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

Yu Bai[1a]

[a]*Monash University*

Abstract

We 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. Implementation issues regarding bandwidth selection, construction of pointwise confidence intervals and testing for overidentifying restrictions are discussed. The finite sample performance of the proposed estimator and test statistic are 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 that economic variables satisfy a set of population moment conditions. They could arise from Euler equations in rational expectation models and could also be obtained from asset pricing models under no arbitrage condition. Since the seminal work by Hansen (1982), the Generalized Method of Moments (GMM) estimation has become one of the most commonly used estimation techniques. It is an important tool in macroeconomics, finance, accounting, and labour economics as well. Hall (2005) provides an excellent early review on both methodological development and empirical applications of GMM.

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.

Indeed, one stand of the vast and growing literature on dealing with parameter instability allows for a smooth evolution of parameters rather than abrupt changes. The time-varying parameters are assumed to be either smooth deterministic functions of scaled time points, as in Cai (2007), Chen and Hong (2012) and Li et al. (2021), or persistent and bounded stochastic processes, as in Giraitis et al. (2014), Giraitis et al. (2016), and Giraitis et al. (2021). These papers have provided theoretical, simulation, and empirical results to justify their estimation methods.

In this paper, we 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. We take a nonparametric approach[2] to remain agnostic as possible on the types of time variation. We show the consistency, derive the uniform consistency rate and pointwise asymptotic normality of the proposed estimator. Implementation issues regarding bandwidth selection, construction of pointwise confidence intervals and testing for overidentifying restrictions are discussed.

[2]The local Generalized Method of Moments estimator itself is not new in the literature, but it has not been used and theoretical properties have not been examined for nonparametric time-varying coefficient moment condition models. Lewbel (2007) proposes a local GMM estimator for nonparametrically estimating unknown functions that are defined by conditional moment restrictions, but his framework only applies to the case when parameters are modeled as deterministic functions of *i.i.d.* random variables. Gospodinov and Otsu (2012) propose a local CU-GMM estimator for time-series models defined by conditional moment restrictions, but parameters are assumed to be constant over time. Smith (2011) and Kuersteiner (2012) also propose a kernel-weighted GMM estimator, but their focus is on how to improve estimation efficiency for constant coefficient unconditional moment condition models.

The finite-sample performance of the estimator is evaluated by a Monte-Carlo study. Using a linear instrumental variable model with time-varying parameters as data generating process, we find that both local CU-GMM estimator and overidentifying restrictions test statistic have satisfactory finite sample performance. To illustrate in practice the use of the estimator and test statistic, we provide an empirical application on asset pricing models with Stochastic Discount Factor (SDF) representation. We find substantial time variation in terms of pricing performance across different sets of testing portfolios. Local CU-GMM estimator generally provides economic gains with higher Sharpe ratios compared to standard CU-GMM estimator for constant coefficient moment condition models.

The paper is organized as follows. Section 2 describes the model and the local CU-GMM estimator. Section 3 presents the main theoretical results. Implementation issues regarding bandwidth selection, construction of pointwise confidence intervals and testing for overidentifying restrictions are discussed in Section 4. Section 5 illustrates the proposed estimator and test statistic in a Monte Carlo study and an empirical application on asset pricing models with SDF representation. Section 6 concludes the paper. The mathematical proofs are presented in the Online Appendix.

NOTATION: $||\cdot||$ is the Euclidean norm. $||\cdot||_p$ is the L_p norm. $||\cdot||_{sp}$ is the spectral norm of a matrix. $x_n = O(y_n)$ states that the deterministic sequence x_n is at most of order y_n. $x_n = O_p(y_n)$ states that the vector of random variables x_n is at most of order y_n in probability, and $x_n = o_p(y_n)$ is of smaller order in probability than y_n. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution. We use C for a generic positive (vector) of constant(s) when convenient. $\mathcal{X}$ denotes the sample space. $\sigma(\mathcal{A})$ denotes the σ-algebra generated by a collection of sets $\mathcal{A}$.

2. Model and estimation

Given observed data $\{w_t : t = 1, 2, \cdots, T\}$, we consider the moment condition model

$$E\left[g(w_t; \theta_t)\right] = 0, \tag{1}$$

specified for each t, where the functional form of $g(\cdot, \cdot)$ is known and $\theta_t = (\theta_{1,t}, \cdots, \theta_{d,t})' \in \Theta \subset \mathbb{R}^d$ is the parameter of interest. In the following, for a generic θ, we will often write $g_t(\theta) = g(w_t; \theta)$ when convenient.

Example 1. *Consider the linear regression model $y_t = x_t'\theta_t + u_t$ with time-varying parameters and endogenous regressors, i.e., $E[x_t u_t] \neq 0$. Suppose that there is a $m \times 1$ ($m \geqslant d$) vector of instruments $z_t = (z_{1,t}, \cdots, z_{m,t})'$. Define the moment conditions $g(w_t; \theta_t) = z_t(y_t - x_t'\theta_t)$, where $w_t = (y_t, x_t', z_t')'$. Then, the validity of instruments (z_t is uncorrelated with u_t) implies* (1).

Example 2. *Consider the linear factor pricing model where the stochastic discount factor (SDF) takes the form: $m_t = 1 - f_t'\theta_t$, and f_t is a $d \times 1$ vector of risk factors. Define $g(w_t; \theta_t) = m_t r_t$, where r_t is a $m \times 1$*

vector of excess returns and $w_t = (r_t', f_t')'$. Then, if m_t is the true SDF, standard arguments (e.g. Hansen and Renault (2009)) imply (1).

We consider the local continuous updated GMM (CU-GMM) estimator for θ_t ($t = 1, 2, \cdots, T$):

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

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

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

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

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

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. $\overline{W}_{T,t}(\theta)$ is a consistent estimator of the long-run covariance matrix of $\overline{g}_{T,t}(\theta)$:

$$W_{T,t}(\theta) = \lim_{T\to\infty} \mathrm{Var}\left(\sqrt{Tb}\,\overline{g}_{T,t}(\theta)\right).$$

Our aim is to establish the consistency and derive asymptotic distribution of (2).

3. Asymptotic theory

We impose the following set of assumptions.

Assumption 3.1. *(i) $\theta_t = \theta_{t,T}$ are triangular arrays of vectors whose elements $(\theta_{\ell,t})$ are functions of scaled time point: $\theta_{\ell,t} = \theta_\ell\left(\frac{t}{T}\right)$, $\ell = 1, 2, \cdots, d$, where $\theta_\ell(\cdot) : [0, 1] \to \mathbb{R}$ is continuously differentiable;*

(ii) $w_t = w_{T,t}$ are triangular arrays of vectors whose elements $(w_{\ell,t}, \ell = 1, 2, \cdots, d)$ are α-mixing (but not necessarily stationary) processes with mixing coefficient

$$\alpha(j) = \sup_{A\in\mathcal{G}^0_{-\infty}, B\in\mathcal{G}^\infty_j} |P(A)P(B) - P(AB)|$$

satisfies $\alpha(j) \leqslant c\phi^j$ with $0 < \phi < 1$ and $c > 0$, where $\mathcal{G}^t_s = \sigma(w_s, \cdots, w_t)$;

(iii) Let $\mathcal{F}^t_s = \sigma(g_s(\theta_s), \cdots, g_t(\theta_t))$, $E\left[g_t(\theta_t)|\mathcal{F}^{t-1}_{-\infty}\right] = 0$.

Assumption 3.2. *$g(\cdot,\theta)$ is measurable for each $\theta \in \mathbb{R}^d$. For each $w_t \in \mathcal{X}$, $g_t(\theta)$ is twice continuously differentiable w.r.t. $\theta \in \Theta$ whose elements ($g_{\ell,t}(\theta), \ell = 1, 2, \cdots, m$) and their associated first ($\frac{\partial g_{\ell,j}(\theta)}{\partial \theta_{\ell_2}}, \ell_2, = 1, 2, \cdots, d$) and second order derivatives ($\frac{\partial^2 g_{\ell,j}(\theta)}{\partial \theta^2_{\ell_2}}, \ell_2, = 1, 2, \cdots, d$) satisfy the following: there exists $\delta > 0$, such that*

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

$\forall \ell = 1, 2, \cdots, m$, $\forall \ell_2 = 1, 2, \cdots, d$;

Assumption 3.3. *For each t, θ_t is an interior point of Θ, a compact subset of $\mathbb{R}^d$, where d is finite. θ_t is the unique solution in Θ to the equation $E[g_t(\theta)] = 0$ for each $t = 1, 2, \cdots, T$.*

Assumption 3.4. *For each $\theta \in \Theta$, define $E[g_t(\theta)g_t'(\theta)] = \Sigma(\frac{t}{T})$ and $E[\frac{\partial^{\overline{d}} g_t(\theta)}{\partial \theta^{\overline{d}}}] = \left(\mu\left(\frac{t}{T}\right)\right)^{\overline{d}}$ for $\overline{d} = 1, 2$.*

(i) Elements in $\Sigma(\,\cdot\,)$ are Lipschitz continuous: for any $u, v \in [0, 1]$, given (ℓ_1, ℓ_2),

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

for some $C > 0$;

(ii) Elements in $(\mu(\cdot))^{\overline{d}}$ are Lipschitz continuous: for any $u, v \in [0, 1]$, given (ℓ_3, ℓ_4),

$$\left|(\mu_{\ell_3,\ell_4}(u))^{\overline{d}} - (\mu_{\ell_3,\ell_4}(v))^{\overline{d}}\right| \leqslant C\,|u - v|,$$

for some $C > 0$.

Assumption 3.5. *(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 $C > 0$,$\nu > 3$,*

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

(ii) $b = b_T$ is a bandwidth parameter such that $b \to 0$, $Tb \to \infty$ and $c_1 T^{\frac{2}{p-2}+\delta'-1} \leqslant b \leqslant c_2 T^{-\delta'}$ for some $c_1, c_2 > 0$, $p = 4 + \delta > 4$ as introduced in Assumption 3.2 and $\delta' > 0$ is arbitrarily small.

