Efficient Derivative Pricing by the Extended Method of Moments,
Supplementary material: APPENDIX B

In this Appendix we provide the proofs of theoretical results and technical Lemmas that have been omitted in the paper. We first give in Section B.1 the proof of Lemma A.1. Then, in Section B.2 we give a detailed proof of consistency of the XMM estimator (see Appendix A.1.3 in the paper). In Sections B.3-B.6 we prove Lemma A.2, Corollary 6, Lemma A.3 and Corollary 8, respectively. In Section B.7 we discuss regularity conditions for XMM estimation when the DGP is the stochastic volatility model of Section 3.2 of the paper. In Section B.8 we derive the risk-neutral distribution in the stochastic volatility model. In Section B.9 we prove Lemma A.4. Finally, in Section B.10 we provide the Fourier transform methods used for option pricing and cross-sectional calibration in the stochastic volatility model. We use the following notation. We denote by K_1 and K_2 the dimensions of functions g_1 and g_2, respectively. Further, $\tilde{g}_2$ denotes function $\tilde{g}_2 = (g_2^{*'}, 1)' = \left(g_2', a', 1\right)'$.

B.1 Proof of Lemma A.1

B.1.1 Conditions for weak convergence of the kernel empirical process

The process $\Psi_T(\theta)$, $\theta \in \Theta$, can be written as:

$$\Psi_T(\theta) = \begin{pmatrix} \sqrt{T}\left(\widehat{E}\left[g_1(\theta)\right] - E\left[g_1(\theta)\right]\right) \\ \sqrt{Th_T^d}\left(\widehat{E}\left[g_2^*(\theta)|x_0\right] - E\left[g_2^*(\theta)|x_0\right]\right) \end{pmatrix}, \quad \theta \in \Theta, \tag{B.1}$$

where g_2^* denotes function $g_2^* = \left(g_2', a'\right)'$. Let us rewrite the second component of $\Psi_T(\theta)$. For $\theta \in \Theta$, let us define (see Assumption A.12):

$$\varphi(\theta) := \varphi(x_0;\theta) = E\left[g_2^*(\theta)|x_0\right] f(x_0),$$

and the corresponding kernel estimator:

$$\widehat{\varphi}(\theta) = \frac{1}{Th_T^d}\sum_{t=1}^{T} g_2^*(y_t;\theta) K\left(\frac{x_t - x_0}{h_T}\right).$$

We have:

$$\begin{aligned}
\sqrt{Th_T^d}\left(\widehat{E}\left[g_2^*(\theta)|x_0\right] - E\left[g_2^*(\theta)|x_0\right]\right) &= \sqrt{Th_T^d}\left(\frac{\widehat{\varphi}(\theta)}{\widehat{f}(x_0)} - \frac{\varphi(\theta)}{f(x_0)}\right) \\
&= \frac{1}{f(x_0)}\sqrt{Th_T^d}\left(\widehat{\varphi}(\theta) - \varphi(\theta)\right) - \frac{E\left[g_2^*(\theta)|x_0\right]}{f(x_0)}\sqrt{Th_T^d}\left(\widehat{f}(x_0) - f(x_0)\right) \\
&\quad - \frac{1}{f(x_0)}\sqrt{Th_T^d}\left(\widehat{E}\left[g_2^*(\theta)|x_0\right] - E\left[g_2^*(\theta)|x_0\right]\right)\left(\widehat{f}(x_0) - f(x_0)\right).
\end{aligned}$$

This can be rewritten as:

$$\begin{aligned}
&\sqrt{Th_T^d}\left(\widehat{E}\left[g_2^*(\theta)|x_0\right] - E\left[g_2^*(\theta)|x_0\right]\right)\left[1 + \frac{1}{f(x_0)}\left(\widehat{f}(x_0) - f(x_0)\right)\right] \\
= \; &\frac{1}{f(x_0)}\sqrt{Th_T^d}\left(\widehat{\varphi}(\theta) - \varphi(\theta)\right) - \frac{E\left[g_2^*(\theta)|x_0\right]}{f(x_0)}\sqrt{Th_T^d}\left(\widehat{f}(x_0) - f(x_0)\right).
\end{aligned} \tag{B.2}$$

Under Assumptions A.5-A.9, we have [see Bosq (1998), Theorem 2.3]:

$$\widehat{f}\left(x_0\right) - f\left(x_0\right) = o_p(1). \tag{B.3}$$

From (B.1)-(B.3) we get:

$$\Psi_T(\theta) = H_0\left(\theta\right)\nu_T^*\left(\theta\right)\left(1 + o_p(1)\right), \qquad \theta \in \Theta, \tag{B.4}$$

where process $\nu_T^*\left(\theta\right)$ is defined by:

$$\nu_T^*\left(\theta\right) = \begin{pmatrix} \sqrt{T}\left(\widehat{E}\left[g_1(\theta)\right] - E\left[g_1(\theta)\right]\right) \\ \sqrt{Th_T^d}\left(\widehat{\varphi}\left(\theta\right) - \varphi\left(\theta\right)\right) \\ \sqrt{Th_T^d}\left(\widehat{f}\left(x_0\right) - f\left(x_0\right)\right) \end{pmatrix}, \quad \theta \in \Theta, \tag{B.5}$$

matrix $H_0(\theta)$ is given by:

$$H_0(\theta) = \begin{pmatrix} Id_{K_1} & 0 & 0 \\ 0 & \frac{1}{f(x_0)} Id_{K_2+L} & -\frac{1}{f(x_0)} E\left[g_2^*(\theta)|x_0\right] \end{pmatrix}, \quad \theta \in \Theta,$$

$K_1 := \dim(g_1)$, $K_2 := \dim(g_2)$ and the $o_p(1)$ term is uniform in $\theta \in \Theta$. The following Lemma shows that process $\nu_T^*\left(\theta\right)$ is asymptotically equivalent to a zero-mean empirical process plus a bias function.

Lemma B.1: *Under Assumptions A.4, A.6, A.8, A.9 and A.12:*

$$\begin{pmatrix} \sqrt{Th_T^d}\left(E\left[\widehat{\varphi}\left(\theta\right)\right] - \varphi\left(\theta\right)\right) \\ \sqrt{Th_T^d}\left(E\left[\widehat{f}\left(x_0\right)\right] - f\left(x_0\right)\right) \end{pmatrix} = \frac{\sqrt{\lim Th_T^{d+2m}}}{m!} w_m \begin{pmatrix} \Delta^m \varphi(x_0;\theta) \\ \Delta^m f(x_0) \end{pmatrix} + o(1), \quad \textit{uniformly in } \theta \in \Theta. \tag{B.6}$$

Proof: From a standard bias expansion and Assumption A.8, we have

$$\begin{aligned} E\left[\widehat{\varphi}\left(\theta\right)\right] - \varphi\left(\theta\right) &= \frac{1}{h_T^d} E\left[\varphi(X;\theta) K\left(\frac{X - x_0}{h_T}\right)\right] - \varphi\left(\theta\right) = \int \left[\varphi\left(x_0 + h_T u;\theta\right) - \varphi(x_0;\theta)\right] K(u) du \\ &= \frac{h_T^m}{m!} \sum_{\alpha:|\alpha|=m} \int \nabla^\alpha \varphi\left(x_0 + h_T \tilde{u};\theta\right) u^\alpha K(u) du, \end{aligned}$$

where $\tilde{u}$ is an intermediary point (depending on u). Since $\nabla^\alpha \varphi$ is uniformly continuous on $\mathcal{X} \times \Theta$ (Assumption A.12), and Θ is compact (Assumption A.4), we have that $\int \nabla^\alpha \varphi\left(x_0 + h_T \tilde{u};\theta\right) u^\alpha K(u) du$ converges to $\nabla^\alpha \varphi\left(x_0;\theta\right) \int u^\alpha K(u) du$, uniformly in $\theta \in \Theta$, for any $\alpha \in \mathbb{N}^d$ with $|\alpha| = m$. A similar argument applies for $E\left[\widehat{f}\left(x_0\right)\right] - f\left(x_0\right)$. Since K is a product kernel of order m (Assumption A.8), the conclusion follows. ■

From (B.4)-(B.6) we deduce:

$$\Psi_T(\theta) = \left[H_0\left(\theta\right)\nu_T\left(\theta\right) + b(\theta) + o(1)\right]\left(1 + o_p(1)\right), \qquad \theta \in \Theta, \tag{B.7}$$

uniformly in $\theta \in \Theta$, where the empirical process $\nu_T\left(\theta\right)$ is defined as:

$$\nu_T\left(\theta\right) = \begin{pmatrix} \sqrt{T}\left(\widehat{E}\left[g_1(\theta)\right] - E\left[g_1(\theta)\right]\right) \\ \sqrt{Th_T^d}\left(\widehat{\varphi}\left(\theta\right) - E\left[\widehat{\varphi}\left(\theta\right)\right]\right) \\ \sqrt{Th_T^d}\left(\widehat{f}\left(x_0\right) - E\left[\widehat{f}\left(x_0\right)\right]\right) \end{pmatrix}, \quad \theta \in \Theta.$$

Lemma A.1 follows if the empirical process $\nu_T(\theta)$ converges weakly:

$$\nu_T(\theta) \Longrightarrow \nu(\theta), \qquad \theta \in \Theta, \tag{B.8}$$

where $\nu(\theta)$ is a Gaussian process on Θ with covariance operator:

$$\Gamma_0(\theta,\tau) = \left(\begin{array}{ccc} S_0(\theta,\tau) & 0 & 0 \\ 0 & w^2 f(x_0) E\left(g_2^*(\theta) g_2^*(\tau)' | x_0\right) & w^2 f(x_0) E\left(g_2^*(\theta)|x_0\right) \\ 0 & w^2 f(x_0) E\left(g_2^*(\tau)' | x_0\right) & w^2 f(x_0) \end{array} \right), \qquad \theta,\tau \in \Theta,$$

and:

$$S_0(\theta,\tau) = \sum_{k=-\infty}^{\infty} Cov\left[g_1(X_t, Y_t;\theta), g_1(X_{t-k}, Y_{t-k};\tau)\right].$$

To prove the weak convergence (B.8) of empirical process $\nu_T(\theta)$, let us note that:

$$\nu_T(\theta) = T^{-1/2} \sum_{t=1}^{T} \left(v_{t,T}(\theta) - E\left[v_{t,T}(\theta)\right]\right), \quad \theta \in \Theta,$$

where

$$v_{t,T}(\theta) = \left(g_1(X_t, Y_t;\theta)', h_T^{-d/2} \widetilde{g}_2(Y_t;\theta)' K\left(\frac{X_t - x_0}{h_T}\right)\right)',$$

and $\widetilde{g}_2$ denotes function $\widetilde{g}_2 = \left(g_2^{*'}, 1\right)'$. From Theorem 10.2 of Pollard (1990), the weak convergence of $\nu_T(\theta)$ to Gaussian process $\nu(\theta)$ over $\Theta \subset \mathbb{R}^p$ compact is implied by the conditions i) and ii) of Proposition B.2 below.

Proposition B.2: *The following conditions are satisfied:*

i) Under Assumptions A.1, A.5-A.15, for any $\theta_1, ..., \theta_n \in \mathbb{R}^p$, $n \in \mathbb{N}$, the vector $\left(\nu_T(\theta_1)', ..., \nu_T(\theta_n)'\right)'$ is asymptotically normally distributed with mean zero, and asymptotic variance-covariance matrix such that:

$$AsCov\left(\nu_T(\theta_i), \nu_T(\theta_j)\right) = \Gamma_0(\theta_i, \theta_j), \quad i, j = 1, ..., n.$$

ii) Under Assumptions A.4, A.5, A.8-A.9 and A.16-A.18, the empirical process $\nu_T(\theta)$ is stochastically equicontinuous, that is, $\forall \varepsilon, \eta > 0\ \exists \delta > 0$:

$$\lim_{T\to\infty} \sup P^* \left[\sup_{\theta,\tau\in\Theta: d(\theta,\tau)<\delta} \|\nu_T(\theta) - \nu_T(\tau)\| > \eta\right] \le \varepsilon,$$

where $d(.,.)$ is a metric on Θ, and P^* denotes the outer probability.

These conditions imply the weak convergence of empirical process ν_T (and, thus, the weak convergence of Ψ_T). Conditions i) and ii) of the previous proposition are verified below in Section B.1.2 and B.1.3, respectively.

B.1.2 Finite dimensional convergence

To prove condition i) of Proposition B.2, we use Cramer-Wold device, and follow an approach similar to Bosq (1998), Theorem 2.3, 3.4, and Tenreiro (1995), Theorem 1.3.10. Let $\lambda = \left(\lambda_1', ..., \lambda_n'\right)' \in$

$\mathbb{R}^{n(K_1+K_2+L+1)}$, and define the zero-mean triangular array:

$$Z_{t,T} = \sum_{i=1}^{n} \frac{1}{\sqrt{T}} \lambda_i^{'} \left(v_{t,T}\left(\theta_i\right) - E\left[v_{t,T}\left(\theta_i\right)\right]\right), \, t \leq T, \quad T \geq 1.$$

Then, we can write:

$$\left(\nu_T(\theta_1)^{'}, ..., \nu_T(\theta_n)^{'}\right) \lambda = \sum_{t=1}^{T} Z_{t,T}.$$

Let $m = m_T$ and $q = q_T$ be sequences of integer numbers such that:

$$m_T = O\left(T^a\right), \quad q_T = O(T^b), \qquad 0 < b < a < 1,$$

and let us define $k = k_T = \lfloor T/(m_T + q_T) \rfloor$. In particular $k_T = O\left(T^{1-a}\right)$. Let us divide the sample into $2k+1$ blocks, whose length is equal to m for blocks $1, 3, ..., 2k-1$, equal to q for blocks $2, 4, ..., 2k$, and equal to $T - k(m+q)$ for the last block. More specifically, define:

$$\begin{aligned}
Y_{1,T} &= Z_{1,T} + ... + Z_{m,T}, \qquad Y_{1,T}^{'} = Z_{m+1,T} + ... + Z_{m+q,T}, \\
Y_{2,T} &= Z_{m+q+1,T} + ... + Z_{2m+q,T}, \qquad Y_{2,T}^{'} = Z_{2m+q+1,T} + ... + Z_{2m+2q,T}, \\
& .... \\
Y_{k,T} &= Z_{(k-1)(m+q)+1,T} + ... + Z_{km+(k-1)q,T}, \quad Y_{k,T}^{'} = Z_{km+(k-1)q+1,T} + ... + Z_{km+kq,T}.
\end{aligned}$$

Thus, we can write:

$$\left(\nu_T(\theta_1)^{'}, ..., \nu_T(\theta_n)^{'}\right) \lambda = \sum_{l=1}^{k} Y_{l,T} + \sum_{l=1}^{k} Y_{l,T}^{'} + Y_T^{''}, \tag{B.9}$$

where $Y_T^{''} = Z_{k(m+q)+1,T} + ... + Z_{T,T}$. Then, we will prove that the first term in the decomposition is asymptotically normal, and that the last two terms are negligible.

a) The first term is asymptotically normal

Lemma B.3: *Under Assumptions A.1, A.5-A.9, A.11-A.14, there exist i.i.d. random variables* $Y_{l,T}^*$, $l = 1, ..., k$, *such that* $Y_{l,T}^* \stackrel{d}{=} Y_{l,T}$, $l = 1, ..., k$, *and* $\sum_{l=1}^{k} Y_{l,T}^* - \sum_{l=1}^{k} Y_{l,T} = o_p(1)$.

