Technical Appendix

for

Conditional Moment Models

under Semi-strong Identification.

Bertille Antoine* and Pascal Lavergne†

*Simon Fraser University. Email: bertille_antoine@sfu.ca.
†Toulouse School of Economics. Email: pascal.lavergne@univ-tlse1.fr
This Technical Appendix contains the proofs of the theoretical results of "Conditional Moments Models under Semi-strong Identification". The assumptions and theoretical results stated in the main paper are first recalled for completeness.

1 Assumptions and Theorems

1.1 Assumptions

Assumption 1. (Global Identifiability)
(i) The parameter space $\Theta$ is compact.
(ii) $\theta_0$ is the unique value in $\Theta$ satisfying (3.1), that is $E[g(Z_i, \theta)|X_i] = 0$ a.s. $\Rightarrow \theta = \theta_0$.

Assumption 2. (Semi-strong Identification)

\[ \tau(X_i, \theta) \equiv E\left[g(Z_i, \theta)|X_i\right] = \sum_{l=1}^{s} r_{l,n}^{-1} r_l(X_i, \theta_1, \ldots, \theta_l), \tag{1.1} \]

where $\theta_l$, $l = 1, \ldots, s$, are vectors of size $p_l$ that form a partition of $\theta$, and $r_{l,n}$ are real sequences such that (i) $r_{1,n} = 1 \to \infty$, (ii) $r_{l,n} = o(r_{l+1,n})$, $l = 1, \ldots, s - 1$, and (iii) $r_n \equiv \max_l [r_{l,n}] = o(\sqrt{n})$.

Assumption 3. (Data Generating Process)
The observations form a rowwise independent triangular array, where the marginal distribution of the continuously distributed $X$ remains unchanged.

Assumption 4. (Regularity of $K$)

$K(\cdot)$ is a symmetric, bounded density on $\mathbb{R}^q$, with integral equal to one. Its Fourier transform is strictly positive on $\mathbb{R}^q$ and non-increasing on $(0, \infty)$.

Assumption 5. (Regularity of $g$)

(i) The families $G_k = \{g^{(k)}(\cdot, \theta) : \theta \in \Theta\}$, $1 \leq k \leq r$, are uniformly Euclidean for an envelope $G$ with $\mathbb{E}G^4(Z) < \infty$.
(ii) Uniformly in $n$, $E g(Z, \theta)g'(Z, \theta)$ is continuous in $\theta$ and $\text{var} [g(Z, \theta_0)|X]$ is almost surely positive definite and bounded away from infinity.
Assumption 6. (Regularity of $\tau$)
The functions $\tau_l(x, \theta_1, \ldots, \theta_l)$, $l = 1, \ldots, s$, satisfy Condition 1.

Condition 1. A function $l(x, \theta)$ satisfies Condition 1 if (i) $\sup_{\theta} \|l(\cdot, \theta)\|f(\cdot)$ is in $L^1 \cap L^2$. (ii) For all $x$, the map $\theta \mapsto l(x, \theta)$ is continuous. (iii) For any $x$, all second partial derivatives of $l(x, \theta)$ exist on a neighborhood $N$ of $\theta_0$ independent of $x$. Each component of $\nabla_{\theta} l(X, \theta_0)\nabla_{\theta} f(X, \theta_0)$ belongs to $L^1 \cap L^2$ and $E\|\nabla_{\theta} l(X, \theta_0)\nabla_{\theta} f(X, \theta_0)\|^2 < \infty$. On the neighborhood $N$ of $\theta_0$, each second-order partial derivative is uniformly Euclidean for a common envelope $H$ with $E H(X) < \infty$.

Assumption 7. (Local Identifiability)
\[ \forall n, \ a' D_n \nabla_{\theta} \tau(X, \theta_0) = 0 \Rightarrow a = 0. \]

1.2 Theorems

Theorem 4.1. (Asymptotic Normality)
Under Assumptions 1–7, $\sqrt{n}(D_n \Delta D_n)^{-1/2}(D_n V D_n) D_n^{-1} \left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{d} N(0, I_p)$.

Theorem 4.2. (Wald Test)
Assume that (i) for any $x$, $g(x, \cdot)$ is differentiable in a neighborhood $N$ of $\theta_0$ independent of $x$, with first derivative Euclidean on this neighborhood for an envelope $L$ with $E L^2(Z) < \infty$, and that (ii) $h(\cdot)$ is continuously differentiable with $\nabla_{\theta} h(\theta_0)$ of full rank $m$.
Then, under the assumptions of Theorem 4.1, $W_n$ is asymptotically chi-square with $m$ degrees of freedom under $H_0$, and $W_n \xrightarrow{p} + \infty$ whenever $h(\theta_0) \neq 0$.

