Supplementary Material for Nonparametric Lack-of-Fit Testing and Consistent Variable Selection

Adriano Zanin Zambom and Michael G. Akritas

State University of Campinas and Penn State University

1. Auxiliary Results

Lemma 1.0.1. *Let $X_1, \ldots, X_n$ be iid[F], and let $\hat{F}_n(x)$ be the corresponding empirical distribution function. Then, for any constant c,*

$$sup_{x_i,x_j}\left\{|F(x_i)-F(x_j)|I\left[|\hat{F}(x_i)-\hat{F}(x_j)|\leq\frac{c}{n}\right]\right\}=O_p\left(\frac{1}{\sqrt{n}}\right).$$

Proof. By the Dvoretzky, Kiefer and Wolfowitz (1956) theorem, we have that $\forall\epsilon\geq 0$,

$$P\left(\sup_x|\hat{F}_n(x)-F(x)|\geq\epsilon\right)\leq Ce^{-2n\epsilon^2}.$$

Therefore, $|\hat{F}(x)-F(x)|=O_p\left(\frac{1}{\sqrt{n}}\right)$ uniformly on x. Hence, writing

$$|F(x_i)-F(x_j)| = |F(x_i)-\hat{F}_n(x_i)+\hat{F}_n(x_i)-F(x_j)+\hat{F}_n(x_j)-\hat{F}_n(x_j)|,$$

it follows that $sup_{x_i,x_j}\left\{|F(x_i)-F(x_j)|I\left[|\hat{F}(x_i)-\hat{F}(x_j)|\leq c/n\right]\right\}$ is less than or equal to

$$\begin{aligned} & sup_{x_i,x_j}\left\{|F(x_i)-\hat{F}_n(x_i)|+|\hat{F}_n(x_j)-F(x_j)|\right\} \\ & +sup_{x_i,x_j}\left\{|\hat{F}_n(x_i)-\hat{F}_n(x_j)|\right\}I\left[|\hat{F}_n(x_i)-\hat{F}_n(x_j)|\leq\frac{c}{n}\right] \\ = \; & O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\frac{1}{n}\right). \end{aligned}$$

This completes the proof of the lemma. □

Lemma 1.0.2. *Let W_i be defined in (2.11) of the paper of reference, and any Lipschitz continuous function $g(x)$,*

$$\frac{1}{p}\sum_{j=1}^{n}g(x_{2j})I(j\in W_i)-g(x_{2i})=O_p\left(\frac{1}{\sqrt{n}}\right),$$

uniformly in $i=1,\ldots,n$.

Proof. First note that by the Lipschitz continuity and the Mean Value Theorem we have

$$|g(x_{2j}) - g(x_{2i})| \leq M|x_{2j} - x_{2i}| \leq M|F_{X_2}(x_{2j}) - F_{X_2}(x_{2i})|/f_{X_2}(\tilde{x}_{ij}),$$

for some constant M, where $\tilde{x}_{ij}$ is between x_{2j} and x_{2i}. Thus,

$$\left|\frac{1}{p}\sum_{j=1}^{n} g(x_{2j})I(j \in W_i) - g(x_{2i})\right| \leq \frac{1}{p}\sum_{j=1}^{n}|g(x_{2j}) - g(x_{2i})| \times$$

$$\times I\left[|\hat{F}_{X_2}(x_{2i}) - \hat{F}_{X_2}(x_{2j})| \leq \frac{p-1}{2n}\right]$$

$$\leq \frac{M}{p}\sum_{j=1}^{n}\frac{|F_{X_2}(x_{2j}) - F_{X_2}(x_{2i})|}{f_{X_2}(\tilde{x}_{ij})} I\left[|\hat{F}_{X_2}(x_{2i}) - \hat{F}_{X_2}(x_{2j})| \leq \frac{p-1}{2n}\right] = O_p\left(\frac{1}{\sqrt{n}}\right),$$

where the last equality follows from Lemma 1.0.1 and the assumption that f_{X_2} remains bounded away from zero. □

Lemma 1.0.3. *Let Y is the response variable, $\tilde{\mathbf{X}} = (\mathbf{1}, \mathbf{X})$, where $\mathbf{X}$ is the $n \times d$ matrix of covariates and $\mathbf{1}$ is the vector of 1, and let e_j denote the $(d+1) \times 1$ vector having 1 in the j-th entry and all other entries 0. In the least squares regression estimation of $(\alpha, \boldsymbol{\beta})^T$ of the model*

$$\mathbf{Y} = \tilde{\mathbf{X}}\begin{pmatrix}\alpha \\ \boldsymbol{\beta}\end{pmatrix} + \epsilon,$$

the weights

$$w_j = e_1^T(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T e_j, \quad j = 1, \dots, n$$

from the estimator $\hat{\alpha} = e_1^T(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T Y$ are such that

$$\sum_{j=1}^{n} w_j = 1.$$

Proof. Note first that $\hat{\boldsymbol{\beta}} = (\bar{\mathcal{X}}^T\bar{\mathcal{X}})^{-1}\bar{\mathcal{X}}^T Y$, where the columns of $\bar{\mathcal{X}}$ are the centered columns of the design matrix.

Since

$$\begin{aligned}\hat{\alpha} &= \bar{Y} - \hat{\beta}_1\bar{X} - \dots \hat{\beta}_d\bar{X}_d \\ &= \left(\frac{1}{n}\mathbf{1^T} - (\bar{\mathbf{X}}, \dots, \bar{\mathbf{X}}_\mathbf{d})(\bar{\mathcal{X}}^\mathbf{T}\bar{\mathcal{X}})^{-\mathbf{1}}\bar{\mathcal{X}}^\mathbf{T}\right)\mathbf{Y}.\end{aligned}$$

The lemma follows from the fact that the weights $(\bar{X},\ldots,\bar{X}_d)(\bar{\mathcal{X}}^T\bar{\mathcal{X}})^{-1}\bar{\mathcal{X}}^T$ sum to zero because $\bar{\mathcal{X}}^T\mathbf{1}=\mathbf{0}$. $\square$

Lemma 1.0.4. *For the local polynomial regression estimator (2.10) of the paper of reference*

$$\hat{m}_1(\mathbf{x}) = \mathbf{e}_1^T\left(\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}}\mathbb{X}_{\mathbf{x}}\right)^{-1}\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}},$$

each of the weights denoted by

$$\tilde{w}(\mathbf{x},\boldsymbol{X}_j) = \mathbf{e}_1^T\left(\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}}\mathbb{X}_{\mathbf{x}}\right)^{-1}\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}}\mathbf{e}_j, \quad j=1,\ldots,n,$$

is of order $O_p\left(\frac{1}{n|H_n|^{1/2}}\right)$.

