ESTIMATING MONOTONE INDEX MODELS WITH NONPARAMETRIC CONTROLS

KOEN JOCHMANS*

CORE AND BROWN UNIVERSITY

JOB-MARKET PAPER

ABSTRACT

This paper considers estimation of the coefficient vector in a semiparametric monotone index model where one needs to condition on control variables to deal with unobserved heterogeneity. Examples that fit this framework are weakly-separable models with sample selection, triangular endogeneity, or a partially-linear index specification. The proposed estimator is based on a local ranking of the observations, given nonparametric estimates of the controls. Rank estimation is conceptually elegant, demands mild shape restrictions that can readily follow from an economic model, and offers robustness against contamination of the data. At the same time, it does not require estimating nuisance functions. Sufficient conditions are given under which the estimator converges to a Gaussian process at the parametric rate. In doing so, distribution theory is derived for two-step estimators that does not require the objective function to be differentiable. These results should prove complementary to the asymptotic theory that underlies the estimators derived from smooth moment conditions. The theory is also generalized to cover three-step estimators whose criterion function depends on the local-rank estimator, by deriving an estimator of a nonparametric transformation model. Simulation experiments serve to illustrate the implementation of the procedure and to evaluate its small-sample effectiveness.

JEL CLASSIFICATION: C14, C34, C35, C41
KEYWORDS: control variable, kernel weight, local-rank estimator, index model, semi- and nonparametric estimation, shape restrictions, two-step and three-step estimation, transformation model, U-process, weighted step function.

REVISIONS: first version: April—May 2010; minor revision: June 2010; this version: November 15, 2010.
Please consult http://sites.google.com/site/koenjochmans/ to ensure that you are in possession of the most recent version.

*Address: Center for Operations Research and Econometrics, U.C. Louvain, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. Tel. +32 10 474329. E-mail: koen.jochmans@uclouvain.be. Or: Department of Economics, Brown University, Box B, Providence, RI 02912, U.S.A. Tel. +1 401 863 2970. E-mail: koen_jochmans@brown.edu.

I INTRODUCTION

There is now a large literature on semiparametric estimation of econometric models. Motivated by a desire to relax functional-form assumptions while simultaneously circumventing the curse of dimensionality, modeling approaches featuring index restrictions are particularly widespread; see Stoker (1986), Powell, Stock, and Stoker (1989), and Ichimura (1993) for seminal work, Powell (1994) and Horowitz (2009) for overviews, and Horowitz and Lee (2002) for a critical assessment of such approaches. A particular class of estimators for these models is based on pairwise comparisons of observations. Rank-based estimation techniques fall into this category, as do pairwise-differencing methods. Inference procedures based on ranks are known to have superior robustness properties over other methods such as (semiparametric) least squares or maximum likelihood, for example. At the same time, their implementation does not require the estimation of nuisance functions or the choice of smoothing and trimming parameters. Examples of rank estimators can be found in Han (1987), Cavanagh and Sherman (1998), and Khan and Tamer (2007), among others.

In many important problems, conditioning on control variables—or, simply, controls for short—is required to obtain moment conditions that identify the parameters of interest. In several such situations, a prior estimation step is needed to construct an empirical counterpart to these identifying restrictions. A prime example of this scenario is the instrumental-variable estimation of a linear structural equation with endogenous regressors via two-stage least-squares; another one is the estimation of a linear model in the presence of sample selection through the familiar Heckman (1979) procedure. This type of methods has been extended to semiparametrically-specified nonlinear models. Ai and Chen (2003), for example, have derived theory for GMM estimators defined by conditional moment restrictions. For pairwise-differencing estimators, Aradillas-López, Honoré, and Powell (2007) have extended the work by Ahn and Powell (1993) and Honoré and Powell (1994) to allow for the inclusion of controls into a class of estimators for nonlinear models that is defined through assumptions of concavity and smoothess on the associated criterion functions.

Here, I consider including nonparametrically-specified controls into rank estimators. I will focus on a modification of a class of estimators proposed by Cavanagh and Sherman (1998). However, the distribution theory will apply more generally to estimators that maximize U-processes of order two; see Jochmans (2010) for details. The resulting estimator is based on a local ranking of observations, which is to be understood as ranking only those observations whose controls are approximately equal. It can be applied to many popular econometric models, including binary-choice models, censored regression models, and transformation models. Because of the step-function nature of the objective function that is inherent to any estimator based on ranks, the large-sample behavior of its maximizer can not be readily established using the results available in the literature. Therefore, I derive distribution theory that takes into account the influence of the first-step nonparametric estimation error while allowing the maximand to be non-differentiable. The theory builds on the work by Sherman (1993,1994a,1994b) and generalizes the aforementioned results of Aradillas-López, Honoré, and Powell (2007). This extension is not merely technical, as the requirements of smoothness and concavity in the latter paper can substantially restrict the scope of the approach if one is unwilling to impose additional parametric structure; see, for example, Honoré and Powell (2005) for a discussion.

The work that is most closely related to what follows is Blundell and Powell (2004), who suggest a three-step estimator for the index coefficients in a semiparametric binary-choice model with triangular endogeneity. While their procedure is also applicable to a more general class of models, the rank approach advocated here essentially sidesteps the need for their second estimation step, which involves a nonparametric regression on both the covariates and nonparametrically generated regressors. In addition to this and the other favorable properties of rank procedures mentioned earlier, the local-rank estimator builds on weaker shape and smoothness restrictions. The theory laid out below is useful in a variety of applications because such weak shape restrictions often follow under mild assumptions that do not involve a parametric specification of link functions or of the distribution of the model's latent components. The structural dynamic-optimization problem of Hong and Shum (2010), for example, is open to local-rank estimation. In their model, the relevant shape restriction involves the first-order conditions to a utility-maximization problem and follows directly from economic principles.

The paper proceeds as follows. I first state the general form of the model and argue for its usefulness by means of three examples. Next, the local-rank estimator is introduced and intuition for its form is provided. An analysis of its large-sample properties follows. Conditions are given under which the estimator is consistent and converges at the parametric rate. The limiting distribution is derived and a consistent estimator of its variance is given. I then turn to the estimation of other parameters of interest in a third estimation round. The usefulness of such an additional estimation step is motivated using a nonparametric transformation model, and asymptotic theory for this estimator is obtained. The paper ends with an overview of results from Monte

Carlo experiments in models with triangular endogeneity or sample selection. Three appendices contain intermediate lemmata, proofs, and a description of the optimization routine used to compute the local-rank estimator.

II THE MODEL AND MOTIVATING EXAMPLES

Let the vector of observable random variables $D \equiv (Y, X, E, Z)$ have distribution P. Define the vector-valued function

$$\vartheta_0(E, Z) \equiv \mathrm{g}(E) - \mu_{\mathrm{a}(E)}(Z)$$

for chosen functions $\mathrm{g} : \mathscr{R}^{\dim(E)} \rightarrow \mathscr{R}^{\dim(\vartheta_0)}$ and $\mathrm{a} : \mathscr{R}^{\dim(E)} \rightarrow \mathscr{R}^{\dim(\vartheta_0)}$; $\mu_{\mathrm{a}(E)}(z)$ denotes the expectation of $\mathrm{a}(E)$, given $Z = z$, and $\dim(A)$ refers to the dimension of a vector A. Suppose that the outcome variable, Y, is a scalar whose mean conditional on realizations of the covariates X and $\vartheta_0(E, Z)$ depends on X only through the linear index $X'\theta_0$. Some overlap between X and (E, Z) will be allowed. Then, in line with the conventional notation for multiple-index models (see, e.g., Ichimura and Lee, 1991),

$$Y = \mu_Y\big(X'\theta_0, \vartheta_0(E, Z)\big) + \xi \tag{2.1}$$

for a conditional-mean function $\mu_Y : \mathscr{R}^{1+\dim(\vartheta_0)} \rightarrow \mathscr{R}$ and a disturbance ξ that has mean zero given realizations of X and $\vartheta_0(E, Z)$. This model is semiparametric because the functions μ_Y and ϑ_0 are unknown. Our intention is to infer θ_0 (up to normalizations) from a random sample without imposing additional parametric structure. The main statistical restriction that will be maintained is a weak-monotonicity condition. It will be assumed that $\mu_Y(x'\theta_0, \vartheta)$ is nondecreasing and nonconstant in $x'\theta_0$ for each value ϑ of the control.

Two special cases of particular interest are covered by these assumptions. The first one has the outcome variable generated as

$$Y = f\big(X'\theta_0 + g[\vartheta_0(E, Z)], U\big), \quad U \perp (X, E, Z), \tag{2.2}$$

where U is a latent disturbance vector, $f : \mathscr{R}^{1+\dim(U)} \rightarrow \mathscr{R}$ is weakly increasing in its first argument, and $g : \mathscr{R}^{\dim(\vartheta_0)} \rightarrow \mathscr{R}$ is smooth, but f and g are otherwise left unspecified. This is a partially-linear-index formulation of the generalized regression model introduced by Han (1987) and, in the absence of $\vartheta_0(E, Z)$, it can be seen to cover many popular models; standard versions of binary-choice models, censored regression models, and duration models are a few examples. The model as specified in (2.2) can be understood to extend Han's in a manner analogous to how Robinson (1988) modified the linear regression model to allow some covariates to affect the outcome variable in a nonparametrically-specified way. Besides the nonlinearity, another difference with Robinson's model is that $\vartheta_0(E,Z)$ depends on the unobserved conditional expectation $\mu_{\mathrm{a}(E)}(Z)$, which will have to be estimated, leading to the presence of generated regressors. If the distribution of U is parametrically specified, this model fits the framework of Aradillas-López, Honoré, and Powell (2007) for various f.

A second special case arises on modifying (2.2) to

$$Y = f(X'\theta_0, U), \quad U \perp X | \vartheta_0(E,Z) = \vartheta \text{ for all } \vartheta, \tag{2.3}$$

which allows $\vartheta_0(E,Z)$ to influence the distribution of Y through the distribution of U. This formulation captures models with nonparametric control functions and can be of use when estimating certain simultaneous-equation systems such as a generalized regression model with sample selection or with endogenous covariates, bringing us closer to the main points of focus in Ahn and Powell (1993) and Blundell and Powell (2004); see the examples below. In all these cases, avoiding parametric specifications on f and the error distribution circumvents several instances of misspecification and allows for generality in the nature of the process under consideration.

To motivate the generic structure of the model it is useful to sketch some situations of practical interest that can be cast into it.

Example 1 (decisions based on expectations)**.** *Suppose that agents choose Y based on observable characteristics X and on their expectations about the realization of a random variable E, given realizations of Z. Unless X and Z are independent, ignoring the effect of expectations on outcomes will generally cause inferential statements on the impact of X to be biased. However, provided that observations on E are available, $\mu_E(z)$ is nonparametrically identified and estimable, and can thus be conditioned upon by the econometrician.*

One potential application of this framework is in market-entry models with few players, where a firm's decision to enter the market depends on its anticipation of the other players' decision ; see, e.g., Berry (1992) and Aradillas-López (2010). Accounting for the impact of expectations on outcomes can be traced back at least as to Manski (1991), who considered the estimation of preference parameters in a parametric discrete-choice setting; see also Ahn and Manski (1993) and Ahn (1997).

The next example is a nonlinear model with sample selection.

Example 2 (endogenous sample selection)**.** *Suppose that Y^* is determined through a model of the form $Y^* = f(X'\theta_0, U_2)$ but that only $Y \equiv EY^*$ is observed along with (X, E, Z), where*

$$E = 1(g(Z) \geq U_1), \quad (U_1, U_2) \perp (X, Z), \quad \text{and the cdf } F_{U_1} \text{ is strictly increasing on } \mathscr{R}.$$

Here, the mean of Y given $(X, E, Z) = (x, 1, z)$ varies with z. Conditioning on $Z = z$ can make the identification and estimation of θ_0 troublesome if X and Z have elements in common. However, independence implies that $U_2|(X, E, Z) = (x, 1, z) \overset{\mathcal{L}}{=} U_2|g(z) \geq U_1$, so that U_2 is i.i.d. given $g(Z) = g(z)$ and $E = 1$. The invertibility of F_{U_1} implies that conditioning on $g(Z) = g(z)$ is identical to conditioning on $\mu_E(Z) = \mu_E(z)$, as $g(z) = F_{U_1}^{-1}(\mu_E(z))$. This leads to a double-index formulation for the mean of Y given realizations of (X, Z) in the subpopulation with $E = 1$, with indices $X'\theta_0$ and $\mu_E(Z)$.

Self selection is a common worry when dealing with microeconomic data; see, e.g., Olley and Pakes (1996) and Borjas (1987) for relevant and well-known applications. The pioneering contributions on the estimation of parametric linear sample-selection models are Gronau (1973) and Heckman (1974, 1979). Various semiparametric alternatives have since then been formulated. A recent suggestion is in Newey (2009), which also contains further references. The approach most closely related to what follows is Ahn and Powell (1993), who derived a pairwise-differenced least-squares estimator while allowing for a nonparametric selection equation as in Example 2. A nonparametric proposal for a model with an additively-separable disturbance was made by Das, Newey, and Vella (2003). In the context of nonlinear models, however, the issue has received relatively little attention; some nonparametric identification results, conditional on selection, are given in Newey (2007).

The final illustration concerns endogeneity bias as induced by simultaneity or by measurement error in covariates, for example.

Example 3 (triangular endogeneity)**.** *Assume that the outcome variable is generated as $Y = f(X'\theta_0, U_2)$, where X partitions as $(X_1', E')'$. Suppose that E depends on $Z = (X_1', X_2')'$ through*

$$E = \mu_E(Z) + U_1,$$

and that the distributional exclusion restriction $U_2|(X, E, Z) = (x, e, z) \overset{\mathcal{L}}{=} U_2|U_1 = u_1$ holds. Then U_2 and E are dependent through their dependence on U_1, rendering a single-index-based estimator inconsistent. Here, $\vartheta_0(E, Z) = E - \mu_E(Z) = U_1$ has an interpretation as an omitted variable. It follows that the conditional mean of Y has a multiple-index representation as in (2.1).

Endogeneity remains a pervasive problem in models with non-additive disturbances; Chesher (2007) outlines some of the most recent attacks. Example 3 is essentially an application of the control-function approach to identification in simultaneous-equation models as put forward by Smith and Blundell (1986) and many others in a parametric setting, by Blundell and Powell (2004) in a semiparametric binary-choice model, and by Chesher (2003) and Imbens and Newey (2009) in a fully nonparametric framework. An extensive overview, a thorough discussion, and many more references are provided by Blundell and Powell (2003).

III LOCAL-RANK ESTIMATION

This section introduces a local-rank estimator to learn about θ_0 from a random sample of observations from P. As with virtually all semiparametric approaches, we will at best be able to identify and estimate θ_0 up to normalizations. Therefore, a constant term is excluded from X and, hereafter, θ_0 refers to its versor, i.e., $\theta_0/\|\theta_0\|$, where $\|\cdot\|$ will be used to indicate both the Euclidean norm and the matrix norm.

To describe the estimator, denote the data by $\{D_i\}_{i=1}^n$, let $V_i \equiv (Y_i, X_i)$, and let $W_i \equiv (E_i, Z_i)$. For a deterministic function $\mathrm{m} : \mathscr{R} \to \mathscr{R}$ that is increasing on $\mathscr{R}$ and for each θ in $\Theta \equiv \{\theta \in \mathscr{R}^{\dim(X)} : \|\theta\| = 1\}$, define the score contribution of the pair of observations (i, j) as

$$\mathrm{s}(V_i, V_j, \theta) \equiv \mathrm{m}(Y_i)\ 1(X_i'\theta > X_j'\theta) + \mathrm{m}(Y_j)\ 1(X_i'\theta < X_j'\theta). \tag{3.1}$$

Also, let $\widehat{\vartheta}(w)$ indicate a nonparametric estimator of $\vartheta_0(w)$. The proposed estimator of θ_0, then, is defined as

$$\widehat{\theta} \equiv \arg\max_{\theta\in\Theta} \widehat{\mathrm{q}}_n(\theta),$$

where the objective function is the following 'weighted average' of score contributions

$$\widehat{\mathrm{q}}_n(\theta) \equiv \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{i<j} \frac{\mathrm{s}(V_i, V_j, \theta)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}} \mathrm{k}\Big(\frac{\widehat{\vartheta}(W_i) - \widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}}\Big) \mathrm{t}(Z_i)\mathrm{t}(Z_j). \tag{3.2}$$

Here, $\mathrm{k} : \mathscr{R}^{\dim(\vartheta_0)} \to \mathscr{R}$ is a chosen symmetric kernel function and σ_{k} is an associated (scalar) bandwidth that goes to zero as n grows large. The function $\mathrm{t} : \mathscr{R}^{\dim(Z)} \to \mathscr{R}^+$ serves to trim away observations for which $\widehat{\vartheta}(w)$ is an unreliable estimator; the necessity for its inclusion will become clear below.

The estimator has an interpretation that explains its form. Monotonicity implies that a ranking of the conditional expectation of $\mathrm{m}(Y)$ given realizations of X and $\vartheta_0(W)$

allows one to deduce an ordering on the associated indices $X'\theta_0$.[1] Moreover,

$$\mu_{\mathrm{m}(Y)}\big(x_1'\theta_0, \vartheta\big) > \mu_{\mathrm{m}(Y)}\big(x_2'\theta_0, \vartheta\big) \Rightarrow x_1'\theta_0 > x_2'\theta_0, \tag{3.3}$$

in obvious notation. Consequently, one fruitful approach to learning about θ_0 is choosing as an estimate that value that best mimics (3.3) in the sample.[2] Denote the density of Y given $(X, \vartheta_0(W))$ by $\mathrm{p}_{Y|X,\vartheta_0(W)}$. Because

$$\mu_{\mathrm{m}(Y)}\big(x'\theta_0, \vartheta\big) = \int \mathrm{m}(y)\ \mathrm{p}_{Y|X,\vartheta_0(W)}(y|x, \vartheta)\ \mathrm{d}y,$$

it is easy to see that the expected score function in (3.1) is concordant with (3.3) when evaluated at θ_0, but not necessarily at any other $\theta \in \Theta$. As the model only enforces an ordering conditional on the control, the score function should be fed only pairs (i, j) for which $\vartheta_0(W_i) - \vartheta_0(W_j)$ lies in a shrinking neighborhood of zero. In (3.2), this is achieved by means of the kernel weights, with nonparametric estimates of the control substituting for the unobserved $\vartheta_0(W_i)$ and $\vartheta_0(W_j)$.

The data-dependent weighting constitutes the main difference with the estimator advocated by Cavanagh and Sherman (1998), whose objective function is recovered from (3.2) on assigning the same weight to each pair of observations in the sample. Accordingly, $\widehat{\theta}$ has an interpretation as a two-step local implementation of their approach to estimating monotone single-index models. Conditions under which this procedure leads to asymptotically-valid inferential statements about θ_0 will be given below.

Before plunging into the large-sample theory, however, a word on the function m. While an obvious choice would be simply to set $\mathrm{m}(Y) = Y$, its presence is not vacious. One attractive feature of specifying $\widehat{\theta}$ in terms of general m is that it covers more robust choices. It is well known that estimation based solely on ranks can enjoy a larger degree of robustness than do other methods, although this robustness will typically come at a cost in terms of efficiency loss.[3] Choosing m thus allows to strike a certain

[1]The ordering need not be complete, as the reverse statement in (3.3) only holds under a strengthening of the assumption of weak monotonicity to invertibility. This would lead us back to what is essentially Blundell and Powell's (2004) model.

[2]The need for location and scale normalizations manifests itself here, as (3.3) conveys no information on an intercept term and continues to be satisfied for all positive-scalar multiples of θ_0. For a discussion on the need for normalizations in semiparametric estimation, see Horowitz (2009).

[3]Not surprisingly, rank estimators generally do not achieve semiparametric efficiency bounds such as those derived for binary-choice models and censored regression models by Chamberlain (1986) and Cosslett (1987), or the efficiency bound for the single-index model computed by Newey (1990) (as cited by Ichimura, 1993). Nevertheless, efficiency can be improved by a weighting approach. Moreover, Subbotin (2008b) has demonstrated that properly-weighted versions of rank estimators can achieve the

balance between robustness and efficiency. To illustrate, the influence of outliers could be dampened by setting $\mathrm{m}(Y) = \underline{y}\ 1(Y < \underline{y}) + Y\ 1(\underline{y} \leq Y \leq \overline{y}) + \overline{y}\ 1(\overline{y} < Y)$ for chosen bound values $\underline{y}$ and $\overline{y}$.[4] As a limiting case, inference about θ_0 could be based solely on the sign of Y.

IV LARGE-SAMPLE PROPERTIES

Let $\mathbb{P}$ be the product measure $\mathrm{P} \otimes \mathrm{P}$ on the product space $\mathrm{supp}(D) \otimes \mathrm{supp}(D)$. By analogy to P, define P_n as the empirical measure generated by independent sampling from P and define $\mathbb{P}_n$ similarly, that is, as the random probability measure that places mass $1/n(n-1)$ on each ordered pair (D_i, D_j). Following the notational conventions from the literature on empirical processes, write $\mathrm{P}[f(\cdot)] = f(\mathrm{P})$ for the expectation of a measurable function f under P. Similarly, refer to the expectation under the product measure as $\mathbb{P}[f(\cdot,\cdot)] = f(\mathrm{P},\mathrm{P})$.

Observe that, for each θ in Θ, $\widehat{\mathrm{q}}_n(\theta)$ is a U-statistic of order two; exploit symmetry to write it compactly as

$$\widehat{\mathrm{q}}_n(\theta) = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{i<j} \widehat{\mathrm{h}}(D_i, D_j, \theta) = \mathbb{P}_n[\, \widehat{\mathrm{h}}(\cdot,\cdot,\theta)].$$

A large block of the available distribution theory for estimators defined as maximizers of U-processes was derived by Sherman (1993, 1994b); see also Pollard (1984) and Arcones and Giné (1993). The problem here differs from his setup in two important respects. The first is the presence of kernel weights in $\widehat{\mathrm{q}}_n(\theta)$, the second is its dependence on a first-step nonparametric estimator.

While, in principle, any nonparametric estimator—such as, inter alia, series-, nearest-neighbor-, or locally-linear regression—could be used to form the weights, here, I work

semiparametric efficiency bound for certain models, including the nonlinear regression model and the binary-choice model. Presumably, a similar argument can be applied here. Croux and Dehon (2010) study robustness and efficiency of rank-based measures of statistical association.

[4] Cavanagh and Sherman (1998) also discussed the use of the rank of Y_i, that is, $\mathrm{m}(Y_i) = \sum_{k=1}^{n} 1(Y_i > Y_k)$; see also Sherman (1994b). Here, the use of the rank function would translate into an objective function of the form

$$\frac{1}{6}\binom{n}{3}^{-1} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq i,j} \frac{1(Y_i > Y_k)\ 1(X_i'\theta > X_j'\theta)}{\sigma_{\mathrm{k}}^{2\dim(\vartheta_0)}} \mathrm{k}\Big(\frac{\widehat{\vartheta}(W_i) - \widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}}\Big) \mathrm{k}\Big(\frac{\widehat{\vartheta}(W_i) - \widehat{\vartheta}(W_k)}{\sigma_{\mathrm{k}}}\Big).$$

Notice the additional weighting to ensure that Y_i is ranked only relative to observations k for which $\widehat{\vartheta}(W_k) - \widehat{\vartheta}(W_i)$ converges to zero as $n \to \infty$. While I restrict m to be deterministic, ruling out the rank function, this case could be dealt with under suitable modifications to the arguments that follow.

with a kernel estimator. It takes the form

$$\widehat{\vartheta}(w) \equiv \mathrm{g}(e) - \widehat{\mu}_{\mathrm{a}(E)}(z) = \mathrm{g}(e) - \frac{\sum_{i=1}^{n} \mathrm{a}(E_i)\, \mathrm{l}\Big(\frac{z-Z_i}{\sigma_\mathrm{l}}\Big)}{\sum_{i=1}^{n} \mathrm{l}\Big(\frac{z-Z_i}{\sigma_\mathrm{l}}\Big)}, \tag{4.1}$$

for a kernel function $\mathrm{l} : \mathscr{R}^{\dim(Z)} \to \mathscr{R}$ and a smoothing parameter σ_l. The Nadaraya-Watson estimator will prove a convenient choice for our purposes. However, the limiting distribution of $\widehat{\theta}$ will not depend on the particular choice for the first-step estimator, so long as it satisfies certain conditions.

Because bias induced by kernel weighting can be dealt with under the usual regularity and smoothness conditions on the kernels and density functions involved, the largest chunk of our subsequent endeavors will be devoted to establishing the impact of first-step estimation error on the asymptotic variance of $\widehat{\theta}$. In doing so, it will be useful to interpret $\widehat{\mathrm{q}}_n(\theta)$ as an approximation to

$$\mathrm{q}_n(\theta) \equiv \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{i<j} \frac{\mathrm{s}(V_i, V_j, \theta)}{\sigma_\mathrm{k}^{\dim(\vartheta_0)}} \mathrm{k}\Big(\frac{\vartheta_0(W_i) - \vartheta_0(W_j)}{\sigma_\mathrm{k}}\Big) \mathrm{t}(Z_i)\mathrm{t}(Z_j) = \mathbb{P}_n[\mathrm{h}(\cdot, \cdot, \theta)].$$

This would be the objective function of choice if $\vartheta_0(w)$ was directly observable; the difference $\widehat{\mathrm{q}}_n(\theta) - \mathrm{q}_n(\theta)$ is entirely due to the noise in $\widehat{\vartheta}(w)$.[5]

4.1 Distribution theory

I begin by stating conditions on the kernel function and bandwidth sequence used in the construction of the first-step estimator. For vectors A and B of equal length, let $|A| \equiv \sum_{j=1}^{\dim(A)} A^{(j)}$ and let $B^{\mathscr{A}} \equiv \sum_{j=1}^{J} (B^{(j)})^{A^{(j)}}$.

Assumption 1. *For a positive integer ℓ, l is a symmetric ℓth-order kernel function. That is, $\mathrm{l}(\eta) = \mathrm{l}(-\eta)$, $\int \mathrm{l}(\eta)\, \mathrm{d}\eta = 1$, $\int \eta^l\, \mathrm{l}(\eta)\, \mathrm{d}\eta = 0$ for $|l| = 0, \ldots, \ell - 1$, and $\int \|\eta^\ell\|\, \|\mathrm{l}(\eta)\|\, \mathrm{d}\eta < \infty$. In addition, l is bounded and $\mathfrak{a}$-Hölder for some $\mathfrak{a} > 0$.*

Assumption 2. *The bandwidth σ_l is nonnegative and proportional to $n^{-\lambda}$, where $\lambda \in (\frac{1}{2\ell}, \frac{1-\epsilon}{2\dim(Z)})$ for some $\epsilon > 0$.*

A kernel that satisfies Assumption 1 may be composed by making use of formulae provided by Müller (1984). As usual, a larger number of regressors requires both a kernel of a higher order and a bandwidth that shrinks to zero more slowly.

The dimension of Z also affects the degree of differentiability that is required from its density, as is apparent from the following assumption.

[5] The trimming in $\mathrm{q}_n(\theta)$ is obsolete because $\vartheta_0(W)$ is assumed known. It is maintained here for convenience, however, as this infeasible criterion function will be of use later on.

Assumption 3. *Let Z have Lebesgue density p_Z and let $\mathscr{Z}$ be a compact subset of* $\mathrm{supp}(Z)$ *so that* $\inf_{z\in\mathscr{Z}} \mathrm{p}_Z(z) > 0$ *and* $\sup_{z\in\mathscr{Z}} \mathrm{p}_Z(z) < \infty$. *Then, for each z in $\mathscr{Z}$,* $\mathrm{p}_Z(z)$ *and* $\mu_{\mathrm{a}(E)}(z)$ *are ℓ-times continuously differentiable with bounded derivatives. In addition, under* P, *the function* a *has an envelope whose fourth moment exists and whose conditional variance given $Z = z$ is continuous in z.*

In addition to imposing smoothness conditions, Assumption 3 introduces a subset of the support of the regressors on which p_Z is known to be bounded away from zero. This is a technical requirement that prevents the denominator of the first-step estimator from getting arbitrarily close to zero. It also avoids $\widehat{\mu}_{\mathrm{a}(E)}(z)$ from converging too slowly due to boundary effects. The demand for all components of Z to be continuous is motivated primarily by notational simplicity. The presence of discrete regressors would require a rewriting of Assumption 3 in terms of conditional densities and a corresponding adjustment to the kernel function in (4.1); see, e.g., Ahn (1997) for details. It is well known that the speed of convergence of nonparametric estimators does not depend on the number of discrete regressors present but does deteriorate with the number of continuously distributed ones.

Assumptions 1–3, in tandem, lead to a uniform rate of convergence and a linear representation result for $\widehat{\mu}_{\mathrm{a}(E)}(z) - \mu_{\mathrm{a}(E)}(z)$.

Lemma 1. *Let Assumptions 1–3 hold. Then*

$$(i)\quad \sup_{z\in\mathscr{Z}} \left\| \widehat{\mu}_{\mathrm{a}(E)}(z) - \mu_{\mathrm{a}(E)}(z) \right\| = \mathcal{O}_p\Big(\sqrt{\frac{n^{\epsilon/2}}{n\sigma_1^{\dim(Z)}}}\Big); \quad \textit{and}$$

$$(ii)\quad \widehat{\mu}_{\mathrm{a}(E)}(z) - \mu_{\mathrm{a}(E)}(z) = \frac{1}{n\sigma_1^{\dim(Z)}} \sum_{i=1}^{n} \frac{[\mathrm{a}(E_i) - \mu_{\mathrm{a}(E)}(z)]}{\mathrm{p}_Z(z)} \mathrm{l}\Big(\frac{z - Z_i}{\sigma_1}\Big) + \mathcal{O}_p\Big(\frac{n^{\epsilon/2}}{n\sigma_1^{\dim(Z)}}\Big)$$

uniformly over $\mathscr{Z}$.

Lemma 1 is similar to Theorem 1 in Aradillas-López, Honoré, and Powell (2007) and will prove useful in handling the sample noise in $\widehat{\vartheta}(w)$.

The next assumption brings us to the second estimation step and is concerned with identification.