2 Preliminary Results

2.1 Convergence Rates for $U$-Processes

In our main proofs, we will often use results on convergence rates for $U$-statistics as derived by Sherman (1994), namely his Corollaries 4, 7, and 8. These results are derived for i.i.d. observations. However, it is easy to see that they extend to our setup of rowwise independent triangular array of data. Indeed, all these results directly derive from his Main Corollary,
an inequality that holds for any finite $n$ under the assumption that the envelope function $F(\cdot)$ satisfies $\mathbb{E} F(Z_n) < \infty$. It is then straightforward to extend Corollaries 4, 7 and 8 of Sherman (1994) to our setup. As an example, we state and prove the first result.

**Corollary 2.1** (Sherman’s (1994) Corollary 4). For a rowwise independent triangular array of observations $\{Z_{1n}, \ldots Z_{nn}\}$, let $\mathcal{F}$ be a class of functions such that $\forall f \in \mathcal{F}$, $\mathbb{E} f(s_1, \ldots, s_{i-1}, \cdot, s_{i+1}, \ldots s_k) > 0$, $i = 1, \ldots k$. Suppose $\mathcal{F}$ is Euclidean for an envelope $F$ satisfying $\limsup_n \mathbb{E} F^2(Z_n) < \infty$. Then

$$
\sup_{\mathcal{F}} |n^{k/2} U_k^n f| = O_p(1), \quad \text{where} \quad U_k^n f \equiv (n)_k^{-1} \sum_{i_k} f(Z_{i_1 n}, \ldots Z_{i_k n})
$$

and $i_k = (i_1, \ldots i_k)$ ranges over the $(n)_k$ ordered $k$-tuples of distinct integers from $\{1, \ldots n\}$.

**Proof.** Sherman’s Main Corollary with $p = 1$ yields that for any $0 < \alpha < 1$ and any $n$

$$
\mathbb{E} \sup_{\mathcal{F}} |n^{k/2} U_k^n f| \leq \Omega_n \left[ \mathbb{E} \sup_{\mathcal{F}} (U_{2n}^k f^2)^\alpha \right]^{1/2},
$$

where $\Omega_n = C_\alpha (\mathbb{E} F^2(Z_n))^{\epsilon/2}$ for some $\epsilon \in (0, 1)$. Now

$$
\limsup_n \Omega_n^2 \left[ \mathbb{E} \sup_{\mathcal{F}} (U_{2n}^k f^2)^\alpha \right] \leq C_\alpha^2 (\mathbb{E} F^2(Z))^\epsilon \limsup_n \left[ \mathbb{E} (U_{2n}^k f^2)^\alpha \right] 
\leq C_\alpha^2 (\mathbb{E} F^2(Z))^\epsilon (\mathbb{E} F^2(Z))^\alpha < \infty.
$$

Conclude from Chebyshev’s inequality. \hfill \Box

### 2.2 Matrices

For a real-valued function $l(\cdot)$, denote by $\mathcal{F}[l](\cdot)$ its Fourier transform, and by $\overline{l}(\cdot)$ its conjugate.

**Lemma 2.2.** Under Assumptions 4 to 7, $D_n V D_n$ and $D_n \Delta D_n$ have eigenvalues uniformly bounded away from zero and infinity.

**Proof.** Let $\delta_n(X) = D_n \nabla \theta \tau(X, \theta_0)$. We have $D_n V D_n = \mathbb{E} [\delta_n(X_1) \delta_n'(X_2) K(X_1 - X_2)]$. Denote convolution by $\ast$. From Assumptions 4, 6 (Condition 1-(i) and (iii)), and the properties
of Fourier transforms,

\[ a' E \left[ \delta_n'(X_1) \delta_n'(X_2) K(X_1 - X_2) \right] a = \int_{\mathbb{R}^q} a' \delta_n(X_1) f(X_1) (f \delta_n a * K) (X_1) \, dX_1 = \int_{\mathbb{R}^q} \mathcal{F} [a' \delta_n f] (t) \mathcal{F} [f \delta_n a * K] (t) \, dt \]

\[ = \int_{\mathbb{R}^q} \mathcal{F} [a' \delta_n f] (t) \mathcal{F} [f \delta_n a] (-t) \, dt = \int_{\mathbb{R}^q} \mathcal{F} [a' \delta_n f] (t) \mathcal{F} [f \delta_n a] (-t) \mathcal{F} [K] (t) \, dt \]

\[ = \int_{\mathbb{R}^q} |\mathcal{F} [a' \delta_n f] (t)|^2 \mathcal{F} [K] (t) \, dt. \]

From the strict positivity of \( \mathcal{F} [K] (t) \), all eigenvalues of \( D_n V D_n \) are non-negative. Since \( \mathcal{F} [K] (t) \leq 1 \ \forall t \), and from Assumption 6, the above quantity is bounded for any \( a \) of norm 1, so that eigenvalues of \( D_n V D_n \) are bounded. Moreover, the minimum eigenvalue is zero iff

\[ \exists a \neq 0 : a' \delta_n(X) f(X) = 0 \text{ a.s.} \iff \exists a \neq 0 : a' \delta_n(X) = 0, \]

which would contradict Assumption 7.

