Local GMM estimation for nonparametric time-varying coefficient moment condition models

Yu Bai[1a]

[a]*Monash University*

Abstract

I develop a local continuously updated GMM estimator for nonparametric time-varying coefficient moment condition models. The uniform consistency rate and the pointwise asymptotic normality of the proposed estimator are derived. The finite sample performance of the proposed estimator is investigated through a Monte Carlo study and an empirical application on asset pricing models with stochastic discount factor (SDF) representation.

Keywords: Continuously Updated GMM, Finite Sample Performance, Stochastic Discount Factor

JEL classification: C10, C13, C14

[1]Corresponding to: Yu Bai, Department of Econometrics and Business Statistics, Monash University, 900 Dandenong Rd, Caulfield East, VIC 3145. Email: yu.bai1@monash.edu.

1. Introduction

Economic theory often implies a set of moment conditions. Since the seminal work by Hansen (1982), the Generalized Method of Moments (GMM) has been widely used in estimation, testing and application of moment condition models. Hall (2005) provides an excellent early review on both methodological development and empirical applications of GMM method.

As parameter instability is pervasive (Stock and Watson (1996)), attempts have been made to handle structural change in the GMM framework. Most of the existing literature focuses on the case in which parameters are assumed to have breakpoints, possibly at unknown dates (Ghysels and Hall (1990), Andrews (1993), Ghysels et al. (1998), Hall et al. (2015)). However, as Hansen (2001) points out, "*it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect*." Indeed, all leading driving forces of structural change, such as technological improvements, climate change, and modifications in the institutional context, may take time to manifest their effects and thus change the parameters of econometric relationships.

In this paper, I develop a local continuously updated GMM (CU-GMM) estimation and inferential theory for nonparametric time-varying coefficient moment condition models. The method extends the original CU-GMM estimator, which is first introduced in Hansen et al. (1996), to the setting in which model parameters are time-varying. Following Cai (2007), Chen and Hong (2012), Giraitis et al. (2014), Giraitis et al. (2016), Giraitis et al. (2021) and Li et al. (2021), I take a nonparametric approach to remain agnostic as possible on the types of time variation. I show the consistency, derive the uniform consistency rate and pointwise asymptotic normality of the proposed estimator. The uniform consistency rate depends on the roughness of the time variation. The pointwise asymptotic variance of the estimator has a simple form, which makes inference straightforward.

The finite-sample performance of the estimator is evaluated by a Monte-Carlo study. Using a MA(1) model with time-varying parameters as data generating process, I find that local CU-GMM estimator has satisfactory finite sample performance. To illustrate in practice the use of the estimator, I provide an empirical application on asset pricing models with Stochastic Discount Factor (SDF) representation. I consider the problem of explaining the size and book-to-market anomalies using the three-factor model as in Fama and French (1993). I find evidence of time variation in SDF model parameters and allowing for time variation improves pricing performance.

The paper is organized as follows. Section 2 describes the framework, the local CU-GMM estimator and presents the main theoretical results. Section 3 illustrates the proposed estimator in a Monte Carlo study and an empirical application. Section 4 concludes the paper. The mathematical proofs are presented in the appendix.

2. Theoretical considerations

2.1. The estimator

Consider the moment conditions at time t:

$$E\left[g(x_t;\theta_{0,t})\right] = 0, \tag{1}$$

where $g : \mathbb{R}^q \times \Theta \to \mathbb{R}^m$ is a vector-valued function, $x_t \in \mathbb{R}^q$ is the data vector, and $\theta_{0,t}$ is the $k \times 1$ vector of true time-varying parameters belonging to the parameter space Θ. The moment function $g(\cdot)$ can be nonlinear in $\theta_{0,t}$.

Let $g_t(\theta_t) = g(x_t;\theta_t)$. GMM criteria are defined as the quadratic forms in terms of the moment indicator $g_t(\theta_t)$ $(t = 1, 2, \cdots, T)$, and their sample average $\overline{g}_{T,t}(\theta_t)$. The local continuous updated GMM (CU-GMM) estimator takes the form

$$\hat{\theta}_t = \arg\min_{\theta_t \in \Theta} Q_{t,T}, \tag{2}$$

where the objective function $Q_{t,T}$ is given by

$$Q_{t,T} = \overline{g}'_{T,t}(\theta_t) W^{-1}_{T,t}(\theta_t) \overline{g}_{T,t}(\theta_t). \tag{3}$$

$\overline{g}_{T,t}(\theta_t)$ is the local average of sample moments:

$$\overline{g}_{T,t}(\theta_t) = \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_t), \tag{4}$$

where $K_t = \sum_{j=1}^{T} k_{jt}$. The kernel weights $k_{jt} = K\left(\frac{j-t}{Tb}\right)$ are computed with a kernel function $K(\cdot)$ and b is a bandwidth parameter. $W_{T,t}(\theta_t)$ is an optimal weight matrix to estimate the parameters in (1), which shall be a consistent estimator of the long-run covariance matrix

$$\lim_{T\to\infty} \operatorname{Var}\left(\frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} b_{jt} g_j(\theta_{0,t})\right),$$

where $K_{2,t} = \sum_{j=1}^{T} b_{jt}^2$.

2.2. Asymptotic theory

To establish the consistency and derive asymptotic distribution of (2), the following regularity conditions are imposed.

Assumption 2.1. *(Data and moments).*

(i) $g_t(\cdot) : \mathbb{R}^q \times \Theta \to \mathbb{R}^m$ *is a known vector-valued function,* $\Theta \subset \mathbb{R}^k$ *is compact,* $\theta_{0,t} \in int(\Theta)$ *and* q, k, m *are all finite;*

(ii) $g_t(\theta)$ *is second order continuously differentiable w.r.t.* $\theta \in \Theta$ *whose elements* $(g_{\ell,t}(\theta), \ell = 1, 2, \cdots, m)$ *and their associated first* $(\frac{\partial g_{\ell,j}(\theta)}{\partial \theta_{\ell_2}}, \ell_2, = 1, 2, \cdots, k)$ *and second order derivatives* $(\frac{\partial^2 g_{\ell,j}(\theta)}{\partial \theta_{\ell_2}^2}, \ell_2, = 1, 2, \cdots, k)$ *satisfy the following:* $\exists \delta > 0$*, such that*

$$\max_{\theta\in\Theta} \max_{1\leqslant t\leqslant T} \left|g_{\ell,t}(\theta)\right|_{4+\delta} < \infty, \quad \max_{\theta\in\Theta} \max_{1\leqslant t\leqslant T} \left|\frac{\partial g_{\ell,t}(\theta)}{\partial \theta_{\ell_2}}\right|_{4+\delta} < \infty, \max_{\theta\in\Theta} \max_{1\leqslant t\leqslant T} \left|\frac{\partial^2 g_{\ell,t}(\theta)}{\partial \theta_{\ell_2}^2}\right|_{4+\delta} < \infty,$$

$\forall \ell = 1, 2, \cdots, k$, $\forall \ell_2 = 1, 2, \cdots, k$;

(iii) $x_t = x_{T,t}$ *is a triangular array of vectors whose elements* $x_{\ell,t}$ *is* α*-mixing with mixing coefficient* $\alpha(h) = O(h^{-\phi})$ *with* $\phi = \frac{(4+\delta)(1+\delta)}{2+\delta}$.

Assumption 2.2. *(Identification). For each t,* $E[g_t(\theta)] \neq 0$, $\forall \theta \in \Theta$ *such that* $\theta \neq \theta_{0,t}$.

Assumption 2.3. *(Smoothness). For given* $\theta \in \Theta$*, define* $E[g_t(\theta)] = \mu(\frac{t}{T})$, $E[g_t(\theta)g_t'(\theta)] = \Sigma(\frac{t}{T})$ *and* $E[\frac{\partial g_t(\theta)}{\partial \theta}] = \mu^d(\frac{t}{T})$.

(i) Elements in $\mu(\cdot)$ *are Hölder continuous with exponent* γ*: for any* $u, v \in (0, 1]$*, given* ℓ*,*

$$|\mu_\ell(u) - \mu_\ell(v)| \leqslant C|u - v|^\gamma,$$

for some $C > 0$ *and* $0 < \gamma \leqslant 1$.

(ii) Elements in $\Sigma(\cdot)$ *are Hölder continuous with exponent* γ'*: for any* $u, v \in (0, 1]$*, given* (ℓ_1, ℓ_2)*,*

$$\left|\Sigma_{\ell_1,\ell_2}(u) - \Sigma_{\ell_1,\ell_2}(v)\right| \leqslant C|u - v|^{\gamma'},$$

for some $C > 0$ *and* $0 < \gamma' \leqslant 1$.

(iii) Elements in $\mu^d(\cdot)$ *are Hölder continuous with exponent* γ''*: for any* $u, v \in (0, 1]$*, given* (ℓ_3, ℓ_4)*,*

$$\left|\mu^d_{\ell_3,\ell_4}(u) - \mu^d_{\ell_3,\ell_4}(v)\right| \leqslant C|u - v|^{\gamma''},$$

for some $C > 0$ *and* $0 < \gamma'' \leqslant 1$.

Assumption 2.4. *(Kernel and bandwidth).*

(i) The kernel weights $k_{jt} := K\left(\frac{j-t}{Tb}\right)$ are computed with a kernel function $K(\cdot) : \mathbb{R} \longrightarrow \mathbb{R}^+$, such that for some $C > 0, \nu > 3$,

$$K(x) \leqslant C(1 + x^\nu)^{-1}, \ |(d/dx)K(x)| \leqslant C(1 + x^\nu)^{-1}, \ \int K(x)dx = 1;$$

(ii) b is a bandwidth parameter such that $b \to 0$, $Tb \to \infty$ and $b = O(T^{\frac{2}{p-2+\delta'}-1})$ for some $\delta' > 0$ and $p = 4 + \delta > 4$ as in Assumption 2.1(ii).

