Estimation of Time Series Models Using Generalized Spectral Distribution

Weifeng Jin

jweifeng@eco.uc3m.es
Universidad Carlos III de Madrid

February 7, 2022

Abstract

Univariate processes with non-fundamental representations have been employed to characterize nonlinear dynamics driven from predictable future innovations. In this paper, we propose a novel estimation technique of general linear time series which are possibly non-invertible and non-causal relying on the dependence structure of residuals. The generalized spectral cumulative function is considered to capture general dependence of non-Gaussian residuals. The loss function is constructed by means of an L_2 distance between the dependence measure in the unrestricted case and the one conjectured in the restricted case under the *iid* assumption. The information at all quantiles is used to achieve model identification. This method yields consistency of estimates of the model parameters without imposing stringent conditions on higher order moments of innovations. Due to non-differentiability of the original loss function, the asymptotic distribution of the estimates is established by using a smoothed cumulative distribution function to approximate the indicator function. Finite sample properties are studied through Monte Carlo simulations. An empirical application of this approach is provided by fitting the daily trading volume of Microsoft stock by autoregressive models with noncausal representation.

Keywords: non-causality, non-invertibility, minimum distance estimation, cumulative distribution function, univariate time series.

JEL Classification: C22, C13.

1 Introduction

Time series models with non-fundamental solutions have drawn considerable attention in the econometrics literature during the last two decades. They are typically represented by noncausal and noninvertible processes models. In Macroeconomics, nonfundamentalness arises from moving average part, namely non-invertibility, has been interpreted as economic agents being endowed with larger information set than econometricians (Hansen and Sargent (1991)). Lippi and Reichlin (1993, 1994) pointed out the importance of exploration of noninvertible moving average representations for the analysis of impulse-response functions

with empirical applications in GNP-unemployment and interest rate-inflation. Leeper et al. (2013) explained noninvertibility as a natural by-product of agent's foresight with an analytical case in tax news. More empirical examples of non-invertible processes applied to modeling forward-looking behavior can be found in Alessi et al. (2011). Noncausal processes have been broadly applied in Engineering, see Tekalp et al. (1986), Gaeta et al. (1997), etc. In Economics and Finance, noncausal linear models are utilized to mimic nonlinear dynamics like locally explosive behavior and asymmetric cycles in time series. For example, a noncausal autoregression (AR) with heavy-tailed innovations could simulate the trajectory of a phase of repetitive upward trends followed by instantaneous drops, which is opposite to the pattern followed by a causal process, see Fig.1.1 where simulated AR(1) processes with root equal to 0.9 and $(0.9)^{-1}$ are depicted. This feature contributes to modelling speculative bubbles in stock markets (Gouriéroux and Zakoïan (2017), Hecq and Voisin (2020)). Moreover, a noncausal process is capable of displaying clustering volatility like ARCH behaviors which are commonly observed in financial data (Breidt et al. (2001)). Lof and Nyberg (2017) take noncausality into account in autoregressions (AR) to improve forecasting of commodity prices. Hecq et al. (2020) highlight gains in ex-post forecasting by proposing a mixed causal-noncausal AR model with inclusion of exogenous regressors. Noncausal autoregression can also be an alternative to fitting non-invertible processes (Lanne and Luoto (2013)) .

The conventional estimation techniques based-on second order moments, like OLS, fail to distinguish causal (invertible) and noncausal (non-invertible) processes due to the fact that all weakly stationary processes admit a casual and invertible representation. The information contained in the variance-covariance matrix of residuals is not sufficient to characterize the serial independence assumption of error terms. There are linear transformations on *iid* data that generate white noise sequences with the same second-order moment structure but not serially independent, like all-pass filter.[1] The logic behind pseudo Gaussian maximum likelihood (ML) is that second-order moments suffice to identify the Gaussian probabilistic structure but are not adequate for other distributions. Hence it is not applicable to noncausal and noninvertible processes driven by non-Gaussian innovations. Gourieroux and Jasiak (2018) have shown that ML method can lead to misspecification of orders of AR models when noncausality is introduced in the process. Therefore, alternative estimation techniques of general time series models are required. In the existing literature, Breid et al. (1991) introduce approximate ML procedure for estimation noncausal processes given full knowledge of the distribution of the innovations. This method achieves efficiency but imposes restrictive assumptions since in most empirical cases the distribution of the innovations is not known. Breidt et al. (2001) use the least absolute deviation (LAD) method to estimate an all-pass time series. They construct an approximate likelihood function from Laplace distribution and show that the method is efficient if and only if the innovations exactly follow that distribution. Later, ML method (Andrews et al. (2006)) and rank-based estimation (Andrews et al. (2007)) have been developed to improve the performance or relax the stringent assumptions required for estimation.

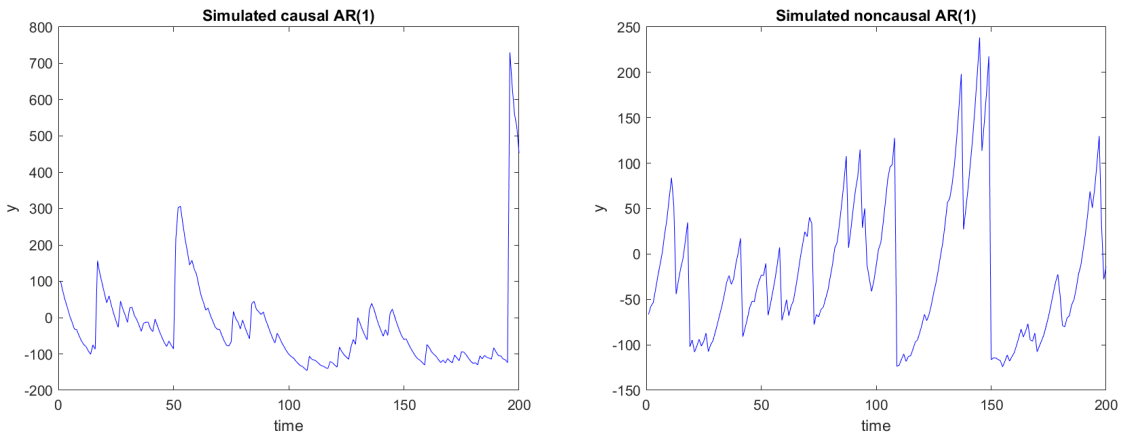
However, all the aforementioned estimation techniques are confined to all-pass time series

[1] That is, an autoregressive moving average model where all of the roots of autoregressive polynomials are the reciprocals of the roots of moving average polynomials

models. As a result, of general AR processes, a two-step procedure is required to conduct analysis based on the residuals sequences. First, fit a causal autoregrssion to the data by Gaussian ML and obtain the residuals. Second, fit the residuals by a purely noncausal all-pass model. In this approach, the validity of the second step depends on whether the residuals from the first step is a white noise sequence but not independent. In addition, the asymptotic analysis of the estimates remains open since the estimation is conducted on the residuals rather than the raw data directly. More recently, Velasco and Lobato (2018) use information from higher order moments to identify general linear time series process, but this method requires finite eighth moment of innovations to achieve the consistency of estimates. Some more recent progress has been made by Velasco (2021) and Cabello (2021) in the estimation in univariate and multivariate processes respectively by measuring independence of innovations through the characteristic functions. All the literature shows that more information on innovation serial independence beyond serial uncorrelation needs to be exploited for the identification and estimation of possibly noncausal noninvertible processes. In this paper, we construct a measure of pairwise independence of innovations based on cumulative distribution function. The distance between the joint distribution function and the product of marginal distribution functions is an indicator of independence of two random variables. This intuitive measure was originally proposed by Hoeffding (1948) and has been extended to m-dimensional random vectors by Blum et al. (1961). Skaug and Tjøstheim (1993) and Delgado (1996) consider tests of first-order and p serial dependence in the time series context grounded on this measure. Hong (1998) proposes a consistent test against all pairwise dependence via empirical distribution function by taking all lags into account. Throughout this paper, following Hong (2000) we adopt a generalized spectral distribution function based on the Fourier transformation of the measure to capture serial dependence. In a similar fashion, Du and Escanciano (2015) construct a distribution free test based on residuals for serial independence.

The loss function for estimating linear time series is constructed by an L_2 distance between the proposed measure of dependence in the unrestricted case and the conjectured one in the restricted case aplied to empirical cumulative function. There are some appealing attributes of our one-step estimation technique compared to other alternatives. First it achieves identification of the model without imposing causality and invertibility. Second, we only impose regularity conditions on the distribution of the innovations without stringent conditions on moments. Unlike other procedures using spectral densities, it does not involve subjective choices of lag windows. Moreover, compared to the approach based on characteristic functions, cumulative distribution function is more robust to outliers and more general in the sense that it can be extended to many types of dependence, for example, τ-quantile independence, conditional mean independence and pairwise independence.

The rest of paper is organized as follows. The second section introduces measure of pairwise independence based on the cumulative distribution function. Section 3 investigates the identification of the model, consistency and asymptotic properties of the proposed estimator under serial independence condition. Section 4 presents results from some Monte Carlo experiments and discusses its finite sample performance. Section 5 illustrates the use of this method by means of an empirical application. Finally, section 6 concludes and discusses



(a) Causal AR(1) process: (0.9) and log normal innovations

(b) Noncausal AR(1) process: $(0.9)^{-1}$ and log normal innovations

Figure 1.1: Simulated processes from causal and noncausal AR(1) models

some possible extensions of our estimates in procedure.

2 Model estimation based on pairwise independence measure

Consider a time series model generated by

$$Y_t = \sum_{j=-\infty}^{\infty} \varphi_j u_{t-j}, \tag{2.1}$$

where $\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of independent identically distributed (iid) innovations with zero mean. Double-sided summation in the representation of infinite moving average (2.1) allows the model to be either noncausal or noninvertible. The stationarity of Y_t is guaranteed under conditions like φ_j being absolutely summable and $\mathbb{E}\,|u_t| < \infty$. The operator $\varphi\,(\theta, L) = \sum_{j=-\infty}^{\infty} \varphi_j\,(\theta)\,L^j$ with coefficients $\varphi_j(\theta)$ and lag operator L defines the generation of Y_t in terms of parameter $\theta \in \Theta \subset \mathbb{R}^d$. Without loss of generality, we assume $\varphi_0(\theta) = 1$ for any $\theta \in \Theta$.

A common example is an autogressive moving average process of order (p, q), abbreviated as ARMA(p, q),

$$\alpha\,(L)\,Y_t = \beta\,(L)\,u_t, \tag{2.2}$$

where $\alpha\,(L) = 1 - \sum_{j=1}^{p} \alpha_j L^j$ is an autoregressive polynomial of order p and $\beta\,(L) = 1 + \sum_{j=1}^{q} \beta_j L^j$ is a moving average polynomial of order q. We allow the roots to both polynomials to lie both inside and outside unit circle, while $\alpha(z)$ and $\beta(z)$ have no common zeroes. The parameter of interest in (2.2) is $\theta = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)' \subset \{\theta \in \mathbb{R}^{p+q} : \alpha(z)\beta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| = 1, \alpha_p \neq 0, \beta_q \neq 0\}$. The restriction defined on the parameter space guarantees the existence of the Laurent expansion of $\alpha^{-1}(L)\beta(L)$, from which the coefficients φ_j in the infinite moving average representation are determined.

The residuals evaluated at any value θ are computed by

$$u_t(\theta) = \varphi^{-1}(\theta, L)Y_t = \varphi^{-1}(\theta, L)\varphi(\theta_0, L)u_t,$$

where $\varphi(\theta, L) = \sum_{j=-\infty}^{\infty} \varphi_j(\theta) L^j$. Evaluation of $\varphi(\theta, L)$ at the true value of the parameter allows to recover the sequence of innovations, i.e., $\varphi_j(\theta_0) = \varphi_j$ and $u_t(\theta_0) = u_t$ when $\theta = \theta_0$. Prior to proceeding with the estimation method based on measure of independence of residuals, we shall introduce some statistical notations to be used in the sequel. The marginal distribution of $u_t(\theta)$ and joint distribution function of $(u_t(\theta), u_{t-j}(\theta))$ are denoted by $F_\theta(x) = P(u_t(\theta) \leq x)$ and $F_{\theta,j}(x, y) = P(u_t(\theta) \leq x, u_{t-j}(\theta) \leq y)$, respectively. Let $f(u)$ be the probability density function (*pdf*) of u_t and $f_j(u, v)$ be the *pdf* of (u_t, u_{t-j}). Θ is a compact set containing the true parameter θ_0 and $I(A)$ is the indicator function of event A taking place. C is a generic positive constant that may vary in different situations.
In order to capture the generic serial dependence of the residual sequence $\{u_t(\theta)\}$ without imposing moment conditions at higher orders, we consider the distance between the joint cumulative distribution function of any pair of residuals $(u_t(\theta), u_{t-j}(\theta))$ and the product of marginal distribution functions at any given $(x, y) \in \mathbb{R}^2$

$$\begin{aligned}\sigma^*_{\theta,j}(x, y) :&= F_{\theta,j}(x, y) - F_\theta(x)F_\theta(y) \\ &= \mathbb{E}\left(I(u_t(\theta) \leq x)I(u_{t-j}(\theta) \leq y)\right) - \mathbb{E}\left(I(u_t(\theta) \leq x)\right)\mathbb{E}\left(I(u_{t-j}(\theta) \leq y)\right), \quad j = 0, \pm 1, \ldots\end{aligned} \tag{2.3}$$

which can also be interpreted as $\mathbb{C}\text{ov}\left(I(u_t(\theta) \leq x), I(u_{t-j}(\theta) \leq y)\right)$, namely, a generalization of the standard covariance between $u_t(\theta)$ and $u_{t-j}(\theta)$ by applying an indicator transformation on the random variables of interest. It is worth noting that $\sigma_{\theta,-j}(x, y) = \sigma_{\theta,j}(y, x)$ for $j \geq 1$, so that, without losing generality we can define our measure of dependence by

$$\sigma_{\theta,j}(x, y) = \sigma^*_{\theta,|j|}(x, y) \text{ for } j = 0, \pm 1, \pm 2, \ldots \quad \forall (x, y) \in \mathbb{R}^2. \tag{2.4}$$

If the dependence decays fast enough as j increases in the sense

$$\sup_{(x,y)\in\mathbb{R}^2} \sum_{j=-\infty}^{\infty} |\sigma_{\theta,j}(x, y)| < \infty,$$

we can define the generalized spectral density based on measure (2.4) at any frequency ω in $[-\pi, \pi]$ with $i = \sqrt{-1}$, by

$$h_\theta(x, y; \omega) := \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_{\theta,j}(x, y) e^{-ij\omega}, \quad \omega \in [-\pi, \pi].$$

and the associated generalized spectral distribution function by

$$H_\theta(x, y; \lambda) = 2\int_0^{\lambda\pi} h_\theta(x, y; \omega) d\omega = \sigma_{\theta,0}(x, y)\lambda + 2\sum_{j=1}^{\infty} \sigma_{\theta,j}(x, y) \frac{\sin j\pi\lambda}{j\pi} \quad \lambda \in [0, 1]. \tag{2.5}$$

The same approach has been applied to test the hypothesis of serial independence against all possible pairwise dependence alternatives in Hong (2000). Some statistics exploiting the same distributional information have been developed by either replacing (x, y) with corresponding quantiles or by defining covariance based on copulas associated with any pair of random

variables of interest. The latter procedure has been used to characterize nonlinear sequential dependence that cannot be fully captured by correlations of higher order moments in Lee and Rao (2011), Kley et al. (2016) and Hagemann (2011).
Under the independent structure on $u_t(\theta)$,

$$h_\theta(x, y; \omega) = \frac{1}{2\pi}\sigma_{\theta,0}(x, y) \quad \forall (x, y) \in \mathbb{R}^2 \tag{2.6}$$

at any frequency ω since $\sigma_{\theta,j}(x, y) = 0$ for all $j \neq 0$ and any given pair $(x, y) \in \mathbb{R}^2$, where

$$\sigma_{\theta,0}(x, y) = F_\theta(x \wedge y) - F_\theta(x)F_\theta(y) = F_\theta(x \wedge y)(1 - F_\theta(x \vee y)).$$

Moreover, the associated generalized spectral distribution function becomes

$$H_\theta(x, y; \lambda) = \sigma_{\theta,0}(x, y)\lambda, \quad \lambda \in [0, 1]. \tag{2.7}$$

3 Model estimation under serial independence

In this section, we study the identification of general linear time series models and investigate the asymptotic properties of the proposed estimate based on the generalized spectral distribution function introduced in Section 2.

3.1 Model identification under serial independence

In this paper, the criterion we adopt to identify the parameter θ in the model is based on the quadratic distance between the generalized spectral distribution function $H_\theta(x, y; \lambda)$ of residuals $u_t(\theta)$ and the counterpart under *iid*-ness.
Given any $\theta \in \Theta$, the residuals $u_t(\theta)$ are computed by

$$u_t(\theta) = \varphi^{-1}(\theta, L)\varphi(\theta_0, L)u_t = \phi(\theta, L)u_t$$

where the linear filter $\phi(\theta, L)$ plays a crucial role in the dependence of sequence of $u_t(\theta)$. If u_t follows non-Gaussian distribution, $u_t(\theta)$ will be serially dependent as long as $\phi(\theta, L) \neq 1$. The conventional methods based on second-order moments fail to discriminate noncausal and noninvertible process from causal and invertible counterparts, as this linear filter can generate uncorrelated but not independent sequences like all-pass models, see e.g. Breidt et al. (2001). In the Gaussian probabilistic structure, being uncorrelated implies independence. Therefore, in order to achieve identification we need to impose following assumption to rule out this possibility.

Assumption 1. *1. Given a compact Θ, for any $\theta \neq \theta_0, \phi(\theta, z) \neq a_0 z^{j_0}$ for any j_0 and some nonzero constant a_0 in a subset of positive measure of $\mathbb{C}$ such that $|z| = 1$.*

2. If $|\phi(\theta, z)|^2 = 1$ a.e. for $z \in \mathbb{C}$ such that $|z| = 1$ for some $\theta \neq \theta_0$, then u_t is non-Gaussian.

Assumption 1.1 guarantees that the true innovation sequence can be only recovered at θ_0. Assumption 1.2 controls for the special case when the innovation is Gaussian and the residual

sequence permits non-unique serially independent solutions. If such a linear filter generating an uncorrelated sequence exists with $\phi(\theta, z) \neq 1$, we have to impose non-Gaussianity on the innovation.

Assumption 2. *For compact Θ and $\mu_0 > 1$,*

$$\sup_{\theta\in\Theta} |\varphi_j(\theta)| + \sup_{\theta\in\Theta} \left|\varphi_j^{(-1)}(\theta)\right| \leq C|j|^{-\mu_0}, \quad j = \pm 1, \pm 2, \ldots$$

We bound coefficients in the filter $\phi(\theta, L)$ uniformly in θ in order to make it summable in absolute value. Together with $\mathbb{E}\,|u_t| < \infty$ this assumption ensures that any residual sequence determined by θ is stationary. Further it allows us to analyze some statistical properties of $u_t(\theta)$ which are time invariant.

Given Assumption 1, when $\theta \neq \theta_0$, $u_t(\theta)$ is not pairwise independent if u_t follows non-Gaussian distribution. $\sigma_{\theta,j}(x, y) \neq 0$ for some $j \neq 0$ and $(x, y) \in \mathbb{R}^2$ since $F\left(u_t(\theta) \leq x, u_{t-j}(\theta) \leq y\right)$ cannot be factorized into a product of two marginal probabilities. In order to exploit all information contained in the distribution of $u_t(\theta)$, the L_2 distance defined on the generalized spectral distribution function is aggregated over (x, y) and frequency λ in Cramér-von Mises criterion,

$$\begin{aligned}
\mathcal{Q}_0\left(\theta\right) :=& L^2\left(H_\theta(x, y; \lambda), H_\theta^{iid}(x, y; \lambda)\right) \\
=& \int_{\mathbb{R}^2} \int_0^1 \left| 2 \sum_{j=1}^{\infty} \sigma_{\theta,j}(x, y) \frac{\sin j\pi\lambda}{j\pi} \right|^2 d\lambda dW(x, y) \\
=& 2 \int_{\mathbb{R}^2} \sum_{j=1}^{\infty} \sigma_{\theta,j}^2(x, y) \frac{1}{(j\pi)^2} dW(x, y),
\end{aligned} \tag{3.1}$$

where the last equality comes immediately from Parseval's identity and for any weighting function W which satisfies the following condition.

Assumption 3. *$W(x, y) = W(x)W(y)$ where W is a probability distribution defined on $\mathbb{R}$, continuous and strictly increasing.*

The unboundedness of the support of W is required for the full characterization of pairwise independence of $(u_t(\theta), u_{t-j}(\theta))$ at any j. The continuous weighting function rules out the special case in which $\sigma_{\theta,j}(x, y) \neq 0$ but $\mathcal{Q}_0(\theta) = 0$, see Hoeffding (1948). The factorization of weighting functions is to simplify the subsequent analysis of estimates based on this population function without sacrificing any power of detecting pairwise dependence. Under Assumption 1-3, u_t being *iid* with zero mean and $\mathbb{E}\,|u_t| < \infty$,

$$\mathcal{Q}_0(\theta) > 0 \text{ when } \theta \neq \theta_0$$

due to some non-degenerated term $\sigma_{\theta,j}^2(x, y)$. By Weierstrass theorem, the continuous non-negative function $Q_0(\theta)$ admits its minimum at 0 in the compact set Θ. Assumption 1 ensures that the minimum can be only attained when $\theta = \theta_0$. Thus, the identification of the

parameter θ_0 in Θ is achieved.