Proof: Let $c_T := E\left[\left(Y_{1,T}\right)^2\right]^{1/2}$ and $0 < \xi_T < c_T$. From Bradley's Lemma [e.g. Bosq (1998), Lemma 1.2], there exist i.i.d. random variables $Y_{l,T}^*$, $l = 1, ..., k$, such that $Y_{l,T}^* \stackrel{d}{=} Y_{l,T}$, $l = 1, ..., k$, and:

$$P\left(\left|Y_{l,T}^* - Y_{l,T}\right| > \xi_T\right) \leq 11\left(c_T/\xi_T\right)^{2/5} \alpha\left(q_T\right)^{4/5}, \qquad l = 1, ..., k,$$

where $\alpha(.)$ are the mixing coefficients of process $\left(X_t^{'}, Y_t^{'}\right)^{'}$. It will be proved below (see Lemma B.4) that $c_T = O\left((m/T)^{1/2}\right)$. Let $\varepsilon > 0$ be given and let $\xi_T := \varepsilon/k_T = o\left((m/T)^{1/2}\right)$. Thus, we have:

$$P\left(\left|Y_{l,T}^* - Y_{l,T}\right| > \varepsilon/k_T\right) = O\left(k_T^{2/5}\left(m/T\right)^{1/5} \alpha\left(q_T\right)^{4/5}\right), \qquad l = 1, ..., k.$$

We deduce:

$$\begin{aligned}
P\left(\left|\sum_{l=1}^{k} Y_{l,T}^* - \sum_{l=1}^{k} Y_{l,T}\right| > \varepsilon\right) &\leq P\left(\sum_{l=1}^{k} \left|Y_{l,T}^* - Y_{l,T}\right| > \varepsilon\right) \leq \sum_{l=1}^{k} P\left(\left|Y_{l,T}^* - Y_{l,T}\right| > \varepsilon/k_T\right) \\
&= O\left(k_T^{7/5}\left(m/T\right)^{1/5} \alpha\left(q_T\right)^{4/5}\right).
\end{aligned}$$

Since $\alpha(.)$ has geometric decay by Assumption A.5, $O\left(k_T^{7/5}(m/T)^{1/5}\alpha(q_T)^{4/5}\right)=o(1)$. The proof is concluded. ■

Thus, we have:

$$\sum_{l=1}^{k} Y_{l,T} = \sum_{l=1}^{k} Y_{l,T}^{*} + o_p(1). \tag{B.10}$$

The asymptotic normality of $\sum_{l=1}^{k} Y_{l,T}^{*}$:

$$\sum_{l=1}^{k} Y_{l,T}^{*} \xrightarrow{d} N\left(0,\sigma^2\right),$$

where $\sigma^2>0$ is given below, is proved by using Liapunov CLT [Billingsley (1965)]. For this purpose we show that:

$$\sum_{l=1}^{k} E\left[\left(Y_{l,T}^{*}\right)^2\right] \to \sigma^2, \qquad \sum_{l=1}^{k} E\left[\left(Y_{l,T}^{*}\right)^3\right] \to 0.$$

These two conditions are verified below in Lemma B.4 and Lemma B.5, respectively.

Lemma B.4: *Under Assumptions A.1, A.5-A.9 and A.11-A.14, we have:*

$$\sum_{l=1}^{k} E\left[\left(Y_{l,T}^{*}\right)^2\right] \to \lambda^{'}\Sigma\lambda,$$

where $\Sigma=(\Sigma_{ij})$ is the matrix with blocks:

$$\Sigma_{ij} = \left(\begin{array}{cc} \sum_{k=-\infty}^{\infty} Cov\left(g_1\left(X_t,Y_t;\theta_i\right), g_1\left(X_{t-k},Y_{t-k};\theta_j\right)\right) & 0 \\ 0 & w^2 f\left(x_0\right) E\left[\widetilde{g}_2\left(Y_t;\theta_i\right)\widetilde{g}_2\left(Y_t;\theta_j\right)^{'}|X_t=x_0\right] \end{array}\right).$$

Proof: We have:

$$\begin{aligned} \sum_{l=1}^{k} E\left[\left(Y_{l,T}^{*}\right)^2\right] &= kV\left[Y_{1,T}^{*}\right] = kV\left[\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{1}{\sqrt{T}}\lambda_i^{'}v_{t,T}\left(\theta_i\right)\right] \\ &= \frac{km}{T}\sum_{i,j=1}^{n}\lambda_i^{'}Cov\left[\frac{1}{\sqrt{m}}\sum_{t=1}^{m}v_{t,T}\left(\theta_i\right), \frac{1}{\sqrt{m}}\sum_{t=1}^{m}v_{t,T}\left(\theta_j\right)\right]\lambda_j. \end{aligned}$$

Since $km/T\to 1$, it is sufficient to prove that:

$$\Sigma_{T,ij} := Cov\left[\frac{1}{\sqrt{m}}\sum_{t=1}^{m}v_{t,T}\left(\theta_i\right), \frac{1}{\sqrt{m}}\sum_{t=1}^{m}v_{t,T}\left(\theta_j\right)\right] \to \Sigma_{ij}, \qquad i,j=1,...,n.$$

Let us write:

$$\Sigma_{T,ij} = \left(\begin{array}{cc} \Sigma_{T,ij}^{11} & \Sigma_{T,ij}^{12} \\ \Sigma_{T,ij}^{21} & \Sigma_{T,ij}^{22} \end{array}\right),$$

where:

$$\Sigma_{T,ij}^{11} = Cov\left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m} g_1\left(X_t, Y_t; \theta_i\right), \frac{1}{\sqrt{m}}\sum_{t=1}^{m} g_1\left(X_t, Y_t; \theta_j\right)\right),$$

$$\Sigma_{T,ij}^{12} = \Sigma_{T,ij}^{21'} = Cov\left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m} g_1\left(X_t, Y_t; \theta_i\right), \frac{1}{\sqrt{m}}\sum_{t=1}^{m} h_T^{-d/2}\widetilde{g}_2\left(Y_t; \theta_j\right) K\left(\frac{X_t - x_0}{h_T}\right)\right),$$

$$\Sigma_{T,ij}^{22} = Cov\left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m} h_T^{-d/2}\widetilde{g}_2\left(Y_t; \theta_i\right) K\left(\frac{X_t - x_0}{h_T}\right), \frac{1}{\sqrt{m}}\sum_{t=1}^{m} h_T^{-d/2}\widetilde{g}_2\left(Y_t; \theta_j\right) K\left(\frac{X_t - x_0}{h_T}\right)\right),$$

and derive the limit of each term for $T \to \infty$.

i) For $\Sigma_{T,ij}^{11}$ we have:

$$\Sigma_{T,ij}^{11} = Cov\left(g_1\left(X_t, Y_t; \theta_i\right), g_1\left(X_t, Y_t; \theta_j\right)\right) + \sum_{l:|l|=1}^{m-1}\left(1 - \frac{|l|}{m}\right) Cov\left(g_1\left(X_t, Y_t; \theta_i\right), g_1\left(X_{t-l}, Y_{t-l}; \theta_j\right)\right).$$

From Assumption A.11 and Cauchy-Schwarz inequality, we get:

$$E\left[\left\|g_1\left(X_t, Y_t; \theta_i\right)\right\|^{\bar{r}}\right] < \infty,$$

for $\bar{r} > 2$. Then, by Davidov inequality [Bosq (1998), Corollary 1.1] and Assumption A.5, we get:

$$\left\|Cov\left(g_1\left(X_t, Y_t; \theta_i\right), g_1\left(X_{t-l}, Y_{t-l}; \theta_j\right)\right)\right\| = O\left(\rho^l E\left[\left\|g_1\left(X_t, Y_t; \theta_i\right)\right\|^{\bar{r}}\right]^{1/\bar{r}} E\left[\left\|g_1\left(X_t, Y_t; \theta_j\right)\right\|^{\bar{r}}\right]^{1/\bar{r}}\right),$$

for some $0 < \rho < 1$. Thus, the cross-autocovariances are summable, and:

$$\lim_{T\to\infty} \Sigma_{T,ij}^{11} = \sum_{l=-\infty}^{\infty} Cov\left(g_1\left(X_t, Y_t; \theta_i\right), g_1\left(X_{t-l}, Y_{t-l}; \theta_j\right)\right).$$

ii) Let us now consider $\Sigma_{T,ij}^{22}$. We have:

$$\begin{aligned}
\Sigma_{T,ij}^{22} &= Cov\left(h_T^{-d/2}\widetilde{g}_2\left(Y_t; \theta_i\right) K\left(\frac{X_t - x_0}{h_T}\right), h_T^{-d/2}\widetilde{g}_2\left(Y_t; \theta_j\right) K\left(\frac{X_t - x_0}{h_T}\right)\right) \\
&\quad + \sum_{l:|l|=1}^{m-1}\left(1 - \frac{|l|}{m}\right) Cov\left(h_T^{-d/2}\widetilde{g}_2\left(Y_t; \theta_i\right) K\left(\frac{X_t - x_0}{h_T}\right), h_T^{-d/2}\widetilde{g}_2\left(Y_{t-l}; \theta_j\right) K\left(\frac{X_{t-l} - x_0}{h_T}\right)\right) \\
&\equiv \Gamma_{0T,ij} + \sum_{l:|l|=1}^{m-1}\left(1 - \frac{|l|}{m}\right)\Gamma_{lT,ij}. \qquad \text{(B.11)}
\end{aligned}$$

Let us first consider the covariance term $\Gamma_{0T,ij}$. The functions $E\left[\widetilde{g}_2\left(Y_t; \theta_j\right)|X_t = .\right] f(.)$ and $E\left[\widetilde{g}_2\left(Y_t; \theta_i\right)\widetilde{g}_2\left(Y_t; \theta_j\right)^{'}|X_t = .\right] f(.)$ are Lebesgue-integrable. Indeed, by applying twice Cauchy-Schwarz inequality, we get:

$$\begin{aligned}
&\int \left\|E\left[g_2^*\left(Y_t; \theta\right) g_2^*\left(Y_t; \tau\right)^{'}|X_t = x\right]\right\| f(x)dx \\
\leq &\int E\left[\left\|g_2^*\left(Y_t; \theta\right)\right\|^2|X_t = x\right]^{1/2} E\left[\left\|g_2^*\left(Y_t; \tau\right)\right\|^2|X_t = x\right]^{1/2} f(x)dx \\
\leq &\left(\int E\left[\left\|g_2^*\left(Y_t; \theta\right)\right\|^2|X_t = x\right] f(x)dx\right)^{1/2}\left(\int E\left[\left\|g_2^*\left(Y_t; \tau\right)\right\|^2|X_t = x\right] f(x)dx\right)^{1/2} \\
= &E\left[\left\|g_2^*\left(Y_t; \theta\right)\right\|^2\right]^{1/2} E\left[\left\|g_2^*\left(Y_t; \tau\right)\right\|^2\right]^{1/2} < \infty,
\end{aligned}$$

by Assumption A.11, and similarly:

$$\int \|E [g_2^* (Y_t; \theta) | X_t = x]\| f(x) dx \leq E \left[\|g_2^* (Y_t; \theta)\|^2 \right]^{1/2} < \infty.$$

Since the functions $E [\widetilde{g}_2 (Y_t; \theta_j) | X_t = .] f(.)$ and $E \left[\widetilde{g}_2 (Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' | X_t = . \right] f(.)$ are Lebesgue-integrable and continuous at $x = x_0$ (Assumption A.12-A.13), we can apply Bochner's Lemma [e.g. Bosq, Lecoutre (1987), p.61] to deduce that:

$$\begin{aligned} E \left[h_T^{-d} \widetilde{g}_2 (Y_t; \theta_j) K \left(\frac{X_t - x_0}{h_T} \right) \right] &= E [\widetilde{g}_2 (Y_t; \theta_j) | X_t = x_0] f (x_0) + o(1), \\ h_T^{-d} E \left[\widetilde{g}_2 (Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' K \left(\frac{X_t - x_0}{h_T} \right)^2 \right] &= w^2 E \left[\widetilde{g}_2 (Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' | X_t = x_0 \right] f (x_0) + o(1). \end{aligned}$$

Then:

$$\begin{aligned} \Gamma_{0T,ij} &= h_T^{-d} E \left[\widetilde{g}_2 (Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' K \left(\frac{X_t - x_0}{h_T} \right)^2 \right] + O \left(h_T^d \right) \\ &= w^2 f (x_0) E \left[\widetilde{g}_2 (Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' | X_t = x_0 \right] + o(1). \end{aligned}$$

Let us now consider the term $\sum_{l:|l|=1}^{m-1} \left(1 - \frac{|l|}{m} \right) \Gamma_{lT,ij}$ in equation (B.11). By repeating the argument used by Bosq (1998), proof of Theorem 2.3, in the case of density estimator, it is possible to prove that Assumptions A.5-A.9, A.11 and A.14 imply (see Section B.1.4 for the detailed derivation):

$$\sum_{l:|l|=1}^{m-1} \left(1 - \frac{|l|}{m} \right) \Gamma_{lT,ij} = o (1).$$

Finally, we conclude:

$$\lim_{T \to \infty} \Sigma_{T,ij}^{22} = w^2 f (x_0) E \left[\widetilde{g}_2 (Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' | X_t = x_0 \right].$$

iii) Finally, let us consider $\Sigma_{T,ij}^{12}$. By a similar argument as for $\Sigma_{T,ij}^{22}$, the cross-terms are negligible. Thus, using Assumptions A.1, A.8, A.9, A.13 and Bochner's Lemma, we get:

$$\begin{aligned} \Sigma_{T,ij}^{12} &= Cov \left(g_1 (X_t, Y_t; \theta_i), h_T^{-d/2} \widetilde{g}_2 (Y_t; \theta_j) K \left(\frac{X_t - x_0}{h_T} \right) \right) + o (1) \\ &= h_T^{-d/2} E \left[g_1 (X_t, Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' K \left(\frac{X_t - x_0}{h_T} \right) \right] + o(1) \\ &= h_T^{d/2} w^2 f(x_0) E \left[g_1 (X_t, Y_t; \theta_i) \widetilde{g}_2 (Y_t; \theta_j)' | X_t = x_0 \right] + o(1) \\ &= o(1). \end{aligned}$$

The proof is concluded. ■

Lemma B.5: *Under the assumptions of Lemma B.4 and Assumption A.15, the Liapunov condition holds:*

$$\sum_{l=1}^{k} E \left[\left(Y_{l,T}^* \right)^3 \right] \to 0.$$

Proof: By Cauchy-Schwarz we have:

$$\sum_{l=1}^{k} E\left[\left(Y_{l,T}^{*}\right)^{3}\right] = kE\left[\left(Y_{1,T}^{*}\right)^{3}\right] \leq \left(kE\left[\left(Y_{1,T}^{*}\right)^{4}\right]\right)^{1/2} \left(kE\left[\left(Y_{1,T}^{*}\right)^{2}\right]\right)^{1/2}.$$

Since $kE\left[\left(Y_{1,T}^{*}\right)^{2}\right] = O(1)$ by Lemma B.4, it is sufficient to prove:

$$kE\left[\left(Y_{1,T}^{*}\right)^{4}\right] = o(1).$$

Since:

$$\left(kE\left[\left(Y_{1,T}^{*}\right)^{4}\right]\right)^{1/4} \leq \sum_{i=1}^{n} \|\lambda_i\| \left(kE\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{m}\left(v_{tT}\left(\theta_i\right) - E\left[v_{tT}\left(\theta_i\right)\right]\right)\right\|^4\right]\right)^{1/4},$$

we have to show:

$$kE\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{m}\left(v_{tT}\left(\theta_i\right) - E\left[v_{tT}\left(\theta_i\right)\right]\right)\right\|^4\right] = o(1), \ \forall i = 1, ..., n. \tag{B.12}$$

We have:

$$kE\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{m}\left(v_{tT}\left(\theta_i\right) - E\left[v_{tT}\left(\theta_i\right)\right]\right)\right\|^4\right] = \frac{k}{T^2}E\left[\left\|\sum_{t=1}^{m} V_{tT,i}\right\|^4\right] \leq \frac{k}{T^2}E\left[\left(\sum_{t=1}^{m}\|V_{tT,i}\|\right)^4\right], \tag{B.13}$$

where $V_{tT,i} := v_{tT}\left(\theta_i\right) - E\left[v_{tT}\left(\theta_i\right)\right]$. Moreover,

$$\begin{aligned} E\left[\left(\sum_{t=1}^{m}\|V_{tT,i}\|\right)^4\right] &= \sum_{t} E\left[\|V_{tT,i}\|^4\right] \\ &+ \sum_{t_1 \neq t_2} E\left[\|V_{t_1T,i}\|^2 \|V_{t_2T,i}\| \left(\|V_{t_1T,i}\| + \|V_{t_2T,i}\|\right)\right] \\ &+ \sum_{t_1 \neq t_2 \neq t_3} E\left[\|V_{t_1T,i}\|^2 \|V_{t_2T,i}\| \|V_{t_3T,i}\|\right] \\ &+ \sum_{t_1 \neq t_2 \neq t_3 \neq t_4} E\left[\|V_{t_1T,i}\| \|V_{t_2T,i}\| \|V_{t_3T,i}\| \|V_{t_4T,i}\|\right], \end{aligned} \tag{B.14}$$

where summations are over $1, ..., m$. Let us now derive the orders of the different terms. Since:

$$\begin{aligned} \|V_{tT,i}\| &\leq \|v_{tT}\left(\theta_i\right)\| + E\left[\|v_{tT}\left(\theta_i\right)\|\right], \\ \|v_{tT}\left(\theta_i\right)\| &\leq \|g_1\left(X_t, Y_t; \theta_i\right)\| + h_T^{-d/2}\|\widetilde{g}_2\left(Y_t; \theta_i\right)\| \left|K\left(\frac{X_t - x_0}{h_T}\right)\right|, \\ E\left[\|v_{tT}\left(\theta_i\right)\|\right] &= O(1), \end{aligned}$$

and by Assumption A.11, the leading terms are either of order $O(1)$ or the terms involving the highest power of $h_T^{-d/2}\|\widetilde{g}_2\left(Y_t; \theta_i\right)\| \left|K\left(X_t - x_0/h_T\right)\right|$.