Assumption 3.1(i) imposes conditions on the smoothness of time-varying parameters. It is required that they are functions of scaled time point t/T and are continuously differentiable. As explained in Robinson (1989), the requirement that time-varying parameter is a function of scaled time point is essential to derive the consistency of the nonparametric estimator, since the amount of local information on which an estimator depends has to increase suitably with sample size T. Assumption 3.1(ii) limits the dependence found in the data w_t. α-mixing is a condition than is weaker than absolute regularity assumed in Li et al. (2021). Assumption 3.1(iii) implies that $(g_t(\theta))_t$ is a martingale difference sequence (M.D.S.). Since structural econometric models (Hansen and Singleton, 1982) are often expressed in terms of conditional moment restrictions, this condition is generally satisfied when model is correctly specified.

Remark 1. *Assumption 3.1(i) also covers cases when the elements in θ_t are the paths of stochastic coefficients. To see this, suppose that $d = 1$ and the scalar θ_t is a realization of a bounded random walk process: $\frac{1}{T^{H-1/2}}\xi_t$, where $\Delta\xi_t = (1-L)^{1-H}v_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and H is the memory parameter. Simple algebra gives $\theta_t = \left(\frac{t}{T}\right)^{H-1/2} \frac{1}{t^{H-1/2}}\xi_t = \left(\frac{t}{T}\right)^{H-1/2} C_t$, where $C_t = \frac{1}{t^{H-1/2}}\xi_t = O_p(1)$ by Theorem 2 in Davydov (1970). This implies that $\theta_t = \theta(t/T) \propto \left(\frac{t}{T}\right)^{H-1/2}$, which is continuously differentiable.*

Assumption 3.2 is standard in the literature (e.g. Hansen (1992)) for the differentiability of $g(w_t;\cdot)$, uniform bounds on $g_t(\theta)$ and associated score. Assumption 3.3 is also a standard condition to ensure the global identification for each θ_t. Assumption 3.4 implies that we allow the variance of moment conditions, expectation of the score and Hessian of moment conditions may be time-varying, but continuous. For this purpose, as for the time-varying parameters (Assumption 3.1(i)), we rescale the time point so that the continuity for both $\Sigma(\cdot)$ and $(\mu(\cdot))^d$ are defined on the set $[0,1]$.

Remark 2. *Assumption 3.4 is related to the nonstationarity in the data w_t. To see this, consider the linear IV model in example 1. Suppose that $d = m = 1$ and (z_t), (u_t) are mutually independent. Then, the variance of moment condition is given by $E\left(z_t^2\right)E\left(u_t^2\right)$. Under Assumption 3.4(i), variances of both (z_t) and (u_t) are allowed to change over time, so they are nonstationary processes. However, changes must be smooth to satisfy the continuity condition. If $x_t = Z_t\Psi_t + v_t$, expectation of the score of the moment condition is given by $\Psi_t E\left(z_t^2\right)$. Assumption 3.4(ii) again implies that variance of (z_t) is allowed to change smoothly over time.*

Assumption 3.5 imposes conditions on kernel function $K(\cdot)$ and bandwidth parameter b. 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). Bounds on b implies that we need a proper degree of smoothing to control the bias-variance trade-off.

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 3.1, 3.2, and 3.5,

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

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

(ii) CLT: under Assumptions 3.1, 3.2, 3.4(i) and 3.5,

$$\frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} k_{jt} g_j(\theta_t) \xrightarrow{d} \mathcal{N}(B_{T,t}, v_0 W_t),$$

where $v_0 = \int K^2(u)du$, $W_t = \mathrm{Var}\left(g_t(\theta_t)\right)$ and $B_{T,t}$ is a bias process of order $T^{1/2}b^{3/2}$ in probability.

Lemma 1(i) provides two types of uniform WLLN results for (4). (5) is the uniform WLLN result over Θ, which is a generalization of the pointwise results presented in Corollary 6(a) in Dendramis et al. (2021). This is needed to prove consistency of the estimator. (6) is the uniform WLLN result over t, where θ_t is given at each t. This is needed to derive the uniform consistency rate of the estimator. Lemma 1(ii) provides pointwise asymptotic normality result. First, it contains a bias term, which vanishes asymptotically if $T^{1/2}b^{3/2} \to 0$. The detailed expression of this term will be made clear in Theorem 1. Of course, $T^{1/2}b^{3/2}$ is at most $O(1)$ as $T \to \infty$. Otherwise, estimation bias dominates and CLT does not apply. Second, the above results imply that optimal weighting matrix $\overline{W}_{T,t}(\theta)$ in (3) takes the form[3]

$$\overline{W}_{T,t}(\theta) = \frac{1}{Tb} \sum_{j=1}^{T} k_{jt}^2 g_j(\theta) g_j'(\theta). \tag{7}$$

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

Theorem 1. Under Assumptions 3.1-3.5,

(i) Consistency:

$$\hat{\theta}_t \xrightarrow{p} \theta_t \quad \text{for each } t,$$
$$\max_{1 \leqslant t \leqslant T} \left\| \hat{\theta}_t - \theta_t \right\| = O_p\left(b + \sqrt{\frac{\log T}{Tb}} \right).$$

(ii) Asymptotic normality:

$$\sqrt{Tb}\, \left(\hat{\theta}_t - \theta_t - b\mu_1 \theta_t^{(1)} \right) \xrightarrow{d} \mathcal{N}\left(0, v_0 \left(G_t' W_t^{-1} G_t \right)^{-1} \right),$$

where $\mu_1 = \int uK(u)du$ and $v_0 = \int K^2(u)du$. $\theta_t^{(1)}$ is the first order derivative of θ_t. G_t and W_t are given

[3]Throughout the paper, we use the noncentered version of the sample covariance matrix estimator. Alternatively, we can follow Hansen et al. (1996) to define our local CU-GMM estimator based on the criteria function $Q_{t,T}^{(c)} = \overline{g}_{T,t}'(\theta) \left(\overline{W}_{T,t}(\theta) - \overline{g}_{T,t}(\theta) \overline{g}_{T,t}'(\theta) \right)^{-1} \overline{g}_{T,t}(\theta)$, but equality of the two estimators follows by $Q_{t,T}^{(c)} = \frac{Q_{t,T}}{1 - Q_{t,T}}$.

by

$$G_t = E\left[\frac{\partial g_t(\theta_t)}{\partial \theta'}\right], \quad W_t = \mathrm{Var}\left(g_t(\theta_t)\right).$$

4. Implementation

4.1. Bandwidth selection

It is well known that the bandwidth plays an essential role in the trade-off between bias and variance. Based on Theorem 1, the asymptotic mean square error (AMSE) of the estimator $\hat{\theta}_t$ is given by

$$\mathrm{AMSE} = \mathrm{Tr}\left(b^2\mu_1^2\left(\theta_t^{(1)}\right)\left(\theta_t^{(1)}\right)' + \frac{v_0\left(G_t'W_t^{-1}G_t\right)^{-1}}{Tb}\right).$$

By minimizing AMSE, we obtain the optimal bandwidth:

$$b_{\mathrm{opt}} = \left\{v_0\mathrm{Tr}\left(\left(G_t'W_t^{-1}G_t\right)^{-1}\right)\mu_1^{-2}\mathrm{Tr}\left(\left(\theta_t^{(1)}\right)\left(\theta_t^{(1)}\right)'\right)^{-1}\right\}^{-1/3} T^{-1/3}.$$

However, the above results are not so useful in practice, since $\left(\theta_t^{(1)}\right)$ is unknown. In practice, we can use the "leave-one-out" cross-validation (CV) method to select the bandwidth. Specifically, a data-driven choice of b is obtained by solving the following minimization problem

$$\hat{b} := \underset{c_1T^{-1/3}\leqslant b\leqslant c_2T^{-1/3}}{\arg\min}\ \tilde{g}_{T,t}'(\theta)\tilde{W}_{T,t}^{-1}(\theta)\tilde{g}_{T,t}(\theta), \tag{8}$$

where $c_1 > 0$, $c_2 > 0$ are suitable positive constants. $\tilde{g}_{T,t}(\theta)$ and $\tilde{W}_{T,t}(\theta)$ are "leave-one-out" version of $\overline{g}_{T,t}(\theta)$ and $\overline{W}_{T,t}(\theta)$, respectively:

$$\tilde{g}_{T,t}(\theta) = \frac{1}{Tb}\sum_{j\neq t} k_{jt}g_j(\theta), \quad \tilde{W}_{T,t}(\theta) = \frac{1}{Tb}\sum_{j\neq t} k_{jt}^2 g_j(\theta)g_j'(\theta).$$

4.2. Pointwise confidence interval

The inference results in Theorem 1 become operational if θ_t in both G_t and W_t are replaced by estimated counterparts. This is provided in the following corollary.

Corollary 1. *Under Assumptions 3.1-3.5, let*

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

Then, it holds that

$$\hat{G}_{T,t} \xrightarrow{p} G_t, \quad \hat{W}_{T,t} \xrightarrow{p} v_0W_t,$$

where $v_0 = \int K^2(u)du$.

4.3. Specification testing

The standard overidentifying restrictions J-test (Hansen, 1982) to assess the adequacy of the model specification can also be easily formulated. Define

$$\hat{V}_{T,t} = \hat{W}_{T,t}^{-1/2} \frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} g_j\left(\hat{\theta}_t\right),$$

and the overidentifying restrictions J-test statistic

$$J_{T,t} = \hat{V}_{T,t}' \hat{V}_{T,t}. \tag{9}$$

Under the null hypothesis at time t:

$$\mathcal{H}_0: \ E\left[g(w_t;\theta_t)\right] = 0,$$

we can derive the limiting distribution of $J_{T,t}$ under $\mathcal{H}_0$. This is formally stated in the next corollary. Notice that, we need an additional assumption $T^{1/2}b^{3/2} \to 0$ so that smoothing bias from the local estimator (Theorem 1)(ii) vanishes asymptotically. Thus, overidentifying restrictions test statistic shall have standard χ^2 asymptotic distribution.

