



Kent Academic Repository

Fan, Yanqin and Guerre, Emmanuel (2016) *Multivariate local polynomial estimators: uniform boundary properties and asymptotic linear representation*. In: *Essays in Honor of Aman Ullah. Advances in Econometrics* . Emerald, pp. 489-537. ISBN 978-1-78560-787-5.

Downloaded from

<https://kar.kent.ac.uk/76313/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1108/S0731-905320160000036023>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

PART VI
NONPARAMETRIC AND
SEMIPARAMETRIC METHODS

1
3
5
7
9
11

MULTIVARIATE LOCAL POLYNOMIAL ESTIMATORS: UNIFORM BOUNDARY PROPERTIES AND ASYMPTOTIC LINEAR REPRESENTATION

13 Yanqin Fan^a and Emmanuel Guerre^b

15 ^a*Department of Economics, University of Washington, Box 353330, Seattle,
WA 98195, USA*

17 ^b*School of Economics and Finance, Queen Mary, University of London,
Mile End Road, London E1 4NS, UK*

19
21

ABSTRACT

23 *The asymptotic bias and variance of a general class of local polynomial*
25 *estimators of M-regression functions are studied over the whole compact*
27 *support of the multivariate covariate under a minimal assumption on the*
29 *support. The support assumption ensures that the vicinity of the bound-*
31 *ary of the support will be visited by the multivariate covariate. The results*
33 *show that like in the univariate case, multivariate local polynomial esti-*
mators have good bias and variance properties near the boundary. For the
local polynomial regression estimator, we establish its asymptotic normal-
ity near the boundary and the usual optimal uniform convergence rate over
the whole support. For local polynomial quantile regression, we establish a
uniform linearization result which allows us to obtain similar results to the

35
37

Essays in Honor of Aman Ullah

Advances in Econometrics, Volume 36, 3–54

Copyright © 2016 by Emerald Group Publishing Limited

All rights of reproduction in any form reserved

ISSN: 0731-9053/doi:10.1108/S0731-905320160000036023

1 *local polynomial regression. We demonstrate both theoretically and numer-*
3 *ically that with our uniform results, the common practice of trimming local*
5 *polynomial regression or quantile estimators to avoid “the boundary effect”*
7 *is not needed.*

9 **Keywords:** Compact support; boundary effect; pseudo-true value;
11 Newton–Kantorovich Theorem; regression discontinuity design; trimming

13 **JEL classifications:** C12; C14; C21

15 1. INTRODUCTION

17 Recent work illustrate the practical relevance of correcting the boundary bias of
19 kernel estimators. For example, Hickman and Hubbard (2015) demonstrate that
21 standard kernel procedures used in the estimation of auction models may not be
23 able to uncover important features due to boundary bias. In the context of regres-
25 sion discontinuity design (RDD), estimating the parameters of interest involves the
27 estimation of two conditional expectations or conditional quantiles at the disconti-
29 nuity point(s), so suffers from the boundary effect if kernel estimators are used. It
is well known in the literature that theoretically one must deal with either the small
denominator problem or the boundary bias for semiparametric estimators and test
statistics involving averages of kernel estimators over all sample points. One way
to address boundary effects is to use boundary kernels as applied in Hickman and
Hubbard (2015). This solution is simple to implement in the univariate case but
can become burdensome in the multivariate case where it may require estimation
of the support or the support is of complicated form. See, for instance, Müller and
Stadtmüller (1999) or Bouezmarni and Rombouts (2010) for the case of a known
support.

Local polynomial estimators are known to have better boundary properties than
the popular kernel estimators – their bias order is the same for interior and boundary
points. This could alleviate the aforementioned technical and practical problems
with kernel estimators. Work in both RDD and semiparametric estimation and
inference using local polynomial estimators have started to appear, see, for exam-
ple, Altonji, Ichimura, and Otsu (2012), Aryal, Gabrielli, and Vuong (2014), Bravo
and Jacho-Chávez (2011), Hoderlein, Su, White, and Yang (2015), Qu and Yoon
(2014), Su and Ullah (2008), and Su and White (2012) for estimation and infer-
ence in semiparametric and nonparametric models; and Hahn, Todd, and Van der
Klaauw (2001), Porter (2003), Frandsen, Frölich, and Melly (2010), Imbens and
Zajonc (2011), and Oka (2009) for the estimation in RDD. However, these works
are either limited to the univariate case or circumvent potential boundary issues by

1 trimming or focusing interest on an inner subset of the covariate support. Notable
2 exceptions are Banerjee (2007) and Kong, Linton, and Xia (2010), who average
3 over the whole support (unit hypercube) of a multivariate covariate to estimate,
4 respectively, regression average derivatives and an additive quantile specification.
5 More theoretical works in this direction are Ruppert and Wand (1994), Gu, Li, and
6 Yang (2015), and Chen and Wu (2013), who deal with pointwise bias and variance
7 expressions for local polynomial regression estimators.

8 Two important properties of local polynomial estimators are crucial to their suc-
9 cessful applications in Econometrics: a precise characterization of their boundary
10 properties including the bias and variance and a uniform asymptotic linear rep-
11 resentation. For univariate covariate, boundary bias and variance expressions are
12 well known, see Fan and Gijbels (1996). For multivariate covariate, they become
13 complicated. The only general paper that deals with boundary bias and variance for
14 multivariate covariates is Ruppert and Wand (1994). Under the boundary assump-
15 tion (A4) of their paper, Ruppert and Wand (1994) establish expressions of bias and
16 variance of local polynomial regression estimators at the boundary points. Specif-
17 ically, Ruppert and Wand (1994) assumption (A4) considers a boundary point x_{∂}
18 belonging to a convex neighborhood \mathcal{C} such that $\inf_{x \in \mathcal{C}} f(x) > 0$, where $f(\cdot)$ is the
19 probability density function (p.d.f.) of the covariate. Imbens and Zajonc (2011)
20 apply results in Ruppert and Wand (1994) in the context of RDD with multiple
21 forcing variables. The local assumption (A4) of Ruppert and Wand (1994) seems
22 difficult to extend to allow for uniform estimation in the neighborhood of the sup-
23 port boundary. The uniform results of Kong et al. (2010) are specific to hypercube
24 supports which, as the half spaces support considered in Gu et al. (2015), may be
25 too restrictive in practice. We propose instead a uniform version of Chen and Wu
26 (2013) which accounts for general boundaries.

27 As seen from Guerre (2000), what matters for consistent estimation is to have
28 “enough” observations near the estimation location, a condition which involves
29 the geometry of the support in a more subtle way and is flexible enough to cover
30 uniform estimation. The first contribution of this paper is to provide expressions for
31 bias and variance of a general class of multivariate local polynomial estimators of
32 M-regression functions, including both regression and conditional quantile func-
33 tions under weaker support assumptions than (A4) in Ruppert and Wand (1994).
34 Specifically, our support condition does not require the support of the covariate
35 to be connected and allows it to have holes. Under the new support condition, we
36 show that the asymptotic order of the bias and variance of the local polynomial
37 estimators of M-regression functions are not affected by boundary.

38 The weaker support condition is made possible by our novel application of the
39 Newton–Kantorovich Theorem¹ to local polynomial estimation. The new approach
40 is in line with White’s (1982) approach and uses a pseudo-true value to center
41 the local polynomial estimator. Because the pseudo-true value minimizes the

1 expectation of the objective function and satisfies a key centered score condition,
it provides a more natural centering than the partial derivatives of the function to
3 be estimated, as in Fan and Gijbels (1996) or Fan, Heckman, and Wand (1995)
and the vast majority of the local polynomial literature. The Newton–Kantorovich
5 Theorem is then used to study the bias, which is defined here as the difference
between the pseudo-true value and the partial derivatives. The first-order vari-
7 ance of the estimator can be defined using the usual sandwich formula taken at
the pseudo-true value. An interesting finding is that the bias expression may not
9 depend on the estimation method, being for instance identical for regression or
quantile local polynomial estimators.

11 Our results for the bias and variance of multivariate local polynomial estimators
have immediate applications in RDD with multiple forcing variables. First, the
13 result for regression could be used to relax Assumption (A4) in Ruppert and Wand
(1994) adopted in Imbens and Zajonc (2011) in the context of RDD with multiple
15 forcing variables. In addition, our result for the conditional quantile could be used
to extend the estimator of quantile treatment effect in RDD with a univariate forcing
17 variable (see, e.g., Frandsen et al., 2010; Oka, 2009), to allow for multiple forcing
variables.

19 The second contribution of this paper is to establish asymptotic normality and
uniform consistency over the whole support for local polynomial estimators of
21 both regression and conditional quantile functions. A simple consistent estima-
tor of the asymptotic variance is also proposed for both models. This requires to
23 establish a new uniform linearization result for local polynomial quantile regres-
sion which holds over the whole support. In sharp contrast to Corollary 2, ii) in
25 Masry (1996), which is concerned with multivariate covariates whose support is
the entire Euclidean space and is uniformly valid over a compact subset of the sup-
27 port, our results deal directly with multivariate covariates with compact supports
and are uniformly valid over the whole supports.

29 Although the order of bias and variance is not affected by boundary, a quali-
tative conclusion of our results is that the variance may nevertheless significantly
31 increase near the boundary as well documented in the univariate case. The intu-
ition is that the small denominator problem of the Kernel estimation method is not
33 specific and should affect any nonparametric methods. Estimating a function in a
narrow area of the support can only be based on few observations so that a high
35 variance should be effected. Hence, trimming to avoid high estimation variance
may make sense in practice. This issue is investigated through a small simulation
37 experiment which considers testing, additive, and single-index specifications. It
suggests that although the small denominator problem cannot be completely ruled
39 out for a simple unit square support, the impact of trimming is mostly negative
with small potential improvements. This is especially true for testing problems

1 where trimming may decrease the power against some standard alternatives, as
 2 well as against more specific boundary alternative which will be poorly detected.

3 The rest of this paper is organized as follows. The next section introduces the
 4 setup, a general class of local polynomial estimators, and the motivating examples.
 5 Section 3 presents our results for the asymptotic bias and variance for the class
 6 of local polynomial estimators in Section 2. Section 3 establishes asymptotic nor-
 7 mality and uniform convergence of the local polynomial regression over the entire
 8 support. A uniform linearization result for the local polynomial quantile regres-
 9 sion is also derived in Section 4. Section 5 presents results from our simulation
 10 experiment. Section 6 offers concluding remarks and Section 7 collects technical
 11 proofs.

12 To close this section, we introduce some notations that will be used throughout
 13 the rest of this paper. Let $\lfloor s \rfloor$ be the lower integer part of the real number s , that is
 14 the unique integer number such that $\lfloor s \rfloor < s \leq \lfloor s \rfloor + 1$. In what follows, $\|\cdot\|$ stands
 15 for the Euclidean or a vector norm and $\mathcal{V}(x, r) = \{z; \|z - x\| \leq r\}$ is the closed
 16 ball with center x and radius r . When M is a matrix, $\|M\| = \text{Tr}^{1/2}(M'M)$ is the
 17 Frobenius norm of M . For symmetric matrices A and B , $A \geq B$ means that $A - B$
 18 is a positive matrix. The indicator function $\mathbb{I}(X \in A)$ takes value 1 when X lies in
 19 A and 0 otherwise.

21 2. M-REGRESSION, MULTIVARIATE LOCAL 22 POLYNOMIAL ESTIMATION, AND MOTIVATING 23 EXAMPLES

24 This section first introduces a general multivariate M-regression and its local
 25 polynomial estimator. It then reviews several examples including the RDD
 26 with multiple forcing variables in Imbens and Zajonc (2011), consistent model
 27 specification testing, additive models, and average derivative estimation, where
 28 estimation of an M-regression either at the boundary or the entire support is needed.

29 Consider a univariate-dependent variable Y and a d -dimensional covariate X .
 30 We assume that the support of X , denoted as \mathcal{X} , is a compact set with a boundary \mathcal{B} .
 31 Let $\rho(\cdot)$ be a loss function and define the associated M-regression of Y on X as

$$32 \mu(X) = \arg \min_{\mu \in \mathbb{R}} \mathbb{E}[\rho(Y - \mu)|X] \quad (1)$$

33 It is assumed that $\mu(X)$ is the unique minimizer of $\mathbb{E}[\rho(Y - \cdot)|X]$. When
 34 $\rho(t) = t^2$, $\mu(X)$ is the regression function $\mathbb{E}[Y|X]$ while when $\rho(t) = (1 - \alpha)t\mathbb{I}(t \leq$
 35 $0) - \alpha t\mathbb{I}(t > 0)$ for some $\alpha \in (0, 1)$, $\mu(X)$ is the α th quantile of the conditional dis-
 36 tribution of Y given X . The α th expectile of Newey and Powell (1987) corresponds
 37 to $\rho(t) = (1 - \alpha)t^2\mathbb{I}(t \leq 0) + \alpha t^2\mathbb{I}(t > 0)$.

1 The p th-order local polynomial estimator of $\mu(\cdot)$ in Eq. (1) is defined as
 2 follows. Let $U(x)$ be the vector which groups the power $x^\pi = x_1^{\pi_1} \times \dots \times x_d^{\pi_d}$
 3 for all non-negative integer numbers π_1, \dots, π_d with $|\pi| = \pi_1 + \dots + \pi_d \leq p$ accord-
 4 ing to the lexicographic order. Let $K(\cdot)$ be a non-negative kernel function and
 5 $h > 0$ a bandwidth. The local polynomial estimator of $\mu(x) = \beta_0(x)$ and its partial
 6 derivatives

$$7 \quad \mu^{(\pi)}(x) = \frac{\partial^{|\pi|} \mu(x)}{\partial x_1^{\pi_1} \dots \partial x_d^{\pi_d}} = \frac{|\pi|!}{\pi_1! \times \dots \times \pi_d!} \beta_\pi(x) \quad (2)$$

8 are denoted as $\hat{\beta}(x; h) = \left(\hat{\beta}_\pi(x; h), |\pi| \leq p \right)'$ with

$$9 \quad \hat{\beta}(x; h) = \arg \min_{\beta} \sum_{i=1}^n \rho(Y_i - U(X_i - x)' \beta) K\left(\frac{X_i - x}{h}\right) \quad (3)$$

10 where a suitable convention is used to break possible ties when the minimizer
 11 $\hat{\beta}(x; h)$ is not unique as in the case of local polynomial quantile regression.

12 M-regressions are not only of interest in their own right but also play important
 13 roles in other contexts including estimation of average treatment effect parameters,
 14 semiparametric models, and consistent model specification testing. Throughout
 15 this paper, we use the RDD in Example 1 to illustrate the usefulness of the bias and
 16 variance expressions established in Section 3 and Examples 2–4 to demonstrate the
 17 usefulness of the uniform results of the type established in this paper in the context
 18 of consistent model specification testing and semiparametric estimation involving
 19 (weighted) averages of multivariate local polynomial estimators by avoiding fixed
 20 trimming of the boundary commonly adopted in existing work.

21 **Example 1: Regression discontinuity design.** Suppose that the support \mathcal{X}
 22 is partitioned into \mathcal{X}_0 and \mathcal{X}_1 , with boundaries \mathcal{B}_0 and \mathcal{B}_1 . Let $\mathcal{B}_{01} = \mathcal{B}_0 \cap \mathcal{B}_1$
 23 be the frontier between \mathcal{X}_0 and \mathcal{X}_1 . Suppose that individuals with $X = X_1$ in \mathcal{X}_1 receive
 24 a treatment and let Y_1 be the associated response. Denote (X_0, Y_0) in $\mathcal{X}_0 \times \mathbb{R}$
 25 the control group variables, and

$$26 \quad \mu_j(x) = \arg \min_{\mu \in \mathbb{R}} \mathbb{E}[\rho(Y_j - \mu) | X_j = x], \quad x \in \mathcal{X}_j \text{ for } j = 0, 1 \quad (4)$$

27 Extending the regression setup of Imbens and Zajonc (2011), we define the
 28 conditional average treatment effect as

$$29 \quad \tau(x) = \mu_1(x) - \mu_0(x), \quad x \in \mathcal{B}_{01} \quad (5)$$

30 When the treatment has no effect on the conditional parameter $\mu(\cdot)$, $\tau(\cdot) = 0$ and
 31 $\tau(\cdot) \neq 0$ otherwise. A potential difficulty with this multivariate setup is that the

1 conditional average treatment effect is a function, as \mathcal{B}_{01} is in general not a single-
 2 ton. Let $x_k \in \mathcal{B}_{01}$ for $k = 1, \dots, K$, where K could be fixed or grow with the sample
 3 size. Simple summaries of the average treatment effect on the boundary \mathcal{B}_{01} such
 4 as

$$5 \quad \tau_A = \frac{1}{K} \sum_{k=1}^K \tau(x_k), \quad \tau_M = \max_{k=1, \dots, K} \tau(x_k), \quad \tau_m = \min_{k=1, \dots, K} \tau(x_k) \quad (6)$$

6 may be of interest. It is also possible to change the discrete set $\{x_1, \dots, x_K\}$ to
 7 the entire frontier \mathcal{B}_{01} but τ_A should be redefined using an integral instead of a
 8 discrete sum.

