla\mathcal{L}_T^{(m)}$ is bounded in probability, equicontinuity of $\mathcal{L}_T^{(m)}(\cdot)$ immediately follows from (58).

The proof of (ii) is identical to the proof of (i), the main difference is that it is based on the boundedness of the third derivative of f_{j_1,j_2}, rather than the first derivative of f_{j_1,j_2}. We omit the details. □

We use equicontinuity now to prove consistency of $(\hat{\boldsymbol{a}}_T, \hat{\boldsymbol{\theta}}_T)$.

PROOF of Proposition 5.1. The proof is split into two stages, we first show that for a large enough T, $(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$ is the unique minimum of the limiting deterministic function function $\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))$ (defined in (57)) and second we show that $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ converges uniformly in probability, together they prove the result. Let us consider the difference

$$\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})) - \mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) \quad = \quad I + II$$

where

$$
\begin{aligned}
I &= \frac{1}{T_m}\sum_{t=m/2}^{T-m/2}\int\left\{\frac{\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}-\log\frac{\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}\right\}d\omega-1\\
II &= \frac{1}{T_m}\sum_{t=m/2}^{T-m/2}\sum_{j_1,j_2=1}^{n}(a_{j_1,0}-a_{j_1})(a_{j_2,0}-a_{j_2})\int J_{t,m}^{(j_1)}(\omega)J_{t,m}^{(j_2)}(-\omega)d\omega.
\end{aligned}
$$

We now show that I and II are positive. We first show that I is positive. Let $x_{t,\omega}=\frac{\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}$, since $x_{t,\omega}$ is positive, it is obvious that $x_{t,\omega}-\log x_{t,\omega}-1$ is a positive function. Moreover $x_{t,\omega}-\log x_{t,\omega}-1$ has a unique minimum when $x_{t,\omega}=1$. Since $I=\frac{1}{T_m}\sum_{t=m/2}^{T-m/2}\int(x_{t,\omega}-\log x_{t,\omega}-1)d\omega$, I is positive and has minimum when $\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)=\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)$. To show that II is positive, we recall from the definition of $\mathcal{J}_{T,m}(\omega)'$ given in Assumption 5.2(ii), that $II=(\boldsymbol{a}-\boldsymbol{a}_0)'\int\mathcal{J}_{T,m}(\boldsymbol{\theta},\omega)\mathcal{J}_{T,m}(\boldsymbol{\theta},\omega)'d\omega(\boldsymbol{a}-\boldsymbol{a}_0)$, where $(\boldsymbol{a}-\boldsymbol{a}_0)=(a_1-a_{1,0},\ldots,a_n-a_{n,0})'$, which is clearly positive. Altogether this gives $\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))-\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))=I+II\geq 0$, and for all values of T, $\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))$ has a minimum at $\boldsymbol{a}=\boldsymbol{a}_0$ and $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. To show that $(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$ is the unique minimum of $\mathbb{E}(\mathcal{L}_T^{(m)}(\cdot))$, suppose by contradiction, there exists another $(\boldsymbol{a}^*,\boldsymbol{\theta}^*)$ such that $\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a}^*,\boldsymbol{\theta}^*))=\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$. Then by using the arguments above, for all values of t and ω we must have $\mathcal{F}_{t,m}(\boldsymbol{\theta}^*,\omega)=\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)$ and $(\boldsymbol{a}^*-\boldsymbol{a}_0)'\int\mathcal{J}_{T,m}(\boldsymbol{\theta},\omega)\mathcal{J}_{T,m}(\boldsymbol{\theta},\omega)'d\omega(\boldsymbol{a}^*-\boldsymbol{a}_0)=0$. However this contradicts Assumption 5.2(ii), therefore we can conclude that for a large T, $(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$ is the unique minimum of $\mathbb{E}(\mathcal{L}_T^{(m)}(\cdot))$.

We now show that $\hat{\boldsymbol{\theta}}_T\stackrel{\mathcal{P}}{\to}\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{a}}_T\stackrel{\mathcal{P}}{\to}\boldsymbol{a}_0$. To prove this we need to show uniform convergence of $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$. Uniform convergence follows from the probabilistic version of the Arzela-Ascoli lemma which means we have to verify (a) pointwise convergence of $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ (b) equicontinuity in probability of $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ and (c) compactness of the parameter space. It is clear from Lemma A.2 that for every $\boldsymbol{\theta}\in\Theta_1\otimes\Theta_2$ and $\boldsymbol{a}\in\Omega$ we have $|\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})-\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))|\stackrel{\mathcal{P}}{\to}0$ as $T\to\infty$ ((a) holds), by Lemma A.3 we have that equicontinuity of $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})$ ((b) holds) and under Assumption 5.1(iii) Θ and Ω are compact sets ((c) holds). Therefore $\sup_{\boldsymbol{a},\boldsymbol{\theta}}|\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})-\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))|\stackrel{\mathcal{P}}{\to}0$. Finally, since $\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)\geq\mathcal{L}_T^{(m)}(\hat{\boldsymbol{a}}_T,\hat{\boldsymbol{\theta}}_T)\stackrel{\mathcal{P}}{\to}\mathbb{E}(\mathcal{L}_T^{(m)}(\hat{\boldsymbol{a}}_T,\hat{\boldsymbol{\theta}}_T))\geq\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$, we have that $|\mathcal{L}_T^{(m)}(\hat{\boldsymbol{a}}_T,\hat{\boldsymbol{\theta}}_T)-\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))|\leq\max\{|\mathcal{L}_T^{(m)}(\hat{\boldsymbol{a}}_T,\hat{\boldsymbol{\theta}}_T)-\mathbb{E}(\mathcal{L}_T^{(m)}(\hat{\boldsymbol{a}}_T,\hat{\boldsymbol{\theta}}_T))|,|\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)-\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))|\}\leq\sup_{\boldsymbol{a},\boldsymbol{\theta}}|\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta})-\mathbb{E}(\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))|\stackrel{\mathcal{P}}{\to}0$. Because $(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$ is the unique minimum of $\mathbb{E}(\mathcal{L}_T^{(m)}(\cdot))$, it follows that $\hat{\boldsymbol{a}}_T\stackrel{\mathcal{P}}{\to}\boldsymbol{a}_0$ and $\hat{\boldsymbol{\theta}}_T\stackrel{\mathcal{P}}{\to}\boldsymbol{a}_0$. □

By approximating $\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$ with the sum of martingale differences, we will use the martingale central limit theorem to prove Theorem 5.1. Using (48) we have

$$
\begin{aligned}
\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0) &= -\frac{1}{T_m}\left\{V_T(\nabla\mathcal{F}_{\boldsymbol{\theta}_0}^{-1})-\mathbb{E}[V_T(\nabla\mathcal{F}_{\boldsymbol{\theta}_0}^{-1})]\right\}=-\frac{1}{T_m}\left[\mathcal{Z}_T^{(m)}+U_T^{(m)}\right],\\
\text{and }\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)\rfloor_j &= \frac{1}{T_m}D_T^{(j)}(\mathcal{F}_{\boldsymbol{\theta}_0}^{-1})=\frac{1}{T_m}\{\mathcal{B}_T^{(m,j)}+C_T^{(m,j)}\}
\end{aligned}
$$

where

$$\mathcal{Z}_T^{(m)} = \frac{1}{2\pi m}\sum_{k=1}^{m}\sum_{j_1,j_2=1}^{n+1}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} x_{s,j_1}x_{s+r,j_2}\underline{b}_{s+m/2-k,m}(r) \times \sum_{i_1,i_2=0}^{s-1,s+r-1}\psi_{i_1,j_1}\psi_{i_2,j_2}\big(\eta_{s-i_1,j_1}\eta_{s+r-i_2,j_2}-\mathbb{E}(\eta_{s-i_1,j_1}\eta_{s+r-i_2,j_2})\big) \tag{61}$$

$$U_T^{(m)} = \frac{1}{2\pi m}\sum_{k=1}^{m}\sum_{j_1,j_2=1}^{n+1}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} x_{s,j_1}x_{s+r,j_2}\underline{b}_{s+m/2-k,m}(r) \times \sum_{i_1\geq s \text{ or } i_2\geq s+r}\psi_{i_1,j_1}\psi_{i_2,j_2}\big(\eta_{s-i_1,j_1}\eta_{s+r-i_2,j_2}-\mathbb{E}(\eta_{s-i_1,j_1}\eta_{s+r-i_2,j_2})\big) \tag{62}$$

(63)

$$\mathcal{B}_T^{(m,j)} = \frac{1}{2\pi m}\sum_{j_2=1}^{n}\sum_{k=1}^{m}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} a_{s+m/2-k,m}(r)x_{s,j_2}x_{s+r,j}\sum_{i=1}^{s_1}\psi_{i,j_2}\eta_{s_1-i,j_2}$$

$$C_T^{(m,j)} = \frac{1}{2\pi m}\sum_{j_2=1}^{n}\sum_{k=1}^{m}\int\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} a_{s+m/2-k,m}(r)x_{s_1,j_2}x_{s_2,j}\sum_{i=s+1}^{\infty}\psi_{i,j_2}\eta_{s-i,j_2}, \tag{64}$$

with $\{\underline{b}_s(k)\}$, $\{a_s(k)\}$ defined in (29) and setting $x_{t,n+1}=1$.

In the following results we show that $\sqrt{T}\nabla_{\boldsymbol{\theta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0) = \frac{-1}{\sqrt{T}}Z_T^{(m)}+o_p(1)$ and $\sqrt{T}\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)\rfloor_j = \frac{1}{\sqrt{T}}B_T^{(m,j)}+o_p(1)$.