Assumption 4. *The vector X has at least one component whose distribution conditional on the remaining* $\dim(X) - 1$ *components and the control has an everywhere positive Lebesgue density, and the support of X given $\vartheta_0(W) = \vartheta$ is not contained in a proper linear subspace of* $\mathscr{R}^{\dim(X)}$ *a.e. ϑ.*

Semiparametric point identification of (scaled) index coefficients, in general, requires stronger conditions on the distribution of the covariates to hold than in parametric problems; see, again, Horowitz (2009) for a discussion. Assumption 4 is a straightforward modification to the support conditions in Manski (1985), Han (1987), Cavanagh and Sherman (1998), and many others. Besides the conventional 'full-rank' condition, which serves to prevent problems of global underidentification, it requires one covariate to have a density with large support, given realizations of the remaining covariates and the control. This is stronger than necessary but implies that the set

$$\Big\{(x_1,x_2)\in\operatorname{supp}(X|\vartheta_0(W)=\vartheta)\times\operatorname{supp}(X|\vartheta_0(W)=\vartheta):\operatorname{sgn}[(x_1-x_2)'\theta]\neq\operatorname{sgn}[(x_1-x_2)'\theta_0]\Big\}$$

has non-zero measure under $\mathbb{P}$ for all ϑ in $\operatorname{supp}(\vartheta_0(W))$ and each θ in Θ except for $\theta=\theta_0$. This will lead to θ_0 uniquely maximizing the large-n probability limit of $\widehat{\mathrm{q}}_n(\theta)$.

Assumption 5 helps to ensure that the objective function is well behaved.

Assumption 5. *The second moment of* $\mathrm{m}(Y)$ *under* P *exists and the function* t *is of the form* $\mathrm{t}(z)=1(z\in\mathscr{Z})\,\mathrm{i}(z)$*, where* $\mathrm{i}:\mathscr{R}^{\dim(Z)}\to\mathscr{R}_0^+$ *is bounded and* ℓ*-times differentiable with bounded derivatives.*

Refer to Assumption 3 to recall that the trimming set serves to keep the kernel weights well defined. The fixed trimming procedure prescribed here comes at a cost in terms of asymptotic efficiency as it implies that a fraction of the data is ignored asymptotically. It is, however, convenient for proving consistency and asymptotic normality of the local-rank estimator and has been applied elsewhere; Ichimura (1993) and Newey (1994a) are two of many examples. Arguably, the analysis below could be adjusted to allow for this fraction to converge to zero slowly with the sample size, as in Stoker (1991), for example.

The second-step kernel and bandwidth are governed by the next two assumptions.

Assumption 6. *For a positive integer* $\mathscr{k}$, k *is a symmetric* $\mathscr{k}$*th-order kernel function. That is,* $\mathrm{k}(\eta)=\mathrm{k}(-\eta)$, $\int\mathrm{k}(\eta)\,\mathrm{d}\eta=1$, $\int\eta^k\mathrm{k}(\eta)\,\mathrm{d}\eta=0$ *for* $|k|=1,\ldots,\mathscr{k}-1$*, and* $\int\|\eta^k\|\,\|\mathrm{k}(\eta)\|\,\mathrm{d}\eta<\infty$ *for* $|k|\in\{0,\mathscr{k}\}$*. In addition,* k *is bounded, of bounded variation, and twice differentiable with bounded derivatives* k' *and* k''*.*

Assumption 7. *The bandwidth* σ_{k} *is nonnegative and proportional to* $n^{-\kappa}$*, where* $\kappa\in\left(\frac{1}{2\mathscr{k}},\frac{1-\epsilon-2\dim(Z)\lambda}{2(\dim(\vartheta_0)+2)}\right)$.

Imposing symmetry on k is natural given that the weight that is assigned to the score contribution of a pair of observations should not depend on the order in which these observations enter s.[6] Assumptions 6 and 7 have a similar purpose as do Assumptions 1 and 2, that is, aid in ensuring the bias induced by kernel-weighting to be asymptotically negligible.

To state the accompanying smoothness condition on the density of the control, and to support our future work, additional notation is useful. Let

$$\tau(d,\theta) \equiv \mathrm{h}(d,\mathrm{P},\theta) \qquad \text{and} \qquad \overline{\tau}(d,\theta) \equiv \lim_{n\to\infty} \tau(d,\theta) = \mathrm{t}(z)\ \overline{\varphi}(v,\vartheta_0(w),\theta),$$

where

$$\overline{\varphi}(v_1,\vartheta,\theta) \equiv \int \mathrm{s}(v_1,v_2,\theta)\ \mathrm{t}(z_2)\ \mathrm{dP}_{(V,Z)|\vartheta_0(W)}(v_2,z_2|\vartheta)\ \mathrm{p}_{\vartheta_0(W)}(\vartheta)$$

and $\mathrm{h}(d,\mathrm{P},\theta)$ refers to the expectation of $\mathrm{h}(\cdot,\cdot,\theta)$ given its first argument. Notice that $\overline{\varphi}(V_i,\vartheta_0(W_i),\theta)$ is the expected score contribution of observation i in the subpopulation for which $\vartheta_0(w) = \vartheta_0(W_i)$ (and $z \in \mathscr{Z}$), scaled by the density of the control evaluated at the same point.

The second-step analog of Assumption 3 now follows.

Assumption 8. *For each θ in Θ, v in $\mathrm{supp}(V)$, and w in $\mathrm{supp}(W)$, the function $\overline{\varphi}(v,\vartheta_0(w),\theta)$ is $(\kappa+1)$-times differentiable in its second argument, and the derivatives are uniformly bounded. Furthermore, the first derivative, $\nabla_\vartheta\overline{\varphi}(v,\vartheta_0(w),\theta)$, is ℓ-times differentiable in z, and the derivatives are uniformly bounded.*

This differentiability condition, in combination with the previous assumptions, implies that $\tau(d,\theta) = \overline{\tau}(d,\theta) + o(1/\sqrt{n})$ uniformly over Θ. The limiting objective function for our problem then is

$$\overline{\mathrm{q}}(\theta) \equiv \operatorname*{plim}_{n\to\infty} \widehat{\mathrm{q}}_n(\theta) = \overline{\tau}(\mathrm{P},\theta),$$

where the convergence is again uniform.

The above conditions suffice for $\widehat{\theta} = \arg\max_{\theta\in\Theta} \widehat{\mathrm{q}}(\theta) \xrightarrow{p} \arg\max_{\theta\in\Theta} \overline{\mathrm{q}}(\theta) = \theta_0$.[7] Theorem 1 summarizes our progress so far.

Theorem 1. *Let Assumptions 1–8 hold. Then $\|\widehat{\theta} - \theta_0\| = o_p(1)$.*

[6] Symmetry of k has additional advantages. First, it implies symmetry of h and thus leads to $\{\mathrm{h}(,\cdot,\cdot,\theta) : \theta \in \Theta\}$ being a U-process with a symmetric 'kernel'; this is convenient for the large-sample analysis. Second, it facilitates the construction of a higher-order kernel. In any case, most garden-variety kernels are symmetric; see, e.g., Li and Racine (2007) for a discussion.

[7] Clearly, the consistency of $\widehat{\theta}$ for θ_0 holds under weaker assumptions; all that is required is that $\tau(d,\theta) = \overline{\tau}(d,\theta) + o(1)$ uniformly over Θ and that θ_0 is the sole global maximizer of $\overline{\mathrm{q}}(\theta)$ on Θ. The higher-order kernel and differentiability conditions, and the undersmoothing will, however, prevent bias terms from appearing in the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$.

Continuing on to the asymptotic distribution of the local-rank estimator requires establishing the impact of the first-step estimation error, which calls for a somewhat more delicate argument. Let

$$\overline{\zeta}(w,\theta) \equiv -\mathrm{t}(z)[\mathrm{a}(e)-\mu_{\mathrm{a}(E)}(z)]'\,\overline{\delta}(z,\theta), \quad \overline{\delta}(z,\theta) \equiv \int \nabla_{\vartheta}\overline{\varphi}(v,\vartheta_0(w),\theta)\;\mathrm{dP}_{(V,W)|Z}(v,w|z).$$

The vector-valued derivative $\nabla_{\vartheta}\overline{\varphi}(v,\vartheta_0(w),\theta)$ is a measure of variability of the expected score to changes in the components of the control; $\overline{\delta}(z,\theta)$ is its expected value given $Z=z$. The more sensitive this latter function is to changes in the control, the greater the extent to which first-step estimation noise affects the asymptotic variance of $\widehat{\theta}$; this point will be made more precise below.

The following lemma shows that $\widehat{\mathrm{q}}_n(\theta)$ asymptotically behaves like the sum of two U-statistics and is key in deriving the limiting distribution of $\sqrt{n}(\widehat{\theta}-\theta_0)$.

Lemma 2. *Let Assumptions 1–8 hold. Then*

$$\widehat{\mathrm{q}}_n(\theta) = \mathrm{q}_n(\theta) + 2\overline{\zeta}(\mathrm{P}_n,\theta) + o_p\Big(\frac{1}{\sqrt{n}}\Big)$$

uniformly over Θ.

Recall that $\mathrm{q}_n(\theta)$ is the infeasible criterion function in which $\widehat{\vartheta}(w)$ does not appear. Lemma 2 thus implies that we can handle the variation that is induced through the first estimation step separately from the analysis of an infeasible estimator that assumes $\vartheta_0(w)$ to be observable.

The proof to asymptotic normality builds on Lemmata A and B in the Appendix. The first of these deals with $\mathrm{q}_n(\theta)$ and uses Assumption 9. The second auxiliary lemma concerns $\overline{\zeta}(w,\theta)$ and relies on Assumption 10.

Assumption 9. *Let $\mathscr{N}$ denote a neighborhood of θ_0. For each d in $\mathrm{supp}(D)$ and θ in $\mathscr{N}$, all mixed second partial derivatives of $\overline{\tau}(d,\theta)$ exist and there exists an integrable function $\mathcal{M}_{\tau}(d)$ so that $\|\nabla_{\theta\theta}\overline{\tau}(d,\theta)-\nabla_{\theta\theta}\overline{\tau}(d,\theta_0)\| \leq \mathcal{M}_{\tau}(d)\ \|\theta-\theta_0\|$. In addition, the moments $\mathrm{P}[\|\nabla_{\theta}\overline{\tau}(\cdot,\theta_0)\|^2]$ and $\mathrm{P}[\|\nabla_{\theta\theta}\overline{\tau}(\cdot,\theta_0)\|]$ exist, and $\mathrm{P}[\nabla_{\theta\theta}\overline{\tau}(\cdot,\theta_0)]$ is negative definite.*

Assumption 10. *For each θ in $\mathscr{N}$ and w in $\mathrm{supp}(W)$, all mixed second derivatives of $\overline{\zeta}(w,\theta)$ exist and there exists an integrable function $\mathcal{M}_{\zeta}(w)$ so that $\|\nabla_{\theta\theta}\overline{\zeta}(w,\theta)-\nabla_{\theta\theta}\overline{\zeta}(w,\theta_0)\| \leq \mathcal{M}_{\zeta}(w)\ \|\theta-\theta_0\|$. In addition, $\mathrm{P}[\|\nabla_{\theta}\overline{\zeta}(\cdot,\theta_0)\|^2]$ and $\mathrm{P}[\|\nabla_{\theta\theta}\overline{\zeta}(\cdot,\theta_0)\|]$ exist.*

These last two assumptions postulate conditions that allow for expansions of $\overline{\tau}(d,\theta)$ and $\overline{\zeta}(w,\theta)$ in a neighborhood of θ_0. They are in line with conventional restrictions which, in the context of rank regressions, first appeared in Assumption A4 of Sherman (1993). Also imposed is the existence of certain moments of the derivatives of $\overline{\tau}(\cdot,\theta)$ and $\overline{\zeta}(\cdot,\theta)$ under P. This allows the application of a standard law of large numbers and a central limit theorem.

All the necessary ingredients are now available to validate the linear representation

$$\sqrt{n}(\widehat{\theta}-\theta_0) = -\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)^{-1}\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\overline{\varsigma}(D_i,\theta_0)+o_p(1),$$

where $\overline{\varsigma}(d,\theta_0) \equiv \nabla_\theta\overline{\tau}(d,\theta_0)+\nabla_\theta\overline{\zeta}(w,\theta_0)$. On noting that $-2\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)^{-1}\overline{\varsigma}(\cdot,\theta_0)$ has zero mean and finite variance under P, the main result of this subsection follows. Theorem 2 provides it.

Theorem 2. *Let Assumptions 1–10 hold. Then*

$$\sqrt{n}(\widehat{\theta}-\theta_0)\xrightarrow{\mathcal{L}}\mathcal{N}\big(0,\Upsilon^{-1}\Sigma\ \Upsilon^{-1}\big),$$

where $\Sigma\equiv\mathrm{P}\big[\overline{\varsigma}(\cdot,\theta_0)\overline{\varsigma}(\cdot,\theta_0)'\big]$ *and* $\Upsilon\equiv\frac{1}{2}\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)$.

The two-step local-rank estimator converges in probability to θ_0 at the parametric rate, and $\sqrt{n}\widehat{\theta}$ converges in distribution to a Gaussian process that is centered at θ_0. The influence-function representation is convenient for evaluating the impact of having to settle with noisy estimates of the control, which is captured by the term $\nabla_\theta\overline{\zeta}(w,\theta_0)$ in $\overline{\varsigma}(d,\theta_0)$. Notice that this adjustment does not depend on the particular form of the first-step estimator used. This is in line with Newey's (1994b) treatment of semiparametric estimators with estimated nuisance functions under conventional smoothness conditions on the objective function. The effect on the asymptotic variance of working with $\widehat{\vartheta}(w)$ rather then with $\vartheta_0(w)$ is apparent from the form of Σ.

On letting $\widetilde{\theta}\equiv\arg\max_{\theta\in\Theta}\mathrm{q}_n(\theta)$, an immediate consequence of the analysis that leads to Theorem 2 is that

$$\sqrt{n}(\widehat{\theta}-\widetilde{\theta}) = -\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)^{-1}\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\nabla_\theta\overline{\zeta}(W_i,\theta_0)+o_p(1),$$

from which the next result follows easily.

Corollary 1. *Let Assumptions 4–9 hold and let* $\widehat{\vartheta}(w)=\vartheta_0(w)$*. Then Theorem 1 still holds and Theorem 2 continues to go through on replacing* $\overline{\varsigma}(d,\theta_0)$ *by* $\nabla_\theta\overline{\tau}(d,\theta_0)$*.*

Corollary 1 essentially provides the asymptotic distribution for the local-rank estimator when the control is observable. This result is of interest in its own right, as it can be applied when dealing with nonseparable versions of Robinson's (1988) partially-linear-index model. Of course, in such a situation, one would work with an objective function from which the trimming functions have been removed.

4.2 Variance estimation

In order to conduct inference an estimator of the asymptotic variance in Theorem 2 is needed. The derivation of such an estimator is a somewhat more cumbersome task than in conventional estimation problems due to the non-smooth nature of the objective function. I follow a kernel-based approach in the spirit of Abrevaya (1999b), among others. An alternative would be to rely on numerical-derivative methods (see, e.g., Pakes and Pollard, 1989 or Sherman, 1993), to use derivatives of a smoothed objective function (as in Chen, 2002), or simply to use the bootstrap, although this latter option would be computationally more demanding.[8]

For ease of notation, let $I(x,w) \equiv (x'\theta_0, \vartheta_0(w)')'$ and write $\mathrm{p}_I(I(x,w))$ for the density of $I(X,W)$ at $I(x,w)$. Define

$$\mathcal{X}(x,w) \equiv \mathrm{t}(z)\ \mu_{\mathrm{t}(Z)}(I(x,w))\Big[x - \frac{\mu_{\mathrm{t}(Z)X}(I(x,w))}{\mu_{\mathrm{t}(Z)}(I(x,w))}\Big]$$

and let $\mathcal{S}(y_1,\iota) \equiv \mathrm{m}(y_1) - \int \mathrm{m}(y_2)\ \mathrm{dP}_{Y|I(X,W)}(y_2|\iota)$; observe that $\mathcal{S}(Y,\iota)$ has mean zero.

Impose the additional regularity conditions below.

Assumption 11. *The functions $\mathcal{S}(y,\iota)$ and $\mathrm{p}_I(\iota)$ are differentiable with respect to ι and the second moment of $\mathrm{t}(Z)X$ under P exists.*

We can then obtain the following result.

Lemma 3. *Let Assumption 11 hold. Then the components of $\overline{\varsigma}(d,\theta)$ are*

$$\begin{aligned}
\nabla_\theta \overline{\tau}(d,\theta_0) &= \quad \mathcal{X}(x,w)\ \mathcal{S}\big(y, I(x,w)\big)\ \mathrm{p}_I\big(I(x,w)\big) \qquad \textit{and} \\
\nabla_\theta \overline{\zeta}(w_1,\theta_0) &= -\int \mathcal{X}(x_2,w_2)\ \mathcal{S}_2\big(y, I(x_2,w_2)\big)'\mathrm{p}_I\big(I(x_2,w_2)\big)\mathrm{dP}_{I(X,W)|Z}(I(x_2,w_2)|z_1) \\
&\qquad \times\ [\mathrm{a}(e_1) - \mu_{\mathrm{a}(E)}(z_1)],
\end{aligned}$$

while $\Upsilon = \int \big[\mathcal{X}(x,w)\ \mathcal{X}(x,w)'\big]\ \mathcal{S}_1\big(y, I(x,w)\big)\ \mathrm{p}_I\big(I(x,w)\big)\ \mathrm{dP}(x,w)$, Here, $\mathcal{S}_1(y,\iota)$ and $\mathcal{S}_2(y,\iota)$ denote the derivatives of $\mathcal{S}(y, I(x,w))$ with respect to the indices, evaluated at ι.

[8] Numerical-derivative methods are known to give unstable results. A strategy based on a smoothed objective function is straightforward to implement; conditions for consistency are easily found. Recent work by Subbotin (2008a) is concerned with validity of the bootstrap in rank estimation problems.

Except for the presence of trimming, the form of $\nabla_\theta \overline{\tau}(d, \theta_0)$ and $\nabla_{\theta\theta} \overline{\tau}(\mathrm{P}, \theta_0)$ are natural generalizations of the building blocks of the asymptotic variance of the estimator in Cavanagh and Sherman (1998); compare them with the expressions in Ichimura and Lee (1991), for example. A further beautification of the formula for the influence function is hindered, however, due to the presence of the term that arises from the nonparametric estimation of $\mu_{\mathrm{a}(E)}(z)$.

Now let $\mathrm{j} : \mathscr{R}^{1+\dim(\vartheta_0)} \to \mathscr{R}$ be a kernel indexed by the bandwidth σ_j. A kernel estimator of $\mathrm{p}_I\big(I(x,w)\big)$, for example, is

$$\widehat{\mathrm{p}}_I\big(\widehat{I}(x,w)\big) \equiv \frac{1}{n\sigma_\mathrm{j}^{\dim(\vartheta_0)+1}} \sum_{i=1}^{n} \mathrm{j}\Big(\frac{\widehat{I}(x,w) - \widehat{I}(X_i, W_i)}{\sigma_\mathrm{j}}\Big)$$

for $\widehat{I}(x,w) \equiv (x'\widehat{\theta}, \widehat{\vartheta}(w)')'$. Estimators of all the objects in Lemma 3 are collected in the Appendix. These can be combined and averaged across observations to obtain the plug-in estimators $\widehat{\Upsilon}$ and $\widehat{\Sigma}$, say.

Suppose that j is constructed in concordance with Assumption 12.

Assumption 12. *The function* j *is twice differentiable with bounded derivatives,* j' *and* j'', $\int \mathrm{j}(\eta)\ \mathrm{d}\eta = 1$, *and* σ_j *is nonnegative and proportional to* n^{-j} *for a positive scalar* j*.*

Then the consistency of the estimator of the asymptotic variance can be established under standard regularity conditions.

Assumption 13. *Both* $\mu_{\mathrm{t}(Z)}(\iota)$ *and* $\mu_{\mathrm{t}(Z)X}(\iota)$ *are once continuously differentiable while* $\mu_{\mathrm{m}(Y)}(\iota)$ *and* $\mathrm{p}_I(\iota)$ *are twice continuously differentiable.*

Theorem 3 states the consistency result.

Theorem 3. *Let Assumptions 1–13 hold. Then*

$$\widehat{\Upsilon}^{-1}\widehat{\Sigma}\ \widehat{\Upsilon}^{-1} \xrightarrow{p} \Upsilon^{-1}\Sigma\ \Upsilon^{-1}$$

provided that $j < \frac{1-\epsilon/2-\lambda\dim(Z)}{2(\dim(\vartheta_0)+3)}$.

The slow rate at which σ_j is restricted to approach zero is due to the presence of the first-step estimator, which retard convergence. This is in contrast to Abrevaya (1999b), where the shrinkage speed of σ_j was dominated by a $\sqrt{n}$-consistent plug-in estimator. Another difference with his result is that the second term in $\overline{\varsigma}(\cdot, \theta_0)$ causes $\widehat{\Upsilon}$ and $\widehat{\Sigma}$ to require the same degree of smoothing to be consistent.

4.3 Comments and extensions

While the focus here has been on a particular estimator, the distribution theory just derived can be applied to a much broader class of estimators. Because the regularity conditions used here do not impose smoothness of the objective function, this class includes rank-based approaches as well as pairwise-difference techniques. Examples of estimators that can be augmented with nonparametric controls in this manner are Han's (1987) maximum-rank correlation estimator, the partial-rank estimator of a duration model with covariate-dependent censoring introduced by Khan and Tamer (2007), and the estimator of Abrevaya (1999a) for the two-period panel data transformation model. It suffices to merely redefine the score function s for both the consistency and asymptotic normality results go through, provided that s is Euclidean with a square-integrable envelope; see Jochmans (2010) for details. Lemma 3 and Theorem 3 generalize on redefining $\mathcal{S}(y, \iota)$ in light of this change.

Similarly, one could pursue modifications of rank estimators that maximize higher-order U-processes to deal with controls. An example of an estimator open to such an exercise is Bhattacharya's (2008) monotone permutations estimator. The key difference between his proposal and Cavanagh and Sherman's (1998) is that it ranks observations within m-tuples of data points, where $m \geq 2$. While such a procedure might improve the accuracy of inferences for $m > 2$—as is suggested by his Monte Carlo evidence—the computational burden rapidly becomes insurmountable as m increases, even for very small n. In addition, it is not clear a priori that such a finite-sample improvement carries over to inference from a kernel-weighted version of a higher-order U-process. The reason is that a local ranking within an m-tuple would require the inclusion of $m - 1$ kernel weights in the objective function.

Throughout the large-sample analysis the bandwidths were taken to be deterministic functions of n. From a practical point of view, however, it would be of interest to have theoretical guidance on choosing the smoothing parameters when dealing with small samples. Allowing for such data-dependent bandwidths is technically challenging as they enter the objective function nonlinearly and their convergence rates are interrelated. These problems are not unique to our framework and many others—including Ahn and Powell (1993), Ahn (1997), and Aradillas-López, Honoré, and Powell (2007)—faced them as well. Common practice so far has been to employ cross-validation techniques in the hope that they would work well; see, e.g., Newey, Powell, and Walker (1990) and Härdle, Hall, and Ichimura (1993). I follow the same strategy below. However, these are not necessarily optimal smoothing choices for estimating the index parameters.

Other questions that are left for future work relate to testing the specification. As an example, test statistic for omitted regressors or for the validity of the index restrictions could be formed by extending the proposals of Fan and Li (1996) for single-index models. The properties of such a test would, however, not follow straightforwardly as $\vartheta_0(w)$ is not estimable at rate $\sqrt{n}$. Similarly, for testing the key monotonicity assumptions underlying the local-rank estimator or Blundell and Powell's (2004) approach, one could pursue modifications of a variety of tests for the shape of a nonparametric regression curve; see, for example, Ghosal, Sen, and van der Vaart (2000).

V THREE-STEP ESTIMATION OF ADDITIONAL PARAMETERS

Besides being of direct use, the estimator just analyzed and the distribution theory underlying it can be helpful in learning about other parameters of interest. I discuss two applications here.

5.1 Transformation models

Many econometric models, and duration models in particular, have an outcome variable that is assumed to be generated through an invertible transformation of covariates and a latent disturbance. One example is Ridder's (1990) generalized accelerated failure-time model. A generic formulation of the transformation model, augmented with controls, is

$$\psi_0(Y) = X'\theta_0 + g[\vartheta_0(W)] + U, \quad U \perp X | \vartheta_0(W) = \vartheta \;\text{ for all }\; \vartheta, \tag{5.1}$$

where $\psi_0 : \mathscr{R} \rightarrow \mathscr{R}$ is an unknown strictly monotonic function, normalized increasing, and the coefficient vector has already been normalized to live in Θ. Notice that (5.1) fits the general specification in (2.1) and so, under Assumptions 1–10, $\widehat{\theta}$ is an asymptotically-linear estimator of the scaled index coefficients in the transformation model. In this subsection our primary interest lies in additionally infering $\psi_0(y)$ at various values y in $\text{supp}(Y)$.

Doing so requires an additional normalization because the location of the distribution of U is not identified. A convenient choice is to set $\psi_0(y_0)$ to zero for some chosen baseline value y_0. Following the discussion in Chen (2002), a local-rank estimator of $\psi_0(y) - \psi_0(y_0) = \psi_0(y)$ is

$$\widehat{\psi}(y) \equiv \arg\max_{\psi \in \Psi} \widehat{\mathrm{q}}_n^y(\psi, \widehat{\theta}),$$

where the parameter space, Ψ, is a compact subset of the real line and, for (ψ, θ) in

$\Psi \times \Theta$,

$$\widehat{\mathrm{q}}^y_n(\psi,\theta) \equiv \binom{n}{2}^{-1} \sum_{i=1}^{N} \sum_{i<j} \frac{\mathrm{s}^y(V_i,V_j,\psi,\theta)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}} \mathrm{k}\Big(\frac{\widehat{\vartheta}(W_i)-\widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}}\Big)\mathrm{t}(Z_i)\mathrm{t}(Z_j). \tag{5.2}$$

This function differs from (3.2) only in the score contributions that are being averaged, which now also depends on an additional unknown parameter. Here,

$$\begin{aligned} \mathrm{s}^y(V_i,V_j,\psi,\theta) \equiv & \frac{1}{2}\big[1(Y_i \geq y) - 1(Y_j \geq y_0)\big]\ 1\big[(X_i - X_j)'\theta \geq \psi\big] \\ & + \frac{1}{2}\big[1(Y_j \geq y) - 1(Y_i \geq y_0)\big]\ 1\big[(X_j - X_i)'\theta \geq \psi\big]. \end{aligned} \tag{5.3}$$

The motivation behind the estimator comes from an insight that is similar to that envoked before. Moreover, because

$$\int 1(y_1 \geq y_2)\ \mathrm{p}_{Y|X,\vartheta_0(W)}(y_1|x_1,\vartheta)\ \mathrm{d}y_1 = 1 - \Pr\big[U \leq \psi_0(y_2) - x_1'\theta_0 - g(\vartheta)\big],$$

it follows that $(x_1 - x_2)'\theta_0 \geq \psi_0(y)$ if

$$\int 1(y_1 \geq y)\ \mathrm{p}_{Y|X,\vartheta_0(W)}(y_1|x_1,\vartheta)\ \mathrm{d}y_1 - \int 1(y_2 \geq y_0)\ \mathrm{p}_{Y|X,\vartheta_0(W)}(y_2|x_2,\vartheta)\ \mathrm{d}y_2 \geq 0;$$

notice that the function g does not appear directly. In (5.2), this implied ordering is enforced on the sample of data using a plug-in estimate of θ_0. The weights again serve to keep the pairwise comparisons in check. So, $\widehat{\psi}(y)$ constitutes a feasible three-step local-rank estimator of $\psi_0(y)$. Given the effort made so far, deriving the pointwise asymptotic behavior of $\widehat{\psi}(y)$ requires little additional work.

Restricting the extra notational burden to a minimum, and keeping the analogy to our old problem as tight as possible, write

$$\widehat{\mathrm{q}}^y_n(\psi,\theta) = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{i<j} \widehat{\mathrm{h}}^y(D_i,D_j,\psi,\theta) = \mathbb{P}_n[\,\widehat{\mathrm{h}}^y(\cdot,\cdot,\psi,\theta)]$$

and let $\mathrm{q}^y_n(\psi,\theta) \equiv \mathbb{P}_n[\mathrm{h}^y(\cdot,\cdot,\psi,\theta)]$, where $\mathrm{h}^y(\cdot,\cdot,\psi,\theta)$ is just $\widehat{\mathrm{h}}^y(\cdot,\cdot,\psi,\theta)$ with $\vartheta_0(W_i)$ replacing $\widehat{\vartheta}(W_i)$ for each $i = 1,\ldots,n$. To establish the large-sample distribution of $\widehat{\psi}(y)$ for fixed y, we will also be needing the functions

$$\tau^y(d,\psi,\theta) \equiv \mathrm{h}^y(d,\mathrm{P},\psi,\theta) \quad \text{and} \quad \overline{\tau}^y(d,\alpha,\theta) \equiv \lim_{n\to\infty} \tau(d,\psi,\theta) = \mathrm{t}(z)\ \overline{\varphi}^y(v,\vartheta_0(w),\psi,\theta),$$

where

$$\overline{\varphi}(v,\vartheta,\theta) \equiv \int \mathrm{s}^y(v_1,v_2,\psi,\theta)\ T(z_2)\ \mathrm{dP}_{(V,Z)|\vartheta_0(W)}(v_2,z_2|\vartheta)\ \mathrm{p}_{\vartheta_0(W)}(\vartheta).$$

As before, $\tau^y(\cdot,\psi,\widehat{\theta})$ is the kernel of the empirical process that drives the asymptotic behavior of $\widetilde{\psi}(y) \equiv \arg\max_{\psi\in\Psi} \mathrm{q}_n^y(\psi,\widehat{\theta})$. Under Assumption 15 below, it converges to $\overline{\tau}^y(d,\psi,\widehat{\theta})$ sufficiently fast so that the bias induced by kernel weighting is a non-issue, asymptotically. The variability of the first-step kernel estimator affects the limiting distribution of $\widehat{\psi}(y)$ through the partial derivatives of

$$\overline{\zeta}^y(w,\psi,\theta) \equiv -\mathrm{t}(z)[\mathrm{a}(e) - \mu_{\mathrm{a}(E)}(z)]'\,\overline{\delta}^y(z,\psi,\theta),$$

where

$$\overline{\delta}^y(z,\psi,\theta) \equiv \int \nabla_\vartheta \overline{\varphi}^y(v,\vartheta_0(w),\psi,\theta)\ \mathrm{dP}_{(V,W)|Z}(v,w|z).$$

The interpretation of these quantities is again clear on noting their resemblance with $\overline{\zeta}(w,\theta)$ and $\overline{\delta}(z,\theta)$ above.

Now, maintain Assumptions 1–10 and consider the following additional restrictions.

