# Supplement to
# Empirical Bayes Regression with Many Regressors


Thomas A. Knox

The University of Chicago Graduate School of Business

James H. Stock

Department of Economics, Harvard University

Mark W. Watson

Department of Economics and
Woodrow Wilson School, Princeton University



Revised December 2003

# A Proofs of Theorems

This part of the supplement contains the proofs omitted from the paper and appendix. All the proofs shown here use the simplifying assumption that $K = \rho T$, where $0 < \rho < 1$, that is, the proofs are for $K = \rho T^{\delta}$ with $\delta = 1$. The integer constraint on $K$ is ignored. Modifications of these proofs for $\delta < 1$ are straightforward, and these modifications are discussed in the final section of this part.

First a word about terminology. References to "Assumption 1," "Lemma 1," and "Theorem 1" are references to the indicated assumption, lemma, or theorem in the paper. Theorems 7 through 10 are stated and proven in this supplement; Theorem 7 is an extension of our results, while Theorems 8 through 10 are foundations needed for our results. This supplement also has its own lemmas, which are called "Lemma S-1," etc.

## A.1 Preliminary Results

We start by collecting some additional definitions. Some of these repeat definitions in the paper and are included for completeness.

**Definition 1**

$$
\begin{aligned}
d_K &\equiv \sqrt{\frac{\sigma_\varepsilon^2}{64}\log K} \\
\hat{b}_{-i} &\equiv \left(\hat{b}_1, \ldots, \hat{b}_{i-1}, \hat{b}_{i+1}, \ldots, \hat{b}_K\right) \\
\phi(u) &\equiv \textit{the univariate normal density with} \\
&\quad \textit{mean } 0 \textit{ and variance } \sigma_\varepsilon^2 \\
f_{iK}(u_i) &\equiv \textit{the marginal likelihood of } \hat{b}_i \textit{ given } b_i \\
&= \int_{u_{-i}} f_K(u)\, du_{-i} \\
\bar{f}_K(x) &\equiv \frac{1}{K}\sum_{i=1}^{K} f_{iK}(x) \\
f_{ijK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) &\equiv \textit{the likelihood of } \hat{b}_i \textit{ and } \hat{b}_j \textit{ given } b_i \textit{ and } b_j \\
&= \int_{u_{-(i,j)}} f_K(u)\, du_{-(i.j)} \\
m_{iK}\left(\hat{b}_i\right) &\equiv \int_{-\infty}^{\infty} f_{iK}\left(\hat{b}_i - b_i\right) dG(b_i) \\
m'_{iK}\left(\hat{b}_i\right) &\equiv \frac{d}{d\hat{b}_i}\int_{-\infty}^{\infty} f_{iK}\left(\hat{b}_i - b_i\right) dG(b_i) \\
\bar{m}_K(x) &\equiv \frac{1}{K}\sum_{i=1}^{K} m_{iK}(x) \\
m_\phi\left(\hat{b}_1\right) &\equiv \int_{-\infty}^{\infty} \phi\left(\hat{b}_1 - b_1\right) dG(b_1)
\end{aligned}
$$

$$
\begin{aligned}
m^C_{ijK}\left(\hat{b}_i \mid \hat{b}_j\right) &\equiv \frac{m_{ijK}\left(\hat{b}_i, \hat{b}_j\right)}{m_{iK}\left(\hat{b}_j\right)} \\
&\equiv \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{ijK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) dG\left(b_i\right) dG\left(b_j\right)}{m_{iK}\left(\hat{b}_j\right)} \\
\bar{m}^C_{iK}\left(x \mid y\right) &\equiv \frac{1}{K-1}\sum_{\substack{j=1,\\ j\neq i}}^{K} m^C_{ijK}\left(x \mid y\right).
\end{aligned}
$$

Note that Assumption 4 implies the following limits:

$$
\begin{aligned}
s_K q_K &\to \infty \\
K s_K^8 &\to \infty \\
s_K^2 \log K &\to \infty \\
\frac{K^{-5/48}\log^2 K}{s_K^2} &\to 0 \\
\frac{K^{-5/96} h_K \log K}{s_K^2} &\to 0 \\
\frac{h_K}{s_K} &\to 0 \\
q_K^2 / K &\to 0.
\end{aligned}
\tag{A.1}
$$

The following lemmas are used in proving the main results. The first three lemmas collect Berry-Esseen-type results about convergence of certain densities and their derivatives to local limits.

**Lemma S-1 (Berry-Esseen Results for Densities)** *Under Assumption 1 and Assumption 6 through Assumption 8 we have the following local limit rate results: $\exists$ finite $C, K_0$ s.t. $\forall\ K \geq K_0$,*

$$
\sup_{i,s}\left|f_{iK}\left(s\right) - \phi\left(s\right)\right| \leq CK^{-\frac{1}{4}}\log K \tag{A.2}
$$

$$
\sup_{i,s}\left|f'_{iK}\left(s\right) - \phi'\left(s\right)\right| \leq CK^{-\frac{1}{8}}\log K \tag{A.3}
$$

*Proof of Lemma S-1*: Let $\eta_{it} = \frac{X_{it}\varepsilon_t}{\sigma_\varepsilon}$. Suppose that the sequence $\eta_{i1}, \eta_{i2}, \ldots$ satisfies Conditions A and C of Part B, where the constants in those conditions do not depend on $i$. Then, by Theorem 8 of Part B and a simple change of scale (recalling that $\sigma_\varepsilon^2$ is bounded away from both zero and infinity by Assumption 6),

$$
\begin{aligned}
\sup_{s}\left|f_{iK}\left(s\right) - \phi\left(s\right)\right| &\leq CK^{-\frac{1}{4}}\log K \qquad \text{(A.4)}\\
\sup_{s}\left|f'_{iK}\left(s\right) - \phi'\left(s\right)\right| &\leq CK^{-\frac{1}{8}}\log K
\end{aligned}
$$

for each $i$, where the constant $C$ does not depend on $i$. It follows that these inequalities hold uniformly in $i = 1, \ldots, K$, *i. e.*, it follows that (A.2) and (A.3) hold. To prove the lemma, it therefore suffices to prove that $\eta_{i1}, \eta_{i2}, \ldots$ satisfy Conditions A and C of Part B with constants that do not depend on $i$.

We first verify that Condition C is satisfied. By Assumption 6, $E[\eta_{it}] = 0$ and

$$\begin{aligned} E\left[\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\eta_{it}\right)^2\right] &= \frac{1}{T\sigma_\varepsilon^2}\sum_{s=1}^{T}\sum_{t=1}^{T}E\left[X_{it}X_{is}\varepsilon_t\varepsilon_s\right] \\ &= \frac{1}{T\sigma_\varepsilon^2}\sum_{t=1}^{T}E\left[X_{it}^2\varepsilon_t^2\right] \\ &= \frac{1}{T}E\left[\sum_{t=1}^{T}X_{it}^2\right] \\ &= 1 \end{aligned} \tag{A.5}$$

where the first equality is by definition, the second is by the m. d. s. property of the $\varepsilon$'s (Assumption 6(a)), the third is by the homoskedasticity of the $\varepsilon$'s (Assumption 6(c)) and the fourth equality is by the orthonormality of the $X$'s (Assumption 1).

Also, by Assumption 6(b),

$$\begin{aligned} \sup_i E\left[\eta_{it}^6\right] &= \sup_i E\left[X_{it}^6\varepsilon_t^6\right] \\ &\leq \sup_i \left(E\left[X_{it}^{12}\right]E\left[\varepsilon_t^{12}\right]\right)^{1/2} \\ &= \left(\sup_i E\left[X_{it}^{12}\right]E\left[\varepsilon_t^{12}\right]\right)^{1/2} \\ &\leq C \end{aligned} \tag{A.6}$$

so the moment conditions in Condition C hold uniformly in $i$. Assumption 7 implies that $\{X_{it}\varepsilon_t,\ t = 1, \ldots, T\}$ has maximal correlation coefficient $\nu'_{in}$ that satisfies $\nu'_{in} \leq De^{-\lambda n}$ for some positive finite constants $D$ and $\lambda$, since

$$\begin{aligned} \nu_n &\equiv \sup_m \sup_{f\in L^2\left(\mathcal{H}_1^m\right), g\in L^2\left(\mathcal{H}_{m+n}^\infty\right)} \left|\mathrm{Corr}\left(f,g\right)\right| \\ &\geq \sup_{f\in L^2\left(\mathcal{G}_{i,1}^m\right), g\in L^2\left(\mathcal{G}_{i,m+n}^\infty\right)} \left|\mathrm{Corr}\left(f,g\right)\right| \\ &\equiv \nu'_{in}, \end{aligned} \tag{A.7}$$

where $\mathcal{H}_a^b$ is the $\sigma$-field generated by the random variables

$$\left\{\left(\varepsilon_s, X_{1s}, \ldots, X_{Ks}\right), s = a, \ldots, b\right\}$$

and $\mathcal{G}_{i,a}^b$ is the $\sigma$-field generated by the random variables $\{X_{is}\varepsilon_s, s = a, \ldots, b\}$. The inequality holds because (as is evident from their respective definitions) $\mathcal{G}_{i,a}^b$ is a sub-$\sigma$-field of $\mathcal{H}_a^b$ for any positive integers $a < b$ and for any $i$. Thus, the first supremum is taken over a larger set, and is therefore (weakly) larger than the second supremum. Now, since $D$ and $\lambda$ obviously do not depend on $i$, the maximal correlation coefficient bound derived above is uniform in $i$. We now need to show that the maximal correlation coefficient bound we have obtained implies a strong mixing coefficient bound as required by Condition C of Part B. But if $\rho_n$ is a maximal correlation coefficient and $\alpha_n$ is the corresponding strong mixing coefficient, we have $\alpha_n \le \frac{1}{4}\rho_n$ (by simple calculation, or see display (1.3) of Bradley (1993)). Thus, the strong mixing coefficients $\alpha_{in}$ of the random variables $\{X_{it}\varepsilon_t, t = 1, \ldots, T\}$ are bounded (uniformly in $i$) by $\frac{1}{4}De^{-\lambda n}$. This verifies that Assumption 1, Assumption 6, and Assumption 7 imply Condition C of Part B.

It remains to show that Condition A of Part B is implied by Assumption 8 uniformly in $i$. Let

$$\psi_{it}^{X\varepsilon}(s) \equiv \int_{-\infty}^{\infty} e^{is\eta_{it}} p_{it}^{\eta}(\eta_{it} \mid \eta_{i,t-1}, \ldots, \eta_{i,1})\, d\eta_{it} \tag{A.8}$$

be the characteristic function of $\eta_{it} \equiv X_{it}\varepsilon_t$ (conditional on all past $\eta_{is}$), and define

$$\psi_{it}^{\varepsilon}(s) \equiv \int_{-\infty}^{\infty} e^{is\varepsilon_t} p_t^{\varepsilon}(\varepsilon_t \mid \eta_{i,t-1}, \ldots, \eta_{i1})\, d\varepsilon_t \tag{A.9}$$

to be the characteristic function of $\varepsilon_t$ (conditional on all past $\eta_{is}$ for some $i$).

By Assumption 8,

$$\sup_{it} \int_{-\infty}^{\infty} \left| \frac{d^2}{d\varepsilon_t^2} p_t^{\varepsilon}(\varepsilon_t \mid \eta_{i,t-1}, \ldots, \eta_{i1}) \right| d\varepsilon_t \le C < \infty$$

and

$$\begin{aligned}
& \sup_{it} p_{it}^{X}(x_{it} | \eta_{i,t-1}, \ldots, \eta_{i1}) \\
= & \sup_{it} \int \left\{ \begin{array}{c} p_{ijt}^{X}(x_{it} | \eta_{ij,t-1}, \ldots, \eta_{ij1}) \\ \times p_{ij}(\eta_{ij,t-1}, \ldots, \eta_{ij1} | \eta_{i,t-1}, \ldots, \eta_{i1}) \end{array} \right\} d\eta_{ij,t-1} \cdots d\eta_{ij1} \\
\le & \ C.
\end{aligned}$$

where $p_{it}^{X}(x_{it} | \eta_{i,t-1}, \ldots, \eta_{i1})$ is the density of $X_{it}$ given all past $\eta_{is}$, and the second display above is by Assumption 8 and the fact that the expectation of a function which is uniformly bounded by a constant is also bounded by that same constant.

We need to show that $\exists \alpha > 0, C_0 < \infty$, and $M_0 < \infty$ such that

$$\forall |s| \ge C_0, \ \sup_{it} \left| \psi_{it}^{X\varepsilon}(s) \right| \le M_0 |s|^{-\alpha}. \tag{A.10}$$

Set $C_0 = 1$. Note that Assumption 8 implies that

$$\begin{aligned}
\sup_{it} |\psi_{it}^{\varepsilon}(s)| &= \sup_{it} \left| \int_{-\infty}^{\infty} e^{is\varepsilon_t} p_t^{\varepsilon}(\varepsilon_t \mid \eta_{i,t-1}, \ldots, \eta_{i1})\, d\varepsilon_t \right| \qquad (A.11) \\
&= \sup_{it} \left| -\frac{1}{is} \int_{-\infty}^{\infty} e^{is\varepsilon_t} \left[ \frac{d}{d\varepsilon_t} p_t^{\varepsilon}(\varepsilon_t \mid \eta_{i,t-1}, \ldots, \eta_{i1}) \right] d\varepsilon_t \right| \\
&= \sup_{it} \left| -\frac{1}{s^2} \int_{-\infty}^{\infty} e^{is\varepsilon_t} \left[ \frac{d^2}{d\varepsilon_t^2} p_t^{\varepsilon}(\varepsilon_t \mid \eta_{i,t-1}, \ldots, \eta_{i1}) \right] d\varepsilon_t \right| \\
&\le |s|^{-2} \sup_{it} \int_{-\infty}^{\infty} \left| e^{is\varepsilon_t} \right| \left| \frac{d^2}{d\varepsilon_t^2} p_t^{\varepsilon}(\varepsilon_t \mid \eta_{i,t-1}, \ldots, \eta_{i1}) \right| d\varepsilon_t
\end{aligned}$$

$$
\begin{aligned}
&= |s|^{-2} \sup_{it} \int_{-\infty}^{\infty} \left| \frac{d^2}{d\varepsilon_t^2} p_t^{\varepsilon} \left( \varepsilon_t \mid \eta_{i,t-1}, \ldots, \eta_{i1} \right) \right| d\varepsilon_t \\
&\leq M_7 \left| s \right|^{-2}
\end{aligned}
$$

where the first equality is by definition, the second is by integration by parts, the third is by another integration by parts, the first inequality is obvious, the fourth equality is by $|e^{-is\varepsilon_t}| = 1$, and the final inequality is by Assumption 8 as noted above.

Let $p_{it}^X(x) \equiv p_{it}^X(x_{it} | \eta_{i,t-1}, \ldots, \eta_{i1})$. By Feller (1971, page 527) (or simply a short calculation), we have the first inequality in the following display, and the rest follow from (A.11) and Assumption 8: for $|s| \geq 1$,

$$
\begin{aligned}
\sup_{it} \left| \psi_{it}^{X\varepsilon}(s) \right| &\leq \sup_{it} \left| \int_{-\infty}^{\infty} \psi_{it}^{\varepsilon}(sx) \, p_{it}^X(x) \, dx \right| \qquad \text{(A.12)} \\
&\leq \sup_{it} \left| \int_{1}^{\infty} \psi_{it}^{\varepsilon}(sx) \, p_{it}^X(x) \, dx \right| + \sup_{it} \left| \int_{1/|s|}^{1} \psi_{it}^{\varepsilon}(sx) \, p_{it}^X(x) \, dx \right| \\
&\quad + \sup_{it} \left| \int_{-1/|s|}^{1/|s|} \psi_{it}^{\varepsilon}(sx) \, p_{it}^X(x) \, dx \right| \\
&\quad + \sup_{it} \left| \int_{-1}^{-1/|s|} \psi_{it}^{\varepsilon}(sx) \, p_{it}^X(x) \, dx \right| \\
&\quad + \sup_{it} \left| \int_{-\infty}^{-1} \psi_{it}^{\varepsilon}(sx) \, p_{it}^X(x) \, dx \right| \\
&\leq \int_{1}^{\infty} \frac{M_7}{|sx|^2} p_{it}^X(x) \, dx + \int_{1/|s|}^{1} \frac{M_6 M_7}{|sx|^2} dx + \int_{-1/|s|}^{1/|s|} M_6 dx \\
&\quad + \int_{-1}^{-1/|s|} \frac{M_6 M_7}{|sx|^2} dx + \int_{-\infty}^{-1} \frac{M_7}{|sx|^2} p_{it}^X(x) \, dx \\
&\leq \frac{2M_7}{|s|^2} + \frac{2M_6}{|s|} + \frac{2M_7 M_6}{|s|^2} + \frac{2M_7 M_6}{|s|}
\end{aligned}
$$

so $\alpha = 1$ and we are finished.

**Lemma S-2 (Berry-Esseen Results for Joint Densities)** *Under Assumption 1 and Assumption 6 through Assumption 8, and with $i \neq j$ always enforced, we have that $\exists$ finite $C, K_0$ s.t. $\forall$ $K \geq K_0$,*

$$
\begin{aligned}
\sup_{i,j,s,u} \left| f_{ijK}(s,u) - \phi(s)\phi(u) \right| &\leq CK^{-\frac{1}{6}} \log K \qquad \text{(A.13)} \\
\sup_{i,j,s,u} \left| \frac{\partial}{\partial s} f_{ijK}(s,u) - \phi'(s)\phi(u) \right| &\leq CK^{-\frac{1}{12}} \log K \qquad \text{(A.14)}
\end{aligned}
$$

*Proof of Lemma S-2*: Let $\eta_{ijt} = \frac{1}{\sigma_\varepsilon} \begin{pmatrix} X_{it}\varepsilon_t \\ X_{jt}\varepsilon_t \end{pmatrix}$. If the sequence of bivariate random variables $\eta_{ij1}, \eta_{ij2}, \ldots$ satisfies Conditions A and C in Part B uniformly in $(i,j)$, then Theorem 9 of Part B holds uniformly in $(i,j)$ and Lemma S-2 follows.

The argument that Assumption 1, Assumption 6, and Assumption 7 imply Condition C of Part B for $\eta_{ij1}, \eta_{ij2}, \ldots$ uniformly in $(i, j)$ parallels the corresponding argument in the proof of Lemma S-1 and is omitted.

It remains only to show that Assumption 8 implies Condition A of Part B uniformly in $(i, j)$. Let

$$\psi_{ijt}^{X\varepsilon}(s_1, s_2) \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\left(s_1\eta_{ijt}^{(1)}+s_2\eta_{ijt}^{(2)}\right)} p_{ijt}^{\eta}(\eta_{ijt} \mid \eta_{ij,t-1}, \ldots, \eta_{ij1})\, d\eta_{ijt}^{(1)} d\eta_{ijt}^{(2)} \tag{A.15}$$

be the characteristic function of $\eta_{ijt}$ conditional on $\eta_{ij,t-1}, \ldots, \eta_{ij1}$. We need to show that $\exists\alpha > 0, C_0 < \infty$, and $M_0 < \infty$ such that

$$\forall\, |s| \geq C_0,\ \sup_{ijt} \left|\psi_{ijt}^{X\varepsilon}(s)\right| \leq M_0 |s|^{-\alpha}. \tag{A.16}$$

First, note that

$$\begin{aligned} & \sup_{ijt} p_{ijt}^{X}(x_i, x_j | \, \eta_{ij,t-1}, \ldots, \eta_{ij1}) \qquad\qquad (A.17)\\ = \ & \sup_{ijt}\left\{p_{ijt}^{X}(x_i | \, x_j, \eta_{ij,t-1}, \ldots, \eta_{ij1})\, p_{ijt}^{X}(x_j | \, \eta_{ij,t-1}, \ldots, \eta_{ij1})\right\}\\ \leq \ & C^2, \end{aligned}$$

where the equality is by standard conditional probability and the inequality is by Assumption 8.

We cannot use the rather simple method of Lemma S-1 above to prove (A.16), because of the possibility that the vectors $(s_1, s_2)$ and $(x_i, x_j)$ might be orthogonal, causing $\frac{1}{(s_1x_i+s_2x_j)^2}$ to be undefined (infinite) even when both $|s|$ and $|x|$ are large.

To avoid this problem, define by $A$ the set of points in the plane such that the angle between the vectors $s$ and $x$ is within $\gamma$ of either $\frac{\pi}{2}$ or $\frac{3\pi}{2}$ (that is, the vectors are "$\gamma$-close" to being orthogonal), and the magnitude of the $x$ vector is between $\frac{1}{|s|}$ and $|s|^{1/4}$. Recall that $(s_1x_i + s_2x_j)^2 = |s|^2 |x|^2 \cos^2\theta_{s,x}$, where $\theta_{s,x}$ is the angle between $s$ and $x$. Also observe that $|\cos\gamma|$ dominates a sawtoothed function of $\gamma$ (draw a line from each zero of $|\cos\gamma|$ to the nearest maximum of $|\cos\gamma|$ to the left of the zero, and another line to the nearest maximum to the right of the zero, and you will have drawn the sawtoothed function). Thus, $\cos^{-2}\left(\frac{\pi}{2} \pm \gamma\right) \leq \frac{\pi^2}{4}\gamma^{-2}$ for $0 < \gamma \leq \frac{\pi}{2}$, and the same expression holds for $\cos^{-2}\left(\frac{3\pi}{2} \pm \gamma\right)$.

Set $C_0 = \left(\frac{\pi}{2}\right)^{-4/3}$, so we will demonstrate that the inequality in display (A.16) holds for $s$ such that $|s| \geq \left(\frac{\pi}{2}\right)^{-4/3}$, which implies that $|s|^{-3/4} \leq \frac{\pi}{2}$. Set $\gamma = |s|^{-3/4}$, so that $\gamma \leq \frac{\pi}{2}$ for $|s| \geq C_0$ and we may apply the cosine inequality developed in the preceding paragraph. Let $B_r$ denote a ball of radius $r$ centered at 0 in $\Re^2$, and let $\Re^2 - B_r$ denote $\Re^2$ excluding this ball. If $r_1 > r_2$, let $B_{r_1} - B_{r_2}$ denote $B_{r_1}$ excluding $B_{r_2}$. Let $p_{ijt}^{X}(x_i, x_j) \equiv p_{it}^{X}(x_{it}, x_{jt}| \, \eta_{ij,t-1}, \ldots, \eta_{ij1})$. By Feller (1971, page 527) (or simply a short calculation), we have the first inequality in the following display, where the rest follow from noting that $\psi_{ijt}^{\varepsilon}$ (the characteristic function of $\varepsilon_t$, conditional on $\eta_{ij,t-1}, \ldots, \eta_{ij1}$) could be subsituted for $\psi_{it}^{\varepsilon}$ in (A.11) with the identical algebraic-tail result (only the conditioning is different, and Assumption 8 applies to both conditioning situations), from (A.17), and from the definitions above.

$$
\begin{aligned}
& \sup_{ijt} \left| \psi_{ijt}^{X\varepsilon}(s_1, s_2) \right| \qquad \text{(A.18)} \\
\leq \;& \sup_{ijt} \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j)\, p_{ijt}^{X}(x_i, x_j)\, dx_i dx_j \right| \\
\leq \;& \sup_{ijt} \left| \int_{\Re^2 - B_{|s|^{1/4}}} \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j)\, p_{ijt}^{X}(x_i, x_j)\, dx \right| + \\
& \sup_{ijt} \left| \int_{B_{1/|s|}} \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j)\, p_{ijt}^{X}(x_i, x_j)\, dx \right| + \\
& \sup_{ijt} \left| \int_{B_{|s|^{1/4}} - B_{1/|s|}} \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j)\, p_{ijt}^{X}(x_i, x_j)\, dx \right| \\
\leq \;& \sup_{ijt} \int_{\Re^2 - B_{|s|^{1/4}}} \left| \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j) \right| p_{ijt}^{X}(x_i, x_j)\, dx + \\
& \sup_{ijt} \int_{B_{1/|s|}} \left| \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j) \right| p_{ijt}^{X}(x_i, x_j)\, dx + \\
& \sup_{ijt} \int_{B_{|s|^{1/4}} - B_{1/|s|}} \left| \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j) \right| p_{ijt}^{X}(x_i, x_j)\, dx \\
\leq \;& \sup_{ijt} \int_{\Re^2 - B_{|s|^{1/4}}} p_{ijt}^{X}(x_i, x_j)\, dx + \\
& \sup_{ijt} \int_{B_{1/|s|}} p_{ijt}^{X}(x_i, x_j)\, dx + \\
& \sup_{ijt} \int_{A} p_{ijt}^{X}(x_i, x_j)\, dx + \\
& \sup_{ijt} \int_{B_{|s|^{1/4}} - B_{1/|s|} - A} \left| \psi_{ijt}^{\varepsilon}(s_1 x_i + s_2 x_j) \right| p_{ijt}^{X}(x_i, x_j)\, dx \\
\leq \;& \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + 2\,|s|^{1/2}\,\gamma + \\
& \sup_{ijt} \int_{B_{|s|^{1/4}} - B_{1/|s|} - A} (s_1 x_i + s_2 x_j)^{-2}\, p_{ijt}^{X}(x_i, x_j)\, dx \\
= \;& \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + 2\,|s|^{1/2}\,\gamma + \\
& \sup_{ijt} \int_{B_{|s|^{1/4}} - B_{1/|s|} - A} |s|^{-2}\,|x|^{-2} \cos^{-2}\theta_{s,x}\, p_{ijt}^{X}(x_i, x_j)\, dx \\
\leq \;& \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + C_3\,|s|^{1/2}\,\gamma + \\
& \frac{\pi^2}{4}\,|s|^{-2}\,\gamma^{-2} \sup_{ijt} \int_{B_{|s|^{1/4}} - B_{1/|s|} - A} |x|^{-2}\, p_{ijt}^{X}(x_i, x_j)\, dx
\end{aligned}
$$

$$
\begin{aligned}
&\le \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + C_3 |s|^{1/2} \gamma + \\
&\quad \frac{\pi^2}{4} |s|^{-2} \gamma^{-2} C_4 \int_{B_{|s|^{1/4}}} \frac{1}{r} dr d\theta \\
&\le \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + C_3 |s|^{1/2} \gamma + \\
&\quad \frac{\pi^2}{4} |s|^{-2} \gamma^{-2} C_4 2\pi \frac{1}{4} \ln(|s|) \\
&\le \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + \frac{C_3}{|s|^{1/4}} + \frac{C_5}{|s|^{1/2}} \ln(|s|) \\
&\le \frac{C}{|s|^{1/2}} + \frac{C_2}{|s|} + \frac{C_3}{|s|^{1/4}} + \frac{C_5}{|s|^{1/3}}
\end{aligned}
$$

where the second-to-last inequality follows from our choice of $\gamma = |s|^{-3\,/\,4}$. Thus, Condition A is satisfied with $\alpha = \frac{1}{4}$, and we are finished.

