Three-stage Semi-Parametric Inference: Control Variables and Differentiability
(Incomplete not to be cited)

Jinyong Hahn* Geert Ridder †

October 2015

Abstract

*Department of Economics, UCLA

†Department of Economics, University of Southern California, Kaprilian Hall, Los Angeles, CA 90089. Electronic correspondence: ridder@usc.edu, http://www-rcf.usc.edu/ ridder/.

1 Introduction

Econometric models with possibly multi-dimensional non-separable errors are now routinely used in applied work. Examples are the linear random coefficients model that is used to estimate average treatment effects and the mixed logit model that generates more credible substitution patterns between alternatives in discrete choice models. In general, assuming that parameters in an econometric model are heterogenous results in a non-separable error model, and therefore non-separable errors are the rule rather than the exception.

In non-separable error models a complication arises if the random errors and one or more of the regressors are dependent. As shown by Hahn and Ridder (2011) even with valid instrumental variables conditional moment restrictions do not recover the population parameters in non-separable error models. An alternative that does identify the population parameters is to use an average moment restriction that conditions on and averages over control variables (CV). These control variables are estimated in a first stage as the residuals in a parametric or non-parametric relation between the endogenous regressors and instruments. In a second step a conditional expectation with the original regressors and the control variables is estimated either parametrically as in Rivers and Vuong (1988) or non-parametrically as in Imbens and Newey (2009). In the latter case the estimated conditional expectation is averaged over the control variable to obtain the average structural function (ASF) that finally can be used as an input in the estimation of a parametric model.

We study aspects of inference for estimators defined by average moment restrictions where we focus on inference for a finite-dimensional parameter vector or statistic. The estimation procedure will consist of at least two steps some of which may involve non-parametric estimation. For instance the Rivers and Vuong (1988) estimator has a parametric first and second step. Below we discuss a semi-parametric control variable estimator that has three stages with the first (estimation of the control variable) and the second (estimation of the ASF) being non-parametric.

We make five contributions. First, we show the usefulness of the path-derivative calculations that were introduced in econometrics by Newey (1994) for multi-step semi-parametric inference with control variables. We derive the asymptotic variance of a semi-parametric CV estimator with a non-parametric first and second step. This variance is not available in the literature. Its derivation builds on our earlier study of three-step semi-parametric estimators in Hahn and Ridder (2013). Their results are however not directly applicable, because the first-stage in which the control variable is estimated is not a conditional expectation (nor a density)[1] and the averaging is over the estimated control variable which leads to a particular V-statistic expression. A second contribution is that we derive the same asymptotic variance by a stochastic expansion. As can be expected this is a major effort that serves two purposes. First, it illustrates that the shortcut method that only involves elementary calculus

[1]This step requires a generalization of Newey's two-step GMM to allow for a first stage that is a conditional empirical cdf.

actually works and, second it allows us to formulate (sufficient) regularity conditions under which the derived asymptotic distribution is valid. For obvious reasons we do not repeat the stochastic expansion for each estimator and statistic considered in this paper, instead using the path-derivative calculations to obtain asymptotic distributions.

Our third contribution is that we consider just and over identification of non-separable error models with endogenous regressors. Matzkin (2003) has shown that if the regressors are exogenous a relation that involves multiple errors is observationally equivalent to a relation with a single error. We show that if the relation between the endogenous regressor and the instrument is monotonic, a relation with multiple errors is observationally equivalent to a relation with two errors. This observation allows us to propose the as as we know first test for overidentifying restrictions in semi-parametric CV estimation. The asymptotic distribution of the test statistic is derived using the path-derivative method.

The fourth contribution also derives from the construction of an observationally equivalent model with two errors. We propose a test for the error dimension and in particular for the representation with a single error as in Matzkin (2003).

The fifth and final contribution is that we consider a key regularity condition for path-derivative calculations, i.e., the differentiability of the second stage conditional expectation. We find that non-differentiability is associated with a breakdown of the asymptotic normality of the semi-parametric estimator. We illustrate this for a case that is of independent interest, i.e., regression on a (constant) propensity score as could occur in covariate corrections in randomized experiments.

2 Inference with an Estimated Control Variable

In this section, we provide tools for asymptotic inference when the control variable is estimated. We begin by reviewing Imbens and Newey (2009), and describe the basic algorithm of identification of the ASF. We then review Hahn and Ridder (2013), and argue that without modification and extension, their result cannot be used for inference when the Imbens and Newey's algorithm is adopted as a basis of the estimation of finite dimensional parameters that are a functional of the ASF. We then proceed to derive that extension.

Consider identification of a nonparametric and nonseparable triangular model, where the dependent variable Y is related to the explanatory variable X through

$$Y = f\left(X, \varepsilon\right) \tag{1}$$

$$X = g\left(Z, V\right) \tag{2}$$

The ε and V are unobserved error terms, and Z is an instrumental variable independent of (ε, V). The dimension of ε is finite but arbitrary, but V is scalar. Imbens and Newey (2009) note that under some regularity conditions, (i) we can write $V = F\left(X \mid Z\right)$ where $F\left(x \mid z\right)$ denotes the conditional CDF of

X given Z;[2] (ii) given V, X and ε are independent of each other; and (iii) the ASF $m(x) = \mathbb{E}[g(x,\varepsilon)]$ can be identified by the equality

$$m(x) = \mathbb{E}[\mu(x,V)] \tag{3}$$

where

$$\mu(x,v) = \mathbb{E}[Y|\, X = x, V = v]\,. \tag{4}$$

They define a control variable as any variable V for which (ii) holds.

