Supplementary Materials: Nonparametric Regression with Common Shocks

Eduardo A. Souza-Rodrigues

In this supplement document, we first present the kernel density estimator, the kernel regression estimator and provide conditions to obtain their asymptotic results (Section A). We proceed with the proofs of the asymptotic results (Section B). We then briefly discuss the choice of the bandwidth (Section C). Finally, we present the probabilistic framework adapted from Andrews [1] that justifies the approach taken in the paper (Section D).

Appendix A. Model, Assumptions and Estimators

In this section we first present the kernel density estimator and its asymptotic properties (Subsection A.1). The asymptotic results for the kernel density are not in the main text. Then, we present the kernel regression and its asymptotic properties (Subsection A.2). Proposition 2 here corresponds to the Proposition 1 in the main text. We restate all the assumptions presented in the paper for completeness.

Appendix A.1. Density Estimator

The Nadaraya-Watson kernel density estimator is

$$\widehat{f}(x) = \frac{1}{nh_n^k} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right). \tag{1}$$

where k is the dimension of X. We assume the following conditions (presented in the main text):

Condition 1. *Let K be the class of all Borel measurable nonnegative bounded real-valued functions $K(u)$ such that (i) $\int K(u)du = 1$; (ii) $\int |K(u)|\, du < \infty$; (iii) $|K(u)|\, \|u\|^k \to 0$ as $\|u\| \to \infty$; (iv) $\kappa = \int K^2(u)du < \infty$; (v) $\sup_u |K(u)| < \infty$; and (vi) $\mu_2 = \int u^2 K(u)du < \infty$.*

Condition 2. *For Q-almost all $c \in \mathcal{C}$, the conditional density $f(x|c)$ is continuous at any point x_0.*

Condition 3. *(i) $h_n \to 0$ as $n \to \infty$ and (ii) $nh_n \to \infty$ as $n \to \infty$.*

Condition 4. *For Q-almost all c, (i) $f(x|c)$ is twice continuously differentiable with respect to x in some neighborhood of x_0, (ii) the second-order derivatives of $f(x|c)$ with respect to x are bounded in this neighborhood.*

Condition 5. *For Q-almost all c, the point x_0 is in the interior of the support of X conditional on $\{C = c\}$, and $f(x_0|c) \geq \xi > 0$, for some finite ξ.*

To derive the asymptotic results for $\widehat{f}(x)$, we first show that the conditional Mean-Squared Error (MSE) converges to zero in probability to obtain consistency. Next, we show that the rate of converge here is the same as the rate of convergence without common shocks. Then we obtain the pointwise asymptotic distribution using the Martingale Difference Sequence Central Limit Theorem (MDS CLT).

Proposition 1. *1. Under Assumptions 1, 2 and Conditions 1–3,*

$$\widehat{f}(x_0) \xrightarrow{p} f(x_0|C), \text{ as } n \to \infty.$$

2. *Under Assumptions 1, 2, and Conditions 1–4,*

$$\widehat{f}(x_0) - f(x_0|C) = O_p\left(n^{-\frac{2}{4+k}}\right).$$

3. *Let Assumptions 1, 2 and Conditions 1–3 hold. Suppose that $\int |K(u)|^{2+\delta} du < \infty$, for some $\delta > 0$. Then, (i) as $n \to \infty$*

$$\sqrt{nh_n^k}\left(\widehat{f}(x_0) - E\left(\widehat{f}(x_0) \mid C\right)\right) \xrightarrow{d} \left(f(x_0|C)\int K^2(u)\,du\right) N(0,1);$$

and (ii) if, in addition, Conditions 4–5 hold and $\sqrt{nh_n^k}h_n^2 \to 0$ as $n \to \infty$, then

$$\sqrt{nh_n^k}\left(\widehat{f}(x_0) - f(x_0|C)\right) \xrightarrow{d} \left(f(x_0|C)\int K^2(u)\,du\right) N(0,1)$$

as $n \to \infty$.

Remark 1. *The probability distribution of $f(x|C)$ is the probability measure induced by the map $f(x|C) : C \to \mathbb{R}_+$ and by Q. I.e., for some $a \in \mathbb{R}$, $\Pr\{f(x|C) \le a\} = Q\{C \in \mathcal{C} : f(x|C) \le a\}$, where $\{C \in \mathcal{C} : f(x|C) \le a\} \in \sigma(C) \subset \mathcal{A}^{\mathbb{N}}$.*

Note that if $C(X) = \sum_{j=1}^{\infty} C_j \phi_j(X) \in \mathcal{L}_2(\mathcal{X})$, as in Remark 1 in the main text, then $f_{X|C(\cdot)} = f_{X|(C_1,C_2,\ldots)}$ and

$$\Pr(X \le x | \{C(\cdot) = c\}) = \int_{-\infty}^{x} f_{X|(C_1,C_2,\ldots)}(\widetilde{x}|c_1, c_2, \ldots)\, d\widetilde{x}$$

takes values between zero and one. This conditional probability is different from

$$\Pr(X \le x | \{c(X) = a\}) = \int_{-\infty}^{x} f_{X|c(X)}(\widetilde{x}|a)\, d\widetilde{x},$$

which takes value zero or one if $C(X)$ is invertible on X.

Remark 2. *Although the condition $\sqrt{nh_n^k}h_n^2 \to 0$ eliminates the random bias $E\left(\widehat{f}(x) \mid C\right) - f(x|C)$ in Proposition 1.3, the difference $f(x|C) - f(x)$ does not die out when $n \to \infty$. The difference can only be eliminated when X and C are independent.*

Appendix A.2. Regression Estimator

The kernel estimator is

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}.$$

In addition to Conditions 1–5, we impose the following (presented in the main text):

Condition 6. *The kernel K is a symmetric function satisfying $\int uK(u)du = 0$.*

Condition 7. *(i) $E[\varepsilon_i \mid X_i, \sigma(C)] = 0$ a.s., and (ii) let $\sigma^2(x,c) = E\left(\varepsilon_i^2 \mid X_i = x, C = c\right)$ and assume $\sigma^2(X,C) < \infty$ a.s..*

Condition 8. *For Q-almost all c, the function $m(x,c)$ is twice continuously differentiable with respect to x in some neighborhood of x_0.*

Next we obtain the asymptotic results for the kernel regression. Let $E\left[\cdot \mid X = \{x_i\}_{i=1}^{n}\right]$ denote the conditional expectation given x_i, $i = 1, ..., n$. The Proposition below is the same presented in the main text.

Proposition 2. *Let Assumptions 1, 2 and Conditions 1–8 hold. Then*

1. $\widehat{m}(x) \xrightarrow{p} m(x, C)$ *as* $n \to \infty$.
2. $\widehat{m}(x) - m(x, C) = O_p\left(n^{-\frac{2}{4+k}}\right)$
3. *Suppose also that* $\int |K(u)|^{2+\delta} du < \infty$ *and* $E\left[|\varepsilon|^{2+\delta}\right] < \infty$, *for some* $\delta > 0$. *Then, (i) as* $n \to \infty$

$$\sqrt{nh_n^k}\left(\widehat{m}(x) - E\left[\widehat{m}(x) \mid X = \{x_i\}_{i=1}^{n}, C\right]\right) \xrightarrow{d} \left(\frac{\sigma^2(x, C)}{f(x|C)}\int K^2(u)\, du\right) N(0, 1);$$

and (ii) if, in addition, $\sqrt{nh_n^k}h_n^2 \to 0$ *as* $n \to \infty$, *then*

$$\sqrt{nh_n^k}\left(\widehat{m}(x) - m(x, C)\right) \xrightarrow{d} \left(\frac{\sigma^2(x, C)}{f(x|C)}\int K^2(u)\, du\right) N(0, 1)$$

as $n \to \infty$.

Appendix B. Proofs of the Propositions

First we present the proofs for the kernel density then, for the kernel regression.