i) We have:

$$
\begin{aligned}
E\left[\|V_{tT,i}\|^4\right] &= O\left(h_T^{-2d}E\left[\|\widetilde{g}_2\left(Y_t;\theta_i\right)\|^4 K\left(\frac{X_t-x_0}{h_T}\right)^4\right]\right) \\
&= O\left(h_T^{-2d}\,\|K\|_\infty^3\,E\left[\|\widetilde{g}_2\left(Y_t;\theta_i\right)\|^4\left|K\left(\frac{X_t-x_0}{h_T}\right)\right|\right]\right).
\end{aligned}
$$

From Assumption A.15:

$$
h_T^{-d}E\left[\|\widetilde{g}_2\left(Y_t;\theta_i\right)\|^4\left|K\left(\frac{X_t-x_0}{h_T}\right)\right|\right] = \int E\left[\|\widetilde{g}_2\left(Y_t;\theta_i\right)\|^4\,|X_t = x_0+h_Tu\right] f(x_0+h_Tu)\,|K(u)|\,du = O(1).
$$

Thus, we get:

$$
E\left[\|V_{tT,i}\|^4\right] = O\left(h_T^{-d}\right). \tag{B.15}
$$

ii) We have:

$$
E\left[\|V_{t_1T,i}\|^3\|V_{t_2T,i}\|\right] = O\left(h_T^{-2d}\,\|K\|_\infty^2\,E\left[\|\widetilde{g}_2\left(Y_{t_1};\theta_i\right)\|^3\|\widetilde{g}_2\left(Y_{t_2};\theta_i\right)\|\left|K\left(\frac{X_{t_1}-x_0}{h_T}\right)\right|\left|K\left(\frac{X_{t_2}-x_0}{h_T}\right)\right|\right]\right).
$$

From Assumption A.15:

$$
\begin{aligned}
& h_T^{-2d}E\left[\|\widetilde{g}_2\left(Y_{t_1};\theta_i\right)\|^3\|\widetilde{g}_2\left(Y_{t_2};\theta_i\right)\|\left|K\left(\frac{X_{t_1}-x_0}{h_T}\right)\right|\left|K\left(\frac{X_{t_2}-x_0}{h_T}\right)\right|\right] \\
= & \int E\left[\|\widetilde{g}_2\left(Y_{t_1};\theta_i\right)\|^3\|\widetilde{g}_2\left(Y_{t_2};\theta_i\right)\|\,|X_{t_1}=x_0+h_Tu, X_{t_2}=x_0+h_Tv\right] f_{t_1,t_2}(x_0+h_Tu,x_0+h_Tv)\,|K(u)K(v)|\,dudv \\
= & O(1).
\end{aligned}
$$

Thus, we get:

$$
E\left[\|V_{t_1T,i}\|^3\|V_{t_2T,i}\|\right] = O\left(1\right). \tag{B.16}
$$

iii) Similarly, we have:

$$
\begin{aligned}
E\left[\|V_{t_1T,i}\|^2\|V_{t_2T,i}\|^2\right] &= O(1), \quad t_1\neq t_2, \\
E\left[\|V_{t_1T,i}\|^2\|V_{t_2T,i}\|\,\|V_{t_3T,i}\|\right] &= O\left(1\right), \quad t_1\neq t_2\neq t_3, \\
E\left[\|V_{t_1T,i}\|\,\|V_{t_2T,i}\|\,\|V_{t_3T,i}\|\,\|V_{t_4T,i}\|\right] &= O\left(1\right), \quad t_1\neq t_2\neq t_3\neq t_4.
\end{aligned} \tag{B.17}
$$

Therefore, from (B.13)-(B.17), we get:

$$
\begin{aligned}
kE\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{m}\left(v_{tT}\left(\theta_i\right)-E\left[v_{tT}\left(\theta_i\right)\right]\right)\right\|^4\right] &= O\left(\frac{k}{T^2}\left(mh_T^{-d}+m^4\right)\right) \\
&= O\left(\frac{m^3}{T}\right),
\end{aligned}
$$

since $km/T\to 1$, $m/T\to 0$, $Th_T^d\to\infty$. (B.12) follows if we choose m such that $m\to\infty$ and $m^3/T\to 0$. ■

From (B.10), Lemma B.4 and Lemma B.5 we conclude that:

$$
\sum_{l=1}^{k} Y_{l,T} \xrightarrow{d} N\left(0,\lambda^{'}\Sigma\lambda\right).
$$

b) The last two terms of the decomposition (B.9) are negligible

Lemma B.6: *Under the assumptions of Lemma B.4,*

$$\sum_{l=1}^{k} Y_{l,T}^{'} = o_p(1), \qquad Y_T^{''} = o_p(1).$$

Proof: The proof is similar to the proof of Theorem 1.3.10 in Tenreiro (1995), p. 14.
i) We have:

$$Y_{l,T}^{'} = \sum_{t=lm+(l-1)q+1}^{l(m+q)} Z_{t,T} = \sum_{i=1}^{n} \lambda_i^{'} \left[\frac{1}{\sqrt{T}} \sum_{t=lm+(l-1)q+1}^{l(m+q)} \left(v_{t,T}\left(\theta_i\right) - E\left[v_{t,T}\left(\theta_i\right)\right]\right) \right] \equiv \sum_{i=1}^{n} \lambda_i^{'} U_{lT,i}.$$

Thus:

$$E\left[\left(\sum_{l=1}^{k} Y_{l,T}^{'}\right)^2\right]^{1/2} = E\left[\left(\sum_{i=1}^{n} \lambda_i^{'} \left(\sum_{l=1}^{k} U_{lT,i}\right)\right)^2\right]^{1/2} \leq \sum_{i=1}^{n} \|\lambda_i\| \, E\left[\left\|\sum_{l=1}^{k} U_{lT,i}\right\|^2\right]^{1/2}.$$

Therefore, it is sufficient to prove:

$$E\left[\left\|\sum_{l=1}^{k} U_{lT,i}\right\|^2\right] = o(1), \quad \forall i. \tag{B.18}$$

We have:

$$E\left[\left\|\sum_{l=1}^{k} U_{lT,i}\right\|^2\right] = kE\left[U_{lT,i}^{'} U_{lT,i}\right] + \sum_{|s|=1}^{k-1} (k-|s|) E\left(U_{lT,i}^{'} U_{l-sT,i}\right),$$

and:

$$kE\left[U_{lT,i}^{'} U_{lT,i}\right] = kTr\left[V\left(U_{lT,i}\right)\right] = \frac{kq}{T} Tr\left[V\left(\frac{1}{\sqrt{q}} \sum_{t=1}^{q} v_{t,T}\left(\theta_i\right)\right)\right] \equiv \frac{kq}{T} Tr\left(\widetilde{\Sigma}_{T,ii}\right).$$

From the proof of Lemma B.4, $\widetilde{\Sigma}_{T,ii} = O(1)$. Since $kq/T = o(1)$, we get:

$$kE\left[U_{lT,i}^{'} U_{lT,i}\right] = o(1).$$

Moreover,

$$\begin{aligned}
\left|\sum_{|s|=1}^{k-1} (k-|s|) E\left(U_{lT,i}^{'} U_{l-sT,i}\right)\right| &\leq k \sum_{|s|=1}^{k-1} \left|E\left(U_{lT,i}^{'} U_{l-s,T,i}\right)\right| \\
&\leq \frac{2kq}{T} \sum_{s=1}^{\infty} \left|E\left(v_{t,T}\left(\theta_i\right)^{'} v_{t-s,T}\left(\theta_i\right)\right) - E\left(v_{t,T}\left(\theta_i\right)\right)^{'} E\left(v_{t-s,T}\left(\theta_i\right)\right)\right| \\
&\leq \frac{2kq}{T} \sum_{s=1}^{\infty} \left\|Cov\left(v_{t,T}\left(\theta_i\right), v_{t-s,T}\left(\theta_i\right)\right)\right\|.
\end{aligned}$$

Using the same argument as in the proof of Lemma B.4 (see also Section B.1.4), we can show that $\sum_{s=1}^{\infty} \|Cov\left(v_{t,T}\left(\theta_i\right), v_{t-s,T}\left(\theta_i\right)\right)\| < \infty$. Thus,

$$\sum_{|s|=1}^{k-1} (k-|s|) E\left(U_{lT,i}^{'} U_{l-sT,i}\right) = o(1).$$

Then (B.18) follows.

ii) We have:

$$Y_T^{''} = \sum_{t=k(m+q)+1}^{T} Z_{t,T} = \sum_{i=1}^{n} \lambda_i^{'} \left[\frac{1}{\sqrt{T}} \sum_{t=k(m+q)+1}^{T} \left(v_{t,T}\left(\theta_i\right) - E\left[v_{t,T}\left(\theta_i\right)\right]\right) \right] \equiv \sum_{i=1}^{n} \lambda_i^{'} U_{T,i},$$

and:

$$E\left[\left(Y_T^{''}\right)^2\right]^{1/2} \leq \sum_{i=1}^{n} \|\lambda_i\| \, E\left[\|U_{T,i}\|^2\right]^{1/2}.$$

Therefore, it is sufficient to prove:

$$E\left[\|U_{T,i}\|^2\right] = o(1), \quad \forall i.$$

We have:

$$E\left[U_{T,i} U_{T,i}^{'}\right] = \frac{T-k(m+q)}{T} V \left[\frac{1}{\sqrt{T-k(m+q)}} \sum_{t=1}^{T-k(m+q)} v_{t,T}\left(\theta_i\right) \right] \to 0,$$

and the proof is concluded. ■

B.1.3 Stochastic equicontinuity

Let us now prove the stochastic equicontinuity of empirical process $\nu_T(\theta)$ [condition ii) in Proposition B.2] along the lines of Theorem 1 in Andrews (1991). Let us introduce the matrix-valued triangular array:

$$W_{t,T} = \begin{pmatrix} Z_t & 0 \\ 0 & h_T^{-d/2} K\left(\frac{X_t - x_0}{h_T}\right) Id_{K_2+L+1} \end{pmatrix}, \; t \leq T, \quad T \geq 1,$$

where Z_t denotes the instrument. We can write:

$$v_{t,T}(\theta) = W_{t,T} \psi\left(Y_t; \theta\right), \quad \theta \in \Theta,$$

where

$$\psi\left(y;\theta\right) = \left(g\left(y;\theta\right)^{'}, \widetilde{g}_2\left(y;\theta\right)^{'}\right)^{'}, \quad \theta \in \Theta.$$

Let $\left\{\psi_j : j \in \mathbb{N}\right\}$ be the basis of $L^2(F_Y)$ introduced in Assumption A.16. Without loss of generality, we can set $\psi_1(y) = 1$. Thus, from Assumption A.16, there exist sequences $\left\{c_j^*(\theta) : j \in \mathbb{N}\right\}$, $\theta \in \Theta$, of vector coefficients such that:

$$\psi\left(y;\theta\right) = \sum_{j=1}^{\infty} c_j^*(\theta) \psi_j(y), \quad y \in \mathcal{Y},$$

for any $\theta \in \Theta$, where

$$\lim_{J\to\infty} \sup_{\theta\in\Theta} \sum_{j=J}^{\infty} \frac{1}{\lambda_j} \left\| c_j^*(\theta) \right\|^2 < \infty.$$

Thus, we have:

$$\begin{aligned} \nu_T(\theta) - \nu_T(\tau) &= \sum_{j=1}^{\infty} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(W_{t,T}\psi_j(Y_t) - E\left[W_{t,T}\psi_j(Y_t) \right] \right) \right) \left[c_j^*(\theta) - c_j^*(\tau) \right] \\ &= \sum_{j=1}^{\infty} \left(T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right) \left[c_j^*(\theta) - c_j^*(\tau) \right], \end{aligned}$$

where $X_{j,tT} := W_{t,T}\psi_j(Y_t) - E\left[W_{t,T}\psi_j(Y_t) \right]$, and:

$$\left\| \nu_T(\theta) - \nu_T(\tau) \right\|^2 \leq \sum_{j=1}^{\infty} \lambda_j \left\| T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right\|^2 \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \left\| c_j^*(\theta) - c_j^*(\tau) \right\|^2. \tag{B.19}$$

Let $d(.,.)$ denote the metric on Θ defined by:

$$d(\theta,\tau) = \left(\sum_{j=1}^{\infty} \left\| c_j^*(\theta) - c_j^*(\tau) \right\|^2 \right)^{1/2}, \quad \theta,\tau \in \Theta.$$

For any $\eta, \delta > 0$, we have:

$$\begin{aligned} & \limsup_{T\to\infty} P^* \left[\sup_{\theta,\tau\in\Theta:\ d(\theta,\tau)\leq\delta} \left\| \nu_T(\theta) - \nu_T(\tau) \right\| > \eta \right] \\ \leq\ & \frac{1}{\eta^2} \limsup_{T\to\infty} E^* \left[\sup_{\theta,\tau\in\Theta:\ d(\theta,\tau)\leq\delta} \left\| \nu_T(\theta) - \nu_T(\tau) \right\|^2 \right] \\ \leq\ & \frac{1}{\eta^2} \left(\sup_{\theta,\tau\in\Theta:\ d(\theta,\tau)\leq\delta} \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \left\| c_j^*(\theta) - c_j^*(\tau) \right\|^2 \right) \limsup_{T\to\infty} \sum_{j=1}^{\infty} \lambda_j E \left[\left\| T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right\|^2 \right], \end{aligned} \tag{B.20}$$

using (B.19). Since:

$$\begin{aligned} E\left[\left\| T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right\|^2 \right] &= Tr\left(E\left[\left(T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right) \left(T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right)' \right] \right) \\ &= Tr\left(E\left[X_{j,tT} X_{j,tT}' \right] \right) + \sum_{|k|=1}^{T-1} \left(1 - \frac{|k|}{T} \right) Tr\left(E\left[X_{j,tT} X_{j,t-k,T}' \right] \right), \end{aligned} \tag{B.21}$$

we can study the asymptotic behaviour of the different terms in the decomposition.