The natural estimator of the ASF $m(x)$ is

$$\widehat{m}(x) = \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(x,\widehat{V}_j)$$

with $\widehat{\mu}$ a nonparametric regression estimator of Y on X and $\widehat{V}$. In this section we will consider the estimation of a functional of the ASF

$$\beta = \mathbb{E}[h(X, m(X)] = \mathbb{E}[h(X, \mathbb{E}_V(\mu(X,V))] \tag{5}$$

with h a known function. The natural estimator is

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} h(X_i, \widehat{m}(X_i)) = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i, \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(X_i,\widehat{V}_j)\right) \tag{6}$$

This estimator has the structure of a two-sample V-statistic (see e.g. Serfling (1980)).

Hahn and Ridder (2013) consider inference for a related estimator that is also a functional of a nonparametrically estimated conditional expectation. They consider estimators that are averages of $h(X_i, \widehat{\mu}(X_i,\widehat{V}_i))$ with $\widehat{\mu}$ the nonparametric regression estimator of Y on $X, \widehat{V}$. This is a single not a double sum as the estimator (6). Moreover Hahn and Ridder only consider the case that $\widehat{V}$ is the residual of a parametric or nonparametric regression of X on Z, i.e., the first-stage model is a regression model with a separable error.

In this section, we provide results that generalize the results we obtained before to double sums and to a first stage with a nonseparable error, i.e., we consider the case that the control variable is obtained from $V = F(X|\, Z)$. As far as we know this has not be discussed in the semiparametric literature. Newey (1994), for example, only considers the case that the first stage estimator is a nonparametric regression.

2.1 Three-step Semi-parametric Control Variable Estimator with a Parametric First Step

We first consider a parametric first stage where $V_i = \varphi(X_i, Z_i, \alpha_*)$ is estimated by $\widehat{V}_i = \varphi(X_i, Z_i, \widehat{\alpha})$ and $\widehat{\alpha}$ is a $\sqrt{n}$-consistent estimator of α_*. Our approach is the computation of the path derivative

[2] Note that this shows that the assumptions of a scalar V, g being non-decreasing in V and Z being independent of V (that has a uniform $[0,1]$ distribution) do hold without loss of generality for the first stage in (2).

of the estimator following the approach in Newey (1994). The path derivative is a total derivative in which the contributions of the various sources of sampling variation enter additively.

If we focus on the contribution of the first-stage estimation, then for accounting purposes, it is useful to adopt the notation in Hahn and Ridder (2013), and write

$$\gamma\left(x,v;\alpha\right)=\mathbb{E}\left[Y\mid X,\varphi\left(X,Z,\alpha\right)=v\right]$$
$$\gamma\left(x,\widehat{V}_j;\widehat{\alpha}\right)=\gamma\left(x,\varphi\left(X,Z,\widehat{\alpha}\right);\widehat{\alpha}\right)$$

This notation emphasizes the two roles of α in the conditional expectation: it enters the control variable directly and it affects the conditional expectation because the distribution of the control variable depends on α. It is the second role of α that is often forgotten in the derivation of the influence function and asymptotic variance.

We will also define

$$G\left(x;\alpha_1,\alpha_2\right)=\mathbb{E}\left[\gamma\left(x,\varphi\left(X_j,Z_j;\alpha_1\right);\alpha_2\right)\right]$$
$$\widehat{G}\left(x;\alpha_1,\alpha_2\right)=\frac{1}{n}\sum_{j=1}^{n}\widehat{\gamma}\left(x,\varphi\left(X_j,Z_j;\alpha_1\right);\alpha_2\right)$$

with $\widehat{\gamma}$ a nonparametric estimator of the conditional mean of Y given $X,\widehat{V}$. An estimator of the ASF is

$$\widehat{m}(x)=\frac{1}{n}\sum_{j=1}^{n}\widehat{\gamma}\left(x,\widehat{V}_j;\widehat{\alpha}\right)=\widehat{G}\left(x;\widehat{\alpha},\widehat{\alpha}\right)$$

Note that the two roles of α are made explicit in $G\left(x,\alpha_1,\alpha_2\right)$ that is obtained by substituting $v=\varphi(x,z,\alpha_1)$ in $\gamma(x,v;\alpha_2)$. Note also that $\mu\left(x,v\right)=\gamma\left(x,v;\alpha_*\right)$, where $\mu\left(x,v\right)=\mathbb{E}\left[Y\mid X=x,\varphi\left(X,Z,\alpha_*\right)=v\right]$. The notation α_1,α_2 is just an expositional device, since $\alpha_1=\alpha_2=\alpha$.[3]

Lemma 1 *Let*

$$\beta_*=\mathbb{E}\left[h\left(X,m\left(X\right)\right)\right]$$
$$\widehat{\beta}=\frac{1}{n}\sum_{i=1}^{n}h\left(X_i,\widehat{m}\left(X_i\right)\right)=\frac{1}{n}\sum_{i=1}^{n}h\left(X_i,\widehat{G}\left(X_i;\widehat{\alpha},\widehat{\alpha}\right)\right)$$

[3] As in Hahn and Ridder (2013), all the results in this section are predicated on the assumption that (i) the derivative $\partial\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)/\partial\alpha_2$ exists, and (ii) we can interchange expectation and differentiation. We consider the implication of non-differentiability later in the paper.

We then have

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\beta}-\beta_*\right) = & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h\left(X_i,m\left(X_i\right)\right)-\mathbb{E}\left[h\left(X,m\left(X\right)\right)\right]\right) \\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{E}_X\left[\tau\left(X\right)\mu\left(X,V_i\right)\right]-\mathbb{E}\left[\tau\left(X\right)m\left(X\right)\right]\right) \\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f\left(X_i\right)f\left(V_i\right)}{f\left(X_i,V_i\right)}\tau\left(X_i\right)\left(Y_i-\mu\left(X_i,V_i\right)\right) \\
& +\left\{\begin{array}{c} \mathbb{E}\left[\mathbb{E}_{\tilde{X}}\left[\tau\left(\tilde{X}\right)\frac{\partial\mu(\tilde{X},V)}{\partial v}\right]\frac{\partial\varphi(X,Z;\alpha_*)}{\partial\alpha}\right] \\ +\mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{f(X)f(V)}{f(X,V)}\right)\tau\left(X\right)\left(Y-\mu\left(X,V\right)\right)\frac{\partial\varphi(X,Z,\alpha_*)}{\partial\alpha}\right] \\ -\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\tau\left(X\right)\frac{\partial\mu(X,V)}{\partial v}\frac{\partial\varphi(X,Z,\alpha_*)}{\partial\alpha}\right] \end{array}\right\}\sqrt{n}\left(\widehat{\alpha}-\alpha_*\right) \\
& +o_p\left(1\right)
\end{aligned}
$$

with

$$\tau\left(X\right)=\frac{\partial h}{\partial m(X)}\left(X,m\left(X\right)\right).$$