**Lemma S-3 (Rates for Additional Densities)** *Under Assumption 1 and Assumption 6 through Assumption 8* $\exists\, C, K_0 < \infty$ *s.t.* $\forall\, K \ge K_0$,

$$
\begin{aligned}
\sup_s \left| \bar{f}_K(s) - \phi(s) \right| &\le CK^{-\frac{1}{4}} \log K && \text{(a)} \\
\sup_{i,\hat{b}_i} \left| m_{iK}\left(\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right| &\le CK^{-\frac{1}{4}} \log K && \text{(b)} \\
\sup_{\hat{b}_i} \left| \bar{m}_K\left(\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right| &\le CK^{-\frac{1}{4}} \log K && \text{(c)} \\
\sup_{i,\hat{b}_i} \left| m'_{iK}\left(\hat{b}_i\right) - m'_\phi\left(\hat{b}_i\right) \right| &\le CK^{-\frac{1}{8}} \log K && \text{(d)} \\
\sup_{\hat{b}_i} \left| \bar{m}'_K\left(\hat{b}_i\right) - m'_\phi\left(\hat{b}_i\right) \right| &\le CK^{-\frac{1}{8}} \log K && \text{(e)} \\
\sup_{i,j,\hat{b}_i,\hat{b}_j} \left| m_{ijK}\left(\hat{b}_i, \hat{b}_j\right) - m_\phi\left(\hat{b}_i\right) m_\phi\left(\hat{b}_j\right) \right| &\le CK^{-\frac{1}{6}} \log K && \text{(f)} \\
\sup_{i,j,\hat{b}_i,\hat{b}_j} \left| \frac{\partial}{\partial \hat{b}_i} m_{ijK}\left(\hat{b}_i, \hat{b}_j\right) - m'_\phi\left(\hat{b}_i\right) m_\phi\left(\hat{b}_j\right) \right| &\le CK^{-\frac{1}{12}} \log K && \text{(g)}
\end{aligned}
$$

*Proof of Lemma S-3*: Part (a) follows from Lemma S-1 and

$$
\begin{aligned}
&\sup_s \left| \bar{f}_K(s) - \phi(s) \right| && \text{(A.19)} \\
&= \sup_s \left| \frac{1}{K} \sum_{i=1}^{K} \left( f_{iK}(s) - \phi(s) \right) \right| \\
&\le \frac{1}{K} \sum_{i=1}^{K} \sup_s \left| f_{iK}(s) - \phi(s) \right| \\
&\le CK^{-1/4} \log K.
\end{aligned}
$$

Part (b) follows from Lemma S-1 and

$$\begin{aligned}
&\sup_{i,\hat{b}_i}\left|m_{iK}\left(\hat{b}_i\right)-m_{\phi}\left(\hat{b}_i\right)\right| \qquad (A.20)\\
=&\sup_{i,\hat{b}_i}\left|\int_{-\infty}^{\infty}\left[f_{iK}\left(\hat{b}_i-b_i\right)-\phi\left(\hat{b}_i-b_i\right)\right]dG\left(b_i\right)\right|\\
\le&\int_{-\infty}^{\infty}\sup_{i,\hat{b}_i}\left|f_{iK}\left(\hat{b}_i-b_i\right)-\phi\left(\hat{b}_i-b_i\right)\right|dG\left(b_i\right)\\
\le& CK^{-1/4}\log K
\end{aligned}$$

Part (c) follows from part (b) in exactly the same way that part (a) follows from Lemma S-1. Part (d) follows from Lemma S-1 and

$$\begin{aligned}
&\sup_{i,\hat{b}_i}\left|m'_{iK}\left(\hat{b}_i\right)-m'_{\phi}\left(\hat{b}_i\right)\right| \qquad (A.21)\\
=&\sup_{i,\hat{b}_i}\left|\frac{d}{d\hat{b}_i}\int_{-\infty}^{\infty}\left[f_{iK}\left(\hat{b}_i-b_i\right)-\phi\left(\hat{b}_i-b_i\right)\right]dG\left(b_i\right)\right|\\
=&\sup_{i,\hat{b}_i}\left|\int_{-\infty}^{\infty}\left[f'_{iK}\left(\hat{b}_i-b_i\right)-\phi'\left(\hat{b}_i-b_i\right)\right]dG\left(b_i\right)\right|\\
\le&\int_{-\infty}^{\infty}\sup_{i,\hat{b}_i}\left|f'_{iK}\left(\hat{b}_i-b_i\right)-\phi'\left(\hat{b}_i-b_i\right)\right|dG\left(b_i\right)\\
\le& CK^{-1/8}\log K
\end{aligned}$$

where we can interchange differentiation and integration because of the uniformly bounded derivatives of both of the likelihood functions (Theorem 8 of Part B implies that $\exists K_0<\infty$ such that $f'_{iK}$ is uniformly bounded for all $K\ge K_0$, under our assumptions). Part (e) follows from Part (d) just as Part (a) follows from Lemma S-1.

Part (f) follows from Lemma S-2 and

$$\begin{aligned}
&\sup_{i,\hat{b}_i,\hat{b}_j}\left|m_{ijK}\left(\hat{b}_i,\hat{b}_j\right)-m_{\phi}\left(\hat{b}_i,\hat{b}_j\right)\right| \qquad (A.22)\\
=&\sup_{i,\hat{b}_i,\hat{b}_j}\left|\int_{-\infty}^{\infty}\left[f_{ijK}\left(\hat{b}_i-b_i,\hat{b}_j-b_j\right)-\phi\left(\hat{b}_i-b_i\right)\phi\left(\hat{b}_j-b_j\right)\right]dG\left(b_i\right)dG\left(b_j\right)\right|\\
\le&\int_{-\infty}^{\infty}\sup_{i,\hat{b}_i,\hat{b}_j}\left|f_{ijK}\left(\hat{b}_i-b_i,\hat{b}_j-b_j\right)-\phi\left(\hat{b}_i-b_i\right)\phi\left(\hat{b}_j-b_j\right)\right|dG\left(b_i\right)dG\left(b_j\right)\\
\le& CK^{-1/6}\log K.
\end{aligned}$$

Part (g) follows from Lemma S-2 and

$$\sup_{i,\hat{b}_i,\hat{b}_j}\left|\frac{\partial}{\partial\hat{b}_j}m_{ijK}\left(\hat{b}_i,\hat{b}_j\right)-\frac{\partial}{\partial\hat{b}_j}m_{\phi}\left(\hat{b}_i,\hat{b}_j\right)\right| \qquad (A.23)$$

$$
\begin{aligned}
&= \sup_{i,\hat{b}_i,\hat{b}_j} \left| \frac{\partial}{\partial \hat{b}_i} \int_{-\infty}^{\infty} \left[ f_{iK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) - \phi\left(\hat{b}_i - b_i\right) \phi\left(\hat{b}_j - b_j\right) \right] dG\left(b_i\right) \right| \\
&= \sup_{i,\hat{b}_i,\hat{b}_j} \left| \int_{-\infty}^{\infty} \left[ \frac{\partial}{\partial \hat{b}_i} f_{iK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) - \phi'\left(\hat{b}_i - b_i\right) \phi\left(\hat{b}_j - b_j\right) \right] dG\left(b_i\right) \right| \\
&\leq \int_{-\infty}^{\infty} \sup_{i,\hat{b}_i,\hat{b}_j} \left| \frac{\partial}{\partial \hat{b}_i} f_{iK}\left(\hat{b}_i - b_i, \hat{b}_j - b_j\right) - \phi'\left(\hat{b}_i - b_i\right) \phi\left(\hat{b}_j - b_j\right) \right| dG\left(b_i\right) \\
&\leq CK^{-1/12} \log K
\end{aligned}
$$

where we can interchange differentiation and integration because of the uniformly bounded derivatives of both of the likelihood functions (Theorem 9 of Part B implies that $\exists K_0 < \infty$ such that $\frac{\partial}{\partial \hat{b}_i} f_{ijK}$ is uniformly bounded for all $K \geq K_0$, under our assumptions).

**Lemma S-4 (Information and Related Bounds)** *Assumption 1, Assumption 2, Assumption 4, and Assumption 6 through Assumption 8 imply that*

$$\sup_x \left|m_\phi(x)\right| < \infty \tag{A.24}$$

$$\sup_x \left|m'_\phi(x)\right| < \infty \tag{A.25}$$

$$\sup_x \left|\bar{m}_K(x)\right| < \infty \tag{A.26}$$

$$\sup_x \left|\bar{m}'_K(x)\right| < \infty \tag{A.27}$$

$$\int_{-\infty}^{\infty} \frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)} dx < \infty \tag{A.28}$$

$$\int_{-\infty}^{\infty} \left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K}\right)^2 \bar{m}_K(x)\, dx \to \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)} dx \tag{A.29}$$

$$\int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x) + s_K}\right)^2 m_\phi(x)\, dx \to \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)} dx \tag{A.30}$$

$$\int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x) + s_K}\right)^2 \bar{m}_K(x)\, dx \to \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)} dx \tag{A.31}$$

$$r_G\left(\hat{b}^{NB}, \phi_K\right) < \infty \tag{A.32}$$

$$r_G\left(\hat{b}^{INB}, \phi_K\right) < \infty \tag{A.33}$$

*Further, if in addition Assumption 3 holds,*

$$\sup_{x,\, \theta\in\Theta} \left|m'_\phi(x;\theta)\right| < \infty \tag{A.34}$$

$$\sup_{x,\, \theta\in\Theta} \left\| \frac{\partial \frac{\sigma_\varepsilon^2 m'(x;\theta,\sigma_e^2)}{m(x;\theta,\sigma_\varepsilon^2)+s_k}}{\partial\theta} \right\| \leq \sigma_\varepsilon^2 \frac{C}{s_k^2} \tag{A.35}$$

$$\sup_{x,\, \theta\in\Theta, \widehat{\sigma}_e^2\in[0.5\sigma_\varepsilon^2, 1.5\sigma_e^2]} \left\| \frac{\partial \frac{\widehat{\sigma}_\varepsilon^2 m'(x;\theta,\widehat{\sigma}_e^2)}{m(x;\theta,\widehat{\sigma}_\varepsilon^2)+s_k}}{\partial \widehat{\sigma}_e^2} \right\| \leq \frac{C}{s_k^2} \tag{A.36}$$

*Proof of Lemma S-4:* We shall prove the above lines in order, beginning with the inequality of expression (A.24).

$$\begin{aligned} \sup_x |m_\phi(x)| &= \sup_x \left| \int \phi(x-b)\, dG(b) \right| \\ &\leq \int \sup_x |\phi(x-b)|\, dG(b) \\ &= \int C_0 dG(b) \\ &= C_0, \end{aligned}$$

where the first equality is by definition, the inequality is due to the convexity of the sup function, the next equality uses $\sup_x |\phi(x-b)| = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}}$ (and employs the definition $C_0 \equiv \frac{1}{\sigma_\varepsilon\sqrt{2\pi}}$), and the final equality is trivial.

To see that the finite-derivative claim (A.25) holds use the following argument:

$$\begin{aligned} \sup_x \left|m'_\phi(x)\right| &= \sup_x \left| \frac{d}{dx} \int \phi(x-b)\, dG(b) \right| \\ &= \sup_x \left| \int \phi'(x-b)\, dG(b) \right| \\ &\leq \int \sup_x |\phi'(x-b)|\, dG(b) \\ &= \int C_0 dG(b) \\ &= C_0 \end{aligned} \tag{A.37}$$

where the first equality is by definition, the second equality is by the form of the normal likelihood, which satisfies the conditions of Dudley (1999, Corollary A.12 on page 394), the inequality is due to the convexity of the sup function, and the third equality follows by the direct calculation that $\sup_x |\phi'(x-b)| = \frac{1}{\sigma_\varepsilon^2\sqrt{2\pi}} e^{-1/2}\ \forall b$, and by defining $C_0 = \frac{1}{\sigma_\varepsilon^2\sqrt{2\pi}} e^{-1/2}$. The fourth equality is trivial, since $C_0$ is a constant with respect to $b$.

Results (A.26) and (A.27) follow from (A.24) and (A.25) together with parts (c) and (e) of Lemma S-3.

We now show that (A.28) holds.

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)} dx &= \int_{-\infty}^{\infty} \left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 m_\phi(x)\, dx \\ &= \int_{-\infty}^{\infty} \left(\frac{x - \hat{b}^{NB}(x)}{\sigma_\varepsilon^2}\right)^2 m_\phi(x)\, dx \\ &\leq \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx \end{aligned} \tag{A.38}$$

$$
\begin{aligned}
& + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \left( \hat{b}^{NB}(x) \right)^2 m_\phi(x)\, dx \\
= \quad & \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx \\
& + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \left( \frac{\int_{-\infty}^{\infty} b\phi(x-b)\, dG(b)}{\int_{-\infty}^{\infty} \phi(x-b)\, dG(b)} \right)^2 m_\phi(x)\, dx \\
\leq \quad & \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx \\
& + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \frac{\int_{-\infty}^{\infty} b^2\phi(x-b)\, dG(b)}{\int_{-\infty}^{\infty} \phi(x-b)\, dG(b)} m_\phi(x)\, dx \\
= \quad & \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx \\
& + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b^2\phi(x-b)\, dG(b)\, dx \\
= \quad & \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx \\
& + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} b^2 \left\{ \int_{-\infty}^{\infty} \phi(x-b)\, dx \right\} dG(b) \\
= \quad & \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} x^2 m_\phi(x)\, dx \\
& + \frac{2}{\sigma_\varepsilon^4} \int_{-\infty}^{\infty} b^2 dG(b) \\
< \quad & \infty,
\end{aligned}
$$

where the first equality follows from trivial manipulation, the second equality is due to the fact that $\hat{b}^{NB}(x) = x - \sigma_\varepsilon^2 \frac{m'_\phi(x)}{m_\phi(x)}$ due to the normal likelihood, the first inequality comes from the fact that $(a-b)^2 \leq 2a^2 + 2b^2$, the third equality is by definition, the second inequality is due to the convexity of the squaring function and an almost-sure-in-$x$ Jensen's inequality result for conditional expectations (since $\hat{b}^{NB}(x)$ is the conditional expectation of $b$ given $x$) (see Billingsley [1995, page 449, equation (34.7)]; note that the distribution for $x$ for which the result holds almost surely is precisely the distribution with density $m_\phi$, so the stated inequality holds), the fourth equality holds since $m_\phi(x) = \int_{-\infty}^{\infty} \phi(x-b)\, dG(b)$, the fifth equality follows from the Tonelli-Fubini theorem, the sixth equality holds since $\int_{-\infty}^{\infty} \phi(x-b)\, dx = 1\ \forall b \in \mathcal{R}$, and the final inequality is a direct result of the moment bounds given in Assumption 2 and Assumption 6.

Now we shall prove the convergence in (A.29).

$$
\int_{-\infty}^{\infty} \left( \frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 \bar{m}_K(x)\, dx \; - \; \int_{-\infty}^{\infty} \frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)} dx \tag{A.39}
$$

$$
= \int_{-\infty}^{\infty} \left( \frac{\bar{m}'_K(x)}{\bar{m}_K(x) + s_K} \right)^2 \left( \bar{m}_K(x) - m_\phi(x) \right) dx \tag{Term I}
$$

$$+\int_{-\infty}^{\infty}\left[\left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)+s_K}\right)^2-\left(\frac{m_\phi'(x)}{m_\phi(x)}\right)^2\right]m_\phi(x)\,dx \qquad \text{(Term II)}$$

follows by adding and subtracting $\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)+s_K}\right)^2 m_\phi(x)\,dx$. First we show that Term I converges to zero. Let $z_K\to\infty$ such that $s_K^{-2}z_K K^{-1/4}\log K\to 0$ and $s_K^{-2}z_K^{-2}\to 0$, which is certainly possible in light of Assumption 4. Then

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)+s_K}\right)^2(\bar{m}_K(x)-m_\phi(x))\,dx \qquad \text{(A.40)}\\
\leq\;& C_0 s_K^{-2}\int_{-\infty}^{\infty}(\bar{m}_K(x)-m_\phi(x))\,dx\\
=\;& C_0 s_K^{-2}\left\{\begin{array}{c}\int_{-z_K}^{z_K}(\bar{m}_K(x)-m_\phi(x))\,dx\\ +\int_{-\infty}^{-z_K}(\bar{m}_K(x)-m_\phi(x))\,dx\\ +\int_{z_K}^{\infty}(\bar{m}_K(x)-m_\phi(x))\,dx\end{array}\right\}\\
\leq\;& C_0 s_K^{-2}\left\{\begin{array}{c}2z_K C K^{-1/4}\log K\\ +\int_{-\infty}^{-z_K}(\bar{m}_K(x)-m_\phi(x))\,dx\\ +\int_{z_K}^{\infty}(\bar{m}_K(x)-m_\phi(x))\,dx\end{array}\right\}\\
\leq\;& C_0 s_K^{-2}\left\{\begin{array}{c}2z_K C K^{-1/4}\log K\\ +C_2 z_K^{-2}\end{array}\right\}\\
\to\;& 0
\end{aligned}$$

where the first inequality follows from the positivity of $\bar{m}_K(x)$ and the boundedness of $\bar{m}_K'(x)$ (from (A.27)), the first equality is trivial, the second inequality follows by using Lemma S-3(c) to get

$$\begin{aligned}
\int_{-z_K}^{z_K}(\bar{m}_K(x)-m_\phi(x))\,dx &\leq \int_{-z_K}^{z_K} C K^{-1/4}\log K\,dx\\
&= 2z_K C K^{-1/4}\log K,
\end{aligned}$$

the third inequality follows from Markov's inequality and the (uniformly in $K$) finite second moments (due to Assumption 2 and Assumption 6) of the distributions with densities $\bar{m}_K$ and $m_\phi$. Finally, the convergence to zero is by the construction of $z_K$.

Second we demonstrate that Term II converges to zero. Use the fact that $(a^2-b^2)=(a-b)^2+2b(a-b)$ to obtain

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left[\left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)+s_K}\right)^2-\left(\frac{m_\phi'(x)}{m_\phi(x)}\right)^2\right]m_\phi(x)\,dx \qquad \text{(A.41)}\\
=&\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)+s_K}-\frac{m_\phi'(x)}{m_\phi(x)}\right)^2 m_\phi(x)\,dx \qquad \text{(Term IIA)}\\
&+2\int_{-\infty}^{\infty}\left(\frac{m_\phi'(x)}{m_\phi(x)}\right)\left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)+s_K}-\frac{m_\phi'(x)}{m_\phi(x)}\right)m_\phi(x)\,dx. \qquad \text{(Term IIB)}
\end{aligned}$$

We deal with Term IIA, then Term IIB. Add and subtract $\frac{m'_\phi(x)}{m_\phi(x)+s_K}$ inside the square, then use the fact that $(a+b) \leq 2a^2+2b^2$, to obtain

$$\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 m_\phi(x)\,dx \tag{A.42}$$

$$\leq 2\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx \tag{Term IIAi}$$

$$+\ 2\int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 m_\phi(x)\,dx. \tag{Term IIAii}$$

Now Term IIAi can be shown to converge to zero by the following argument, which uses the fact that $\left(\frac{a}{b}-\frac{c}{d}\right)^2=\left(\frac{a}{b}\left(\frac{d-b}{d}\right)+\frac{a-c}{d}\right)^2 \leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2+2\left(\frac{a-c}{d}\right)^2$ to obtain the first inequality below:

$$\begin{aligned}
& 2\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx \qquad \text{(A.43)}\\
\leq\ & 4\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K}\right)^2\left(\frac{m_\phi(x)-\bar{m}_K(x)}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx\\
& +4\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K(x)-m'_\phi(x)}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx\\
\leq\ & 4\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K}\right)^2\left(\frac{CK^{-1/4}\log K}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx\\
& +4\int_{-\infty}^{\infty}\left(\frac{CK^{-1/8}\log K}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx\\
\leq\ & 4s_K^{-4}C_2K^{-1/2}\log K\ +\ 4s_K^{-2}C^2K^{-1/4}\log K\\
\to\ & 0
\end{aligned}$$

where the second inequality follows from Lemma S-3(c,e) and the third inequality follows from the nonnegativity of $\bar{m}_K(x)$ and the boundedness (by (A.27)) of $\bar{m}'_K(x)$. Finally, the convergence to zero is by rate bounds given in equation (A.1).

To see that Term IIAii converges to zero, note that $\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)}\right)^2$ clearly converges to zero pointwise (since $s_K \to 0$ by Assumption 4). Thus, if there is an integrable (with respect to the measure having density $m_\phi$) dominating function for $\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)}\right)^2$, then we may apply the Dominated Convergence Theorem and be finished. But since $(a-b)^2 \leq 2a^2+2b^2$, we have that

$$\begin{aligned}
& \left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 \qquad \text{(A.44)}\\
\leq\ & 2\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2+2\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2\\
\leq\ & 2\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2+2\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2
\end{aligned}$$

$$= \quad 4\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2$$

and we know from our above proof of the inequality (A.28) that

$$\int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 m_\phi(x)\,dx \;=\; \int_{-\infty}^{\infty}\frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)}dx \;<\; \infty.$$

Thus, Term IIAii converges to zero.

We must still deal with Term IIB:

$$\begin{aligned} & 2\int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)}\right)m_\phi(x)\,dx \qquad \text{(A.45)}\\ \leq\quad & 2\left(\int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 m_\phi(x)\,dx\right)^{1/2}\\ & \times\left(\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K(x)}{\bar{m}_K(x)+s_K}-\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 m_\phi(x)\,dx\right)^{1/2} \end{aligned}$$

and we see that the first factor is, by our above proof of (A.28), simply a finite constant, while the second factor is just the square root of Term IIA. But this means that Term IIB converges to zero, so we have demonstrated the convergence displayed in (A.29).

To show that the convergence (A.30) holds, simply note that $\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2 \rightarrow \left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2$ pointwise certainly holds, and further that $\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2 \leq \left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2$ and $\int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)}\right)^2 m_\phi(x)\,dx = \int_{-\infty}^{\infty}\frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)}dx < \infty$, so we may apply the Dominated Convergence Theorem to obtain the desired conclusion.

To demonstrate the convergence (A.31) simply add and subtract $\int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx$ to obtain

$$\begin{aligned} & \int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2\bar{m}_K(x)\,dx \;-\; \int_{-\infty}^{\infty}\frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)}dx \qquad \text{(A.46)}\\ =\quad & \int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2\left(\bar{m}_K(x)-m_\phi(x)\right)dx\\ & +\int_{-\infty}^{\infty}\left(\frac{m'_\phi(x)}{m_\phi(x)+s_K}\right)^2 m_\phi(x)\,dx \;-\; \int_{-\infty}^{\infty}\frac{\left(m'_\phi(x)\right)^2}{m_\phi(x)}dx\\ \rightarrow\quad & 0 \end{aligned}$$

where the convergence to zero follows because the first integral can be treated in exactly the same way that Term I above was, while the difference of the second and third integrals goes to zero by the convergence (A.30) shown immediately above.

The finiteness of the Bayes risk in the Gaussian problem, as claimed in (A.32), can be shown as follows:

$$r_G\left(\hat{b}^{NB},\phi_K\right) \qquad \text{(A.47)}$$

$$
\begin{aligned}
&= \rho \int \int \frac{1}{K} \sum_{i=1}^{K} \left( \hat{b}_i^{NB} \left( \hat{b}_i \right) - b_i \right)^2 \phi_K \left( \hat{b} - b \right) d\hat{b} dG(b) \\
&= \rho \frac{1}{K} \sum_{i=1}^{K} \int \int \left( \hat{b}_i^{NB} \left( \hat{b}_i \right) - b_i \right)^2 \phi \left( \hat{b}_i - b_i \right) d\hat{b}_i dG(b_i) \\
&= \rho \int \int \left( \hat{b}_1^{NB} \left( \hat{b}_1 \right) - b_1 \right)^2 \phi \left( \hat{b}_1 - b_1 \right) d\hat{b}_1 dG(b_1) \\
&\leq \rho \int \int b_1^2 \phi \left( \hat{b}_1 - b_1 \right) d\hat{b}_1 dG(b_1) \\
&= \rho \int b_1^2 \left\{ \int \phi \left( \hat{b}_1 - b_1 \right) d\hat{b}_1 \right\} dG(b_1) \\
&= \rho \int b_1^2 dG(b_1) \\
&< \infty
\end{aligned}
$$

where the first equality is by definition, the second equality is trivial, the third equality follows because each term in the sum of the expression on the second line is identical, and we may rename all of the variables we are integrating over $b_1$ and $\hat{b}_1$, then take the average over $K$ such identical terms, and the first inequality holds because $\hat{b}^{NB}$ is a Bayes decision rule, so it produces a Bayes risk no higher than that of any other decision rule; in particular, it performs no worse, in Bayes-risk terms, than the constant estimator 0. The fourth equality follows from the Tonelli-Fubini theorem. Finally, the second inequality holds by Assumption 2.

Result (A.33) follows from $|r_G \left( \hat{b}^{INB}, \phi_K \right) - r_G(\widehat{b}^{NB}, \phi_K)| \to 0$ as shown in the proof of Theorem 1(a).

To see that the finite-derivative claim (A.34) holds in the parametric case, use an argument analogous to that used in demonstrating (A.25):

$$
\begin{aligned}
\sup_{x,\, \theta \in \Theta} \left| m_\phi' (x; \theta) \right| &= \sup_{x,\, \theta \in \Theta} \left| \frac{d}{dx} \int \phi (x - b) \, dG(b; \theta) \right| \qquad \text{(A.48)} \\
&= \sup_{x,\, \theta \in \Theta} \left| \int \phi' (x - b) \, dG(b; \theta) \right| \\
&\leq \sup_{\theta \in \Theta} \int \sup_{x} \left| \phi' (x - b) \right| dG(b; \theta) \\
&= \sup_{\theta \in \Theta} \int C_0 dG(b; \theta) \\
&= C_0.
\end{aligned}
$$

A trivial application of the quotient rule, and a recognition of the fact that $m \geq 0$, shows that (A.35) follows from sup-bounds on the two derivatives $\frac{\partial m'\left(x;\sigma_\varepsilon^2,\theta\right)}{\partial \theta}$ and $\frac{\partial m\left(x;\sigma_\varepsilon^2,\theta\right)}{\partial \theta}$. We demonstrate the bound for $m$; the bound for $m'$ is obtained in an entirely similar fashion.

$$
\sup_{x,\theta} \left\| \frac{\partial m \left( x; \sigma_\varepsilon^2, \theta \right)}{\partial \theta} \right\| \qquad \text{(A.49)}
$$

$$
\begin{aligned}
&= \sup_{x,\theta} \left\| \int_{-\infty}^{\infty} \phi\left(x-y;\sigma_\varepsilon^2\right) \frac{\partial g(y,\theta)}{\partial \theta} dy \right\| \\
&\le \sup_{x,\theta} \int_{-\infty}^{\infty} \phi\left(y-x;\sigma_\varepsilon^2\right) \left\| \frac{\partial g(y,\theta)}{\partial \theta} \right\| dy \\
&= \sup_{x,\theta} \int_{-\infty}^{\infty} \phi\left(z;\sigma_\varepsilon^2\right) \left\| \frac{\partial g(z+x,\theta)}{\partial \theta} \right\| dz \\
&\le \int_{-\infty}^{\infty} \phi\left(z;\sigma_\varepsilon^2\right) \sup_{x,\theta} \left\| \frac{\partial g(z+x,\theta)}{\partial \theta} \right\| dz \\
&\le C \int_{-\infty}^{\infty} \phi\left(z;\sigma_\varepsilon^2\right) dz \\
&= C,
\end{aligned}
$$

where the first equality is by definition and the regularity of the integrand (so that the derivative may be passed under the integral, see Richard M. Dudley (*Uniform Central Limit Theorems*, 1999, Corollary A.10)), and the first inequality is by the convexity of the norm, having exploited the symmetry of the normal density function and applied Jensen's inequality after regarding the integral as the expectation of a function of a normally distributed random variable. The second equality is by a change of variables, the second inequality is by the convexity of the supremum function, and the third inequality is by Assumption 3(b). The final inequality is due to the fact that a normal density integrates to one. Note that $|\phi'|$ is symmetric, so exactly the same logic applies to $m'$.