Lemma B.7: *Under Assumptions A.5, A.8-A.9 and A.17-A.18,*

$$\begin{aligned} E\left[X_{j,tT} X_{j,tT}' \right] &= \begin{pmatrix} V\left[Z_t \psi_j(Y_t) \right] & 0 \\ 0 & w^2 E\left[\psi_j(Y_t)^2 \,|X_t = x_0 \right] f(x_0) Id_{K_2+L+1} \end{pmatrix} + u_{j,T}, \\ E\left[X_{j,tT} X_{j,t-k,T}' \right] &= \begin{pmatrix} Cov\left(Z_t\psi_j(Y_t), Z_{t-k}\psi_j(Y_{t-k}) \right) & 0 \\ 0 & 0 \end{pmatrix} + u_{j,kT}, \qquad k \neq 0, \end{aligned}$$

where $V\left[Z_t\psi_j\left(Y_t\right)\right]:=E\left[Z_tZ_t'\psi_j\left(Y_t\right)^2\right]-E\left[Z_t\psi_j\left(Y_t\right)\right]E\left[Z_t\psi_j\left(Y_t\right)\right]'$, $Cov\left(Z_t\psi_j\left(Y_t\right),Z_{t-k}\psi_j\left(Y_{t-k}\right)\right):=$ $E\left[Z_tZ_{t-k}'\psi_j\left(Y_t\right)\psi_j\left(Y_{t-k}\right)\right]-E\left[Z_t\psi_j\left(Y_t\right)\right]E\left[Z_{t-k}\psi_j\left(Y_{t-k}\right)\right]'$, *and:*

$$\sup_j\|u_{j,T}\|=o\left(1\right),\qquad \sup_j\sum_{k=1}^{T}\|u_{j,kT}\|=o\left(1\right).$$

Proof: i) We have:

$$E\left[X_{j,tT}X_{j,tT}'\right]=\begin{pmatrix} E\left[Z_tZ_t'\psi_j\left(Y_t\right)^2\right]-E\left[Z_t\psi_j\left(Y_t\right)\right]E\left[Z_t\psi_j\left(Y_t\right)\right]' & 0 \\ 0 & h_T^{-d}V\left[\psi_j\left(Y_t\right)K\left(\frac{X_t-x_0}{h_T}\right)\right]Id_{K_2+L+1}\end{pmatrix}.$$

Let us consider the lower right block. The term:

$$h_T^{-d}E\left[\psi_j\left(Y_t\right)K\left(\frac{X_t-x_0}{h_T}\right)\right]=\int E\left[\psi_j\left(Y_t\right)|X_t=x_0+h_Tu\right]f\left(x_0+h_Tu\right)K(u)du,$$

is bounded uniformly in $j\in\mathbb{N}$ from Assumption A.17 and the Cauchy-Schwartz inequality. Moreover, from standard bias expansion and Assumption A.17:

$$h_T^{-d}E\left[\psi_j\left(Y_t\right)^2K\left(\frac{X_t-x_0}{h_T}\right)^2\right]=\int\varphi_j(x_0+h_Tu)K\left(u\right)^2du=w^2\varphi_j(x_0)+O\left(\sup_j\left\|D^2\varphi_j\right\|_\infty h_T^2\right).$$

Thus:

$$h_T^{-d}V\left[\psi_j\left(Y_t\right)K\left(\frac{X_t-x_0}{h_T}\right)\right]=w^2E\left[\psi_j\left(Y_t\right)^2|X_t=x_0\right]f(x_0)+o(1),$$

uniformly in $j\in\mathbb{N}$.

ii) We have:

$$E\left[X_{j,tT}X_{j,t-kT}'\right]=\begin{pmatrix}\Omega_{kT,j}^{11} & 0\\ 0 & \Omega_{kT,j}^{22}\end{pmatrix},$$

where:

$$\begin{aligned}\Omega_{kT,j}^{11} &= E\left[Z_tZ_{t-k}'\psi_j\left(Y_t\right)\psi_j\left(Y_{t-k}\right)\right]-E\left[Z_t\psi_j\left(Y_t\right)\right]E\left[Z_{t-k}\psi_j\left(Y_{t-k}\right)\right]',\\ \Omega_{kT,j}^{22} &= h_T^{-d}Cov\left[\psi_j\left(Y_t\right)K\left(\frac{X_t-x_0}{h_T}\right),\psi_j\left(Y_{t-k}\right)K\left(\frac{X_{t-k}-x_0}{h_T}\right)\right]Id_{K_2+L+1}.\end{aligned}$$

Let us consider $\Omega_{kT,j}^{22}$. We can use the same arguments as in the proof of Lemma B.4 to get bounds uniform in $j\in\mathbb{N}$ from Assumptions A.5, A.8, A.9 and A.18. Thus,

$$\sum_{k=1}^{T-1}\left|h_T^{-d}Cov\left[\psi_j\left(Y_t\right)K\left(\frac{X_t-x_0}{h_T}\right),\psi_j\left(Y_{t-k}\right)K\left(\frac{X_{t-k}-x_0}{h_T}\right)\right]\right|=o(1),$$

uniformly in $j\in\mathbb{N}$. The proof is concluded. ∎

From Davidov inequality, we have:

$$\left\|Cov\left(Z_t\psi_j\left(Y_t\right),Z_{t-k}\psi_j\left(Y_{t-k}\right)\right)\right\|\leq const\cdot\rho^kE\left[\left\|Z_t\psi_j\left(Y_t\right)\right\|^r\right]^{2/r}, \tag{B.22}$$

uniformly in $j \in \mathbb{N}$, for some $0 < \rho < 1$ and $r > 2$ as in Assumption A.16. Thus, from Lemma B.7 and equations (B.21), (B.22) we get:

$$
\begin{aligned}
& \lim\sup_{T\to\infty} \sum_{j=1}^{\infty} \lambda_j E\left[\left\| T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right\|^2\right] \\
\leq \quad & \sum_{j=1}^{\infty} \lambda_j \left\{ Tr\left(V\left[Z_t \psi_j\left(Y_t\right)\right]\right) + C_1 E\left[\left\|Z_t \psi_j\left(Y_t\right)\right\|^r\right]^{2/r} + C_2 E\left[\psi_j\left(Y_t\right)^2 |X_t = x_0\right]\right\},
\end{aligned}
$$

for some constants $C_1, C_2 < \infty$. Now:

$$
Tr\left(V\left[Z_t \psi_j\left(Y_t\right)\right]\right) = Tr\left(E\left[Z_t Z_t^{'} \psi_j\left(Y_t\right)^2\right]\right) - Tr\left(E\left[Z_t \psi_j\left(Y_t\right)\right] E\left[Z_t \psi_j\left(Y_t\right)\right]^{'}\right).
$$

Since:

$$
Tr\left(E\left[Z_t Z_t^{'} \psi_j\left(Y_t\right)^2\right]\right) = E\left[\left\|Z_t \psi_j\left(Y_t\right)\right\|^2\right],
$$

and:

$$
Tr\left(E\left[Z_t \psi_j\left(Y_t\right)\right] E\left[Z_t \psi_j\left(Y_t\right)\right]^{'}\right) = \left\|E\left[Z_t \psi_j\left(Y_t\right)\right]\right\|^2 \leq E\left[\left\|Z_t \psi_j\left(Y_t\right)\right\|^2\right],
$$

we have:

$$
Tr\left(V\left[Z_t \psi_j\left(Y_t\right)\right]\right) \leq 2E\left[\left\|Z_t \psi_j\left(Y_t\right)\right\|^2\right] \leq 2E\left[\left\|Z_t \psi_j\left(Y_t\right)\right\|^r\right]^{2/r}.
$$

Thus, we get:

$$
\begin{aligned}
\lim\sup_{T\to\infty} \sum_{j=1}^{\infty} \lambda_j E\left[\left\| T^{-1/2} \sum_{t=1}^{T} X_{j,tT} \right\|^2\right] & \leq C_3 \sum_{j=1}^{\infty} \lambda_j \left\{ E\left[\left\|Z_t \psi_j\left(Y_t\right)\right\|^r\right]^{2/r} + E\left[\psi_j\left(Y_t\right)^2 |X_t = x_0\right]\right\} \\
& = C_4 < \infty,
\end{aligned}
$$

for some constants $C_3, C_4 < \infty$ from Assumption A.16. We deduce from (B.20):

$$
\lim\sup_{T\to\infty} P^* \left[\sup_{\theta,\tau\in\Theta:\ d(\theta,\tau)\leq\delta} \left\|\nu_T\left(\theta\right) - \nu_T\left(\tau\right)\right\| > \eta\right] \leq C_4 \frac{1}{\eta^2} \left(\sup_{\theta,\tau\in\Theta:\ d(\theta,\tau)\leq\delta} \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \left\|c_j^*\left(\theta\right) - c_j^*\left(\tau\right)\right\|^2\right).
$$

The conclusion follows from:

$$
\lim_{\delta\to 0} \sup_{\theta,\tau\in\Theta:\ d(\theta,\tau)\leq\delta} \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \left\|c_j^*\left(\theta\right) - c_j^*\left(\tau\right)\right\|^2 = 0,
$$

which is proved by Andrews (1991), Equation (2.6).

B.1.4 Bounds on covariance terms

In this section we derive a bound for the covariance term $\sum_{l:|l|=1}^{m-1} \left(1 - \frac{|l|}{m}\right) \Gamma_{lT,ij}$ in equation (B.11). This is done by deriving two bounds for the covariance terms. i) For this purpose, let us define functions:

$$
\begin{aligned}
\phi_i\left(x\right) &= E\left[\widetilde{g}_2\left(Y_t;\theta_i\right) |X_t = x\right] f(x), \\
\phi_{l,ij}\left(x,\xi\right) &= E\left[\widetilde{g}_2\left(Y_t;\theta_i\right) \widetilde{g}_2\left(Y_{t-l};\theta_j\right)^{'} |X_t = x, X_{t-l} = \xi\right] f_{t,t-l}(x,\xi).
\end{aligned}
$$

We can write:

$$
\begin{aligned}
& Cov\left(\widetilde{g}_2\left(Y_t;\theta_i\right)K\left(\frac{X_t-x_0}{h_T}\right),\widetilde{g}_2\left(Y_{t-l};\theta_j\right)K\left(\frac{X_{t-l}-x_0}{h_T}\right)\right) \\
= & \int\int \phi_{l,ij}\left(x,\xi\right)K\left(\frac{x-x_0}{h_T}\right)K\left(\frac{\xi-x_0}{h_T}\right)dxd\xi \\
& -\int \phi_i\left(x\right)K\left(\frac{x-x_0}{h_T}\right)dx\int \phi_j\left(\xi\right)^{'}K\left(\frac{\xi-x_0}{h_T}\right)d\xi \\
= & h_T^{2d}\left(\int\int \phi_{l,ij}\left(x_0+h_Tu,x_0+h_Tv\right)K\left(u\right)K\left(v\right)dudv\right. \\
& \left.-\int \phi_i\left(x_0+h_Tu\right)K\left(u\right)du\int \phi_j\left(x_0+h_Tv\right)^{'}K\left(v\right)dv\right).
\end{aligned}
$$

From Assumptions A.6-A.7 and A.14, and by the Cauchy-Schwartz inequality, function ϕ_i and $\phi_{l,ij}$ are bounded uniformly in $l \in \mathbb{N}$. We get:

$$
\left\|\Gamma_{lT,ij}\right\| \leq \left(\sup_l \left\|\phi_{l,ij}\right\|_\infty + \left\|\phi_i\right\|_\infty \left\|\phi_j\right\|_\infty\right)\left\|K\right\|_{L^1}^2 h_T^d \equiv C_1 h_T^d. \tag{B.23}
$$

ii) From the strong mixing property (Assumption A.5) and Davidov inequality:

$$
\begin{aligned}
& \left\|Cov\left(\widetilde{g}_2\left(Y_t;\theta_i\right)K\left(\frac{X_t-x_0}{h_T}\right),\widetilde{g}_2\left(Y_{t-l};\theta_j\right)K\left(\frac{X_{t-l}-x_0}{h_T}\right)\right)\right\| \\
\leq & \; const\cdot\rho^l\cdot E\left[\left\|\widetilde{g}_2\left(Y_t;\theta_i\right)K\left(\frac{X_t-x_0}{h_T}\right)\right\|^{\bar{r}}\right]^{1/\bar{r}} E\left[\left\|\widetilde{g}_2\left(Y_{t-l};\theta_j\right)K\left(\frac{X_{t-l}-x_0}{h_T}\right)\right\|^{\bar{r}}\right]^{1/\bar{r}},
\end{aligned}
$$

for some $0 < \rho < 1$ and $\bar{r} > 2$. Moreover we have:

$$
E\left[\left\|\widetilde{g}_2\left(Y_t;\theta_i\right)K\left(\frac{X_t-x_0}{h_T}\right)\right\|^{\bar{r}}\right] \leq \left\|K\right\|_\infty^{\bar{r}} E\left[\left\|\widetilde{g}_2\left(Y_t;\theta_i\right)\right\|^{\bar{r}}\right] =: \left\|K\right\|_\infty^{\bar{r}} c_i < \infty,
$$

from Assumptions A.8 and A.11. We deduce:

$$
\left\|\Gamma_{lT,ij}\right\| \leq C_2\rho^l h_T^{-d}, \tag{B.24}
$$

for some constant $C_2 < \infty$ (that depends on i and j).

Let us now define $L_T = \lfloor h_T^{-d/2}\rfloor \to \infty$. From (B.23) and (B.24) we have:

$$
\begin{aligned}
\left\|\sum_{l:|l|=1}^{m-1}\left(1-\frac{|l|}{m}\right)\Gamma_{lT,ij}\right\| & \leq 2\sum_{l=1}^{m-1}\left\|\Gamma_{lT,ij}\right\| \\
& \leq 2\left(\sum_{l=1}^{L_T} C_1 h_T^d + \sum_{l=L_T+1}^{\infty} C_2\rho^l h_T^{-d}\right) \\
& = 2\left(C_1 L_T h_T^d + \frac{C_2}{1-\rho}h_T^{-d}\rho^{L_T+1}\right) \\
& \leq const\left(1/L_T + L_T^2\rho^{L_T+1}\right) \to 0.
\end{aligned}
$$

We deduce $\sum_{l:|l|=1}^{m-1}\left(1-\frac{|l|}{m}\right)\Gamma_{lT,ij} = o(1)$.

B.2 Proof of consistency

In this Section we prove that $P\left[\left\|\widehat{\theta}_T^* - \theta_0^*\right\| \geq \varepsilon\right] \to 0$, as $T \to \infty$, for any $\varepsilon > 0$. We have:

$$
\begin{aligned}
P\left[\left\|\widehat{\theta}_T^* - \theta_0^*\right\| \geq \varepsilon\right] &\leq P\left[\inf_{\theta^* \in \Theta \times B: \left\|\theta^* - \theta_0^*\right\| \geq \varepsilon} Q_T(\theta^*) \leq Q_T\left(\widehat{\theta}_T^*\right)\right] \\
&\leq P\left[\inf_{\theta^* \in \Theta \times B: \left\|\theta^* - \theta_0^*\right\| \geq \varepsilon} Q_T(\theta^*) \leq Q_T(\theta_0^*)\right]. \qquad \text{(B.25)}
\end{aligned}
$$

Let us derive the orders of the RHS term and the LHS term inside the probability. Write the criterion as:

$$
Q_T(\theta^*) = \left[\Psi_T(\theta) + m_T(\theta^*)\right]' \Omega \left[\Psi_T(\theta) + m_T(\theta^*)\right], \quad \theta^* \in \Theta \times B. \qquad \text{(B.26)}
$$

Then, since $\Psi_T(\theta_0) = O_p(1)$ from Lemma A.1, and $m_T(\theta_0^*) = 0$, we get:

$$
Q_T(\theta_0^*) = O_p(1). \qquad \text{(B.27)}
$$

Let us now derive the order of $\inf_{\theta^* \in \Theta \times B: \left\|\theta^* - \theta_0^*\right\| \geq \varepsilon} Q_T(\theta^*)$. From Lemma A.1 and the Continuous Mapping Theorem [CMT, Billingsley (1968)], we have:

$$
\begin{aligned}
\sup_{\theta \in \Theta} \Psi_T(\theta)' \Omega \Psi_T(\theta) &= O_p(1), \\
\sup_{\theta^* \in \Theta \times B} m_T(\theta^*)' \Omega \Psi_T(\theta) &= O_p(\sqrt{T}).
\end{aligned}
$$

From (B.26) it follows that:

$$
Q_T(\theta^*) = m_T(\theta^*)' \Omega m_T(\theta^*) + O_p(\sqrt{T}),
$$

uniformly in $\theta^* \in \Theta \times B$. Now, let $\lambda > 0$ be the smallest eigenvalue of Ω (Assumption A.20). We get:

$$
\begin{aligned}
& m_T(\theta^*)' \Omega m_T(\theta^*) \\
\geq\ & T\lambda \left(\left\|E\left[g_1(Y_t, X_t; \theta)\right]\right\|^2 + h_T^d \left\|E\left[g_2(Y_t; \theta) \mid X_t = x_0\right]\right\|^2 + h_T^d \left\|E\left[a(Y_t; \theta) \mid X_t = x_0\right] - \beta\right\|^2 \right) \\
\geq\ & T h_T^d \lambda \left(\left\|E\left[g_1(Y_t, X_t; \theta)\right]\right\|^2 + \left\|E\left[g_2(Y_t; \theta) \mid X_t = x_0\right]\right\|^2 + \left\|E\left[a(Y_t; \theta) \mid X_t = x_0\right] - \beta\right\|^2 \right),
\end{aligned}
$$

for T large, and any $\theta^* \in \Theta \times B$. From continuity of moment functions (Assumption A.19), compactness of $\Theta \times B$ (Assumption A.4) and global identification (Assumption A.2), we have:

$$
\inf_{\theta^* \in \Theta \times B: \left\|\theta^* - \theta_0^*\right\| \geq \varepsilon} m_T(\theta^*)' \Omega m_T(\theta^*) \geq C T h_T^d,
$$

for a constant $C = C_\varepsilon > 0$. From bandwidth Assumption A.9, we have $\sqrt{T} = o\left(T h_T^d\right)$. Thus, we get:

$$
\inf_{\theta^* \in \Theta \times B: \left\|\theta^* - \theta_0^*\right\| \geq \varepsilon} Q_T(\theta^*) \geq \frac{1}{2} C T h_T^d, \qquad \text{(B.28)}
$$

with probability approching 1. Since $T h_T^d \to \infty$ from Assumption A.9, and by using (B.25), (B.27) and (B.28), the conclusion follows.