Proof. Recall from (2.10) of the paper of reference that

$$\mathbb{X}_{\mathbf{x}} = \begin{pmatrix} 1 & (\mathbf{X}-\mathbf{x})^T & \mathrm{vech}^T\left\{(\mathbf{X}-\mathbf{x})(\mathbf{X}-\mathbf{x})^T\right\} & \ldots \\ \vdots & \vdots & \vdots & \ldots \\ 1 & (\mathbf{X}_n-\mathbf{x})^T & \mathrm{vech}^T\left\{(\mathbf{X}_n-\mathbf{x})(\mathbf{X}_n-\mathbf{x})^T\right\} & \ldots \end{pmatrix},$$

$\mathbb{W}_{\mathbf{x}} = diag\{K_H(\mathbf{X}_1-\mathbf{x}),\ldots,K_H(\mathbf{X}_n-\mathbf{x})\}$, and that the dimensions of $\mathbb{X}_{\mathbf{x}}$ are $n\times\gamma_d$.

Now, it is easy to see that $\frac{1}{n}\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}}$ is a $\gamma_d\times n$ matrix with column j given by

$$\mathbf{V}_j := \frac{1}{n}\begin{pmatrix} K_{H_n}(\mathbf{X}_j-\mathbf{x}) \\ K_{H_n}(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_j-\mathbf{x}) \\ K_{H_n}(\mathbf{X}_j-\mathbf{x})\mathrm{vech}\{(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_j-\mathbf{x})^T\} \\ \vdots \end{pmatrix}. \tag{1.1}$$

Let $\mathbf{1}$ be the vector of ones (of dimension defined by the context) and $D_f(\mathbf{x})$

the vector of partial derivatives of $f(\mathbf{x})$. As in Ruppert and Wand (1994), define

$$N_{\mathbf{x}} = \begin{pmatrix} \nu_{\mathbf{x},11} & \nu_{\mathbf{x},12} & \nu_{\mathbf{x},13} & \cdots \\ \nu_{\mathbf{x},21} & \nu_{\mathbf{x},22} & \nu_{\mathbf{x},23} & \cdots \\ \nu_{\mathbf{x},31} & \nu_{\mathbf{x},32} & \nu_{\mathbf{x},33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \int \begin{bmatrix} 1 \\ \mathbf{u} \\ \text{vech}(\mathbf{u}\mathbf{u}^T) \\ \vdots \end{bmatrix} \begin{bmatrix} 1 & \mathbf{u} & \text{vech}^T(\mathbf{u}\mathbf{u}^T) & \ldots \end{bmatrix} K(\mathbf{u})d\mathbf{u},$$

and

$$Q_{\mathbf{x}} = \int K(\mathbf{u}) \begin{bmatrix} 0 & \mathbf{u}^T & 0 & \cdots \\ \mathbf{u} & 0 & \mathbf{u}\text{vech}^T(\mathbf{u}\mathbf{u}^T) & \cdots \\ 0 & \text{vech}(\mathbf{u}\mathbf{u}^T)\mathbf{u}^T & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \{D_f^T(\mathbf{x})H_n^{1/2}\mathbf{u}\}d\mathbf{u}.$$

It is known (Ruppert and Wand, 1994) that $Q_{\mathbf{x}} = O(tr(H_n^{1/2}))$, where $tr(H_n)$ is the trace of matrix H_n.

For $\ell = 2, \ldots, q$, let C_ℓ be a matrix whose each element is of the same order of a product of ℓ elements of $H_n^{1/2}$. For example, C_2 can be defined (see proof of Theorem 3.1 Ruppert and Wand, 1994) as the $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$ matrix such that

$$\text{vech}(H_n^{1/2}\mathbf{u}\mathbf{u}^T H_n^{1/2}) = C_2\text{vech}(\mathbf{u}\mathbf{u}^T),$$

for all d-vectors $\mathbf{u}$.

Extending the formulas of $\left(n^{-1}\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}}\mathbb{X}_{\mathbf{x}}\right)^{-1}$ in Ruppert and Wand (1994) to a more general case, we have that

$$\mathbf{e}_1^T\left(n^{-1}\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}}\mathbb{X}_{\mathbf{x}}\right)^{-1} = O_p\left(\mathbf{1}^T\text{diag}\{1, H_n^{-1/2}, C_2^{-1}, C_3^{-1}, \ldots\}\right).$$

Also, each column V_j of $\frac{1}{n}\mathbb{X}_{\mathbf{x}}^T\mathbb{W}_{\mathbf{x}}$ defined in (1.1) is of order

$$O_p\left(\frac{1}{n|H_n|^{1/2}}\text{diag}\{1, H_n^{1/2}, C_2, C_3, \ldots\}\mathbf{1}\right).$$

Therefore, noting that $\mathbf{1}$ is a $\gamma_d \times 1$ vector and that each weight $w(\mathbf{x}, \mathbf{X}_j)$ is computed by $\mathbf{e}_1^T \left(n^{-1}\mathbb{X}_{\mathbf{x}}^T \mathbb{W}_{\mathbf{x}} \mathbb{X}_{\mathbf{x}}\right)^{-1} \mathbf{V}_j$, we have that

$$\begin{aligned} & \mathbf{e}_1^T \left(n^{-1}\mathbb{X}_{\mathbf{x}}^T \mathbb{W}_{\mathbf{x}} \mathbb{X}_{\mathbf{x}}\right)^{-1} \mathbf{V}_j \\ = \; & O_p\left(\mathbf{1}^T \mathrm{diag}\{1, H_n^{-1/2}, C_2^{-1}, C_3^{-1}, \ldots\}\right) \times \\ & \qquad \times O_p\left(\frac{1}{n|H_n|^{1/2}} \mathrm{diag}\{1, H_n^{1/2}, C_2, C_3, \ldots\}\mathbf{1}\right) \\ = \; & O_p\left(\frac{1}{n|H_n|^{1/2}}\right), \end{aligned}$$

completing the proof. □

Lemma 1.0.5. *For a symmetric, positive definite bandwidth matrix $H_n^{1/2}$, we have*