Proposition A.1 *Suppose Assumptions 2.1, 5.1 and 5.2(i) are satisfied and $U_T^{(m)}$ and $W_T^{(m)}$ satisfy (61). Then we have*

$$[\mathbb{E}(|U_T^{(m)}|^2)]^{1/2} \leq K \tag{65}$$

where K is independent of T and $\mathcal{Z}_T^{(m)} = \sum_{d=1}^{T} Z_d^{(m)}$, $\{Z_d^{(m)}\}_d$ are martingale difference vectors with

$$Z_d^{(m)} = \sum_{j_1,j_2=1}^{n+1}\left(\{\eta_{d,j_1}\eta_{d,j_2}-\mathbb{E}[\eta_{d,j_1}\eta_{d,j_2}]\}A_{d,d}^{(j_1,j_2)} + 2\eta_{d,j_1}\sum_{d_1<d}\eta_{d_1,j_2}A_{d,d_1}^{(j_1,j_2)}\right) \tag{66}$$

$\psi_{-i,j_i}=0$ for $i\geq 1$, and

$$A_{d,d_1}^{(j_1,j_2)} = \frac{1}{2\pi m}\sum_{k=1}^{m}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} x_{s,j_1}x_{s+r,j_2}\psi_{s+d,j_1}\psi_{s+r+d_1,j_2}\underline{b}_{s+m/2-k,m}(r). \tag{67}$$

For all d, d_2, j_1 and j_2 we have

$$\sup_d\sum_{d_2=1}^{d}|A_{d,d_2}^{(j_1,j_2)}|^2<\infty \quad \textit{and} \quad \sup_d|A_{d,d}^{(j_1,j_2)}|<\infty. \tag{68}$$

PROOF. We first prove (65). To make the notation in the proof less cumbersome, we prove the result for $X_t = x_{t,1}\alpha_{t,1}$, where $\alpha_{t,1}$ satisfies Assumption 2.1. The proof for the general case is the same. We observe $U_T^{(m)}$ contains all the terms in $\nabla_{\boldsymbol{\theta}}\mathcal{L}^{(m)}$ which have $\{\eta_t\}_{t\leq 0}$. We partition $U_T^{(m)} = R_{T,1} + R_{T,2}$, where

$$R_{T,1} = \frac{2}{2\pi m}\sum_{k=1}^{m}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} x_{s,1}x_{s+r,1}\underline{b}_{s+m/2-k,m}(r)\sum_{i_1\geq s}\sum_{i_2=0}^{s+r-1}\psi_{i_1,1}\psi_{i_2,1}\big(\eta_{s-i_1,1}\eta_{s+r-i_2,1}-\mathbb{E}(\eta_{s-i_1,1}\eta_{s+r-i_2,1})\big),$$

$$R_{T,2} = \frac{1}{2\pi m}\sum_{k=1}^{m}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k} x_{s,1}x_{s+r,1}\underline{b}_{s+m/2-k,m}(r)\sum_{i_1\geq s}\sum_{i_2\geq s+r}\psi_{i_1,1}\psi_{i_2,1}\big(\eta_{s-i_1,1}\eta_{s+r-i_2,1}-\mathbb{E}(\eta_{s-i_1,1}\eta_{s+r-i_2,1})\big)$$

We bound each term. Studying $R_{T,1}$ we see that due to the ranges in the summands, $\eta_{s-i_1,1}$ and $\eta_{s+r-i_2,1}$ are independent, therefore the second moment of $(\mathbb{E}|R_{T,1}|^2)^{1/2}$ is bounded by

$$\begin{aligned}
&(\mathbb{E}|R_{T,1}|^2)^{1/2}\\
\leq\ & \frac{2}{m}\mathbb{E}(\eta_{0,1}^2)^{1/2}\sup_t|x_{t,1}|^2\sum_{k=1}^{m}\sup_s\sum_{r=-k}^{m-k}|\underline{b}_{s,m}(r)|\sum_{s=k}^{T-m+k}\sum_{i_1\geq s}|\psi_{i_1,1}|\left\{\mathbb{E}\left(\sum_{i_2=1}^{s+r}\psi_{i_2,1}\eta_{s_2-i_2,1}\right)^2\right\}^{1/2}\\
\leq\ & 2\mathbb{E}(\eta_{0,1}^2)\sup_t|x_{t,1}|^2\sup_s\sum_r|\underline{b}_{s,m}(r)|\Big\{\sum_{s=1}^{\infty}|s\psi_{s,1}|\Big\}\Big\{\sum_{i=0}^{\infty}|\psi_{i,1}|^2\Big\}^{1/2}<\infty,
\end{aligned}$$

where to obtain the last inequality in the above we use that $\sum_{s=k}^{T-m+k}\sum_{i_1\geq s}|\psi_{i_1,1}|\leq\sum_{i=1}^{\infty}|i\cdot\psi_{i,1}|$. To bound $(\mathbb{E}|R_{T,2}|^2)^{1/2}$, it is straightforward to show

$$(\mathbb{E}|R_{T,2}|^2)^{1/2} \leq \sup_t|x_{t,1}|^2\max(\mathbb{E}|\eta_0^4|^{1/2},\mathbb{E}|\eta_0^2|)\Big\{\sup_s\sum_r|\underline{b}_{s,m}(r)|\sum_{s=0}^{\infty}|s\psi_{s,1}|\Big\}^2<\infty$$

Using the above bounds we have $(\mathbb{E}|U_T^{(m)}|^2)^{1/2}<\infty$ thus giving (65). To prove (**??**), we change indices and let $d_1 = s-i_1$ and $d_2 = s+r-i_2$, substituting this into $\mathcal{Z}_T^{(m)}$ gives

$$\mathcal{Z}_T^{(m)} = \frac{1}{2\pi m}\sum_{k=1}^{m}\sum_{j_1,j_2=1}^{n}\sum_{s=k}^{T-m+k}\sum_{r=-k}^{m-k}\underline{b}_{s+m/2-k,m}(r)\sum_{d=0}^{s}\sum_{d_1=0}^{s+r}\psi_{s+d,j_1}\psi_{s+r+d_1,j_2}\big(\eta_{d,j_1}\eta_{d_1,j_2}-\mathbb{E}(\eta_{d,j_1}\eta_{d_1,j_2})\big).$$

Using the above and partitioning into the cases $d_1<d$ and $d_1=d$ we obtain the martingale difference decomposition in (66).

Finally from the definition of $A_{d,d_2}^{(j_1,j_2)}$ we have

$$\begin{aligned}
\sum_{d_2=1}^{d}|A_{d,d_2}^{(j_1,j_2)}|^2 \leq\ & \frac{1}{m^2}\sum_{k_1,k_2=1}^{m}\sum_{d_2=1}^{d}\sum_{s=k_1}^{T-m+k_1}\sum_{u=k_2}^{T-m+k_2}\sum_{r_1=-k_1}^{m-k_1}\sum_{r_2=-k_2}^{m-k_2}|x_{s,j_1}x_{s+r_1,j_2}x_{u,j_1}x_{u+r_2,j_2}|\\
& |\underline{b}_{s+m/2-k_1,m}(r_1)|\cdot|\underline{b}_{s+m/2-k_2,m}(r_2)|\psi_{s+d,j_1}\psi_{s+r+d,j_2}\psi_{u+d,j_1}\psi_{u+r_2+d,j_2}\\
\leq\ & \sup_{t,j}|x_{t,j}|^4(\sup_s\sum_r|\underline{b}_{s,m}(r)|)^2\sup_j(\sum_i|\psi_{i,j}|^2)\sup_j(\sum_i|\psi_{i,j}|)^2,
\end{aligned}$$

thus giving $\sup_d\sum_{d_2=1}^{d}|A_{d,d_2}^{(j_1,j_2)}|^2<\infty$ (the first part of (68)) we use a similar method to show $\sup_d|A_{d,d}^{(j_1,j_2)}|<\infty$. □

We state an analogous result for $\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)\rfloor_j = \frac{1}{T_m}\{\mathcal{B}_T^{(m,j)}+C_T^{(m,j)}\}$.

Proposition A.2 *Suppose Assumptions 2.1, 5.1 and 5.2 are satisfied and* $\mathcal{B}_T^{(m,j)}$ *and* $C_T^{(m,j)}$ *satisfy (64). Then we have*

$$(\mathbb{E}|C_T^{(m,j)}|^2)^{1/2} \leq K \tag{69}$$

and $\mathcal{B}_T^{(m,j)} = \sum_{d=1}^{T} B_d^{(m,j)}$, *where* $\{B_d^{(m,j)}\}$ *are martingale difference with* $B_d^{(m,j)} = \sum_{j_1=1}^{n+1} \eta_{d,j_1} D_d^{(j,j_1)}$ *and*

$$D_d^{(j,j_1)} = \frac{1}{2\pi m} \sum_{k=1}^{m} \sum_{s=k}^{T-m+k} \sum_{r=-k}^{m-k} x_{s,j} x_{s+r,j_1} a_{s+m/2-k,m}(r) \psi_{s+d,j_1}. \tag{70}$$

Furthermore $D_d^{(j,j_1)}$ *is bounded by*

$$\sup_{j,j_2} |D_d^{(j,j_1)}| \leq \frac{1}{2\pi} \sup_{t,j} |x_{t,j}|^2 \sup_{s} \sum_{r} |a_{s,m}(r)| \sup_{j} \sum_{s=1}^{\infty} |\psi_{s,j}|. \tag{71}$$

PROOF. It is straightforward, hence we omit the details. □

In order to use the martingale central limit theorem we show that the martingales differences $\{W_d^{(m)}\}$ and $\{B_d^{(m,j)}\}$ have a finite fourth moment.

Lemma A.4 *Suppose Assumptions 2.1, 5.1 and 5.2(i) are satisfied. Let* $Z_d^{(m)}$ *be defined as in (66) and* $B_d^{(m,j)}$ *be defined as in (70). Then, for some finite* K, *we have* $\sup_d \mathbb{E}(|Z_d^{(m)}|^4) \leq K$ *and* $\sup_d \mathbb{E}(|B_d^{(m,j)}|^4) \leq K$.

PROOF. In order to bound $\mathbb{E}(|Z_d^{(m)}|^4)$ we note that $Z_d^{(m)} = X_d^{(m)} + Y_d^{(m)}$, where $X_d^{(m)} = 2\sum_{j_1,j_2=1}^{n+1} \eta_{d,j_1} \sum_{d_1<d} \eta_{d_1,j_2} A_{d,d_1}^{(j_1,j_2)}$ and $Y_d^{(m)} = \sum_{j_1,j_2=1}^{n+1} \{\eta_{d,j_1}\eta_{d,j_2} - \mathbb{E}[\eta_{d,j_1}\eta_{d,j_2}]\} A_{d,d}^{(j_1,j_2)}$. Now the fourth moment of the inner summand in $X_d^{(m)}$ is

$$\begin{aligned}
\mathbb{E}\big\{\eta_{d,j_1} \sum_{d_2<d} \eta_{d_2,j_2} A_{d,d_2}^{(j_1,j_2)}\big\}^4 &\leq \mathbb{E}(\eta_{d,j_1}^4) \sum_{d_2<d,d_3<d} \mathbb{E}[\eta_{d_2,j_2}^2 \eta_{d_3,j_2}^2](A_{d,d_2}^{(j_1,j_2)})^2 (A_{d,d_3}^{(j_1,j_2)})^2 \\
&\leq \max_{j}(\mathbb{E}[\eta_{d,j}^4]^2, \mathbb{E}[\eta_{d,j}^2]^4)\big(\sum_{d_1<d} (A_{d,d_1}^{(j_1,j_2)})^2\big)^2.
\end{aligned}$$

Hence $\mathbb{E}(|X_d^{(m)}|^4) \leq 2(n+1)\max_j[(\mathbb{E}(\eta_{d,j}^4))^2, (\mathbb{E}(\eta_{d,j}^2))^4]\big(\sum_{d_1<d}(A_{d,d_1}^{(j_1,j_2)})^2\big)^2$. Since $n < \infty$ by using (68), we have $\mathbb{E}(|X_d^{(m)}|^4) < \infty$. Using the same arguments we can show that $\mathbb{E}(|Y_d^{(m)}|^4) < \infty$, this gives $\mathbb{E}(|Z_d^{(m)}|^4) \leq K$. Finally we note that since $B_d^{(m,j)} = \sum_{j_2=0}^{n} \eta_{d,j_2} D_d^{(j,j_2)}$, by using (71) we have $\mathbb{E}(|B_d^{(m,j)}|^4) < \infty$, as desired. □

We use the following lemma to prove asymptotic normality of the estimator. In the lemma below we show that the likelihood of the derivative can be written in terms of a martingale term and a term which is of lower order.