Proof. In Appendix. ■

In the expression for the influence function the first two terms are pure variance terms that account for the double sum. The third term accounts for variation due to the estimation of the conditional mean as in all semi-parametric two-step estimators. The final three terms represent variability due to the first stage estimation with the first accounting for the variation in the control variable and the second and third for the variation in the conditional expectation. The penultimate term is 0 if $x=g(z,v)$ has a unique solution for z for all x,v. This holds if g is monotonic in z and z is scalar or a one-dimensional index.

2.2 Two-step Control Variable Estimator with a Nonparametric CDF in First Step

As an intermediate result we present an extension of Newey (1994, Proposition 4). In our notation, Newey's proposition is applicable when the parameter of interest is estimated by

$$\widehat{\beta}=\frac{1}{n}\sum_{i=1}^{n}h\left(W_i,\widehat{V}_i\right)$$

if $\widehat{V}_i$ is an estimate of the conditional expectation of X given Z evaluated at Z_i. We extend his result to the case where $\widehat{V}_i=\widehat{F}\left(X_i\right|Z_i)$ with $\widehat{F}\left(x\right|z)$ a nonparametric estimator of the conditional CDF of X given $Z=z$. To our knowledge, this sort of result is not explicitly available in the literature, so there is an independent interest in establishing this. The result is useful for parametric control variable estimation as in Rivers and Vuong (1988). Consider the binary logit model with independent variables X and and the control variable $\widehat{V}$

$$\Pr(Y=1|X,\widehat{V}) = \frac{e^{\beta X+\gamma\widehat{V}}}{1+e^{\beta X+\gamma\widehat{V}}} \equiv p(X,\widehat{V})$$

The control variable estimator of β can be expressed as

$$\widehat{\beta}-\beta = \frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{V}_{11}X_i+\mathcal{V}_{12}\widehat{V}_i\right)(Y_i-p(X_i,\widehat{V}_i))+o_p(1)$$

with $\mathcal{V}_{11}$ and $\mathcal{V}_{12}$ components of the inverse of the information matrix corresponding to β,γ. The right hand side is an average as above with $W=(Y\ X\ Z)'$.

Lemma 2 *Let* $W=(Y\ X\ Z)'$ *and* $V=F(X|Z)$

$$\begin{aligned}\beta_* &= \mathbb{E}\left[h\left(W,V\right)\right]\\ \widehat{V}_i &= \widehat{F}\left(X_i|\,Z_i\right)\\ \widehat{\beta} &= \frac{1}{n}\sum_{i=1}^{n}h\left(W_i,\widehat{V}_i\right)\end{aligned}$$

We then have

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}-\beta_*\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h\left(W_i,V_i\right)-\mathbb{E}\left[h\left(W,V\right)\right]\right)\\ &+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta\left(X_i,Z_i\right)-\overline{\delta}\left(Z_i\right)\right)+o_p\left(1\right)\end{aligned}$$

where

$$\begin{aligned}\delta\left(x,z\right) &= \mathbb{E}\left[\left.\frac{\partial h\left(W,V\right)}{\partial v}1\left(x\leq X\right)\right|Z=z\right]\\ \overline{\delta}\left(z\right) &= \mathbb{E}\left[\left.\frac{\partial h\left(W,V\right)}{\partial v}F\left(X|\,Z\right)\right|Z=z\right]\end{aligned}$$

Proof. In Appendix. ■

2.3 Three-step Semi-parametric Control Variable Estimator with a Nonparametric CDF in the First Step

We now consider the estimator in (6) with $\widehat{V}_i=\widehat{F}\left(X_i|\,Z_i\right)$. Combining Lemmas 1 and 2 leads us to:

Lemma 3 *Let* $W=(Y\ X\ Z)'$ *and* $m(x)=\mathbb{E}[\mu(x,V)]$, $\widehat{m}(x)=\frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(x,\widehat{V}_j)$ *with* $V=F(X|Z)$ *and* $\widehat{V}_i=\widehat{F}(X_i|Z_i)$ *and*

$$\begin{aligned}\beta_* &= \mathbb{E}\left[h\left(X,m\left(X\right)\right)\right]\\ \widehat{\beta} &= \frac{1}{n}\sum_{i=1}^{n}h\left(X_i,\widehat{m}\left(X_i\right)\right)\end{aligned}$$

We then have

$$\sqrt{n}\left(\widehat{\beta}-\beta_*\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h\left(X_i, m\left(X_i\right)\right) - \mathbb{E}\left[h\left(X, m\left(X\right)\right)\right]\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{E}_X\left[\tau\left(X\right)\mu\left(X, V_i\right)\right] - \mathbb{E}\left[\tau\left(X\right)m\left(X\right)\right]\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f\left(X_i\right)f\left(V_i\right)}{f\left(X_i, V_i\right)}\tau\left(X_i\right)\left(Y_i - \mu\left(X_i, V_i\right)\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta\left(X_i, Z_i\right) - \overline{\delta}\left(Z_i\right)\right) + o_p\left(1\right)$$