1. *The determinant of $H_n^{1/2}$ is equal to the product of the eigen-values of $H_n^{1/2}$.*
2. *Define the norm $||H_n^{1/2}||$ to be the maximum of its eigenvalues. Given $\mathbf{X}$, for the weights used in local polynomial regression (2.10) of the paper of reference*

$$w(\boldsymbol{X}_{1i}, \boldsymbol{X}_{1j}) = \mathbf{e}_1^T \left(\mathbb{X}_{\mathbf{X}_{1i}}^T \mathbb{W}_{\mathbf{X}_{1i}} \mathbb{X}_{\mathbf{X}_{1i}}\right)^{-1} \mathbb{X}_{\mathbf{X}_{1i}}^T \mathbb{W}_{\mathbf{X}_{1i}} \mathbf{e}_j,$$

we have that

$$\sum_{j=1}^{n} w(\boldsymbol{X}_{1i}, \boldsymbol{X}_{1j})||\boldsymbol{X}_{1j} - \boldsymbol{X}_{1i}|| = O(||H_n^{1/2}||). \tag{1.2}$$

Proof. 1. Note that the bandwidthmatrix $H_n^{1/2}$ is positive definite, therefore there exists an eigen value decomposition $H_n^{1/2} = V\Lambda V^{-1}$ such that V is a orthogonal matrix with columns corresponding to the eigenvectors of $H_n^{1/2}$ and Λ is a diagonal matrix with elements corresponding to the eigenvalues of $H_n^{1/2}$. Thus,

$$|H_n^{1/2}| = |V\Lambda V^{-1}| = |V||\Lambda||V^{-1}| = |VV^{-1}||\Lambda| = |\Lambda|.$$

2. Let b be such that $K(\mathbf{x}) = K(\mathbf{x})I(||\mathbf{x}|| \leq \sqrt{d-1}b)$. Such a b exists by the

assumption that the density K has bounded support. By noting that

$$\begin{aligned} & K\left(H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})\right)||\mathbf{X}_{1j}-\mathbf{X}_{1i}|| \\ = \; & K\left(H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})\right)||H_n^{1/2}H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})|| \\ \leq \; & K\left(H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})\right)||H_n^{1/2}||||H_n^{-1/2}(vX_{1i}-\mathbf{X}_{1j})|| \\ \leq \; & K\left(H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})\right)||H_n^{1/2}||\sqrt{d-1}b, \end{aligned}$$

we have that

$$\begin{aligned} & K_{H_n}\left(H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})\right)||\mathbf{X}_{1j}-\mathbf{X}_{1i}|| \\ = \; & O\left(K_{H_n}\left(H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})\right)||H_n^{1/2}||\sqrt{d-1}b\right) \\ = \; & K_{H_n}\left(H_n^{-1/2}(\mathbf{X}_{1i}-\mathbf{X}_{1j})\right)O\left(||H_n^{1/2}||\right). \end{aligned} \tag{1.3}$$

Let $V_{\mathbf{X}_{1i}}$ be the $n\times 1$ vector with j-th entry $||\mathbf{X}_{1i}-\mathbf{X}_{1j}||O\left(||H_n^{1/2}||\right)$. From the definition of the weights, the left hand side of (1.2) is equal to

$$\begin{aligned} & \mathbf{e}_1^T\left(\mathbb{X}_{\mathbf{X}_{1i}}^T\mathbb{W}_{\mathbf{X}_{1i}}\mathbb{X}_{\mathbf{X}_{1i}}\right)^{-1}\mathbb{X}_{\mathbf{X}_{1i}}^T\mathbb{W}_{\mathbf{X}_{1i}}V_{\mathbf{X}_{1i}} \\ = \; & O\left(||H_n^{1/2}||\right)\mathbf{e}_1^T\left(\mathbb{X}_{\mathbf{X}_{1i}}^T\mathbb{W}_{\mathbf{X}_{1i}}\mathbb{X}_{\mathbf{X}_{1i}}\right)^{-1}\mathbb{X}_{\mathbf{X}_{1i}}^T\mathbb{W}_{\mathbf{X}_{1i}}\mathbf{1} \\ = \; & O\left(||H_n^{1/2}||\right). \end{aligned}$$

The first equality follows from the fact that the each entry j of the $n\times 1$ vector $\mathbb{W}_{\mathbf{X}_{1i}}V_{\mathbf{X}_{1i}}$ is equal to (1.3), and the last equality follows from the fact that the weights sum to 1, proved in Lemma 1.0.3. This completes the proof.

□

2. Proofs of Theorems

Proof of Theorem 1. First note that the vector of $(n-p+1)p$ constructed "observations" in the augmented one-way ANOVA design can be constructed as

$$\hat{\boldsymbol{\xi}}_V = (\hat{\xi}_j, j\in W_{(p-1)/2+1},\ldots,\hat{\xi}_j, j\in W_{n-(p-1)/2})'. \tag{2.1}$$

This vector of constructed "observations" is composed by observations of $\hat{\xi}$ in blocks, according to the windows W_i, which are described in (2.11) of the paper of reference. Thus that the test statistic can be written as

$$\text{MST - MSE} = \hat{\boldsymbol{\xi}}_V' A \hat{\boldsymbol{\xi}}_V$$

where

$$A = \frac{np-1}{n(n-1)p(p-1)} \oplus_{i=1}^{n} \underline{\mathrm{J}}_p - \frac{1}{n(n-1)p} \underline{\mathrm{J}}_{np} - \frac{1}{n(p-1)} \underline{\mathrm{I}}_{np}, \tag{2.2}$$

where $\underline{\mathrm{I}}_r$ is a identity matrix of dimension r, $\underline{\mathrm{J}}_r$ is a rxr matrix of 1's and $\oplus$ is the Kronecker sum or direct sum.

Under H_0 in (2.3) of the paper of reference we write

$$\begin{aligned} \hat{\xi}_i &= Y_i - \hat{m}_1(\mathbf{X}_{1i}) + m_1(\mathbf{X}_{1i}) - m_1(\mathbf{X}_{1i}) = \xi_i - (\hat{m}_1(\mathbf{X}_{1i}) - m_1(\mathbf{X}_{1i})) \\ &= \xi_i - \Delta_{m_1}(\mathbf{X}_{1i}), \end{aligned}$$

where $\Delta_{m_1}(\mathbf{X}_{1i})$ is defined implicitly in the above relation. Thus, $\hat{\boldsymbol{\xi}}_V$ is decomposed as $\hat{\boldsymbol{\xi}}_V = \boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1 V}$, where $\boldsymbol{\xi}_V$ and $\boldsymbol{\Delta}_{m_1 V}$ are defined as in (2.1) but using ξ_i and $\Delta_{m_1}(\mathbf{X}_{1i})$, respectively, instead of $\hat{\xi}_i$. Thus $\sqrt{n}$(MST - MSE) can be written as

$$\sqrt{n}\hat{\boldsymbol{\xi}}_V' A \hat{\boldsymbol{\xi}}_V = \sqrt{n}\boldsymbol{\xi}_V' A \boldsymbol{\xi}_V - \sqrt{n} 2 \boldsymbol{\xi}_V' A \boldsymbol{\Delta}_{m_1 V} + \sqrt{n} \boldsymbol{\Delta}_{m_1 V}' A \boldsymbol{\Delta}_{m_1 V}, \tag{2.3}$$