Assumption 2.1 imposes conditions on data and moment functions. It is required that L_p norm for the moment functions and associated first and second order derivatives to be uniformly bounded with $p > 4$. The mixing coefficient is required to decay at a polynomial rate with $\phi > 2$. It follows immediately from Theorem 14.1 in Davidson (1994) that $\{g_t\}$ is also α-mixing with mixing coefficient bounded by $\alpha(h)$. Assumption 2.2 is a standard condition to ensure global identification for each $\theta_{0,t}$.

Assumption 2.3 imposes smoothness conditions for $E[g_t(\theta)]$, $E[g_t(\theta)g_t'(\theta)]$ and $E[\frac{\partial g_t(\theta)}{\partial \theta}]$. In particular, Assumption 2.3(i) is indirectly related to the smoothness of time-varying parameters. To see this, consider first the MA(1) model with time-varying parameters: $y_t = \epsilon_t + \theta_t \epsilon_{t-1}$, where ϵ_t has zero mean and unit variance. Suppose that two moments $E(y_t^2) = 1 + \theta_t^2$, $E(y_t y_{t-1}) = \theta_t$ are used to estimate the parameter. Then, it is quite straightforward to see that the continuity condition on $\mu(\cdot)$ implies that $\left|\theta_t - \theta_j\right| \leqslant C(\frac{|t-j|}{T})^\gamma$. For the empirical application in section 3.2, when f_t is a scalar, if $E(R_t) = \mu^R(t/T)$ and $E(R_t f_t') = \mu^{Rf}(t/T)$ are Hölder continuous with exponent γ, the assumption is satisfied if $\left|\lambda_{0t} - \lambda_{0j}\right| \leqslant C(\frac{|t-j|}{T})^\gamma$ and $\left|\lambda_{1t} - \lambda_{1j}\right| \leqslant C(\frac{|t-j|}{T})^\gamma$.

In a linear instrumental variable (IV) framework, Giraitis et al. (2021) have considered two special cases when $\gamma = 1$ and $\gamma = \frac{1}{2}$. When $\gamma = 1$, the condition implies that $\theta_t = \theta(t/T)$ is Lipschitz continuous on $(0, 1]$. The case analyzed in Li et al. (2021) when $\theta_t = \theta(t/T)$ is twice continuously differentiable is a special case when $\gamma = 1$. When $\gamma = \frac{1}{2}$, $\theta_t = \theta(t/T)$ is a Hölder continuous function with exponent $1/2$. Consider the case when $\overline{\theta}_t$ is generated from a persistent and bounded stochastic process $\frac{1}{\sqrt{T}} u_t$, $t = 1, 2, \cdots, T$, where u_t are such that $u_t - u_{t-1} = \nu_t$, where ν_t is an *i.i.d.* process. As shown in Dendramis et al. (2021), such a process satisfies the condition $\left|\overline{\theta}_t - \overline{\theta}_s\right| \leqslant \xi_{ts}\left(\frac{|t-s|}{T}\right)^{1/2}$, where ξ_{ts} has a thin-tailed distribution: $\mathbb{P}\Big(|\xi_{ts}| > \omega\Big) \leqslant \exp\Big(-c_0|\omega|^\alpha\Big)$, $\omega > 0$, for some $c_0 > 0$, $\alpha > 0$, which does not depend on t, s and T. Then, for a realization of the process like θ_t, one could always find a generic C such that the Hölder continuous condition is satisfied. As γ approaches 0, $\theta_t = \theta(t/T)$ becomes rough. Assumptions 2.3(ii)-(iii) are smoothness conditions for the derivative and second order moments, which are needed to establish Theorem 1. Assumption 2.4 imposes conditions on kernel function and bandwidth parameter. Examples of kernel functions satisfying this assumption include $K(x) = \frac{1}{2}I\{|x| \leqslant 1\}$ (uniform kernel), $K(x) = \frac{3}{4}(1 - x^2)I\{|x| \leqslant 1\}$ (Epanechnikov kernel) and $K(x) \propto \exp(-cx^\alpha)$ with $c > 0$, $\alpha > 0$ (exponential-type kernel).

The large sample properties of consistency and pointwise asymptotic normality of local CU-GMM estimator defined in (2) rely on a law of large numbers (WLLN) and central limit theorem (CLT) for (4). These are provided in the following lemma.

Lemma 1. (i) WLLN: under Assumptions 2.1, 2.2, and 2.4,

$$\max_{\theta_t \in \Theta} \left\| \overline{g}_{T,t}(\theta_t) - E(g_t(\theta_t)) \right\| = O_p\Big((Tb)^{-1/2} + b^{\gamma}\Big), \tag{5}$$

$$\max_{1 \leqslant t \leqslant T} \left\| \overline{g}_{T,t}(\theta_t) - E(g_t(\theta_t)) \right\| = O_p\Big((Tb)^{-1/2}\sqrt{\log T} + b^{\gamma}\Big), \tag{6}$$

where $0 < \gamma \leqslant 1$;

(ii) CLT: under Assumptions 2.1, 2.2, 2.3(i)-(ii) and 2.4, suppose that $T^{\frac{1}{2}} b^{\frac{1}{2}+\gamma} \to 0$,

$$\frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \xrightarrow{d} \mathcal{N}(0, W_t),$$

where $W_t = \mathrm{Var}\left(g_t(\theta_{0,t})\right)$.

Lemma 1(i) provides uniform WLLN results for (4). In particular, the uniform WLLN (over Θ) ((5)) result is a generalization of the pointwise results presented in Corollary 6 in Dendramis et al. (2021), which is needed to prove consistency of the estimator. Lemma 1(ii) provides pointwise asymptotic normality result. It is shown that, under Assumption 2.1(iii), provided that mixing coefficient decays fast enough (Assumption 2.1(iii)), covariance terms of (4) vanishes asymptotically. This also implies that optimal weighting matrix $W_{T,t}(\theta_t)$ takes the form

$$W_{T,t}(\theta_t) = \frac{1}{K_{2,t}} \sum_{j=1}^{T} k_{jt}^2 g_j(\theta_t) g_j'(\theta_t), \tag{7}$$

where $K_{2,t} = \sum_{j=1}^{T} k_{jt}^2$.

Theorem 1 establishes the uniform consistency and (pointwise) asymptotic normality for the local CU-GMM estimator $\hat{\theta}_t$.

Theorem 1. (i) Consistency: under Assumptions 2.1, 2.2, 2.3(i) and 2.4,

$$\hat{\theta}_t \xrightarrow{p} \theta_{0,t} \quad \text{for each } t,$$

$$\max_{1 \leqslant t \leqslant T} \left\| \hat{\theta}_t - \theta_{0,t} \right\| = O_p\Big((b^{\gamma} + \sqrt{\frac{\log T}{Tb}}\Big).$$

(ii) Asymptotic normality: under Assumptions 2.1-2.4, suppose that $T^{\frac{1}{2}}b^{\frac{1}{2}+\gamma} \to 0$,

$$\frac{K_t}{K_{2,t}^{1/2}} (G_t' W_t^{-1} G_t)^{1/2} (\hat{\theta}_t - \theta_{0,t}) \xrightarrow{d} \mathcal{N}(0, I_k).$$

where $K_t = \sum_{j=1}^{T} k_{jt}$, $K_{2,t} = \sum_{j=1}^{T} k_{jt}^2$. G_t and W_t are given by

$$G_t = E\Big[\frac{\partial g_t(\theta_{0,t})}{\partial \theta'}\Big], \quad W_t = \mathrm{Var}\,(g_t(\theta_{0,t})).$$

Several comments are in order. First, the uniform consistency rate is related to the degree of smoothness γ. Clearly, when γ becomes lower, the convergence rate gets distorted. Second, since $K_t = O(Tb)$, $K_{2,t} = O(Tb)$, the convergence rate of our estimator is proportional to the square root of Tb. The condition $T^{\frac{1}{2}}b^{\frac{1}{2}+\gamma} \to 0$ is to ensure that the smoothing bias term introduced by the nonparametric estimator vanishes asymptotically. Thus, CLT can be applied to the dominating term to have a proper asymptotic distribution of the estimator. Finally, the asymptotic variance-covariance matrix of the estimator only involves the variance term and covariance term drops out asymptotically. This makes pointwise inference quite straightforward.

The above inference results become operational if $\theta_{0,t}$ in both G_t and W_t are replaced by estimated counterparts. This is provided in the following corollary.