Assumption 14. *There exist values y_{L} and y_{U} in $\mathrm{supp}(Y)$ so that, for some $\varepsilon > 0$, $[\psi_0(y_{\mathrm{L}}-\varepsilon),\psi_0(y_{\mathrm{U}}+\varepsilon)] \subset \Psi$, and Ψ is a known compact interval of $\mathscr{R}$.*

Assumption 15. *For each y in $[y_{\mathrm{L}},y_{\mathrm{U}}]$, ψ in Ψ, v in $\mathrm{supp}(V)$, and w in $\mathrm{supp}(W)$, $\overline{\varphi}^y(v,\vartheta_0(w),\psi,\theta)$ is $(\mathscr{k}+1)$-times differentiable in its second argument over an $\mathcal{O}_p(1/\sqrt{n})$ neighborhood of θ_0, with the derivatives being uniformly bounded. Furthermore, the first derivative vector, $\nabla_\vartheta\overline{\varphi}^y(v,\vartheta_0(w),\theta)$, is $\mathscr{l}$-times differentiable in z, and the derivatives are uniformly bounded.*

Assumption 16. *Let $\mathscr{N}_y$ denote a neighborhood of $(\psi_0(y),\theta_0)$. For each y in $[y_{\mathrm{L}},y_{\mathrm{U}}]$, d in $\mathrm{supp}(D)$, and (ψ,θ) in $\mathscr{N}_y$, all mixed third partial derivatives of $\overline{\tau}^y(d,\psi,\theta)$ exist and there is an integrable function $\mathcal{M}_\tau^y(d)$ so that $\|\nabla_{\psi\psi}\overline{\tau}^y(d,\psi,\theta) - \nabla_{\psi\psi}\overline{\tau}^y(d,\psi_0(y),\theta)\| \leq \mathcal{M}_\tau^y(d)\|\psi-\psi_0(y)\|$. In addition, $\mathrm{P}[\nabla_\psi\overline{\tau}^y(\cdot,\psi_0(y),\theta_0)^2]$, $\mathrm{P}[\|\nabla_{\psi\psi}\overline{\tau}^y(\cdot,\psi_0(y),\theta_0)\|]$ and $\mathrm{P}[\|\nabla_{\psi\theta}\overline{\tau}^y(\cdot,\psi_0(y),\theta_0)\|]$ exist, and $\mathrm{P}[\nabla_{\psi\psi}\overline{\tau}^y(\cdot,\psi_0(y),\theta_0)] < 0$.*

Assumption 17. *For each y in $[y_{\mathrm{L}},y_{\mathrm{U}}]$, w in $\mathrm{supp}(W)$, and (ψ,θ) in $\mathscr{N}_y$, all mixed third partial derivatives of $\overline{\zeta}^y(w,\psi,\theta)$ exist and there exists an integrable function $\mathcal{M}_\zeta^y(w)$ so that $\|\nabla_{\psi\psi}\overline{\zeta}^y(w,\psi,\theta) - \nabla_{\psi\psi}\overline{\zeta}^y(w,\psi_0(y),\theta)\| \leq \mathcal{M}_\zeta^y(w)\ \|\psi-\psi_0(y)\|$. In addition, the moments $\mathrm{P}[\|\nabla_\psi\overline{\zeta}^y(\cdot,\psi_0(y),\theta_0)\|^2]$, $\mathrm{P}[\|\nabla_{\psi\theta}\overline{\zeta}^y(\cdot,\psi_0(y),\theta_0)\|]$, and $\mathrm{P}[\|\nabla_{\psi\psi}\overline{\zeta}^y(\cdot,\psi_0(y),\theta_0)\|]$ exist.*

The first of these assumptions, imposing compactness of the parameter space, is standard when analyzing estimators that have no closed-form solution.[9] The need for Assumption

[9]The same assumption was made on Θ, albeit implicitely. One of the attractive features of the scale normalization on the index coefficients maintained here is that it implies (i) Θ to be compact; and (ii) θ_0 to be interior to Θ.

15 has already been discussed. Assumptions 16 and 17 comprise smoothness conditions and the existence of moments analogous to Assumptions 9 and 10, guaranteeing limit quantities to be well defined. The need for mixed third- rather than second partial derivatives of $\overline{\tau}^y(d,\psi,\theta)$ to exist stems from the presence of θ.

These assumptions imply that, for each y in $[y_{\mathrm{L}}, y_{\mathrm{U}}]$, $\operatorname{plim}_{n\to\infty} \widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta}) = \overline{\mathrm{q}}_n^y(\psi,\theta_0)$ for $\overline{\mathrm{q}}_n^y(\psi,\theta_0) \equiv \overline{\tau}^y(\mathrm{P},\psi,\theta_0)$ uniformly over Ψ. The limiting function is continuous in ψ and reaches its unique global maximum on Ψ at $\psi_0(y)$. This statement follows immediately from Assumption 4, the same assumption that was previously envoked for the consistency of $\widehat{\theta}$. Consequently, we have that

$$\widehat{\psi}(y) = \arg\max_{\psi\in\Psi} \widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta}) \xrightarrow{p} \arg\max_{\psi\in\Psi} \overline{\mathrm{q}}_n^y(\psi,\theta_0) = \psi_0(y);$$

$\widehat{\psi}(y)$ is consistent for $\psi_0(y)$ for each y in $[y_{\mathrm{L}}, y_{\mathrm{U}}]$.

Furthermore, by smoothness of the objective function, the use of a bias-reducing kernel, and the $\sqrt{n}$-consistency of $\widehat{\theta}$, the estimation error in $\widehat{\psi}(y)$ asymptotically behaves like the sample average of a zero-mean random variable. Moreover,

$$\widehat{\psi}(y) - \psi_0(y) = -\nabla_{\psi\psi}\overline{\tau}^y\big(\mathrm{P},\psi_0(y),\theta_0\big)^{-1}\frac{2}{n}\sum_{i=1}^n \big[\overline{\varsigma}^y(D_i,\psi_0(y),\theta_0) + \overline{\varrho}^y(D_i,\psi_0(y),\theta_0)\big],$$

up to $o_p(1/\sqrt{n})$, for functions $\overline{\varsigma}^y\big(d,\psi_0(y),\theta_0\big) \equiv \nabla_\psi\overline{\tau}^y\big(d,\psi_0(y),\theta_0\big) + \nabla_\psi\overline{\zeta}^y\big(w,\psi_0(y),\theta_0\big)$ and

$$\overline{\varrho}^y\big(d,\psi_0(y),\theta_0\big) \equiv \frac{1}{2}\nabla_{\psi\theta}\overline{\tau}^y\big(\mathrm{P},\psi_0(y),\theta_0\big)\ \upsilon(d,\theta_0),$$

where $\upsilon(d,\theta_0)$ is shorthand for the influence function of $\widehat{\theta}$ evaluated at d. This latter term renders the asymptotic variance of our current problem more complicated than before and arises because of the additional noise induced by having to estimate θ_0 next to $\vartheta_0(w)$. Nevertheless, Assumptions 16 and 17 imply that, when multiplied by $\sqrt{n}$, the sample average above converges to a zero-mean random variable whose variance is finite. From this, the next asymptotic-normality result follows.

Theorem 4. *Let Assumptions 1–10 and 14–17 hold. Then, for each y in $[y_{\mathrm{L}}, y_{\mathrm{U}}]$, we have that (i) $\|\widehat{\psi}(y) - \psi_0(y)\| = o_p(1)$; and that (ii)*

$$\sqrt{n}\big(\widehat{\psi}(y) - \psi_0(y)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \Upsilon(y)^{-1}\Sigma(y)\ \Upsilon(y)^{-1}\big),$$

where $\Sigma(y) \equiv \mathrm{P}\big[\big(\overline{\varsigma}^y(\cdot,\psi_0(y),\theta_0) + \overline{\varrho}^y(\cdot,\psi_0(y),\theta_0)\big)\big(\overline{\varsigma}^y(\cdot,\psi_0(y),\theta_0) + \overline{\varrho}^y(\cdot,\psi_0(y),\theta_0)\big)'\big]$ and $\Upsilon(y) \equiv \frac{1}{2}\nabla_{\psi\psi}\overline{\tau}^y(\mathrm{P},\psi_0(y),\theta_0)$.

Under regularity conditions, a consistent estimator of the asymptotic variance can again be obtained via nonparametric techniques, using a plug-in estimator for $\upsilon(D_i, \theta_0)$, $i = 1, \ldots, n$. Theorem 4 deals with pointwise asymptotics. The result can be strengthened to hold uniformly over Ψ by considering a strategy based on rearrangement as proposed by Chernozhukov, Fernández-Val, and Galichon (2009). This additional step is required because a higher-order kernel k was used to eliminate asymptotic bias. Because such functions have to take on negative values on subsets of their support, it can be shown that $\widehat{\psi}(y)$, as a function of y, is no longer guaranteed to be monotonic.

Our analysis of the transformation model ends with a derivative result to Theorem 4 that parallels Corollary 1.

Corollary 2. *Let Assumptions 4–9 and Assumptions 14–16 hold, and let* $\widehat{\vartheta}(w) = \vartheta_0(w)$. *Then Theorem 4(i) still holds and Theorem 4(ii) continues to go through on replacing* $\overline{\varsigma}^y(d, \psi_0(y), \theta_0)$ *by* $\nabla_\psi \overline{\tau}^y(d, \psi_0(y), \theta_0)$.

5.2 Policy parameters

Many quantities of interest have as an elementary building block $\mu_{\mathrm{f}(Y)}(X'\theta_0, \vartheta_0(W))$ for a function $\mathrm{f} : \mathscr{R} \to \mathscr{R}$ that will depend on the problem at hand. While such type of parameters may be identified nonparametrically, knowledge of the index structure allows for dimensionality reduction in estimation and a relaxation of support conditions required for identification.

One important area of application is in triangular models. There, policy parameters typically take the form of partial means over the control. As an illustration, the average structural function at $X = x$ (Stock, 1989; Blundell and Powell, 2003, 2004) is defined as

$$\widetilde{\mu}_Y(x) \equiv \int \mu_Y(x'\theta_0, \vartheta)\ \mathrm{p}_{\vartheta_0(W)}(\vartheta)\ \mathrm{d}\vartheta \neq \int \mu_Y(x'\theta_0, \vartheta)\ \mathrm{p}_{\vartheta_0(W)|X}(\vartheta|x)\ \mathrm{d}\vartheta = \mu_Y(x)$$

and provides the expected value of the outcome for exogenously determined values of the covariates. This function can aid in the construction of counterfactual quantities or summary measures, by averaging $\widetilde{\mu}_Y(X)$ over a chosen distribution for the covariates in an additional step, for example. In models of the form in (2.2) or (2.3), another parameter that can be recovered by marginal integration over the control is the quantile structural function (Imbens and Newey, 2009). The value of the αth-quantile structural function at $X = x$ is that q_α that solves

$$\alpha = \widetilde{\mu}_{1(Y \le q_\alpha)}(x), \quad \widetilde{\mu}_{1(Y \le q)}(x) \equiv \int \mu_{1(Y \le q)}(x'\theta_0, \vartheta)\ \mathrm{p}_{\vartheta_0(W)}(\vartheta)\ \mathrm{d}\vartheta.$$

This second example also illustrates how index restrictions allow a relaxation of support conditions that are needed to ensure that $\mu_{\mathrm{f}(Y)}(x'\theta_0, \vartheta_0(W))$ is identified over the entire support of the density of the control; see Imbens and Newey (2009) for the original discussion. The method of marginal integration has been considered by Newey (1994a) and Linton and Nielsen (1995). Mammen, Rothe, and Schienle (2010) recently extended these results to averages over generated regressors. Our model falls into this latter category.

While our primary motivation for the inclusion of $\vartheta_0(W)$ into the analysis was to merely control for heterogenous effects, knowledge of $\mu_{\mathrm{f}(Y)}(X'\theta_0, \vartheta_0(W))$ can also be of use to learn about the impact of the control on the outcome variable. Recall Manski's (1991) approach to infering the effect of expectations on outcomes (cfr. Example 1). While in his model—as in Ahn's (1997)—this influence was specified to run through index parameters, here, counterfactual analyses can be performed semiparametrically. For example, the expected ceteris paribus effect of a change in expectations on outcomes at $(X, W) = (x, w)$ is $\nabla_\vartheta \mu_Y(x'\theta_0, \vartheta_0(w))$. Summary measures for the population or policy-relevant variables can again be formed by looking at the mean or the quantiles of the distribution of $\nabla_\vartheta \mu_Y(X'\theta_0, \vartheta_0(W))$ obtained on integrating out the covariates or the control using a chosen distribution.

VI SMALL-SAMPLE ASSESSMENT

To shed some light on the practical implementation of the local-rank estimator, and to position it against alternative techniques, Monte Carlo experiments are useful. Here, I report on results from applications to a model with an endogenous covariate and a model with sample selection.

6.1 A triangular model

The prime example of a control-function application is the estimation of the index coefficients in a linear simultaneous-equation model. I generated data from the following underlying model. Outcomes Y and E are related through

$$Y = X_1\theta_0^{(1)} + E\theta_0^{(2)} + U_1, \qquad E = X_1\gamma_0^{(1)} + X_2\gamma_0^{(2)} + U_2 \tag{6.1}$$

for disturbances (U_1, U_2) and regressors (X_1, X_2). These random variables were drawn as

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 1 & \rho_U \\ \rho_U & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 1 & \rho_X \\ \rho_X & 1 \end{pmatrix}, \tag{6.2}$$

respectively. This configuration is as in Example 3, with the conditional-mean functions for the outcome variables linear in parameters. The variable E is endogenous in the equation for Y unless $\rho_U = 0$. The disturbance U_2 can be interpreted as an omitted regressor which captures unobserved heterogeneity across units. So, here, $Z = (X_1, X_2)'$ and

$$\vartheta_0(E, Z) = U_2 = E - X_1\gamma_0^{(1)} + X_2\gamma_0^{(2)}$$

is the control function.

The designs for which I report results in Tables 1–3 below have $\theta_0 = (.7071, -.7071)'$, so $\|\theta_0\| = 1$, estimated from a sample of size $n = 100$. The strength of the dependence between the latent variables and the covariates considered are all combinations of ρ_U and ρ_X in $\{-.50, -.25, .25, .50\}$. To manipulate the explanatory power of the instrumental variables I vary the concentration parameter (Basmann, 1963), μ_0^2, from 100 to 20. Given that there are two elements in Z, these values lead to an F-statistic of 50 and 10, respectively (evaluated at the true γ_0). This latter value is commonly taken as the rule-of-thumb cut-off between weak and strong instruments in two-stage least squares (2SLS) regressions; see Stock, Wright, and Yogo (2002) for a motivation and formal derivation. To keep the explanatory power of the instruments fixed across simulation runs I used

$$\gamma_0 = \sqrt{\frac{\mu_0^2}{\sum_{i=1}^n \pi_0'(X_{1i}, X_{2i})'(X_{1i}, X_{2i})\pi_0}}\, \pi_0$$

to generate observations on E. Here, π_0 is a bivariate coefficient vector that was set to either $(2,2)'$ (balanced design; Table 1), $(2,1)'$ (skewed-left design; Table 2), or $(1,2)'$ (skewed-right design; Table 3). The variation in π_0 enables shifting the relative importance of X_1 and X_2 as sources of exogenous variation in E, and thus influences the degree to which the regressors in the equation for Y covary.

For this setup, 2SLS ($\widehat{\theta}_{\text{2SLS}}$) is the optimally-weighted GMM estimator, and it gives a useful benchmark to evaluate the performance of other techniques against. The other estimators are the kernel-weighted pairwise-differenced least-squares estimator of Ahn and Powell (1993) ($\widehat{\theta}_{\text{AP}}$), its nonlinear analog proposed by Blundell and Powell (2004) ($\widehat{\theta}_{\text{BP}}$), and the local-rank estimator ($\widehat{\theta}_{\text{RK}}$) for $\mathrm{m}(Y) = Y$.[10] Including the Ahn-Powell estimator allows evaluating the impact of introducing kernel-weights on the one hand, and the efficiency cost of avoiding the linearity assumption on the other.

[10] Recall that 2SLS is equivalent to adding the residual from a least-squares regression of E on Z to the second-step regression model. For 2SLS and the Ahn-Powell estimator, I rescaled the point-estimates by their norm to ensure that they lie in Θ. For 2SLS, a constant term was also included in the first- and second estimation step.

Table 1: Monte Carlo results for the triangular model in (6.1)–(6.2); balanced design ($\gamma_0 \propto (2,2)'$)

$\mu_0^2 = 100$

		BIAS				STD				IQR				RMSE				RMAE			
ρ_X	ρ_U	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$
−.50	.50	.0039	−.0215	−.1695	−.0449	.0755	.0599	.1242	.0618	.0750	.0802	.1792	.0833	.0756	.0636	.2101	.0764	.2272	.2237	.4146	.2486
		.0227	−.0161	−.1234	−.0377	.1425	.0574	.0762	.0562	.0757	.0759	.1172	.0736	.1442	.0596	.1450	.0676	.2518	.2183	.3550	.2370
−.50	.25	−.0017	.0146	−.1041	.0476	.0712	.0612	.1201	.0767	.0748	.0798	.1574	.1008	.0712	.0629	.1589	.0903	.2244	.2230	.3442	.2669
		.0080	.0208	−.0769	.0568	.0925	.0663	.0852	.0868	.0748	.0835	.1230	.1121	.0929	.0695	.1147	.1037	.2307	.2296	.3045	.2799
−.50	−.25	.0018	−.0139	−.1533	−.0226	.0726	.0551	.1232	.0575	.0757	.0727	.1763	.0758	.0726	.0568	.1966	.0617	.2252	.2116	.3968	.2207
		.0176	−.0095	−.1126	−.0180	.1295	.0540	.0785	.0543	.0759	.0702	.1204	.0710	.1306	.0549	.1373	.0572	.2445	.2086	.3422	.2145
−.50	−.50	.0046	−.0211	−.1693	−.0449	.0754	.0589	.1249	.0612	.0749	.0786	.1822	.0833	.0755	.0625	.2104	.0759	.2277	.2213	.4143	.2475
		.0224	−.0159	−.1232	−.0377	.1380	.0563	.0763	.0554	.0756	.0744	.1193	.0739	.1398	.0585	.1449	.0670	.2512	.2159	.3545	.2357
−.25	.50	.0036	−.0127	−.1339	−.0347	.0704	.0537	.1050	.0562	.0691	.0708	.1418	.0747	.0705	.0552	.1702	.0661	.2171	.2080	.3696	.2301
		.0192	−.0085	−.1026	−.0291	.1295	.0527	.0693	.0524	.0697	.0686	.1017	.0678	.1310	.0534	.1238	.0599	.2380	.2052	.3247	.2215
−.25	.25	−.0011	.0098	−.0822	.0400	.0625	.0576	.0998	.0732	.0687	.0749	.1239	.0959	.0625	.0584	.1292	.0834	.2126	.2146	.3097	.2556
		.0060	.0151	−.0635	.0477	.0782	.0613	.0748	.0813	.0688	.0773	.1019	.1047	.0784	.0631	.0981	.0942	.2172	.2192	.2802	.2661
−.25	−.25	.0020	−.0092	−.1219	−.0178	.0648	.0512	.1041	.0543	.0694	.0675	.1384	.0714	.0648	.0520	.1603	.0572	.2141	.2026	.3555	.2118
		.0141	−.0055	−.0939	−.0138	.1129	.0508	.0710	.0519	.0696	.0659	.1024	.0679	.1138	.0511	.1177	.0537	.2297	.2008	.3142	.2070
−.25	−.50	.0044	−.0123	−.1340	−.0347	.0684	.0530	.1060	.0563	.0687	.0697	.1418	.0753	.0685	.0544	.1708	.0662	.2171	.2064	.3697	.2299
		.0188	−.0082	−.1025	−.0292	.1237	.0518	.0696	.0521	.0694	.0677	.1020	.0686	.1251	.0525	.1239	.0597	.2369	.2035	.3245	.2211
.25	.50	.0015	−.0083	−.0964	−.0225	.0803	.0452	.0800	.0469	.0633	.0587	.0990	.0608	.0803	.0459	.1252	.0520	.2145	.1895	.3149	.2024
		.0227	−.0053	−.0786	−.0190	.1517	.0447	.0578	.0446	.0640	.0575	.0781	.0572	.1534	.0450	.0976	.0484	.2398	.1879	.2854	.1971
.25	.25	−.0005	.0076	−.0631	.0297	.0606	.0531	.0754	.0669	.0633	.0695	.0920	.0854	.0606	.0536	.0983	.0732	.2043	.2054	.2718	.2392
		.0070	.0119	−.0518	.0359	.0831	.0557	.0602	.0728	.0633	.0712	.0790	.0914	.0834	.0569	.0794	.0811	.2109	.2088	.2517	.2470
.25	−.25	.0012	−.0067	−.0892	−.0112	.0701	.0454	.0791	.0485	.0636	.0603	.0983	.0645	.0701	.0459	.1192	.0498	.2094	.1902	.3054	.1977
		.0166	−.0038	−.0729	−.0081	.1286	.0452	.0586	.0471	.0639	.0593	.0789	.0623	.1296	.0453	.0935	.0478	.2282	.1891	.2780	.1947
.25	−.50	.0029	−.0078	−.0967	−.0228	.0782	.0449	.0810	.0476	.0639	.0584	.1003	.0627	.0782	.0456	.1261	.0528	.2150	.1888	.3153	.2038
		.0225	−.0049	−.0787	−.0193	.1452	.0444	.0583	.0449	.0646	.0574	.0791	.0591	.1469	.0447	.0979	.0489	.2394	.1873	.2856	.1982
.50	.50	−.0033	−.0095	−.0847	−.0190	.1133	.0414	.0718	.0432	.0630	.0536	.0884	.0563	.1134	.0425	.1111	.0472	.2279	.1821	.2960	.1925
		.0341	−.0071	−.0705	−.0162	.1971	.0407	.0537	.0411	.0633	.0524	.0716	.0535	.2000	.0413	.0886	.0442	.2628	.1801	.2711	.1880
.50	.25	−.0013	.0083	−.0575	.0257	.0702	.0519	.0675	.0649	.0623	.0671	.0832	.0820	.0702	.0526	.0887	.0698	.2079	.2031	.2587	.2330
		.0119	.0124	−.0481	.0315	.1165	.0545	.0549	.0700	.0624	.0689	.0723	.0869	.1171	.0559	.0730	.0767	.2210	.2066	.2412	.2400
.50	−.25	−.0018	−.0073	−.0789	−.0091	.0984	.0428	.0710	.0462	.0634	.0564	.0881	.0604	.0984	.0434	.1061	.0471	.2202	.1848	.2878	.1923
		.0258	−.0046	−.0658	−.0064	.1695	.0424	.0542	.0451	.0637	.0554	.0724	.0586	.1714	.0427	.0852	.0455	.2473	.1835	.2645	.1897
.50	−.50	−.0020	−.0091	−.0852	−.0191	.1104	.0413	.0726	.0437	.0638	.0538	.0899	.0577	.1104	.0423	.1119	.0477	.2285	.1818	.2966	.1934
		.0348	−.0067	−.0707	−.0162	.1966	.0406	.0541	.0414	.0644	.0526	.0728	.0549	.1996	.0412	.0890	.0445	.2643	.1798	.2714	.1887

$\mu_0^2 = 20$

		BIAS				STD				IQR				RMSE				RMAE			
ρ_X	ρ_U	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$	$\hat\theta_{2SLS}$	$\hat\theta_{AP}$	$\hat\theta_{BP}$	$\hat\theta_{RK}$
−.50	.50	−.0165	.1245	−.2285	.1587	.1569	.0655	.1480	.0751	.1640	.0895	.2110	.0971	.1578	.1406	.2723	.1755	.3344	.3566	.4836	.4002
		.0590	.1649	−.1531	.2169	.2799	.1011	.0949	.1219	.1677	.1369	.1165	.1604	.2860	.1934	.1801	.2488	.3800	.4091	.4012	.4672
−.5	.25	−.0159	.0578	−.2746	.0792	.1566	.0642	.1226	.0735	.1624	.0853	.1695	.0975	.1574	.0864	.3007	.1081	.3338	.2678	.5247	.2995
		.0671	.0710	−.1828	.0965	.2969	.0784	.0775	.0921	.1664	.1024	.0819	.1214	.3044	.1057	.1985	.1334	.3903	.2892	.4308	.3259
−.50	−.25	−.0108	−.0507	−.3434	−.0500	.1542	.0589	.0999	.0642	.1565	.0786	.1432	.0860	.1546	.0777	.3576	.0814	.3305	.2523	.5860	.2572
		.0935	−.0431	−.2181	−.0426	.3391	.0530	.0418	.0565	.1672	.0685	.0556	.0743	.3517	.0683	.2221	.0708	.4157	.2395	.4670	.2432
−.50	−.50	−.0124	−.0922	−.3653	−.0993	.1613	.0578	.0935	.0586	.1562	.0770	.1356	.0788	.1617	.1088	.3771	.1153	.3347	.3095	.6044	.3191
		.1080	−.0780	−.2271	−.0838	.3638	.0468	.0413	.0457	.1764	.0600	.0490	.0597	.3795	.0910	.2308	.0955	.4309	.2862	.4769	.2939
−.25	.50	−.0146	.1177	−.2033	.1540	.1514	.0645	.1376	.0760	.1580	.0879	.1919	.0998	.1521	.1342	.2454	.1717	.3270	.3473	.4565	.3945
		.0548	.1538	−.1415	.2084	.2684	.0975	.0843	.1212	.1608	.1308	.1138	.1605	.2739	.1821	.1647	.2411	.3695	.3957	.3841	.4581
−.25	.25	−.0130	.0540	−.2497	.0760	.1511	.0633	.1157	.0729	.1566	.0844	.1580	.0964	.1517	.0832	.2752	.1053	.3271	.2620	.5003	.2949
		.0629	.0662	−.1717	.0921	.2836	.0764	.0717	.0903	.1599	.0999	.0819	.1181	.2904	.1011	.1861	.1290	.3800	.2820	.4168	.3197
−.25	−.25	−.0081	−.0462	−.3170	−.0463	.1520	.0572	.0960	.0622	.1506	.0764	.1348	.0829	.1522	.0735	.3312	.0776	.3255	.2447	.5630	.2509
		.0933	−.0394	−.2077	−.0396	.3340	.0520	.0425	.0552	.1621	.0674	.0572	.0727	.3467	.0652	.2120	.0680	.4104	.2332	.4557	.2380
−.25	−.50	−.0108	−.0836	−.3384	−.0918	.1611	.0553	.0898	.0562	.1508	.0731	.1281	.0745	.1614	.1002	.3501	.1077	.3301	.2957	.5817	.3077
		.1126	−.0714	−.2173	−.0782	.3686	.0456	.0372	.0447	.1714	.0583	.0508	.0579	.3854	.0847	.2205	.0901	.4303	.2750	.4662	.2848
.25	.50	−.0168	.1174	−.1754	.1476	.1690	.0608	.1196	.0716	.1673	.0817	.1596	.0934	.1698	.1322	.2123	.1640	.3402	.3458	.4233	.3859
		.0743	.1521	−.1281	.1974	.3075	.0922	.0744	.1136	.1747	.1212	.1022	.1487	.3164	.1779	.1481	.2278	.3954	.3927	.3636	.4457
.25	.25	−.0167	.0527	−.2199	.0716	.1685	.0607	.1021	.0703	.1651	.0807	.1343	.0925	.1694	.0804	.2424	.1003	.3403	.2572	.4694	.2873
		.0891	.0640	−.1583	.0863	.3362	.0735	.0568	.0866	.1751	.0949	.0761	.1131	.3478	.0975	.1682	.1223	.4134	.2764	.3985	.3105
.25	−.25	−.0192	−.0431	−.2818	−.0421	.1745	.0535	.0868	.0585	.1564	.0699	.1175	.0774	.1756	.0687	.2949	.0721	.3413	.2362	.5309	.2420
		.1267	−.0371	−.1928	−.0363	.3994	.0493	.0412	.0526	.1770	.0622	.0556	.0685	.4190	.0617	.1972	.0639	.4507	.2258	.4391	.2305
.25	−.50	−.0269	−.0777	−.3003	−.0833	.1930	.0491	.0818	.0511	.1539	.0642	.1126	.0675	.1949	.0920	.3112	.0977	.3514	.2839	.5480	.2928
		.1459	−.0674	−.2020	−.0720	.4303	.0409	.0367	.0412	.1863	.0520	.0504	.0537	.4544	.0788	.2053	.0829	.4692	.2655	.4494	.2727
.50	.50	−.0234	.1218	−.1677	.1465	.1930	.0585	.1116	.0677	.1865	.0785	.1464	.0899	.1944	.1351	.2014	.1614	.3620	.3511	.4129	.3841
		.1108	.1582	−.1245	.1957	.3739	.0907	.0700	.1089	.2019	.1181	.0956	.1420	.3899	.1824	.1429	.2239	.4428	.3995	.3572	.4434
.50	.25	−.0287	.0540	−.2099	.0706	.1989	.0594	.0961	.0681	.1831	.0790	.1250	.0899	.2010	.0802	.2308	.0980	.3648	.2572	.4585	.2847
		.1325	.0652	−.1535	.0850	.4123	.0723	.0543	.0840	.2071	.0933	.0728	.1094	.4330	.0974	.1628	.1195	.4659	.2765	.3922	.3076
.50	−.25	−.0413	−.0436	−.2689	−.0409	.2080	.0513	.0827	.0569	.1695	.0668	.1110	.0748	.2120	.0673	.2814	.0701	.3679	.2342	.5186	.2379
		.1601	−.0378	−.1871	−.0353	.4628	.0469	.0403	.0514	.2011	.0593	.0545	.0663	.4897	.0602	.1914	.0623	.4935	.2238	.4325	.2270
.50	−.50	−.0474	−.0782	−.2861	−.0805	.2199	.0459	.0782	.0487	.1612	.0597	.1050	.0647	.2250	.0906	.2966	.0941	.3744	.2832	.5349	.2877
		.1714	−.0680	−.1959	−.0700	.4789	.0380	.0362	.0394	.2087	.0483	.0491	.0516	.5087	.0780	.1992	.0804	.5046	.2650	.4426	.2687

$n = 100$, $\theta_0 = (.7071, -.7071)'$; standard-normal kernels; cross-validated bandwidths; 10,000 simulation runs.

To ensure that a proper comparison between the kernel-based estimators can be made, all were computed using the same kernel k and bandwidth σ_{k}, and the same estimates of $\vartheta_0(E, Z)$. As is agreed upon in the literature, bias-reducing kernels only give a worthwhile improvement over kernels of order two for reasonably large samples; see, e.g., Jones and Signorini (1997). In additional simulation experiments not reported on here I reached the same conclusion. Thus, because $n = 100$, the results below were obtained by means of the standard-normal density function as kernel in the first and second estimation step, with each of their arguments scaled down by their empirical standard deviation.[11] The bandwidths σ_{l} and σ_{k} were obtained using least-squares cross-validation methods; see, again, Li and Racine (2007) for details. The cross-validated σ_{k} relates to a nonparametric mean regression of Y on X_1, E, and $\widehat{\vartheta}(E, Z)$. The resulting estimates also serve as inputs for $\widehat{\theta}_{\mathrm{BP}}$.