The matrix \( D_n \Delta D_n \) is the variance matrix of \( E \left[ \delta_n(X_0) K(X - X_0) | X \right] g(Z, \theta_0) \), so that it is singular iff there exists \( a \neq 0 \) such that

\[ a' E \left[ \delta_n(X_0) K(X - X_0) | X \right] g(Z, \theta_0) = 0 \ \text{a.s.} \]

Given that \( g(Z, \theta_0) \) cannot be identically zero by Assumption 5-(ii), this is equivalent to

\[ a' E \left[ \delta_n(X_0) K(X - X_0) | X \right] = a' (\delta_n f * K)(X) = 0 \ \text{a.s.} \]

\[ \iff \mathcal{F} [a' \delta_n f] (t) \mathcal{F} [K] (t) = 0 \ \forall t \in \mathbb{R}^q \iff a' \delta_n(X) = 0 \ \text{a.s.} \]

which would contradict Assumption 7.

\[ \square \]

3 Main Proofs

3.1 Proof of Theorem 4.1

For the sake of simplicity, we detail most of our arguments only for the simplest case (3.3), and explain how they easily adapt to the general case (3.2), that is in three directions: (1)
there may be only one rate for the whole parameter, (2) there may be more than two rates, and (3) the slowest rate \( r_{1n} \) could diverge.

(i) \textit{Approximation of the criterion.} Let write our criterion \( Q_n(\theta) = M_n(\theta)/\sigma_n^2(\theta) \) where

\[
M_n(\theta) = \frac{1}{2n(n-1)} \sum_{i \neq j} g'(Z_i, \theta)g(Z_j, \theta)K(X_i - X_j)
\]

and \( \sigma_n^2(\theta) = \frac{1}{n} \sum_i g'(Z_i, \theta)g(Z_i, \theta) \).

From Assumption 5-(i) and Corollary 2.1, \( \sup_{\theta} |\sigma_n^2(\theta) - \text{E} g'(Z, \theta)g(Z, \theta)| = O_p(n^{-1/2}) \). Moreover, from Assumptions 1 and 5, \( \sigma^2(\theta) \equiv \text{E} g'(Z, \theta)g(Z, \theta) \) is uniformly bounded away from zero and infinity. Hence \( Q_n(\theta) = \frac{M_n(\theta)}{\sigma_n(\theta)} (1 + O_p(n^{-1/2})) \) uniformly in \( \theta \) and

\[
Q_n(\theta) - Q_n(\theta_0) = [M_n(\theta) - M_n(\theta_0)] \frac{1}{\sigma^2(\theta)} (1 + O_p(n^{-1/2})) + M_n(\theta_0) \left[ \frac{1}{\sigma^2(\theta)} - \frac{1}{\sigma^2(\theta_0)} \right] (1 + O_p(n^{-1/2}))
\]

uniformly in \( \theta \), since \( M_n(\theta_0) \) is a degenerate second-order \( U \)-statistic.

(ii) \textit{Consistency of \( \tilde{\alpha}_n \).} The parameter value \( \theta_0 \) is the unique minimizer of \( \text{E} M_n(\theta) \). Indeed, reason as in the proof of Lemma 2.2 to get that

\[
\text{E} M_n(\theta) = \frac{1}{2} \text{E} \left[ \tau'(X_1, \theta)\tau(X_2, \theta)K(X_1 - X_2) \right]
\]

\[
= \frac{1}{2} \sum_{k=1}^r \left\{ \int_{\mathbb{R}^q} |\mathcal{F} [\tau^{(k)}(\cdot, \theta)f(\cdot)] (t)|^2 \mathcal{F} [K] (t) dt \right\} \geq 0 , \quad (3.2)
\]

so that by Assumption 1

\[
\text{E} M_n(\theta) = 0 \iff \mathcal{F} [\tau^{(k)}(\cdot, \theta)f(\cdot)] (t) = 0 \quad \forall t \in \mathbb{R}^q , \ k = 1, \ldots r
\]

\[
\iff \text{E} [g(Z, \theta)|X] = 0 \quad \text{a.s.} \iff \theta = \theta_0 .
\]

By Assumption 6, \( \text{E} M_n(\theta) \) is continuous in \( \theta \), and then in \( \alpha \). Hence from Assumption 1, \( \forall \varepsilon > 0, \exists \mu > 0 \) such that \( \inf_{\|\alpha - \alpha_0\| \geq \varepsilon} \text{E} M_n(\theta) \geq \mu \). The family \( \{g'(Z_1, \theta)g(Z_2, \theta)K(X_1 - X_2) : \theta \in \Theta\} \) is uniformly Euclidean for an envelope \( F(\cdot) \) such that \( \text{E} F(Z_1, Z_2) \) from Assumptions 4 and 5. By Corollary 7 of Sherman (1994), \( \sup_{\theta \in \Theta} |M_n(\theta) - \text{E} M_n(\theta)| = O_p(n^{-1/2}) \). Hence

\[
\inf_{\|\alpha - \alpha_0\| \geq \varepsilon} M_n(\theta) - M_n(\theta_0) \geq \mu + O_p(n^{-1/2}) .
\]