11 **Example 2: Significance testing.** Consider the null and alternative hypotheses:

$$13 \quad H_0 : \mathbb{P}(\mu(X) = 0) = 1, \quad H_1 : \mathbb{P}(\mu(X) = 0) < 1$$

15 To avoid boundary issues, trimming is often used to construct test statistics for H_0
 16 versus H_1 . Let \mathcal{X}_c be an inner subset of the support \mathcal{X} of X , which is for instance
 17 obtained by selecting those x in \mathcal{X} at a distance c from the boundary \mathcal{B} , where c
 18 is a trimming parameter. The null and alternative hypotheses are

$$21 \quad H_{0c} : \mathbb{P}(\mu(X) \mathbb{I}[X \in \mathcal{X}_c] = 0) = 1$$

$$22 \quad H_{1c} : \mathbb{P}(\mu(X) \mathbb{I}[X \in \mathcal{X}_c] = 0) < 1$$

23 Note that the alternative $H_{1,0} = H_1$ contains all the alternatives H_{1c} with $c > 0$,
 24 so that trimming may give a test which is not consistent against the alternatives in
 25 H_1 but not in H_{1c} . A possible test statistic is

$$27 \quad \hat{t}_c^2 = \frac{1}{n} \sum_{i=1}^n \hat{\mu}^2(X_i) \mathbb{I}[X_i \in \mathcal{X}_c] \quad (7)$$

31 where $\hat{\mu}(X_i)$ is a local polynomial estimator of $\mu(\cdot)$ in Eq. (1). As is well docu-
 32 mented in the testing literature, such test statistics can also be applied to residuals
 33 to test for more general model specification. The statistic \hat{t}_c^2 is an average version
 34 of the integral test statistic of Härdle and Mammen (1993) for testing specification
 35 of the regression model. A study of the asymptotic behavior of \hat{t}_0^2 for a regression
 36 null hypothesis with a univariate covariate can be found in Li (2005). A similar
 37 test was developed for a linear regression null hypothesis with dependent data by
 38 Hjellvik, Yao, and Tjøstheim (1998). It follows from these authors that for some
 39 Δ_c and σ_c^2 which depend upon the distribution of the observations,

$$41 \quad nh^{d/2} \hat{t}_c^2 - h^{-d/2} \Delta_c \xrightarrow{d} \mathcal{N}(0, \sigma_c^2) \quad (8)$$

1 under the null. The asymptotic mean $h^{-d/2} \Delta_c$ and variance σ_c^2 can be consistently
 2 estimated provided that $h^{-d/2}$ does not diverge too fast, so that Eq. (8) leads
 3 to a rejection region: $nh^{d/2} \hat{\Delta}_c^2 - h^{-d/2} \hat{\Delta}_c \geq \hat{\sigma}_c^2 \Phi(1 - \alpha)$, where $\Phi(1 - \alpha)$ is the
 4 standard normal $(1 - \alpha)$ th quantile.

5 An alternative approach is Fan and Li (1996), who propose a test statistic which
 6 is asymptotically centered, so that $\hat{\Delta}_c$ is not needed. In our context, this would
 7 lead to a test statistic of the form:

$$8 \quad \hat{T}_c = \sum_{i=1}^n Y_i \hat{\mu}_{-i}(X_i) \mathbb{I}[X_i \in \mathcal{X}_c] \quad (9)$$

9 where $\hat{\mu}_{-i}(X_i)$ is a local polynomial leave-one-out estimation of $\mu(X_i)$. But \hat{T}_c is
 10 asymptotically normal under the null $E[Y\mu(X)\mathbb{I}[X \in \mathcal{X}_c]] = 0$, which is equivalent
 11 to H_{0c} in the regression case but not necessarily for alternative choices of $\rho(\cdot)$,
 12 so that Eq. (7) should be preferred for general $\rho(\cdot)$.

13 **Example 3: Additive specification estimation.** A useful dimension reduction
 14 technique to estimate a function $\mu(\cdot)$ depending upon a high-dimensional covariate
 15 is to impose an additive structure on μ . Suppose that $X = (X'_1, X'_2)'$ and that the
 16 function $\mu(\cdot)$ of Eq. (1) has an additive decomposition,
 17

$$18 \quad \mu(X) = m_1(X_1) + m_2(X_2)$$

19 A popular method for estimating $m_1(\cdot)$ is the marginal integration method of Linton
 20 and Nielsen (1995). Let \mathcal{X}_{2c} be an inner subset of the support \mathcal{X}_2 of X_2 . An estimator
 21 of
 22

$$23 \quad \mu_{1,c}(x_1) = m_1(x_1) + \mathbb{E}[m_2(X_2)\mathbb{I}(X_2 \in \mathcal{X}_{2c})]$$

24 is

$$25 \quad \hat{\mu}_{1,c}(x_1) = \frac{\sum_{i=1}^n \mathbb{I}(X_{2i} \in \mathcal{X}_{2c}) \hat{\mu}_{-i}(x_1, X_{2i})}{\sum_{i=1}^n \mathbb{I}(X_{2i} \in \mathcal{X}_{2c})} \quad (10)$$

26 In a regression setup, Linton and Nielsen (1995) consider an integral version of
 27 Eq. (10) where $\hat{\mu}_{-i}(\cdot, \cdot) = \hat{\mu}(\cdot, \cdot)$ is a standard kernel regression estimator. The role
 28 of the trimming set \mathcal{X}_{2c} is to avoid boundary effects. In a quantile setup and for
 29 the support $\mathcal{X}_2 = [0, 1]^{d_2}$, Kong et al. (2010) show that trimming is not needed
 30 when using local polynomial estimation. In both papers, the convergence rate of
 31 the additive component is shown to be faster than the ones usually obtained for
 32 the estimation of an unrestricted function $\mu(\cdot)$.
 33
 34
 35
 36
 37
 38
 39
 40
 41

Example 4: Average derivative estimation. Another popular dimension reduction approach is based on the single-index specification

$$\mu(X) = g(X'\beta)$$

where β is a d -dimensional vector and $g(\cdot)$ a real-valued function. The slope parameter β can be identified, up to a scaling coefficient, using the average derivatives of $\mu(X)$ since

$$\mathbb{E} \left[\frac{\partial \mu(X)}{\partial X} \right] = \beta \mathbb{E} [g^{(1)}(X'\beta)]$$

where $g^{(1)}(\cdot)$ is the derivative of $g(\cdot)$. See Härdle and Stoker (1989) or Powell, Stock, and Stoker (1989) for the regression case and Chaudhuri, Doksum, and Samarov (1997) for quantile functions. As seen from Altonji et al. (2012), average derivatives are also of independent interest in microeconomic issues. Direct methods average an estimator of $\partial \mu(X)/\partial X$ over \mathcal{X}_c as in

$$\widehat{M}_c^{(1)} = \frac{1}{\sum_{i=1}^n \mathbb{I}[X_i \in \mathcal{X}_c]} \sum_{i=1}^n \mathbb{I}[X_i \in \mathcal{X}_c] \frac{\widehat{\partial \mu}}{\partial X}(X_i) \quad (11)$$

Chaudhuri, Doksum, and Samarov (1997) consider local polynomial estimators of $\partial \mu(X)/\partial X$ in a quantile setup while Li, Lu, and Ullah (2003), Banerjee (2007), and Altonji et al. (2012) implement local polynomial for regressions. Chaudhuri et al. (1997) use a weighting function instead of trimming and Altonji et al. (2012) consider $\mathcal{X}_c = [0.5, 3.5]$ for a support $\mathcal{X} = [0, 4]$.

3. THE BIAS AND VARIANCE OF MULTIVARIATE LOCAL POLYNOMIAL ESTIMATORS

In this section, we establish results for the asymptotic bias and variance of multivariate local polynomial estimators of an M-regression that are valid uniformly over \mathcal{X} . First, we introduce the main assumptions including our new support condition and contrast it with Assumption (A4) in Ruppert and Wand (1994), who use it to develop pointwise results.

3.1. Main Assumptions

A key issue that we have to deal with is that $\hat{\beta}(x; h)$ is an estimator of the pseudo-true value $\bar{\beta}(x; h)$ defined as

$$\bar{\beta}(x; h) = \arg \min_{\beta} \mathbb{E} \left[\rho(Y - U(X - x)' \beta) K \left(\frac{X - x}{h} \right) \right] \quad (12)$$

and that $\bar{\beta}(x; h)$ may differ from $\beta(x)$, the vector with entries

$$\beta_{\pi}^*(x) = \begin{cases} 0 & \text{if the partial derivative } \mu^{(\pi)}(x) \text{ does not exist} \\ \beta_{\pi}(x) \text{ defined in Eq. (2)} & \text{otherwise} \end{cases}$$

The first goal of this paper is to study the bias term $[\bar{\beta}(x; h) - \beta(x)]$ over the support \mathcal{X} of X . This will be done under the assumptions introduced below.

Assumption R. (*Loss function $\rho(\cdot)$*).

(i) For each $x \in \mathcal{X}$, $\mu \in \mathbb{R} \mapsto \mathbb{E}[\rho(Y - \mu) | X = x] = R(\mu | x)$ is twice continuously differentiable with respect to μ . The second order derivative $R^{(2)}(\cdot | \cdot)$ satisfies, for some $\epsilon > 0$,

$$\inf_{x \in \mathcal{X}} \inf_{\mu \in [\mu(x) - \epsilon, \mu(x) + \epsilon]} R^{(2)}(\mu | x) > 0, \sup_{x \in \mathcal{X}} |R^{(2)}(\mu(x) | x)| < \infty, \text{ and} \\ |R^{(2)}(\mu | x) - R^{(2)}(\mu' | x)| < C |\mu - \mu'|$$

for all real numbers μ and μ' in $[\mu(x) - \epsilon, \mu(x) + \epsilon]$. (ii) $\rho(\cdot)$ is continuous and there is a finite collection of intervals (a_j, a_{j+1}) with $\bigcup_{j=0}^J (a_j, a_{j+1})$ such that $\rho(\cdot)$ is continuously differentiable with derivative $\rho^{(1)}(\cdot)$ over each (a_j, a_{j+1}) . Moreover,

$$\sup_{x \in \mathcal{X}} \mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 | X = x \right] < \infty$$

The functions $(\mu, x) \in \mathbb{R} \times \mathcal{X} \mapsto R^{(2)}(\mu | x)$ and $\mathbb{E} \left[(\rho^{(1)}(Y - \mu))^2 | X = x \right]$ are continuous.

Assumption S. (*Smoothness of $\mu(\cdot)$*).

There is a smoothness index $s > 0$ such that either S1 or S2 below holds:

S1: $s > 1$, $\mu(\cdot)$ is $\lfloor s \rfloor$ times differentiable and for some $L > 0$, the partial derivatives of order π with $|\pi| = \lfloor s \rfloor$ satisfy: for all $x, x' \in \mathcal{X}$,

$$|\mu^{(\pi)}(x') - \mu^{(\pi)}(x)| \leq L \|x' - x\|^{s - \lfloor s \rfloor}$$

1 **S2:** s is an integer and $\mu(\cdot)$ is s times continuously differentiable.

3 **Assumption X.** (\mathcal{X} -boundary).

5 (i) The marginal probability density function $f(\cdot)$ of the d -dimensional X is continuously differentiable and bounded away from 0 on its support \mathcal{X} . (ii) There are some κ_0, κ_1 in $(0, 1]$ such that for any $x \in \mathcal{X}$ and all ϵ in $(0, \kappa_0]$, there is a $x' \in \mathcal{X}$ satisfying

$$7 \quad \mathcal{V}(x', \kappa_1 \epsilon) \subset \mathcal{V}(x, \epsilon) \cap \mathcal{X} \quad (13)$$

9 (iii) \mathcal{X} is a compact subset of \mathbb{R}^d .

11 **Assumption K.** (Kernel function).

13 The kernel function $K(\cdot)$ is non-negative and Lipschitz, that is, $|K(x) - K(x')| \leq L \|x - x'\|$ for any $x, x' \in \mathbb{R}^d$. The kernel function $K(\cdot)$ has a compact support and is bounded away from 0 over the unit ball $\mathcal{V}(0, 1)$. The bandwidth $h = h_n \rightarrow 0$ as $n \rightarrow \infty$.

17 A brief discussion of the assumptions is in order. Assumption R-(i) is important to ensure that $\mu(x)$ in Eq. (1) and the pseudo-true value $\bar{\beta}(x; h)$ in Eq. (12) are unique as implied by $R^{(2)}(\cdot | x) > 0$ and Assumption K. Indeed this implies that the Hessian matrix

$$23 \quad \frac{\partial}{\partial \beta \partial \beta'} \mathbb{E} \left[\rho(Y - U(X - x)\beta) K\left(\frac{X - x}{h}\right) \right]$$

25 is strictly positive and remains so when $h \rightarrow 0$. The additional Lipschitz condition on $R^{(2)}(\cdot | x)$ is used to study $\bar{\beta}(x; h)$ when $h \rightarrow 0$. Assumption R-(i) clearly holds when $\rho(t) = t^2$. In the quantile case with α in $(0, 1)$, $\rho(t) = (1 - \alpha)t\mathbb{I}(t \leq 0) - \alpha t\mathbb{I}(t > 0)$ and $R^{(2)}(\mu | x) = f(\mu | x)$, where $f(\cdot | x)$ denotes the conditional pdf of Y given $X = x$, so that Assumption R-(i) holds when $\inf_{x \in \mathcal{X}} f(\mu(x) | x) > 0$ as standard in quantile estimation and if $\mu \mapsto f(\mu | x)$ is differentiable or Lipschitz in the vicinity of $\mu(x)$. Assumption R-(ii) is used to study the variance of the local polynomial estimator.

33 Assumption S describes some smoothness conditions for $\mu(\cdot)$. Assumption S1 is from Chaudhuri (1991) or Masry (1996). It allows for a noninteger smoothness index s as necessary when considering some power functions like $|x|^s$ whose derivatives of order $\lfloor s \rfloor$ can have a singular behavior in the vicinity of the origin. Assumption S2 is slightly stronger and implies Assumption S1 when the support \mathcal{X} of the covariate is compact. Note that it is not assumed that $p < s$ as it is also clear from the definition of the pseudo-true value $\beta_\pi^*(x)$ which allows for $p \geq s$. Indeed, nonexistence of the partial derivative $\mu^{(\pi)}(x)$ in the definition of $\beta_\pi^*(x)$

1 should be understood as $|\pi| > \lfloor s \rfloor$ under Assumption S1 and as $|\pi| > p$ under
 Assumption S2. Assumption K is rather standard in local polynomial estimation.

3 Assumption X is our key support condition. Assumption X-(i, iii) are standard
 but Assumption X-(ii) seems to be new to the best of our knowledge. Assumption
 5 X-(ii) holds for hypercubes $[a, b]^d$ or hyperrectangles, spheres, or supports delimit-
 7 ed by smooth boundaries but also for more irregular support shapes. It relaxes
 the usual intuition of a connected support with no hole and delimited by a smooth
 9 boundary. In particular, the support \mathcal{X} of the covariate X does not need to be
 connected and can have holes. The latter possibility extends Assumption (A4) in
 Ruppert and Wand (1994), who assume that there is a nontrivial convex set $\mathcal{C} \subset \mathcal{X}$
 11 with nonempty interior containing x . This restricts the shape of \mathcal{X} since there can-
 not be a sequence of holes with vanishing size converging to x . In sharp contrast,
 13 under Assumption X-(ii), the boundary of \mathcal{X} can be very irregular in the vicinity
 of x with many peaks, as illustrated by an example below. In addition, Assump-
 15 tion X-(ii) is suitable for uniform or global studies of local polynomial estimation,
 while Ruppert and Wand (1994) only consider pointwise estimation of a regression
 17 function $\mu(x)$ for a given x .

The intuition behind Assumption X-(ii) is that a local polynomial estimator
 19 performs well provided that there are many observations close to x , say up to
 a distance $h \rightarrow 0$ given by a bandwidth, to estimate $\mu(x)$. This will hold if it is
 21 possible to find a sequence of balls in the support \mathcal{X} with a radius proportional
 to h which will converge to x . The key point here is that these balls do not need
 23 to be centered at x , which would be impossible when x is on the boundary of \mathcal{X} ,
 but can be centered at an interior point $x' \neq x$ of \mathcal{X} . This is the intuitive content of
 25 Eq. (13), where ϵ plays the role of a bandwidth. For $\epsilon > 0$ small enough, Eq. (13)
 means that there is a ball $\mathcal{V}(x', \kappa_1 \epsilon)$ in the support \mathcal{X} which is also in the vicinity
 27 set $\mathcal{V}(x, \epsilon)$ of x . The fact that the constants κ_0 and κ_1 in Eq. (13) do not depend
 upon x is essential to establish the uniformity results in this paper.

29 To compare and contrast with Assumption (A4) in Ruppert and Wand (1994),
 consider the case $d=2$ for the sake of brevity. Like Assumption (A4) in
 31 Ruppert and Wand (1994), Assumption X-(ii) also excludes the support defined
 below:

$$33 \quad \mathcal{X}_b = \{x = (x_1, x_2) : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq x_1^2\}$$

35 due to the origin $o = (0, 0)'$ and the fast decrease of x_2 when $x_1 \rightarrow 0$. Indeed balls
 in $\mathcal{V}(o, \epsilon) \cap \mathcal{X}_b$ have a radius which is of order ϵ^2 and therefore not compatible
 37 with Eq. (13). On the other hand, a support like

$$39 \quad \mathcal{X}_c = \{x = (x_1, x_2) : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq b(x_1)\}, \quad \text{with } b(x_1) \geq \underline{b}(x_1),$$

$$41 \quad \text{where } \underline{b}(x_1) = C \sum_{k=1}^{\infty} |x_1|^k \mathbb{I} \left(x_1 \in \left[\frac{1}{k+1/2}, \frac{1}{k} \right] \right) \quad \text{for some } C > 0$$

1 will satisfy Assumption X-(ii) due to the restriction $b(x_1) \geq \underline{b}(x_1)$ with $\underline{b}(x_1) =$
 2 $C|x_1|$ on $[(k + 1/2)^{-1}, k^{-1}]$. Indeed, it is sufficient to consider, for all $k \geq 1$,
 3 those x' with $x'_1 \in [(k + 1/2)^{-1}, k^{-1}]$ and $x'_2 \in [0, Ck^{-1}]$ and balls $\mathcal{V}(x', \epsilon)$ in the
 4 rectangle $[(k + 1/2)^{-1}, k^{-1}] \times [0, Ck^{-1}]$ to show that Eq. (13) holds. On the other
 5 hand, when $b(x_1) = \underline{b}(x_1)$, \mathcal{X}_c does not satisfy Assumption (A4) in Ruppert and
 6 Wand (1994), due to the irregular behavior of $\underline{b}(x_1)$ when $x_1 \rightarrow 0$. Indeed since
 7 $\underline{b}(x_1) = 0$ on all $((k + 1)^{-1}, (k + 1/2)^{-1})$, there is no nontrivial convex set in \mathcal{X}_c
 8 which contains the origin o .

3.2. The Bias

11

Under the assumptions introduced in Section 3.1, it is possible to obtain the
 12 orders of the bias terms: $[\tilde{\beta}_\pi(x, h) - \beta_\pi(x)]$ uniformly over $x \in \mathcal{X}$, including the
 13 boundary of \mathcal{X} .