It is worth noting that population distance function (3.1) can be interpreted as an infinite weighted sum of pairwise dependence measure $\int_{\mathbb{R}^2} \sigma^2_{\theta,j}(x,y)dW(x,y)$, which is a generalization of test statistic proposed by Skaug and Tjøstheim (1993) by replacing joint distribution $F_{\theta,j}(x,y)$ by any function satisfying Assumption 3 as a weighting function. The summand is down weighted for higher order lags by the factor $(j\pi)^{-2}$. An equivalent criterion can be constructed by replacing general covariance by copula covariance defined as

$$\sigma^c_{\theta,j}(u_1,u_2) = \mathbb{C}\text{ov}\left(I\left(U_t(\theta) \leq u_1, U_{t-j}(\theta) \leq u_2\right)\right)$$

where $U_t(\theta)$ is the cumulative distribution function of $u_t(\theta)$ and (u_1,u_2) is defined in the interval $[0,1]$. The copula covariance is invariant to monotonic transformations and free of subjective choice in the weighting functions. However the effect of estimating $U_t(\theta)$ is not trivial in the asymptotic analysis.

3.2 Asymptotic properties of estimates under serial independence

In this section, we investigate the asymptotic properties of the estimates based on the sample counterpart of population distance function $\mathcal{Q}_0(\theta)$ using residuals. In practice, since we only observe T finite samples, the computed residuals are approximated by a truncation in infinite moving average representation,

$$\hat{u}_t(\theta) = \varphi^{-1}(\theta, L)Y_t I\{1 \leq t \leq T\}$$

where the information lost from the part

$$\delta_T(\theta) =: u_t(\theta) - \hat{u}_t(\theta) = \left(\sum_{j=t}^{\infty} \varphi_j^{(-1)}(\theta) + \sum_{j=-\infty}^{t-T-1} \varphi_j^{(-1)}(\theta)\right) Y_{t-j}$$

can be shown to be asymptotically negligible as Assumption 2 guarantees that coefficients $\varphi_j^{(-1)}(\theta)$ decay at a sufficient rate when $j \to \infty$ uniformly in θ. Based on the sequence of residuals $\hat{u}_t(\theta)$ for any θ, the sample loss function can be constructed as

$$\hat{\mathcal{Q}}_T(\theta) = 2\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\frac{1}{(j\pi)^2}\int_{\mathbb{R}^2} \hat{\sigma}^2_{\theta,j}(x,y)dW(x,y) \tag{3.2}$$

where

$$\hat{\sigma}_{\theta,j}(x,y) = \frac{1}{T-j}\sum_{t=j+1}^{T} I\left(\hat{u}_t(\theta) \leq x\right) I\left(\hat{u}_{t-j}(\theta) \leq y\right) - \frac{1}{(T-j)^2}\sum_{t=j+1}^{T} I\left(\hat{u}_t(\theta) \leq x\right) \sum_{t=j+1}^{T} I\left(\hat{u}_{t-j}(\theta) \leq y\right)$$

and $(1 - \frac{j}{T})$ is a finite sample correction. Compared to the distance criterion defined on the generalized spectral density function $h(x,y;\lambda)$, this criterion has the appealing advantage of avoiding any subjective choice of smoothing functions which can be a delicate issue in finite

samples. The proposed estimator of θ_0 is defined as the minimum of $\hat{\mathcal{Q}}_T(\theta)$

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}}\, \hat{\mathcal{Q}}_T(\theta)$$

A deeper analysis on the dependence of residual process is required prior to the analysis on the estimates. Assumption 2 plus the condition of u_t having a continuous *pdf* guarantee the mixing condition of process if the moving average representation of $u_t(\theta)$ is one-sided, i.e. purely causal or noncausal, or contains only finite lags and leads. The problem stays unclear when we allow for two sided infinite summation. Nevertheless, we are able to draw a uniform bound for the generalized covariance at any $j \in \mathbb{Z}$ under the following extra smoothness condition on the distribution of u_t.

Assumption 4. *The innovation u_t admits a density $f(u)$ with the first order derivative $f^{(1)}(u)$ that is Lebesgue integrable and has a^{th} order bounded moment, i.e. $\int_{\mathbb{R}} \left| f^{(1)}(u) \right| du < \infty$ and $\int_{\mathbb{R}} \left| u^a f^{(1)}(u) \right| du < \infty$.*

Lemma 3.1. *Assume u_t is iid, mean zero, $\mathbb{E}\,|u_t| < \infty$, satisfying Assumption 4 with $a = 1$, then, under Assumption 2, we have*

$$|\sigma_{\theta,j}(x,y)| \leq C j^{1-\mu_0}$$

uniformly in $(x,y) \in \mathbb{R}^2$ and $\theta \in \Theta$, for $\mu_0 > 1$ and $C < \infty$

Lemma 3.1 provides the geometric decay of the covariance at any percentiles in the distribution of residuals in a similar fashion as mixing condition. The condition on the derivative of density can be regarded as a slightly stronger version of uniformly boundedness of $f(u)$ and $\mathbb{E}\,|u_t| < \infty$ to contain the "thickness" of the tail of the distributions.

The uniform convergence of the estimator is shown by the Theorem 2.1 in Newey (1991) given that our loss function is not differentiable. The stochastic equicontinuity of $\hat{\mathcal{Q}}_t(\theta)$ can be deduced from the following Lemma 3.2 and Assumption 5.

Lemma 3.2. *Assume u_t is iid, mean zero, $\mathbb{E}\,(u_t)^2 < \infty$, satisfying Assumption 4 with $a = 2$, under Assumption 2, for $\theta \in \Theta$,*

$$\mathbb{E}\left| \hat{F}_{\theta,j}(x,y) - F_{\theta,j}(x,y) \right|^2 \leq C\left(1 \wedge \frac{j}{T-j}\right) + C\frac{j^{2-\mu_0}}{T-j} + C\frac{\ln T}{(T-j)^{\mu_0-1}}$$

uniformly in (x,y), for $\mu_0 > 1$ and $C < \infty$, $j = 1, 2, \ldots$

Assumption 5. *The filter $\phi(\theta; z)$ is differentiable with the first order derivative $\phi^{(1)}(\theta; z) := \frac{\partial}{\partial\theta}\phi^{-1}(\theta; z) = \sum_{j=-\infty}^{\infty} \phi_j^{(1)}(\theta) z^j$ such that there exists a $\mu_1 > 1$,*

$$\sup_{\theta \in \Theta} \left\| \phi_j^{(1)}(\theta) \right\| \leq C|j|^{-\mu_1}, \quad j = \pm 1, \pm 2, \ldots$$

Assumption 5 imposes further restrictions on the smoothness of the linear filter $\phi_j(\theta)$ to achieve uniform boundedness of the derivative of the expectation of empirical *cdf* which

plays a crucial role in achieving stochastic equicontinuity and consistency. Then, the following theorem provides us the consistency.

Theorem 3.3. *Assume $\{u_t\}$ is iid with zero mean and $\mathbb{E}\left(u_t\right)^2 < \infty$ and satisfies Assumption 4 with $a = 2$, $\theta_0 \in \Theta$, $\mu_0 > 3, \mu_1 > 1$, Under Assumptions 1-3 and Assumption 5, as $T \to \infty$,*

$$\hat{\theta}_T \longrightarrow_p \theta_0.$$

In the investigation of asymptotic normality of the estimates based on the sample counterpart of population function. The non-differentiability of indicator function does not enable us to derive the asymptotic distribution based on the linear expansion of the score around the true value. Neither we can adapt the method in quantile regression to this estimate as the function is not convex. Therefore, we first approximate the indicator function with a smoothed cumulative distribution function $\Lambda(u)$ with positive and uniformly bounded *pdf* $\lambda(u)$ together by means of a smoothing parameter h such that

$$\Lambda(\frac{z}{h}) \to I\left(z > 0\right) \text{ for } |z| > 0 \text{ when } h \to 0$$

The positiveness of *pdf* $\lambda(u)$ ensures no loss of information in the transformation procedure. The new smoothed loss function is obtained by replacing the indicator function with the smoothed *cdf* in the original formula,

$$\tilde{Q}_T\left(\theta; h\right) = 2\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\frac{1}{(j\pi)^2}\int_{\mathcal{R}^2}\tilde{\sigma}^2_{\theta,j}(x, y; h)dW(x, y),$$

where

$$\tilde{\sigma}_{\theta,j}(x, y; h) = \tilde{F}_{\theta,j}\left(x, y; h\right) - \tilde{F}_{\theta,j}\left(x, \infty; h\right)\tilde{F}_{\theta,j}\left(\infty, y; h\right)$$
$$\tilde{F}_{\theta,j}\left(x, y; h\right) = \frac{1}{T-j}\sum_{t=j+1}^{T}\Lambda\left(\frac{x - \hat{u}_t(\theta)}{h}\right)\Lambda\left(\frac{y - \hat{u}_{t-j}(\theta)}{h}\right).$$

The corresponding estimator from the smoothed version is defined by

$$\tilde{\theta}^h_T = \underset{\theta\in\Theta}{\operatorname{argmin}}\,\tilde{Q}_T\left(\theta; h\right).$$

We start with the simple case when h is fixed and positive. The identification of the parameter in the model is fulfilled for any given $h > 0$ as $\Lambda(\frac{\cdot}{h})$ is a transformation from the class that is totally revealing, see Stinchcombe and White (1998).

Theorem 3.4. *Let $\{u_t\}$ be iid with zero mean, $\mathbb{E}\left|u_t\right| < \infty, \theta_0 \in \Theta$, and Λ be a strictly increasing cdf defined on unbounded support with density λ uniformly bounded by C, $\mu_0 > 3, \mu_1 > 1$, under Assumptions 1-3 and 5, as $T \to \infty$,*

$$\tilde{\theta}^h_T \longrightarrow_p \theta_0 \textit{ for any fixed } h > 0$$

The proof of the consistency for $\tilde{\theta}^h_T$ given any fixed positive h is similar to the consis-

tency theorem in Velasco (2021) by replacing the characteristic function with our proposed smoothed *cdf* Λ. The choice of smoothing parameter h does not affect consistency. Before analyzing the asymptotic distribution of $\tilde{\theta}_T^h$, it is useful to define following variables to simplify the notations,

$$
\begin{aligned}
e_t^h &:= \int_{\mathbb{R}} \left(\Lambda\left(\frac{x-u_t}{h}\right) - \varphi^h(x) \right) \lambda^h(x) dW(x) \\
\nu_t^h &:= \int_{\mathbb{R}} \left(\Lambda\left(\frac{x-u_t}{h}\right) - \varphi^h(x) \right) \mu^h(x) dW(x) \\
E_{t-1}^h &:= \sum_{j=1}^{\infty} \frac{1}{j^2} \phi_{-j}^{(1)}(\theta_0) e_{t-j}^h \\
V_{t-1}^h &:= \sum_{j=1}^{\infty} \frac{1}{j^2} \phi_{j}^{(1)}(\theta_0) \nu_{t-j}^h,
\end{aligned}
$$

where

$$
\varphi^h(x) := \mathbb{E}\left(\Lambda\left(\frac{x-u_t}{h}\right)\right) \quad \lambda^h(x) := \frac{1}{h}\mathbb{E}\left(\lambda\left(\frac{x-u_t}{h}\right)\right) \quad \mu^h(x) := \mathbb{E}\left(u_t \Lambda\left(\frac{x-u_t}{h}\right)\right).
$$

As described above, $\{e_t^h\}, \{\nu_t^h\}$ are *iid* sequences with mean zero with their corresponding variance $\{\sigma_{e;h}^2, \sigma_{\nu;h}^2\}$ and covariance $\sigma_{e\nu;h}^2$. Likewise, $\{e_t^h V_{t-1}^h\}$, $\{\nu_t^h E_{t-1}^h\}$ are martingale difference sequences conditional on the $\sigma-$field generated by $\{u_{t-j}, j \geq 1\}$ set.

$$
\Sigma_{0,a} := \sum_{j=1}^{\infty} j^{-2a} \phi_j^{(1)}(\theta_0) \phi_j^{(1)}(\theta_0)' \quad \Sigma_{0,a}^* := \sum_{j=1}^{\infty} j^{-2a} \phi_{-j}^{(1)}(\theta_0) \phi_{-j}^{(1)}(\theta_0)'
$$

and $\Sigma_{0,a}^\dagger := \sum_{j=1}^{\infty} j^{-2a} \phi_j^{(1)}(\theta_0) \phi_{-j}^{(1)}(\theta_0)'$ for $a = 1, 2$.

$$
\begin{aligned}
H_{1,h} &= \left(\Sigma_{0,2} + \Sigma_{0,2}^*\right) \sigma_{e;h}^2 \sigma_{\nu;h}^2 + \left(\Sigma_{0,2}^\dagger + \Sigma_{0,2}^{\dagger'}\right) \sigma_{e\nu;h}^2 \\
H_{0,h} &= \left(\Sigma_{0,1} + \Sigma_{0,1}^*\right) \rho_1^h \rho_2^h + \left(\Sigma_{0,1}^\dagger + \Sigma_{0,1}^{\dagger'}\right) \left(\rho_{12}^h\right)^2,
\end{aligned}
$$

where

$$
\rho_1^h = \int_{\mathbb{R}} \left(\mu^h(x)\right)^2 dW(x) \quad \rho_2^h = \int_{\mathbb{R}} \left(\lambda^h(x)\right)^2 dW(x) \quad \rho_{12}^h = \int_{\mathbb{R}} \mu^h(x) \lambda^h(x) dW(x)
$$

Finally, two further assumptions need to be used for asymptotic normality.

Assumption 6. *The filter $\phi(\theta) = \sum_{j=-\infty}^{\infty} \phi_j(\theta)$ with three derivatives $\phi^{(a)}(\theta)$ satisfies following condition:*

$$
\sup_{\theta \in \Theta} \left\| \phi_j^{(a)}(\theta) \right\| < C\,|j|^{-\eta_a} \quad \textit{with } \eta_a > 1
$$

for a=1,2,3 and C.

Assumption 7.

1. *The smoothed cumulative distribution function $\Lambda(u)$ admits uniformly bounded positive probability density function $\lambda(u)$ with differentiable first order and second order derivatives $\dot{\lambda}(u)$ and $\ddot{\lambda}(u)$ uniformly bounded by some constants C.*

2. *$H_{0,h}$ is positive definite.*

Stronger conditions on smoothness of linear filter $\phi(\theta; z)$ and $cdf\,\Lambda$ are imposed for the analysis of the score and Hessian matrix of $\tilde{\mathcal{Q}}_T(\theta; h)$. Finite third order moment of innovations and uniform boundedness of density function λ together with its higher order derivatives are necessary for the convergence of the aforementioned score and Hessian matrix. Assumption 7.2 ensures the components of covariance-variance matrix is invertible.

Theorem 3.5. *Let $\{u_t\}$ be iid with zero mean, $\mathbb{E}\left|u_t\right|^3 < \infty$ and , $\mu_0 > 3, \mu_1 > 1, \theta_0 \in \Theta$, Under Assumptions 1-3 and 6-7, as $T \to \infty$,*

$$T^{1/2}\left(\tilde{\theta}_T^h - \theta_0\right) \longrightarrow_d \mathcal{N}\left(0, H_{0,h}^{-1} H_{1,h} H_{0,h}^{-1}\right)$$

Now we set h arbitrarily close to zero to numerically approximate the asymptotic distribution of the original estimator based on the indicator transformation of the residuals.
As $h \to 0$,

$$\varphi^h(x) \to F(x), \quad \lambda^h(x) \to f(x),$$

and $\mu^h(x) \to \mu(x) \equiv \mathbb{E}\left(u_t I\left(u_t \leq x\right)\right)$. The asymptotic variance becomes

$$\begin{aligned} H_1 &= \left(\Sigma_{0,2} + \Sigma_{0,2}^*\right)\sigma_e^2\sigma_\nu^2 + \left(\Sigma_{0,2}^\dagger + \Sigma_{0,2}^{\dagger'}\right)\sigma_{e\nu}^2 \\ H_0 &= \left(\Sigma_{0,1} + \Sigma_{0,1}^*\right)\rho_1\rho_2 + \left(\Sigma_{0,1}^\dagger + \Sigma_{0,1}^{\dagger'}\right)\left(\rho_{12}\right)^2, \end{aligned}$$

with $\{\sigma_e, \sigma_\nu, \sigma_{e\nu}, \rho_1, \rho_2, \rho_{12}\}$ being the limits of $\{\sigma_{e;h}, \sigma_{\nu;h}, \sigma_{e\nu;h}, \rho_1^h, \rho_2^h, \rho_{12}^h\}$ when $h \to 0$. The identification of this model is still valid under the same structure as h goes to zero. This immediately follows from

$$\mathcal{Q}_0(\theta; h) = \mathcal{Q}_0(\theta) + O(h^2).$$

Some extra care need to be taken on the smoothing parameter h to preserve the classical rate $T^{1/2}$ in the application of CLT. In effect, the rate of convergence of $\frac{\partial}{\partial\theta}\tilde{\sigma}_{\theta,j}(x, y; h)$ needs to be controlled unchanged as $h \to 0$. From the bias square and variance

$$\mathbb{E}\left\|\frac{\partial}{\partial\theta}\tilde{\sigma}_{\theta_0,j}(x, y; h) - \frac{\partial}{\partial\theta}\sigma_{\theta_0,j}(x, y)\right\|^2 = O\left(h^4 + h^{-1}(T-j)^{-1}\right),$$

it can be concluded that h has to go to zero but not faster than T^{-1} to guarantee the approximation effect is asymptotically negligible. The other restriction comes from the non-centering error

$$\varphi^h(x)\varphi^h(y) = F(x)F(y) + O(h^2).$$

This bias must be $o(T^{-1/2})$ to be negligible in the asymptotic distribution. Hence, h cannot go to zero too slowly, implying that $h = o(T^{-1/4})$. Induced from preceding assumption for fixed h, we propose following conditions for the limiting behavior of $\tilde{\theta}_T^h$ as $h \to 0$.

Assumption 8. *1. The innovation $\{u_t\}$ admits uniformly bounded probability density function $f(u)$ with differentiable derivatives $f^{(a)}(u)$ of order a uniformly bounded by some constants C for a =1,2.*

2. H_0 is positive definite.

Theorem 3.6. *Under Assumptions 1-3 and Assumptions 6-8,* $\{u_t\}$ *iid with zero mean,* $\mathbb{E}\left|u_t\right|^3 < \infty, \theta_0 \in \Theta,\ \mu_0 > 3, \mu_1 > 1,$ *as* $T \to \infty, h \to 0$

$$T^{1/2}\left(\tilde{\theta}_T^h - \theta_0\right) \longrightarrow_p \mathcal{N}\left(0, H_0^{-1} H_1 H_0^{-1}\right)$$

Theorem 3.6 states the asymptotic distribution of the estimates based on the smoothed *cdf* when Λ approaches the indicator function as $T \to \infty$. It can numerically represent the asymptotic behavior of $\hat{\theta}_T$ obtained from the sample loss function based on the empirical *cdf*.

The asymptotic variance in the causal and invertible models can be simplified to

$$\kappa \Sigma_{0,1}^{-1} \Sigma_{0,2} \Sigma_{0,1}^{-1}$$

with $\kappa = \sigma_e^2 \sigma_\nu^2 (\rho_1 \rho_2)^{-2}$ since some components like $\Sigma_{0,a}^*$ and $\Sigma_{0,a}^\dagger$ are degenerated for $a = 1, 2$. The potential gain in the efficiency of the estimates can reach $\kappa \Sigma_{0,0}^{-1}$ by replacing spectral *cdf* with spectral *pdf* in the loss function, in which the natural down weights j^{-2} on the higher lags are not imposed. It can be shown $\Sigma_{0,1}^{-1} \Sigma_{0,2} \Sigma_{0,1}^{-1} - \Sigma_{0,0}^{-1}$ is positive semidefinite.

Remark.

Boldin et al. (1997) introduce a sign-based estimation method of causal AR models where the information is exploited from the generalized autocovariance

$$\mathbb{E}\left(\operatorname{sgn}\left(u_t(\theta)\right) \operatorname{sgn}\left(u_{t-j}(\theta)\right)\right) = \mathbb{E}\left(\left(2I\left(u_t(\theta) > 0\right) - 1\right)\left(2I\left(u_{t-j}(\theta) > 0\right) - 1\right)\right)$$

at any lag $j \geq 1$ with condition that $F(0) = 1/2$. The proposed approach in this paper could be regarded as an extension of sign-based estimator in the sense it takes into account all the percentiles of the distribution by replacing $(0, 0)$ by any $(x, y) \in \mathbb{R}^2$

$$\int_{\mathbb{R}^2} \mathbb{E}\left(\operatorname{sgn}\left(u_t(\theta), x\right) \operatorname{sgn}\left(u_{t-j}(\theta), y\right)\right) dW(x, y),$$

where

$$\begin{aligned} \operatorname{sgn}\left(u_t(\theta), x\right) &= 2I\left(u_t(\theta) > x\right) - 2\left(1 - F_\theta(x)\right) \\ &= -2I\left(u_t(\theta) \leq x\right) + F_\theta(x), \end{aligned}$$

and replacing the weighting factor on higher lags by exponential terms arising from the derivative of the population generalized covariance. Uniform analysis can be carried out on the basis of this adaptation.