Using arguments similar to those used in the proof of Lemma 3.1 in Wang, Akritas and Van Keilegom (2008), it can be shown that if $\sigma^2(., x_2)$, defined in condition e) of the paper of reference, is Lipschitz continuous and $E(\epsilon_i^4) < \infty$ then, under H_0 and as n $\to \infty$,

$$n^{1/2}[\boldsymbol{\xi}_V' A \boldsymbol{\xi}_V - \boldsymbol{\xi}_V' A_d \boldsymbol{\xi}_V] \xrightarrow{p} 0, \tag{2.4}$$

where $A_d = diag\{B_1, ..., B_n\}$, with $B_i = \frac{1}{n(p-1)}[\underline{\mathrm{J}}_p - \underline{\mathrm{I}}_p]$, and moreover, it follows that $\sqrt{n}\boldsymbol{\xi}_V' A \boldsymbol{\xi}_V$ is asymptotically normal distributed. It remains to derive its asymptotic variance and to show that the other two terms in (2.3) converge to zero in probability. Using (2.4) it suffices to find the asymptotic variance of $\sqrt{n}\boldsymbol{\xi}_V' A_d \boldsymbol{\xi}_V$. Since $E(\boldsymbol{\xi}_V' A_d \boldsymbol{\xi}_V) = 0$ its variance equals $E[(\sqrt{n}\boldsymbol{\xi}_V' A_d \boldsymbol{\xi}_V)^2]$. To find this we first evaluate its conditional expectation, $E[(\sqrt{n}\boldsymbol{\xi}_V' A_d \boldsymbol{\xi}_V)^2 | \{X_{2j}\}_{j=1}^n]$, given $X_{21}, \ldots, X_{2n}$. Recalling the notation $\sigma^2(., x_2) = E(\xi^2 | X_2 = x_2)$, we have

$$\frac{1}{n(p-1)^2}\sum_{i_1,i_2}^{n}\sum_{j_1\neq l_1}^{n}\sum_{j_2\neq l_2}^{n} E(\xi_{j_1}\xi_{l_1}\xi_{j_2}\xi_{l_2}|\{X_{2j}\}_{j=1}^{n})I(j_s\in W_{i_s}, l_s\in W_{i_s}, s=1,2)$$

$$= \frac{2}{n(p-1)^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j\neq l}^{n}\sigma^2(.,x_{2j})\sigma^2(.,x_{2l})I(j,l\in W_{i_1}\cap W_{i_2}) \quad (2.5)$$

$$= \frac{2}{n(p-1)^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j\neq l}^{n}\sigma^2(.,x_{2j})\left(\sigma^2(.,x_{2j})+O_p\left(\frac{p}{\sqrt{n}}\right)\right)I(j,l\in W_{i_1}\cap W_{i_2})$$

$$= \frac{2}{n(p-1)^2}\sum_{j=1}^{n}\sigma^4(.,x_{2j})\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{l\neq j}^{n}I(j,l\in W_{i_1}\cap W_{i_2})+O_p\left(\frac{p^2}{n^{1/2}}\right)$$

$$= \frac{2}{n(p-1)^2}\sum_{j=1}^{n}\sigma^4(.,x_{2j})2(1+2^2+3^2+...+(p-1)^2)+O_p\left(\frac{p^2}{n^{1/2}}\right)$$

$$= \frac{2}{n(p-1)^2}\frac{p(p-1)(2p-1)}{3}\sum_{j=1}^{n}\sigma^4(.,x_{2j})+O_p\left(\frac{p^2}{n^{1/2}}\right),$$

where the third equality follows from Lemma 1.0.2 using the assumption that $\sigma^2(.,x_2)$ is Lipschitz continuous and the second last inequality results from the fact that if $1\leq |j_1-j_2| = s\leq p-1$, then they are $(p-s)^2$ pairs of windows whose intersection includes j_1 and j_2. Taking limits as $n\to\infty$ it is seen that

$$E\left(n^{1/2}\boldsymbol{\xi}_V' A_d\boldsymbol{\xi}_V\,\middle|\, X_2=x_2\right)^2 \overset{a.s.}{\longrightarrow} \frac{2(2p-1)}{3(p-1)}E(\sigma^4(.,X_2)) = \frac{2(2p-1)}{3(p-1)}\tau^2. \quad (2.6)$$

From relation (2.5) it is easily seen that $E[(\sqrt{n}\boldsymbol{\xi}_V' A_d\boldsymbol{\xi}_V)^2|\{X_{2j}\}_{j=1}^n]$ remains bounded, and thus $\mathrm{Var}(n^{1/2}\xi_V' A\xi_V)$ also converges to the same limit by the Dominated Convergence Theorem. Hence, $n^{1/2}\xi_V' A\xi_V$ converges in distribution to the designated normal distribution. That the second and third terms in (2.3) converge in probability to zero are shown in Lemmas 3.0.6, 3.0.7, respectively. □

Proof of Theorem 2. **Part 1: Local Additive Alternatives**

Note that we can write $\hat{\xi}_j = Y_j - \hat{m}_1(\mathbf{X}_{1j})$ as

$$\begin{aligned}\hat{\xi}_j &= Y_j - m_1(\mathbf{X}_{1j}) - n^{-1/4}\tilde{m}_2(X_{2j}) - [\hat{m}_1(\mathbf{X}_{1j}) - m_1(\mathbf{X}_{1j})] + n^{-1/4}\tilde{m}_2(X_{2j})\\ &= \xi_j - \Delta_{m_1}(\mathbf{X}_{1j}) + n^{-1/4}\tilde{m}_2(X_{2j}), \quad (2.7)\end{aligned}$$

and therefore

$\hat{\boldsymbol{\xi}}_V = \boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} + n^{-1/4}\tilde{\mathbf{m}}_{2V}$, where $\boldsymbol{\xi}_V$, $\boldsymbol{\Delta}_{m_1V}$ and $\tilde{\mathbf{m}}_{2V}$ are defined as in (2.1) but using ξ_i, $\Delta_{m_1}(\mathbf{X}_{1i})$ and $\tilde{m}_2(X_{2i})$, respectively, instead of $\hat{\xi}_i$. Thus, we can write

$$\begin{aligned}\sqrt{n}(MST - MSE) &= \sqrt{n}\hat{\boldsymbol{\xi}}_V' A\hat{\boldsymbol{\xi}}_V = \sqrt{n}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V})' A(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V}) + \\ &+ \sqrt{n}2n^{-1/4}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V})' A\tilde{\mathbf{m}}_{2V} \\ &\quad +\sqrt{n}(n^{-1/4})^2\tilde{\mathbf{m}}_{2V}' A\tilde{\mathbf{m}}_{2V}. \end{aligned} \tag{2.8}$$