15

Theorem 1. *Suppose Assumptions K, R, S1, and X hold with $p \geq \lfloor s \rfloor$. Then for*
 16 *all $\pi \in \mathbb{N}^d$ with $|\pi| \leq \lfloor s \rfloor$ and h small enough,*

17

$$\sup_{x \in \mathcal{X}} |\tilde{\beta}_\pi(x, h) - \beta_\pi(x)| \leq CLh^{s-|\pi|}$$

19

20 Theorem 1 extends existing bias results for x in inner subsets of \mathcal{X} . See for instance
 21 Chaudhuri (1991) and Guerre and Sabbah (2012) for the quantile case. Theorem
 22 1 therefore shows that, thanks to Assumption X, the order of the bias of the local
 23 polynomial estimator is not affected by boundary. This contrasts with Nadaraya–
 24 Watson Kernel estimators and is a key reason for preferring local polynomial
 25 methods as argued by Fan and Gijbels (1996) for univariate local polynomial
 26 regression. The proof of Theorem 1 works by checking that the tentative limit
 27 $\tilde{\beta}_\pi(x)$ approximately satisfies the first-order condition of the minimization (12).
 28 Since Assumption R ensures that the Hessian of the objective function of Eq. (12) is
 29 full-rank, the Newton–Kantorovich Theorem is then used to show that $\tilde{\beta}_\pi(x, h) -$
 30 $\beta_\pi(x) = O(Lh^{s-|\pi|})$ uniformly over \mathcal{X} as stated in the theorem. As far as we know,
 31 this approach is new in the context of local polynomial estimation.

33

Theorem 2 gives a uniform expansion of $[\tilde{\beta}_\pi(x, h) - \beta_\pi(x)]$ under the stronger
 34 smoothness Assumption S2 which allows for a better description of the bias.
 35 Theorem 2 also completes Theorem 1 by considering the case where
 36 $p \leq s - 1 = \lfloor s \rfloor$. Define

37

$$\Omega_1(x, h) = \int \mathbb{I}(x + hz \in \mathcal{X}) U(z)U(z)' K(z) dz \quad (14)$$

39

41

$$\text{Sp}(\Omega_1^{-1}(x, h)) = \text{the largest eigenvalue of } \Omega_1^{-1}(x, h)$$

$$b_{p+1}(x, h) = \sum_{\pi \in \mathbb{N}^d: |\pi|=p+1} \kappa_\pi(x, h) \beta_\pi(x), \quad \text{and} \quad (15)$$

$$\kappa_\pi(x, h) = \int \mathbb{I}(x + hz \in \mathcal{X}) z^\pi U(z) K(z) dz$$

Let e_π be the π th element of the canonical basis, that is, the vector with a 1 in the π th lexicographic position and 0 elsewhere.

Theorem 2. *Suppose Assumptions K, R, S2, and X hold and that p satisfies $p \leq \lfloor s \rfloor = s - 1$. Then for all $\pi \in \mathbb{N}^d$ with $|\pi| \leq p$,*

$$\bar{\beta}_\pi(x, h) = \beta_\pi(x) + h^{p+1-|\pi|} e'_\pi \Omega_1(x, h)^{-1} b_{p+1}(x, h) + o(h^{p+1-|\pi|})$$

uniformly in $x \in \mathcal{X}$ with $\max_{x \in \mathcal{X}} \text{Sp}(\Omega_1(x, h)^{-1}) < \infty$ and $\max_{x \in \mathcal{X}} \|b_{p+1}(x, h)\| < \infty$ for h small enough.

A first noticeable fact is that the leading term in the bias expansion in Theorem 2 is independent of the loss function $\rho(\cdot)$. The term $e'_\pi \Omega_1(x, h)^{-1} b_{p+1}(x, h)$ multiplying $h^{p+1-|\pi|}$ is identical to the one obtained for regression local polynomial estimators which has been already studied by Ruppert and Wand (1994) and more recently by Gu et al. (2015). Second, the bias boundary effect is captured through the matrix $\Omega_1(x, h)$ and the vector $\kappa_\pi(x, h)$. When x is an inner point of \mathcal{X} , or, more precisely, when $x + hz$ lies in \mathcal{X} for all z in the support of the kernel function, $\Omega_1(x, h) = \Omega_1$ and $\kappa_\pi(x, h) = \kappa_\pi$ with

$$\Omega_1 = \int U(z) U(z)' K(z) dz \quad \text{and} \quad \kappa_\pi = \int z^\pi U(z) K(z) dz \quad (16)$$

in which case $b_{p+1}(x, h) = b_{p+1}(x)$ with,

$$b_{p+1}(x) = \sum_{\pi \in \mathbb{N}^d: |\pi|=p+1} \kappa_\pi \beta_\pi(x)$$

An important issue is whether the term $e'_\pi \Omega_1(x, h)^{-1} b_{p+1}(x, h)$ in front of $h^{p+1-|\pi|}$ vanishes or not. This has been recently discussed for the regression case and a symmetric kernel $K(\cdot)$ by Gu et al. (2015) and, as noted above, their results can also be applied to a more general $\rho(\cdot)$. When $p + 1 - |\pi|$ is odd, the exact order $h^{p+1-|\pi|}$ holds over the whole support except for those unlikely x such that $b_{p+1}(x, h) = 0$. In this case, Theorem 2 can be used in conjunction with Theorem 4 below to propose an optimal bandwidth for inner x . The situation differs when $p + 1 - |\pi|$ is even. When $K(\cdot)$ is symmetric, $\int z^{\pi_1} z^{\pi_2} K(z) dz = 0$

1 when $|\pi_1|$ is even and $|\pi_2|$ odd, so that reorganizing the entries of $U(z)$ gives a
 2 block-diagonal Ω . As shown by Gu et al. (2015), this implies that $e'_\pi \Omega_1^{-1} b(x) = 0$
 3 when $p + 1 - |\pi|$ is even. Because $\Omega_1(x, h)$ is not similar to a block-diagonal
 4 matrix when x lies on the boundary \mathcal{B} , it is unlikely to have $e'_\pi \Omega_1(x, h)^{-1} b(x, h) = 0$
 5 unless the partial derivatives $\mu^{(\pi)}(x)$ take some very specific values. Hence for
 6 inner x , the bias has a smaller order $o(h^{p+1-|\pi|})$ while for x on the boundary the
 7 bias will be $O(h^{p+1-|\pi|})$. This shows that the bias can be slightly larger near the
 8 boundary. Under some additional regularity conditions, it is possible to show that
 9 the bias is $O(h^{p+2-|\pi|})$ for inner x .

Let

$$11 \quad \mathbf{R}_{f,\mu}^{(2)}(x) = R^{(2)}(\mu(x)|x)f(x), \quad \kappa_{\pi,j} = \int z_j z^\pi U(z)K(z)dz \quad \text{and}$$

$$13 \quad \mathbf{b}_{p+1}(x) = \sum_{j=1}^d \sum_{\pi \in \mathbb{N}^d: |\pi|=p+1} \kappa_{\pi,j} \frac{1}{\mathbf{R}_{f,\mu}^{(2)}(x)} \frac{\partial \mathbf{R}_{f,\mu}^{(2)}(x)}{\partial x_j} \beta_\pi(x)$$

15
 16 **Proposition 3.** *Suppose Assumptions K, R, S2, and X hold and that p satisfies*
 17 *$p \leq \lfloor s \rfloor = s - 1$. Assume in addition that $K(\cdot)$ is symmetric and that $\mathbf{R}_{f,\mu}^{(2)}(x)$,*
 18 *$\mu^{(p+1)}(x)$ are continuously differentiable over \mathcal{X} . Then if $p + 1 - |\pi|$ is even*
 19 *and if x is in the interior of \mathcal{X} ,*

AQ:9

$$21 \quad \bar{\beta}_\pi(x, h) = \beta_\pi(x) + h^{p+2-|\pi|} e'_\pi \Omega_1^{-1} (b_{p+2}(x) + \mathbf{b}_{p+1}(x)) + o(h^{p+2-|\pi|})$$

23
 24 Proposition 3 suggests that for an even $p + 1 - |\pi|$, the bias grows from
 25 $O(h^{p+2-|\pi|})$ for those x at a distance $O(h)$ to the higher order $O(h^{p+1-|\pi|})$ when
 26 x is a boundary point. As in Gu et al. (2015), Ruppert and Wand (1994) or when
 27 $d = 1$, it is possible to study the bias for a sequence $x_h = x + hc$ in the interior of
 28 \mathcal{X} and x in \mathcal{B} under additional simplifying assumptions on the boundary shape.
 29 Note also that the higher-order expansion of Proposition 3 now depends on the
 30 choice of $\rho(\cdot)$ through the partial derivatives of $\mathbf{R}_{f,\mu}^{(2)}(x)$ in the term $\mathbf{b}_{p+1}(x)$ of the
 31 expansion.

3.3. The Variance

33
 34
 35 Consider now the variance of the local polynomial estimator. As is well known,
 36 the partial derivative estimators $\hat{\beta}_\pi(x; h)$ converge with a different rate and should
 37 be first standardized with the diagonal matrix

$$41 \quad H = \text{Diag}(h^{|\pi|}, |\pi| \leq p)$$

1 Standard arguments as in Huber (1967) then suggest that the variance of
 2 $H\left(\hat{\beta}(x, h) - \bar{\beta}(x, h)\right)$ is close to $\bar{V}(x, h)/nh^d$, where $\bar{V}(x, h)$ has the usual
 3 sandwich form

$$5 \quad \bar{V}(x, h) = \bar{\mathbf{R}}^{(2)}(x, h)^{-1} \bar{\mathbf{S}}(x, h) \bar{\mathbf{R}}^{(2)}(x, h)^{-1}$$

7
 8 with

$$11 \quad \bar{\mathbf{R}}^{(2)}(x, h) = h^{-d} \mathbb{E} \left[\mathbf{R}^{(2)}(U(X-x)' \bar{\beta}(x, h) | X) U \left(\frac{X-x}{h} \right) \right. \\
 12 \quad \left. \times U \left(\frac{X-x}{h} \right)' K \left(\frac{X-x}{h} \right) \right] \\
 13 \\
 14 \quad \bar{\mathbf{S}}(x, h) = h^{-d} \mathbb{E} \left[(\rho^{(1)}(Y - U(X-x)' \bar{\beta}(x, h)))^2 U \left(\frac{X-x}{h} \right) \right. \\
 15 \quad \left. \times U \left(\frac{X-x}{h} \right)' K^2 \left(\frac{X-x}{h} \right) \right]$$

21 where the superscript “-” indicates dependence on $\bar{\beta}(x, h)$. Since Assumptions
 22 R-(i) and X-(i) give, for $\Omega_1(x, h)$ in Eq. (14),

$$25 \quad \bar{\mathbf{R}}^{(2)}(x, h) \geq Ch^{-d} \mathbb{E} \left[U \left(\frac{X-x}{h} \right) U \left(\frac{X-x}{h} \right)' K \left(\frac{X-x}{h} \right) \right] \\
 26 \\
 27 \quad = C \int U(z) U(z)' K(z) f(x + hz) dz \geq C \Omega_1(x, h)$$

31 and because $\max_{(x, h) \in \mathcal{X} \times [0, \infty)} \text{Sp}(\Omega_1(x, h)^{-1}) < \infty$ as stated in Theorem 2,
 32 $\bar{\mathbf{R}}^{(2)}(x, h)$ has an inverse for all $x \in \mathcal{X}$ and $h \geq 0$, so that $\bar{V}(x, h)$ is well defined.

33 Consider the following approximations for $\bar{V}(x, h)$:

$$35 \quad V(x, h) = \frac{\mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 | X = x \right]}{(R^{(2)}(\mu(x)|x))^2 f(x)} \Omega(x, h)^{-1} \quad \text{and} \\
 36 \\
 37 \\
 38 \quad V(x) = \frac{\mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 | X = x \right]}{(R^{(2)}(\mu(x)|x))^2 f(x)} \Omega^{-1}$$

41

1 where

$$\begin{aligned}
 3 \quad \Omega_2(x, h) &= \int \mathbb{I}(x + hz \in \mathcal{X}) U(z)U(z)'K^2(z)dz \\
 5 \quad \Omega(x, h)^{-1} &= \Omega_1(x, h)^{-1}\Omega_2(x, h)\Omega_1(x, h)^{-1} \\
 7 \quad \Omega_2 &= \int U(z)U(z)'K^2(z)dz, \quad \Omega^{-1} = \Omega_1^{-1}\Omega_2\Omega_1^{-1}
 \end{aligned}$$

9 and $\Omega_1(x, h)$ and Ω_1 are as in Eqs. (14) and (16), respectively. The next theorem shows that $V(x, h)$ is a suitable approximation for $\bar{V}(x, h)$ over \mathcal{X} , while $V(x)$ is a suitable approximation for $\bar{V}(x, h)$ over interior subsets of \mathcal{X} .

13 **Theorem 4.** *Suppose Assumptions K, R, X hold together with Assumptions S1 or S2. Then $\max_{x \in \mathcal{X}} \text{Sp}(V(x, h)) = O(1)$ and*

$$17 \quad \sup_{x \in \mathcal{X}} \|\bar{V}(x, h) - V(x, h)\| = o(1)$$

19 Consider a subset \mathcal{X}_0 of \mathcal{X} such that $\bigcup_{x \in \mathcal{X}_0} \mathcal{V}(x, \epsilon) \subset \mathcal{X}$ for some $\epsilon > 0$. Then

$$23 \quad \sup_{x \in \mathcal{X}_0} \|\bar{V}(x, h) - V(x)\| = o(1)$$

25 Theorem 4 shows that the asymptotic variance $\bar{V}(x, h)$ of $(nh^d)^{1/2}H(\hat{\beta}(x, h) - \bar{\beta}(x, h))$ stays bounded over \mathcal{X} so that the order of $H(\hat{\beta}(x, h) - \bar{\beta}(x, h))$ should be $(nh^d)^{-1/2}$ for all $x \in \mathcal{X}$. Combining this result with Theorem 1 gives that $[\hat{\beta}_\pi(x, h) - \beta_\pi(x)]$ should be of order $((nh^d)^{-1/2} + h^s)/h^{-|\pi|}$ for all π with $|\pi| \leq s$ and for all $x \in \mathcal{X}$ including the boundary, that is, there is no boundary effect for the consistency rate of $\hat{\beta}_\pi(x, h)$.

31 As shown as Theorem 4, the variance boundary effect arises because the limit of $\Omega(x, h)^{-1}$ may differ from Ω^{-1} when x is close to the boundary as when $x = x_\partial + hc$ where x_∂ lies on the boundary. In the univariate case, Ruppert and Wand (1994) mentioned an increase by a factor 4 when x goes to the boundary and $K(\cdot)$ is a uniform kernel over $[-1, 1]$. The situation can be much worse in a multidimensional case since the variance increase due to boundary is not bounded and can be made arbitrarily large by considering a support which is very narrow in the vicinity of some estimation point x . For instance, consider a bivariate covariate $x = (x_1, x_2)'$ and the ray of the unit disk determined by angle $\bar{\theta} > 0$, that is,

$$41 \quad \mathcal{X}_{\bar{\theta}} = \{(x_1, x_2)'; x_1 = r \cos \theta, x_2 = r \sin \theta, (r, \theta) \in [0, 1] \times [0, \bar{\theta}]\}$$

1 and estimation at the vertex $o=(0,0)$. Suppose that $K(x_1, x_2)=K(r)$ with
 2 $\int_0^1 K(r)rdr = 1/(2\pi)$ and $U(x)=[1, x_1, x_2]'$. Then for h small enough and $\bar{\theta}$ small,
 3 the symmetric $\Omega_j(o, h)$, $j=1, 2$ satisfy

$$\begin{aligned}
 & \left[\begin{array}{cc} \frac{\bar{\theta}}{2\pi} \left(\bar{\theta} - \frac{\bar{\theta}^3}{6} \right) \int_0^1 r K^j(r)dr & \frac{\bar{\theta}^2}{2} \int_0^1 r K^j(r)dr \\ \times \left(\bar{\theta} - \frac{\bar{\theta}^3}{3} \right) \int_0^1 r^2 K^j(r)dr & \frac{\bar{\theta}^2}{2} \int_0^1 r^2 K^j(r)dr \\ \times & \times \frac{\bar{\theta}^3}{3} \int_0^1 r^2 K^j(r)dr \end{array} \right] + o(\bar{\theta}^3)
 \end{aligned}$$

15 which is such $\Omega_j(o, h)$ goes to 0 when $\bar{\theta} \rightarrow 0$. Further calculations yield that all
 16 the diagonal entries of

$$\Omega(o, h) = \Omega_1(o, h)^{-1} \Omega_2(o, h) \Omega_1(o, h)^{-1}$$

18 diverge as $1/\bar{\theta}$ diverges suggesting estimation of all derivatives becomes imprecise.
 19 In other words, although Assumption X implies that $\Omega(x, h)^{-1}$ stays bounded when
 20 x varies over \mathcal{X} , this example suggests that $\Omega(x, h)^{-1}$ can be large especially when
 21 the boundary takes the shape of such small angle ray with center x . By contrast, the
 22 leading bias term $\Omega_1(x, h)^{-1} b_{p+1}(x, h)$ from Theorem 2 is probably less affected
 23 by such small denominator problems. Indeed, by Eqs. (14) and (15),
 24

$$\Omega_1(x, h)^{-1} b_{p+1}(x, h) = \sum_{\pi \in \mathbb{N}^d: |\pi|=p+1} \Omega_1(x, h)^{-1} \kappa_\pi(x, h) \beta_\pi(x)$$

26 and a large $\Omega_1(x, h)^{-1}$ can be compensated by a small $\kappa_\pi(x, h)$.

33 4. LOCAL POLYNOMIAL REGRESSION AND 34 QUANTILE REGRESSION

35 In this section, we focus on two specific M-regressions, the local polynomial
 36 regression and local polynomial quantile regression. For each case, we establish
 37 asymptotic normality and uniform convergence rate. For local polynomial quantile,
 38 we first derive a uniform Bahadur representation valid over the whole support of
 39 the covariate X . We illustrate its usefulness via Examples 2–4.
 40

4.1. Local Polynomial Regression

Consider a regression model with a heteroscedastic error term,

$$Y_i = m(X_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i | X_i] = 0, \quad \text{Var}(\varepsilon_i | X_i) = \sigma^2(X_i) \quad (17)$$

and the quadratic loss function $\rho(t) = t^2$. We shall assume that

Assumption E.

The variables $(Y_i, X_i)'$, $i = 1, \dots, n$, are i.i.d. The conditional p.d.f. of ε_i given X_i denoted as $f(e|x)$ is continuous with respect to e and x . The variance function $\sigma^2(\cdot) = \text{Var}(\varepsilon_i | X_i = \cdot)$ is bounded away from 0 and continuous over \mathcal{X} . Moreover, $\sup_{x \in \mathcal{X}} \mathbb{E}[|\varepsilon_i|^{2+\nu} | X_i = x] < \infty$ for some $\nu > 0$.