Tables 1–3 give the bias, standard deviation (STD), interquartile range (IQR), root mean-squared error (RMSE), and root mean-absolute error (RMAE) of the estimators considered. The numbers were obtained over 10,000 Monte Carlo runs. Only the local-rank estimator requires an optimization procedure. To obtain the point estimates I modified the maximum-score algorithm of Manski and Thompson (1986). The procedure is described in the Appendix and was found to perform well. For the case of only two regressors, as here, the procedure is guaranteed to find a global maximizer of $\widehat{\mathrm{q}}_n(\theta)$.

Start with the balanced design with the concentration parameter set to 100. Both $\widehat{\theta}_{\mathrm{2SLS}}$ and $\widehat{\theta}_{\mathrm{AP}}$ have very small bias, and none consistently outperforms the other in terms of this measure. The Blundell-Powell estimator and the local-rank estimator have a larger bias throughout, with that of the former by far being the largest. Nevertheless, the average of the local-rank estimates is still very close to the true parameter values. When looking at the STD, $\widehat{\theta}_{\mathrm{AP}}$ and $\widehat{\theta}_{\mathrm{RK}}$ perform best. The standard error of 2SLS goes up by a factor of as much as four compared to the ones of these two approaches. The STD of $\widehat{\theta}_{\mathrm{BP}}$ also tends to be larger than that of the other kernel-based estimators. In terms of IQR, all of $\widehat{\theta}_{\mathrm{2SLS}}$, $\widehat{\theta}_{\mathrm{AP}}$, and $\widehat{\theta}_{\mathrm{RK}}$ are roughly equally precise. The Blundell-Powell estimator has the highest mid spread throughout. For the combined measures of bias and variability, that is, RMSE and RMAE, kernel-weighted least-squares does best; $\widehat{\theta}_{\mathrm{BP}}$ is on the other side of the spectrum. The local-rank estimator performs well according to both statistics of estimator risk, consistently reporting numbers that are close to those of the Ahn-Powell estimator. In allmost all designs, too, it outperforms 2SLS.

[11] It is well established that the choice of the particular form of the kernel matters far less than does the choice of the bandwidth in nonparametric estimation. Indeed, in additional experiments, I found very similar results as given here when using the quartic kernel and the cosine kernel.

Table 2: Monte Carlo results for the triangular model in (6.1)–(6.2); skewed-left design ($\gamma_0 \propto (2,1)'$)

$\mu_0^2 = 100$

		BIAS				STD				IQR				RMSE				RMAE			
ρ_X	ρ_U	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$
−.50	.50	−.0050	.0531	−.0719	.0817	.0975	.0714	.0909	.0828	.0743	.0964	.1064	.1138	.0976	.0890	.1159	.1163	.2391	.2664	.2894	.3088
		.0220	.0677	−.0569	.1030	.1677	.0905	.0682	.1078	.0745	.1133	.0901	.1423	.1692	.1131	.0888	.1491	.2618	.2899	.2637	.3398
−.50	.25	−.0049	.0213	−.0972	.0367	.1055	.0568	.0905	.0663	.0752	.0743	.1134	.0859	.1056	.0607	.1328	.0757	.2431	.2190	.3201	.2449
		.0302	.0272	−.0775	.0442	.1937	.0629	.0650	.0743	.0753	.0791	.0894	.0943	.1961	.0685	.1011	.0864	.2745	.2274	.2883	.2560
−.50	−.25	−.0051	−.0172	−.1292	−.0166	.1180	.0466	.0987	.0505	.0751	.0612	.1365	.0660	.1181	.0496	.1626	.0532	.2521	.1969	.3622	.2042
		.0492	−.0139	−.1003	−.0133	.2456	.0447	.0660	.0480	.0762	.0585	.0988	.0630	.2504	.0468	.1201	.0498	.3050	.1929	.3200	.1993
−.50	−.50	−.0090	−.0286	−.1374	−.0345	.1398	.0478	.1012	.0487	.0736	.0627	.1407	.0647	.1401	.0557	.1707	.0597	.2610	.2082	.3724	.2159
		.0640	−.0246	−.1058	−.0302	.2820	.0441	.0661	.0438	.0755	.0583	.1001	.0589	.2892	.0505	.1248	.0532	.3245	.2009	.3274	.2070
−.25	.50	−.0044	.0417	−.0624	.0740	.0913	.0693	.0832	.0842	.0704	.0907	.0973	.1152	.0914	.0809	.1039	.1121	.2291	.2527	.2743	.3015
		.0180	.0534	−.0500	.0926	.1517	.0843	.0646	.1062	.0703	.1029	.0843	.1403	.1528	.0998	.0817	.1409	.2475	.2712	.2527	.3284
−.25	.25	−.0035	.0162	−.0858	.0327	.0961	.0557	.0831	.0664	.0708	.0726	.1016	.0856	.0961	.0580	.1194	.0740	.2320	.2140	.3030	.2415
		.0249	.0215	−.0696	.0396	.1741	.0604	.0617	.0730	.0710	.0762	.0824	.0929	.1759	.0641	.0930	.0831	.2589	.2206	.2757	.2509
−.25	−.25	−.0045	−.0129	−.1145	−.0139	.1146	.0456	.0905	.0492	.0706	.0588	.1202	.0644	.1147	.0474	.1459	.0511	.2441	.1922	.3418	.1999
		.0440	−.0099	−.0908	−.0108	.2316	.0444	.0627	.0471	.0722	.0569	.0903	.0618	.2357	.0455	.1103	.0483	.2918	.1893	.3054	.1958
−.25	−.50	−.0077	−.0206	−.1213	−.0280	.1360	.0456	.0921	.0464	.0693	.0593	.1218	.0609	.1362	.0501	.1523	.0542	.2529	.1971	.3505	.2056
		.0575	−.0173	−.0957	−.0244	.2652	.0431	.0625	.0427	.0711	.0563	.0902	.0564	.2713	.0464	.1143	.0492	.3104	.1919	.3119	.1985
.25	.50	−.0049	.0325	−.0480	.0606	.0968	.0671	.0708	.0837	.0675	.0862	.0850	.1123	.0969	.0745	.0855	.1034	.2266	.2423	.2498	.2874
		.0205	.0420	−.0393	.0759	.1615	.0780	.0582	.1015	.0675	.0952	.0759	.1316	.1627	.0886	.0702	.1268	.2478	.2563	.2340	.3093
.25	.25	−.0074	.0117	−.0684	.0268	.1181	.0542	.0711	.0655	.0681	.0700	.0863	.0834	.1183	.0554	.0986	.0707	.2367	.2089	.2745	.2357
		.0317	.0164	−.0570	.0330	.2009	.0574	.0556	.0705	.0685	.0726	.0730	.0890	.2034	.0597	.0797	.0778	.2678	.2136	.2538	.2430
.25	−.25	−.0141	−.0095	−.0914	−.0096	.1549	.0427	.0773	.0466	.0688	.0546	.0984	.0610	.1555	.0438	.1197	.0475	.2585	.1846	.3073	.1932
		.0548	−.0069	−.0750	−.0069	.2651	.0420	.0569	.0451	.0698	.0534	.0785	.0591	.2707	.0426	.0941	.0457	.3080	.1826	.2796	.1903
.25	−.50	−.0209	−.0142	−.0963	−.0204	.1843	.0409	.0779	.0428	.0670	.0518	.0992	.0555	.1855	.0433	.1239	.0475	.2737	.1831	.3137	.1918
		.0719	−.0117	−.0788	−.0176	.3029	.0393	.0565	.0402	.0690	.0500	.0782	.0526	.3113	.0410	.0969	.0439	.3311	.1797	.2846	.1866
.50	.50	−.0104	.0318	−.0418	.0546	.1310	.0680	.0658	.0831	.0686	.0859	.0787	.1073	.1314	.0750	.0780	.0994	.2428	.2433	.2389	.2818
		.0338	.0413	−.0345	.0690	.2098	.0782	.0554	.0987	.0690	.0948	.0714	.1239	.2125	.0885	.0652	.1204	.2729	.2568	.2254	.3017
.50	.25	−.0158	.0112	−.0609	.0242	.1594	.0541	.0662	.0645	.0698	.0691	.0811	.0829	.1601	.0552	.0899	.0689	.2583	.2080	.2616	.2330
		.0473	.0158	−.0513	.0302	.2479	.0570	.0531	.0690	.0703	.0717	.0698	.0879	.2524	.0591	.0738	.0753	.2963	.2124	.2436	.2398
.50	−.25	−.0361	−.0094	−.0819	−.0085	.2262	.0410	.0721	.0454	.0705	.0524	.0913	.0584	.2291	.0421	.1091	.0461	.2986	.1809	.2921	.1899
		.0829	−.0069	−.0681	−.0059	.3301	.0403	.0544	.0440	.0721	.0511	.0745	.0569	.3403	.0408	.0871	.0444	.3517	.1789	.2677	.1872
.50	−.50	−.0477	−.0141	−.0865	−.0179	.2517	.0384	.0724	.0407	.0684	.0489	.0913	.0533	.2562	.0410	.1128	.0445	.3149	.1779	.2983	.1856
		.1018	−.0119	−.0718	−.0154	.3681	.0369	.0540	.0383	.0708	.0472	.0737	.0507	.3819	.0387	.0898	.0413	.3750	.1746	.2727	.1810

$\mu_0^2 = 20$

		BIAS				STD				IQR				RMSE				RMAE			
ρ_X	ρ_U	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$
−.50	.50	−.0342	.1400	−.2056	.1559	.2165	.0577	.1260	.0629	.2195	.0769	.1713	.0826	.2192	.1514	.2411	.1681	.3899	.3753	.4558	.3956
		.1601	.1875	−.1458	.2123	.4484	.0958	.0723	.1066	.2530	.1249	.1014	.1388	.4761	.2106	.1627	.2376	.5025	.4340	.3849	.4613
−.50	.25	−.0422	.0627	−.2451	.0750	.2211	.0596	.1071	.0671	.2114	.0799	.1443	.0901	.2250	.0865	.2675	.1006	.3919	.2689	.4954	.2897
		.1754	.0760	−.1713	.0909	.4757	.0743	.0590	.0844	.2511	.0971	.0763	.1117	.5070	.1063	.1812	.1240	.5179	.2914	.4146	.3148
−.50	−.25	−.0598	−.0516	−.3048	−.0465	.2249	.0515	.0929	.0589	.1984	.0680	.1335	.0789	.2327	.0729	.3186	.0750	.3940	.2451	.5521	.2475
		.1810	−.0449	−.2027	−.0403	.5036	.0456	.0415	.0521	.2378	.0592	.0589	.0686	.5351	.0640	.2069	.0658	.5281	.2326	.4502	.2348
−.50	−.50	−.0658	−.0937	−.3232	−.0908	.2274	.0469	.0884	.0517	.1841	.0615	.1267	.0682	.2367	.1048	.3351	.1045	.3947	.3079	.5685	.3045
		.1827	−.0804	−.2113	−.0781	.5119	.0372	.0375	.0405	.2340	.0478	.0528	.0532	.5435	.0886	.2145	.0880	.5328	.2857	.4596	.2825
−.25	.50	−.0340	.1343	−.1949	.1528	.2170	.0586	.1221	.0646	.2134	.0771	.1633	.0840	.2197	.1465	.2300	.1659	.3867	.3679	.4440	.3918
		.1552	.1783	−.1400	.2068	.4419	.0951	.0718	.1075	.2426	.1224	.0995	.1394	.4684	.2021	.1573	.2330	.4950	.4235	.3775	.4555
−.25	.25	−.0400	.0599	−.2356	.0732	.2185	.0598	.1043	.0674	.2048	.0794	.1392	.0898	.2221	.0846	.2577	.0995	.3871	.2653	.4857	.2873
		.1686	.0725	−.1668	.0885	.4662	.0739	.0556	.0843	.2435	.0955	.0754	.1109	.4957	.1036	.1758	.1223	.5093	.2868	.4088	.3116
−.25	−.25	−.0570	−.0488	−.2952	−.0445	.2231	.0516	.0902	.0583	.1905	.0676	.1258	.0770	.2303	.0710	.3087	.0733	.3894	.2414	.5434	.2446
		.1787	−.0424	−.1987	−.0385	.4984	.0462	.0413	.0520	.2295	.0592	.0572	.0675	.5294	.0627	.2030	.0647	.5229	.2296	.4458	.2325
−.25	−.50	−.0633	−.0883	−.3138	−.0868	.2282	.0469	.0854	.0510	.1783	.0614	.1199	.0668	.2368	.1000	.3252	.1007	.3916	.2995	.5602	.2982
		.1827	−.0761	−.2076	−.0750	.5083	.0378	.0370	.0404	.2290	.0484	.0515	.0525	.5401	.0850	.2109	.0852	.5288	.2787	.4556	.2773
.25	.50	−.0447	.1304	−.1777	.1490	.2438	.0580	.1140	.0649	.2299	.0773	.1517	.0840	.2479	.1427	.2111	.1625	.4062	.3626	.4242	.3870
		.1980	.1718	−.1307	.2004	.4926	.0935	.0694	.1069	.2830	.1206	.0968	.1355	.5309	.1956	.1480	.2271	.5348	.4158	.3649	.4484
.25	.25	−.0554	.0575	−.2187	.0712	.2458	.0593	.0979	.0674	.2192	.0785	.1288	.0888	.2519	.0826	.2396	.0980	.4073	.2615	.4679	.2847
		.2111	.0695	−.1583	.0860	.5186	.0731	.0541	.0838	.2862	.0938	.0732	.1084	.5598	.1008	.1673	.1201	.5489	.2821	.3982	.3083
.25	−.25	−.0750	−.0460	−.2767	−.0425	.2475	.0505	.0847	.0569	.1985	.0659	.1157	.0755	.2586	.0684	.2894	.0710	.4077	.2365	.5260	.2401
		.2063	−.0401	−.1906	−.0368	.5370	.0457	.0404	.0509	.2583	.0582	.0555	.0667	.5753	.0608	.1949	.0628	.5513	.2255	.4366	.2287
.25	−.50	−.0830	−.0829	−.2939	−.0820	.2500	.0448	.0801	.0488	.1825	.0584	.1097	.0642	.2634	.0942	.3047	.0954	.4097	.2904	.5422	.2900
		.2030	−.0720	−.1993	−.0713	.5422	.0366	.0363	.0392	.2465	.0467	.0501	.0510	.5789	.0808	.2026	.0813	.5520	.2712	.4464	.2706
.50	.50	−.0629	.1314	−.1708	.1482	.2771	.0570	.1095	.0636	.2581	.0761	.1451	.0829	.2841	.1432	.2029	.1613	.4339	.3637	.4158	.3859
		.2601	.1732	−.1271	.1990	.5555	.0927	.0676	.1050	.3661	.1192	.0942	.1345	.6133	.1964	.1439	.2250	.5886	.4172	.3596	.4468
.50	.25	−.0740	.0576	−.2110	.0703	.2722	.0587	.0946	.0664	.2375	.0782	.1242	.0884	.2821	.0822	.2312	.0967	.4293	.2609	.4596	.2826
		.2556	.0695	−.1544	.0847	.5669	.0725	.0530	.0827	.3490	.0935	.0722	.1077	.6218	.1004	.1632	.1184	.5897	.2814	.3932	.3059
.50	−.25	−.0974	−.0459	−.2678	−.0415	.2703	.0494	.0822	.0565	.2125	.0644	.1108	.0735	.2873	.0674	.2801	.0701	.4304	.2350	.5175	.2387
		.2303	−.0402	−.1866	−.0360	.5727	.0445	.0399	.0507	.2902	.0568	.0546	.0651	.6172	.0599	.1908	.0622	.5784	.2241	.4319	.2276
.50	−.50	−.1039	−.0823	−.2841	−.0806	.2670	.0432	.0777	.0481	.1908	.0567	.1041	.0633	.2865	.0930	.2945	.0939	.4280	.2889	.5330	.2880
		.2138	−.0717	−.1949	−.0703	.5670	.0351	.0361	.0388	.2636	.0454	.0490	.0504	.6060	.0798	.1982	.0803	.5700	.2700	.4415	.2691

$n = 100$, $\theta_0 = (.7071, -.7071)'$; standard-normal kernels; cross-validated bandwidths; 10,000 simulation runs.

Table 3: Monte Carlo results for the triangular model in (6.1)–(6.2); skewed-right design ($\gamma_0 \propto (1,2)'$)

$\mu_0^2 = 100$

		BIAS				STD				IQR				RMSE				RMAE			
ρ_X	ρ_U	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$
−.50	.50	−.0057	.0235	−.0552	.1187	.0730	.0769	.1598	.1236	.0863	.1008	.1909	.1933	.0732	.0804	.1690	.1714	.2336	.2530	.3468	.3733
		.0018	.0343	−.0228	.1520	.0721	.0871	.1292	.1580	.0860	.1086	.1753	.2428	.0721	.0936	.1312	.2192	.2324	.2651	.3188	.4127
−.50	.25	−.0035	.0113	−.1062	.0592	.0717	.0675	.1603	.0897	.0853	.0892	.2154	.1203	.0718	.0685	.1923	.1074	.2328	.2336	.3748	.2924
		.0046	.0185	−.0647	.0710	.0800	.0717	.1322	.1015	.0852	.0928	.1692	.1349	.0801	.0740	.1472	.1238	.2352	.2392	.3307	.3075
−.50	−.25	.0017	−.0142	−.1940	−.0310	.0718	.0643	.1626	.0696	.0855	.0858	.2456	.0942	.0718	.0658	.2531	.0762	.2332	.2282	.4511	.2460
		.0116	−.0083	−.1292	−.0239	.0934	.0635	.1044	.0651	.0860	.0828	.1509	.0865	.0941	.0640	.1661	.0693	.2416	.2252	.3758	.2372
−.50	−.50	.0031	−.0223	−.2211	−.0637	.0717	.0684	.1600	.0771	.0842	.0914	.2483	.1075	.0717	.0720	.2729	.1000	.2320	.2385	.4755	.2863
		.0141	−.0153	−.1462	−.0516	.1009	.0664	.0997	.0678	.0847	.0862	.1411	.0907	.1019	.0681	.1770	.0852	.2436	.2330	.3919	.2683
−.25	.50	−.0041	.0142	−.0349	.1007	.0621	.0685	.1168	.1141	.0746	.0893	.1355	.1774	.0623	.0699	.1219	.1522	.2168	.2348	.2948	.3490
		.0013	.0220	−.0176	.1253	.0618	.0751	.0964	.1415	.0744	.0934	.1284	.2142	.0619	.0783	.0980	.1890	.2162	.2424	.2778	.3798
−.25	.25	−.0024	.0066	−.0735	.0485	.0616	.0607	.1191	.0817	.0748	.0796	.1469	.1099	.0616	.0611	.1399	.0950	.2164	.2201	.3182	.2733
		.0034	.0122	−.0519	.0575	.0666	.0635	.0907	.0911	.0746	.0815	.1249	.1205	.0667	.0647	.1045	.1077	.2181	.2236	.2882	.2852
−.25	−.25	.0017	−.0086	−.1371	−.0231	.0623	.0573	.1250	.0621	.0753	.0756	.1728	.0823	.0623	.0580	.1855	.0663	.2174	.2138	.3810	.2290
		.0089	−.0039	−.1007	−.0177	.0786	.0572	.0830	.0591	.0756	.0741	.1231	.0771	.0791	.0574	.1305	.0617	.2237	.2124	.3309	.2226
−.25	−.50	.0030	−.0120	−.1564	−.0466	.0630	.0597	.1256	.0671	.0737	.0780	.1763	.0926	.0630	.0609	.2006	.0817	.2168	.2189	.4008	.2569
		.0109	−.0068	−.1142	−.0383	.0842	.0594	.0794	.0612	.0740	.0757	.1195	.0816	.0849	.0598	.1391	.0722	.2252	.2166	.3448	.2445
.25	.50	−.0028	.0132	−.0308	.0770	.0533	.0593	.0818	.0957	.0645	.0770	.0982	.1431	.0533	.0608	.0874	.1228	.2010	.2182	.2526	.3107
		.0013	.0190	−.0215	.0935	.0539	.0647	.0709	.1158	.0643	.0801	.0922	.1655	.0539	.0674	.0741	.1489	.2010	.2246	.2408	.3337
.25	.25	−.0012	.0054	−.0572	.0357	.0534	.0529	.0820	.0704	.0647	.0698	.0988	.0936	.0534	.0532	.1000	.0789	.2011	.2049	.2722	.2482
		.0034	.0097	−.0455	.0423	.0608	.0551	.0666	.0774	.0646	.0711	.0865	.1003	.0609	.0559	.0806	.0882	.2035	.2077	.2529	.2571
.25	−.25	.0022	−.0058	−.0940	−.0144	.0561	.0488	.0868	.0526	.0649	.0638	.1110	.0687	.0562	.0491	.1279	.0546	.2038	.1963	.3154	.2069
		.0086	−.0024	−.0755	−.0107	.0770	.0488	.0636	.0508	.0651	.0629	.0880	.0660	.0774	.0489	.0987	.0519	.2109	.1955	.2853	.2030
.25	−.50	.0033	−.0068	−.1039	−.0291	.0577	.0492	.0875	.0531	.0637	.0629	.1131	.0703	.0578	.0496	.1358	.0606	.2039	.1970	.3269	.2192
		.0112	−.0034	−.0832	−.0244	.0879	.0494	.0620	.0500	.0640	.0619	.0874	.0650	.0886	.0495	.1037	.0557	.2142	.1960	.2940	.2121
.50	.50	−.0023	.0185	−.0329	.0666	.0533	.0575	.0713	.0863	.0617	.0738	.0864	.1211	.0534	.0604	.0785	.1090	.1986	.2174	.2400	.2918
		.0020	.0243	−.0254	.0808	.0572	.0636	.0619	.1041	.0617	.0779	.0802	.1393	.0573	.0681	.0669	.1318	.1994	.2252	.2286	.3127
.50	.25	−.0008	.0072	−.0552	.0309	.0540	.0509	.0714	.0659	.0625	.0669	.0878	.0851	.0540	.0514	.0903	.0728	.1994	.2011	.2602	.2385
		.0048	.0111	−.0456	.0369	.0706	.0532	.0584	.0718	.0625	.0684	.0768	.0907	.0708	.0543	.0741	.0807	.2038	.2043	.2428	.2464
.50	−.25	.0024	−.0065	−.0835	−.0118	.0614	.0453	.0758	.0493	.0627	.0589	.0958	.0633	.0614	.0458	.1127	.0507	.2045	.1892	.2965	.1989
		.0118	−.0036	−.0688	−.0086	.0974	.0451	.0571	.0477	.0630	.0579	.0779	.0610	.0981	.0453	.0894	.0485	.2159	.1881	.2712	.1957
.50	−.50	.0031	−.0083	−.0906	−.0234	.0669	.0447	.0764	.0471	.0618	.0573	.0971	.0618	.0670	.0454	.1185	.0526	.2058	.1884	.3054	.2034
		.0158	−.0054	−.0745	−.0198	.1150	.0443	.0562	.0445	.0624	.0562	.0776	.0580	.1160	.0446	.0933	.0487	.2224	.1868	.2781	.1978

$\mu_0^2 = 20$

		BIAS				STD				IQR				RMSE				RMAE			
ρ_X	ρ_U	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$	$\widehat{\theta}_{2SLS}$	$\widehat{\theta}_{AP}$	$\widehat{\theta}_{BP}$	$\widehat{\theta}_{RK}$
−.50	.50	−.0165	.1189	−.2720	.1675	.1486	.0700	.1780	.0855	.1627	.0950	.2506	.1107	.1495	.1380	.3251	.1881	.3282	.3512	.5308	.4121
		.0323	.1578	−.1592	.2327	.2134	.1049	.1689	.1378	.1645	.1426	.1237	.1871	.2158	.1895	.2321	.2705	.3483	.4022	.4346	.4846
−.50	.25	−.0122	.0574	−.3282	.0864	.1427	.0691	.1406	.0815	.1622	.0921	.1889	.1060	.1433	.0898	.3570	.1188	.3246	.2730	.5740	.3133
		.0425	.0717	−.1980	.1067	.2346	.0840	.1323	.1031	.1636	.1105	.0775	.1348	.2384	.1105	.2381	.1484	.3588	.2952	.4601	.3426
−.50	−.25	−.0024	−.0542	−.4085	−.0573	.1396	.0665	.1056	.0717	.1594	.0895	.1377	.0975	.1396	.0858	.4220	.0918	.3221	.2648	.6392	.2743
		.0685	−.0450	−.2381	−.0478	.2760	.0601	.0798	.0629	.1664	.0773	.0427	.0829	.2843	.0751	.2511	.0790	.3836	.2507	.4925	.2579
−.50	−.50	−.0008	−.0978	−.4345	−.1137	.1392	.0667	.0956	.0672	.1572	.0887	.1265	.0894	.1391	.1184	.4449	.1321	.3197	.3216	.6591	.3421
		.0761	−.0812	−.2489	−.0936	.2890	.0543	.0499	.0515	.1652	.0680	.0355	.0654	.2988	.0977	.2539	.1069	.3890	.2956	.5003	.3117
−.25	.50	−.0129	.1088	−.2205	.1590	.1340	.0688	.1577	.0862	.1481	.0946	.2263	.1137	.1346	.1287	.2711	.1809	.3118	.3371	.4791	.4021
		.0264	.1418	−.1454	.2170	.1920	.1003	.1080	.1348	.1489	.1360	.1282	.1836	.1938	.1737	.1811	.2555	.3291	.3825	.3981	.4684
−.25	.25	−.0085	.0517	−.2757	.0802	.1311	.0668	.1294	.0794	.1474	.0892	.1797	.1046	.1313	.0845	.3045	.1128	.3100	.2635	.5263	.3043
		.0365	.0642	−.1818	.0980	.2123	.0799	.0835	.0986	.1491	.1050	.0863	.1298	.2154	.1025	.2001	.1390	.3397	.2833	.4309	.3307
−.25	−.25	.0007	−.0469	−.3543	−.0507	.1310	.0632	.1018	.0676	.1457	.0852	.1415	.0909	.1310	.0787	.3686	.0845	.3097	.2529	.5952	.2622
		.0632	−.0391	−.2221	−.0426	.2591	.0581	.0420	.0600	.1505	.0750	.0531	.0785	.2667	.0700	.2260	.0736	.3668	.2410	.4713	.2480
−.25	−.50	.0023	−.0839	−.3793	−.1001	.1300	.0631	.0939	.0635	.1435	.0835	.1339	.0838	.1300	.1050	.3907	.1185	.3071	.3008	.6159	.3225
		.0734	−.0706	−.2323	−.0837	.2798	.0530	.0353	.0503	.1513	.0663	.0461	.0635	.2892	.0883	.2350	.0976	.3756	.2791	.4820	.2964
.25	.50	−.0110	.1079	−.1777	.1491	.1353	.0634	.1285	.0789	.1450	.0861	.1746	.1039	.1357	.1252	.2193	.1687	.3102	.3338	.4285	.3887
		.0342	.1387	−.1276	.1999	.2105	.0926	.0807	.1231	.1471	.1232	.1109	.1633	.2133	.1667	.1510	.2348	.3351	.3767	.3663	.4492
.25	.25	−.0072	.0497	−.2267	.0734	.1346	.0623	.1090	.0735	.1448	.0833	.1468	.0970	.1348	.0797	.2515	.1039	.3101	.2555	.4770	.2919
		.0474	.0607	−.1606	.0887	.2383	.0745	.0660	.0903	.1480	.0972	.0813	.1181	.2430	.0961	.1737	.1266	.3506	.2739	.4027	.3155
.25	−.25	−.0008	−.0421	−.2945	−.0437	.1392	.0567	.0910	.0611	.1430	.0753	.1247	.0810	.1392	.0706	.3082	.0751	.3118	.2391	.5427	.2468
		.0805	−.0357	−.1983	−.0373	.2987	.0523	.0424	.0548	.1513	.0672	.0568	.0715	.3094	.0633	.2027	.0663	.3859	.2288	.4453	.2347
.25	−.50	−.0009	−.0747	−.3154	−.0857	.1427	.0545	.0860	.0561	.1391	.0711	.1216	.0745	.1427	.0924	.3269	.1024	.3117	.2825	.5616	.2988
		.0935	−.0642	−.2082	−.0733	.3230	.0465	.0374	.0456	.1510	.0580	.0521	.0587	.3362	.0792	.2115	.0863	.3976	.2643	.4563	.2775
.50	.50	−.0137	.1142	−.1691	.1465	.1539	.0601	.1171	.0725	.1573	.0812	.1565	.0960	.1545	.1291	.2057	.1634	.3268	.3412	.4159	.3845
		.0559	.1471	−.1244	.1956	.2673	.0905	.0738	.1149	.1613	.1187	.1018	.1504	.2730	.1727	.1446	.2268	.3690	.3862	.3587	.4437
.50	.25	−.0125	.0516	−.2139	.0713	.1535	.0604	.1005	.0699	.1569	.0800	.1319	.0922	.1540	.0794	.2363	.0998	.3273	.2553	.4631	.2862
		.0727	.0625	−.1551	.0859	.3025	.0728	.0566	.0865	.1639	.0937	.0759	.1124	.3111	.0960	.1652	.1219	.3891	.2740	.3946	.3094
.50	−.25	−.0115	−.0425	−.2761	−.0417	.1629	.0534	.0854	.0587	.1508	.0700	.1159	.0761	.1633	.0683	.2890	.0720	.3300	.2354	.5255	.2415
		.1098	−.0366	−.1903	−.0358	.3646	.0492	.0411	.0528	.1668	.0624	.0556	.0677	.3808	.0613	.1947	.0638	.4272	.2252	.4362	.2302
.50	−.50	−.0136	−.0755	−.2952	−.0821	.1678	.0497	.0814	.0518	.1462	.0646	.1130	.0688	.1683	.0904	.3062	.0971	.3306	.2810	.5433	.2913
		.1246	−.0655	−.1997	−.0709	.3893	.0418	.0370	.0421	.1658	.0526	.0515	.0549	.4087	.0777	.2031	.0825	.4400	.2630	.4469	.2715

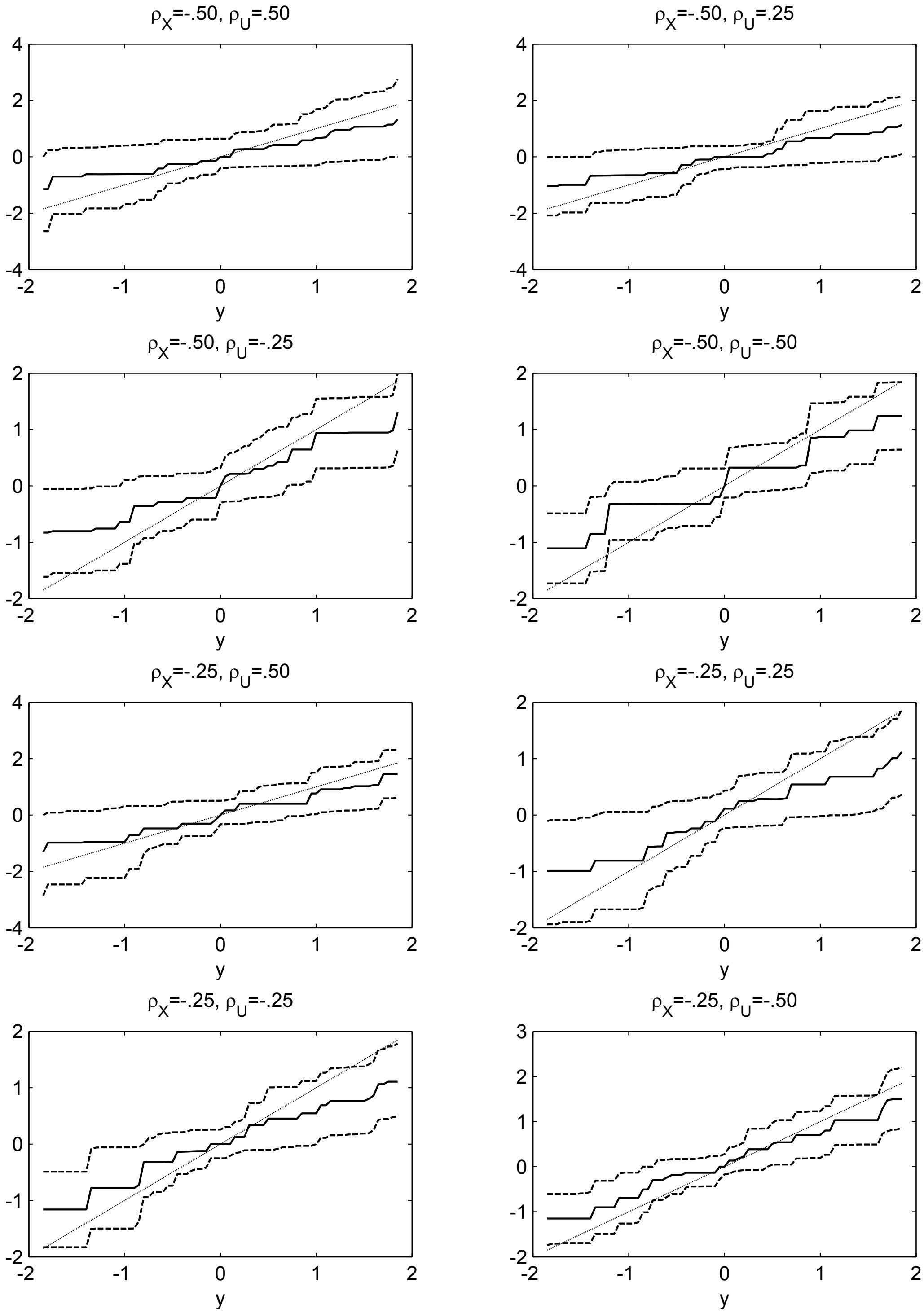
$n = 100$, $\theta_0 = (.7071, -.7071)'$; standard-normal kernels; cross-validated bandwidths; 10,000 simulation runs.

