APPENDIX D: SUPPLEMENT

D.1. *Prior Sensitivity and Robustness Checks*

Figure 5 shows how the results in Figure 3 change when the prior hyper-parameters (a, γ) take the following values: $\{(15, 0.25), (15, 0.5), (15, 1), (10, 0.5), (20, 0.5)\}$.

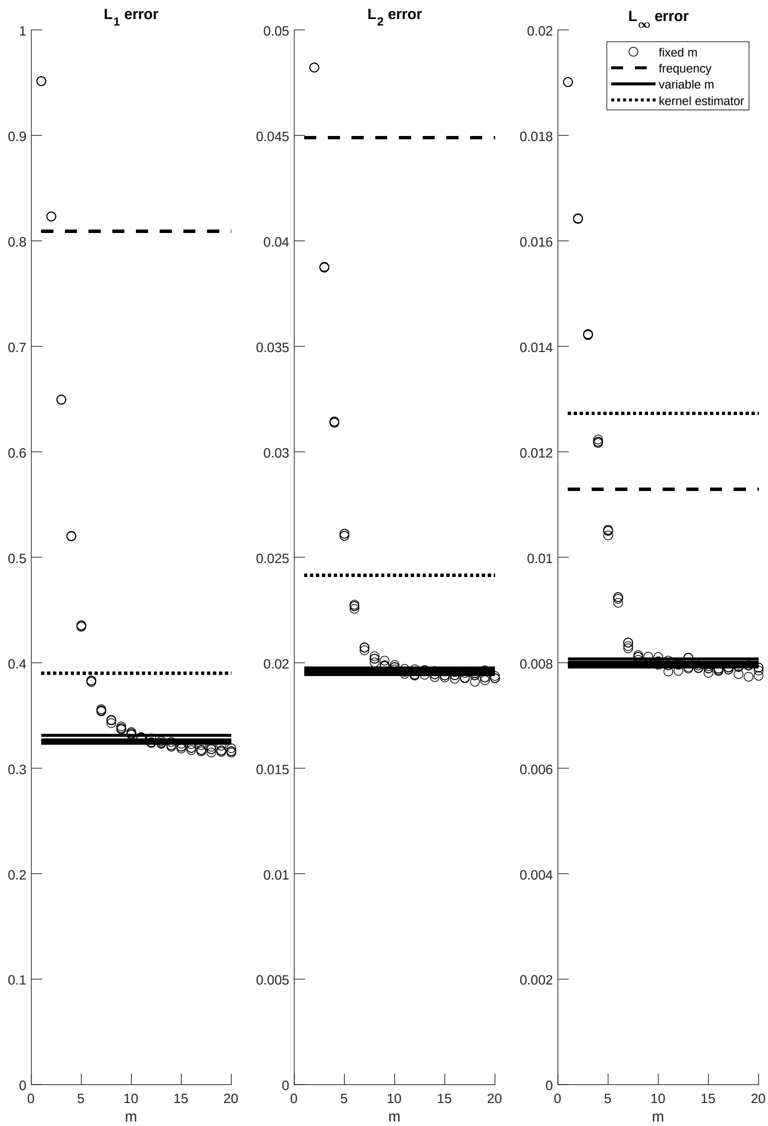

FIGURE 5.— Average Estimation Errors under Different Priors

Note that some lines from these five different experiments coincide (the estimation results for fixed m are not affected by changes in γ) and some lines for variable m are almost indistinguishable from each other. As can be seen from the figure, the estimation results are not sensitive to moderate variations in (a, γ).

D.2. *Proofs and Auxiliary Results for Lower Bounds*

LEMMA 3 *For q_j, q_l, $i \neq l$ defined in (16), the L_1 distance is bounded below by const$\cdot\Gamma_n$.*

PROOF: Let us establish several facts about g_r in the definition of q_j. For any $(\tilde{y}, x) \in [0,1]^d$, there exists $r(\tilde{y}, x)$ such that

$$g_r(\tilde{y}, x) = 0, \ \forall r \neq r(\tilde{y}, x). \tag{34}$$

For $(\tilde{y}, x) \in B_r$, $r(\tilde{y}, x) = r$ and for $(\tilde{y}, x) \notin \cup_{r=1}^{\bar{m}} B_r$, $r(\tilde{y}, x)$ can have an arbitrary value. Thus,

$$\begin{aligned} d_{L_1}(q_j, q_l) &= \sum_y \int \left| \int_{A_y} \left[\sum_{r=1}^{\bar{m}} (w_r^j - w_r^l) g_r(\tilde{y}, x) \right] d\tilde{y} \right| dx \\ &= \sum_y \int \left| \int_{A_y} (w_{r(\tilde{y},x)}^j - w_{r(\tilde{y},x)}^l) g_{r(\tilde{y},x)}(\tilde{y}, x) d\tilde{y} \right| dx. \end{aligned}$$

From $h_i = (2/N_i) \cdot R_i$ for $i \in \{1, \ldots, d_y\}$, where R_i is a positive integer, and the definitions of g, g_r, and A_y, it follows that for fixed $y \in \mathcal{Y}$ and $x \in [0,1]^{d_x}$, $(w_{r(\tilde{y},x)}^j - w_{r(\tilde{y},x)}^l) g_{r(\tilde{y},x)}(\tilde{y}, x)$ does not change the sign as $\tilde{y}$ changes within A_y ($r(\tilde{y}, x)$ is the same $\forall \tilde{y} \in A_y$ by the choice of c_i^r and h_i). Therefore,

$$\begin{aligned} d_{L_1}(q_j, q_l) &= \int \int \left| (w_{r(\tilde{y},x)}^j - w_{r(\tilde{y},x)}^l) g_{r(\tilde{y},x)}(\tilde{y}, x) \right| d\tilde{y} dx \\ &= \sum_{r=1}^{\bar{m}} \int_{B_r} \left| (w_{r(z)}^j - w_{r(z)}^l) g_{r(z)}(z) \right| dz \\ &= \sum_{r=1}^{\bar{m}} |w_r^j - w_r^l| \int_{B_r} |g_r(z)| \, dz. \end{aligned} \tag{35}$$

Finally, using a change of variables in (35), Lemma 2, and $m_i h_i > 1/2$, we get

$$\begin{aligned} d_{L_1}(q_j, q_l) &= \sum_{r=1}^{\bar{m}} 1\{w_r^j \neq w_r^l\} \cdot \Gamma_n \cdot \prod_{i=1}^d h_i \cdot \left[\int_{-1/2}^{1/2} |g(u)| du \right]^d \\ &\geq \Gamma_n \cdot \prod_{i=1}^d m_i h_i \cdot \left[\int_{-1/2}^{1/2} |g(u)| du \right]^d /8 \\ &\geq \Gamma_n \cdot \left[\int_{-1/2}^{1/2} |g(u)| du/2 \right]^d /8. \end{aligned}$$

Q.E.D.

LEMMA 4 *For $\Gamma_n \to 0$ and $\bar{m} \geq 8$ and a sufficiently small c_0 in the definition of g, condition (15) in Lemma (1) holds for all sufficiently large n.*

PROOF: By Lemma 2, it suffices to show that

$$d_{KL}(Q_j^n, Q_0^n) = n \cdot d_{KL}(q_j, q_0) < (\bar{m} \log 2)/64. \tag{36}$$

First, note that for any $z \in [0,1]^d$, the density in the definition of q_j

$$g_0(z) + \sum_{r=1}^{\bar{m}} w_r^j g_r(z) \geq \underline{g}_0 - \Gamma_n \left[\max_{u \in [-1/2,1/2]} g(u)\right]^d \geq \underline{g}_0/2 > 0 \tag{37}$$

for all sufficiently large n, where $\underline{g}_0 = \min_{z \in [0,1]^d} g_0(z) > 0$ by the assumption on g_0.

By (47) in Lemma 6 and non-negativity of the Kullback-Leibler divergence

$$\begin{aligned} d_{KL}(q_j, q_0) &\leq d_{KL}\left(g_0 + \sum_{r=1}^{\bar{m}} w_r^j g_r, g_0\right) \\ &\leq d_{KL}\left(g_0 + \sum_{r=1}^{\bar{m}} w_r^j g_r, g_0\right) + d_{KL}\left(g_0, g_0 + \sum_{r=1}^{\bar{m}} w_r^j g_r\right) \\ &= \int_{\mathbb{R}^d} \log\left(g_0(z) + \sum_{r=1}^{\bar{m}} w_r^j g_r(z)\right)\left(\sum_{r=1}^{\bar{m}} w_r^j g_r(z)\right) dz \\ &= \int_{[0,1]^d} \log\left(g_0(z) + \sum_{r=1}^{\bar{m}} w_r^j g_r(z)\right)\left(\sum_{r=1}^{\bar{m}} w_r^j g_r(z)\right) dz, \end{aligned} \tag{38}$$

where the last equality follows from $g_r(z) = 0$ outside $[0,1]^d$. The integrand of the last integral is bounded above by $2\underline{g}_0^{-1}\left(\sum_{r=1}^{\bar{m}} w_r^j g_r(z)\right)^2$, which follows from the logarithm inequality, $1 - 1/u \leq \log u \leq u - 1$, $\forall u > 0$, and (37). Thus,

$$\begin{aligned} d_{KL}(q_j, q_0) &\leq 2\underline{g}_0^{-1} \int \left[\sum_{r=1}^{\bar{m}} w_r^j g_r(z)\right]^2 dz \\ &= 2\underline{g}_0^{-1} \int \sum_{r=1}^{\bar{m}} w_r^j (g_r(z))^2 dz \\ &\leq 2\underline{g}_0^{-1}\bar{m} \int (g_1(z))^2 dz = 2\underline{g}_0^{-1}\Gamma_n^2 \prod_i (m_i h_i) \left[\int_{-1/2}^{1/2} g(u)^2 du\right]^d \\ &\leq 2\underline{g}_0^{-1}\Gamma_n^2 \left[\int_{-1/2}^{1/2} g(u)^2 du\right]^d \leq 2\underline{g}_0^{-1}\Gamma_n^2 c_0^{2d}, \end{aligned} \tag{39}$$

where the first equality holds since $g_r(z) g_l(z) = 0, \forall r \neq l$. Finally,

$$\bar{m} = \prod_{i=1}^d m_i \geq 2^{-d} \prod_{i=1}^d h_i^{-1}$$

$$
\begin{aligned}
&= 2^{-d}\prod_{i\in J_*}(N_i/2)\cdot\prod_{i\in J_*^c, i\le d_y}\left(\Gamma_n^{-\beta_i^{-1}}/\varrho_i\right)\cdot\prod_{i\in J_*^c, i> d_y}\left(\Gamma_n^{-\beta_i^{-1}}\right)\\
&\ge 2^{-d}\prod_{i\in J_*}(N_i/2)\cdot\prod_{i\in J_*^c, i\le d_y}\left(\Gamma_n^{-\beta_i^{-1}}/2\right)\cdot\prod_{i\in J_*^c, i> d_y}\left(\Gamma_n^{-\beta_i^{-1}}\right)\\
&= 2^{-d-d_y}\cdot N_{J_*}\cdot\Gamma_n^{-\beta_{J_*^c}^{-1}} = 2^{-d-d_y}n\Gamma_n^2\\
&\ge 2^{-d-d_y}n\cdot d_{KL}(q_j,q_0)/(2\underline{g}_0^{-1}c_0^{2d}),
\end{aligned}
$$

where the first inequality holds by definitions of $\bar{m}$ and m_i, the second equality by definition of h_i, the second inequality by restrictions on ϱ_i, and the last inequality by (39). The last inequality implies (36) if

$$
c_0 \le [\underline{g}_0 2^{-(d+d_y+7)}\log 2]^{1/(2d)}.
$$

Q.E.D.