where

$$\tau\left(X\right) = \frac{\partial h}{\partial m(X)}\left(X, m\left(X\right)\right)$$
$$\xi\left(W\right) = \mathbb{E}_{\tilde{X}}\left[\tau\left(\tilde{X}\right)\frac{\partial\mu\left(\tilde{X}, V\right)}{\partial v}\right] + \frac{\partial}{\partial v}\left(\frac{f\left(X\right)f\left(V\right)}{f\left(X, V\right)}\right)\tau\left(X\right)\left(Y - \mu\left(X, V\right)\right)$$
$$-\frac{f\left(X\right)f\left(V\right)}{f\left(X, V\right)}\tau\left(X\right)\frac{\partial\mu\left(X, V\right)}{\partial v}$$
$$\delta\left(x, z\right) = \mathbb{E}\left[\xi\left(W\right)1\left(x \leq X\right)\middle| Z = z\right]$$
$$\overline{\delta}\left(z\right) = \mathbb{E}\left[\xi\left(W\right)F\left(X\middle| Z\right)\middle| Z = z\right]$$

Proof. In Appendix. ■

2.4 Average Derivatives

We now consider the extension of Lemmas 1 and 3 to two types of average derivatives. The first average derivative is

$$\beta_* = \mathbb{E}\left[\frac{\partial m}{\partial x}(X)\right] \tag{7}$$

with estimator

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{m}\left(X_i\right)}{\partial x} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{G}\left(X_i;\widehat{\alpha},\widehat{\alpha}\right)}{\partial x}$$

The Lemma below gives the asymptotically linear representation

Lemma 4 *Let*

$$\beta_* = \mathbb{E}\left[\frac{\partial m}{\partial x}(X)\right]$$
$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{m}\left(X_i\right)}{\partial x}$$

with

$$\widehat{m}(x) = \frac{1}{n}\sum_{j=1}^{n}\widehat{\mu}(X_j, \widehat{V}_j) \qquad \mu(x,v) = \mathbb{E}[Y|X=x, V=v]$$

then if $V = \varphi(X, Z, \alpha_*)$ *and* $\widehat{V}_i = \varphi(X_i, Z_i, \hat{\alpha})$ *(parametric first stage) and the density of* X *is 0 at the boundary of the support*

$$\begin{aligned}
\sqrt{n}(\widehat{\beta} - \beta_*) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial m\left(X_i\right)}{\partial x} - \mathbb{E}\left[\frac{\partial m\left(X\right)}{\partial x}\right]\right) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{E}_X\left[\frac{\partial \mu\left(X, V_i\right)}{\partial x}\right] - \mathbb{E}\left[\frac{\partial m\left(X\right)}{\partial x}\right]\right) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\frac{\partial f(X_i)}{\partial x}f\left(V_i\right)}{f\left(X_i, V_i\right)}\left(Y_i - \mu\left(X_i, V_i\right)\right) \\
&+ \left\{\begin{array}{c} -\mathbb{E}\left[\mathbb{E}_{\tilde{X}}\left[\frac{\frac{\partial f(\tilde{X})}{\partial x}}{f(\tilde{X})}\frac{\partial \mu(\tilde{X}, V)}{\partial v}\right]\frac{\partial \varphi(X, Z; \alpha_*)}{\partial \alpha}\right] \\ -\mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{\frac{\partial f(X)}{\partial x}f(V)}{f(X,V)}\right)\left(Y - \mu(X,V)\right)\frac{\partial \varphi(X,Z,\alpha_*)}{\partial \alpha}\right] + \mathbb{E}\left[\frac{\frac{\partial f(X)}{\partial x}f(V)}{f(X,V)}\frac{\partial \mu(X,V)}{\partial v}\frac{\partial \varphi(X,Z,\alpha_*)}{\partial \alpha}\right]\end{array}\right\}\sqrt{n}(\widehat{\alpha} - \alpha_*) \\
&+ o_p\left(1\right).
\end{aligned}$$

If $V = F(X|X)$ *and* $\widehat{V}_i = \widehat{F}(X_i|Z_i)$ *(nonparametric first stage), the final term on the right-hand side is (other terms are the same)*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta\left(X_i, Z_i\right) - \overline{\delta}\left(Z_i\right)\right)$$

with

$$\begin{aligned}
\delta\left(x, z\right) &= \mathbb{E}\left[\xi\left(W\right)1\left(x \leq X\right)\middle| Z = z\right] \\
\overline{\delta}\left(z\right) &= \mathbb{E}\left[\xi\left(W\right)F\left(X\middle| Z\right)\middle| Z = z\right]
\end{aligned}$$

and

$$\xi(W) = -\mathbb{E}_{\tilde{X}}\left[\frac{\frac{\partial f(\tilde{X})}{\partial x}}{f\left(\tilde{X}\right)}\frac{\partial \mu\left(\tilde{X}, V\right)}{\partial v}\right] - \frac{\partial}{\partial v}\left(\frac{\frac{\partial f(X)}{\partial x}f\left(V\right)}{f\left(X,V\right)}\right)\left(Y - \mu\left(X,V\right)\right) + \frac{\frac{\partial f(X)}{\partial x}f\left(V\right)}{f\left(X,V\right)}\frac{\partial \mu\left(X,V\right)}{\partial v}$$

Proof. In Appendix. ■

Imbens and Newey (2009) consider another average derivative for (1)

$$\beta_* = \mathbb{E}\left[\frac{\partial f(X, \varepsilon)}{\partial x}\right]$$

This average derivative is identified by

$$\beta_* = \mathbb{E}\left[\frac{\partial \mu(X,V)}{\partial x}\right]$$

With a parametric first stage this is close to the the three-step estimators considered in Hahn and Ridder (2013) but their result does not cover the nonparametric first stage.