Appendix B.1. Proofs for the Density Estimator

First, we need the following Lemma, which is a version of the Dominated Convergence Theorem:

Lemma 1. *Under Assumptions 1, 2 and Conditions 1–3(i), we have for Q-almost all* $c \in \mathcal{C}$ *and for any* $r \geq 0$

$$\begin{aligned} \frac{1}{h_n^k}E\left[K^r\left(\frac{X - x_0}{h_n}\right) \mid C = c\right] &= \frac{1}{h_n^k}\int K^r\left(\frac{x - x_0}{h_n}\right) f(x|c)\, dx \\ &\longrightarrow f(x_0|c)\int K^r(x)\, dx \end{aligned} \tag{2}$$

as $n \to \infty$.

Proof. The proof is similar to Corollary 2 in the Appendix A.2.6 in Pagan and Ullah [2], which is based on their Lemma 1 in the same section. The only difference here is that we have to substitute their density $f(x)$ by the conditional density $f(x|c)$. ☐

Because Lemma 1 is valid for Q-almost all c, the almost-sure convergence implies convergence in probability. Hence, we have the following Corollary:

Corollary 1. *Under Assumptions 1, 2 and Conditions 1–3(i), we have for any* $r \geq 0$

$$E\left[\frac{1}{h_n^k}K^r\left(\frac{X - x_0}{h_n}\right) \mid C\right] \xrightarrow{p} f(x_0|C)\int K^r(x)\, dx \tag{3}$$

as $n \to \infty$, *where both* $E\left[\frac{1}{h_n^k}K^r\left(\frac{X-x_0}{h_n}\right) \mid C\right]$ *and* $f(x_0|C)$ *are* $\sigma(C)$*-measurable random variables.*

Next, we prove Proposition 1.1–1.3.

Proof of Proposition 1.1. By definition,

$$MSE\left(\widehat{f}\left(x_0\right)\mid C\right)=\left[E\left(\widehat{f}\left(x\right)\mid C\right)-f\left(x|C\right)\right]^2+Var\left(\widehat{f}\left(x_0\right)\mid C\right).$$

We start with the bias term. For Q-almost all c,

$$\begin{aligned}E\left(\widehat{f}\left(x_0\right)|C=c\right) &= \frac{1}{nh_n^k}\sum_{i=1}^{n}E\left(K\left(\frac{X_i-x_0}{h_n}\right)|C=c\right)\\ &= \frac{1}{h_n^k}E\left(K\left(\frac{X_i-x_0}{h_n}\right)|C=c\right)\\ &\longrightarrow f\left(x_0|c\right)\int K\left(u\right)du, \quad as\ n\to\infty\\ &= f\left(x_0|c\right),\end{aligned}$$

by Lemma 1 and Condition 1(i). The second equality above comes from Assumption 1 (conditional i.i.d. observations). This Q-almost sure convergence implies the convergence in probability of $E\left(\widehat{f}\left(x_0\right)|C\right)$ to $f\left(x_0|C\right)$ and, hence,

$$E\left(\widehat{f}\left(x_0\right)\mid C\right)-f\left(x_0|C\right)\xrightarrow{p}0.$$

Next, we look at the variance term, $Var\left(\widehat{f}\left(x_0\right)\mid C\right)$. Again, for Q-almost all c,

$$\begin{aligned}Var\left(\widehat{f}\left(x_0\right)\mid C=c\right) &= \frac{1}{nh_n^k}\int K^2\left(\frac{x-x_0}{h_n}\right)f\left(x|c\right)dx\\ &\quad -\left[\frac{1}{n}\int K\left(\frac{x-x_0}{h_n}\right)f\left(x|c\right)dx\right]^2\\ &\longrightarrow 0\times\left(f\left(x_0|c\right)\int K^2\left(u\right)du\right)-0\times\left[f\left(x_0|c\right)\right]^2\\ &= 0\end{aligned}$$

as $n\to\infty$, by Lemma 1 and Conditions 1–3. Therefore, we have $Var\left(\widehat{f}\left(x_0\right)\mid C\right)\xrightarrow{p}0$ and, as a result $MSE\left(\widehat{f}\left(x_0\right)\mid C\right)\xrightarrow{p}0$.

Now, for any $\varepsilon>0$, define the event $A_n=\left\{\left|\widehat{f}\left(x_0\right)-f\left(x_0|C\right)\right|>\varepsilon\right\}$. Then, by Markov inequality,

$$\begin{aligned}\Pr\left(A_n\mid\sigma(C)\right) &\leq \frac{1}{\varepsilon^2}E\left[\left(\widehat{f}\left(x_0\right)-f\left(x_0|C\right)\right)^2\mid\sigma(C)\right]\\ &= \frac{1}{\varepsilon^2}MSE\left(\widehat{f}\left(x_0\right)\mid C\right)\xrightarrow{p}0.\end{aligned}$$

So, the $\sigma(C)$-measurable random variable $\Pr\left(A_n\mid\sigma(C)\right)$ is $o_p(1)$. Moreover, $\Pr\left(A_n\mid\sigma(C)\right)\leq 1$ and $\int_C 1dQ\left(C\right)=1$. Hence, by the Dominated Convergence Theorem,

$$\Pr\left(A_n\right)=E_C\left[\Pr\left(A_n\mid\sigma(C)\right)\right]\longrightarrow 0,$$

where E_C denotes the expectation taken over Q. □

Proof of Proposition 1.2. The argument follows the standard proof modified to consider the conditional density function. Under the Assumptions 1, 2, and Conditions 1–4, by taking a (Q-almost sure) Taylor expansion we obtain the conditional Mean Squared Errors:

$$MSE\left(\widehat{f}\left(x_0\right)\mid C\right)=\left[\frac{h_n^4}{4}\mu_2\left(\left[D_x^2 f\left(x_0|C\right)\right]^2\right)+O_p\left(h_n^2\right)\right]+\left[\frac{\kappa}{nh_n^k}+O_p\left(\frac{1}{nh_n^k}\right)\right],$$

where $D_x^2 f(x|c)$ denotes the second derivative of $f(x|c)$ with respect to x; $\mu_2 = \int u^2 K(u)du$; and $\kappa = \int K^2(u)du$. The bias term—the first term on the right hand side—is of order $O_p(h_n^2)$ and the variance term—the second term in the right hand side—is of order $O_p(1/nh_n^k)$. Define for any $\varepsilon > 0$ the event

$$A_n = \left\{ \left| \widehat{f}(x_0) - f(x_0|C) \right| > M r_n \right\},$$

where $r_n^2 = \max\left\{ h_n^4, \left(nh_n^k\right)^{-1} \right\}$. Then, by Chebychev inequality,

$$\begin{aligned} \Pr(A_n \mid \sigma(C)) &\leq \frac{1}{M^2 r_n^2} E\left[\left(\widehat{f}(x_0) - f(x_0|C) \right)^2 \mid \sigma(C) \right] \\ &= \frac{1}{M^2 r_n^2} MSE\left(\widehat{f}(x_0) \mid C \right) \xrightarrow{p} 0, \end{aligned}$$

as $M \to \infty$. Hence, by the Dominated Convergence Theorem, $\Pr(A_n) = E_C[\Pr(A_n \mid \sigma(C))] \longrightarrow 0$, as $M \to \infty$. By choosing $h_n \propto n^{-\frac{1}{k+4}}$, the desired result follows:

$$\widehat{f}(x_0) - f(x_0|C) = O_p\left(n^{-\frac{2}{4+k}} \right).$$

□

Proof of Proposition 1.3. We apply the Corollary 3.1 of Hall and Heyde [3] (p. 59) using a conditional Liapounov Condition in place of the conditional Lindeberg Condition to obtain the results. We prove (i) first. Define

$$S_n = \frac{\widehat{f}(x_0) - E\left(\widehat{f}(x_0) \mid C\right)}{\left[Var\left(\widehat{f}(x_0) \mid C\right) \right]^{1/2}} = \sum_{1 \leq i \leq n} \frac{w_i - E(w_i \mid C)}{\left[nVar(w_i \mid C) \right]^{1/2}} = \sum_{1 \leq i \leq n} L_{i,n},$$

where $w_i = h_n^{-k} K\left(\frac{X_i - x_0}{h_n}\right)$, and $L_{i,n} = \frac{w_i - E(w_i \mid C)}{[nVar(w_i \mid C)]^{1/2}}$. Note that $Var(w_i \mid C) > 0$ almost surely because $Var(w_i \mid C = c) > 0$ for Q-almost all c.