LEMMA 5 *For $j\in\{1,\ldots,M\}$, a part of the density in the definition of q_j, $f_j = \sum_{r=1}^{\bar{m}} w_r^j g_r \in \mathcal{C}^{\beta_1^*,\ldots,\beta_d^*,L}$ with $L=1$ for any sufficiently small constant c_0 in the definition of g.*

PROOF: Consider $k=(k_1,\ldots,k_d)$ and $z,\Delta z\in\mathbb{R}^d$ such that for some $i\in\{1,\ldots,d\}$, $\Delta z_i\neq 0$, for any $l\neq i$, $\Delta z_l = 0$, $\sum_{l=1}^d k_l/\beta_l^* < 1$, and $\sum_{l=1}^d k_l/\beta_l^* + 1/\beta_i^* \ge 1$ so that

$$
0 \le \beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*) \le 1. \tag{40}
$$

For $r(\cdot)$ defined in (34),

$$
\begin{aligned}
D^k f_j(z) &= w_{r(z)}\Gamma_n\prod_{l=1}^d g^{(k_l)}((z_l-c_l^{r(z)})/h_l)/h_l^{k_i}\\
&= B_i\cdot w_{r(z)} h_i^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)}\prod_{l=1}^d g^{(k_l)}((z_l-c_l^{r(z)})/h_l),
\end{aligned}\tag{41}
$$

where $B_i\in\{1,1/2,\varrho_i^{-\beta_i^*}\}\subset(0,1]$. From Tsybakov (2008), (2.33)-(2.34), for any sufficiently small c_0 and $s\le \max_l \beta_l^*+1$,

$$
\max_z |g^{(s)}(z)| \le 1/8. \tag{42}
$$

This imply that

$$
|g^{(k_i)}((z_i+\Delta z_i - c_i^r)/h_i) - g^{(k_i)}((z_i-c_i^r)/h_i)| \le |\Delta z_i|/(8h_i). \tag{43}
$$

First, let us consider the case when $r(z) = r(z + \Delta z)$ and $|\Delta z_i| \le h_i$. From (41), (42), and (43),

$$|D^k f_j(z + \Delta z) - D^k f_j(z)| \le h_i^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)} 8^{-d} |\Delta z_i / h_i|$$
$$= 8^{-d} |\Delta z_i|^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)} \left| \frac{\Delta z_i}{h_i} \right|^{1-\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)}$$
$$\le |\Delta z_i|^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)}, \tag{44}$$

where the last inequality follows from $\Delta z_i \le h_i$ and (40).

Second, consider the case when $r(z) = r(z + \Delta z)$ and $|\Delta z_i| > h_i$. Similarly to the previous case but without using (43),

$$|D^k f_j(z + \Delta z) - D^k f_j(z)| \le 2 \cdot 8^{-d} h_i^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)} \le |\Delta z_i|^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)}.$$

Third, consider the case when $r(z) \ne r(z+\Delta z)$ and $|\Delta z_i| \le h_i/2$. If $w_{r(z)} = w_{r(z+\Delta z)} = 0$ or $z, z + \Delta z \notin \cup_{r=1}^{\bar{m}} B_r$

$$|D^k f_j(z + \Delta z) - D^k f_j(z)| = D^k f_j(z + \Delta z) = D^k f_j(z) = 0.$$

If $w_{r(z)} \ne w_{r(z+\Delta z)}$ or if one of z and $z + \Delta z$ is not in $\cup_{r=1}^{\bar{m}} B_r$, then without a loss of generality suppose that $w_{r(z)} = 1$ or that $z + \Delta z \notin \cup_{r=1}^{\bar{m}} B_r$. Let $|\Delta z_i^\star| \in [0, |\Delta z_i|]$ and $\Delta z^\star = (0, \ldots, 0, \Delta z_i^\star, 0, \ldots, 0)$ be such that $z + \Delta z^\star$ is a boundary point of $B_{r(z)}$. Then, $D^k f_j(z + \Delta z^\star) = 0$ and (44) imply

$$|D^k f_j(z + \Delta z) - D^k f_j(z)| = |D^k f_j(z)| = |D^k f_j(z + \Delta z^\star) - D^k f_j(z)|$$
$$\le |\Delta z_i^\star|^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)} \le |\Delta z_i|^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)}.$$

If $w_{r(z)} = w_{r(z+\Delta z)} = 1$ and $z, z + \Delta z \in \cup_{r=1}^{\bar{m}} B_r$ then by construction of f_j and g

$$|D^k f_j(z + \Delta z) - D^k f_j(z)| = |D^k f_j(z + \Delta z + 0.5h_i) - D^k f_j(z + 0.5h_i)|$$
$$\le |\Delta z_i|^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)},$$

where the last inequality follows from (44).

Finally, when $r(z) \ne r(z + \Delta z)$ and $\Delta z_i > h_i/2$,

$$|D^k f_j(z + \Delta z) - D^k f_j(z)| \le |D^k f_j(z + \Delta z)| + |D^k f_j(z)|$$
$$\le 2 \cdot 8^{-d} h_i^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)}$$
$$\le |\Delta z_i|^{\beta_i^*(1-\sum_{l=1}^d k_l/\beta_l^*)}.$$

Now, let us consider a general Δz such that for $\Delta z_i \neq 0$, $\sum_{l=1}^d k_l/\beta_l^* + 1/\beta_i^* \geq 1$.

$$\begin{aligned}
&|D^k f_j(z+\Delta z) - D^k f_j(z)| \\
&\leq \sum_{i=1}^d \big| D^k f_j(z_1,\dots,z_{i-1},z_i+\Delta z_i,\dots,z_d+\Delta z_d) \\
&\qquad - D^k f_j(z_1,\dots,z_i,z_{i+1}+\Delta z_{i+1},\dots,z_d+\Delta z_d)\big|.
\end{aligned}$$

The preceding argument applies to every term in this sum and, thus, $f_j \in \mathcal{C}^{\beta_1^*,\dots,\beta_d^*,1}$.

Q.E.D.

LEMMA 6 *Let $f_i : \tilde{\mathcal{Y}} \times \mathcal{X} \to \mathbb{R}$, $i \in \{1,2\}$, be densities with respect to a product measure $\lambda \times \mu$ on $\tilde{\mathcal{Y}} \times \mathcal{X} \subset \mathbb{R}^d$. For a finite set $\mathcal{Y}$, let $\{A_y,\, y \in \mathcal{Y}\}$ be a partition of $\tilde{\mathcal{Y}}$ and let $p_i(y,x) = \int_{A_y} f_i(\tilde{y},x) d\lambda(\tilde{y})$. Then,*

$$d_{L_1}(p_1,p_2) \leq d_{L_1}(f_1,f_2) \tag{45}$$

$$d_H(p_1,p_2) \leq d_H(f_1,f_2) \tag{46}$$

$$d_{KL}(p_1,p_2) \leq d_{KL}(f_1,f_2). \tag{47}$$

Also, if for given (y,x), $f_2(\tilde{y},x) > 0$ for any $\tilde{y} \in A_y$, then

$$\inf_{\tilde{y}\in A_y} \frac{f_1(\tilde{y},x)}{f_2(\tilde{y},x)} \leq \frac{p_1(y,x)}{p_2(y,x)} \leq \sup_{\tilde{y}\in A_y} \frac{f_1(\tilde{y},x)}{f_2(\tilde{y},x)}. \tag{48}$$

PROOF: Trivially,

$$\begin{aligned}
d_{L_1}(p_1,p_2) &= \sum_y \int \left| \int_{A_y} (f_1(\tilde{y},x) - f_2(\tilde{y},x)) d\tilde{y} \right| d\mu(x) \\
&\leq \sum_y \int \int_{A_y} |f_1(\tilde{y},x) - f_2(\tilde{y},x)| d\lambda(\tilde{y}) d\mu(x) = d_{L_1}(f_1,f_2).
\end{aligned}$$

By Holder inequality,

$$\begin{aligned}
d_H(p_1,p_2) &= 2\left(1 - \sum_y \int \sqrt{\int 1_{A_y}(\tilde{y}_1) f_1(\tilde{y}_1,x) d\lambda(\tilde{y}_1) \cdot \int 1_{A_y}(\tilde{y}_2) f_2(\tilde{y}_2,x) d\lambda(\tilde{y}_2) d\mu(x)}\right) \\
&\leq 2\left(1 - \sum_y \int \int 1_{A_y}(\tilde{y}) \sqrt{f_1(\tilde{y},x) f_2(\tilde{y},x)} d\lambda(\tilde{y}) d\mu(x)\right) = d_H(f_1,f_2).
\end{aligned}$$

For fixed (y,x),

$$\int_{A_y} (f_1(\tilde{y},x)/p_1(y,x)) \log \frac{f_1(\tilde{y},x)/p_1(y,x)}{f_2(\tilde{y},x)/p_2(y,x)} d\lambda(\tilde{y}) \geq 0$$

since the Kullback-Leibler divergence is nonnegative. Thus,

$$\int_{A_y} f_1(\tilde{y},x)\log\frac{f_1(\tilde{y},x)}{f_2(\tilde{y},x)}d\lambda(\tilde{y}) \geq \int_{A_y} f_1(\tilde{y},x)\log\frac{p_1(y,x)}{p_2(y,x)}d\lambda(\tilde{y}) = p_1(y,x)\log\frac{p_1(y,x)}{p_2(y,x)}.$$

This inequality integrated with respect to $d\mu(x)$ and summed over y implies (47). The last claim follows from

$$f_2(\tilde{y},x)\inf_{\tilde{z}\in A_y}\frac{f_1(\tilde{z},x)}{f_2(\tilde{z},x)} \leq f_1(\tilde{y},x) \leq f_2(\tilde{y},x)\sup_{\tilde{z}\in A_y}\frac{f_1(\tilde{z},x)}{f_2(\tilde{z},x)}.$$

Q.E.D.

LEMMA 7 *For Γ_n, h_i, ϱ_i, and β_i^* defined in Section 4.2, (i) $\beta_i^* \geq \beta_i$ for $i = 1,\ldots,d$ and (ii) $\varrho_i \in (1,2]$ for $i \in J_*^c \cap \{1,\ldots,d_y\}$.*

PROOF: For $i \notin J_*$, $\beta_i^* = \beta_i$ by definition. For $i \in J_*$, from the definition of Γ_n,

$$\Gamma_n \leq \left[\frac{N_{J_*}/N_i}{n}\right]^{\frac{1}{2+\beta_{J_*^c}^{-1}+\beta_i^{-1}}} = \Gamma_n^{\frac{2+\beta_{J_*^c}^{-1}}{2+\beta_{J_*^c}^{-1}+\beta_i^{-1}}} N_i^{\frac{-1}{2+\beta_{J_*^c}^{-1}+\beta_i^{-1}}},$$

which implies $N_i^{-\beta_i} \geq \Gamma_n$. By the definition of β_i^*, $N_i^{-\beta_i^*} = \Gamma_n$ and, thus, $\beta_i^* \geq \beta_i$.

For $i \in J_*^c$, from the definition of Γ_n,

$$\left[\frac{N_{J_*}N_i}{n}\right]^{\frac{1}{2+\beta_{J_*^c}^{-1}-\beta_i^{-1}}} \geq \left[\frac{N_{J_*}}{n}\right]^{\frac{1}{2+\beta_{J_*^c}^{-1}}},$$

which implies

$$N_i \geq \left[\frac{N_{J_*}}{n}\right]^{\frac{2+\beta_{J_*^c}^{-1}-\beta_i^{-1}}{2+\beta_{J_*^c}^{-1}}} = \Gamma_n^{-\beta_i^{-1}} \implies$$
$$\Gamma_n^{\beta_i^{-1}} \geq \frac{1}{N_i},$$

and, therefore, $\Gamma_n^{\beta_i^{-1}} N_i \geq 1$. Next, define

$$\varrho_i = \frac{\left\lfloor \Gamma_n^{\beta_i^{-1}} N_i/2 \right\rfloor + 1}{\Gamma_n^{\beta_i^{-1}} N_i/2}.$$

Then $\varrho_i \in (1,2]$ as $\Gamma_n^{\beta_i^{-1}} N_i \geq 1$. *Q.E.D.*

D.3. *Proofs of Posterior Contraction Results*

D.3.1. *Proof of Theorem 4 for $J^c \neq \emptyset$*

Define $\beta = d_{J^c}\left[\sum_{k\in J^c}\beta_k^{-1}\right]^{-1}$, $\beta_{\min} = \min_{j\in J^c}\beta_j$, and $\sigma_n = [\tilde\epsilon_n/\log(1/\tilde\epsilon_n)]^{1/\beta}$. For ε defined in (22)-(23), b and τ defined in (17), and a sufficiently small $\delta > 0$, let $a_0 = \{(8\beta + 4\varepsilon + 8 + 8\beta/\beta_{\min})/(b\delta)\}^{1/\tau}$, $a_{\sigma_n} = a_0\{\log(1/\sigma_n)\}^{1/\tau}$, and $b_1 > \max\{1, 1/2\beta\}$ satisfying $\tilde\epsilon_n^{b_1}\{\log(1/\tilde\epsilon_n)\}^{5/4} \leq \tilde\epsilon_n$. Then, the proofs of Theorems 4 and 6 in Shen et al. (2013) imply the following two claims for each $y_J = k \in \mathcal{Y}_J$ under the assumptions of Section C.1.