Corollary 2. *Under Assumptions 3.1-3.5 and $\mathcal{H}_0$, suppose that $T^{1/2}b^{3/2} \to 0$, we have*

$$J_{T,t} \xrightarrow{d} \chi^2_{m-d}.$$

5. Numerical studies

5.1. A simulated example

In this Section we evaluate the finite sample performance of the local CU-GMM estimator and the local J-test. We consider the linear IV model with time-varying parameters as data generating process (DGP):

$$y_t = x_t\theta_t + \gamma z_{1,t} + \varepsilon_{yt}, \tag{10}$$

where x_t is a scalar process generated by

$$x_t = z_{1,t} + z_{2,t} + \varepsilon_{xt}, \tag{11}$$

where $z_{1,t}$ and $z_{2,t}$ are the instruments. $z_t = (z_{1,t}, z_{2,t})'$ is generated according to $z_{1,t} = \frac{e^1_{zt}+e^0_{zt}}{\sqrt{2}}$, $z_{2,t} = \frac{e^2_{zt}+e^0_{zt}}{\sqrt{2}}$, where $(e^0_{zt}, e^1_{zt}, e^2_{zt})' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3)$. By construction, $z_{1,t}$ and $z_{2,t}$ all have unit variance with correlation equal to 0.5. The disturbance vector $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{xt})'$ is generated in the same way as z_t: $\varepsilon_{yt} = \frac{e^1_{\varepsilon t}+e^0_{\varepsilon t}}{\sqrt{2}}$, $\varepsilon_{xt} = \frac{e^2_{\varepsilon t}+e^0_{\varepsilon t}}{\sqrt{2}}$, where $(e^0_{\varepsilon t}, e^1_{\varepsilon t}, e^2_{\varepsilon t})' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3)$. Furthermore, we set $\gamma = 0$ so that both $z_{1,t}$ and $z_{2,t}$ are valid instruments.

For the time-varying parameter θ_t, we consider five different specifications. In DGPs 1-2, they are generated according to: (1) $\theta_t = 2(t/T) + \exp\big(-16(t/T - 0.5)^2\big) - 1$; (2) $\theta_t = (t/T)^2 + 0.4\cos(4\pi t/T)$. For DGPs 3-5, we consider θ_t as the realization of stochastic process: $\tilde{\theta}_t = \frac{\xi_t}{T^{H-1/2}}$, where $(1 - L)^{H-1}\xi_t = \epsilon_t$ and ϵ_t is the standard normal *i.i.d.* noise. We set $H = 1.25$ for DGP3, $H = 1$ for DGP4, and $H = 0.75$ for DGP5. A graphical illustration for the paths coefficients of θ_t when $T = 1000$ is provided in Figure 1. It is worth mentioning that remark 1 only holds asymptotically, and we do see paths become rougher as H decreases.

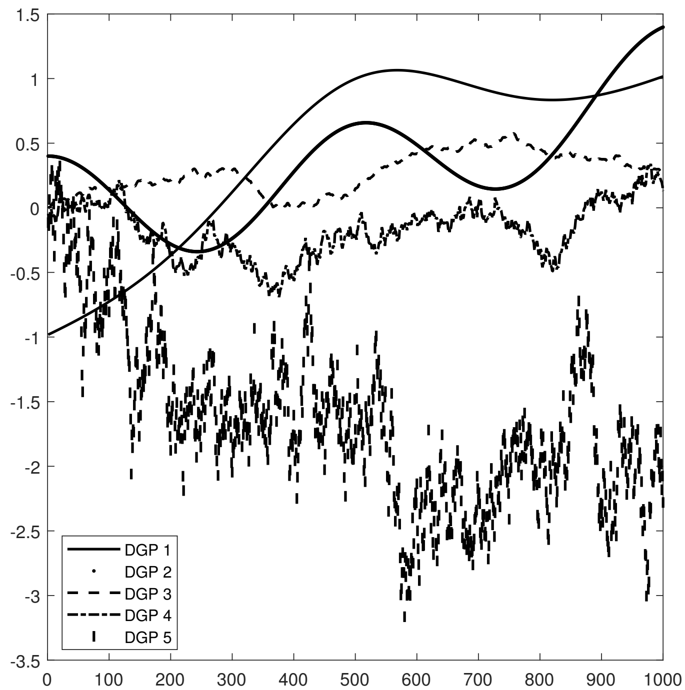

Figure 1: Plots of paths coefficients of θ_t when $T = 1000$ from DGPs 1-5.

To implement the local CU-GMM estimation, an Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{|u|\leqslant 1}$ is used. In terms of bandwidth selection, we first set $b = cT^{-1/3}$ with c ranging from 0.2 to 1.8 with step size equal to 0.2. The leave-one-out cross-validation method, detailed in section 4.1, is then used to select the optimal bandwidth. We consider three sample sizes: $T = 200, 400, 800$, with $R = T$ 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): RMSE $= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\theta}_t - \theta_t)^2}$ and the 95% coverage probability (CP), which is the estimated probability that the true θ_t lies in the interval $(\hat{\theta}_t - 1.96\mathrm{sd}(\hat{\theta}_t), \hat{\theta}_t + 1.96\mathrm{sd}(\hat{\theta}_t))$, where $\mathrm{sd}(\hat{\theta}_t)$ is the estimated variance of the estimator obtained from the associated asymptotic distributions. For the overidentifying restrictions J-test, we report the rejection frequencies for the middle point $t = \lfloor T/2 \rfloor$, taking 5% as the significance level.

Table 1 reports RMSEs, CP and rejection frequencies for the local J-test at the point $t = \lfloor T/2 \rfloor$ from local CU-GMM estimator. In all, the performance are quite satisfactory as RMSEs decrease and CP increase as sample size increases. The size of local J-test also gets closer to nominal value (5%) as sample size increases. We observe higher RMSEs and lower CP in DGP 5 compared to DGPs 1-4. From Figure 1, we see that paths from DGP 5 are rough, which may be harder to estimate in finite sample. In addition, as explained in remark 1, the path coefficients in DGP 5 behave like $\theta_t \propto (t/T)^{1/4}$, while they are proportional to $(t/T)^{3/4}$ for DGP 3 and $(t/T)^{1/2}$ for DGP 4. As H decreases, the overall degree of time variation is also lower.

To explore the power properties of the local J-test, we let γ in (10) vary from 0.01 to 1.8, with step size equal to 0.1 from 0.01 to 1 and 0.2 from 1 to 1.8. We then calculate the rejection frequencies for the middle point $t = \lfloor T/2 \rfloor$, taking 5% as the significance level, over 1,000 replications. The optimal bandwidth is again selected based on the leave-one-out cross-validation method. The power curves appear in Figure 2. Each plot represents a different DGP and different markers in each plot indicate different sample size. It can be clearly seen from the figure that local J-test has good power, since it increases and gradually approaches 1 as γ and sample size increase.

Table 1: Small sample properties of the local CU-GMM estimator: Average RMSFE, coverage probability and rejection frequencies for the local J-test

DGP	$T = 200$			$T = 400$			$T = 800$		
	RMSE	CP	J-test	RMSE	CP	J-test	RMSE	CP	J-test
1	0.283	0.80	0.01	0.180	0.85	0.04	0.129	0.88	0.04
2	0.274	0.80	0.01	0.199	0.85	0.02	0.128	0.88	0.05
3	0.334	0.80	0.01	0.188	0.85	0.02	0.128	0.88	0.03
4	0.325	0.79	0.01	0.204	0.83	0.03	0.134	0.87	0.04
5	0.421	0.67	0.01	0.287	0.68	0.02	0.235	0.67	0.03

Notes: The table reports RMSEs, 95% coverage probabilities (CP) and rejection frequencies for the local J-test at the point $t = \lfloor T/2 \rfloor$ from local CU-GMM estimator.

5.2. *An empirical application*

Consider the asset pricing model with stochastic discount factor (SDF) representation:

$$E\left(r_t\left(1 - f_t'\theta_t\right)\right) = 0_N, \tag{12}$$

where $r_t = (r_{1,t}, \cdots, r_{N,t})'$ is a $N \times 1$ vector of portfolios' excess returns, $f_t = (f_{1,t}, \cdots, f_{d,t})'$ is a $d \times 1$ vector of pricing factors, and 0_N is a $N \times 1$ vector of zeros. The coefficients vector θ_t is assumed to be time-varying.

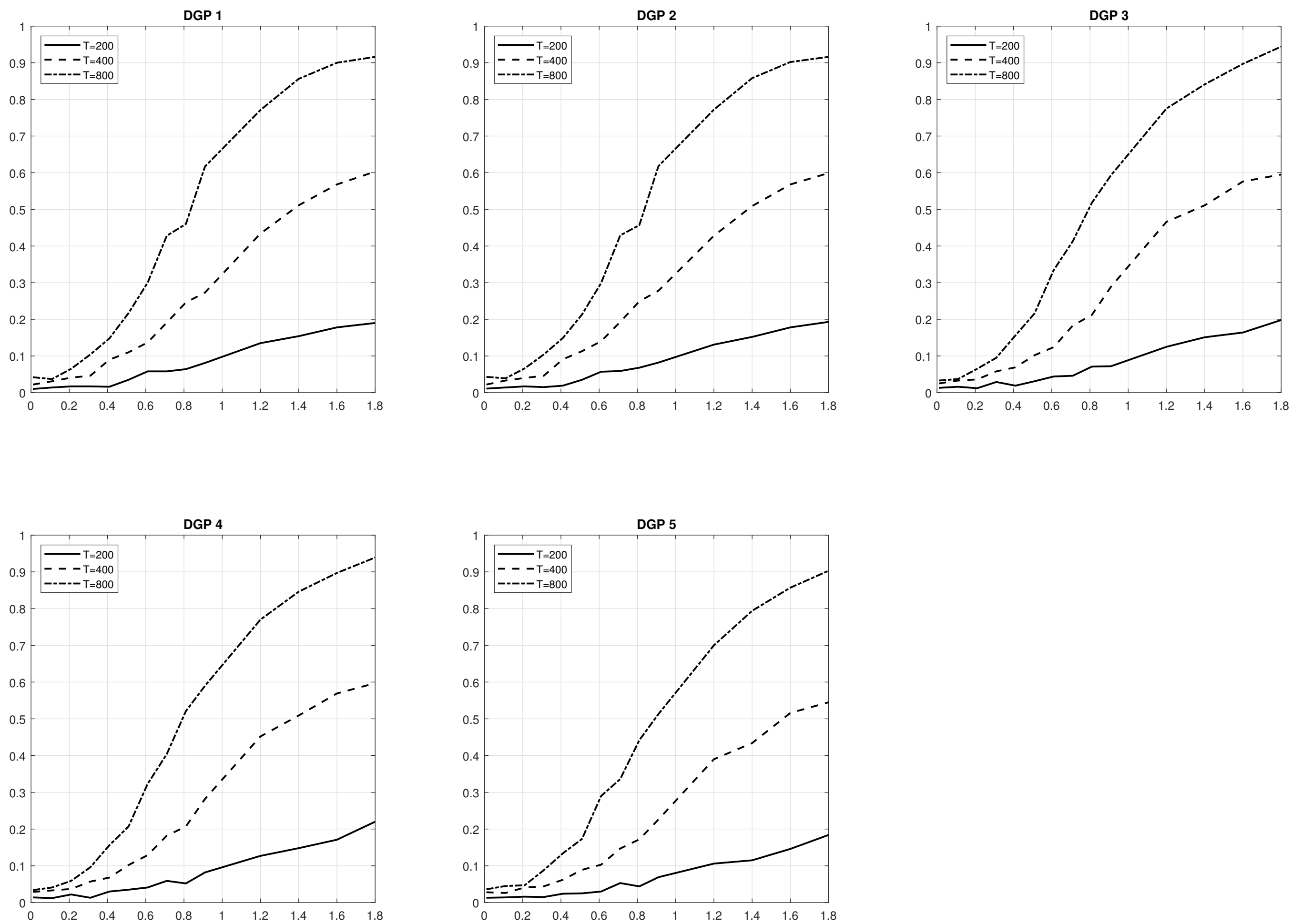

Figure 2: Power curve for the local J-test at the point $t = \lfloor T/2 \rfloor$.