On weakening the instrument strength by setting $\mu_0^2 = 20$ all estimators report higher bias numbers, but their relative performance remains largely unaltered. Moreover, $\widehat{\theta}_{\text{2SLS}}$ and $\widehat{\theta}_{\text{AP}}$ continue to provide the most accurate point estimates, on average; $\widehat{\theta}_{\text{RK}}$ remains close to $\widehat{\theta}_{\text{AP}}$ throughout. The Blundell-Powell estimator has the largest bias across all entries. Lowering the concentration parameter affects the variability of the estimators too. This effect is particularly noticable on 2SLS, whose STD and IQR go up by 100% or more. This is well documented in the literature. The other estimators' volatility is far less influenced by this design change. In fact, in some cases, their STD and IQR can be seen to decrease. Measured by their RMSE and RMAE, the Ahn-Powell estimator and the local-rank estimator perform best overal, as they also did when $\mu_0^2 = 100$. This implies that, when evaluating performance in terms of estimator risk, these estimators are to be prefered over the optimal GMM estimator for the linear model with endogenous regressors, at least in the designs considered here.

When the main source of exogenous variation in E comes from X_1, the covariate in the main equation of interest, and μ_0^2 is reset to 100, $\widehat{\theta}_{\text{2SLS}}$ and $\widehat{\theta}_{\text{AP}}$ tend to report slightly higher biases. The average error of the Blundell-Powell estimator, in contrast, has a tendency to decrease relative to the balanced design. The local-rank estimator reacts more erratic to this parameter shift, with its average across simulation runs sometimes being further away from the true value, and sometimes closer to it. Nevertheless, the bias remains reasonably small and comparable in magnitude to that of $\widehat{\theta}_{\text{AP}}$. When looking at the STD and IQR we can see that the kernel-based estimators behave differently than does 2SLS. The latter's precision decreases, as would be expected; the former's does so to a much smaller extent and actually decrease in many of the situations. In terms of RMSE, $\widehat{\theta}_{\text{AP}}$ and $\widehat{\theta}_{\text{RK}}$ report the best numbers, often drastically superior to $\widehat{\theta}_{\text{2SLS}}$. Now, also $\widehat{\theta}_{\text{BP}}$ positions itself competitively against 2SLS. The same pattern, although less pronounced, emerges when looking at the RMAE.

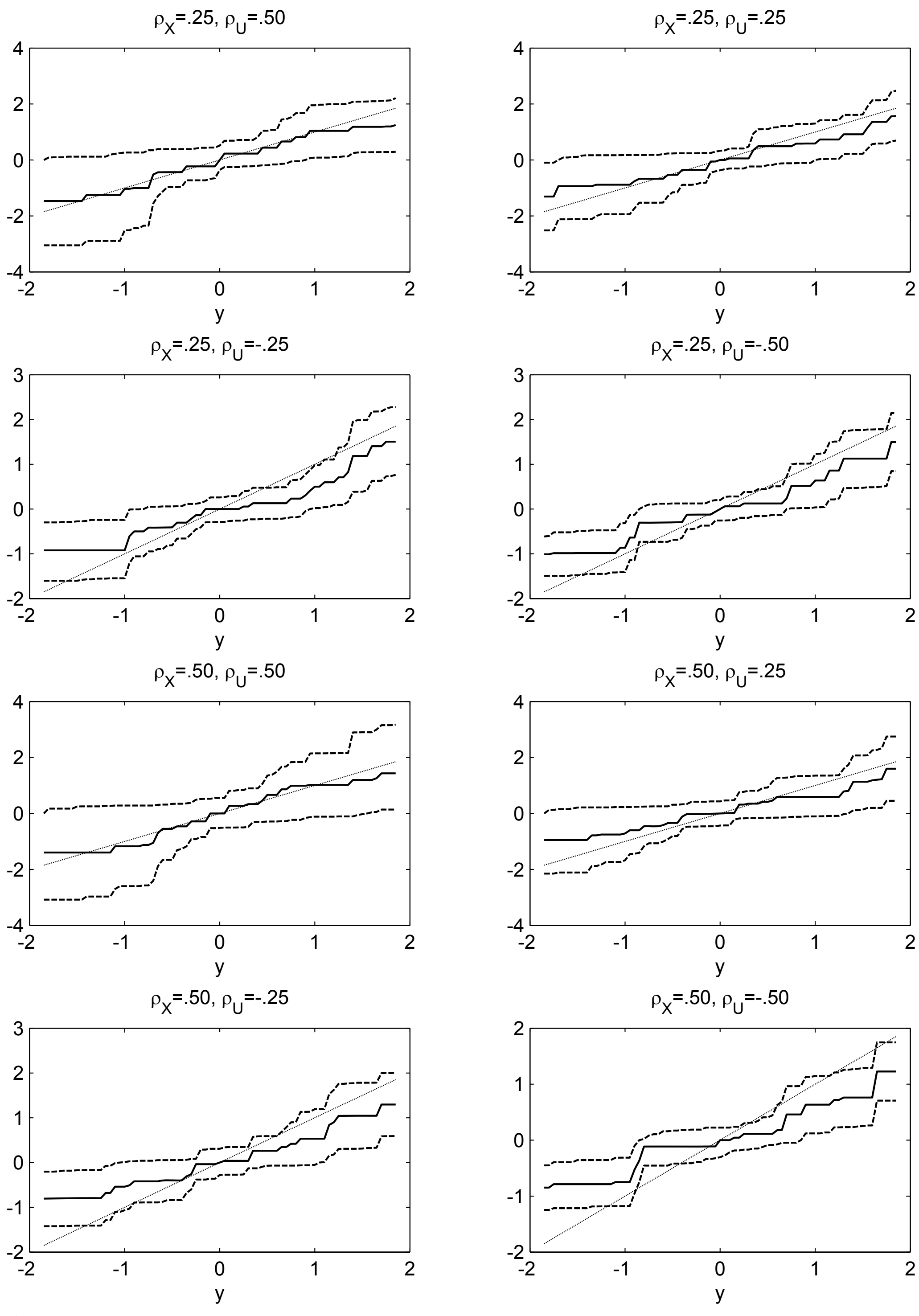
Lowering μ_0^2 to 20 has a similar effect as it did in the balanced case. That is, the bias increases and there is a mixed effect on the precision of the estimators. A look at the results for the skewed-right design in Table 3 reveals them to be in line with what has been observed before. The most important difference compared to Table 2 is that 2SLS tends to be less variable. This should be no surprise as, here, more of the exogenous variation in E comes from X_2, the instrumental variable that was excluded from the equation for Y. So, we find that the local-rank performs solid across the designs considered, positioning itself competitively against the alternative procedures considered.

Figure 1: Estimates of the transformation function in the triangular model



$\widehat{\psi}(y)$ (solid), $\psi_0(y)$ (dotted), pointwise 95% confidence bound (dashed) using an estimated standard error obtained over 200 bootstrap replications.

Figure 2: Estimates of the transformation function in the triangular model (contd.)



$\widehat{\psi}(y)$ (solid), $\psi_0(y)$ (dotted), pointwise 95% confidence bound (dashed) using an estimated standard error obtained over 200 bootstrap replications.

Figures 1–2 collect estimates of the transformation function in the balanced design with $\mu_0^2 = 100$. For each of the combinations of ρ_X and ρ_U considered in Table 1 one estimate of ψ_0 was calculated at equally-distant points in $[-1.85, 1.85]$, with a step size of .05, using $\widehat{\theta}_{\text{RK}}$ as the first-step estimator.[12] The kernel function and bandwidth used to perform the weighting in the second estimation step were taken to be the same as those in the first step, and $y_0 = 0$ for graphical elegance. Pointwise 95% confidence bounds are also reported. The standard errors used to form the bounds were obtained as the standard deviation of the empirical distribution function of the estimator obtained over 200 bootstrap replications.

Given the small sample size, the estimator does fairly well. All graphs suggest $\widehat{\psi}$ to be fairly antisymmetric around y_0, overestimating (underestimating) $\psi_0(y)$ for $y < y_0$ ($y > y_0$). Not surprisingly, $\|\widehat{\psi}(y) - \psi_0(y)\|$ is largest for y furthest away from y_0. Given the design, the empirical density of y has its largest concentration of mass around zero, implying that very few observations contribute to $\widehat{q}_n^y(\psi, \widehat{\theta})$ for y in the tails of the aformentioned distribution. This leads to $\widehat{\psi}$ being essentially flat on the edges of the interval considered. The little information in the data about the transformation function in these areas is also reflected in the standard errors, with the confidence intervals tending to become more wide as y moves further away from y_0. Nevertheless, the confidence bounds appear informative about the shape of ψ_0. They present clear evidence against ψ_0 being highly-nonlinear and tend to contain ψ_0 on a large subset of the interval considered.

6.2 A sample-selection model

The second Monte Carlo experiment is centered around the linear sample-selection model

$$Y = E \times (X_1\theta_0^{(1)} + X_2\theta_0^{(2)} + U_1), \qquad E = 1(X_1\gamma_0^{(1)} + X_2\gamma_0^{(2)} + X_3\gamma_0^{(3)} \geq -U_2). \quad (6.3)$$

Here, the disturbances (U_1, U_2) were again standard-normal with correlation ρ_U, as in (6.4). The designs considered vary in the dependence between these disturbances and in the correlation between the regressors, which were drawn as

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N} \begin{pmatrix} 1 & \rho_{X_1X_2} & \rho_{X_1X_3} \\ \rho_{X_1X_2} & 1 & \rho_{X_2X_3} \\ \rho_{X_1X_3} & \rho_{X_1X_2} & 1 \end{pmatrix}. \quad (6.4)$$

In this setup, which fits into Example 2, $Z = (X_1, X_2, X_3)'$ and

$$\vartheta_0(e, z) = \mu_E(z) = \Pr[E = 1 | Z = z],$$

[12]The Blundell-Powell estimator could be used here as well, but its higher variability is transmitted into the rank estimator of the transformation function.

which is free of E.

Throughout, θ_0 was kept fixed at $(.7071, -.7071)'$ and γ_0 was set as in the previous Monte Carlo exercise, with $\pi_0 = (1, 1, 1,)'$ and $\mu_0^2 = 300$. Several combinations of $\rho_U, \rho_{X_1X_2}, \rho_{X_1X_3}$, and $\rho_{X_2X_3}$ in $\{-.50, -.25, .25, .50\}$ were considered. For each design point, θ_0 was again estimated $10,000$ times from a sample of $n = 500$. With the chosen parameter constellations, this resulted in an average sample size of about 250 observations for the second estimation step. The standard deviation on this effective sample size ranged from 70 to 90 observations across designs. In Tables 4 and 5, the column for $\widetilde{n}$ contains the average effective number of observations used in the second estimation step; the standard deviation around this mean is stated in parenthesis below each entry.

The estimators reported on are the three kernel-weighted estimators from before, that is, $\widehat{\theta}_{\text{AP}}$, $\widehat{\theta}_{\text{BP}}$, and $\widehat{\theta}_{\text{RK}}$. The Ahn-Powell estimator, which was designed with the linear sample-selection model in mind, would be the optimal choice from this set. The choice for the kernel functions and the data-driven procedure to select the bandwidths from the previous Monte Carlo experiment was maintained here.

Tables 4 and 5 show that the Ahn-Powell estimator tends to perform well, reporting solid numbers throughout. Overal, the results do not lean in favor of one estimator in particular. The local-rank estimator again closely mimics the kernel-weighted least-squares estimator, with the differences between their respective bias and spread being consistently small. In several design, the Blundell-Powell also performs well. In other cases, however, it is heavily biased and behaves very volatile. As a consequence, also its RMSE and RMAE takes on large values in such cases. The numbers are particularly worrysome when X_1 and X_2 are negatively correlated. The performance is worst when $\rho_{X_1X_2}$ equals $-.50$, the strongest negative correlation considered. In such cases the bias of $\widehat{\theta}_{\text{BP}}$ can be as much as 35 times larger than that of the others. Similarly, its variance inflates by a factor of 10 compared to those of $\widehat{\theta}_{\text{AP}}$ and $\widehat{\theta}_{\text{RK}}$. No such variability across the designs is observed for these latter two estimators.

Thus, the local-rank estimator performs well. It was found not to be dominated by the optimal GMM estimator for the triangular model and its performance was similar to that of the Ahn-Powell estimator, both when estimating the triangular model and the sample-selection model. It compares favorably Blundell and Powell (2004), which is the most general alternative currently available, being more stable in performance across designs, and often much less biased and far less volatile. The local-rank estimator thus seems a strong candidate for the estimation of weakly-separable models with controls.

Table 4: Monte Carlo results for the sample-selection model in (6.3)–(6.4)

$\rho_{X_1X_2}$	$\rho_{X_1X_3}$	$\rho_{X_2X_3}$	ρ_U	$\bar{n}$ (STD)	BIAS $\hat\theta_{AP}$	BIAS $\hat\theta_{BP}$	BIAS $\hat\theta_{RK}$	STD $\hat\theta_{AP}$	STD $\hat\theta_{BP}$	STD $\hat\theta_{RK}$	IQR $\hat\theta_{AP}$	IQR $\hat\theta_{BP}$	IQR $\hat\theta_{RK}$	RMSE $\hat\theta_{AP}$	RMSE $\hat\theta_{BP}$	RMSE $\hat\theta_{RK}$	RMAE $\hat\theta_{AP}$	RMAE $\hat\theta_{BP}$	RMAE $\hat\theta_{RK}$
.50	.50	.25	.50	250.97	.0249	.0662	.0549	.0483	.0524	.0623	.0631	.0676	.0783	.0543	.0844	.0830	.2098	.2682	.2580
				(66.65)	.0295	.0788	.0622	.0519	.0666	.0691	.0682	.0829	.0886	.0597	.1031	.0929	.2172	.2900	.2702
.50	.50	.25	.25	251.03	.0112	.0434	.0329	.0504	.0557	.0566	.0666	.0724	.0718	.0516	.0706	.0654	.2028	.2392	.2270
				(66.62)	.0152	.0519	.0368	.0525	.0661	.0596	.0691	.0825	.0764	.0547	.0840	.0700	.2067	.2543	.2325
.50	.50	.25	−.25	251.16	−.0191	−.0138	−.0166	.0535	.0616	.0512	.0697	.0776	.0658	.0568	.0631	.0538	.2099	.2202	.2056
				(66.70)	−.0149	−.0083	−.0137	.0508	.0608	.0483	.0663	.0751	.0622	.0530	.0613	.0502	.2050	.2172	.2001
.50	.50	.25	−.50	250.76	−.0334	−.0442	−.0414	.0532	.0633	.0518	.0695	.0792	.0691	.0628	.0772	.0663	.2220	.2434	.2323
				(65.61)	−.0284	−.0368	−.0364	.0485	.0555	.0464	.0638	.0708	.0611	.0561	.0666	.0589	.2132	.2310	.2218
.25	.50	.25	.50	251.01	.0282	.0768	.0595	.0501	.0617	.0650	.0654	.0784	.0832	.0574	.0985	.0881	.2159	.2896	.2655
				(68.41)	.0334	.0948	.0680	.0545	.0820	.0734	.0713	.0998	.0954	.0640	.1253	.1001	.2245	.3180	.2796
.25	.50	.25	.25	251.03	.0125	.0502	.0354	.0530	.0669	.0601	.0699	.0836	.0755	.0544	.0836	.0698	.2084	.2600	.2346
				(68.45)	.0169	.0627	.0400	.0555	.0822	.0639	.0729	.0975	.0813	.0580	.1034	.0754	.2128	.2799	.2412
.25	.50	.25	−.25	251.18	−.0213	−.0150	−.0185	.0556	.0743	.0547	.0715	.0920	.0715	.0596	.0758	.0577	.2152	.2393	.2132
				(68.59)	−.0166	−.0071	−.0152	.0526	.0730	.0513	.0678	.0890	.0672	.0551	.0733	.0535	.2097	.2359	.2070
.25	.50	.25	−.50	250.79	−.0373	−.0497	−.0454	.0557	.0770	.0552	.0727	.0954	.0728	.0670	.0916	.0715	.2296	.2628	.2406
				(67.43)	−.0316	−.0392	−.0396	.0502	.0696	.0488	.0660	.0842	.0638	.0593	.0799	.0628	.2197	.2480	.2289
−.25	.50	.25	.50	251.13	.0397	.1397	.0735	.0586	.1199	.0747	.0769	.1468	.0973	.0708	.1841	.1048	.2405	.4027	.2911
				(72.85)	.0480	.2515	.0870	.0664	.2470	.0889	.0871	.2643	.1159	.0819	.3525	.1243	.2540	.5162	.3119
−.25	.50	.25	.25	251.12	.0174	.0958	.0431	.0635	.1484	.0719	.0832	.1771	.0929	.0658	.1766	.0839	.2299	.3839	.2580
				(72.81)	.0242	.2080	.0505	.0676	.2903	.0783	.0881	.2762	.1017	.0718	.3571	.0931	.2367	.4908	.2679
−.25	.50	.25	−.25	251.16	−.0293	−.0345	−.0246	.0666	.1963	.0678	.0881	.2294	.0894	.0728	.1993	.0721	.2382	.3858	.2384
				(72.85)	−.0225	.0864	−.0195	.0619	.3519	.0624	.0819	.2366	.0827	.0659	.3623	.0653	.2300	.4468	.2295
−.25	.50	.25	−.50	250.83	−.0517	−.1104	−.0598	.0668	.2017	.0685	.0878	.2439	.0912	.0845	.2299	.0909	.2584	.4118	.2705
				(71.60)	−.0430	.0186	−.0505	.0582	.3596	.0581	.0767	.1966	.0774	.0724	.3600	.0770	.2439	.4353	.2536
−.50	.50	.25	.50	251.19	.0535	.1641	.0893	.0701	.1692	.0863	.0912	.1444	.1116	.0882	.2357	.1242	.2703	.4633	.3189
				(75.43)	.0671	.9206	.1100	.0834	.4085	.1076	.1083	.4879	.1413	.1071	1.0072	.1539	.2909	.9653	.3482
−.50	.50	.25	.25	250.6	.0251	.1138	.0538	.0766	.1990	.0862	.1023	.2065	.1125	.0806	.2292	.1016	.2549	.4462	.2842
				(74.22)	.0355	.9980	.0651	.0834	.4493	.0962	.1118	.5284	.1268	.0907	1.0945	.1161	.2658	1.0070	.2985
−.50	.50	.25	−.25	251.17	−.0403	−.0361	−.0315	.0830	.2611	.0853	.1099	.3725	.1132	.0922	.2636	.0910	.2678	.4584	.2680
				(75.42)	−.0296	1.0559	−.0239	.0751	.6001	.0768	.0995	.7334	.1019	.0807	1.2144	.0804	.2555	1.0488	.2554
−.50	.50	.25	−.50	251.16	−.0730	−.1315	−.0767	.0837	.2748	.0855	.1103	.4270	.1154	.1111	.3046	.1149	.2977	.4948	.3050
				(75.46)	−.0584	1.0565	−.0627	.0698	.6865	.0698	.0912	1.0100	.0924	.0910	1.2599	.0938	.2755	1.0602	.2812
.50	.25	.25	.50	251.02	.0217	.0440	.0522	.0470	.0521	.0623	.0617	.0657	.0793	.0518	.0682	.0813	.2040	.2362	.2539
				(68.41)	.0259	.0518	.0591	.0503	.0606	.0691	.0660	.0751	.0894	.0566	.0797	.0909	.2105	.2503	.2654
.50	.25	.25	.25	250.64	.0114	.0233	.0310	.0483	.0547	.0547	.0639	.0702	.0673	.0496	.0595	.0629	.1989	.2178	.2222
				(67.42)	.0150	.0289	.0347	.0501	.0599	.0574	.0663	.0754	.0718	.0523	.0665	.0671	.2025	.2258	.2274
.50	.25	.25	−.25	251.12	−.0162	−.0288	−.0163	.0501	.0582	.0485	.0649	.0727	.0639	.0526	.0649	.0512	.2024	.2228	.2010
				(68.53)	−.0125	−.0233	−.0135	.0482	.0548	.0463	.0622	.0675	.0601	.0498	.0595	.0482	.1985	.2152	.1961
.50	.25	.25	−.50	251.05	−.0286	−.0540	−.0389	.0490	.0595	.0488	.0635	.0751	.0648	.0568	.0804	.0624	.2114	.2522	.2258
				(68.54)	−.0245	−.0461	−.0342	.0457	.0509	.0446	.0588	.0652	.0579	.0518	.0687	.0562	.2044	.2377	.2166
.25	.25	.25	.50	251.08	.0244	.0493	.0555	.0480	.0582	.0646	.0612	.0731	.0839	.0538	.0763	.0851	.2078	.2502	.2591
				(70.49)	.0290	.0597	.0633	.0519	.0727	.0726	.0660	.0851	.0953	.0594	.0941	.0963	.2152	.2678	.2721
.25	.25	.25	.25	250.67	.0125	.0258	.0331	.0504	.0628	.0583	.0669	.0808	.0728	.0519	.0679	.0670	.2035	.2323	.2294
				(69.42)	.0165	.0333	.0375	.0524	.0702	.0617	.0697	.0879	.0782	.0549	.0777	.0722	.2076	.2422	.2356
.25	.25	.25	−.25	251.16	−.0180	−.0328	−.0177	.0519	.0685	.0513	.0686	.0839	.0677	.0549	.0759	.0543	.2073	.2397	.2073
				(70.62)	−.0140	−.0255	−.0146	.0497	.0628	.0486	.0655	.0772	.0638	.0516	.0678	.0507	.2028	.2304	.2019
.25	.25	.25	−.50	251	−.0324	−.0608	−.0423	.0512	.0699	.0524	.0661	.0868	.0699	.0606	.0926	.0673	.2189	.2693	.2341
				(70.62)	−.0276	−.0505	−.0369	.0474	.0589	.0472	.0606	.0742	.0620	.0548	.0776	.0599	.2109	.2519	.2235
−.25	.25	.25	.50	251.2	.0334	.0723	.0660	.0544	.1089	.0731	.0698	.1285	.0945	.0638	.1307	.0985	.2274	.3247	.2797
				(75.35)	.0400	.1131	.0777	.0607	.1669	.0860	.0773	.1656	.1103	.0727	.2016	.1159	.2384	.3724	.2982
−.25	.25	.25	.25	250.69	.0159	.0345	.0388	.0584	.1221	.0686	.0764	.1401	.0880	.0605	.1269	.0788	.2194	.3126	.2486
				(74.20)	.0215	.0705	.0454	.0615	.1728	.0744	.0805	.1616	.0954	.0651	.1866	.0872	.2252	.3454	.2576
−.25	.25	.25	−.25	251.17	−.0236	−.0606	−.0219	.0604	.1420	.0636	.0808	.1551	.0836	.0648	.1544	.0673	.2258	.3342	.2305
				(75.43)	−.0181	−.0183	−.0171	.0572	.1890	.0597	.0761	.1357	.0784	.0600	.1898	.0621	.2196	.3300	.2233
−.25	.25	.25	−.50	251.14	−.0432	−.1068	−.0529	.0605	.1389	.0646	.0802	.1607	.0858	.0743	.1753	.0835	.2431	.3635	.2596
				(75.50)	−.0363	−.0599	−.0449	.0545	.1790	.0561	.0715	.1241	.0739	.0654	.1887	.0719	.2316	.3399	.2449
−.50	.25	.25	.50	251.25	.0415	.0991	.0765	.0625	.1918	.0829	.0821	.2202	.1072	.0750	.2158	.1128	.2472	.4279	.2999
				(78.26)	.0510	.3554	.0929	.0720	.4355	.1019	.0935	.5058	.1301	.0882	.5621	.1379	.2623	.6312	.3243
−.50	.25	.25	.25	250.64	.0196	.0337	.0460	.0674	.2338	.0798	.0872	.2933	.1039	.0702	.2361	.0921	.2365	.4379	.2696
				(77.18)	.0273	.3631	.0556	.0722	.5275	.0889	.0932	.6233	.1151	.0772	.6403	.1049	.2445	.6595	.2820
−.50	.25	.25	−.25	251.22	−.0298	−.1304	−.0280	.0711	.2609	.0774	.0949	.3809	.1015	.0771	.2917	.0823	.2460	.4789	.2553
				(78.41)	−.0221	.2780	−.0209	.0664	.6445	.0714	.0881	.5845	.0933	.0700	.7018	.0744	.2376	.6643	.2454
−.50	.25	.25	−.50	251.13	−.0545	−.2072	−.0647	.0718	.2526	.0776	.0943	.3861	.1017	.0901	.3267	.1010	.2677	.5131	.2851
				(78.43)	−.0446	.1970	−.0533	.0627	.6529	.0656	.0818	.3743	.0853	.0769	.6819	.0845	.2521	.6446	.2658

$n = 500$, $\theta_0 = (.7071, -.7071)'$; standard-normal kernels; cross-validated bandwidths; 10,000 simulation runs.

Table 5: Monte Carlo results for the sample-selection model in (6.3)–(6.4) (contd.)