For $i \geq 1$, let $\mathcal{F}_{i,n}$ denote the σ-field generated by $\sigma(C)$ and $(X_1, ..., X_i)$. Then $\{L_{i,n}, \mathcal{F}_{i,n} : i \geq 1\}$ is a triangular array of Martingale Difference Sequence, because $\{w_i : i \geq 1\}$ are i.i.d. conditional on $\sigma(C)$, and hence $E(w_i \mid \mathcal{F}_{i-1,n}) = E(w_i \mid \sigma(C)) = E(w_i \mid C)$, Q-almost surely. Therefore $E(L_{i,n} \mid \mathcal{F}_{i-1,n}) = 0$, Q-a.s.. Moreover,

$$\begin{aligned} {}_{1 \leq i \leq n} E\left(L_{i,n}^2 \mid \mathcal{F}_{i-1,n} \right) &= n^{-1} \sum_{1 \leq i \leq n} E\left(\left[\frac{w_i - E(w_i \mid C)}{[Var(w_i \mid C)]^{1/2}} \right]^2 \mid \mathcal{F}_{i-1,n} \right) \\ &= n^{-1} \sum_{1 \leq i \leq n} E\left(\left[\frac{w_i - E(w_i \mid C)}{[Var(w_i \mid C)]^{1/2}} \right]^2 \mid \sigma(C) \right) \\ &= E\left(\frac{[w_i - E(w_i \mid C)]^2}{[Var(w_i \mid C)]} \mid \sigma(C) \right) = 1, \end{aligned}$$

where the first equality follows from the definition of $\mathcal{F}_{i-1,n}$ and the third follows from Assumption 1 (and all equalities hold Q-almost surely).

A conditional Liapounov condition holds because

$$
\begin{aligned}
{}_{1\leq i\leq n}E\left(\left|L_{i,n}\right|^{2+\delta}\mid\mathcal{F}_{i-1,n}\right) &= \sum_{1\leq i\leq n}E\left(\left|\frac{w_i-E\left(w_i\mid C\right)}{\left[nVar\left(w_i\mid C\right)\right]^{1/2}}\right|^{2+\delta}\mid\sigma\left(C\right)\right)\\
&= nE\left(\left|\frac{w_i-E\left(w_i\mid C\right)}{\left[nVar\left(w_i\mid C\right)\right]^{1/2}}\right|^{2+\delta}\mid\sigma\left(C\right)\right)\\
&= n\frac{1}{\left[nVar\left(w_i\mid C\right)\right]^{1+\delta/2}}E\left(\left|w_i-E\left(w_i\mid C\right)\right|^{2+\delta}\mid\sigma\left(C\right)\right)\\
&\leq n\frac{1}{\left[nVar\left(w_i\mid C\right)\right]^{1+\delta/2}}2^{1+\delta}E\left(\left|w_i\right|^{2+\delta}\mid\sigma\left(C\right)\right)\\
&= n\frac{2^{1+\delta}}{\left[nVar\left(w_i\mid C\right)\right]^{1+\delta/2}}h_n^{-k(1+\delta)}\\
&\quad\times E\left(h_n^{-k}\left|K\left(\frac{X_i-x_0}{h_n}\right)\right|^{2+\delta}\mid\sigma\left(C\right)\right)\\
&= \frac{2^{1+\delta}}{\left[h_n^kVar\left(w_i\mid C\right)\right]^{1+\delta/2}}\left(nh_n^k\right)^{-\delta/2}\\
&\quad\times E\left(h_n^{-k}\left|K\left(\frac{X_i-x_0}{h_n}\right)\right|^{2+\delta}\mid\sigma\left(C\right)\right),
\end{aligned}
$$

where all the equalities and inequalities hold Q-almost surely. The first equality comes from the definition of $\mathcal{F}_{i-1,n}$, and the second equality from the fact that $\{w_i : i\geq 1\}$ are i.i.d. conditional on $\sigma(C)$. The inequality uses the fact $E\left|a-Ea\right|^{2+\delta}\leq 2^{1+\delta}E\left|a\right|^{2+\delta}$.

Now note that (i)

$$
\begin{aligned}
h_n^kVar\left(w_i\mid C=c\right) &= nh_n^kVar\left(\widehat{f}\left(x_0\right)\mid C=c\right)\\
&\to \left(f\left(x_0|c\right)\int K^2\left(u\right)du\right)>0, \qquad (4)
\end{aligned}
$$

as $n\to\infty$ for Q-almost all c. Hence, $\left[h_n^kVar\left(w_i\mid C\right)\right]^{-1}=O_p(1)$. Also, (ii) for Q-almost all c,

$$
E\left(h_n^{-k}\left|K\left(\frac{X_i-x_0}{h_n}\right)\right|^{2+\delta}\mid C=c\right)\to f\left(x_0|c\right)\int\left|K\left(x\right)\right|^{2+\delta}dx<\infty
$$

as $n\to\infty$, implying $E\left(h_n^{-k}\left|K\left(\frac{X_i-x_0}{h_n}\right)\right|^{2+\delta}\mid\sigma\left(C\right)\right)=O_p(1)$. And, finally, (iii) $\left(nh_n^k\right)^{-\delta/2}\to 0$. By collecting (i)–(iii), we obtain

$$
{}_{1\leq i\leq n}E\left(\left|L_{i,n}\right|^{2+\delta}\mid\mathcal{F}_{i-1,n}\right)\xrightarrow{p}0
$$

Hence all the conditions for the MSD CLT are satisfied. Therefore, we obtain

$$
S_n=\frac{\widehat{f}\left(x_0\right)-E\left(\widehat{f}\left(x_0\right)\mid C\right)}{\left[Var\left(\widehat{f}\left(x_0\right)\mid C\right)\right]^{1/2}}\xrightarrow{d}N\left(0,1\right)
$$

Moreover, we note that

$$\left(nh_n^k\right)^{1/2} \frac{\widehat{f}\left(x_0\right) - E\left(\widehat{f}\left(x_0\right) \mid C\right)}{\left[f\left(x_0|C\right) \int K^2\left(u\right) du\right]^{1/2}} = \frac{\widehat{f}\left(x_0\right) - E\left(\widehat{f}\left(x_0\right) \mid C\right)}{\left[Var\left(\widehat{f}\left(x_0\right) \mid C\right)\right]^{1/2}} \times \left(nh_n^k\right)^{1/2} \frac{\left[Var\left(\widehat{f}\left(x_0\right) \mid C\right)\right]^{1/2}}{\left[f\left(x_0|C\right) \int K^2\left(u\right) du\right]^{1/2}}.$$

And, by Equation (4), for Q-almost all c,

$$\lim_{n\to\infty} \left[\left(nh_n^k\right) Var\left(\widehat{f}\left(x_0\right) \mid C = c\right)\right] = f\left(x_0|c\right) \int K^2\left(u\right) du,$$

implying,

$$\left(nh_n^k\right)^{1/2} \frac{\left[Var\left(\widehat{f}\left(x_0\right) \mid C\right)\right]^{1/2}}{\left[f\left(x_0|C\right) \int K^2\left(u\right) du\right]^{1/2}} \xrightarrow{p} 1,$$

as $n \to \infty$. And so

$$\left(nh_n^k\right)^{1/2} \frac{\widehat{f}\left(x_0\right) - E\left(\widehat{f}\left(x_0\right) \mid C\right)}{\left[f\left(x_0|C\right) \int K^2\left(u\right) du\right]^{1/2}} \xrightarrow{d} N\left(0,1\right).$$

To prove (ii) we eliminate the bias term using $\left(nh_n\right)^{1/2} h_n^2 \to 0$, since, from the usual Taylor expansion (again, conditioned on the event $\{C = c\}$ and using Q-almost sure results to obtain "in-probability" results), the bias term is of order $o_p(h_{nn}^2)$. □

Appendix B.2. Proofs for the Regression Estimator

Define $D_x^j m\left(x,c\right)$ to be the j-th partial derivative of $m\left(x,c\right)$ with respect to x. Define also $\sigma^2\left(x,c\right) = E\left(\varepsilon_i^2 \mid X_i = x, C = c\right)$, and let $E\left[\cdot\mid X = \{x_i\}_{i=1}^n\right]$ denote the conditional expectation given x_i, $i = 1, ..., n$. Let $\mu_2 = \int u^2 K(u) du$ and $\kappa = \int K^2(u) du$. In order to obtain the results of this section, we use Lemmas 2 and 3 presented below.