In terms of testing portfolios, we consider 6 portfolios formed on size and book-to-market, 6 portfolios formed on size and momentum, and 10 industry portfolios. Pricing factors used include the excess returns on the value-weighted equity market portfolio (Mkt), as well as returns on the small minus big (SMB) portfolio, the high minus low (HML) portfolio, the robust minus weak (RMW) portfolio and the conservative minus aggressive (CMA) portfolio. Thus, we have three different asset pricing models: CAPM ($f_t = \text{Mkt}_t$), FF3 (Fama-French 3-factor model (Fama and French, 1993), where $f_t = (\text{Mkt}_t, \text{SMB}_t, \text{HML}_t)'$), and FF5 (Fama-French 5-factor model (Fama and French, 2015), where $f_t = (\text{Mkt}_t, \text{SMB}_t, \text{HML}_t, \text{RMW}_t, \text{CMA}_t)'$). All data are taken from Ken French's online data library. The data is monthly and spans the period 1963:7-2024:2, for a total of 728 observations.

As in the previous subsection, an Epanechnikov kernel $K(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{|u|\leqslant 1}$ is used. In terms of bandwidth selection, we first set $b = cT^{-1/3}$ with c ranging from 0.2 to 1.8 with step size equal to 0.2. The leave-one-out cross-validation method, detailed in section 4.1, is then used to select the optimal bandwidth. Figure 3 provides time-varying p-value from J-test in (9). Each row represents one set of testing portfolios. In each row, the left plot is for CAPM model, the middle plot is for FF3 model, and the right plot is for FF5 model. Overall, across all sets of testing portfolios and pricing models, the time variation is substantial. For size and book-to-market portfolios, we are more likely to reject the null hypothesis that FF3 model is correctly specified particularly at the 10% significance level. For size and momentum portfolios, we are more likely to reject the null hypothesis that CAPM model is correctly specified at the 5% significance level particularly during the period 1975-1985.

We now assess whether using the local CU-GMM estimator provides economic gains in a dynamic asset allocation framework. To build the portfolio at time T_0, we first run the regression of fitted SDF and on excess returns:

$$\hat{M}_t = a + b'r_t + \varepsilon_t, \quad t = 1, 2, \cdots, T_0,$$

where $\hat{M}_t = 1 - f_t'\hat{\theta}_{T_0}$ and $\hat{\theta}_{T_0}$ is obtained from one-sided Epanechnikov kernel with bandwidth selected by leave-one-out cross-validation. Then, the weights for portfolio i at time T_0 is formed as $\omega_i = \frac{\hat{b}_i}{\sum_i \hat{b}_i}$. The initial estimation sample runs from 1963:M7 to 1997:M12 ($T_0 = 414$). The portfolio weights are estimated recursively using an expanding window.

In terms of evaluation of portfolio performance, the first measure we consider is the Sharpe ratio (SR), which is calculated as the ratio of the average portfolio excess returns to its standard deviation: $\frac{1}{T-T_0+1}\sum_{t=T_0+1}^{T}\frac{r_{p,t}^F - r_{f,t}}{\sigma^F}$, where $r_{p,t}^F = \omega_{t-1}'r_t$ and σ^F denotes the standard deviation of $r_{p,t}^F - r_{f,t}$. The risk free rate $r_{f,t}$ is the one-month treasury bill rate (from Ibbotson Associates). Apart from SR, we also calculated a risk-adjusted abnormal return relative to a benchmark strategy as follows (Goetzmann et al., 2007):

$$\text{GISW} = \frac{1}{1-\overline{\gamma}}\left[\log\left(\frac{1}{T-T_0+1}\sum_{t=T_0+1}^{T}\left[\frac{1+r_{p,t}^F}{1+r_{f,t}}\right]^{1-\overline{\gamma}}\right) - \log\left(\frac{1}{T-T_0+1}\sum_{t=T_0+1}^{T}\left[\frac{1+r_{p,t}^B}{1+r_{f,t}}\right]^{1-\overline{\gamma}}\right)\right],$$

where $\overline{\gamma}$ is the investor's degree of relative risk aversion and we set it to 5. The benchmark portfolio return $r^B_{p,t}$ is constructed based on equal weighted (EW) strategy in which portfolio weights are equal to $\omega_{i,t} = 1/N$ for all $t = T_0, \cdots, T-1$.

Table 2 reports the performance measures discussed above. Apart from the asset allocation strategy from local CU-GMM and EW strategy, we also consider the strategy in which portfolio weights are obtained from standard CU-GMM where coefficients in (12) are assumed to be constant over time. Across all sets of testing portfolios, local CU-GMM generally delivers higher SR compared to standard CU-GMM. The highest SR is obtained from local-GMM for portfolios formed on size and momentum when FF3 model is used. However, gains in terms of GISW are really small compared to the EW strategy and they are only positive for the CAPM model when testing portfolios are either formed on size and book-to-market or size and momentum.

Table 2: Out-of-sample performance assessment: Performance measures

	EW	CAPM		FF3		FF5	
		Ccoef	TVP	Ccoef	TVP	Ccoef	TVP
		6 portfolios formed on size and book-to-market					
SR	0.138	0.145	0.145	0.098	0.130	0.231	0.115
GISW		0.001	0.001	-0.016	-0.005	-0.034	-0.049
		6 portfolios formed on size and momentum					
SR	0.135	0.141	0.141	0.093	0.248	0.130	0.120
GISW		0.001	0.001	-0.004	-0.002	-0.057	-0.121
		10 industry portfolios					
SR	0.155	0.145	0.145	0.109	0.122	0.061	0.096
GISW		-0.001	-0.001	-0.002	-0.020	-0.049	-0.301

Notes: The table reports the performance measures from portfolios constructed using the equal weighted (EW) strategy, and strategies based on CAPM, FF3 and FF5 asset pricing models. For each asset pricing models, we consider portfolio weights obtained from both constant coefficient (Ccoef) models and time-varying parameter (TVP) models. For Ccoef, parameters are estimated using the standard CU-GMM estimator. For TVP, parameters are estimated using the local CU-GMM estimator. SR denotes the Sharpe ratio. GISW denotes the risk-adjusted abnormal return relative to the EW strategy based on Goetzmann et al. (2007).

6. Conclusions

In this paper, we develop a local continuously updated GMM estimator for nonparametric time-varying coefficient moment condition models. The uniform consistency is shown and the pointwise asymptotic normality of the proposed estimator is derived. Implementation issues regarding bandwidth selection, construction of pointwise confidence intervals and testing for overidentifying restrictions are discussed.

The finite sample properties of the estimator and overidentifying restrictions J-test are evaluated in a Monte-Carlo study. The results show that performance from local CU-GMM estimator, in terms of both biases and coverage rates, is satisfactory. The overidentifying restrictions J-test also has good finite sample performance in terms of size and power. We then illustrate the methods by an empirical application on