Lemma 5 *Let*

$$\beta_* = \mathbb{E}\left[\frac{\partial \mu(X,V)}{\partial x}\right]$$
$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \widehat{\mu}(X_i,\widehat{V}_i)}{\partial x}$$

then if $V = \varphi(X, Z, \alpha_*)$ *and* $\widehat{V}_i = \varphi(X_i, Z_i, \hat{\alpha})$ *(parametric first stage) and the density of* X, Z *and of* X, V *is 0 at the boundary of the support for* X *for all values* z *and* v *in the support of* Z *and* V

$$\begin{aligned}\sqrt{n}(\widehat{\beta} - \beta_*) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial \mu\left(X_i,V_i\right)}{\partial x} - \mathbb{E}\left[\frac{\partial \mu\left(X,V\right)}{\partial x}\right]\right) \\ & - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\frac{\partial f(X_i,Z_i)}{\partial x}}{f(X_i,Z_i)}(Y_i - \mu(X_i,V_i)) \\ & - \left\{\begin{array}{c}\mathbb{E}\left[\frac{\frac{\partial f(X,Z)}{\partial x}}{f(X,Z)}\frac{\partial \mu(X,V)}{\partial v}\frac{\partial \varphi(X,Z;\alpha_*)}{\partial \alpha}\right] + \mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{\frac{\partial f(X,V)}{\partial x}}{f(X,V)}\right)(Y - \mu(X,V))\frac{\partial \varphi(X,Z,\alpha_*)}{\partial \alpha}\right] \\ -\mathbb{E}\left[\frac{\frac{\partial f(X,V)}{\partial x}}{f(X,V)}\frac{\partial \mu(X,V)}{\partial v}\frac{\partial \varphi(X,Z,\alpha_*)}{\partial \alpha}\right]\end{array}\right\}\sqrt{n}(\widehat{\alpha} - \alpha_*) \\ & + o_p\left(1\right).\end{aligned}$$

If $V = F(X|X)$ *and* $\widehat{V}_i = \widehat{F}(X_i|Z_i)$ *(nonparametric first stage), the final term on the right-hand side is (other terms are the same)*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta\left(X_i,Z_i\right) - \overline{\delta}\left(Z_i\right)\right)$$

with

$$\delta\left(x,z\right) = \mathbb{E}\left[\xi\left(W\right)1\left(x \leq X\right)\middle| Z = z\right]$$
$$\overline{\delta}\left(z\right) = \mathbb{E}\left[\xi\left(W\right)F\left(X\middle| Z\right)\middle| Z = z\right]$$

and

$$\xi(W) = -\frac{\frac{\partial f(X,Z)}{\partial x}}{f(X,Z)}\frac{\partial \mu(X,V)}{\partial v} - \frac{\partial}{\partial v}\left(\frac{\frac{\partial f(X,V)}{\partial x}}{f\left(X,V\right)}\right)(Y - \mu(X,V)) + \frac{\frac{\partial f(X,V)}{\partial x}}{f\left(X,V\right)}\frac{\partial \mu\left(X,V\right)}{\partial v}$$

Proof. In Appendix. ■

3 Estimation of Parametric Models with Nonseparable Errors and Continuous Endogenous Regressors

The identification method of Imbens and Newey (2009) identifies the Average Structural Function (ASF) $m(x) = \mathbb{E}[f(x,\varepsilon)]$. In this section we will consider the estimation of parametric models with nonseparable errors, i.e.,

$$Y = f(X, \varepsilon; \beta)$$

The independent variables are $X = (X_1 \ Z_1')'$ with X_1 possibly correlated with ε, but Z_1 independent of ε. The instrument is denoted by $Z = (Z_1' \ Z_2)'$. In this section we take X_1 as scalar but the extension to multivariate X_1 if we have an instrument vector Z_2 of the same dimension only involves more complicated notation. The first stage is nonparametric

$$X_1 = F^{-1}(V|Z)$$

The assumption on the random errors ε, V is

Assumption 1 (Joint independence)

$$\varepsilon, V \perp Z$$

The assumption $V \perp Z$ holds by construction.

If we allow that ε is multidimensional, then in general we will not be able to identify the joint distribution of ε even if X is exogenous. However the ASF is still identified. The ASF can be considered as the reduced form of the model. The proposed estimator minimizes the average difference between the ASF and the parametric model. In the case that ε is scalar and $f(x, \varepsilon; \beta)$ is monotonic in ε the conditional distribution of Y given $X = x$ is identified by the same method that identifies the ASF. In that case the reduced form conditional CDF of Y given $X = x$ can be compared to that implied by the parametric model. In the sequel we only consider the case that $f(x, \varepsilon; \beta)$ has a multidimensional error and is not necessarily monotonic in ε.

As an example we consider a logit model with a single continuous endogenous regressor

$$Y^* = \beta' X - \varepsilon$$

where ε has a logistic distribution and X and ε are possibly correlated. We observe

$$Y = I(Y^* \geq 0)$$

If X and ε are independent then

$$\mathbb{E}[Y|X = x] = \frac{e^{\beta' x}}{1 + e^{\beta' x}} \equiv r(x; \beta) \tag{8}$$

If in the spirit of Rivers and Vuong (1988) we are prepared to assume that

$$\varepsilon | X = x, V = v \stackrel{d}{=} \varepsilon | V = v \stackrel{d}{=} \gamma k(v) + \eta$$

where η has a logistic distribution then

$$\Pr(Y = 1 | X = x, S = s, V = v) = \frac{e^{\beta' x + \gamma k(v)}}{1 + e^{\beta' x + \gamma k(v)}}$$

and the parameters can be estimated by two-step MLE after estimating $\hat{V}_i$ in the first step. Section 2.2 gives the result needed to compute the influence function and asymptotic variance of this estimator. The support condition is not needed for this estimator.