Lemma 2. *Suppose Assumptions 1 and 2 and Conditions 1–8 hold. Then, for Q-almost all c,*

$$\begin{aligned} E\left(\widehat{g}\left(x\right)|C = c\right) &= m\left(x,c\right) f\left(x|c\right) + \frac{h_n^2}{2}\mu_2 \\ &\quad \times \left[f\left(x|c\right) D_x^2 m\left(x,c\right) + m\left(x,c\right) D_x^2 f\left(x|c\right) + 2D_x m\left(x,c\right) D_x f\left(x|c\right)\right] \\ &\quad + o(h_n^2), \end{aligned}$$

$$Var\left(\widehat{g}\left(x\right)|C = c\right) = \frac{\left(m^2\left(x,c\right) + \sigma^2\left(x,c\right)\right)}{nh_n^k} f\left(x|c\right) \left(\int K^2(u) du\right) + O\left(\frac{1}{n}\right)$$

and

$$Cov\left(\widehat{g}\left(x\right), \widehat{f}\left(x\right)|C = c\right) = \frac{m\left(x,c\right) f\left(x|c\right)}{nh_n^k} \left(\int K^2(u) du\right) + O\left(\frac{1}{n}\right)$$

Proof. The proof is similar to Lemma 3.1 in Pagan and Ullah [2], except that (i) whenever Pagan and Ullah take the expectation $E\left(\cdot\right)$, we take the conditional $E\left(\cdot|C = c\right)$; (ii) whenever they take the conditional $E\left[\cdot\mid X = \{x_i\}_{i=1}^n\right]$, we take $E\left[\cdot\mid X = \{x_i\}_{i=1}^n, C\right]$ (and similarly for $Var(\cdot)$); and (iii) we substitute $m(x)$, $f(x)$ and $\sigma(x)$ in their proof by $m(x,c)$, $f(x|c)$ and $\sigma(x,c)$, respectively. Then, every step in their proof goes through for Q-almost all c and the desired result follows. □

From Lemma 2 above we obtain the approximated bias and variance of $\widehat{m}(x)$.

Lemma 3. *Suppose Assumptions 1 and 2 and Conditions 1–8 hold. Then,*

$$\begin{aligned} E\left(\widehat{m}(x)\,|C\right)-m(x,C) &= \mu_2\frac{\left[f(x|C)\,D_x^2 m(x,C)+2D_x m(x,C)\,D_x f(x|C)\right]}{2f(x|C)}h_n^2 \\ &\quad +O_p\left(\frac{1}{nh_n}\right)+o_p(h_n^2) \end{aligned}$$

and

$$Var\left(\widehat{m}(x)\,|C\right)=\frac{\sigma^2(x,C)}{nh_n^k f(x|C)}\left(\int K^2(u)du\right)+o_p\left(\frac{1}{nh_n^k}\right)$$

Proof. Once more, the proof is similar to Theorem 3.2 in Pagan and Ullah [2]. We first mimic their proof obtaining the approximation bias and variance by substituting the unconditional terms by the proper conditional versions (i.e., conditional on $\{C=c\}$, for Q-almost all c), and obtaining the corresponding expressions for $E\left(\widehat{m}(x)\,|C=c\right)-m(x,c)$ and $Var\left(\widehat{m}(x)\,|C=c\right)$. Since the result holds for Q-almost all c, we can obtain the "in-probability" version of the result for the $\sigma(C)$-measurable random variables $E\left(\widehat{m}(x)\,|C\right)-m(x,C)$ and $Var\left(\widehat{m}(x)\,|C\right)$. □

Next we prove Proposition 2.1–2.3.

Proof of Proposition 2.1. As usual, we write the kernel estimator as

$$\widehat{m}(x)=\frac{\frac{1}{nh_n^k}\sum_{i=1}^n Y_i K\left(\frac{X_i-x}{h_n}\right)}{\frac{1}{nh_n^k}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)}=\frac{\widehat{g}(x)}{\widehat{f}(x)}. \tag{5}$$

Define $g(x,c)\equiv m(x,c)\,f(x|c)$. It is enough to show that $\widehat{g}(x)\to_p g(x,C)$, because

$$\widehat{m}(x)=\frac{\widehat{g}(x)}{\widehat{f}(x)}\xrightarrow{p}\frac{g(x,C)}{f(x|C)}=m(x,C).$$

We adopt an approach similar to the proof of Proposition 1.1. For Q-almost all c,

$$\begin{aligned} E\left(\widehat{g}(x)\,|C=c\right) &= E\left(E\left[\widehat{g}(x)\ |\ X=\{x_i\}_{i=1}^n,C\right]\ |\ C=c\right) \\ &= E\left(E\left[\frac{1}{nh_n^k}\Sigma_{i=1}^n Y_i K\left(\frac{X_i-x}{h_n}\right)\ |\ X=\{x_i\}_{i=1}^n,C\right]\ |\ C=c\right). \end{aligned}$$

Now, note that

$$\begin{aligned} & E\left[\frac{1}{nh_n^k}\Sigma_{i=1}^n Y_i K\left(\frac{X_i-x}{h_n}\right)\ |\ X=\{x_i\}_{i=1}^n,C\right] \\ =\ & \frac{1}{nh_n^k}\Sigma_{i=1}^n E\left[Y_i K\left(\frac{X_i-x}{h_n}\right)\ |\ X=\{x_i\}_{i=1}^n,C\right] \\ =\ & \frac{1}{nh_n^k}\Sigma_{i=1}^n\left[m(x_i,C)\,K\left(\frac{x_i-x}{h_n}\right)\right] \end{aligned}$$

where the second equality follows from the Law of Iterated Expectation. And, so,

$$\begin{aligned} E\left(\widehat{g}\left(x\right)|C=c\right) &= E\left(\frac{1}{nh_n^k}\Sigma_{i=1}^n\left[m\left(x_i,C\right)K\left(\frac{x_i-x}{h_n}\right)\right] \mid C=c\right) \\ &= E\left(\frac{1}{h_n^k}m\left(x_i,C\right)K\left(\frac{x_i-x}{h_n}\right) \mid C=c\right) \\ &= \frac{1}{h_n^k}\int K\left(\frac{x_1-x}{h_n}\right)m\left(x_1,c\right)f\left(x_1|c\right)dx_1 \\ &\to \left[m\left(x,c\right)f\left(x|c\right)\right]\int K\left(\frac{x_1-x}{h_n}\right)dx_1 \\ &= \left[m\left(x,c\right)f\left(x|c\right)\right] \\ &= g\left(x,c\right) \end{aligned}$$

as $n \to \infty$, by Lemma 1. The second equality above comes from Assumption 1 (conditional i.i.d. observations) and the fourth equality from Condition 1(i). As usual, this Q-almost sure convergence implies the convergence in probability of $E\left(\widehat{g}\left(x\right)|C\right)$ to $g\left(x,C\right)$ and, hence,

$$E\left(\widehat{g}\left(x\right) \mid C\right)-g\left(x,C\right)\xrightarrow{p}0.$$