First, there exists a partition $\{U_{j|k}, j = 1,\dots,K\}$ of $\{\tilde x \in \tilde{\mathcal{X}} : ||\tilde x|| \leq 2a_{\sigma_n}\}$, such that for $j = 1,\dots,N$, $U_{j|k}$ is contained within an ellipsoid with center $\mu^\star_{j|k}$ and radii $\{\sigma_n^{\beta/\beta_i}\tilde\epsilon_n^{2b_1}, i \in J^c\}$

$$U_{j|k} \subset \left\{\tilde x : \sum_{i=1}^{d_{J^c}}\left[(\tilde x_i - \mu^\star_{j|k,i})/(\sigma_n^{\beta/\beta_{d_J+i}}\tilde\epsilon_n^{2b_1})\right]^2 \leq 1\right\};$$

for $j = N+1,\dots,K$, $U_{j|k}$ is contained within an ellipsoid with radii $\{\sigma_n^{\beta/\beta_i}, i \in J^c\}$, and $1 \leq N < K \leq C_1\sigma_n^{-d_{J^c}}\{\log(1/\tilde\epsilon_n)\}^{d_{J^c}+d_{J^c}/\tau}$, where $C_1 > 0$ does not depend on n and y_J.

Second, for each $k \in \mathcal{Y}_J$ there exist $\alpha^\star_{j|k}$, $j = 1,\dots,K$, with $\alpha^\star_{j|k} = 0$ for $j > N$, and $\mu^{x\star}_{jk} \in U_{j|k}$ for $j = N+1,\dots,K$ such that for a positive constant C_2 and $\sigma^\star_{J^c} = \{\sigma_n^{\beta/\beta_i}$ for $i \in J^c\}$,

$$d_H\left(f_{0|J}(\cdot|k), f^\star_{|J}(\cdot|k)\right) \leq C_2\sigma_n^\beta, \tag{49}$$

where $f^\star_{|J}$ is defined in (33). Constant C_2 is the same for all $k \in \mathcal{Y}_J$ since all the bounds on $f_{0|J}$ assumed in Section C.1 are uniform over k.

Note also that our smoothness definition is different from the one used by Shen et al. (2013). In Lemmas 8 and 9 we show that our smoothness definition ($f_{0|J} \in \mathcal{C}^{L,\beta_{d_J+1},\dots,\beta_d}$) delivers an anisotropic Taylor expansion with bounds on remainder terms such that the argument on p. 637 of Shen et al. (2013) goes through.

Third, by Lemma 12, which is an extension of a part of Proposition 1 in Shen et al. (2013), there exists a constant $B_0 > 0$ such that for all $y_J \in \mathcal{Y}_J$

$$F_{0|J}\left(||\tilde X|| > a_{\sigma_n}|y_J\right) \leq B_0\sigma_n^{4\beta+2\varepsilon}\underline{\sigma}_n^8, \tag{50}$$

where

$$\underline{\sigma}_n = \min_{i\in J^c}\sigma_n^{\beta/\beta_i}.$$

For $m = N_J K$ we define $\theta^\star$ and $S_{\theta^\star}$ as:

$$\theta^\star = \Big\{ \{\mu_1^\star, \ldots, \mu_m^\star\} = \left\{(k, \mu_{j|k}^\star),\ j = 1, \ldots, K,\ k \in \mathcal{Y}_J\right\},$$
$$\{\alpha_1^\star, \ldots, \alpha_m^\star\} = \left\{\alpha_{jk}^\star = \alpha_{j|k}^\star \pi_{0J}(k),\ j = 1, \ldots, K,\ k \in \mathcal{Y}_J\right\},$$
$$\sigma_J^{\star 2} = \{\sigma_i^{\star 2} = 1/[64 N_i^2 \beta \log(1/\sigma_n)],\ i \in J\}$$
$$\sigma_{J^c}^\star = \{\sigma_i^\star = \sigma_n^{\beta/\beta_i},\ i \in J^c\}, \Big\}$$

$$S_{\theta^\star} = \Big\{ \{\mu_1, \ldots, \mu_m\} = \{(\mu_{jk,J}, \mu_{jk,J^c}),\ j = 1, \ldots, K,\ k \in \mathcal{Y}_J\},$$
$$\mu_{jk,J^c} \in U_{j|k},\quad \mu_{jk,i} \in \left[k_i - \frac{1}{4N_i}, k_i + \frac{1}{4N_i}\right],\ i \in J,$$
$$\sigma_i^2 \in \left(0, \sigma_i^{\star 2}\right),\ i \in J,$$
$$\sigma_i^2 \in \left(\sigma_i^{\star 2}(1 + \sigma_n^{2\beta})^{-1}, \sigma_i^{\star 2}\right),\ i \in J^c,$$
$$(\alpha_1, \ldots, \alpha_m) = \{\alpha_{jk},\ j = 1, \ldots, K,\ k \in \mathcal{Y}_J\} \in \Delta^{m-1},$$
$$\sum_{r=1}^{m} |\alpha_r - \alpha_r^\star| \le 2\sigma_n^{2\beta}, \qquad \min_{j \le K, k \in \mathcal{Y}_J} \alpha_{jk} \ge \frac{\sigma_n^{2\beta + d_{J^c}}}{2m^2} \Big\},$$

where Δ^{m-1} denotes the m-dimensional simplex.

The rest of the proof of the Kullback-Leibler thickness condition follows the general argument developed for mixture models in Ghosal and van der Vaart (2007) and Shen et al. (2013) among others. First, we will show that for $m = N_J K$ and $\theta \in S_{\theta^\star}$, the Hellinger distance $d_H^2(p_0(\cdot,\cdot), p(\cdot,\cdot|\theta, m))$ can be bounded by $\sigma_n^{2\beta}$ up to a multiplicative constant. Second, we construct bounds on the ratios $p(\cdot,\cdot|\theta, m)/p_0(\cdot,\cdot)$ and combine them with the bound on the Hellinger distance using Lemma 11. Finally, we will show that the prior puts sufficient probability on $m = N_J K$ and $S_{\theta^\star}$.

For $f_{|J}^\star$ defined in (33), let us define

$$p_{|J}^\star(y_I, x|y_J) = \int_{A_{y_I}} f_{|J}^\star(\tilde{y}_I, x|y_J) d\tilde{y}_I.$$

For $m = N_J K$ and $\theta \in S_{\theta^\star}$, we can bound the Hellinger distance between the DGP and the model as follows,

$$d_H^2(p_0(\cdot,\cdot), p(\cdot,\cdot|\theta, m)) = d_H^2(p_{0|J}(\cdot|\cdot)\pi_0(\cdot), p(\cdot,\cdot|\theta, m))$$
$$\le d_H^2(p_{0|J}(\cdot|\cdot)\pi_{0J}(\cdot), p_{|J}^\star(\cdot|\cdot)\pi_{0J}(\cdot)) + d_H^2(p_{|J}^\star(\cdot|\cdot)\pi_{0J}(\cdot), p(\cdot,\cdot|\theta, m)).$$

It follows from (49) and Lemma 6 linking distances between probability mass functions and corresponding latent variable densities that the first term on the right hand side of this inequality is bounded by $(C_2)^2\sigma_n^{2\beta}$. Combining this result with the bound on $d_H^2(p^\star_{|J}(\cdot|\cdot)\pi_{0J}(\cdot), p(\cdot,\cdot|\theta,m))$ from Lemma 13 we obtain

$$(51) \qquad d_H^2(p_0(\cdot,\cdot), p(\cdot,\cdot|\theta,m)) \lesssim \sigma_n^{2\beta},$$

where "$\lesssim$" denotes less or equal up to a multiplicative positive constant relation.

Next, for $\theta \in S_{\theta^\star}$ and $m = N_J K$, let us consider lower bounds on the ratio $p(y_J, y_I, x|\theta, m)/p_0(y_J, y_I, x)$. In Lemma 16, we show that lower bounds on the ratio $f_J(y_J, \tilde{x}|\theta, m)/f_{0|J}(\tilde{x}|y_J)\pi_0(y_J)$ imply the following bounds for all sufficiently large n: for any $x \in \mathcal{X}$ with $\|x\| \le a_{\sigma_n}$,

$$(52) \qquad \frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} \ge C_3 \frac{\sigma_n^{2\beta}}{2m^2} \equiv \lambda_n,$$

for some constant $C_3 > 0$; and for any $x \in \mathcal{X}$ with $\|x\| > a_{\sigma_n}$,

$$(53) \qquad \frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} \ge \exp\left\{-\frac{8||x||^2}{\sigma_n^2} - C_4 \log n\right\},$$

for some constant $C_4 > 0$. Consider all sufficiently large n such that $\lambda_n < e^{-1}$ and (52) and (53) hold. Then, for any $\theta \in S_{\theta^\star}$,

$$\begin{aligned}
(54) \qquad & \sum_{y\in\mathcal{Y}} \int_{\mathcal{X}} \left(\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right)^2 \mathbf{1}\left\{\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} < \lambda_n\right\} p_0(y_J, y_I, x) dx \\
&= \sum_{y\in\mathcal{Y}} \int_{\tilde{\mathcal{X}}} \left(\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right)^2 \\
&\qquad \mathbf{1}\left\{\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} < \lambda_n\right\} \mathbf{1}\left\{\tilde{y}_I \in A_{y_I}\right\} f_{0J}(y_J, \tilde{x}) d\tilde{x} \\
&= \sum_{y\in\mathcal{Y}} \int_{\tilde{\mathcal{X}}} \left(\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right)^2 \\
&\qquad \mathbf{1}\left\{\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} < \lambda_n, ||x|| > a_{\sigma_n}, \tilde{y}_I \in A_{y_I}\right\} f_{0J}(y_J, \tilde{x}) d\tilde{x} \\
&\le \sum_{y\in\mathcal{Y}} \int_{\{\tilde{x}: ||x|| > a_{\sigma_n}\}} \left(\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right)^2 \mathbf{1}\left\{\tilde{y}_I \in A_{y_I}\right\} f_{0J}(y_J, \tilde{x}) d\tilde{x} \\
&\le \sum_{y\in\mathcal{Y}} \int_{\{\tilde{x}: ||x|| > a_{\sigma_n}\}} \left[\frac{128}{\sigma_n^4}||x||^4 + 2(C_4 \log n)^2\right] f_{0|J}(\tilde{x}|y_J) \mathbf{1}\left\{\tilde{y}_I \in A_{y_I}\right\} d\tilde{x} \pi_{0J}(y_J) \\
&\le \sum_{y_J\in\mathcal{Y}_J} \int_{\{\tilde{x}: ||\tilde{x}|| > a_{\sigma_n}\}} \left[\frac{128}{\sigma_n^4}||\tilde{x}||^4 + 2(C_4 \log n)^2\right] f_{0|J}(\tilde{x}|y_J) d\tilde{x} \pi_{0J}(y_J)
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{128}{\underline{\sigma}_n^4} \sum_{y_J \in \mathcal{Y}_J} E_{0|y_J}\left(\left\|\tilde{X}\right\|^8\right)^{1/2} \left(F_{0|y_J}\left(\left\|\tilde{X}\right\| > a_{\sigma_n}\right)\right)^{1/2} \pi_{0J}(y_J) \\
&\quad + 2(C_4 \log n)^2 B_0 \sigma_n^{4\beta+2\varepsilon} \underline{\sigma}_n^8 \\
&\leq C_5 \sigma_n^{2\beta+\varepsilon}
\end{aligned}
$$

for some constant $C_5 > 0$ and all sufficiently large n, where the last inequality holds by the tail condition in (17), (50), and $(\log n)^2 \sigma_n^{2\beta+\varepsilon} \underline{\sigma}_n^8 \to 0$.