Again, applying the quotient rule, and recalling that $m \geq 0$, result (A.36) is implied by sup-bounds for $m'$, $\frac{\partial m'}{\partial \hat{\sigma}_\varepsilon^2}$, and $\frac{\partial m}{\partial \hat{\sigma}_\varepsilon^2}$ over the above region of $\hat{\sigma}_\varepsilon^2$. But the restrictions on the region make this trivial: it is clear from inspection that the functions concerned are bounded as long as $\hat{\sigma}_\varepsilon^2$ is bounded away from zero or $\infty$, which it evidently is.

**Lemma S-5 (Lower Bounds for Densities)** *Under Assumption 1, Assumption 2, and Assumption 4 through Assumption 8* $\exists C_1, C_2, K_0 < \infty$ *s.t.* $\forall\, K \geq K_0$

$$\inf_{x \in [-d_K, d_K]} m_\phi(x) \geq C_1 K^{-1/32} \tag{a}$$

$$\inf_{i,\, x \in [-d_K, d_K]} m_{iK}(x) \geq C_2 K^{-1/32} \tag{b}$$

$$\sup_{x \in [-d_K, d_K]} \left\{ \frac{1}{m_\phi(x)} \right\} \leq \frac{1}{C_1} K^{1/32} \tag{c}$$

$$\sup_{i,\, x \in [-d_K, d_K]} \left\{ \frac{1}{m_{iK}(x)} \right\} \leq \frac{1}{C_2} K^{1/32} \tag{d}$$

*where* $d_K$ *is defined in Definition 1 above.*

*Proof of Lemma S-5*: We prove the parts of the lemma in order, from (a) to (d).

$$\inf_{x \in [-d_K, d_K]} m_\phi(x) = \inf_{x \in [-d_K, d_K]} \int_{-\infty}^{\infty} \phi(x-u)\, dG(u)$$

$$
\begin{aligned}
&\geq \int_{-\infty}^{\infty}\left\{\inf_{x \in [-d_K, d_K]} \phi(x-u)\right\} dG(u) \\
&= \int_{-\infty}^{\infty}\left\{\inf_{x \in [-d_K, d_K]} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2}(x-u)^2}\right\} dG(u) \\
&\geq \int_{-d_K}^{d_K}\left\{\inf_{x \in [-d_K, d_K]} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2}(x-u)^2}\right\} dG(u) \\
&\geq \int_{-d_K}^{d_K} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2}4d_K^2} dG(u) \\
&= \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-\frac{1}{2\sigma_\varepsilon^2}4d_K^2} \int_{-d_K}^{d_K} dG(u) \\
&\geq \frac{1}{D\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2}4d_K^2\right) \int_{-d_K}^{d_K} dG(u) \\
&\geq \frac{1}{D\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2}4d_K^2\right) \left[1 - \frac{C_3}{d_K^2}\right] \\
&\geq C_1 \exp\left(-\frac{1}{2\sigma_\varepsilon^2}4d_K^2\right) \\
&= C_1 \exp\left(-\frac{1}{2\sigma_\varepsilon^2}4\sigma_\varepsilon^2\frac{\log K}{64}\right) \\
&= C_1 K^{-1/32}
\end{aligned}
$$

where the first equality is by definition, the first inequality holds because the infimum function is concave, the second equality is again by definition, the second inequality is due to the fact that the integrand is nonnegative, the third inequality follows because the infimum of the normal density over $x$ and $u$ both in $[-d_K, d_K]$ can be no smaller than if $x$ and $u$ were $2d_K$ apart, the third equality follows because the integrand in the previous line is not a function of $u$, the fourth inequality follows from the fact that $\sigma_\varepsilon^2 \leq D < \infty$ by Assumption 6, the fifth inequality is by the existence of a second moment of $G$ (Assumption 2) and Markov's inequality, the sixth equality holds because, for $K$ beyond some $K_0$ large enough, we can simply note that $1 - \frac{C_3}{d_K^2}$ is greater than some positive constant, the fourth equality is by the definition of $d_K$ in Definition 1, and the last equality follows by simple calculation.

To see that the inequality of part (b) holds, note that $\inf_{i,\, x \in [-d_K, d_K]} m_{iK}(x) \geq \inf_{x \in [-d_K, d_K]} m_\phi(x) - CK^{-1/4}\log K \geq C_1 K^{-1/32} - CK^{-1/4}\log K$, where the first inequality comes from Lemma S-3(b) and the second comes from part (a) of this lemma, which we just proved. Now, for $K$ beyond some $K_0$ which is sufficiently large, we can simply absorb the $K^{-1/4}\log K$ term into the constant on the $K^{-1/32}$ term, because the latter goes to zero more slowly. Doing so, we obtain the inequality of part (b). To prove (c) and (d), simply invert the relations in (a) and (b).

**Lemma S-6 (Kernel MSE Rates)** *Under Assumption 1, Assumption 2, and Assumption 4 through Assumption 9* $\exists\, C, K_0 < \infty$ *s.t.* $\forall\, K \geq K_0$*,*

$$
\sup_{i, |\hat{b}_i| \leq d_K} E\left\{\left[\hat{m}_{iK}\left(\hat{b}_i\right) - \bar{m}_K\left(\hat{b}_i\right)\right]^2 \mid \hat{b}_i\right\} \qquad \text{(a)}
$$

$$\leq C\left(\begin{array}{c}\frac{1}{h_K(K-1)}+h_K^2+K^{-13/48}\log^2 K\\ +2K^{-13/96}h_K\log K\ +\ K^{-1/2}\log^2 K\end{array}\right)$$

$$\sup_{i,\left|\hat{b}_i\right|\leq d_K} E\left\{\left[\hat{m}'_{iK}\left(\hat{b}_i\right)-\bar{m}'_K\left(\hat{b}_i\right)\right]^2\mid\hat{b}_i\right\} \tag{b}$$

$$\leq C\left(\begin{array}{c}\frac{1}{h_K^3(K-1)}+h_K^2+K^{-5/48}\log^2 K\\ +2K^{-5/96}h_K\log K\ +\ K^{-1/4}\log^2 K\end{array}\right)$$

$$\sup_{i,\left|\hat{b}_i\right|\leq d_K} \Pr\left[\left|\breve{l}_i\left(\hat{b}_i\right)\right|>q_K\mid\hat{b}_i\right] \tag{c}$$

$$\to 0$$

*Proof of Lemma S-6*: If we have $\hat{b}_i\in[-d_K,d_K]$ then

$$\begin{aligned}
&E\left\{\left[\hat{m}_{iK}\left(\hat{b}_i\right)-\bar{m}_K\left(\hat{b}_i\right)\right]^2\mid\hat{b}_i\right\} && \text{(A.50)}\\
&= Var\left(\hat{m}_{iK}\left(\hat{b}_i\right)\mid\hat{b}_i\right)\\
&\quad+\left(E\left[\hat{m}_{iK}\left(\hat{b}_i\right)\mid\hat{b}_i\right]-\bar{m}_K\left(\hat{b}_i\right)\right)^2\\
&\leq Var\left(\hat{m}_{iK}\left(\hat{b}_i\right)\mid\hat{b}_i\right) && \text{(Term I)}\\
&\quad+2\left(E\left[\hat{m}_{iK}\left(\hat{b}_i\right)\mid\hat{b}_i\right]-m_\phi\left(\hat{b}_i\right)\right)^2 && \text{(Term II)}\\
&\quad+2\left(m_\phi\left(\hat{b}_i\right)-\bar{m}_K\left(\hat{b}_i\right)\right)^2. && \text{(Term III)}
\end{aligned}$$

First consider Term I. Now,

$$\begin{aligned}
&Var\left(\hat{m}_{iK}\left(\hat{b}_i\right)\mid\hat{b}_i\right) && \text{(A.51)}\\
=\ & Var\left(\frac{1}{h_K(K-1)}\sum_{j\neq i}w\left(\frac{\hat{b}_i-\hat{b}_j}{h_K}\right)\mid\hat{b}_i\right)\\
\leq\ & \frac{1}{h_K^2(K-1)^2}\sum_{j\neq i}\sum_{n\neq i}\left|Cov\left(w\left(\frac{\hat{b}_i-\hat{b}_j}{h_K}\right),w\left(\frac{\hat{b}_i-\hat{b}_n}{h_K}\right)\mid\hat{b}_i\right)\right|\\
\leq\ & \frac{1}{h_K^2(K-1)^2}\sum_{j\neq i}\sum_{n\neq i}\tau(|j-n|)\sqrt{E\left(w\left(\frac{\hat{b}_i-\hat{b}_j}{h_K}\right)^2\mid\hat{b}_i\right)}\times\\
&\sqrt{E\left(w\left(\frac{\hat{b}_i-\hat{b}_j}{h_K}\right)^2\mid\hat{b}_i\right)}\\
\leq\ & \frac{C}{h_K(K-1)}
\end{aligned}$$

where the first equality is by definition, the first inequality is familiar, and the second inequality holds by the following argument: first, note that the $\sigma$-fields generated by random variables which are linear combinations of the elements of the $T \times 1$ vectors $\underline{X}_i$ (where the weights in the linear combinations are fixed) are certainly sub-$\sigma$-fields of the $\sigma$-fields generated by the $\underline{X}_i$ vectors themselves (intuitively, we may lose information by taking linear combinations, but we will certainly never gain information by doing so). Thus, if we recall the notation of Assumption 9, and in addition define $\mathcal{G}_a^c$ as the $\sigma$-field generated by the random variables $\left\{\hat{b}_i = \frac{1}{\sqrt{T}}\underline{X}_i \cdot \varepsilon + b_i : a \le i \le c\right\}$, we see that for any $b$ and for any $\varepsilon$, $\mathcal{G}_a^c$ is a sub-$\sigma$-field of $\mathcal{F}_a^c$, so that for any $\varepsilon$ and for any $b$,

$$\begin{aligned}\tau(n) &\ge \sup_m \sup_{x \in \mathcal{F}_1^m, y \in \mathcal{F}_{m+n}^\infty} \left|Corr\left(x,\ y \mid \underline{X}_j, \varepsilon\right)\right| \\ &\ge \sup_m \sup_{x \in \mathcal{G}_1^m(\varepsilon), y \in \mathcal{G}_{m+n}^\infty(\varepsilon)} \left|Corr\left(x,\ y \mid \frac{1}{\sqrt{T}}\underline{X}_j \cdot \varepsilon + b_j\right)\right|\end{aligned}$$

but this means that the mixing coefficients $\tau(n)$ apply to the $\hat{b}_i$ as well as the $\underline{X}_i$, since we can certainly use the following bound (letting the $\sigma$-fields generated by the $\hat{b}_i$ be denoted $\mathcal{G}_a^c$ as before):

$$\begin{aligned}&\sup_m \sup_{x \in \mathcal{G}_1^m, y \in \mathcal{G}_{m+n}^\infty} \left|Corr\left(x,\ y \mid \hat{b}_j, \varepsilon\right)\right| \\ =\ &\sup_m \sup_{x \in \mathcal{G}_1^m, y \in \mathcal{G}_{m+n}^\infty} E\left[\left|Corr\left(x,\ y \mid \hat{b}_j, \varepsilon, b\right)\right|\right] \\ \le\ &\sup_{\varepsilon, b} \sup_m \sup_{x \in \mathcal{G}_1^m, y \in \mathcal{G}_{m+n}^\infty} \left|Corr\left(x,\ y \mid \hat{b}_j, \varepsilon, b\right)\right| \\ \le\ &\tau(n)\end{aligned}$$

which allows us to make use of Doukhan (1994, Theorem 3 (5) on page 9) to conclude that, if $u$ and $v$ are $\mathcal{G}_1^m$-measurable and $\mathcal{G}_{m+n}^\infty$-measurable random variables, respectively, and if $E[u^2],\ E[v^2] < \infty$, then

$$|Cov(u, v)| \le \tau(n)\sqrt{E[u^2]}\sqrt{E[v^2]}.$$

The final inequality in display (A.51) follows from a change of variables (so that the argument of the kernel is no longer scaled by $h_K$), the boundedness of the kernel, and the summability of the $\tau(n)$ (from Assumption 9).

Next, turn to Term II in display (A.50). Because $|a|\,|b-c| + |ab - c^2| \ge |c|\,|a-c|$,

$$\begin{aligned}&\sup_{i, x \in [-d_K, d_K], y \in \Re} m_\phi(x)\left|\bar{m}_{iK}^C(y|x) - m_\phi(y)\right| \\ =\ &\sup_{i, x \in [-d_K, d_K], y \in \Re} m_\phi(x)\left|\frac{1}{(K-1)}\sum_{j \ne i}\left(m_{ijK}^C(y|x) - m_\phi(y)\right)\right| \\ \le\ &\sup_{i, x \in [-d_K, d_K], y \in \Re} \frac{1}{(K-1)}\sum_{j \ne i} m_\phi(x)\left|m_{ijK}^C(y|x) - m_\phi(y)\right| \\ \le\ &\frac{1}{(K-1)}\sum_{j \ne i}\sup_{i, x \in [-d_K, d_K], y \in \Re} m_\phi(x)\left|m_{ijK}^C(y|x) - m_\phi(y)\right|\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{(K-1)}\sum_{j\neq i}\sup_{i,x\in[-d_K,d_K],y\in\Re}\left\{\begin{array}{c} m_{ijK}^{C}(y|x)\,|m_{iK}(x)-m_{\phi}(x)| \\ +|m_{ijK}(y,x)-m_{\phi}(y)\,m_{\phi}(x)| \end{array}\right\} \\
&\leq \frac{1}{(K-1)}\sum_{j\neq i}\left\{\begin{array}{c} \sup_{i,x\in[-d_K,d_K],y\in\Re} m_{ijK}^{C}(y|x)\,|m_{iK}(x)-m_{\phi}(x)| \\ +\ \sup_{i,x\in[-d_K,d_K],y\in\Re}|m_{ijK}(y,x)-m_{\phi}(y)\,m_{\phi}(x)| \end{array}\right\} \\
&\leq \left\{\frac{1}{(K-1)}\sum_{j\neq i}\sup_{i,x\in[-d_K,d_K],y\in\Re} m_{ijK}^{C}(y|x)\,|m_{iK}(x)-m_{\phi}(x)|\right\} \\
&\quad + CK^{-1/6}\log K \\
&\leq \left\{\frac{1}{(K-1)}\sum_{j\neq i}\left[\sup_{i,x\in[-d_K,d_K],y\in\Re} m_{ijK}^{C}(y|x)\right]\left[\sup_{i,x\in[-d_K,d_K],y\in\Re}|m_{iK}(x)-m_{\phi}(x)|\right]\right\} \\
&\quad + CK^{-1/6}\log K \\
&\leq \left\{\frac{1}{(K-1)}\sum_{j\neq i}\sup_{i,x\in[-d_K,d_K],y\in\Re} m_{ijK}^{C}(y|x)\,CK^{-1/4}\log K\right\} \\
&\quad + CK^{-1/6}\log K \\
&\leq \left\{\frac{1}{(K-1)}\sum_{j\neq i}C^{*}K^{1/32}CK^{-1/4}\log K\right\} \\
&\quad + CK^{-1/6}\log K \\
&= C^{*}K^{1/32}CK^{-1/4}\log K\ +\ CK^{-1/6}\log K \\
&\leq C_2^{*}K^{-1/6}\log K
\end{aligned}$$

where the first equality is by definition, the first inequality is by the convexity of the absolute value function, the second inequality is by the convexity of the sup function, the third inequality is by the fact stated immediately above the display, the fourth inequality is again by the convexity of the sup function, the fifth inequality is by Lemma S-3(f), the sixth inequality is again by the properties of the sup function (the sup of the product may never be greater than the product of the sups when the arguments are nonnegative), the seventh inequality follows from Lemma S-3(b), and the eighth inequality follows because $m_{ijK}^{C}(y|x) = \frac{m_{ijK}(y,x)}{m_{iK}(x)}$ by definition, the density $m_{\phi}$ is bounded (by (A.24)), and Lemma S-3(f) implies that then the numerator of the fraction is bounded. But the denominator is bounded by $\frac{1}{C_2}K^{1/32}$, as we know from Lemma S-5(d). The second equality follows since the summand does not depend on $j$, and the final inequality is due to the fact that $K^{(1/32)-(1/4)}\log K = K^{-7/32}\log K$ converges to zero more quickly than $K^{-1/6}\log K$, so the $K^{-7/32}\log K$ term may be absorbed into the constant on the $K^{-1/6}\log K$ term for sufficiently large $K$.

But from the above calculation, it follows that

$$\begin{aligned}
&\sup_{i,x\in[-d_K,d_K],y\in\Re}\left|\bar{m}_{iK}^{C}(y|x)-m_{\phi}(y)\right| \qquad \text{(A.52)} \\
&= \sup_{i,x\in[-d_K,d_K],y\in\Re}\frac{m_{\phi}(x)}{m_{\phi}(x)}\left|\bar{m}_{iK}^{C}(y|x)-m_{\phi}(y)\right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \left\{ \sup_{i,x\in[-d_K,d_K],y\in\Re} \frac{1}{m_\phi(x)} \right\} \\
&\quad \times \left\{ \sup_{i,x\in[-d_K,d_K],y\in\Re} m_\phi(x) \left| \bar{m}^C_{iK}(y|x) - m_\phi(y) \right| \right\} \\
&\leq \frac{1}{C_1} K^{1/32} C_2^* K^{-1/6} \log K \\
&= C_3^* K^{-13/96} \log K
\end{aligned}
$$

where the first equality follows by multiplication by one, the first inequality is due to the fact that the product of the sups is always at least as large as the sup of the products (for nonnegative arguments), the second inequality follows from the above calculation and Lemma S-5(c), and the second equality comes from a trivial computation. Now, the result immediately above implies that

$$
\begin{aligned}
&\sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| E\left[ \hat{m}_{iK}\left(\hat{b}_i\right) \mid \hat{b}_i \right] - m_\phi\left(\hat{b}_i\right) \right| \qquad \text{(A.53)} \\
&= \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \begin{array}{c} \frac{1}{(K-1)} \sum_{j\neq i} \int_{-\infty}^{\infty} w(z)\, m^C_{ijK}\left(\hat{b}_i - h_K z|\hat{b}_i\right) dz \\ -m_\phi\left(\hat{b}_i\right) \end{array} \right| \\
&= \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \int_{-\infty}^{\infty} w(z)\, \bar{m}^C_{iK}\left(\hat{b}_i - h_K z|\hat{b}_i\right) dz - m_\phi\left(\hat{b}_i\right) \right| \\
&= \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \int_{-\infty}^{\infty} w(z) \left[ \bar{m}^C_{iK}\left(\hat{b}_i - h_K z|\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right] dz \right| \\
&\leq \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \int_{-\infty}^{\infty} w(z) \left| \bar{m}^C_{iK}\left(\hat{b}_i - h_K z|\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right| dz \\
&\leq \int_{-\infty}^{\infty} w(z) \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \bar{m}^C_{iK}\left(\hat{b}_i - h_K z|\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right) \right| dz \\
&\leq \int_{-\infty}^{\infty} w(z) \sup_{i,\,\hat{b}_i\in[-d_K,d_K]} \left| \bar{m}^C_{iK}\left(\hat{b}_i - h_K z|\hat{b}_i\right) - m_\phi\left(\hat{b}_i - h_K z\right) \right| dz \\
&\quad + \int_{-\infty}^{\infty} w(z) \sup_{\hat{b}_i\in[-d_K,d_K]} \left| m_\phi\left(\hat{b}_i - h_K z\right) - m_\phi\left(\hat{b}_i\right) \right| dz \\
&\leq \int_{-\infty}^{\infty} w(z) \sup_{i,x\in[-d_K,d_K],y\in\Re} \left| \bar{m}^C_{iK}(y|x) - m_\phi(y) \right| dz \\
&\quad + \int_{-\infty}^{\infty} w(z)\, C_4^* h_K z dz \\
&\leq \int_{-\infty}^{\infty} w(z)\, C_3^* K^{-13/96} \log K dz \; + \; C_4^* h_K \int_{-\infty}^{\infty} z w(z)\, dz \\
&\leq C_3^* K^{-13/96} \log K \; + \; C_5^* h_K
\end{aligned}
$$

where the first equality is by definition, the second equality follows from the linearity of integration, the third equality is due to the fact that $m_\phi(x) = \int_{-\infty}^{\infty} m_\phi(x)\, w(z)\, dz$

as long as $w$ is a probability density, the first inequality is by the convexity of the absolute value function, the second inequality is by the convexity of the sup function, the third inequality follows from adding and subtracting $m_\phi\left(\hat{b}_i - h_K z\right)$ and then applying the triangle inequality (and the convexity of the sup function again), the fourth inequality follows because $m_\phi$ is uniformly Lipschitz continuous (since $\phi$ has a uniformly bounded first derivative, and we can safely interchange the derivative and the integral), the fifth inequality follows from the calculation immediately above the current one, and the sixth inequality is due to the fact that the kernel $w$ has a bounded second, and thus first, moment (as well as the fact that it integrates to one, in the case of the first of the two terms).

Finally, consider Term III in display (A.50) – we know that

$$\sup_{i,\hat{b}_i}\left(\bar{m}_K\left(\hat{b}_i\right) - m_\phi\left(\hat{b}_i\right)\right)^2 \le CK^{-\frac{1}{2}}\log^2 K \tag{A.54}$$

from Lemma S-3(c) above.

Substituting the bounds (A.51), (A.53), and (A.54) into display (A.50) yields part (a) of this lemma.

In the case of part (b) of this lemma, concerning the derivative of the density, everything is entirely similar, except that the rates slow somewhat, as we use the rates from Lemma S-3 which pertain to the derivatives, rather than those which pertain to the densities themselves.

We now have only to demonstrate part (c) of the lemma.

$$\begin{aligned}
& \sup_{i,\left|\hat{b}_i\right|\le d_K} \Pr\left[\left|\breve{l}_i\left(\hat{b}_i\right)\right| > q_K \mid \hat{b}_i\right] \qquad \text{(A.55)} \\
\le\; & \sup_{i,\left|\hat{b}_i\right|\le d_K} \Pr\left[\left|\hat{m}'_{iK}\left(\hat{b}_i\right)\right| > s_K q_K \mid \hat{b}_i\right] \\
\le\; & \sup_{i,\left|\hat{b}_i\right|\le d_K} \left(\frac{1}{s_K q_K}\right)^2 E\left[\left(\hat{m}'_{iK}\left(\hat{b}_i\right)\right)^2 \mid \hat{b}_i\right] \\
\le\; & C_0\left(\frac{1}{s_K q_K}\right)^2 \\
\to\; & 0
\end{aligned}$$

where the first inequality follows from the definition of $\breve{l} = \frac{\hat{m}'}{\hat{m}+s_K}$ and the nonnegativity of $\hat{m}$, the second inequality is a direct application of Markov's inequality, and the third inequality follows from part (b) of this lemma and the uniform boundedness of $\bar{m}_K$ (from (A.26)). Finally, the convergence to zero is by Assumption 4 and the rate results in display (A.1). Q.E.D.

## A.2 Proof of Lemma 1

Lemma 1 is Lemma S-4 above, for the Gaussian model. The additional Assumption 6 through Assumption 8 are not needed in this case because $\widehat{b} \sim N(0, \sigma_\varepsilon^2 I_K)$.

## A.3 Proof of Lemma 2

Lemma 2 is Lemma S-6 above, for the Gaussian model. The additional Assumption 6 through Assumption 9 are not needed in this case because $\widehat{b} \sim N(0, \sigma_\varepsilon^2 I_K)$.

## A.4 Proof of Theorem 2

The proof of Theorem 2(b) (given in the text) required the result

$$\lim_{K\to\infty} \sup_{\|b\|_2 \le M} \left| R(b, \widehat{b}^{NSEB}) - R(b, \widehat{b}^{NB}_{\widehat{G}_K}) \right| = 0.$$

This is shown in the proof of Theorem 6 below, under a more general set of conditions; see the limit in display (A.122). Note that the limit in (A.122) is proven without reference to Theorem 2(b), though Theorem 2(b) is used later in the proof of Theorem 6, so our logic is not circular.

## A.5 Proof of Theorem 4

Before we begin, note that Assumption 7 implies the following summability inequalities (due to the exponentially decaying upper bound on the $\nu_n$)

$$\begin{aligned} \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \nu_n &\le D < \infty \\ \sum_{n=1}^{\infty} n\nu_n &\le D < \infty \end{aligned} \tag{A.56}$$

which evidently yield (since $\nu_n$ is nonnegative by definition)

$$\sum_{n=1}^{\infty} \nu_n \le D < \infty. \tag{A.57}$$

*Proof of Part (a)*: First, we show unbiasedness:

$$\begin{aligned} E\left[\hat{\sigma}_\varepsilon^2\right] &= \frac{1}{T-K} E\left[\varepsilon'\varepsilon\right] - \frac{1}{T(T-K)} E\left[\varepsilon' X X' \varepsilon\right] \\ &= \frac{T}{T-K}\sigma_\varepsilon^2 - \frac{1}{T(T-K)} \sum_{i=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{is} X_{it} \varepsilon_s \varepsilon_t\right] \\ &= \frac{T}{T-K}\sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{T(T-K)} \sum_{i=1}^{K} E\left[\sum_{t=1}^{T} X_{it}^2\right] \\ &= \frac{T}{T-K}\sigma_\varepsilon^2 - \frac{KT\sigma_\varepsilon^2}{T(T-K)} \\ &= \sigma_\varepsilon^2 \end{aligned} \tag{A.58}$$

where the first equality follows easily from the definition of $\hat{\sigma}_\varepsilon^2$, the second is by Assumption 6 and calculation, for the first and second terms respectively, and the third equality is due to the fact that $E[X_{is}X_{it}\varepsilon_s\varepsilon_t] = 0$ if $s \neq t$, and $= \sigma_\varepsilon^2 E[X_{it}^2]$ if $s = t$. The fourth equality follows from the fact that $\sum_{t=1}^T X_{it}^2 = T$ and from trivial summation. The final equality follows by simple cancellation. Note that the unbiasedness does *not* depend on Assumption 7, although Assumption 7 is crucial later in the proof.

Before entering into the proof of the squared-error bound, it is worth noting that, for any random variables $Z_1, \ldots, Z_m$ each of which has an $n^{th}$ absolute moment, and if $p_i \geq 0$, $\sum_{i=1}^m p_i = n$, then, by iterating Hölder's inequality,

$$\left|E\left[\Pi_{i=1}^m Z_i^{p_i}\right]\right| \leq \Pi_{i=1}^m \left(E\left[|Z_i|^n\right]\right)^{p_i\,/\,n}. \tag{A.59}$$

This fact will prove useful when we apply Assumption 6 to various cross-moments below, and we will not explicitly cite it when it is applied, in the interest of brevity.