From the above we have $\mathcal{Z}_T^{(m)} = O_p(T^{-1/2})$, $\mathcal{B}_T^{(m,j)} = O_p(T^{-1/2})$, and by using (61),

$$\begin{aligned}
\sqrt{T}\nabla_{\boldsymbol{\theta}} \mathcal{L}_T^{(m)}(\boldsymbol{a}_0, \boldsymbol{\theta}_0) &= T^{-1/2}\mathcal{Z}_T^{(m)} + O_p(T^{-1/2}) \\
\sqrt{T}\nabla_{a_j} \mathcal{L}_T^{(m)}(\boldsymbol{a}_0, \boldsymbol{\theta}_0) &= T^{-1/2}\mathcal{B}_T^{(m,j)} + O_p(T^{-1/2}),
\end{aligned} \tag{72}$$

as $T \to \infty$. We use this to prove Theorem 5.1.

Proposition A.3 *Suppose Assumptions 2.1, 5.1 and 5.2 hold, then we have*

$$\sqrt{T_m}\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,W_T^{(m)}), \text{ as } T\to\infty, \tag{73}$$

where $W_T^{(m)}$ is defined in (22).

PROOF. Using (72), Propositions A.1 and A.2 we have

$$\sqrt{T}\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0) = T^{-1/2}\sum_{d=1}^{T}\mathcal{S}_d + O_p(T^{-1/2}),$$

where $\mathcal{S}_d' = (B_d^{(m,1)},\ldots,B_d^{(m,n)},Z_d^{(m)})$ are martingale differences defined in Propositions A.1 and A.2. Clearly $T^{-1/2}\sum_{d=1}^{T}\mathcal{S}_d$ is the dominating term in $\sqrt{T}\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$, hence we prove the result by showing asymptotic normality of $T^{-1/2}\sum_{d=1}^{T}\mathcal{S}_d$ using the martingale central limit theorem and the Cramer-Wold device. Let $\boldsymbol{c}=(c_1,\ldots,c_{2n+q})$ be any non-zero vector and define the random variable

$$Q_d = \sum_{j=1}^{n}c_jB_d^{(m,j)} + \sum_{j=n+1}^{2n+q+1}c_j(Z_d^{(m)})_{j-n}.$$

To apply the martingale central limit theorem we need to verify that the variance of $T^{-1/2}\sum_{d=1}^{T}Q_d$ is finite and that Lindeberg's condition is satisfied (see (Hall & Heyde, 1980), Theorem 3.2). Using Lemma A.4 we have that $\frac{1}{T}\text{var}(\sum_{d=1}^{T}Q_d)=O(1)$. To verify Lindeberg's condition, we require that for all $\delta>0$, $L_T=\frac{1}{T}\sum_{d=1}^{T}\mathbb{E}(Q_d^2I(T^{-1/2}|Q_d|>\delta)|\mathcal{F}_{d-1}) \xrightarrow{\mathcal{P}} 0$ as $T\to\infty$, where $I(\cdot)$ is the identity function and $\mathcal{F}_{d-1}=\sigma(Q_{d-1},Q_{d-2},\ldots,Q_1)$. By using Hölder's and Markov's inequality and Lemma A.4 we obtain the following bound $L_T\le\frac{1}{T^2\delta}\sum_{d=1}^{T}\mathbb{E}(Q_d^4|\mathcal{F}_{d-1})=O(\frac{1}{T})\to 0$, as $T\to\infty$. Finally we need to show that

$$\frac{1}{T}\sum_{d=1}^{T}\mathbb{E}(Q_d^2|\mathcal{F}_{d-1}) = \frac{1}{T}\sum_{d=1}^{T}\left[\mathbb{E}(Q_d^2|\mathcal{F}_{d-1})-\mathbb{E}(Q_d^2)\right] + \frac{1}{T}\sum_{d=1}^{T}\mathbb{E}(Q_d^2) \xrightarrow{\mathcal{P}} \underline{c}'W_T^{(m)}\underline{c} \tag{74}$$

where $\underline{c}'=(c_1,\ldots,c_{n+q})$. From the definition of Q_d^2 in terms of $Z_d^{(m)}$ and $\{B_d^{(m,j)}\}$ which are truncated versions of $\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)$, it is straightforward to show that $|\frac{1}{T_m}\sum_{d=1}^{T}\mathbb{E}(Q_d^2)-\underline{c}'W_T^{(m)}\underline{c}|\to 0$ as $T\to\infty$. Therefore it remains to show

$$P_T = \frac{1}{T_m}\sum_{d=1}^{T}\left(\mathbb{E}(Q_d^2|\mathcal{F}_{d-1})-\mathbb{E}(Q_d^2)\right) \xrightarrow{\mathcal{P}} 0,$$

which will give us (74). We will show that $\mathbb{E}(P_T^2)\to 0$, which gives the above. To do this we note that $\mathbb{E}(P_T)=0$ and

$$\text{var}(P_T) = \frac{1}{T_m^2}\sum_{d=1}^{T}\text{var}(\mathbb{E}(Q_d^2|\mathcal{F}_{d-1})) + \frac{2}{T_m^2}\sum_{d_1>d_2}^{T}\text{cov}(\mathbb{E}(Q_{d_1}^2|\mathcal{F}_{d_1-1}),\mathbb{E}(Q_{d_2}^2|\mathcal{F}_{d_2-1})). \tag{75}$$

Now by using the Cauchy Schwartz inequality and conditional expectation arguments for $\mathcal{F}_{d_2}\subset\mathcal{F}_{d_1}$ we have

$$\text{cov}(\mathbb{E}(Q_{d_1}^2|\mathcal{F}_{d_1-1}),\mathbb{E}(Q_{d_2}^2|\mathcal{F}_{d_2-1})) \le \left[\mathbb{E}\big(\mathbb{E}(Q_{d_2}^2|\mathcal{F}_{d_2-1})-\mathbb{E}(Q_{d_2}^2)\big)^2\right]^{1/2}\left[\mathbb{E}\big(\mathbb{E}(Q_{d_1}^2|\mathcal{F}_{d_2-1})-\mathbb{E}(Q_{d_1}^2)\big)^2\right]^{1/2}. \tag{76}$$

We now show that $\left[\mathbb{E}\big(\mathbb{E}(Q_{d_1}^2|\mathcal{F}_{d_2-1}) - \mathbb{E}(Q_{d_1}^2)\big)^2\right] \to 0$ as $d_1 - d_2 \to \infty$. Define the sigma-algebra $\mathcal{G}_d = \sigma(\eta_{d,1}, \eta_{d,n+1}, \eta_{d-1,1}, \ldots)$. it is straightaforward to show that $\mathcal{F}_d \subset \mathcal{G}_d$. Therefore $\mathbb{E}\left[\mathbb{E}(Q_{d_1}^2|\mathcal{F}_{d_2-1})^2\right] \leq \mathbb{E}\left[\mathbb{E}(Q_{d_1}^2|\mathcal{G}_{d_2-1})^2\right]$ which gives

$$\mathbb{E}\left[\mathbb{E}(Q_{d_1}^2|\mathcal{F}_{d_2-1}) - \mathbb{E}(Q_{d_1}^2)\right]^2 \leq \mathbb{E}\left[\mathbb{E}(Q_{d_1}^2|\mathcal{G}_{d_2-1})^2\right] - \left[\mathbb{E}(Q_{d_1}^2)\right]^2.$$

From the definition of Q_{d_1} (in terms of $\{\eta_t\}_{t\geq 1}$) and using Propositions A.1 and A.2, we have that $\mathbb{E}\left[\mathbb{E}(Q_{d_1}^2|\mathcal{G}_{d_2-1})^2\right] - \left[\mathbb{E}(Q_{d_1}^2)\right]^2 \to 0$ as $d_2 \to \infty$, and

$$\sup_{d_1} \sum_{d_2=1}^{d_1} \left(\mathbb{E}\left[\mathbb{E}(Q_{d_1}^2|\mathcal{G}_{d_2-1})^2\right] - \left[\mathbb{E}(Q_{d_1}^2)\right]^2\right)^{1/2} < \infty.$$

Substituting the above into (76) and (75) we have $\mathrm{var}(P_T) = O(T^{-1})$, hence we have shown (74), and the conditions of the martingale central limit theorem are satisfied. Thus we obtain (73). □

PROOF of Theorem 5.1. To prove (24) we use (19). We note that by Lemma A.3(ii), $\nabla^2 \mathcal{L}_T^{(m)}$ is pointwise convergent and equicontinuous on probability. Together with the compactness of the parameter spaces Ω and $\Theta_1 \otimes \Theta_2$ this implies that $\sup_{\boldsymbol{a},\boldsymbol{\theta}} |\nabla^2 \mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}) - \mathbb{E}(\nabla^2 \mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\theta}))| \overset{\mathcal{P}}{\to} 0$, therefore $|\nabla^2 \mathcal{L}_T^{(m)}(\bar{\boldsymbol{a}}_T, \bar{\boldsymbol{\theta}}_T) - \mathbb{E}(\nabla^2 \mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))| \overset{\mathcal{P}}{\to} 0$. Given that $\mathbb{E}(\nabla_{\boldsymbol{a}} \nabla_{\boldsymbol{\theta}} \mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0)) = 0$, and $\mathbb{E}(\nabla^2_{\boldsymbol{\theta}} \mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$ and $\mathbb{E}(\nabla^2_{\boldsymbol{a}} \mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$ are non-singular, then $|\nabla^2 \mathcal{L}_T^{(m)}(\bar{\boldsymbol{a}}_T, \bar{\boldsymbol{\theta}}_T)^{-1} - \mathbb{E}(\nabla^2 \mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))^{-1}| \overset{\mathcal{P}}{\to} 0$. This together with (73) and (19) proves (24). □

A.3 Proofs in Section 5.3

We now consider the sampling properties of the two-stage scheme. We observe that the derivatives of $\mathcal{L}_T^{(1)}(\boldsymbol{a},\boldsymbol{\Sigma})$ evaluated at the true parameter are

$$\nabla_{\boldsymbol{\Sigma}} \mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0) = \frac{1}{T}\sum_{t=1}^{T} \frac{(\sigma_t(\boldsymbol{\Sigma}_0) - X_t^2)\mathbf{X}_t}{\sigma_t(\boldsymbol{\Sigma}_0)^2} \quad \text{and} \quad \mathbb{E}\big(\nabla^2_{\boldsymbol{\Sigma}} \mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0)\big) = \frac{1}{T}\sum_{t=1}^{T} \frac{\mathbf{X}_t \mathbf{X}_t'}{\sigma_t(\boldsymbol{\Sigma}_0)^3}$$

where $\mathbf{X}_t = (x_{t,1}^2, \ldots, x_{t,n}^2)'$.