Furthermore, as $\lambda_n < e^{-1}$,

$$
\begin{aligned}
&\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)} \mathbf{1}\left\{\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} < \lambda_n\right\} \\
&\leq \left(\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right)^2 \mathbf{1}\left\{\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} < \lambda_n\right\}
\end{aligned}
$$

and, therefore,

$$
\begin{aligned}
(55) \qquad &\sum_{y \in \mathcal{Y}} \int_{\mathcal{X}} \log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)} \mathbf{1}\left\{\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} < \lambda_n\right\} p_0(y_J, y_I, x) dx \\
&\leq C_5 \sigma_n^{2\beta+\varepsilon}.
\end{aligned}
$$

Inequalities (51), (54), and (55) combined with Lemma 11 imply

$$
E_0\left(\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right) \leq A\tilde{\epsilon}_n^2, \; E_0\left(\left[\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right]^2\right) \leq A\tilde{\epsilon}_n^2
$$

for any $\theta \in S_{\theta^\star}$, $m = N_J K$, and some positive constant A (details are provided in Lemma 17).

By Lemma 18 for all sufficiently large n, $s = 1 + 1/\beta + 1/\tau$, and some $C_6 > 0$,

$$
\begin{aligned}
&\Pi(\mathcal{K}(p_0, \tilde{\epsilon}_n)) \geq \Pi(m = N_J K, \theta \in S_{\theta^\star}) \\
&\geq \exp\left[-C_6 N_J \tilde{\epsilon}_n^{-d_{J^c}/\beta} \{\log(n)\}^{d_{J^c} s + \max\{\tau_1, 1, \tau_2/\tau\}}\right].
\end{aligned}
$$

The last expression of the above display is bounded below by $\exp\{-Cn\tilde{\epsilon}_n^2\}$ for any $C > 0$, $\tilde{\epsilon}_n = \left[\frac{N_J}{n}\right]^{\beta/(2\beta+d_{J^c})} (\log n)^{t_J}$, any $t_J > (d_{J^c} s + \max\{\tau_1, 1, \tau_2/\tau\})/(2 + d_{J^c}/\beta)$, and all sufficiently large n. Since the inequality in the definition of t_J is strict, the claim of the theorem follows.

When $J = \emptyset$ and $N_J = 1$, the preceding argument delivers the claim of the theorem if we add an artificial discrete coordinate with only one possible value to the vector of observables.

D.3.2. *Proof of Theorem 4 for $J^c = \emptyset$*

In this case, the proof from the previous subsection can be simplified as follows. For $m = N_J$ and for any $\beta > 0$ we define $\theta^\star$ and $S_{\theta^\star}$ as

$$\theta^\star = \Big\{ \{\mu_1^\star, \dots, \mu_m^\star\} = \{k,\ k \in \mathcal{Y}_J\}, \\ \{\alpha_1^\star, \dots, \alpha_m^\star\} = \{\alpha_k^\star,\ k \in \mathcal{Y}_J\} = \{\pi_0(k)\}_{k \in \mathcal{Y}_J}, \\ \sigma^{\star 2} = \{\sigma_i^{\star 2} = \frac{1}{64 N_i^2 \beta \log(1/\sigma_n)},\ i \in J\} \Big\},$$

$$S_{\theta^\star} = \Big\{ \{\mu_1, \dots, \mu_m\} = \{\mu_k,\ k \in \mathcal{Y}_J\},\ \mu_{k,i} \in \left[k_i - \frac{1}{4N_i}, k_i + \frac{1}{4N_i}\right],\ i \in J, \\ \sigma = \{\sigma_i \in (0, \sigma_i^\star),\ i \in J\}, \\ \{\alpha_j,\ j = 1, \dots, m\} = \{\alpha_k,\ k \in \mathcal{Y}_J\} \in \Delta^{m-1}, \\ \sum_{k \in \mathcal{Y}_J} |\alpha_k - \alpha_k^\star| \leq 2\sigma_n^{2\beta}, \qquad \min_{k \in \mathcal{Y}_J} \alpha_k \geq \frac{\sigma_n^{2\beta}}{2m^2} \Big\}.$$

For $m = N_J$ and $\theta \in S_{\theta^\star}$, a simplification of the proof of Lemma 13 delivers

$$d_H^2(p_0(\cdot), p(\cdot|\theta, m)) \leq 2 \max_{k \in \mathcal{Y}_J} \int_{A_k^c} \phi(\tilde{y}_J; \mu_k, \sigma) d\tilde{y}_J + \sum_{k \in \mathcal{Y}_J} |\alpha_k^\star - \alpha_k| \lesssim \sigma_n^{2\beta}.$$

A simplification of derivations in Lemma 16 show that for all $y_J \in \mathcal{Y}_J$

$$\frac{p(y_J|\theta, m)}{p_0(y_J)} \geq \frac{1}{2} \frac{\sigma_n^{2\beta}}{2m^2} \equiv \lambda_n.$$

Then, for any $\theta \in S_{\theta^\star}$

$$\sum_{y_J \in \mathcal{Y}_J} \left(\log \frac{p_0(y_J)}{p(y_J|\theta, m)} \right)^2 \mathbf{1} \left\{ \frac{p(y_J|\theta, m)}{p_0(y_J)} < \lambda_n \right\} p_0(y_J) = 0$$

$$\sum_{y_J \in \mathcal{Y}_J} \left(\log \frac{p_0(y_J)}{p(y_J|\theta, m)} \right) \mathbf{1} \left\{ \frac{p(y_J|\theta, m)}{p_0(y_J)} < \lambda_n \right\} p_0(y_J) = 0$$

as $\frac{p(y_J|\theta,m)}{p_0(y_J)} \geq \lambda_n$ for all $y_J \in \mathcal{Y}_J$. As $\lambda_n \to 0$, by Lemma 11 for $\lambda_n < \lambda_0$, both $E_0(\log \frac{p_0(y_J)}{p(y_J|\theta,m)})$ and $E_0([\log \frac{p_0(y_J)}{p(y_J|\theta,m)}]^2)$ are bounded by $C_7 \log(1/\lambda_n)^2 \sigma_n^{2\beta} \leq A\tilde{\epsilon}_n^2$ for some constant A. By the simplification of Lemma 18 for this particular case for all sufficiently large n and some $C_8 > 0$,

$$\Pi(\mathcal{K}(p_0, \tilde{\epsilon}_n)) \geq \Pi(m = N_J, \theta \in S_{\theta^\star}) \geq \exp\left[-C_8 N_J \{\log(n)\}^{\max\{\tau_1, 1\}}\right].$$

The last expression of the above display is bounded below by $\exp\{-Cn\tilde{\epsilon}_n^2\}$ for any $C > 0$, $\tilde{\epsilon}_n = \left[\frac{N_J}{n}\right]^{1/2} (\log n)^{t_J}$, any $t_J > \max\{\tau_1, 1\}/2$, and all sufficiently large n. Since the inequality in the definition of t_J is strict, the claim of the theorem follows.

D.3.3. *Auxiliary Results for Posterior Contraction Rates*

For a multi-index $k = (k_1, \ldots, k_d) \in \mathbb{Z}_+^d$, let $k! = \prod_{i=1}^d k_i!$, and for $z \in \mathbb{R}^d$, let $z^k = \prod_{i=1}^d z_i^{k_i}$.

LEMMA 8 *(Anisotropic Taylor Expansion) For $f \in \mathcal{C}^{\beta_1,\ldots,\beta_d,L}$ and $r \in \{1,\ldots,d\}$*

$$f(x_1+y_1,\ldots,x_d+y_d) = \sum_{k\in I^r} \frac{y^k}{k!} D^k f(x_1,\ldots,x_r,x_{r+1}+y_{r+1},\ldots,x_d+y_d) \tag{56}$$

$$+\sum_{l=1}^{r}\sum_{k\in \bar{I}^l} \frac{y^k}{k!}\Big(D^k f(x_1,\ldots,x_l+\zeta_l^k, x_{l+1}+y_{l+1},\ldots,x_d+y_d) \tag{57}$$

$$- D^k f(x_1,\ldots,x_l,x_{l+1}+y_{l+1},\ldots,x_d+y_d)\Big), \tag{58}$$

where $\zeta_l^k \in [x_l, x_l+y_l] \cup [x_l+y_l, x_l]$,

$$I^l = \left\{ k = (k_1,\ldots,k_l,0,\ldots,0) \in \mathbb{Z}_+^d : \; k_i \le \Big\lfloor \beta_i(1-\sum_{j=1}^{i-1} k_j/\beta_j)\Big\rfloor_s, \; i = 1,\ldots,l \right\},$$

$$\bar{I}^l = \left\{ k \in I^l : \; k_l = \Big\lfloor \beta_l(1-\sum_{j=1}^{l-1} k_j/\beta_j)\Big\rfloor_s \right\},$$

and the differences in derivatives in (57)-(58) are bounded by

$$L\left|\zeta_l^k\right|^{\beta_l(1-\sum_{i=1}^d k_i/\beta_i)}.$$

PROOF: The lemma is proved by induction. For $r = 1$, (56)-(58) is a standard univariate Taylor expansion of $f(x+y)$ in the first argument around $(x_1, x_2+y_2, \ldots, x_d+y_d)$. Suppose (56)-(58) holds for some $r \in \{1,\ldots,d\}$. Then, let us show that (56)-(58) holds for $r+1$. For that, consider a univariate Taylor expansion of $D^k f$ in (56). The following notation will be useful. Let $e_i \in \mathbb{R}^d$, $i = 1,\ldots,d$, be such that $e_{ij} = 1$ for $i = j$ and $e_{ij} = 0$ for $i \ne j$ and $k_{r+1}^* = \lfloor \beta_{r+1}(1-\sum_{j=1}^r k_j/\beta_j)\rfloor_s$. Then,

$$\begin{aligned}
&D^k f(x_1,\ldots,x_r,x_{r+1}+y_{r+1},\ldots,x_d+y_d) = \\
&\sum_{k_{r+1}=0}^{k_{r+1}^*} \frac{y_{r+1}^{k_{r+1}}}{k_{r+1}!} D^{k+k_{r+1}\cdot e_{r+1}} f(x_1,\ldots,x_{r+1},x_{r+2}+y_{r+2},\ldots,x_d+y_d) \\
&+ \frac{y_{r+1}^{k_{r+1}^*}}{k_{r+1}^*!}\Big(D^{k+k_{r+1}^*\cdot e_{r+1}} f(x_1,\ldots,x_r,x_{r+1}+\zeta_{r+1}^{k+k_{r+1}^*\cdot e_{r+1}}, x_{r+2}+y_{l+2},\ldots,x_d+y_d) \\
&- D^{k+k_{r+1}^*\cdot e_{r+1}} f(x_1,\ldots,x_r,x_{r+1},x_{r+2}+y_{l+2},\ldots,x_d+y_d)\Big).
\end{aligned}$$

Inserting this expansion into (56) delivers the result for $r+1$.

Q.E.D.

Lemma 9 *Let $R(x,y)$ denote the remainder term in the anisotropic Taylor expansion ((57)-(58) for $r=d$). Suppose $f \in \mathcal{C}^{\beta_1,\dots,\beta_d,L}$ and L satisfies (20)-(21). Let $\sigma = \{\sigma_i = \sigma_n^{\beta/\beta_i},\ i=1,\dots,d\}$ and $\sigma_n \to 0$. Then, for all sufficiently large n,*

$$\int |R(x,y)| \phi(y;0,\sigma) dy \lesssim L(x)\sigma_n^{\beta}.$$

Proof: Note that $|R(x,y)|$ is bounded by a sum of the following terms over $k \in \bar{I}^l$ and $l \in \{1,\dots,d\}$

$$\begin{aligned}
&\frac{y^k}{k!}\Big| D^k f(x_1,\dots,x_l+\zeta_l^k, x_{l+1}+y_{l+1},\dots,x_d+y_d) \\
&\qquad - D^k f(x_1,\dots,x_l,x_{l+1}+y_{l+1},\dots,x_d+y_d)\Big| \\
&\leq \frac{y^k}{k!} L\left(x+(0,\dots,0,y_{l+1:d}),\zeta_l^k e_l\right)\left|\zeta_l^k\right|^{\beta_l(1-\sum_{i=1}^d k_i/\beta_i)} \\
&\leq \tilde{L}(x)\exp\left\{\tau_0 ||y_{l+1:d}||^2\right\}\exp\left\{\tau_0||\zeta_l^k||^2\right\}\left|\zeta_l^k\right|^{\beta_l(1-\sum_{i=1}^d k_i/\beta_i)} \\
&\leq \tilde{L}(x)\frac{y^k}{k!}\exp\left\{\tau_0||y||^2\right\}|y_l|^{\beta_l(1-\sum_{i=1}^d k_i/\beta_i)},
\end{aligned}$$

where we used inequalities (4), (20), and (21) and that $\left|\zeta_l^k\right| \leq |y_l|$.