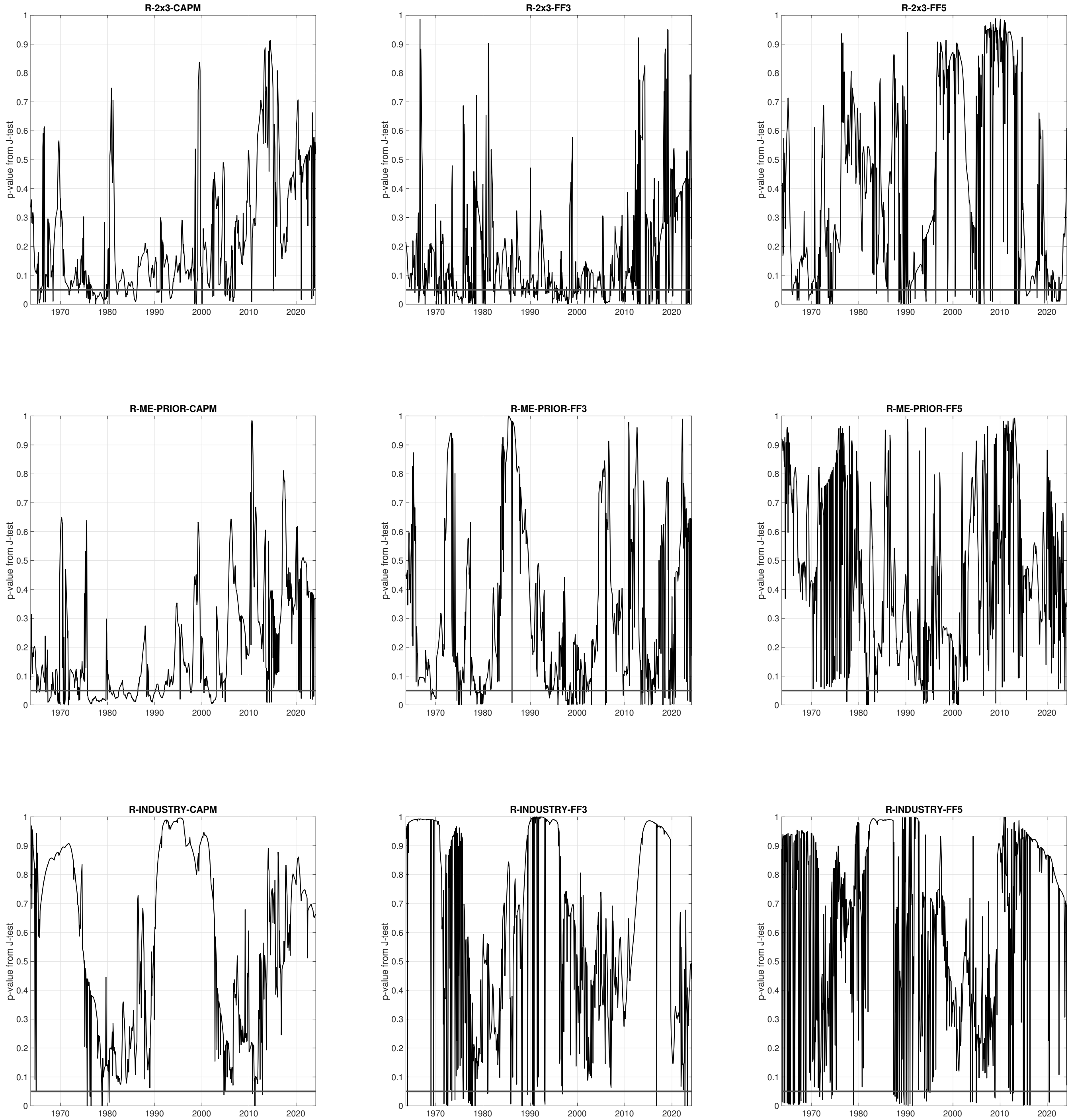

Figure 3: Time-varying p-value from J-test. The title of each plot is structured as "R-X-Y", where X={2x3,ME-PRIOR,INDUSTRY} and Y={CAPM,FF3,FF5}. 2x3: 6 portfolios formed on size and book-to-market; ME-PRIOR: 6 portfolios formed on size and momentum; INDUSTRY: 10 industry portfolios. CAPM: $f_t = \text{Mkt}_t$; FF3: $f_t = (\text{Mkt}_t, \text{SMB}_t, \text{HML}_t)'$; FF5: $f_t = (\text{Mkt}_t, \text{SMB}_t, \text{HML}_t, \text{RMW}_t, \text{CMA}_t)'$.

asset pricing models with SDF representation for the cross-section of equity portfolios. We find substantial time variation in terms of pricing performance across different sets of testing portfolios. Allowing for time variation in SDF parameters generally provides economic gains with higher Sharpe ratios.

Acknowledgements

Previous versions of this paper were circulated with the titles "Time-varying GMM estimation" and "GMM estimation for moment condition models with time-varying parameters". I am very grateful to the editor Prof. Robert Taylor, a co-editor, and two anonymous referees for their constructive comments which help to improve the quality of the paper significantly. I also would like to thank Prof. Massimiliano Marcellino as well as seminar participants at the IAAE 2023 Annual Conference and 2023 ESAM for their suggestions. Financial support from MIUR-PRIN Bando 2017 prot. 2017TA7TYC and ARC Discovery Project DP210100476 are greatly acknowledged.

References

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

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.

Davydov, Y.A., 1970. The invariance principle for stationary processes. Theory of Probability & Its Applications 15, 487–498.

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

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

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of financial economics 116, 1–22.

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.

Goetzmann, W., Ingersoll, J., Spiegel, M., Welch, I., 2007. Portfolio performance manipulation and manipulation-proof performance measures. The Review of Financial Studies 20, 1503–1546.

Gospodinov, N., Otsu, T., 2012. Local gmm estimation of time series models with conditional moment restrictions. Journal of Econometrics 170, 476–490.

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.

Hansen, B.E., 1992. Consistent covariance matrix estimation for dependent heterogeneous processes. Econometrica: Journal of the Econometric Society , 967–972.

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., Renault, E., 2009. Pricing kernels and stochastic discount factors. Encyclopedia of quantitative finance , 1418–1427.

Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica: Journal of the Econometric Society , 1269–1286.

Kuersteiner, G.M., 2012. Kernel-weighted gmm estimators for linear time series models. Journal of Econometrics 170, 399–421.

Lewbel, A., 2007. A local generalized method of moments estimator. Economics Letters 94, 124–128.

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

Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters. Springer.

Smith, R.J., 2011. Gel criteria for moment condition models. Econometric Theory 27, 1192–1235.

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

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

The online appendix is organized as follows. Section A provides proofs of Lemma 1, Theorem 1 and Corollaries 1 and 2. Section B presents auxiliary results and their proofs.

NOTATION: $\|\cdot\|$ is the Euclidean norm. $\|\cdot\|_p$ is the L_p norm. $\|\cdot\|_{sp}$ is the spectral norm of a matrix. $x_n = O(y_n)$ states that the deterministic sequence x_n is at most of order y_n. $x_n = O_p(y_n)$ states that the vector of random variables x_n is at most of order y_n in probability, and $x_n = o_p(y_n)$ is of smaller order in probability than y_n . The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution. We use C for a generic positive (vector) of constant(s) when convenient. $\sigma(\mathcal{A})$ denotes the σ-algebra generated by a collection of sets $\mathcal{A}$.

A Proof of main results

A.1 Proof of Lemma 1

Proof of (i). By Triangular inequality,

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

(6) can be obtained using

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

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

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

Under Assumption 3.5(ii), for sufficiently small $\varepsilon > 0$, it holds $(T^2b)^{1/p}(Tb)^{\varepsilon-1} \leqslant (Tb)^{-1/2}$, this implies that

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

For $M_{T,t}^{(2)}(\theta_t)$, we have

$$\left\|M_{T,t}^{(2)}(\theta_t)\right\| \leqslant \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left\|E(g_j(\theta_t)) - E(g_t(\theta_t))\right\| \leqslant \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left\|E(g_j(\theta_t)) - E\left(g_t(\theta_t)\right)\right\|. \tag{A.1}$$

By mean-value theorem, we have

$$g_j(\theta_t) = g_j(\theta_j) + \frac{\partial g_j(\overline{\theta}_t)}{\partial \theta'}\left(\theta_t - \theta_j\right),$$

where $\overline{\theta}_t$ lies between θ_t and θ_j. Then, by continuing from (A.1), we have

$$\left\|M_{T,t}^{(2)}(\theta_t)\right\| \leqslant \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left\|E\left(\frac{\partial g_j(\overline{\theta}_t)}{\partial \theta'}\right)\right\| \left\|\theta_t - \theta_j\right\| \leqslant C\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left(\frac{|j-t|}{T}\right) \approx Cb\int K(u)du = O(b),$$

because $\max_{1\leqslant t\leqslant T}\left\|E\left(\frac{\partial g_t(\overline{\theta}_t)}{\partial \theta'}\right)\right\| < \infty$ (Assumption 3.3) and the fact that $\theta(t/T)$ is continuously differentiable on $[0,1]$ also implies that it is Lipschitz continuous. This completes the proof of (6).

(5) can be obtained similarly by first noticing that

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

Then, the results follow from Lemma B1(1a) and Lemma B1(2a).
Proof of (ii). Write

$$\begin{aligned}
\frac{1}{\sqrt{Tb}}\sum_{j=1}^{T} k_{jt} g_j(\theta_t) &= \frac{1}{\sqrt{Tb}}\sum_{j=1}^{T} k_{jt}\Big(g_j(\theta_t) - g_j(\theta_j) + g_j(\theta_j)\Big) \\
&= \frac{1}{\sqrt{Tb}}\sum_{j=1}^{T} k_{jt}\left(g_j(\theta_t) - g_j(\theta_j)\right) + \frac{1}{\sqrt{Tb}}\sum_{j=1}^{T} k_{jt} g_j(\theta_j) \\
&:= B_{T,t} + V_{T,t}.
\end{aligned}$$

We first show that

$$V_{T,t} \xrightarrow{d} \mathcal{N}\left(0, v_0 W_t\right),$$

where $W_t = \mathrm{Var}\left(g_t(\theta_t)\right)$[1]. We will proceed by assuming $m = 1$, since the case when $m > 1$ follows immediately from Cramér–Wold device.

Since $g_t(\theta_t)$ is a martingale difference sequence (M.D.S.), by the central limit theorem (CLT)

[1] W_t depends on θ_t, but we write W_t for convenience, and change the notation to $W_t(\cdot)$ when needed.

for M.D.S (e.g. Theorem 3.2 in Hall and Heyde (1980)), we need to verify that

$$\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 g_j^2(\theta_j) \xrightarrow{p} v_0 W_t, \tag{A.2}$$

$$\max_j \left| \frac{1}{\sqrt{Tb}} k_{jt} g_j(\theta_j) \right| \xrightarrow{p} 0, \quad E\left(\max_j \ \frac{1}{Tb} k_{jt}^2 g_j^2(\theta_j)\right) < \infty. \tag{A.3}$$

Proof of (A.2): Write

$$\begin{aligned}
\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 g_j^2(\theta_j) &= \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 \left(g_j^2(\theta_j) - E\left(g_j^2(\theta_j)\right)\right) + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 E\left(g_j^2(\theta_j)\right) \\
&:= j_T^{(1)} + j_T^{(2)}.
\end{aligned}$$

Observe that $g_j^2(\theta_j) - E\left(g_j^2(\theta_j)\right)$ also satisfies Assumptions 3.1(ii) and 3.2. Then, by Lemma B1(i)(a), we have $\left|j_T^{(1)}\right| = O_p\left((Tb)^{-1/2}\right) = o_p(1)$. For $j_T^{(2)}$, we have

$$\begin{aligned}
\left|j_T^{(2)}\right| &\leqslant \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 \left|E(g_j^2(\theta_j)) - E(g_t^2(\theta_j))\right| + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2\, E\left(g_t^2(\theta_j)\right) \\
&:= j_T^{(21)} + j_T^{(22)}.
\end{aligned}$$

It follows by Assumption 3.4(i) that

$$j_T^{(21)} \leqslant C\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 \left(\frac{|j-t|}{T}\right) \approx Cb\int K^2(u)du = O(b),$$

where the approximation follows from the Riemann sum approximation of an integral. For $j_T^{(22)}$, by applying mean-value theorem, we have (as $Tb \to \infty$)

$$\begin{aligned}
j_T^{(22)} &= \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2\, E\left(g_t^2(\theta_t)\right) + \frac{2}{Tb}\sum_{j=1}^{T} k_{jt}^2\, E\left(g_t(\bar{\theta}_t)\frac{\partial g_t(\bar{\theta}_t)}{\partial \theta'}\right)\left(\theta_j - \theta_t\right) \\
&= v_0 W_t + \frac{2}{Tb}\sum_{j=1}^{T} k_{jt}^2\, E\left(g_t(\bar{\theta}_t)\frac{\partial g_t(\bar{\theta}_t)}{\partial \theta'}\right)\theta_t^{(1)}\left(\frac{|j-t|}{T}\right) \\
&= v_0 W_t + O(b) = v_0 W_t + o(1),
\end{aligned}$$

where the second equality follows from Taylor series expansion of θ_j around θ_t. This established (A.2).