In general the Average Structural Function (ASF)

$$m(x) = \mathbb{E}[f(x, \varepsilon; \beta))]$$

is identified by

$$m(x) = \int_0^1 \mathbb{E}\left[Y \mid X = x, V = v\right] \mathrm{d}v$$

This requires that the support of V is $[0, 1]$. For some x the control variable V may take values in a subset of $[0, 1]$ and in the estimation these values of X will be excluded.[4]

Assumption 2 (Support) *There is a subset of the support of X denoted by $\mathcal{X}$ such that V has support $[0, 1]$ for all $x \in \mathcal{X}$. The parameters β are identified if the support of X is restricted to $\mathcal{X}$.*

The estimator of the ASF is

$$\widehat{m}(x) = \frac{1}{n} \sum_{j=1}^{n} \widehat{\mu}\left(x, \widehat{V}_j\right)$$

with

$$V_j = \widehat{F}\left(X_{1j} \mid Z_j\right)$$

and $\widehat{\mu}(x, v)$ the nonparametric series estimator of $\mu(x, v) = \mathbb{E}\left[Y \mid X = x, V = v\right]$

The estimator of the parametric model is the solution to

$$\min_{\beta} \sum_{i=1}^{n} 1_{\mathcal{X}}(X_i)\left(\widehat{m}(X_i) - r(X_i; \beta)\right)^2$$

Under the assumptions of Lemma 2 in Hahn and Ridder (2014) this estimator is asymptotically equivalent to (from now on we ignore the restriction of the observations to $\mathcal{X}$

$$\sqrt{n}(\widehat{\beta} - \beta_*) = \left(\mathbb{E}\left[\frac{\partial r}{\partial \beta}(X; \beta_*)\frac{\partial r}{\partial \beta'}(X; \beta_*)\right]\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\widehat{m}(X_i) - m(X_i)\right) \frac{\partial r}{\partial \beta}(X_i; \beta_*) + o_p(1) \qquad (9)$$

The asymptotic distribution of $\widehat{\beta}$ can be established with the help of Lemma 1 for the parametric and Lemma 3 for the nonparametric first stage.

Lemma 6 *Let*[5] *$W = (Y\ X'\ Z')'$ and $m(x) = \mathbb{E}[\mu(x, V)]$, $\widehat{m}(x) = \frac{1}{n}\sum_{j=1}^{n} \widehat{\mu}(x, \widehat{V}_j)$. If the estimator $\widehat{\beta}$ satisfies (41), then if $V = \varphi(X_1, Z, \alpha)$ and $\widehat{V}_i = \varphi(X_{1i}, Z_i, \widehat{\alpha})$ (parametric first stage) and $A =$*

[4] Our inference results are also useful if β is interval identified. This is however beyond the scope of the present paper.

[5] Note that X and Z share the vector Z_1. For instance, the joint density of X, Z is that of X_1, Z_1, Z_2. This is left implicit in the expressions.

$\left(\mathbb{E}\left[\frac{\partial r}{\partial \beta}(X;\beta_*)\frac{\partial r}{\partial \beta'}(X;\beta_*)\right]\right)^{-1}$

$$\begin{aligned}\sqrt{n}(\widehat{\beta}-\beta_*) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A\left(\mathbb{E}_X\left[\mu(X,V_i)\frac{\partial r}{\partial \beta}(X;\beta_*)\right]-\mathbb{E}\left[m(X)\frac{\partial r}{\partial \beta}(X;\beta_*)\right]\right)\\ &+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i-\mu(X_i,V_i))\frac{\partial r}{\partial \beta}(X_i;\beta_*)\\ &+A\left\{\begin{array}{c}\mathbb{E}\left[\mathbb{E}_{\tilde{X}}\left[\frac{\partial r}{\partial \beta}(\widetilde{X};\beta_*)\frac{\partial \mu(\tilde{X},V)}{\partial v}\right]\frac{\partial \varphi(X_1,Z;\alpha_*)}{\partial \alpha'}\right]\\ +\mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{f(X)f(V)}{f(X,V)}\right)\frac{\partial r}{\partial \beta}(X;\beta_*)\left(Y-\mu\left(X,V\right)\right)\frac{\partial \varphi(X_1,Z,\alpha_*)}{\partial \alpha'}\right]\\ -\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial r}{\partial \beta}(X;\beta_*)\frac{\partial \mu(X,V)}{\partial v}\frac{\partial \varphi(X_1,Z,\alpha_*)}{\partial \alpha'}\right]\end{array}\right\}\sqrt{n}(\widehat{\alpha}-\alpha_*)+o_p(1)\end{aligned}$$

If $V = F(X_1|Z)$ and $\widehat{V}_i = \widehat{F}(X_{1i}|Z_i)$ (non-parametric first stage) then the contribution of the first stage nonparametric estimation is

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A(\delta(X_{1i},Z_i)-\overline{\delta}(Z_i))$$

with for

$$\begin{aligned}\xi(W) =& \mathbb{E}_{\tilde{X}}\left[\frac{\partial r}{\partial \beta}(\widetilde{X};\beta_*)\frac{\partial \mu\left(\tilde{X},V\right)}{\partial v}\right]+\frac{\partial}{\partial v}\left(\frac{f\left(X\right)f\left(V\right)}{f\left(X,V\right)}\right)\frac{\partial r}{\partial \beta}(X;\beta_*)\left(Y-\mu\left(X,V\right)\right)\\ &-\frac{f\left(X\right)f\left(V\right)}{f\left(X,V\right)}\frac{\partial r}{\partial \beta}(X;\beta_*)\frac{\partial \mu\left(X,V\right)}{\partial v}\end{aligned}$$

we define

$$\delta(x_1,z) = \mathbb{E}\left[\xi(W)1(x_1 \le X_1)|Z=z\right]$$
$$\overline{\delta}(z) = \mathbb{E}\left[\xi(W)F(X_1|Z)|Z=z\right]$$

The other terms of the influence function are equal to those in the parametric case.

Proof. In Appendix. ■

The expression that we obtain for the parametric first-stage is the same as that derived in Hahn and Ridder (2014) using a stochastic expansion. That argument is much more complicated and the path derivative calculation shows how we can obtain the influence function directly without having to go through the derivation of the expansion. The stochastic expansion will provide regularity conditions, e.g. restrictions on the order of the series estimator as a function of the number of observations.