Lemma A.5 *Suppose Assumptions 2.1, 5.1 and 5.2 are satisfied then Let* $\mathcal{L}_T^{(1)}(\boldsymbol{a},\boldsymbol{\Sigma})$ *and* $\tilde{\boldsymbol{\Sigma}}_T$ *be defined as in (13). Then we have*

$$\big((\tilde{\boldsymbol{a}}_T - \boldsymbol{a}_0), (\tilde{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_0)\big) = O_p(T^{-1/2}), \tag{77}$$

$$\sqrt{T}\nabla \mathcal{L}_T^{(1)}(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0) = \frac{1}{\sqrt{T}} Z_T^{(1)} + \frac{1}{\sqrt{T}} U_T^{(1)}, \tag{78}$$

where $\sup_T \mathbb{E}|U_T^{(1)}| < \infty$, $\mathcal{Z}_T^{(1)} = \sum_{d=1}^{T} Z_d^{(1)}$ *and* $\{Z_d^{(1)}\}$ *are martingale differences with*

$$Z_d^{(1)} = \sum_{j_1,j_2=0}^{n} \big([\eta_{d,j_1}\eta_{d,j_2} - \mathbb{E}(\eta_{d,j_1}\eta_{d,j_2})] B_{d,d}^{(j_1,j_2)} + 2\eta_{d,j_1} \sum_{d_2<d} \eta_{d_2,j_2} B_{d,d_2}^{(j_1,j_2)}\big) \tag{79}$$

$$B_{d,d_2}^{(j_1,j_2)} = \sum_{s=1}^{T} \psi_{s-d,j_1}\psi_{s-d_2,j_2} \frac{x_{s,j_1} x_{s,j_2}}{\sigma_t(\boldsymbol{\Sigma}_0)} \boldsymbol{X}_s. \tag{80}$$

PROOF. By using analogous results to Lemmas A.2 and A.3 and the same technique of as the proof of Proposition 5.1 it can be shown that $\big((\tilde{\boldsymbol{a}}_T - \boldsymbol{a}_0), (\tilde{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_0)\big) \xrightarrow{\mathcal{P}} 0$. Further this gives (77). Using Propositions A.1 and A.2 we can show that $\sqrt{T}\nabla\mathcal{L}_T(\boldsymbol{a}_0, \boldsymbol{\Sigma}_0)$ and (78) is dominated by a term which is the sum of martingale differences, which is given in (78). □

PROOF of Proposition 5.2 By using the same arguments in Propositions A.3 and Theorem 5.1 we have (30).

We now consider the properties of $\mathcal{L}_T^{(m)}(\boldsymbol{a}_0, \boldsymbol{\vartheta}_0, \tilde{\boldsymbol{\Sigma}}_T)$. To do this we compare $\mathcal{L}_T^{(m)}(\boldsymbol{a}_0, \boldsymbol{\vartheta}_0, \tilde{\boldsymbol{\Sigma}}_T)$ with $\mathcal{L}_T^{(m)}(\boldsymbol{a}_0, \boldsymbol{\vartheta}_0, \boldsymbol{\Sigma}_0)$

Lemma A.6 *Suppose Assumptions 2.1, 5.1 and 5.2 are satisfied,* $\mathbb{E}(Z_t(\omega)) = 0$ *and for any deterministic sequence* $\{a_k\}$*, with* $\sup_k |a_k| < \infty$ *we have* $\sup_\omega \textit{var}\big(\frac{1}{n}\sum_{t=1}^n a_t Z_t(\omega)\big) = O(T^{-1})$*, then for any* $k \geq 1$ *we have*

$$\frac{1}{\sqrt{T}} \int \sum_{t=1}^{T} a_t Z_t(\omega)\big\{\nabla_{\boldsymbol{\vartheta}}\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0, \tilde{\boldsymbol{\Sigma}}_T, \omega) - \nabla_{\boldsymbol{\vartheta}}\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0, \boldsymbol{\Sigma}_0, \omega)\big\}d\omega \xrightarrow{\mathcal{P}} 0 \tag{81}$$

$$\frac{1}{\sqrt{n}} \int \sum_{t=1}^{n} a_t Z_t(\omega)\bigg(\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0, \tilde{\boldsymbol{\Sigma}}_T, \omega)^k} - \frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0, \boldsymbol{\Sigma}_0, \omega)^k}\bigg)d\omega \xrightarrow{\mathcal{P}} 0 \quad k > 0. \tag{82}$$

PROOF. To prove the result we first recall that from the definition of $\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}, \omega)$, that

$$\sum_{t=1}^{T} a_t Z_t(\omega)\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}, \omega) = \sum_{j=1}^{n+1} \sigma_j \sum_{t=1}^{T} a_t Z_t(\omega) \int_{-\pi}^{\pi} I_{t,m}^{(j)}(\lambda) f_j(\boldsymbol{\vartheta}, \omega - \lambda) d\lambda. \tag{83}$$

Now by using the above, (77) and differentiating $\mathcal{F}_{t,m}$ with respect to $\boldsymbol{\vartheta}$ we immediately obtain (81). We now prove (82). We observe that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} a_t Z_t(\omega)\bigg(\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0, \tilde{\boldsymbol{\Sigma}}_T, \omega)^k} - \frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0, \boldsymbol{\Sigma}_0, \omega)^k}\bigg) = A + B + C,$$

where

$$\begin{aligned}
A &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} a_t Z_t(\omega) \frac{\big[\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \tilde{\boldsymbol{\Sigma}}_T, \omega)^k - \mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega)^k\big]}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega)^{2k}} \\
B &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} a_t Z_t(\omega) \frac{\big[\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \tilde{\boldsymbol{\Sigma}}_T, \omega)^k - \mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega)^k\big]^2}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega)^{3k}} \\
C &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} a_t Z_t(\omega) \frac{\big[\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \tilde{\boldsymbol{\Sigma}}_T, \omega)^k - \mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega)^k\big]^3}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega)^{3k}\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \tilde{\boldsymbol{\Sigma}}_T, \omega)^{3k}}.
\end{aligned}$$

Now by using (83) and (77), and the stated assumptions $\mathbb{E}(\frac{1}{T}\sum_{t=1}^T a_t Z_t(\omega)) = 0$ and $\sup_\omega \text{var}\big(\frac{1}{T}\sum_{t=1}^T a_t Z_t(\omega)\big) = O(T^{-1})$ and that $\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega) \geq \delta$ (due to Assumption 5.1(i)), we have $A = O_p(T^{-1/2})$ and $B = O_p(T^{-1})$. Finally, to obtain a bound for C we note that

$$|C| \;\leq\; \frac{1}{\delta^{6k}\sqrt{n}} \sum_{t=1}^{n} |a_t| \cdot |Z_t(\omega)| \cdot \big|\big[\mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \tilde{\boldsymbol{\Sigma}}_T, \omega)^k - \mathcal{F}_{t,m}(\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_0, \omega)^k\big]^3\big|,$$

now by using the same methods as above it can be shown that $C = O_p(T^{-1})$. Thus we obtain (82). $\square$

We use the result above to prove the results below.

Lemma A.7 *Suppose the assumptions in Theorem 5.2 are satisfied. Then we have*

$$\sup_{\tilde{\boldsymbol{\theta}}}\Big|\frac{1}{T_m}\int\sum_{t=m/2}^{T-m/2}\Big(\log[\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T,\omega)]-\log[\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)]\Big)d\omega\Big|\xrightarrow{\mathcal{P}}0 \tag{84}$$

$$\sup_{\tilde{\boldsymbol{\theta}},\boldsymbol{a}}\Big|\frac{1}{T_m}\int\sum_{t=m/2}^{T-m/2}I_{t,m}(\boldsymbol{a},\omega)\Big(\frac{\nabla^{k_1}\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)^{k_2}}-\frac{\nabla^{k_1}\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T,\omega)^{k_2}}\Big)\Big|\xrightarrow{\mathcal{P}}0 \tag{85}$$

$$\frac{1}{\sqrt{T_m}}\int\sum_{t=m/2}^{T-m/2}\Re[J_{t,m}^{(X)}(\omega)J_{t,m}^{(j)}(-\omega)]\Big(\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T,\omega)}-\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)}\Big)d\omega\xrightarrow{\mathcal{P}}0 \tag{86}$$

$$\frac{1}{\sqrt{T_m}}\int\sum_{t=m/2}^{T-m/2}\big(I_{t,m}(\boldsymbol{a}_0,\omega)-\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T,\omega)\big)\Big(\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T,\omega)^2}-\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)^2}\Big)\nabla_{\tilde{\boldsymbol{\theta}}}\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T,\omega)d\omega\xrightarrow{\mathcal{P}}0 \tag{87}$$

and

$$\frac{1}{\sqrt{T_m}}\int\sum_{t=m/2}^{T-m/2}\frac{\big(I_{t,m}(\boldsymbol{a}_0,\omega)-\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T,\omega)\big)}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega))}\big\{\nabla_{\tilde{\boldsymbol{\theta}}}\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T,\omega)-\nabla_{\tilde{\boldsymbol{\theta}}}\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)\big\}d\omega\xrightarrow{\mathcal{P}}0 \tag{88}$$

PROOF. To prove (84) we use the mean value theorem and that $\mathcal{F}_{t,m}\geq\delta>0$ to obtain

$$\big|\log(\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T,\omega))-\log(\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega))\big|\leq\frac{1}{\delta}\big|\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega)-\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T,\omega)\big|.$$

Now by substituting the above into the sum in (84) and using (77) we obtain (84). Using a similar argument to the proof given in Lemma A.6 together with (77) we obtain (85). Now noting that $\mathbb{E}\Re[J_{t,m}^{(X)}(\omega)J_{t,m}^{(j)}(-\omega)]=0$, $\mathbb{E}\big(I_{t,m}(\boldsymbol{a}_0,\omega)-\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T,\omega)\big)=0$ and using Lemma A.1 we can apply (82) to obtain (86) and (87). Finally by using (81) we have (88). $\square$

We can now show obtain approximations of $\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T)$ and its derivatives. In the following lemma we show that $\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T)$ is uneffected by whether the estimated $\tilde{\boldsymbol{\Sigma}}_T$ or true $\boldsymbol{\Sigma}_0$ is used. On the other hand, estimating $\tilde{\boldsymbol{\Sigma}}_T$, changes of the asymptotic variance of $\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T)$.