Corollary 1. *Under Assumptions 2.1, 2.2, 2.3(i) and 2.4, let*

$$\hat{G}_{T,t} = \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} \frac{\partial g_j(\hat{\theta}_t)}{\partial \theta'}, \quad \hat{W}_{T,t} = \frac{1}{K_{2,t}} \sum_{j=1}^{T} k_{jt}^2 \tilde{g}_j(\hat{\theta}_t) \tilde{g}_j'(\hat{\theta}_t),$$

where $K_t = \sum_{j=1}^{T} k_{jt}$, $K_{2,t} = \sum_{j=1}^{T} k_{jt}^2$ *and* $\tilde{g}_j(\hat{\theta}_t) = g_j(\hat{\theta}_t) - \frac{1}{T}\sum_{j=1}^{T} g_j(\hat{\theta}_t)$. *Then, it holds that*

$$\hat{G}_{T,t} \xrightarrow{p} G_t, \quad \hat{W}_{T,t} \xrightarrow{p} W_t.$$

3. Numerical studies

3.1. A simulated example

Consider a MA(1) model with time-varying parameters as the data generating process (DGP):

$$y_t = \epsilon_t + \theta_{0,t}\epsilon_{t-1}, \quad t = 1, 2, \cdots, T, \tag{8}$$

where the innovation ϵ_t is generated from the standard Normal distribution. For the time-varying parameter $\theta_{0,t}$, I consider three cases. In Case i, $\theta_{0,t}$ is generated according to $\theta_{0,t} = 0.4\cos(6\pi t/T) + 0.55$. Since $\theta_{0,t} = \theta(\frac{t}{T})$ is Lipschitz continuous in $(0, 1]$, this is the case when $\gamma = 1$. In Cases ii and iii, I first generate

a_t according to $a_t = a_{t-1} + v_t$, where $(1 - L)^{d-1} v_t = \epsilon_t$ and ϵ_t is the standard normal *i.i.d.* noise. Then, $\theta_{0,t}$ is generated according to $\theta_{0,t} = \rho a_t / \max_t |a_t|$, where $\rho = 0.9$. I set $d = 1$ in Case ii and $d = 0.75$ in Case iii. As explained in Giraitis et al. (2014), γ is equal to 0.5 in Case ii and 0.25 in Case iii. Two moment conditions ($E(y_t^2)$ and $E(y_t y_{t-1})$) are used to estimate the scalar parameter $\theta_{0,t}$ in each t.

The estimators are computed using the Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. Since in all three cases γ is known, I set the bandwidth according to $b = cT^{-\frac{1}{2\gamma+1}}$ to ensure a maximum allowable convergence rate of the estimator while the condition needed for CLT: $T^{\frac{1}{2}} b^{\frac{1}{2}+\gamma} \to 0$, is satisfied at the same time. I consider three sample sizes: $T = 250, 500, 1000$, with $R = 500$ as the burn-in period. The Monte Carlo analysis is based on 1,000 replications.

The performance of the estimators is evaluated by the root mean squared error (RMSE): RMSFE $= \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{\theta}_t - \theta_t)^2}$ and the 95% coverage probability (CP), which is the estimated probability that the true θ_t lies in the interval $(\hat{\theta}_t - 1.96\text{sd}(\hat{\theta}_t), \hat{\theta}_t + 1.96\text{sd}(\hat{\theta}_t))$, where $\text{sd}(\hat{\theta}_t)$ is the estimated variance of the estimator obtained from the associated asymptotic distributions.

Table 1 reports RMSE and coverage rate of the estimator over various choices of c. A number of comments can be made. First, when $\{\theta_{0,t}\}$ is smooth (Case i), results can be sensitive to the choice of normalizing constant c. Larger values of c ($c \geqslant 1$) results in larger RMSFEs, and coverage probabilities are also very low. Second, when $\{\theta_{0,t}\}$ is less smooth (Cases ii and iii), RMSFEs are lower for larger values of c, but coverage probabilities are generally not sensitive to the choice of c. Finally, as predicted by Theorem 1, uniform consistency rate depends on γ, so RMSFEs are lower when $\{\theta_{0,t}\}$ becomes rough. When $T = 1000$, the lowest RMSFEs in Case i is 0.136, which increases to 0.14 in Case ii and further to 0.212 in Case iii.

3.2. An empirical application

Consider the asset pricing model with beta representation:

$$E\left[R_t\right] = 1_N \gamma_0 + \beta \phi$$

where R_t is a $N \times 1$ vector of gross return, f_t is a $k \times 1$ vector of pricing factors with mean μ_f and variance Ω_f, γ_0 is the zero-beta rate and γ_1 is the risk premia with $\phi = \gamma_1 - \mu_f$.

Since asset specific beta is defined as $\beta = E\left[R_t (f_t - \mu_f)'\right] \Omega_f^{-1}$, by plugging it back to the asset pricing equations and rearranging terms, we obtain the asset pricing model with stochastic discount factor (SDF) representation:

$$E\left[R_t (\lambda_0 + f_t' \lambda_1)\right] = 1_N, \tag{9}$$

where $\lambda_0 = \frac{1 + \mu_f' \Omega_f^{-1} \phi}{\gamma_0}$ and $\lambda_1 = -\frac{\Omega_f^{-1} \phi}{\gamma_0}$. The coefficients in SDF representation λ_0, λ_1 are likely to be time-varying if Ω_f is changing over time. For instance, consumption growth (a factor in consumption based asset pricing models) has been documented to exhibit time-varying second order moments (see, Bansal

Table 1: Small sample properties of the local CU-GMM estimator for $\theta_{0,t}$: Average RMSFE and coverage probability

Case		$T = 250$		$T = 500$		$T = 1000$	
	c	RMSFE	CP	RMSFE	CP	RMSFE	CP
i	0.4	0.198	0.87	0.160	0.88	0.136	0.88
	0.6	0.180	0.85	0.153	0.84	0.138	0.80
	8	0.185	0.81	0.169	0.75	0.158	0.65
	1	0.197	0.74	0.197	0.61	0.186	0.50
	1.2	0.216	0.65	0.220	0.50	0.215	0.39
	1.4	0.230	0.58	0.242	0.42	0.238	0.31
	1.6	0.245	0.52	0.259	0.36	0.257	0.27
ii	0.4	0.274	0.86	0.239	0.87	0.210	0.87
	0.6	0.241	0.86	0.211	0.87	0.185	0.86
	0.8	0.213	0.86	0.192	0.85	0.173	0.85
	1	0.202	0.85	0.181	0.85	0.164	0.84
	1.2	0.192	0.85	0.174	0.84	0.158	0.82
	1.4	0.186	0.84	0.168	0.83	0.155	0.80
	1.6	0.184	0.83	0.165	0.82	0.154	0.78
iii	0.4	0.339	0.84	0.340	0.84	0.311	0.85
	0.6	0.316	0.84	0.294	0.85	0.276	0.85
	0.8	0.296	0.85	0.278	0.85	0.254	0.84
	1	0.281	0.84	0.257	0.84	0.238	0.84
	1.2	0.260	0.83	0.243	0.83	0.227	0.83
	1.4	0.251	0.83	0.236	0.83	0.218	0.83
	1.6	0.249	0.83	0.226	0.83	0.212	0.82

Notes: The table reports RMSFE and 95% coverage probabilities (CP) from local CU-GMM estimator. The bandwidth is parametrized as $b = cT^{-\frac{1}{2\gamma+1}}$.

and Yaron (2004)). The market factor also contains the time-varying second order moments (see, Engle et al. (2013)).

I estimate the asset pricing model given in (9), but the coefficients are now allowed to be time-varying: $(\lambda_{0t}, \lambda_{1t})$. With regard to testing portfolios and pricing factors, I obtain 6 portfolios formed on Size and Book-to-Market from Ken French's online data library. Pricing factors used include the excess return on the value-weighted equity market portfolio (Mkt) from CRSP, the small minus big (SMB) portfolio and the high minus low (HML) portfolio from Fama and French (1993), which are again taken from French's website. The data is monthly and spans the period 1954:01-2023:7, for a total of 835 observations.

To implement the local CU-GMM estimation, a Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ is used. Since γ is unknown in the empirical application, I assume $\gamma = 1$ and set $b = cT^{-1/3}$ to ensure the maximum allowable convergence rate of the estimator. c is selected over a course grid of width 0.1 from 0.4 to 1.6 by minimizing the pricing error: $Q(\mathrm{M_{TVP}}) = e'(\mathrm{M_{TVP}})e(\mathrm{M_{TVP}})$, where $e(\mathrm{M_{TVP}}) = \frac{1}{T}\sum_{t=1}^{T} r_{it}(\hat{\lambda}_{0t} + f_t'\hat{\lambda}_{1t}) - 1$ for each portfolio $i = 1, 2, \cdots, 6$. c selected from this criteria is 1.4. Thus, the bandwidth parameter used to obtain final estimation results is $b = 1.4T^{-1/3}$. Figure 1 provides plots of time-varying estimates for λ_{1t} with the associated 95% pointwise confidence interval.

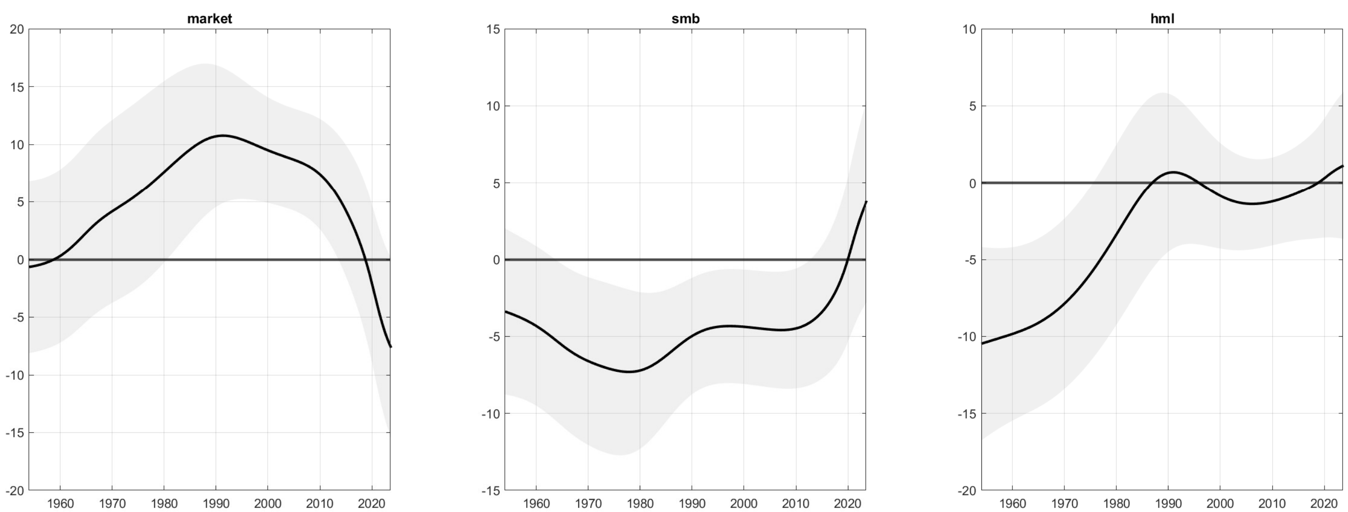

Figure 1: Time-varying estimates of λ_{1t} from three-factor model.