3.3 Standard error calculation

Generally there is no closed-form expression of integration in the variance, i.e. H_0 and H_1. We propose to estimate the components in asymptotic variance directly for the sake of time-saving in comparison with bootstrap method. Following the definitions of e_t and ν_t, they

can be replaced by their sample counterparts:

$$\hat{e}_t := \int_{\mathbb{R}} \left(I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right) - \hat{F}(x)\right) \hat{f}(x) dW(x)$$
$$\hat{\nu}_t := \int_{\mathbb{R}} \left(I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right) - \hat{F}(x)\right) \hat{\mu}(x) dW(x)$$

where

$$\begin{aligned}
\hat{F}(x) &= \frac{1}{T}\sum_{t=1}^{T} I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right) \\
\hat{f}(x) &= \frac{\hat{F}(x) - \hat{F}(x')}{x - x'} \text{ for a properly chosen sequence of } x \\
\text{or } &= \frac{1}{Th}\sum_{t=1}^{T} \lambda\left(\frac{x - \hat{u}_t(\hat{\theta}_T)}{h}\right) \text{ for a sufficiently small } h \text{ and smooth density } \lambda \\
\hat{\mu}(x) &= \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t(\hat{\theta}_T) I\left(\hat{u}_t(\hat{\theta}_T) \leq x\right).
\end{aligned}$$

The estimation of the density function can be also computed by any consistent kernel density. We use numerical integration algorithm to calculate the integration with smoothed weighting function $W(x)$. Alternatively, one can choose the empirical *cdf* as weighting function to replace integration by averages at data points,

$$\begin{aligned}
\hat{e}_t :=& \frac{1}{T}\sum_{s=1}^{T} \left(I\left(\hat{u}_t(\hat{\theta}_T) \leq \hat{u}_s(\hat{\theta}_T)\right) - \hat{F}\left(\hat{u}_s(\hat{\theta}_T)\right)\right) \hat{f}(\hat{u}_s\left(\hat{\theta}_T\right) \\
\hat{\nu}_t :=& \frac{1}{T}\sum_{s=1}^{T} \left(I\left(\hat{u}_t(\hat{\theta}_T) \leq \hat{u}_s(\hat{\theta}_T)\right) - \hat{F}\left(\hat{u}_s(\hat{\theta}_T)\right)\right) \hat{\mu}\left(\hat{u}_s(\hat{\theta}_T)\right),
\end{aligned}$$

where $\{\hat{\sigma}_e^2, \hat{\sigma}_\nu^2, \hat{\sigma}_{e\nu}^2\}$ are the sample element of covariance-variance matrix of $(\hat{e}_t, \hat{\nu}_t)$. The estimates of $\{\rho_1, \rho_2, \rho_{12}\}$ involve numerical integration of $\{(\hat{\mu}(x))^2, \left(\hat{f}(x)\right)^2, \hat{\mu}(x)\hat{f}(x)\}$ with respect to x over a chosen $W(x)$. The derivative of the linear filter can be obtained from the model once the order is determined and $\{\Sigma_{0,a}, \Sigma_{0,a}^*, \Sigma_{0,a}^\dagger\}_{a=1,2}$ are estimated by plugging $\hat{\theta}_T$ in the corresponding expressions.

4 Simulation

In this section we carry out some Monte Carlo simulations to investigate finite sample properties of the proposed estimates with different innovation distributions and sample sizes. Recall that the method works regardless of subjective choice of weighting function W, but in practice it can be affected by scaling. The indicator in the loss function is scale-free but the weights imposed on $\hat{\sigma}_{\theta,j}(x, y)$ could differ before and after rescaling the residuals $u_t(\theta)$. For example, if W is set as standard normal distribution, residuals $u_t(\theta)$ whose values lie outside interval $(-3, 3)$ will be assigned very trivial weights as a result of 3-σ rule of thumb. In this case, the estimates would lose efficiency since they do not fully exploit the information contained in the extremes (tails) of the residuals. To overcome this issue, we propose two

approaches.
In the first one we standardize the residuals by dividing the original sequence by its standard deviation

$$u_t^*(\theta) := \frac{u_t(\theta)}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(u_t(\theta) - \overline{u(\theta)}\right)^2}},$$

where $\overline{u(\theta)}$ is the sample mean of the residual sequence. The standard normal distribution is chosen to be W (Logistic distribution can also be a good candidate and it does not make much difference compared with Gaussian distribution). The detailed calculation of the loss function in finite samples can be found in the Appendix. It can be shown that this standardization does not change the asymptotic properties of the proposed estimator. The second approach relies on fitting the weighting functions using the empirical distribution functions of residuals which play a role as an automatic rescaling scheme for each sequence of residuals respectively. The loss function can be simplified to

$$2\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{\sigma}^2_{\theta,j}(u_s(\theta), u_t(\theta)).$$

Both approaches allow to avoid numerical integration and subjective choice of rescaling parameter.
In the first experiment, we consider AR(1) processes with *iid* innovations generated by uniform distribution $U_{[-5,5]}$, t-distribution t_3 and centered chi-square distribution χ_5. These choices of innovations basically cover various distributions with symmetry and asymmetry ($\kappa_3 = 0, 0$ and $\sqrt{\frac{8}{5}}$), bounded and heavy-tailed property ($\kappa_4 = -\frac{6}{5}, \infty$ and $\frac{12}{5}$).[2] We try sample sizes T=100 and 200. The parameter in AR(1) models are 0.4 (0.4^{-1}) and 0.9 (0.9^{-1}) in the causal (noncausal) cases. The simulation results over 100 replications are reported in Table.1, which includes percentage of correct root identification in 100 replications (PCI), bias of the estimates (Bias) and mean square errors (MSE) given the sample size, distribution of innovations and true parameters of AR(1) models in both approaches.
As can be observed, when the innovation has either heavy tail or asymmetry, our proposed method works better. That indicates we gain more information from skewness and kurtosis. This result coincides with the estimation technique proposed by Velasco and Lobato (2018) exploiting information from higher order moments. The percentage of correct root identification increases and MSE of estimates decreases as the sample size increases. Generally both approaches either using standardization or weighting by empirical *cdf* work well. In the case of t_3 and χ_5, the approach by standard normal *cdf* outperforms the one by empirical *cdf* but in the other way around when the innovation follows uniform distribution. Hence, we suggest the approach based on standard normal *cdf* in the empirical example as long as there is no strong evidence of the processes being generated by a uniform distributed innovation. We also report the results of estimation when the true parameter is close to unity. As expected, it can be more difficult to discriminate the location of the root from being outside or inside the unit circle since they are similar in magnitude.

In the second experiment we estimate an AR(2) process generated by centered χ_5 in-

[2] κ_3 is skewness and κ_4 is kurtosis.

novations. In this simulation, we look into the identification of causality and noncausality since it can have different combinations of causality and noncausality when the order of autoregression is greater than 1, such as mixed causal and noncausal processes discussed in Hecq et al. (2016). We pick $\theta_{0,1} = 0.4$ and $\theta_{0,2} = 0.8$ as two parameters of polynomial $(1-\theta_{0,1}L)(1-\theta_{0,2}L)$ in the causal case and we construct other three types of processes. The sample size varies from 100 to 200 and the replication is 100. The results are displayed in Table 2. In this simple setup we have four types of processes in terms of causality: two are mixed causal and noncausal AR(2), one is purely causal AR(2) and the other one is purely noncausal AR(2). So far we only discuss about identification of number of noncausal roots included in the process. Hence within the group of mixed causal-noncausal models, we are unable to discriminate them but it may be feasible after we develop confidence interval of the estimates. The first row in the Table reports the percentage of correct identification of number of noncausal roots in the process and the second row presents us the percentage of identification which detects the existence of noncausality. As shown in the result, we can observe that the proposed method works better in the processes with noncausality than causal case. In the purely casual case, it tends to work less satisfactorily in a relative small sample like T=100 but the proportion of correct identification is greatly improved when the sample size increases to 200.

5 Empirical application

In this section we apply our method to analyze 753 daily trading volume of Microsoft (MSFT) stock from 6/3/1996 to 5/26/1999. Breidt et al. (2001) has argued that a noncausal AR(1) model fits the data better than a causal AR(1) model by the diagnostics of residuals computed from both model respectively. We remove the heteroskedasticity and drift by taking the logarithm and demeaning the sequence (see fig.**??**). The ADF test indicates rejection of existence of unit root in the resulting sequence. The partial autocorrelation of the sequence suggests that AR model with order 1,2 or 3 are appropriate (see fig.**??**). Here we fit the data by an AR process of order 1 as the correlation at lags 2,3 are close to insignificant. The proposed method yields result

$$\hat{u}_t = Y_t - 1.7953Y_{t-1}$$

By factorizing the AR(1) polynomial, we find the noncausal root in this sequence. The stationary solution would be $Y_t = (1.7953)^{-1}Y_{t+1} + u_t$. In another word, the investment in this stock is more influenced by the forward-looking behavior of agents. The clustering phenomenon in the process can be explained as a result of foresight of uncertainty of stock market held by investors. Intuitively, the perception of a more volatile market environment in the future would lead to more variant investment strategies of agents of different types. Instead of fitting the date with AR-ARCH model, with noncausal AR(1) model we can avoid estimating more parameters. It also matches the result by Breidt et al. (2001) that noncausal processes can mimic ARCH-type behavior. Here we plot autocorrelation function (ACF) of squared value of residuals $\{\tilde{u}_t^2\}$ and $\{\hat{u}_t^2\}$ from causal model using Gaussian MLE

($\tilde{u}_t = Y_t - 0.5854Y_{t-1}$) and noncausal model using our approach respectively. The upper part of the Figure.6.2 displays a clear correlation in the residuals from causal AR(1) model at first lag, indicating ARCH would be an appropriate alternative to characterize the residuals as we observe significant correlation of the residuals at the first lag both in absolute and squared value. The lower part of the Figure.6.2 shows residuals generated from noncausal model does not show evident sign of correlation on variance of the residuals. Apart from volatility clustering behaviour, noncausal autoregression can also characterize the dependence between different percentiles that conventional GARCH models cannot explain. These nonlinear dynamics observed in Microsoft stock data may be linked to "informational heterogeneity" of investors in the financial market.

5.1 Another application on bubbles

6 Discussion: Measure of dependence under martingale difference sequence

The assumption on the innovation used to generate data process can be relaxed to martingale difference sequence(mds), which is more commonly observed in empirical data. It can broaden the class of time series including conditional heteroskedasticity(ARCH).
Denote the σ-field generated by the past sequence of u_t by $I_{t-1} = \sigma(u_{t-1}, u_{t-2}, \dots)$. By the definition of mds, we obtain

$$\mathbb{E}(u_t | I_{t-1}) = \mathbb{E}(u_t | \sigma(u_{t-1}, u_{t-2}, \dots)) = 0$$

By theorem in Bierens (1982), this infinite dimensional conditional restriction can be expressed equivalently as

$$\mathbb{E}\left(u_t e^{ixu_{t-j}}\right) = 0 \text{ for } j \geq 1 \quad \forall x \in \mathbb{R}$$

The exponential function can be replaced by indicator function $I(\cdot)$ or any other parametric family considered in Escanciano (2006). Here we choose indicator function to be consistent with the notation introduced throughout this paper and also its simplicity in computation. We define the following measure of dependence

$$\gamma_{\theta,j}(x) = \mathbb{E}(u_t(\theta) I(u_{t-j}(\theta) \leq x)) \text{ for } j \geq 1, \forall x \in \mathbb{R}$$

and $\gamma_{\theta,j}(x) = \gamma_{\theta,|j|}(x)$ for $j < 0$. The corresponding spectral density and distribution based on this measure are

$$\begin{aligned} d_\theta(x;\omega) =& \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j(x) e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \\ D_\theta(x;\lambda) =& \gamma_{\theta,0}(x)\lambda + 2\sum_{j=1}^{\infty} \gamma_{\theta,j}(x) \frac{\sin j\pi\lambda}{j\pi} \quad \lambda \in [0,1]. \end{aligned}$$

Following the same approach in *iid* case, the population loss function is constructed by a L_2 distance of the generalized *cdf* in the unrestricted case and the conjectured one in the

restricted case

$$\begin{aligned}\mathcal{Q}_0^{mds}(\theta) :&= L^2\left(D_\theta(x;\lambda), D_\theta^{mds}(x;\lambda)\right)\\ &= 2\int_{\mathbb{R}^2}\sum_{j=1}^{\infty}\gamma_{\theta,j}^2(x)\frac{1}{(j\pi)^2}dW(x)\end{aligned}$$

The study on the identification and estimation of the model is left to further research.

Table 1: Comparison of estimates using empirical *cdf* and standard normal *cdf* under different distributions of innovations: study in AR(1) case

			W:	empirical *cdf*				standard normal *cdf*			
u_t	T		θ_0:	0.4	0.4^{-1}	0.9	0.9^{-1}	0.4	0.4^{-1}	0.9	0.9^{-1}
$U_{[-5,5]}$	100	PCI		67.00%	84.00%	53.00%	54.00%	62.00%	77.00%	45.00%	41.00%
		Bias		-0.0148	0.3112	-0.0254	0.0421	-0.0211	0.5532	-0.0405	0.0598
		MSE		0.0092	0.5532	0.0067	0.0146	0.0122	0.09017	0.0076	0.0177
	200	PCI		87.00%	86.00%	64.00%	63.00%	82.00%	78.00%	50.00%	54.00%
		Bias		-0.0101	0.0507	-0.0130	0.0155	-0.0169	0.0964	-0.0203	0.0206
		MSE		0.0046	0.2240	0.0031	0.0047	0.0054	0.3443	0.0037	0.0049
t_3	100	PCI		60.00%	69.00%	69.00%	57.00%	81.00%	86.00%	75.00%	72.00%
		Bias		-0.0089	0.1421	-0.0079	0.0170	-0.0145	0.0792	-0.0162	0.0257
		MSE		0.0131	0.4052	0.0036	0.0079	0.0102	0.2954	0.0035	0.0075
	200	PCI		76.00%	75.00%	63.00%	59.00%	90.00%	84.00%	71.00%	76.00%
		Bias		-0.0105	0.1740	0.0121	0.0053	-0.0087	0.1285	0.0023	0.0162
		MSE		0.0037	0.2849	0.0026	0.0043	0.0034	0.2090	0.0024	0.0040
$\chi_5 - 5$	100	PCI		94.00%	94.00%	71.00%	69.00%	94.00%	93.00%	78.00%	73.00%
		Bias		-0.0035	0.1921	-0.0372	0.0550	-0.0062	0.2085	-0.0377	0.0699
		MSE		0.0078	0.6751	0.0062	0.0149	0.0071	0.6660	0.0063	0.0181
	200	PCI		99.00%	98.00%	80.00%	76.00%	99.00%	99.00%	81.00%	77.00%
		Bias		-0.0101	0.1082	0.0060	0.0134	-0.0085	0.1674	0.0034	0.0025
		MSE		0.0045	0.2281	0.0032	0.0041	0.0045	0.3838	0.0031	0.0043

PCI: percentage of correct root identification using our method
Bias: computed by $\hat{\mathbb{E}}_T(\hat{\theta}_T) - \theta_0$
MSE (mean squared error): computed by sum of $\hat{\mathbb{V}\text{ar}}(\hat{\theta}_T)$and Bias^2

Table 2: Estimates of AR(2) generated by innovations following $\chi_5 - 5$

	$\chi^2(5) - 5$							
	T= 100				T= 200			
θ_0 :	$(0.4, 0.8)$	$(0.4^{-1}, 0.8^{-1})$	$(0.4^{-1}, 0.8)$	$(0.4, 0.8^{-1})$	$(0.4, 0.8)$	$(0.4^{-1}, 0.8^{-1})$	$(0.4^{-1}, 0.8)$	$(0.4, 0.8^{-1})$
PCI	59.00%	81.00%	84.00%	85.00%	80.00%	94.00%	95.00%	90.00%
PN	41.00%	95.00%	96.00%	85.00%	20.00%	100.00%	98.00%	99.00%

PCI: percentage of correct root identification including the number of roots lying inside unit circle
PN: percentage of detecting existence of noncausality in the process. *i.e.* There is at least one root lying inside unit circle.

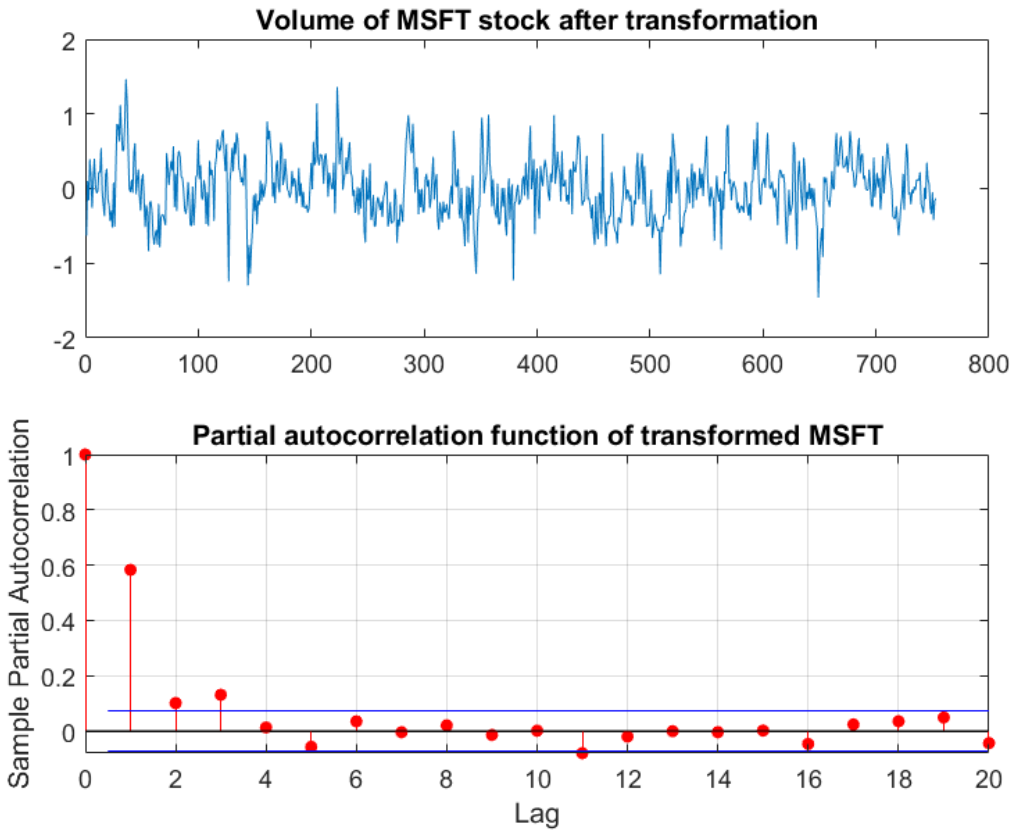


Figure 6.1: Microsoft daily trading volume from 6/3/1993 to 5/26/1999

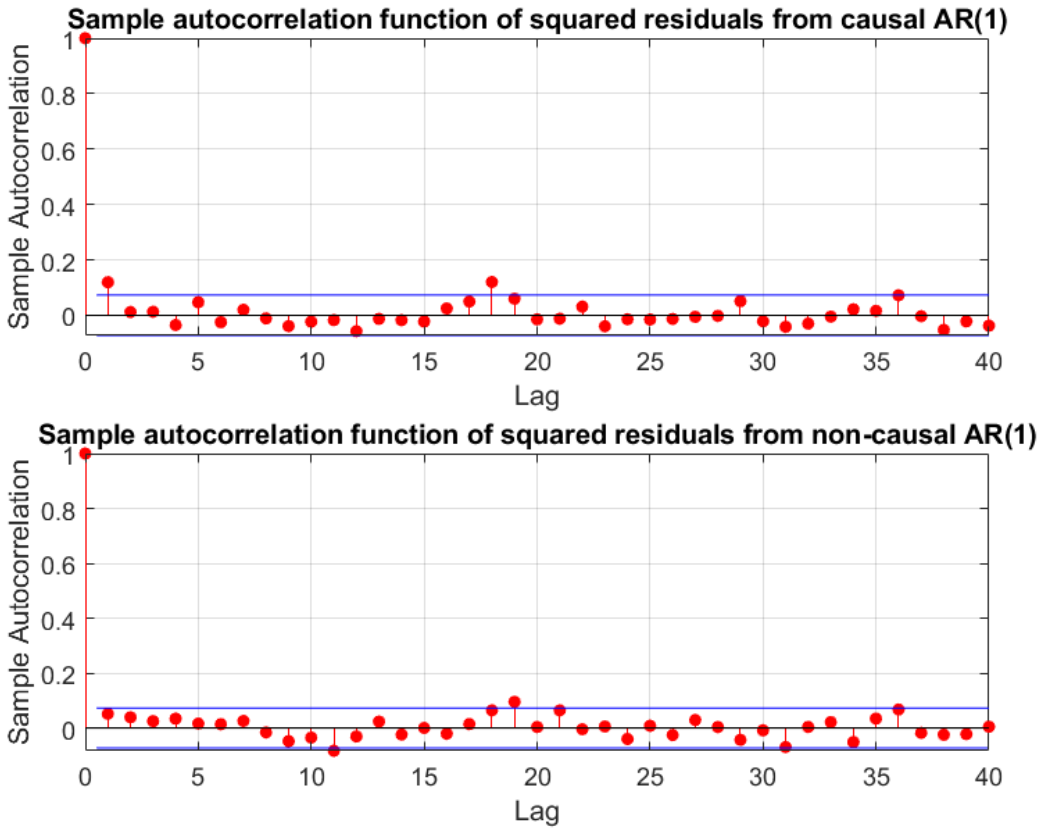


Figure 6.2: Diagnostics of residuals from both causal and non-causal models: a comparison in ACF of residuals in squared values

References

Alessi, L., Barigozzi, M. and Capasso, M. (2011) Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, **79**, 16–47.

Andrews, B., Davis, R. A. and Breidt, F. J. (2006) Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, **97**, 1638–1659.

Andrews, B., Davis, R. A., Breidt, F. J. et al. (2007) Rank-based estimation for all-pass time series models. *The Annals of Statistics*, **35**, 844–869.

Bierens, H. J. (1982) Consistent model specification tests. *Journal of Econometrics*, **20**, 105–134.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961) *Distribution free tests of independence based on the sample distribution function*. Sandia Corporation.

Boldin, M. V., Simonova, G. I. and Ti_urin, I. N. (1997) *Sign-based methods in linear statistical models*, vol. 162. American Mathematical Soc.

Breid, F. J., Davis, R. A., Lh, K.-S. and Rosenblatt, M. (1991) Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, **36**, 175–198.

Breidt, F. J., Davis, R. A., Trindade, A. A. et al. (2001) Least absolute deviation estimation for all-pass time series models. *The Annals of Statistics*, **29**, 919–946.

Brown, B. M. (1971) Martingale central limit theorems. *The Annals of Mathematical Statistics*, 59–66.