Now, using unbiasedness,

$$\begin{aligned}
& E\left[\left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right)^2\right] \qquad\qquad (A.60) \\
= \; & E\left[\left(\hat{\sigma}_\varepsilon^2\right)^2\right] - \sigma_\varepsilon^4 \\
= \; & E\left[\frac{1}{(T-K)^2}\left(\varepsilon'\left(I - \frac{XX'}{T}\right)\varepsilon\right)\left(\varepsilon'\left(I - \frac{XX'}{T}\right)\varepsilon\right)\right] - \sigma_\varepsilon^4 \\
= \; & E\left[\frac{1}{(T-K)^2}\left[\begin{array}{c}(\varepsilon'\varepsilon)^2 - 2(\varepsilon'\varepsilon)\left(\varepsilon'\frac{XX'}{T}\varepsilon\right) + \\ \left(\varepsilon'\frac{XX'}{T}\varepsilon\right)\left(\varepsilon'\frac{XX'}{T}\varepsilon\right)\end{array}\right]\right] - \sigma_\varepsilon^4 \\
= \; & \left\{\begin{array}{c}
E\left[\frac{1}{(T-K)^2}\sum_{s=1}^T\sum_{t=1}^T \varepsilon_s^2\varepsilon_t^2\right] \\
+E\left[\frac{-2}{T(T-K)^2}\sum_{t=1}^T\sum_{i=1}^K\sum_{s=1}^T\sum_{u=1}^T \varepsilon_t^2 X_{is}X_{iu}\varepsilon_s\varepsilon_u\right] \\
+E\left[\begin{array}{c}\frac{1}{T^2(T-K)^2}\sum_{i=1}^K\sum_{j=1}^K\sum_{s=1}^T\sum_{t=1}^T\sum_{u=1}^T\sum_{w=1}^T \\ X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w\end{array}\right]
\end{array}\right\} \\
& -\sigma_\varepsilon^4
\end{aligned}$$

and we shall address each of the three expectations in the final expression in turn. The first expectation can be written as

$$\begin{aligned}
E\left[\frac{1}{(T-K)^2}\sum_{s=1}^T\sum_{t=1}^T \varepsilon_s^2\varepsilon_t^2\right] &= \frac{1}{(T-K)^2}\sum_{s=1}^T\sum_{t=1}^T E\left[\varepsilon_s^2\varepsilon_t^2\right] \qquad (A.61) \\
&= \frac{1}{(T-K)^2}\sum_{s=1}^T\left[(T-1)\,\sigma_\varepsilon^4 + E\left[\varepsilon_s^4\right]\right] \\
&= \frac{T(T-1)}{(T-K)^2}\sigma_\varepsilon^4 + \frac{\sum_{s=1}^T E\left[\varepsilon_s^4\right]}{(T-K)^2}.
\end{aligned}$$

so that

$$\left|E\left[\frac{1}{(T-K)^2}\sum_{s=1}^T\sum_{t=1}^T \varepsilon_s^2\varepsilon_t^2\right] - \frac{T(T-1)}{(T-K)^2}\sigma_\varepsilon^4\right| \tag{A.62}$$

$$
\begin{aligned}
&\leq \frac{\sum_{s=1}^{T} E\left[\varepsilon_s^4\right]}{(T-K)^2} \\
&\leq \frac{DT}{(T-K)^2} \\
&\leq \frac{C_1}{K}
\end{aligned}
$$

where the first inequality is clear, the second follows from Assumption 6, and the third is by the asymptotic nesting which has been assumed, in which $K$ is an asymptotically constant fraction of $T$.

The second expectation may be evaluated as follows:

$$
\begin{aligned}
& E\left[\frac{-2}{T(T-K)^2}\sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{s=1}^{T}\sum_{u=1}^{T}\varepsilon_t^2 X_{is}X_{iu}\varepsilon_s\varepsilon_u\right] \qquad \text{(A.63)} \\
= & \frac{-2}{T(T-K)^2}\sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{s=1}^{T}\sum_{u=1}^{T} E\left[\varepsilon_t^2 X_{is}X_{iu}\varepsilon_s\varepsilon_u\right]
\end{aligned}
$$

Now we must break out six cases:

Case 1: $s = t = u$.

This contributes, in *absolute value*,

$$
\begin{aligned}
& \frac{2}{T(T-K)^2}\sum_{i=1}^{K}\sum_{t=1}^{T} E\left[X_{it}^2\varepsilon_t^4\right] \qquad \text{(A.64)} \\
\leq & \frac{2D}{T(T-K)^2}\sum_{i=1}^{K}\sum_{t=1}^{T} E\left[X_{it}^2\right] \\
= & \frac{2D}{T(T-K)^2}\sum_{i=1}^{K} E\left[\sum_{t=1}^{T} X_{it}^2\right] \\
= & \frac{2KTD}{T(T-K)^2} \\
\leq & \frac{C_2}{K}
\end{aligned}
$$

where the first inequality is by Assumption 6, the first equality is by the linearity of expectations, the second equality is by $\sum_{t=1}^{T} X_{it}^2 = T$, and the second inequality is by the asymptotic nesting, in which $K$ is asymptotically a constant fraction of $T$.

Case 2: All time subscripts are distinct.

Then $E\left[X_{is}X_{iu}\varepsilon_s\varepsilon_u\varepsilon_t^2\right] = 0$ by using the m. d. s. property of $\varepsilon$, or, if $t$ is the greatest subscript, by the homoskedastic m. d. s. property of $\varepsilon$ followed by the m. d. s. property.

Case 3: $s = t > u$ or $u = t > s$.

Suppose w. l. o. g. that $s = t > u$; then

$$\begin{aligned} & \left|E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right]\right| && \text{(A.65)} \\ = & \left|E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right] - E\left[X_{it}\varepsilon_t^3\right] E\left[X_{iu}\varepsilon_u\right]\right| \\ \leq & \nu_{t-u}\left(E\left[X_{it}^2\varepsilon_t^6\right]\right)^{1/2}\left(E\left[X_{iu}^2\varepsilon_u^2\right]\right)^{1/2} \\ \leq & M_1 \nu_{t-u} && \text{(A.66)} \end{aligned}$$

where the equality follows from the fact that $E\left[X_{iu}\varepsilon_u\right] = 0$, and the first inequality follows from Doukhan (1994, Theorem 3 (5) on page 9), which states that, if $r$ and $z$ are $\mathcal{H}_1^m$-measurable and $\mathcal{H}_{m+n}^\infty$-measurable random variables, respectively, and if $E\left[r^2\right],\ E\left[z^2\right] < \infty$, then

$$|Cov\,(r,z)| \leq \nu_n \sqrt{E\left[r^2\right]}\sqrt{E\left[z^2\right]}.$$

Note in particular that the moment bounds hold by Assumption 6, and that the $\sigma$-field $\mathcal{H}_a^b$ is generated by the random variables $\{X_a, \varepsilon_a, \ldots, X_b, \varepsilon_b\}$. The second inequality follows from the uniform (over $i, u$, and $t$) bounds on the moments guaranteed by Assumption 6. Thus, the absolute value of the contribution of these terms to the expectation is

$$\begin{aligned} & \frac{2}{T\,(T-K)^2}\left|\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T} E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right]\right| && \text{(A.67)} \\ \leq & \frac{2}{T\,(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T}\left|E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right]\right| \\ \leq & \frac{2M_1}{T\,(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T}\nu_{t-u} \\ = & \frac{2M_1}{T\,(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{n=1}^{T-u}\nu_n \\ \leq & \frac{2M_1}{T\,(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T}\sum_{n=1}^{\infty}\nu_n \\ \leq & \frac{2M_1 C^*}{T\,(T-K)^2}\sum_{i=1}^{K}\sum_{u=1}^{T} 1 \\ = & \frac{2M_1 C^* K T}{T\,(T-K)^2} \\ \leq & \frac{C_3}{K} \end{aligned}$$

where the first inequality is by the triangle inequality, the second inequality is by the preceding display, the first equality is by setting $n = t - u$, the third inequality is due to the fact that $\nu_n \geq 0$ by definition, the fourth inequality is by expression (A.57), and second equality is by trivial summation, and the final inequality is due to the asymptotic nesting, in which $K$ is asymptotically a constant fraction of $T$.

Case 4: $s = t < u$.

Here $E\left[X_{it}\varepsilon_t^3 X_{iu}\varepsilon_u\right] = 0$ by the m. d. s. property of $\varepsilon$.

Case 5: $s = u > t$.

Here $E\left[X_{is}^2\varepsilon_s^2\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{is}^2\varepsilon_t^2\right]$ by the homoskedastic m. d. s. property of $\varepsilon$, and

$$\begin{aligned} & \left|E\left[X_{is}^2\varepsilon_t^2\right] - E\left[X_{is}^2\right]E\left[\varepsilon_t^2\right]\right| \\ \leq\; & \nu_{s-t}\left(E\left[X_{is}^4\right]\right)^{1/2}\left(E\left[\varepsilon_t^4\right]\right)^{1/2} \\ \leq\; & M_2\nu_{s-t} \end{aligned} \tag{A.68}$$

where the first inequality is that of Doukhan (1994, Theorem 3 (5) on page 9) used above (noting that Assumption 6 guarantees moment existence), while the second follows from Assumption 6's uniform (over $i$, $s$, and $t$) bound on the moments. But $E\left[X_{is}^2\right]E\left[\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{is}^2\right]$. Thus

$$\begin{aligned} & \left|\begin{array}{l} \frac{-2}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T E\left[X_{is}^2\varepsilon_s^2\varepsilon_t^2\right] \\ -\left(\frac{-2\sigma_\varepsilon^4}{T(T-K)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T E\left[X_{is}^2\right]\right)\end{array}\right| \\ \leq\; & \frac{2}{T\left(T-K\right)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T\left|E\left[X_{is}^2\varepsilon_s^2\varepsilon_t^2\right] - \sigma_\varepsilon^4 E\left[X_{is}^2\right]\right| \\ \leq\; & \frac{2M_2}{T\left(T-K\right)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=t+1}^T\nu_{s-t} \\ \leq\; & \frac{2M_2}{T\left(T-K\right)^2}\sum_{i=1}^K\sum_{t=1}^T\sum_{n=1}^\infty\nu_n \\ \leq\; & \frac{2M_2C^*}{T\left(T-K\right)^2}\sum_{i=1}^K\sum_{t=1}^T 1 \\ =\; & \frac{2M_2C^*KT}{T\left(T-K\right)^2} \\ \leq\; & \frac{C_4}{K}. \end{aligned} \tag{A.69}$$

Case 6: $t > s = u$.

Here $E\left[X_{is}^2\varepsilon_s^2\varepsilon_t^2\right] = \sigma_\varepsilon^4 E\left[X_{is}^2\right]$ by the homoskedastic m. d. s. property of $\varepsilon$ as given in Assumption 6. Thus the overall contribution of these terms is:

$$\frac{-2\sigma_\varepsilon^4}{T\left(T-K\right)}\sum_{i=1}^K\sum_{t=1}^T\sum_{s=1}^{t-1}E\left[X_{is}^2\right]. \tag{A.70}$$

Now that we have broken out the cases, we may pull them back together again:

$$\left| \begin{array}{c} E\left[\frac{-2}{T(T-K)^2}\sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{s=1}^{T}\sum_{u=1}^{T}\varepsilon_t^2 X_{is}X_{iu}\varepsilon_s\varepsilon_u\right] \\ -\left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right) \end{array} \right| \tag{A.71}$$

$$\begin{aligned}
&\leq \frac{C_1+C_2+C_3+C_4}{K} + \left| \begin{array}{c} \frac{-2\sigma_\varepsilon^4}{T(T-K)^2}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=t+1}^{T}E\left[X_{is}^2\right] \\ +\frac{-2\sigma_\varepsilon^4}{T(T-K)}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left[X_{is}^2\right] \\ -\left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right) \end{array} \right| \\
&\leq \frac{C_1+C_2+C_3+C_4}{K} + \frac{2\sigma_\varepsilon^4}{T(T-K)^2}\sum_{i=1}^{K}\sum_{t=1}^{T}E\left[X_{it}^2\right] + \\
&\left|\frac{-2\sigma_\varepsilon^4}{T(T-K)^2}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{is}^2\right] - \left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right)\right| \\
&= \frac{C_1+C_2+C_3+C_4}{K} + \frac{2\sigma_\varepsilon^4}{T(T-K)^2}\sum_{i=1}^{K}E\left[\sum_{t=1}^{T}X_{it}^2\right] + \\
&\left|\frac{-2\sigma_\varepsilon^4}{T(T-K)^2}\sum_{i=1}^{K}\sum_{t=1}^{T}E\left[\sum_{s=1}^{T}X_{is}^2\right] - \left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right)\right| \\
&= \frac{C_1+C_2+C_3+C_4}{K} + \frac{2\sigma_\varepsilon^4 K}{(T-K)^2} + \\
&\left|\frac{-2\sigma_\varepsilon^4 KT^2}{T(T-K)^2} - \left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right)\right| \\
&= \frac{C_1+C_2+C_3+C_4}{K} + \frac{2\sigma_\varepsilon^4 K}{(T-K)^2} \\
&\leq \frac{C_5}{K}
\end{aligned}$$

Finally, the third expectation is

$$E\left[ \begin{array}{c} \frac{1}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{u=1}^{T}\sum_{w=1}^{T} \\ X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w \end{array} \right] \tag{A.72}$$

$$\begin{aligned} &= \frac{1}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{u=1}^{T}\sum_{w=1}^{T} \\ &\quad E\left[X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w\right] \end{aligned}$$

for which we will need to consider eleven different cases.

Case 1: $s = t = u = w$.

$E\left[X_{is}^2 X_{js}^2 \varepsilon_s^4\right] \leq DE\left[X_{is}^2 X_{js}^2\right] \leq D^2$ by Assumption 6 (and nonnegativity, for the first inequality), so the absolute value of the contribution of these terms is

$$\frac{1}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{s=1}^{T}E\left[X_{is}^2 X_{js}^2 \varepsilon_s^4\right] \tag{A.73}$$

$$
\begin{aligned}
&\leq \frac{D^2}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{s=1}^{T} 1 \\
&= \frac{K^2 D^2}{T (T-K)^2} \\
&\leq \frac{C_6}{K}.
\end{aligned}
$$

Case 2: All time subscripts are distinct.

Here $E[X_{is} X_{it} X_{ju} X_{jw} \varepsilon_s \varepsilon_t \varepsilon_u \varepsilon_w] = 0$ by the m. d. s. property of $\varepsilon$ (Assumption 6).

Case 3: $s = t > u > w$ (w. l. o. g.; also, $t = u > w > s$, etc.).

The terms here are of the form $E[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w]$. Noting that, by the m. d. s. property of $\varepsilon$ (Assumption 6), $E[X_{ju} \varepsilon_u X_{jw} \varepsilon_w] = 0$, we see that

$$
\begin{aligned}
& \left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w\right] \right| \qquad \text{(A.74)} \\
=\; & \left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w\right] - E\left[X_{it}^2 \varepsilon_t^2\right] E\left[X_{ju} \varepsilon_u X_{jw} \varepsilon_w\right] \right| \\
\leq\; & \nu_{t-u} \left(E\left[X_{it}^4 \varepsilon_t^4\right]\right)^{1/2} \left(E\left[X_{ju}^2 \varepsilon_u^2 X_{jw}^2 \varepsilon_w^2\right]\right)^{1/2} \\
\leq\; & M_3 \nu_{t-u}
\end{aligned}
$$

where the first inequality follows by the observation preceding the display, the second inequality is by Doukhan (1994, Theorem 3 (5) on page 9) and Assumption 7, and the third inequality follows from Assumption 6.

Observing that $E[X_{jw} \varepsilon_w] = 0$ by the m. d. s. property of $\varepsilon$ (Assumption 6), we also obtain

$$
\begin{aligned}
& \left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w\right] \right| \qquad \text{(A.75)} \\
\leq\; & \left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w\right] - E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u\right] E\left[X_{jw} \varepsilon_w\right] \right| \\
\leq\; & \nu_{u-w} \left(E\left[X_{it}^4 \varepsilon_t^4 X_{ju}^2 \varepsilon_u^2\right]\right)^{1/2} \left(E\left[X_{jw}^2 \varepsilon_w^2\right]\right)^{1/2} \\
\leq\; & M_4 \nu_{u-w}
\end{aligned}
$$

in an entirely similar fashion.

Thus, the total absolute-value of the contribution of the terms covered by this case satisfies

$$
\begin{aligned}
& \left| \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w\right] \right| \qquad \text{(A.76)} \\
\leq\; & \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} \left| E\left[X_{it}^2 \varepsilon_t^2 X_{ju} \varepsilon_u X_{jw} \varepsilon_w\right] \right| \\
\leq\; & \frac{1}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{w=1}^{T} \sum_{u=w+1}^{T} \sum_{t=u+1}^{T} \min\left\{M_3 \nu_{t-u}, M_4 \nu_{u-w}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{\max\{M_3, M_4\}}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{w=1}^{T}\sum_{u=w+1}^{T}\sum_{t=u+1}^{T} \min\{\nu_{t-u}, \nu_{u-w}\} \\
&\leq \frac{D\max\{M_3, M_4\}}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{w=1}^{T}\sum_{u=w+1}^{T}\sum_{t=u+1}^{T} \min\left\{e^{-\frac{\lambda}{2}(t-u)}, e^{-\frac{\lambda}{2}(u-w)}\right\} \\
&= \frac{D\max\{M_3, M_4\}}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{w=1}^{T}\sum_{u=w+1}^{T} \left\{ \begin{array}{c} \sum_{t=u+1}^{2u-w} e^{-\frac{\lambda}{2}(u-w)} \\ + \sum_{t=2u-w+1}^{T} e^{-\frac{\lambda}{2}(t-u)} \end{array} \right\} \\
&= \frac{D\max\{M_3, M_4\}}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{w=1}^{T} \left\{ \begin{array}{c} \sum_{u=w+1}^{T} (u-w) e^{-\frac{\lambda}{2}(u-w)} \\ + \sum_{u=w+1}^{T}\sum_{t=2u-w+1}^{T} e^{-\frac{\lambda}{2}(t-u)} \end{array} \right\} \\
&= \frac{D\max\{M_3, M_4\}}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{w=1}^{T} \left\{ \begin{array}{c} \sum_{n=1}^{T-w} n e^{-\frac{\lambda}{2}n} \\ + \sum_{n=1}^{T-w}\sum_{m=n+1}^{T-(n+w)} e^{-\frac{\lambda}{2}m} \end{array} \right\} \\
&\leq \frac{D\max\{M_3, M_4\}}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{w=1}^{T} \left\{ \begin{array}{c} \sum_{n=1}^{\infty} n e^{-\frac{\lambda}{2}n} \\ + \sum_{n=1}^{\infty}\sum_{m=n+1}^{\infty} e^{-\frac{\lambda}{2}m} \end{array} \right\} \\
&\leq \frac{2D^2\max\{M_3, M_4\}}{T^2 (T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{w=1}^{T} 1 \\
&= \frac{2D^2\max\{M_3, M_4\} K^2}{T (T-K)^2} \\
&\leq \frac{C_7}{K}
\end{aligned}
$$

where the first inequality is clear, the second inequality follows from the two bounds given above (since each of the previous two bounds holds for each term, the smaller of them must hold), the third inequality is evident, the fourth inequality follows from Assumption 7, the first equality follows from the fact that $t-u > u-w$ if and only if $t > 2u-w$ (and the monotonicity of $e^{-n}$), the second equality follows from trivial summation and rearrangement, the third equality follows upon setting $m = t-u$ and $n = u-w$, the fifth inequality follows from the nonnegativity of $\nu_n$ and $n$, the sixth inequality follows directly from the bounds of expression (A.56), the third equality is due to simple summation, and the seventh and final inequality follows from the asymptotic nesting, in which $K$ is asymptotically a constant fraction of $T$.

Case 4: $s > t = u > w$ (w. l. o. g.; this case includes all terms whose time subscripts take this form, with one greatest, two equal, and one least).

Here we obtain, by the m. d. s. property of $\varepsilon$ (Assumption 6), $E\left[X_{is}\varepsilon_s X_{it} X_{jt} \varepsilon_t^2 X_{jw} \varepsilon_w\right] = 0$.

Case 5: $s > t > u = w$ (w. l. o. g.; this case includes all terms whose time subscripts take this form, with one greatest, one intermediate, and two equal and minimal).

Just as in Case 4, $E\left[X_{is}\varepsilon_s X_{it}\varepsilon_t X_{ju}^2\varepsilon_u^2\right] = 0$ by the m. d. s. property of $\varepsilon$.

Case 6: $s = t = u > w$ (w. l. o. g.; this case includes all terms which have three equal time subscripts and one lesser time subscript).

Noting that $E\left[X_{jw}\varepsilon_w\right] = 0$ by the m. d. s. property of $\varepsilon$ (Assumption 6), we have that

$$\begin{aligned} & \left|E\left[X_{it}^2 X_{jt}\varepsilon_t^3 X_{jw}\varepsilon_w\right]\right| && \text{(A.77)} \\ \leq\ & \left|E\left[X_{it}^2 X_{jt}\varepsilon_t^3 X_{jw}\varepsilon_w\right] - E\left[X_{it}^2 X_{jt}\varepsilon_t^3\right] E\left[X_{jw}\varepsilon_w\right]\right| \\ \leq\ & \nu_{t-w}\left(E\left[X_{it}^4 X_{jt}^2\varepsilon_t^6\right]\right)^{1/2}\left(E\left[X_{jw}^2\varepsilon_w^2\right]\right)^{1/2} \\ \leq\ & M_5\nu_{t-w} \end{aligned}$$

where the first inequality follows from the preceding observation, the second is due to Doukhan (1994, Theorem 3 (5) on page 9) and Assumption 7, and the final inequality is due to the uniform moment bounds of Assumption 6.

The total absolute-value contribution of the terms handled in this case is thus

$$\begin{aligned} & \left|\frac{1}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}\sum_{w=t+1}^{T} E\left[X_{it}^2 X_{jt}\varepsilon_t^3 X_{jw}\varepsilon_w\right]\right| && \text{(A.78)} \\ \leq\ & \frac{1}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}\sum_{w=t+1}^{T}\left|E\left[X_{it}^2 X_{jt}\varepsilon_t^3 X_{jw}\varepsilon_w\right]\right| \\ \leq\ & \frac{M_5}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}\sum_{w=t+1}^{T}\nu_{t-w} \\ \leq\ & \frac{DM_5}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T} 1 \\ =\ & \frac{DM_5K^2}{T(T-K)^2} \\ \leq\ & \frac{C_8}{K} \end{aligned}$$

where the first inequality is standard, the second follows from the above bound on each summand, the third inequality follows from the summability condition of expression (A.57), the equality follows by simple summation, and the final inequality is due to the asymptotic nesting, in which $K$ is a constant fraction of $T$ asymptotically.

Case 7: $s > t = u = w$ (w. l. o. g.; this case handles all terms which have three equal time subscripts and one greater time subscript).

By the m. d. s. property of the $\varepsilon$, we have that $E\left[X_{is}\varepsilon_s X_{it} X_{jt}^2\varepsilon_t^3\right] = 0$.

Case 8: $s = t > u = w$ (note that here, we deal only with the specific subscript ordering given).

We have that $E\left[X_{it}^2\varepsilon_t^2 X_{ju}^2\varepsilon_u^2\right] = \sigma_\varepsilon^2 E\left[X_{it}^2 X_{ju}^2 \varepsilon_u^2\right]$, $E\left[X_{it}^2\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{it}^2\right]$, and $E\left[X_{ju}^2\varepsilon_u^2\right] = \sigma_\varepsilon^2 E\left[X_{ju}^2\right]$, so

$$\begin{aligned} & \left|E\left[X_{it}^2\varepsilon_t^2 X_{ju}^2\varepsilon_u^2\right] - \sigma_\varepsilon^4 E\left[X_{it}^2\right] E\left[X_{ju}^2\right]\right| \\ = \; & \sigma_\varepsilon^2 \left|E\left[X_{it}^2 X_{ju}^2\varepsilon_u^2\right] - \sigma_\varepsilon^2 E\left[X_{it}^2\right] E\left[X_{ju}^2\right]\right| \\ = \; & \sigma_\varepsilon^2 \left|E\left[X_{it}^2 X_{ju}^2\varepsilon_u^2\right] - E\left[X_{it}^2\right] E\left[X_{ju}^2\varepsilon_u^2\right]\right| \\ \leq \; & \sigma_\varepsilon^2 \nu_{t-u} \left(E\left[X_{it}^4\right]\right)^{1/2} \left(E\left[X_{ju}^4\varepsilon_u^4\right]\right)^{1/2} \\ \leq \; & M_6 \nu_{t-u} \end{aligned} \tag{A.79}$$

where the first two equalities follow from the observations made immediately above the display, the first inequality comes from Hall and Heyde (1980, Theorem A.6 on page 278) and Assumption 7, and the final inequality comes from the uniform moment bounds given by Assumption 6.

We can now bound the absolute value of the difference between the total contribution of the terms handled by this case and

$$\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T} E\left[X_{it}^2\right] E\left[X_{ju}^2\right]. \tag{A.80}$$

This is done as follows:

$$\begin{aligned} & \left| \begin{array}{l} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T} E\left[X_{it}^2\varepsilon_t^2 X_{ju}^2\varepsilon_u^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \end{array} \right| \\ \leq \; & \frac{1}{T^2\left(T-K\right)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T} \left|E\left[X_{it}^2\varepsilon_t^2 X_{ju}^2\varepsilon_u^2\right] - \sigma_\varepsilon^4 E\left[X_{it}^2\right] E\left[X_{ju}^2\right]\right| \\ \leq \; & \frac{M_6}{T^2\left(T-K\right)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=u+1}^{T} \nu_{t-u} \\ \leq \; & \frac{M_6}{T^2\left(T-K\right)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{n=1}^{\infty} \nu_n \\ \leq \; & \frac{M_6 D}{T^2\left(T-K\right)^2} \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T} 1 \\ = \; & \frac{M_6 D K^2}{T\left(T-K\right)^2} \\ \leq \; & \frac{C_9}{K} \end{aligned} \tag{A.81}$$

where the first inequality is standard, the second is by the above bound on each summand, the third follows from nonnegativity of the $\nu_n$ and setting

$n = t - u$, the fourth follows from expression (A.57), the equality is by simple summation, and the final inequality is due to the asymptotic nesting, in which $K$ is an asymptotically constant fraction of $T$.

Case 9: $u = w > s = t$ (note that here, we deal only with the specific subscript ordering given).

This is identical to Case 8 above, except that we get the time terms $t < u$ rather than the terms $t > u$. Thus, the result is that the total contribution of the terms to which this case applies is close to

$$\frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{u-1} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \tag{A.82}$$

in the sense of the following bound:

$$\begin{aligned} & \left| \begin{array}{l} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{u-1} E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{u-1} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \end{array} \right| \\ \leq \; & \frac{C_{10}}{K}. \end{aligned} \tag{A.83}$$

Case 10: $s = u > t = w$ (note that here, we deal only with the specific ordering of the subscripts given).

$E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{is} X_{js} X_{it} X_{jt} \varepsilon_t^2\right]$ by the homoskedastic m. d. s. property of $\varepsilon$ (from Assumption 6). Now, $E\left[X_{it} X_{jt} \varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{it} X_{jt}\right]$ for the same reason, so

$$\begin{aligned} & \left| E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] - \sigma_\varepsilon^4 E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \right| \\ = \; & \left| \sigma_\varepsilon^2 E\left[X_{is} X_{js} X_{it} X_{jt} \varepsilon_t^2\right] - \sigma_\varepsilon^2 E\left[X_{it} X_{jt} \varepsilon_t^2\right] E\left[X_{is} X_{js}\right] \right| \\ = \; & \sigma_\varepsilon^2 \left| E\left[X_{is} X_{js} X_{it} X_{jt} \varepsilon_t^2\right] - E\left[X_{it} X_{jt} \varepsilon_t^2\right] E\left[X_{is} X_{js}\right] \right| \\ \leq \; & \sigma_\varepsilon^2 \nu_{s-t} \left(E\left[X_{it}^2 X_{jt}^2 \varepsilon_t^4\right]\right)^{1/2} \left(E\left[X_{is}^2 X_{js}^2\right]\right)^{1/2} \\ \leq \; & M_7 \nu_{s-t} \end{aligned} \tag{A.84}$$

where the first two equalities are by the identities noted immediately above the display, the first inequality is by Doukhan (1994, Theorem 3 (5) on page 9) and Assumption 7, and the final inequality is by the uniform moment bounds given in Assumption 6. Thus, we can obtain, exactly as in Cases 8 and 9 above, a bound on the absolute value of the difference between the total contribution of the terms handled here and the object

$$\frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \tag{A.85}$$

where the bound is:

$$\left| \begin{array}{l} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \end{array} \right| \tag{A.86}$$

$$\leq \quad \frac{C_{11}}{K}.$$

Case 11: $t = w > s = u$ (note that here, we deal only with the specific ordering of the subscripts given).