Similarly, from Lemma 2 above, we have for Q-almost all c

$$\begin{aligned} \left(nh_n^k\right)Var\left(\widehat{g}\left(x\right) \mid C=c\right) &= \left[\sigma^2\left(x,c\right)+m^2\left(x,c\right)\right]f\left(x|c\right) \\ &\quad \times\int K^2\left(\frac{x_1-x}{h_n}\right)dx_1+o(h_n^k) \end{aligned}$$

where $\sigma^2\left(x,c\right)=E\left(\varepsilon_i^2 \mid X_i=x,C=c\right)$. Hence, $Var\left(\widehat{g}\left(x\right) \mid C=c\right)\to 0$ as $n\to\infty$, for Q-almost all c, which implies $Var\left(\widehat{g}\left(x\right) \mid C\right)\xrightarrow{p}0$ as $n\to\infty$. Using the Markov inequality conditioned on $\sigma\left(C\right)$ and, then, the Dominated Convergence Theorem in the same way we did in the proof of Proposition 1.1, the desired result follows

$$\widehat{g}\left(x\right)-g\left(x,C\right)\xrightarrow{p}0,\text{ as }n\to\infty.$$

□

Proof of Proposition 2.2. The results follow directly from Lemmas 2 and 3. Similarly to Proposition 1.1, define for any $\varepsilon>0$ the event

$$A_n=\left\{\left|\widehat{m}\left(x\right)-m\left(x,C\right)\right|>Mr_n\right\},$$

where $r_n^2=\max\left\{h_n^4,\left(nh_n^k\right)^{-1}\right\}$. Then, by Chebychev inequality,

$$\begin{aligned} \Pr\left(A_n \mid \sigma(C)\right) &\leq \frac{1}{M^2r_n^2}E\left[\left(\widehat{m}\left(x\right)-m\left(x,C\right)\right)^2 \mid \sigma(C)\right] \\ &= \frac{1}{M^2r_n^2}MSE\left(\widehat{m}\left(x\right) \mid C\right)\xrightarrow{p}0, \end{aligned}$$

as $M\to\infty$. The limit follows from Lemma 3. Hence, by the Dominated Convergence Theorem,

$$\Pr\left(A_n\right)=E_C\left[\Pr\left(A_n \mid \sigma(C)\right)\right]\longrightarrow 0,\quad\text{as }M\to\infty.$$

By choosing $h_n \propto n^{-\frac{1}{k+4}}$, we obtain

$$\widehat{m}(x) - m(x, C) = O_p\left(n^{-\frac{2}{k+4}}\right).$$

□

Proof of Proposition 2.3. Again, we apply the Corollary 3.1 of Hall and Heyde [3] (p. 59) using a conditional Liapounov Condition in place of the conditional Lindeberg Condition. We prove (i) first. Note that

$$\widehat{m}(x) = \frac{1}{\widehat{f}(x)}\left(\frac{1}{nh_n^k}\Sigma_{i=1}^n K\left(\frac{x_i - x}{h_n}\right)[m(x_i, C) + \varepsilon_i]\right)$$

where $m(x_i, C) = E[Y \mid X = x_i, C]$, since $E[\varepsilon_i \mid X, C] = 0$ a.s.. Also,

$$E\left[\widehat{m}(x) \mid X = \{x_i\}_{i=1}^n, C\right] = \frac{1}{\widehat{f}(x)}\left(\frac{1}{nh_n^k}\Sigma_{i=1}^n K\left(\frac{x_i - x}{h_n}\right) m(x_i, C)\right).$$

As a consequence,

$$\begin{aligned}
&\left(nh_n^k\right)^{1/2}\left(\widehat{m}(x) - E\left[\widehat{m}(x) \mid X = \{x_i\}_{i=1}^n, C\right]\right)\\
=\;& \frac{1}{\widehat{f}(x)}\left[\left(nh_n^k\right)^{-1/2}\sum_{1\le i\le n} K\left(\frac{x_i - x}{h_n}\right)\varepsilon_i\right]\\
=\;& \frac{1}{\widehat{f}(x)}\left[\sum_{1\le i\le n} L_{i,n}\right]
\end{aligned}$$

where we define $L_{i,n} = (nh_{nn})^{-1/2} K_i\varepsilon_i$, with $K_i = K\left(\frac{x_i - x}{h_n}\right)$. Because $\widehat{f}(x) - f(x|C) = o_p(1)$, we concentrate on the second term above.

For $i \geq 1$, let $\mathcal{F}_{i,n}$ denote the σ-field generated by $\sigma(C)$ and $(Z_1, ..., Z_i)$, where $Z_i = (Y_i, X_i)$. Then $\{L_{i,n}, \mathcal{F}_{i,n} : i \geq 1\}$ is a triangular array of Martingale Difference Sequence, because

$$\{K_i\varepsilon_i : i \geq 1\} = \left\{K\left(\frac{X_i - x}{h_n}\right)[Y_i - m(X_i, C)] : i \geq 1\right\}$$

are i.i.d. conditional on $\sigma(C)$ by Assumption 1, and hence

$$\begin{aligned}
E\left(K_i\varepsilon_i \mid \mathcal{F}_{i-1,n}\right) &= E\left(K_i\varepsilon_i \mid \sigma(C)\right)\\
&= E\left(K_i E\left[\varepsilon_i \mid X = \{x_i\}_{i=1}^n, C\right] \mid \sigma(C)\right)\\
&= 0.
\end{aligned}$$

Therefore $E\left(L_{i,n} \mid \mathcal{F}_{i-1,n}\right) = 0$. Moreover,

$$\begin{aligned}
{}_{1\le i\le n}E\left(L_{i,n}^2 \mid \mathcal{F}_{i-1,n}\right) &= \sum_{1\le i\le n} E\left(\left[\left(nh_n^k\right)^{-1/2} K_i\varepsilon_i\right]^2 \mid \sigma(C)\right)\\
&= E\left(h_n^{-k}K_i^2\varepsilon_i^2 \mid \sigma(C)\right)\\
&= \frac{1}{h_n^k}E\left(K_i^2 E\left[\varepsilon_i^2 \mid X = \{x_i\}_{i=1}^n, C\right] \mid \sigma(C)\right)\\
&= \frac{1}{h_n^k}E\left(K_i^2\sigma^2(x_i, C) \mid \sigma(C)\right).
\end{aligned}$$

Note that, for Q-almost all c,

$$\begin{aligned} \frac{1}{h_n^k} E\left(K_i^2 \sigma^2\left(x_i, C\right) \mid C=c\right) &= \frac{1}{h_n^k} \int K^2\left(\frac{x_1-x}{h_n}\right) \sigma^2\left(x_1, c\right) f\left(x_1 \mid c\right) d x_1 \\ &\longrightarrow \left[\sigma^2(x, c) f(x \mid c)\right] \int K^2(u)\, d u \end{aligned}$$

as $n \rightarrow \infty$, by Lemma 1. Hence, we have the convergence in probability result

$$_{1 \leq i \leq n} E\left(L_{i, n}^2 \mid \mathcal{F}_{i-1, n}\right) \xrightarrow{p} \left[\sigma^2(x, C) f(x \mid C)\right] \int K^2(u)\, d u .$$

A conditional Liapounov condition holds because

$$\begin{aligned} _{1 \leq i \leq n} E\left(\left|L_{i, n}\right|^{2+\delta} \mid \mathcal{F}_{i-1, n}\right) &= n E\left(\left|L_{i, n}\right|^{2+\delta} \mid \sigma(C)\right) \\ &= n E\left(\left|\left(n h_n^k\right)^{-1 / 2} K_i \varepsilon_i\right|^{2+\delta} \mid \sigma(C)\right) \\ &= \left(n h_n^k\right)^{-\delta / 2} h_n^{-k} E\left(\left|K_i \varepsilon_i\right|^{2+\delta} \mid \sigma(C)\right) \\ &\leq \left(n h_n^k\right)^{-\delta / 2} E\left(\left|\varepsilon_i\right|^{2+\delta} \mid \sigma(C)\right) h_n^{-k} E\left(\left|K_i\right|^{2+\delta} \mid \sigma(C)\right) \end{aligned}$$

where the first equality comes from the definition of $\mathcal{F}_{i-1,n}$, and the fact that $\{L_{i,n} : i \geq 1\}$ are i.i.d. conditional on $\sigma(C)$.