For a quadratic loss function $\rho(\cdot)$, the local polynomial regression estimator in Eq. (3) is

$$\begin{aligned} \hat{\beta}(x, h) &= \left(\sum_{i=1}^n U(X_i - x) U(X_i - x)' K \left(\frac{X_i - x}{h} \right) \right)^{-1} \\ &\quad \times \sum_{i=1}^n U(X_i - x) Y_i K \left(\frac{X_i - x}{h} \right) \\ &= H^{-1} \left(\sum_{i=1}^n U \left(\frac{X_i - x}{h} \right) U \left(\frac{X_i - x}{h} \right)' K \left(\frac{X_i - x}{h} \right) \right)^{-1} \\ &\quad \times \sum_{i=1}^n U \left(\frac{X_i - x}{h} \right) Y_i K \left(\frac{X_i - x}{h} \right) \end{aligned} \quad (18)$$

where H is the diagonal matrix with entries $h^{|\pi|}$. The first entry of $\hat{\beta}(x, h)$, say $\hat{m}_h(x)$, is an estimator of the regression function $m(x)$, whereas the other entries estimate its partial derivatives. The pseudo-true value $\bar{\beta}(x, h)$ from Eq. (12) can be computed explicitly as

$$\begin{aligned} \bar{\beta}(x, h) &= \left(\mathbb{E} \left[U \left(\frac{X - x}{h} \right) U \left(\frac{X - x}{h} \right)' K \left(\frac{X - x}{h} \right) \right] \right)^{-1} \\ &\quad \times \mathbb{E} \left[\mu(X) U \left(\frac{X - x}{h} \right) K \left(\frac{X - x}{h} \right) \right] \end{aligned}$$

1 The asymptotic variance $V(x, h)$ is

$$3 \quad V(x, h) = \frac{\sigma^2(x)}{f(x)} \Omega(x, h)^{-1}$$

5

which can be estimated using

7

$$\widehat{V}(x, h) = \widehat{\sigma}^2(x) \widehat{\Omega}_1(x, h)^{-1} \widehat{\Omega}_2(x, h) \widehat{\Omega}_1(x, h)^{-1}$$

9

where

11

$$13 \quad \widehat{\Omega}_j(x, h) = \frac{1}{nh^d} \sum_{i=1}^n U\left(\frac{X_i - x}{h}\right) U\left(\frac{X_i - x}{h}\right)' K^j\left(\frac{X_i - x}{h}\right), \quad j = 1, 2,$$

15

17

$$\widehat{\sigma}^2(x) = \frac{\sum_{i=1}^n \left(Y_i - \widehat{\beta}_0(x, h)\right)^2 K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

19

which is consistent but with a large $O(h)$ boundary bias. Note that both $\widehat{\beta}(x, h)$ and $\widehat{V}(x, h)$ depend upon an inverse matrix which may not exist. The next lemma shows that $\widehat{\beta}(x, h)$ and $\widehat{V}(x, h)$ are well defined with a probability tending to 1.

21

23

Lemma 5. *Suppose that Assumptions K and X hold and that $h = h_n \rightarrow 0$ with $\log n / (nh^d) = o(1)$. Then, for $j = 1, 2$,*

25

27

$$\sup_{x \in \mathcal{X}} \left\| \frac{1}{nh^d} \sum_{i=1}^n U\left(\frac{X_i - x}{h}\right) U\left(\frac{X_i - x}{h}\right)' K^j\left(\frac{X_i - x}{h}\right) \right.$$

29

$$\left. - \int U(z) U'(z) K^j(z) f(x + hz) dz \right\|$$

31

$$= O_{\mathbb{P}}\left(\left(\frac{\log n}{nh^d}\right)^{1/2}\right) = o_{\mathbb{P}}(1)$$

33

(19)

35

Moreover, for $j = 1$, the limit matrix has an inverse and

37

$$\max_{x \in \mathcal{X}} \text{Sp} \left(\left(\int U(z) U'(z) K(z) f(x + hz) dz \right)^{-1} \right) = O(1)$$

39

41

The next two propositions show that standard asymptotic normality and uniform convergence results hold over the whole support. Because, as is well-known,

1 the process $(nh^d)^{1/2} H \left(\hat{\beta}(\cdot; h) - \bar{\beta}(\cdot; h) \right)$ is not tight so that convergence in dis-
 3 tribution to a Gaussian process cannot hold in usual functional sense, Proposition 6
 5 considers the asymptotic normality of $(nh^d)^{1/2} \widehat{V}(x_n, h)^{-1/2} H \left(\hat{\beta}(x_n; h) - \bar{\beta}(x_n; h) \right)$,
 7 where the sequence $\{x_n\} \subset \mathcal{X}$ can go to the boundary of the support \mathcal{X} . Note that
 Proposition 6 also establishes the consistency of the variance estimator $\widehat{V}(x_n, h)$
 near the boundary. These asymptotic normality and consistency results imply that
 each

$$9 \quad (nh^d)^{1/2} h^{|\pi|} \left(\hat{\beta}_\pi(x_n; h) - \bar{\beta}_\pi(x_n; h) \right)$$

11 is asymptotically normal and gives an estimate of its asymptotic variance.

13 **Proposition 6.** *Suppose that Assumptions E, K, X, and S1 or S2 hold, that*
 $h = h_n \rightarrow 0$ with $\log n / (nh^d) = o(1)$ and that $\{x_n\} \subset \mathcal{X}$ is a deterministic
 15 *sequence. Then*

$$17 \quad (nh^d)^{-1/2} \widehat{V}(x_n, h)^{-1/2} H \left(\hat{\beta}(x_n; h) - \bar{\beta}(x_n; h) \right)$$

19 *converges in distribution to a standard multivariate normal with*

$$21 \quad \widehat{V}(x_n, h) = V(x_n, h) + o_p(1)$$

23 **Proposition 7.** *Suppose that Assumptions E, K, X, and S1 or S2 hold, that*
 $h = h_n \rightarrow 0$ with $h^{-d} = O(n^{(v-2)/v} / \log n)$. Then

$$27 \quad \sup_{x \in \mathcal{X}} \left\| H \left(\hat{\beta}(x; h) - \beta(x) \right) \right\| = O_p \left(\left(\frac{\log n}{nh^d} \right)^{1/2} + h^s \right)$$

29 **Example 1 (Cont'd).** In this example, the two different regression functions are
 31 estimated using two different and independent samples of observations (X_{ji}, Y_{ji}) ,
 $j = 0, 1$, with the same sample size n and the same bandwidth for the sake of
 33 *simplicity. The estimator of the average treatment effect in Eq. (5) is then*

$$35 \quad \hat{\tau}_h(x) = \hat{\mu}_{1h}(x) - \hat{\mu}_{2h}(x), \quad x \in \mathcal{B}_{01}$$

37 Proposition 7 implies that $\hat{\tau}_h(x)$ converges uniformly to $\tau(x)$ over B_{01} with
 a rate $\left[(\log n / (nh^d))^{1/2} + h^s \right]$ extending the pointwise result in Imbens and
 39 Zajonc (2011). Proposition 6 easily extends to a vector $(\hat{\tau}_h(x_1), \dots, \hat{\tau}_h(x_K))'$
 41 which will be asymptotically independent. It follows that the estimation $\hat{\tau}_{Ah}$ of
 the average treatment effect mean τ_A as in Eq. (6) is asymptotically normal

1 with a bias which can be derived from Theorems 1 and 2 and an asymptotic
 2 variance of order $1/(nh^d)$ obtained by averaging the ones from Proposition 6.
 3 Approximation for the asymptotic distribution of $\hat{\tau}_{Mh}$ and $\hat{\tau}_{mh}$ can also be
 4 easily obtained under the additional condition that the maximum and mini-
 5 mum of the $\tau(x_k)$'s are achieved for one x_k , say x_M and x_m . In this case,
 6 $(nh^d)^{1/2}(\hat{\tau}_{Mh} - \hat{\tau}_h(x_M))$ and $(nh^d)^{1/2}(\hat{\tau}_{mh} - \hat{\tau}_h(x_m))$ are both $o_p(1)$ so that the
 7 asymptotic distribution of $(nh^d)^{1/2}(\hat{\tau}_{Mh} - \tau_M)$ and $(nh^d)^{1/2}(\hat{\tau}_{mh} - \tau_m)$ are the
 8 ones of $(nh^d)^{1/2}(\hat{\tau}_h(x_M) - \tau(x_M))$ and $(nh^d)^{1/2}(\hat{\tau}_h(x_m) - \tau(x_m))$. Interestingly,
 9 under the assumption of no treatment effect, the estimator $\hat{\tau}_h(x)$ is asymptotically
 10 unbiased because the same bandwidth is used for the treated and control samples.
 11 This may considerably simplify testing.

12 **Examples 2, 3, and 4 (Cont'd).** Although Propositions 6 and 7 do not apply to
 13 these Examples, the proof of these results suggests that our bias and variance results
 14 are sufficient to extend existing results which involve trimming to the bounded
 15 support case under Assumption X. For the full support test statistic \hat{t}_0 in Eq. (7),
 16 establishing that $nh^{d/2}\hat{t}_0$ converges in distribution to a centered normal distribution
 17 can be done with minor modifications of the arguments of Hjellvik et al. (1998).
 18 Studying the full support marginal integration estimator $\hat{\mu}_{1,0}(x_1)$ in Eq. (10) easily
 19 follows from Kong et al. (2010) while the full support average derivatives estimator
 20 $\hat{M}_0^{(1)}$ in Eq. (11) can be studied following Chaudhuri et al. (1997), Li et al. (2003),
 21 or Banerjee (2007).

22

23 4.2. Local Polynomial Quantile Regression

24 Consider the family of loss functions

$$25 \rho_\alpha(t) = t[\alpha - \mathbb{I}(t \leq 0)], \quad \alpha \in [\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$$

26 so that

$$27 \rho_\alpha^{(1)}(t) = \alpha - \mathbb{I}(t \leq 0)$$

28 For this choice of loss functions, $\hat{\beta}_0(\alpha|x, h) = \hat{\beta}_0(x, h)$ is an estimator of the condi-
 29 tional quantile function $Q(\alpha|x)$ of Y_i given $X_i = x$ and $\hat{\beta}_\pi(\alpha|x, h) = \hat{\beta}_\pi(x, h)$
 30 estimates the partial derivative $\partial^{|\pi|} Q(\alpha|x) / \partial x^\pi$. We shall use the following
 31 standard assumption.

32 Assumption F .

33 The variables $(Y_i, X_i)_{i=1}^n$ are i.i.d. The conditional p.d.f. of Y given $X = x$
 34 denoted as $f(y|x)$ is continuous and differentiable with respect to y

1 such that $f(y|x) > 0$ for all $(x, y) \in \mathcal{X} \times \mathbb{R}$, $\sup_{(x,y) \in \mathcal{X} \times \mathbb{R}} \left| \frac{\partial f(y|x)}{\partial y} \right| < \infty$, and
 3 $\sup_{(x,y) \in \mathcal{X} \times \mathbb{R}} \left| \frac{\partial f(y|x)}{\partial x} \right| < \infty$.

5 Under Assumption F, the minimizers $Q(\alpha|x) = \mu(x)$ and $\bar{\beta}(\alpha|x, h) = \bar{\beta}(x, h)$
 7 of Eqs. (1) and (12) are unique. Let $F(\cdot|\cdot)$ be the cumulative distribution
 9 function of Y given X . The functions $R_\alpha^{(1)}(\mu|x)$, $\mathbb{E} \left[(\rho_\alpha^{(1)}(Y - \mu))^2 | X = x \right]$,
 $\mathbb{E} \left[(\rho_\alpha^{(1)}(Y - Q(\alpha|x)))^2 | X = x \right]$, $R_\alpha^{(2)}(\mu|x)$, $V(\alpha|x, h)$, and $V(\alpha|x)$ are:

11

$$R_\alpha^{(1)}(\mu|x) = F(\mu|x) - \alpha, \quad R_\alpha^{(2)}(\mu|x) = f(\mu|x)$$

13

$$\mathbb{E} \left[(\rho_\alpha^{(1)}(Y - \mu))^2 | X = x \right] = \mathbb{E} \left[(\mathbb{I}(Y \leq \mu) - \alpha)^2 | X = x \right]$$

15

$$\mathbb{E} \left[(\rho_\alpha^{(1)}(Y - Q(\alpha|x)))^2 | X = x \right] = \alpha(1 - \alpha)$$

17

$$V(\alpha|x, h) = \frac{\alpha(1 - \alpha)}{f^2(Q(\alpha|x)|x) f(x)} \Omega(x, h)^{-1}, \quad \text{and}$$

19

$$V(\alpha|x) = \frac{\alpha(1 - \alpha)}{f^2(Q(\alpha|x)|x) f(x)} \Omega^{-1}$$

21

23 Hence, Assumption R follows from Assumption F. Since

25

$$\frac{1}{f(Q(\alpha|x)|x)} = \frac{\partial Q(\alpha|x)}{\partial \alpha}$$

27

29 a possible estimator of the asymptotic variance $V(\alpha|x, h)$ is

31

$$\widehat{V}(\alpha|x, h) = \alpha(1 - \alpha) \left(\frac{\partial \widehat{Q}(\alpha|x)}{\partial \alpha} \right)^2 \Omega(x, h)^{-1},$$

33

35 where $\frac{\partial \widehat{Q}(\alpha|x)}{\partial \alpha} = \frac{\hat{\beta}_0(\alpha + \eta|x, h) - \hat{\beta}_0(\alpha - \eta|x, h)}{2\eta}$, $\eta = \eta_n \rightarrow 0$

37

39 An important difference between conditional regression and quantile estima-
 41 tion is that the local polynomial quantile regression is neither explicit nor linear
 with respect to the Y_i . An additional step is needed to show that $\hat{\beta}(\alpha|x, h)$ is

1 asymptotically linear, see, for example, Chaudhuri (1991), Su and Xiao (2009), or
 2 Guerre and Sabbah (2012). To this aim, define

$$\begin{aligned}
 \widehat{S}(\alpha|x, h) &= \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n (\mathbb{I}(Y_i \leq U(X_i - x)' \bar{\beta}(\alpha|x, h)) - \alpha) \\
 &\quad \times U \left(\frac{X_i - x}{h} \right) K \left(\frac{X_i - x}{h} \right), \\
 \widehat{J}(\alpha|x, h) &= \frac{1}{nh^d} \sum_{i=1}^n f(U(X_i - x)' \bar{\beta}(\alpha|x, h) | X_i) U \left(\frac{X_i - x}{h} \right) \\
 &\quad \times U' \left(\frac{X_i - x}{h} \right) K \left(\frac{X_i - x}{h} \right), \quad \text{and} \\
 \bar{J}(\alpha|x, h) &= \mathbb{E} [\widehat{J}(\alpha|x, h)]
 \end{aligned}$$

17 Lemma 5 and Assumption F ensure that $\widehat{J}(\alpha|x, h)$ has an inverse asymptotically.
 18 The next linearization proposition is an extension of Guerre and Sabbah (2012,
 19 Theorem 2), which allows for estimation location x close or on the boundary of
 20 \mathcal{X} . Note that under Assumption F, the smoothness index s in Assumption S can be
 21 taken greater than 1 as assumed in all the results below.

23 **Proposition 8.** *Suppose that Assumptions K, F, X, and S1 or S2 hold for some*
 24 *$s \geq 1$, that $h = h_n \rightarrow 0$ with $\log n / (nh^d) = o(1)$. Then*

$$\begin{aligned}
 &\sup_{(\alpha, x) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}} \left\| H \left(\hat{\beta}(\alpha|x, h) - \bar{\beta}(\alpha|x, h) \right) + \frac{\widehat{J}(\alpha|x, h)^{-1} \widehat{S}(\alpha|x, h)}{(nh^d)^{1/2}} \right\| \\
 &= O_p \left(\left(\frac{\log n}{nh^d} \right)^{\frac{3}{4}} \right) \text{ and} \\
 &\sup_{(\alpha, x) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}} \left\| H \left(\hat{\beta}(\alpha|x, h) - \bar{\beta}(\alpha|x, h) \right) + \frac{\bar{J}(\alpha|x, h)^{-1} \widehat{S}(\alpha|x, h)}{(nh^d)^{1/2}} \right\| \\
 &= O_p \left(\left(\frac{\log n}{nh^d} \right)^{\frac{3}{4}} \right)
 \end{aligned}$$

39 Approximating $H \left(\hat{\beta}(\alpha|x, h) - \bar{\beta}(\alpha|x, h) \right)$ with $\widehat{J}(\alpha|x, h)^{-1} \widehat{S}(\alpha|x, h)$ as in the
 40 first equation gives an approximation similar to that for a local polynomial regres-
 41 sion estimator. Using the leading term $\bar{J}(\alpha|x, h)^{-1} \widehat{S}(\alpha|x, h)$ shows that the local

1 polynomial quantile estimator is asymptotically a sum which can be handled with
 2 standard limit theorems. A similar result can be established for a larger class of
 3 loss functions $\rho(\cdot)$ such as the one used in Powell and Newey (1987). However, the
 4 rate $(\log n/(nh^d))^{3/4}$ is typical of the quantile check function which is not twice
 5 continuously differentiable and better rates hold for smoother $\rho(\cdot)$.

6 This linearization result is the key tool to establish a Central Limit Theorem
 7 and uniform consistency for local polynomial quantile regression. The next two
 8 propositions parallel Propositions 6 and 7 for local polynomial regression. Like
 9 Theorems 1, 2, and 4, these results show that the boundary effect can be weak for
 10 local polynomial quantile regression estimators.