Lemma A.8 *Suppose the assumptions Theorem 5.2 are satisfied and let $\mathcal{Q}_T(\boldsymbol{a}_0,\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0)$ be defined as in (32). Then we have*

$$\sup_{\boldsymbol{a},\boldsymbol{\vartheta}}\big|\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T)-\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0)\big|\xrightarrow{\mathcal{P}}0, \tag{89}$$

$$\sup_{\boldsymbol{a},\boldsymbol{\vartheta}}\big|\nabla^2\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T)-\nabla^2\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0)\big|\xrightarrow{\mathcal{P}}0, \tag{90}$$

$$\sqrt{T}\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T) = \sqrt{T}\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0) + o_p(1) \tag{91}$$

$$\sqrt{T}\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T) = \sqrt{T}\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0) + \sqrt{T}\mathcal{Q}_T\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0) + o_p(1) \tag{92}$$

where $\mathcal{Q}_T$ is defined in (32).

PROOF. By using (84) and (85) we immediately obtain (89) and (90). By using (86) we immediately obtain (91). To prove (92) we observe that by expanding $\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma})$ and using (87) and (86) we have

$$\sqrt{T}\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T) = \sqrt{T}\nabla_{\boldsymbol{\vartheta}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0) + \sqrt{T}\mathcal{K}_T(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0) + o_p(1)$$

where

$$\begin{aligned}
\sqrt{T}\mathcal{K}_T(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0) &= \frac{1}{\sqrt{T}}\int\sum_{t=m/2}^{T-m/2}\frac{\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\theta}_{2,T},\omega)-\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0,\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0,\omega)^2}\nabla_{\boldsymbol{\vartheta}}\mathcal{F}_{t,m}(\boldsymbol{\theta}_{1,0},\boldsymbol{\Sigma}_0,\omega) \\
&= \frac{1}{\sqrt{T}}\sum_{j=1}^{n+1}(\hat{\sigma}_j-\sigma_{j,0})\int\sum_{t=m/2}^{T-m/2}\frac{h_j^{(t,m)}(\omega)}{\mathcal{F}_{t,m}(\boldsymbol{\vartheta}_0,\boldsymbol{\Sigma}_0,\omega)^2}\nabla_{\boldsymbol{\vartheta}}\mathcal{F}_{t,m}(\boldsymbol{\vartheta},\boldsymbol{\Sigma}_0,\omega) \\
&= \sqrt{T}\mathcal{Q}_T\nabla_{\boldsymbol{\Sigma}}\mathcal{L}_T^{(1)}(\boldsymbol{a}_0,\boldsymbol{\Sigma}_0) + o_p(1),
\end{aligned}$$

where $\mathcal{Q}_T$ is defined in (32). Thus we obtain the desired result. □

Theorem A.1 *Suppose the assumptions Theorem 5.2 are satisfied, and let $\tilde{W}_T^{(m)}$ be defined as in Theorem 5.2. Then we have*

$$\sqrt{T}\nabla\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0,\tilde{W}_T^{(m)}\big)$$

PROOF. To prove the result we substitute (72) into Lemma A.8 to obtain

$$\begin{aligned}
\sqrt{T}\nabla_{\boldsymbol{a}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T) &= T^{-1/2}\sum_{d=1}^{T}B_d^{(m,j)} + o_p(1) \\
\sqrt{T}\nabla_{\tilde{\boldsymbol{\theta}}}\mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\vartheta}_0,\tilde{\boldsymbol{\Sigma}}_T) &= T^{-1/2}\sum_{d=1}^{T}Z_d^{(m)} + T^{-1/2}\sum_{d=1}^{T}\mathcal{Q}_T Z_d^{(1)} + o_p(1),
\end{aligned}$$

where $B_d^{(m,j)}$ and $Z_d^{(1)}$ are defined as in (70) and (80), and $Z_d^{(m)}$ is defined as in (66), except it does not include the derivatives with respect to $\boldsymbol{\Sigma}$. Now from the definitions we observe that $\{B_d^{(m,j)}\}$ and $\{Z_d^{(m)}+\mathcal{Q}_T Z_d^{(1)}\}$ are martingale differences, hence by using the same method of proof given in Proposition A.3, we have the result. □

PROOF of Theorem 5.2 The proof follows by making a Taylor series expansion of $\mathcal{L}_T^{(m)}(\boldsymbol{a},\boldsymbol{\vartheta},\tilde{\boldsymbol{\Sigma}}_T)$ with respect to $\boldsymbol{a}$ and $\boldsymbol{\vartheta}$ and using Theorem A.1. □

A.4 Approximations to the covariance operator

A.4.1 General approximations

In results in this section are used to prove Proposition 5.3. However, the results may also be of independent interest.

In this section we will consider functions which admit the representation $G(\omega) = \sum_{j=-\infty}^{\infty} g(k)\exp(ik\omega)$, where $\{g(k)\}$ is a square summable sequence, hence $G(\cdot) \in L_2[0,2\pi]$, where $L_2[0,2\pi]$ denotes all functions defined on $[0,2\pi]$. Many of the results below, require that the sequence $\{g(k)\}$ is absolutely summable. Sufficient conditions for this can be obtained in terms of $G(\cdot)$. More precisely, since $(ik)^r g(k) = \int_0^{2\pi} \frac{d^r G(\omega)}{d\omega^r}\exp(ik\omega)d\omega$, if $\frac{d^r G(\omega)}{d\omega^r} \in L_2[0,2\pi]$, then $\sum_k |k^{r-1}| \cdot |g(k)| < \infty$. Hence if the derivative of $G(\cdot)$ is square integrable, then this implies absolute summability of $\{g(k)\}$. Let $\|G\|_2 = \int |G(x)|^2 dx$, where $G : [0,2\pi] \to \infty$.

We begin by considering some of the properties of the operators $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$ and $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$, defined in (25) and (34) respectively. We recall that $\mathrm{var}(\nabla \mathcal{L}_T^{(m)}(\boldsymbol{a}_0,\boldsymbol{\theta}_0))$ can be written in terms of $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$ (see (28)). The operator $\Gamma_{\mathbf{x}_s,\mathbf{x}_u}$ has a similar interpretation. Define the (unobserved) time series $X_t(s) = \sum_{j=1}^n \alpha_{t,j}x_{s,j} + \varepsilon_t$ and $J_m^{(X)}(s,\omega) = \frac{1}{\sqrt{2\pi m}}\sum_{k=1}^m X_{s-m/2+k}(s)\exp(ik\omega)$, using this, it can be shown that

$$\begin{aligned}
&\mathrm{cov}\Big(\int_0^{2\pi}\mathcal{G}_s(\omega)\sum_{r_1=-\infty}^{\infty} X_s(s)X_{s+r_1}(s)\exp(ir_1\omega)d\omega, \int_0^{2\pi}\mathcal{H}_u(\omega)\sum_{r_2=-\infty}^{\infty} X_u(u)X_{u+r_2}(u)\exp(ir_2\omega)d\omega\Big) \\
&= \Gamma^{(0)}_{\mathbf{x}_s,\mathbf{x}_u}(g_s,\bar{h}_u)(s-u) \qquad (93)
\end{aligned}$$

where $g_s = \{g_s(r) = \int \mathcal{G}_s(\omega)\exp(ir\omega)d\omega\}$, $\bar{h}_s = \{h_s(r) = \int \mathcal{H}_s(\omega)\exp(r\omega)d\omega\}$. Similar interpretations are true for $\Gamma^{(j_1,j_2)}_{\mathbf{x}_s,\mathbf{x}_u}(g_s,h_u)(s-u)$ and $\Gamma^{(j)}_{\mathbf{x}_s,\mathbf{x}_u}(g_s,h_u)(s-u)$.

The proof of Proposition 5.3 requires the proposition below, where we show that if the kernels $\mathcal{K}^{(\mathbf{j})}$ satisfy a type of Lipschitz condition, then we can approximate $\Gamma^{(k_1,k_2)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$ with $\Gamma_{\mathbf{x}_s,\mathbf{x}_u}$. We start by obtain upper bounds for the kernels $\mathcal{K}^{(\mathbf{j})}$. Recalling the definition of $\rho_2(\cdot)$ and $\rho_4(\cdot)$ in (54), let $\rho_4(k_1,k_2,k_3) = \kappa_3 \sup_{j_1,j_2,j_3}\sum_i \cdot|a_{i+k_1,j_1}|\cdot|a_{i+k_2,j_2}|\cdot|a_{i+k_3,j_3}|$. Using these bounds it is clear that for $\mathbf{j} \in \{0, j, (j_1,j_2)\}$ we have $\sup|\mathcal{K}^{(\mathbf{j})}_{\mathbf{x}_{s_1},\mathbf{x}_{s_2},\mathbf{x}_{s_3},\mathbf{x}_{s_4}}(k,r_1,r_2)| \le \tilde{\mathcal{K}}^{(\mathbf{j})}(k,r_1,r_2)$, where

$$\begin{aligned}
\mathcal{K}^{(0)}(k,r_1,r_2) &= (n+1)\sup_{s,j}|x_{s,j}|^4\Big(\rho_2(k)\rho_2(k+r_1-r_2)+\rho_2(k-r_2)\rho_2(k+r_2)+\kappa_4(r_1,k,k-r_2)\Big) \\
\tilde{\mathcal{K}}^{(j)}(k,r_1,r_2) &= (n+1)\sup_{s,j}|x_{s,j}|^4\rho_3(k,r) \text{ and } \tilde{\mathcal{K}}^{(j_1,j_2)}(k,r_1,r_2) = (n+1)\sup_{s,j}|x_{s,j}|^4\rho_2(k).
\end{aligned}$$

For all $\mathbf{j} \in \{0, j, (j_1,j_2)\}$, these kernels can be used as the Lipschitz bounds

$$\big|\mathcal{K}^{(\mathbf{j})}_{\mathbf{x}_{s_1},\mathbf{x}_{s_2},\mathbf{x}_{s_3},\mathbf{x}_{s_4}}(k,r_1,r_2) - \mathcal{K}^{(\mathbf{j})}_{\mathbf{x}_{s_5},\mathbf{x}_{s_6},\mathbf{x}_{s_7},\mathbf{x}_{s_8}}(k,r_1,r_2)\big| \le \frac{1}{\hat{N}}\Big\{\sum_{i=1}^4 |s_i - s_{i+4}|\Big\}\tilde{\mathcal{K}}^{(\mathbf{j})}(k,r_1,r_2). \qquad (94)$$