I now assess whether allowing for time variation leads to improvement in the pricing performance. As the Hansen-Jagannathan (HJ) distance (Hansen and Jagannathan (1997)) is widely used to evaluate the goodness of fit of an SDF model, I follow Cui et al. (2022) to compute the HJ-R^2:

$$\mathrm{HJ} - R^2 = 1 - \frac{Q(\mathrm{M_{TVP}})}{Q(\mathrm{M_C})},$$

where $Q(\mathrm{M_{TVP}}) = e'(\mathrm{M_{TVP}})e(\mathrm{M_{TVP}})$ is the pricing error from time-varying parameter models as defined above and $Q(\mathrm{M_C})$ can be computed similarly. A positive HJ-R^2 value indicates that model with time-varying parameters outperforms the constant coefficients benchmark. The value I obtain is 17.42%. Therefore, allowing for time variation in (9) leads to improvement in the pricing performance.

4. Conclusions

In this paper, I develop a local continuously updated GMM estimator for nonparametric time-varying coefficient moment condition models. The asymptotic properties of the proposed estimator are established. The uniform consistency rate depends on the roughness of the time variation. The pointwise asymptotic variance of the estimator has a simple form, which makes inference straightforward.

The finite sample properties of the estimator are evaluated in a Monte-Carlo study. The results show that performances from local CU-GMM estimator, in terms of both biases and coverage rates, are satisfactory. I then illustrate the methods by an empirical application on asset pricing models with SDF representation for the cross-section of equity portfolios. I find evidence of time variation in the SDF parameters. By allowing for time variation improves the pricing performance.

References

Andrews, D.W., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica: Journal of the Econometric Society , 821–856.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. The journal of Finance 59, 1481–1509.

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163–188.

Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica: Journal of the econometric society 80, 1157–1183.

Cui, L., Feng, G., Hong, Y., 2022. Regularized gmm for time-varying models with applications to asset pricing. International Economic Review, *forthcoming* .

Davidson, J., 1994. Stochastic limit theory: An introduction for econometricians. OUP Oxford.

Dendramis, Y., Giraitis, L., Kapetanios, G., 2021. Estimation of time-varying covariance matrices for large datasets. Econometric Theory 37, 1100–1134.

Engle, R.F., Ghysels, E., Sohn, B., 2013. Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics 95, 776–797.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of financial economics 33, 3–56.

Ghysels, E., Guay, A., Hall, A., 1998. Predictive tests for structural change with unknown breakpoint. Journal of Econometrics 82, 209–233.

Ghysels, E., Hall, A., 1990. A test for structural stability of euler conditions parameters estimated via the generalized method of moments estimator. International Economic Review , 355–364.

Giraitis, L., Kapetanios, G., Marcellino, M., 2021. Time-varying instrumental variable estimation. Journal of Econometrics 224, 394–415.

Giraitis, L., Kapetanios, G., Wetherilt, A., Žikeš, F., 2016. Estimating the dynamics and persistence of financial networks, with an application to the sterling money market. Journal of Applied Econometrics 31, 58–84.

Giraitis, L., Kapetanios, G., Yates, T., 2014. Inference on stochastic time-varying coefficient models. Journal of Econometrics 179, 46–65.

Hall, A.R., 2005. Generalized method of moments. OUP Oxford.

Hall, A.R., Li, Y., Orme, C.D., Sinko, A., 2015. Testing for structural instability in moment restriction models: an info-metric approach. Econometric Reviews 34, 286–327.

Hall, P., Heyde, C.C., 1980. Martingale limit theory and its application. Academic press.

Hansen, B.E., 2001. The new econometrics of structural change: dating breaks in us labour productivity. Journal of Economic perspectives 15, 117–128.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica: Journal of the econometric society , 1029–1054.

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative gmm estimators. Journal of Business & Economic Statistics 14, 262–280.

Hansen, L.P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. The Journal of Finance 52, 557–590.

Li, H., Zhou, J., Hong, Y., 2021. Estimating and testing for smooth structural changes in moment condition models. Mimeo .

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

Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business & Economic Statistics 14, 11–30.

In the proofs, the following properties of the weights b_{jt} (See (6.16) in Giraitis et al. (2014)) are repeated used: as $b \to 0$, $Tb \to \infty$,

$$K_t = \sum_{j=1}^{T} k_{jt} = O(Tb), \quad K_{2,t} = \sum_{j=1}^{T} k_{jt}^2 = O(Tb).$$

Appendix A. Proof of main results

Appendix A.1. Proof of Lemma 1

Proof of (i). By Triangular inequality,

$$\begin{aligned}\left\|\frac{1}{K_t}\sum_{j=1}^{T} k_{jt}\Big(g_j(\theta_t) - E(g_t(\theta_t))\Big)\right\| &\leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} k_{jt}\Big(g_j(\theta_t) - E(g_j(\theta_t))\Big)\right\| \\ &\quad + \left\|\frac{1}{K_t}\sum_{j=1}^{T} k_{jt}\Big(E(g_j(\theta_t)) - E(g_t(\theta_t))\Big)\right\| \\ &= M_{T,t}^{(1)}(\theta_t) + M_{T,t}^{(2)}(\theta_t).\end{aligned}$$

(6) can be obtained using

$$\max_t \left\|\overline{g}_{T,t}(\theta_t)\right\| \leqslant \max_t \left\|M_{T,t}^{(1)}(\theta_t)\right\| + \max_t \left\|M_{T,t}^{(2)}(\theta_t)\right\|.$$

It follows immediately from Lemma B1(1b) that, for any $\varepsilon > 0$, $p > 2$,

$$\max_t \left\|M_{T,t}^{(1)}(\theta_t)\right\| = O_p((Tb)^{-1/2}\log^{1/2} T + (T^2 b)^{1/p}(Tb)^{\varepsilon-1}).$$

Under Assumption 2.4(ii), for sufficiently small $\varepsilon > 0$, it holds $(T^2b)^{1/p}(Tb)^{\varepsilon-1} \leqslant (Tb)^{-1/2}$, this implies that

$$\max_t \left\|M_{T,t}^{(1)}(\theta_t)\right\| = O_p((Tb)^{-1/2}\sqrt{\log T}).$$

By Assumption 2.1(iv) and 2.4(i), it follows from Lemma B1(2b) that (by setting $r = 1$)

$$\max_t \left\|M_{T,t}^{(2)}(\theta_t)\right\| = O_p(b^{\gamma}).$$

This completes the proof of (6). (5) can be obtained similarly by first noticing that

$$\max_{\theta_t\in\Theta} \left\|\overline{g}_{T,t}(\theta_t)\right\| \leqslant \max_{\theta_t\in\Theta} \left\|M_{T,t}^{(1)}(\theta_t)\right\| + \max_{\theta_t\in\Theta} \left\|M_{T,t}^{(2)}(\theta_t)\right\|.$$

Then, the results follow from Lemma B1(1a) and Lemma B1(2a).

Proof of (ii). We first show that $W_t = \mathrm{Var}\left(g_t(\theta_{0,t})\right) + o(1)$. Write

$$\mathrm{Var}\left(\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})\right) = \frac{1}{K_{2,t}}\sum_{j=1}^{T} k_{jt}^2 \mathrm{Var}\left(g_j(\theta_{0,t})\right) + \frac{2}{K_{2,t}}\sum_{k'=2}^{T}\sum_{j=1}^{k'-1} k_{k't}k_{jt}\mathrm{cov}(g_{k'}(\theta_{0,t}), g_j(\theta_{0,t})).$$

By Assumption 2.2(ii), it is easy to verify that

$$\frac{1}{K_{2,t}}\sum_{j=1}^{T} k_{jt}^2 \mathrm{Var}\left(g_j(\theta_{0,t})\right) \leqslant \left\{\max_j \ \mathrm{Var}\left(g_j(\theta_{0,t})\right)\right\}\frac{1}{K_{2,t}}\sum_{j=1}^{T} k_{jt}^2 = O(1).$$

We need to show that

$$\frac{1}{K_{2,t}}\sum_{k'=2}^{T}\sum_{j=1}^{k'-1} k_{k't}k_{jt}\mathrm{cov}(g_{k'}(\theta_{0,t}), g_j(\theta_{0,t})) = o(1).$$