Cabello, M. A. (2021) Identification and estimation of linear structural varma models using pairwise dependence measures. *Working paper*.

Delgado, M. A. (1996) Testing serial independence using the sample distribution function. *Journal of Time Series Analysis*, **17**, 271–285.

Doukhan, P. and Louhichi, S. (1999) A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications*, **84**, 313–342.

Du, Z. and Escanciano, J. C. (2015) A nonparametric distribution-free test for serial independence of errors. *Econometric Reviews*, **34**, 1011–1034.

Escanciano, J. C. (2006) Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, **101**, 531–541.

Gaeta, M., Briolle, F. and Esparcieux, P. (1997) Blind separation of sources applied to convolutive mixtures in shallow water. In *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, 340–343. IEEE.

Gourieroux, C. and Jasiak, J. (2018) Misspecification of noncausal order in autoregressive processes. *Journal of Econometrics*, **205**, 226–248.

Gouriéroux, C. and Zakoïan, J.-M. (2017) Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 737–756.

Hagemann, A. (2011) Robust spectral analysis. *arXiv preprint arXiv:1111.1965*.

Hansen, L. P. and Sargent, T. J. (1991) Two difficulties in interpreting vector autoregressions. *Rational expectations econometrics*, **1**, 77–119.

Hecq, A., Issler, J. V. and Telg, S. (2020) Mixed causal–noncausal autoregressions with exogenous regressors. *Journal of Applied Econometrics*, **35**, 328–343.

Hecq, A., Lieb, L. and Telg, S. (2016) Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 307–331.

Hecq, A. and Voisin, E. (2020) Forecasting bubbles with mixed causal-noncausal autoregressive models. *Econometrics and Statistics*.

Hoeffding, W. (1948) A non-parametric test of independence. *The annals of mathematical statistics*, 546–557.

Hong, Y. (1998) Testing for pairwise serial independence via the empirical distribution function. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60**, 429–453.

— (2000) Generalized spectral tests for serial dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**, 557–574.

Kley, T., Volgushev, S., Dette, H., Hallin, M. et al. (2016) Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli*, **22**, 1770–1807.

Lanne, M. and Luoto, J. (2013) Autoregression-based estimation of the new keynesian phillips curve. *Journal of Economic Dynamics and Control*, **37**, 561–570.

Lee, J. and Rao, S. S. (2011) The quantile spectral density and comparison based tests for nonlinear time series. *arXiv preprint arXiv:1112.2759*.

Leeper, E. M., Walker, T. B. and Yang, S.-C. S. (2013) Fiscal foresight and information flows. *Econometrica*, **81**, 1115–1145.

Lippi, M. and Reichlin, L. (1993) The dynamic effects of aggregate demand and supply disturbances: Comment. *The American Economic Review*, **83**, 644–652.

— (1994) Var analysis, nonfundamental representations, blaschke matrices. *Journal of Econometrics*, **63**, 307–325.

Lof, M. and Nyberg, H. (2017) Noncausality and the commodity currency hypothesis. *Energy Economics*, **65**, 424–433.

Newey, W. K. (1991) Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 1161–1167.

Newey, W. K. and McFadden, D. (1994) Large sample estimation and hypothesis testing. *Handbook of econometrics*, **4**, 2111–2245.

Skaug, H. J. and Tjøstheim, D. (1993) A nonparametric test of serial independence based on the empirical distribution function. *Biometrika*, **80**, 591–602.

Stinchcombe, M. B. and White, H. (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, **14**, 295–325.

Tekalp, A., Kaufman, H. and Woods, J. (1986) Identification of image and blur parameters for the restoration of noncausal blurs. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **34**, 963–972.

Velasco, C. (2021) Estimation of time series using residuals dependence measures. *Working paper*.

Velasco, C. and Lobato, I. N. (2018) Frequency domain minimum distance inference for possibly noninvertible and noncausal arma models. *The Annals of Statistics*, **46**, 555–579.

7 Appendix: Proofs

7.1 Proof to Lemma 3.1

Define

$$F_\theta(x) := \mathbb{E}\left(\hat{F}_{\theta,j}(x,\infty)\right) = \mathbb{E}\left(I\left(u_t(\theta) \le x\right)\right) = \mathbb{E}\left(F\left(x - u_t^{(0)}(\theta)\right)\right)$$

where $\phi_0(\theta) = 1$ for all $\theta \in \Theta$,

$$u_t(\theta) = \sum_{j=-\infty}^{\infty} \phi_j(\theta) u_{t-j}, \quad u_t^{(0)}(\theta) = u_t(\theta) - u_t = \sum_{j \neq 0} \phi_j(\theta) u_{t-j}.$$

For any $j > 0$,

$$\begin{aligned}
\mathbb{E}\left(\hat{F}_{\theta,j}(x,y)\right) &= \mathbb{E}\left(I\left(u_t(\theta) \le x\right) I\left(u_{t-j}(\theta) \le y\right)\right) \\
&= \mathbb{E}\left(I\left(u_t \le x - u_t^{(0)}(\theta)\right) I\left(\phi_{-j}(\theta) u_t \le y - u_{t-j}^{(-j)}(\theta)\right)\right) \\
&= \mathbb{E}\left(I\left(u_t + \phi_j u_{t-j} \le x - u_t^{(0,j)}(\theta)\right) I\left(\phi_{-j}(\theta) u_t + u_{t-j} \le y - u_{t-j}^{(0,-j)}(\theta)\right)\right) \\
&= \mathbb{E}\left(\int I\left(z + \phi_j(\theta) w \le x - u_t^{(0,j)}(\theta)\right) I\left(\phi_{-j}(\theta) z + w \le y - u_{t-j}^{(0,-j)(\theta)}\right) f(z) f(w) dz dw\right) \\
&= \mathbb{E}\left(\int I\left(u \le x - u_t^{(0,j)}(\theta)\right) I\left(v \le y - u_{t-j}^{(0,-j)}(\theta)\right) f_{u,v}^{(j)}(u,v) du dv\right) \\
&= \mathbb{E}\left(F_{u,v}^{(j)}\left(x - u_t^{(0,j)}(\theta), y - u_{t-j}^{(0,-j)}(\theta)\right)\right)
\end{aligned}$$

where the last second equality comes from the change of variables

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} z + \phi_j(\theta) w \\ \phi_{-j}(\theta) z + w \end{pmatrix}, \quad \begin{pmatrix} z \\ w \end{pmatrix} = \frac{1}{1 - \phi_j(\theta)\phi_{-j}(\theta)} \begin{pmatrix} u - \phi_j(\theta) v \\ v - \phi_{-j}(\theta) u \end{pmatrix}.$$

The Jacobian equals

$$\left|\frac{d(z,w)}{d(u,v)}\right| = \frac{1}{\left(1 - \phi_j(\theta)\phi_{-j}(\theta)\right)^2} \begin{vmatrix} 1 & -\phi_j(\theta) \\ -\phi_{-j}(\theta) & 1 \end{vmatrix} = \frac{\left|1 - \phi_j(\theta)\phi_{-j}(\theta)\right|}{\left(1 - \phi_j(\theta)\phi_{-j}(\theta)\right)^2} = \frac{1}{1 - \phi_j(\theta)\phi_{-j}(\theta)} > 0.$$

For sufficiently large j, $|\phi_k(\theta)\phi_{-k}(\theta)| < 1$ for all $k \ge j$ and for relatively small j, there is always a compact set of $\theta \in \Theta$ such that $|\phi_j(\theta)\phi_{-j}(\theta)| < 1$ since $\phi_j(\theta)$ is zero at the true parameter value for $j \neq 0$.

By applying Mean Value Theorem on f, we can get

$$\begin{aligned}
f_{u,v}^{(j)}(u,v) &= \frac{1}{1-\phi_j(\theta)\phi_{-j}(\theta)} f\left(\frac{u-\phi_j(\theta)v}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) f\left(\frac{v-\phi_{-j}(\theta)u}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) \\
&= \left(1+O\left(\phi_j(\theta)\phi_{-j}(\theta)\right)\right) f\left(\frac{u-\phi_j(\theta)v}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) f\left(\frac{v-\phi_{-j}(\theta)u}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) \\
&= f(u)f(v)\left(1+O\left(\phi_j(\theta)\phi_{-j}(\theta)\right)\right) \\
&\quad + O\left(u\phi_{-j}(\theta) f\left(\frac{u-\phi_j(\theta)v}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) \dot{f}\left(\frac{v-\phi_{-j}(\theta)u\eta}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right)\right) \\
&\quad + O\left(v\phi_j(\theta) f\left(\frac{v-\phi_{-j}u}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) \dot{f}\left(\frac{u-\phi_j(\theta)v\eta}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right)\right)
\end{aligned}$$

for $\eta \in (0,1)$. Under Assumption 4 with $a=1$,

$$F_{u,v}^{(j)}(x,y) = F(x)F(y) + O\left(\phi_j(\theta)+\phi_{-j}(\theta)\right)$$

uniformly in $(x,y) \in \mathbb{R}^2$.
Before proceeding to next step, we define trucated versions of $u_{t,m}^{(0,j)}(\theta)$ and $u_{t-j,m}^{(0)}(\theta)$ for some $j/2 \le m \le j$,

$$\begin{aligned}
u_t^{(0,j)}(\theta) =& u_{t,m}^{(0,j)}(\theta) + \varepsilon_{t,m}^{(0,j)}(\theta) \\
u_{t-j}^{(0)}(\theta) =& u_{t-j,m}^{(0)}(\theta) + \varepsilon_{t-j,m}^{(0)}(\theta)
\end{aligned}$$

where

$$u_{t,m}^{(0,j)}(\theta) = u_{t,m}^{(0)}(\theta) = \sum_{k=-m}^{m} \phi_k(\theta) u_{t-k}$$

and for each $\theta \in \Theta$ and $\mu_0 > 1$,

$$\mathbb{E}\left|\varepsilon_{t,m}^{(0,j)}(\theta)\right| \le \mathbb{E}\left|u_t\right| \left(\sum_{k=m+1}^{\infty} |\phi_k(\theta)| + \sum_{k=-\infty}^{-m-1} |\phi_k(\theta)|\right) \le C(m+1)^{1-\mu_0} < Cj^{1-\mu_0}$$

Using this representation,

$$\begin{aligned}
&\mathbb{E}\left(F_{u,v}^{(j)}\left(x-u_t^{(0,j)}(\theta), y-u_{t-j}^{(0,-j)}(\theta)\right)\right) \\
=& \mathbb{E}\left(F\left(x-u_t^{(0,j)}(\theta)\right) F\left(y-u_{t-j}^{(0,-j)}(\theta)\right) + O\left(\phi_j(\theta)+\phi_{-j}(\theta)\right)\right) \\
=& \mathbb{E}\left(F\left(x-u_{t,m}^{(0,j)}(\theta)\right) F\left(y-u_{t-j,m}^{(0,-j)}(\theta)\right)\right) + \eta_1 \mathbb{E}\left|\varepsilon_{t,m}^{(0,j)}(\theta) F\left(y-u_{t-j}^{(0,-j)}(\theta)\right)\right| \\
&+ \eta_2 \mathbb{E}\left|\varepsilon_{t-j,m}^{(0,-j)}(\theta) F\left(x-u_t^{(0,j)}(\theta)\right)\right| + O\left(\phi_j(\theta)+\phi_{-j}(\theta)\right) \qquad \text{where } \eta_1, \eta_2 \in (0,1) \\
=& \mathbb{E}\left(F\left(x-u_{t,m}^{(0,j)}(\theta)\right) F\left(y-u_{t-j,m}^{(0,-j)}(\theta)\right)\right) + O(j^{1-\mu_0}) \\
=& \mathbb{E}\left(F\left(x-u_{t,m}^{(0,j)}(\theta)\right)\right) \mathbb{E}\left(F\left(y-u_{t-j,m}^{(0,-j)}(\theta)\right)\right) + O(j^{1-\mu_0}) \\
=& \mathbb{E}\left(F\left(x-u_t^{(0,j)}(\theta)\right)\right) \mathbb{E}\left(F\left(y-u_{t-j}^{(0,-j)}(\theta)\right)\right) + O(j^{1-\mu_0}) \\
=& F_\theta(x)F_\theta(y) + O(j^{1-\mu_0})
\end{aligned}$$

The second equality follows immediately from the application of MVT and the fourth one comes from independence after designed truncation on the residuals. The last second equality in induced from redoing truncation and filling the gaps at lag j and leads $-j$. Therefore, we conclude uniformly in θ and $(x, y) \in \mathbb{R}^2$

$$|\sigma_{\theta,j}(x, y)| \leq O(j^{1-\mu_0})$$

7.2 Proof to Lemma 3.2

To analyze the varaince of $\hat{\sigma}_{\theta,j}(x, y)$, we start with the joint moment of 4 indicators involved in the computation in a similar manner with Lemma 3.1.
Assume $t > t'$, $j > 0$ and $t - t' > 2j$, so $t - t' - j > (t - t')/2$,

$$\begin{aligned}
&\mathbb{E}\left(I\left(u_{t'-j}(\theta) \leq y\right) I\left(u_{t'}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right) I\left(u_t(\theta) \leq x\right)\right) \\
=&\mathbb{E}\left(\int \begin{pmatrix} I\left(z_1 + \phi_{-j}(\theta)z_2 + \phi_{t'-t}(\theta)z_3 + \phi_{t'-t-j}(\theta)z_4 \leq y - u_{t'-j}^{(0,-j,t'-t-j,t'-t)}(\theta)\right) \\ I\left(\phi_j(\theta)z_1 + z_2 + \phi_{t'-t+j}(\theta)z_3 + \phi_{t'-t}(\theta)z_4 \leq x - u_{t'}^{(0,j,t'-t+j,t'-t)}(\theta)\right) \\ I\left(\phi_{t-t'}(\theta)z_1 + \phi_{t-t'-j}(\theta)z_2 + z_3 + \phi_{-j}(\theta)z_4 \leq y - u_{t-j}^{(0,-j,t-t'-j,t-t')}(\theta)\right) \\ I\left(\phi_{t-t'+j}(\theta)z_1 + \phi_{t-t'}(\theta)z_2 + \phi_j(\theta)z_3 + z_4 \leq x - u_t^{(0,j,t-t'+j,t-t')}(\theta)\right) \end{pmatrix} \begin{matrix} f(z_1)f(z_2)f(z_3)f(z_4) \\ dz_1dz_2dz_3dz_4 \end{matrix}\right) \\
=&\mathbb{E}\left(\int \begin{pmatrix} I\left(u_1 \leq y - u_{t'-j}^{(0,-j,t'-t-j,t'-t)}(\theta)\right) I\left(u_2 \leq x - u_t^{(0,j,t'-t+j,t'-t)}(\theta)\right) \\ I\left(u_3 \leq y - u_{t-j}^{(0,-j,t-t'-j,t-t')}(\theta)\right) I\left(u_4 \leq x - u_t^{(0,j,t-t'+j,t-t')}(\theta)\right) \end{pmatrix} f_u(u)du\right) \\
=&\mathbb{E}\left(F_u\left(y - u_{t'-j}^{(0,-j,t'-t-j,t'-t)}(\theta), x - u_t^{(0,j,t'-t+j,t'-t)}(\theta), y - u_{t-j}^{(0,-j,t-t'-j,t-t')}(\theta), x - u_t^{(0,j,t-t'+j,t-t')}(\theta)\right)\right)
\end{aligned}$$

where

$$\begin{aligned}
u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} &= \begin{pmatrix} z_1 + \phi_{-j}(\theta)z_2 + \phi_{t'-t}(\theta)z_3 + \phi_{t'-t-j}(\theta)z_4 \\ \phi_j(\theta)z_1 + z_2 + \phi_{t'-t+j}(\theta)z_3 + \phi_{t'-t}(\theta)z_4 \\ \phi_{t-t'}(\theta)z_1 + \phi_{t-t'-j}(\theta)z_2 + z_3 + \phi_{-j}(\theta)z_4 \\ \phi_{t-t'+j}(\theta)z_1 + \phi_{t-t'}(\theta)z_2 + \phi_j(\theta)z_3 + z_4 \end{pmatrix} \\
&= \begin{pmatrix} 1 & \phi_{-j}(\theta) & \phi_{t'-t}(\theta) & \phi_{t'-t-j}(\theta) \\ \phi_j(\theta) & 1 & \phi_{t'-t+j}(\theta) & \phi_{t'-t}(\theta) \\ \phi_{t-t'}(\theta) & \phi_{t-t'-j}(\theta) & 1 & \phi_{-j}(\theta) \\ \phi_{t+t'+j}(\theta) & \phi_{t-t'}(\theta) & \phi_j(\theta) & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}
\end{aligned}$$

and

$$
\begin{aligned}
&\begin{pmatrix} 1 & \phi_{-j}(\theta) & \phi_{t'-t}(\theta) & \phi_{t'-t-j}(\theta) \\ \phi_j(\theta) & 1 & \phi_{t'-t+j}(\theta) & \phi_{t'-t}(\theta) \\ \phi_{t-t'}(\theta) & \phi_{t-t'-j}(\theta) & 1 & \phi_{-j}(\theta) \\ \phi_{t+t'+j}(\theta) & \phi_{t-t'}(\theta) & \phi_j(\theta) & 1 \end{pmatrix}^{-1} \\
=&\begin{pmatrix} \begin{pmatrix} 1 & \phi_{-j}(\theta) \\ \phi_j(\theta) & 1 \end{pmatrix}^{-1} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 & \phi_{-j}(\theta) \\ \phi_j(\theta) & 1 \end{pmatrix}^{-1} \end{pmatrix} (1 + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)) \\
=&\frac{1}{1-\phi_j(\theta)\phi_{-j}(\theta)} \begin{pmatrix} 1 & -\phi_j(\theta) & 0 & 0 \\ -\phi_{-j}(\theta) & 1 & 0 & 0 \\ 0 & 0 & 1 & -\phi_j(\theta) \\ 0 & 0 & -\phi_{-j}(\theta) & 1 \end{pmatrix} (1 + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)).
\end{aligned}
$$

Hence the Jacobian is

$$
\frac{1}{\left(1-\phi_{t-t'}(\theta)\phi_{t'-t}(\theta)\right)^2} \left(1 + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)\right)
$$

and by applying MVT to the argument of each f in the linear mappings $z_a = z_a(u), a = 1, 2, 3, 4$, and only keeping components involved with $1, \phi_j(\theta), \phi_{-j}(\theta)$,

$$
\begin{aligned}
&f_u(u) \\
=&\frac{1}{\left(1-\phi_j(\theta)\phi_{-j}(\theta)\right)^2} f\left(\frac{u_1-\phi_j u_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) f\left(\frac{-\phi_j u_1+u_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) f\left(\frac{u_3-\phi_j u_4}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) f\left(\frac{-\phi_j u_3+u_4}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) \\
&\times \left(1 + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)\right) + g(u) O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right) \\
=&f_{12}\left(u_1, u_2\right) f_{34}\left(u_3, u_4\right) \left(1 + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)\right) + g(u) O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)
\end{aligned}
$$

where

$$
f_{12}\left(u_1, u_2\right) = \frac{1}{1-\phi_j(\theta)\phi_{-j}(\theta)} f\left(\frac{u_1-\phi_j u_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) f\left(\frac{-\phi_j u_1+u_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right) \left(1 + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)\right),
$$

$f_{12} = f_{34}$ due to stationarity and $g(u)$ is integrable in $u \in \mathbb{R}^4$ under Assumption 4 with $a = 2$.