Now note that (i) $\left(n h_n^k\right)^{-\delta/2} \rightarrow 0$; and (ii) $E\left(\left|\varepsilon_i\right|^{2+\delta}\right) < \infty$, implying

$$E\left(\left|\varepsilon_i\right|^{2+\delta} \mid \sigma(C)\right) < \infty, \quad Q\text{-a.s.}.$$

Hence $E\left(\left|\varepsilon_i\right|^{2+\delta} \mid \sigma(C)\right) = O_{as}(1)$. And finally, (iii) for Q-almost all c,

$$E\left(h_n^{-k}\left|K\left(\frac{X_i-x}{h_n}\right)\right|^{2+\delta} \mid C=c\right) \rightarrow f(x \mid c) \int |K(x)|^{2+\delta} d x<\infty$$

as $n \rightarrow \infty$, implying $E\left(h_n^{-k} K^{2+\delta}\left(\frac{X_i-x_0}{h_n}\right) \mid \sigma(C)\right) = O_{as}(1)$. Therefore, by collecting (i)–(iii), we obtain

$$_{1 \leq i \leq n} E\left(\left|L_{i, n}\right|^{2+\delta} \mid \mathcal{F}_{i-1, n}\right) \xrightarrow{as} 0 .$$

Hence all the conditions for the MSD CLT are satisfied. And so,

$$\sum_{1 \leq i \leq n} L_{i, n} \xrightarrow{d} N\left(0,\left(\sigma^2(x, C) f(x \mid C)\right)\left(\int K^2(u)\, d u\right)\right) .$$

Moreover, since $\widehat{f}(x) \xrightarrow{p} f(x \mid C)$, then,

$$\frac{1}{\widehat{f}(x)} \sum_{1 \leq i \leq n} L_{i, n} \xrightarrow{d} N\left(0,\left(\frac{\sigma^2(x, C)}{f(x \mid C)} \int K^2(u)\, d u\right)\right) .$$

To prove (ii) we need to show that if $\left(n h_n^k\right)^{1/2} h_n^2 \rightarrow 0$ as $n \rightarrow \infty$, then

$$\left(n h_n^k\right)^{1 / 2}\left(E\left[\widehat{m}(x) \mid X=\left\{x_i\right\}_{i=1}^n, C\right]-m(x, C)\right) \xrightarrow{p} 0 .$$

First of all, note that

$$
\begin{aligned}
& (nh_{nn})^{1/2} \left(E\left[\widehat{m}(x) \mid X = \{x_i\}_{i=1}^{n}, C \right] - m(x, C) \right) \\
= & \quad \frac{1}{\widehat{f}(x)} \left[\left(nh_n^k \right)^{-1/2} \sum_{1 \le i \le n} K\left(\frac{x_i - x}{h_n} \right) \left[m(x_i, C) - m(x, C) \right] \right] \qquad (6)
\end{aligned}
$$

and, again, because $\widehat{f}(x) - f(x|C) = o_p(1)$, we only need to prove that the second term of (6) is $o_p(1)$. We first note that for Q-almost all c

$$
\begin{aligned}
& E\left[\left(nh_n^k \right)^{-1/2} \sum_{1 \le i \le n} K\left(\frac{x_i - x}{h_n} \right) \left[m(x_i, C) - m(x, C) \right] \mid C = c \right] \\
= & \quad \left(nh_n^k \right)^{1/2} E\left[h_n^{-k} K\left(\frac{x_i - x}{h_n} \right) \left[m(x_i, C) - m(x, C) \right] \mid C = c \right] \\
= & \quad \left(nh_n^k \right)^{1/2} \left[h_n^{-k} \int K\left(\frac{x_1 - x}{h_n} \right) \left[m(x_1, c) - m(x, c) \right] f(x_1|c)\, dx_1 \right].
\end{aligned}
$$

Using a changing-in-variable argument and a Taylor expansion of $[m(x_1, c) - m(x, c)]\, f(x_1|c)$ with respect to x_1 about x we obtain:

$$
\begin{aligned}
& h_n^{-k} \int K\left(\frac{x_1 - x}{h_n} \right) \left[m(x_1, c) - m(x, c) \right] f(x_1|c)\, dx_1 \\
= & \quad \int K(x + h_n u) \left[m(x + h_n u, c) - m(x, c) \right] f(x + h_n u|c)\, du \\
= & \quad \frac{h_n^2}{2} \left(f(x|c)\, D_x^2 m(x, c) + D_x m(x, c)\, D_x f(x|c) \right) \left(\int u^2 K(u)\, du \right) \\
& + o(h_n^2)
\end{aligned}
$$

implying

$$
\begin{aligned}
& \left(nh_n^k \right)^{1/2} \left[h_n^{-k} \int K\left(\frac{x_1 - x}{h_n} \right) \left[m(x_1, c) - m(x, c) \right] f(x_1|c)\, dx_1 \right] \\
= & \quad \left(nh_n^k \right) \left[\frac{h_n^2}{2} \left(f(x|c)\, D_x^2 m(x, c) + D_x m(x, c)\, D_x f(x|c) \right) \left(\int u^2 K(u)\, du \right) + o(h_n^2) \right] \\
= & \quad \left(nh_n^k \right) h_n^2 \frac{\left(f(x|c)\, D_x^2 m(x, c) + D_x m(x, c)\, D_x f(x|c) \right)}{2} \left(\int u^2 K(u)\, du \right) + o\left(\left(nh_n^k \right) h_n^2 \right) \\
\to & \quad 0
\end{aligned}
$$

as $n \to \infty$. Therefore,

$$
E\left[\left(nh_n^k \right)^{-1/2} \sum_{1 \le i \le n} K\left(\frac{x_i - x}{h_n} \right) \left[m(x_i, C) - m(x, C) \right] \mid C \right] \xrightarrow{as} 0.
$$

Next, look at the conditional variance of the term in the brackets in (6)

$$
\begin{aligned}
& Var\left[\left(nh_n^k\right)^{-1/2}\sum_{1\le i\le n}K\left(\frac{x_i-x}{h_n}\right)\left[m\left(x_i,C\right)-m\left(x,C\right)\right]\mid C=c\right] \\
= & \left(nh_n^k\right)^{-1}\sum_{1\le i\le n}E\left[K_i^2\left[m\left(x_i,C\right)-m\left(x,C\right)\right]^2\mid C=c\right] \\
& -\left(nh_n^k\right)^{-1}\left(\sum_{1\le i\le n}E\left[K_i\left[m\left(x_i,C\right)-m\left(x,C\right)\right]\mid C=c\right]\right)^2 \\
= & h_n^{-k}E\left[K_i^2\left[m\left(x_i,C\right)-m\left(x,C\right)\right]^2\mid C=c\right]-nh_n^{-k}\left(E\left[K_i\left[m\left(x_i,C\right)-m\left(x,C\right)\right]\mid C=c\right]\right)^2
\end{aligned}
$$

by Assumption 1 (i.i.d. data conditional on $\sigma(C)$). Now, from Lemma 1 again, the first term above is such that, for Q-almost all c,

$$
h_n^{-k}E\left[K_i^2\left[m\left(x_i,C\right)-m\left(x,C\right)\right]^2\mid C=c\right]\to\left(\left[m\left(x,c\right)-m\left(x,c\right)\right]^2f\left(x|c\right)\right)\int K^2\left(u\right)du=0
$$

as $n\to\infty$, and, from previous argument, the second term is such that, for Q-almost all c,