This is entirely similar to Case 10, except that the bound is derived on the distance between the total contribution of the terms handled here and

$$\frac{\sigma_\varepsilon^4}{T^2 (T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right]. \tag{A.87}$$

The bound is:

$$\left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \end{array} \right| \tag{A.88}$$
$$\leq \quad \frac{C_{12}}{K}.$$

Now we may pull all the cases back together again to get:

$$\left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{u=1}^{T} \sum_{w=1}^{T} \\ E\left[X_{is} X_{it} X_{ju} X_{jw} \varepsilon_s \varepsilon_t \varepsilon_u \varepsilon_w\right] \\ -\frac{\sigma_\varepsilon^4 K^2}{(T-K)^2} \end{array} \right| \tag{A.89}$$

$$\begin{aligned}
\leq \quad & \frac{C_7 + C_8 + C_9}{K} \\
& + \left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=u+1}^{T} E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=u+1}^{T} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \end{array} \right| \\
& + \left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{u-1} E\left[X_{it}^2 \varepsilon_t^2 X_{ju}^2 \varepsilon_u^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{u-1} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \end{array} \right| \\
& + \left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=t+1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \end{array} \right| \\
& + \left| \begin{array}{c} \frac{1}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{is} X_{js} \varepsilon_s^2 X_{it} X_{jt} \varepsilon_t^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{t-1} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \end{array} \right| \\
& + \left| \begin{array}{c} \frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{is} X_{js}\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} E\left[X_{it} X_{jt}\right] E\left[X_{it} X_{jt}\right] \end{array} \right| \\
& + \left| \begin{array}{c} \frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{u=1}^{T} \sum_{t=1}^{T} E\left[X_{it}^2\right] E\left[X_{ju}^2\right] \\ -\frac{\sigma_\varepsilon^4}{T^2(T-K)^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{T} E\left[X_{it}^2\right] E\left[X_{jt}^2\right] \\ -\frac{\sigma_\varepsilon^4 K^2}{(T-K)^2} \end{array} \right| \\
\leq \quad & \frac{C_7 + C_8 + C_9 + C_{10} + C_{11} + C_{12}}{K}
\end{aligned}$$

$$
\begin{aligned}
&+\left|\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}X_{jt}\right]E\left[X_{is}X_{js}\right]\right| \\
&+\left|\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}E\left[X_{it}X_{jt}\right]E\left[X_{it}X_{jt}\right]\right| \\
&+\left|\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}E\left[X_{it}^2\right]E\left[X_{jt}^2\right]\right| \\
&+\left|\begin{array}{c}\frac{\sigma_\varepsilon^4}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{u=1}^{T}\sum_{t=1}^{T}E\left[X_{it}^2\right]E\left[X_{ju}^2\right] \\ -\frac{\sigma_\varepsilon^4K^2}{(T-K)^2}\end{array}\right| \\
\leq\ & \frac{C_7+C_8+C_9+C_{10}+C_{11}+C_{12}+C_{13}+C_{14}}{K} \\
&+\left|\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}X_{jt}\right]E\left[X_{is}X_{js}\right]\right| \\
\leq\ & \frac{C_7+C_8+C_9+C_{10}+C_{11}+C_{12}+C_{13}+C_{14}}{K} \\
&+\left|\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}X_{jt}\right]E\left[X_{is}X_{js}\right]\right| \\
&+\left|\frac{\sigma_\varepsilon^4}{T^2\left(T-K\right)^2}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}^2\right]E\left[X_{is}^2\right]\right| \\
\leq\ & \frac{C_7+C_8+C_9+C_{10}+C_{11}+C_{12}+C_{13}+C_{14}+C_{15}}{K} \\
=\ & \frac{C_{16}}{K}.
\end{aligned}
$$

We may finally combine the bounds we have obtained for each of the three expectations involved, and notice that $\sigma_\varepsilon^4 = \frac{T^2-2KT+K^2}{(T-K)^2}\sigma_\varepsilon^4$, to yield:

$$
\begin{aligned}
&\left|E\left[\left(\hat{\sigma}_\varepsilon^2\right)^2\right]-\sigma_\varepsilon^4\right| \qquad \text{(A.90)} \\
\leq\ &\left|E\left[\frac{1}{\left(T-K\right)^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\varepsilon_s^2\varepsilon_t^2\right]-\frac{T\left(T-1\right)}{\left(T-K\right)^2}\sigma_\varepsilon^4\right| \\
&+\left|\frac{T}{\left(T-K\right)^2}\sigma_\varepsilon^4\right| \\
&+\left|\begin{array}{c}E\left[\frac{-2}{T(T-K)^2}\sum_{t=1}^{T}\sum_{i=1}^{K}\sum_{s=1}^{T}\sum_{u=1}^{T}\varepsilon_t^2X_{is}X_{iu}\varepsilon_s\varepsilon_u\right] \\ -\left(\frac{-2KT}{(T-K)^2}\sigma_\varepsilon^2\right)\end{array}\right| \\
&+\left|\begin{array}{c}\frac{1}{T^2(T-K)^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{u=1}^{T}\sum_{w=1}^{T} \\ E\left[X_{is}X_{it}X_{ju}X_{jw}\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_w\right] \\ -\frac{\sigma_\varepsilon^4K^2}{(T-K)^2}\end{array}\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{C_1}{K} + \left|\frac{T}{(T-K)^2}\sigma_\varepsilon^4\right| + \frac{C_5}{K} + \frac{C_{16}}{K} \\
&\leq \frac{C_1}{K} + \frac{C_{17}}{K} + \frac{C_5}{K} + \frac{C_{16}}{K} \\
&\leq \frac{C}{K}
\end{aligned}
$$

which completes the proof of Theorem 4(a). Q.E.D.

*Proof of (b)*:

$$
\begin{aligned}
(T/K)R\left(\hat{b}, b; f_K\right) &= \frac{1}{K}E\left[\left(\hat{b}-b\right)'\left(\hat{b}-b\right)\right] \\
&= \frac{1}{K}E\left[\left(\frac{1}{\sqrt{T}}X'\varepsilon\right)'\left(\frac{1}{\sqrt{T}}X'\varepsilon\right)\right] \\
&= \frac{1}{KT}E\left[\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}X_{it}X_{is}\varepsilon_t\varepsilon_s\right] \\
&= \frac{1}{KT}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}X_{is}\varepsilon_t\varepsilon_s\right] \\
&= \frac{1}{KT}\sum_{i=1}^{K}\sum_{t=1}^{T}E\left[X_{it}^2\varepsilon_t^2\right] \\
&= \frac{\sigma_\varepsilon^2}{KT}\sum_{i=1}^{K}E\left[\sum_{t=1}^{T}X_{it}^2\right] \\
&= \frac{\sigma_\varepsilon^2}{KT}\sum_{i=1}^{K}T \\
&= \sigma_\varepsilon^2
\end{aligned} \tag{A.91}
$$

where the first four equalities follow from simple calculation, the fifth equality holds because if $s \neq t$, then $E\left[X_{is}X_{it}\varepsilon_s\varepsilon_t\right] = 0$ by the m. d. s. property of $\varepsilon$ (from Assumption 6), the sixth equality is due to the fact that $E\left[X_{it}^2\varepsilon_t^2\right] = \sigma_\varepsilon^2 E\left[X_{it}^2\right]$, the seventh equality follows from $\sum_{t=1}^{T} X_{it}^2 = T$ (by Assumption 1), and the eighth equality follows from simple summation. Since $(T/K)R\left(\hat{b}, b\right)$ does not depend on $b$, we have $(T/K)r_G\left(\hat{b}, f_K\right) = \sigma_\varepsilon^2$. Q. E. D.

## A.6 Proof of Theorem 5

The proof of part (a) exactly mimics the proof of Theorem 1(a) using the more general results in Lemma S-4 above.

The proof of part (b) differs somewhat from the proof of Theorem 1(b) and we provide the details here.

Let $\hat{b}_i^{ISEB} = \hat{b}_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}$; this is the infeasible simple empirical Bayes estimator based on the true average marginal $\bar{m}_K$. Now write

$$r_G\left(\hat{b}^{NSEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \tag{A.92}$$

$$= r_G\left(\hat{b}^{NSEB}, f_K\right) - r_G\left(\hat{b}^{ISEB}, f_K\right) \tag{Term I}$$

$$+ r_G\left(\hat{b}^{ISEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \tag{Term II}$$

We will show that both Term I and Term II go to zero. Recall that $r_G\left(\hat{b}^{NB}, \phi_K\right) < \infty$ from Lemma S-4. First, let us deal with the easier term, Term II:

$$\begin{aligned}
& r_G\left(\hat{b}^{ISEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \qquad \text{(A.93)}\\
= \; & \rho \int\int \frac{1}{K}\sum_{i=1}^{K}\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_K\left(\hat{b} - b\right) d\hat{b}dG_K(b) \\
& - \rho \int\int \frac{1}{K}\sum_{i=1}^{K}\left(\hat{b}_i^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \phi_K\left(\hat{b} - b\right) d\hat{b}dG_K(b) \\
= \; & \frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i dG(b_i) \\
& - \frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{b}_i^{NB}\left(\hat{b}_i\right) - b_i\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i dG(b_i) \\
= \; & \frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{b}_1^{ISEB}(x) - y\right)^2 f_{iK}(x-y)\, dxdG(y) \\
& - \frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{b}_1^{NB}(x) - y\right)^2 \phi(x-y)\, dxdG(y) \\
= \; & \rho\int\int\left(\hat{b}_1^{ISEB}(x) - y\right)^2 \bar{f}_K(x-y)\, dxdG(y) \\
& - \rho\int\int\left(\hat{b}_1^{NB}(x) - y\right)^2 \phi(x-y)\, dxdG(y)
\end{aligned}$$

The second equality comes from the fact that both $\hat{b}_i^{NB}$ and $\hat{b}_i^{ISEB}$ depend only on $\hat{b}_i$ and not on $\hat{b}_{-i}$, while the third equality follows from the fact that the functional form of both $\hat{b}_i^{NB}$ and $\hat{b}_i^{ISEB}$ is the same for each $i$.

From the bounds in Lemma S-3, and the definition of $\hat{b}_i^{ISEB}$, we see that the first integrand in the final line converges to the second pointwise. Thus, if we can produce a dominating function for the first integrand, we may apply the Dominated Convergence Theorem and be finished. But

$$\begin{aligned}
& \left(\hat{b}_1^{ISEB}(x) - y\right)^2 \bar{f}_K(x-y) \qquad \text{(A.94)}\\
= \; & \left((x-y) - \sigma_\varepsilon^2 \frac{\bar{m}_K'(x)}{\bar{m}_K(x) \; + \; s_K}\right)^2 \bar{f}_K(x-y)
\end{aligned}$$

$$\leq \quad 2(x-y)^2 \bar{f}_K(x-y) \;+\; 2\sigma_\varepsilon^4 \left( \frac{\bar{m}_K'(x)}{\bar{m}_K(x) \;+\; s_K} \right)^2 \bar{f}_K(x-y)$$

The first term in this bound is evidently integrable by Assumption 6. The second term is integrable by Lemma S-4, so we have shown that Term II converges to zero.

Before dealing with the more challenging Term I, it is worth noting that we will never encounter a "division by zero" problem, because of three facts: 1) the limit density $m_\phi(x)$ is positive everywhere (since it is a convolution of a normal density with the prior); 2) the approximation error given by Lemma S-3 is less than $s_K$ by construction (for large enough $K$, see Assumption 4); 3) the denominator of $\hat{b}_i^{ISEB}$ is thus always greater than $m_\phi$, since we take $\bar{m}_K(x) > m_\phi(x) - s_K$ (for large enough $K$) and add $s_K$ to both sides of the inequality, so that we never encounter difficulties.

Now we shall demonstrate that Term I converges to zero. Our proof is in the spirit of Bickel *et al.* (1993, p. 405 ff) and van der Vaart (1988, p. 169 ff), but we extend their approaches to handle cross-sectional dependence of the $\hat{b}_i$ and to deal with a nonconstant sequence of likelihoods.

Note that, from (2.10) through (2.13) of the body of this paper, we can write $\hat{l}_{iK}(x) = \hat{l}\left(\hat{b}_i\,;\,\hat{b}_{-i}\right)$, where the function $\hat{l}(\cdot,\cdot)$ does not depend on $i$. This representation, and a similar one for $\breve{l}$, are adopted here. Although $\hat{\sigma}_\varepsilon^2$ depends on the full data, rather than only $\hat{b}$, this dependence is suppressed for notational convenience; the treatment below does, however, account for the fact that $\hat{\sigma}_\varepsilon^2$ is a function of the full data (see display (A.99)). Now,

$$r_G\left(\hat{b}^{ISEB}, f_K\right) - r_G\left(\hat{b}^{NSEB}, f_K\right) \tag{A.95}$$

$$= \frac{1}{K}\sum_{i=1}^K \rho \int\!\!\int \left[ \begin{array}{c} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 \\ -\left(\hat{b}_i^{NSEB}\left(\hat{b}\right) - b_i\right)^2 \end{array} \right] f_K\left(\hat{b}-b\right) d\hat{b}\, dG_K(b)$$

$$= \frac{1}{K}\sum_{i=1}^K \rho \int\!\!\int \left[ \begin{array}{c} \left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 - \\ \left(\hat{b}_i - b_i - \hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right)^2 \end{array} \right] f_K\left(\hat{b}-b\right) d\hat{b}\, dG_K(b)$$

$$= \frac{1}{K}\sum_{i=1}^K \rho \int\!\!\int \left[ \begin{array}{c} -\left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 + \\ 2\left(\hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)\left(\hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right) \end{array} \right]$$

$$\times f_K\left(\hat{b}-b\right) d\hat{b}\, dG_K(b)$$

$$= -\frac{1}{K}\sum_{i=1}^K \rho \int\!\!\int \left( \hat{\sigma}_\varepsilon^2 \hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \sigma_\varepsilon^2 \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \;+\; s_K} \right)^2 f_K\left(\hat{b}-b\right) d\hat{b}\, dG_K(b)$$

$$+\frac{1}{K}\sum_{i=1}^K \rho \int\!\!\int 2\left( \hat{b}_i - b_i - \sigma_\varepsilon^2 \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \;+\; s_K} \right) \times$$

$$\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)f_K\left(\hat{b}-b\right)d\hat{b}dG_K\left(b\right)$$

$$\begin{aligned}
\leq\ & -\frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K\left(b\right)\\
& +\frac{2}{K}\sum_{i=1}^{K}\rho\left[\begin{array}{c}\left(\int\int\left(\hat{b}_i-b_i-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K\left(b\right)\right)^{\frac{1}{2}}\times\\ \left(\int\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K\left(b\right)\right)^{\frac{1}{2}}\end{array}\right]\\
\leq\ & -\frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K\left(b\right)\\
& +2\left\{\frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{b}_i-b_i-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K\left(b\right)\right\}^{\frac{1}{2}}\times\\
& \left\{\frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K\left(b\right)\right\}^{\frac{1}{2}}
\end{aligned}$$

In the above, the first and second equalities are by definition; the third is a consequence of the fact that $a^2-b^2 = 2a\left(a-b\right)-\left(a-b\right)^2$; the fourth equality simply breaks out the two terms of the integrand; the first inequality is an application of Hölder's inequality, and the second is an application of the Cauchy-Schwarz inequality.

From the final bound given above, we see that if we can show that

$$\begin{aligned}
& \frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left\{\begin{array}{c}\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2\\ \times f_K\left(\hat{b}-b\right)\end{array}\right\}d\hat{b}dG_K\left(b\right) \quad \text{(A.96)}\\
\rightarrow\ & 0
\end{aligned}$$

then we will have demonstrated that Term I converges to zero, because

$$\begin{aligned}
& \frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left\{\begin{array}{c}\left(\hat{b}_i-b_i-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2\\ \times f_K\left(\hat{b}-b\right)\end{array}\right\}d\hat{b}dG_K\left(b\right) \qquad \text{(A.97)}\\
=\ & r_G\left(\hat{b}^{ISEB},f_K\right)\\
\rightarrow\ & r_G\left(\hat{b}^{NB},\phi_K\right)\\
<\ & \infty
\end{aligned}$$

from the fact that Term II converges to zero, and, as noted above, $r_G\left(\hat{b}^{NB}, \phi_K\right) < \infty$.

Using $(a+b)^2 \leq 2a^2 + 2b^2$, and adding and subtracting $\sigma_\varepsilon^2 \hat{l}\left(\hat{b}_i ; \hat{b}_{-i}\right)$, we have

$$\frac{1}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{\sigma}_\varepsilon^2\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_\varepsilon^2\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2\times \quad (A.98)$$

$$f_K\left(\hat{b}-b\right)d\hat{b}dG_K(b)$$

$$\leq \frac{2}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2\hat{l}^2\left(\hat{b}_i;\hat{b}_{-i}\right)f_K\left(\hat{b}-b\right)d\hat{b}dG_K(b) \quad \text{(Term A)}$$

$$+\frac{2\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2\times \quad \text{(Term B)}$$

$$f_K\left(\hat{b}-b\right)d\hat{b}dG_K(b)$$

which separates the problem of nonparametric score estimation from the problem of estimating the residual variance. Now, Term A satisfies

$$\begin{aligned}
& \frac{2}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2\hat{l}^2\left(\hat{b}_i;\hat{b}_{-i}\right)f_K\left(\hat{b}-b\right)d\hat{b}dG_K(b) \quad (A.99)\\
\leq\ & \frac{2\rho q_K^2}{K}\sum_{i=1}^{K}\int\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K(b)\\
=\ & 2\rho q_K^2\int\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K(b)\\
=\ & 2\rho q_K^2 Var\left[\hat{\sigma}_\varepsilon^2\right]\\
\leq\ & 2\rho q_K^2\frac{C}{K}\\
\rightarrow\ & 0
\end{aligned}$$

where the first inequality comes from the truncation of our estimator $\hat{l}$ of the score, and the second inequality comes from Theorem 4(a). The convergence of the final bound to zero is by construction: by display (A.1), $\frac{q_K^2}{K} \rightarrow 0$. Thus, Term A converges to zero.

Now consider Term B. Define

$$D_i \equiv \left\{\hat{b} : \left|\hat{b}_i\right| \leq \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K} \text{ and } \left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| \leq q_K\right\},$$

and let $E_{D_i}[(\cdot)] \equiv \int_{\hat{b}\in D_i}(\cdot)\, m_K\left(\hat{b}\right)d\hat{b}$ (so that the area of integration is restricted, but in a way which may differ for each $i$). Now,

$$\frac{2\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\rho\int\int\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}dG_K(b) \quad (A.100)$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \ + \ s_K}\right)^2 m_K\left(\hat{b}\right) d\hat{b}$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int E\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \ + \ s_K}\right)^2 \mid \hat{b}_i\right] m_{iK}\left(\hat{b}_i\right) d\hat{b}_i$$

$$\leq \frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \ + \ s_K}\right)^2 m_{iK}\left(\hat{b}_i\right) \times \tag{Term Bi}$$

$$\left(\Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i\right) + \Pr\left(\left|\hat{b}_i\right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K} \mid \hat{b}_i\right)\right) d\hat{b}_i$$

$$+ \ \frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K} E_{D_i}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \ + \ s_K}\right)^2\right] \tag{Term Bii}$$

where the first equality is by the Tonelli-Fubini Theorem, the second is by definition, and the inequality follows from the truncation of the estimated score function according to the definition of $D_i$.

Consider Term Bi first.

$$\frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \ + \ s_K}\right)^2 m_{iK}\left(\hat{b}_i\right) \times \tag{A.101}$$

$$\left(\begin{array}{c} \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i\right) + \\ \Pr\left(\left|\hat{b}_i\right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K} \mid \hat{b}_i\right) \end{array}\right) d\hat{b}_i$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \ + \ s_K}\right)^2 m_{iK}\left(\hat{b}_i\right)$$

$$\times \Pr\left(\left|\hat{b}_i\right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K} \mid \hat{b}_i\right) d\hat{b}_i$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\int_{-\infty}^{\infty}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) \ + \ s_K}\right)^2 m_{iK}\left(\hat{b}_i\right) \times$$

$$\Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i\right) d\hat{b}_i$$

$$\leq \frac{2\rho\sigma_\varepsilon^4}{K s_K^2} C \sum_{i=1}^{K} \Pr\left(\left|\hat{b}_i\right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K}\right)$$

$$+ \ 2\rho\sigma_\varepsilon^4 \int_{|x|>d_K}\left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) \ + \ s_K}\right)^2 \bar{m}_K(x)\, dx$$

$$
\begin{aligned}
& + \sup_{i,\left|\hat{b}_i\right| \leq d_K} \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i\right) \\
& \times 2\rho\sigma_\varepsilon^4 \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) \ + \ s_K}\right)^2 \bar{m}_K(x)\,dx \\
\leq \quad & \frac{2\rho\sigma_\varepsilon^4}{K s_K^2} C \sum_{i=1}^{K} \left\{ \begin{array}{c} \Pr\left(\left|\hat{b}_i\right| > \sqrt{\frac{\sigma_\varepsilon^2}{256}\log K}\right) \\ + \Pr\left(\hat{\sigma}_\varepsilon^2 \leq \frac{1}{2}\sigma_\varepsilon^2\right) \end{array} \right\} \\
& + 2\rho\sigma_\varepsilon^4 \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) \ + \ s_K}\right)^2 \bar{m}_K(x)\,dx \\
& + \sup_{i,\left|\hat{b}_i\right| \leq d_K} \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i\right) \\
& \times 2\rho\sigma_\varepsilon^4 \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) \ + \ s_K}\right)^2 \bar{m}_K(x)\,dx \\
\leq \quad & C_2 s_K^{-2} \left\{\frac{1}{\log K} + \frac{1}{K}\right\} \\
& + 2\rho\sigma_\varepsilon^4 \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) \ + \ s_K}\right)^2 \bar{m}_K(x)\,dx \\
& + \sup_{i,\left|\hat{b}_i\right| \leq d_K} \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i\right) \\
& \times 2\rho\sigma_\varepsilon^4 \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x) \ + \ s_K}\right)^2 \bar{m}_K(x)\,dx
\end{aligned}
$$

The equality is trivial; the first inequality follows from the boundedness of $\left(\bar{m}_K'(x)\right)^2$ (by (A.27)), the second inequality follows by $\Pr(A) \leq \Pr(B) + \Pr(C)$ whenever $A \subset (B \cup C)$ (if $\left|\hat{b}_i\right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K}$, then either $\left|\hat{b}_i\right| > \sqrt{\frac{\sigma_\varepsilon^2}{256}\log K}$ or $\hat{\sigma}_\varepsilon^2 \leq \frac{1}{2}\sigma_\varepsilon^2$, or both), and the third inequality follows from Chebyshev's inequality and the variance bound of Theorem 4(a) (along with $\sigma_\varepsilon^2 > 0$ by Assumption 6). Of the terms in the final expression, we see immediately that the first term converges to zero by Assumption 4 and the second converges to zero by the uniform integrability of its integrand, which was shown in the course of the proof that Term II converges to zero, and by $d_K \to \infty$. The third term converges to zero by Lemmas S-4 and S-6(c). Thus, Term Bi converges to zero.

Finally, turn to Term Bii. Consider the $i^{th}$ term of the average which makes up Term Bii. We define $E_{D_i}^{out}[(\cdot)]$ to be $\int_{\hat{b}\in D_i \cap \{\hat{\sigma}_\varepsilon^2 > 2\sigma_\varepsilon^2\}} (\cdot)\, m_K\left(\hat{b}\right) d\hat{b}$ and $E_{D_i}^{cond}\left[(\cdot) \mid \hat{b}_i\right]$ to be $\int_{\hat{b}\in D_i} (\cdot)\, m_K\left(\hat{b}_{-i} \mid \hat{b}_i\right) d\hat{b}_{-i}$ (so that both the probability measure and the area of integration depend on $\hat{b}_i$). Since $\left(\frac{a}{b} - \frac{c}{d}\right)^2 = \left(\frac{a}{b}\left(\frac{d-b}{d}\right) + \frac{a-c}{d}\right)^2 \leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2 + 2\left(\frac{a-c}{d}\right)^2$ we have:

$$
\begin{aligned}
&E_{D_i}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\right] \qquad \text{(A.102)}\\
&\leq 2E^{out}_{D_i}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right)^2\right]+2E^{out}_{D_i}\left[\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\right]\\
&\quad+\int_{-d_K}^{d_K}E^{cond}_{D_i}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2|\hat{b}_i\right]m_{iK}\left(\hat{b}_i\right)d\hat{b}_i\\
&\leq 2C\left(q_K^2+s_K^{-2}\right)\Pr\left(\hat{\sigma}^2_\varepsilon>2\sigma^2_\varepsilon\right)\\
&\quad+2\int_{-d_K}^{d_K}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2E^{cond}_{D_i}\left[\left(\frac{\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right)+s_K}\right)^2|\hat{b}_i\right]m_{iK}\left(\hat{b}_i\right)d\hat{b}_i\\
&\quad+2\int_{-d_K}^{d_K}E^{cond}_{D_i}\left[\left(\frac{\hat{m}'_{iK}\left(\hat{b}_i\right)-\bar{m}'_K\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right)+s_K}\right)^2|\hat{b}_i\right]m_{iK}\left(\hat{b}_i\right)d\hat{b}_i\\
&\leq 2\left(q_K^2+s_K^{-2}\right)\frac{C_2}{K}\\
&\quad+\frac{2}{s_K^2}\int_{-d_K}^{d_K}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\times \qquad \text{(Term Biia)}\\
&\qquad E_{D_i}\left[\left(\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)\right)^2|\hat{b}_i\right]m_{iK}\left(\hat{b}_i\right)d\hat{b}_i\\
&\quad+\frac{2}{s_K^2}\int_{-d_K}^{d_K}E_{D_i}\left[\left(\hat{m}'_{iK}\left(\hat{b}_i\right)-\bar{m}'_K\left(\hat{b}_i\right)\right)^2|\hat{b}_i\right]m_{iK}\left(\hat{b}_i\right)d\hat{b}_i. \qquad \text{(Term Biib)}
\end{aligned}
$$

where the first inequality follows from splitting the area of integration and $2a^2+2b^2\geq(a+b)^2$ (and from the definition of $d_K$), the second inequality is due to the truncation of our score estimator and the boundedness of $\left(\bar{m}'_K(x)\right)^2$ (by (A.27)), and applying the fact that $\left(\frac{a}{b}-\frac{c}{d}\right)^2=\left(\frac{a}{b}\left(\frac{d-b}{d}\right)+\frac{a-c}{d}\right)^2\leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2+2\left(\frac{a-c}{d}\right)^2$ to the third term in the previous line. The final inequality is by Theorem 4(a) (for the first term) and is clear for the other two terms. The first term of the final line converges to zero (uniformly in $i$) due to Assumption 4, so we consider only the remaining two terms.