For all sufficiently large n such that $\tau_0 < 0.5/\max_i \sigma_i^2$,

$$\begin{aligned}
&\int \left|\tilde{L}(x)\frac{y^k}{k!}\exp\left\{\tau_0||y||^2\right\}|y_l|^{\beta_l(1-\sum_{i=1}^d k_i/\beta_i)}\right|\phi(y;0,\sigma)dy \\
&\lesssim \tilde{L}(x)\prod_{i=1}^{l-1}\int |y_i|^{k_i}\phi(y_i;0;\sigma_i\sqrt{2})dy_i \cdot \int y_l^{k_l}|y_l|^{\beta_l(1-\sum_{i=1}^d k_i/\beta_i)}\phi(y_l;0;\sigma_l\sqrt{2})dy_l \\
&\lesssim \tilde{L}(x)\sigma_1^{k_1}\cdots\sigma_{l-1}^{k_{l-1}}\sigma_l^{k_l+\beta_l(1-\sum_{i=1}^d k_i/\beta_i)} \\
&= \tilde{L}(x)\sigma_n^{k_1\beta/\beta_1}\cdots\sigma_n^{k_l\beta/\beta_l}\sigma_n^{\frac{\beta}{\beta_l}\beta_l(1-\sum_{i=1}^d k_i/\beta_i)} = \tilde{L}(x)K_2\sigma_n^{\beta},
\end{aligned}$$

where we use $\int |z|^{\rho}\phi(z,0,\omega)dz \lesssim \omega^{\rho}$ and $k_{l+1}=\cdots=k_d=0$ for $k \in \bar{I}_l$. Thus, the claim of the lemma follows.

Q.E.D.

Lemma 10 *Suppose density $f_0 \in \mathcal{C}^{\beta_1,\dots,\beta_d,L}$ with a constant envelope L has support on $[0,1]^d$ and $f_0(z) \geq \underline{f} > 0$. Then, $f_{0|J} \in \mathcal{C}^{\beta_{d_{J^c}},\dots,\beta_d,L/\underline{f}}$.*

Proof: For $\tilde{x},\Delta\tilde{x} \in \mathcal{X}$, $y_J \in \mathcal{Y}_J$, and some $\tilde{y}_J^* \in A_{y_J}$, by the mean value theorem,

$$D^k f_{0|J}(\tilde{x}+\Delta\tilde{x}|y_J) - D^k f_{0|J}(\tilde{x}|y_J) =$$

$$= \frac{1}{\pi_{0J}(y_J)} \int_{A_{y_J}} \left(D^{0,\ldots,0,k} f_0(\tilde{y}_J, \tilde{x} + \Delta\tilde{x}) - D^{0,\ldots,0,k} f_0(\tilde{y}_J, \tilde{x}) \right) d\tilde{y}_J$$
$$= \frac{1/N_J}{\pi_{0J}(y_J)} \left(D^{0,\ldots,0,k} f_0(\tilde{y}_J^*, \tilde{x} + \Delta\tilde{x}) - D^{0,\ldots,0,k} f_0(\tilde{y}_J^*, \tilde{x}) \right)$$

and the claim of the lemma follows from the definition of $\mathcal{C}^{\beta_1,\ldots,\beta_d,L}$ and $\pi_{0J}(y_J) \geq \underline{f}/N_J$. *Q.E.D.*

LEMMA 11 *There is a $\lambda_0 \in (0,1)$ such that for any $\lambda \in (0, \lambda_0)$ and any two conditional densities $p, q \in \mathcal{F}$, a probability measure P on $\mathcal{Z}$ that has a conditional density equal to p, and d_h defined with the distribution on $\mathcal{X}$ implied by P,*

$$P \log \frac{p}{q} \leq d_h^2(p,q) \left(1 + 2\log\frac{1}{\lambda}\right) + 2P\left\{ \left(\log\frac{p}{q}\right) 1\left(\frac{q}{p} \leq \lambda\right)\right\},$$

$$P\left(\log\frac{p}{q}\right)^2 \leq d_h^2(p,q)\left(12 + 2\left(\log\frac{1}{\lambda}\right)^2\right) + 8P\left\{\left(\log\frac{p}{q}\right)^2 1\left(\frac{q}{p}\leq\lambda\right)\right\},$$

PROOF: The proof is exactly the same as the proof of Lemma 4 of Shen et al. (2013), which in turn, follows the proof of Lemma 7 in Ghosal and van der Vaart (2007). *Q.E.D.*

LEMMA 12 *Under the assumptions and notation of Section 4.3, for for some $B_0 \in (0, \infty)$ and any $y_J \in \mathcal{Y}_J$,*

$$F_{0|J}\left(||\tilde{X}|| > a_{\sigma_n} \middle| y_J\right) \leq B_0 \sigma_n^{4\beta+2\varepsilon} \underline{\sigma}_n^8.$$

PROOF: Note that in the proof of Proposition 1 of Shen et al. (2013) it is shown that $a_{\sigma_n}^{STG} > a$, where $a_0^{STG} = \{(8\beta + 4\varepsilon + 16)/(b\delta)\}^{1/\tau}$ and $a_{\sigma_n}^{STG} = a_0^{STG} \log(1/\sigma_n)^{1/\tau}$. As $a_0 > a_0^{STG}$ and $a_{\sigma_n} > a_{\sigma_n}^{STG}$, therefore $a_{\sigma_n} > a$. Define $E_{\sigma_n}^* = \left\{\tilde{x} \in \mathbb{R}^{d_{J^c}} \,:\, f_{0|J}(\tilde{x}|y_J) \geq \sigma_n^{(4\beta+2\varepsilon+8\beta/\beta_{\min})/\delta}\right\}$. Note that by construction of s_2 in proof of Proposition 1 of Shen et al. (2013) and as $\sigma_n < s_2$ it follows that

$$\frac{(4\beta + 2\varepsilon + 8)}{b\delta} \log\left(\frac{1}{\sigma_n}\right) \geq \frac{1}{b}\log \bar{f}_0 \implies \sigma_n^{-\frac{(4\beta+2\varepsilon+8)}{\delta}} \geq \bar{f}_0.$$

For $\tilde{x} \in E_{\sigma_n}^*$,

$$f_{0|J}(\tilde{x}|y_J) \geq \sigma_n^{(4\beta+2\varepsilon+8\beta/\beta_{\min})/\delta} = \sigma_n^{(8\beta+4\varepsilon+8\beta/\beta_{\min}+8)/\delta}\sigma_n^{-(4\beta+2\varepsilon+8)/\delta}$$
$$\geq \bar{f}_0 \sigma_n^{(8\beta+4\varepsilon+8\beta/\beta_{\min}+8)/\delta} = \bar{f}_0 \sigma_n^{a_0^\tau b} = \bar{f}_0 \exp\left\{-b a_0^\tau \log(\frac{1}{\sigma_n})\right\}$$
$$= \bar{f}_0 \exp\left\{-b\left(a_0(\log(\frac{1}{\sigma_n})^{1/\tau})\right)^\tau\right\} = \bar{f}_0 \exp\left\{-b a_{\sigma_n}^\tau\right\}.$$

As $a_{\sigma_n} > a$ and as $f_{0|J}(\tilde{x}|y_J) \geq \bar{f}_0 \exp\{-ba^{\tau}_{\sigma_n}\}$, then the tail condition (17) is satisfied only if $||\tilde{x}|| < a_{\sigma_n}$. Therefore, $E^*_{\sigma_n} \subset \left\{\tilde{x} \in \mathbb{R}^{d_J} \,:\, ||\tilde{x}|| \leq a_{\sigma_n}\right\}$. As in the proof of Proposition 1 of Shen et al. (2013), by Markov's inequality,

$$\begin{aligned}
&F_{0|J}\left(||\tilde{X}|| > a_{\sigma_n}|y_J\right) \leq F_{0|J}(E^{*,c}_{\sigma_n}|y_J)\\
&= F_{0|J}\left(f_{0|J}(\tilde{x}|y_J)^{-\delta} > \sigma_n^{-(4\beta+2\varepsilon+8\beta/\beta_{\min})}|y_J\right)\\
&\leq B_0\sigma_n^{4\beta+2\varepsilon+8\beta/\beta_{\min}} = B_0\sigma_n^{4\beta+2\varepsilon}\underline{\sigma}_n^{8}
\end{aligned}$$

as desired since $\sigma_n^{\beta/\beta_{\min}} = \underline{\sigma}_n$ and the tail condition on $f_{0|J}(\cdot|y_J)$, (17), implies the existence of a $\delta > 0$ small enough such that $E_{0|J}(f_{0|J}^{-\delta}) \leq B_0 < \infty$ for any $y_J \in \mathcal{Y}_J$. *Q.E.D.*

LEMMA 13 *Under the assumptions and notation of Section 4.3, for $m = KN_J$ and any $\theta \in S_{\theta^\star}$*

$$d_H^2(p^\star_{|J}(\cdot|\cdot)\pi_0(\cdot), p(\cdot,\cdot|\theta,m)) \lesssim \sigma_n^{2\beta}.$$

PROOF: Let us define

$$f_J(y_J,\tilde{x}|\theta,m) = \int_{A_{y_J}} f(\tilde{y}_J,\tilde{x}|\theta,m)d\tilde{y}_J.$$

Then,

$$\begin{aligned}
&d_H^2(p^\star_{|J}(\cdot|\cdot)\pi_0(\cdot), p(\cdot,\cdot|\theta,m)) \leq d_{L_1}(p^\star_{|J}(\cdot|\cdot)\pi_0(\cdot), p(\cdot,\cdot|\theta,m))\\
&\leq d_{L_1}(f^\star_{|J}(\cdot|\cdot)\pi_0(\cdot), f_J(\cdot,\cdot|\theta,m))\\
&= \sum_{y_J\in\mathcal{Y}_J}\int_{\tilde{\mathcal{X}}}\Bigg|\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^K \alpha^\star_{j|k}\pi_0(k)\mathbf{1}\{k=y_J\}\phi(\tilde{x},\mu^\star_{j|k},\sigma^\star_{J^c})\\
&\qquad\qquad - \alpha_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\cdot\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})\Bigg|d\tilde{x}\\
&\leq \sum_{y_J\in\mathcal{Y}_J}\int_{\tilde{\mathcal{X}}}\left|\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^K \alpha^\star_{jk}\mathbf{1}\{k=y_J\}\phi(\tilde{x},\mu^\star_{j|k},\sigma^\star_{J^c}) - \alpha^\star_{jk}\mathbf{1}\{k=y_J\}\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})\right|d\tilde{x}\\
&+ \sum_{y_J\in\mathcal{Y}_J}\int_{\tilde{\mathcal{X}}}\Bigg|\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^K \alpha^\star_{jk}\mathbf{1}\{k=y_J\}\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})\\
&\qquad\qquad - \alpha_{jk}\int_{A_{y_J}}\phi(\tilde{y},\mu_{jk,J},\sigma_J)d\tilde{y}\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})\Bigg|d\tilde{x},
\end{aligned}$$

where the first inequality follows from $d_H^2(\cdot,\cdot) \leq d_{L_1}(\cdot,\cdot)$, the second inequality holds by Lemma 6, and the last inequality is obtained by the triangle inequality.

Let's explore the two parts of the right hand side in the last inequality independently. First,

$$\sum_{y_J\in\mathcal{Y}_J}\int_{\tilde{\mathcal{X}}}\left|\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\alpha^\star_{jk}\mathbf{1}\{k=y_J\}\phi(\tilde{x},\mu^\star_{j|k},\sigma^\star_{J^c})-\alpha^\star_{jk}\mathbf{1}\{k=y_J\}\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})\right|d\tilde{x}$$
$$\leq\sum_{y_J\in\mathcal{Y}_J}\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\alpha^\star_{jk}\mathbf{1}\{k=y_J\}\int_{\tilde{\mathcal{X}}}\left|\phi(\tilde{x},\mu^\star_{j|k},\sigma^\star_{J^c})-\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})\right|d\tilde{x}$$
$$\leq\max_{j\leq N,k\in\mathcal{Y}_J}d_{L_1}(\phi(\cdot;\mu^\star_{j|k},\sigma^\star_{J^c}),\phi(\cdot,\mu_{jk,J^c},\sigma_{J^c}))\lesssim\sigma_n^{2\beta},$$

where the fact that $\alpha^\star_{j,k}=0$ for $j>N$ by design is used to get $j\leq N$ rather than $j\leq K$ in the max subscript. The last inequality is proved in Lemma 14.