11 **Proposition 9.** *Suppose that Assumptions K, F, X, and S1 or S2 hold*
 12 *with $s \geq 1$, that $h = h_n \rightarrow 0$ and $\eta = \eta_n \rightarrow 0$ with $\log^3 n/(nh^d) = o(1)$,*
 13 *$h + (\log n/(nh^d))^{1/2} = o(\eta)$ and that $\{x_n\} \subset \mathcal{X}$ is a deterministic sequence.*
 14 *Then*

$$(nh^d)^{-1/2} \widehat{V}(\alpha|x_n, h)^{-1/2} H\left(\widehat{\beta}(\alpha|x_n; h) - \overline{\beta}(\alpha|x_n; h)\right)$$

17 *converges in distribution to a multivariate normal with*

$$\widehat{V}(\alpha|x_n, h) = V(\alpha|x_n, h) + o_p(1)$$

21 **Proposition 10.** *Suppose that Assumptions K, F, X, and S1 or S2 hold, that*
 22 *$h = h_n \rightarrow 0$ with $\log^3 n/(nh^d) = o(1)$. Then*

$$\sup_{(\alpha, x) \in [\underline{\alpha}, \overline{\alpha}] \times \mathcal{X}} \left\| H\left(\widehat{\beta}(\alpha|x; h) - \beta(\alpha|x; h)\right) \right\| = O_p\left(\left(\frac{\log n}{nh^d}\right)^{1/2} + h^s\right)$$

25 **Example 1 (Cont'd).** In the quantile setup, the estimation of the average treatment
 26 effect in Eq. (5) becomes

$$\widehat{\tau}_h(\alpha|x) = \widehat{Q}_{1h}(\alpha|x) - \widehat{m}_{2h}(\alpha|x), \quad \alpha \in [0, 1], \quad x \in \mathcal{B}_{01}$$

31 Proposition 10 yields that $\widehat{\tau}_h(\alpha|x)$ converges uniformly to $\tau(\alpha|x)$ over $[\underline{\alpha}, \overline{\alpha}] \times$
 32 \mathcal{B}_{01} with a rate $(\log n/(nh^d))^{1/2} + h^s$ extending Imbens and Zajonc (2011) to
 33 quantile setup. The indicators in Eq. (6) can be computed for each quantile levels.
 34 For a given α , they behave as in the regression case. When α varies in $[\underline{\alpha}, \overline{\alpha}]$,
 35 these indicators should be considered as stochastic processes whose asymptotic
 36 distribution can be derived using the asymptotic expansion stated in Proposition 8.

37 **Example 2 (Cont'd).** Proposition 8 is useful to obtain a suitable approximation
 38 of the test statistic \widehat{t}_c^2 . Let $\widetilde{\mu}(x) = \widetilde{Q}(\alpha|x)$ be the leading term of the conditional

1 quantile local polynomial estimator $\hat{\mu}(x) = \widehat{Q}(\alpha|x)$

$$3 \quad \tilde{\mu}(x) = \overline{Q}_h(\alpha|x) - e'_1 \frac{\overline{J}(\alpha|x, h)^{-1} \widehat{S}(\alpha|x, h)}{(nh^d)^{1/2}} \quad (20)$$

7 where $e'_1 = [1, 0, \dots, 0]$ and $\overline{Q}_h(\alpha|x) = e'_1 \overline{\beta}(\alpha|x; h)$. The “linearized” version of \hat{t}_c^2 is

$$9 \quad \tilde{t}_c^2 = \frac{1}{n} \sum_{i=1}^n \tilde{\mu}^2(X_i) \mathbb{I}[X_i \in \mathcal{X}_c]$$

11 Proposition 8 and the triangular inequality yield that

$$\begin{aligned} 13 \quad nh^{d/2} (\hat{t}_c^2 - \tilde{t}_c^2) &= \left((nh^{d/2})^{1/2} [\hat{t}_c - \tilde{t}_c] \right) \times \left((nh^{d/2})^{1/2} [\hat{t}_c + \tilde{t}_c] \right) \\ 15 \quad &= (nh^{d/2})^{1/2} O_p \left(\left(\frac{\log n}{nh^d} \right)^{\frac{3}{4}} \right) \\ 17 \quad &\quad \times \left((nh^{d/2})^{1/2} 2\hat{t}_c + (nh^{d/2})^{1/2} O_p \left(\left(\frac{\log n}{nh^d} \right)^{\frac{3}{4}} \right) \right) \\ 19 \quad &= O_p \left(\left(\frac{\log^3 n}{nh^{2d}} \right)^{\frac{1}{4}} \right) \times \left((nh^{d/2})^{1/2} 2\hat{t}_c + O_p \left(\left(\frac{\log^3 n}{nh^{2d}} \right)^{\frac{1}{4}} \right) \right) \end{aligned}$$

25 Consider the null hypothesis. Solving the first-order condition for $\overline{\beta}(\alpha|x; h)$ gives $\overline{\beta}(\alpha|x; h) = 0$ when $Q(\alpha|x) = 0$ for all x in \mathcal{X}_c . As a consequence, \tilde{t}_c^2 is a quadratic form similar for the one obtained in the regression framework, but with the centered variables $\mathbb{I}(Y_i \leq 0) - \alpha$ instead of the regression error terms. It follows that \tilde{t}_c^2 will satisfy Eq. (8) with a proper choice of standardizing constants. This gives,

$$31 \quad nh^{d/2} (\hat{t}_c^2 - \tilde{t}_c^2) = O_p \left(\left(\frac{\log^3 n}{nh^{3d}} \right)^{\frac{1}{4}} \right)$$

33 which goes to 0 when $(\log^3 n)/(nh^{3d}) \rightarrow 0$, a condition which will also ensure that \hat{t}_c^2 is asymptotically normal as in Eq. (8) and that normal critical values can be used to perform the test.

37 **Example 3 (Cont'd).** Defining a leave-one-out version of the linear $\tilde{\mu}(x_1, x_2)$ as in Eq. (20) yields for the marginal integration estimator (10) of $\mu_1(x_1)$ and applying

1 Proposition 8 suggests,

$$3 \hat{\mu}_{1,c}(x_1) = \frac{\sum_{i=1}^n \mathbb{I}(X_{2i} \in \mathcal{X}_{2c}) \tilde{\mu}_i(x_1, X_{2i})}{\sum_{i=1}^n \mathbb{I}(X_{2i} \in \mathcal{X}_{2c})} + O_p\left(\left(\frac{\log n}{nh^d}\right)^{\frac{3}{4}}\right)$$

5 with a leading term which can be studied as in Kong et al. (2010), for any $c \geq 0$.
 7 The estimator error $(nh^{d_1})^{1/2}(\hat{\mu}_{1,c}(x_1) - \mu_{1,c}(x_1))$ will satisfies a CLT if the con-
 9 tribution of the linearization remainder term is negligible, yielding the necessary
 condition,

$$11 (nh^{d_1})^{1/2} \left(\frac{\log n}{nh^d}\right)^{\frac{3}{4}} = \left(\frac{\log^3 n}{nh^{3d-2d_1}}\right)^{\frac{1}{4}} \rightarrow 0$$

13 **Example 4 (Cont'd).** The case of the average derivatives estimator is similar to
 15 Example 3. However, the contribution of the linearization remainder term must
 take into account an additional $1/h$ due to the estimation of a derivative. The $n^{1/2}$
 17 asymptotic normality of the estimator requests a more drastic bandwidth condition,

$$19 n^{1/2} \frac{1}{h} \left(\frac{\log n}{nh^d}\right)^{\frac{3}{4}} = \left(\frac{\log^3 n}{nh^{3d+4}}\right)^{\frac{1}{4}} \rightarrow 0$$

21 The asymptotic bias and variance of the estimator can be obtained from Chaudhuri
 23 et al. (1997) and our bias and variance results to account for the good boundary
 25 properties of the conditional quantile local polynomial estimator.

27 5. SIMULATION EXPERIMENTS

29 Our theoretical results have illustrated the good bias and variance boundary prop-
 31 erties of local polynomial estimation for general loss functions. The boundary bias
 has the same order as the minimax bias obtained for worst case specifications. The
 33 behavior of the variance is similar, with an order which is not affected by the bound-
 ary. However, for the variance, the constant in front of the order may be larger for
 35 points close to the boundary than for points away from the boundary. Preventing
 poor estimation induced by such areas may justify trimming in Examples 2, 3, and
 37 4. The purpose of this section is to use a small simulation experiment to illustrate
 how trimming influences inference in various settings.

39 We will use the setup of Examples 2, 3, and 4. To avoid nonlinearity issues, we
 will consider the regression model,

$$41 Y_i = \mu(X_{1i}, X_{2i}) + \varepsilon_i, \quad i = 1, \dots, 1,000$$

Table 1. Trimming Values Used in the Experiment.

Trimming 1 ($c = 0$)	Trimming 2 ($c = 0.1$)	Trimming 3 ($c = 0.2$)
$(X_{1i}, X_{2i}) \in [0, 1]^2$	$(X_{1i}, X_{2i}) \in [0.1, 0.9]^2$	$(X_{1i}, X_{2i}) \in [0.2, 0.8]^2$

where $\{[X_{1i}, X_{2i}]\}_{i=1}^n$ are i.i.d. Uniform over $[0, 1]^2$ and $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. centered normal with standard deviation 0.1. The choice of the regression function $\mu(\cdot)$ will vary across examples. The number of replications is 5,000. The regression local polynomial estimators of order 1 (linear) and order 2 (quadratic) will be considered with the kernel:

$$K(x) = (1 - x^2) \mathbb{I}(x \in [-1, 1])$$

Three trimming values will be investigated, see Table 1, where trimming 1 corresponds to no trimming.

5.1. Example 2: Significance Testing

Instead of Eq. (7), the more popular test statistic \widehat{T}_c of Eq. (9) is used. This test statistic can be written as a quadratic form $Y_n' W Y_n$, where the symmetric matrix W depends on the local polynomial estimator, the bandwidth h , and the trimming parameter c . Its variance can be estimated using

$$\widehat{\sigma}_c^2 = 2 \sum_{i \leq i \neq j \leq n} W_{ij}^2 \widehat{\varepsilon}_i^2 \widehat{\varepsilon}_j^2$$

where the residuals $\widehat{\varepsilon}_i$ are computed from the local polynomial estimation. The considered bandwidths are

$$h^2 \in \{0.01, 0.04, 0.07, 0.10, 0.13, 0.16, 0.19\}$$

We estimated the 90%, 95%, and 99% critical values by computing the test statistic $\widehat{T}_c / \widehat{\sigma}_c$ for each level of trimming and each bandwidth over 5,000 replications of the null model: $\mu(\cdot) = 0$. This did not show specific impact of trimming with simulated critical values reasonably close to their nominal counterparts.

The simulated 90% critical values were used to study the power of the tests using 1,000 replications of the model for the two regression functions below:

$$m_1(x_1, x_2) = 0.1K\left(\frac{x_1}{0.1}\right) - 0.1K\left(\frac{1-x_2}{0.1}\right) \text{ and}$$

$$m_2(x_1, x_2) = -0.05 \cos(6\pi x_1) + 0.05 \cos(6\pi x_2)$$

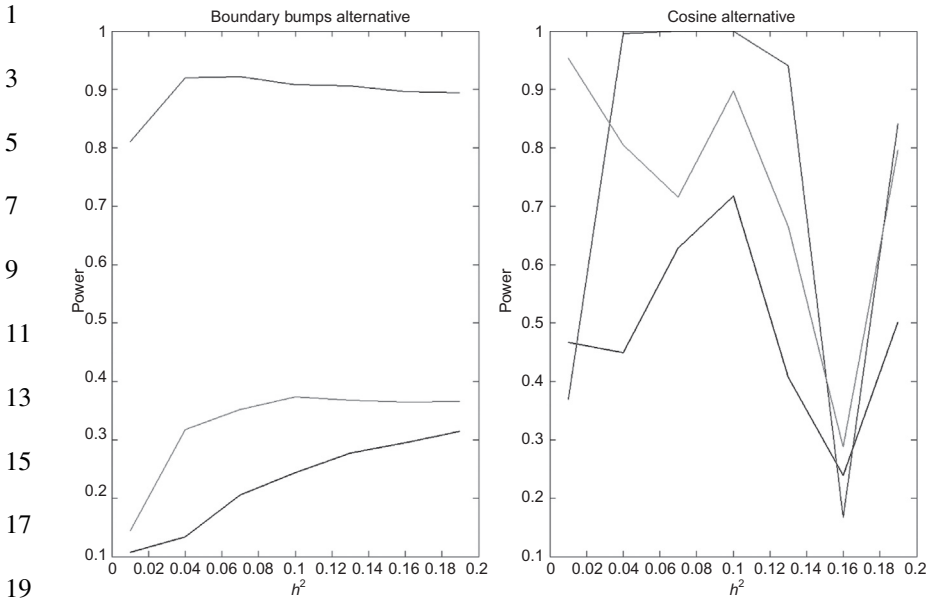


Fig. 1. Power of the Three Trimming 90% Tests as a Function of h^2 . Left: Alternative $m_1(\cdot)$. Right: Alternative $m_2(\cdot)$. Note: Blue: No trimming. Green: Trimming 2. Red: Trimming 3.

The alternative $m_1(\cdot)$ consists of two bumps along the boundaries $x_1 = 0$ and $x_2 = 1$ which are difficult to detect with the trimming tests. On the contrary, the alternative $m_2(\cdot)$ violates the null for most x in the support $[0, 1]^2$. Fig. 1 reports the results of the simulation experiment. As expected the power of the tests against alternative $m_1(\cdot)$ deteriorates with the level of trimming and the test without trimming clearly dominates. The evidence is less clear for alternative $m_2(\cdot)$ which periodicity induces an irregular bandwidth behavior. However, only the test without trimming achieves a power close to 1. Considering a quadratic local polynomial estimator gives a much less powerful test for the considered bandwidths. This surprising finding is however in line with the theoretical results of Guerre and Lavergne (2002) which shows that lower order methods can have good power properties in a minimax framework.

AQ:4

5.2. Example 3: Additive Specification Estimation

In this experiment, the regression function is set to $m(x_1) + m(x_2)$ with $m(x_1) = \sin(2\pi x_1)$. The local linear smoother $\hat{\mu}_{1,c}(x_1)$ from Eq. (10) is an estimator of

$$\mu_{1,c}(x_1) = m(x_1) + \mathbb{E}[m(X_2)\mathbb{I}(X_2 \in [c, 1 - c])]$$

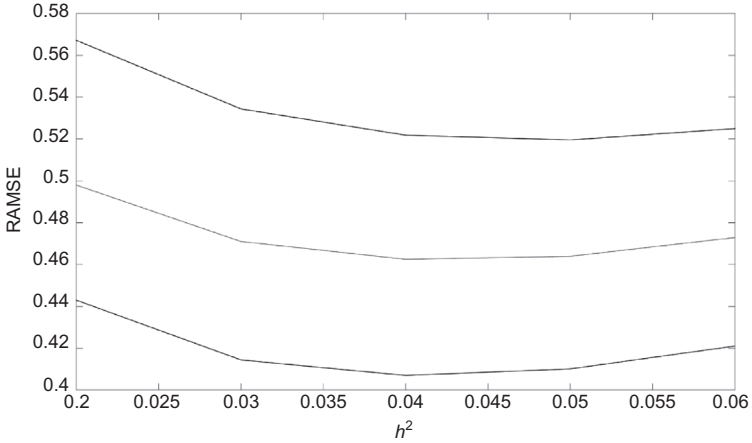


Fig. 2. Square Root Average Mean-Squared Error of $\widehat{m}_c(\cdot)$ as a Function of h^2 . Note: Blue: No trimming. Green: Trimming 2. Red: Trimming 3.

which includes an expectation term. This term can be removed by imposing an identification restriction such as $m(1/2) = 0$. The proposed estimator of $m(x_1)$ is therefore²

$$\widehat{m}_c(x_1) = \widehat{\mu}_{1,c}(x_1) - \widehat{\mu}_{1,c}(1/2)$$

The local linear estimator performs poorly and the reported results are for the local quadratic estimator. The considered bandwidths are smaller than the ones used in Example 1. They are:

$$h^2 \in \{0.02, 0.03, \dots, 0.06\}$$

The performance of the estimator is evaluated using the square root average mean-squared error (RAMSE):

$$RAMSE = \left[\frac{1}{11} \sum_{j=0}^{10} \left(\widehat{m}_c\left(\frac{j}{10}\right) - m_c\left(\frac{j}{10}\right) \right)^2 \right]^{\frac{1}{2}}$$

Fig. 2 shows that the estimator without trimming clearly dominates. RAMSE is around 0.40 at best, which is quite big but not surprising since the best bandwidth is $h^2 = 0.04$ implying that only 40 observations are used to estimate each $\mu_{1,c}(j/10)$.

5.3. Example 4: Average Derivative Estimation

For this example, a single-index specification is used and the regression function is

$$\mu(x) = \sin^3\left(\frac{\pi}{2}(x' - [1/2, 1/2])\beta_0\right), \quad \beta'_0 = [1, 1/2]$$

The parameter of interest is $\delta_0 = \beta_{02}/\beta_{01} = 1/2$ which is estimated using a ratio of average derivatives

$$\widehat{\delta}_c = \frac{\sum_{i=1}^n \widehat{\partial\mu_i/\partial x_2}(X_i)\mathbb{I}[X_i \in \mathcal{X}_c]}{\sum_{i=1}^n \widehat{\partial\mu_i/\partial x_1}(X_i)\mathbb{I}[X_i \in \mathcal{X}_c]}$$

The performance of $\widehat{\delta}_c$ is measured using the square root mean squared error (RMSE). Because estimation of derivatives yields a bigger variance of order $1/(nh^3)$ compared to estimation of a regression for which the variance order is $1/(nh^2)$, the RMSE of $\widehat{\delta}_c$ could be more sensitive than the RAMSE of Example 3

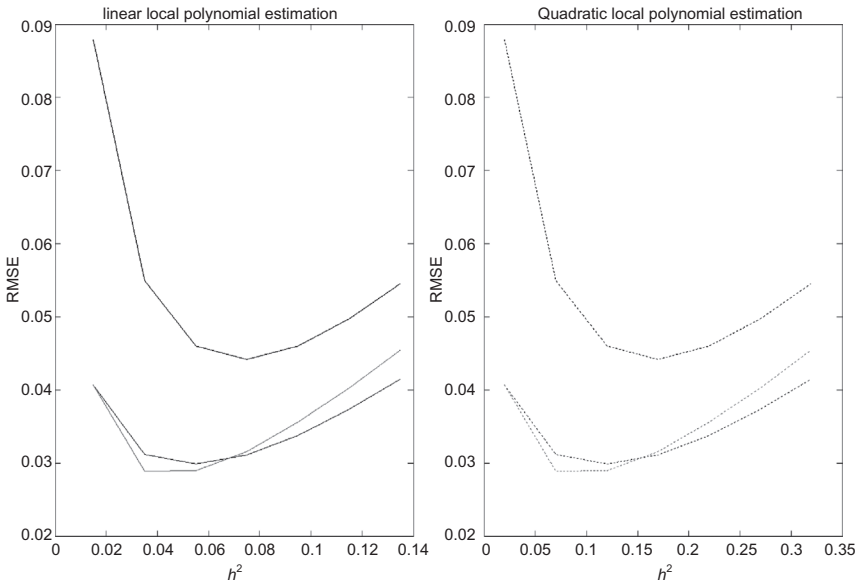


Fig. 3. Square Root Mean-Squared Error for the Three Trimming Estimators as a Function of h^2 . Left: Linear Local Polynomial. Right: Quadratic Local Polynomial. Note: Blue: No trimming. Green: Trimming 2. Red: Trimming 3.