4 Application - Test of Overidentification in Nonparametric and Nonseparable Triangular Model

[JH: I shortened the material from the CMRX file. Perhaps too much.]

Despite plethora of literature on identification of nonparametric and nonseparable triangular models, we find no literature on the basic issue of specification testing, i.e., the test of overidentification in the linear model counterparts. In this section, we try to bridge the gap by presenting such a test. The idea is based on a "construction" of pseudo-models under the assumption that there is certain monotonicity in the first stage. We note that if there are two instruments available both satisfying the monotonicty assumption, then the pseudo-ASF's identified by both instruments should coincide under correct specification.

4.1 Creation of a Nonparametric Triangular Simultaneous Equations Model

Suppose that we are given a joint population distribution of a triplet (Y, X, Z). If we define the random variable ε by

$$\varepsilon = F\left(Y\middle| X\right)$$

with $F\left(Y\middle| X\right)$ the conditional CDF of a scalar Y given X, then

$$Y = f\left(X, \varepsilon\right) \qquad \varepsilon \perp X$$

with $f\left(X, \varepsilon\right) = F^{-1}\left(\varepsilon\middle| X\right)$. Matzkin (2003) used this to identify a structural relation between Y and X. If the structural error ε^* is multidimensional, then we cannot recover the structural relation from the inverse conditional CDF, but if X and ε^* are independent we can recover the average structural function (ASF) $E\left[Y\middle| X = x\right] = E\left[f^*\left(x, \varepsilon^*\right)\right]$.

We present a similar argument, and argue that it is possible to 'create' a nonparametric triangular system as discussed in e.g. Imbens (2006), Matzkin (2008), and Imbens and Newey (2009). The triangular simultaneous equations model is given by the two equations

$$Y = f\left(X, \varepsilon\right)$$
$$X = g\left(Z, V\right)$$

where the instrument Z is independent of the vector (ε^*, V).

We begin with the construction of the first stage, i.e., the relation between X and Z. The most general model for the relation between a dependent variable X and a vector of independent variables Z is

$$X = g\left(Z, V\right) \qquad\qquad Z \perp V \tag{10}$$

with $g\left(z, v\right)$ monotone in v for (almost) all values of z. This model can be 'constructed' from the joint distribution of X, Z by defining the random variable V as

$$V = G\left(X\middle| Z\right) \tag{11}$$

with $G\left(X\middle| Z\right)$ the conditional CDF of a scalar X given Z. Because

$$\Pr(V \leq v|Z = z) = \Pr(G(X|Z) \leq v|Z = z) = \Pr(X \leq G^{-1}(v|z)|Z = z) = v \tag{12}$$

the error V has a uniform distribution that is independent of Z. Upon inversion we obtain (10).

Now, let $H(y|z,v) = \Pr(Y \le y| Z = z, V = v)$ be the conditional CDF of Y given (Z,V), and define $U = H(Y|Z,V)$. By the same argument as above we have that U has a uniform distribution and is independent of (Z,V). Hence if we define $h(z,v,u) = H^{-1}(u|z,v)$, we have

$$Y = h(Z,V,U) \qquad U \perp Z,V \tag{13}$$

Note that by construction g,h are increasing in their last argument and that Z,U,V are mutually independent.

To construct a triangular simultaneous equations model we would like to invert (10) with respect to Z and express Z as a function of (X,V). The simplest case is that $g(z,v)$ is strictly monotonic, without loss of generality strictly increasing, in z for (almost all) v. This is equivalent to assuming that the joint distribution of X,Z is such that if $z > z'$ then $G(x|z) < G(x|z')$ for all x in the union of the supports of these distributions, i.e., the distribution of X given $Z = z$ is increasing in z if we order distributions according to first-order stochastic dominance. Because monotonicity is equivalent to $G(x|z)$ being decreasing in z for all x we can check whether this assumption holds.

If $g(z,v)$ is strictly increasing in z for (almost all) v, then

$$Z = g^{-1}(X,V) \tag{14}$$

where g^{-1} is the inverse with respect to the first argument. Note that in (14) X and V are not independent. Substitution in equation (13) gives

$$Y = h\left(g^{-1}(X,V),V,U\right) = f(X,\varepsilon) \tag{15}$$

with ε the vector (U,V). Hence we have constructed a triangular system (15) and (10) with errors (ε,V) that are independent of Z.[6]

Under monotonicity in Z we have that

$$Z \perp (\varepsilon,V) \quad \Rightarrow \quad X \perp \varepsilon \mid V$$

Therefore for the constructed model

$$\begin{aligned} E[Y|X=x,V=v] &= E[f(X,\varepsilon)|X=x,V=v] \\ &= E\left[f(x,\varepsilon)|Z=g^{-1}(x,v),V=v\right] = E[f(x,\varepsilon)|V=v] \end{aligned}$$

Note that if the structural relation is

$$Y = f^*(X,\varepsilon^*)$$

[6]The result in this paper used to be in the working paper version of Hahn and Ridder (2011), which was taken out in the published version.

and Z is a valid instrument, i.e., $Z \perp (\varepsilon^*, V)$, then under monotonicity the same argument applies, so that

$$E_{\varepsilon^*}\left[f^*(x, \varepsilon^*) \middle| V = v\right] = \int E\left[Y \middle| X = x, V = v\right] \mathrm{d}v = E_{\varepsilon}\left[f(x, \varepsilon) \middle| V = v\right]$$

See Kasy (2013) for related discussion.

4.2 Intuition

To understand the intuition behind our construction we first consider the construction of a linear triangular simultaneous equations model. In this model the inversion of conditional CDF's is replaced by linear projections. Also because the errors are additively separable we do not have to deal with multidimensional random errors.