Second,

$$\sum_{y_J\in\mathcal{Y}_J}\int_{\tilde{\mathcal{X}}}\Bigg|\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\alpha^\star_{jk}\mathbf{1}\{k=y_J\}\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})$$
$$-\alpha_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})\Bigg|d\tilde{x}$$
$$=\sum_{j=1}^{K}\left(\sum_{y_J\in\mathcal{Y}_J}\left|\sum_{k\in\mathcal{Y}_J}\alpha^\star_{jk}\mathbf{1}\{k=y_J\}-\alpha_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right|\int_{\tilde{\mathcal{X}}}\phi(\tilde{x},\mu_{jk,J^c},\sigma_{J^c})d\tilde{x}\right)$$
$$=\sum_{j=1}^{K}\sum_{y_J\in\mathcal{Y}_J}\left|\sum_{k\in\mathcal{Y}_J}\alpha^\star_{jk}\mathbf{1}\{k=y_J\}-\alpha_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right|$$
$$\leq\sum_{y_J\in\mathcal{Y}_J}\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\left|\alpha^\star_{jk}\mathbf{1}\{k=y_J\}-\alpha^\star_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right|$$
$$+\sum_{y_J\in\mathcal{Y}_J}\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\left|\alpha^\star_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J-\alpha_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right|$$
$$\leq\sum_{y_J\in\mathcal{Y}_J}\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\alpha^\star_{jk}\left|\mathbf{1}\{k=y_J\}-\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right|$$
$$+\sum_{y_J\in\mathcal{Y}_J}\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}|\alpha^\star_{jk}-\alpha_{jk}|\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J$$
$$=\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\left(\alpha^\star_{jk}\sum_{y_J\in\mathcal{Y}_J}\left|\mathbf{1}\{k=y_J\}-\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right|\right)$$
$$+\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\left(|\alpha^\star_{jk}-\alpha_{jk}|\sum_{y_J\in\mathcal{Y}_J}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right)$$
$$\leq\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\alpha^\star_{jk}\left[\int_{A_k^c}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J+\sum_{y_J\neq k}\int_{A_{y_J}}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\right]$$

$$+\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\left|\alpha^\star_{jk}-\alpha_{jk}\right|$$

$$=\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\alpha^\star_{jk}\cdot 2\int_{A_k^c}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J+\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\left|\alpha^\star_{jk}-\alpha_{jk}\right|$$

$$\leq 2\max_{j\leq N,k\in\mathcal{Y}_J}\int_{A_k^c}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J+\sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\left|\alpha^\star_{jk}-\alpha_{jk}\right|\lesssim\sigma_n^{2\beta}.$$

The last inequality follows from Lemma 15 and the definition of $S_{\theta^\star}$.

Q.E.D.

LEMMA 14 *Under the assumptions and notation of Section 4.3,*

$$\max_{j\leq N,k\in\mathcal{Y}_J} d_{L_1}(\phi(\cdot;\mu^\star_{j|k},\sigma^\star_{J^c}),\phi(\cdot,\mu_{jk,J^c},\sigma_{J^c}))\lesssim\sigma_n^{2\beta}.$$

PROOF: Fix some $j\leq N$ and $k\in\mathcal{Y}_J$. It is known that

$$d_{L_1}(\phi(\cdot;\mu^\star_{j|k},\sigma^\star_{J^c}),\phi(\cdot,\mu_{jk,J^c},\sigma_{J^c}))\leq 2\sqrt{d_{KL}(\phi(\cdot;\mu^\star_{j|k},\sigma^\star_{J^c}),\phi(\cdot,\mu_{jk,J^c},\sigma_{J^c}))}$$

and

$$d_{KL}(\phi(\cdot;\mu^\star_{j|k},\sigma^\star_{J^c}),\phi(\cdot,\mu_{jk,J^c},\sigma_{J^c}))=\sum_{i\in J^c}\frac{\sigma_i^2}{\sigma_i^{\star 2}}-1-\log\frac{\sigma_i^2}{\sigma_i^{\star 2}}+\frac{(\mu^\star_{j|k,i}-\mu_{jk,i})^2}{\sigma_i^{\star 2}}.$$

From the definition of $S_{\theta^\star}$,

$$\sum_{i\in J^c}\frac{(\mu^\star_{j|k,i}-\mu_{jk,i})^2}{\sigma_i^{\star 2}}\leq\tilde{\epsilon}_n^{4b_1}\leq\sigma_n^{4\beta}.$$

Since $\sigma_i^2\in(\sigma_i^{\star 2}(1+\sigma_n^{2\beta})^{-1},\sigma_i^{\star 2})$ and the fact that $|z-1-\log z|\lesssim|z-1|^2$ for z in a neighborhood of 1, we have for all sufficiently large n

$$\left|\frac{\sigma_i^2}{\sigma_i^{\star 2}}-1-\log\frac{\sigma_i^2}{\sigma_i^{\star 2}}\right|\lesssim\left(1-\frac{\sigma_i^2}{\sigma_i^{\star 2}}\right)^2\lesssim\sigma_n^{4\beta}.$$

The three inequalities derived above imply the claim of the lemma.

Q.E.D.

LEMMA 15 *Under the assumptions and notation of Section 4.3, for $\theta\in S_{\theta^\star}$,*

$$\max_{j\leq N,k\in\mathcal{Y}_J}\int_{A_k^c}\phi(\tilde{y}_J,\mu_{jk,J},\sigma_J)d\tilde{y}_J\lesssim\sigma_n^{2\beta}.$$

PROOF: Fix $j \leq N$, $k \in \mathcal{Y}_J$, and $\theta \in S_{\theta^\star}$. Since $\mu_{jk,i} \in \left[k_i - \frac{1}{4N_i}, k_i + \frac{1}{4N_i}\right]$,

$$\int_{A_k^c} \phi(\tilde{y}_J, \mu_{jk,J}, \sigma_J) d\tilde{y}_J \leq \sum_{i \in J} Pr\left(\tilde{y}_i \notin \left[k_i - \frac{1}{2N_i}, k_i + \frac{1}{2N_i}\right]\right)$$
$$\leq \sum_{i \in J} Pr\left(\tilde{y}_i \notin \left[\mu_{jk,i} - \frac{1}{4N_i}, \mu_{jk,i} + \frac{1}{4N_i}\right]\right)$$
$$= 2\sum_{i \in J} \int_{-\infty}^{-\frac{1}{4N_i\sigma_i}} \phi(\tilde{y}_i, 0, 1) d\tilde{y}_i$$
$$\leq 2\sum_{i \in J} \exp\left\{-\frac{1}{2(4N_i\sigma_i)^2}\right\} \leq 2\sum_{i \in J} \sigma_n^{2\beta} \lesssim \sigma_n^{2\beta},$$

where the last inequality follows from the restrictions on σ_J in $S_{\theta^\star}$ and the penultimate inequality follows from a bound on the normal tail probability derived below.

If $\tilde{Y}_i$ has $N(0,1)$ distribution, then the moment generating function is $M(\theta) = \exp\{\theta^2/2\}$. Note that $\exp\{\theta(\tilde{Y}_i - (4N_i\sigma_i)^{-1})\} \geq 1$ when $\tilde{Y}_i \leq (4N_i\sigma_i)^{-1}$ and $\theta \leq 0$, therefore,

$$\int_{-\infty}^{-\frac{1}{4N_i\sigma_i}} \phi(\tilde{y}_i, 0, 1) d\tilde{y}_i \leq \inf_{\theta \leq 0} \mathbb{P} \exp\left\{\theta(\tilde{Y}_i - (4N_i\sigma_i)^{-1})\right\}$$
$$= \inf_{\theta \leq 0} \exp\left\{-\theta(4N_i\sigma_i)^{-1}\right\} M(\theta)$$
$$= \inf_{\theta \leq 0} \exp\left\{-\theta(4N_i\sigma_i)^{-1}\right\} \exp\left\{\theta^2/2\right\} = \exp\left\{-(4N_i\sigma_i)^{-2}/2\right\}.$$

Q.E.D.

LEMMA 16 *Under the assumptions and notation of Section 4.3, for any $(y_J, y_I) \in \mathcal{Y}$, some constants $C_3, C_4 > 0$ and all sufficiently large n,*

$$\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} \geq C_3 \frac{\sigma_n^{2\beta}}{m^2} \equiv \lambda_n, \tag{59}$$

when $\|x\| \leq a_{\sigma_n}$ and

$$\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} \geq \exp\left\{-\frac{8\|x\|^2}{\sigma_n^2} - C_4 \log n\right\} \tag{60}$$

when $\|x\| > a_{\sigma_n}$.

PROOF: By assumption (17), $f_{0|J}(\tilde{x}|y_J) \leq \bar{f}_0$, and $\pi_{0J}(y_J) \leq 1$ for all $(\tilde{x}, y_J)$. Therefore,

$$\frac{f_J(y_J, \tilde{x}|\theta, m)}{f_{0|J}(\tilde{x}|y_J)\pi_{0J}(y_J)} \geq \bar{f}_0^{-1} f_J(y_J, \tilde{x}|\theta, m) \tag{61}$$

Let $k^\star = y_J$. Then, by Lemma 15, for any $j \in \{1, \ldots, K\}$,

$$\int_{A_{y_J}} \phi(\tilde{y}_J; \mu_{jk^*,J}, \sigma_J) d\tilde{y}_J \geq \frac{1}{2}$$

for all n large enough as $\sigma_n \to 0$.

For any $\tilde{x} \in \tilde{\mathcal{X}}$ with $\|\tilde{x}\| \leq 2a_{\sigma_n}$, by the construction of sets $U_{j|k^\star}$, there exists $j^\star \in \{1, \ldots, K\}$ such that $\tilde{x}, \mu_{j^\star|k^\star} \in U_{j^\star|k^\star}$ and for all sufficiently large n, $\sum_{i \in J^c}(\tilde{x}_i - \mu_{j^\star|k^\star,i})^2/\sigma_i^2 \leq 4$. Then,

$$\begin{aligned}
\phi(\tilde{x}, \mu_{j^\star|k^\star}, \sigma_{J^c}) &= (2\pi)^{-d_{J^c}/2} \prod_{i \in J^c} \sigma_i^{-1} \exp\left\{-0.5 \sum_{i \in J^c} (\tilde{x}_i - \mu_{j^\star|k^\star,i})^2/\sigma_i^2\right\} \\
&\geq (2\pi)^{-d_{J^c}/2} \sigma_n^{-d_{J^c}} e^{-2}.
\end{aligned}$$

Thus,

$$\begin{aligned}
f_J(y_J, \tilde{x}|\theta) &= \sum_{k \in \mathcal{Y}_J} \sum_{j=1}^{K} \alpha_{jk} \int_{A_{y_J}} \phi(\tilde{y}_J, \mu_{jk,J}, \sigma_J) d\tilde{y}_J \phi(\tilde{x}, \mu_{jk,J^c}, \sigma_{J^c}) \\
&\geq \alpha_{j^\star k^\star} \phi(\tilde{x}, \mu_{j^\star k^\star, J^c}, \sigma_{J^c}) \int_{A_{y_J}} \phi(\tilde{y}_J, \mu_{j^\star k^\star, J}, \sigma_J) d\tilde{y}_J
\end{aligned}$$

and for $C_3 = \bar{f}_0^{-1}(2\pi)^{-d_{J^c}/2} e^{-2}/8$,

$$\begin{aligned}
\frac{f_J(y_J, \tilde{x}|\theta, m)}{f_{0|J}(\tilde{x}|y_J)\pi_{0J}(y_J)} &\geq \bar{f}_0^{-1} \cdot \min_{j \leq K, k \in \mathcal{Y}_J} \alpha_{jk} \cdot (2\pi)^{-d_{J^c}/2} \sigma_n^{-d_{J^c}} e^{-2} \cdot \frac{1}{2} \\
&\geq 2C_3 \frac{\sigma_n^{2\beta}}{m^2} = 2\lambda_n. \qquad (62)
\end{aligned}$$