1 to an increase of the variance near the boundary. The local linear and quadratic
 2 regression estimators perform similarly and both are reported here. The considered
 3 bandwidths are

5 $h^2 \in \{0.015, 0.035, \dots, 0.135\}$ for linear local estimation and

7 $h^2 \in \{0.02, 0.07, \dots, 0.32\}$ for local quadratic estimation

9 The highest trimming estimator is dominated by lower trimming ones, for both
 10 local linear and quadratic estimation. The trimming 1 and 2 estimators behave
 11 similarly but trimming 2 seems slightly better than no trimming when looking at
 12 the optimal performance. This holds for both local linear and quadratic estimators.
 13 Since the bias of the local quadratic estimator is smaller, this behavior suggests that
 14 there is an optimal level of trimming possibly due to an increase of variance near the
 15 boundary. However, the potential gain seems very small in this experiment (Fig. 3).

AQ:5

17

6. CONCLUSION

19

21

23

25

27

29

31

33

In this paper, we have investigated the boundary and uniform asymptotic properties of multivariate local polynomial estimators of M-regression functions under a weak condition on the *compact* support of the multivariate covariate. This is made possible by a pseudo-true value approach based on a novel application of the Newton–Kantorovich Theorem in our context. Compared with Assumption (A4) in Ruppert and Wand (1994) who use it to establish pointwise boundary properties of local polynomial regression estimators, our support condition allows for more general support shapes, in particular, it allows the support of the covariate to be non-connected and have holes. Compared with the uniform result in Corollary 2, ii) in Masry (1996), our results deal directly with multivariate covariates with compact support and are uniformly valid over the entire support. As such they should be useful in contexts where estimation or testing require (weighted) averages of multivariate nonparametric estimators with compactly supported covariate as in Examples 2–4.

35

7. PROOF SECTION

37

39

41

Note that Assumption X ensures that for any $h > 0$ small enough, any $x \in \mathcal{X}$, there is a $x_h \in \mathcal{X}$ such that $\mathcal{V}(x_h, \kappa_1 h) \subset \mathcal{V}(x, h) \cap \mathcal{X}$. We will use this equivalent statement of Eq. (13) throughout this section. Also we will use C to denote a generic positive constant whose value may differ in different places.

We start with Lemma 11 which will be used in many proofs below.

1 **Lemma 11.** Under Assumptions K and X, there is a $C > 1$ such that the eigenvalues of $\int U(z)U'(z)K^j(z)f(x + hz)dz$ and $\Omega_j(x, h)$, $j = 1, 2$, are in $[1/C, C]$
 3 for all $x \in \mathcal{X}$ and $h \geq 0$ small enough.

5 **Proof of Lemma 11.** It is sufficient to consider $j = 1$. Since

$$7 \quad \int U(z)U'(z)K(z)f(x + hz)dz \leq C\Omega_j(x, h) \leq C \int U(z)U'(z)K(z)dz$$

9 the eigenvalues are in $[0, C]$ for all $x \in \mathcal{X}$ and $h \geq 0$. We now show that the eigenvalues can be bounded from below by $1/C$. Assumptions K and X-(i) give, uniformly in $x \in \mathcal{X}$,

$$13 \quad \int U(z)U'(z)K(z)f(x + hz)dz \geq C\Omega(x, h)$$

$$15 \quad \geq C \int U(z)U(z)' \mathbb{I}(x + hz \in \mathcal{X}, z \in \mathcal{V}(0, 1)) dz$$

17 Assume that h is small enough. Assumption X-(ii) gives, with $(t - x)/h = z$,

$$21 \quad \int U(z)U(z)' \mathbb{I}(x + hz \in \mathcal{X}, z \in \mathcal{V}(0, 1)) dz$$

$$23 \quad = \int U\left(\frac{t-x}{h}\right)U\left(\frac{t-x}{h}\right)' \mathbb{I}(t \in \mathcal{V}(x, h) \cap \mathcal{X}) \frac{dt}{h^d}$$

$$25 \quad \geq \int U\left(\frac{t-x}{h}\right)U\left(\frac{t-x}{h}\right)' \mathbb{I}(t \in \mathcal{V}(x_h, \kappa_1 h)) \frac{dt}{h^d}$$

$$27 \quad \geq \int U(z)U(z)' \mathbb{I}\left(z \in \mathcal{V}\left(\frac{x_h - x}{h}, \kappa_1\right)\right) dz$$

29 where $\|(x_h - x)/h\| \leq 1$. Hence, the eigenvalues of $\int U(z)U'(z)K(z)f(x + hz)dz$
 31 and $\Omega_j(x, h)$ are larger than

$$33 \quad \inf_{y \in \mathcal{V}(0, 1)} \min_{b: b'b=1} b' \left(\int U(z)U(z)' \mathbb{I}(z \in \mathcal{V}(y, \kappa_1)) dz \right) b$$

35 Suppose now that this lower bound is equal to 0. This implies that there is a
 37 sequence $y_n \in \mathcal{V}(0, 1)$ and b_n with $b_n'b_n = 1$ such that

$$39 \quad \int (U(z)'b_n)^2 \mathbb{I}(z \in \mathcal{V}(y_n, \kappa_1)) dz \rightarrow 0$$

41

1 By compactness and continuity of $(b, y) \mapsto \int (U(z)'b)^2 \mathbb{I}(z \in \mathcal{V}(y, \kappa_1)) dz$, this
 3 implies that there is a $y \in \mathcal{V}(0, 1)$ and b with $b'b = 1$ such that

$$5 \quad \int (U(z)'b)^2 \mathbb{I}(z \in \mathcal{V}(y, \kappa_1)) dz = 0$$

7 Hence, $U(z)'b = 0$, but this is impossible since $b \neq 0$. Hence, the eigenvalues of
 9 $\int U(z)U'(z)K(z)f(x + hz)dz$ are in $[1/C, C]$ for all $x \in \mathcal{X}$ and h small enough. ■

11 7.1. Theorems 1, 2, and Proposition 3

13 A technical challenge comes from the fact that $\bar{\beta}(x, h)$ is not explicit but defined
 15 through Eq. (12). The next lemma is the key tool to study the bias term
 17 $[\bar{\beta}(x, h) - \beta(x)]$ when using the first-order condition which characterizes $\bar{\beta}(x, h)$.
 In this lemma, D is an integer number and $\|\cdot\|$ stands for the Euclidean norm over
 \mathbb{R}^D or for the associated operator norm.

19 **Lemma 12. (Newton–Kantorovich).** *Let $\mathcal{F}(\cdot): \mathbb{R}^D \rightarrow \mathbb{R}$ be a twice contin-*
uously differentiable convex function with a unique minimizer $\bar{\mathbf{b}}$. Suppose
 21 *that*

- 23 1. *There is $\mathbf{b}^* \in \mathbb{R}^D$ such that $\|\mathcal{F}^{(1)}(\mathbf{b}^*)\| \leq \eta$ and $\left\| [\mathcal{F}^{(2)}(\mathbf{b}^*)]^{-1} \right\| \leq C_0$;*
- 25 2. *$\|\mathcal{F}^{(2)}(b) - \mathcal{F}^{(2)}(b')\| \leq C_1 \|b - b'\|$ for all $b, b' \in \mathbb{R}^D$;*
3. *$C_0^2 C_1 \eta \leq 1/4$.*

27 *Then $\|\mathbf{b}^* - \bar{\mathbf{b}}\| \leq 2C_0\eta$.*

29 **Proof of Lemma 12.** This follows from conclusion 3 in the Newton–Kantorovich
 Theorem stated in Gragg and Tapia (1974). ■

31 It is convenient to rescale β with H^{-1} and to set $b = H^{-1}\beta$, where H is a
 diagonal matrix with entries $h^{|\pi|}$. This gives in particular

$$33 \quad U(X_i - x)' \beta = U \left(\frac{X_i - x}{h} \right)' H \beta = U \left(\frac{X_i - x}{h} \right)' b$$

37 Lemma 12 will be applied for $\bar{\mathbf{b}} = H \bar{\beta}(x; h) = \bar{b}(x; h)$, where $\bar{\beta}(x; h)$ is as in
 Eq. (12) and we now define a candidate \mathbf{b}^* . Define

$$39 \quad s_p = \begin{cases} s & \text{for the proof of Theorem 1} \\ p & \text{for the proof of Theorem 2} \end{cases}$$

41

1 and $b^*(x, h) = (b_{\pi}^*(x, h), |\pi| \leq p)'$ with

$$3 \quad b_{\pi}^*(x, h) = \begin{cases} \frac{h^{|\pi|} \pi_1! \cdots \pi_d! \mu^{(\pi)}(x)}{|\pi|!} & \text{for } |\pi| \leq \lfloor s_p \rfloor \\ 0 & \text{for } \lfloor s_p \rfloor < |\pi| \leq p \end{cases}$$

7 In other words, when proving Theorem 1, $b^*(x, h)$ completes the entries
 9 $h^{|\pi|} \mu^{(\pi)}(x) / (\pi_1! \cdots \pi_d!)$, $|\pi| \leq \lfloor s \rfloor$, with entries equal to 0 whereas in the proof
 11 of Theorem 2 the entries of $b^*(x, h)$ are all the $h^{|\pi|} \mu^{(\pi)}(x) / (\pi_1! \cdots \pi_d!)$, $|\pi| \leq p$.
 13 The Taylor formula implies, under Assumptions S and X which implies that \mathcal{X} has
 no isolated points,

$$15 \quad \max_{(x,z) \in \mathcal{X} \times \text{Supp } K, x+hz \in \mathcal{X}} |\mu(x+hz) - U(z)' b^*(x, h)| \leq C h^{s_p} \quad (21)$$

17 with $C = C'L$ under Assumption S1.

19 Define now

$$21 \quad \mathbf{R}_h(b|x) = \int \mathbf{R}(U(z)'b|x+hz) f(x+hz) K(z) dz$$

23 where $x+hz$ stands for X . The Lebesgue Dominated Convergence Theorem
 25 gives under Assumptions R, X, and K that $b \mapsto \mathbf{R}_h(b|x)$ is twice continuously
 27 differentiable with first and second derivatives

$$29 \quad \mathbf{R}_h^{(1)}(b|x) = \int \mathbf{R}^{(1)}(U(z)'b|x+hz) U(z) f(x+hz) K(z) dz \quad \text{and}$$

$$31 \quad \mathbf{R}_h^{(2)}(b|x) = \int \mathbf{R}^{(2)}(U(z)'b|x+hz) U(z) U(z)' f(x+hz) K(z) dz$$

33 The next lemma shows that the matrix $\mathbf{R}_h^{(2)}(b|x)$ satisfies some of the conditions
 35 of Lemma 12.

37 **Lemma 13.** *Under Assumptions K, R, and X and for $h \leq \kappa_0/\kappa_1$, there is a $C > 1$
 39 such that the eigenvalues of $\mathbf{R}_h^{(2)}(b|x)$ are in $[1/C, \infty)$ for all b and all $x \in \mathcal{X}$
 41 and, for all b, b' and all $x \in \mathcal{X}$, $\|\mathbf{R}_h^{(2)}(b|x) - \mathbf{R}_h^{(2)}(b'|x)\| \leq C \|b - b'\|$.*

1 **Proof of Lemma 13.** The bound for $\left\| \mathbf{R}_h^{(2)}(b|x) - \mathbf{R}_h^{(2)}(b'|x) \right\|$ follows from
 3 Assumption R:

$$\begin{aligned}
 & \left\| \mathbf{R}_h^{(2)}(b|x) - \mathbf{R}_h^{(2)}(b'|x) \right\| \\
 & \leq \int |R^{(2)}(U(z)'b|x+hz) - R^{(2)}(U(z)'b'|x+hz)| \\
 & \quad \times \|U(z)U(z)'\| f(x+hz)K(z)dz \\
 & \leq C \int |U(z)'(b-b')| \|U(z)\|^2 f(x+hz)K(z)dz \leq C \\
 & \quad \times \int \|b-b'\| \|U(z)\|^3 f(x+hz)K(z)dz \leq C \|b-b'\|
 \end{aligned}$$

17 since the support of $K(\cdot)$ is compact and $f(\cdot)$ is bounded. For the lower bound of
 19 the eigenvalues of $\mathbf{R}_h^{(2)}(b|x)$, observe

$$\mathbf{R}_h^{(2)}(b|x) \succeq C \int U(z)U(z)'K(z)f(x+hz)dz$$

25 so that the result follows from Lemma 11.

27 Let us now return to the proof of Theorems 1 and 2. Equation (12) gives
 $\bar{\beta}(x; h) = \arg \min_{\beta} \mathbf{R}_h(H^{-1}\beta|x)$ so that $\bar{b}(x; h)$ satisfies the first-order condition:

$$\mathbf{R}_h^{(1)}(\bar{b}(x; h)|x) = 0 \quad \text{for all } x \in \mathcal{X} \tag{22}$$

33 We now study $\mathbf{R}_h^{(1)}(b^*(x; h)|x)$ which satisfies:

$$\begin{aligned}
 \mathbf{R}_h^{(1)}(b^*(x; h)|x) &= \int U(z)R^{(1)}(U(z)'b^*(x; h)|x+hz) f(x+hz)K(z)dz \\
 &= \int U(z)R^{(1)}(\mu(x+hz) + U(z)'b^*(x; h) - \mu(x+hz)|x+hz) \\
 & \quad \times f(x+hz)K(z)dz
 \end{aligned}$$

1 Hence, Eq. (21), Assumption R, and Eq. (1) give, uniformly in $x \in \mathcal{X}$

$$\begin{aligned}
 3 \quad & \left\| \mathbf{R}_h^{(1)}(b^*(x; h)|x) \right\| \leq \left\| \int U(z) R^{(1)}(\mu(x + hz)|x + hz) f(x + hz) K(z) dz \right\| \\
 5 \quad & + C \max_{(x, x+hz, z) \in \mathcal{X}^2 \times \text{Supp } K} \left| \mu(x + hz) - U(z)' b^*(x, h) \right| \\
 7 \quad & \leq Ch^{s_p} \\
 9 \quad &
 \end{aligned}$$

11 Then Lemma 12 shows that Theorem 1 is proved.

13 For Theorem 2, recall that $s_p = p$. A Taylor expansion of order $p + 1$ gives that uniformly in $x \in \mathcal{X}$

$$17 \quad \mu(x + hz) - U(z)' b^*(x, h) = h^{p+1} \sum_{|\pi|=p+1} \frac{\pi_1! \cdots \pi_d!}{(p+1)!} z^\pi \mu^{(\pi)}(x) + o(h^{p+1})$$

21 Recall that $\mathbf{R}_{f, \mu}^{(2)}(x) = R^{(2)}(\mu(x)|x) f(x)$ is bounded away from 0. Hence, Eqs. (22), (1) and standard uniform expansions give, for $\mathbb{I}_{\mathcal{X}}(x + hz) = \mathbb{I}(x + hz \in \mathcal{X})$,

$$\begin{aligned}
 25 \quad 0 &= \int U(z) R^{(1)}(\mu(x + hz) + U(z)' \bar{b}(x; h) - \mu(x + hz)|x + hz) \\
 27 \quad & \times f(x + hz) K(z) dz = \underbrace{\int U(z) R^{(1)}(\mu(x + hz)|x + hz) f(x + hz) K(z) dz}_{=0} \\
 29 \quad & + \int \left(\mathbf{R}_{f, \mu}^{(2)}(x + hz) + o(1) \right) U(z) (U(z)' \bar{b}(x; h) \\
 31 \quad & - \mu(x + hz)) \mathbb{I}_{\mathcal{X}}(x + hz) K(z) dz \\
 33 \quad & = \left(\mathbf{R}_{f, \mu}^{(2)}(x) + o(1) \right) \int U(z) U(z)' (\bar{b}(x; h) - b^*(x; h)) \mathbb{I}_{\mathcal{X}}(x + hz) K(z) dz \\
 35 \quad & + \left(\mathbf{R}_{f, \mu}^{(2)}(x) + o(1) \right) \int U(z) \left\{ h^{p+1} \sum_{|\pi|=p+1} \frac{\pi_1! \cdots \pi_d!}{(p+1)!} z^\pi \mu^{(\pi)}(x) \right\} \\
 37 \quad & \times \mathbb{I}_{\mathcal{X}}(x + hz) K(z) dz + o(h^{p+1}) \\
 39 \quad & \\
 41 \quad &
 \end{aligned}$$

1 Rearranging gives,

$$\begin{aligned}
3 \quad \bar{b}(x; h) - b^*(x; h) &= h^{p+1} \left(\int U(z)U(z)'K(z)\mathbb{I}_{\mathcal{X}}(x + hz)dz \right)^{-1} \\
5 &\quad \times \int U(z) \left\{ \sum_{|\pi|=p+1} \frac{\pi_1! \cdots \pi_d!}{(p+1)!} z^\pi \mu^{(\pi)}(x) \right\} \mathbb{I}_{\mathcal{X}}(x + hz)K(z)dz \\
7 &\quad + o(h^{p+1})
\end{aligned}$$

9 showing that Theorem 2 is proved.