By Theorem 1, $\sqrt{n}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V})' A(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V}) \xrightarrow{d} N(0, [2p(2p-1)\tau^2]/[3(p-1)])$. That $\sqrt{n}2n^{-1/4}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V})' A\tilde{\mathbf{m}}_{2V} \xrightarrow{p} 0$ and $\sqrt{n}(n^{-1/4})^2\tilde{\mathbf{m}}_{2V}' A\tilde{\mathbf{m}}_{2V} \xrightarrow{p} pV(\tilde{m}_2(X_2))$ are shown in Lemma 3.0.8 and Lemma 3.0.9, respectively. This completes the proof of part 1.

Part 2: Local General Alternatives

Working as in (2.7) we can write $\hat{\boldsymbol{\xi}}_V = \boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} + n^{-1/4}\tilde{\mathbf{m}}_{2V} + n^{-1/4}\tilde{\mathbf{m}}_{12V}$, where $\boldsymbol{\xi}_V$, $\boldsymbol{\Delta}_{m_1V}$, $\tilde{\mathbf{m}}_{2V}$ and $\tilde{\mathbf{m}}_{12V}$ are defined as in (2.1) but using ξ_i, $\Delta_{m_1}(\mathbf{X}_{1i})$, $\tilde{m}_2(X_{2i})$ and $\tilde{m}_{12}(\mathbf{X}_{1i}, X_{2i})$, respectively, instead of $\hat{\xi}_i$. Thus $\sqrt{n}(MST - MSE)$ is

$$\begin{aligned}\sqrt{n}\hat{\boldsymbol{\xi}}_V' A\hat{\boldsymbol{\xi}}_V &= \sqrt{n}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - n^{-1/4}\tilde{\mathbf{m}}_{2V})' A(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - n^{-1/4}\tilde{\mathbf{m}}_{2V}) \\ &\quad +\sqrt{n}2n^{-1/4}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - n^{-1/4}\tilde{\mathbf{m}}_{2V})' A\tilde{\mathbf{m}}_{12V} \\ &\quad +\sqrt{n}(n^{-1/4})^2\tilde{\mathbf{m}}_{12V}' A\tilde{\mathbf{m}}_{12V}. \end{aligned} \tag{2.9}$$

By Part 1 of the theorem, $\sqrt{n}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - n^{-1/4}\tilde{\mathbf{m}}_{2V})' A(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - n^{-1/4}\tilde{\mathbf{m}}_{2V})$ converges in distribution to

$$N(pVar(m_2(X_2)), [2p(2p-1)\tau^2]/[3(p-1)]).$$

Hence, it is enough to show that $\sqrt{n}(n^{-1/4})^2\tilde{\mathbf{m}}_{12V}' A\tilde{\mathbf{m}}_{12V} \xrightarrow{p} pVar(\tilde{m}_{12}(\mathbf{X}_1, X_2))$ and $\sqrt{n}2n^{-1/4}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - n^{-1/4}\tilde{\mathbf{m}}_{2V})' A\tilde{\mathbf{m}}_{12V} \xrightarrow{p} 2pCov(\tilde{m}_2(X_2), \tilde{m}_{12}(\mathbf{X}_1, X_2))$. These are shown in Lemmas 3.0.11 and 3.0.10, respectively. □

3. Some Detailed Derivations

Lemma 3.0.6. *The second term in (2.3) converges in probability to zero, i.e.*

$$T_{2n} := \sqrt{n}\boldsymbol{\xi}_V^T A\boldsymbol{\Delta}_{m_1V} \xrightarrow{p} 0.$$

Proof. After some algebra it can be seen that

$$\begin{aligned} T_{2n} &= \frac{n^{-1/2}(np-1)}{(n-1)p(p-1)}\sum_{i=1}^{n}\sum_{j\in W_i}\xi_j\sum_{k\in W_i}\Delta_{m_1}(\mathbf{X}_k) \\ &\quad -\frac{n^{-1/2}p}{(n-1)}\sum_{i=1}^{n}\xi_i\sum_{j=1}^{n}\Delta_{m_1}(\mathbf{X}_j)-\frac{n^{-1/2}p}{(p-1)}\sum_{i=1}^{n}\xi_i\Delta_{m_1}(\mathbf{X}_i). \end{aligned} \tag{3.1}$$

We will show that each of the three terms above converge in probability to zero conditionally on $\mathbf{U} = \{\mathbf{X}, Z\}$, and thus also unconditionally. Note that, because all windows W_i are of finite size (p), the first term on the right hand side of (3.1) can be written as a finite (p^2) sum of terms each of which is similar to the last term in (3.1). Thus, it suffices to show that the last and second terms of (3.1) converge to zero. For notational simplicity, all expectations and variances in this proof are to be understood as conditional on $\mathbf{U} = \{\mathbf{X}, Z\}$. For the second term in (3.1), note that $n^{-1/2}\sum_{i=1}^{n}\xi_i$ remains bounded in probability, and therefore, its convergence to zero will follow if we show that $n^{-1}\sum_{k=1}^{n}\Delta_{m_1}(\mathbf{X}_k)\xrightarrow{p} 0$. For later use, we will actually show that

$$\frac{1}{n^{3/4}}\sum_{k=1}^{n}\Delta_{m_1}(\mathbf{X}_k)=\frac{1}{n^{3/4}}\sum_{k=1}^{n}(\hat{m}_1(\mathbf{X}_k)-m_1(\mathbf{X}_k))\xrightarrow{p} 0. \tag{3.2}$$

By Theorem 6 in Masry (1996), it follows that

$$\sup_{\mathbf{x}}|\hat{m}_1(\mathbf{x})-m_1(\mathbf{x})| = O\left(\left(\frac{log(n)}{n\lambda_i^{d_1}}\right)^{\frac{1}{2}}\right)+O\left(\lambda_i^{q+1}\right). \tag{3.3}$$

Thus

$$\frac{1}{n^{3/4}}\sum_{k=1}^{n}\Delta_{m_1}(\mathbf{X}_k) = O\left(n^{1/4}\left(\frac{log(n)}{n\lambda_i^{d_1}}\right)^{\frac{1}{2}}\right)+O\left(n^{1/4}\lambda_i^{q+1}\right)=o(1),$$

where the last equality follows from the assumptions of the Theorem 2.0.1.