B.3 Proof of Lemma A.2

We use the following Lemma.

Lemma B.8: *Under Assumptions A.1-A.20 and A.24:* $\left\|\widehat{\theta}_T^* - \theta_0^*\right\| = O_p\left(1/\sqrt{Th_T^d}\right)$.

Proof: We follow the approach in the proof of Lemma A1 in Stock, Wright (2000). Since $\widehat{\theta}_T^*$ is the minimizer of Q_T we have:

$$Q_T\left(\widehat{\theta}_T^*\right) - Q_T\left(\theta_0^*\right) = \left[\Psi_T(\widehat{\theta}_T) + m_T(\widehat{\theta}_T^*)\right]' \Omega \left[\Psi_T(\widehat{\theta}_T) + m_T(\widehat{\theta}_T^*)\right] - \Psi_T(\theta_0)'\Omega\Psi_T(\theta_0) \leq 0,$$

that is,

$$m_T(\widehat{\theta}_T^*)'\Omega m_T(\widehat{\theta}_T^*) + 2m_T(\widehat{\theta}_T^*)'\Omega\Psi_T(\widehat{\theta}_T) + d_{1,T} \leq 0,$$

where $d_{1,T} = \Psi_T(\widehat{\theta}_T)'\Omega\Psi_T(\widehat{\theta}_T) - \Psi_T(\theta_0)'\Omega\Psi_T(\theta_0)$. By using:

$$\begin{aligned} m_T(\widehat{\theta}_T^*)'\Omega m_T(\widehat{\theta}_T^*) &\geq \lambda\left\|m_T(\widehat{\theta}_T^*)\right\|^2, \\ m_T(\widehat{\theta}_T^*)'\Omega\Psi_T(\widehat{\theta}_T) &\geq -\left\|m_T(\widehat{\theta}_T^*)\right\|\left\|\Omega\Psi_T(\widehat{\theta}_T)\right\|, \end{aligned}$$

we deduce:

$$\left\|m_T(\widehat{\theta}_T^*)\right\|^2 - 2d_{2,T}\left\|m_T(\widehat{\theta}_T^*)\right\| + d_{3,T} \leq 0, \tag{B.29}$$

where:

$$d_{2,T} = \left\|\Omega\Psi_T(\widehat{\theta}_T)\right\|/\lambda \quad \text{and} \quad d_{3,T} = d_{1,T}/\lambda = \left[\Psi_T(\widehat{\theta}_T)'\Omega\Psi_T(\widehat{\theta}_T) - \Psi_T(\theta_0)'\Omega\Psi_T(\theta_0)\right]/\lambda.$$

Inequality (B.29) implies:

$$\left\|m_T(\widehat{\theta}_T^*)\right\| \leq d_{2,T} + \left(d_{2,T}^2 - d_{3,T}\right)^{1/2}.$$

Let us now derive the order of the RHS. From Lemma A.1 and CMT we have:

$$\begin{aligned} d_{2,T} &\leq \sup_{\theta\in\Theta}\left\|\Omega\Psi_T(\theta)\right\|/\lambda = O_p(1), \\ |d_{3,T}| &\leq 2\sup_{\theta\in\Theta}\left|\Psi_T(\theta)'\Omega\Psi_T(\theta)\right|/\lambda = O_p(1). \end{aligned}$$

We get $\left\|m_T(\widehat{\theta}_T^*)\right\| = O_p(1)$. Define:

$$G(\theta^*) = \left(E\left[g_1(X_t, Y_t;\theta)\right]', E\left[g_2(Y_t;\theta)|x_0\right]', E\left[a(Y_t;\theta) - \beta|x_0\right]'\right),$$

for $\theta^* \in \Theta\times B$. Since $\|m_T(\theta^*)\|^2 \geq Th_T^d\|G(\theta^*)\|^2$, $\theta^* \in \Theta\times B$, we deduce: $\left\|G(\widehat{\theta}_T^*)\right\| = O_p\left(1/\sqrt{Th_T^d}\right)$.

By the mean-value theorem we can write [1]:

$$\left\|\frac{\partial G}{\partial\theta^{*\prime}}(\widetilde{\theta}_T^*)\left(\widehat{\theta}_T^* - \theta_0^*\right)\right\| = O_p\left(1/\sqrt{Th_T^d}\right),$$

[1] More precisely, the mean-value theorem is applied separately for any component of function G, and the intermediary point $\tilde{\theta}_T^*$ can differ across components.

where $\widetilde{\theta}_T^*$ is between $\widehat{\theta}_T^*$ and θ_0^*. Since $\widehat{\theta}_T^*$ converges to θ_0^* by consistency (Section B.2), and $\partial G/\partial\theta^{*'}(\theta^*)$ is continuous by Assumption A.24, we have:

$$\frac{\partial G}{\partial\theta^{*'}}(\widetilde{\theta}_T^*) \xrightarrow{p} \frac{\partial G}{\partial\theta^{*'}}(\theta_0^*),$$

where $\partial G/\partial\theta^{*'}(\theta_0^*)$ has full rank, by the local identification condition in Assumption A.3. The conclusion follows. ■

Let us now prove Lemma A.2. From Lemma B.8, it is enough to show that $\text{plim}_{T\to\infty} \frac{\partial \widehat{g}_T}{\partial\theta^{*'}}\left(\bar{\theta}_T^*\right) R_T = J_0$, for any $\bar{\theta}_T^*$ such that $\left\|\bar{\theta}_T^* - \theta_0^*\right\| = O_p\left(1/\sqrt{Th_T^d}\right)$. We have:

$$\frac{\partial \widehat{g}_T}{\partial\theta^{*'}}\left(\bar{\theta}_T^*\right) R_T = \begin{pmatrix} \widehat{E}\left[\frac{\partial g_1}{\partial\theta'}\left(\bar{\theta}_T\right)\right] R_{1,Z} & h_T^{-d/2}\widehat{E}\left[\frac{\partial g_1}{\partial\theta'}\left(\bar{\theta}_T\right)\right] R_{2,Z} & 0 \\ h_T^{d/2}\widehat{E}\left[\frac{\partial g_2}{\partial\theta'}\left(\bar{\theta}_T\right)|x_0\right] R_{1,Z} & \widehat{E}\left[\frac{\partial g_2}{\partial\theta'}\left(\bar{\theta}_T\right)|x_0\right] R_{2,Z} & 0 \\ h_T^{d/2}\widehat{E}\left[\frac{\partial a}{\partial\theta'}\left(\bar{\theta}_T\right)|x_0\right] R_{1,Z} & \widehat{E}\left[\frac{\partial a}{\partial\theta'}\left(\bar{\theta}_T\right)|x_0\right] R_{2,Z} & -Id_L \end{pmatrix}.$$

Thus, we have to show:

$$\begin{aligned} &\text{i)} \quad \widehat{E}\left[\frac{\partial g_1}{\partial\theta'}\left(\bar{\theta}_T\right)\right] \xrightarrow{p} E\left[\frac{\partial g_1}{\partial\theta'}\left(\theta_0\right)\right], \\ &\text{ii)} \quad \widehat{E}\left[\frac{\partial g_2^*}{\partial\theta'}\left(\bar{\theta}_T\right)|x_0\right] \xrightarrow{p} E\left[\frac{\partial g_2^*}{\partial\theta'}\left(\theta_0\right)|x_0\right], \\ &\text{iii)} \quad h_T^{-d/2}\widehat{E}\left[\frac{\partial g_1}{\partial\theta'}\left(\bar{\theta}_T\right)\right] R_{2,Z} \xrightarrow{p} 0 . \end{aligned}$$

Let us now prove these results.

i) From Assumptions A.4, A.5, A.21 and A.22, the ULLN [see Potscher, Prucha (1989), Corollary 1] implies that $\widehat{E}\left[\frac{\partial g_1}{\partial\theta'}(\theta)\right] \xrightarrow{p} E\left[\frac{\partial g_1}{\partial\theta'}(\theta)\right]$ uniformly in $\theta\in\Theta$. Moreover, $E\left[\frac{\partial g_1}{\partial\theta'}(\theta)\right]$ is continuous w.r.t. θ by Assumption A.24. Then i) follows.

ii) Let $g_{2,i}^*$ denote the i-th component of function g_2^*, for $i = 1,\cdots,K_2+L$. We have:

$$\widehat{E}\left[\frac{\partial g_{2,i}^*}{\partial\theta}(\bar{\theta}_T)|x_0\right] = \widehat{E}\left[\frac{\partial g_{2,i}^*}{\partial\theta}(\theta_0)|x_0\right] + \widehat{E}\left[\frac{\partial^2 g_{2,i}^*}{\partial\theta\partial\theta'}(\ddot{\theta}_T)|x_0\right](\bar{\theta}_T - \theta_0),$$

where $\ddot{\theta}_T$ is between $\bar{\theta}_T$ and θ_0. Under Assumptions A.5-A.9 and A.23 one can show that $\widehat{E}\left[\frac{\partial g_{2,i}^*}{\partial\theta}(\theta_0)|x_0\right] \xrightarrow{p} E\left[\frac{\partial g_{2,i}^*}{\partial\theta}(\theta_0)|x_0\right]$ and $\widehat{E}\left[\frac{\partial^2 g_{2,i}^*}{\partial\theta\partial\theta'}(\ddot{\theta}_T)|x_0\right] = O_p(1)$. Then ii) follows.

iii) Let $g_{1,i}$ denote the i-th component of function g_1, $i = 1,...,K_1$. We have:

$$h_T^{-d/2}\widehat{E}\left[\frac{\partial g_{1,i}}{\partial\theta'}\left(\bar{\theta}_T\right)\right] R_{2,Z} = \frac{1}{\sqrt{Th_T^d}}\sqrt{T}\widehat{E}\left[\frac{\partial g_{1,i}}{\partial\theta'}\left(\theta_0\right)\right] R_{2,Z} + \frac{1}{\sqrt{T}h_T^d}\sqrt{Th_T^d}\left(\bar{\theta}_T - \theta_0\right)'\widehat{E}\left[\frac{\partial^2 g_{1,i}}{\partial\theta\partial\theta'}\left(\ddot{\theta}_T\right)\right] R_{2,Z}, \tag{B.30}$$

where $\ddot{\theta}_T$ is between $\bar{\theta}_T$ and θ_0. Let us derive the orders of the two terms in the RHS of (B.30). For the first one:

$$\sqrt{T}\widehat{E}\left[\frac{\partial g_{1,i}}{\partial\theta'}\left(\theta_0\right)\right] R_{2,Z} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial g_{1,i}}{\partial\theta'}\left(Y_t, X_t;\theta_0\right) R_{2,Z},$$

where $E\left[\partial g_{1,i}/\partial\theta^{'}\left(Y_t,X_t;\theta_0\right)R_{2,Z}\right]=0$. From Assumptions A.5 and A.22, the CLT for mixing processes [e.g. Herrndorf (1984), Corollary 1] implies:

$$\sqrt{T}\widehat{E}\left[\frac{\partial g_{1,i}}{\partial\theta^{'}}\left(\theta_0\right)\right]R_{2,Z}=O_p(1).$$

Let us now consider the second term in (B.30). From Assumptions A.4, A.5, A.22 and A.24, the ULLN implies:

$$\widehat{E}\left[\frac{\partial^2 g_{1,i}}{\partial\theta\partial\theta^{'}}\left(\ddot{\theta}_T\right)\right]=O_p\left(1\right).$$

Thus, from (B.30) we get:

$$h_T^{-d/2}\widehat{E}\left[\frac{\partial g_{1,i}}{\partial\theta}\left(\bar{\theta}_T\right)\right]R_{2,Z}=O_p\left(\frac{1}{\sqrt{Th_T^d}}+\frac{1}{\sqrt{T}h_T^d}\right)=o_p\left(1\right),$$

from bandwidth condition in Assumption A.9. The proof is concluded.

B.4 Proof of Corollary 6

If the bandwidth is such that $\bar{c}=\lim Th_T^{2m+d}=0$, from (A.6) the optimal weighting matrix for given instrument is $\Omega=V_0^{-1}$. The proof that $Z^*=E\left(\frac{\partial g^{'}}{\partial\theta}\left(Y;\theta_0\right)|X\right)W\left(X\right)$ is still an optimal instrument is similar to the proof of Proposition 3, replacing $M\left(Z,c,a\right)$ with $V\left(Z,a\right)=\frac{w^2}{f_X(x_0)}e^{'}\left(J_{0,Z}^{\prime}\Sigma_0^{-1}J_{0,Z}\right)^{-1}e$, which is the asymptotic variance of $\hat{\beta}_T$. Thus, the bias-free kernel nonparametric efficiency bound is $\mathcal{B}\left(a,x_0\right)=\frac{w^2}{f_X(x_0)}e^{'}\left(J_0^{*'}\Sigma_0^{-1}J_0^{*}\right)^{-1}e$. Corollary 6 follows from the block inversion formula.

B.5 Proof of Lemma A.3

B.5.1 Asymptotic expansion of the concentrated objective function

Since the conditional moment restrictions are satisfied asymptotically, we have $\widehat{\lambda}_T\overset{p}{\to}0$, when $T\to\infty$. Therefore, we can consider the second-order asymptotic expansion of function $\mathcal{L}_T^c(\theta,\lambda)$ in a neighbourhood of $\theta=\theta_0$, $\lambda=0$. Let us first derive the expansion w.r.t. λ. We have:

$$\begin{aligned}\log\widehat{E}\left(\exp\lambda^{'}g_2(\theta)|x_0\right)&\simeq\log\left[1+\lambda^{'}\widehat{E}\left(g_2(\theta)|x_0\right)+\frac{1}{2}\lambda^{'}\widehat{E}\left(g_2(\theta)g_2(\theta)^{'}|x_0\right)\lambda\right]\\&\simeq\lambda^{'}\widehat{E}\left(g_2(\theta)|x_0\right)+\frac{1}{2}\lambda^{'}\widehat{V}\left(g_2(\theta)|x_0\right)\lambda.\end{aligned}$$

Therefore, we can asymptotically concentrate w.r.t. λ:

$$\lambda\simeq-\widehat{V}\left(g_2(\theta)|x_0\right)^{-1}\widehat{E}\left(g_2(\theta)|x_0\right),\tag{B.31}$$

and the asymptotic expansion of the concentrated objective function becomes:

$$\mathcal{L}_T^c(\theta)\simeq\frac{1}{T}\sum_{t=1}^{T}\widehat{E}\left(g(\theta)|x_t\right)^{'}\widehat{V}\left(g(\theta)|x_t\right)^{-1}\widehat{E}\left(g(\theta)|x_t\right)+\frac{1}{2}h_T^d\widehat{E}\left(g_2(\theta)|x_0\right)^{'}\widehat{V}\left(g_2(\theta)|x_0\right)^{-1}\widehat{E}\left(g_2(\theta)|x_0\right).$$

Criterion $\mathcal{L}_T^c(\theta)$ multiplied by T is asymptotically equivalent to the criterion of the kernel moment estimator (see Definition 4) with optimal instrument and weighting matrix.

Let us now consider the expansion around $\theta = \theta_0$. We have:

$$\widehat{E}\left(g(\theta)|x_t\right) \simeq \widehat{E}\left(g(\theta_0)|x_t\right) + E\left(\frac{\partial g}{\partial \theta'}(\theta_0) \mid x_t\right)(\theta - \theta_0), \qquad \widehat{V}\left(g(\theta)|x_t\right) \simeq V\left(g(\theta_0) \mid x_t\right),$$

and similarly for the expectations of function g_2. Thus, we get:

$$\begin{aligned}\mathcal{L}_T^c(\theta) \;\simeq\; & \frac{1}{T}\sum_{t=1}^{T}\left\{\widehat{E}\left(g|x_t\right) + E\left(\frac{\partial g}{\partial \theta'} \mid x_t\right)(\theta - \theta_0)\right\}' V\left(g \mid x_t\right)^{-1}\left\{\widehat{E}\left(g|x_t\right) + E\left(\frac{\partial g}{\partial \theta'} \mid x_t\right)(\theta - \theta_0)\right\} \\ & + \frac{1}{2}h_T^d\left\{\widehat{E}\left(g_2|x_0\right) + E\left(\frac{\partial g_2}{\partial \theta'} \mid x_0\right)(\theta - \theta_0)\right\}' V\left(g_2 \mid x_0\right)^{-1}\left\{\widehat{E}\left(g_2|x_0\right) + E\left(\frac{\partial g_2}{\partial \theta'} \mid x_0\right)(\theta - \theta_0)\right\},\end{aligned}$$

where functions g, g_2 are evaluated at θ_0.