11 For Proposition 3, performing a Taylor expansion of order $p + 2$ and arguing
13 as above gives, for inner x ,

$$15 \quad \bar{b}(x; h) - b^*(x; h) = \frac{(1 + o(1))}{\mathbf{R}_{f,\mu}^{(2)}(x)} \Omega^{-1} \mathbf{B}(x; h) + o(h^{p+2})$$

17 where

$$\begin{aligned}
19 \quad \mathbf{B}(x; h) &= \int U(z) \left\{ h^{p+1} \mathbf{R}_{f,\mu}^{(2)}(x + hz) \sum_{|\pi|=p+1} \frac{\pi_1! \cdots \pi_d!}{(p+1)!} z^\pi \mu^{(\pi)}(x) \right\} K(z)dz \\
21 &\quad + \int U(z) \left\{ h^{p+2} \mathbf{R}_{f,\mu}^{(2)}(x) \sum_{|\pi|=p+2} \frac{\pi_1! \cdots \pi_d!}{(p+1)!} z^\pi \mu^{(\pi)}(x) \right\} K(z)dz \\
23 &= h^{p+1} \mathbf{R}_{f,\mu}^{(2)}(x) b_{p+1}(x) + h^{p+2} \sum_{j=1}^d \frac{\partial \mathbf{R}_{f,\mu}^{(2)}(x)}{\partial x_j} \mu^{(\pi)}(x) \int z_j z^\pi U(z) K(z) dz \\
25 &\quad + o(h^{p+2}) + h^{p+2} \mathbf{R}_{f,\mu}^{(2)}(x) b_{p+2}(x)
\end{aligned}$$

27 This gives the expansion of the proposition since $e'_\pi \Omega_1^{-1} b_{p+1}(x) = 0$ by Gu et al.
29 (2015). ■

33 7.2. Theorem 4

35 Let

$$\begin{aligned}
37 \quad \mathbf{R}^{(2)}(x, h) &= h^{-d} \mathbb{E} \left[\mathbf{R}^{(2)}(\mu(x)|X) U \left(\frac{X-x}{h} \right) U \left(\frac{X-x}{h} \right)' K \left(\frac{X-x}{h} \right) \right], \\
39 \quad \mathbf{S}(x, h) &= h^{-d} \mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 U \left(\frac{X-x}{h} \right) U \left(\frac{X-x}{h} \right)' K^2 \left(\frac{X-x}{h} \right) \right]
\end{aligned}$$

41

1 Theorems 1 and 2, Eq. (21), give that

$$3 \max_{(x,z) \in \mathcal{X} \times \text{Supp } K, x+hz \in \mathcal{X}} |U(hz)' \beta(x, h) - \mu(x + hz)| = o(1)$$

7 Therefore $\max_{x,h} \text{Sp} \left(h^{-d} \mathbb{E} \left[U \left(\frac{X-x}{h} \right) U \left(\frac{X-x}{h} \right)' K^j \left(\frac{X-x}{h} \right) \right] \right) < \infty$, $j = 1, 2$,
 9 Assumption R-(ii), and uniform continuity of $\mu(\cdot)$ over \mathcal{X} give when $h \rightarrow 0$,

$$11 \max_{x \in \mathcal{X}} \left\| \overline{\mathbf{R}}^{(2)}(x, h) - \mathbf{R}^{(2)}(x, h) \right\| = o(1), \quad \max_{x \in \mathcal{X}} \left\| \overline{\mathbf{S}}(x, h) - \mathbf{S}(x, h) \right\| = o(1)$$

15 A standard change of variable gives

$$17 \mathbf{R}^{(2)}(x, h) = \int \mathbf{R}^{(2)}(\mu(x)|x + hz) U(z)U(z)' K(z) f(x + hz) dx$$

$$19 = \mathbf{R}^{(2)}(\mu(x)|x) f(x) \Omega_1(x, h) + o(1),$$

$$21 \mathbf{S}(x, h) = \int \mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 | X = x + hz \right] U(z)U(z)' K^2(z) f(x + hz) dx$$

$$23 = \mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 | X = x \right] f(x) \Omega_2(x, h) + o(1)$$

27 uniformly over \mathcal{X} by Assumption R-(ii) and X-(i). Hence, Lemma 11 yields that
 $\overline{\mathbf{V}}(x, h)$ is equal to

$$31 \left(\mathbf{R}^{(2)}(\mu(x)|x) f(x) \Omega_1(x, h) \right)^{-1} \mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 | X = x \right]$$

$$33 \times f(x) \Omega_2(x, h) \left(\mathbf{R}^{(2)}(\mu(x)|x) f(x) \Omega_1(x, h) \right)^{-1} + o(1)$$

$$35 = \frac{\mathbb{E} \left[(\rho^{(1)}(Y - \mu(X)))^2 | X = x \right]}{\left(\mathbf{R}^{(2)}(\mu(x)|x) \right)^2 f(x)} \Omega(x, h)^{-1} + o(1)$$

39 uniformly over \mathcal{X} , that is the first approximation in the theorem. The second
 41 approximation follows since $\Omega(x, h) = \Omega$ for all x in subset \mathcal{X}_0 of \mathcal{X} as in the
 theorem provided h is small enough. ■

7.3. Lemma 5, Propositions 6, and 7

The proof of these results makes use of the Bernstein inequality, which states that for independent centered real random variables Z_i with $|Z_i| \leq M$,

$$\mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \right| \geq t \right) \leq 2 \exp \left(- \frac{t^2/2}{\frac{1}{n} \sum_{i=1}^n \text{Var}(Z_i) + \frac{1}{3} \frac{Mt}{\sqrt{n}}} \right) \text{ for any } t \geq 0 \quad (23)$$

Proof of Lemma 5. It is sufficient to consider $j=1$. To prove Eq. (19), it is sufficient to show that for any $\epsilon > 0$ large enough, all π with $|\pi| \leq 2p$, $K_\pi(z) = z^\pi K(z)$ and $r_n = (\log n / (nh^d))^{1/2}$,

$$\mathbb{P} \left(\sup_{x \in \mathcal{X}} \left| \frac{1}{nh^d} \sum_{i=1}^n K_\pi \left(\frac{X_i - x}{h} \right) - \int K_\pi(z) f(x + hz) dz \right| \geq r_n \epsilon \right) \rightarrow 0 \quad (24)$$

By Assumptions K, X and $1/h^d = O(n/\log n)$, there is a $\delta = \delta_n = n^{-a}$ such that

1. There is an integer number $J_n = O(n^b)$, $b > 0$, and some $x_j \in \mathcal{X}$ such that $\mathcal{X} = \bigcup_{j=1}^{J_n} \mathcal{V}_X(x_j, \delta_n)$, where $\mathcal{V}_X(x_j, \delta_n) = \mathcal{V}(x_j, \delta_n) \cap \mathcal{X}$;
2. For all x, x' with $\|x - x'\| \leq \delta_n$ and all i and n , $\left| K_\pi \left(\frac{X_i - x}{h} \right) - K_\pi \left(\frac{X_i - x'}{h} \right) \right| \leq h^d r_n \epsilon / 3$;
3. For all $x, x' \in \mathcal{X}$ with $\|x - x'\| \leq \delta_n$ and all n , $\left| \int K_\pi(z) f(x + hz) dz - \int K_\pi(z) f(x' + hz) dz \right| \leq r_n \epsilon / 3$.

This gives

$$\begin{aligned} & \sup_{x \in \mathcal{X}} \left| \frac{1}{nh^d} \sum_{i=1}^n K_\pi \left(\frac{X_i - x}{h} \right) - \int K_\pi(z) f(x + hz) dz \right| \\ & \leq \max_{j=1, \dots, J_n} \left| \frac{1}{nh^d} \sum_{i=1}^n K_\pi \left(\frac{X_i - x_j}{h} \right) - \int K_\pi(z) f(x_j + hz) dz \right| \\ & \quad + \max_{j=1, \dots, J_n} \sup_{x \in \mathcal{V}_X(x_j, \delta_n)} \left| \frac{1}{nh^d} \sum_{i=1}^n \left(K_\pi \left(\frac{X_i - x}{h} \right) - K_\pi \left(\frac{X_i - x_j}{h} \right) \right) \right| \\ & \quad + \max_{j=1, \dots, J_n} \sup_{x \in \mathcal{V}_X(x_j, \delta_n)} \left| \int K_\pi(z) (f(x) - f(x_j + hz)) dz \right| \\ & \leq \max_{j=1, \dots, J_n} \left| \frac{1}{nh^d} \sum_{i=1}^n K_\pi \left(\frac{X_i - x_j}{h} \right) - \int K_\pi(z) f(x_j + hz) dz \right| + \frac{2}{3} r_n \epsilon \end{aligned}$$

1 Hence, Eq. (24) holds if

$$3 \mathbb{P} \left(\max_{j=1, \dots, J_n} \left| \frac{1}{nh} \sum_{i=1}^n K_\pi \left(\frac{X_i - x_j}{h} \right) - \int K_\pi(z) f(x_j + hz) dz \right| \geq r_n \frac{\epsilon}{3} \right) \rightarrow 0$$

7 Since an elementary change of variables gives

$$9 \mathbb{E} \left[\frac{1}{h^d} K_\pi \left(\frac{X_i - x_j}{h} \right) \right] = \int K_\pi(z) f(x_j + hz) dz \text{ and}$$

$$11 \text{Var} \left(\frac{1}{h^{d/2}} K_\pi \left(\frac{X_i - x_j}{h} \right) \right) \leq \mathbb{E} \left[\frac{1}{h^d} K_\pi^2 \left(\frac{X_i - x_j}{h} \right) \right] \leq \frac{C}{h^d}$$

15 Equation (24) follows from the Bonferoni inequality and Eq. (23) which give

$$19 \mathbb{P} \left(\max_{j=1, \dots, J_n} \left| \frac{1}{nh^d} \sum_{i=1}^n K_\pi \left(\frac{X_i - x_j}{h} \right) - \int K_\pi(z) f(x_j + hz) dz \right| \geq r_n \frac{\epsilon}{3} \right)$$

$$21 \leq \sum_{j=1}^{J_n} \mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n h^{-\frac{d}{2}} K_\pi \left(\frac{X_i - x_j}{h} \right) - \int K_\pi(z) f(x_j + hz) dz \right| \geq \log^{1/2} n \frac{\epsilon}{3} \right)$$

$$23 \leq 2J_n \exp \left(-\frac{\epsilon^2 \log n}{C + C/(nh^d)^{1/2}} \right) \leq 2 \exp \left(-\frac{\epsilon^2 \log n - b \log n}{C} \right) \rightarrow 0$$

27 for $\epsilon^2 > b$. Hence, Eq. (19) is proved. The existence of the inverse matrix stated in the lemma and the uniform bound for its spectral radius follow from Lemma 11.

31 ■

33 **Proof of Proposition 6.** Let

$$37 \widehat{M}_n(x, h) = \frac{1}{nh^d} \sum_{i=1}^n U \left(\frac{X_i - x}{h} \right) U' \left(\frac{X_i - x}{h} \right) K \left(\frac{X_i - x}{h} \right) \text{ and}$$

$$39 M(x, h) = h^{-d} \mathbb{E} \left[U \left(\frac{X_i - x}{h} \right) U' \left(\frac{X_i - x}{h} \right) K \left(\frac{X_i - x}{h} \right) \right]$$

41

1 We first show that $(nh^d)^{1/2}V(x_n, h)^{-1/2}H\left(\hat{\beta}(x_n; h) - \bar{\beta}(x_n; h)\right)$ is asymptotically
 3 a standard normal. Observe that

$$\begin{aligned}
 & (nh^d)^{1/2}V(x_n, h)^{-1/2}H\left(\hat{\beta}(x_n; h) - \bar{\beta}(x_n; h)\right) \\
 &= V(x_n, h)^{-1/2}\widehat{M}_n(x_n, h)^{-1}\frac{\sum_{i=1}^n U\left(\frac{X_i-x_n}{h}\right)\xi_i(x_n; h)K\left(\frac{X_i-x_n}{h}\right)}{(nh^d)^{1/2}} \\
 &+ V(x_n, h)^{-1/2}\widehat{M}_n(x_n, h)^{-1}\frac{\sum_{i=1}^n U\left(\frac{X_i-x_n}{h}\right)\varepsilon_i K\left(\frac{X_i-x_n}{h}\right)}{(nh^d)^{1/2}}
 \end{aligned}$$

11 where

$$13 \quad \xi_i(x; h) = m(X_i) - U(X_i - x)' \bar{\beta}(x; h)$$

15 Since, for all $x \in \mathcal{X}$, $\bar{\beta}(x; h)$ satisfies the first-order condition

$$17 \quad \mathbb{E}\left[U(X_i - x)\left(m(X_i) - U(X_i - x)' \bar{\beta}(x; h)\right) K\left(\frac{X_i - x}{h}\right)\right] = 0 \quad (25)$$

19 the variables $U\left(\frac{X_i-x_n}{h}\right)\xi_i(x_n; h)K\left(\frac{X_i-x_n}{h}\right)$ are centered. Moreover, Theorems 1
 21 and 2 together with Eq. (21), Assumptions X-(iii) and K and Lemma 5 give that

$$\begin{aligned}
 & \text{Var}\left(h^{-d/2}U(X_i - x)\left(m(X_i) - U(X_i - x)' \bar{\beta}(x; h)\right) K\left(\frac{X_i - x}{h}\right)\right) \\
 &= O(h^{2s}) \int U(z)U(z)'K^2(z)f(x + hz)dz = o(1)
 \end{aligned}$$

29 uniformly in $x \in \mathcal{X}$. Hence,

$$31 \quad V(x_n, h)^{-1/2}\widehat{M}_n(x_n, h)^{-1}\frac{\sum_{i=1}^n U\left(\frac{X_i-x_n}{h}\right)\xi_i(x_n; h)K\left(\frac{X_i-x_n}{h}\right)}{(nh^d)^{1/2}} = o_p(1)$$

33 and it is sufficient to show that

$$35 \quad V(x_n, h)^{-1/2}\widehat{M}_n(x_n, h)^{-1}\frac{\sum_{i=1}^n U\left(\frac{X_i-x_n}{h}\right)\varepsilon_i K\left(\frac{X_i-x_n}{h}\right)}{(nh^d)^{1/2}} \xrightarrow{d} N(0, I_d)$$

39 This follows from Theorem 4, Assumption E, and the Lindeberg Central Limit
 41 Theorem for triangular arrays. To complete the proof of the proposition, it is now
 sufficient to show that $\widehat{V}(x_n, h) = V(x_n, h) + o_p(1)$. This follows from Lemma 5

1 and $\hat{\sigma}^2(x_n) = \sigma^2(x_n) + o_p(1)$ as established now. Let

$$5 \quad \tilde{\sigma}^2(x) = \frac{\sum_{i=1}^n (Y_i - \mu(x))^2 K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad \text{and} \quad \bar{\sigma}^2(x) = \frac{\sum_{i=1}^n \sigma^2(X_i) K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

9 The asymptotic normality above, Theorems 1 and 2, yields that

$$11 \quad |\hat{\sigma}(x_n) - \tilde{\sigma}(x_n)| \leq \left(\frac{\sum_{i=1}^n \left(\hat{\beta}_0(x_n, h) - \mu(x_n) \right)^2 K\left(\frac{X_i - x_n}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x_n}{h}\right)} \right)^{1/2}$$

$$13 \quad = \left| \hat{\beta}_0(x_n, h) - \mu(x_n) \right| = O_p \left(\frac{1}{(nh^d)^{1/2}} + h^s \right) = o_p(1)$$

19

21 Let $\delta_i = \varepsilon_i^2 - \sigma^2(X_i)$. Observe that

$$23 \quad \tilde{\sigma}^2(x_n) = \bar{\sigma}^2(x_n) + O_p \left(\frac{\sum_{i=1}^n (\mu(X_i) - \mu(x_n))^2 K\left(\frac{X_i - x_n}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x_n}{h}\right)} \right)$$

$$25 \quad + O_p \left(\frac{\sum_{i=1}^n \delta_i K\left(\frac{X_i - x_n}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x_n}{h}\right)} \right)$$

29

31 with, by uniform continuity of $\mu(\cdot)$ over the compact \mathcal{X} and since $h \rightarrow 0$,

33

$$35 \quad \frac{\sum_{i=1}^n (\mu(X_i) - \mu(x_n))^2 K\left(\frac{X_i - x_n}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x_n}{h}\right)} = o(1)$$

37

39 Let $\mathbb{E}_n[\cdot]$ be the conditional expectation given X_1, \dots, X_n and $i = \sqrt{-1}$. Then, under Assumption E and assuming w.l.o.g. that $\nu \leq 2$, Assumptions X-(i) and K, and by standard manipulations involving uniform $O(\cdot)$ terms, we have for any t ,

41

$$\begin{aligned}
& \left| \mathbb{E}_n \left[\exp \left(it \frac{\sum_{i=1}^n \delta_i K \left(\frac{X_i - x_n}{h} \right)}{\sum_{i=1}^n K \left(\frac{X_i - x_n}{h} \right)} \right) \right] \right| = \left| \prod_{i=1}^n \mathbb{E}_n \left[\exp \left(it \frac{\delta_i K \left(\frac{X_i - x_n}{h} \right)}{\sum_{i=1}^n K \left(\frac{X_i - x_n}{h} \right)} \right) \right] \right| \\
& = \left| \prod_{i=1}^n \exp \left\{ \ln \left(1 - it \mathbb{E}_n \left[\frac{\delta_i K \left(\frac{X_i - x_n}{h} \right)}{\sum_{i=1}^n K \left(\frac{X_i - x_n}{h} \right)} \right] \right. \right. \right. \\
& \quad \left. \left. \left. + |t|^{1+\nu/2} O \left(\mathbb{E}_n \left[\left| \frac{\delta_i K \left(\frac{X_i - x_n}{h} \right)}{\sum_{i=1}^n K \left(\frac{X_i - x_n}{h} \right)} \right|^{1+\nu/2} \right] \right) \right) \right\} \right| \\
& \leq \exp \left(-C |t|^{1+\nu/2} \sum_{i=1}^n \left(\frac{K \left(\frac{X_i - x_n}{h} \right)}{\sum_{i=1}^n K \left(\frac{X_i - x_n}{h} \right)} \right)^{1+\nu/2} \right) = \exp \left(O_p \left((nh)^{-\nu/2} \right) \right) \xrightarrow{p} 1
\end{aligned}$$

15 This implies that

$$\frac{\sum_{i=1}^n \delta_i K \left(\frac{X_i - x_n}{h} \right)}{\sum_{i=1}^n K \left(\frac{X_i - x_n}{h} \right)} \xrightarrow{d} 0 \text{ and then } \xrightarrow{p} 0$$

21 so that $\hat{\sigma}^2(x_n) = \bar{\sigma}^2(x_n) + o_p(1)$, with $\bar{\sigma}^2(x_n) = \sigma^2(x_n) + o_p(1)$ by uniform
 23 continuity of $\sigma(\cdot)$ over the compact \mathcal{X} . Hence, $\hat{\sigma}^2(x_n) = \sigma^2(x_n) + o_p(1)$. ■