Proof of (A.3): Observe that for $\varepsilon > 0$, we have

$$P\left(\max_j \left|\frac{1}{\sqrt{Tb}} k_{jt} g_j(\theta_j)\right| \geqslant \varepsilon\right) \leqslant \varepsilon^{-2}(Tb)^{-1} \sum_{j=1}^{T} k_{jt}^2 E\left[g_j^2(\theta_j) \mathbb{1}\left(\left|g_j(\theta_j)\right| \geqslant \sqrt{Tb}\varepsilon\right)\right] \to 0,$$

as $Tb \to \infty$, since if Assumption 3.2 holds, then $E\left[g_j^2(\theta_j)\mathbb{1}\left(\left|g_j(\theta_j)\right| \geqslant \sqrt{Tb}\varepsilon\right)\right] \to 0$ as $Tb \to \infty$, which follows from Theorem 12.10 in Davidson (1994), together with the fact that $(Tb)^{-1}\sum_{j=1}^{T} k_{jt}^2 = O(1)$. This establishes (A.3).

We next move on to the analysis of $(Tb)^{-1/2}B_{T,t}$. By mean-value theorem, we have

$$\begin{aligned}
(Tb)^{-1/2}B_{T,t} &= \frac{1}{Tb} \sum_{j=1}^{T} k_{jt}\left(g_j(\theta_t) - g_j(\theta_j)\right) \\
&= \frac{1}{Tb} \sum_{j=1}^{T} k_{jt} \frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'}\left(\theta_t - \theta_j\right) \\
&= \frac{1}{Tb} \sum_{j=1}^{T} k_{jt}\left(\frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'} - E\left(\frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'}\right)\right)\theta_t^{(1)}\left(\frac{|j-t|}{T}\right) + \frac{1}{Tb} \sum_{j=1}^{T} k_{jt} E\left(\frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'}\right)\theta_t^{(1)}\left(\frac{|j-t|}{T}\right) \\
&:= B_{T,t}^{(1)} + B_{T,t}^{(2)},
\end{aligned}$$

where the third equality follows again from Taylor series expansion of θ_j at θ_t. For $B_{T,t}^{(1)}$, since $\frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'} - E\left(\frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'}\right)$ also satisfies Assumptions 3.1(ii) and 3.2, by Lemma B1(2)(a), we have $\left|B_{T,t}^{(1)}\right| = O_p(b) = o_p(1)$. Following similar analysis as for (A.2), we have

$$\begin{aligned}
B_{T,t}^{(2)} &= \frac{1}{Tb} \sum_{j=1}^{T} k_{jt}\left(E\left(\frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'}\right) - E\left(\frac{\partial g_t(\bar{\theta}_t)}{\partial \theta'}\right) + E\left(\frac{\partial g_t(\bar{\theta}_t)}{\partial \theta'}\right)\right)\theta_t^{(1)}\left(\frac{|j-t|}{T}\right) \\
&= O_p(b) + E\left(\frac{\partial g_t(\bar{\theta}_t)}{\partial \theta'}\right)\theta_t^{(1)} \frac{1}{Tb} \sum_{j=1}^{T} k_{jt}\left(\frac{|j-t|}{T}\right) = O_p(b).
\end{aligned}$$

This implies that $B_{T,t} = O_p\left(T^{1/2}b^{3/2}\right)$.

A.2 Proof of Theorem 1

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

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

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

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

We first prove the consistency of the estimator. Let

$$Q_t(\theta) = \left(E\left[g_t(\theta)\right]\right)' (v_0 W_t(\theta))^{-1} \left(E\left[g_t(\theta)\right]\right),$$

where $v_0 = \int K^2(u)du$ and $W_t(\theta) = \mathrm{Var}(g_t(\theta))$. In view of Theorem 2.1 in Newey and McFadden (1994), it is sufficient to verify that

(i) Θ is compact (assumed in Assumption 3.3);

(ii) $Q_t(\theta)$ is uniquely minimized at θ_t (implies by Assumption 3.3);

(iii) $Q_t(\theta)$ is continuous in Θ (implied in Assumption 3.2));

(iv) Uniform consistency:

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

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

$$\max_{\theta\in\Theta} \left\|\overline{g}_{T,t}(\theta) - E(g_t(\theta))\right\| \xrightarrow{p} 0, \tag{A.4}$$

$$\max_{\theta\in\Theta} \left\|\overline{W}_{T,t}^{-1}(\theta) - v_0^{-1} W_t^{-1}(\theta)\right\| \xrightarrow{p} 0. \tag{A.5}$$

(A.4) is exactly (5) in Lemma 1(i). For (A.5), notice that

$$\max_{\theta\in\Theta} \left\|\overline{W}_{T,t}^{-1}(\theta) - v_0^{-1} W_t^{-1}(\theta)\right\| \leqslant \max_{\theta\in\Theta} \|W_t(\theta)\|_{sp}^{-1} \max_{\theta\in\Theta} \left\|v_0 W_t(\theta) - \overline{W}_{T,t}(\theta)\right\|_{sp} v_0^{-1} \max_{\theta\in\Theta} \left\|\overline{W}_{T,t}(\theta)\right\|^{-1}.$$

Recall that, for each $\theta \in \Theta$,

$$\begin{aligned}\overline{W}_{T,t}(\theta) &= \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 \left(g_j(\theta)g_j'(\theta) - E\left(g_j(\theta)g_j'(\theta)\right)\right) + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 E\left(g_j(\theta)g_j'(\theta)\right) \\ &= \overline{W}_{T,t}^{(1)}(\theta) + \overline{W}_{T,t}^{(2)}(\theta)\end{aligned}$$

Following same arguments as used for (A.2), we have $\max_{\theta\in\Theta} \left\|\overline{W}_{T,t}^{(1)}(\theta)\right\| = O_p\left((Tb)^{-1/2}\right)$, $\overline{W}_{T,t}^{(2)}(\theta) = v_0 W_t(\theta) + o(1)$. This further implies that $\max_{\theta\in\Theta} \left\|v_0 W_t(\theta) - \overline{W}_{T,t}(\theta)\right\|_{sp} = O_p\left((Tb)^{-1/2}\right) = o_p(1)$. Together with Assumption 3.2 (which implies that both $\max_{\theta\in\Theta} \|W_t(\theta)\|_{sp}^{-1}$ and $\max_{\theta\in\Theta} \left\|\overline{W}_{T,t}(\theta)\right\|^{-1}$ are $O_p(1)$), we have

$$\max_{\theta\in\Theta} \left\|\overline{W}_{T,t}^{-1}(\theta) - v_0^{-1} W_t^{-1}(\theta)\right\| = o_p(1), \tag{A.6}$$

which establish (A.5).

By expanding the first-order condition of $\frac{\partial Q_{t,T}(\hat{\theta}_t)}{\partial\theta} = 0$ around θ_t, we have

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

where $\overline{\theta}_t$ lies between $\hat{\theta}_t$ and θ_t. By rearranging terms, we have

$$\begin{aligned}\hat{\theta}_t - \theta_t &= -\Big(\frac{\partial^2 Q_{t,T}(\overline{\theta}_t)}{\partial\theta\partial\theta'}\Big)^{-1}\frac{\partial Q_{t,T}(\theta_t)}{\partial\theta} \\ &= -\Big(\frac{\partial^2 Q_{t,T}(\theta_t)}{\partial\theta\partial\theta'}\Big)^{-1}\frac{\partial Q_{t,T}(\theta_t)}{\partial\theta} + \left[\Big(\frac{\partial^2 Q_{t,T}(\theta_t)}{\partial\theta\partial\theta'}\Big)^{-1} - \Big(\frac{\partial^2 Q_{t,T}(\overline{\theta}_t)}{\partial\theta\partial\theta'}\Big)^{-1}\right]\frac{\partial Q_{t,T}(\theta_t)}{\partial\theta}.\end{aligned} \tag{A.7}$$

We need to show that

$$\max_{1\leqslant t\leqslant T}\left\|\left(\frac{\partial^2 Q_{t,T}(\theta_t)}{\partial\theta\partial\theta'}\right)^{-1} - \left(\frac{\partial^2 Q_{t,T}(\overline{\theta}_t)}{\partial\theta\partial\theta'}\right)^{-1}\right\|_{sp} = o_p(1), \tag{A.8}$$

$$\max_{1\leqslant t\leqslant T}\left\|\frac{\partial^2 Q_{t,T}(\theta_t)}{\partial\theta\partial\theta'}\right\|_{sp} = O_p(1). \tag{A.9}$$

Then, uniform consistency rate and asymptotic normality are determined by $\frac{\partial Q_{t,T}(\theta_t)}{\partial\theta}$. 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:

$$\begin{aligned}\frac{\partial Q_{t,T}(\theta_t)}{\partial\theta} &= 2\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\frac{\partial g_j(\theta_t)}{\partial\theta'}\Big]'\overline{W}_{T,t}^{-1}(\theta_t)\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}g_j(\theta_t)\Big] \\ &\quad + \Big(A_{2,1,t}(\theta_t), \cdots, A_{2,d,t}(\theta_t)\Big)' \\ &= A_{1,t}(\theta_t) + A_{2,t}(\theta_t).\end{aligned}$$

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

$$A_{2,\ell_1,t}(\theta_t) = \left[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}g_j(\theta_t)\right]'\frac{\partial\overline{W}_{T,t}^{-1}(\theta_t)}{\partial\theta_{\ell_1}}\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}g_j(\theta_t)\Big],$$

where

$$\frac{\partial\overline{W}_{T,t}^{-1}(\theta_t)}{\partial\theta_{\ell_1}} = -\overline{W}_{T,t}^{-1}(\theta_t)\frac{\partial\overline{W}_{T,t}(\theta_t)}{\partial\theta_{\ell_1}}\overline{W}_{T,t}^{-1}(\theta_t). \tag{A.10}$$

We will show that

$$\max_{1\leqslant t\leqslant T} \left\|A_{1,t}(\theta_t)\right\| = O_p\Big(b + (Tb)^{-1/2}\sqrt{\log T}\Big) \tag{A.11}$$

$$\max_{1\leqslant t\leqslant T} \left|A_{2,\ell_1,t}(\theta_t)\right| = o_p(1), \quad \text{for } \ \ell_1 = 1, 2, \cdots, d, \tag{A.12}$$

which implies that the dominating term is $A_{1,t}(\theta_t)$, while $A_{2,t}(\theta_t)$ is smaller order term.