By Corollary A.2 in Hall and Heyde (1980) with $p = q = 4 + \delta$ for some $\delta > 0$, we have

$$\begin{aligned}
\left\|\frac{1}{K_{2,t}}\sum_{k'=2}^{T}\sum_{j=1}^{k'-1} k_{k't}k_{jt}\mathrm{cov}(g_{k'}(\theta_{0,t}), g_j(\theta_{0,t}))\right\| &\leqslant \frac{1}{K_{2,t}}\sum_{k'=2}^{T}\sum_{j=1}^{k'-1} k_{k't}k_{jt}\left\|\mathrm{cov}(g_{k'}(\theta_{0,t}), g_j(\theta_{0,t}))\right\| \\
&\leqslant \frac{C}{K_{2,t}}\sum_{k'}\sum_{r} K\Big(\frac{k'-t}{Tb}\Big)K\Big(\frac{k'+r-t}{Tb}\Big)\alpha^{\frac{2+\delta}{4+\delta}}(r) \\
&\leqslant \frac{C}{(Tb)^{1+\delta}}\sum_{r} K\Big(\frac{k'+r-t}{Tb}\Big)r^{-(1+\delta)} = O((Tb)^{-\delta}) = o(1).
\end{aligned}$$

Define $\mathcal{F}_{j-1} = \sigma(g_s, s \leqslant j-1)$. Write

$$\begin{aligned}
\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) &= \frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} k_{jt}\Big(g_j(\theta_{0,t}) - E(g_j(\theta_{0,t})|\mathcal{F}_{j-1}) + E(g_j(\theta_{0,t})|\mathcal{F}_{j-1})\Big) \\
&= \frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} k_{jt} g_j^*(\theta_{0,t}) + \frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} k_{jt} E(g_j(\theta_{0,t})|\mathcal{F}_{j-1}).
\end{aligned}$$

In view of Cramer-Wold device, it is sufficient to show that, for $a \in \mathbb{R}^m$, where $\|a\| = 1$, the following holds

$$\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} k_{jt} a' g_j^*(\theta_{0,t}) \xrightarrow{d} \mathcal{N}(0, a'W_t a), \tag{A.1}$$

$$\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} k_{jt} E(g_j(\theta_{0,t})|\mathcal{F}_{j-1}) = O_p\Big(T^{\frac{1}{2}} b^{\frac{1}{2}+\gamma}\Big), \tag{A.2}$$

Proof of (A.1). Since the analysis is based on fixed $\theta_{0,t}$, we shall drop this for notation simplicity. Define

$V_j = \text{Var}\,(g_j(\theta_{0,t})|\mathcal{F}_{j-1})$ and $\xi_{tj} = K_{2,t}^{-1/2} a' g_j^* V_j^{-1/2}$. Since $\{g_t^*(\theta_{0,t})\}_t$ is a martingale difference sequence (M.D.S.), by the central limit theorem (CLT) for M.D.S (see Corollary 3.1 in Hall and Heyde (1980)), we need to verify that, for any $\epsilon > 0$,

$$j_T := \sum_{j=1}^{T} k_{jt}^2 E[\xi_{tj}^2|\mathcal{F}_{j-1}] \xrightarrow{p} a'a \tag{A.3}$$

$$\nu_{T,\epsilon} := \sum_{j=1}^{T} E[k_{jt}^2 \xi_{tj}^2 \mathbb{I}(k_{jt}^2 \xi_{tj}^2 > \epsilon)|\mathcal{F}_{j-1}] \xrightarrow{p} 0. \tag{A.4}$$

First, notice that

$$j_T = \sum_{j=1}^{T} k_{jt}^2 E[\xi_{tj}^2|\mathcal{F}_{j-1}] = \|a\|^2 \frac{1}{K_{2,t}} \sum_{j=1}^{T} k_{jt}^2 E[(g_j^*)^2|\mathcal{F}_{j-1}] V_j^{-1} = \|a\|^2.$$

This verifies (A.3), since $E[(g_j^*)^2|\mathcal{F}_{j-1}] = \text{Var}\,(g_j(\theta_{0,t})) = \text{Var}\,(g_t(\theta_{0,t})) + o(1)$ by Assumption 2.3(ii) (as $T \to \infty$). To show (A.4), bound $\nu_{T,\epsilon} \leqslant \epsilon^{-1} E[k_{jt}^4 \xi_{tj}^4|\mathcal{F}_{j-1}]$. Note that the weights k_{jt} are uniformly bounded in t, j, and

$$E[\xi_{tj}^4|\mathcal{F}_{j-1}] = K_{2,t}^{-2} E[a' g_j^*]^4 V_j^{-2} \leqslant K_{2,t}^{-2} \|a\|^4 E\left\|g_j^*\right\|^4 \left\|V_j\right\|_{sp}^{-2} = O_p(1),$$

by Assumption 2.1(iii). This implies that

$$\nu_{T,\epsilon} \leqslant O_p(1) K_{2,t}^{-2} \sum_{j=1}^{T} k_{jt}^4 = o_p(1),$$

since $K_{2,t}^{-2} \sum_{j=1}^{T} k_{jt}^4 \leqslant K_{2,t}^{-2} \sum_{j=1}^{T} k_{jt}^2 = O((Tb)^{-1}) = o(1)$.

Proof of (A.2). We have

$$\begin{aligned} E\left\|\frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} k_{jt} E(g_j(\theta_{0,t})|\mathcal{F}_{j-1})\right\| &\leqslant \frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} k_{jt} E\left\|E(g_j(\theta_{0,t})|\mathcal{F}_{j-1}) - E[g_j(\theta_{0,t})] + E[g_j(\theta_{0,t})]\right\| \\ &\leqslant \frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} k_{jt}\Big(E[g_j(\theta_{0,t})] - E[g_t(\theta_{0,t})] + E[g_t(\theta_{0,t})]\Big) = O_p\Big(T^{\frac{1}{2}} b^{\frac{1}{2}+\gamma}\Big), \end{aligned}$$

where the second inequality follows first from the Ibragimov's mixing inequality for α-mixing process (see Theorem 14.2 in Davidson (1994)) by setting $p = r = 1$ and then by applying Lemma C1(2)(a), together with the fact that $E[g_t(\theta_{0,t})] = 0$. This completes the proof.

Appendix A.2. Proof of Theorem 1

The local CU-GMM estimator is defined by (3):

$$\hat{\theta}_t = \arg\min_{\theta_t \in \Theta} \; Q_{t,T},$$

where the criteria function $Q_{t,T}$ is given by

$$Q_{t,T}(\theta_t) = \overline{g}'_{T,t}(\theta_t) W^{-1}_{T,t}(\theta_t) \overline{g}_{T,t}(\theta_t).$$

We first prove the consistency of the estimator. Let

$$Q_t(\theta_t) = \Big(E\big[g_t(\theta_t)\big]\Big)' W_t^{-1}(\theta_t) \Big(E\big[g_t(\theta_t)\big]\Big),$$

where $W_t = \mathrm{Var}(g_t(\theta_t))$. In view of Theorem 2.1 in Newey and McFadden (1994), it is sufficient to show that

(i) Θ is compact (assumed in Assumption 2.1(i));

(ii) $Q_t(\theta_t)$ is uniquely minimized at $\theta_{0,t}$ (implies by Assumption 2.2);

(iii) $Q_t(\theta_t)$ is continuous in Θ (implied in Assumption 2.1(ii));

(iv) Uniform consistency:

$$\max_{\theta_t \in \Theta} \left| Q_{t,T}(\theta_t) - Q_t(\theta_t) \right| \xrightarrow{p} 0.$$

Thus, it remains to show (iv), which follows from

$$\max_{\theta_t \in \Theta} \left\| \overline{g}_{T,t}(\theta_t) - E(g_t(\theta_t)) \right\| \xrightarrow{p} 0, \tag{A.5}$$

$$\max_{\theta_t \in \Theta} \left\| W^{-1}_{T,t}(\theta_t) - W_t^{-1}(\theta_t) \right\| \xrightarrow{p} 0. \tag{A.6}$$

(A.5) is exactly (5) in Lemma 1(i). For (A.6), notice that

$$\max_{\theta_t \in \Theta} \left\| W^{-1}_{T,t}(\theta_t) - W_t^{-1}(\theta_t) \right\| \leqslant \max_{\theta_t \in \Theta} \|W_t(\theta_t)\|^{-1}_{sp} \max_{\theta_t \in \Theta} \left\| W_t(\theta_t) - W_{T,t}(\theta_t) \right\|_{sp} \max_{\theta_t \in \Theta} \left\| W_{T,t}(\theta_t) \right\|^{-1}.$$

Observe that summand in $W_{T,t}(\theta_t)$ satisfies the same assumption as $g_j(\theta_t)$. Similar arguments as used in the proofs of Lemma 1(i) lead to $\max_{\theta_t \in \Theta} \left\| W_t(\theta_t) - W_{T,t}(\theta_t) \right\|_{sp} = o_p(1)$. Together with Assumption 2.1(ii), we have

$$\max_{\theta_t \in \Theta} \left\| W^{-1}_{T,t}(\theta_t) - W_t^{-1}(\theta_t) \right\| = o_p(1), \tag{A.7}$$

which establish (A.6).