By assumption (18), for any $x \in \mathcal{X}$, any $y_J \in \mathcal{Y}_J$, and all sufficiently large n,

$$\int_{A_{y_I}} f_{0|J}(\tilde{x}|y_J)\pi_{0J}(y_J) d\tilde{y}_I \leq 2 \int_{A_{y_I} \cap \{\tilde{y}_I : \|\tilde{y}_I\| \leq a_{\sigma_n}\}} f_{0|J}(\tilde{x}|y_J)\pi_{0J}(y_J) d\tilde{y}_I. \qquad (63)$$

For any $x \in \mathcal{X}$ with $\|x\| \leq a_{\sigma_n}$ and $\tilde{y}_I \in A_{y_I} \cap \{\tilde{y}_I : \|\tilde{y}_I\| \leq a_{\sigma_n}\}$, we have $\|\tilde{x}\| \leq 2a_{\sigma_n}$ and

$$\begin{aligned}
\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} &= \frac{\int_{A_{y_I}} f_J(y_J, \tilde{x}|\theta, m) d\tilde{y}_I}{\int_{A_{y_I}} f_{0|J}(\tilde{x}|y_J)\pi_{0J}(y_J) d\tilde{y}_I} \\
&\geq \frac{\int_{A_{y_I} \cap \{\tilde{y}_I : \|\tilde{y}_I\| \leq a_{\sigma_n}\}} f_J(y_J, \tilde{x}|\theta, m) d\tilde{y}_I}{2\int_{A_{y_I} \cap \{\tilde{y}_I : \|\tilde{y}_I\| \leq a_{\sigma_n}\}} f_{0|J}(\tilde{x}|y_J)\pi_{0J}(y_J) d\tilde{y}_I} \geq \lambda_n, \qquad (64)
\end{aligned}$$

where the first inequality follows from (63) and the second one from (62) combined with Lemma 6.

Next, let us bound $f_J(y_J, \tilde{x}|\theta, m)/f_{0|J}(\tilde{x}|y_J)\pi_0(y_J)$ from below for $\tilde{x} \in \tilde{\mathcal{X}}$ such that $\|x\| > a_{\sigma_n}$ and $\|\tilde{y}_I\| \leq a_{\sigma_n}$. For any $j \leq K$ and $k \in \mathcal{Y}_J$, $||\tilde{x} - \mu_{jk,J^c}||^2 \leq 2(||\tilde{x}||^2 + ||\mu_{jk,J^c}||^2) \leq 16||x||^2$ as $||\mu_{jk,J^c}|| \leq 2a_{\sigma_n}$ by construction of $U_{j|k}$ and $2||x|| > ||\tilde{x}||$. Then

$$\phi(\tilde{x}, \mu_{jk,J^c}, \sigma_{J^c}) = (2\pi)^{-d_{J^c}/2} \prod_{i \in J^c} \sigma_i^{-1} \exp\left\{-0.5 \sum_{i \in J^c} (\tilde{x}_i - \mu_{jk,i})^2/\sigma_i^2\right\}$$

$$\geq (2\pi)^{-d_{J^c}/2}\sigma_n^{-d_{J^c}} \exp\left\{-\frac{8||x||^2}{\underline{\sigma}_n^2}\right\}.$$

Then, for n large enough

$$f_J(y_J, \tilde{x}|\theta, m) = \sum_{k\in\mathcal{Y}_J}\sum_{j=1}^{K}\alpha_{jk}\int_{A_{y_J}}\phi(\tilde{y}_J, \mu_{jk,J}, \sigma_J)d\tilde{y}_J\phi(\tilde{x}, \mu_{jk,J^c}, \sigma_{J^c})$$
$$\geq (2\pi)^{-d_{J^c}/2}\sigma_n^{-d_{J^c}} \exp\left\{-\frac{8||x||^2}{\underline{\sigma}_n^2}\right\}\sum_{j=1}^{K}\alpha_{jk}\sum_{k\in\mathcal{Y}_J}\int_{A_{y_J}}\phi(\tilde{y}_J, \mu_{jk,J}, \sigma_J)d\tilde{y}_J$$
$$\geq (2\pi)^{-d_{J^c}/2}\sigma_n^{-d_{J^c}} \exp\left\{-\frac{8||x||^2}{\underline{\sigma}_n^2}\right\}\frac{1}{2}K\min_{j,k}\alpha_{jk}.$$

Combining this inequality with (61), we get

$$\frac{f_J(y_J, \tilde{x}|\theta, m)}{f_{0|J}(\tilde{x}|y_J)\pi_{0J}(y_J)} \geq \frac{1}{2}(2\pi)^{-d_{J^c}/2}\bar{f}_0^{-1}\sigma_n^{-d_{J^c}}K\frac{\sigma_n^{2\beta+d_{J^c}}}{2m^2}\exp\left\{-\frac{8||x||^2}{\underline{\sigma}_n^2}\right\}$$
$$\geq \exp\left\{-\frac{8||x||^2}{\underline{\sigma}_n^2} - C_4\log n\right\} \tag{65}$$

for sufficiently large C_4 because $|\log\left[K\sigma_n^{2\beta}/m^2\right]| \lesssim \log n$.

Thus, for $||x|| > a_{\sigma_n}$, (65) and the first inequality in (64), which holds for any $x \in \mathcal{X}$, deliver

$$\frac{p(y_J, y_I, x|\theta, m)}{p_0(y_J, y_I, x)} \geq \exp\left\{-\frac{8||x||^2}{\underline{\sigma}_n^2} - C_4\log n\right\}. \tag{66}$$

Q.E.D.

LEMMA 17 *Under the assumptions and notation of Section 4.3, for $\lambda_n < \lambda_0$, where λ_0 is defined in Lemma 11,*

$$E_0\left(\left[\log\frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right]^2\right) \leq A\tilde{\epsilon}_n^2$$
$$E_0\left(\left[\log\frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right]\right) \leq A\tilde{\epsilon}_n^2$$

PROOF:

$$E_0\left(\left[\log\frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right]^2\right)$$
$$\leq d_H^2(p_0(\cdot,\cdot), p(\cdot,\cdot|\theta, m))\left(12 + 2\left(\log\frac{1}{\lambda_n}\right)^2\right)$$
$$+ 8P\left\{\left(\log\frac{p_0(\cdot,\cdot)}{p(\cdot,\cdot|\theta, m)}\right)^2\mathbf{1}\left\{\frac{p(\cdot,\cdot|\theta, m)}{p_0(\cdot,\cdot)} < \lambda_n\right\}\right\}$$

$$\lesssim \sigma_n^{2\beta}(12 + 2\log(1/\lambda_n)^2) + \sigma_n^{2\beta+\epsilon} \lesssim \log(1/\lambda_n)^2\sigma_n^{2\beta},$$

where first inequality is derived using Lemma 11 and penultimate inequality is derived using inequalities (51) and (55). Similarly,

$$\begin{aligned}
&E_0\left(\log \frac{p_0(y_J, y_I, x)}{p(y_J, y_I, x|\theta, m)}\right) \\
&\leq d_H^2(p_0(\cdot,\cdot), p(\cdot,\cdot|\theta,m))\left(1 + 2\left(\log\frac{1}{\lambda_n}\right)\right) \\
&\quad + 2P\left\{\left(\log\frac{p_0(\cdot,\cdot)}{p(\cdot,\cdot|\theta,m)}\right)\mathbf{1}\left\{\frac{p(\cdot,\cdot|\theta,m)}{p_0(\cdot,\cdot)} < \lambda_n\right\}\right\} \\
&\lesssim \sigma_n^{2\beta}(1 + 2\log(1/\lambda_n)) + \sigma_n^{2\beta+\epsilon} \lesssim \log(1/\lambda_n)\sigma_n^{2\beta}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\log(1/\lambda_n)\sigma_n^{2\beta} \leq \log(1/\lambda_n)^2\sigma_n^{2\beta} &= \log\left(\frac{2N_JK^2}{\sigma_n^{2\beta}}\right)^2 \tilde{\epsilon}_n^2(\log(\tilde{\epsilon}_n^{-1}))^{-2} \\
&\leq \left(\frac{\log[2N_J^2(C_1\sigma_n^{-d_{Jc}}\{\log(\tilde{\epsilon}_n^{-1})\}^{d_{Jc}+d_{Jc}/\tau})^2\sigma_n^{-2\beta}]}{\log(\tilde{\epsilon}_n^{-1})}\right)^2 \tilde{\epsilon}_n^2,
\end{aligned}$$

where the term multiplying $\tilde{\epsilon}_n^2$ on the right hand side is bounded by Assumption 5 ($N_J = o(n^{1-\nu})$) and definitions of $\tilde{\epsilon}_n$ and σ_n. *Q.E.D.*

LEMMA 18 *Under the assumptions and notation of Section 4.3, for all sufficiently large n, $s = 1 + 1/\beta + 1/\tau$, and some $C_6 > 0$*

$$\Pi(m = N_JK, \theta \in S_{\theta^\star}) \geq \exp\left[-C_6 N_J \tilde{\epsilon}_n^{-d_{Jc}/\beta}\{\log(n)\}^{d_{Jc}s + \max\{\tau_1, 1, \tau_2/\tau\}}\right].$$

PROOF: First, consider the prior probability of $m = N_JK$. By (3) for some $C_{61} > 0$,

$$\begin{aligned}
(67) \qquad &\Pi(m = N_JK) \propto \exp[-\gamma N_JK(\log N_JK)^{\tau_1}] \\
&\geq \exp[-C_{61}N_J\tilde{\epsilon}_n^{-d_{Jc}/\beta}\{\log(1/\tilde{\epsilon}_n)\}^{sd_{Jc}}(\log n)^{\tau_1}] \\
&\geq \exp[-C_{61}N_J\tilde{\epsilon}_n^{-d_{Jc}/\beta}\{\log(n)\}^{sd_{Jc}+\tau_1}]
\end{aligned}$$

as $N_J = o(n^{1-\nu})$ by (24) and $\tilde{\epsilon}_n^{-1} < n$.

Second, consider the prior on $\{\alpha_{jk}\}$. There exist (j_0, k_0) such that $\alpha_{j_0k_0}^\star \geq \frac{1}{m}$ and suppose that $|\alpha_{jk}^\star - \alpha_{jk}| \leq \frac{\sigma_n^{2\beta}}{m^2}$ for all $(j,k) \neq (j_0,k_0)$. Then,

$$|\alpha_{j_0k_0}^\star - \alpha_{j_0k_0}| = \left|\sum_{(jk)\neq(j_0k_0)} \alpha_{jk}^\star - \alpha_{jk}\right| \leq (m-1)\frac{\sigma_n^{2\beta}}{m^2} \leq \frac{\sigma_n^{2\beta}}{m}$$

$$\alpha_{j_0k_0} \geq \alpha^\star_{j_0k_0} - \frac{\sigma_n^{2\beta}}{m} \geq \frac{1-\sigma_n^{2\beta}}{m} \geq \frac{\sigma_n^{2\beta+d_{J^c}}}{2m^2}.$$

Furthermore,

$$\sum_{j=1}^{K}\sum_{k\in\mathcal{Y}_J} |\alpha_{jk} - \alpha^\star_{jk}| \leq (m-1)\frac{\sigma_n^{2\beta}}{m^2} + \frac{\sigma_n^{2\beta}}{m} \leq 2\sigma_n^{2\beta}.$$