In the proofs below we require the following bounds, which are finite under Assumption 5.1

$$\mathcal{K}_1^{(\mathbf{j})} = \begin{cases} \sum_{k,r_1}|\tilde{\mathcal{K}}^{(0)}(k,r_1,r_2)| & \mathbf{j}=0 \\ \sum_{k,r_1}|\tilde{\mathcal{K}}^{(j)}(k,r_1,r_2)| & \mathbf{j}=j \\ \sum_k |\tilde{\mathcal{K}}^{(j_1,j_2)}(k,r_1,r_2)| & \mathbf{j}=(j_1,j_2) \end{cases} \qquad \mathcal{K}_2^{(\mathbf{j})} = \begin{cases} \sum_{k,r_1}|k\tilde{\mathcal{K}}^{(\mathbf{0})}(k,r_1,r_2)| & \mathbf{j}=0 \\ \sum_{k,r_1}|k\tilde{\mathcal{K}}^{(j)}(k,r_1,r_2)| & \mathbf{j}=j \\ \sum_k |k\tilde{\mathcal{K}}^{(j_1,j_2)}(k,r_1,r_2)| & \mathbf{j}=(j_1,j_2) \end{cases} \qquad (95)$$

Proposition A.4 *Let* $\Gamma^{(k_1,k_2)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$ *and* $\Gamma_{\mathbf{x}_s,\mathbf{x}_u}$ *be defined as in (25) and (34). Let* $H_s(\omega)=\sum_{k=-\infty}^{\infty}h_s(k)\exp(ik\omega)$ *and* $G_s(\omega)=\sum_{k=-\infty}^{\infty}g_s(k)\exp(ik\omega)$*. Then we have*

$$\begin{aligned}&\Big|\frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=k_2}^{T-m+k_2}\Big(\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(g_{s+k_1},h_{u+k_2})(s,u)-\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}(g_{s+k_1},h_{u+k_2})(s-u)\Big)\Big|\\ \leq\ & C\Big\{\frac{1}{\hat{N}}+\frac{1}{m}\Big\}\Big\{\sup_s\|\frac{d^2H_s(\omega)}{d\omega^2}\|_2\sup_u\|\frac{d^2G_u(\omega)}{d\omega^2}\|_2\Big\}\mathcal{K}_1^{(\mathbf{j})},\end{aligned}$$

where $\mathbf{j}\in\{0,j,(j_1,j_2)\}$ *and* $C<\infty$.

PROOF. The main differences between $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$ and $\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}$ are the kernels used and the limits of the summands. Systematically replacing the kernels $\mathcal{K}^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_{s+r_1},\mathbf{x}_u,\mathbf{x}_{u+r_2}}(k,r_1,r_2)$ with $\mathcal{K}^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s,\mathbf{x}_u,\mathbf{x}_u}(k,r_1,r_2)$ and the summands $\sum_{r_1=-k_1}^{m-k_1}\sum_{r_2=-k_2}^{m-k_2}$ with $\sum_{r_1=-\infty}^{\infty}\sum_{r_2=-\infty}^{\infty}$, and using (94) gives

$$\Big|\frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=k_2}^{T-m+k_2}\big(\Gamma^{(k_1,k_2)}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(g_{s-k_1},h_{u-k_2})(s,u)-\Gamma_{\mathbf{x}_s,\mathbf{x}_u}(g_{s-k_1},h_{u-k_2})(s-u)\big)\Big|\leq I+II+III \tag{96}$$

where

$$\begin{aligned}I&=\frac{1}{m^2T_m\hat{N}}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=k_2}^{T-m+k_2}\sum_{r_1,r_2=-\infty}^{\infty}(|r_1|+|r_2|)|g_{s-k_1}(r_1)|\cdot|h_{u-k_2}(r_2)|\cdot\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2)\\ II&=\frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{d_1=0}^{T-m}\sum_{d_2=0}^{T-m}\sum_{r_1=-\infty}^{\infty}\sum_{r_2<-k_2,r_2>m-k_2}|g_{d_1}(r_1)|\cdot|h_{d_2}(r_2)|\cdot\tilde{\mathcal{K}}^{(\mathbf{j})}(d_1+k_1-d_2-k_2,r_1,r_2)\\ III&=\frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{d_1=0}^{T-m}\sum_{d_2=0}^{T-m}\sum_{r_1<-k_1,r_1>m-k_1}\sum_{r_2=-\infty}^{\infty}|g_{d_1}(r_1)|\cdot|h_{d_2}(r_2)|\cdot\tilde{\mathcal{K}}^{(\mathbf{j})}(d_1+k_1-d_2-k_2,r_1,r_2)\end{aligned}$$

We will derive bounds for I,II and III in the case $\mathbf{j}=0$, similar bounds can be obtained for the $\mathbf{j}=j$ and $\mathbf{j}=(j_1,j_2)$. We first bound I. By using the Cauchy-Schwartz inequality we have

$$\begin{aligned}&\sum_{r_1,r_2=-\infty}^{\infty}|r_1|\cdot|g_{s-k_1}(r_1)|\cdot|h_{u-k_2}(r_2)|\cdot\tilde{\mathcal{K}}^{(0)}(s-u,r_1,r_2)\\ \leq\ &\Big(\sum_{r_1=-\infty}^{\infty}|r_1\cdot g_{s-k_1}(r_1)|^2\Big)^{1/2}\Big(\sum_{r_1=-\infty}^{\infty}|\tilde{\mathcal{K}}^{(0)}(s-u,r_1,r_2)|^2\Big)^{1/2}\leq\sup_s\|\frac{dG_s(\omega)}{d\omega}\|_2\Big(\sum_{r_1=-\infty}^{\infty}|\tilde{\mathcal{K}}^{(0)}(s-u,r_1,r_2)|^2\Big)^{1/2}\end{aligned}$$

The same bound can be obtained for $\sum_{r_1,r_2=-\infty}^{\infty}|r_1|\cdot|g_{s-k_1}(r_2)|\cdot|h_{u-k_2}(r_2)|\cdot\tilde{\mathcal{K}}^{(0)}(s-u,r_1,r_2)$, substituting both these bounds into I and noting that $\sum_{u=-\infty}^{\infty}\big(\sum_{r_2=-\infty}^{\infty}|\tilde{\mathcal{K}}^{(0)}(s-u,r_1,r_2)|^2\big)^{1/2}\leq\sum_{u=-\infty}^{\infty}\sum_{r_2=-\infty}^{\infty}|\tilde{\mathcal{K}}^{(0)}(s-u,r_1,r_2)|=\mathcal{K}_1^{(\mathbf{j})}$, we have

$$\begin{aligned}I&\leq\frac{1}{m^2T_m\hat{N}}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sup_s\|\frac{dG_s(\omega)}{d\omega}\|_2\sup_u\|\frac{dH_s(\omega)}{d\omega}\|_2\sum_{u=k_2}^{T-m+k_2}\Big(\sum_{r_1=-\infty}^{\infty}+\sum_{r_1=-\infty}^{\infty}\Big)\tilde{\mathcal{K}}^{(0)}(s-u,r_1,r_2)\\ &\leq\frac{K}{\hat{N}}\sup_s\|\frac{dG_s(\omega)}{d\omega}\|_2\sup_u\|\frac{dH_s(\omega)}{d\omega}\|_2\mathcal{K}_1^{(0)},\end{aligned}$$

where K is a finite constant. We now bound II. By using (95) we have

$$II \leq \frac{1}{mT_m}\sup_{s,r}|g_s(r)|\mathcal{K}_1^{(0)}\sum_{d_2=0}^{T-m}\sum_{k_2=1}^{m}\sum_{r_2<-k_2,r_2>m-k_2}|h_{d_2}(r_2)|. \quad (97)$$

Studying the inner sum of the above we observe that

$$\sum_{k_2=1}^{m}\sum_{r_2<-k_2,r_2>m-k_2}|h_{d_2}(r_2)| \leq \sum_{r_2=-\infty}^{\infty}|r_2 h_{d_2}(r_2)| \leq \sup_s\|\frac{d^2H_s(\omega)}{d\omega^2}\|_2.$$

Substituting the above and $\sup_{s,r}|g_s(r)| \leq \sup_s\|\frac{dG_s(\omega)}{d\omega}\|_2$ into (97) gives

$$II \leq \frac{1}{m}\mathcal{K}_1^{(0)}\Big\{\sup_s\|\frac{dH_s(\omega)}{d\omega}\|_2\sup_u\|\frac{dG_u(\omega)}{d\omega}\|_2\Big\}.$$

Using a similar argument we obtain the same bound for II. Altogether these bounds for I, II and III give (96). To prove the result for $\mathbf{j}\in\{j,(j_1,j_2)\}$ is similar and we omit the details. □

Proposition A.5 *Suppose the assumptions in Proposition A.4 hold,* $\sup_{s,\omega}|\frac{dH_s(\omega)}{d\omega}| < \infty$ *and for all* s *and* u *we have*

$$\sup_\omega|H_s(\omega)-H_u(\omega)| \leq K\frac{|s-u|}{\hat{N}} \quad \sup_\omega|G_s(\omega)-G_u(\omega)| \leq K\frac{|s-u|}{\hat{N}}, \quad (98)$$

where $K<\infty$*. Then we have*

$$\begin{aligned}&\Big|\frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\Big(\sum_{u=k_2}^{T-m+k_2}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}(g_{s-k_1},h_{u-k_2})(s-u)-\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_{s-k_1},h_{s-k_2})(s-u)\Big)\Big|\\ \leq\ & K\Big\{\frac{1}{\hat{N}}+\frac{1}{m}\Big\}\big\{(2\sup_{s,\omega}|H_s(\omega)|\sup_{s,\omega}|\frac{dH_s(\omega)}{d\omega}|)^{1/2}+\sup_s\|\frac{H_s(\omega)}{d\omega}\|_2\big\}\sup_u\|\frac{dG_u(\omega)}{d\omega}\|_2\mathcal{K}_2^{(\mathbf{j})} \quad (99)\end{aligned}$$

and

$$\begin{aligned}&\Big|\frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_{s-k_1},h_{s-k_2})(s-u)-\frac{1}{T}\sum_{s=1}^{T}\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_s,h_s)(s-u)\Big|\\ \leq\ & K\Big\{\frac{m}{\hat{N}}\Big\}\big\{\sup_{s,\omega}|\frac{dH_s(\omega)}{d\omega}|^{1/2}\sup_u\|\frac{dG_u(\omega)}{d\omega}\|_2\mathcal{K}_2^{(\mathbf{j})}. \quad (100)\end{aligned}$$

PROOF. We first prove (99), where the summand $\sum_{u=k_2}^{T-m+k_2}$ changes to $\sum_{u=-\infty}^{\infty}$ and $\Gamma_{\mathbf{x}_s,\mathbf{x}_u}$ changes to $\Gamma_{\mathbf{x}_s,\mathbf{x}_s}$. We first observe

$$\begin{aligned}&\Big|\frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\Big\{\sum_{u=k_2}^{T-m+k_2}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}(g_{s-k_1},h_{u-k_2})(s-u)-\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_{s-k_1},h_{s-k_2})(s-u)\Big\}\Big|\\ \leq\ & I+II, \quad (101)\end{aligned}$$

where

$$\begin{aligned}I &= \frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=k_2}^{T-m+k_2}\big|\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}(g_{s-k_1},h_{u-k_2})(s-u)-\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_{s-k_1},h_{s-k_2})(s-u)\big|\\ II &= \frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u<k_2,u>T-m+k_2}\big|\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_{s-k_1},h_{s-k_2})(s-u)\big|.\end{aligned}$$