To derive uniform consistency rate and asymptotic normality of the estimator, we rely on Taylor series expansion. Write $\theta_{0,t}$ as the true value and consider a first-order Taylor series expansion of $\frac{\partial Q_{t,T}(\hat{\theta}_t)}{\partial \theta} = 0$ around $\theta_{0,t}$,

$$\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial \theta} + \frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial \theta \partial \theta'}(\hat{\theta}_t - \theta_{0,t}) = 0,$$

where θ_{t^*} lies between $\hat{\theta}_t$ and $\theta_{0,t}$. By rearranging terms, we have

$$\begin{aligned}\hat{\theta}_t - \theta_{0,t} &= -\Big(\frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial \theta \partial \theta'}\Big)^{-1} \frac{\partial Q_{t,T}(\theta_{0,t})}{\partial \theta} \\ &= -\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial \theta \partial \theta'}\Big)^{-1} \frac{\partial Q_{t,T}(\theta_{0,t})}{\partial \theta} + \left[\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial \theta \partial \theta'}\Big)^{-1} - \Big(\frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial \theta \partial \theta'}\Big)^{-1}\right] \frac{\partial Q_{t,T}(\theta_{0,t})}{\partial \theta}. \qquad \text{(A.8)}\end{aligned}$$

We need to show that

$$\left\| \Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial \theta \partial \theta'}\Big)^{-1} - \Big(\frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial \theta \partial \theta'}\Big)^{-1} \right\|_{sp} = o_p(1), \quad \forall t \qquad \text{(A.9)}$$

$$\max_{1 \leqslant t \leqslant T} \left\| \frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial \theta \partial \theta'} \right\|_{sp} = O_p(1). \qquad \text{(A.10)}$$

Then, uniform consistency rate and asymptotic normality are determined by $\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial \theta}$. Thus, we need a detailed expansion for the first and second order derivatives for the criteria function $Q_{t,T}$.

Let us first compute the score:

$$\begin{aligned}\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial \theta} &= 2\left[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} \frac{\partial g_j(\theta_{0,t})}{\partial \theta'}\right]' W_{T,t}^{-1}(\theta_{0,t}) \left[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})\right] \\ &\quad + \Big(A_{2,1,t}(\theta_{0,t}), \cdots, A_{2,k,t}(\theta_{0,t})\Big)' \\ &= A_{1,t}(\theta_{0,t}) + A_{2,t}(\theta_{0,t}).\end{aligned}$$

The ℓ_1th elements in $A_{2,t}(\theta_{0,t})$ is given by

$$A_{2,\ell_1,t}(\theta_{0,t}) = \left[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})\right]' \frac{\partial W_{T,t}^{-1}(\theta_{0,t})}{\partial \theta_{\ell_1}} \left[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})\right],$$

where

$$\frac{\partial W_{T,t}^{-1}(\theta_{0,t})}{\partial \theta_{\ell_1}} = -W_{T,t}^{-1}(\theta_{0,t}) \frac{\partial W_{T,t}(\theta_{0,t})}{\partial \theta_{\ell_1}} W_{T,t}^{-1}(\theta_{0,t}). \qquad \text{(A.11)}$$

We need to show that

$$\max_{1\leqslant t\leqslant T}\left\|A_{1,t}(\theta_{0,t})\right\| = O_p\Big((b^\gamma + (Tb)^{-1/2}\sqrt{\log T}\Big) \tag{A.12}$$

$$\left|A_{2,\ell_1,t}(\theta_{0,t})\right| = o_p(1), \quad \text{for } \ell_1 = 1,2,\cdots,k,\ t = 1,2,\cdots,T. \tag{A.13}$$

Proof of (A.13). By (A.6), we have

$$\left\|W_{T,t}^{-1}(\theta_{0,t}) - W_t^{-1}\right\| = o_p(1). \tag{A.14}$$

Next, we consider

$$\frac{\partial W_{T,t}(\theta_{0,t})}{\partial\theta_{\ell_1}} = \frac{1}{K_{2,t}}\sum_{j=1}^{T}k_{jt}^2\Big(g_j(\theta_{0,t})\big(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}}\big)' + \big(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}}\big)g_j'(\theta_{0,t})\Big).$$

Observe that both squared weights k_{jt} and any (a,b)th elements in $g_j(\theta_{0,t})\big(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}}\big)'$ satisfy Assumptions B1-B2, by Lemma B1(1a), we obtain

$$\left\|\frac{1}{K_{2,t}}\sum_{j=1}^{T}k_{jt}^2\Big(g_j(\theta_{0,t})(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}})' - E\Big[g_j(\theta_{0,t})(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}})'\Big]\Big)\right\| = O_p((Tb)^{-1/2}) = o_p(1).$$

Notice that

$$\begin{aligned}\left\|W_{t,d_1}\right\| &= \left\|\frac{1}{K_{2,t}}\sum_{j=1}^{T}k_{jt}^2E\Big[g_j(\theta_{0,t})(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}})'\Big]\Big)\right\| \\ &\leqslant \frac{1}{K_{2,t}}\sum_{j=1}^{T}k_{jt}^2E\left\|g_j(\theta_{0,t})\big(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}}\big)'\right\| \\ &\leqslant \frac{1}{K_{2,t}}\sum_{j=1}^{T}k_{jt}^2\Big\{E\left\|g_j(\theta_{0,t})\right\|\Big\}^{1/2}\Big\{E\left\|\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}}\right\|\Big\}^{1/2} < \infty,\end{aligned}$$

which follows from Assumption 2.1(iii). This implies that

$$\frac{\partial W_{T,t}(\theta_{0,t})}{\partial\theta_{\ell_1}} = W_{t,d_1} + W_{t,d_1}' + o_p(1).$$

Thus, we have, by continuing from (A.11),

$$\left\|\frac{\partial W_{T,t}^{-1}(\theta_{0,t})}{\partial\theta_{\ell_1}}\right\| \leqslant \|W_t\|_{sp}^{-2}\left\|\frac{\partial W_{T,t}(\theta_{0,t})}{\partial\theta_{\ell_1}}\right\| + o_p(1) = O_p(1).$$

By Lemma 1(i), we have

$$\left\| \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \right\| = O_p\Big(b^{\gamma} + (Tb)^{-1/2}\Big) = o_p(1),$$

which implies (A.13)

$$\Big|A_{2,\ell_1,t}(\theta_{0,t})\Big| \leqslant \left\| \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \right\|^2 \left\| \frac{\partial W_{T,t}^{-1}(\theta_{0,t})}{\partial \theta_{\ell_1}} \right\| + o_p(1) = o_p(1).$$

Proof of (A.12). Define

$$\begin{aligned} G_{D,t} &= \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} \Big(\frac{\partial g_j(\theta_{0,t})}{\partial \theta'} - E\Big[\frac{\partial g_j(\theta_{0,t})}{\partial \theta'}\Big] \Big) \\ W_{D,t} &= W_{T,t}^{-1}(\theta_{0,t}) - W_t^{-1}, \end{aligned}$$

We have shown in (A.14) that $\|W_{D,t}\| = o_p(1)$, which holds for all t. Similarly, observe that any (a,b)th elements in $\frac{\partial g_j(\theta_{0,t})}{\partial \theta'}$ also satisfy Assumption B1, by applying Lemma B1(1b), for sufficient small ε, we have

$$\max_{1\leqslant t\leqslant T} \|G_{D,t}\| = O_p((Tb)^{-1/2} \sqrt{\log T}) = o_p(1).$$

By Assumption 2.3(iii), applying Lemma B1(2a), we have

$$\frac{1}{K_t} \sum_{j=1}^{T} k_{jt} E\Big[\frac{\partial g_j(\theta_{0,t})}{\partial \theta'}\Big] = E\Big[\frac{\partial g_t(\theta_{0,t})}{\partial \theta'}\Big] + o(1),$$

Define $G_t = E\Big[\frac{\partial g_t(\theta_{0,t})}{\partial \theta'}\Big]$. Let us rewrite $A_{1,t}(\theta_{0,t})$:

$$\begin{aligned} A_{1,t}(\theta_{0,t}) = {} & G_t' W_t^{-1} \Big(\frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \Big) + G_{D,t}' W_{D,t} \Big(\frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \Big) \\ & + G_{D,t}' W_t^{-1} \Big(\frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \Big) + G_t' W_{D,t} \Big(\frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \Big). \end{aligned}$$

Clearly, the dominating term is the first one. Then, we have

$$\max_{1\leqslant t\leqslant T} \|A_{1,t}(\theta_{0,t})\| \leqslant \Big(\max_{1\leqslant t\leqslant T} \|G_t\|_{sp} \Big) \max_{1\leqslant t\leqslant T} \|W_t\|_{sp}^{-1} \left(\max_{1\leqslant t\leqslant T} \left\| \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) \right\| \right) + o_p(1) = O_p\Big(b^{\gamma} + (Tb)^{-1/2} \sqrt{\log T}\Big).$$

Consider now the second order derivatives of the criteria function:

$$\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'} = \Big[\tfrac{\partial A_{1,t}(\theta_{0,t})}{\partial\theta_1} \quad \cdots \quad \tfrac{\partial A_{1,t}(\theta_{0,t})}{\partial\theta_k}\Big]_{k\times k} + \begin{bmatrix} \frac{\partial A_{2,1,t}(\theta_{0,t})}{\partial\theta'} \\ \vdots \\ \frac{\partial A_{2,k,t}(\theta_{0,t})}{\partial\theta'} \end{bmatrix}_{k\times k}$$
$$= B_{1,t}(\theta_{0,t}) + B_{2,t}(\theta_{0,t})$$

We will show that

$$\max_{1\leqslant t\leqslant T}\left\|\frac{\partial A_1(\theta_{0,t})}{\partial\theta_{\ell_2}}\right\| = O_p(1), \quad \ell_2 = 1, \cdots, k \tag{A.15}$$
$$\left\|\frac{\partial A_{2,\ell_2}(\theta_{0,t})}{\partial\theta'}\right\| = o_p(1), \quad \ell_2 = 1, \cdots, k, \;\; t = 1, 2, \cdots, T. \tag{A.16}$$