It then follows that

$$\begin{aligned}
&\Pi\left(\sum_{j=1}^{K}\sum_{k\in\mathcal{Y}_J} |\alpha_{jk} - \alpha^\star_{jk}| \leq 2\sigma_n^{2\beta}, \min_{j\leq K, k\in\mathcal{Y}_J} \alpha_{jk} \geq \frac{\sigma_n^{2\beta+d_{J^c}}}{2m^2}\right)\\
&\geq \Pi\left(|\alpha_{jk} - \alpha^\star_{jk}| \leq \frac{\sigma_n^{2\beta}}{m^2}, \alpha_{jk} \geq \frac{\sigma_n^{2\beta}}{2m^2},\ (j,k) \in \{1,\ldots,K\}\times\mathcal{Y}_J \setminus \{(j_0,k_0)\}\right)\\
&\geq \exp\left\{-C_{62}N_J K \log(N_J K/\sigma_n^\beta)\right\},
\end{aligned}$$

where the last inequality is derived in the proof of Lemma 10 in Ghosal and van der Vaart (2007) for some $C_{62} > 0$ (see, also, Lemma 6.1 in Ghosal et al. (2000)). Note that

$$\begin{aligned}
(68)\qquad & K\log(N_J K/\sigma_n^\beta)\\
&\leq \tilde{\epsilon}_n^{-d_{J^c}/\beta} \log(\tilde{\epsilon}_n^{-1})^{d_{J^c}s} \log(N_J \tilde{\epsilon}_n^{-d_{J^c}/\beta - 1} \log(\tilde{\epsilon}_n^{-1})^{d_{J^c}s+1})\\
&\lesssim \tilde{\epsilon}_n^{-d_{J^c}/\beta} \log(n)^{d_{J^c}s+1}.
\end{aligned}$$

Assumption (9) on the prior for σ_i implies that for $i \in J$

$$\begin{aligned}
(69)\qquad & \prod_{i=1}^{d_J} \Pi(\sigma_i^{-2} \geq 32N_i^2\beta\log\sigma_n^{-1})\\
&\geq \prod_{i=1}^{d_J}\left(a_6(64N_i^2\beta\log\sigma_n^{-1})^{a_7}\exp\left\{-a_9(64N_i^2\beta\log\sigma_n^{-1})^{1/2}\right\}\right)\\
&\geq \exp\left\{-C_{63}N_J\log(\sigma_n^{-1})\right\} \geq \exp\left\{-C_{64}N_J\log(n)\right\},
\end{aligned}$$

and for $i \in J^c$,

$$\begin{aligned}
(70)\qquad & \prod_{i=1}^{d_{J^c}} \Pi\left(\sigma_{i,n}^{-2} \leq \sigma_i^{-2} \leq \sigma_{i,n}^{-2}(1+\sigma_n^{2\beta})\right) \geq \prod_{i=1}^{d_{J^c}}\left(a_6(\sigma_{i,n}^{-2})^{a_7}\sigma_n^{2a_8\beta}\exp\left\{-a_9\sigma_{i,n}^{-1}\right\}\right)\\
&\geq \prod_{i=1}^{d_{J^c}}\exp\left\{-C_{65}\sigma_{i,n}^{-1}\right\} = \prod_{i=1}^{d_{J^c}}\exp\left\{-C_{65}\sigma_n^{-\beta/\beta_i}\right\} \geq \exp\left\{-C_{65}d_{J^c}\sigma_n^{-d_{J^c}}\right\}\\
&\geq \exp\left\{-C_{66}\tilde{\epsilon}_n^{-d_{J^c}/\beta}\log(n)^{d_{J^c}/\beta}\right\}.
\end{aligned}$$

Assumption (10) on the prior for μ_{jk} implies

$$(71)\qquad \prod_{j=1}^{K}\prod_{k\in\mathcal{Y}_J}\prod_{i\in J}\Pi\left(\mu_{jk,i} \in \left[k_i - \frac{1}{4N_i}, k_i + \frac{1}{4N_i}\right]\right)$$

$$\geq \left(a_{11} 2^{-d_J} N_J^{-1} \exp\left\{-a_{12}\right\}\right)^{N_J K}$$
$$\geq \exp\left\{-C_{67} N_J K \log(N_J)\right\}$$
$$\geq \exp\left\{-C_{68} N_J \tilde{\epsilon}_n^{-d_{J^c}/\beta} \log(n)^{d_{J^c} s+1}\right\}$$

and

$$\prod_{j=1}^{K} \prod_{k \in \mathcal{Y}_J} \Pi\left(\mu_{jk,J^c} \in U_{j|k}\right) \geq \left(a_{11} \exp\left\{-a_{12} a_{\sigma_n}^{\tau_2}\right\} \min_{j,k} Vol(U_{j|k})\right)^{N_J K} \tag{72}$$
$$= \left(a_{11} \exp\left\{-a_{12} a_{\sigma_n}^{\tau_2}\right\} \sigma_n^{d_{J^c}} \tilde{\epsilon}_n^{2b_1 d_{J^c}}\right)^{N_J K}$$
$$\geq \exp\left\{-C_{69} N_J \tilde{\epsilon}_n^{-d_{J^c}/\beta} \log(n)^{d_{J^c} s+\max\{1,\tau_2/\tau\}}\right\}.$$

It follows from (67) - (72), that for all sufficiently large n and some $C_6 > 0$,

$$\Pi(\mathcal{K}(p_0, \tilde{\epsilon}_n)) \geq \Pi(m = N_J K, \theta \in S_{\theta^\star})$$
$$\geq \exp[-C_6 N_J \tilde{\epsilon}_n^{-d_{J^c}/\beta} \{\log(n)\}^{d_{J^c} s+\max\{\tau_1,1,\tau_2/\tau\}}].$$

Q.E.D.

LEMMA 19 *For $H \in \mathbb{N}$, $0 < \underline{\sigma} < \overline{\sigma}$, and $\overline{\mu} > 0$, let us define a sieve*

$$\mathcal{F} = \{p(y, x|\theta, m) : m \leq H,\ \mu_j \in [-\overline{\mu}, \overline{\mu}]^d, j = 1, \ldots, m, \quad \sigma_i \in [\underline{\sigma}, \overline{\sigma}], i = 1, \ldots, d\}. \tag{73}$$

For $0 < \epsilon < 1$ and $\underline{\sigma} \leq 1$,

$$M_e(\epsilon, \mathcal{F}, d_{L_1}) \leq H \cdot \left\lceil \frac{12\overline{\mu} d}{\underline{\sigma}\epsilon} \right\rceil^{Hd} \cdot \left\lceil \frac{15}{\epsilon} \right\rceil^{H} \cdot \left\lceil \frac{\log(\overline{\sigma}/\underline{\sigma})}{\log(1+\epsilon/[12d])} \right\rceil^{d}.$$

For all sufficiently large H, large $\overline{\sigma}$ and small $\underline{\sigma}$,

$$\Pi(\mathcal{F}^c) \leq H^2 d \exp\{-a_{13} \overline{\mu}^{\tau_3}\} + \exp\{-\gamma H (\log H)^{\tau_1}\}$$
$$+ d a_1 \exp\{-a_2 \underline{\sigma}^{-2a_3}\} + d a_4 \exp\{-2a_5 \log \overline{\sigma}\}.$$

PROOF: The proof is similar to proofs of related results in Norets and Pati (2017), Shen et al. (2013), and Ghosal and van der Vaart (2001) among others.

Let us begin with the first claim. For a fixed value of m, define set S_μ^m to contain centers of $|S_\mu^m| = \lceil 12\overline{\mu} d/(\underline{\sigma}\epsilon) \rceil$ equal length intervals partitioning $[-\overline{\mu}, \overline{\mu}]$. Let S_α^m be an $\epsilon/3$-net of Δ^{m-1} in the L_1 distance ($\forall \alpha \in \Delta^{m-1}$, $\exists \tilde{\alpha} \in S_\alpha^m$, $d_{L_1}(\alpha, \tilde{\alpha}) \leq \epsilon/3$). From Lemma A.4 in Ghosal and van der Vaart (2001), the cardinality of S_α^m, is bounded as follows

$$|S_\alpha^m| \leq [15/\epsilon]^m.$$

Define $S_\sigma = \{\sigma^l, l=1,\ldots,\lceil \log(\overline{\sigma}/\underline{\sigma})/(\log(1+\epsilon/(12d)\rceil, \sigma^1 = \underline{\sigma}, (\sigma^{l+1}-\sigma^l)/\sigma^l = \epsilon/(12d)\}$.

Let us show that

$$S_{\mathcal{F}} = \{p(y,x|\theta,m):\ m\leq H,\ \alpha\in S_\alpha^m,\ \sigma_i\in S_\sigma,\ \mu_{ji}\in S_\mu^m,\ j\leq m,\ i\leq d\}$$

is an ϵ-net for $\mathcal{F}$ in d_{L_1}. For a given $p(\cdot|\theta,m)\in\mathcal{F}$ with $\sigma^{l_i}\leq\sigma_i\leq\sigma^{l_i+1}$, $i=1,\ldots,d$, find $\tilde\alpha\in S_\alpha^m$, $\tilde\mu_{ji}\in S_{\mu^x}^m$, and $\tilde\sigma_i=\sigma_{l_i}\in S_\sigma$ such that for all $j=1,\ldots,m$ and $i=1,\ldots,d$

$$|\mu_{ji}-\tilde\mu_{ji}|\leq\frac{\underline{\sigma}\epsilon}{12d},\ \sum_j|\alpha_j-\tilde\alpha_j|\leq\frac{\epsilon}{3},\ \frac{|\sigma_i-\tilde\sigma_i|}{\tilde\sigma_i}\leq\frac{\epsilon}{12d}.$$

By Lemma 6, $d_{L_1}(p(\cdot|\theta,m),p(\cdot|\tilde\theta,m))\leq d_{L_1}(f(\cdot|\theta,m),f(\cdot|\tilde\theta,m))$. Similarly to the proof of Proposition 3.1 in Norets and Pelenis (2014) or Theorem 4.1 in Norets and Pati (2017),

$$\begin{aligned}
&d_{L_1}(f(\cdot|\theta,m),f(\cdot|\tilde\theta,m))\leq\sum_j|\alpha_j-\tilde\alpha_j|+2\max_{j=1,\ldots,m}||\phi(\cdot;\mu_j,\sigma)-\phi(\cdot;\tilde\mu_j,\tilde\sigma||_1\\
&\leq\epsilon/3+4\sum_{i=1}^d\left\{\frac{|\mu_{ji}-\tilde\mu_{ji}|}{\min(\sigma_i,\tilde\sigma_i)}+\frac{|\sigma_i-\tilde\sigma_i|}{\min(\sigma_i,\tilde\sigma_i)}\right\}\leq\epsilon.
\end{aligned}$$

This concludes the proof for the covering number.

The proof of the upper bound on $\Pi(\mathcal{F}^c)$ is the same as the corresponding proof of Theorem 4.1 in Norets and Pati (2017), except here the coordinate specific scale parameters and slightly different notation for the prior tail condition (11) lead to dimension d appearing in front of some of the terms in the bound.

Q.E.D.

LEMMA 20 *Consider* $\epsilon_n=(N_J/n)^{\beta_{Jc}/(2\beta_{Jc}+1)}(\log n)^{t_J}$ *and* $\tilde\epsilon_n=(N_J/n)^{\beta_{Jc}/(2\beta_{Jc}+1)}(\log n)^{\tilde t_J}$ *with* $t_J>\tilde t_J+\max\{0,(1-\tau_1)/2\}$ *and* $\tilde t_J>t_{J0}$, *where* t_{J0} *is defined in* (25). *Define* $\mathcal{F}_n$ *as in* (73) *with* $\epsilon=\epsilon_n, H=n\epsilon_n^2/(\log n)$, $\underline{\alpha}=e^{-nH}$, $\underline{\sigma}=n^{-1/(2a_3)}$, $\overline{\sigma}=e^n$, *and* $\overline{\mu}=n^{1/\tau_3}$. *Then, for some constants* $c_1,c_3>0$ *and every* $c_2>0$, $\mathcal{F}_n$ *satisfies* (28) *and* (29) *for all large* n.

PROOF: From Lemma 19,

$$\log M_e(\epsilon_n,\mathcal{F}_n,\rho)\leq c_1H\log n=c_1n\epsilon_n^2.$$

Also,

$$\Pi(\mathcal{F}_n^c)\leq H^2\exp\{-a_{13}n\}+\exp\{-\gamma H(\log H)^{\tau_1}\}$$

$$+ a_1 \exp\{-a_2 n\} + a_4 \exp\{-2a_5 n\}.$$

Hence, $\Pi(\mathcal{F}_n^c) \leq e^{-(c_2+4)n\tilde{\epsilon}_n^2}$ for any c_2 if $\epsilon_n^2(\log n)^{\tau_1 - 1}/\tilde{\epsilon}_n^2 \to \infty$, which holds for $t_J > \tilde{t}_J + \max\{0, (1-\tau_1)/2\}$.

Q.E.D.