B.5.2 Asymptotic expansion of $\widehat{\theta}_T$

We have:

$$E\left(\frac{\partial g}{\partial \theta'} \mid x_t\right)(\theta - \theta_0) = E\left(\frac{\partial g}{\partial \theta'} \mid x_t\right) R_1\left(\eta_1^* - \eta_{1,0}^*\right).$$

We get:

$$\begin{aligned}\mathcal{L}_T^c(\eta^*) \;\simeq\; & \frac{1}{T}\sum_{t=1}^{T}\left\{\widehat{E}\left(g|x_t\right) + E\left(\frac{\partial g}{\partial \theta'}|x_t\right) R_1\left(\eta_1^* - \eta_{1,0}^*\right)\right\}' V\left(g|x_t\right)^{-1}\left\{\widehat{E}\left(g|x_t\right) + E\left(\frac{\partial g}{\partial \theta'}|x_t\right) R_1\left(\eta_1^* - \eta_{1,0}^*\right)\right\} \\ & + \frac{1}{2}h_T^d\left\{\widehat{E}\left(g_2|x_0\right) + E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_1\left(\eta_1^* - \eta_{1,0}^*\right) + E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2\left(\eta_2^* - \eta_{2,0}^*\right)\right\}' \\ & \cdot V\left(g_2|x_0\right)^{-1}\left\{\widehat{E}\left(g_2|x_0\right) + E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_1\left(\eta_1^* - \eta_{1,0}^*\right) + E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2\left(\eta_2^* - \eta_{2,0}^*\right)\right\}.\end{aligned}$$

The asymptotic expansion of $\widehat{\eta}_{1,T}^*$ is obtained from the maximization of the first term in $\mathcal{L}_T^c(\eta^*)$, since the contribution of the second term is asymptotically negligible. We get:

$$\begin{aligned}\sqrt{T}\left(\widehat{\eta}_{1,T}^* - \eta_{1,0}^*\right) \;\simeq\; & -\left[\frac{1}{T}\sum_{t=1}^{T} R_1' E\left(\frac{\partial g'}{\partial \theta}|x_t\right) V\left(g|x_t\right)^{-1} E\left(\frac{\partial g}{\partial \theta'}|x_t\right) R_1\right]^{-1} \\ & \cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T} R_1' E\left(\frac{\partial g'}{\partial \theta}|x_t\right) V\left(g|x_t\right)^{-1}\int g(y;\theta_0)\widehat{f}(y|x_t)dy.\end{aligned}$$

Thus:

$$\begin{aligned}\sqrt{T}\left(\widehat{\eta}_{1,T}^* - \eta_{1,0}^*\right) \;\simeq\; & -\left(R_1' E\left[E\left(\frac{\partial g'}{\partial \theta}|x_t\right) V\left(g|x_t\right)^{-1} E\left(\frac{\partial g}{\partial \theta'}|x_t\right)\right] R_1\right)^{-1} \\ & \cdot\sqrt{T}\int\int R_1' E\left(\frac{\partial g'}{\partial \theta}|x\right) V\left(g|x\right)^{-1} g(y;\theta_0)\widehat{f}(y,x)dxdy \\ \simeq\; & -\left(R_1' E\left[E\left(\frac{\partial g'}{\partial \theta}|x_t\right) V\left(g|x_t\right)^{-1} E\left(\frac{\partial g}{\partial \theta'}|x_t\right)\right] R_1\right)^{-1} \\ & \cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T} R_1' E\left(\frac{\partial g'}{\partial \theta}|x_t\right) V\left(g|x_t\right)^{-1} g(y_t;\theta_0).\end{aligned}$$

The bias term induced by the kernel estimator is asymptotically negligible since $Th_T^{d+2m} = o(1)$. The asymptotic expansion of $\widehat{\eta}^*_{2,T}$ can be deduced from the maximization of the second component of $\mathcal{L}^c_T(\eta^*)$. Estimator $\widehat{\eta}^*_{2,T}$ converges at a nonparametric rate, and terms involving $\left(\widehat{\eta}^*_{1,T} - \eta^*_{1,0}\right)$ can be neglected. We get:

$$\begin{aligned}\sqrt{Th_T^d}\left(\widehat{\eta}^*_{2,T} - \eta^*_{2,0}\right) &\simeq -\left[R_2^{'} E\left(\frac{\partial g_2^{'}}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1} E\left(\frac{\partial g_2}{\partial \theta^{'}}|x_0\right) R_2\right]^{-1} \\ &\quad \cdot R_2^{'} E\left(\frac{\partial g_2^{'}}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1} \sqrt{Th_T^d}\int g_2(y;\theta_0)\widehat{f}(y|x_0)dy.\end{aligned}$$

Then, point i) of Lemma A.3 is proved.

B.5.3 Asymptotic expansion of $\widehat{\lambda}_T$

We have from (B.31):

$$\widehat{\lambda}_T \simeq -\widehat{V}\left(g_2(\widehat{\theta}_T)|x_0\right)^{-1}\widehat{E}\left(g_2(\widehat{\theta}_T)|x_0\right) \simeq -V\left(g_2(\theta_0)|x_0\right)^{-1}\widehat{E}\left(g_2(\widehat{\theta}_T)|x_0\right).$$

Moreover,

$$\begin{aligned}\widehat{E}\left(g_2(\widehat{\theta}_T)|x_0\right) &\simeq \int g_2(y;\theta_0)\widehat{f}(y|x_0)dy + E\left(\frac{\partial g_2}{\partial \theta^{'}}|x_0\right)\left(\widehat{\theta}_T - \theta_0\right) \\ &\simeq \int g_2(y;\theta_0)\widehat{f}(y|x_0)dy + E\left(\frac{\partial g_2}{\partial \theta^{'}}|x_0\right) R_2\left(\widehat{\eta}^*_{2,T} - \eta^*_{2,0}\right) \\ &\quad \text{(since the contribution of } \widehat{\eta}^*_{1,T} - \eta^*_{1,0} \text{ is asymptotically negligible)} \\ &\simeq (Id - M)\int g_2(y;\theta_0)\widehat{f}(y|x_0)dy,\end{aligned}$$

where M is the matrix in (A.19). Then:

$$\widehat{\lambda}_T \simeq -V\left(g_2(\theta_0)|x_0\right)^{-1}(Id - M)\int g_2(y;\theta_0)\widehat{f}(y|x_0)dy,$$

and point ii) of Lemma A.3 is proved.

B.6 Proof of Corollary 8

From Appendix A.1.4, equation (A.5), the asymptotic distribution of the optimal kernel moment estimator of θ_0^* is such that:

$$\left(\sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0})', \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0})', \sqrt{Th_T^d}\left(\hat{\beta}_T - \beta_0\right)\right)' = -\left(J_0'\Omega J_0\right)^{-1} J_0'\Omega\hat{g}_T(\theta_0^*) + o_p(1), \tag{B.32}$$

where $\Omega = V_0^{-1}$, matrix V_0 is given in (A.2), and:

$$J_0 = \begin{pmatrix} E\left(\frac{\partial g_1}{\partial \theta'}\right) R_1 & 0 & 0 \\ 0 & E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2 & 0 \\ 0 & E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_2 & -Id_L \end{pmatrix} =: \begin{pmatrix} E\left(\frac{\partial g_1}{\partial \theta'}\right) R_1 & 0 \\ 0 & J_0^* \end{pmatrix}.$$

For $Z = E\left(\frac{\partial g(Y;\theta_0)'}{\partial \theta}|X\right) V\left(g(Y;\theta_0)|X\right)^{-1}$, we have:

$$E\left(\frac{\partial g_1}{\partial \theta'}\right) = E\left[E\left(\frac{\partial g'}{\partial \theta}|X\right) V\left(g|X\right)^{-1} E\left(\frac{\partial g}{\partial \theta'}|X\right)\right] = V\left(g_1\right).$$

Thus:

$$\left(J_0'\Omega J_0\right)^{-1} J_0'\Omega = \left[\begin{array}{cc} \left(R_1'E\left[E\left(\frac{\partial g'}{\partial \theta}|X\right) V\left(g|X\right)^{-1} E\left(\frac{\partial g}{\partial \theta'}|X\right)\right] R_1\right)^{-1} R_1' & 0 \\ 0 & \left(J_0^{*\prime}\Sigma_0^{-1}J_0^*\right)^{-1} J_0^{*\prime}\Sigma_0^{-1} \end{array}\right].$$

We get:

$$\left(J_0'\Omega J_0\right)^{-1} J_0'\Omega \hat{g}_T(\theta_0^*) = \left[\begin{array}{c} \left(R_1'E\left[E\left(\frac{\partial g'}{\partial \theta}|X\right) V\left(g|X\right)^{-1} E\left(\frac{\partial g}{\partial \theta'}|X\right)\right] R_1\right)^{-1} R_1'\sqrt{T}\hat{E}\left[g_1\right] \\ \left(J_0^{*\prime}\Sigma_0^{-1}J_0^*\right)^{-1} J_0^{*\prime}\Sigma_0^{-1}\sqrt{Th_T^d}\hat{E}\left[g_2^*|x_0\right] \end{array}\right]. \tag{B.33}$$

Let us now compute $\xi := \left(J_0^{*\prime}\Sigma_0^{-1}J_0^*\right)^{-1} J_0^{*\prime}\Sigma_0^{-1}\hat{E}\left[g_2^*|x_0\right]$. Let us denote $G := E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2$, $A := E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_2$. Then $\xi = (\xi_1', \xi_2')' \in \mathbb{R}^{s^*} \times \mathbb{R}^L$ solves $J_0^{*\prime}\Sigma_0^{-1}\left(\hat{E}\left[g_2^*|x_0\right] - J_0^*\xi\right) = 0$, that is

$$\left(\begin{array}{cc} G'\Sigma_0^{11} + A'\Sigma_0^{21} & G'\Sigma_0^{12} + A'\Sigma_0^{22} \\ -\Sigma_0^{21} & -\Sigma_0^{22} \end{array}\right)\left(\begin{array}{c} \hat{E}\left[g_2|x_0\right] - G\xi_1 \\ \hat{E}\left[a-\beta_0|x_0\right] - A\xi_1 + \xi_2 \end{array}\right) = 0,$$

where Σ_0^{ij}, $i,j = 1,2$, denote the blocks of Σ_0^{-1}. Solving for ξ_2 in the second block equation, we get:

$$\xi_2 = -\hat{E}\left[a-\beta_0|x_0\right] + A\xi_1 - \left(\Sigma_0^{22}\right)^{-1}\Sigma_0^{21}\left(\hat{E}\left[g_2|x_0\right] - G\xi_1\right).$$

By replacing in the first block equation, and using $\Sigma_{0,11}^{-1} = \Sigma_0^{11} - \Sigma_0^{12}\left(\Sigma_0^{22}\right)^{-1}\Sigma_0^{21}$ and $\left(\Sigma_0^{22}\right)^{-1}\Sigma_0^{21} = -\Sigma_{0,21}\Sigma_{0,11}^{-1}$ from the formulas of the inverse of a block matrix, we get:

$$\xi_1 = \left(G'\Sigma_{0,11}^{-1}G\right)^{-1} G'\Sigma_{0,11}^{-1}\hat{E}\left[g_2|x_0\right],$$

and:

$$\begin{aligned} \xi_2 &= -\hat{E}\left[a-\beta_0|x_0\right] + \Sigma_{0,21}\Sigma_{0,11}^{-1}\hat{E}\left[g_2|x_0\right] \\ &\quad + \left[A - \Sigma_{0,21}\Sigma_{0,11}^{-1}G\right]\left(G'\Sigma_{0,11}^{-1}G\right)^{-1} G'\Sigma_{0,11}^{-1}\hat{E}\left[g_2|x_0\right]. \end{aligned}$$

Thus, using (B.32), (B.33) and $\Sigma_{0,11} = V\left(g_2|x_0\right)$, $\Sigma_{0,21} = Cov\left(a, g_2|x_0\right)$, we get:

$$\begin{aligned} \sqrt{T}(\hat{\eta}_{1,T} - \eta_{1,0}) &= -\left(R_1'E\left[E\left(\frac{\partial g'}{\partial \theta}|X\right) V\left(g|X\right)^{-1} E\left(\frac{\partial g}{\partial \theta'}|X\right)\right] R_1\right)^{-1} R_1'\sqrt{T}\hat{E}\left[g_1\right] + o_p(1), \\ \sqrt{Th_T^d}(\hat{\eta}_{2,T} - \eta_{2,0}) &= -\left(R_2'E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2\right)^{-1} \\ &\quad R_2'E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1}\sqrt{Th_T^d}\hat{E}\left[g_2|x_0\right] + o_p(1), \end{aligned}$$

and:

$$
\begin{aligned}
& \sqrt{Th_T^d}\left(\hat{\beta}_T-\beta_0\right) \\
= \; & \sqrt{Th_T^d}\hat{E}\left[a-\beta_0|x_0\right]-Cov\left(a,g_2|x_0\right)V\left(g_2|x_0\right)^{-1}\sqrt{Th_T^d}\hat{E}\left[g_2|x_0\right] \\
& -\left[E\left(\frac{\partial a}{\partial\theta'}|x_0\right)R_2-Cov\left(a,g_2|x_0\right)V\left(g_2|x_0\right)^{-1}E\left(\frac{\partial g_2}{\partial\theta'}|x_0\right)R_2\right] \\
& \left(R_2'E\left(\frac{\partial g_2'}{\partial\theta}|x_0\right)V\left(g_2|x_0\right)^{-1}E\left(\frac{\partial g_2}{\partial\theta'}|x_0\right)R_2\right)^{-1}R_2'E\left(\frac{\partial g_2'}{\partial\theta}|x_0\right)V\left(g_2|x_0\right)^{-1}\sqrt{Th_T^d}\hat{E}\left[g_2|x_0\right]+o_p(1).
\end{aligned}
$$

The asymptotic expansions for $\hat{\eta}_{1,T}$, $\hat{\eta}_{2,T}$ correspond to the asymptotic expansions of the XMM estimators $\hat{\eta}_{1,T}^*$, $\hat{\eta}_{2,T}^*$ in Lemma A.3 (i). The conclusion follows.

B.7 Regularity conditions in the stochastic volatility model

In this Section, we discuss the technical regularity assumptions for the XMM estimator (see Appendix A.1.1 in the paper) when the DGP P_0 is compatible with the stochastic volatility model (3.6)-(3.8). They concern the stationary distribution (Section B.7.1) and the existence of moments (Section B.7.2).