Consider now the last term in (3.1). Because the weights $\tilde{w}(\mathbf{X}_i, \mathbf{X}_j)$ of the local polynomial regression sum to 1 (Lemma 1.0.3), we can write

$$\begin{aligned}\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\xi_i\Delta_{m_1}(\mathbf{X}_i) &= n^{-1/2}\sum_{i=1}^{n}\xi_i(\hat{m}(\mathbf{X}_i)-m(\mathbf{X}_i)) \\ &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{w}(\mathbf{X}_i,\mathbf{X}_j)(m(\mathbf{X}_j)+\xi_j-m(\mathbf{X}_i))\xi_i \\ &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{w}(\mathbf{X}_i,\mathbf{X}_j)(m(\mathbf{X}_j)-m(\mathbf{X}_i))\xi_i \\ &\quad +n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{w}(\mathbf{X}_i,\mathbf{X}_j)\xi_j\xi_i. \qquad (3.4)\end{aligned}$$

The first term of the right hand side of (3.4) has zero expectation, so it suffices to show that its variance goes to zero. To this end, we write

$$\begin{aligned}&Var(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\tilde{w}(\mathbf{X}_i,\mathbf{X}_j)(m_1(\mathbf{X}_j)-m_1(\mathbf{X}_i))) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\tilde{w}(\mathbf{X}_i,\mathbf{X}_{j_1})\tilde{w}(\mathbf{X}_i,\mathbf{X}_{j_2})\times \\ &\qquad \times(m_1(\mathbf{X}_{j_1})-m_1(\mathbf{X}_i))(m_1(\mathbf{X}_{j_2})-m_1(\mathbf{X}_i))Var(\xi_i) \\ &\leq \frac{M}{n}\sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\tilde{w}(\mathbf{X}_i,\mathbf{X}_{j_1})\tilde{w}(\mathbf{X}_i,\mathbf{X}_{j_2})\times \\ &\qquad \times(c||\mathbf{X}_{j_1}-\mathbf{X}_i||c||\mathbf{X}_{j_2}-\mathbf{X}_i||) \\ &= Mc^2O(||H_n^{1/2}||)O(||H_n^{1/2}||)=o(1),\end{aligned}$$

for some constants M and c, where the inequality holds because $m_1(\cdot)$ is Lipschitz continuous, the last equality follows from Lemma 1.0.4, and the second to last equality from Lemma 1.0.5. Thus, by the assumptions of Theorem 2.0.1, the first term of the right hand side of (3.4) goes in probability to zero. To show that the second term in (3.4) also goes to 0 in probability, we will show that its second moment goes to zero. To this end, we write

$$
\begin{aligned}
&E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j\tilde{w}(\mathbf{X}_i,\mathbf{X}_j)\right]^2\\
&=E\left[\frac{1}{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\xi_{i_1}\xi_{i_2}\xi_{j_1}\xi_{j_2}\tilde{w}(\mathbf{X}_{i_1},\mathbf{X}_{j_1})\tilde{w}(\mathbf{X}_{i_2},\mathbf{X}_{j_2})\right]\\
&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E(\xi_i^2\xi_j^2)\tilde{w}(\mathbf{X}_i,\mathbf{X}_j)^2\\
&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E(\xi_i^2\xi_j^2)\tilde{w}(\mathbf{X}_i,\mathbf{X}_i)\tilde{w}(\mathbf{X}_j,\mathbf{X}_j)\\
&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E(\xi_i^2\xi_j^2)\tilde{w}(\mathbf{X}_i,\mathbf{X}_j)\tilde{w}(\mathbf{X}_j,\mathbf{X}_i)\\
&+\frac{1}{n}\sum_{i=1}^{n}E(\xi_i^4)\tilde{w}(\mathbf{X}_i,\mathbf{X}_i)\tilde{w}(\mathbf{X}_i,\mathbf{X}_i)\\
&=O\left(\frac{1}{n|H_n|}\right)+O\left(\frac{1}{n|H_n|}\right)+O\left(\frac{1}{n|H_n|}\right)+O\left(\frac{1}{n^2|H_n|}\right)\\
&=o(1), \qquad (3.5)
\end{aligned}
$$

by the fact that $E(\xi_i^4)$ is bounded and that each weight is of the order $\frac{1}{n|H_n|^{1/2}}$ (Lemma 1.0.4). Thus, by the assumptions of Theorem 2.0.1, the second term of the right hand side of (3.4) goes in probability to zero.

This completes the proof of Lemma 3.0.6. □

Lemma 3.0.7. *The third term in (2.3) converges in probability to zero, i.e.*

$$T_{3n}=\sqrt{n}\boldsymbol{\Delta}_{m_1V}^T A\boldsymbol{\Delta}_{m_1V}\xrightarrow{p}0.$$

Proof. Similarly to Lemma 3.0.6, we can write

$$
\begin{aligned}
T_{3n} &= \frac{\sqrt{n}(np-1)}{n(n-1)p(p-1)}\sum_{i=1}^{n}\left(\sum_{j\in W_i}\Delta_{m_1}(\mathbf{X}_j)\right)^2\\
&- \frac{\sqrt{n}p}{n(n-1)}\left(\sum_{i=1}^{n}\Delta_{m_1}(\mathbf{X}_i)\right)^2-\frac{\sqrt{n}p}{n(p-1)}\sum_{i=1}^{n}\Delta_{m_1}^2(\mathbf{X}_i). \qquad (3.6)
\end{aligned}
$$

we have to show that each of the three terms on the right hand side of (3.6) converges to zero in probability Again, because all windows W_i are of finite size (p), the first term on the right hand side of (3.6) can be written as a finite (p^2) sum of terms each of which is similar to the last term in (3.6). Thus, it suffices to show that the last and second terms of (3.6) converge to zero.

Recall that (Masry, 1996)

$$\sup_{\mathbf{x}} |\hat{m}_1(\mathbf{x}) - m_1(\mathbf{x})| = O\left(\left(\frac{log(n)}{n\lambda_i^{d_1}}\right)^{\frac{1}{2}}\right) + O\left(\lambda_i^{q+1}\right).$$

Replacing Δ_{m_1} by its order, the second term in (3.6) is or order

$$O\left(\frac{\sqrt{n}}{n^2}\left(n\left(\frac{log(n)}{n\lambda_i^{(d-1)}}\right)^{\frac{1}{2}} + n\lambda_i^{q+1}\right)^2\right)$$
$$= O\left(\left(n^{1/4}\left(\frac{log(n)}{n\lambda_i^{d_1}}\right)^{\frac{1}{2}} + n^{1/4}\lambda_i^{q+1}\right)^2\right) = o(1)$$

where the last equality follows from the assumptions of the theorem.