6
Using (3.1), $\inf_{||\alpha - \alpha_0|| \geq \varepsilon} Q_n(\theta) - Q_n(\theta_0) \geq C + O_p(n^{-1/2})$, for some $C > 0$. Since $Q_n(\theta) \leq Q_n(\theta_0)$, it follows that $\forall \varepsilon > 0$ $Pr [||\hat{\alpha}_n - \alpha_0|| < \varepsilon] \to 1$.

Extension: If the rate $r_n$ for $\alpha$ diverges, then use Hoeffding’s decomposition of $M_n(\theta)$ and Corollary 2.1 to obtain $\sup_{\theta \in \Theta} |M_n(\theta) - E M_n(\theta)| = O_p(n^{-1}) + O_p(n^{-1/2}r^{-1}_n)$ and note that $\forall \varepsilon > 0$, $\exists \mu > 0$ such that $\inf_{||\alpha - \alpha_0|| \geq \varepsilon} E M_n(\theta) \geq r^{-2}_n \mu$. Then proceed as above.

(iii) Consistency of $\tilde{\beta}_n$. Apply Hoeffding’s decomposition to $M_n(\theta) - M_n(\theta_0)$. The second order $U$-process in the decomposition of $M_n(\theta) - M_n(\theta_0)$ is $O_p(n^{-1})$ uniformly over $\theta$ by Assumption 5 and Corollary 7 of Sherman (1994). Consider the first-order $U$-process $\mathbb{P}_n l_\theta$, where $l_\theta(Z_i) = E[l_\theta(Z_i, Z_j) | Z_i] + E[l_\theta(Z_i, Z_j) | Z_j] - 2 E[l_\theta(Z_i, Z_j)]$,

$$
l_\theta(Z_i, Z_j) = \left(\frac{1}{2}\right) (g'(Z_i, \theta)g(Z_j, \theta) - g'(Z_i, \theta_0)g(Z_j, \theta_0)) h^{-q} K \left(\frac{(X_i - X_j)}{h}\right)
+ \left(\frac{1}{2}\right) g'(Z_i, \theta) (g(Z_j, \theta) - g(Z_j, \theta_0)) K (X_i - X_j)
+ \left(\frac{1}{2}\right) g'(Z_i, \theta) (g(Z_j, \theta) - g(Z_j, \theta_0))' (g(Z_j, \theta) - g(Z_j, \theta_0)) K (X_i - X_j)
= l_{1\theta}(Z_i, Z_j) + l_{2\theta}(Z_i, Z_j) + l_{3\theta}(Z_i, Z_j),$$

and $l_{1\theta}(Z_i, Z_j) = l_{2\theta}(Z_j, Z_i)$ by the symmetry of $K(\cdot)$. Now

$$
2 E[l_{1\theta}(Z_i, Z_j) | Z_i]
= g'(Z_i, \theta_0) \left\{ E[\tau_\alpha(X, \alpha) K (X_i - X) | X_i] + r^{-1}_n E[\tau_\beta(X, \theta) K (X_i - X) | X_i] \right\}.
$$

The $U$-process based on the second part of (3.3) is $O_p(r^{-1}_n n^{-1/2})$ uniformly in $\theta$. Using Assumption 6, the first term in (3.3) admits uniformly for $\alpha$ in a $o(1)$ neighborhood of $\alpha_0$ the expansion

$$
g'(Z_i, \theta_0) \left[ \int_{\mathbb{R}^q} \nabla'_\alpha \tau_{\alpha_0}(x, \alpha_0) f(x) K (X_i - x) \, dx \right] (\alpha - \alpha_0)
+ \frac{1}{2} g'(Z_i, \theta_0) \sum_{k,l=1}^p \left( \alpha^{(k)} - \alpha_0^{(k)} \right) \left( \alpha^{(l)} - \alpha_0^{(l)} \right)
\left[ \int_{\mathbb{R}^q} H_{\alpha^{(k)}, \alpha^{(l)}} \tau_{\alpha}(x, \alpha_0) f(x) K (X_i - x) \, dx \right] + R_n(\hat{\alpha}(Z_i, \alpha)).
$$