By integrating the joint density over u, we get

$$
F_u\left(x\right) = F_{12}\left(x_1, x_2\right) F_{34}\left(x_3, x_4\right) + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right)
$$

uniformly in $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$.
Therefore,

$$
\begin{aligned}
&\mathbb{E}\left(I\left(u_{t'-j}(\theta) \le y\right) I\left(u_{t'}(\theta) \le x\right) I\left(u_{t-j}(\theta) \le y\right) I\left(u_t(\theta) \le x\right)\right) \\
=&\mathbb{E}\left(\begin{array}{c} F_{12}\left(x - u_{t'}^{(0,j,t'-t,t'-t+j)}(\theta), y - u_{t'-j}^{(0,-j,t'-t,t'-t-j)}(\theta)\right) \\ F_{34}\left(x - u_{t}^{(0,j,t-t',t-t'+j)}(\theta), y - u_{t-j}^{(0,-j,t-t',t-t'-j)}(\theta)\right) \end{array}\right) + O\left(\phi_{t-t'}(\theta) + \phi_{t'-t}(\theta)\right) \\
=&\mathbb{E}\left(F_{12}\left(x - u_t^{(0,j)}(\theta), y - u_{t-j}^{(0,-j)}(\theta)\right)\right)^2 + O\left(\left|t-t'\right|^{1-\mu_0}\right) \\
=&F_{\theta,j}(x,y)^2 + O\left(\left|t-t'\right|^{1-\mu_0}\right)
\end{aligned}
$$

The second last equality comes from truncation, independence and refilling the truncation, the same tricks used in the previous proof.
Then,

$$
\mathbb{V}\mathrm{ar}\left(\hat{F}_{\theta,j}(x,y)\right) = \frac{1}{(T-j)^2} \sum_{t=1+j}^{T} \sum_{t'+j}^{T} \mathbb{C}\mathrm{ov}\left(I\left(u_{t-j}(\theta) \le y\right) I\left(u_t(\theta) \le x\right), I\left(u_{t'-j}(\theta) \le y\right) I\left(u_{t'}(\theta) \le x\right)\right)
$$

where the covariance equals

$$
\begin{aligned}
&\mathbb{E}\left(I\left(u_{t'}(\theta) \le x\right) I\left(u_{t'-j}(\theta) \le y\right) I\left(u_t(\theta) \le x\right) I\left(u_{t-j}(\theta) \le y\right)\right) \\
&- \mathbb{E}\left(I\left(u_{t'}(\theta) \le x\right) I\left(u_{t'-j}(\theta) \le y\right)\right) \mathbb{E}\left(I\left(u_t(\theta) \le x\right) I\left(u_{t-j}(\theta) \le y\right)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left|\hat{F}_{\theta,j}(x,y) - F_{\theta,j}(x,y)\right|^2 \\
\le& C\left(1 \wedge \frac{(T-j)j}{(T-j)^2}\right) + C\frac{1}{(T-j)^2} \sum_{t=j+1}^{T} \sum_{t':|t-t'|>2j} \left|t-t'\right|^{1-\mu_0} \\
\le& C\left(1 \wedge \frac{j}{(T-j)}\right) + C\frac{j^{2-\mu_0}}{T-j} + C\frac{\log T}{(T-j)^{\mu_0-1}}
\end{aligned}
$$

7.3 Proof to Theorem 3.3

First we need to show pointwise convergence of $\hat{\mathcal{Q}}_T(\theta)$ to $\mathcal{Q}_0(\theta)$ for each $\theta \in \Theta$,i.e.

$$
\hat{\mathcal{Q}}_T(\theta) - \mathcal{Q}_0(\theta) = o_p(1) \text{ for each } \theta \in \Theta.
$$

We first approximate population loss function by the

$$
\mathcal{Q}_T(\theta) = \sum_{j=1}^{T-1} \left(1 - j/T\right) \sigma_{\theta,j}^2(x,y)/(j\pi)^2 dW(x,y)
$$

$$
\begin{aligned}
\left|\mathcal{Q}_0(\theta) - \mathcal{Q}_T(\theta)\right| &= \left|\sum_{j=T}^{\infty} \int_{\mathbb{R}^2} \sigma_{\theta,j}^2(x,y)/(j\pi)^2 dW(x,y) + \sum_{j=1}^{T-1} \int_{\mathbb{R}^2} \frac{j}{T} \sigma_{\theta,j}^2(x,y)/(j\pi)^2 dW(x,y)\right| \\
&\le \left|\sum_{j=T}^{\infty} \int_{\Gamma} \sigma_{\theta,j}^2(x,y)/(j\pi)^2 dW(x,y)\right| + \left|\frac{1}{\pi^2} \sum_{j=1}^{T-1} \int_{\Gamma} \frac{1}{T} \sigma_{\theta,j}^2(x,y)/j dW(x,y)\right|
\end{aligned}
$$

Note that $|\sigma_{\theta,j}(x,y)|$ is uniformly bounded by 2 in $(\theta,x,y)\in\Theta\times\mathbb{R}^2$ for each $j=1,2,\dots T$. Hence

$$\begin{aligned}\sup_{\theta\in\Theta}|\mathcal{Q}_0(\theta)-\mathcal{Q}_T(\theta)| \leq& \frac{1}{\pi^2}\sum_{j=T}^{\infty}\int_{\mathbb{R}^2}\sup_{(\theta,x,y)\in\Theta\times\mathbb{R}^2}|\sigma_{\theta,j}(x,y)|^2/(j\pi)^2 dW(x,y)\\ &+\frac{1}{T\pi^2}\sum_{j=1}^{T-1}\int_{\mathbb{R}^2}\sup_{(\theta,x,y)\in\Theta\times\mathbb{R}^2}|\sigma_{\theta,j}(x,y)|^2/j dW(x,y)\\ \leq& C\sum_{j=T}^{\infty}\frac{1}{(j\pi)^2}+\frac{C}{T}\sum_{j=1}^{T-1}\frac{1}{j}\\ \leq& \frac{C}{T}+\frac{C\ln(T-1)}{T}=o(1)\end{aligned}$$

From the above statement, showing $\hat{\mathcal{Q}}_T(\theta)-\mathcal{Q}_0(\theta)=o_p(1)$ is equivalent to showing

$$|\hat{\mathcal{Q}}_T(\theta)-\mathcal{Q}_T(\theta)|=o_p(1) \tag{7.1}$$

First we define $z_t(\theta,x):=I\left(u_t(\theta)\leq x\right)-F_\theta(x)$ and

$$\bar{\sigma}_{\theta,j}(x,y)=\frac{1}{T-j}\sum_{t=j+1}^{T}z_t\left(\theta,x\right)z_{t-j}\left(\theta,y\right)$$

Since

$$\begin{aligned}\hat{\sigma}^2_{\theta,j}(x,y)-\sigma^2_{\theta,j}(x,y)=&\hat{\sigma}^2_{\theta,j}(x,y)-\bar{\sigma}^2_{\theta,j}(x,y)+\bar{\sigma}^2_{\theta,j}(x,y)-\sigma^2_{\theta,j}(x,y)\\ \hat{\sigma}^2_{\theta,j}(x,y)-\bar{\sigma}^2_{\theta,j}(x,y)=&\left|\hat{\sigma}_{\theta,j}(x,y)-\bar{\sigma}_{\theta,j}(x,y)\right|^2+2\left(\hat{\sigma}_\theta(x,y)-\bar{\sigma}_{\theta,j}(x,y)\right)\bar{\sigma}_{\theta,j}(x,y)\end{aligned}$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}&\mathbb{E}\left|\left(\hat{\sigma}_{\theta,j}(x,y)-\bar{\sigma}_{\theta,j}(x,y)\right)\bar{\sigma}_{\theta,j}(x,y)\right|\\ \leq&\left\{\mathbb{E}\left|\left(\hat{\sigma}_{\theta,j}(x,y)-\bar{\sigma}_{\theta,j}(x,y)\right)\right|^2\mathbb{E}\left|\bar{\sigma}_{\theta,j}(x,y)\right|^2\right\}^{1/2}\\ \leq& C\left(T-j\right)^{-1}\end{aligned}$$

The last inequality comes from $|\bar{\sigma}_{\theta,j}(x,y)|\leq 2$ and

$$\begin{aligned}\left(T-j\right)^4\mathbb{E}\left|\hat{\sigma}_{\theta,j}(x,y)-\bar{\sigma}_{\theta,j}(x,y)\right|^2\leq&\left\{\mathbb{E}\left|\sum_{t=j+1}^{T}z_t(\theta,x)\right|^4\mathbb{E}\left|\sum_{t=j+1}^{T}z_{t-j}(\theta,y)\right|^4\right\}^{1/2}\\ \leq& C\left(T-j\right)^2\end{aligned}$$

The proof of above result uses the Marcinkiewicz-Zygmund inequality in Theorem 1 of Doukhan and Louhichi (1999) for a sequence of weakly dependent random variable centered at expectation.
Similarly,

$$\left|\bar{\sigma}^2_{\theta,j}(x,y)-\sigma^2_{\theta,j}(x,y)\right|=\left|\bar{\sigma}_{\theta,j}(x,y)-\sigma_{\theta,j}(x,y)\right|^2+2\left(\bar{\sigma}_{\theta,j}(x,y)-\sigma_{\theta,j}(x,y)\right)\sigma_{\theta,j}(x,y)$$

so that

$$
\begin{aligned}
&\mathbb{E}\left|\bar{\sigma}_{\theta,j}^2(x,y)-\sigma_{\theta,j}^2(x,y)\right| \\
\leq&\,\mathbb{E}\left|\bar{\sigma}_{\theta,j}(x,y)-\sigma_{\theta,j}(x,y)\right|^2+2\left\{\mathbb{E}\left|\bar{\sigma}_{\theta,j}(x,y)-\sigma_{\theta,j}(x,y)\right|^2\mathbb{E}\left|\sigma_{\theta,j}(x,y)\right|^2\right\}^{1/2} \\
\leq&\,C\left(1\wedge\frac{j}{T-j}\right)+C\frac{j^{2-\mu_0}}{T-j}+C\frac{\ln T}{(T-j)^{\mu_0-1}} \\
&+C\left(1\wedge\frac{j}{T-j}\right)^{1/2}+C\frac{j^{1-\mu_0/2}}{(T-j)^{1/2}}+C\frac{(\ln T)^{1/2}}{(T-j)^{\mu_0/2-1/2}}
\end{aligned}
$$

by Lemma 3.2. Hence

$$
\begin{aligned}
\mathbb{E}\left|\hat{\mathcal{Q}}_T(\theta)-\mathcal{Q}_T(\theta)\right| =&\,\mathbb{E}\left|\sum_{j=1}^{T-1}\int_{\mathbb{R}^2}\left(\hat{\sigma}_{\theta,j}^2(x,y)-\sigma_{\theta,j}^2(x,y)\right)(1-\frac{j}{T})\frac{1}{(j\pi)^2}dW(x,y)\right| \\
\leq&\,\frac{1}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\int_{\mathbb{R}^2}\left\{\mathbb{E}\left|\hat{\sigma}_{\theta,j}^2(x,y)-\bar{\sigma}_{\theta,j}^2(x,y)\right|+\mathbb{E}\left|\bar{\sigma}_{\theta,j}^2(x,y)-\sigma_{\theta,j}^2(x,y)\right|\right\}dW(x \\
\leq&\,\frac{1}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\int_{\mathbb{R}^2}\mathbb{E}\left|\bar{\sigma}_{\theta,j}^2(x,y)-\sigma_{\theta,j}^2(x,y)\right|dW(x,y) \\
&+\frac{1}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\int_{\mathbb{R}^2}\mathbb{E}\left|\hat{\sigma}_{\theta,j}(x,y)-\bar{\sigma}_{\theta,j}(x,y)\right|^2dW(x,y) \\
&+\frac{2}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\int_{\mathbb{R}^2}\mathbb{E}\left|\left(\hat{\sigma}_{\theta,j}(x,y)-\bar{\sigma}_{\theta,j}(x,y)\right)\bar{\sigma}_{\theta,j}(x,y)\right|dW(x,y) \\
=&\,A+B+C
\end{aligned}
$$

For some $\mu_0>1$,

$$
\begin{aligned}
A\leq&\frac{C}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\int_{\mathbb{R}^2}\left(\left(1\wedge\frac{j}{T-j}\right)+\frac{j^{2-\mu_0}}{T-j}+\frac{\ln T}{(T-j)^{\mu_0-1}}\right)dW(x,y) \\
&+\frac{C}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\int_{\mathbb{R}^2}\left(\left(1\wedge\frac{j}{T-j}\right)^{1/2}+\frac{j^{1-\mu_0/2}}{(T-j)^{1/2}}+\frac{(\ln T)^{1/2}}{(T-j)^{\mu_0/2-1/2}}\right)dW(x,y) \\
\leq&\frac{C}{T}\sum_{j=1}^{T-1}\frac{1}{j^2}\left(T-j\wedge j\right)+\frac{C}{T}\sum_{j=1}^{T-1}\left(j^{-\mu_0}+\frac{1}{j^2}(\ln T)\left(T-j\right)^{2-\mu_0}\right) \\
&+\frac{C}{T}\sum_{j=1}^{T-1}\frac{1}{j^2}\left((T-j)^{1/2}\wedge(j(T-j))^{1/2}\right)+\frac{C}{T}\sum_{j=1}^{T-1}\left(T-j\right)^{1/2}j^{-1-\mu_0/2}+\frac{1}{j^2}\left(\ln T\right)^{1/2}\left(T-j\right)^{1/2-\mu_0/2} \\
=&\,o(1)\text{ as }T\to\infty.
\end{aligned}
$$

$$
B\leq\frac{C}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\left(T-j\right)^{-1}\leq\frac{C}{T}\sum_{j=1}^{T-1}\frac{1}{j^2}\leq\frac{C}{T}=o(1).
$$

Similar for C,

$$
C\leq\frac{C}{\pi^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\left(T-j\right)^{-1}\leq\frac{C}{T}=o(1).
$$

The proof of pointwise convergence of $\hat{\mathcal{Q}}_T(\theta)$ to $\mathcal{Q}_T(\theta)$ is completed.
As the result of non-differentiability of the objective function, we need to show stochastic equicontinuity of $\hat{\mathcal{Q}}_T(\theta)$. For the definition of stochastic equicontinuity see chapter 36 section 2.7 in Newey and McFadden (1994).
By definition, equivalently we have to prove

$$\sup_{\bar{\theta}\in\Theta(\epsilon,\eta)} \left|\hat{\mathcal{Q}}_T(\bar{\theta}) - \hat{\mathcal{Q}}_T(\theta)\right| = \Delta\left(\epsilon,\eta\right) = o_p(1) \text{ as } T \geq T_0(\epsilon,\eta),$$

which can be decomposed into

$$\begin{aligned} \mathbb{E}\left|\hat{\mathcal{Q}}_T(\bar{\theta}) - \mathcal{Q}_T(\bar{\theta})\right| &= o(1) \\ \sup_{\theta\in\Theta(\epsilon,\eta)} \left|\mathcal{Q}_T(\bar{\theta}) - \mathcal{Q}_T(\theta)\right| &= o(1) \end{aligned}$$

The first part can be shown in the similar way as previous discussion and the second part immediately follows from the continuity of population function $\mathcal{Q}_T(\theta)$ in θ.
Theorem 2.1 of Newey (1991) confirms uniform convergence in probability of objective function in a compact set of parameter Θ. The fundamental consistency theorem for extremum estimators implies

$$\hat{\theta}_T \longrightarrow_p \theta_0 \text{ as } T \to \infty$$

7.4 Proof to Theorem 3.4

To obtain the asymptotic distribution of the proposed estimator, due to non-differentiability of the objective function, we first investigate the asymptotic behaviour of the estimator based on the smoothed cumulative distribution function.
In this part, we approximate empirical (joint) cumulative function of residuals u_t by

$$\tilde{F}_{\theta,j}(x,y;h) = \frac{1}{T-j}\sum_{t=j+1}^{T} \Lambda\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right),$$

where Λ is a continuous cdf with twice differentiable pdf λ and h is a smoothing parameter so as to make $\Lambda(\frac{z}{h})$ converge to $I\left(-z \leq 0\right)$ for any $|z| > 0$ as h goes to 0. Based on this approximation, the general correlation of residuals defined in the section 2 becomes

$$\tilde{\sigma}_{\theta,j}(x,y;h) = \tilde{F}_{\theta,j}(x,y;h) - \tilde{F}_{\theta,j}(x,\infty;h)\tilde{F}_{\theta,j}(\infty,y;h),$$

and its corresponding objective function

$$\tilde{Q}_T(\theta;h) = 2\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}\tilde{\sigma}^2_{\theta,j}(x,y;h)\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}dW(x,y)$$

The score can be computed by

$$\frac{\partial}{\partial\theta}\tilde{Q}_T(\theta;h) = 4\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\tilde{\sigma}_{\theta,j}(x,y;h)\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}dW(x,y),$$

where

$$
\begin{aligned}
&\frac{\partial \tilde{\sigma}_{\theta,j}(x,y;h)}{\partial \theta} \\
=&\frac{\partial \tilde{F}_{\theta,j}(x,y;h)}{\partial \theta} - \frac{\partial \tilde{F}_{\theta,j}(x,\infty;h)}{\partial \theta}\tilde{F}_{\theta,j}(\infty,y;h) - \tilde{F}_{\theta,j}(x,\infty;h)\frac{\partial \tilde{F}_{\theta,j}(\infty,y;h)}{\partial \theta} \\
=&\frac{1}{T-j}\sum_{t=1+j}^{T}\left\{\lambda\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\frac{1}{h}\left(-u_t^{(1)}(\theta)\right) + \Lambda\left(\frac{x-u_t(\theta)}{h}\right)\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\frac{1}{h}\left(-u_{t-j}^{(1)}(\theta\right.\right. \\
&+ \tilde{F}_{\theta,j}(\infty,y;h)\frac{1}{T-j}\sum_{t=1+j}^{T}\lambda\left(\frac{x-u_t(\theta)}{h}\right)\frac{1}{h}u_t^{(1)}(\theta) + \tilde{F}_{\theta,j}(x,\infty;h)\frac{1}{T-j}\sum_{t=1+j}^{T}\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\frac{1}{h}u_{t-j}^{(1)}(\theta),
\end{aligned}
$$

with

$$
u_t^{(1)}(\theta) = \sum_{j=-\infty}^{\infty}\phi_j^{(1)}(\theta)u_{t-j}
$$

By Taylor expansion of $\frac{\partial}{\partial\theta}\tilde{Q}_T(\theta;h)$ at $\theta = \tilde{\theta}_T^h$[3], we have

$$
0 = \frac{\partial}{\partial\theta}\tilde{Q}_T(\tilde{\theta}_T^h;h) = \frac{\partial}{\partial\theta}\tilde{Q}_T^h(\theta_0;h) + \frac{\partial^2}{\partial\theta\partial\theta'}\tilde{Q}_T(\bar{\theta}_T^h)\left(\tilde{\theta}_T^h - \theta_0\right), \text{ for some } \bar{\theta}_T^h \in (\theta_0, \tilde{\theta}_T^h)
$$

Given consistency of $\tilde{\theta}_T^h$, so as $\bar{\theta}_T$, we have

$$
T^{1/2}\left(\tilde{\theta}_T^h - \theta_0\right) = -\left(\frac{\partial^2}{\partial\theta\partial\theta'}\tilde{Q}_T(\theta_0) + o_p(1)\right)^{-1} T^{1/2}\frac{\partial}{\partial\theta}\tilde{Q}_T(\theta_0)
$$

First we would like to replace the score by

$$
\widehat{\frac{\partial}{\partial\theta}\tilde{Q}_T(\theta)} = 4\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\tilde{\sigma}_{\theta,j}\widehat{(x,y;h)}\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}dW(x,y)
$$

where

$$
\tilde{\sigma}_{\theta,j}\widehat{(x,y;h)} = \frac{1}{T-j}\sum_{t=1+j}^{T}\left(\Lambda\left(\frac{x-u_t(\theta)}{h}\right) - \varphi_\theta^h(x)\right)\left(\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right) - \varphi_\theta^h(y)\right)
$$

where

$$
\varphi_\theta^h(x) = \mathbb{E}\left(\Lambda\left(\frac{x-u(\theta)}{h}\right)\right)
$$

[3]The minimizer of $\tilde{Q}_T(\theta;h)$

$$
\begin{aligned}
&\mathbb{E}\left\|\frac{\partial}{\partial\theta}\tilde{Q}_T(\theta)-\widehat{\frac{\partial}{\partial\theta}\tilde{Q}_T(\theta)}\right\| \\
=&4\,\mathbb{E}\left\|\int_{\mathcal{R}^2}\sum_{t=1+j}^{T}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\left(\tilde{\sigma}_{\theta,j}(x,y;h)-\widehat{\tilde{\sigma}_{\theta,j}(x,y;h)}\right)\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}dW(x,y)\right\| \\
\leq&4\sum_{t=1+j}^{T}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\int_{\mathcal{R}^2}\mathbb{E}\left\|\left(\tilde{\sigma}_{\theta,j}(x,y;h)-\widehat{\tilde{\sigma}_{\theta,j}(x,y;h)}\right)\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}\right\|dW(x,y) \\
\leq&4\sum_{t=1+j}^{T}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\int_{\mathcal{R}^2}\left(\mathbb{E}\left|\left(\tilde{\sigma}_{\theta,j}(x,y;h)-\widehat{\tilde{\sigma}_{\theta,j}(x,y;h)}\right)\right|^2\mathbb{E}\left\|\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}\right\|^2\right)^{1/2}dW(x,y) \\
=&O(\frac{C}{T-j})
\end{aligned}
$$

as

$$
\mathbb{E}\left|\left(\tilde{\sigma}_{\theta,j}(x,y;h)-\widehat{\tilde{\sigma}_{\theta,j}(x,y;h)}\right)\right|^2\leq C(T-j)^{-2}
$$

and

$$
\mathbb{E}\left\|\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}\right\|^2\leq C
$$

The convergence of Hessian matrix evaluated at $\theta=\theta_0$ in probability is induced from the uniform boundedness of the third order derivative of the objective function and consistency of the estimator.