Proof of (A.12). We first establish a bound for (A.10). First, recall that

$$\overline{W}_{T,t}(\theta_t) = \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 g_j(\theta_t) g_j'(\theta_t).$$

Then, using similar steps as in the proof of (A.5) and applying Lemma B1(1)(b), we obtain

$$\max_{1\leqslant t\leqslant T} \left\|\overline{W}_{T,t}^{-1}(\theta_t) - v_0^{-1}W_t^{-1}(\theta_t)\right\| = O_p\Big((Tb)^{-1/2}\sqrt{\log T} + b\Big) = o_p(1). \tag{A.13}$$

Next, we consider

$$\frac{\partial \overline{W}_{T,t}(\theta_t)}{\partial\theta_{\ell_1}} = \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2\Big(g_j(\theta_t)\big(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\big)' + \big(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\big)g_j'(\theta_t)\Big).$$

Write

$$\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 g_j(\theta_t)\left(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\right)' = \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2\left(g_j(\theta_t)\left(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\right)' - E\left(g_j(\theta_t)\left(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\right)'\right)\right) + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 E\left(g_j(\theta_t)\left(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\right)'\right)$$

Observe that any elements in $g_j(\theta_t)\big(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\big)' - E\left(g_j(\theta_t)\left(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\right)'\right)$ satisfy Assumptions B1-B2, by Lemma B1(1b), we obtain

$$\max_{1\leqslant t\leqslant T}\left\|\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2\Big(g_j(\theta_t)\big(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\big)' - E\Big[g_j(\theta_t)\big(\frac{\partial g_j(\theta_t)}{\partial\theta_{\ell_1}}\big)'\Big]\Big)\right\| = O_p((Tb)^{-1/2}\sqrt{\log T}) = o_p(1).$$

Next, notice that

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

which follows from Assumption 3.2. This implies that

$$\frac{\partial \overline{W}_{T,t}(\theta_t)}{\partial \theta_{\ell_1}} = W_{t,d_1} + W'_{t,d_1} + o_p(1),$$

which holds uniformly over t. Thus, we have, by continuing from (A.10),

$$\max_{1\leqslant t\leqslant T} \left\| \frac{\partial \overline{W}_{T,t}^{-1}(\theta_t)}{\partial \theta_{\ell_1}} \right\| \leqslant v_0^{-2} \max_{1\leqslant t\leqslant T} \|W_t(\theta_t)\|_{sp}^{-1} \max_{1\leqslant t\leqslant T} \left\| \frac{\partial \overline{W}_{T,t}(\theta_t)}{\partial \theta_{\ell_1}} \right\|_{sp} \max_{1\leqslant t\leqslant T} \|W_t(\theta_t)\|^{-1} + o_p(1) = O_p(1).$$

By Lemma 1(i), we have

$$\max_{1\leqslant t\leqslant T} \left\| \frac{1}{Tb} \sum_{j=1}^{T} k_{jt} g_j(\theta_t) \right\| = O_p\Big(b + (Tb)^{-1/2} \sqrt{\log T}\Big) = o_p(1).$$

This implies (A.12)

$$\max_{1\leqslant t\leqslant T} \left| A_{2,\ell_1,t}(\theta_t) \right| \leqslant \max_{1\leqslant t\leqslant T} \left\| \frac{1}{Tb} \sum_{j=1}^{T} k_{jt} g_j(\theta_t) \right\|_{sp} \max_{1\leqslant t\leqslant T} \left\| \frac{\partial \overline{W}_{T,t}^{-1}(\theta_t)}{\partial \theta_{\ell_1}} \right\| \max_{1\leqslant t\leqslant T} \left\| \frac{1}{Tb} \sum_{j=1}^{T} k_{jt} g_j(\theta_t) \right\| + o_p(1) = o_p(1).$$

Proof of (A.11). Define

$$\begin{aligned} G_{D,t} &= \frac{1}{Tb} \sum_{j=1}^{T} k_{jt} \Big(\frac{\partial g_j(\theta_t)}{\partial \theta'} - E\Big[\frac{\partial g_j(\theta_t)}{\partial \theta'}\Big] \Big) \\ W_{D,t} &= \overline{W}_{T,t}^{-1}(\theta_t) - v_0^{-1} W_t^{-1}(\theta_t), \end{aligned}$$

We have shown in (A.13) that $\max_{1\leqslant t\leqslant T} \left\|W_{D,t}\right\| = o_p(1)$,. Similarly, observe that any (a,b)th elements in $\frac{\partial g_j(\theta_t)}{\partial \theta'}$ also satisfy Assumptions B1-B2, by applying Lemma B1(1b), we obtain

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

Following similar steps in the proof of (A.2), we could show that

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

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

$$A_{1,t}(\theta_t) = v_0^{-1} G_t' W_t^{-1}\Big(\frac{1}{Tb}\sum_{j=1}^{T} k_{jt} g_j(\theta_t)\Big) + G_{D,t}' W_{D,t}\Big(\frac{1}{Tb}\sum_{j=1}^{T} k_{jt} g_j(\theta_t)\Big)$$
$$+ v_0^{-1} G_{D,t}' W_t^{-1}\Big(\frac{1}{Tb}\sum_{j=1}^{T} k_{jt} g_j(\theta_t)\Big) + G_t' W_{D,t}\Big(\frac{1}{Tb}\sum_{j=1}^{T} k_{jt} g_j(\theta_t)\Big).$$

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

$$\max_{1\leqslant t\leqslant T} \big\|A_{1,t}(\theta_t)\big\| \leqslant \Big(\max_{1\leqslant t\leqslant T} \|G_t\|_{sp}\Big)\, v_0^{-1} \max_{1\leqslant t\leqslant T} \|W_t(\theta_t)\|_{sp}^{-1} \left(\max_{1\leqslant t\leqslant T} \left\|\frac{1}{Tb}\sum_{j=1}^{T} k_{jt} g_j(\theta_t)\right\|\right) + o_p(1) = O_p\Big(b + (Tb)^{-1/2}\sqrt{\log T}\Big).$$

Consider now the second order derivatives of the criteria function:

$$\frac{\partial^2 Q_{t,T}(\theta_t)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial A_{1,t}(\theta_t)}{\partial\theta_1} & \cdots & \frac{\partial A_{1,t}(\theta_t)}{\partial\theta_d} \end{bmatrix}_{d\times d} + \begin{bmatrix} \frac{\partial A_{2,1,t}(\theta_t)}{\partial\theta'} \\ \vdots \\ \frac{\partial A_{2,d,t}(\theta_t)}{\partial\theta'} \end{bmatrix}_{d\times d}.$$

We will show that

$$\max_{1\leqslant t\leqslant T}\left\|\frac{\partial A_1(\theta_t)}{\partial\theta_{\ell_2}}\right\| = O_p(1), \quad \ell_2 = 1,\cdots,d \tag{A.14}$$
$$\max_{1\leqslant t\leqslant T}\left\|\frac{\partial A_{2,\ell_2}(\theta_t)}{\partial\theta'}\right\| = o_p(1), \quad \ell_2 = 1,\cdots,d. \tag{A.15}$$

Proof of (A.14). Consider

$$\frac{\partial A_{1,t}(\theta_t)}{\partial\theta_{\ell_2}} = \left[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right]' \overline{W}_{T,t}^{-1}(\theta_t)\left[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt} g_j(\theta_t)\right]$$
$$+ \left[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right]' \frac{\partial\overline{W}_{T,t}^{-1}(\theta_t)}{\partial\theta_{\ell_2}}\left[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt} g_j(\theta_t)\right]$$
$$+ \left[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right]' \overline{W}_{T,t}^{-1}(\theta_t)\left[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right]$$
$$= B_{11,t}(\theta_t) + B_{12,t}(\theta_t) + B_{13,t}(\theta_t).$$

We need to find bounds for the above three terms. First, we write

$$\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'} = \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left(\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'} - E\left(\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right)\right) + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt} E\left(\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right).$$

Following again similar steps as in the proof of either Lemma 1(i), we have, $\forall \ell_2$,

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

and

$$\frac{1}{Tb}\sum_{j=1}^{T}k_{jt}E\left(\frac{\partial^2 g_j(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right) = E\left(\frac{\partial^2 g_t(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right)+o(1).$$

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

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

Clearly, the dominating term is $B_{13,t}(\theta_t)$: $\forall \ell_2$,

$$\max_{1\leqslant t\leqslant T}\left\|B_{13,t}(\theta_t)\right\| \leqslant \max_{1\leqslant t\leqslant T}\left\|E\left(\frac{\partial^2 g_t(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right)\right\|_{sp} v_0^{-1}\ \max_{1\leqslant t\leqslant T}\|W_t(\theta_t)\|_{sp}^{-1}\ \max_{1\leqslant t\leqslant T}\left\|E\left(\frac{\partial^2 g_t(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\right)\right\|+o_p(1) = O_p(1).$$

Summing up, we get: $\forall \ell_2$,

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

Proof of (A.15). Consider

$$\begin{aligned}\frac{\partial A_{2,\ell_2,t}(\theta_t)}{\partial\theta} &= 2\left[\frac{1}{Tb}\sum_{j=1}^{T}k_{jt}g_j(\theta_t)\right]'\frac{\partial\overline{W}_{T,t}^{-1}(\theta_t)}{\partial\theta_{\ell_2}}\left[\frac{1}{Tb}\sum_{j=1}^{T}k_{jt}\frac{\partial g_j(\theta_t)}{\partial\theta'}\right]\\ &\quad+\Big[A_{2,1,1,t}(\theta_t)\cdots A_{2,d,1,t}(\theta_t)\Big]_{1\times d},\end{aligned}$$

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

$$A_{2,\ell_4,1,t}(\theta_t) = \left[\frac{1}{Tb}\sum_{j=1}^{T}k_{jt}g_j(\theta_t)\right]'\frac{\partial^2\overline{W}_{T,t}(\theta_t)}{\partial\theta_{\ell_1}\partial\theta_{\ell_4}}\left[\frac{1}{Tb}\sum_{j=1}^{T}k_{jt}g_j(\theta_t)\right].$$

Since both elements above involves $\frac{1}{Tb}\sum_{j=1}^{T}k_{jt}g_j(\theta_t)$, similar arguments as above leads to (A.15), which concludes the claim. Again, by triangular inequality, we establish (A.9):

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

We now move to (A.8):

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

We need to show:

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

These bounds follow immediately by letting $\bar{\theta}_t \xrightarrow{p} \theta_t$ uniformly over t (by the uniform consistency of $\hat{\theta}_t$) and (A.9).