Term Biia satisfies

$$
\begin{aligned}
&\frac{2}{s_K^2}\int_{-d_K}^{d_K}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\times \qquad \text{(A.103)}\\
&\qquad E_{D_i}\left[\left(\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)\right)^2|\hat{b}_i\right]m_{iK}\left(\hat{b}_i\right)d\hat{b}_i
\end{aligned}
$$

$$\leq \frac{2}{s_K^2} C \begin{pmatrix} \frac{1}{h_K(K-1)} + h_K^2 + K^{-13/48} \log^2 K \\ +2K^{-13/96} h_K \log K + K^{-1/2} \log^2 K \end{pmatrix}$$
$$\times \int_{-\infty}^{\infty} \left( \frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K} \right)^2 m_{iK}\left(\hat{b}_i\right) d\hat{b}_i$$

by Lemma S-6(a), since omitting the restriction to $\hat{b} \in D_i$ can only make the expectation larger. Also, Term Biib satisfies

$$\frac{2}{s_K^2} \int_{-d_K}^{d_K} E_{D_i} \left[ \left( \hat{m}'_{iK}\left(\hat{b}_i\right) - \bar{m}'_K\left(\hat{b}_i\right) \right)^2 | \hat{b}_i, \hat{b} \in D_i \right] m_{iK}\left(\hat{b}_i\right) d\hat{b}_i$$
$$\leq \frac{2}{s_K^2} C \left( \frac{1}{h_K^3 (K-1)} + h_K^2 + K^{-5/48} \log^2 K + 2K^{-5/96} h_K \log K + K^{-1/4} \log^2 K \right)$$

by Lemma S-6(b), by the same logic. Thus we have that

$$\begin{aligned} &\text{Term Bii} \\ \leq\ & \frac{4\rho\sigma_\varepsilon^4}{s_K^2} C \begin{pmatrix} \begin{pmatrix} \frac{1}{h_K(K-1)} + h_K^2 + K^{-13/48} \log^2 K \\ +\ 2K^{-13/96} h_K \log K \ +\ K^{-1/2} \log^2 K \end{pmatrix} \\ \times \int_{-\infty}^{\infty} \left( \frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K} \right)^2 m_{iK}\left(\hat{b}_i\right) d\hat{b}_i \\ + \begin{pmatrix} \frac{1}{h_K^3(K-1)} + h_K^2 + K^{-5/48} \log^2 K \\ +\ 2K^{-5/96} h_K \log K \ +\ K^{-1/4} \log^2 K \end{pmatrix} \end{pmatrix} \\ \to\ & 0 \end{aligned} \tag{A.104}$$

by display (A.1) and Lemma S-4 (which applies upon averaging), and we are finished with the proof that $r_G\left(\hat{b}^{NSEB}, f_K\right) - r_G\left(\hat{b}^{NB}, \phi_K\right) \to 0$.

Proof of Theorem 5(b)(ii):

We show

$$\lim_{K\to\infty} \inf_{\tilde{b}} \left\{ \sup_{f_K} r_G\left(\tilde{b}, f_K\right) \right\} = r_G\left(\hat{b}^{NB}, \phi\right) \tag{A.105}$$

where the supremum is taken over the set of likelihoods satisfying the assumptions of the theorem. First,

$$\lim \inf_{K\to\infty} \inf_{\tilde{b}} \left\{ \sup_{f_K} r_G\left(\tilde{b}, f_K\right) \right\} \geq r_G\left(\hat{b}^{NB}, \phi\right) \tag{A.106}$$

follows from $\sup_{f_K} r_G\left(\tilde{b}, f_K\right) \geq r_G\left(\tilde{b}, \phi\right)$ $\forall \tilde{b}, \forall K$, so that $\forall K$,

$$\begin{aligned} \inf_{\tilde{b}} \left\{ \sup_{f_K} r_G\left(\tilde{b}, f_K\right) \right\} &\geq \inf_{\tilde{b}} \left\{ r_G\left(\tilde{b}, \phi\right) \right\} \\ &= r_G\left(\hat{b}^{NB}, \phi\right). \end{aligned} \tag{A.107}$$

Further,

$$\inf_{\tilde{b}} \left\{ \sup_{f_K} r_G\left(\tilde{b}, f_K\right) \right\} \leq \sup_{f_K} r_G\left(\hat{b}^{INB}, f_K\right) \tag{A.108}$$

so that

$$\begin{aligned}
\lim_{K\to\infty}\sup\inf_{\tilde{b}}\left\{\sup_{f_K} r_G\left(\tilde{b}, f_K\right)\right\} &\le \lim_{K\to\infty}\sup\sup_{f_K} r_G\left(\hat{b}^{INB}, f_K\right) \quad \text{(A.109)}\\
&= r_G\left(\hat{b}^{NB}, \phi\right).
\end{aligned}$$

The last equality is obviously the key to the entire result. It holds because each of the bounding constants in the proof that Terms II and III converge to zero in part (i) depends on the primitive bounding constants in the assumptions. Thus, as long as these primitive bounding constants are fixed, the bounds in the proof of part (i) hold uniformly for $f_K$ satisfying the assumptions, and the final equality above follows.

## A.7 Proof of Theorem 6

We first show that $\left|R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right)\right| \to 0$, uniformly in $\|b\|_2 \le M$ for all $f_K$ satisfying the assumptions of the theorem. (Note that setting $f_K = \phi_K$ yields the result needed for the Theorem 2.) Let $\hat{b}_i^{ISEB} = \hat{b}_i - \sigma_\varepsilon^2 \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i)+s_K}$; this is the infeasible simple empirical Bayes estimator based on the average marginal

$$\bar{m}_K(x) = \frac{1}{K}\sum_{j=1}^{K} \bar{f}_K(x - b_j)$$

(with respect to the empirical c. d. f. of the true $b_j$). Likewise, we have the modified definitions

$$\begin{aligned}
m_\phi(x) &= \frac{1}{K}\sum_{j=1}^{K}\phi(x-b_j)\\
m_{iK}(x) &= \frac{1}{K}\sum_{j=1}^{K} f_{iK}(x-b_j).
\end{aligned}$$

With these definitions, and imposing $\|b\|_2 \le M$ uniformly along the $K$ sequence, we note that Lemmas S-1 through S-6 hold uniformly along the $K$ sequence. This is a critical conclusion, and is used throughout the following. The reason we may make this observation is that the only feature of the prior used in the lemmas is that the variance is bounded, and our assumption that $\|b\|_2 \le M$ implies that the (empirical) variance of the $b_i$ is uniformly bounded along the $K$ sequence.

Now write

$$\begin{aligned}
&R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right) && \text{(A.110)}\\
=&R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{ISEB};\ f_K\right) && \text{(Term I)}\\
&+R\left(b,\ \hat{b}^{ISEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right) && \text{(Term II)}
\end{aligned}$$

We will show that both Term I and Term II go to zero.

Note that $\sup_{\|b\|_2 \le M} \limsup_{K\to\infty} R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right) < \infty$ (for any $M < \infty$, where $\|b\|^2 \equiv \frac{1}{K}\sum_{i=1}^K b_i^2$) from reasoning exactly similar to that of Lemma S-4. First, let us deal with the easier term, Term II:

$$
\begin{aligned}
&\left|R\left(b,\ \hat{b}^{ISEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right)\right| \qquad \text{(A.111)}\\
= &\left|\begin{array}{c} \rho\int \frac{1}{K}\sum_{i=1}^K \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_K\left(\hat{b} - b\right) d\hat{b} \\ -\rho\int \frac{1}{K}\sum_{i=1}^K \left(\hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right) - b_i\right)^2 \phi_K\left(\hat{b} - b\right) d\hat{b}\end{array}\right|\\
= &\left|\begin{array}{c} \frac{1}{K}\sum_{i=1}^K \rho\int \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 f_{iK}\left(\hat{b}_i - b_i\right) d\hat{b}_i \\ -\frac{1}{K}\sum_{i=1}^K \rho\int \left(\hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right) - b_i\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i\end{array}\right|\\
\le\ &\rho\frac{1}{K}\sum_{i=1}^K \int \left|\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 - \left(\hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right) - b_i\right)^2\right|\\
&\times\phi\left(\hat{b}_i - b_i\right) d\hat{b}_i\\
&+\rho\frac{1}{K}\sum_{i=1}^K \int \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2\\
&\times\left|f_{iK}\left(\hat{b}_i - b_i\right) - \phi\left(\hat{b}_i - b_i\right)\right| d\hat{b}_i\\
\le\ &\rho\frac{1}{K}\sum_{i=1}^K \int \left|\begin{array}{c} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right)\right)^2 \\ +2\left(\hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right) - b_i\right) \\ \times\left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right)\right)\end{array}\right| \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i\\
&+\rho\frac{1}{K}\sum_{i=1}^K \int_{-z_K}^{z_K} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 C K^{-1/4}\log K d\hat{b}_i\\
&+\rho\frac{1}{K}\sum_{i=1}^K \int_{\left|\hat{b}_i\right| > z_K} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2\\
&\times\left(f_{iK}\left(\hat{b}_i - b_i\right) + \phi\left(\hat{b}_i - b_i\right)\right) d\hat{b}_i\\
\le\ &\rho\frac{1}{K}\sum_{i=1}^K \int \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right)\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i\\
&+2\rho\frac{1}{K}\sum_{i=1}^K \left(\int \left(\hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right) - b_i\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i\right)^{1/2}\\
&\times\left(\int \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}^{NB}_{\tilde{G}_K,i}\left(\hat{b}_i\right)\right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i\right)^{1/2}\\
&+\rho\frac{1}{K}\sum_{i=1}^K \int_{-z_K}^{z_K} \left(\hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i\right)^2 C K^{-1/4}\log K d\hat{b}_i
\end{aligned}
$$

$$
\begin{aligned}
& +\rho \frac{1}{K} \sum_{i=1}^{K} \int_{\left|\hat{b}_i\right|>z_K} \left( \hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i \right)^2 \\
& \times \left( f_{iK}\left(\hat{b}_i - b_i\right) + \phi\left(\hat{b}_i - b_i\right) \right) d\hat{b}_i \\
\leq \quad & \rho \frac{1}{K} \sum_{i=1}^{K} \int \left( \hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) \right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i \\
& + 2\rho \left( \frac{1}{K} \sum_{i=1}^{K} \int \left( \hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) - b_i \right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i \right)^{1/2} \\
& \times \left( \frac{1}{K} \sum_{i=1}^{K} \int \left( \hat{b}_i^{ISEB}\left(\hat{b}_i\right) - \hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) \right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i \right)^{1/2} \\
& +\rho \frac{1}{K} \sum_{i=1}^{K} \int_{-z_K}^{z_K} \left( \hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i \right)^2 C K^{-1/4} \log K d\hat{b}_i \\
& +\rho \frac{1}{K} \sum_{i=1}^{K} \int_{\left|\hat{b}_i\right|>z_K} \left( \hat{b}_i^{ISEB}\left(\hat{b}_i\right) - b_i \right)^2 \\
& \times \left( f_{iK}\left(\hat{b}_i - b_i\right) + \phi\left(\hat{b}_i - b_i\right) \right) d\hat{b}_i
\end{aligned}
$$

$$
\begin{aligned}
\leq \quad & \rho \frac{\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int \left( \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K} - \frac{m_\phi'\left(\hat{b}_i\right)}{m_\phi\left(\hat{b}_i\right)} \right)^2 \phi\left(\hat{b}_i - b_i\right) db_i \\
& + 2\rho \left( \frac{1}{K} \sum_{i=1}^{K} \int \left( \hat{b}_{\tilde{G}_K,i}^{NB}\left(\hat{b}_i\right) - b_i \right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i \right)^{1/2} \\
& \times \left( \frac{\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int \left( \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K} - \frac{m_\phi'\left(\hat{b}_i\right)}{m_\phi\left(\hat{b}_i\right)} \right)^2 \phi\left(\hat{b}_i - b_i\right) d\hat{b}_i \right)^{1/2} \\
& + \rho \frac{2}{K} \sum_{i=1}^{K} \int_{-z_K}^{z_K} \left( \hat{b}_i - b_i \right)^2 C K^{-1/4} \log K d\hat{b}_i \\
& + \rho \frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-z_K}^{z_K} \left( \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K} \right)^2 C K^{-1/4} \log K d\hat{b}_i \\
& + \rho \frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{\left|\hat{b}_i\right|>z_K} \left( \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right) + s_K} \right)^2 \\
& \times \left( f_{iK}\left(\hat{b}_i - b_i\right) + \phi\left(\hat{b}_i - b_i\right) \right) d\hat{b}_i \\
& + \rho \frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{\left|\hat{b}_i\right|>z_K} \left( \hat{b}_i - b_i \right)^2
\end{aligned}
$$

$$\times \left( f_{iK}\left(\hat{b}_i - b_i\right) + \phi\left(\hat{b}_i - b_i\right)\right) d\hat{b}_i$$

The second equality comes from the fact that both $\hat{b}^{NB}_{\tilde{G}_K,i}$ and $\hat{b}^{ISEB}_i$ depend only on $\hat{b}_i$ and not on $\hat{b}_{-i}$, the first inequality is trivial, the second inequality follows from $a^2 - b^2 = (a-b)^2 + 2b(a-b)$ applied to the integrand of the first term of the preceding expression, splitting up the area of integration of the second term of the preceding expression, and noting that $\left| f_{iK}\left(\hat{b}_i - b_i\right) - \phi\left(\hat{b}_i - b_i\right)\right| \leq f_{iK}\left(\hat{b}_i - b_i\right) + \phi\left(\hat{b}_i - b_i\right)$, the third inequality follows from applying the triangle inequality and then Hölder's inequality to the first term of the preceding display, the fourth inequality follows from applying the Cauchy-Schwarz inequality to the second term of the preceding display, the fifth inequality follows from rewriting the first two terms of the preceding display and using $(a+b)^2 \leq 2a^2 + 2b^2$ on the last two terms of the preceding display, and the last expression converges to zero by the Dominated Convergence Theorem through Lemma S-4 (for the first two terms), and $z_K = s_K^{-2}$ (for the remaining terms) along with Chebyshev's inequality and the observation that $\left(\frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 \leq C s_K^{-2}$ for sufficiently large $K$ by Lemma S-3(e), the boundedness of $m'_\phi$, and the nonnegativity of $\bar{m}_K$.

Now we shall demonstrate that Term I converges to zero. Our proof is in the spirit of Bickel *et al.* (1993, p. 405 ff) and van der Vaart (1988, p. 169 ff), but we extend their approaches to handle cross-sectional dependence of the $\hat{b}_i$ and to deal with a nonconstant sequence of likelihoods.

Note that, from (2.10) through (2.13) of the body of this paper, we can write $\hat{l}_{iK}(x) = \hat{l}\left(\hat{b}_i\,;\,\hat{b}_{-i}\right)$, where the function $\hat{l}(\cdot,\cdot)$ does not depend on $i$. This representation, and a similar representation for $\breve{l}$, are adopted here. Although $\hat{\sigma}^2_\varepsilon$ depends on the full data, rather than only $\hat{b}$, this dependence is suppressed for notational convenience; the treatment below does, however, account for the fact that $\hat{\sigma}^2_\varepsilon$ is a function of the full data (see display (A.116)). Now,

$$\begin{aligned}
& R\left(b,\, \hat{b}^{ISEB};\, f_K\right) - R\left(b,\, \hat{b}^{NSEB};\, f_K\right) \qquad \text{(A.112)} \\
= \;& \frac{1}{K}\sum_{i=1}^{K} \rho \int \begin{bmatrix} \left(\hat{b}^{ISEB}_i\left(\hat{b}_i\right) - b_i\right)^2 \\ -\left(\hat{b}^{NSEB}_i\left(\hat{b}\right) - b_i\right)^2 \end{bmatrix} f_K\left(\hat{b} - b\right) d\hat{b} \\
= \;& \frac{1}{K}\sum_{i=1}^{K} \rho \int \begin{bmatrix} \left(\hat{b}_i - b_i - \sigma^2_\varepsilon \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 - \\ \left(\hat{b}_i - b_i - \hat{\sigma}^2_\varepsilon \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right)\right)^2 \end{bmatrix} f_K\left(\hat{b} - b\right) d\hat{b} \\
= \;& \frac{1}{K}\sum_{i=1}^{K} \rho \int \begin{bmatrix} -\left(\hat{\sigma}^2_\varepsilon \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma^2_\varepsilon \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)^2 + \\ 2\left(\hat{b}_i - b_i - \sigma^2_\varepsilon \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right)\left(\hat{\sigma}^2_\varepsilon \hat{l}\left(\hat{b}_i; \hat{b}_{-i}\right) - \sigma^2_\varepsilon \frac{\bar{m}'_K(\hat{b}_i)}{\bar{m}_K(\hat{b}_i) + s_K}\right) \end{bmatrix} \\
& \times f_K\left(\hat{b} - b\right) d\hat{b}
\end{aligned}$$

$$
\begin{aligned}
= & -\frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\;+\;s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b} \\
& +\frac{1}{K}\sum_{i=1}^{K}\rho\int 2\left(\hat{b}_i-b_i-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\;+\;s_K}\right)\times \\
& \qquad \left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\;+\;s_K}\right) f_K\left(\hat{b}-b\right)d\hat{b} \\
\leq & -\frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\;+\;s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b} \\
& +\frac{2}{K}\sum_{i=1}^{K}\rho\left[\begin{array}{c}\left(\int\left(\hat{b}_i-b_i-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}\right)^{\frac{1}{2}}\times \\ \left(\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}\right)^{\frac{1}{2}}\end{array}\right] \\
\leq & -\frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\;+\;s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b} \\
& +2\left\{\frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\hat{b}_i-b_i-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\;+\;s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}\right\}^{\frac{1}{2}}\times \\
& \left\{\frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\;+\;s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}\right\}^{\frac{1}{2}}
\end{aligned}
$$

In the above, the first and second equalities are by definition; the third is a consequence of the fact that $a^2 - b^2 \;=\; 2a\,(a-b) - (a-b)^2$; the fourth equality simply breaks out the two terms of the integrand; the first inequality is an application of Hölder's inequality, and the second is an application of the Cauchy-Schwarz inequality.

From the final bound given above, we see that if we can show that

$$
\begin{aligned}
& \frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\begin{array}{c}\hat{\sigma}_{\varepsilon}^{2}\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) \\ -\sigma_{\varepsilon}^{2}\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\end{array}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b} \\
\to & \;\; 0
\end{aligned}
\tag{A.113}
$$

then we will have demonstrated that Term I converges to zero, because

$$\begin{aligned}
&\frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\hat{b}_i - b_i - \sigma_\varepsilon^2\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 f_K\left(\hat{b}-b\right)d\hat{b} \quad \text{(A.114)}\\
=\ & R\left(b,\ \hat{b}^{ISEB};\ f_K\right)\\
<\ & \infty \text{ uniformly along the } K,\ T \text{ sequence}
\end{aligned}$$

from the fact that Term II $= R\left(b,\ \hat{b}^{ISEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right)$ converges to zero, and, as noted above, $\sup_{\|b\|\leq M}\limsup_{K\to\infty} R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right)\ <\ \infty$.

Using $(a+b)^2\ \leq\ 2a^2\ +\ 2b^2$, and adding and subtracting $\sigma_\varepsilon^2\hat{l}\left(\hat{b}_i\ ;\ \hat{b}_{-i}\right)$, we have

$$\begin{aligned}
&\frac{1}{K}\sum_{i=1}^{K}\rho\int\left(\hat{\sigma}_\varepsilon^2\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \sigma_\varepsilon^2\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 \times \quad \text{(A.115)}\\
&\qquad f_K\left(\hat{b}-b\right)d\hat{b}\\
\leq\ & \frac{2}{K}\sum_{i=1}^{K}\rho\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2\hat{l}^2\left(\hat{b}_i;\hat{b}_{-i}\right)f_K\left(\hat{b}-b\right)d\hat{b} \quad \text{(Term A)}\\
&+\frac{2\sigma_\varepsilon^4}{K}\sum_{i=1}^{K}\rho\int\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right) - \frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 \times \quad \text{(Term B)}\\
&\qquad f_K\left(\hat{b}-b\right)d\hat{b}
\end{aligned}$$

which separates the problem of nonparametric score estimation from the problem of estimating the residual variance. Now, Term A satisfies

$$\begin{aligned}
&\frac{2}{K}\sum_{i=1}^{K}\rho\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2\hat{l}^2\left(\hat{b}_i;\hat{b}_{-i}\right)f_K\left(\hat{b}-b\right)d\hat{b} \quad \text{(A.116)}\\
\leq\ & \frac{2\rho q_K^2}{K}\sum_{i=1}^{K}\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}\\
=\ & 2\rho q_K^2\int\left(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\right)^2 f_K\left(\hat{b}-b\right)d\hat{b}\\
=\ & 2\rho q_K^2 Var\left[\hat{\sigma}_\varepsilon^2\right]\\
\leq\ & 2\rho q_K^2\frac{C}{K}\\
\to\ & 0
\end{aligned}$$

where the first inequality comes from the truncation of our estimator $\hat{l}$ of the score, and the second inequality comes from Theorem 4(a). The convergence of the final bound to zero is by construction: by display (A.1), $\frac{q_K^2}{K}\ \to\ 0$. Thus, Term A converges to zero.

Now consider Term B. Define

$$D_i \equiv \left\{ \hat{b} : \left| \hat{b}_i \right| \leq \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \text{ and } \left| \breve{l}\left( \hat{b}_i ; \hat{b}_{-i} \right) \right| \leq q_K \right\},$$

and let $E_{D_i}[(\cdot)] \equiv \int_{\hat{b} \in D_i} (\cdot) f_K \left( \hat{b} - b \right) d\hat{b}$ (so that the area of integration is restricted, but in a way which may differ for each $i$). Now,

$$\frac{2\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \rho \int \left( \hat{l}\left( \hat{b}_i ; \hat{b}_{-i} \right) - \frac{\bar{m}_K'\left( \hat{b}_i \right)}{\bar{m}_K\left( \hat{b}_i \right) + s_K} \right)^2 f_K\left( \hat{b} - b \right) d\hat{b} \tag{A.117}$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int E\left[ \left( \hat{l}\left( \hat{b}_i ; \hat{b}_{-i} \right) - \frac{\bar{m}_K'\left( \hat{b}_i \right)}{\bar{m}_K\left( \hat{b}_i \right) + s_K} \right)^2 \mid \hat{b}_i, b \right] f_{iK}\left( \hat{b}_i - b_i \right) d\hat{b}_i$$

$$\leq \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left( \frac{\bar{m}_K'\left( \hat{b}_i \right)}{\bar{m}_K\left( \hat{b}_i \right) + s_K} \right)^2 f_{iK}\left( \hat{b}_i - b_i \right) \times \tag{Term Bi}$$

$$\left( \Pr\left( \left| \breve{l}\left( \hat{b}_i ; \hat{b}_{-i} \right) \right| > q_K \mid \hat{b}_i, b \right) + \Pr\left( \left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid \hat{b}_i, b \right) \right) d\hat{b}_i$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} E_{D_i}\left[ \left( \hat{l}\left( \hat{b}_i ; \hat{b}_{-i} \right) - \frac{\bar{m}_K'\left( \hat{b}_i \right)}{\bar{m}_K\left( \hat{b}_i \right) + s_K} \right)^2 \mid b \right] \tag{Term Bii}$$

where the first equality is by definition, and the inequality follows from the truncation of the estimated score function according to the definition of $D_i$.

Consider Term Bi first.

$$\frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left( \frac{\bar{m}_K'\left( \hat{b}_i \right)}{\bar{m}_K\left( \hat{b}_i \right) + s_K} \right)^2 f_{iK}\left( \hat{b}_i - b_i \right) \times \tag{A.118}$$

$$\left( \begin{array}{c} \Pr\left( \left| \breve{l}\left( \hat{b}_i ; \hat{b}_{-i} \right) \right| > q_K \mid \hat{b}_i, b \right) + \\ \Pr\left( \left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid \hat{b}_i, b \right) \end{array} \right) d\hat{b}_i$$

$$= \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left( \frac{\bar{m}_K'\left( \hat{b}_i \right)}{\bar{m}_K\left( \hat{b}_i \right) + s_K} \right)^2 f_{iK}\left( \hat{b}_i - b_i \right)$$

$$\times \Pr\left( \left| \hat{b}_i \right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128} \log K} \mid \hat{b}_i, b \right) d\hat{b}_i$$

$$+ \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^{K} \int_{-\infty}^{\infty} \left( \frac{\bar{m}_K'\left( \hat{b}_i \right)}{\bar{m}_K\left( \hat{b}_i \right) + s_K} \right)^2 f_{iK}\left( \hat{b}_i - b_i \right) \times$$

$$
\begin{aligned}
& \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i, b\right) d\hat{b}_i \\
\leq\ & \frac{2\rho\sigma_\varepsilon^4}{Ks_K^2} C \sum_{i=1}^K \Pr\left(\left|\hat{b}_i\right| > \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{128}\log K} \mid b\right) \\
& + \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^K \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)\ +\ s_K}\right)^2 f_{iK}(x-b_i)\,dx \\
& + \sup_{i,\left|\hat{b}_i\right|\leq d_K} \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i, b\right) \\
& \times \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^K \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)\ +\ s_K}\right)^2 f_{iK}(x-b_i)\,dx \\
\leq\ & \frac{2\rho\sigma_\varepsilon^4}{Ks_K^2} C \sum_{i=1}^K \left\{ \begin{array}{c} \Pr\left(\left|\hat{b}_i\right| > \sqrt{\frac{\sigma_\varepsilon^2}{256}\log K}\right) \\ +\Pr\left(\hat{\sigma}_\varepsilon^2 > \frac{1}{2}\sigma_\varepsilon^2\right) \end{array} \right\} \\
& + \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^K \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)\ +\ s_K}\right)^2 f_{iK}(x-b_i)\,dx \\
& + \sup_{i,\left|\hat{b}_i\right|\leq d_K} \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i, b\right) \\
& \times \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^K \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)\ +\ s_K}\right)^2 f_{iK}(x-b_i)\,dx \\
\leq\ & C_2 s_K^{-2} \left\{\frac{1}{\log K} + \frac{1}{K}\right\} \\
& + \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^K \int_{|x|>d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)\ +\ s_K}\right)^2 f_{iK}(x-b_i)\,dx \\
& + \sup_{i,\left|\hat{b}_i\right|\leq d_K} \Pr\left(\left|\breve{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right| > q_K \mid \hat{b}_i, b\right) \\
& \times \frac{2\rho\sigma_\varepsilon^4}{K} \sum_{i=1}^K \int_{-d_K}^{d_K} \left(\frac{\bar{m}_K'(x)}{\bar{m}_K(x)\ +\ s_K}\right)^2 f_{iK}(x-b_i)\,dx
\end{aligned}
$$

The equality is trivial; the first inequality follows from the boundedness of $\left(\bar{m}_K'(x)\right)^2$ (by (A.27)), the second inequality follows by $\Pr(A) \leq \Pr(B) + \Pr(C)$ whenever $A \subset (B \cap C)$, and the third inequality follows from Chebyshev's inequality and the variance bound of Theorem 4(a) (along with $\sigma_\varepsilon^2 > 0$). Of the terms in the final expression, we see immediately that the first term converges to zero by Assumption 4 and the second converges to zero by the uniform integrability of its integrand, which was shown in the course of the proof that Term II converges to zero, and by $d_K \to \infty$. The third term converges to zero by Lemmas S-4 and S-6(c). Thus, Term Bi converges to zero.