Similarly, the third term in (3.6) is or order

$$O\left(\frac{\sqrt{n}}{n}n\left(\left(\frac{log(n)}{n\lambda_i^{d_1}}\right)^{\frac{1}{2}} + \lambda_i^{q+1}\right)^2\right)$$
$$= O\left(\left(n^{1/4}\left(\frac{log(n)}{n\lambda_i^{d_1}}\right)^{\frac{1}{2}} + n^{1/4}\lambda_i^{q+1}\right)^2\right) = o(1)$$

This completes the proof of Lemma 3.0.7. □

Lemma 3.0.8. *The second term in (2.8) converges in probability to zero, i.e.*

$$\sqrt{n}2n^{-1/4}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V})'A\tilde{\mathbf{m}}_{2V} \xrightarrow{p} 0.$$

Proof. By the definition of the matrix A, we can write $(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V})' A \tilde{\mathbf{m}}_{2V}$ as

$$\frac{np-1}{n(n-1)p(p-1)} \sum_{i=1}^{n} \left[\sum_{j=1}^{n} \tilde{m}_2(X_{2j}) I(j \in W_i) \right] \left[\sum_{k=1}^{n} (\xi_k - \Delta_{m_1}(\mathbf{X}_{1k})) I(k \in W_i) \right]$$
$$- \frac{1}{n(n-1)p} \left[p \sum_{i=1}^{n} \tilde{m}_2(X_{2i}) \right] \left[p \sum_{i=1}^{n} (\xi_i - \Delta_{m_1}(\mathbf{X}_{1i})) \right]$$
$$- \frac{p}{n(p-1)} \sum_{i=1}^{n} \tilde{m}_2(X_{2i})(\xi_i - \Delta_{m_1}(\mathbf{X}_{1i})).$$

Using Lemma 1.0.2 and the fact that $\tilde{m}_2(\cdot)$ is Lipschitz continuous, the sum in the first term can be expressed as

$$\begin{aligned} & p \sum_{i=1}^{n} [\tilde{m}_2(X_{2i}) + O(n^{-1/2})] \left[\sum_{k=1}^{n} (\xi_k - \Delta_{m_1}(\mathbf{X}_{1k})) I(k \in W_i) \right] \leq \\ \leq \; & p \sum_{k=1}^{n} \left[\sum_{i=1}^{n} \tilde{m}_2(X_{2i}) I(i \in W_k) \right] (\xi_k - \Delta_{m_1}(\mathbf{X}_{1k})) \\ & + p^2 O(n^{-1/2}) \sum_{k=1}^{n} |(\xi_k - \Delta_{m_1}(\mathbf{X}_{1k}))| \\ = \; & p^2 \sum_{k=1}^{n} \tilde{m}_2(X_{2k})(\xi_k - \Delta_{m_1}(\mathbf{X}_{1k})) + O_p(p^2 n^{1/2}), \end{aligned}$$

so that

$$\begin{aligned} \sqrt{n} n^{-1/4} \tilde{\mathbf{m}}_{2V}' A (\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V}) \; &= \; \frac{n^{1.25} p}{n-1} \left[\frac{1}{n} \sum_{i=1}^{n} \tilde{m}_2(X_{2i})(\xi_i - \Delta_{m_1}(\mathbf{X}_{1i})) \right] \\ & - \frac{n^{1.25} p}{n-1} \left[\frac{1}{n} \sum_{i=1}^{n} \tilde{m}_2(X_{2i}) \right] \left[\frac{1}{n} \sum_{i=1}^{n} (\xi_i - \Delta_{m_1}(\mathbf{X}_{1i})) \right] + O_p\left(\frac{1}{n^{1/4}} \right). \end{aligned}$$

Using the fact that $E(\tilde{m}_2(X_{2i})) = E(\tilde{m}_2(X_{2i})\xi_i) = E(\xi_i) = 0$, relation (3.2) and also that $n^{-3/4} \sum_{i=1}^{n} \tilde{m}_2(X_{2i}) \Delta_{m_1}(\mathbf{X}_{1i}) \xrightarrow{p} 0$, as is shown in a similar way to (3.2), completes the proof of the lemma. □

Lemma 3.0.9. *The third term in (2.8) converges in probability to $pV(\tilde{m}_2(X_2))$, i.e.*

$$\sqrt{n}(n^{-1/4})^2 \tilde{\mathbf{m}}_{2V}' A \tilde{\mathbf{m}}_{2V} \xrightarrow{p} pV(\tilde{m}_2(X_2)).$$

Proof. Writing

$$\begin{aligned}\tilde{m}_{2V}' A \tilde{m}_{2V} &= \frac{np}{n-1}\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{m}_2^2(X_{2i})\right] - \left[\frac{1}{n}\sum_{i=1}^{n}\tilde{m}_2(X_{2i})\right]^2\right\} + O_p\left(\frac{1}{n^{1/2}}\right)\\ &= p\left\{E\tilde{m}_2^2(X_2) - [E\tilde{m}_2(X_2)]^2\right\} + O_p\left(\frac{1}{n^{1/2}}\right),\end{aligned}$$

it follows that

$$\sqrt{n}(n^{-1/4})^2 \tilde{m}_{2V}' A \tilde{m}_{2V} = p\mathrm{Var}(\tilde{m}_2(X_2)) + O_p\left(\frac{1}{n^{1/2}}\right),$$

which completes the proof. □

Lemma 3.0.10. *The second term in (2.9) converges in probability to* $2pCov(\tilde{m}_2(X_2), \tilde{m}_{12}(\mathbf{X}_1, X_2))$*, i.e.*

$$\sqrt{n}2n^{-1/4}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - n^{-1/4}\tilde{\mathbf{m}}_{2V})' A \tilde{\mathbf{m}}_{12V} \xrightarrow{p} 2pCov(\tilde{m}_2(X_2), \tilde{m}_{12}(\mathbf{X}_1, X_2)).$$