We now bound I. To bound inside the summation in I we use using (94) and the definition of $\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}$ to obtain

$$\begin{aligned}
&\left|\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}(g_{s-k_1},h_{u-k_2})^{(}s-u) - \Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_{s-k_1},h_{s-k_2})(s-u)\right| \\
\leq & \sum_{r_1=-\infty}^{\infty}|g_{s+k_1}(r_1)|\sum_{r_2=-\infty}^{\infty}\left\{\left|h_{u-k_2}(r_2)-h_{s-k_2}(r_2)\right| + \frac{|s-u|}{\hat{N}}|h_{u-k_2}(r_2)|\right\}\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2) \quad (102)
\end{aligned}$$

We start by examining $\sum_{r_2=-\infty}^{\infty}\left|h_{u-k_2}(r_2)-h_{s-k_2}(r_2)\right|\tilde{\mathcal{K}}(s-u,r_1,r_2)$, which is the first term in (102). By using the Cauchy-Schwartz inequality we have and that $\left\{\sum_{r_2=-\infty}^{\infty}\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2)^2\right\}^{1/2} \leq \sum_{r_2=-\infty}^{\infty}\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2)$ gives

$$\sum_{r_2=-\infty}^{\infty}\left|h_{u-k_2}(r_2)-h_{s-k_2}(r_2)\right|\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2) \leq \left\{\sum_{r_2=-\infty}^{\infty}\left|h_{u-k_2}(r_2)-h_{s-k_2}(r_2)\right|^2\right\}^{1/2}\sum_{r_2=-\infty}^{\infty}\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2)$$

Recalling that $h_s(r) = \int H_s(\omega)\exp(ir\omega)d\omega$ we have by Parseval's theorem and (98) that

$$\sum_{r=-\infty}^{\infty}\left|h_{u+k_2}(r)-h_{s+k_2}(r)\right|^2 = \int_0^{2\pi}|H_{s+k_2}(\omega)-H_{u+k_2}(\omega)|^2 d\omega \leq H\frac{|u-s|}{\hat{N}},$$

where $H = (2\sup_{s,\omega}|H_s(\omega)|\sup_{s,\omega}|\frac{dH_s(\omega)}{d\omega}|)^{1/2}$. Substituting the above into (102) gives

$$\begin{aligned}
&\left|\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}(g_{s-k_1},h_{u-k_2})^{(}s-u) - \Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(g_{s-k_1},h_{s-k_2})(s-u)\right| \\
\leq & \frac{|s-u|}{\hat{N}}\sum_{r_1=-\infty}^{\infty}|g_{s-k_1}(r_1)|\left(H+\sum_{r_2=-\infty}^{\infty}|h_{u-k_2}(r_2)|\right)\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2).
\end{aligned}$$

Substituting the above into I, and noting that $\sum_{r_2=-\infty}^{\infty}|h_{u-k_2}(r_2)| \leq \sup_u\|\frac{dH_u(\omega)}{d\omega}\|$ and $\sum_{r_1=-\infty}^{\infty}|g_{s-k_1}(r_2)| \leq \sup_s\|\frac{dG_s(\omega)}{d\omega}\|$ we have

$$\begin{aligned}
I \leq & \frac{1}{m^2\hat{N}T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{r_1=-\infty}^{\infty}|g_{s-k_1}(r_1)|\left(H+\sum_{r_2=-\infty}^{\infty}|h_{u-k_2}(r_2)|\right)\sum_{u=k_2}^{T-m+k_2}|s-u|\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2) \\
\leq & \hat{N}^{-1}\sup_s\|\frac{dG_s(\omega)}{d\omega}\|_2\left(H+\sup_s\|\frac{dH_s(\omega)}{d\omega}\|_2\right)\sum_{u=-\infty}^{\infty}|s-u|\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2) \\
\leq & \hat{N}^{-1}\sup_s\|\frac{dG_s(\omega)}{d\omega}\|_2\left(H+\sup_s\|\frac{dH_s(\omega)}{d\omega}\|_2\right)\mathcal{K}_2^{(\mathbf{j})}.
\end{aligned}$$

Now we bound II. Using similar methods to those used in the proof of Proposition A.4 we have

$$\begin{aligned}
II \leq & \frac{1}{m^2T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=-\infty}^{\infty}|u|\sum_{r_1,r_2=-\infty}^{\infty}|g_{s+k_1}(r_1)||h_{s+k_2}(r_2)|\tilde{\mathcal{K}}(s-u,r_1,r_2) \\
\leq & \frac{1}{m}\sup_s\|\frac{dH_s(\omega)}{d\omega}\|_2\sup_s\|G_s(\omega)\|_2\mathcal{K}_2^{(\mathbf{j})}.
\end{aligned}$$

Substituting the bounds for I and II into (101) and noting that $\|G_s(\omega)\|_2 \leq \|\frac{dG_s(\omega)}{d\omega}\|_2$ gives (99).

The proof of (100) uses the same methods used to bound I, hence we omit the details. □

A.4.2 Proofs in Section 5.4

We first outline the proof of Lemma 5.1.

PROOF of Lemma 5.1 The result can be shown by using the same arguments as those used in Dahlhaus (2000), where it is shown that a version of the Whittle likelihood containing the pre-periodogram and the localised spectrum is asymptotically equivalent to the Gaussian likelihood. As the proof is identical to the proof given in Dahlhaus (2000), we omit the details. □

In the following proposition we obtain a series of approximations to (28) which we use to approximate $\mathrm{var}(\sqrt{T}\nabla\mathcal{L}_T^{(m)}(\boldsymbol{\alpha}_0,\boldsymbol{\theta}_0))$. The proof of Proposition 5.3 follows immediately from the proposition below.

Proposition A.6 *Suppose Assumptions 2.1, 5.1 and 5.2 hold, $\hat{N}$ is defined as in (36), the operators $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}$ and $\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_u}$ are defined as in (25) and (34) respectively and*

$$\sup_j \int \Big|\frac{d^2 f_j(\boldsymbol{\vartheta}_0,\omega)}{d\omega^2}\Big|^2 d\omega < \infty \quad \text{and} \quad \sup_j \int \Big|\frac{d^2 \nabla_{\boldsymbol{\theta}} f_j(\boldsymbol{\vartheta}_0,\omega)}{d\omega^2}\Big|^2 d\omega < \infty. \tag{103}$$

Then we have

$$\Big|\frac{1}{m^2 T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\Big\{\sum_{u=k_2}^{T-m+k_2}\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\ldots,\mathbf{x}_T}(d^{(1)}_{s-m/2+k_1},d^{(2)}_{u-m/2+k_2})(s,u) - \sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(d^{(1)}_{s-m/2+k_1},d^{(2)}_{s-m/2+k_2})(s-u)\Big\}\Big| \leq K\Big\{\frac{1}{\hat{N}}+\frac{1}{m}+\frac{1}{T_m}\Big\} \tag{104}$$

$$\Big|\frac{1}{m^2 T_m}\sum_{k_1,k_2=1}^{m}\sum_{s=k_1}^{T-m+k_1}\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(d^{(1)}_{s-m/2+k_1},d^{(2)}_{s-m/2+k_2})(s-u) - \frac{1}{T}\sum_{s=1}^{T}\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(d^{(1)}_s,d^{(2)}_s)(s-u)\Big| \leq K\Big(\frac{m}{\hat{N}}+\frac{m}{T}\Big) \tag{105}$$

$$\Big|\frac{1}{T}\sum_{s=1}^{T}\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(d^{(1)}_s,d^{(2)}_s)(s-u) - \frac{N}{T}\int_0^{T/N}\sum_{u=-\infty}^{\infty}\Gamma^{(\mathbf{j})}_{\mathbf{x}(v),\mathbf{x}(v)}(d(v)^{(1)},d(v)^{(2)})(u)dv\Big| \leq K\Big(\frac{m}{\hat{N}}+\frac{1}{\hat{N}}+\frac{1}{m}\Big) \tag{106}$$

$$\Big|\frac{1}{T_m}\sum_{s=m/2}^{T-m/2}\frac{\nabla\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)\nabla\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)'}{(\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega))^2} - \frac{N}{T}\int_0^{T/N}\frac{\nabla\mathcal{F}(u,\boldsymbol{\theta}_0,\omega)\nabla\mathcal{F}(u,\boldsymbol{\theta}_0,\omega)'}{(\mathcal{F}(u,\boldsymbol{\theta}_0,\omega))^2}du\Big| \leq K\frac{m}{\hat{N}} \tag{107}$$

where $\mathbf{j}\in\{0,j,(j_1,j_2)\}$, $d^{(i)}_s(k)\in\{a_s(k),\underline{b}_s(k)\}$ (see (29) for their definition), $d^{(i)}(u,k)\in\{a(u,k),\underline{b}(u,k)\}$ (see (35) for their definition) and K is a finite constant.

PROOF. The proof follows from Propositions A.4 and A.5 We need to verify the conditions in these propositions which we do below. □

Under Assumptions 5.1 and 5.2 we have $\sup_{t,m,\omega}|\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)| < \infty$, $\sup_{t,m,\omega}|\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}| < \infty$ and $\sup_{t,m,\omega}|\nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}| < \infty$, therefore $\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\cdot)$, $\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\cdot)^{-1}$ and $\nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\cdot)^{-1}$ belong to

$L_2[0,2\pi]$. We recall that Proposition A.6 is based on the sequences $\{a_s(k)\}_k$ and $\{\underline{b}_s(k)\}_k$ (defined in (29)), the above means

$$\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1} = \sum_{k=-\infty}^{\infty} a_s(k)\exp(ik\omega) \quad \text{and} \quad \nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1} = \sum_{k=-\infty}^{\infty} \underline{b}_s(k)\exp(ik\omega). \tag{108}$$

We use the following lemma to prove Lemma A.2.