Proof of (A.15). Consider

$$\begin{aligned}\left\|\frac{\partial A_{1,t}(\theta_{0,t})}{\partial\theta_{\ell_2}}\right\| &= \left[\frac{1}{K_t}\sum_{j=1}^T k_{jt}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\right]' W_{T,t}^{-1}(\theta_{0,t})\Big[\frac{1}{K_t}\sum_{j=1}^T k_{jt} g_j(\theta_{0,t})\Big] \\ &\quad + \left[\frac{1}{K_t}\sum_{j=1}^T k_{jt}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\right]'\frac{\partial W_{T,t}^{-1}(\theta_{0,t})}{\partial\theta_{\ell_2}}\Big[\frac{1}{K_t}\sum_{j=1}^T k_{jt} g_j(\theta_{0,t})\Big] \\ &\quad + \left[\frac{1}{K_t}\sum_{j=1}^T k_{jt}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\right]' W_{T,t}^{-1}(\theta_{0,t})\Big[\frac{1}{K_t}\sum_{j=1}^T k_{jt}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big] \\ &= B_{11,t}(\theta_{0,t}) + B_{12,t}(\theta_{0,t}) + B_{13,t}(\theta_{0,t}).\end{aligned}$$

We need to find bounds for the above three terms. First, observe that any (a, b)th elements in $\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}$ satisfy Assumption B1, we apply Lemma B1(1b) to obtain (for some sufficient small $\epsilon > 0$)

$$\max_{1\leqslant t\leqslant T}\left\|\frac{1}{K_t}\sum_{j=1}^T k_{jt}\Big(\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'} - E\Big[\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big]\Big)\right\| = O_p\big((Tb)^{-1/2}\sqrt{\log T}\big) = o_p(1).$$

Second, by Assumption 2.3(iii)[2], applying Lemma B1(2b), we have

$$\frac{1}{K_t}\sum_{j=1}^T k_{jt}E\Big[\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big] \longrightarrow E\Big[\frac{\partial^2 g_t(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big],$$

[2]Assumption 2.3(iii) is stated in terms of first-order derivative. It still holds for second-order derivative since $E[g_t(\theta_t)]$ is twice continuously differentiable.

uniformly over t. Finally, observe that both $B_{11,t}(\theta_{0,t})$ and $B_{12,t}(\theta_{0,t})$ involve $\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})$, following the arguments used to establish (A.12) and (A.13), it is straightforward to verify that

$$\max_{1\leqslant t\leqslant T}\left\|B_{11,t}(\theta_{0,t})\right\| = o_p(1), \quad \max_{1\leqslant t\leqslant T}\left\|B_{12,t}(\theta_{0,t})\right\| = o_p(1).$$

Clearly, the dominating term is $B_{13,t}(\theta_{0,t})$:

$$\max_{1\leqslant t\leqslant T}\left\|B_{13,t}(\theta_{0,t})\right\| \leqslant \left(\max_{1\leqslant t\leqslant T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} E\Big[\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big]\right\|_{sp}\right)\max_{1\leqslant t\leqslant T}\|W_t\|_{sp}^{-1}\left(\max_{1\leqslant t\leqslant T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} E\Big[\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big]\right\|\right)+o_p(1) = O_p(1).$$

Summing up, we get:

$$\max_{1\leqslant t\leqslant T}\left\|\frac{\partial A_{1,t}(\theta_{0,t})}{\partial\theta_{\ell_2}}\right\| = O_p(1).$$

Proof of (A.16). Consider

$$\begin{aligned}\frac{\partial A_{2,\ell_2,t}(\theta_{0,t})}{\partial\theta'} = {}& 2\Big[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})\Big]'\frac{\partial W_{T,t}^{-1}(\theta_{0,t})}{\partial\theta_{\ell_2}}\Big[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt}\frac{\partial g_j(\theta_{0,t})}{\partial\theta'}\Big]\\ &+\Big[A_{2,1,1,t}(\theta_{0,t})\cdots A_{2,k,1,t}(\theta_{0,t})\Big]_{1\times k},\end{aligned}$$

where a typical element $A_{2,\ell_4,1,t}(\theta_{0,t})$, $\ell_4 = 1, 2, \cdots, k$ is given by

$$A_{2,\ell_4,1,t}(\theta_{0,t}) = \Big[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})\Big]'\frac{\partial^2 W_{T,t}(\theta_{0,t})}{\partial\theta_{\ell_1}\partial\theta_{\ell_4}}\Big[\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})\Big].$$

Since both elements above involves $\frac{1}{K_t}\sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t})$, similar arguments as above leads to (A.16), which concludes the claim. Again, by triangular inequality, we establish (A.10):

$$\max_{1\leqslant t\leqslant T}\left\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\right\|_{sp} \leqslant \max_{1\leqslant t\leqslant T}\left\|B_{1,t}(\theta_{0,t})\right\|_{sp} + \max_{1\leqslant t\leqslant T}\left\|B_{2,t}(\theta_{0,t})\right\|_{sp} = O_p(1)$$

We now move to (A.9):

$$\left\|\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big)^{-1} - \Big(\frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial\theta\partial\theta'}\Big)^{-1}\right\|_{sp} \leqslant \left\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\right\|_{sp}^{-1}\left\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial\theta\partial\theta'}\right\|_{sp}\left\|\frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial\theta\partial\theta'}\right\|_{sp}^{-1}.$$

We need to show:

$$\left\| \frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial\theta\partial\theta'} \right\|_{sp} = o_p(1), \quad \forall t$$

$$\max_{1\leqslant t\leqslant T} \left\| \frac{\partial^2 Q_{t,T}(\theta_{t^*})}{\partial\theta\partial\theta'} \right\|_{sp}^{-1} = O_p(1).$$

These bounds follow immediately by letting $\theta_{t^*} \xrightarrow{p} \theta_{0,t}$ (by the consistency of $\hat{\theta}_t$) and (A.10).

Uniform consistency rate: By continuing from (A.8), we obtain the consistency results:

$$\begin{aligned}\max_{1\leqslant t\leqslant T} \left\| \hat{\theta}_t - \theta_{0,t} \right\| &\leqslant \max_{1\leqslant t\leqslant T} \left\| \frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'} \right\|_{sp}^{-1} \max_{1\leqslant t\leqslant T} \left\| \frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta} \right\| + o_p(1) \\ &= O_p\Big(b^\gamma + (Tb)^{-1/2}\sqrt{\log T}\Big)\end{aligned}$$

CLT: for $t = \lfloor \tau T \rfloor$, $0 < \tau < 1$, we can rewrite the estimator as

$$\begin{aligned}\frac{K_t}{K_{2,t}^{1/2}} (\hat{\theta}_t - \theta_{0,t}) &= -\left(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta^2} \right)^{-1} \frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta} + o_p(1) \\ &= -(G_t' W_t^{-1} G_t)^{-1} G_t' W_t^{-1} \frac{K_t}{K_{2,t}^{1/2}} \sum_{j=1}^{T} k_{jt} g_j(\theta_{0,t}) + o_p(1).\end{aligned}$$

By Lemma 1(ii), together with Slutsky's theorem, we obtain

$$\frac{K_t}{K_{2,t}^{1/2}} (G_t' W_t^{-1} G_t)^{1/2} (\hat{\theta}_t - \theta_{0,t}) \xrightarrow{d} N(0, I_k),$$

provided that $T^{\frac{1}{2}} b^{\frac{1}{2}+\gamma} \to 0$. This completes the proof.

Appendix A.3. Proof of Corollary 1

By triangular inequality,

$$\begin{aligned}\left\| \hat{G}_{T,t} - G_t \right\| &\leqslant \left\| \hat{G}_{T,t} - \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} \frac{\partial g_j(\theta_{0,t})}{\partial\theta'} \right\| + \left\| \frac{1}{K_t} \sum_{j=1}^{T} k_{jt} \frac{\partial g_j(\theta_{0,t})}{\partial\theta'} - G_t \right\| \\ &= G_{T,t,1} + G_{T,t,2}.\end{aligned}$$

Following the proof of Lemma 1(i), we can show that $\left\|G_{T,t,2}\right\| = o_p(1)$. For $G_{T,t,1}$, notice that, by mean-value theorem,

$$\left\|G_{T,t,1}\right\| \leqslant \frac{1}{K_t}\sum_{j=1}^{T} k_{jt}\left\|\frac{\partial g_j(\hat{\theta}_t)}{\partial\theta'} - \frac{\partial g_j(\theta_{0,t})}{\partial\theta'}\right\| \leqslant \frac{1}{K_t}\sum_{j=1}^{T} k_{jt}\left\|\frac{\partial^2 g_j(\theta_{t^*})}{\partial\theta_\ell\partial\theta'}\right\|_{sp}\left\|\hat{\theta}_t - \theta_{0,t}\right\|,$$

which holds for all $\ell = 1, 2, \cdots, k$. Since $\max_{\theta\in\Theta}\ \max_j \left\|\frac{\partial^2 g_j(\theta)}{\partial\theta_\ell\partial\theta'}\right\|_{sp} < \infty$ and by consistency we have that $\left\|\hat{\theta}_t - \theta_{0,t}\right\| = o_p(1)$, this completes the proof. The proof for $\hat{W}_{T,t}$ follows similarly.