25 **Proof of Proposition 7.** Let $K_\pi(z) = z^\pi K(z)$,

$$\begin{aligned}
27 \quad S_n(x; h) &= \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n (\xi_i(x; h) + \varepsilon_i) K_\pi \left(\frac{X_i - x}{h} \right), \\
29 \quad \xi_i(x; h) &= m(X_i) - U \left(\frac{X_i - x}{h} \right)' H \bar{\beta}(x; h)
\end{aligned}$$

33 Lemma 5 implies that it is sufficient to show that

$$35 \quad \sup_{x \in \mathcal{X}} |S_n(x; h)| = O_p \left(\log^{1/2} n \right) \quad (26)$$

37 Define, for $\eta_i = \varepsilon_i \mathbb{I}(|\varepsilon_i| < \tau_n) - \mathbb{E}[\varepsilon_i \mathbb{I}(|\varepsilon_i| < \tau_n) | X_i]$,

$$41 \quad s_n(x; h) = \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n (\xi_i(x; h) + \eta_i) K_\pi \left(\frac{X_i - x}{h} \right)$$

1 The Chebychev inequality gives, for $\tau_n = C_\tau n^{1/\nu}$,

3
$$\mathbb{P} \left(\max_{i=1, \dots, n} |\varepsilon_i| \geq \tau_n \right) \leq \sum_{i=1}^n \mathbb{P} (|\varepsilon_i| \geq \tau_n) \leq \frac{\mathbb{E} [|\varepsilon_i|^\nu]}{C_\tau^\nu}$$

5

7 which can be made arbitrarily small by increasing C_τ . Arguing as in the proof
of Lemma 5 gives, when $\max_{i=1, \dots, n} |\varepsilon_i| \geq \tau_n$ and since $-\mathbb{E} [\varepsilon_i \mathbb{I} (|\varepsilon_i| < \tau_n) | X_i] =$
9 $\mathbb{E} [\varepsilon_i \mathbb{I} (|\varepsilon_i| \geq \tau_n) | X_i]$,

11
$$\sup_{x \in \mathcal{X}} |S_n(x; h) - s_n(x; h)| = \sup_{x \in \mathcal{X}} \left| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbb{E} [\varepsilon_i \mathbb{I} (|\varepsilon_i| \geq \tau_n) | X_i] K_\pi \left(\frac{X_i - x}{h} \right) \right|$$

13
$$\leq \sup_{x \in \mathcal{X}} \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbb{E} \left[\frac{|\varepsilon_i|^\nu \mathbb{I} (|\varepsilon_i| \geq \tau_n)}{\tau_n^{\nu-1}} | X_i \right] \left| K_\pi \left(\frac{X_i - x}{h} \right) \right|$$

15
$$= O_p \left(\frac{(nh^d)^{1/2}}{n^{1-1/\nu}} \right) = o_p \left(\frac{1}{n^{1/2-1/\nu}} \right) = o_p(1)$$

17

19 since $\nu > 2$ by Assumption E. Therefore, it is sufficient to show that

21
$$\sup_{x \in \mathcal{X}} |s_n(x; h)| = O_p (\log^{1/2} n) \tag{27}$$

23

to prove Eq. (26). Observe that the expectation of the summands in $s_n(x; h)$ is 0
25 by Eq. (25) and the definition of η_i . Equation (25) also gives

27
$$H \bar{\beta}(x; h) = \left(\mathbb{E} \left[U \left(\frac{X_i - x}{h} \right) U \left(\frac{X_i - x}{h} \right)' K \left(\frac{X_i - x}{h} \right) \right] \right)^{-1}$$

29
$$\times \mathbb{E} \left[U \left(\frac{X_i - x}{h} \right) m(X_i) K \left(\frac{X_i - x}{h} \right) \right]$$

31

33 so that Assumption K gives that $\|H (\bar{\beta}(x; h) - \bar{\beta}(x'; h))\| \leq Ch^{-1} \|x - x'\|$ for
all $x, x' \in \mathcal{X}$. This also gives, by Eq. (21) which ensures that $\max_{x \in \mathcal{X}} |\xi_i(x; h)| \leq C$,

35
$$\left\| (\xi_i(x; h) + \eta_i) K_\pi \left(\frac{X_i - x}{h} \right) - (\xi_i(x'; h) + \eta_i) K_\pi \left(\frac{X_i - x'}{h} \right) \right\| \leq C \frac{\tau_n}{h} \|x - x'\|$$

37

39 for all $i = 1, \dots, n$ and all n , all $x, x' \in \mathcal{X}$,

41
$$|s_n(x; h) - s_n(x'; h)| \leq C \frac{n^{1/2} \tau_n}{h^{d/2+1}} \|x - x'\|$$

1 This together Assumptions K, X and $1/h^d = O(n/\log n)$, there is a $\delta = \delta_n = n^{-a}$
 2 such that

- 3 1. There is an integer number $J_n = O(n^b)$, $b > 0$, and some $x_j \in \mathcal{X}$ such that
 4 $\mathcal{X} = \bigcup_{j=1}^{J_n} \mathcal{V}_X(x_j, \delta_n)$, where $\mathcal{V}_X(x_j, \delta_n) = \mathcal{V}(x_j, \delta_n) \cap \mathcal{X}$;
 5 2. For all x, x' with $\|x - x'\| \leq \delta_n$ and all i and n , $|s_n(x; h) - s_n(x'; h)| \leq 1$.

6 Hence,

$$\begin{aligned} 7 \sup_{x \in \mathcal{X}} |s_n(x; h)| &\leq \max_{j=1, \dots, J_n} |s_n(x_j; h)| + \max_{j=1, \dots, J_n} \sup_{x \in \mathcal{V}_X(x_j, \delta_n)} |s_n(x; h) - s_n(x_j; h)| \\ 8 &\leq \max_{j=1, \dots, J_n} |s_n(x_j; h)| + 1 \end{aligned}$$

9 As a consequence, Eq. (27) holds if $\max_{j=1, \dots, J_n} |s_n(x_j; h)| = O_p(\log^{1/2} n)$.
 10 The Bonferoni inequality and Eq. (23) give, using $\max_{x \in \mathcal{X}} |\xi_i(x; h)| \leq C$ and
 11 $h^{-d} = O(n^{1-2/\nu} \log n)$,

$$\begin{aligned} 12 \mathbb{P} \left(\max_{j=1, \dots, J_n} |s_n(x_j; h)| \geq t \log^{1/2} n \right) &\leq \sum_{j=1}^{J_n} \mathbb{P} (|s_n(x_j; h)| \geq t \log^{1/2} n) \\ 13 &\leq 2J_n \exp \left(- \frac{t \log n}{C + \left(\frac{h^{-d} \tau_n^2 \log n}{n} \right)^{1/2} t} \right) = \exp \left(- \frac{(t - C) \log n}{C'} \right) \rightarrow 0 \end{aligned}$$

14 for $t > C$. This ends the proof of Proposition 7. ■

28 7.4. Propositions 8, 9 and 10

29 **Proof of Proposition 8.** Lemma 5, Theorems 1 and 2 which give Eq. (21), imply
 30 that the conclusions of Lemmas A.1, A.2 and A.3 in Guerre and Sabbah (2012)
 31 are true when $x \in \mathcal{X}$. Hence, the first equation in Proposition 8 follows from
 32 minor modifications of the proof of Theorem 2 in Guerre and Sabbah (2012).
 33 The second equation follows from $\max_{(\alpha, x) \in [\underline{q}, \bar{q}] \times \mathcal{X}} \|\widehat{J}(\alpha|x, h) - \bar{J}(\alpha|x, h)\| =$
 34 $O_{\mathbb{P}} \left((\log n / (nh^d))^{1/2} \right)$ and $\max_{(\alpha, x) \in [\underline{q}, \bar{q}] \times \mathcal{X}} \|\widehat{S}(\alpha|x, h)\| = O_p(\log^{1/2} n)$ as estab-
 35 lished below. ■

36 **Proof of Proposition 10.** Theorems 1 and 2 which give Eq. (21) are sufficient to
 37 show that the conclusion of Lemma A.3 in Guerre and Sabbah (2012) holds for
 38 $x \in \mathcal{X}$, that is

$$\max_{(\alpha, x) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}} \|\widehat{S}(\alpha|x, h)\| = O_p(\log^{1/2} n)$$

Hence, Proposition 8 and Lemma 5 give

$$\begin{aligned} & \sup_{(\alpha, x) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}} \left\| (nh^d)^{1/2} H \left(\widehat{\beta}(\alpha|x, h) - \bar{\beta}(\alpha|x, h) \right) \right\| \\ &= O_p \left(\max_{(\alpha, x) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}} \left\| \frac{\widehat{S}(\alpha|x, h)}{(nh^d)^{1/2}} \right\| \right) + O_p \left(\left(\frac{\log n}{nh^d} \right)^{3/4} \right) = O_p \left(\left(\frac{\log n}{nh^d} \right)^{1/2} \right) \end{aligned}$$

Hence, Theorems 1 and 2 show that the Proposition is proved. \blacksquare

Proof of Proposition 9. As in the proof of Proposition 6, the key issue here is to show that $\widehat{V}(\alpha|x_n, h) = V(\alpha|x_n, h) + o_p(1)$, and Lemma 5 shows that is sufficient to show that $\partial \widehat{Q}(\alpha|x_n) / \partial \alpha$ is consistent. This follows from Proposition 10, the fact that $s \geq 1$ and the choice of η which gives under Assumption F which ensures that $(\alpha, x) \mapsto \partial Q(\alpha|x) / \partial \alpha$ is continuous,

$$\begin{aligned} \frac{\partial \widehat{Q}(\alpha|x_n)}{\partial \alpha} &= \frac{\widehat{Q}(\alpha + \eta|x_n) - \widehat{Q}(\alpha - \eta|x_n)}{2\eta} \\ &= \frac{Q(\alpha + \eta|x_n) - Q(\alpha - \eta|x_n)}{2\eta} + O_p \left(\frac{1}{\eta} \left(h + \left(\frac{\log n}{nh^d} \right)^{1/2} \right) \right) \\ &= \frac{\partial Q(\alpha|x_n)}{\partial \alpha} + o_p(1) \end{aligned}$$

NOTES

1. We refer interested readers to Gragg and Tapia (1974) for a complete statement of the Newton–Kantorovich Theorem and to Lemma 7.2 of this paper for the part used in this paper.

2. When $m(x_1) = \sin(2\pi x_1)$, $\mathbb{E}[m(X_2)\mathbb{I}(X_2 \in [c, 1-c])] = 0$ so that $\hat{\mu}_{1,c}(x_1)$ could also be used. However identifying restrictions such as $m(1/2) = 0$ seems to be more popular and does not involve the trimming parameter. Unreported simulation results suggest however that $\hat{\mu}_{1,c}(\cdot)$ may have better performances than $\widehat{m}_c(\cdot)$.

ACKNOWLEDGMENTS

We thank Ruixuan Liu for drawing our attention to Chen and Wu (2013) and Shih-Tang Hwu for pointing out an error in a preliminary draft. All remaining

1 errors are ours. Yanqin Fan is grateful to Aman Ullah for introducing him to
 3 nonparametric econometrics and for his continued support throughout his career.

5 REFERENCES

- 7 Altonji, J. G., Ichimura, H., & Otsu, T. (2012). Estimating derivatives in nonseparable models with
 9 limited dependent variables. *Econometrica*, *80*, 1701–1719.
- AQ:6 Aryal, G., Gabrielli, M. F., & Vuong, Q. (2014). *Semiparametric estimation of first-price auction*
 8 *models*. Working Paper.
- AQ:7 Banerjee, A. N. (2007). A method of estimating the average derivative. *Journal of Econometrics*, *136*,
 11 65–88.
- Bouezmarni, T., & Rombouts, J. (2010). Nonparametric density estimation for multivariate bounded
 13 data. *Journal of Statistical Planning and Inference*, *140*, 139–152.
- Bravo, F., & Jacho-Chávez, D. T. (2011). Empirical likelihood for efficient semiparametric average
 15 treatment effects. *Econometric Reviews*, *30*, 1–24.
- Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local bahadur
 17 representation. *The Annals of Statistics*, *11*, 760–777.
- Chaudhuri, P., Doksum, K., & Samarov, A. (1997). On average derivative quantile regression.
 19 *The Annals of Statistics*, *25*, 715–744.
- Chen, M.-Y., & Wu, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation.
 21 *Journal of the American Statistical Association*, *108*, 1421–1434.
- Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall/CRC.
- 23 Fan, J., Heckman, N. E., & Wand, M. P. (1995). Local polynomial kernel regression for generalized
 25 linear models and quasi-likelihood functions. *Journal of the American Statistical Association*,
 90, 141–150.
- Fan, Y., & Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric
 27 functional forms. *Econometrica*, *64*, 865–890.
- Frandsen, B. R., Frölich, M., & Melly, B. (2010). *Quantile treatment effects in the regression*
 29 *discontinuity design*. Working Paper.
- Gragg, W. B., & Tapia, R. A. (1974). Optimal error bounds for the Newton–Kantorovich Theorem.
 31 *SIAM Journal on Numerical Analysis*, *11*, 10–13.
- Gu, J., Li, Q., & Yang, J.-C. (2015). Multivariate local polynomial kernel estimators: Leading bias and
 33 asymptotic distribution. *Econometric Reviews*, *34*, 979–1010.
- Guerre, E. (2000). Design adaptive nearest neighbor regression estimation. *Journal of Multivariate*
 35 *Analysis*, *75*, 219–244.
- Guerre, E., & Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in
 37 regression models. *Econometric Theory*, *18*, 1139–1171.
- Guerre, E., & Sabbah, C. (2012). Uniform bias study and Bahadur representation for local polynomial
 39 estimators of the conditional quantile function. *Econometric Theory*, *28*, 87–129.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects
 41 with a regression-discontinuity design. *Econometrica*, *69*, 201–209.
- Härdle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits.
The Annals of Statistics, *21*, 1926–1947.
- Härdle, W., & Stoker, T. H. (1989). Investigating smooth multiple regression by the method of average
 derivatives. *Journal of the American Statistical Association*, *84*, 986–995.
- Hickman, B. R., & Hubbard, T. P. (2015). Replacing sample trimming with boundary correction in
 nonparametric estimation of first-price auctions. *Journal of Applied Econometrics*, *30*, 739–762.


- 1 Hjellvik, J., Yao, Q., & Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference*, 68, 295–321.
- 3 Hoderlein, S., Su, L., White, H., & Yang, T. T. (2015). [Testing for monotonicity in unobservables under unconfoundedness. Working Paper](#)
- 5 Huber, P. J. (1967). The behavior of maximum likelihood estimation under nonstandard conditions. *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1). University of California Press.
- 7 ~~Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615–635.~~
- 9 Imbens, G. W., & Zajonc, T. (2011). **Regression discontinuity design with multiple forcing variables.** Working Paper.
- Kong, E., Linton, O., & Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, 26, 1529–1564.
- 11 Li, C.-S. (2005). Using local linear kernel smoothers to test the lack of fit of nonlinear regression models. *Statistical Methodology*, 2, 267–284.
- 13 Li, Q., Lu, X., & Ullah, A. (2003). Multivariate local polynomial regression for estimating average derivatives. *Journal of Nonparametric Statistics*, 15, 607–624.
- 15 Linton, O., & Nielsen, J. P. (1995). A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration. *Biometrika*, 82, 93–100.
- 17 Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17, 571–599.
- 19 Müller, H. G., & Stadtmüller, U. (1999). Multivariate boundary kernels and a continuous least squares principle. *Journal of the Royal Statistical Society B*, 61, 439–458.
- Newey, W., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55, 819–847.
- 21 Newey, W. K., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55, 819–847.
- 23 Oka, T. (2009). *Nonparametric quantile regression with discontinuities*. Working Paper.
- Porter, J. (2003). Estimation in the regression discontinuity model. [Manuscript, Harvard University](#)
- 25 Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57, 1403–1430.
- 27 Qu, Z., & Yoon, J. (2014). [Nonparametric estimation and inference on conditional quantile processes. Working Paper](#)
- 29 Ruppert, D., & Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22, 1346–1370.
- Su, L., & Ullah, A. (2008). Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics*, 144, 193–218.
- 31 Su, L., & White, H. (2012). Conditional independence specification testing for dependent processes with local polynomial quantile regression. *Advances in Econometrics*, 29, 355–434.
- 33 Su, L., & Xiao, Z. (2009). Testing structural change in conditional distribution. Unpublished Manuscript.
- 35 White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.

37

39

41

AUTHOR QUERY FORM

	<p>Book: AECO-V036-3611227</p> <p>Chapter: CH011</p>	<p>Please e-mail or fax your responses and any corrections to:</p> <p>E-mail:</p> <p>Fax:</p>
---	--	--

Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

Disk use

Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:



- Scanning (parts of) your article Rekeying (parts of) your article
 Scanning the artwork







Bibliography

If discrepancies were noted between the literature list and the text references, the following may apply:

- The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.**
- UNCITED REFERENCES: This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.**

Queries and/or remarks

Location of Article	Query / remark	Response
AQ1:	The right running head is amended to fit the space. Please check the amendment for correctness.	
AQ2:	The reference citations 'Härdle and Stocker (1989) or Powell, Stock, and Stocker (1989)' have been changed to 'Härdle and Stoker (1989) or Powell, Stock, and Stoker (1989)' to match with the reference list. Please check.	

1 3 5	AQ3:	The following reference is cited but not provided in the reference list: Powell and Newey (1987). Kindly provide complete reference details for this reference.	
7 9	AQ4:	Figs. 1–3 have been submitted as colour images; however, the captions have been reworded to ensure that they are meaningful when your article is reproduced both in colour and in black and white. Please check and correct if necessary.	
11	AQ5:	Please check the citation of Fig. 3.	
13 15	AQ6:	Please note the following references were not cited in text: Imbens and Lemieux (2008). Kindly cite them, if there are no citations then they need to be removed from the list.	
17 19	AQ7:	Please provide working paper number for the references 'Aryal, Gabrielli, & Vuong (2014); Frandsen, Frölich, & Melly (2010); Hoderlein, Su, White, & Yang (2015); Imbens, & Zajonc (2011); Oka (2009); Qu, & Yoon (2014)'.	
21	AQ8:	Please provide complete details for the reference 'Porter (2003)'.	
23	AQ9:	Please check the settings and numbering of all enunciations (e.g. Lemma, Theorem, Proposition) for correctness.	

25

27

29

31

33

35

37

39

41