Finally, turn to Term Bii. Consider the $i^{th}$ term of the average which makes up

Term Bii. We define

$$E_{D_i}^{out}\left[(\cdot)\mid b\right] \equiv \int_{\hat{b}\in D_i\cap\{\hat{\sigma}_\varepsilon^2>2\sigma_\varepsilon^2\}} (\cdot)\, f_K\left(\hat{b}\mid b\right) d\hat{b}$$

$$E_{D_i}^{cond}\left[(\cdot)\mid \hat{b}_i,\ b\right] \equiv \int_{\hat{b}\in D_i} (\cdot)\, f_K\left(\hat{b}_{-i}\mid \hat{b}_i, b\right) d\hat{b}_{-i}$$

(so that both the probability measure and the area of integration depend on $\hat{b}_i$). Since $\left(\frac{a}{b}-\frac{c}{d}\right)^2=\left(\frac{a}{b}\left(\frac{d-b}{d}\right)+\frac{a-c}{d}\right)^2\leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2+2\left(\frac{a-c}{d}\right)^2$ we have:

$$E_{D_i}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\mid b\right] \tag{A.119}$$

$$\leq 2E_{D_i}^{out}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)\right)^2\mid b\right]+2E_{D_i}^{out}\left[\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\mid b\right]$$

$$+\int_{-d_K}^{d_K} E_{D_i}^{cond}\left[\left(\hat{l}\left(\hat{b}_i;\hat{b}_{-i}\right)-\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2|\hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i$$

$$\leq 2C\left(q_K^2+s_K^{-2}\right)\Pr\left(\hat{\sigma}_\varepsilon^2>2\sigma_\varepsilon^2\right)$$

$$+2\int_{-d_K}^{d_K}\left\{\begin{array}{c}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\\ E_{D_i}^{cond}\left[\left(\frac{\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right)+s_K}\right)^2|\hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right)\end{array}\right\}d\hat{b}_i$$

$$+2\int_{-d_K}^{d_K} E_{D_i}^{cond}\left[\left(\frac{\hat{m}_{iK}'\left(\hat{b}_i\right)-\bar{m}_K'\left(\hat{b}_i\right)}{\hat{m}_{iK}\left(\hat{b}_i\right)+s_K}\right)^2|\hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i$$

$$\leq 2\left(q_K^2+s_K^{-2}\right)\frac{C_2}{K}$$

$$+\frac{2}{s_K^2}\int_{-d_K}^{d_K}\left(\frac{\bar{m}_K'\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2\times \tag{Term Biia}$$

$$E_{D_i}\left[\left(\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)\right)^2|\hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i$$

$$+\frac{2}{s_K^2}\int_{-d_K}^{d_K} E_{D_i}\left[\left(\hat{m}_{iK}'\left(\hat{b}_i\right)-\bar{m}_K'\left(\hat{b}_i\right)\right)^2|\hat{b}_i, b\right] f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i. \tag{Term Biib}$$

where the first inequality follows from splitting the area of integration and $2a^2+2b^2\geq(a+b)^2$ (and from the definition of $d_K$), the second inequality is due to the truncation of our score estimator and the boundedness of $\left(\bar{m}_K'(x)\right)^2$ (by (A.27)), and applying the fact that $\left(\frac{a}{b}-\frac{c}{d}\right)^2=\left(\frac{a}{b}\left(\frac{d-b}{d}\right)+\frac{a-c}{d}\right)^2\leq 2\left(\frac{a}{b}\right)^2\left(\frac{d-b}{d}\right)^2+2\left(\frac{a-c}{d}\right)^2$ to the third term in the previous line. The final inequality is by Theorem 4(a) (for the first term) and is clear for the other two terms. The first term of the final line converges to zero (uniformly in $i$) due to Assumption 4, so we consider only the remaining two terms.

Term Biia satisfies

$$\begin{aligned} & \frac{2}{s_K^2}\int_{-d_K}^{d_K}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 \times \\ & \qquad E_{D_i}\left[\left(\bar{m}_K\left(\hat{b}_i\right)-\hat{m}_{iK}\left(\hat{b}_i\right)\right)^2|\hat{b}_i,b\right]f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i \\ \leq\ & \frac{2}{s_K^2}C\left(\begin{array}{c}\frac{1}{h_K(K-1)}+h_K^2+K^{-13/48}\log^2 K \\ +\ 2K^{-13/96}h_K\log K\ +\ K^{-1/2}\log^2 K\end{array}\right) \\ & \times\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)\ +\ s_K}\right)^2 f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i \end{aligned} \tag{A.120}$$

by Lemma S-6(a), since omitting the restriction to $\hat{b}\in D_i$ can only make the expectation larger, and conditioning on any $b$ such that $\|b\|^2\leq M<\infty$ will not change the results of that lemma, as can be easily verified. Also, Term Biib satisfies

$$\begin{aligned} & \frac{2}{s_K^2}\int_{-d_K}^{d_K}E_{D_i}\left[\left(\hat{m}'_{iK}\left(\hat{b}_i\right)-\bar{m}'_K\left(\hat{b}_i\right)\right)^2|\hat{b}_i,\hat{b}\in D_i,b\right]f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i \\ \leq\ & \frac{2}{s_K^2}C\left(\begin{array}{c}\frac{1}{h_K^3(K-1)}+h_K^2+K^{-5/48}\log^2 K \\ +\ 2K^{-5/96}h_K\log K\ +\ K^{-1/4}\log^2 K\end{array}\right) \end{aligned}$$

by Lemma S-6(b), by the same logic. Thus we have that

$$\begin{aligned} & \text{Term Bii} \\ \leq\ & \frac{4\rho\sigma_\varepsilon^4}{s_K^2}C\left(\begin{array}{c}\left(\begin{array}{c}\frac{1}{h_K(K-1)}+h_K^2+K^{-13/48}\log^2 K \\ +\ 2K^{-13/96}h_K\log K\ +\ K^{-1/2}\log^2 K\end{array}\right) \\ \times\int_{-\infty}^{\infty}\left(\frac{\bar{m}'_K\left(\hat{b}_i\right)}{\bar{m}_K\left(\hat{b}_i\right)+s_K}\right)^2 f_{iK}\left(\hat{b}_i-b_i\right)d\hat{b}_i \\ +\left(\begin{array}{c}\frac{1}{h_K^3(K-1)}+h_K^2+K^{-5/48}\log^2 K \\ +\ 2K^{-5/96}h_K\log K\ +\ K^{-1/4}\log^2 K\end{array}\right)\end{array}\right) \\ \to\ & 0 \end{aligned} \tag{A.121}$$

by display (A.1) and Lemma S-4 (after averaging), and we are finished with the proof that

$$\left|R\left(b,\ \hat{b}^{NSEB};\ f_K\right)-R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right)\right|\ \to\ 0. \tag{A.122}$$

Examination of the proof shows that this results holds uniformly over all $f_K$ satisfying Assumption 6 through Assumption 9 with the same constants and over all $b$ is the set $\|b\|_2 \leq M$. (The only restriction on $b$ was that the "prior" $\widetilde{G}_K$ had a finite variance.) Thus,

$$\lim_{K\to\infty} \sup_{\|b\|_2\leq M} \left| R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right)\right| = 0 \tag{A.123}$$

and

$$\lim_{K\to\infty} \sup_{\|b\|_2\leq M} \sup_{f_K} \left| R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right)\right| = 0. \tag{A.124}$$

Now

$$\begin{aligned}
& \lim_{K\to\infty} \sup_{\|b\|_2\leq M} \sup_{f_K} \left| R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - \inf_{\tilde{b}\in\mathcal{B}} R(b,\widetilde{b},\phi_K)\right| && \text{(A.125)} \\
\leq\ & \lim_{K\to\infty} \sup_{\|b\|_2\leq M} \sup_{f_K} \left| R\left(b,\ \hat{b}^{NSEB};\ f_K\right) - R(b,\hat{b}^{NB}_{\tilde{G}_K};\ \phi_K)\right| \\
& + \lim_{K\to\infty} \sup_{\|b\|_2\leq M} \left| R\left(b,\ \hat{b}^{NB}_{\tilde{G}_K};\ \phi_K\right) - R(b,\hat{b}^{NSEB};\ \phi_K)\right| \\
& + \lim_{K\to\infty} \sup_{\|b\|_2\leq M} \left| R\left(b,\ \hat{b}^{NSEB};\ \phi_K\right) - \inf_{\tilde{b}\in\mathcal{B}} R(b,\widetilde{b},\phi_K)\right| \\
=\ & 0,
\end{aligned}$$

where the final equality obtains from (A.124) applied the first term after the inequality, (A.123) with $f_K = \phi_K$ applied to the second term, and Theorem 2(b) applied to the final term (note that Theorem 2(b) relies on the limit (A.122), but that the limit (A.122) is proven above without reference to Theorem 2(b), so our logic here is not circular).

## A.8 Including Additional Regressors

Section 4.3 of the text considers the regression model $y = U\gamma + X\beta + \varepsilon$, where $U$ is a $T \times L$ matrix of regressors, $\gamma$ is a $L \times 1$ vector of regressor coefficients, and $L$ is fixed as $T, K \to \infty$ (where $K = \rho T$). When $U^T X = 0$, then $\hat{b} - b = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t \varepsilon_t$, so that Theorem 1 through Theorem 6 continue to hold without modification. The following theorem shows that the results in Theorem 1 and Theorem 2 continue to hold when $U^T X \neq 0$ if the regressors are nicely behaved.

**Theorem 7**: Let $T_0$, $C_0$, and $C_1$ denote finite constants. Suppose that, for $T > T_0$, (i) $E\left\{\text{tr}\left[\left(\frac{1}{T}Z^T M_X Z\right)^{-1}\right]\right\} \leq C_0$ and (ii) $E\left\{\text{tr}\left[\left(\frac{1}{T}Z^T P_X Z\right)\left(\frac{1}{T}Z^T M_X Z\right)^{-1}\right]\right\} \leq C_1$, where $P_X = X\left(X^T X\right)^{-1} X^T$ and $M_X = I - P_X$. Then in the Gaussian regression model,

(a) given Assumption 1 through Assumption 4,

$$r_G\left(\hat{b}^{PEB}_{\hat{\gamma}}\right) - r_G\left(\hat{b}^{NB}_{\gamma}\right) \ \to\ 0; \tag{A.126}$$

(b) given Assumption 1, Assumption 2, and Assumption 4,

$$r_G\left(\hat{b}_{\hat{\gamma}}^{NSEB}\right) - r_G\left(\hat{b}_{\gamma}^{NB}\right) \rightarrow 0; \tag{A.127}$$

(c) given Assumption 1 and Assumption 4,

$$\sup_{\|b\|_2 \le M}\left|R\left(b, \hat{b}_{\hat{\gamma}}^{NSEB}\right) - \inf_{\tilde{b}_\gamma \in \mathcal{B}} R\left(b, \tilde{b}_\gamma\right)\right| \rightarrow 0 \tag{A.128}$$

for all $M < \infty$, where $\mathcal{B}$ is defined as in Theorem 2.

*Proof of Theorem 7:* The proof of this theorem utilizes the results in Theorem 5 and Theorem 6. Notice that the results in Theorem 7 are special cases of the results listed in Theorem 5 and Theorem 6, and thus we just need to verify that all of the steps in those theorems continue to hold under the assumptions of this theorem. Comparing the conditions, we see that every assumption is satisfied except the assumption

$$\frac{1}{K-1}\sum_{\substack{j=1,\\ j\neq i}}^{K}\sum_{\substack{n=1,\\ n\neq i}}^{K}\tau\left(j,n|i\right) \le C$$

for some absolute constant $C$, where

$$\tau\left(j,n|i\right) = \sup_{f,g\in L^2(P)}\left|Corr\left(f\left(\widehat{b_j}\right), g\left(\widehat{b_n}\right)|\widehat{b_i}\right)\right|.$$

Note that the above is actually a relaxation of the assumption used, and is the condition actually relevant in display (A.51). Since $\left(\widehat{b_j}, \widehat{b_n}|\widehat{b_i}\right)$ are jointly Gaussian (by the joint Gaussianity of $\left(\widehat{b_j}, \widehat{b_n}, \widehat{b_i}\right)$), we may apply Rozanov (1967, Theorem 10.1 [page 181]) to conclude that

$$\tau\left(j,n|i\right) = \left|Corr\left(\widehat{b_j}, \widehat{b_n}|\widehat{b_i}\right)\right|.$$

Thus, our problem is reduced to that of proving the summability of the absolute values of the elements of the correlation matrix, conditional on $\widehat{b_i}$, of the $\widehat{b_{-i}}$. First let us consider the matrix $Cov\left(\widehat{b_{-i}}|\widehat{b_i}\right)$. From the familiar formula for conditioning under joint Gaussianity, we will be able to obtain this readily as soon as we have found a simple form for $Cov\left(\widehat{b}\right)$. To simplify notation, we assume $\sigma_\varepsilon^2 = 1$ in the following.

$$Cov\left(\widehat{b}\right) = E\left[I_K + \frac{X^T U}{T}\left(\frac{U^T M_X U}{T}\right)^{-1}\frac{U^T X}{T}\right], \tag{A.129}$$

where the equality follows from Theil (1971, Exercise 7.3, page 146), using $\frac{X^T X}{T} = I_K$.

Now we are ready to apply the standard Gaussian conditioning formula to obtain:

$$
\begin{aligned}
& Cov\left(\widehat{b_{-i}}|\widehat{b_i}\right) \\
= \; & E\left[\begin{array}{c} I_{K-1} + \frac{X_{-i}^T U}{T}\left(\frac{U^T M_X U}{T}\right)^{-1}\frac{U^T X_{-i}}{T} \\ -\frac{1}{1+r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_i}\left[\begin{array}{c} r_1^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_1 \\ \vdots \\ r_{i-1}^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_{i-1} \\ r_{i+1}^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_{i+1} \\ \vdots \\ r_K^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_K \end{array}\right]\left[\begin{array}{c} r_1^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_1 \\ \vdots \\ r_{i-1}^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_{i-1} \\ r_{i+1}^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_{i+1} \\ \vdots \\ r_K^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_K \end{array}\right]^T \end{array}\right]
\end{aligned}
$$

where $X_{-i}$ is the $X$ matrix with column $i$ deleted, and $r_j = \frac{1}{T}\sum_{t=1}^{T} x_{jt}u_t$. By the Rozanov result cited above, it suffices to demonstrate that

$$
\frac{1}{K-1}\sum_{\substack{j=1,\\ j\neq i}}^{K}\sum_{\substack{n=1,\\ n\neq i}}^{K}\frac{\left|Cov\left(\widehat{b_j},\widehat{b_n}|\widehat{b_i}\right)\right|}{\sqrt{Var\left(\widehat{b_j}|\widehat{b_i}\right)Var\left(\widehat{b_n}|\widehat{b_i}\right)}} \leq C
$$

for some absolute constant $C$. To do so, first observe that

$$
\begin{aligned}
& Var\left(\widehat{b_j}|\widehat{b_i}\right) && \text{(A.130)} \\
= \; & E\left[1 + r_j^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_j - \frac{\left(r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_j\right)^2}{1 + r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_i}\right] \\
\geq \; & E\left[\begin{array}{c} 1 + r_j^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_j \\ -\frac{r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_i r_j^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_j}{1+r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_i} \end{array}\right] && \text{(A.131)} \\
\geq \; & E\left[1 + r_j^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_j - r_j^T\left(\frac{U^T M_X U}{T}\right)^{-1} r_j\right] && \text{(A.132)} \\
= \; & 1 && \text{(A.133)}
\end{aligned}
$$

where the first inequality follows from the generalized Cauchy-Schwarz inequality (see Magnus and Neudecker (1999, page 200, problem 1)) and the fact that $\left(\frac{U^T M_X U}{T}\right)^{-1}$ is positive semidefinite. The second inequality comes from simple calculation. The above display shows that

$$\frac{1}{K-1}\sum_{\substack{j=1,\\ j\neq i}}^{K}\sum_{\substack{n=1,\\ n\neq i}}^{K}\frac{\left|Cov\left(\widehat{b_j},\widehat{b_n}|\widehat{b_i}\right)\right|}{\sqrt{Var\left(\widehat{b_j}|\widehat{b_i}\right)Var\left(\widehat{b_n}|\widehat{b_i}\right)}}$$

$$\leq \frac{1}{K-1}\sum_{\substack{j=1,\\ j\neq i}}^{K}\sum_{\substack{n=1,\\ n\neq i}}^{K}\left|Cov\left(\widehat{b_j},\widehat{b_n}|\widehat{b_i}\right)\right|.$$

It will be convenient to define $Abs\left(A\right)=\left\{\left|a_{ij}\right|\right\}$. Then

$$\frac{1}{K-1}\sum_{\substack{j=1,\\ j\neq i}}^{K}\sum_{\substack{n=1,\\ n\neq i}}^{K}\left|Cov\left(\widehat{b_j},\widehat{b_n}|\widehat{b_i}\right)\right| \tag{A.134}$$

$$= \frac{1}{K-1}\underline{1}^T Abs\left(Cov\left(\widehat{b_{-i}}|\widehat{b_i}\right)\right)\underline{1}$$

$$\leq \frac{1}{K-1}E\left[\begin{array}{c}\underline{1}^T Abs\left(I_{K-1}\right)\underline{1}+\\ \underline{1}^T Abs\left(\frac{X_{-i}^T U}{T}\left(\frac{U^T M_X U}{T}\right)^{-1}\frac{U^T X_{-i}}{T}\right)\underline{1}\end{array}\right] \tag{A.135}$$

$$+ \frac{1}{K-1}E\left[\underline{1}^T Abs\left(\frac{1}{1+r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1}r_i}R_{-i}R_{-i}^T\right)\underline{1}\right]$$

where $R_{-i}^T=\left[\begin{array}{ccc} r_1^T\left(\frac{U^T M_X U}{T}\right)^{-1}r_1 & \cdots & r_K^T\left(\frac{U^T M_X U}{T}\right)^{-1}r_K\end{array}\right]$ with the $i^{th}$ element removed, the equality is by the definition of the function $Abs\left(\cdot\right)$ and the inequality follows from the triangle inequality applied elementwise and from the distributive property of matrix multiplication (and, of course, from the fact that the expectation of the absolute value exceeds the absolute value of the expectation).

Now, $\frac{1}{K-1}E\left[\underline{1}^T Abs\left(I_{K-1}\right)\underline{1}\right]=\frac{1}{K-1}\underline{1}^T I_{K-1}\underline{1}=\frac{K-1}{K-1}=1$, since $I_{K-1}$ has only 1 and 0 as elements. Further,

$$\frac{1}{K-1}E\left[\underline{1}^T Abs\left(\frac{1}{1+r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1}r_i}R_{-i}R_{-i}^T\right)\underline{1}\right] \tag{A.136}$$

$$= \frac{1}{K-1}E\left[\begin{array}{c}\frac{1}{1+r_i^T\left(\frac{U^T M_X U}{T}\right)^{-1}r_i}\\ \times\left(\sum_{\substack{j=1,\\ j\neq i}}^{K}\left|r_j^T\left(\frac{U^T M_X U}{T}\right)^{-1}r_i\right|\right)^2\end{array}\right] \tag{A.137}$$

$$\leq \frac{1}{K-1} E \left[ \begin{array}{c} \frac{K-1}{1 + r_i^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_i} \\ \times \sum^K_{\substack{j=1, \\ j \neq i}} \left( r_j^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_i \right)^2 \end{array} \right] \tag{A.138}$$

$$\leq E \left[ \begin{array}{c} \frac{1}{1 + r_i^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_i} \\ \times \sum^K_{\substack{j=1, \\ j \neq i}} \left( r_j^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_j \right) \left( r_i^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_i \right) \end{array} \right] \tag{A.139}$$

$$\leq E \left[ \sum^K_{\substack{j=1, \\ j \neq i}} r_j^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_j \right] \tag{A.140}$$

$$\leq E \left[ \sum_{j=1}^K r_j^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_j \right] \tag{A.141}$$

$$= E \left[ tr \left( \frac{U^T P_X U}{T} \left( \frac{U^T M_X U}{T} \right)^{-1} \right) \right] \tag{A.142}$$

$$\leq C_1. \tag{A.143}$$

In the above display, the first equality is due to matrix multiplication, $|ab| = |a|\,|b|$, and

$$1 + r_i^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_i \ > \ 0. \tag{A.144}$$

The first inequality is by Cauchy-Schwarz, the second inequality is by the cancellation of $K-1$ and $\frac{1}{K-1}$ and by the generalized Cauchy-Schwarz inequality (see Magnus and Neudecker (1999, page 200, problem 1)), the third inequality is by the fact that $\frac{a}{1+a} \leq 1$ for $a \geq 0$, and the fourth inequality is due to the fact that $r_i^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_i \geq 0$. The second equality is by the usual method of rearranging the order of matrix multiplication within a trace and the definition of the $r_j$ (we also exploit the fact that $\frac{X^T X}{T} = I_K$), and the final inequality is by the assumption of the theorem. Finally,

$$\frac{1}{K-1} E \left[ \underline{1}^T Abs \left( \frac{X_{-i}^T U}{T} \left( \frac{U^T M_X U}{T} \right)^{-1} \frac{U^T X_{-i}}{T} \right) \underline{1} \right] \tag{A.145}$$

$$= \frac{1}{K-1} E \left[ \sum^K_{\substack{j=1, \\ j \neq i}} \sum^K_{\substack{n=1, \\ n \neq i}} \left| r_j^T \left( \frac{U^T M_X U}{T} \right)^{-1} r_n \right| \right] \tag{A.146}$$

$$\leq \frac{1}{K-1} E\left[ \sum_{\substack{j=1,\\ j\neq i}}^{K} \sum_{\substack{n=1,\\ n\neq i}}^{K} \left\{ \begin{array}{c} \sqrt{r_j^T \left(\frac{U^T M_X U}{T}\right)^{-1} r_j} \\ \times \sqrt{r_n^T \left(\frac{U^T M_X U}{T}\right)^{-1} r_n} \end{array} \right\} \right] \quad (A.147)$$

$$= E\left[ \left( \sum_{\substack{j=1,\\ j\neq i}}^{K} \sqrt{\frac{r_j^T \left(\frac{U^T M_X U}{T}\right)^{-1} r_j}{K-1}} \right)^2 \right] \quad (A.148)$$

$$\leq E\left[ (K-1) \sum_{\substack{j=1,\\ j\neq i}}^{K} \frac{r_j^T \left(\frac{U^T M_X U}{T}\right)^{-1} r_j}{K-1} \right] \quad (A.149)$$

$$= E\left[ \sum_{\substack{j=1,\\ j\neq i}}^{K} r_j^T \left(\frac{U^T M_X U}{T}\right)^{-1} r_j \right] \quad (A.150)$$

$$\leq E\left[ \sum_{j=1}^{K} r_j^T \left(\frac{U^T M_X U}{T}\right)^{-1} r_j \right] \quad (A.151)$$

$$= E\left[ tr\left( \frac{U^T P_X U}{T} \left(\frac{U^T M_X U}{T}\right)^{-1} \right) \right] \quad (A.152)$$

$$\leq C_1. \quad (A.153)$$

The logic of the display above is as follows: the first equality holds by matrix multiplication and the definition of the $r_J$; the first inequality is by the generalized Caucy-Schwarz inequality (see Magnus and Neudecker (1999, page 200, problem 1)); the second equality follows from simply regrouping terms; the second inequality is due to Cauchy-Schwarz; the third equality is by the cancellation of $K-1$ and $\frac{1}{K-1}$; the third inequality is due to the fact that $r_i^T \left(\frac{U^T M_X U}{T}\right)^{-1} r_i \geq 0$; the fourth equality is by the usual trace trick and the definition of the $r_j$ (we also exploit the fact that $\frac{X^T X}{T} = I_K$); the final inequality is by the assumption of the theorem.

Combining the above bounds yields:

$$\frac{1}{K-1} \sum_{\substack{j=1,\\ j\neq i}}^{K} \sum_{\substack{n=1,\\ n\neq i}}^{K} \frac{\left| Cov\left(\widehat{b}_j, \widehat{b}_n | \widehat{b}_i\right) \right|}{\sqrt{Var\left(\widehat{b}_j | \widehat{b}_i\right) Var\left(\widehat{b}_n | \widehat{b}_i\right)}} \quad (A.154)$$

$$\leq 1 + 2C_1. \quad (A.155)$$

which gives the desired bound in terms of an absolute constant $C = 1 + 2C_1$.

### A.9 The Case $K = \rho T^{\delta}$, $\delta < 1$

The bound on the mean squared error of the standard residual variance estimator given in Theorem 4(a) depends on $K$. An examination of the proof of Theorem 4(a) above shows that this bound could actually be stated in terms of $T$; however, if $\delta < 1$, the bound as stated (in terms of $K$) is weaker than the corresponding bound with $T$ substituted for $K$, and certainly remains valid.

In the proofs above, note that all of the risks include the factor $\rho$. This is a consequence of the fact that, under the simplifying assumption that $\delta = 1$, $\rho = \frac{K}{T}$. When $\delta < 1$, all of the risks in these proofs actually include the factor $\frac{K}{T} = \rho T^{\delta - 1}$. Thus, scaling the risks by $\frac{T}{K} = T^{1-\delta}$ just returns us to the situation considered in the proofs, in which each risk includes the (constant) factor $\rho$. Every other element of the proofs is completely unchanged, so the results continue to hold with the stated scaling of the risks.

# References


[1] Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, New York: Springer-Verlag.

[2] Billingsley, P. (1995), *Probability and Measure* (Third Edition), New York: Wiley.

[3] Bradley, R. (1993), "Equivalent Mixing Conditions for Random Fields," *The Annals of Probability*, 21, 1921-1926.

[4] Doukhan, P. (1994), *Mixing: Properties and Examples*, New York: Springer-Verlag.

[5] Dudley, R. (1999), *Uniform Central Limit Theory*, New York: Cambridge University Press.

[6] Feller, W. (1971), *An Introduction to Probability Theory and Its Applications, Volume 2* (Second Editon), New York: Wiley.

[7] Hall. P. and C. C. Heyde (1980), *Martingale Limit Theory and Its Applications*, New York: Academic Press.

[8] Magnus, J. and H. Neudecker (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Revised Edition), New York: Wiley.

[9] Theil, H. (1971), *Principles of Econometrics*, New York: Wiley.

[10] Rozanov, Y. (1967), *Stationary Random Processes*, San Francisco: Holden-Day.

[11] van der Vaart, A. (1988), *Statistical Estimation in Large Parameter Spaces*, Amsterdam: Stichting Mathematisch Centrum.

# B Berry-Esseen Theorems for Densities and Their Derivatives

This part provides Berry-Esseen-type theorems for the densities and derivatives of densities of univariate and bivariate random variables. These theorems are referred to below as local limit results.

The presentation proceeds in two steps. First, local limit results are proven assuming that a Berry-Esseen theorem (for c.d.f.'s) and a smoothness condition hold. Because Berry-Esseen theorems hold under a variety of primitive conditions, this provides general conditions under which the local limit results hold. Second, it is shown that a (multivariate) Berry-Esseen theorem does in fact hold for averages of strongly mixing random variables satisfying certain moment and mixing-rate conditions. This theorem is an adaptation of Tikhomirov's (1980) univariate result.

## B.1 Local Limit Result

Let $\eta_1, \eta_2, \ldots$ be a sequence of $m$-dimensional random variables with mean zero, finite second moments, and finite long-run covariance matrix. Without loss of generality, let

$$\lim_{n\to\infty} E\left[\left(\sqrt{\frac{1}{n}}\sum_{i=1}^{n}\eta_i\right)\left(\sqrt{\frac{1}{n}}\sum_{i=1}^{n}\eta_i\right)'\right] = I_m.$$

Let $J_n$ denote the distribution function of $\sqrt{\frac{1}{n}}\sum_{i=1}^{n}\eta_i$ and let $\Phi$ denote the $m$-variate standard normal distribution function. The local limit results will be proven under the following two conditions.

**Condition A**: The random variables $\eta_1, \eta_2, \ldots$ have conditional characteristic functions $\psi_1, \psi_2, \ldots$ (so that $\psi_i$ is the characteristic function of the distribution of $\eta_i$ conditional on $\eta_j,\ 1 \le j < i$) with the property that

$$\begin{aligned} \exists\, \alpha &> 0,\ C_0 < \infty,\ M_0 < \infty \text{ s. t.} \\ \sup_i |\psi_i(t)| &\le M_0 |t|^{-\alpha} \quad \forall |t| \ge C_0. \end{aligned} \tag{B.1}$$

Condition A is weaker than requiring, for instance, that the conditional densities of the $\eta_i$ be uniformly bounded, or even that any of them be bounded, though it does rule out discreteness.