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta^2}\tilde{Q}_T(\theta) \\
=&4\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\left\{\left(\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}\right)\left(\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}\right)'+\left(\tilde{\sigma}_{\theta,j}(x,y;h)\frac{\partial^2\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta\partial\theta'}\right)\right\}dW(x,y)
\end{aligned}
$$

For simplicity, we would conduct analysis of uniform boundedness of the third order moments in single dimension.

$$
\begin{aligned}
&\frac{\partial^3}{\partial\theta^3}\tilde{Q}_T(\theta) \\
=&4\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\left\{3\left(\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}\right)\frac{\partial^2\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta^2}+\left(\tilde{\sigma}_{\theta,j}(x,y;h)\frac{\partial^3\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta^3}\right)\right\}dW(x,y)
\end{aligned}
$$

where

$$
\begin{aligned}
&\frac{\partial \tilde{\sigma}_{\theta,j}(x,y;h)}{\partial \theta} \\
=&\frac{\partial \tilde{F}_{\theta,j}(x,y;h)}{\partial \theta} - \frac{\partial \tilde{F}_{\theta,j}(x,\infty;h)}{\partial \theta}\tilde{F}_{\theta,j}(\infty,y;h) - \tilde{F}_{\theta,j}(x,\infty;h)\frac{\partial \tilde{F}_{\theta,j}(\infty,y;h)}{\partial \theta} \\
&\frac{\partial^2 \tilde{\sigma}_{\theta,j}(x,y;h)}{\partial \theta^2} \\
=&\frac{\partial^2 \tilde{F}_{\theta,j}(x,y;h)}{\partial \theta^2} - \frac{\partial^2 \tilde{F}_{\theta,j}(x,\infty;h)}{\partial \theta^2}\tilde{F}_{\theta,j}(\infty,y;h) - 2\frac{\partial \tilde{F}_{\theta,j}(x,\infty;h)}{\partial \theta}\frac{\partial \tilde{F}_{\theta,j}(\infty,y;h)}{\partial \theta} \\
&- \tilde{F}_{\theta,j}(x,\infty;h)\frac{\partial^2 \tilde{F}_{\theta,j}(\infty,y;h)}{\partial \theta^2} \\
&\frac{\partial^3 \tilde{\sigma}_{\theta,j}(x,y;h)}{\partial \theta^3} \\
=&\frac{\partial^3 \tilde{F}_{\theta,j}(x,y;h)}{\partial \theta^3} - \frac{\partial^3 \tilde{F}_{\theta,j}(x,\infty;h)}{\partial \theta^3}\tilde{F}_{\theta,j}(\infty,y;h) - 3\frac{\partial^2 \tilde{F}_{\theta,j}(x,\infty;h)}{\partial \theta^2}\frac{\partial \tilde{F}_{\theta,j}(\infty,y;h)}{\partial \theta} \\
&- 3\frac{\partial \tilde{F}_{\theta,j}(x,\infty;h)}{\partial \theta}\frac{\partial^2 \tilde{F}_{\theta,j}(\infty,y;h)}{\partial \theta^2} - \tilde{F}_{\theta,j}(x,\infty;h)\frac{\partial^3 \tilde{F}_{\theta,j}(\infty,y;h)}{\partial \theta^3}
\end{aligned}
$$

with

$$
\begin{aligned}
&\frac{\partial \tilde{F}_{\theta,j}(x,y;h)}{\partial \theta} \\
=&\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\lambda\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(-\frac{u_t^{(1)}(\theta)}{h}\right)+\Lambda\left(\frac{x-u_t(\theta)}{h}\right)\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{-u_{t-j}^{(1)}(\theta)}{h}\right)\right. \\
&\frac{\partial^2 \tilde{F}_{\theta,j}(x,y;h)}{\partial \theta^2} \\
=&\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\dot{\lambda}\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{u_t^{(1)}(\theta)}{h}\right)^2+\Lambda\left(\frac{x-u_t(\theta)}{h}\right)\dot{\lambda}\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{u_{t-j}^{(1)}(\theta)}{h}\right)^2\right. \\
&+\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\lambda\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{-u_t^{(2)}(\theta)}{h}\right)+\Lambda\left(\frac{x-u_t(\theta)}{h}\right)\lambda\left(\frac{y-u_{t-j}}{h}\right)\left(\frac{-u_{t-j}^{(2)}(\theta)}{h}\right)\right. \\
&+2\frac{1}{T-j}\sum_{t=j+1}^{T}\lambda\left(\frac{x-u_t(\theta)}{h}\right)\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{u_t^{(1)}(\theta)u_{t-j}^{(1)}(\theta)}{h^2}\right) \\
&\frac{\partial^3 \tilde{F}_{\theta,j}(x,y;h)}{\partial \theta^3} \\
=&\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\ddot{\lambda}\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{-u_t^{(1)}(\theta)}{h}\right)^3+\Lambda\left(\frac{x-u_t(\theta)}{h}\right)\ddot{\lambda}\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{-u_{t-j}^{(1)}(\theta)}{h}\right)\right. \\
&+\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\dot{\lambda}\left(\frac{x-u_t(\theta)}{h}\right)\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(\frac{u_t^{(1)}(\theta)}{h}\right)^2\frac{u_{t-j}^{(1)}(\theta)}{h}\right\} \\
&+\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\lambda\left(\frac{x-u_t(\theta)}{h}\right)\dot{\lambda}\left(\frac{y-u_{t-j}(\theta)}{h}\right)\frac{u_t^{(1)}(\theta)}{h}\left(\frac{u_{t-j}^{(1)}(\theta)}{h}\right)^2\right\} \\
&+\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\dot{\lambda}\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)3\left(\frac{u_t^{(1)}(\theta)}{h}\frac{u_t^{(2)}(\theta)}{h}\right)\right\} \\
&+\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\Lambda\left(\frac{x-u_t(\theta)}{h}\right)\dot{\lambda}\left(\frac{y-u_{t-j}(\theta)}{h}\right)3\left(\frac{u_{t-j}^{(1)}(\theta)}{h}\frac{u_{t-j}^{(2)}(\theta)}{h}\right)\right\} \\
&-\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\lambda\left(\frac{x-u_t(\theta)}{h}\right)\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)u_t^{(3)}(\theta)+\Lambda\left(\frac{x-u_t(\theta)}{h}\right)\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)u_{t-j}^{(3)}(\theta)\right\} \\
&+3\frac{1}{T-j}\sum_{t=j+1}^{T}\left\{\lambda\left(\frac{x-u_t(\theta)}{h}\right)\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(u_t^{(2)}(\theta)u_t^{(1)}(\theta)+u_t^{(1)}(\theta)u_{t-j}^{(2)}(\theta)\right)\right\}
\end{aligned}
$$

Provided Assumption 6 and boundedness of $\mathbb{E}\left|u_t\right|^3$, after applying Minkowski's inequality, we can show

$$
\begin{aligned}
&\mathbb{E}\sup_{\theta\in\Theta}\left\|u_t^{(a)}(\theta)\right\|^b \\
\leq&\mathbb{E}\sup_{\theta\in\Theta}\left\|\sum_{j=-\infty}^{\infty}\phi_j^{(a)}(\theta)\left|u_t\right|\right\|^b \leq C_b\sup_{\theta\in\Theta}\sum_{j=-\infty}^{\infty}\left\|\phi_j^{(a)}(\theta)\right\|^b\mathbb{E}\left|u_t\right|^b<\infty \text{ for } a,b=1,2,3.
\end{aligned}
$$

By Hölder's inequality,

$$
\begin{aligned}
&\mathbb{E}\sup_{\theta\in\Theta}\left\|\left(u_t^{(a_1)}(\theta)\right)\left(u_{t-j}^{(a_2)}(\theta)\right)\right\| \\
\leq&\left(\mathbb{E}\sup_{\theta\in\Theta}\left\|u_t^{(a_1)}(\theta)\right\|^{b_1}\right)^{1/b_1}\left(\mathbb{E}\sup_{\theta\in\Theta}\left\|u_{t-j}^{(a_2)}(\theta)\right\|^{b_2}\right)^{1/b_2}<\infty \text{ for } a_1,a_2=1,2,b_1,b_2>1 \text{ and } j=0,\pm1,\pm2,\ldots \\
&\mathbb{E}\sup_{\theta\in\Theta}\left\|\left(u_t^{(1)}(\theta)\right)^m\left(u_{t-j}^{(1)}(\theta)\right)^n\right\| \\
\leq&\left(\mathbb{E}\sup_{\theta\in\Theta}\left\|\left(u_t^{(1)}(\theta)\right)^m\right\|^2\right)^{1/2}\left(\mathbb{E}\sup_{\theta\in\Theta}\left\|\left(u_{t-j}^{(1)}(\theta)\right)^n\right\|^2\right)^{1/2} \text{ for } m,n=1,2 \text{ and } j=0,\pm1,\pm2,\ldots
\end{aligned}
$$

Following above inequalities together with the uniform boundedness of density of the smoothed cumulative functions and its derivatives of order 1 and 2, we are able to prove

$$
\begin{aligned}
&\mathbb{E}\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2}\left\|\frac{\partial^a\tilde{F}_{\theta,j}(x,y;h)}{\partial\theta^a}\right\|<\infty, \text{ and} \\
&\mathbb{E}\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2}\left\|\frac{\partial^a\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta^a}\right\|<\infty, \text{ for } a=1,2,3 \text{ and } j=\pm1,\pm2,\ldots
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\mathbb{E}\left\|\sup_{\theta\in\Theta}\frac{\partial^3}{\partial\theta^3}\tilde{Q}_T(\theta)\right\| \\
\leq&4\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\mathbb{E}\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2}\left\|3\left(\frac{\partial\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta}\right)\frac{\partial^2\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta^2}\right\|dW(x,y) \\
&+4\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{(j\pi)^2}\mathbb{E}\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2}\left\|\tilde{\sigma}_{\theta,j}(x,y;h)\frac{\partial^3\tilde{\sigma}_{\theta,j}(x,y;h)}{\partial\theta^3}\right\|dW(x,y) \\
\leq&C\sum_{j=1}^{T-1}\frac{1}{(j\pi)^2}<\infty
\end{aligned}
$$

The score evaluated at $\theta=\theta_0$ is

$$
\begin{aligned}
&\frac{\partial\tilde{\sigma}_{\theta_0,j}(x,y;h)}{\partial\theta} \\
=&\frac{1}{T-j}\sum_{t=1+j}^{T}\left\{\lambda\left(\frac{x-u_t}{h}\right)\Lambda\left(\frac{y-u_{t-j}}{h}\right)\frac{1}{h}\left(-u_t^{(1)}(\theta_0)\right)+\Lambda\left(\frac{x-u_t}{h}\right)\lambda\left(\frac{y-u_{t-j}}{h}\right)\frac{1}{h}\left(-u_{t-j}^{(1)}(\theta_0)\right)\right\} \\
&+\tilde{F}_{\theta_0,j}(\infty,y;h)\frac{1}{T-j}\sum_{t=1+j}^{T}\lambda\left(\frac{x-u_t}{h}\right)\frac{1}{h}u_t^{(1)}(\theta_0)+\tilde{F}_{\theta_0,j}(x,\infty;h)\frac{1}{T-j}\sum_{t=1+j}^{T}\lambda\left(\frac{y-u_{t-j}}{h}\right)\frac{1}{h}u_{t-j}^{(1)}(\theta_0)
\end{aligned}
$$

with

$$
\begin{aligned}
u_t(\theta_0)&=u_t, \\
u_t^{(1)}(\theta_0)&=\sum_{j=-\infty}^{\infty}\phi_j^{(1)}(\theta_0)u_{t-j}
\end{aligned}
$$

Given $\mathbb{E}(u_t) = 0$, we have that, for $j = 1, 2, \ldots$

$$
\begin{aligned}
&\frac{\partial \tilde{\sigma}_{\theta_0,j}(x,y;h)}{\partial \theta} \\
\longrightarrow_p &- \frac{1}{h}\phi_0^{(1)}(\theta_0)\,\mathbb{E}\left(u_t\lambda\left(\frac{x-u_t}{h}\right)\right)\mathbb{E}\left(\Lambda\left(\frac{y-u_t}{h}\right)\right) - \frac{1}{h}\phi_j^{(1)}(\theta_0)\,\mathbb{E}\left(u_t\Lambda\left(\frac{y-u_t}{h}\right)\right)\mathbb{E}\left(\lambda\left(\frac{x-u_t}{h}\right)\right) \\
&- \frac{1}{h}\phi_{-j}^{(1)}(\theta_0)\,\mathbb{E}\left(u_t\Lambda\left(\frac{x-u_t}{h}\right)\right)\mathbb{E}\left(\lambda\left(\frac{y-u_t}{h}\right)\right) - \frac{1}{h}\phi_0^{(1)}(\theta_0)\,\mathbb{E}\left(u_t\lambda\left(\frac{y-u_t}{h}\right)\right)\mathbb{E}\left(\Lambda\left(\frac{x-u_t}{h}\right)\right) \\
&+ \frac{1}{h}\phi_0^{(1)}(\theta_0)\,\mathbb{E}\left(\Lambda\left(\frac{y-u_t}{h}\right)\right)\mathbb{E}\left(u_t\lambda\left(\frac{x-u_t}{h}\right)\right) + \frac{1}{h}\phi_0^{(1)}(\theta_0)\,\mathbb{E}\left(\Lambda\left(\frac{x-u_t}{h}\right)\right)\mathbb{E}\left(u_t\lambda\left(\frac{y-u_t}{h}\right)\right) \\
= &- \frac{1}{h}\phi_j^{(1)}(\theta_0)\,\mathbb{E}\left(u_t\Lambda\left(\frac{y-u_t}{h}\right)\right)\mathbb{E}\left(\lambda\left(\frac{x-u_t}{h}\right)\right) - \frac{1}{h}\phi_{-j}^{(1)}(\theta_0)\,\mathbb{E}\left(u_t\Lambda\left(\frac{x-u_t}{h}\right)\right)\mathbb{E}\left(\lambda\left(\frac{y-u_t}{h}\right)\right) \\
= &- \phi_j^{(1)}(\theta_0)\mu^h(y)\lambda^h(x) - \phi_{-j}^{(1)}(\theta_0)\mu^h(x)\lambda^h(y)
\end{aligned}
$$

As smoothing parameter h is sufficiently small, Λ can perform like indicator function. Hence,

$$
\begin{aligned}
\frac{1}{h}\,\mathbb{E}\left(\lambda\left(\frac{x-u_t}{h}\right)\right) &= \frac{1}{h}\int \lambda\left(\frac{x-u}{h}\right) f(u)du \to f(x) \\
\frac{1}{h}\,\mathbb{E}\left(u_t\lambda\left(\frac{x-u_t}{h}\right)\right) &= \frac{1}{h}\int u\lambda\left(\frac{x-u}{h}\right) f(u)du \to xf(x) \\
\mathbb{E}\left(\Lambda\left(\frac{y-u_t}{h}\right)\right) &= \int \Lambda\left(\frac{y-u}{h}\right) f(u)du = \frac{1}{h}\int F(u)\,\lambda\left(\frac{y-u}{h}\right) du \to F(y) \\
\mathbb{E}\left(u_t\Lambda\left(\frac{y-u_t}{h}\right)\right) &= \int u\Lambda\left(\frac{y-u}{h}\right) f(u)du = \frac{1}{h}\int u\lambda\left(\frac{y-u}{h}\right) F(u)\,du - \int F(u)\,\Lambda\left(\frac{y-u}{h}\right) du \\
&\to F(y)\,y - \int_{-\infty}^{y} F(u)\,du \\
&= F(y)\,y - F(y)\,y + \int_{-\infty}^{y} uf(u)\,du = \mathbb{E}\left(uI\left(u \le y\right)\right) \equiv \mu(y)
\end{aligned}
$$

We treat the derivative of $\tilde{\sigma}_{\theta_0,j}(x,y)$ derived from indicator function as $\frac{\partial \tilde{\sigma}_{\theta_0,j}(x,y;0)}{\partial \theta}$ which is defined as $\lim_{h\to 0} \frac{\partial \tilde{\sigma}_{\theta_0,j}(x,y;h)}{\partial \theta}$

$$
\begin{aligned}
&\frac{\partial \tilde{\sigma}_{\theta_0,j}(x,y)}{\partial \theta} \\
\to_p &- \phi_j^{(1)}(\theta_0)\,\mathbb{E}\left(uI\left(u \le y\right)\right) f(x) - \phi_{-j}^{(1)}(\theta_0)\,\mathbb{E}\left(uI\left(u \le x\right)\right) f(y) \\
= &- \phi_j^{(1)}(\theta_0)\mu(y)f(x) - \phi_{-j}^{(1)}(\theta_0)\mu(x)f(y)
\end{aligned}
$$

The limit coincides with the limit using "generalized" derivative of indicator function. Similarly, for any fixed h,

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta^2}\tilde{Q}_T(\theta;h)\\
\to_p &4\int_{\mathcal{R}^2}\sum_{j=1}^{\infty}\frac{1}{(j\pi)^2}\left\{\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{\theta_0,j}(x,y;h)}{\partial\theta}\right)\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{\theta_0,j}(x,y;h)}{\partial\theta}\right)'\right\}dW(x,y)\\
=&\frac{4}{\pi^2}\sum_{j=1}^{\infty}\phi_j^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)'\frac{1}{j^2}\int_{\mathcal{R}}\left(\mu^h(y)\right)^2dW(y)\int_{\mathcal{R}}\left(\lambda^h(x)\right)^2dW(x)\\
&+\frac{4}{\pi^2}\sum_{j=1}^{\infty}\phi_{-j}^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'\frac{1}{j^2}\int_{\mathcal{R}}\left(\mu^h(x)\right)^2dW(x)\int_{\mathcal{R}}\left(\lambda^h(y)\right)^2dW(y)\\
&+\frac{4}{\pi^2}\sum_{j=1}^{\infty}\left(\phi_j^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'+\phi_{-j}^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)'\right)\frac{1}{j^2}\left(\int_{\mathcal{R}}\mu^h(y)\lambda^h(y)dW(y)\right)^2\\
=&\frac{4}{\pi^2}\left(\Sigma_{0,1}+\Sigma_{0,1}^*\right)\rho_1^h\rho_2^h+\frac{4}{\pi^2}\left(\Sigma_{0,1}^\dagger+\Sigma_{0,1}^{\dagger'}\right)\left(\rho_{12}^h\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&T^{1/2}\frac{\partial}{\partial\theta}\widehat{\tilde{Q}_T(\theta_0;h)}\\
=&-\frac{4}{\pi^2}\frac{1}{T^{1/2}}\sum_{t=2}^{T}\int_{\mathcal{R}}\left(\Lambda\left(\frac{x-u_t}{h}\right)-\varphi^h(x)\right)\frac{1}{h}\mathbb{E}\left(\lambda\left(\frac{x-u_t}{h}\right)\right)dW(x)\\
&\times\left\{\sum_{j=1}^{t-1}\frac{1}{j^2}\phi_j^{(1)}(\theta_0)\int_{\mathcal{R}}\left(\Lambda\left(\frac{y-u_{t-j}}{h}\right)-\varphi^h(y)\right)\mathbb{E}\left(u_t\Lambda\left(\frac{y-u_t}{h}\right)\right)dW(y)\right\}\\
&-\frac{4}{\pi^2}\frac{1}{T^{1/2}}\sum_{t=2}^{T}\int_{\mathcal{R}}\left(\Lambda\left(\frac{x-u_t}{h}\right)-\varphi^h(x)\right)\mathbb{E}\left(u_t\Lambda\left(\frac{x-u_t}{h}\right)\right)dW(x)\\
&\times\left\{\sum_{j=1}^{t-1}\frac{1}{j^2}\phi_{-j}^{(1)}(\theta_0)\int_{\mathcal{R}}\left(\Lambda\left(\frac{y-u_{t-j}}{h}\right)-\varphi^h(y)\right)\frac{1}{h}\mathbb{E}\left(\lambda\left(\frac{y-u_t}{h}\right)\right)dW(y)\right\}+o_p(1)\\
=&-\frac{4}{T^{1/2}}\sum_{t=2}^{T}e_t^hV_{t-1}^h-\frac{4}{T^{1/2}}\sum_{t=2}^{T}\nu_tE_{t-1}^h+o_p(1)
\end{aligned}
$$

We can apply CLT for martingale difference sequence Brown (1971) on $e_t^hV_{t-1}^h$ and $\nu_tE_{t-1}^h$. We obtain

$$T^{1/2}\frac{\partial}{\partial\theta}\tilde{Q}_T(\theta_0;h)\longrightarrow_p\mathcal{N}\left(0,\frac{16}{\pi^4}H_{1,h}\right)$$

where $H_{1,h}$ is

$$\left\{\sigma_{e;h}^2\sigma_{\nu;h}^2\left(\Sigma_{0,2}+\Sigma_{0,2}^*\right)+\sigma_{e\nu;h}^2\left(\Sigma_{0,2}^\dagger+\Sigma_{0,2}^{\dagger'}\right)\right\}$$

Since $\mathbb{E}\left(e_t\right) = \mathbb{E}\left(\nu_t\right) = 0$,

$$\begin{aligned}
\sigma_{e;h}^2 &= \mathbb{E}\left(\left(e_t^h\right)^2\right) = \mathbb{E}\left(\left(\int_{\mathcal{R}}\left(\Lambda\left(\frac{x-u_t}{h}\right) - \varphi^h(x)\right)\lambda^h(x)dW(x)\right)^2\right) \\
&= \iiint\left(\Lambda\left(\frac{x-z}{h}\right) - \varphi^h(x)\right)\lambda^h(x)dW(x)\left(\Lambda\left(\frac{y-z}{h}\right) - \varphi^h(y)\right)\lambda^h(y)dW(y)f(z)dz \\
&= \iiint\Lambda\left(\frac{x-z}{h}\right)\Lambda\left(\frac{y-z}{h}\right)f(z)dz\lambda^h(x)dW(x)\lambda^h(y)dW(y) - \left(\int_{\mathcal{R}}\varphi^h(x)\lambda^h(x)dW(x)\right)^2 \\
&\to \int_{\mathcal{R}^2} F\left(x\wedge y\right)f(x)f(y)dW(x,y) - \left(\int_{\mathcal{R}} F(x)f(x)dW(x)\right)^2 \equiv \sigma_e^2 \text{ as } h\to 0
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sigma_{e\nu;h}^2 &= \mathbb{E}\left(e_t^h\nu_t^h\right) \\
&= \iiint\Lambda\left(\frac{x-z}{h}\right)\Lambda\left(\frac{y-z}{h}\right)f(z)dz\lambda^h(x)dW(x)\mu^h(y)dW(y) \\
&\quad - \left(\int_{\mathcal{R}}\varphi^h(x)\lambda^h(x)dW(x)\right)\left(\int_{\mathcal{R}}\varphi^h(y)\mu^h(y)dW(y)\right) \\
&\to \int_{\mathcal{R}^2} F\left(x\wedge y\right)f(x)\mu(y)dW(x,y) - \int_{\mathcal{R}} F(x)f(x)dW(x)\left(\int_{\mathcal{R}} F(y)\mu(y)dW(y)\right) \equiv \sigma_{e\nu}^2 \text{ as } h\to 0
\end{aligned}$$

$$\begin{aligned}
\sigma_{\nu,h}^2 &= \mathbb{E}\left(\left(\nu_t^h\right)^2\right) = \mathbb{E}\left(\left(\int_{\mathcal{R}}\left(\Lambda\left(\frac{x-u_t}{h}\right) - \varphi^h(x)\right)\mu^h(x)dW(x)\right)^2\right) \\
&= \iiint\Lambda\left(\frac{x-z}{h}\right)\Lambda\left(\frac{y-z}{h}\right)f(z)dz\mu^h(x)dW(x)\mu^h(y)dW(y) \\
&\quad - \left(\int_{\mathcal{R}}\varphi^h(x)\mu^h(x)dW(x)\right)^2 \\
&\to \int_{\mathcal{R}^2} F\left(x\wedge y\right)f(x)\mu(x)\mu(y)dW(x,y) - \left(\int_{\mathcal{R}} F(x)\mu(x)dW(x)\right)^2 \equiv \sigma_\nu^2 \text{ as } h\to 0
\end{aligned}$$

8 Numerical Calculus

8.1 Joint distribution and marginal distribution

Recall $u_t(\theta) = \varphi^{-1}(\theta, L)\varphi(\theta_0, L)u_t = \phi(\theta, L)u_t = \sum_{j=-\infty}^{\infty} \phi_j(\theta)u_{t-j}$. For simplicity, we normalize $\phi_0(\theta) = 1$. The joint distribution of $(u_t(\theta), u_{t-j}(\theta))$ evaluated at (x, y) is

$$
\begin{aligned}
&P\left(u_t(\theta) \le x, u_{t-j}(\theta) \le y\right) \\
=&P\left(\sum_{k=-\infty}^{\infty} \phi_k(\theta)u_{t-k} \le x, \sum_{k=-\infty}^{\infty} \phi_k(\theta)u_{t-j-k} \le y\right) \\
=&\mathbb{E}\left(P\left(u_t + \phi_j(\theta)u_{t-j} \le x - \sum_{k\neq 0,j} \phi_k(\theta)u_{t-k}, u_{t-j} + \phi_{-j}(\theta)u_t \le y - \sum_{k\neq 0,-j} \phi_k(\theta)u_{t-j-k}|\Omega_{t,j}\right)\right) \\
=&\mathbb{E}\left(\int_{-\infty}^{x-\sum_{k\neq 0,j}\phi_k(\theta)u_{t-k}} \int_{-\infty}^{y-\sum_{k\neq 0,-j}\phi_k(\theta)u_{t-j-k}} f_{z_1,z_2|\Omega_{t,j}}(t_1, t_2)dt_1dt_2\right) \\
=&\mathbb{E}\left(\iint_{\mathcal{D}} f_{u_1,u_2|\Omega_{t,j}}(u_1, u_2)du_1du_2\right) \\
=&\frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|}\mathbb{E}\left(\int_{-\infty}^{x-\sum_{k\neq 0,j}\phi_k(\theta)u_{t-k}} \int_{-\infty}^{y-\sum_{k\neq 0,-j}\phi_k(\theta)u_{t-j-k}} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\right.
\end{aligned}
$$

where $z_1 = u_t + \phi_j(\theta)u_{t-j}$, $z_2 = u_{t-j} + \phi_{-j}(\theta)u_t$, $u_t = \frac{z_1-\phi_j(\theta)z_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}$, $u_{t-j} = \frac{z_2-\phi_{-j}(\theta)z_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}$ and

$$\mathcal{D} = \{(t_1, t_2) \in \mathcal{R}^2 : t_1 + \phi_j(\theta)t_2 \le x - \sum_{k\neq 0,j} \phi_k(\theta)u_{t-k}, t_2 + \phi_{-j}(\theta)t_1 \le y - \sum_{k\neq 0,-j} \phi_k(\theta)u_{t-j-k}\}$$

$\Omega_{t,j} = \sigma(\dots, u_{t-j-1}, u_{t-j+1}, \dots, u_{t-1}, u_{t+1}, \dots)$ is information set generated by all innovations but u_t, u_{t-j}

Furthermore we assume $1 - \phi_j(\theta)\phi_{-j}(\theta) \neq 0$.