Lemma A.9 *Suppose that Assumptions 2.1, 5.1 and 5.2 hold. Then we have*

$$\sup_{t,m,\boldsymbol{\theta},\omega}\left|\frac{d\nabla^k_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}{d\omega}\right| < \infty, \tag{109}$$

$$\sup_{t,m,\boldsymbol{\theta},\omega}\left|\frac{d\nabla^k_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)^{-1}}{d\omega}\right| < \infty, \tag{110}$$

and

$$\sup_{t,m,\boldsymbol{\theta}}\sum_{r=-\infty}^{\infty}\left|\int \nabla^k_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)^{-1}\exp(ir\omega)d\omega\right| < \infty. \tag{111}$$

PROOF. In order to prove the result we will use that Assumption 5.1(iii) implies

$$\sup_j \sum_{r=0}^{\infty}|r^k\nabla_{\boldsymbol{\theta}}c_j(\boldsymbol{\theta},r)| < \infty. \tag{112}$$

To prove (109), we observe $\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega) = \sum_{j=1}^{n+1}\sum_{r=-(m-1)}^{m-1}\exp(ir\omega)\nabla_{\boldsymbol{\theta}}c_j(\boldsymbol{\theta},r)\frac{1}{m}\sum_{k=1}^{m-|r|}x_{t-m/2+k}x_{t-m/2+k+r}$, now differentiating this wrt ω gives

$$\frac{d\nabla^k_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}{d\omega} = \sum_{j=1}^{n+1}\sum_{r=-(m-1)}^{m-1} ir\exp(ir\omega)\nabla^k_{\boldsymbol{\theta}}c_j(\boldsymbol{\theta},r)\frac{1}{m}\sum_{k=1}^{m-|r|}x_{t-m/2+k}x_{t-m/2+k+r}.$$

Hence using (112) gives (109). We now prove (110) for $k=0$. Under Assumption 5.1(i), $\inf_{\boldsymbol{\theta},\omega}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega) > \delta > 0$ (for some δ). Therefore, this together with (109) gives

$$\sup_{t,m,\boldsymbol{\theta},\omega}|\frac{d\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)^{-1}}{d\omega}| = \sup_{t,m,\boldsymbol{\theta},\omega}|\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)^2}\frac{d\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)}{d\omega})| < \infty,$$

hence (110) holds for $k=0$. The proof of (110) for $k>0$, is similar and we omit the details. To prove (111), we note that (110) implies $\frac{d\nabla^k_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\cdot)^{-1}}{d\omega} \in L_2[0,2\pi]$, hence the Fourier coefficients of $\nabla^k_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta},\omega)^{-1}$ are absolutely summable, thus giving (111). □

In order for us to apply Propositions A.4 and A.5 to $\Gamma^{(k_1,k_2),(\mathbf{j})}_{\mathbf{x}_1,\dots,\mathbf{x}_T}(\underline{b}_{s+m/2-k_1},\underline{b}_{u+m/2-k_2})(s,u)$ we require boundedness of $\frac{d^2\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}}{d\omega^2}$ and $\frac{d^2\nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}}{d\omega^2}$, which is shown in the following lemma.

Lemma A.10 *Suppose Assumptions 2.1, 5.1 and 5.2 and (103) hold. Then we have*

$$\sup_{t,m,\omega}\left|\frac{d^2\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1}}{d\omega^2}\right| < \infty \quad \sup_{t,m,\omega}\left|\frac{d^2\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1}}{d\omega^2}\right| < \infty$$

PROOF. We first note that (103) implies

$$\sup_{j_1,j_2}\sum_r (r^2|c_{j_1,j_2}(\boldsymbol{\theta}_0,r)|)^2<\infty \quad \text{and} \quad \sup_{j_1,j_2}\sum_r |r^2\nabla_{\boldsymbol{\theta}} c_{j_1,j_2}(\boldsymbol{\theta}_0,r)|^2<\infty,$$

which implies

$$\sup_{t,m,\omega}|\frac{d^2\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}{d\omega^2}|<\infty \quad \text{and} \quad \sup_{t,m,\omega}|\frac{d^2\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}{d\omega^2}|<\infty. \tag{113}$$

By expanding $\frac{d^2\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1}}{d\omega^2}$ and $\frac{d^2\nabla_{\boldsymbol{\theta}}\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1}}{d\omega^2}$ and using (109) and (113) we obtain the result. □

Lemma A.11 *Suppose Assumptions 5.1(i,iii) and 5.2(i) holds. Let $\mathcal{F}_{t,m}$ and $\hat{N}$ be defined as in (15) and (36). Then we have*

$$\left|\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}-\frac{1}{\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)}\right|\leq\frac{K}{\hat{N}}|s-t|, \tag{114}$$

$$\left|\nabla\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1}-\nabla\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}\right|\leq\frac{K}{\hat{N}}|s-t| \tag{115}$$

PROOF. To prove (114), we use (15), Assumptions 5.1(iii) and 5.2(i) to obtain

$$\left|\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)-\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)\right|=\left|\sum_{j=1}^{n+1}\int_{-\pi}^{\pi}\left\{I_{t,m}^{(j)}(\lambda)-I_{s,m}^{(j)}(\lambda)\right\}f_j(\boldsymbol{\theta}_0,\omega-\lambda)d\lambda\right|\leq K\frac{|t-s|}{\hat{N}}, \tag{116}$$

for some finite constant K. Now, under Assumption 5.1(i), $\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)$ is uniformly bounded away from zero over all ω, t and m. Therefore by using (116) we obtain (114). We use a similar method to prove (115). □

Proof of Proposition A.6(i,ii) To prove Proposition A.6(i) we apply Propositions A.4 and A.5, equation (99), where depending on the block matrix in $W_T^{(m)}$ we are considering, we let $G_s(\omega)$ and $H_s(\omega)$ take either one of the functions

$$\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}=\sum_{k=-\infty}^{\infty}a_s(k)\exp(ik\omega), \quad \nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}=\sum_{k=-\infty}^{\infty}\underline{b}_s(k)\exp(ik\omega).$$

To verify the conditions in Propositions A.4 and A.5, for $\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}$ and $\nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)^{-1}$, we note that Lemma A.10 implies $\|\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\cdot)^{-1}\|_2\in L_2[0,\pi]$ and $\|\nabla_{\boldsymbol{\theta}}\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\cdot)^{-1}\|_2\in L_2[0,\pi]$ (verfying the conditions in Propositions A.4) and Lemma A.11 verifies the condition in Proposition A.5. Altogether this gives Proposition A.6(i).

To prove Proposition A.6(ii) we apply Proposition A.5, (100). The proof uses identical arguments to those given above, hence we omit the details. □

We use the following lemma to prove Proposition A.6(iii,iv).

Lemma A.12 *Suppose Assumptions 5.1(i,iii) and 5.2(i) holds. Let $\mathcal{F}_{t,m}$, $\hat{N}$ and $\mathcal{F}(u,\cdot)$ be defined as in (15), (36) and (35). Then we have*

$$\left|\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega) - \mathcal{F}(\frac{t}{N},\boldsymbol{\theta}_0,\omega)\right| \leq K\big(\frac{m}{\hat{N}} + \frac{1}{m}\big), \tag{117}$$

$$\left|\frac{1}{\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)} - \frac{1}{\mathcal{F}(\frac{t}{N},\boldsymbol{\theta}_0,\omega)}\right| \leq K\big(\frac{m}{\hat{N}} + \frac{1}{m}\big), \quad \left|\nabla\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1} - \nabla\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)^{-1}\right| \leq K\big(\frac{m}{\hat{N}} + \frac{1}{m}\big). \tag{118}$$

PROOF. We first prove (117). We note that by using Assumption 5.1(iii) (with $k = 0$) we have $\sum_{r=-\infty}^{\infty}|r\nabla_{\boldsymbol{\theta}}^{k}c_{j_1,j_2}(\boldsymbol{\theta}_0,r)| < \infty$. Therefore by using this, the definitions of $\mathcal{F}_{t,m}$ and $\mathcal{F}(u,\cdot)$ given in (15) and (35) and a similar method of proof to Lemma A.11 we have (117). To prove (118) we use that under Assumption 5.1(i), $\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)$ is uniformly bounded away from zero over all ω, t and m and similar proof to the proof of (117), we omit the details. □

Proof of Proposition A.6(iii,iv) We first prove (iii). We will use similar methods to those used in (102). We will prove it for $d_s^{(1)}(k) := a_s(k)$, $d_s^{(2)}(k) := a_s(k)$, $d^{(1)}(u,k) := a(u,k)$ and $d^{(2)}(u,k)^{(2)}$ (the proof for the other cases are identical). The difference between the two operators is

$$\begin{aligned}
&\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(a_s,a_s)(k) - \Gamma^{(\mathbf{j})}_{\mathbf{x}(\frac{s}{N}),\mathbf{x}(\frac{s}{N})}(a(\frac{s}{N}),a(\frac{s}{N}))(k)\\
\leq\ & \sum_{r_1=-\infty}^{\infty}\Big(|a_s(r_1)| + |a(\frac{s}{N},r_1)|\Big)\sum_{r_2=-\infty}^{\infty}\big|a_s(r_2) - a(\frac{s}{N},r_2)\big|\tilde{\mathcal{K}}^{(\mathbf{j})}(k,r_1,r_2)\\
\leq\ & \sum_{r_1=-\infty}^{\infty}\big(|a_s(r_1)| + |a(\frac{s}{N},r_2)|\big)\Big\{\sum_{r_2=-\infty}^{\infty}\big|a_s(r_2) - a(\frac{s}{N},r_2)|^2\Big\}^{1/2}\Big\{\sum_{r_2=-\infty}^{\infty}\tilde{\mathcal{K}}^{(\mathbf{j})}(s-u,r_1,r_2)^2\Big\}^{1/2}.
\end{aligned}$$

Therefore by using the same arguments given in the proof of Proposition A.4, and using (118) we have

$$\sum_{r=-\infty}^{\infty}\big|a_s(r) - \tilde{a}(\frac{s}{N},r)\big|^2 = \int_0^{2\pi}\big|\frac{1}{\mathcal{F}_{s,m}(\boldsymbol{\theta}_0,\omega)} - \frac{1}{\mathcal{F}(\frac{s}{N},\boldsymbol{\theta}_0,\omega)}\big|^2 d\omega \leq K\big(\frac{m}{\hat{N}} + \frac{1}{m}\big)^2.$$

Substituting the above into the difference below gives

$$\begin{aligned}
&\big|\frac{1}{T}\sum_{s=1}^{T}\sum_{u=-\infty}^{\infty}\Big(\Gamma^{(\mathbf{j})}_{\mathbf{x}_s,\mathbf{x}_s}(a_s,a_s)(s-u) - \Gamma^{(\mathbf{j})}_{\mathbf{x}(\frac{s}{N}),\mathbf{x}(\frac{s}{N})}(a(u),a(u))(k)\Big)\big|\\
\leq\ & K\big|\frac{m}{\hat{N}} + \frac{1}{m}\big|\big(\sup_s\|\frac{d\mathcal{F}_{t,m}(\boldsymbol{\theta}_0,\omega)}{d\omega}\|_2 + \|\frac{d\mathcal{F}(u,\boldsymbol{\theta}_0,\omega)}{d\omega}\|_2\big)\tilde{\mathcal{K}}_1^{(\mathbf{j})},
\end{aligned}$$

finally replacing $\sum_{s=1}^{T}$ with and integral yields (iii). It is straightfoward to prove (iv) by using (115), we omit the details. □