					BIAS			STD			IQR			RMSE			RMAE		
$\rho_{X_1X_2}$	$\rho_{X_1X_3}$	$\rho_{X_2X_3}$	ρ_U	$\frac{\bar{n}}{(\text{STD})}$	$\widehat{\theta}_{\text{AP}}$	$\widehat{\theta}_{\text{BP}}$	$\widehat{\theta}_{\text{RK}}$	$\widehat{\theta}_{\text{AP}}$	$\widehat{\theta}_{\text{BP}}$	$\widehat{\theta}_{\text{RK}}$	$\widehat{\theta}_{\text{AP}}$	$\widehat{\theta}_{\text{BP}}$	$\widehat{\theta}_{\text{RK}}$	$\widehat{\theta}_{\text{AP}}$	$\widehat{\theta}_{\text{BP}}$	$\widehat{\theta}_{\text{RK}}$	$\widehat{\theta}_{\text{AP}}$	$\widehat{\theta}_{\text{BP}}$	$\widehat{\theta}_{\text{RK}}$
.50	−.25	.25	.50	251.20	.0229	.0102	.0490	.0467	.0587	.0613	.0601	.0716	.0765	.0520	.0596	.0785	.2035	.2145	.2474
				(72.80)	.0271	.0154	.0557	.0507	.0594	.0687	.0644	.0743	.0858	.0574	.0614	.0884	.2107	.2178	.2589
.50	−.25	.25	.25	251.08	.0104	−.0162	.0280	.0491	.0633	.0553	.0646	.0760	.0696	.0502	.0654	.0620	.1994	.2204	.2192
				(72.84)	.0142	−.0101	.0317	.0515	.0649	.0584	.0667	.0735	.0734	.0534	.0656	.0665	.2032	.2164	.2243
.50	−.25	.25	−.25	250.73	−.0137	−.0663	−.0153	.0483	.0644	.0469	.0643	.0798	.0606	.0502	.0925	.0494	.1991	.2706	.1976
				(71.73)	−.0102	−.0557	−.0128	.0472	.0577	.0449	.0621	.0674	.0573	.0483	.0802	.0466	.1962	.2526	.1931
.50	−.25	.25	−.50	249.70	−.0262	−.0877	−.0357	.0474	.0629	.0468	.0628	.0816	.0612	.0541	.1079	.0588	.2077	.3007	.2193
				(71.96)	−.0223	−.0739	−.0316	.0447	.0488	.0430	.0586	.0649	.0551	.0500	.0886	.0534	.2016	.2772	.2110
.25	−.25	.25	.50	251.20	.0230	.0104	.0493	.0468	.0629	.0623	.0610	.0776	.0780	.0522	.0637	.0795	.2044	.2219	.2488
				(75.35)	.0272	.0164	.0562	.0507	.0649	.0700	.0653	.0807	.0871	.0576	.0669	.0897	.2115	.2258	.2606
.25	−.25	.25	.25	251.15	.0105	−.0158	.0286	.0490	.0680	.0568	.0653	.0820	.0730	.0501	.0698	.0636	.1998	.2290	.2222
				(75.40)	.0143	−.0093	.0326	.0513	.0650	.0603	.0675	.0793	.0774	.0532	.0656	.0685	.2036	.2246	.2277
.25	−.25	.25	−.25	250.71	−.0145	−.0690	−.0163	.0493	.0721	.0494	.0647	.0874	.0638	.0513	.0998	.0520	.2009	.2788	.2021
				(74.23)	−.0109	−.0569	−.0133	.0479	.0626	.0470	.0622	.0735	.0604	.0492	.0846	.0489	.1977	.2590	.1972
.25	−.25	.25	−.50	249.65	−.0274	−.0909	−.0378	.0484	.0703	.0496	.0630	.0892	.0660	.0556	.1150	.0623	.2098	.3081	.2247
				(74.51)	−.0233	−.0751	−.0331	.0454	.0595	.0451	.0586	.0707	.0594	.0510	.0958	.0559	.2032	.2830	.2154
−.25	−.25	.25	.50	251.29	.0269	.0044	.0535	.0511	.0967	.0686	.0658	.1081	.0861	.0577	.0968	.0870	.2152	.2660	.2599
				(81.71)	.0322	.0210	.0624	.0561	.1169	.0796	.0715	.1121	.0964	.0646	.1188	.1011	.2239	.2738	.2747
−.25	−.25	.25	.25	250.72	.0131	−.0291	.0301	.0533	.1051	.0640	.0694	.1160	.0816	.0549	.1090	.0707	.2085	.2792	.2344
				(80.60)	.0177	−.0092	.0358	.0561	.1272	.0692	.0723	.1105	.0873	.0588	.1275	.0779	.2133	.2759	.2419
−.25	−.25	.25	−.25	249.57	−.0173	−.0957	−.0184	.0556	.1165	.0604	.0728	.1327	.0797	.0582	.1508	.0632	.2138	.3368	.2225
				(81.11)	−.0128	−.0654	−.0140	.0534	.1261	.0571	.0699	.1058	.0752	.0549	.1420	.0587	.2095	.3095	.2164
−.50	−.25	.25	−.50	251.14	−.0330	−.1220	−.0435	.0539	.1129	.0596	.0700	.1338	.0780	.0632	.1662	.0737	.2237	.3618	.2428
				−81.86	−.0279	−.0868	−.0371	.0499	.1207	.0531	.0642	.0990	.0692	.0572	.1487	.0648	.2153	.3265	.2310
−.50	−.25	.25	.50	251.32	.0315	−.0004	.0554	.0562	.1413	.0739	.0725	.1506	.0921	.0644	.1413	.0924	.2282	.3192	.2683
				(85.81)	.0382	.0512	.0664	.0623	.2243	.0874	.0799	.1580	.1060	.0731	.2300	.1098	.2389	.3517	.2857
−.50	−.25	.25	.25	250.82	.0151	−.0482	.0332	.0597	.1583	.0735	.0770	.1710	.0913	.0616	.1654	.0807	.2209	.3431	.2502
				(84.70)	.0209	.0168	.0410	.0630	.2535	.0808	.0811	.1587	.0986	.0664	.2541	.0906	.2266	.3623	.2596
−.50	−.25	.25	−.25	249.53	−.0207	−.1334	−.0199	.0617	.1731	.0708	.0808	.1996	.0889	.0651	.2186	.0735	.2258	.4050	.2394
				(85.22)	−.0151	−.0356	−.0140	.0587	.2989	.0666	.0768	.1482	.0837	.0607	.3009	.0680	.2204	.4007	.2324
−.50	−.25	.25	−.50	251.26	−.0383	−.1671	−.0466	.0601	.1651	.0691	.0797	.1996	.0917	.0713	.2349	.0834	.2373	.4273	.2572
				(85.97)	−.0319	−.0731	−.0386	.0548	.2691	.0612	.0720	.1334	.0805	.0634	.2788	.0724	.2272	.3994	.2436
.50	−.50	.25	.50	251.19	.0315	−.0437	.0506	.0492	.0745	.0588	.0642	.0887	.0737	.0585	.0864	.0776	.2162	.2543	.2476
				(75.43)	.0371	−.0338	.0578	.0555	.0736	.0669	.0705	.0799	.0832	.0667	.0810	.0884	.2265	.2421	.2603
.50	−.50	.25	.25	251.14	.0149	−.0743	.0285	.0504	.0744	.0546	.0674	.0919	.0682	.0525	.1051	.0615	.2043	.2891	.2183
				(75.41)	.0191	−.0609	.0323	.0537	.0650	.0582	.0705	.0759	.0719	.0570	.0890	.0666	.2097	.2681	.2240
.50	−.50	.25	−.25	250.75	−.0166	−.1278	−.0162	.0498	.0743	.0477	.0647	.0933	.0596	.0525	.1479	.0504	.2038	.3591	.1995
				(74.22)	−.0129	−.1022	−.0136	.0483	.0628	.0454	.0621	.0670	.0562	.0500	.1200	.0474	.2001	.3237	.1946
.50	−.50	.25	−.50	249.67	−.0325	−.1464	−.0375	.0481	.0698	.0460	.0629	.0923	.0594	.0581	.1622	.0594	.2160	.3831	.2204
				(74.59)	−.0281	−.1166	−.0335	.0448	.0469	.0420	.0576	.0629	.0533	.0529	.1257	.0537	.2084	.3421	.2117
.25	−.50	.25	.50	251.23	.0260	−.0265	.0481	.0485	.0773	.0610	.0626	.0918	.0755	.0550	.0817	.0777	.2099	.2456	.2463
				(78.26)	.0308	−.0174	.0550	.0534	.0761	.0687	.0677	.0868	.0846	.0616	.0780	.0880	.2182	.2380	.2583
.25	−.50	.25	.25	251.15	.0126	−.0570	.0277	.0505	.0805	.0564	.0662	.0962	.0722	.0520	.0986	.0628	.2032	.2736	.2207
				(78.37)	.0167	−.0453	.0318	.0532	.0688	.0599	.0689	.0834	.0767	.0558	.0824	.0678	.2078	.2561	.2264
.25	−.50	.25	−.25	250.76	−.0148	−.1136	−.0163	.0497	.0822	.0492	.0655	.1008	.0627	.0519	.1402	.0518	.2027	.3414	.2020
				(77.17)	−.0111	−.0899	−.0134	.0485	.0760	.0470	.0631	.0755	.0594	.0497	.1177	.0488	.1994	.3094	.1972
.25	−.50	.25	−.50	249.63	−.0292	−.1329	−.0378	.0487	.0788	.0491	.0634	.1027	.0649	.0568	.1545	.0620	.2127	.3661	.2244
				(77.45)	−.0250	−.1056	−.0332	.0456	.0596	.0447	.0586	.0731	.0580	.0520	.1212	.0557	.2058	.3279	.2152
−.25	−.50	.25	.50	251.35	.0257	−.0305	.0474	.0511	.1036	.0659	.0667	.1129	.0814	.0572	.1080	.0812	.2139	.2753	.2510
				(85.81)	.0309	−.0093	.0554	.0559	.1353	.0758	.0722	.1067	.0904	.0638	.1356	.0939	.2223	.2737	.2642
−.25	−.50	.25	.25	250.77	.0124	−.0661	.0260	.0534	.1120	.0635	.0693	.1206	.0789	.0548	.1300	.0686	.2082	.3046	.2305
				(84.72)	.0169	−.0395	.0314	.0560	.1380	.0680	.0720	.1039	.0837	.0585	.1436	.0749	.2127	.2900	.2370
−.25	−.50	.25	−.25	249.51	−.0162	−.1272	−.0171	.0542	.1198	.0592	.0703	.1371	.0764	.0565	.1747	.0616	.2103	.3695	.2197
				(85.22)	−.0119	−.0858	−.0129	.0523	.1436	.0565	.0675	.1002	.0721	.0536	.1672	.0579	.2065	.3350	.2143
−.25	−.50	.25	−.50	248.55	−.0292	−.1519	−.0388	.0537	.1164	.0591	.0690	.1377	.0767	.0611	.1914	.0707	.2187	.3957	.2367
				(85.10)	−.0244	−.1048	−.0330	.0500	.1381	.0535	.0639	.0950	.0685	.0556	.1734	.0628	.2113	.3531	.2261
−.50	−.50	.25	.50	251.42	.0285	−.0314	.0460	.0546	.1342	.0695	.0699	.1388	.0839	.0616	.1378	.0834	.2226	.3099	.2542
				(91.03)	.0346	.0115	.0551	.0602	.2038	.0803	.0763	.1334	.0938	.0694	.2041	.0973	.2322	.3221	.2685
−.50	−.50	.25	.25	251.36	.0131	−.0707	.0257	.0578	.1460	.0699	.0756	.1543	.0874	.0592	.1622	.0745	.2164	.3384	.2399
				(91.01)	.0184	−.0176	.0324	.0605	.2202	.0760	.0791	.1331	.0929	.0632	.2209	.0826	.2213	.3382	.2475
−.50	−.50	.25	−.25	250.97	−.0191	−.1480	−.0193	.0590	.1562	.0685	.0763	.1828	.0876	.0620	.2152	.0712	.2204	.4059	.2355
				(89.98)	−.0139	−.0743	−.0135	.0567	.2290	.0654	.0728	.1289	.0826	.0584	.2407	.0668	.2157	.3757	.2294
−.50	−.50	.25	−.50	249.59	−.0339	−.1687	−.0384	.0588	.1507	.0687	.0778	.1776	.0885	.0679	.2262	.0787	.2312	.4244	.2487
				(90.24)	−.0280	−.0938	−.0314	.0543	.2144	.0621	.0712	.1174	.0789	.0610	.2340	.0695	.2222	.3841	.2373

$n = 500$, $\theta_0 = (.7071, -.7071)'$; standard-normal kernels; cross-validated bandwidths; 10,000 simulation runs.

Euclidean properties. The class of functions $\mathscr{H} \equiv \{\mathrm{s}(\cdot,\cdot,\theta)\mathrm{t}(\cdot)\mathrm{t}(\cdot)\mathrm{k}((\cdot - \cdot)/\sigma_{\mathrm{k}}) : \theta \in \Theta\}$ with $\sigma_{\mathrm{k}} > 0$ and $\lim_{n\to\infty} \sigma_{\mathrm{k}} = 0$ is Euclidean for the envelope $\mathcal{H}(\cdot,\cdot) \equiv \sup_{\eta\in\mathscr{R}^{\dim(\vartheta_0)}} \|\mathrm{k}(\eta)\| \; \|[\mathrm{m}(\cdot) + \mathrm{m}(\cdot)]\mathrm{t}(\cdot)\mathrm{t}(\cdot)\|$. To see this, notice that $\mathscr{H}$ is a subclass of the class $\overline{\mathscr{H}} \equiv \{\mathrm{s}(\cdot,\cdot,\theta)\mathrm{t}(\cdot)\mathrm{t}(\cdot)\mathrm{k}((\cdot - \cdot)/\sigma) : \theta \in \Theta, \sigma > 0\} = \mathscr{S}\mathscr{K}$, with $\mathscr{S} \equiv \{\mathrm{s}(\cdot,\cdot,\theta)\mathrm{t}(\cdot)\mathrm{t}(\cdot) : \theta \in \Theta\}$ and $\mathscr{K} \equiv \{\mathrm{k}((\cdot - \cdot)/\sigma) : \sigma > 0\}$. By Assumption 6, the class $\mathscr{K}$ is Euclidean for the constant envelope $\sup_{\eta\in\mathscr{R}^{\dim(\vartheta_0)}} \|\mathrm{k}(\eta)\|$; see Example 2.10 in Pakes and Pollard (1989). Likewise, the class $\mathscr{S}$ is Euclidean for the envelope $\|[\mathrm{m}(\cdot) + \mathrm{m}(\cdot)]\mathrm{t}(\cdot)\mathrm{t}(\cdot)\|$. This follows by Assumption 5 together with the discussion in Cavanagh and Sherman (1998). The envelope $\mathcal{H}(\cdot,\cdot)$, then, follows from Lemma 2.14 in Pakes and Pollard (1989).

To prevent some expressions in the Appendix from becoming overly lenghty, it is useful to define ahead the functions $\mathcal{H}'(\cdot,\cdot) \equiv \sup_{\eta\in\mathscr{R}^{\dim(\vartheta_0)}} \|\mathrm{k}'(\eta)\| \; \|[\mathrm{m}(\cdot) + \mathrm{m}(\cdot)]\mathrm{t}(\cdot)\mathrm{t}(\cdot)\|$ and $\mathcal{H}''(\cdot,\cdot) \equiv \sup_{\eta\in\mathscr{R}^{\dim(\vartheta_0)}} \|\mathrm{k}''(\eta)\| \; \|[\mathrm{m}(\cdot) + \mathrm{m}(\cdot)]\mathrm{t}(\cdot)\mathrm{t}(\cdot)\|$, where notation has been abused slightly to keep the analogy to $\mathcal{H}(\cdot,\cdot)$ transparent. Notice that these functions are well behaved because, by Assumption 6, both k' and k'' are bounded.

Lemma A. *Let Assumptions 5–9 hold. Then* $\mathrm{q}_n(\theta) - \mathrm{q}_n(\theta_0)$ *is*

$$(\theta-\theta_0)'\frac{\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)}{2}(\theta-\theta_0)+\frac{(\theta-\theta_0)'}{\sqrt{n}}\Big[2\sqrt{n}\nabla_{\theta}\overline{\tau}(\mathrm{P}_n,\theta_0)+o_p(1)\Big]+o_p(\|\theta-\theta_0\|^2)+o_p\Big(\frac{1}{n}\Big)$$

uniformly over $o_p(1/\sqrt{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}n}) = o_p(1)$ *neighborhoods of* θ_0.

Lemma B. *Let Assumption 10 hold. Then*

$$\overline{\zeta}(\mathrm{P}_n,\theta) - \overline{\zeta}(\mathrm{P}_n,\theta_0) = \frac{(\theta-\theta_0)'}{\sqrt{n}}\big[\sqrt{n}\nabla_{\theta}\overline{\zeta}(\mathrm{P}_n,\theta_0)\big] + o_p(\|\theta-\theta_0\|^2)$$

uniformly over $o_p(1)$ *neighborhoods of* θ_0.

Estimators of the components of the influence function. Rosenblatt-Parzen kernel estimates of $\mathrm{p}_I(I(x,z))$ and its first derivative are

$$\widehat{\mathrm{p}}_I(\widehat{I}(x,z)) \equiv \frac{1}{n\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+1}}\sum_{i=1}^{n}\mathrm{j}\Big(\frac{\widehat{I}(x,w)-\widehat{I}(X_i,W_i)}{\sigma_{\mathrm{j}}}\Big) \qquad \text{and} \tag{A.1}$$

$$\widehat{\mathrm{p}}_I'(\widehat{I}(x,w)) \equiv \frac{1}{n\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+2}}\sum_{i=1}^{n}\mathrm{j}'\Big(\frac{\widehat{I}(x,w)-\widehat{I}(X_i,W_i)}{\sigma_{\mathrm{j}}}\Big), \tag{A.2}$$

respectively.

Nadaraya-Watson estimates of $\mu_{\mathrm{t}(Z)}(I(x,w))$, $\mu_{\mathrm{t}(Z)X}(I(x,w))$, and $\mu_{\mathrm{m}(Y)}(I(x,w))$ are given by

$$\widehat{\mu}_{\mathrm{t}(Z)}(\widehat{I}(x,w)) \equiv \frac{1}{n\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+1}} \sum_{i=1}^{n} \frac{\mathrm{t}(Z_i)}{\widehat{\mathrm{p}}_I\big(\widehat{I}(x,w)\big)} \mathrm{j}\Big(\frac{\widehat{I}(x,w) - \widehat{I}(X_i,W_i)}{\sigma_{\mathrm{j}}}\Big), \tag{A.3}$$

$$\widehat{\mu}_{\mathrm{t}(Z)X}(\widehat{I}(x,w)) \equiv \frac{1}{n\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+1}} \sum_{i=1}^{n} \frac{\mathrm{t}(Z_i)X_i}{\widehat{\mathrm{p}}_I\big(\widehat{I}(x,w)\big)} \mathrm{j}\Big(\frac{\widehat{I}(x,w) - \widehat{I}(X_i,W_i)}{\sigma_{\mathrm{j}}}\Big), \text{ and} \tag{A.4}$$

$$\widehat{\mu}_{\mathrm{m}(Y)}(\widehat{I}(x,w)) \equiv \frac{1}{n\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+1}} \sum_{i=1}^{n} \frac{\mathrm{m}(Y_i)}{\widehat{\mathrm{p}}_I\big(\widehat{I}(x,w)\big)} \mathrm{j}\Big(\frac{\widehat{I}(x,w) - \widehat{I}(X_i,W_i)}{\sigma_J}\Big), \tag{A.5}$$

respectively

Then $\widehat{\mathcal{X}}(x,w) \equiv \mathrm{t}(z)\ \widehat{\mu}_{\mathrm{t}(Z)}(\widehat{I}(x,w))\Big[x - \frac{\widehat{\mu}_{\mathrm{t}(Z)X}(\widehat{I}(x,w))}{\widehat{\mu}_{\mathrm{t}(Z)}(\widehat{I}(x,w))}\Big]$ and

$$\widehat{\mathcal{S}}\big(y,\widehat{I}(x,w)\big) \equiv \mathrm{m}(y) - \widehat{\mu}_{\mathrm{m}(Y)}(\widehat{I}(x,w)) \tag{A.6}$$

constitute plug-in estimates of $\mathcal{X}(x,w)$ and $\mathcal{S}(y,I(x,w))$, respectively. On differentiating this latter quantity, we find estimates of $\mathcal{S}_1(y,I(x,w))$ and $\mathcal{S}_2(y,I(x,w))$. For the former quantity, for example, the estimate $\widehat{\mathcal{S}}_1(y,\widehat{I}(x,w))$ is

$$\frac{1}{n\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+2}} \sum_{i=1}^{n} \frac{\mathrm{m}(Y_i)}{\widehat{\mathrm{p}}_I(\widehat{I}(x,w))} \mathrm{j}_1\Big(\frac{\widehat{I}(x,w) - \widehat{I}(X_i,W_i)}{\sigma_{\mathrm{j}}}\Big) - \frac{\widehat{\mu}_{\mathrm{m}(Y)}(\widehat{I}(x,w))\ \widehat{\mathrm{p}}_{I1}(\widehat{I}(x,w))}{\widehat{\mathrm{p}}_I(\widehat{I}(x,w))} \tag{A.7}$$

where j_1 and $\widehat{\mathrm{p}}_{I1}$ are the first components of j' and $\widehat{\mathrm{p}}'_I$, respectively. Observe that the derivatives of $\widehat{\mathcal{S}}(y,\widehat{I}(x,w))$ are free of y.

Finally, for $\nu(x,w) \equiv \mathcal{X}(x,w)\ \mathcal{S}_2(y,I(x,w))\mathrm{p}_I(I(x,w))$ and $\mu_{\nu(X,W)}(z)$, the necessary kernel estimates are

$$\widehat{\nu}(x,w) \equiv \widehat{\mathcal{X}}(x,w)\ \widehat{\mathcal{S}}_2(y,\widehat{I}(x,w))\ \widehat{\mathrm{p}}_I(\widehat{I}(x,w)) \text{ and} \tag{A.8}$$

$$\widehat{\mu}_{\nu(X,W)}(z) \equiv \sum_{i=1}^{n} \widehat{\nu}(X_i,W_i)\ \omega_i(z), \tag{A.9}$$

where $\omega_i(z) \equiv \mathrm{l}\Big(\frac{z-Z_i}{\sigma_{\mathrm{l}}}\Big)/\sum_{i=1}^{N}\mathrm{l}\Big(\frac{z-Z_i}{\sigma_{\mathrm{l}}}\Big)$.

Lemma C. *Let the conditions for Lemma 3 and Assumptions 12–13 hold. Then, if* $j < \frac{1-\epsilon/2-\lambda\dim(Z)}{2(\dim(\vartheta_0)+3)}$*, the kernel estimators in (A.1)–(A.9) consistently estimate their population counterparts.*

Proof of Lemma 1. The proof follows from standard kernel-smoothing arguments; see, e.g., Collomb and Härdle (1986) and Aradillas-López (2010). □

Proof of Theorem 1. Given random sampling, and by the construction of Θ, showing consistency of $\widehat{\theta}$ for θ_0 amounts to verifying that (i) $\sup_{\theta\in\Theta}\left\|\widehat{\mathrm{q}}_n(\theta)-\overline{\mathrm{q}}(\theta)\right\| = o_p(1)$; and that (ii) $\overline{\mathrm{q}}(\theta)$ is continuous and reaches its unique global maximum on Θ when evaluated at θ_0; see, e.g., Newey and McFadden (1994).

To show (i), observe that the triangle inequality provides the bound

$$\begin{aligned}\sup_{\theta\in\Theta}\left\|\widehat{\mathrm{q}}_n(\theta)-\overline{\mathrm{q}}(\theta)\right\| \le & \sup_{\theta\in\Theta}\left\|\widehat{\mathrm{q}}_n(\theta)-\mathrm{q}_n(\theta)\right\| + \sup_{\theta\in\Theta}\left\|\mathrm{q}_n(\theta)-\mathrm{q}(\theta)\right\| \\ & + \sup_{\theta\in\Theta}\left\|\mathrm{q}(\theta)-\overline{\mathrm{q}}(\theta)\right\|,\end{aligned} \tag{B.1}$$

where $\mathrm{q}(\theta) \equiv \mathbb{P}[\mathrm{h}(\cdot,\cdot,\theta)]$.

The first right-hand side term in (B.1) captures the estimation error in the controls. Recall that $\widehat{\mathrm{q}}_n(\theta)-\mathrm{q}_n(\theta) = \mathbb{P}_n\big[\,\widehat{\mathrm{h}}(\cdot,\cdot,\theta)-\mathrm{h}(\cdot,\cdot,\theta)\big]$ or, equivalently,

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\mathrm{s}(V_i,V_j,\theta)\mathrm{t}(Z_i)\mathrm{t}(Z_j)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}}\Big[\mathrm{k}\Big(\frac{\widehat{\vartheta}(W_i)-\widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}}\Big)-\mathrm{k}\Big(\frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}\Big)\Big].$$

Take a first-order expansion of $\mathrm{k}\Big(\frac{\widehat{\vartheta}(W_i)-\widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}}\Big)$ around $\frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}$ and next exploit symmetry to write $\sup_{\theta\in\Theta}\|\widehat{\mathrm{q}}_n(\theta)-\mathrm{q}_n(\theta)\|$ as

$$\sup_{\theta\in\Theta}\left\|\binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\mathrm{s}(V_i,V_j,\theta)\mathrm{t}(Z_i)\mathrm{t}(Z_j)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}}\;[\widehat{\vartheta}(W_i)-\vartheta_0(W_i)]'\;\mathrm{k}'(*)\right\|,$$

where $\mathrm{k}'(*)$ is k' evaluated in a $\dim(\vartheta_0)$-vector that lies inbetween $\frac{\widehat{\vartheta}(W_i)-\widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}}$ and $\frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}$; the $*$-notation will be reserved for such purposes throughout. Then

$$\begin{aligned}\sup_{\theta\in\Theta}\left\|\widehat{\mathrm{q}}_n(\theta)-\mathrm{q}_n(\theta)\right\| &\le 2\;\mathbb{P}_n[\mathcal{H}'(\cdot,\cdot)]\;\frac{\sup_{z\in\mathscr{Z}}\|\widehat{\mu}_{\mathrm{a}(E)}(z)-\mu_{\mathrm{a}(E)}(z)\|}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}} \\ &= \frac{1}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}}\mathcal{O}_p\Big(\sqrt{\frac{n^{\epsilon/2}}{n\sigma_{\mathrm{l}}^{\dim(Z)}}}\Big) = o_p(1).\end{aligned}$$

The first step follows by Assumption 6 and the Euclidean properties of the class $\mathscr{S}$ above, and the next two transitions follow by Lemma 1(i) and by Assumptions 2 and 7, respectively.

The second right-hand side term in (B.1) involves a zero-mean U-process of order two. Because the class $\mathscr{H}$ is Euclidean for an envelope whose second moment exists by Assumption 5, Corollary 7 of Sherman (1994b) can be applied to get

$$\sup_{\theta\in\Theta}\Big\|\mathrm{q}_n(\theta)-\mathrm{q}(\theta)\Big\| = \sup_{\theta\in\Theta}\Big\|\mathbb{P}_n\big[\mathrm{h}(\cdot,\cdot,\theta)\big]-\mathbb{P}\big[\mathrm{h}(\cdot,\cdot,\theta)\big]\Big\| = \frac{1}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}}\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big) = \mathcal{o}_p(1),$$

with the last transition following again by Assumption 7.

For the non-stochastic term in (B.1), finally, recall that $\mathrm{q}(\theta)=\tau(\mathrm{P},\theta)$. Standard kernel-smoothing arguments, as validated by Assumptions 6–8, can be used to show that $\sup_{\theta\in\Theta}\|\tau(\mathrm{P},\theta)\| = \sup_{\theta\in\Theta}\|\overline{\tau}(\mathrm{P},\theta)\| + \mathcal{o}(1)$. Because such arguments will be used at various subsequent stages of the Appendix, it is detailed only here. First, rewrite $\tau(d,\theta)$ as a kernel-weighted average of $\overline{\varphi}(v,\vartheta_0(W),\theta)$ under P, i.e.,

$$\tau(d,\theta) = \mathrm{t}(z)\int \frac{\overline{\varphi}(v,\vartheta,\theta)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}}\ \mathrm{k}\Big(\frac{\vartheta_0(w)-\vartheta}{\sigma_{\mathrm{k}}}\Big)\ \mathrm{d}\vartheta.$$

Next, observe that, by a mean-value expansion of $\overline{\varphi}(v,\vartheta,\theta)$ around $\vartheta_0(w)$ followed by a change of variable from ϑ to $\eta\equiv\frac{\vartheta_0(w)-\vartheta}{\sigma_{\mathrm{k}}}$,

$$\begin{aligned}
&\sup_{\theta\in\Theta}\Big\|\tau(d,\theta)-\overline{\tau}(d,\theta)\Big\| \\
\le&\sup_{\theta\in\Theta}\Big\|\mathrm{t}(z)\ \overline{\varphi}(v,\vartheta_0(w),\theta)-\overline{\tau}(d,\theta)\Big\| + \sigma_{\mathrm{k}}\ \sup_{\theta\in\Theta}\Big\|\mathrm{t}(z)\int\nabla_{\vartheta}\overline{\varphi}(v,*,\theta)\ \eta\ \mathrm{k}(\eta)\ \mathrm{d}\eta\Big\| \\
\le&\ \sigma_{\mathrm{k}}\ \mathrm{t}(z)\int\sup_{\theta\in\Theta}\big\|\nabla_{\vartheta}\overline{\varphi}(v,*,\theta)\big\|\ \big\|\eta\big\|\ \big\|\mathrm{k}(\eta)\big\|\ \mathrm{d}\eta = \mathcal{O}(\sigma_{\mathrm{k}}).
\end{aligned}$$

Then, by dominated convergence and Assumption 7,

$$\sup_{\theta\in\Theta}\Big\|\mathrm{q}(\theta)-\overline{\mathrm{q}}(\theta)\Big\| = \mathcal{O}(\sigma_{\mathrm{k}}) = \mathcal{o}(1).$$

This establishes (i).

Assumption 4 ensures (ii). This can be shown by small modifications to the argument in Cavanagh and Sherman (1998); see also Manski (1985) and Han (1987), among others. Because the details are standard and lengthy, they are omitted here. □

Proof of Lemma 2. Let $\mathrm{d}_n(\theta) \equiv \widehat{\mathrm{q}}_n(\theta) - \mathrm{q}_n(\theta)$. The proof then boils down to characterizing $\mathrm{d}_n(\theta)$ up to $\mathscr{o}_p(1/\sqrt{n})$. The point of departure is a second-order expansion of $\widehat{\mathrm{q}}_n(\theta)$ around $\frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}$. On invoking symmetry,

$$\mathrm{d}_n(\theta) = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{j\neq i} \frac{\mathrm{s}(V_i,V_j,\theta)\mathrm{t}(Z_i)\mathrm{t}(Z_j)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}} \, [\widehat{\vartheta}(W_i) - \vartheta(W_i)]' \, \mathrm{k}'\Big(\frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}\Big)$$

up to a remainder term $\mathrm{r}_n(\theta)$, say, that captures the contribution of

$$\Big[\frac{\widehat{\vartheta}(W_i)-\widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}} - \frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}\Big]' K''(*) \Big[\frac{\widehat{\vartheta}(W_i)-\widehat{\vartheta}(W_j)}{\sigma_{\mathrm{k}}} - \frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}\Big]$$

to $\mathrm{d}_n(\theta)$. The remainder term can be ignored for our purposes because

$$\mathrm{r}_n(\theta) \leq 2 \; \mathbb{P}_n[\mathcal{H}''(\cdot,\cdot)] \; \frac{\sup_{z\in\mathscr{Z}} \|\widehat{\mu}_{\mathrm{a}(E)}(z) - \mu_{\mathrm{a}(E)}(z)\|^2}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+2}} = \mathscr{o}_p\Big(\frac{1}{\sqrt{n}}\Big).$$

The inequality follows from the Euclidean properties of the class $\mathscr{S}$ together with Assumption 6, and the rate of convergence can be seen to hold on combining Lemma 1(i) with Assumption 7.