The $U$-statistic based on the first element of (3.4) is $\|\alpha - \alpha_0\| O_p(n^{-1/2})$. The one based on the second element of (3.4) is $\|\alpha - \alpha_0\|^2 O_p(n^{-1/2})$. The remaining term is a $U$-process.
of the form \((\alpha - \alpha_0)' C_n(\alpha) (\alpha - \alpha_0)\), where \(C_n\) has typical element

\[
\frac{1}{2n} \sum_{i=1}^{n} g'(Z_i, \theta_0) \left[ \int_{\mathbb{R}^d} (H_{\alpha^{(k)}(x)} \tau_\alpha(x, \bar{\alpha}) - H_{\alpha^{(k)}(x)} \tau_\alpha(x, \alpha_0)) f(x) K(X_i - x) \, dx \right],
\]

where \(||\bar{\alpha} - \alpha_0|| \leq ||\alpha - \alpha_0||\). The above function has a squared integrable envelope from Assumptions 5 and 6, and approaches zero when \(\alpha\) tends to \(\alpha_0\). Use Corollary 8 of Sherman (1994) to obtain that the remaining term is an \(||\alpha - \alpha_0||^2 o_p(n^{-1/2})\). Use similar arguments for \(2 E[|_3(\bar{Z}, Z_j)| Z]\). We thus obtain that uniformly in \(\beta\) and uniformly over \(o(1)\) neighborhoods of \(\alpha_0\)

\[
M_n(\theta) - M_n(\theta_0) = E M_n(\theta) + ||\alpha - \alpha_0|| O_p(n^{-1/2}) + ||\alpha - \alpha_0||^2 o_p(1) + O_p(n^{-1/2} r_n^{-1}). \tag{3.5}
\]

From Assumption 6 and a Taylor expansion of \(\tau_\alpha(X, \alpha)\), for \(\alpha\) in a \(o(1)\) neighborhood of \(\alpha_0\),

\[
E M_n(\theta) \geq E [\tau'_\alpha(X_1, \alpha) \tau_\alpha(X_2, \alpha) K(X_1 - X_2)]
\geq (\alpha - \alpha_0)' E [\nabla_\alpha \tau_\alpha(X_1, \alpha_0) \nabla'_\alpha \tau_\alpha(X_2, \alpha_0) K(X_1 - X_2)] (\alpha - \alpha_0) + o(||\alpha - \alpha_0||^2).
\]

Since the above matrix is positive definite, see Lemma 2.2, then \(\forall \varepsilon > 0, \exists \mu > 0\) such that \(\inf E M_n(\theta) \geq \mu ||\alpha - \alpha_0||^2\). This and (3.5) imply that for some \(C > 0\)

\[
\inf_{||\alpha - \alpha_0|| \geq \varepsilon r_n^{-1}} M_n(\theta) - M_n(\theta_0) \geq \mu r_n^{-2} + o_p(r_n^2).
\]

Now (3.1) implies that for some \(C > 0\)

\[
\inf_{||\alpha - \alpha_0|| \geq \varepsilon r_n^{-1}} Q_n(\theta) - Q_n(\theta_0) \geq C r_n^{-2} + o_p(r_n^2).
\]

Since \(Q_n(\bar{\theta}) \leq Q_n(\theta_0), ||\bar{\alpha}_n - \alpha_0|| = o_p(r_n^{-1})\).

For \(||\alpha - \alpha_0|| = o(r_n^{-1}), \) (3.5) yields \(M_n(\theta) - M_n(\theta_0) = E M_n(\theta) + o_p(r_n^{-2})\). As

\[
E M_n(\theta) \geq r_n^{-2} E [\tau'_\beta(X_1, \theta) \tau_\beta(X_2, \theta) K(X_1 - X_2)],
\]

and the latter is continuous in \(\beta\), we obtain that \(\forall \varepsilon > 0, \exists \mu > 0\) such that \(\inf_{||\beta - \beta_0|| \geq \varepsilon} E M_n(\theta) \geq \mu r_n^{-2} ||\beta - \beta_0||^2\), and then that for some \(C > 0\)

\[
\inf_{||\bar{\alpha}_n - \alpha_0|| = o(r_n^{-1}), ||\beta - \beta_0|| \geq \varepsilon} Q_n(\theta) - Q_n(\theta_0) \geq C r_n^{-2} + o_p(r_n^{-2}).
\]
Since \( Q_n(\hat{\theta}) \leq Q_n(\theta_0) \), this in turn yields \( \| \hat{\beta}_n - \beta_0 \| = o_p(1) \).

Extension: If there are more than two rates, e.g. the case where \( \theta = (\alpha, \beta, \lambda) \) with corresponding rates \( (1, r_{2n}, r_{3n}) \), proceed as in Part (iii) to show first that \( \| \hat{\beta}_n - \beta_0 \| = o_p(r_{3n}) \) and then that \( \| \hat{\lambda}_n - \lambda_0 \| = o_p(1) \).