Taking the derivative of above joint probability w.r.t θ, we obtain

$$
\begin{aligned}
&\frac{\partial F_{\theta,j}(x, y)}{\partial\theta} \\
=&\frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|}\mathbb{E}\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\frac{\partial m(\theta)}{\partial\theta}\right) \\
&+\frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|}\mathbb{E}\left\{\int_{-\infty}^{m(\theta)} \frac{\partial}{\partial\theta}\left(\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right)dt_1\right\} \\
&+\frac{\operatorname{sgn}\left(1-\phi_j(\theta)\phi_{-j}(\theta)\right)\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)}{|1-\phi_j(\theta)\phi_{-j}(\theta)|} \\
&\times\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\} \\
=&D_1(\theta, x, y) + D_2(\theta, x, y) + D_3(\theta, x, y)
\end{aligned}
\tag{8.1}
$$

where

$$m(\theta) = x - \sum_{k\neq 0,j} \phi_k(\theta) u_{t-k}$$
$$n(\theta) = y - \sum_{k\neq 0,-j} \phi_k(\theta) u_{t-j-k}$$

and $D_3(\theta, x, y)$ can be re-expressed as following form for further study,

$$D_3(\theta, x, y) = \operatorname{sgn}\left(1 - \phi_j(\theta)\phi_{-j}(\theta)\right) \left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta) + \phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right) F_{\theta,j}(x, y)$$

Regarding the first component in the first order derivative

$$\begin{aligned}
&D_1(\theta, x, y) \\
=&\mathbb{E}\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta) - \phi_j(\theta)t_2}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{t_2 - \phi_{-j}(\theta)m(\theta)}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) dt_2 \frac{\partial m(\theta)}{\partial \theta}\right) \frac{1}{|1 - \phi_j(\theta)\phi_{-j}(\theta)|} \\
=&\mathbb{E}\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta) - \phi_j(\theta)t_2}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{t_2 - \phi_{-j}(\theta)m(\theta)}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) dt_2 (-\sum_{k\neq 0,j} \phi_k^{(1)}(\theta) u_{t-k})\right) \frac{1}{|1 - \phi_j(\theta)\phi_{-j}(\theta)|}
\end{aligned}$$

Before we continue analyzing second component, we first study the following derivative

$$\begin{aligned}
&\frac{\partial}{\partial \theta}\left(\int_{-\infty}^{n(\theta)} f(\frac{t_1 - \phi_j(\theta)t_2}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{t_2 - \phi_{-j}(\theta)t_1}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) dt_2\right) \\
= \quad & f(\frac{t_1 - \phi_j(\theta)n(\theta)}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{n(\theta) - \phi_{-j}(\theta)t_1}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) \left(-\sum_{k\neq 0,-j} \phi_k^{(1)}(\theta) u_{t-j-k}\right) \\
& + \int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1 - \phi_j(\theta)t_2}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) \frac{\partial}{\partial \theta}(\frac{t_1 - \phi_j(\theta)t_2}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{t_2 - \phi_{-j}(\theta)t_1}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) dt_2 \\
& + \int_{-\infty}^{n(\theta)} f(\frac{t_1 - \phi_j(\theta)t_2}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) f^{(1)}(\frac{t_2 - \phi_{-j}(\theta)t_1}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) \frac{\partial}{\partial \theta}(\frac{t_2 - \phi_{-j}(\theta)t_1}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) dt_2
\end{aligned}$$

where

$$\frac{\partial}{\partial \theta}(\frac{t_1 - \phi_j(\theta)t_2}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) = \frac{\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta) + \phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right) t_1 - \left(\phi_j^{(1)}(\theta) + \phi_j^2(\theta)\phi_{-j}^{(1)}(\theta)\right) t_2}{\left(1 - \phi_j(\theta)\phi_{-j}(\theta)\right)^2}$$
$$\frac{\partial}{\partial \theta}(\frac{t_2 - \phi_{-j}(\theta)t_1}{1 - \phi_j(\theta)\phi_{-j}(\theta)}) = \frac{\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta) + \phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right) t_2 - \left(\phi_{-j}^{(1)}(\theta) + \phi_{-j}^2(\theta)\phi_j^{(1)}(\theta)\right) t_1}{\left(1 - \phi_j(\theta)\phi_{-j}(\theta)\right)^2}$$

Then we can move on to

$$
\begin{aligned}
&|1-\phi_j(\theta)\phi_{-j}(\theta)|\, D_2(\theta,x,y)\\
=&\,\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\frac{\partial}{\partial\theta}\left(\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right)dt_1\right\}\\
=&\,\mathbb{E}\left\{\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})' f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})' dt_2dt_1\right\}
\end{aligned}
$$

The partial derivative of $F_{\theta,j}(x,y)$ evaluated at true value $\theta=\theta_0$, that is ,

$$
\begin{aligned}
&\phi_j(\theta_0)=0 \text{ for } j=\pm1,\pm2,\ldots\\
&\phi_0(\theta_0)=1
\end{aligned}
$$

Under this condition, $m(\theta_0)=x$, $n(\theta_0)=y$ and $D_1(\theta_0,x,y)$ can be simplified to

$$
\begin{aligned}
D_1(\theta_0,x,y)&=\mathbb{E}_{\Omega_{t,j}}\left(\int_{-\infty}^{y} f(x)f(t_2)dt_2(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta_0)u_{t-k})\right)\\
&=f(x)F(y)\,\mathbb{E}_{\Omega_{t,j}}\left(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta_0)u_{t-k}\right)\\
&=f(x)F(y)\times 0=0
\end{aligned}
$$

$$
\begin{aligned}
D_2(\theta_0,x,y)&=\mathbb{E}_{\Omega_{t,j}}\left\{\int_{-\infty}^{x} f(t_1)f(y)dt_1\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\right\}\\
&\quad+\mathbb{E}_{\Omega_{t,j}}\left\{\int_{-\infty}^{x}\int_{-\infty}^{y} f^{(1)}(t_1)\times(-\phi_j^{(1)}(\theta_0)t_2)\times f(t_2)dt_2dt_1\right\}\\
&\quad+\mathbb{E}_{\Omega_{t,j}}\left\{\int_{-\infty}^{x}\int_{-\infty}^{y} f(t_1)f^{(1)}(t_2)\times(-\phi_{-j}^{(1)}(\theta_0)t_1)dt_2dt_1\right\}\\
&=0-\phi_j^{(1)}(\theta_0)f(x)\int_{-\infty}^{y} f(t_2)t_2dt_2-\phi_{-j}^{(1)}(\theta_0)f(y)\int_{-\infty}^{x} f(t_1)t_1dt_1
\end{aligned}
$$

$$
D_3(\theta_0,x,y)=0
$$

$$
\begin{aligned}
\frac{\partial F_{\theta_0,j}(x,y)}{\partial\theta}&=D_1(\theta_0,x,y)+D_2(\theta_0,x,y)\\
&=-\phi_j^{(1)}(\theta_0)f(x)\int_{-\infty}^{y} f(u)udu-\phi_{-j}^{(1)}(\theta_0)f(y)\int_{-\infty}^{x} f(u)udu
\end{aligned}
\tag{8.2}
$$

Following similar procedures, we can also derive the expression for partial derivative of

marginal cdf of $u_t(\theta)$

$$
\begin{aligned}
F_\theta(x) &= P\left(u_t(\theta) \le x\right) \\
&= P\left(\sum_{j=-\infty}^{\infty} \phi_j(\theta) u_{t-j} \le x\right) \\
&= \mathbb{E}\left\{P\left(u_t \le x - \sum_{j\neq 0} \phi_j(\theta) u_{t-j} | \Omega_t\right)\right\} \\
&= \mathbb{E}\left(\int_{-\infty}^{x-\sum_{j\neq 0}\phi_j(\theta)u_{t-j}} f(u)du\right) \\
&= \mathbb{E}\left(F(x - \sum_{j\neq 0} \phi_j(\theta) u_{t-j})\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial F_\theta(x)}{\partial \theta} &= \mathbb{E}\left(\frac{\partial}{\partial \theta} F(x - \sum_{j\neq 0} \phi_j(\theta) u_{t-j})\right) \\
&= \mathbb{E}\left(f(x - \sum_{j\neq 0} \phi_j(\theta) u_{t-j})(-\sum_{j\neq 0} \phi_j^{(1)}(\theta) u_{t-j})\right) \\
&= -\sum_{j\neq 0} \phi_j^{(1)}(\theta)\, \mathbb{E}\left(f(x - \sum_{j\neq 0} \phi_j(\theta) u_{t-j}) u_{t-j}\right)
\end{aligned}
$$

The following two conditions are sufficient for uniform boundedness of $\frac{\partial F_\theta(x)}{\partial \theta}$:

$$
\sum_{j\neq 0} \sup_{\theta\in\Theta} \left\|\phi_j^{(1)}(\theta)\right\| < \infty
$$

$$
\mathbb{E}_{\Omega_t}\left(f(x - \sum_{j\neq 0} \phi_j(\theta) u_{t-j}) u_{t-j}\right) < \infty
$$

The first is guaranteed by Assumption.7 and second can be induced from uniform boundedness of marginal density function of innovations $f(\cdot)$ and $\mathbb{E}(|u_t|) < \infty$.
To bound $\frac{\partial F_{\theta,j}(x,y)}{\partial \theta}$, we need to bound D_1, D_2 uniformly respectively,

$$
\left\|\sup_{\theta\in\Theta} \sup_{(x,y)\in\mathcal{R}^2} D_1(\theta, x, y)\right\| < \infty
$$

$$
\left\|\sup_{\theta\in\Theta} \sup_{(x,y)\in\mathcal{R}^2} D_2(\theta, x, y)\right\| < \infty
$$

First we deal with $D_1(\theta, x, y)$,

$$
\begin{aligned}
&\left\| \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} D_1(\theta,x,y) \right\| \\
=&\left\| \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \mathbb{E}\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 (-\sum_{k\neq 0,j} \phi_k^{(1)}(\theta) u_{t-k}) \frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|} \right) \right| \\
\leq& \mathbb{E}\left(C \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|} \left| \int_{-\infty}^{n(\theta)} f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 \right| \sup_{\theta\in\Theta} \left\| -\sum_{k\neq 0,j} \phi_k^{(1)}(\theta) \right\| |u_{t-k}| \right) \\
=& C\,\mathbb{E}\left(\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left| \max\{F(\frac{n(\theta)-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)}), 1-F(\frac{n(\theta)-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\} \right| \sup_{\theta\in\Theta} \left\| -\sum_{k\neq 0,j} \phi_k^{(1)}(\theta) \right\| |u_{t-k}| \right) \\
\leq& \sup_{\theta\in\Theta} \sum_{k\neq 0,j} \|\phi_k^{(1)}(\theta)\|\, \mathbb{E}(|u_{t-k}|) < \infty
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \|D_2(\theta,x,y)\| \\
\leq& \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left\| \mathbb{E}\left\{ \int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_1 \left(-\sum_{k\neq 0,-j} \phi_k^{(1)}(\theta) u_{t-j-k} \right) \right\} \eta_j(\theta) \right\| \\
+& \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left\| \mathbb{E}\left\{ \int_{-\infty}^{m(\theta)} \int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) \frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 dt_1 \right\} \eta_j(\theta) \right| \\
+& \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left\| \mathbb{E}\left\{ \int_{-\infty}^{m(\theta)} \int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) \frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 dt_1 \right\} \eta_j(\theta) \right| \\
=& D_{21} + D_{22} + D_{23}
\end{aligned}
$$

where

$$\eta_j(\theta) = \frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|}$$

D_{21} is uniformly bounded following similar steps as one in $D_1(\theta, x, y)$

$$
\begin{aligned}
&D_{21} \\
\leq& C\,\mathbb{E}\left(\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left| \max\{F(\frac{m(\theta)-\phi_{-j}(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)}), 1-F(\frac{m(\theta)-\phi_{-j}(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\} \right| \sup_{\theta\in\Theta} \left\| -\sum_{k\neq 0,-j} \phi_k^{(1)}(\theta) \right\| |u_{t-j-k} \right. \\
\leq& \sup_{\theta\in\Theta} \sum_{k\neq 0,-j} \|\phi_k^{(1)}(\theta)\|\, \mathbb{E}(|u_{t-j-k}|) < \infty
\end{aligned}
$$

Here we assume

$$0 < \frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|} < C$$

This is implied by mixing condition for residual sequence(like $\sup_{\theta\in\Theta} \phi_j(\theta) < C|j|^{\mu}$ for a $\mu > 1$ for all $j = \pm 1, \pm 2, \ldots$.)

To prove boundedness condition of D_{22} and D_{23} ,

$$
\begin{aligned}
&D_{22} \\
&= \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left\| \mathbb{E}\left\{ \int_{-\infty}^{m(\theta)} \int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) \frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 dt_1 \right\} \eta_j(\theta) \right\| \\
&\leq \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left| \mathbb{E}\left\{ \int_{-\infty}^{m(\theta)} \int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) t_1 f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 dt_1 \right\} \eta_j(\theta) \right| \\
&\quad \cdot \sup_{\theta\in\Theta} \left\| \frac{\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2} \right\| \\
&\quad + \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2} \left| \mathbb{E}\left\{ \int_{-\infty}^{m(\theta)} \int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) t_2 f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 dt_1 \right\} \eta_j(\theta) \right| \\
&\quad \cdot \sup_{\theta\in\Theta} \left\| \frac{\phi_j^{(1)}(\theta)+\phi_j^2(\theta)\phi_{-j}^{(1)}(\theta)}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2} \right\|
\end{aligned}
$$

The uniform boundedness of $\left\| \frac{\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2} \right\|$ and $\left\| \frac{\phi_j^{(1)}(\theta)+\phi_j^2(\theta)\phi_{-j}^{(1)}(\theta)}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2} \right\|$ is guaranteed by assumption 7 and mixing condition.
Now we need to discuss the rest,

$$
\begin{aligned}
&\int_{-\infty}^{m(\theta)} \int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) t_1 f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}) dt_2 dt_1 \frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|} \\
&= \iint_{\mathcal{D}} f^{(1)}(u_1)(u_1+\phi_j(\theta)u_2) f(u_2) du_1 du_2 \\
&= \iint_{\mathcal{D}} f^{(1)}(u_1) u_1 f(u_2) du_1 du_2 + \phi_j(\theta) \iint_{\mathcal{D}} u_2 f(u_2) df(u_1) du_2 \\
&\leq \int_{-\infty}^{\infty} \int_{-\infty}^{n(\theta)-\phi_{-j}(\theta)u_1} |f^{(1)}(u_1)u_1| f(u_2) du_2 du_1 + |\phi_j(\theta)| \int_{-\infty}^{\infty} \int_{f^{-1}(U_1)\leq m(\theta)-\phi_j(\theta)u_2} |u_2| f(u_2) dU_1 du_2 \\
&= \int_{-\infty}^{\infty} |f^{(1)}(u_1)u_1| F(n(\theta)-\phi_{-j}(\theta)u_1) du_1 + |\phi_j(\theta)| \int_{-\infty}^{\infty} f(u_2)|u_2| \left| \int_{f^{-1}(U_1)\leq m(\theta)-\phi_j(\theta)u_2} dU_1 \right| du_2 \\
&\leq \int_{-\infty}^{\infty} |f^{(1)}(u_1)u_1| du_1 + C\,|\phi_j(\theta)| \int_{-\infty}^{\infty} f(u_2)|u_2| du_2
\end{aligned}
$$

where $U_1 = f(u_1)$.
The first inequality comes from the expansion of zones of a nonnegative integrand:

$$
\begin{aligned}
&\mathcal{D} = \{(u_1,u_2)\in\mathcal{R}^2 : u_1+\phi_j(\theta)u_2 \leq m(\theta), u_2+\phi_{-j}(\theta)u_1 \leq n(\theta)\} \\
&\subseteq \left\{(u_1,u_2)\in\mathcal{R}^2 : u_1\in\mathcal{R}, u_2 \leq n(\theta)-\phi_{-j}(\theta)u_1\right\}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathcal{D} = \{(u_1,u_2)\in\mathcal{R}^2 : u_1+\phi_j(\theta)u_2 \leq m(\theta), u_2+\phi_{-j}(\theta)u_1 \leq n(\theta)\} \\
&\subseteq \left\{(u_1,u_2)\in\mathcal{R}^2 : u_2\in\mathcal{R}, u_1 \leq m(\theta)-\phi_j(\theta)u_2\right\}
\end{aligned}
$$

The second inequality follows from uniform boundedness of $F(\cdot)$ and $f(\cdot)$.

Once we have above inequality, we can show

$$
\begin{aligned}
&\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2}\left|\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})t_1 f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|}\right| \\
\leq &\sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2}\left|\mathbb{E}\iint_{\mathcal{D}} f^{(1)}(u_1)u_1 f(u_2)du_1du_2\right| + \sup_{\theta\in\Theta,(x,y)\in\mathcal{R}^2}\left|\mathbb{E}\left|\phi_j(\theta)\right|\iint_{\mathcal{D}} u_2 f(u_2)df(u_1)du_2\right| \\
\leq &\sup_{\theta\in\Theta}\int_{-\infty}^{\infty}\left|f^{(1)}(u)u\right|du + C\sup_{\theta\in\Theta}|\phi_j(\theta)|\int_{-\infty}^{\infty} f(u)|u|du \\
<&\infty
\end{aligned}
\tag{8.3}
$$

as long as

$$\int_{-\infty}^{\infty}\left|f^{(1)}(u)u\right|du<\infty$$

and

$$\int_{-\infty}^{\infty} f(u)\,|u|\,du<\infty$$

Following same steps, we can show D_{22} and D_{23} are uniformly bounded by some constant C.