Proof. By the definition of the matrix A, we can write

$$\begin{aligned}&\sqrt{n}n^{-1/4}(\boldsymbol{\xi}_V - \boldsymbol{\Delta}_{m_1V} - \rho_{1n}\tilde{\mathbf{m}}_{2V})' A \tilde{\mathbf{m}}_{12V} = \sqrt{n}n^{-1/4}\frac{np-1}{n(n-1)p(p-1)} \times\\ &\quad \times \sum_{i=1}^{n}\left[\sum_{j=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1j}, X_{2j}) I(j \in W_i)\right] \times\\ &\quad \times \left[\sum_{k=1}^{n}(\xi_k - \Delta_{m_1}(\mathbf{X}_{1k}) - n^{-1/4}\tilde{m}_2(X_{2k})) I(k \in W_i)\right]\\ &-\sqrt{n}n^{-1/4}\frac{1}{n(n-1)p}\left[p\sum_{i=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1i}, X_{2i})\right] \times\\ &\quad \left[p\sum_{i=1}^{n}(\xi_i - \Delta_{m_1}(\mathbf{X}_{1i}) - n^{-1/4}\tilde{m}_2(X_{2i}))\right]\\ &-\sqrt{n}n^{-1/4}\frac{p}{n(p-1)}\sum_{i=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1i}, X_{2i})(\xi_i - \Delta_{m_1}(\mathbf{X}_{1i}) - \rho_{1n}\tilde{m}_2(X_{2i})). \quad (3.7)\end{aligned}$$

Noting that $n^{-3/4}\sum_{i=1}^{n}\xi_i \tilde{m}_{12}(\mathbf{X}_{1i}, X_{2i}) \xrightarrow{p} 0$, and $\frac{1}{n^{3/4}}\sum_{i=1}^{n}\Delta_{m_1}(\mathbf{X}_{1i}) m_{12}(\mathbf{X}_{1i}, X_{2i}) \xrightarrow{p} 0$, which follows by arguments similar to (3.2), the third term in (3.7) goes in probability to $[p/(p-1)]E(\tilde{m}_2(X_2)\tilde{m}_{12}(\mathbf{X}_1, X_2))$. Also, using (3.2), and the facts $E(\tilde{m}_{12}(\mathbf{X}_1, X_2)) = 0$, and $n^{-3/4}\sum_{i=1}^{n}\xi_i = o_p(1)$, the second term in (3.7) goes to $pE(\tilde{m}_2(X_2))E(\tilde{m}_{12}(\mathbf{X}_1, X_2))$ in probability. Next, the component of the first term in (3.7) that corresponds to

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1j}, X_{2j})(\xi_k - \Delta_{m_1}(\mathbf{X}_{1k}))I(j \in W_i)I(k \in W_i)$$

goes to zero in probability by arguments similar to those used for the last term in (3.7). Set $\bar{m}_2^i(X_{2i}) = \frac{1}{p}\sum_{j=1}^{n}\tilde{m}_2(X_{2j})I(j \in W_i)$ and $\bar{m}_{12}^i(., X_{2i}) = \frac{1}{p}\sum_{j=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1j}, X_{2i})I(j \in W_i)$, so that

$$\frac{1}{p}\sum_{j=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1j}, X_{2j})I(j \in W_i) = \bar{m}_{12}^i(., X_{2i}) + o_p(1),$$
$$\frac{1}{p}\sum_{j=1}^{n}\tilde{m}_2(X_{2j})I(j \in W_i) = \bar{m}_2^i(X_{2i}) + o_p(1).$$

The remaining component of the first term in (3.7) can be written as

$$\begin{aligned}
&\frac{(np-1)}{n(n-1)p(p-1)}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1j}, X_{2j})\tilde{m}_2(X_{2k})I(j \in W_i)I(k \in W_i)\\
= \ &\frac{(np-1)p}{(n-1)(p-1)}\frac{1}{n}\sum_{i=1}^{n}\bar{m}_{12}^i(., X_{2i})\bar{m}_2^i(X_{2i}) + o_p(1)\\
&\xrightarrow{p} \frac{p^2}{p-1}E\left[\tilde{m}_{12}(\mathbf{X}_1, X_2)\tilde{m}_2(X_2)\right],
\end{aligned}$$

completing the proof. □

Lemma 3.0.11. *The third term in (2.9) converges in probability to* $pVar(\tilde{m}_{12}(\mathbf{X}_1, X_2))$, *i.e.*

$$\sqrt{n}(n^{-1/4})^2\tilde{\mathbf{m}}_{12V}' A\tilde{\mathbf{m}}_{12V} \xrightarrow{p} pVar(\tilde{m}_{12}(\mathbf{X}_1, X_2)).$$

Proof. Note that we can write $\sqrt{n}(n^{-1/4})^2\tilde{\mathbf{m}}'_{12V}A\tilde{\mathbf{m}}_{12V}$ as

$$\frac{(np-1)}{n(n-1)p(p-1)}\sum_{i=1}^{n}\left[\sum_{j=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1j},X_{2j})I(j\in W_i)\right]\times \left[\sum_{k=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1k},X_{2k}))I(k\in W_i)\right] -\frac{p}{n(n-1)}\left[\sum_{i=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1i},X_{2i})\right]\left[\sum_{i=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1i},X_{2i})\right] -\frac{p}{n(p-1)}\sum_{i=1}^{n}\tilde{m}_{12}(\mathbf{X}_{1i},X_{2i})^2. \tag{3.8}$$

Clearly, the third term in (3.8) goes to $[p/(p-1)]E[\tilde{m}_{12}(\mathbf{X}_1,X_2)^2]$ in probability, and the second term in (3.8) goes to $p[E(\tilde{m}_{12}(\mathbf{X}_1,X_2))]^2$ in probability. Using the same notation as in lemma 3.0.10, the first term in (3.8) is equal to

$$\frac{(np-1)p}{n(n-1)(p-1)}\sum_{i=1}^{n}\left[\bar{m}_{12}^{i}(.,X_{2i})\right]^2+o_p(1)\xrightarrow{p}\frac{p^2}{p-1}E[(\tilde{m}_{12}(\mathbf{X}_{1i},X_{2i}))^2],$$

completing the proof. □

Acknowledgment

This research was partially supported by CAPES/Fulbright grant 15087657, FAPESP 2012/22603-6 and 2012/10808-2, and NSF grants DMS-0805598 and DMS-1209059.

References

Dvoretzky, A.; Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, **27** (3), 642669.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of time series analysis*, **17**, 571-599.

Ruppert, D. and Wand, M. P. (1994). Multivariate Locally Weighted Least Squares Regression. *The Annals of Statistics*, **22**, 1346-1370.

Department of Statistics, State University of Campinas, Rua Sérgio Buarque de Holanda, 651. Cidade Universitária "Zeferino Vaz" Distr. Barão Geraldo. Campinas, São Paulo, Brasil - 13083-859.

E-mail: zambom@ime.unicamp.br

Department of Statistics, Penn State University, 326 Thomas Bldg., University Park, PA 16802, USA.

E-mail: mga@stat.psu.edu