B.7.1 Stationary distribution

Let us consider process $\left\{X_t=\left(\tilde{r}_t,\sigma_t^2\right):t\in\mathbb{Z}\right\}$, where the dynamics of $\tilde{r}_t=r_t-r_{f,t}$ and σ_t^2 under the DGP P_0 are defined in Section 3.2. Markov process X_t is exponential affine:

$$
\begin{aligned}
E_0\left[e^{-z'X_{t+1}}\mid \underline{X_t}\right] &= E_0\left[e^{-u\tilde{r}_{t+1}-v\sigma_{t+1}^2}\mid \underline{X_t}\right]=E_0\left[e^{-(\gamma_0u+v)\sigma_{t+1}^2}E_0\left[e^{-u\sigma_{t+1}\varepsilon_{t+1}}\mid\left(\sigma_t^2\right),\underline{X_t}\right]\mid \underline{X_t}\right] \\
&= E_0\left[e^{-\left(\gamma_0u+v-\frac{1}{2}u^2\right)\sigma_{t+1}^2}\mid\sigma_t^2\right]=\exp\left[-a_0\left(\gamma_0u+v-\frac{1}{2}u^2\right)\sigma_t^2-b_0\left(\gamma_0u+v-\frac{1}{2}u^2\right)\right] \\
&= \exp\left[-A(z)'X_t-B(z)\right],
\end{aligned}
$$

where $A(z)=\left(0,a_0\left(\gamma_0u+v-\frac{1}{2}u^2\right)\right)'$, $B(z)=b_0\left(\gamma_0u+v-\frac{1}{2}u^2\right)$, for $z=(u,v)'\in\mathbb{C}^2$ such that $\mathrm{Re}\left(\gamma_0u+v-\frac{1}{2}u^2\right)>-1/c_0$, and functions a_0 and b_0 are defined in Section 3.2.

i) Strict stationarity and geometric strong mixing

From Proposition 2 in Gouriéroux, Jasiak (2006), the ARG process $\left(\sigma_t^2\right)$ is stationary if $0\leq\rho_0<1$, with marginal invariant distribution such that $\left[\left(1-\rho_0\right)/c_0\right]\sigma_t^2\sim\gamma\left(\delta_0\right)$, where $\gamma\left(\delta_0\right)$ denotes the gamma distribution with parameter δ_0. Thus, when $\rho_0<1$, process (X_t) admits the marginal invariant distribution:

$$
f(x) = \frac{1}{\sigma}\phi\left(\frac{\tilde{r}-\gamma_0\sigma^2}{\sigma}\right)\frac{\left[\left(1-\rho_0\right)/c_0\right]^{\delta_0}}{\Gamma\left(\delta_0\right)}e^{-\frac{1-\rho_0}{c_0}\sigma^2}\left(\sigma^2\right)^{\delta_0-1}\quad,\; x=\left(\tilde{r},\sigma^2\right)\in\mathbb{R}\times\mathbb{R}^+=\mathcal{X}. \tag{B.34}
$$

To prove that (X_t) is geometrically strongly mixing, we use Proposition 4.2 in Darolles, Gouriéroux, Jasiak (2006), and verify the condition:

$$
\lim_{h\to\infty}\frac{\partial A}{\partial z'}(0)^h=0. \tag{B.35}
$$

We have:

$$\frac{\partial A}{\partial z'}(0) = \left(\begin{array}{cc} 0 & 0 \\ \gamma_0 \rho_0 & \rho_0 \end{array} \right).$$

Condition (B.35) is satisfied if $\rho_0 < 1$. Thus, with $Y_t = \left(X_{t+1}, ..., X_{t+\bar{h}}\right)'$ for given $\bar{h} \in \mathbb{N}$, we conclude that Assumption A.5 is satisfied if $0 \leq \rho_0 < 1$.

ii) Smoothness of the marginal distribution

The stationary distribution f in (B.34) is in $C^{\infty}(\mathcal{X})$. Moreover, we have:

$$f(x) \leq C_1 e^{-\frac{1-\rho_0}{c_0}\sigma^2} \left(\sigma^2\right)^{\delta_0 - 3/2}, \qquad x \in \mathcal{X},$$

for a constant $C_1 > 0$. Thus, $\|f\|_\infty < \infty$ if, and only if, $\delta_0 \geq 3/2$. Moreover, we have the following Lemma B.9.

Lemma B.9: *$\|D^m f\|_\infty < \infty$ if, and only if, $\delta_0 \geq 3/2 + m$.*
Proof: Let $m \in \mathbb{N}$. Since $f \in C^m(\mathcal{X})$, to prove $\|D^m f\|_\infty < \infty$, it is sufficient to show that any partial derivative of order m of function f is bounded at the boundary of $\mathcal{X}$, that is, for $\tilde{r} \to \pm\infty$, $\sigma^2 \to \infty$, $\sigma^2 \to 0$. From (B.34), let us write:

$$f(\tilde{r}, \sigma^2) = C\phi\left[h(\tilde{r}, \sigma^2)\right] e^{-\lambda\sigma^2} \left(\sigma^2\right)^{\delta_0 - 3/2}, \qquad \left(\tilde{r}, \sigma^2\right) \in \mathcal{X},$$

where $\lambda = (1 - \rho_0)/c_0$, $C = \lambda^{\delta_0} / \left[\Gamma(\delta_0)\right]$, and:

$$h(\tilde{r}, \sigma^2) := \frac{\tilde{r} - \gamma_0 \sigma^2}{\sigma}.$$

The function h is such that:

$$\begin{aligned} \frac{\partial h}{\partial \tilde{r}}(\tilde{r}, \sigma^2) &= \frac{1}{\sigma}, \\ \frac{\partial h}{\partial \sigma^2}(\tilde{r}, \sigma^2) &= \frac{-\gamma_0 \sigma - \left(\tilde{r} - \gamma_0 \sigma^2\right)/2\sqrt{\sigma^2}}{\sigma^2} = -\gamma_0 \frac{1}{\sigma} - \frac{1}{2\sigma^2} h(\tilde{r}, \sigma^2). \end{aligned}$$

We deduce that:
i) Any partial derivative of f is a linear combination of functions of the type:

$$\phi^{(n)}\left[h(\tilde{r}, \sigma^2)\right] h(\tilde{r}, \sigma^2)^k e^{-\lambda\sigma^2} \left(\sigma^2\right)^l, \quad n, k \in \mathbb{N}, \; l \in \mathbb{R}.$$

ii) Since function $h \longmapsto \phi^{(n)}(h) h^k$ is bounded, for any $n, k \in \mathbb{N}$, it is sufficient to prove that partial derivatives of order m are bounded for $\sigma^2 \to 0$. This is the case if, and only if, the smallest power l of σ^2, which occurs in partial derivatives of order m, is non-negative.
iii) The smallest power of σ^2 is featured by $\partial^m f / \partial \left(\sigma^2\right)^m$ and is $l = \delta_0 - 3/2 - m$. We conclude that $\|D^m f\|_\infty < \infty$ if, and only if, $\delta_0 \geq 3/2 + m$. ■

Thus, Assumption A.6 is satisfied if $\delta_0 \geq 3/2 + m$. For instance, for $m = 2$, we get $\delta_0 \geq 7/2$.

B.7.2 Existence of moments

For expository purpose, let us assume that the actively traded derivatives at date t_0 have times-to-maturity $h_j = 1$ (and moneyness strikes k_j), for $j = 1, ..., n$, and that we are interested in estimating the price of the derivative with time-to-maturity $h = 1$ and moneyness strike k. The moment function $g_2^*(y;\theta)$, is given by:

$$g_2^*(y_t;\theta) = e^{-r_{f,t+1}} e^{-\theta_1-\theta_2\sigma_{t+1}^2-\theta_3\sigma_t^2-\theta_4\tilde{r}_{t+1}} \begin{pmatrix} e^{r_{f,t+1}} \\ e^{r_{t+1}} \\ (e^{r_{t+1}}-k_1)^+ \\ \vdots \\ (e^{r_{t+1}}-k_n)^+ \\ (e^{r_{t+1}}-k)^+ \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ c_{t_0}(k_1,1) \\ \vdots \\ c_{t_0}(k_n,1) \\ 0 \end{pmatrix},$$

where $y_t = \left(\tilde{r}_{t+1}, \sigma_{t+1}^2, \sigma_t^2\right)'$ and $r_{t+1} = \tilde{r}_{t+1} + r_{f,t+1}$. The relevant variables are $Y_t = \left(\tilde{r}_{t+1}, \sigma_{t+1}^2, \sigma_t^2\right)$, and $X_t = \left(\tilde{r}_t, \sigma_t^2\right)$, respectively. Note that function g_2^* does not depend on $\tilde{r}_t$ and thus we have dropped this variable from Y_t. The following Lemma B.10 provides a condition for $E_0\left[\|g_2^*(Y_t;\theta_0)\|^4\right] < \infty$ (see Assumption A.11).

Lemma B.10: *The function $g_2^*(.;\theta_0)$ is such that $E_0\left[\|g_2^*(Y_t;\theta_0)\|^4\right] < \infty$ if, and only if:*

$$\theta_0 \in \Gamma = \left\{(\theta_1,\theta_2,\theta_3,\theta_4)' \in \mathbb{R}^4 \mid \theta_3 > -1/4c_0,\ \theta_2 > -\frac{1}{4c_0}\frac{1-\rho_0+4c_0\theta_3}{1+4c_0\theta_3} - \gamma_0\theta_4 + 2\theta_4^2 + (2+\gamma_0-4\theta_4)^+\right\}.$$

Proof: Since $(e^r - s)^+ \le e^r$, for any $r, s \in \mathbb{R}$, condition $E_0\left[\|g_2^*(Y_t;\theta)\|^4\right] < \infty$ is satisfied if, and only if:

$$E_0\left[e^{-4\theta_1-4\theta_2\sigma_{t+1}^2-4\theta_3\sigma_t^2-4\theta_4\tilde{r}_{t+1}}\right] < \infty \quad , \quad E_0\left[e^{-4\theta_1-4\theta_2\sigma_{t+1}^2-4\theta_3\sigma_t^2-4(\theta_4-1)\tilde{r}_{t+1}}\right] < \infty. \qquad \text{(B.36)}$$

We have:

$$\begin{aligned} E_0\left[e^{-4\theta_1-4\theta_2\sigma_{t+1}^2-4\theta_3\sigma_t^2-4\theta_4\tilde{r}_{t+1}}\right] &= E_0\left[e^{-4\theta_1-4(\theta_2+\gamma_0\theta_4)\sigma_{t+1}^2-4\theta_3\sigma_t^2} E_0\left(e^{-4\theta_4\sigma_{t+1}\varepsilon_{t+1}} | \sigma_{t+1}^2, \sigma_t^2\right)\right] \\ &= E_0\left[e^{-4\theta_1-4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\sigma_{t+1}^2-4\theta_3\sigma_t^2}\right] \\ &= e^{-4\theta_1} E_0\left[e^{-4\theta_3\sigma_t^2} E_0\left(e^{-4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\sigma_{t+1}^2} | \sigma_t^2\right)\right] \\ &= e^{-4\theta_1-b_0\left(4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\right)} E_0\left[e^{-\left[4\theta_3+a_0\left(4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\right)\right]\sigma_t^2}\right], \end{aligned}$$

if:

$$1 + 4c_0\left(\theta_2 + \gamma_0\theta_4 - 2\theta_4^2\right) > 0.$$

Moreover, since $[(1-\rho_0)/c_0]\,\sigma_t^2 \sim \gamma(\delta_0)$, we have:

$$E_0\left[e^{-\left[4\theta_3+a_0\left(4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\right)\right]\sigma_t^2}\right] = \frac{1}{\left(1+\frac{c_0}{1-\rho_0}\left[4\theta_3+a_0\left(4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\right)\right]\right)^{\delta_0}},$$

if:

$$1 + \frac{c_0}{1-\rho_0}\left[4\theta_3 + a_0\left(4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\right)\right] > 0.$$

We deduce that conditions (B.36) are satisfied if, and only if:

$$\begin{aligned} 1+4c_0\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right) &> 0, \\ 1+\frac{c_0}{1-\rho_0}\left[4\theta_3+a_0\left(4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\right)\right] &> 0, \\ 1+4c_0\left(\theta_2+\gamma_0\left(\theta_4-1\right)-2\left(\theta_4-1\right)^2\right) &> 0, \\ 1+\frac{c_0}{1-\rho_0}\left[4\theta_3+a_0\left(4\left(\theta_2+\gamma_0\left(\theta_4-1\right)-2\left(\theta_4-1\right)^2\right)\right)\right] &> 0. \end{aligned} \tag{B.37}$$

Since $\theta_2+\gamma_0\left(\theta_4-1\right)-2\left(\theta_4-1\right)^2=\theta_2+\gamma_0\theta_4-2\theta_4^2+\left(4\theta_4-\gamma_0-2\right)$, and function a_0 is increasing, we can distinguish between two parameter regions to solve system (B.37). i) First case: $4\theta_4-\gamma_0-2\geq 0$, that is, $\theta_4\geq(2+\gamma_0)/4$. In this region system (B.37) is equivalent to:

$$\begin{aligned} 1+4c_0\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right) &> 0, \\ 1+\frac{c_0}{1-\rho_0}\left[4\theta_3+a_0\left(4\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)\right)\right] &> 0. \end{aligned}$$

Let us introduce the new variable $x=4c_0\left(\theta_2+\gamma_0\theta_4-2\theta_4^2\right)$. Then, this system becomes:

$$\begin{aligned} x &> -1, \\ 1+\frac{4c_0\theta_3}{1-\rho_0}+\frac{\rho_0}{1-\rho_0}\frac{x}{1+x} &> 0. \end{aligned}$$

By transforming the second equation, we get:

$$\begin{aligned} x &> -1, \\ \left(1+4c_0\theta_3\right)x &> -\left(1-\rho_0+4c_0\theta_3\right). \end{aligned}$$

There is no solution for which $1+4c_0\theta_3<0$. Indeed, the second equation becomes:

$$x<-\frac{1-\rho_0+4c_0\theta_3}{1+4c_0\theta_3}=-1+\frac{\rho_0}{1+4c_0\theta_3}<-1,$$

which is incompatible with the first equation. Instead, for $1+4c_0\theta_3>0$, the second equation becomes:

$$x>-\frac{1-\rho_0+4c_0\theta_3}{1+4c_0\theta_3}=-1+\frac{\rho_0}{1+4c_0\theta_3},$$

and implies the first equation. To summarize, a first region of solutions is:

$$\theta_4\geq(2+\gamma_0)/4 \quad , \quad 1+4c_0\theta_3>0 \quad , \quad x>-\frac{1-\rho_0+4c_0\theta_3}{1+4c_0\theta_3},$$

that is,

$$\theta_4\geq(2+\gamma_0)/4 \quad , \quad \theta_3>-1/4c_0 \quad , \quad \theta_2>-\frac{1}{4c_0}\frac{1-\rho_0+4c_0\theta_3}{1+4c_0\theta_3}-\gamma_0\theta_4+2\theta_4^2.$$

ii) Second case: $4\theta_4-\gamma_0-2<0$, that is, $\theta_4<(2+\gamma_0)/4$. In this region, system (B.37) is equivalent to:

$$\begin{aligned} 1+4c_0\left(\theta_2+\gamma_0\left(\theta_4-1\right)-2\left(\theta_4-1\right)^2\right) &> 0, \\ 1+\frac{c_0}{1-\rho_0}\left[4\theta_3+a_0\left(4\left(\theta_2+\gamma_0\left(\theta_4-1\right)-2\left(\theta_4-1\right)^2\right)\right)\right] &> 0. \end{aligned}$$

By introducing the new variable $y = 4c_0\left(\theta_2 + \gamma_0(\theta_4 - 1) - 2(\theta_4 - 1)^2\right)$, and repeating the same argument as above, we get the second region of solutions:

$$\theta_4 < (2+\gamma_0)/4 \qquad , \quad 1 + 4c_0\theta_3 > 0 \quad , \;\; y > -\frac{1-\rho_0+4c_0\theta_3}{1+4c_0\theta_3},$$

that is,

$$\theta_4 < (2+\gamma_0)/4 \quad , \quad \theta_3 > -1/4c_0 \qquad , \quad \theta_2 > -\frac{1}{4c_0}\frac{1-\rho_0+4c_0\theta_3}{1+4c_0\theta_3} - \gamma_0\theta_4 + 2\theta_4^2 + 2 + \gamma_0 - 4\theta_4.$$

■

From Lemma B.10, the condition $E_0\left[\|g_2^*(Y_t;\theta_0)\|^4\right] < \infty$ is satisfied, whenever the risk premia parameters θ_2^0 and θ_3^0 for stochastic volatility are above some thresholds. In particular, the lower bound for θ_2^0 depends on θ_3^0 and θ_4^0. Imposing the no-arbitrage restriction $\theta_4^0 = \gamma_0 + 1/2$, the inequality constraints become:

$$\theta_3^0 > -1/4c_0, \qquad \theta_2^0 > -\frac{1}{4c_0}\frac{1-\rho_0+4c_0\theta_3^0}{1+4c_0\theta_3^0} + \gamma_0/2 + 3(-\gamma_0)^+ + 1/4.$$

These constraints are satisfied for the parameter values used in Section 3.4 iii).