(iv) Rate of convergence and asymptotic normality. Apply once again Hoeffding’s decomposition to \( M_n(\theta) - M_n(\theta_0) \) as in the previous part. The second order \( U \)-process in this decomposition is \( o_p(n) \) uniformly over \( o(1) \) neighborhoods of \( \theta_0 \) from Assumption 5 and Corollary 8 of Sherman (1994). To treat the first-order empirical process \( \mathbb{P}_n \mathbb{E} [ l_{i\theta}(Z_i, Z_j) | Z_i] \), use this time a Taylor expansion in \( \theta \), that is,

\[
2 \mathbb{E} [ l_{i\theta}(Z_i, Z_j) | Z_i] = g_n'(Z_i, \theta_0) \mathbb{E} \left[ (g_n(Z, \theta) - g_n(Z, \theta_0)) K(X_i - x) \right] (\theta - \theta_0) + \frac{1}{2} g_n''(Z_i, \theta_0) \sum_{k,l=1}^{p} \left( \theta^{(k)} - \theta_0^{(k)} \right) \left( \theta^{(l)} - \theta_0^{(l)} \right) \left[ \int \mathbb{H}_x(K, \theta_0) f(x) K(X_i - x) \right] + R_{1n}(Z_i, \theta). \tag{3.6}
\]

Use the structure of \( \tau(\cdot, \cdot) \) to decompose the first element of (3.6) into

\[
g_n'(Z_i, \theta_0) \left[ \int \mathbb{E}_{\alpha} \mathbb{E}_{\alpha} \mathbb{E}_{\alpha} \left( \nabla_{\alpha} \tau_{\alpha}(x, \alpha_0) + r_{n}^{-1} \nabla_{\alpha} \tau_{\alpha}(x, \theta_0) \right) f(x) K(X_i - x) \right] (\alpha - \alpha_0)
\]

\[
+ r_{n}^{-1} g_n'(Z_i, \theta_0) \left[ \int \mathbb{E}_{\alpha} \mathbb{E}_{\alpha} \mathbb{E}_{\alpha} \left( \nabla_{\beta} \tau_{\beta}(x, \theta_0) f(x) K(X_i - x) \right) \right] (\beta - \beta_0) .
\]

Use the same reasoning as in Part (ii) to conclude that the corresponding U-statistic can be written as \( n^{-1/2} A_n D_n^{-1} \theta \), where \( A_n = O_p(1) \). The U-statistic based on the second element of (3.6) can be decomposed as

\[
(\alpha - \alpha_0)' B_{n \alpha \alpha} (\alpha - \alpha_0) + 2 r_{n}^{-1} (\alpha - \alpha_0)' B_{n \alpha \beta} (\beta - \beta_0) + r_{n}^{-1} (\beta - \beta_0)' B_{n \beta \beta} (\beta - \beta_0),
\]

where \( B_{n \alpha \alpha}, B_{n \alpha \beta}, \) and \( B_{n \beta \beta} \) are \( O_p(n^{-1/2}) \), so it is an \( \| \alpha - \alpha_0 \|^2 O_p(n^{-1/2}) + \| \beta - \beta_0 \|^2 O_p(n^{-1/2} r_{n}^{-1}) + \| \alpha - \alpha_0 \| | \beta - \beta_0 \| O_p(n^{-1/2}) \). The empirical process in the remaining term can be shown to be of a smaller order similarly to what was done in Part (iii), using Assumption 6 and
Corollary 8 of Sherman (1994). For the empirical process $\mathbb{P}_n E \left[ l_{\theta} (Z_i, Z_j) | Z_i \right]$, use a similar expansion, the fact that

$$E \left[ (g(Z_i, \theta) - g(Z_i, \theta_0))' \left[ \int_{\mathbb{R}^q} \nabla'_\theta \tau(x, \theta_0) f(x) K (X_i - x) \, dx \right] \right] \to 0$$

and

$$E \left[ (g(Z_i, \theta) - g(Z_i, \theta_0))' \left[ \int_{\mathbb{R}^q} \nabla'_\beta \tau(x, \theta_0) f(x) K (X_i - x) \, dx \right] \right] \to 0$$

as $\theta - \theta_0 \to 0$, and Corollary 8 of Sherman (1994) to conclude that it is of a smaller order than $\mathbb{P}_n E \left[ l_{\theta} (Z_i, Z_j) | Z_i \right]$ uniformly in $\theta$ in a $o(1)$ neighborhood of $\theta_0$.

Let us now gather the results. Adopting the reparametrization $\gamma = D_n^{-1} \theta$ yields

$$M_n(\theta) - M_n(\theta_0) = \frac{1}{\sqrt{n}} A_n' (\gamma - \gamma_0) + \| \gamma - \gamma_0 \|^2 o_p(1) + o_p(n^{-1}),$$

uniformly for $\gamma$ in a $o(r_n^{-1})$ neighborhood of $\gamma_0$. For $\theta$ in a $o(1)$ neighborhood of $\theta_0$, $\sigma^2(\theta) = \sigma^2(\theta_0) + o(1)$, which is bounded from below by Assumption 5. Equation (3.1) thus implies

$$Q_n(\theta) - Q_n(\theta_0) = \frac{E M_n(\theta)}{\sigma^2(\theta_0)} + \frac{n^{-1/2}}{\sigma^2(\theta_0)} A_n' (\gamma - \gamma_0) + \| \gamma - \gamma_0 \|^2 o_p(1) + o_p(n^{-1}). \quad (3.7)$$