Next, recall that Lemma 1(ii) implies that

$$\widehat{\vartheta}(W_i) - \vartheta_0(W_i) = -\frac{1}{n\sigma_{\mathrm{l}}^{\dim(Z)}} \sum_{k=1}^{n} \frac{[\mathrm{a}(E_k) - \mu_{\mathrm{a}(E)}(Z_i)]}{\mathrm{p}_Z(Z_i)} \mathrm{l}\Big(\frac{Z_i - Z_k}{\sigma_{\mathrm{l}}}\Big) + \mathcal{O}_p\Big(\frac{n^{\epsilon/2}}{n\sigma_{\mathrm{l}}^{\dim(Z)}}\Big)$$

for each $Z_i \in \mathscr{Z}$. Plug this expression into $\mathrm{d}_n(\theta)$ and use Assumptions 1–3 and 6–7 to write

$$\mathrm{d}_n(\theta) = \frac{1}{3}\binom{n}{3}^{-1} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq i,j} \dot{\mathrm{b}}(D_i,D_j,D_k,\theta) + \mathscr{o}_p\Big(\frac{1}{\sqrt{n}}\Big), \tag{B.2}$$

where $\dot{\mathrm{b}}(D_i,D_j,D_k,\theta)$ is defined as

$$-\frac{\mathrm{s}(V_i,V_j,\theta)\mathrm{t}(Z_i)\mathrm{t}(Z_j)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}\sigma_{\mathrm{l}}^{\dim(Z)}} \; \frac{[\mathrm{a}(E_k) - \mu_{\mathrm{a}(E)}(Z_i)]'}{\mathrm{p}_Z(Z_i)} \mathrm{k}'\Big(\frac{\vartheta_0(W_i)-\vartheta_0(W_j)}{\sigma_{\mathrm{k}}}\Big) \mathrm{l}\Big(\frac{Z_i - Z_k}{\sigma_{\mathrm{l}}}\Big).$$

The influence of the remainder term in the linear representation of $\widehat{\vartheta}(w) - \vartheta_0(w)$ on $\mathrm{d}_n(\theta)$ is

$$\mathrm{r}_n(\theta) = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{j\neq i} \frac{\mathrm{s}(V_i,V_j,\theta)\mathrm{t}(Z_i)\mathrm{t}(Z_j)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}} \mathrm{k}'\Big(\frac{\vartheta_0(W_i - \vartheta_0(W_j))}{\sigma_{\mathrm{k}}}\Big) \mathcal{O}_p\Big(\frac{n^{\epsilon/2}}{n\sigma_{\mathrm{l}}^{\dim(Z)}}\Big),$$

on recycling constructive notation. In (B.2), $\mathrm{r}_n(\theta)$ is absorbed into the $\scriptstyle\mathcal{O}_p(1/\sqrt{n})$ term; argue as in the proof to Theorem 1 to see that

$$\sup_{\theta\in\Theta}\big\|\mathrm{r}_n(\theta)\big\| \le \frac{2\ \mathbb{P}_n[\mathcal{H}'(\cdot,\cdot)]}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}}\mathcal{O}_p\Big(\frac{n^{\epsilon/2}}{n\sigma_{\mathrm{l}}^{\dim(Z)}}\Big) = o_p\Big(\frac{1}{\sqrt{n}}\Big).$$

Likewise, uniformly over Θ, the contributions of the terms with $k = i$ or $k = j$ to $\mathrm{d}_n(\theta)$ are bounded by

$$\begin{aligned}\Big\|\binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\dot{\mathrm{b}}(D_i,D_j,D_i,\theta)}{n}\Big\| &\le \frac{4\ \mathrm{l}(0)\sum_{i=1}^{n}\sum_{j\neq i}\mathcal{H}'(D_i,D_j)\ \|\mathrm{a}(E_i)-\mu_{\mathrm{a}(E)}(Z_i)\|}{n^2(n-1)\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}\sigma_{\mathrm{l}}^{\dim(Z)}} \\ &= \mathcal{O}_p\Big(\frac{n^{[\dim(\vartheta_0)+1]\kappa+\dim(Z)\lambda}}{n}\Big) = o_p\Big(\frac{1}{\sqrt{n}}\Big)\end{aligned}$$

and

$$\begin{aligned}\Big\|\binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\dot{\mathrm{b}}(D_i,D_j,D_j,\theta)}{n}\Big\| &\le \frac{4\ \mathrm{l}(0)\sum_{i=1}^{n}\sum_{j\neq i}\mathcal{H}'(D_i,D_j)\ \|\mathrm{a}(E_j)-\mu_{\mathrm{a}(E)}(Z_j)\|}{n^2(n-1)\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}\sigma_{\mathrm{l}}^{\dim(Z)}} \\ &= \mathcal{O}_p\Big(\frac{n^{[\dim(\vartheta_0)+1]\kappa+\dim(Z)\lambda}}{n}\Big) = o_p\Big(\frac{1}{\sqrt{n}}\Big),\end{aligned}$$

respectively, and are thus asymptotically negligible.

To make further progress it is useful to 'symmetrize' the third-order U-statistic in (B.2). To this end, let

$$\begin{aligned}\mathrm{b}(D_i,D_j,D_k,\theta) \equiv\ & \dot{\mathrm{b}}(D_i,D_j,D_k,\theta)+\dot{\mathrm{b}}(D_i,D_k,D_j,\theta)+\dot{\mathrm{b}}(D_j,D_i,D_k,\theta) \\ &+\dot{\mathrm{b}}(D_k,D_i,D_j,\theta)+\dot{\mathrm{b}}(D_j,D_k,D_i,\theta)+\dot{\mathrm{b}}(D_k,D_j,D_i,\theta)\end{aligned}$$

and rewrite (B.2) as

$$\mathrm{d}_n(\theta) = \frac{1}{3}\binom{n}{3}^{-1}\sum_{\iota_3}\mathrm{b}(D_i,D_j,D_k,\theta) + o_p\Big(\frac{1}{\sqrt{n}}\Big),$$

where $\iota_3 = (i,j,k)$ ranges over the $n(n-1)(n-2)$ ordered triplets of distinct integers from the set $\{1,2,\ldots,n\}$. It is immediately verified that $\|\mathrm{b}(\mathrm{P},\mathrm{P},\mathrm{P},\theta)\|^2/n$ is $\mathcal{O}(1)$ so that

$$\mathrm{d}_n(\theta) = \frac{1}{3}\mathrm{b}(\mathrm{P},\mathrm{P},\mathrm{P},\theta) + \frac{1}{n}\sum_{i=1}^{n}\big[\mathrm{b}(D_i,\mathrm{P},\mathrm{P},\theta)-\mathrm{b}(\mathrm{P},\mathrm{P},\mathrm{P},\theta)\big] + o_p\Big(\frac{1}{\sqrt{n}}\Big) \qquad \text{(B.3)}$$

by Lemma A.3 in Ahn and Powell (1993). (B.3) further simplifies upon calculation of the expectations involved, which requires evaluating each of the six components of $\mathrm{b}(D_i, \mathrm{P}, \mathrm{P}, \theta)$.

The contribution of the first four components of $\mathrm{b}(\cdot, \cdot, \cdot, \theta)$ to $\mathrm{b}(D_i, \mathrm{P}, \mathrm{P}, \theta)$ is asymptotically negligible. To see this, consider the first of these components, $\dot{\mathrm{b}}_{i,j,k}(D_i, \mathrm{P}, \mathrm{P}, \theta)$, in obvious shorthand notation; the remaining three can be dealt with similarly. Observe that, uniformly over Θ,

$$\begin{aligned}\left\|\dot{\mathrm{b}}_{i,j,k}(D_i, \mathrm{P}, \mathrm{P}, \theta)\right\| \leq \mathrm{t}(Z_i) & \left\|\int \frac{[\mathrm{a}(e) - \mu_{\mathrm{a}(E)}(Z_i)]}{\mathrm{p}_Z(Z_i)\sigma_{\mathrm{l}}^{\dim(Z)}} \mathrm{l}\Big(\frac{Z_i - z}{\sigma_{\mathrm{l}}}\Big) \mathrm{p}_{E|Z}(e|z)\ \mathrm{p}_Z(z) \mathrm{d}(e, z)\right\| \\ \times \qquad & \left\|\int -\frac{\mathrm{s}(V_i, v, \theta)\mathrm{t}(z)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}} \mathrm{k}'\Big(\frac{\vartheta_i - \vartheta}{\sigma_{\mathrm{k}}}\Big)\ \mathrm{p}_{D|\vartheta}(d|\vartheta)\mathrm{p}_\vartheta(\vartheta)\ \mathrm{d}(d, \vartheta)\right\|.\end{aligned}$$

The explicit conditioning on $\vartheta_i \equiv \vartheta_0(W_i)$ is feasible as $\vartheta_0(w)$ is noise-free. By iterated expectations and Assumptions 1–3, the first of these right-hand side terms is

$$\mathrm{t}(Z_i)\left\|\int \frac{[\mu_{\mathrm{a}(E)}(z) - \mu_{\mathrm{a}(E)}(Z_i)]}{\mathrm{p}_Z(Z_i)\sigma_{\mathrm{l}}^{\dim(Z)}} \mathrm{l}\Big(\frac{Z_i - z}{\sigma_{\mathrm{l}}}\Big) \mathrm{p}_Z(z)\ \mathrm{d}z\right\| = \mathcal{O}(\sigma_{\mathrm{l}}^{\ell}) = o\Big(\frac{1}{\sqrt{n}}\Big),$$

as can be shown using standard arguments. Next, iterate expectations on the second right-hand side term and use the definition of $\overline{\varphi}$ to write

$$\sup_{\theta\in\Theta}\left\|\dot{\mathrm{b}}_{i,j,k}(D_i, \mathrm{P}, \mathrm{P}, \theta)\right\| \leq o\Big(\frac{1}{\sqrt{n}}\Big)\ \sup_{\theta\in\Theta}\left\|\int \frac{\overline{\varphi}(V_i, \vartheta, \theta)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}} \mathrm{k}'\Big(\frac{\vartheta_i - \vartheta}{\sigma_{\mathrm{k}}}\Big)\ \mathrm{d}\vartheta\right\|.$$

On changing variable from ϑ to $\eta = \frac{\vartheta_i - \vartheta}{\sigma_{\mathrm{k}}}$ and integrating by parts,

$$\begin{aligned}\sup_{\theta\in\Theta}\left\|\int \frac{\overline{\varphi}(V_i, \vartheta, \theta)}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)+1}} \mathrm{k}'\Big(\frac{\vartheta_i - \vartheta}{\sigma_{\mathrm{k}}}\Big)\ \mathrm{d}\vartheta\right\| &= \sup_{\theta\in\Theta}\left\|\int \nabla_\vartheta\overline{\varphi}(V_i, \vartheta_i - \eta\sigma_{\mathrm{k}}, \theta)\mathrm{k}(\eta)\ \mathrm{d}\eta\right\| \\ &= \sup_{\theta\in\Theta}\left\|\nabla_\vartheta\overline{\varphi}(V_i, \vartheta_0(W_i), \theta)\right\| + \mathcal{O}(\sigma_{\mathrm{k}}^{\kappa}),\end{aligned}$$

where the last transition follows again by a κth-order expansion and Assumption 8.[13] Deduce from this and Assumption 7 that

$$\begin{aligned}&\sup_{\theta\in\Theta}\|\dot{\mathrm{b}}_{i,j,k}(D_i, \mathrm{P}, \mathrm{P}, \theta)\| = o\Big(\frac{1}{\sqrt{n}}\Big), &&\sup_{\theta\in\Theta}\|\dot{\mathrm{b}}_{j,i,k}(\mathrm{P}, D_i, \mathrm{P}, \theta)\| = o\Big(\frac{1}{\sqrt{n}}\Big), \\ &\sup_{\theta\in\Theta}\|\dot{\mathrm{b}}_{i,k,j}(D_i, \mathrm{P}, \mathrm{P}, \theta)\| = o\Big(\frac{1}{\sqrt{n}}\Big), &&\sup_{\theta\in\Theta}\|\dot{\mathrm{b}}_{k,i,j}(\mathrm{P}, D_i, \mathrm{P}, \theta)\| = o\Big(\frac{1}{\sqrt{n}}\Big),\end{aligned} \tag{B.4}$$

[13] The term $\frac{\overline{\varphi}(V_i, \vartheta_i - \eta\sigma_{\mathrm{k}}, \theta)}{\sigma_{\mathrm{k}}}\mathrm{k}(\eta)\big|_{-\infty}^{+\infty}$ vanishes because $\mathrm{k}(\eta) \overset{\|\eta\|\to\infty}{\longrightarrow} 0$ by Assumption 6.

so that it remains only to work out $\dot{\mathrm{b}}_{j,k,i}(\mathrm{P},\mathrm{P},D_i,\theta)$ and $\dot{\mathrm{b}}_{k,j,i}(\mathrm{P},\mathrm{P},D_i,\theta)$.

Iterate expectations and argue as in the previous paragraph to write $\dot{\mathrm{b}}_{j,k,i}(\mathrm{P},\mathrm{P},D_i,\theta)$ as

$$-\int \mathrm{t}(Z)\frac{[\mathrm{a}(E_i)-\mu_{\mathrm{a}(E)}(z)]'}{\sigma_{\mathrm{l}}^{\dim(Z)}\mathrm{p}_Z(z)}\Big[\nabla_\vartheta\overline{\varphi}(v,\vartheta_0(w),\theta)+\mathcal{O}(\sigma_{\mathrm{k}}^{\mathcal{k}})\Big]\mathrm{l}\Big(\frac{z-Z_i}{\sigma_{\mathrm{l}}}\Big)\,\mathrm{dP}(v,w).$$

Integrate (V,E) against the density $\mathrm{p}_{(V,E)|Z}$ and recall that $\mathcal{O}(\sigma_{\mathrm{k}}^{\mathcal{k}})=o(1/\sqrt{n})$. Next, use an $\mathcal{l}$th-order expansion around Z_i, a change of variable, and Assumption 2 to see that

$$\begin{aligned}\dot{\mathrm{b}}_{j,k,i}(\mathrm{P},\mathrm{P},D_i,\theta)&=-\mathrm{t}(Z_i)\big[\mathrm{a}(E_i)-\mu_{\mathrm{a}(E)}(Z_i)\big]'\,\overline{\delta}(Z_i,\theta)+o\Big(\frac{1}{\sqrt{n}}\Big)\\&=\overline{\zeta}(W_i,\theta)+o\Big(\frac{1}{\sqrt{n}}\Big),\end{aligned}\tag{B.5}$$

Also, because $\dot{\mathrm{b}}_{k,j,i}(\mathrm{P},\mathrm{P},D_i,\theta)$ has an identical structure,

$$\begin{aligned}\dot{\mathrm{b}}_{k,j,i}(\mathrm{P},\mathrm{P},D_i,\theta)&=-\mathrm{t}(Z_i)\big[\mathrm{a}(E_i)-\mu_{\mathrm{a}(E)}(Z_i)\big]'\,\overline{\delta}(Z_i,\theta)+o\Big(\frac{1}{\sqrt{n}}\Big)\\&=\overline{\zeta}(W_i,\theta)+o\Big(\frac{1}{\sqrt{n}}\Big),\end{aligned}\tag{B.6}$$

by the same reasoning.

Combine (B.4), (B.5), and (B.6) with (B.3) to obtain

$$\mathrm{d}_n(\theta)=-\frac{2}{3}\mathrm{b}(\mathrm{P},\mathrm{P},\mathrm{P},\theta)+\frac{2}{n}\sum_{i=1}^{n}\overline{\zeta}(W_i,\theta)+o_p\Big(\frac{1}{\sqrt{n}}\Big).$$

The proof is complete on noting that

$$\mathrm{b}(\mathrm{P},\mathrm{P},\mathrm{P},\theta)=o\Big(\frac{1}{\sqrt{n}}\Big);$$

use the fact that $\int \mathrm{a}(e)\ \mathrm{p}_{E|Z}(e|z)\ \mathrm{d}e=\mu_{\mathrm{a}(E)}(z)$ to deduce that the dominant term in both (B.5) and (B.6) has mean zero conditional on $Z=Z_i$. □

Proof of Lemma A. Let $\mathrm{d}_n(\theta)\equiv\mathrm{q}_n(\theta)-\mathrm{q}_n(\theta_0)$ and let $\mathrm{d}(\theta)\equiv\mathbb{P}[\mathrm{d}_n(\theta)]$. Then, by an application of a Hoeffding decomposition (see, e.g., Serfling, 1980; Sherman, 1993,1994b),

$$\mathrm{d}_n(\theta)=\mathrm{d}(\theta)+2\big[\tau(\mathrm{P}_n,\theta)-\tau(\mathrm{P}_n,\theta_0)-\mathrm{d}(\theta)\big]+\mathbb{P}_n\big[\mathrm{r}(\cdot,\cdot,\theta)\big],\tag{B.7}$$

where, on letting $\mathrm{b}(D_i, D_j, \theta) \equiv [\mathrm{h}(D_i, D_j, \theta) - \mathrm{h}(D_i, D_j, \theta_0)]$, the remainder takes the form

$$\mathrm{r}(D_i, D_j, \theta) \equiv \mathrm{d}(\theta) + \mathrm{b}(D_i, D_j, \theta) - \mathrm{b}(D_i, \mathrm{P}, \theta) - \mathrm{b}(\mathrm{P}, D_j, \theta).$$

Observe that $\mathrm{d}(\theta) = \tau(\mathrm{P}, \theta) - \tau(\mathrm{P}, \theta_0)$. Furthermore, by the arguments used in the proof of Lemma 2,

$$\tau(d, \theta) = \mathrm{t}(z)\ \overline{\varphi}(v, \vartheta_0(w), \theta) + \mathcal{O}(\sigma_{\mathrm{k}}^{\not k}) = \overline{\tau}(d, \theta) + o\Big(\frac{1}{\sqrt{n}}\Big) \tag{B.8}$$

uniformly over Θ. The current proof then parallels the proofs in Sherman (1993) and Jochmans (2010), with some modifications.

fix θ in $\mathscr{N}$. Call upon the differentiability of $\overline{\tau}(d, \theta)$ as postulated under Assumption 9 to expand $\overline{\tau}(d, \theta)$ around θ_0. Then

$$\begin{aligned} \tau(d, \theta) - \tau(d, \theta_0) = \quad & (\theta - \theta_0)' \nabla_\theta \overline{\tau}(d, \theta_0) \\ & + \frac{1}{2} (\theta - \theta_0)' \nabla_{\theta\theta} \overline{\tau}(d, \theta_0)(\theta - \theta_0) \\ & + \frac{1}{2} (\theta - \theta_0)' \big[\nabla_{\theta\theta} \overline{\tau}(d, *) - \nabla_{\theta\theta} \overline{\tau}(d, \theta_0)\big] (\theta - \theta_0) + o\Big(\frac{1}{\sqrt{n}}\Big) \end{aligned} \tag{B.9}$$

on linking $\overline{\tau}(d, \theta) - \overline{\tau}(d, \theta_0)$ to $\tau(d, \theta) - \tau(d, \theta_0)$ through (B.8).

Envoke the Lipschitz condition in Assumption 9, take expectations, and use the fact that $\nabla_\theta \overline{\tau}(\mathrm{P}, \theta_0) = 0$ by the first-order condition for a maximum of the limiting objective function to see that

$$\mathrm{d}(\theta) = (\theta - \theta_0)' \frac{\nabla_{\theta\theta} \overline{\tau}(\mathrm{P}, \theta_0)}{2} (\theta - \theta_0) + o(\|\theta - \theta_0\|^2) + o\Big(\frac{1}{\sqrt{n}}\Big) \tag{B.10}$$

uniformly over $o_p(1)$ neighborhoods of θ_0.

Subtract (B.10) from (B.9) and average across observations. Then

$$\tau(\mathrm{P}_n, \theta) - \tau(\mathrm{P}_n, \theta_0) - \mathrm{d}(\theta) = \frac{(\theta - \theta_0)'}{\sqrt{n}} \Big[\sqrt{n} \nabla_\theta \overline{\tau}(\mathrm{P}_n, \theta_0) + o_p(1)\Big] + o_p(\|\theta - \theta\|^2) \tag{B.11}$$

uniformly over $o_p(1)$ neighborhoods of θ_0 because

$$\big\|(\theta - \theta_0)' [\nabla_{\theta\theta} \overline{\tau}(\mathrm{P}_n, *) - \nabla_{\theta\theta} \overline{\tau}(\mathrm{P}_n, \theta_0)](\theta - \theta_0)\big\| \leq \mathcal{M}_\tau(\mathrm{P}_n) \big\|\theta - \theta_0\big\|^3$$

and $(\theta - \theta_0)' [\nabla_{\theta\theta} \overline{\tau}(\mathrm{P}_n, \theta_0) - \nabla_{\theta\theta} \overline{\tau}(\mathrm{P}, \theta_0)](\theta - \theta_0) = o_p(1)$ by Assumption 9 (that is, the integrability of the Lipschitz constant) and a law of large numbers, respectively.

Combine the Euclidean properties of the class $\mathscr{H}$ with Corollary 17 and Corollary 21 in Nolan and Pollard (1987) to see that the class $\{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}\mathrm{r}(\cdot,\cdot,\theta) : \theta \in \Theta\}$ is Euclidean for an envelope whose second moment under $\mathbb{P}$ exists. Further observe that $\mathrm{r}(\cdot,\cdot,\theta)$ is P-degenerate on $\mathrm{supp}(D) \otimes \mathrm{supp}(D)$, that $\mathrm{r}(\cdot,\cdot,\theta_0) = 0$, and that $\|\mathrm{r}(\cdot,\cdot,\theta_0)\|$ is bounded by a multiple of $[\mathcal{H}(\mathrm{P},\mathrm{P})+\mathcal{H}(\cdot,\cdot)]/\sigma_K^{\dim(\vartheta_0)}$; refer to the bound as $\mathcal{H}(\cdot,\cdot)/\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}$. Apply Theorem 3 in Sherman (1994a) with, in his notation, $\Theta_n = \Theta$, $\gamma_n = 1$, and any $\alpha \in (0,1)$ to see that

$$\mathbb{P}_n[\mathrm{r}(\cdot,\cdot,\theta)] = \mathcal{O}_p\Big(\frac{1}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}n}\Big) = o_p(1)$$

uniformly over Θ. Reset $\gamma_n = \mathbb{P}_n[\mathcal{H}(\cdot,\cdot)]/\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}$ and let $\delta_n = 1/\big(\sqrt{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}n}\big)$. Then, on setting α sufficiently close to unity, by another application of the same theorem, in tandem with Assumption 7,

$$\mathbb{P}_n[\mathrm{r}(\cdot,\cdot,\theta)] = \frac{\mathbb{P}_n[\mathcal{H}(\cdot,\cdot)]}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}}\ \mathcal{O}_p\Big(\frac{(\gamma_n\delta_n)^\alpha}{n}\Big) = o_p\Big(\frac{1}{n}\Big) \tag{B.12}$$

uniformly over $\mathcal{O}_p\Big(1/\sqrt{\sigma_K^{\dim(\vartheta_0)}n}\Big) = o_p(1)$ neighborhoods of θ_0.

Plug (B.10), (B.11), and (B.12) into (B.7). The proof is complete on collecting terms. □

Proof of Lemma B. Let $\mathrm{d}_n(\theta) \equiv \overline{\zeta}(\mathrm{P}_n,\theta) - \overline{\zeta}(\mathrm{P}_n,\theta_0)$. Envoke Assumption 10 to expand $\overline{\zeta}(w,\theta)$ around θ_0. On averaging,

$$\mathrm{d}_n(\theta) = (\theta-\theta_0)'\ \nabla_\theta\overline{\zeta}(\mathrm{P}_n,\theta_0) + (\theta-\theta_0)'\frac{\nabla_{\theta\theta}\overline{\zeta}(\mathrm{P}_n,*)}{2}(\theta-\theta_0).$$

Refer to the Lipschitz continuity in Assumption 10 and a law of large numbers to dispense with the quadratic term. Rearrange to complete the proof. □

Proof of Theorem 2. Combine Lemma 2 with Lemmata A and B. On collecting terms,

$$\widehat{\mathrm{q}}_n(\theta) = \mathrm{q}_n(\theta_0) + (\theta-\theta_0)'\frac{\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)}{2}(\theta-\theta_0) + (\theta-\theta_0)'\frac{2\mathcal{D}_n}{\sqrt{n}} + o_p(\|\theta-\theta_0\|^2) + o_p\Big(\frac{1}{n}\Big)$$

uniformly over $\mathcal{O}_p\Big(1/\sqrt{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}n}\Big)$ neighborhoods of θ_0, where $\mathcal{D}_n \equiv \sqrt{n}\overline{\varsigma}(\mathrm{P}_n) + o_p(1)$. $\sqrt{n}$-consistency follows immediately from Theorem 1 in Sherman (1994a). Next, refer to Assumptions 9 and 10 to see that $\mathcal{D}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma)$. Then, because $\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)$ is negative definite, the asymptotic-distribution result follows on applying Theorem 2 in Sherman (1994a). □

Proof of Corollary 1. Kill the randomness in the first-step estimates, i.e., set $\widehat{\mu}_{\mathrm{a}(E)}(z)$ to $\mu_{\mathrm{a}(E)}(z)$ for each z in $\mathscr{Z}$. Then Lemma 2 becomes superfluous and the result follows from the proof to Theorem 2 on replacing $\mathcal{D}_n$ by $\sqrt{n}\nabla_\theta\overline{\tau}(\mathrm{P}_n,\theta_0)+\mathcal{o}_p(1)$. □

Proof of Lemma 3. The strategy followed is similar to the arguments that lead to Theorem 4 in Sherman (1993) and to those used in the proof of Theorem 1 in Abrevaya (1999b). Observe that $\overline{\tau}(d_1,\theta)$ can be written as

$$\mathrm{t}(z_1)\int_{x'\theta<x_1'\theta}\mathrm{t}(z)\ \mathcal{S}\big(y_1,I(x_2,w_1)\big)\mathrm{p}_I\big(I(x_2,w_1)\big)\ \mathrm{p}_{X,Z|I(X,W)}\big(x_2,z_2|I(x_2,w_1)\big)\ \mathrm{d}(x_2,z_2)$$
$$+\mathrm{t}(z_1)\int\mathrm{t}(z_2)\int\mathrm{m}(y_2)\ \mathrm{dP}_I(I(x_2,w_2))\ \mathrm{p}_I\big(I(x_2,w_1)\big)\mathrm{p}_{X,Z|I(X,W)}\big(x_2,z_2|I(x_2,w_1)\big)\ \mathrm{d}(x_2,z_2).$$

The second term is free of θ. Hence, on letting $\mathrm{u}^{(i)}$ be the unit vector with ith-element equal to one, because $\nabla_{\theta^{(i)}}\overline{\tau}(d_1,\theta_0)=\lim_{h\to0}\frac{1}{h}[\overline{\tau}\big(d_1,\theta_0+h\mathrm{u}^{(i)}\big)-\overline{\tau}\big(d_1,\theta_0\big)]$ by definition, $\nabla_{\theta^{(i)}}\overline{\tau}(d_1,\theta_0)$ equals

$$\mathrm{t}(z_1)\int\mathrm{t}(z_2)(x_1^{(i)}-x_2^{(i)})\ \mathcal{S}\big(y_1,I(x_2,w_1)\big)\ \mathrm{p}_I\big(I(x_1,w_2)\big)\mathrm{p}_{X,Z|I(X,W)}\big(x_2,z_2|I(x_1,w_1)\big)\ \mathrm{d}(x_2,z_2)$$

for each $i=1,\dots,\dim(X)$. Perform the integration and stack the components to get the expression for $\nabla_\theta\overline{\tau}(d_1,\theta_0)$ as stated in the lemma.

Next, rearrange and differentiate under the integral to get

$$\begin{aligned}\nabla_\theta\overline{\zeta}(w_1,\theta_0)&=\int\nabla_{\theta\vartheta}\overline{\tau}(d_2,\theta_0)\ \mathrm{dP}_{D|Z}(d_2|z_1)\ \big[\mu_{\mathrm{a}(E)}(z_1)-\mathrm{a}(e_1)\big]\\&=\int\mathcal{X}(x_2,w_2)\ \mathcal{S}_2\big(y_2,I(x_2,w_2)\big)'\ \mathrm{p}_I\big(I(x_2,w_2)\big)\ \mathrm{dP}_{X,W|Z}(x_2,w_2|z_1)\\&\quad\times\big[\mu_{\mathrm{a}(E)}(z_1)-\mathrm{a}(e_1)\big].\end{aligned}$$

The second transition above follows again from an application of the moment condition $\int\mathcal{S}(y,I(x,w))\ \mathrm{dP}_{I(X,W)}(I(x,w))=0$.

Finally, calculations similar to those used to arrive at the expression for $\nabla_\theta\overline{\tau}(d,\theta_0)$ in combination with the existence of the second moment of $\mathrm{t}(Z)X$ lead to the expression for the second-derivative term $\nabla_{\theta\theta}\overline{\tau}(\mathrm{P},\theta_0)$. □

Proof of Lemma C. Let $\mathscr{D}\equiv\mathrm{supp}(Y)\times\mathrm{supp}(X)\times\mathrm{supp}(E)\times\mathscr{Z}$. To see that $\widehat{\mathrm{p}}_I(\widehat{I}(x,w))$ is consistent for $\mathrm{p}_I(I(x,w))$, apply the triangle inequality to get that

$$\big\|\widehat{\mathrm{p}}_I\big(\widehat{I}(x,w)\big)-\mathrm{p}_I\big(I(x,w)\big)\big\|\leq\big\|\widehat{\mathrm{p}}_I\big(\widehat{I}(x,w)\big)-\widehat{\mathrm{p}}_I\big(I(x,w)\big)\big\|+\big\|\widehat{\mathrm{p}}_I\big(I(x,w)\big)-\mathrm{p}_I\big(I(x,w)\big)\big\|$$

uniformly over $\mathscr{D}$. For the first right-hand side term,

$$\begin{aligned}
&\sup_{d\in\mathscr{D}}\Big\|\frac{1}{n\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+1}}\sum_{i=1}^{n}\mathrm{j}\Big(\frac{\widehat{I}(X_i,W_i)-\widehat{I}(x,w)}{\sigma_{\mathrm{j}}}\Big)-\mathrm{j}\Big(\frac{I(X_i,W_i)-I(x,w)}{\sigma_{\mathrm{j}}}\Big)\Big\|\\
\leq&\frac{1}{n}\sum_{i=1}^{n}\Big\|\mathrm{j}'(\,*\,)\Big\|\ \frac{1}{\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+2}}\ \max\Big\{\big\|X_i-x\big\|\ \big\|\widehat{\theta}-\theta_0\big\|,\ 2\sup_{z\in\mathscr{Z}}\big\|\widehat{\mu}_{\mathrm{a}(E)}(z)-\mu_{\mathrm{a}(E)}(z)\big\|\Big\}\\
=&\frac{1}{\sigma_{\mathrm{j}}^{\dim(\vartheta_0)+2}}\ \max\Big\{\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big),\ \mathcal{O}_p\Big(\sqrt{\frac{n^{\epsilon/2}}{n\sigma_{\mathrm{l}}^{\dim(Z)}}}\Big)\Big\}=o_p(1)
\end{aligned}$$

by Assumptions 2 and 12, Lemma 1, and Theorem 2 provided that j is smaller than $(1-\epsilon/2-\lambda\dim(Z))/2(\dim(\vartheta_0)+2)$. The second right-hand side term is free of generated regressors and thus $o_p(1)$ uniformly over $\mathscr{D}$ by standard arguments; see, e.g., Silverman (1978).

Showing that $\widehat{\mu}_{\mathrm{t}(Z)}(\widehat{I}(x,w))$, $\widehat{\mu}_{\mathrm{t}(Z)X}(\widehat{I}(x,w))$, and $\widehat{\mu}_{\mathrm{m}(Y)}(\widehat{I}(x,w))$ are consistent then reduces to proving the consistency of their numerators. Because of Assumption 13, this follows by the same arguments as in the previous paragraph. This is so because the presence of $\mathrm{t}(Z_i)$, $\mathrm{t}(Z_i)X_i$, and $\mathrm{m}(Y_i)$ creates no additional difficulty. The consistency of $\widehat{\mathcal{X}}(x,w)$ for $\mathcal{X}(x,w)$, then, follows from Slutsky's theorem. The same conclusion can be drawn for $\widehat{\mathcal{S}}(y,\widehat{I}(x,w))$.