Appendix B. Auxiliary results

Let $f(x_t, \theta) = f_t(\theta) : \mathbb{R} \times \Theta \to \mathbb{R}$ be a scalar-valued function. Under Assumptions C1-C2, we shall obtain the uniform bounds for sums

$$S_{T,t}(\theta) := \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(f_j(\theta) - Ef_j(\theta)), \tag{B.1}$$

$$\Delta_{T,t}(\theta) := \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(Ef_j^r(\theta) - Ef_t^r(\theta)), \tag{B.2}$$

for some $1 < r \leqslant 2$.

Assumption B1. *(i) Θ is compact;*

(ii) The stochastic process x_t is an α-mixing (but not necessarily stationary) sequence with the mixing coefficients α_k satisfying $\alpha(h) = O(h^{-\gamma})$ for some $\gamma > 2$;

(iii) $f_t(\theta)$ is continuously differentiable w.r.t $\theta \in \Theta$ and

$$\max_{\theta\in\Theta}\max_{1\leqslant t\leqslant T} E|f_t(\theta)|^{p_1} < \infty, \quad \max_{\theta\in\Theta}\max_{1\leqslant t\leqslant T} E\left|\frac{\partial f_t(\theta)}{\partial\theta'}\right|^{p_2} < \infty$$

for some $p_1, p_2 > 2$;

(iv) For any $\theta \in \Theta$, $\{Ef_t(\theta)\}_t$ satisfies the following

$$\left|Ef_j^r(\theta) - Ef_t^r(\theta)\right| \leqslant C\left(\frac{|j-t|}{T}\right)^\gamma, \quad j, t = 1, 2, \cdots, T,$$

where $1 < r \leqslant 2$, $0 < \gamma \leqslant 1$ and the positive constant C does not depend on j, t and T.

Assumption B2. *The weights k_{jt} are computed with a kernel function*

$$k_{jt} = K\left(\frac{j-t}{Tb}\right),$$

where $b \to 0$, $Tb \to \infty$. $K(x)$, $x \in \mathbb{R}$ is a non-negative continuous function satisfying

$$K(x) \leqslant C(1+x^{\nu})^{-1}, \ |(d/dx)K(x)| \leqslant C(1+x^{\nu})^{-1},$$

for some $C > 0$ and $\nu > 3$.

Lemma B1. *Let $f(x_t, \theta) = f_t(\theta) : \mathbb{R} \times \Theta \to \mathbb{R}$ be a scalar-valued function. Under Assumptions B1-B2, we have*

(1) (a) For any sequence $1 \leqslant t = t_T \leqslant T$, as $b \to 0$, $Tb \to \infty$,

$$\max_{\theta \in \Theta} \left| S_{T,t}(\theta) \right| = O_p((Tb)^{-1/2}).$$

(b) If $b = O(T^{-\delta})$ for some $\delta > 0$, then for any $\varepsilon > 0$, $p > 2$,

$$\max_{1 \leqslant t \leqslant T} \left| S_{T,t}(\theta) \right| = O_p((Tb)^{-1/2} \log^{1/2} T + (T^2 b)^{1/p} (Tb)^{\varepsilon - 1}).$$

(2) (a) For any sequence $1 \leqslant t = t_T \leqslant T$, as $T \to \infty$,

$$\max_{\theta \in \Theta} \left| \Delta_{T,t}(\theta) \right| = O_p(b^{\gamma}).$$

(b) If $b = O(T^{-\delta})$ for some $\delta > 0$, then for any $\varepsilon > 0$, $p > 2$,

$$\max_{1 \leqslant t \leqslant T} \left| \Delta_{T,t}(\theta) \right| = O_p(b^{\gamma}).$$

Proof. (1) (b) is (51) in Dendramis et al. (2021). For a given θ, (a) is (48) in Dendramis et al. (2021)[3]. However, we have a different rate of decay for mixing coefficients. In viewing of their proofs, it is essentially to verify Lemma A1 in Dendramis et al. (2021). Since $m^* := \max_{1 \leqslant t \leqslant T} \left(E|f_t|^p\right)^{1/p} < \infty$ and $\sum_{j=1}^{\infty} \alpha_j^{1-2/p} \leqslant C \sum_{j=1}^{\infty} j^{-\gamma(1-2/p)} < \infty$ for some $p > 2$, the results follow immediately.

In the next step, we show that, results in (48) from Dendramis et al. (2021) hold uniformly over θ. We follow the steps in **?** (Proof of Theorem 4.2). Let $\delta > 0$. Since Θ is compact, there exists a finite covering of Θ, $\Theta \subset \cup_{j=1}^{K} \Theta_j$, where $\Theta_j = \Theta_{\delta}(\theta_j)$ is the sphere of radius δ about θ_j and $K \equiv K(\delta)$. It

[3] The results presented in Dendramis et al. (2021) are expressed in terms of $H = Tb$.

follows that, for each $\varepsilon > 0$,

$$P\Big[\max_{\theta\in\Theta}\big|S_{T,t}(\theta)\big| > (Tb)^{-1/2}\varepsilon\Big] \leqslant P\Big[\max_{1\leqslant j\leqslant K}\max_{\theta\in\Theta_j}\big|S_{T,t}(\theta)\big| > (Tb)^{-1/2}\varepsilon\Big]$$
$$\leqslant \sum_{j=1}^{K} P\Big[\max_{\theta\in\Theta_j}\big|S_{T,t}(\theta)\big| > (Tb)^{-1/2}\varepsilon\Big].$$

We will bound each probability in the above summand. For $\theta \in \Theta_j$, by triangular inequality,

$$\begin{aligned}\big|S_{T,t}(\theta)\big| &= \left|\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big(f_j(\theta) - f_j(\theta_j) + f_j(\theta_j) - Ef_j(\theta_j) + Ef_j(\theta_j) - Ef_j(\theta)\big)\right| \\ &\leqslant \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big|f_j(\theta) - f_j(\theta_j)\big| + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big|f_j(\theta_j) - Ef_j(\theta_j)\big| + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big|Ef_j(\theta_j) - Ef_j(\theta)\big|.\end{aligned}$$

Observe that $f_t(\cdot)$ is differentiable, by mean-value theorem,

$$\big|f_j(\theta) - f_j(\theta_j)\big| \leqslant c_j\big|\theta - \theta_j\big| \leqslant \delta c_j, \quad \big|Ef_j(\theta_j) - Ef_j(\theta)\big| \leqslant \overline{c}_j\big|\theta - \theta_j\big| \leqslant \delta\overline{c}_j,$$

where

$$c_j = \frac{\partial f_j(\theta^*)}{\partial\theta'}, \quad \overline{c}_j = E\Big[\frac{\partial f_j(\theta^{**})}{\partial\theta'}\Big],$$

for some θ^*, θ^{**} lie between θ and θ_j. Thus, we have

$$\begin{aligned}\max_{\theta\in\Theta_j}\big|S_{T,t}(\theta)\big| &\leqslant \delta\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(c_j - \overline{c}_j)\Big] + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big|f_j(\theta_j) - Ef_j(\theta_j)\big| + 2\delta\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}c_j \\ &\leqslant \delta\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(c_j - \overline{c}_j)\Big] + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big|f_j(\theta_j) - Ef_j(\theta_j)\big| + 2\delta\overline{C}\end{aligned}$$

where $\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}c_j \leqslant \overline{C}$ which is implies by Assumption B1(iii). It follows that

$$\begin{aligned}P\Big[\max_{\theta\in\Theta_j}\big|S_{T,t}(\theta)\big| > (Tb)^{-1/2}\varepsilon\Big] &\leqslant P\Bigg[\delta\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(c_j - \overline{c}_j)\Big] + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big|f_j(\theta_j) - Ef_j(\theta_j)\big| \\ &\qquad > (Tb)^{-1/2}\varepsilon - 2\delta\overline{C}\Bigg] \\ &\leqslant P\Bigg[\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(c_j - \overline{c}_j)\Big] + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\big|f_j(\theta_j) - Ef_j(\theta_j)\big| > (Tb)^{-1/2}\frac{\varepsilon}{2}\Bigg],\end{aligned}$$

where the second inequality follows by letting $\delta \leqslant 1$ such that $(Tb)^{-1/2}\varepsilon - 2\delta\overline{C} < (Tb)^{-1/2}\frac{\varepsilon}{2}$. Letting

$\theta^* = \theta^{**}$, by applying (48) in Dendramis et al. (2021), we have

$$\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(c_j - \bar{c}_j) = O_p((Tb)^{-1/2}), \quad \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left|f_j(\theta_j) - Ef_j(\theta_j)\right| = O_p((Tb)^{-1/2}).$$

Then, since $K = K(\delta)$ is finite, we can choose T_0, such that

$$P\Big[\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(c_j - \bar{c}_j)\Big] + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left|f_j(\theta_j) - Ef_j(\theta_j)\right| > (Tb)^{-1/2}\frac{\varepsilon}{2}\Big] \leqslant \frac{\varepsilon}{K}$$

holds for all $T \geqslant T_0$. Then

$$P\Big[\max_{\theta\in\Theta}\left|S_{T,t}(\theta)\right| > (Tb)^{-1/2}\varepsilon\Big] \leqslant \varepsilon,$$

which establishes the results.

(2) Notice that, when t is at the interior point,

$$\left|\Delta_{T,t}(\theta)\right| \leqslant C\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\Big(\frac{|j-t|}{T}\Big)^{\gamma} \approx Cb^{\gamma}\int u^{\gamma}K(u)du = O(b^{\gamma}),$$

where the approximation follows from Riemann sum approximation of an integral. The results hold for all t. The proof of (2)(a) follows similar as in (1) by utilizing the compactness of Θ, so we omit. The case when t is at the boundary point is also similar.

□