**Condition B**: A Berry-Esseen theorem holds for $\eta_1, \eta_2, \ldots$, that is,

$$\exists\, \beta > 0,\ \mu < \infty,\ M_1 < \infty \ \ s.t. \ \sup_{z\in R^m} |J_n(z) - \Phi(z)| \le M_1 n^{-\beta} \log^{\mu}(n) \tag{B.2}$$

Typical Berry-Esseen theorems specify a particular constant $M_1$ and have $\beta = \frac{1}{2}$ and $\mu = 0$ (*c. f.* Feller (1971), F. Götze (1991), Hall and Heyde (1980)). However, the local limit results here do not rely on these specific values, and so are proven for the more general statement (B.2).

**Lemma S-7 (Cramér 1937)** *Suppose $t, \zeta, \chi \in \Re^m$. If $\psi(t)$ is a characteristic function such that $|\psi(t)| \leq \nu < 1$ for all $|t| \geq M$, then we have for $|t| < M$*

$$|\psi(t)| \leq 1 - (1-\nu)^2 \frac{|t|^2}{8M^2} \tag{B.3}$$

*Proof of Lemma S-7*: A terse proof for $m = 1$ is to be found on page 26 of Cramér (1937). However, it is given in an expanded form here for the reader's convenience.

Recall first that for scalars $A$ and $B$, $\cos A \leq \frac{3}{4} + \frac{1}{4}\cos(2A)$ (since $\frac{1}{4}\cos(2A) - \cos(A) + \frac{3}{4} = \frac{1}{2}\cos^2(A) - \cos(A) + \frac{1}{2} = \frac{1}{2}(\cos(A) - 1)^2 \geq 0$) and $\sin(A - B) = \sin(A)\cos(B) - \sin(B)\cos(A)$.

$$\begin{aligned}
|\psi(t)|^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it'(\zeta-\chi)} dJ(\zeta)\, dJ(\chi) \qquad \text{(B.4)}\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i\left[\sin(t'\zeta)\cos(t'\chi) - \sin(t'\chi)\cos(t'\zeta)\right] + \\
&\quad \left[\cos(t'(\zeta-\chi))\right] dJ(\zeta)\, dJ(\chi)\\
&= i\left[\begin{array}{c} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sin(t'\zeta)\cos(t'\chi)\, dJ(\zeta)\, dJ(\chi) - \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sin(t'\chi)\cos(t'\zeta)\, dJ(\zeta)\, dJ(\chi) \end{array}\right]\\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(t'(\zeta-\chi))\, dJ(\zeta)\, dJ(\chi)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(t'(\zeta-\chi))\, dJ(\zeta)\, dJ(\chi)\\
&\leq \frac{3}{4} + \frac{1}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(2t'(\zeta-\chi))\, dJ(\zeta)\, dJ(\chi)\\
&= \frac{3}{4} + \frac{1}{4}\left\{\begin{array}{c} i\left[\begin{array}{c} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sin(2t'\zeta)\cos(2t'\chi)\, dJ(\zeta)\, dJ(\chi) - \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sin(2t'\chi)\cos(2t'\zeta)\, dJ(\zeta)\, dJ(\chi) \end{array}\right] + \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(2t'(\zeta-\chi))\, dJ'(\zeta)\, dJ(\chi) \end{array}\right\}\\
&= \frac{3}{4} + \frac{1}{4}|\psi(2t)|^2
\end{aligned}$$

where the first equality holds because the square of a complex number's modulus equals its product (in complex multiplication) with its complex conjugate, the second equality holds by the trigonometric identity recalled above, the third is merely a rearrangement, and the fourth follows from the fact that the two integrals in the imaginary part are equal, so their difference is zero (we may simply relabel the variables). The inequality comes from the trigonometric inequality noted above, and the last two equalities are simply the first four "in reverse." Thus, we have that for $|t| \in \left[\frac{M}{2}, M\right)$

$$|\psi(t)|^2 \leq 1 - \frac{1}{4}(1-\nu)^2 \tag{B.5}$$

and we may repeat this argument to show that for $|t| \in \left[\frac{M}{2^q}, \frac{M}{2^{q-1}}\right)$ (for any integer $q \geq 1$)

$$|\psi(t)|^2 \leq 1 - \left(\frac{1}{4}\right)^q (1-\nu)^2 < 1 - (1-\nu)^2 \frac{|t|^2}{4M^2} \tag{B.6}$$

so that, in the same region,

$$|\psi(t)| \leq 1 - (1-\nu)^2 \frac{|t|^2}{8M^2} \tag{B.7}$$

as can be seen by squaring the righthand side of the inequality immediately above and comparing it to the rightmost quantity of the inequality in the previous display. Now, $q$ is arbitrary, so the desired conclusion must hold for $|t| \in (0, M)$. But we know that $\psi(0) = 1$, so we are finished. Q.E.D.

**Theorem 8 (Univariate Local Limit Theorem)** *Suppose that Conditions A and B hold, and let $m = 1$. Then $\forall\, d \in N$, $\exists\, B(d) < \infty$ and $n_0(d) < \infty$ such that, $\forall n > n_0(d)$,*

$$\sup_{z \in R} \left| j_n^{(d)}(z) - \phi^{(d)}(z) \right| \leq B(d)\, n^{-\beta \,/\, 2^{d+1}} \left[\log(n)\right]^{\mu \,/\, 2^{d+1}}$$

*where $j_n^{(d)}$ is the $d^{th}$ derivative, $d = 0, 1, 2, \ldots$, of the density of $\frac{\eta_1 + \eta_2 + \cdots + \eta_n}{\sqrt{n}}$ (this density will be shown to exist for sufficiently large $n$ as part of the proof), and $\phi^{(d)}$ is the $d^{th}$ derivative of the standard normal density.*

*Proof of Theorem 8*: We will proceed by induction on the order $d$ of the derivative to be taken. First we will prove that the result holds for $d = 0$. Note that if the first derivative of $j_n$ is uniformly bounded for all $n$ greater than some given $n_0$, then $j_n$ will clearly satisfy a Lipschitz condition uniformly beyond $n_0$, that is, we will have $\sup_{z,w \in \Re} \sup_{n \geq n_0} |j_n(z) - j_n(w)| \leq B_0 |z - w|$ for some $B_0 < \infty$. Clearly, $\phi$ satisfies such a smoothness condition (all its derivatives exist and are bounded). Let $B_1 < \infty$ be the Lipschitz constant for $\phi$. Now, if the Lipschitz constant holds for $j_n$,

$$\begin{aligned}
\sup_{z \in R} |j_n(z) - \phi(z)| \leq\; & \sup_{z \in R} \left| j_n(z) - \frac{\int_z^{z+r} j_n(w)\, dw}{r} \right| + \\
& \sup_{z \in R} \frac{1}{r} \left| \int_z^{z+r} j_n(w)\, dw - \int_z^{z+r} \phi(w)\, dw \right| + \\
& \sup_{z \in R} \left| \phi(z) - \frac{\int_z^{z+r} \phi(w)\, dw}{r} \right| \\
\leq\; & \sup_{z \in R} |j_n(z) - j_n(c_f)| + \\
& \sup_{z \in R} \frac{1}{r} \left| \int_z^{z+r} j_n(w)\, dw - \int_z^{z+r} \phi(w)\, dw \right| + \\
& \sup_{z \in R} |\phi(z) - \phi(c_\phi)| \\
\leq\; & B_0 |z - c_f| + \\
& \frac{1}{r} \sup_{z \in \Re} |J_n(z+r) - J_n(z) - \Phi(z+r) + \Phi(z)| + \\
& B_1 |z - c_\phi|
\end{aligned} \tag{B.8}$$

$$
\begin{aligned}
&\leq \quad 2\left(B_0+B_1\right) r \;+\; \frac{2}{r} \sup_{z \in \Re} \left|J_n(z)-\Phi(z)\right| \\
&\leq \quad \left(B_0+B_1\right) r \;+\; \frac{2 M_1}{r n^{\beta}} \log^{\mu}(n) \\
&\leq \quad 2\left[2^{1/2}\left(B_0+B_1\right)^{1/2} M_1^{1/2}\right] n^{-\beta/2} \log^{\mu/2}(n)
\end{aligned}
$$

where $c_f, c_\phi \in [z, z+r]$ by the mean value theorem. The second term in the fifth inequality uses Condition B, and the final inequality follows by setting $r = \sqrt{\frac{2M_1}{B_0+B_1}} n^{-\beta/2} \log^{\mu/2}(n)$, which is certainly permitted, since $r$ is arbitrary.

Thus the lemma is proven with $B(0) = 2\left[2^{1/2}\left(B_0+B_1\right)^{1/2} M_1^{1/2}\right]$ if we can show that the Lipschitz condition holds.

This is where Cramér's (1937) lemma is useful: consider general $d \in N$. By the Fourier inversion theorem,

$$
\begin{aligned}
\left|j_n^{(d)}(z)\right| \quad &\leq \quad \int_{-\infty}^{\infty} |t|^d \left\{\prod_{i=1}^{n} \left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt \qquad \text{(B.9)} \\
&\leq \quad \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t|^d \left\{\prod_{i=1}^{n} \left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt \;+\; \int_{\delta\sqrt{n}}^{\infty} |t|^d \left\{\prod_{i=1}^{n} \left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt \\
&\quad\; + \int_{-\infty}^{-\delta\sqrt{n}} |t|^d \left\{\prod_{i=1}^{n} \left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt.
\end{aligned}
$$

Now choose $\delta = \max\left\{\left(2M_0\right)^{1/\alpha}, C_0\right\}$, where the parameters refer to Condition A. Then, by Condition A, $\forall t > \delta$, $\sup_i \left|\psi_i(t)\right| \leq \frac{1}{2} \equiv \nu$. We can now apply Cramér's lemma to obtain $\sup_i \left|\psi_i(t)\right| \leq 1-(1-\nu)^2 \frac{|t|^2}{8\delta^2} \;\forall\; |t| < \delta$. Thus, we have

$$
\begin{aligned}
\sup_{n \geq n_0} \left|j_n^{(d)}(z)\right| \quad &\leq \quad \sup_{n \geq n_0} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t|^d \left(1-(1-\nu)^2 \frac{t^2/n}{8\delta^2}\right)^n dt \;+ \qquad \text{(B.10)} \\
&\quad\quad \sup_{n \geq n_0} \nu^{n-m} \int_{\delta\sqrt{n}}^{\infty} |t|^d \left(\frac{M_0}{\left|t/\sqrt{n}\right|^{\alpha}}\right)^m dt \;+ \\
&\quad\quad \sup_{n \geq n_0} \nu^{n-m} \int_{-\infty}^{-\delta\sqrt{n}} |t|^d \left(\frac{M_0}{\left|t/\sqrt{n}\right|^{\alpha}}\right)^m dt \\
&\leq \quad \sup_{n \geq n_0} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t|^d \exp\left(-Ct^2\right) dt \;+ \\
&\quad\quad 2 \sup_{n \geq n_0} \nu^{n-m} M_0^m n^{\alpha m/2} \int_{-\infty}^{-\delta\sqrt{n}} |t|^{d-\alpha m} dt \\
&\leq \quad \sup_{n \geq n_0} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t| \exp\left(-Ct^2\right) dt \;+ \\
&\quad\quad 2 \sup_{n \geq n_0} M_0^{\lceil (d+2)/\alpha \rceil} \nu^{n-\lceil (d+2)/\alpha \rceil} n^{(d+2)/2} \int_{\delta}^{\infty} |t|^{-2} dt \\
&\leq \quad M_2(d) \;<\; \infty,
\end{aligned}
$$

where the second inequality follows by setting $C = \frac{(1-\nu)^2}{8\delta^2}$ and observing that

$$\exp\left(-Ct^2\right) = \sup_n \left(1 - \frac{Ct^2}{n}\right)^n,$$

and the third inequality follows by choosing $m = \left\lceil \frac{d+2}{\alpha} \right\rceil$, where $\lceil \cdot \rceil$ is the least greater (or equal) integer function. The desired Lipschitz condition is now verified. We have only to set $d = 0$ to prove the existence and boundedness of the density itself for sufficiently large $n$ in an identical fashion.

We have now proven the $d = 0$ case. To prove higher-$d$ cases, we now simply substitute $j_n^{(d)}$ for $j_n$ and $\phi^{(d)}$ for $\phi$ in (B.8) and, following the steps in (B.8), we obtain

$$\begin{aligned} & \sup_{z \in \Re} \left| j_n^{(d)}(z) - \phi^{(d)}(z) \right| \\ \leq \quad & \left(B_0(d) + B_1(d)\right) r \; + \; \frac{2}{r} \sup_{z \in \Re} \left| j_n^{(d-1)}(z) - \phi^{(d-1)}(z) \right| \end{aligned} \tag{B.11}$$

where the Lipschitz condition holds as a consequence of (B.10). It is readily verified that the bound in the statement of the theorem satisfies the recursion (B.11), and since we have shown the bound for $d = 0$, we are finished. Q. E. D.

**Theorem 9 (Bivariate Local Limit Theorem)** *Suppose that Conditions A and B hold with $m = 2$. Let $j_n^{(0)}(z, w)$ denote the density of $\sqrt{\frac{1}{n}}\sum_{i=1}^n \eta_i$, and let $\phi(z, w)$ denote the bivariate standard normal density. Then $\forall\, d \in N$, $\exists\, B(d) < \infty$ such that*

$$\sup_{z \in R, w \in R} \left| j_n^{(d)}(z, w) - \phi^{(d)}(z, w) \right| \;\leq\; B(d)\, n^{-\beta \,/\, \left(3 \times 2^d\right)} \left[\log(n)\right]^{\mu \,/\, \left(3 \times 2^d\right)}$$

*where $j_n^{(d)}$ is the $d^{th}$ derivative, $d = 0, 1, 2, \ldots$, of $j_n^{(0)}(z, w)$ with respect to $z$ (this density will be shown to exist for sufficiently large $n$ as part of the proof) and $\phi^{(d)}$ is the $d^{th}$ derivative of $\phi(z, w)$ with respect to $z$.*

*Proof of Theorem 9*: We will proceed by induction on the order $d$ of the derivative to be taken. First we will prove that the result holds for $d = 0$. Note that if the first derivative of $j_n$ with respect to $z$ is uniformly bounded for all $n$ greater than some given $n_0$, then $j_n$ will clearly satisfy a Lipschitz condition in $z$ uniformly beyond $n_0$, that is, we will have $\sup_{z,w,u,v \in \Re} \sup_{n \geq n_0} \left| j_n(z, w) - j_n(u, v) \right| \leq B_0 \left| \begin{pmatrix} z \\ w \end{pmatrix} - \begin{pmatrix} u \\ v \end{pmatrix} \right|$ and for some $B_0 < \infty$. Now, if such a uniform Lipschitz condition holds, we can easily show that, since $\phi$ clearly satisfies such a smoothness condition (all its derivatives exist and are bounded),

$$\begin{aligned} & \sup_{z, w \in R} \left| j_n(z, w) - \phi(z, w) \right| \\ \leq \quad & \sup_{z, w \in \Re} \left| j_n(z, w) - \frac{\int_z^{z+r} \int_w^{w+r} j_n(u, v)\, du dv}{r^2} \right| \; + \end{aligned} \tag{B.12}$$

$$
\begin{aligned}
& \sup_{z,w\in\Re} \frac{1}{r^2}\left|\int_z^{z+r}\int_w^{w+r} j_n(u,v)\,dudv - \int_z^{z+r}\int_w^{w+r}\phi(u,v)\,dudv\right| \; + \\
& \sup_{z,w\in\Re}\left|\phi(z,w) - \frac{\int_z^{z+r}\int_w^{w+r}\phi(u,v)\,dudv}{r^2}\right| \\
\leq \; & \sup_{z,w\in\Re}\left|j_n(z,w) - j_n(c_f,d_f)\right| \; + \\
& \sup_{z,w\in\Re} \frac{1}{r^2}\left|\int_z^{z+r}\int_w^{w+r} j_n(u,v)\,dudv - \int_z^{z+r}\int_w^{w+r}\phi(u,v)\,dudv\right| \; + \\
& \sup_{z,w\in\Re}\left|\phi(z,w) - \phi(c_\phi,d_\phi)\right| \\
\leq \; & B_0\left|\begin{pmatrix} z \\ w\end{pmatrix} - \begin{pmatrix} c_f \\ d_f\end{pmatrix}\right| \; + \; \frac{2}{r^2}\sup_{z,w\in\Re}\left|J_n(z,w) - \Phi(z,w)\right| \; + \\
& B_1\left|\begin{pmatrix} z \\ w\end{pmatrix} - \begin{pmatrix} c_\phi \\ d_\phi\end{pmatrix}\right| \\
\leq \; & (B_0+B_1)\,r \; + \; \frac{2M_1}{r^2 n^\beta}\log^\mu(n) \\
\leq \; & \left[\left(4^{1/3}+4^{1/6}\right)(B_0+B_1)^{2/3} M_1^{1/3}\right] n^{-\beta/3}\log^{\mu/3}(n)
\end{aligned}
$$

where $c_f, c_\phi \in [z, z+r]$ and $d_f, d_\phi \in [w, w+r]$ by the mean value theorem. The second term in the fourth inequality uses Condition B, and the final inequality follows by setting $r = \sqrt[3]{\frac{4M_1}{B_0+B_1}}n^{-\beta/3}\log^{\mu/3}(n)$, which is certainly permitted, since $r$ is arbitrary.

Thus the lemma is proven with $B(0) = \left[\left(4^{1/3}+4^{1/6}\right)(B_0+B_1)^{2/3} M_1^{1/3}\right]$ if we can show that the Lipschitz condition holds.

This is where Cramér's (1937) lemma is useful: consider general $d \in N$. By the Fourier inversion theorem,

$$
\begin{aligned}
\left|j_n^{(d)}(z,w)\right| \; & \leq \; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |t_1|^d \left\{\prod_{i=1}^n \left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt_1 dt_2 \\
& \leq \; \int_{B_{\delta\sqrt{n}}} |t_1|^d \left\{\prod_{i=1}^n \left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt \\
& \quad + \int_{\Re^2 - B_{\delta\sqrt{n}}} |t|^d \left\{\prod_{i=1}^n \left|\psi_i\left(\frac{t}{\sqrt{n}}\right)\right|\right\} dt
\end{aligned} \tag{B.13}
$$

where $B_{\delta\sqrt{n}} \equiv \{t : |t| \leq \delta\sqrt{n}\}$, and, by choosing $\delta = \max\left\{(2M_0)^{1/\alpha}, C_0\right\}$, we will have $\nu = \frac{1}{2}$ such that $\sup_i |\psi_i(t)| \leq \nu \; \forall \; |t| \geq \delta$. But then Cramér's lemma proves that we have $\sup_i |\psi_i(t)| \leq 1 - (1-\nu)^2 \frac{|t|^2}{8\delta^2} \; \forall \; |t| < \delta$. Thus, we have

$$
\begin{aligned}
& \sup_{n\geq n_0}\left|j_n^{(d)}(z,w)\right| \\
\leq \; & \sup_{n\geq n_0}\int_{B_{\delta\sqrt{n}}} |t_1|^d \left(1 - (1-\nu)^2 \frac{|t|^2/n}{8\delta^2}\right)^n dt \; +
\end{aligned} \tag{B.14}
$$

$$
\begin{aligned}
& \sup_{n \geq n_0} \nu^{n-m} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t_1|^d \left( \frac{M_0}{|t \,/\, \sqrt{n}|^\alpha} \right)^m dt \\
\leq \;& \sup_{n \geq n_0} \int_{B_{\delta\sqrt{n}}} |t_1|^d \exp\left(-C\,|t|^2\right) dt \;\; + \\
& \sup_{n \geq n_0} M_0^m \nu^{n-m} n^{\alpha m \,/\, 2} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t_1|^d \, |t|^{-\alpha m} \, dt \\
\leq \;& \sup_{n \geq n_0} \int_{B_{\delta\sqrt{n}}} |t_1|^d \exp\left(-C\,|t|^2\right) dt \;\; + \\
& \sup_{n \geq n_0} \dot{M}_0^{\lceil (d+4) \,/\, \alpha \rceil} \nu^{n - \lceil (d+4) \,/\, \alpha \rceil} n^{(d+4) \,/\, 2} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t_1|^d \, |t|^{-(d+4)} \, dt \\
\leq \;& \sup_{n \geq n_0} \int_{\Re^2} |t_1|^d \exp\left(-C\,|t|^2\right) dt \;\; + \\
& \sup_{n \geq n_0} M_0^{\lceil (d+4) \,/\, \alpha \rceil} \nu^{n - \lceil (d+4) \,/\, \alpha \rceil} n^{(d+4) \,/\, 2} \int_{\Re^2 - B_{\delta\sqrt{n}}} |t|^{-4} \, dt \\
\leq \;& \sup_{n \geq n_0} \int_{\Re^2} |t_1| \exp\left(-C\,|t|^2\right) dt \;\; + \\
& \sup_{n \geq n_0} M_0^{\lceil (d+4) \,/\, \alpha \rceil} \nu^{n - \lceil (d+4) \,/\, \alpha \rceil} n^{(d+4) \,/\, 2} \int_{\Re^2 - B_{\delta\sqrt{n}}} r^{-3} dr d\theta \\
\leq \;& M_2 \; + \; \sup_{n \geq n_0} M_0^{\lceil (d+4) \,/\, \alpha \rceil} \nu^{n - \lceil (d+4) \,/\, \alpha \rceil} n^{(d+2) \,/\, 2} 2\pi \int_\delta^\infty r^{-3} dr \\
\leq \;& M_3 \; < \; \infty
\end{aligned}
$$

where the second inequality follows by setting $C = \frac{(1-\nu)^2}{8\delta^2}$ and noting that

$$
\exp\left(-C\,|t|^2\right) \;=\; \sup_n \left( 1 - \frac{(1-\nu)^2 \, |t|^2}{8\delta^2 n} \right)^n,
$$

and the third inequality follows by setting $m = \left\lceil \frac{d+4}{\alpha} \right\rceil$, where $\lceil \cdot \rceil$ is the least integer greater function. The fourth inequality is due to the fact that $|t_1|^2 = ((1\;0\;\cdots\;0)\,t)^2 \;\leq\; |t|^2$ by the Cauchy-Schwarz inequality, so that $|t_1| \leq |t|$, and the fifth inequality follows upon a change of variables into polar coordinates. So the $d^{\text{th}}$ derivative of $j_n$ w. r. t. $z$ is bounded. An identical proof in which $|t_2|$ is substituted for $|t_1|$ shows that the $d^{\text{th}}$ derivative of $j_n$ w. r. t. $w$ is bounded. Together, these two bounds produce the Lipschitz condition. We have only to set $d = 0$ in the proof above to prove the existence and boundedness of the density itself for sufficiently large $n$.

We have now proven the $d = 0$ case. To prove higher-$d$ cases, we now simply substitute $j_n^{(d)}$ for $j_n$ and $\phi^{(d)}$ for $\phi$ in (B.12), and, only slightly modifying the steps in (B.12), we obtain

$$
\begin{aligned}
& \sup_{z,w \in R} \left| j_n^{(d)}(z,w) - \phi^{(d)}(z,w) \right| \qquad \text{(B.15)} \\
\leq \;& \sup_{z,w \in \Re} \left| j_n^{(d)}(z,w) - \frac{\int_z^{z+r} j_n^{(d)}(u,w)\,du}{r} \right| \; +
\end{aligned}
$$

$$
\begin{aligned}
& \sup_{z,w\in\Re} \frac{1}{r} \left| \int_z^{z+r} j_n^{(d)}(u,w)\,du - \int_z^{z+r} \phi^{(d)}(u,w)\,du \right| + \\
& \sup_{z,w\in\Re} \left| \phi(z,w) - \frac{\int_z^{z+r} \phi^{(d)}(u,w)\,du}{r} \right| \\
\leq \quad & (B_0(d) + B_1(d))\, r + \frac{2}{r} \sup_{z,w\in\Re} \left| j_n^{(d-1)}(z,w) - \phi^{(d-1)}(z,w) \right| .
\end{aligned}
$$

Note that the second term of the upper bound is, as a function of $r$, only an inverse, rather than an inverse squared, because we have taken only a first partial derivative, rather than a cross or second partial. It is readily verified that the bound in the statement of the theorem satisfies the recursion (B.15), and since we have shown the bound for $d = 0$, we are finished. Q. E. D.

## B.2 Multivariate Berry-Esseen Theorem Under Strong Mixing

We now provide a multivariate Berry-Esseen theorem that applies to sequences $\{\eta_i\}_{i=1}^{\infty}$ of random variables in $\Re^m$ which satisfy a strong mixing condition and a moment condition.

**Definition 2** *A sequence of random variables $\eta_1, \eta_2, \ldots$ will be said to be* strongly mixing *with coefficients $\alpha(n)$ if*

$$
\alpha(n) = \sup_{k, A\in\mathcal{F}_1^k, B\in\mathcal{F}_{k+n}^{\infty}} |P(AB) - P(A)P(B)| \qquad \text{(B.16)}
$$

*where $\mathcal{F}_a^b$ is the $\sigma$-algebra generated by $\eta_j$, $j \in \{a, a+1, \ldots b\}$.*

**Condition C**. Let $\eta_1, \eta_2, \ldots$ be a sequence of $\Re^m$-valued random variables with $E[\eta_i] = 0$, $\sup_i E\left[|\eta_i|^{4+\gamma}\right] < \infty$ for some $\gamma > 0$, and $\alpha(n) \leq M_3 e^{-\beta n}$ for some $M_3 < \infty$ and some $\beta > 0$.

Theorem 10 provides a $m$-variate Berry-Esseen theorem which holds under Condition C. The results of this theorem satisfy Condition B. Thus the local limit results given above hold under, in particular, Conditions A and C.

This section concludes with the statement of this theorem, which is minor modification of a result of Tikhomirov (1980).

**Theorem 10** *Suppose Condition C is satisfied. Then there is a constant $C_2$ depending only on $m, \beta, \gamma, M_3$ such that*

$$
\sup_{z\in\Re^m} |J_n(z) - \Phi(z)| \leq C_2 n^{-1/2} \log n \qquad \text{(B.17)}
$$

*where $J_n(z) = \Pr(S_{n,1} \leq z_1, S_{n,2} \leq z_2, \ldots, S_{n,m} \leq z_m)$, with $S_n = H_n^{-1} \sum_{i=1}^n \eta_i$, where $H_n \equiv$ Cholesky $\left\{E\left[\left(\sum_{i=1}^n \eta_i\right)\left(\sum_{i=1}^n \eta_i\right)'\right]\right\}$ (the Cholesky factor of the term given).*

*Proof:* Consider Tikhomirov's (1980) Theorem 4. Although this is a univariate result, we see that minor modifications make it applicable to $\Re^m$. Namely, the ordinary differential equation that Tikhomirov derives for the characteristic function becomes a system of ordinary differential equations, and the solution, naturally, becomes a multivariate characteristic function. However, the structure of his proof remains exactly the same; his lemmas transfer naturally to the multivariate case.

# References


[1] Cramér, H. (1937), *Random Variables and Probability Distributions*, Cambridge UK: Cambridge Tracts.

[2] Feller, W. (1971), *An Introduction to Probability Theory and Its Applications, Volume 2* (Second Editon), New York: Wiley.

[3] Götze, F. (1991), "On the rate of convergence of the multivariate CLT," *Annals of Probability*, pp. 724-739.

[4] Hall. P. and C. C. Heyde (1980), *Martingale Limit Theory and Its Applications*, New York: Academic Press.

[5] Tikhomirov, A. (1980), "On the convergence rate in the central limit theorem for weakly dependent random variables," *Theory of Probability and Its Applications*, 25, pp. 790-809.