For the derivative estimates $\widehat{\mathcal{S}}_j\big(y,\widehat{I}(x,w)\big)$, $j=1,2$,

$$\sup_{d\in\mathscr{D}}\big\|\widehat{\mathcal{S}}_j\big(y,\widehat{I}(x,w)\big)-\widehat{\mathcal{S}}_j\big(y,I(x,w)\big)\big\|+\sup_{d\in\mathscr{D}}\big\|\widehat{\mathcal{S}}_j\big(y,I(x,w)\big)-\mathcal{S}_j\big(y,I(x,w)\big)\big\|$$

is an upper bound for $\sup_{d\in\mathscr{D}}\big\|\widehat{\mathcal{S}}_j(y,\widehat{I}(x,w))-\mathcal{S}_j(y,I(x,w))\big\|$. Under Assumption 13, the second part of this bound can again be dispensed with by following Silverman (1978). For the first part, the only new terms involve $\widehat{\mathrm{p}}_{Ij}(\widehat{I}(x,w))$. So, it remains to establish that $\sup_{d\in\mathscr{D}}\|\widehat{\mathrm{p}}_{Ij}(\widehat{I}(x,w))-\mathrm{p}_{Ij}(I(x,w))\|=o_p(1)$ for $j=1,2$. Because j'' exists and is bounded, a mean-value expansion—again in combination with Assumptions 2 and 12, Lemma 1, and Theorem 2—provides the result.

finally, turn to $\sup_{d\in\mathscr{D}}\big\|\widehat{\mu}_{\widehat{\nu}(X,W)}(z)-\mu_{\nu(X,W)}(z)\big\|$, which is no greater than

$$\sup_{d\in\mathscr{D}}\big\|\widehat{\mu}_{\widehat{\nu}(X,W)}(z)-\widehat{\mu}_{\nu(X,W)}(z)\big\|+\sup_{z\in\mathscr{Z}}\big\|\widehat{\mu}_{\nu(X,W)}(z)-\mu_{\nu(X,W)}(z)\big\|.$$

Because $\widehat{\nu}(x,w)\xrightarrow{p}\nu(x,w)$ uniformly over $\mathscr{D}$ by Slutsky's theorem, and because l is bounded (Assumption 1),

$$\sup_{d\in\mathscr{D}}\left\|\widehat{\mu}_{\widehat{\nu}(X,W)}(z)-\widehat{\mu}_{\nu(X,W)}(z)\right\|=\sup_{d\in\mathscr{D}}\left\|\sum_{i=1}^{n}\omega_i(z)\ \left[\widehat{\nu}(X_i,W_i)-\nu(X_i,W_i)\right]\right\|$$
$$\leq\sum_{i=1}^{n}\ \|\omega_i(z)\|\ \sup_{d\in\mathscr{D}}\left\|\widehat{\nu}(x,w)-\nu(x,w)\right\|=\mathcal{O}_p(1).$$

Observe that $\widehat{\mu}_{\nu(X,W)}(z)$ is again a Nadaraya-Watson estimator of $\mu_{\nu(X,W)}(z)$. Deduce that $\sup_{z\in\mathscr{Z}}\left\|\widehat{\mu}_{\nu(X,W)}(z)-\mu_{\nu(X,W)}(z)\right\|=\mathcal{O}_p(1)$ to complete the proof. □

Proof of Theorem 3. Let $\upsilon(d,\theta)\equiv-\Upsilon^{-1}\overline{\varsigma}(d,\theta)$. Combine the kernel estimates of the components of this influence function to obtain the plug-in estimates $\widehat{\upsilon}(D_i,\widehat{\theta})$, $i=1,\ldots,n$. Then we can write $\Upsilon^{-1}\Sigma\ \Upsilon^{-1}=\mathrm{P}[\upsilon(\cdot,\theta_0)\upsilon(\cdot,\theta_0)']$ and $\widehat{\Upsilon}^{-1}\widehat{\Sigma}\ \widehat{\Upsilon}^{-1}=\mathrm{P}_n[\widehat{\upsilon}(\cdot,\widehat{\theta})\widehat{\upsilon}(\cdot,\widehat{\theta})']$.

By Lemma 1, Lemma C, Theorem 2, and Slutsky's theorem,

$$\mathrm{P}_n\big[\widehat{\upsilon}(\cdot,\widehat{\theta})\widehat{\upsilon}(\cdot,\widehat{\theta})'\big]=\mathrm{P}_n\big[\upsilon(\cdot,\theta_0)\upsilon(\cdot,\theta_0)'\big]+\mathcal{O}_p(1).\tag{B.13}$$

Also, Assumptions 9–10 and the law of large numbers imply that

$$\mathrm{P}_n\big[\upsilon(\cdot,\theta_0)\upsilon(\cdot,\theta_0)'\big]\xrightarrow{p}\mathrm{P}\big[\upsilon(\cdot,\theta_0)\upsilon(\cdot,\theta_0)'\big].\tag{B.14}$$

Put together, (B.13) and (B.14) yield $\widehat{\Upsilon}^{-1}\widehat{\Sigma}\ \widehat{\Upsilon}^{-1}=\Upsilon^{-1}\Sigma\ \Upsilon^{-1}+\mathcal{O}_p(1)$. The proof is complete. □

Proof of Theorem 4. Fix y in $[y_{\mathrm{L}},y_{\mathrm{U}}]$; then $\psi_0(y)$ is interior to Ψ which is a compact interval by Assumption 14. By Assumption 4, $\overline{\mathrm{q}}^y(\psi,\theta_0)$ is continuous in ψ and uniquely maximized at $\psi_0(y)$. The proof to these claims is again identical as to when the control is absent. For the consistency of $\widehat{\psi}(y)$ for $\psi_0(y)$ it remains only to establish the uniform convergence of $\widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta})$ to $\overline{\mathrm{q}}^y(\psi,\theta_0)$.

Let $\mathrm{q}^y(\psi,\theta)\equiv\mathbb{P}[\mathrm{h}^y(\cdot,\cdot,\psi,\theta)]$. Apply the triangle inequality to obtain

$$\sup_{\psi\in\Psi}\left\|\widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta})-\overline{\mathrm{q}}^y(\psi,\theta_0)\right\|\leq\sup_{\psi\in\Psi}\left\|\widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta})-\mathrm{q}_n^y(\psi,\widehat{\theta})\right\|+\sup_{\psi\in\Psi}\left\|\mathrm{q}_n^y(\psi,\widehat{\theta})-\mathrm{q}^y(\psi,\widehat{\theta})\right\|$$
$$+\sup_{\psi\in\Psi}\left\|\mathrm{q}^y(\psi,\widehat{\theta})-\mathrm{q}^y(\psi,\theta_0)\right\|+\sup_{\psi\in\Psi}\left\|\mathrm{q}^y(\psi,\theta_0)-\overline{\mathrm{q}}^y(\psi,\theta_0)\right\|.$$

Observe that the class $\{\mathrm{s}^y(\cdot,\cdot,\psi,\theta) : \psi \in \Psi, \theta \in \Theta\}$ is Euclidean for the constant envelope of unity; use Example 2.11 in Pakes and Pollard (1989). Consequently, by applying the same arguments as in the proof to Theorem 1,

$$\sup_{\psi\in\Psi}\left\|\widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta}) - \mathrm{q}_n^y(\psi,\widehat{\theta})\right\| = \mathscr{o}_p(1) \quad \text{and} \quad \sup_{\psi\in\Psi}\left\|\mathrm{q}_n^y(\psi,\widehat{\theta}) - \mathrm{q}^y(\psi,\widehat{\theta})\right\| = \mathscr{o}_p(1).$$

Next, by Theorem 2 and Assumption 4, $\|\widehat{\theta}-\theta_0\| = \mathcal{O}_p(1/\sqrt{n})$ and $\mathrm{q}^y(\psi,\theta)$ is continuous in θ. Hence,

$$\sup_{\psi\in\Psi}\left\|\mathrm{q}^y(\psi,\widehat{\theta}) - \mathrm{q}^y(\psi,\theta_0)\right\| = \frac{1}{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}}\mathcal{O}_p\Big(\frac{1}{\sqrt{n}}\Big) = \mathscr{o}_p(1),$$

on employing Assumption 7. For the remaining component, finally, use Assumptions 6, 7, and 15 to obtain

$$\sup_{\psi\in\Psi}\left\|\mathrm{q}^y(\psi,\theta_0) - \overline{\mathrm{q}}^y(\psi,\theta_0)\right\| = \mathcal{O}(\sigma_{\mathrm{k}}) = \mathscr{o}(1)$$

by applying the usual trick. Thus, $\sup_{\psi\in\Psi}\|\widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta}) - \overline{\mathrm{q}}^y(\psi,\theta_0)\| = \mathscr{o}_p(1)$; statement (i) of Theorem 4 follows.

The proof of Theorem 4(ii) proceeds in three steps. first, recall the symmetry of $\mathrm{h}^y(\cdot,\cdot,\alpha,\theta)$ and the Euclidean properties of the classes $\mathscr{K}$ and $\{\mathrm{s}^y(\cdot,\cdot,\alpha,\theta) : \psi \in \Psi, \theta \in \Theta\}$. Then, for each y in $[y_{\mathrm{L}}, y_{\mathrm{U}}]$,

$$\widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta}) - \mathrm{q}_n^y(\psi,\widehat{\theta}) = \frac{2}{n}\sum_{i=1}^{n}\overline{\zeta}^y(W_i,\psi,\widehat{\theta}) + \mathscr{o}_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{B.15}$$

uniformly over Ψ. Notice that (B.15) can be shown by applying the same arguments as those contained in the proof of Lemma 2.

To handle $\mathrm{q}_n^y(\psi,\widehat{\theta})$, proceed as in the proof of Lemma A. fix (ψ,θ) in $\mathscr{N}_y$, define the functions $\mathrm{d}_n^y(\psi,\theta) \equiv \mathrm{q}_n^y(\psi,\theta) - \mathrm{q}_n^y(\psi_0(y),\theta)$ and $\mathrm{d}^y(\psi,\theta) \equiv \mathbb{P}[\mathrm{d}_n^y(\psi,\theta)]$, and apply a Hoeffding decomposition. The resulting approximation is

$$\mathrm{d}_n^y(\psi,\theta) = \mathrm{d}^y(\psi,\theta) + 2\big[\tau^y(\mathrm{P}_n,\psi,\theta) - \tau^y(\mathrm{P}_n,\psi_0(y),\theta) - \mathrm{d}^y(\psi,\theta)\big]$$

and the remainder term that can be dispensed with in the usual way. Start with $\mathrm{d}^y(\psi,\theta) = \tau^y(\mathrm{P},\psi,\theta) - \tau^y(\mathrm{P},\psi_0(y),\theta)$. Envoke Assumptions 6, 7, and 15 to write

$\tau^y(d,\psi,\theta)$ as $\overline{\tau}^y(d,\psi,\theta) + \mathcal{o}(1/\sqrt{n})$. Taylor-expand $\overline{\tau}^y(d,\psi,\theta)$ around $\psi_0(y)$ and then around θ_0. On taking expectations,

$$\begin{aligned}\mathrm{d}^y(\psi,\theta) &= (\psi-\psi_0(y))\nabla_{\psi\theta}\overline{\tau}^y(\mathrm{P},\psi_0(y),\theta_0)(\theta-\theta_0) + (\psi-\psi_0(y))^2\frac{\nabla_{\psi\psi}\overline{\tau}^y(\mathrm{P},\psi_0(y),\theta_0)}{2}\\ &\quad + \mathcal{o}((\psi-\psi_0(y))^2) + \mathcal{o}_p(\|\theta-\theta_0\|^2) + \mathcal{o}\Big(\frac{1}{\sqrt{n}}\Big)\end{aligned}$$

uniformly over $\mathcal{o}_p(1)$ neighborhoods of $(\psi_0(y),\theta_0)$; make use of the Lipschitz condition in Assumption 16 and notice that $\nabla_\psi\overline{\tau}^y(\mathrm{P},\psi_0(y),\theta_0)=0$. On evaluating at $\widehat{\theta}$,

$$\begin{aligned}\mathrm{d}^y(\psi,\widehat{\theta}) &= (\psi-\psi_0(y))\nabla_{\psi\theta}\overline{\tau}^y(\mathrm{P},\psi_0(y),\theta_0)\upsilon(\mathrm{P}_n,\theta_0) + (\psi-\psi_0(y))^2\frac{\nabla_{\psi\psi}\overline{\tau}^y(\mathrm{P},\psi_0(y),\theta_0)}{2}\\ &\quad + \mathcal{o}((\psi-\psi_0(y))^2) + \mathcal{o}_p\Big(\frac{1}{\sqrt{n}}\Big)\end{aligned} \tag{B.16}$$

follows by applying Theorem 2 and absorbing all terms that are asymptotically negligible into the $\mathcal{o}_p(1/\sqrt{n})$ term. Similarly,

$$\begin{aligned}\tau^y(\mathrm{P}_n,\psi,\theta) - \tau^y(\mathrm{P}_n,\psi_0(y),\theta) - \mathrm{d}^y(\psi,\widehat{\theta}) &= (\psi-\psi_0(y))\nabla_\psi\overline{\tau}^y(\mathrm{P}_n,\psi_0(y),\theta_0)\\ &\quad + \mathcal{o}((\psi-\psi_0(y))^2) + \mathcal{o}_p\Big(\frac{1}{\sqrt{n}}\Big)\end{aligned} \tag{B.17}$$

uniformly over $\mathcal{o}_p(1)$ neighborhoods of $\psi_0(y)$.

Finally, consider $\overline{\zeta}^y(\mathrm{P}_n,\psi,\widehat{\theta})$. Taylor-expand around $\psi_0(y)$ and θ_0, in turn. Use the Lipschitz condition and the finiteness of the population moments in Assumption 17 to dispense with $\nabla_{\psi\psi}\overline{\zeta}^y(\mathrm{P}_n,\psi_0(y),\theta_0)$. Because $\|\widehat{\theta}-\theta_0\| = \mathcal{O}_p(1/\sqrt{n})$ by Theorem 2 and $\nabla_{\psi\theta}\overline{\zeta}^y(\mathrm{P}_n,\psi_0(y),\theta_0) \xrightarrow{p} \nabla_{\psi\theta}\overline{\zeta}^y(\mathrm{P},\psi_0(y),\theta_0) = 0$ by the law of large numbers,

$$\begin{aligned}\overline{\zeta}^y(\mathrm{P}_n,\psi,\widehat{\theta}) - \overline{\zeta}^y(\mathrm{P}_n,\psi_0(y),\widehat{\theta}) &= (\psi-\psi_0(y))\big[\nabla_\psi\overline{\zeta}^y(\mathrm{P}_n,\psi_0(y),\theta_0) + \mathcal{o}_p(1)\big]\\ &\quad + \mathcal{o}_p\big((\psi-\psi_0(y))^2\big).\end{aligned} \tag{B.18}$$

uniformly over $\mathcal{o}_p(1)$ neighborhoods of $\psi_0(y)$.

Combine (B.15)–(B.18) and rearrange to see that, uniformly over $\mathcal{o}_p(1/\sqrt{\sigma_{\mathrm{k}}^{\dim(\vartheta_0)}n})$ neighborhoods of $\psi_0(y)$, $\widehat{\mathrm{q}}_n^y(\psi,\widehat{\theta}) - \widehat{\mathrm{q}}_n^y(\psi_0(y),\widehat{\theta})$ equals

$$(\psi-\psi_0(y))^2\frac{\nabla_{\psi\psi}\overline{\tau}^y(\mathrm{P},\psi_0(y),\theta_0)}{2} + (\psi-\psi_0(y))\frac{2\mathcal{D}_n^y}{\sqrt{n}} + \mathcal{o}_p((\psi-\psi_0(y))^2) + \mathcal{o}_p\Big(\frac{1}{n}\Big),$$

where $\mathcal{D}_n^y \equiv \sqrt{n}[\overline{\varsigma}^y(\mathrm{P}_n,\psi_0(y),\theta_0) + \overline{\varrho}^y(\mathrm{P}_n,\psi_0(y),\theta_0)] + \mathcal{o}_p(1)$. By Assumptions 16 and 17, $\mathcal{D}_n^y \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma(y))$. The proof is complete on unleashing Theorems 1 and 2 in Sherman (1994a), in turn, on the above expression. □

Proof of Corollary 2. The result follows on making the same modifications to the proof of Theorem 4 as exposed in the proof of Corollary 1. □

APPENDIX C: COMPUTATIONAL DETAILS[14]

The optimization routine consists of a user-determined maximum number of rounds. Each such round consists of a series of iterations, followed by a series of stability checks, with maxima again set by the user.

A single iteration, say the kth, proceeds as follows. For an initial value for $\widehat{\theta}$, $\theta_k \in \Theta$ say, an orthonormal set of dim(X)-vectors $\delta_k = (\delta_k^{(1)}, \cdots, \delta_k^{(\dim(X)-1)})$—each of which is orthogonal to θ_k—is drawn. The great circles connecting θ_k to a $\delta_k^{(l)}$ $(l = 1, \ldots, \dim(X)-1)$, that is, the sets of points $\theta_k(\lambda)$ for $\lambda \in [0, 2\pi)$ and

$$\theta_k^{(l)}(\lambda) \equiv \cos(\lambda)\ \theta_k + \sin(\lambda)\ \delta_k^{(l)},$$

provide a collection of orthogonal search directions along Θ. Next, $\widehat{Q}_N(\theta)$ is sequentially maximized along each of these directions.[15] The first sequence starts at θ_k and delivers $\theta_k^{(1)}$, the lth sequence starts at $\theta_k^{(l-1)}$ and delivers $\theta_k^{(l)}$.

If $\widehat{\mathrm{q}}_n(\theta_k^{(\dim(X)-1)}) > \widehat{\mathrm{q}}_n(\theta_k)$, the $(k+1)$th iteration commences with starting value $\theta_{k+1} = \theta_k^{(\dim(X)-1)}$. This process continues until a given set of search directions provides no further increase in the objective function. The final point estimate that results from this routine is declared a trial maximizer of $\widehat{\mathrm{q}}_n(\theta)$.

Next, the trial maximizer is subjected to a number of stability checks. These are necessary because the behavior of $\widehat{\mathrm{q}}_n(\theta)$ is only investigated along a finite number of search directions, so that the trial maximizer may, in fact, be only a local maximizer. A stability check consists of drawing at random an orthonormal set of search directions from Θ—each again being orthogonal to the trial maximizer—and evaluating the objective function along these directions. If an increase in $\widehat{\mathrm{q}}_n(\theta)$ is found, the check is terminated and the algorithm reverts to the next round, iterating around the new point estimate. If all checks are passed, the trial maximizer is declared stable and called $\widehat{\theta}$. As the number of stability checks increases to infinity, the randomization in drawing

[14]The algorithm discussed in this section is a modified version of the optimization routine for the maximum-score estimator introduced by Manski and Thompson (1986).

[15]Because in any given direction $\widehat{\mathrm{q}}_n(\theta)$ takes at most $2n(n-1)+2$ different values (see below), there is generally a subinterval of $[0, 2\pi)$ on which the objective function is maximized. Any of these points may be chosen as the new (intermediate) maximizer, but the mean or median of this subinterval seem natural choices. The results in the main text where obtained by using the latter.

search directions ensures that $\widehat{\theta}$ will be the global maximizer of $\widehat{\mathrm{q}}_n(\theta)$ with probability approaching one.

Evaluating the objective function along the great circle connecting θ_k and $\delta_k^{(l)}$ boils down to computing $\widehat{\mathrm{q}}_n(\theta_k^{(l)}(\lambda))$ for $\lambda \in [0, \pi)$. For $\lambda \in [\pi, 2\pi)$, the corresponding values of the objective function follow immediately. This is so because $\widehat{\mathrm{q}}_n(\theta_k^{(l)}(\lambda + \pi)) = \widehat{\mathrm{q}}_n(-\theta_k^{(l)}(\lambda))$ and

$$\widehat{\mathrm{q}}_n(-\theta_k^{(l)}(\lambda)) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \mathrm{m}(Y_i) - \widehat{\mathrm{q}}_n(\theta_k^{(l)}(\lambda)). \tag{C.1}$$

The first right-hand side term in the above expression does not contain any unknown parameters and $\widehat{\mathrm{q}}_n(\theta_k^{(l)}(\lambda))$ has already been computed. Because the weights do not depend on λ, it suffices to focus on the dynamics of $\mathrm{s}(\cdot, \cdot, \theta_k^{(l)}(\lambda))$.

Fix a pair of observations (i, j) and consider λ_{ij}, the solution to $(X_i - X_j)'\theta_k^{(l)}(\lambda) = 0$ on $[0, \pi)$. If $(X_i - X_j)'\theta_k$ is nonzero, λ_{ij} is unique and partitions $[0, \pi)$ into two subintervals, each on which $\mathrm{d}(D_i, D_j, \theta_k^{(l)}(\lambda))$ is constant. Moreover,

$$(X_i - X_j)'\theta_k > 0 \Rightarrow \begin{cases} (X_i - X_j)'\theta_k^{(l)}(\lambda) > 0 & \text{if } \lambda \in [0, \lambda_{ij}) \\ (X_i - X_j)'\theta_k^{(l)}(\lambda) < 0 & \text{if } \lambda \in (\lambda_{ij}, \pi) \end{cases}$$

while

$$(X_i - X_j)'\theta_k < 0 \Rightarrow \begin{cases} (X_i - X_j)'\theta_k^{(l)}(\lambda) < 0 & \text{if } \lambda \in [0, \lambda_{ij}) \\ (X_i - X_j)'\theta_k^{(l)}(\lambda) > 0 & \text{if } \lambda \in (\lambda_{ij}, \pi) \end{cases}$$

If $(X_i - X_j)'\theta_k = 0$, then $\lambda_{ij} = 0$ and

$$\begin{aligned} (X_i - X_j)'\theta_k > 0 &\Rightarrow (X_i - X_j)'\delta_k^{(l)} > 0 \text{ for all } \lambda \in [0, \lambda), \\ (X_i - X_j)'\theta_k < 0 &\Rightarrow (X_i - X_j)'\delta_k^{(l)} < 0 \text{ for all } \lambda \in [0, \lambda). \end{aligned}$$

It then follows that, for a given λ, $\mathrm{s}\big(D_i, D_j, \theta_k^{(l)}(\lambda)\big)$ equals

$$\begin{aligned} &\mathrm{m}(Y_i)\Big[1\big((\lambda_{ij} - \lambda)(X_i - X_j)'\theta_k > 0\big) + 1\big((X_i - X_j)'\delta_k^{(l)} > 0\big)1\big((X_i - X_j)'\theta_k = 0\big)\Big] \\ &+\mathrm{m}(Y_j)\Big[1\big((\lambda_{ij} - \lambda)(X_i - X_j)'\theta_k < 0\big) + 1\big((X_i - X_j)'\delta_k^{(l)} < 0\big)1\big((X_i - X_j)'\theta_k = 0\big)\Big]. \end{aligned}$$

Because $\lambda_{ij} = \lambda_{ji}$, there are at most $n(n-1)/2$ unique such λ_{ij}. They partition $[0, \pi)$ into $n(n-1) + 1$ intervals, each on which the objective function is constant in λ. The dynamics of the score contributions as a function of λ displayed above, together with (C.1), make it easy to compute $\widehat{\mathrm{q}}_n(\theta)$ on the entire interval $[0, 2\pi)$ and along any given direction.

REFERENCES

Abrevaya, J. (1999a). Leapfrog estimation of a fixed-effects models with unknown transformation of the dependent variable. *Journal of Econometrics*, 93:203–228.

Abrevaya, J. (1999b). Rank regression for current-status data: Asymptotic normality. *Statistics and Probability Letters*, 43:275–287.

Ahn, H. (1997). Semiparametric estimation of a single-index model with nonparametrically generated regressors. *Econometric Theory*, 13:3–31.

Ahn, H. and C. F. Manski. (1993). Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. *Journal of Econometrics*, 56:291–321.

Ahn, H. and J. L. Powell. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58:3–29.

Ai, C. and X. Chen. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71:1795–1843.

Aradillas-López, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics*, 157:409–431.

Aradillas-López, A., B. E. Honoré, and J. L. Powell. (2007). Pairwise difference estimation with nonparametric control variables. *International Economic Review*, 48:1119–1158.

Arcones, M. A. and E. Giné. (1993). Limit theorems for U-processes. *Annals of Probability*, 21:1494–1542.

Basmann, R. L. (1963). Remarks concerning the application of exact finite sample distribution functions of GCL estimators in econometric statistical inference. *Journal of the American Statistical Association*, 58:943–976.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, 60:889–917.

Bhattacharya, D. (2008). A permutation-based approach to estimating monotone index models. *Econometric Theory*, 24:795–807.

Blundell, R. W. and J. L. Powell. (2003). Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics*, volume II. Econometric Society. Cambridge University Press.

Blundell, R. W. and J. L. Powell. (2004). Endogeneity in semiparametric binary response models. *Review of Economic Studies*, 71:655–679.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 77:531–553.

Cavanagh, C. and R. P. Sherman. (1998). Rank estimators for monotonic index models. *Journal of Econometrics*, 84:351–381.

Chamberlain, G. (1986). Asymptotic efficiency in semiparametric models with censoring. *Journal of Econometrics*, 32, 189–218.

Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, 70:1683–1697.

Chernozhukov, V., I. Fernández-Val, and A. Galichon. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96:559–575.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71:1405–1441.

Chesher, A. (2007). Identification of nonadditive structural functions. In R. W. Blundell, W. K. Newey, and T. Persson (eds.), *Advances In Economics and Econometrics*, volume III. Econometric Society. Cambridge University Press.

Collomb, G. and W. Härdle. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression from dependent observations. *Stochastic Processes and their Applications*, 23:77–98.

Cosslett, S. R. (1987). Efficiency bounds for distribution-free estimators of the binary choice and censored regression models. *Econometrica*, 55:559–586.

Croux, C. and C. Dehon. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods and Applications*, forthcoming.

Das, M., W. K. Newey, and F. Vella. (2003). Nonparametric estimation of sample selection models. *Review of Economic Studies*, 70:33–58.

Fan, Y. and Q. Li. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, 64:865–890.

Ghosal, S., A. Sen, and A. W. van der Vaart. (2000). Testing monotonicity of regression. *Annals of statistics*, 28:1054–1082.

Gronau, R. (1973). The effect of children on the housewife's value of time. *Journal of Political Economy*, 81:S168–S199.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics*, 35:303–316.

Härdle, W., P. Hall, and H. Ichimura. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, 21:157–178.

Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, 42:679–693.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47:153–161.

Hong, H. and M. Shum. (2010). Pairwise-difference estimation of a dynamic optimization model. *Review of Economic Studies*, 77:273–304.

Honoré, B. E. and J. L. Powell. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, 64:241–278.

Honoré, B. E. and J. L. Powell. (2005). Pairwise difference estimators for nonlinear models. In D. W. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models*. Cambridge University Press.

Horowitz. J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*. Lecture Notes in Statistics. Springer-Verlag.

Horowitz, J. L. and S. Lee. (2002). Semiparametric methods in applied econometrics: Do the models fit the data? *Statistical Modelling*, 2:3–22.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58:71–120.

Ichimura, H. and L.-F. Lee. (1991). Semiparametric least squares estimation of multiple index models: Single equation estimation. In W. A. Barnett, J. L. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

Imbens, G. W. and W. K. Newey. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77:1481–1512.

Jochmans, K. (2010). Inference from kernel-weighted U-processes. Unpublished manuscript.

Jones, M. and D. Signorini. (1997). A comparison of higher-order bias kernel density estimators. *Journal of the American Statistical Association*, 92:1063–1073.

Khan, S. and E. Tamer. (2007). Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics*, 136:251–280.

Linton, O., and J. P. Nielsen. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82:93-100.

Li, Q. and J. S. Racine. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Mammen, E., C. Rothe, and M. Schienle. (2010). Nonparametric regression with nonparametrically generated covariates. Unpublished manuscript.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27:313–333.

Manski, C. F. (1991). Nonparametric estimation of expectations in the analysis of discrete choice under uncertainty. In W. A. Barnett, J. L. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

Manski, C. F. and T. S. Thompson. (1986). Operational characteristics of maximum score estimation. *Journal of Econometrics*, 32:85–108.

Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *Annals of Statistics*, 12:766–774.

Newey, W. K. (1990). Efficient estimation of semiparametric models via moment restrictions. Unpublished manuscript.

Newey, W. K. (1994a). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10:233–253.

Newey, W. K. (1994b). The asymptotic variance of semiparametric estimators. *Econometrica*, 62:1349–1382.

Newey, W. K. (2007). Nonparametric continuous/discrete choice models. *International Economic Review*, 48:1429–1439.

Newey, W. K. (2009). Two-step series estimation of sample selection models. *Econometrics Journal*, 12:S217–S229.

Newey, W. K. and D. L. McFadden. (1994). Large sample estimation and hypothesis testing. In R. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, volume IV. Elsevier Science.

Newey, W. K., J. L. Powell, and J. R. Walker. (1990). Semiparametric estimation of selection models: Some empirical results. *American Economic Review: Papers and Proceedings*, 80:324–328.

Nolan, D. and D. Pollard. (1987). U-processes: Rates of convergence. *Annals of Statistics*, 15:780–799.

Olley, G. S. and A. Pakes. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64:1263–1297.

Pakes, A. and D. Pollard. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57:1027–1057.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag.

Powell, J. L. (1994). Estimation of semiparametric models. In R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, volume IV. Elsevier Science.

Powell, J. L., J. H. Stock, and T. M. Stoker. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57:1403–1430.

Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies*, 57:167–181.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, 56:931–954.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61:123–137.

Sherman, R. P. (1994a). U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory*, 10:372–395.

Sherman, R. P. (1994b). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Annals of Statistics*, 22:439–459.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics*, 6:177–184.

Smith, R. J. and R. W. Blundell. (1986). An exogeneity test for simultaneous equations tobit models with an application to labor supply. *Econometrica*, 54:679–686.

Stock, J. H. (1989). Nonparametric policy analysis. *Journal of the American Statistical Association*, 84:567–575.

Stock, J. H., J. H. Wright, and M. Yogo. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, 20:518–529.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica*, 54:1461–1481.

Stoker, T. M. (1991). Equivalence of direct, indirect, and slope estimators of average derivatives. In W. A. Barnett, J. L. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

Subbotin, V. (2008a). Asymptotic and bootstrap properties of rank regressions. Unpublished manuscript.

Subbotin, V. (2008b). Weighted rank estimators. Unpublished manuscript.