$$
\begin{aligned}
& nh_n^{-k}\left(E\left[K_i\left[m\left(x_i,C\right)-m\left(x,C\right)\right]\mid C=c\right]\right)^2 \\
= & nh_n^k\left(E\left[h_n^{-k}K_i\left[m\left(x_i,C\right)-m\left(x,C\right)\right]\mid C=c\right]\right)^2 \\
= & nh_n^k\left[h_n^2\frac{\left(f\left(x|c\right)D_x^2m\left(x,c\right)+D_xm\left(x,c\right)D_xf\left(x|c\right)\right)}{2}\left(\int u^2K\left(u\right)du\right)+o(h_n^2)\right]^2 \\
\to & 0
\end{aligned}
$$

as $n\to\infty$. Therefore, for Q-almost all c,

$$
Var\left[\left(nh_n^k\right)^{-1/2}\sum_{1\le i\le n}K\left(\frac{x_i-x}{h_n}\right)\left[m\left(x_i,C\right)-m\left(x,C\right)\right]\mid C=c\right]=o(1)
$$

and, so, both the conditional (on $\{C=c\}$) bias and variance of the term in the brackets in (6) converge to zero Q-almost surely. Markov inequality and the Dominated Convergence Theorem then imply

$$
\left(nh_n^k\right)^{-1/2}\sum_{1\le i\le n}K\left(\frac{x_i-x}{h_n}\right)\left[m\left(x_i,C\right)-m\left(x,C\right)\right]\xrightarrow{p}0.
$$

Therefore,

$$
\left(nh_n^k\right)^{1/2}\left(E\left[\widehat{m}\left(x\right)\mid X=\{x_i\}_{i=1}^n,C\right]-m\left(x,C\right)\right)\xrightarrow{p}0.
$$

And, hence,

$$
\left(nh_n^k\right)^{1/2}\left(\widehat{m}\left(x\right)-m\left(x,C\right)\right)\xrightarrow{d}\left(\frac{\sigma^2\left(x,C\right)}{f\left(x|C\right)}\int K^2\left(u\right)du\right)N\left(0,1\right).
$$

□

Appendix C. The Choice of the Bandwidth

The conditional Approximated Mean Integrated Squared Errors (CAMISE) for the kernel density estimator is:

$$
CAMISE\left(\widehat{f}\mid C\right)=\left[\frac{h_n^4}{4}\mu_2\left(\int\left[D_x^2f\left(x|C\right)\right]^2dx\right)\right]+\left[\frac{\kappa}{nh_n^k}\right].
$$

The unconditional Approximated Mean Integrated Squared Errors (AMISE) is:

$$AMISE\left(\widehat{f}\right) = \left[\frac{h_n^4}{4}\mu_2\left(\int\left[D_x^2 f\left(x|C\right)\right]^2 dxdQ\right)\right] + \left[\frac{\kappa}{nh_n^k}\right].$$

Provided $\left(\int\left[D_x^2 f\left(x|C\right)\right]^2 dxdQ\right) > 0$, the choice of h_n that minimizes $AMISE$ is

$$h_n = \left(\frac{\kappa}{\mu_2\left(\int\left[D_x^2 f\left(x|C\right)\right]^2 dxdQ\right)}\right)^{-\frac{1}{4+k}} n^{-\frac{1}{4+k}}.$$

Such a choice of h_n is unfeasible however. A plug-in estimator would require estimating both $D_x^2 f\left(x|C\right)$ for various C and the distribution of C using a single cross-sectional data. Cross-validation methods also would not estimate h_n by the same reason kernel density does not estimate $f\left(x\right)$ in the presence of common shocks.

On the other hand, the choice of h_n that minimizes the $CAMISE$ is feasible. It is the $\sigma(C)$-measurable random variable

$$h_n\left(C\right) = h\left(C\right) n^{-\frac{1}{4+k}}$$

where

$$h\left(C\right) = \left(\frac{\mu_2\left(\int\left[D_x^2 f\left(x|C\right)\right]^2 dx\right)}{\kappa}\right)^{\frac{1}{4+k}}$$

Provided $D_x^2 f\left(x|c\right)$ is finite for almost all (x,c), we have that $h\left(C\right) = O_p\left(1\right)$ and, so, $h_n\left(C\right) = O_p\left(n^{-\frac{1}{4+k}}\right)$. One should expect that both a plug-in estimator (using, e.g., $D_x^2\widehat{f}$) and a cross-validation method would estimate the random bandwidth $h_n\left(C\right)$. Usual concerns in the literature about how to select the (conditional version of) h_n are present here, but for brevity we do not investigate the topic further.

The same reasoning applies to the kernel regression estimator. But in this case the conditional $AMISE$ is

$$CAMISE\left(\widehat{m}|\,C\right) = \frac{h_n^4}{4}\mu_2\left(\int\phi^2\left(x,C\right)dx\right) + \frac{\kappa}{nh_n^k}\left(\int\frac{\sigma^2\left(x,C\right)}{f\left(x|C\right)}dx\right)$$

where

$$\phi\left(x,C\right) = \frac{f\left(x|C\right)D_x^2 m\left(x,C\right) + 2D_x m\left(x,C\right)D_x f\left(x|C\right)}{f\left(x|C\right)}.$$

The corresponding choice of h_n is

$$h_n\left(C\right) = h\left(C\right) n^{-\frac{1}{k+4}}$$

where

$$h\left(C\right) = \left[\frac{k\mu_2\left(\int\phi^2\left(x,C\right)dx\right)}{4\kappa\left(\int\frac{\sigma^2(x,C)}{f(x|C)}dx\right)}\right]^{\frac{1}{k+4}}.$$

We have that $h\left(C\right) = O_p\left(1\right)$ provided that $f\left(x|c\right), D_x f\left(x|c\right), D_x m\left(x,c\right)$ and $D_x^2 m\left(x,c\right)$ are finite for almost all (x,c); and that both $f\left(x|c\right)$ and $\sigma^2\left(x,c\right)$ are strictly positive for almost all (x,c).

Appendix D. Probabilistic Framework

This section is based on Section 7 of Andrews [1]. Let γ denote some unit in the population and let Γ be the set of all units in the population. Assume Γ is an arbitrary topological space.

For each unit $\gamma \in \Gamma$, $Y(\gamma) \in \mathcal{Y}$ ($\subseteq \mathbb{R}$) is the regression dependent variable, $X(\gamma) \in \mathcal{X}$ ($\subseteq \mathbb{R}^k$) is a vector of regressors, $S(\gamma) \in \mathcal{S}$ ($\subseteq \mathbb{R}^{d_s}$, with $d_s \in \mathbb{N}$) is the vector of supplementary variables that may include other characteristics of the population unit γ, and $C(\gamma) = C \in \mathcal{C}$ is the common shock across units. Define the vector $W(\gamma) = (Y(\gamma), X(\gamma), S(\gamma), C) \in \mathcal{W}$, where $\mathcal{W} \subseteq \mathcal{Y} \times \mathcal{X} \times \mathcal{S} \times \mathcal{C}$.

Throughout this section, let $(\Omega, \mathcal{F}, \mathcal{P})$ denote a probability space and $\omega \in \Omega$. For each unit $\gamma \in \Gamma$, the vector $W(\gamma)$ is a random element defined on the (common) probability space $(\Omega, \mathcal{F}, \mathcal{P})$. In other words, $W(\gamma, \omega)$ is an $\mathcal{F}$-measurable function mapping from $(\Omega, \mathcal{F})$ to $(\mathcal{W}, \mathcal{A})$, where $\mathcal{A}$ is assumed to be the (product) Borel sigma-field. Define $\mathcal{P}_{W(\gamma)}$ as the probability measure defined on $\mathcal{A}$ induced by $\mathcal{P}$ and $W(\gamma, \omega)$. I.e., $\mathcal{P}_{W(\gamma)}$ is the probability distribution of $W(\gamma)$.