B.8 ARG risk-neutral dynamics

In this section, we derive the dynamics of the ARG stochastic volatility model under the risk-neutral distribution Q defined by the sdf $M_{t,t+1}(\theta_0) = e^{-r_{f,t+1}}\exp\left(-\theta_1^0 - \theta_2^0\sigma_{t+1}^2 - \theta_3^0\sigma_t^2 - \theta_4^0\tilde{r}_{t+1}\right)$. In Section B.7.1, we derived the historical conditional moment generating function of $X_{t+1} = \left(\tilde{r}_{t+1}, \sigma_{t+1}^2\right)$:

$$E_0\left[\exp\left(-u\tilde{r}_{t+1} - v\sigma_{t+1}^2\right)|x_t\right] = \exp\left[-a_0\left(\gamma_0 u + v - \frac{1}{2}u^2\right)\sigma_t^2 - b_0\left(\gamma_0 u + v - \frac{1}{2}u^2\right)\right]. \quad \text{(B.38)}$$

Let us compute the risk-neutral conditional moment generating function of $\left(\tilde{r}_{t+1}, \sigma_{t+1}^2\right)$. We have:

$$\begin{aligned} E_0^Q\left[\exp\left(-u\tilde{r}_{t+1} - v\sigma_{t+1}^2\right)|x_t\right] &= E_0\left[M_{t,t+1}(\theta_0)\exp\left(-u\tilde{r}_{t+1} - v\sigma_{t+1}^2\right)|x_t\right]/E_0\left[M_{t,t+1}(\theta_0)|x_t\right] \\ &= e^{-\theta_1^0 - \theta_3^0\sigma_t^2}E_0\left[\exp\left(-\left(u+\theta_4^0\right)\tilde{r}_{t+1} - \left(v+\theta_2^0\right)\sigma_{t+1}^2\right)|x_t\right] \\ &= \exp\left\{-\left[a_0\left(\gamma_0\left(u+\theta_4^0\right) + \left(v+\theta_2^0\right) - \frac{1}{2}\left(u+\theta_4^0\right)^2\right) + \theta_3^0\right]\sigma_t^2\right. \\ &\quad \left. -b_0\left(\gamma_0\left(u+\theta_4^0\right) + \left(v+\theta_2^0\right) - \frac{1}{2}\left(u+\theta_4^0\right)^2\right) - \theta_1^0\right\}, \end{aligned}$$

by using $E_0\left[M_{t,t+1}(\theta_0)|x_t\right] = e^{-r_{f,t+1}}$ and (B.38). From equations (3.9) we have:

$$\begin{aligned} \gamma_0\left(u+\theta_4^0\right) + \left(v+\theta_2^0\right) - \frac{1}{2}\left(u+\theta_4^0\right)^2 &= u\left(\gamma_0 - \theta_4^0\right) + v - \frac{1}{2}u^2 + \theta_4^0\gamma_0 + \theta_2^0 - \frac{(\theta_4^0)^2}{2} \\ &= -\frac{1}{2}u + v - \frac{1}{2}u^2 + \lambda_2^0, \end{aligned}$$

where $\lambda_2^0 = \theta_2^0 + \gamma_0^2/2 - 1/8$, and:

$$\theta_1^0 = -b_0(\lambda_2^0), \quad \theta_3^0 = -a_0(\lambda_2^0).$$

Thus, we get:

$$E_0^Q\left[\exp\left(-u\tilde{r}_{t+1}-v\sigma^2_{t+1}\right)|x_t\right] \quad = \quad \exp\left[-a_0^*\left(-\frac{1}{2}u+v-\frac{1}{2}u^2\right)\sigma_t^2-b_0^*\left(-\frac{1}{2}u+v-\frac{1}{2}u^2\right)\right], \tag{B.39}$$

where:

$$a_0^*(u)=a_0(u+\lambda_2^0)-a_0(\lambda_2)=\frac{\rho_0^* u}{1+c_0^* u},$$
$$b_0^*(u)=b_0(u+\lambda_2^0)-b_0(\lambda_2^0)=\delta_0^*\log(1+c_0^* u),$$

with:

$$\begin{aligned}\rho_0^* &= \frac{\rho_0}{\left(1+c_0\lambda_2^0\right)^2}=\frac{\rho_0}{\left[1+c_0\left(\theta_2^0+\gamma_0^2/2-1/8\right)\right]^2},\\ \delta_0^* &= \delta_0,\\ c_0^* &= \frac{c_0}{1+c_0\lambda_2^0}=\frac{c_0}{1+c_0\left(\theta_2^0+\gamma_0^2/2-1/8\right)}.\end{aligned}$$

By comparing (B.38) and (B.39), we deduce that, under the risk neutral distribution, the returns follow a stochastic volatility model with risk premium parameter $\gamma_0^*=-\frac{1}{2}$ and ARG stochastic volatility with parameters $\rho_0^*,\delta_0^*,c_0^*$.

B.9 Proof of Lemma A.4

We have to show that:

$$P_0^Q\left[\sigma^2_{t+1}+\cdots+\sigma^2_{t+h}\geq z|\sigma^2_{t+h}=s,\sigma_t^2=\sigma_0^2\right] \text{ is increasing w.r.t. } s, \text{for any } z.$$

This condition is implied by:

$$P_0^Q\left[\sigma^2_{t+1}+\cdots+\sigma^2_{t+h-1}\geq z|\sigma^2_{t+h}=s,\sigma_t^2=\sigma_0^2\right] \text{ is increasing w.r.t. } s, \text{for any } z. \tag{B.40}$$

Since the ARG process is time-reversible, condition (B.40) is equivalent to:

$$P_0^Q\left[\sigma^2_{t+1}+\cdots+\sigma^2_{t+h-1}\geq z|\sigma_t^2=s,\sigma^2_{t+h}=\sigma_0^2\right] \text{ is increasing w.r.t. } s, \text{for any } z. \tag{B.41}$$

To show (B.41) we use the stochastic representation of Markov process (σ_t^2):

$$\sigma^2_{t+1}=g(\sigma_t^2,u_{t+1}), \tag{B.42}$$

where the innovation u_{t+1} is independent of $\underline{\sigma_t^2}$. By l-fold compounding of function g w.r.t. the first argument, we have $\sigma^2_{t+l}=g_l(\sigma_t^2,u_{t+1},\cdots,u_{t+l})$, say, and $\sigma^2_{t+1}+\cdots+\sigma^2_{t+h-1}=G(\sigma_t^2,u_{t+1},\cdots,u_{t+h-1})$, where $G(\sigma_t^2,u_{t+1},\cdots,u_{t+h-1})=g(\sigma_t^2,u_{t+1})+\cdots+g_{h-1}(\sigma_t^2,u_{t+1},\cdots,u_{t+h-1})$. Condition (B.41) becomes:

$$P_0^Q\left[G(s,u_{t+1},\cdots,u_{t+h-1})\geq z|\sigma^2_{t+h}=\sigma_0^2\right] \text{ is increasing w.r.t. } s, \text{for any } z.$$

This condition is satisfied if function G is increasing w.r.t. the first argument, that is, if the function g in the stochastic representation (B.42) is increasing w.r.t. the first argument. The latter condition is equivalent to σ^2_{t+1} being stochastically increasing in σ_t^2 under Q.

Finally, let us show that σ^2_{t+1} is stochastically increasing in σ_t^2 under Q for the ARG process. This follows from the gamma-Poisson mixture representation of the ARG process:

$$\sigma^2_{t+1}/c_0^* \quad |\zeta_{t+1}\sim\gamma(\delta_0^*+\zeta_{t+1}), \quad \zeta_{t+1}|\sigma_t^2\sim\mathcal{P}(\rho_0^*\sigma_t^2/c_0^*),$$

where γ and $\mathcal{P}$ denote gamma and Poisson distributions, respectively. Then σ^2_{t+1} is stochastically increasing in ζ_{t+1}, and ζ_{t+1} is stochastically increasing in σ_t^2. The conclusion follows.

B.10 Calibration of the parametric stochastic volatility model

In this Section we describe the computation by Fourier transform methods of the option prices in the parametric stochastic volatility model of Section 2.6 i) in the paper. These Fourier transform methods are used for the cross-sectional calibration of the model parameters. The risk-neutral distribution Q is given in equations (2.15)-(2.16). The option price is such that:

$$c_t(h,k) = B(t,t+h)E^Q\left[(\exp R_{t,h} - k)^+ |\sigma_t^2\right] = E^Q\left[\left(\exp \tilde{R}_{t,h} - \tilde{k}\right)^+ |\sigma_t^2\right],$$

where $\tilde{R}_{t,h} = \tilde{r}_{t+1} + \cdots + \tilde{r}_{t+h}$ is the cumulated excess return of the underlying asset between t and $t+h$ and $\tilde{k} = B(t,t+h)k$ is the discounted moneyness of the option. Let us introduce the variable $s := \log(\tilde{k})$ and define the function:

$$\phi(s) = e^{\alpha s}E^Q\left[\left(\exp \tilde{R}_{t,h} - e^s\right)^+ |\sigma_t^2\right], \quad s \in \mathbb{R},$$

for a given $\alpha > 0$ (for expository purpose we omit the dependence of function ϕ on time-to-maturity h and current volatility value σ_t^2). Following Carr and Madan (1999), the Fourier transform of ϕ is (see below):

$$\hat{\phi}(u) = \int_{-\infty}^{\infty} e^{-ius}\phi(s)\,ds = \frac{\Phi(iu-\alpha-1)}{\alpha^2+\alpha-u^2-iu(2\alpha+1)}, \quad u \in \mathbb{R}, \tag{B.43}$$

where

$$\Phi(z) = E^Q\left[\exp\left(-z\tilde{R}_{t,h}\right)|\sigma_t^2\right].$$

For the ARG model, function Φ is given by (see below):

$$\Phi(z) = \exp\left[-A_h\sigma_t^2 - B_h\right], \tag{B.44}$$

where $A_h = A_h(z)$ and $B_h = B_h(z)$ are defined recursively by:

$$\begin{aligned} A_h &= a_0^*(w + A_{h-1}) \quad , \quad A_1 = a_0^*(w), \\ B_h &= B_{h-1} + b_0^*(w + A_{h-1}) \quad , \quad B_1 = b_0^*(w), \end{aligned}$$

$w = -z(1+z)/2$, $a_0^*(u) = \frac{\rho_0^* u}{1+c_0^* u}$, $b_0^*(u) = \delta_0^* \log(1+c_0^* u)$. By inverse Fourier transform, we get the option price:

$$c_t(h,k) = \frac{e^{-\alpha s}}{2\pi}\int_{-\infty}^{\infty} e^{ius}\hat{\phi}(u)du,$$

where $s = \log(B(t,t+h)k)$ in the RHS. Since function $\phi(s)$ is real valued, we have $\hat{\phi}(-u) = \overline{\hat{\phi}(u)}$. It follows

$$c_t(h,k) = \frac{e^{-\alpha s}}{\pi}\,\mathrm{Re}\int_0^{\infty} e^{ius}\hat{\phi}(u)du. \tag{B.45}$$

To compute the integral (B.45), we introduce a finite upper integration boundary $\Lambda > 0$ and we discretize the resulting integral over $[0,\Lambda]$. More precisely, let $\Lambda > 0$ be such that $\left|\hat{\phi}(u)\right|$ is small for $u > \Lambda$. Define the grid $u_k = (\Lambda/N)(k-1)$, for $k = 1, ..., N$, where $N \in \mathbb{N}$ is the number of grid points. Then we have:

$$\begin{aligned} c_t(h,k) &\simeq \frac{e^{-\alpha s}}{\pi}\,\mathrm{Re}\int_0^{\Lambda} e^{ius}\hat{\phi}(u)du \\ &\simeq \frac{\Lambda e^{-\alpha s}}{\pi}\,\mathrm{Re}\,\frac{1}{N}\sum_{k=1}^{N} e^{i\frac{\Lambda s}{N}(k-1)}\hat{\phi}_k, \end{aligned}$$

where $\hat{\phi}_k := \hat{\phi}(u_k)$.

To summarize, the algorithm to compute $c_t(h,k)$ is as follows:

1. Compute the coefficients

$$\hat{\phi}_k := -\frac{1}{2}\frac{\exp\left[-A_h\sigma_t^2 - B_h\right]}{w_k},\ k = 1, 2, ..., N,$$

where

$$\begin{aligned} A_h &= a_0^*(w_k + A_{h-1}) \quad , \quad A_1 = a_0^*(w_k), \\ B_h &= B_{h-1} + b_0^*(w_k + A_{h-1}) \quad , \quad B_1 = b_0^*(w_k), \end{aligned}$$

$w_k = -\frac{1}{2}\left(\alpha^2 + \alpha - u_k^2 - iu_k(2\alpha+1)\right)$, $u_k = (\Lambda/N)(k-1)$.

2. Compute the inverse Fourier transform of the coefficients

$$c_t(h,k) = \frac{\Lambda e^{-\alpha s}}{\pi}\operatorname{Re}\frac{1}{N}\sum_{k=1}^{N} e^{i\frac{\Lambda s}{N}(k-1)}\hat{\phi}_k.$$

Proof of Equation (B.43): We have

$$\begin{aligned} \hat{\phi}(u) &= \int_{-\infty}^{\infty} e^{-(iu-\alpha)s} E_t^Q\left[\left(e^{\tilde{R}_{t,h}} - e^s\right)^+\right] ds \\ &= E_t^Q\left[\int_{-\infty}^{\infty} e^{-(iu-\alpha)s}\left(e^{\tilde{R}_{t,h}} - e^s\right)^+ ds\right] \\ &= E_t^Q\left[e^{\tilde{R}_{t,h}}\int_{-\infty}^{\tilde{R}_{t,h}} e^{-(iu-\alpha)s} ds - \int_{-\infty}^{\tilde{R}_{t,h}} e^{-(iu-\alpha-1)s} ds\right] \\ &= E_t^Q\left[-\frac{1}{iu-\alpha}e^{-(iu-\alpha-1)\tilde{R}_{t,h}} + \frac{1}{iu-\alpha-1}e^{-(iu-\alpha-1)\tilde{R}_{t,h}}\right] \\ &= \frac{1}{\alpha^2 + \alpha - u^2 - iu(2\alpha+1)} E_t^Q\left[e^{-(iu-\alpha-1)\tilde{R}_{t,h}}\right], \end{aligned}$$

where $E_t^Q[.] = E^Q[.|\sigma_t^2]$.

■

Proof of Equation (B.44): Under the risk-neutral distribution Q we have $\tilde{r}_t = -\frac{1}{2}\sigma_t^2 + \sigma_t\varepsilon_t$, where $\varepsilon_t \sim IIN(0,1)$ and $\left(\sigma_t^2\right)$ follows an ARG process independent of (ε_t) with parameters $\rho_0^*, \delta_0^*, c_0^*$. Thus:

$$\begin{aligned} \Phi(z) &= E_t^Q\left[\exp\left(\frac{z}{2}\sigma_{t,t+h}^2 - z\left(\sigma_{t+1}\varepsilon_{t+1} + ... + \sigma_{t+h}\varepsilon_{t+h}\right)\right)\right] \\ &= E_t^Q\left[\exp\left(\frac{1}{2}\left(z + z^2\right)\sigma_{t,t+h}^2\right)\right], \end{aligned}$$

where $\sigma_{t,t+h}^2 := \sigma_{t+1}^2 + ... + \sigma_{t+h}^2$. From standard results for affine processes in discrete time [e.g., Darolles, Gouriéroux, Jasiak (2006)], equation (B.44) follows. ■

References

[1] Andrews, D. (1991): An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables, Journal of Multivariate Analysis, 38, 187–203.

[2] Billingsley, P. (1965): Ergodic Theory and Information. New York: John Wiley.

[3] Billingsley, P. (1968): Convergence of Probability Measures. New York: John Wiley.

[4] Bosq, D. (1998): Nonparametric Statistics for Stochastic Processes: Estimation and Prediction. New York, Springer.

[5] Bosq, D., and P. Lecoutre (1987): Theorie de l'estimation fonctionnelle. Paris: Economica.

[6] Carr, P., and D. Madan (1999): Option Pricing and the Fast Fourier Transform, Journal of Computational Finance, 2, 61-73.

[7] Darolles, S., C. Gouriéroux, and J. Jasiak (2006): Structural Laplace Transform and Compound Autoregressive Models, Journal of Time Series Analysis, 27, 477-504.

[8] Gouriéroux, C., and J. Jasiak (2006): Autoregressive Gamma Process, Journal of Forecasting, 25, 129-152.

[9] Herrndorf, N. (1984): A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables, Annals of Probability, 12, 141-153.

[10] Pollard, D. (1990): Empirical Processes: Theory and Applications. CBMS Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

[11] Potscher, B., and I. Prucha (1989): A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes, Econometrica, 57, 675-683.

[12] Stock, J., and J. Wright (2000): GMM with Weak Identification, Econometrica, 68, 1055-1096.

[13] Tenreiro, C. (1995): Estimation fonctionnelle: Applications aux tests d'adéquation et de paramètre constant, Ph.D. Thesis, University of Paris VI.