Moreover, as $\nabla_\theta E M_n(\theta_0) = 0$,

$$\frac{E M_n(\theta) - E M_n(\theta_0)}{\sigma^2(\theta_0)} = \left[ (\theta - \theta_0)' \nabla_\theta E M_n(\theta_0) + \frac{1}{2} (\theta - \theta_0)' V (\theta - \theta_0) + R_1 \right] \sigma^{-2}(\theta_0) = \frac{1}{2\sigma^2(\theta_0)} (\gamma - \gamma_0)' D_n V D_n (\gamma - \gamma_0) + o(\| \gamma - \gamma_0 \|^2) \geq C \| \gamma - \gamma_0 \|^2,$$

for some $C > 0$, by Assumption 6 and since $D_n V D_n$ has eigenvalues bounded away from zero by Lemma 2.2. Use now (3.7) to deduce that $\| \tilde{\gamma} - \gamma_0 \|^2 = O_p(n^{-1/2})$ by Theorem 1 of Sherman (1993), see also the extension of Lavergne and Patilea (2010) that allows to deal with a varying limit criterion. Therefore

$$Q_n(\theta) = Q_n(\theta_0) + \frac{1}{\sqrt{n}} A_n' (\gamma - \gamma_0) + \frac{1}{2} (\gamma - \gamma_0)' D_n V D_n (\gamma - \gamma_0) + o_p(n^{-1}),$$

uniformly over $O(n^{-1/2})$ neighborhoods of $\gamma_0$, and $\sqrt{n} (D_n V D_n) (\tilde{\gamma}_n - \gamma_0) + A_n = o_p(1)$ from Theorem 2 of Sherman (1993). By Lemma 2.2, the variance $D_n \Delta D_n$ of $A_n$ is non-singular and bounded, and by a standard central limit theorem for triangular arrays, $(D_n \Delta D_n)^{-1/2} A_n$ is asymptotically normal with mean zero and variance identity. This concludes the proof.
3.2 Proof of Theorem 4.2

To simplify the exposition, most of the proof is performed with only two groups of parameters, i.e. $\theta = (\alpha', \beta')'$ and (3.3) holds, and we explain how this generalizes to more complex setups. Following Antoine and Renault (2012), we proceed in two main steps. First, we define an equivalent formulation of $H_0$ as $\tilde{h}(\theta_0) = 0$, where $\tilde{h}$ is a linear transformation of $h$ that separates the restrictions into (i) restrictions that involve components of $\alpha$ only and are therefore strongly identified, and (ii) restrictions that gathers the remaining restrictions. Second, we show that the Wald test statistic for testing $\tilde{h}(\theta) = 0$ is numerically equal to the Wald statistic for testing $H_0$ and we derive its asymptotic behavior. The two extreme cases where all restrictions are identified at the same rate, whether strongly or weakly, do not require the first step.

The space $I_1 \equiv \text{col} (\nabla_\theta h(\theta_0)) \cap \text{col} (\nabla_\theta \alpha')$ is the space of tested restrictions that are identified at the standard rate $\sqrt{n}$. Let its dimension be $m_1$ and $\epsilon_i$, $i = 1, \ldots, m_1$, be vectors of $\mathbb{R}^{m_1}$ such that $[\nabla_\theta h(\theta_0)\epsilon_i]_{i=1}^{m_1}$ is a basis of $I_1$. The remaining $(m - m_1)$ directions that are identified at the slower rate $r_n$ belongs to the space $I_2 \equiv \text{col} (\nabla_\theta h(\theta_0)) \cap [I_1]^\perp$. Let similarly $\epsilon_i$, $i = m_1 + 1, \ldots, m$ be vectors of $\mathbb{R}^{m}$ such that $[\nabla_\theta h(\theta_0)\epsilon_i]_{i=m_1+1}^{m}$ is a basis of $I_R$. Consider the invertible matrix $H' = [\epsilon_1 \epsilon_2 \cdots \epsilon_s]$. Then $H_0 : h(\theta_0) = 0$ is equivalent to $\tilde{h}(\theta_0) = 0$ with $\tilde{h}(\theta) = Hh(\theta)$.

Extension: In cases where there are more than two rates and $\theta = (\theta_1, \ldots, \theta_s)$, one should define $s$ spaces $I_l$, $l = 1, \ldots, s$, such that $I_l$ is the (possibly empty) space of tested directions that are identified at rate faster or equal than $r_{ln}$. The vectors $\epsilon_i$, $i = 1, \ldots, m$ and the matrix $H$ are thus defined similarly as above.