Samples are obtained by drawing indices of units $\{\gamma_i : i \geq 1\}$ randomly from Γ according to some conditional distribution $\mathcal{P}_\Gamma$ on Γ. The indices $\{\gamma_i : i \geq 1\}$ are defined on $(\Omega, \mathcal{F}, \mathcal{P})$ the same probability space as $\{W(\gamma) : \gamma \in \Gamma\}$. Following Andrews [1] we assume

Assumption A.1. $\{\gamma_i : i \geq 1\}$ *are i.i.d. indices independent of* $\{W(\gamma) : \gamma \in \Gamma\}$.

Assumption A.1 allows for proportional sampling by taking the distribution $\mathcal{P}_\Gamma$ to be uniform on Γ. But some units can be over-sampled when these distribution is not uniform. Still, the crucial restriction of Assumption A.1 is to not allow the sampling scheme to depend on the characteristics of the unit. In this sense, there is no sample selection in this framework.

Define $W_i = W(\gamma_i)$, and note that $\{W_i : i \geq 1\}$ is a subordinated stochastic process (i.e., subordinated to $\{W(\gamma) : \gamma \in \Gamma\}$ via the directing process $\{\gamma_i : i \geq 1\}$). This process $\{W_i : i \geq 1\}$ is defined on the probability space $(\mathcal{W}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}}, \mathcal{P}^{\mathbb{N}})$, where $\mathcal{W}^{\mathbb{N}}$ is the product space, $\mathcal{A}^{\mathbb{N}}$ is the product Borel sigma-field on $\mathcal{W}^{\mathbb{N}}$ and $\mathcal{P}^{\mathbb{N}}$ is the probability measure on $(\mathcal{W}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$ induced by $\mathcal{P}$, $\{W(\gamma) : \gamma \in \Gamma\}$ and $\{\gamma_i : i \geq 1\}$.

Given the sampling scheme specified in Assumption A.1, the random elements $\{W_i : i = 1, 2, ...\}$ are exchangeable. That is, $\left(W_{\pi(1)}, ..., W_{\pi(n)}\right)$ has the same distribution as $(W_1, ..., W_n)$ for every finite permutation π of $(1, ..., n)$ for all $n \geq 2$. As a consequence, we can apply de Finetti's theorem and conclude that $\{W_i : i \geq 1\}$ is i.i.d. given a sub-sigma field of $\mathcal{A}^{\mathbb{N}}$. But before presenting this well-known result, we need to introduce more notation.

Based on Meyer [4] (Chapter VIII) and Pollard [5], define the following sub-σ-fields on $\mathcal{W}^{\mathbb{N}}$:

i. Denote $\mathcal{J}_n$ the collection of all sets in $\mathcal{A}^{\mathbb{N}}$ whose indicator functions are n-symmetric. Then $\mathcal{J}_n$ is a σ-field, the sequence $\{\mathcal{J}_n : n \geq 1\}$ forms a decreasing filtration on $\mathcal{A}^{\mathbb{N}}$ and $\mathcal{J} = \cap_{n \geq 1} \mathcal{J}_n$ is called the *symmetric σ-field*.
ii. Denote $\mathcal{A}_n = \sigma(W_1, W_2, ..., W_n)$, the σ-field generated by the random elements $(W_1, W_2, ..., W_n)$. Then $\{\mathcal{A}_n : n \geq 1\}$ forms an increasing filtration on $\mathcal{A}^{\mathbb{N}}$ and $\mathcal{A}^{\mathbb{N}} = \cup_{n \geq 1} \mathcal{A}_n$.
iii. Denote $\mathcal{G}_n = \sigma(W_{n+1}, W_{n+2}, ...)$, the σ-field generated by the random elements $(W_{n+1}, W_{n+2}, ...)$. Then $\{\mathcal{G}_n : n \geq 1\}$ forms a decreasing filtration on $\mathcal{A}^{\mathbb{N}}$ and $\mathcal{G} = \cap_{n \geq 1} \mathcal{G}_n$ is called the *tail σ-field*.

The de Finetti's theorem implies the following Lemma (for a proof, see Meyer [4] (Theorem VIII-T5, p. 151), or Pollard [5] (Theorem 52, p.161))

Lemma 4. *Suppose Assumption A.1 holds. Then (i)* $\{W_i : i = 1, 2, ...\}$ *are exchangeable and are i.i.d. conditional on the symmetric sub-σ-field* $\mathcal{J} \subset \mathcal{A}^{\mathbb{N}}$.

This Lemma states that, for all sets $A_i \in \mathcal{A}$, and for every $m \in \mathbb{N}$,

$$\mathcal{P}^{\mathbb{N}}(w_1 \in A_1, ..., w_m \in A_m | \mathcal{J}) = \mathcal{P}^{\mathbb{N}}(w_1 \in A_1 | \mathcal{J}) \times ... \times \mathcal{P}^{\mathbb{N}}(w_m \in A_m | \mathcal{J}) \tag{7}$$

$\mathcal{P}^{\mathbb{N}}$-almost surely, where $\mathcal{P}^{\mathbb{N}}(\cdot | \mathcal{J})$ is the conditional expectation given the sub-sigma-field $\mathcal{J} \subset \mathcal{A}^{\mathbb{N}}$.

We want to obtain Assumption 1 stated in the main text instead, i.e., we want the sequence $\{W_i : i \geq 1\}$ to be i.i.d. conditional on the σ-field $\sigma(C) \subset \mathcal{A}^{\mathbb{N}}$. It is simple to show that this is indeed the case under Assumption A.1 by two steps. First, we note that Lemma 4 also holds for the *tail* sub-σ-field $\mathcal{G} = \cap_{n\geq 1}\mathcal{G}_n$. Then, we note that $\mathcal{G} = \sigma(C)$, because, by definition, $\mathcal{G} = \cap_{n\geq 1}\mathcal{G}_n = \cap_{n\geq 1}\sigma(W_{n+1}, W_{n+2}, ...)$ which equals the sub-sigma-field generated by the common elements of $(W_{n+1}, W_{n+2}, ...)$ for all $n \geq 1$. I.e., $\cap_{n\geq 1}\sigma(W_{n+1}, W_{n+2}, ...) = \sigma(C)$.

The first step is a well-known result and can be obtained by applying Meyer's [4] Lemma VIII-T2 and Theorem VIII-T3 (p. 150).

Theorem 1. *(i) The tail σ-field $\mathcal{G}$ is contained in the symmetric σ-field $\mathcal{J}$.*

(ii) If $A \in \mathcal{J}$, then there exists an element $B \in \mathcal{G}$ such that $A = B$ almost surely.

As a result, despite the fact $\mathcal{G} \subsetneq \mathcal{J}$, we can always find for any $\mathcal{J}$-measurable element, a corresponding $\mathcal{G}$-measurable element that equals the former almost surely. Therefore, Equation (7) holds almost surely when substituting the $\mathcal{J}$-measurable elements by the corresponding $\mathcal{G}$-measurable elements.

Given that (i) Lemma 4 holds for $\mathcal{G}$ in place of $\mathcal{J}$, by Theorem 1, and that (ii) $\sigma(C) = \mathcal{G}$, we obtain the following result:

Claim 1. *Suppose Assumption A.1 holds. Then $\{W_i : i = 1, 2, ...\}$ are exchangeable and are i.i.d. conditional on the σ-field $\sigma(C) \subset \mathcal{A}^{\mathbb{N}}$.*

References

1. Andrews, D.W.K. Cross-section regression with common shocks. *Econometrica* **2005**, *73*, 1551–1585.
2. Pagan, A.R.; Ullah, A. *Nonparametric Econometrics*; Cambridge University Press: New York, NY, USA, 1999.
3. Hall, P.; Heyde, C.C. *Martingale Limit Theory and Its Applications*; Academic Press: New York, NY, USA, 1980.
4. Meyer, P.A. *Probability and Potentials*; Blaisdell Publishing Co.: New York, NY, USA, 1966.
5. Pollard, D. *A User's Guide to Measure Theoretic Probability*; Cambridge University Press: New York, NY, USA, 2002.