Following above calculation, the second order partial derivative of $F_{\theta,j}(x,y)$ is

$$\frac{\partial^2}{\partial\theta\partial\theta'}F_{\theta,j}(x,y)=\frac{\partial D_1(\theta,x,y)}{\partial\theta'}+\frac{\partial D_2(\theta,x,y)}{\partial\theta'}+\frac{\partial D_3(\theta,x,y)}{\partial\theta'}$$

where

$$
\begin{aligned}
&\frac{\partial D_1(\theta,x,y)}{\partial\theta'} \\
=&\mathbb{E}\left\{\left(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta)u_{t-k}\right)\frac{\partial}{\partial\theta'}\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right)\eta_j(\theta)\right\} \\
&+\mathbb{E}\left\{\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right)\left(\frac{\partial}{\partial\theta'}(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta)u_{t-k})\right)\eta_j(\theta)\right\} \\
&+\mathbb{E}\left\{\left(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta)u_{t-k}\right)\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right)\frac{\partial\eta_j(\theta)}{\partial\theta'}\right\}
\end{aligned}
$$

For the first term,

$$
\begin{aligned}
&\frac{\partial}{\partial\theta'}\left(\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right) \\
=&f(\frac{m(\theta)-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)' \\
&+\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\left(\frac{\partial}{\partial\theta'}(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\right)f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2 \\
&+\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\left(\frac{\partial}{\partial\theta'}(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\right)dt_2
\end{aligned}
$$

where

$$\frac{\partial}{\partial\theta}(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}) = \frac{\left(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta)u_{t-k}\right)(1-\phi_j(\theta)\phi_{-j}(\theta)) + m(\theta)\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)-\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2} - \frac{\left(\phi_j^{(1)}(\theta)-\phi_j^2(\theta)\phi_{-j}^{(1)}(\theta)\right)t_2}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2}$$

$$\frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)}) = \frac{\left(\phi_j(\theta)\phi_{-j}^2(\theta)-\phi_{-j}(\theta)\right)\left(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta)u_{t-k}\right) - \left(\phi_{-j}^{(1)}(\theta)-\phi_{-j}^2(\theta)\phi_j^{(1)}(\theta)\right)m(\theta)}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2} + \frac{\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)t_2}{(1-\phi_j(\theta)\phi_{-j}(\theta))^2}$$

For the second term in the partial derivative of $D_1(\theta,x,y)$,

$$\begin{aligned}&\frac{\partial}{\partial\theta'}(-\sum_{k\neq 0,j}\phi_k^{(1)}(\theta)u_{t-k})\\ =&-\sum_{k\neq 0,j}\phi_k^{(2)}(\theta)u_{t-k}\end{aligned}$$

The third term,

$$\begin{aligned}\frac{\partial\eta_j(\theta)}{\partial\theta'} &= \frac{\partial}{\partial\theta'}\frac{1}{|1-\phi_j(\theta)\phi_{-j}(\theta)|}\\ &= \frac{\mathrm{sgn}(1-\phi_j(\theta)\phi_{-j}(\theta))\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)'}{|1-\phi_j(\theta)\phi_{-j}(\theta)|}\end{aligned}$$

Regarding the derivative of $D_2(\theta,x,y)$, we need to calculate three parts:

$$\begin{aligned}&\frac{\partial D_2(\theta,x,y)}{\partial\theta'}\\ =&\eta_j(\theta)\frac{\partial}{\partial\theta'}\mathbb{E}\left\{\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\right\}\\ &+\eta_j(\theta)\frac{\partial}{\partial\theta'}\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}\left(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right)f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\\ &+\eta_j(\theta)\frac{\partial}{\partial\theta'}\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}\left(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right)dt_2dt_1\right\}\\ &+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\right\}\frac{\partial\eta_j(\theta)}{\partial\theta'}\\ &+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}\left(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right)f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\frac{\partial\eta_j(\theta)}{\partial\theta'}\\ &+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}\left(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right)dt_2dt_1\right\}\frac{\partial\eta_j(\theta)}{\partial\theta'}\end{aligned}$$

We calculate these three components one by one:

$$
\begin{aligned}
&\frac{\partial}{\partial\theta'}\mathbb{E}\left\{\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\right\}\\
=&\mathbb{E}\left\{\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\frac{\partial}{\partial\theta'}\left(\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\right)\right\}\\
&+\mathbb{E}\left\{\left(\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\right)\frac{\partial}{\partial\theta'}\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\right\}\\
=&\mathbb{E}\left\{\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\int_{-\infty}^{m(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta'}(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\right\}\\
&+\mathbb{E}\left\{\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta'}(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\right\}\\
&+\mathbb{E}\left\{f(\frac{m(\theta)-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)\left(-\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta)u_{t-j-k}\right)'\right\}\\
&+\mathbb{E}\left\{\left(\int_{-\infty}^{m(\theta)} f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_1\right)\left(-\sum_{k\neq 0,-j}\phi_k^{(2)}(\theta)u_{t-j-k}\right)\right\}
\end{aligned}
$$

The second term

$$
\begin{aligned}
&\frac{\partial}{\partial\theta'}\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}\left(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)}\right)f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\\
=&\mathbb{E}\left\{\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\Big|_{t_1=m(\theta)}f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\frac{\partial m(\theta)}{\partial\theta}\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\frac{\partial}{\partial\theta}\left(\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right)dt_1\right\}\\
=&\mathbb{E}\left\{\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\Big|_{t_1=m(\theta)}f(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\frac{\partial m(\theta)}{\partial\theta}\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\left(f^{(1)}(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\Big|_{t_2=n(\theta)}f(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial n(\theta)}{\partial\theta}\right)dt_1\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(2)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta'}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right.\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial^2}{\partial\theta\partial\theta'}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta'}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\right.
\end{aligned}
$$

The third term:

$$
\begin{aligned}
&\frac{\partial}{\partial\theta}\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\\
=&\mathbb{E}\left\{\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\Big|_{t_1=m(\theta)}dt_2\frac{\partial m(\theta)}{\partial\theta}\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\frac{\partial}{\partial\theta}\left(\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right)dt_1\right\}\\
=&\mathbb{E}\left\{\int_{-\infty}^{n(\theta)} f(\frac{m(\theta)-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)m(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\Big|_{t_1=m(\theta)}dt_2\frac{\partial m(\theta)}{\partial\theta'}\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\left(f(\frac{t_1-\phi_j(\theta)n(\theta)}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{n(\theta)-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\Big|_{t_2=n(\theta)}\frac{\partial n(\theta)}{\partial\theta'}\right)dt_1\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f^{(1)}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta'}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\right.\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial^2}{\partial\theta\partial\theta'}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(1)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2dt_1\right\}\\
&+\mathbb{E}\left\{\int_{-\infty}^{m(\theta)}\int_{-\infty}^{n(\theta)} f(\frac{t_1-\phi_j(\theta)t_2}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})\frac{\partial}{\partial\theta'}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})f^{(2)}(\frac{t_2-\phi_{-j}(\theta)t_1}{1-\phi_j(\theta)\phi_{-j}(\theta)})dt_2\right.
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial D_3(\theta,x,y)}{\partial\theta'}\\
=&\frac{\partial\,\mathrm{sgn}\,(1-\phi_j(\theta)\phi_{-j}(\theta))}{\partial\theta'}\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)F_{\theta,j}(x,y)\\
&+\mathrm{sgn}\,(1-\phi_j(\theta)\phi_{-j}(\theta))\frac{\partial\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)}{\partial\theta'}F_{\theta,j}(x,y)\\
&+\mathrm{sgn}\,(1-\phi_j(\theta)\phi_{-j}(\theta))\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)\frac{\partial F_{\theta,j}(x,y)}{\partial\theta'}\\
=&\mathrm{sgn}\,(1-\phi_j(\theta)\phi_{-j}(\theta))\left(\phi_j^{(2)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(2)}(\theta)+\phi_j^{(1)}(\theta)\phi_{-j}^{(1)}(\theta)'+\phi_{-j}^{(1)}(\theta)\phi_j^{(1)}(\theta)'\right)F_{\theta,j}(x,y)\\
&+\mathrm{sgn}\,(1-\phi_j(\theta)\phi_{-j}(\theta))\left(\phi_j^{(1)}(\theta)\phi_{-j}(\theta)+\phi_j(\theta)\phi_{-j}^{(1)}(\theta)\right)\frac{\partial F_{\theta,j}(x,y)}{\partial\theta'}
\end{aligned}
$$

The second order partial derivative of $F_{\theta,j}(x,y)$ evaluated at $\theta=\theta_0$ is

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\theta'}F_{\theta_0,j}(x,y)\\
=&\frac{\partial D_1(\theta_0,x,y)}{\partial\theta'}+\frac{\partial D_2(\theta_0,x,y)}{\partial\theta'}+\frac{\partial D_3(\theta_0,x,y)}{\partial\theta'}
\end{aligned}
\tag{8.4}
$$

where

$$
\begin{aligned}
\frac{\partial D_1(\theta_0,x,y)}{\partial\theta'} &= \left(\sum_{k\neq 0,j}\phi_k^{(1)}(\theta_0)(\phi_{k-j}^{(1)}(\theta_0))'\right)\mathbb{E}(u_t^2)\left(f(x)f(y)+f^{(1)}(x)F(y)\right)\\
\frac{\partial D_2(\theta_0,x,y)}{\partial\theta'} &= f^{(1)}(y)F(x)\left(\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta_0)(\phi_k^{(1)}(\theta_0))'\,\mathbb{E}(u_t^2)\right)\\
&\quad + f(x)f(y)\left(\sum_{k\neq 0,-j}\phi_k^{(1)}(\theta_0)(\phi_k^{(1)}(\theta_0))'\,\mathbb{E}(u_t^2)\right)\\
&\quad + f^{(1)}(x)\left(\int_{-\infty}^{y}u^2f(u)du\right)\left(\phi_j^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)'\right)+f(x)\left(\int_{-\infty}^{y}f(u)udu\right)\left(-\phi_j^{(2)}(\theta_0)\right)\\
&\quad + \left(\int_{-\infty}^{x}f^{(1)}(u)udu\right)F(y)\left(\phi_j^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'+\phi_{-j}^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)'\right)\\
&\quad + \left(\int_{-\infty}^{x}f^{(1)}(u)udu\right)\left(\int_{-\infty}^{y}f^{(1)}(u)udu\right)\left(\phi_j^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'\right)\\
&\quad + \left(\int_{-\infty}^{x}f^{(1)}(u)udu\right)\left(\int_{-\infty}^{y}f^{(1)}(u)udu\right)\left(\phi_{-j}^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)'\right)\\
&\quad + \left(\int_{-\infty}^{y}f^{(1)}(u)udu\right)F(x)\left(\phi_j^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'+\phi_{-j}^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)'\right)\\
&\quad + f(y)\left(\int_{-\infty}^{x}f(u)udu\right)\left(-\phi_{-j}^{(2)}(\theta_0)\right)+f^{(1)}(y)\left(\int_{-\infty}^{x}u^2f(u)du\right)\left(\phi_{-j}^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'\right)\\
\frac{\partial D_3(\theta_0,x,y)}{\partial\theta'} &= \left(\phi_j^{(1)}(\theta_0)\phi_{-j}^{(1)}(\theta_0)'+\phi_{-j}^{(1)}(\theta_0)\phi_j^{(1)}(\theta_0)'\right)F_{\theta_0,j}(x,y)
\end{aligned}
$$

Asymptotic distribution: We first carry out the analysis of the asymptotic study of the estimator based on its approximation using joint cdf and marginal cdf.

$$
\begin{aligned}
&\frac{\partial}{\partial\theta}\mathcal{Q}_T(\theta_0)\\
=&\int_{\mathcal{R}^2}\sum_{j=1}^{T-1}(1-\frac{j}{T})\frac{1}{(j\pi)^2}2\sigma_{\theta_0,j}(x,y)\frac{\partial\sigma_{\theta_0,j}(x,y)}{\partial\theta}dW(x,y)\\
=&2\sum_{j=1}^{T-1}(1-\frac{j}{T})\frac{1}{(j\pi)^2}\int_{\mathcal{R}^2}\sigma_{\theta_0,j}(x,y)\frac{\partial\sigma_{\theta_0,j}(x,y)}{\partial\theta}dW(x,y)\\
=&2\sum_{j=1}^{T-1}(1-\frac{j}{T})\frac{1}{(j\pi)^2}\int_{\mathcal{R}^2}\sigma_{\theta_0,j}(x,y)\left(\frac{\partial F_{\theta_0,,j}(x,y)}{\partial\theta}-\frac{\partial F_{\theta_0}(x)}{\partial\theta}F_{\theta_0}(y)-F_{\theta_0}(x)\frac{\partial F_{\theta_0}(y)}{\partial\theta}\right)dW(x,y)\\
=&2\sum_{j=1}^{T-1}(1-\frac{j}{T})\frac{1}{(j\pi)^2}\int_{\mathcal{R}^2}\sigma_{\theta_0,j}(x,y)\left(-\phi_j^{(1)}(\theta_0)f(x)\int_{-\infty}^{y}f(u)udu-\phi_{-j}^{(1)}(\theta_0)f(y)\int_{-\infty}^{x}f(u)udu\right)dW(x,y)\\
=&0
\end{aligned}
$$

since $\sigma_{\theta_0,j}(x,y)=0$ for all $(x,y)\in\mathcal{R}^2$ and $j=\pm1,\pm2,\ldots$ **Hessian matrix**:

$$
\begin{aligned}
&\frac{\partial^2}{\partial\theta\partial\theta'}\mathcal{Q}_T(\theta)\\
=&2\sum_{j=1}^{T-1}(1-\frac{j}{T})\frac{1}{(j\pi)^2}\int_{\mathcal{R}^2}\left(\frac{\partial\sigma_{\theta,j}(x,y)}{\partial\theta}\frac{\partial\sigma_{\theta,j}(x,y)}{\partial\theta'}+\sigma_{\theta,j}(x,y)\frac{\partial^2\sigma_{\theta,j}(x,y)}{\partial\theta\partial\theta'}\right)dW(x,y)
\end{aligned}
$$

where

$$\frac{\partial \sigma_{\theta,j}(x,y)}{\partial \theta} = \frac{\partial F_{\theta,t,j}(x,y)}{\partial \theta} - \frac{\partial F_{\theta,t}(x)}{\partial \theta} F_{\theta,t-j}(y) - F_{\theta,t}(x) \frac{\partial F_{\theta,t-j}(y)}{\partial \theta}$$

$$\begin{aligned}\frac{\partial^2 \sigma_{\theta,j}(x,y)}{\partial \theta \partial \theta'} = & \frac{\partial^2 F_{\theta,t,j}(x,y)}{\partial \theta \partial \theta'} - \frac{\partial^2 F_{\theta,t}(x)}{\partial \theta \partial \theta'} F_{\theta,t-j}(y) - \frac{\partial F_{\theta,t}(x)}{\partial \theta} \frac{\partial F_{\theta,t-j}(y)}{\partial \theta'} \\ & - \frac{\partial F_{\theta,t-j}(y)}{\partial \theta} \frac{\partial F_{\theta,t}(x)}{\partial \theta'} - F_{\theta,t}(x) \frac{\partial^2 F_{\theta,t-j}(y)}{\partial \theta \partial \theta'}\end{aligned}$$

Hessian matrix at $\theta = \theta_0$:

$$\begin{aligned}
&\frac{\partial^2}{\partial \theta \partial \theta'} \mathcal{Q}_T(\theta_0) \\
=& 2 \sum_{j=1}^{T-1} (1 - \frac{j}{T}) \frac{1}{(j\pi)^2} \int_{\mathcal{R}^2} \left(\frac{\partial \sigma_{\theta_0,j}(x,y)}{\partial \theta} \frac{\partial \sigma_{\theta_0,j}(x,y)}{\partial \theta'} + \sigma_{\theta_0,j}(x,y) \frac{\partial^2 \sigma_{\theta_0,j}(x,y)}{\partial \theta \partial \theta'} \right) dW(x,y) \\
=& 2 \sum_{j=1}^{T-1} (1 - \frac{j}{T}) \frac{1}{(j\pi)^2} \int_{\mathcal{R}^2} \left(\frac{\partial \sigma_{\theta_0,j}(x,y)}{\partial \theta} \frac{\partial \sigma_{\theta_0,j}(x,y)}{\partial \theta'} \right) dW(x,y) \\
=& 2 \sum_{j=1}^{T-1} (1 - \frac{j}{T}) \frac{1}{(j\pi)^2} \left(\phi_j^{(1)}(\theta_0) \phi_j^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f^2(x) (\int_{-\infty}^{y} f(u) u du)^2 \right) dW(x,y) \\
&+ 2 \sum_{j=1}^{T-1} (1 - \frac{j}{T}) \frac{1}{(j\pi)^2} \left(\phi_{-j}^{(1)}(\theta_0) \phi_{-j}^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f^2(y) (\int_{-\infty}^{x} f(u) u du)^2 \right) dW(x,y) \\
&+ 2 \sum_{j=1}^{T-1} (1 - \frac{j}{T}) \frac{1}{(j\pi)^2} \left(\phi_j^{(1)}(\theta_0) \phi_{-j}^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f(x) f(y) (\int_{-\infty}^{y} f(u) u du) (\int_{-\infty}^{x} f(u) u du) \right) dW(x,y) \\
&+ 2 \sum_{j=1}^{T-1} (1 - \frac{j}{T}) \frac{1}{(j\pi)^2} \left(\phi_{-j}^{(1)}(\theta_0) \phi_j^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f(x) f(y) (\int_{-\infty}^{y} f(u) u du) (\int_{-\infty}^{x} f(u) u du) \right) dW(x,y) \\
=& 2 \sum_{j=1}^{\infty} \frac{1}{(j\pi)^2} \left(\phi_j^{(1)}(\theta_0) \phi_j^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f^2(x) (\int_{-\infty}^{y} f(u) u du)^2 \right) dW(x,y) \\
&+ 2 \sum_{j=1}^{\infty} \frac{1}{(j\pi)^2} \left(\phi_{-j}^{(1)}(\theta_0) \phi_{-j}^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f^2(y) (\int_{-\infty}^{x} f(u) u du)^2 \right) dW(x,y) \\
&+ 2 \sum_{j=1}^{\infty} \frac{1}{(j\pi)^2} \left(\phi_j^{(1)}(\theta_0) \phi_{-j}^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f(x) f(y) (\int_{-\infty}^{y} f(u) u du) (\int_{-\infty}^{x} f(u) u du) \right) dW(x,y) \\
&+ 2 \sum_{j=1}^{\infty} \frac{1}{(j\pi)^2} \left(\phi_{-j}^{(1)}(\theta_0) \phi_j^{(1)}(\theta_0)' \right) \int_{\mathcal{R}^2} \left(f(x) f(y) (\int_{-\infty}^{y} f(u) u du) (\int_{-\infty}^{x} f(u) u du) \right) dW(x,y) + o(1)
\end{aligned}$$

8.2 Calculation in Simulation

The objective function in Monte Carlo experiments can be computed as follows:

$$\hat{\mathcal{Q}}_T(\theta) = 2 \int_{\Gamma} \sum_{j=1}^{T-1} \hat{\sigma}_{\theta,j}^2(x,y) (1 - \frac{j}{T}) \frac{1}{(j\pi)^2} dW(x,y)$$

where

$$\hat{\sigma}_{\theta,j}^2(x,y) = \hat{F}_{\theta,j}^2(x,y) - 2\hat{F}_{\theta,j}(x,y)\hat{F}_{\theta,j}(x,\infty)\hat{F}_{\theta,j}(\infty,y) + \hat{F}_{\theta,j}^2(x,\infty)\hat{F}_{\theta,j}^2(\infty,y),$$

$$\hat{F}_{\theta,j}(x,y) = \frac{1}{T-j}\sum_{t=j+1}^{T} I(u_t(\theta) \le x) I(u_{t-j}(\theta) \le y).$$

The integration of $\hat{\sigma}^2_{\theta,j}(x,y)$ over weighting function W can be decomposed into following three components:

$$\begin{aligned}
&\int \hat{F}^2_{\theta,j}(x,y) dW(x,y) \\
=&\iint \frac{1}{(T-j)^2}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T} I(u_t(\theta) \le x) I(u_{t-j}(\theta) \le y) I(u_s(\theta) \le x) I(u_{s-j}(\theta) \le y) dW(x) dW(y) \\
=&\frac{1}{(T-j)^2}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T} \int I(u_t(\theta) \le x) I(u_s(\theta) \le x) dW(x) \int I(u_{t-j}(\theta) \le y) I(u_{s-j}(\theta) \le y) dW(y) \\
=&\frac{1}{(T-j)^2}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T} \int_{\max\{u_t(\theta),u_s(\theta)\}}^{\infty} dW(x) \int_{\max\{u_{t-j}(\theta),u_{s-j}(\theta)\}}^{\infty} dW(y) \\
=&\frac{1}{(T-j)^2}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T} \left(1 - W\left(\max\{u_t(\theta), u_s(\theta)\}\right)\right)\left(1 - W\left(\max\{u_{t-j}(\theta), u_{s-j}(\theta)\}\right)\right)
\end{aligned}$$

where W is set to be a continuous probability measure.
Similarly,

$$\begin{aligned}
&\int \hat{F}_{\theta,j}(x,y) \hat{F}_{\theta,j}(x,\infty) \hat{F}_{\theta,j}(\infty,y) dW(x,y) \\
=&\iint \frac{1}{(T-j)^3}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T}\sum_{r=j+1}^{T} I(u_t(\theta) \le x) I(u_{t-j}(\theta) \le y) I(u_s(\theta) \le x) I(u_{r-j}(\theta) \le y) dW(x) dW(y) \\
=&\frac{1}{(T-j)^3}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T}\sum_{r=j+1}^{T} \int I(u_t(\theta) \le x) I(u_s(\theta) \le x) dW(x) \int I(u_{t-j}(\theta) \le y) I(u_{r-j}(\theta) \le y) dW(y) \\
=&\frac{1}{(T-j)^3}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T}\sum_{r=j+1}^{T} \left(1 - W\left(\max\{u_t(\theta), u_s(\theta)\}\right)\right)\left(1 - W\left(\max\{u_{t-j}(\theta), u_{r-j}(\theta)\}\right)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\int \hat{F}^2_{\theta,j}(x,\infty) \hat{F}^2_{\theta,j}(\infty,y) dW(x,y) \\
=&\frac{1}{(T-j)^4}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T}\sum_{m=j+1}^{T}\sum_{n=j+1}^{T} \iint I(u_t(\theta) \le x) I(u_s(\theta) \le x) I(u_{m-j}(\theta) \le y) I(u_{n-j}(\theta) \le y) dW(x,y) \\
=&\frac{1}{(T-j)^4}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T}\sum_{m=j+1}^{T}\sum_{n=j+1}^{T} \left(1 - W\left(\max\{u_t(\theta), u_s(\theta)\}\right)\right)\left(1 - W\left(\max\{u_{m-j}(\theta), u_{n-j}(\theta)\}\right)\right)
\end{aligned}$$