The Wald statistics based on $h(\cdot)$ and $\tilde{h}(\cdot)$ are numerically equal, since

$$ W_n(h) = n h'(\tilde{\theta}_n) \left[ \nabla_\theta h(\tilde{\theta}_n)V_n^{-1}\Delta_n V_n^{-1}\nabla_\theta h(\tilde{\theta}_n) \right]^{-1} h(\tilde{\theta}_n) $$

$$ = n (Hh)'(\tilde{\theta}_n) \left[ H\nabla_\theta h(\tilde{\theta}_n)V_n^{-1}\Delta_n V_n^{-1}\nabla_\theta h(\tilde{\theta}_n)H' \right]^{-1} Hh(\tilde{\theta}_n) $$

$$ = n \tilde{h}'(\tilde{\theta}_n) \left[ \nabla_\theta \tilde{h}(\tilde{\theta}_n) V_n^{-1}\Delta_n V_n^{-1}\nabla_\theta \tilde{h}(\tilde{\theta}_n) \right]^{-1} \tilde{h}(\tilde{\theta}_n) \equiv \tilde{W}_n. $$

11
Hence we can equivalently prove our theorem for $\tilde{W}_n$. Now this statistic equals

$$\begin{align*}
\left[\sqrt{n}D_n^{-1}1\right]' \left(\left(\frac{\partial}{\partial \theta} h(\tilde{\theta}_n) \right) (D_n V_n D_n)^{-1} (D_n \Delta D_n) (D_n V_n D_n)^{-1} \left(\frac{\partial}{\partial \theta} h(\tilde{\theta}_n) \right) D_n^{-1}\right)^{-1} \\
\times \left[\sqrt{n}D_n^{-1}1\right],
\end{align*}$$

where

$$\tilde{D}_n = \begin{pmatrix} I_{s_1} & 0 \\ 0 & r_n I_{s_2} \end{pmatrix}.$$ 

From the consistency of $\tilde{\theta}$, Assumption 5, the assumption on the derivative of $g(\cdot, \theta)$ on $\mathcal{N}$, and Hoeffding’s decomposition,

$$D_n (V_n - V) D_n = O_p(r_n^2 n^{-1}) + O_p(r_n n^{-1/2}) = o_p(1),$$

and similarly $D_n (\Delta_d - \Delta) D_n = o_p(1)$. From Lemma 2.2, $D_n V D_n$ and $D_n \Delta D_n$ have finite and non-vanishing eigenvalues. From a mean-value expansion of $\tilde{h}$ around $\theta_0$,

$$\tilde{h}(\tilde{\theta}_n) = \tilde{h}(\theta_0) + \nabla' \tilde{h}(\tilde{\theta}_n)(\tilde{\theta}_n - \theta_0)$$

$$\Leftrightarrow \sqrt{n}D_n^{-1}(\tilde{h}(\tilde{\theta}_n) - \tilde{h}(\theta_0)) = \left[\tilde{D}_n^{-1} \nabla' \tilde{h}(\tilde{\theta}_n) D_n \right] \left[\sqrt{n}D_n^{-1} \left(\tilde{\theta}_n - \theta_0\right)\right],$$

with $\|\tilde{\theta}_n - \theta_0\| \leq \|\tilde{\theta}_n - \theta_0\|$. The desired result then follows from Theorem 4.1 if $D_n \nabla\tilde{h}(\tilde{\theta}_n) \tilde{D}_n^{-1}$, and then $\tilde{D}_n^{-1} \nabla' \tilde{h}(\tilde{\theta}_n) D_n$, converges to a full rank matrix. Finally,

$$D_n \nabla\tilde{h}(\tilde{\theta}_n) \tilde{D}_n^{-1} = D_n \nabla\tilde{h}(\tilde{\theta}_n) H' \tilde{D}_n^{-1}$$

$$= \begin{pmatrix} I_{p_1} & 0 \\ 0 & r_n I_{p_2} \end{pmatrix} \begin{pmatrix} \nabla h(\tilde{\theta}_n)_{\epsilon_1} \\ r_n^{-1} \nabla h(\tilde{\theta}_n)_{\epsilon_1} \end{pmatrix}_{i=1,\ldots,s_1}$$

$$= \begin{pmatrix} \nabla h(\tilde{\theta}_n)_{\epsilon_1} \\ r_n^{-1} \nabla h(\tilde{\theta}_n)_{\epsilon_1} \end{pmatrix}_{i=s_1+1,\ldots,s}$$

$$= \begin{pmatrix} \nabla h(\tilde{\theta}_n)_{\epsilon_1} \\ r_n^{-1} \nabla h(\tilde{\theta}_n)_{\epsilon_1} \end{pmatrix} \nabla h(\tilde{\theta}_n)_{\epsilon_1}_{i=s_1+1,\ldots,s}$$

$$\rightarrow \begin{pmatrix} \nabla h(\tilde{\theta}_n)_{\epsilon_1} \\ 0 \end{pmatrix} \nabla h(\tilde{\theta}_n)_{\epsilon_1}_{i=s_1+1,\ldots,s}$$

by the continuous mapping theorem, and this matrix is full rank by construction.
References