<u>Uniform consistency rate</u>: By continuing from (A.7), we obtain the consistency results:

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

<u>CLT</u>: Based on the above analysis, we can rewrite the estimator as

$$\begin{aligned}\sqrt{Tb}\,\left(\hat{\theta}_t - \theta_t\right) &= -\left(\frac{\partial^2 Q_{t,T}(\theta_t)}{\partial\theta\theta'}\right)^{-1} \frac{\partial Q_{t,T}(\theta_t)}{\partial\theta} + o_p(1) \\ &= -(G_t'\,(v_0 W_t)^{-1}\,G_t)^{-1} G_t'\,(v_0 W_t)^{-1}\,\frac{1}{\sqrt{Tb}}\sum_{j=1}^{T} k_{jt}\, g_j(\theta_t) + o_p(1).\end{aligned}$$

Then, we have

$$\begin{aligned}&\sqrt{Tb}\left(\hat{\theta}_t - \theta_t + (G_t'\,(v_0 W_t)^{-1}\,G_t)^{-1} G_t'\,(v_0 W_t)^{-1}\,\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left(g_j(\theta_j) - g_j(\theta_t)\right)\right) \\ &= -(G_t'\,(v_0 W_t)^{-1}\,G_t)^{-1} G_t'\,(v_0 W_t)^{-1}\,\frac{1}{\sqrt{Tb}}\sum_{j=1}^{T} k_{jt}\, g_j(\theta_j) + o_p(1)\end{aligned}$$

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

$$\sqrt{Tb}\,\left(\hat{\theta}_t - \theta_t - b\mu_1\theta_t^{(1)}\right) \xrightarrow{d} \mathcal{N}\left(0, v_0\left(G_t' W_t^{-1} G_t\right)^{-1}\right),$$

where $\mu_1 = \int uK(u)du$ and $v_0 = \int K^2(u)du$. $\theta_t^{(1)}$ is the first order derivative of θ_t. G_t and W_t are given by

$$G_t = E\left[\frac{\partial g_t(\theta_t)}{\partial \theta'}\right], \quad W_t = \mathrm{Var}\left(g_t(\theta_t)\right).$$

This completes the proof.

A.3 Proof of Corollary 1

By triangular inequality,

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

In the previous section, we have shown that $\left\|G_{T,t,2}\right\| = o_p(1)$. For $G_{T,t,1}$, notice that, by mean-value theorem,

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

which holds for all $\ell = 1, 2, \cdots, d$. Since $\max_{\theta\in\Theta} \max_{1\leqslant t\leqslant T} \left\|\frac{\partial^2 g_t(\theta)}{\partial \theta_\ell \partial \theta'}\right\|_{sp} < \infty$ and by uniform consistency we have that $\max_{1\leqslant t\leqslant T} \left\|\hat{\theta}_t - \theta_t\right\| = o_p(1)$, this completes the proof.

For $\hat{W}_{T,t}$, by triangular inequality,

$$\begin{aligned}
\left\|\hat{W}_{T,t} - v_0 W_t\right\| &\leqslant \left\|\hat{W}_{T,t} - \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 g_j(\theta_t) g_j'(\theta_t)\right\| + \left\|\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}^2 g_j(\theta_t) g_j'(\theta_t) - v_0 W_t\right\| \\
&= W_{T,t,1} + W_{T,t,2}.
\end{aligned}$$

In the previous section, we have shown that $\left\|W_{T,t,2}\right\| = o_p(1)$. Following similar analysis as for $G_{T,t,1}$, we could show that $\left\|W_{T,t,1}\right\| = o_p(1)$. This completes the proof.

A.4 Proof of Corollary 2

Consider the following decomposition of $V_{T,t}$:

$$
\begin{aligned}
V_{T,t} &= \left(\hat{W}_{T,t}^{-1/2} - v_0^{-1/2} W_t^{-1/2} + v_0^{-1/2} W_t^{-1/2}\right) \frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} k_{jt} g_j(\hat{\theta}_t) \\
&= v_0^{-1/2} W_t^{-1/2} \frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} k_{jt} g_j(\theta_t) + W_t^{-1/2} G_t \sqrt{Tb} \Bigg(-(G_t' W_t^{-1} G_t)^{-1} G_t' W_t^{-1} \frac{1}{Tb} \sum_{j=1}^{T} k_{jt} g_j(\theta_t) \Bigg) + o_p(1) \\
&= v_0^{-1/2} W_t^{-1/2} \left(I_m - G_t (G_t' W_t^{-1} G_t)^{-1} G_t' W_t^{-1}\right) \frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} k_{jt} g_j(\theta_t) + o_p(1) \\
&= v_0^{-1/2} W_t^{-1/2} \left(I_m - G_t (G_t' W_t^{-1} G_t)^{-1} G_t' W_t^{-1}\right) \frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} k_{jt} \left(g_j(\theta_j) + \frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'} (\theta_t - \theta_j)\right) + o_p(1) \\
&= v_0^{-1/2} W_t^{-1/2} \left(I_m - G_t (G_t' W_t^{-1} G_t)^{-1} G_t' W_t^{-1}\right) \frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} k_{jt} g_j(\theta_j) + o_p(1),
\end{aligned}
$$

where $v_0 = \int K^2(u)du$. The second equality follows first from the fact that $\hat{W}_{T,t}^{-1/2}$ is a consistent estimator of $W_t^{-1/2}$ and the expansion of each $g_j(\hat{\theta}_t)$ around true $g_j(\theta_t)$. The fourth equality follows from the expansion of each $g_j(\theta_t)$ around $g_j(\theta_j)$. The last equality follows by the assumption $T^{1/2} b^{3/2} \to 0$ so the smoothing bias vanishes asymptotically.

Recall that $v_0^{-1/2} W_t^{-1/2} \frac{1}{\sqrt{Tb}} \sum_{j=1}^{T} k_{jt} g_j(\theta_j)$ converges to the standard normal distribution and the fact that $I_m - G_t (G_t' W_t^{-1} G_t)^{-1} G_t' W_t^{-1}$ is idempotent of rank $m - d$. Then, the results follow immediately from Rao et al. (1973)(p.186).

B Auxiliary results

Definition B1. *The random function $f(x, \theta) : \mathbb{R} \times \Theta \to \mathbb{R}$ satisfies the standard measurability and differentiability conditions on $\mathbb{R} \times \Theta \to \mathbb{R}$ if*

(1) for each $\theta \in \Theta$, $f(\cdot, \theta)$ is measurable;

(2) for each $x \in \mathbb{R}$, $f(x, \cdot)$ is twice continuously differentiable on Θ.

We shall obtain the uniform bounds for sums

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

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

for $r = 1, 2$.

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 $\alpha(j)$ satisfying $\alpha(j) \leqslant c\phi^j$ with $0 < \phi < 1$ and $c > 0$;

(iii) $f(x_t, \theta) = f_t(\theta)$ satisfies the standard measurability and differentiability conditions as in Definition B1 and

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

for some $p > 2$;

(iv) For any $\theta \in \Theta$, $E\left(f_t(\theta)\right)^r = \mu^r(t/T)$ satisfies the following

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

for $r = 1, 2$ 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*

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

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

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

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

Lemma B1. *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 $c_1 T^{\frac{2}{p}+\delta-1} \leqslant b \leqslant c_2 T^{-\delta}$ for some $\delta > 0$, $c_1, c_2 > 0$, $p > 2$ as in Assumption B1(iii), 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 $T \to \infty$,

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

(b) *If* $c_1 T^{\frac{2}{p}+\delta-1} \leqslant b \leqslant c_2 T^{-\delta}$ *for some* $\delta > 0$, $c_1, c_2 > 0$, $p > 2$ *as in Assumption B1(iii), then for any* $\varepsilon > 0$, $p > 2$,

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

Proof. (1) (b) is (51) in Dendramis et al. (2021). For a given θ, (a) is (48) in Dendramis et al. (2021)[2]. In the next step, we show that, results in (48) from Dendramis et al. (2021) hold uniformly over θ. We follow the steps in Wooldridge (1994). 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 follows that, for each $\varepsilon > 0$,

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

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

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

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

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

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

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

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

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

$$P\Big[\max_{\theta\in\Theta_j}\left|S_{T,t}(\theta)\right| > (Tb)^{-1/2}\varepsilon\Big] \leqslant P\left[\delta\Big[\frac{1}{Tb}\sum_{j=1}^{T} k_{jt}(c_j-\overline{c}_j)\Big] + \frac{1}{Tb}\sum_{j=1}^{T} k_{jt}\left|f_j(\theta_j) - Ef_j(\theta_j)\right| \right.$$

$$\left. > (Tb)^{-1/2}\varepsilon - 2\delta\overline{C}\right]$$

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

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

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

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

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

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,

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

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.

□

References

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.

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

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

Rao, C.R., Rao, C.R., Statistiker, M., Rao, C.R., Rao, C.R., 1973. Linear statistical inference and its applications. volume 2. Wiley New York.

Wooldridge, J.M., 1994. Estimation and inference for dependent processes. Handbook of econometrics 4, 2639–2738.