Suppose that we are given a joint population distribution of a triplet (Y, X, Z). Let

$$X = \alpha_X Z + V$$

be the population linear projection of X on Z, and let

$$Y = \alpha_Y Z + \rho V + U$$

be the population linear projection of Y on Z and V. By construction, we have $E[ZV] = 0$, $E[ZU] = 0$, and $E[VU] = 0$. In other words, Z, V, and U are uncorrelated.

Assume that $\alpha_X \neq 0$, which is testable. We may then write

$$Z = \frac{1}{\alpha_X} X - \frac{1}{\alpha_X} V$$

If we substitute this into the equation for Y, we obtain

$$Y = \frac{\alpha_Y}{\alpha_X} X - \frac{\alpha_Y}{\alpha_X} V + \rho V + U$$

If we define $\beta = \frac{\alpha_Y}{\alpha_X}$ and $\varepsilon = -\frac{\alpha_Y}{\alpha_X} V + \rho V + U$, we obtain a linear 'model'

$$Y = \beta X + \varepsilon$$

where the error ε is such that $E[X\varepsilon] = \left(\rho - \frac{\alpha_Y}{\alpha_X}\right) \sigma_V^2 \neq 0$ and $E[Z\varepsilon] = 0$.

4.3 Test of Overidentification

Our construction provides a basis of specification test when two or more instruments are available. If we have two valid instruments Z_1 and Z_2 available, and if the first stages constructed with Z_1 and Z_2 are both monotonic, then we can construct the second-stage equations from these two different first stages. If there is a structural relation $Y = f^*(X, \varepsilon^*)$ then for values of x where both first-stage errors V_1 and V_2 have support $[0,1]$ we can recover the ASF of the structural model by

$$M_1(x) = \int_0^1 \mathbb{E}[Y|\, X = x, V_1 = v_1]\, \mathrm{d}v_1$$

or by

$$M_2(x) = \int_0^1 \mathbb{E}[Y|\, X = x, V_2 = v_2]\, \mathrm{d}v_2$$

Equality of these ASF's holds if the relation f^* is structural. Therefore our construction suggests a nonparametric overidentification test.

One possibility is to test whether $\frac{\partial M_1(x)}{\partial x} = \frac{\partial M_2(x)}{\partial x}$ on the average, i.e., we may want to test

$$H_0 : \mathbb{E}\left[\frac{\partial M_1(X)}{\partial x} - \frac{\partial M_2(X)}{\partial x}\right] = 0$$

Although one may prefer a test along the line of $H_0 : \mathbb{E}\left[(M_1(X) - M_2(X))^2\right] = 0$, the derivative based approach may seem preferable because of its natural interpretability in view of the linear model; in the linear simultaneous equations model with two instruments, a test of overidentification is equivalent to testing whether the two IV estimators of the slope coefficients (i.e., the derivative of the ASF) are equal to each other. In particular, we can let

$$\widehat{M_1}(x) = \frac{1}{n}\sum_{j=1}^{n} \widehat{\mu}\left(x, \widehat{V}_{1,j}\right)$$

$$\widehat{M_2}(x) = \frac{1}{n}\sum_{j=1}^{n} \widehat{\mu}\left(x, \widehat{V}_{2,j}\right)$$

where $\widehat{V}_1 = \widehat{F}(X|\, Z_1)$ and $\widehat{V}_2 = \widehat{F}(X|\, Z_1)$, and use the test statistic

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \widehat{M_1}(X_i)}{\partial x} - \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \widehat{M_2}(X_i)}{\partial x}$$

(Extension of) Lemma 4 provides a natural way of assessing the asymptotic variance of the test statistic under the null.

5 Cautionary Tale

As in Hahn and Ridder (2013), all the results in Section 2 this section are predicated on the assumption that (i) the derivative $\partial\gamma(x, \varphi(x, z, \alpha_*); \alpha_*)/\partial\alpha_2$ exists, and (ii) we can interchange expectation and differentiation. Hahn and Ridder (2013) state "Analysis of the case that the derivative does not exist, which may render $\sqrt{n}$-consistency or asymptotic normality infeasible, is beyond the scope of this paper." In this section, we consider an implication of the non-differentiability. We do this by reporting and discussing an anomaly in one of the most popular estimators of the Average Treatment Effect

(ATE). The estimator under consideration estimates the ATE based on a non-parametric regressions $\mathbb{E}\left[y|D=1,\varphi\left(x\right)\right]$, $\mathbb{E}\left[y|D=0,\varphi\left(x\right)\right]$, where the vector x contains covariates that are not affected by the intervention, and $\varphi\left(x\right)=\Pr(D=1|x)$ denotes the probability of selection or propensity score. This estimator was suggested and analyzed by Heckman, Ichimura, and Todd (HIT) (1998). Hahn and Ridder (2013) argued that under certain regularity conditions, it has the same asymptotic distribution as the estimators proposed by Hahn (1998) and Hirano, Imbens and Ridder (2003). In this paper, we note that a regularity condition in Hahn and Ridder (2013) is violated when the propensity score is constant, and we discuss the implication of this violation for the HIT estimator. Our conclusion is that the HIT estimator is not asymptotically normal in this case.

In the remainder of this paper, the y_0, y_1 denote the control and treated outcome, respectively. The treatment indicator is D and $y = Dy_1 + (1-D)y_0$ is the observed outcome. The parameter of interest is $\beta = \mathbb{E}[y_1 - y_0]$. We omit the i subscript when obvious. The result in this paper may be of more than just theoretical interest. If we consider a randomized experiment with $\varphi\left(x\right) = .5$, i.e., half of the experimental units is assigned at random to treatment and half to control, then, if covariates x unaffected by treatment are available, the best possible variance of an ATE estimator is

$$V_{cov} = 2\mathbb{E}[V(y_1|x)] + 2\mathbb{E}[V(y_0|x)] + V(\mathbb{E}[y_1 - y_0|x])$$

The variance of the estimator that does not use the covariates is

$$V_{nocov} = 2V(y_1) + 2V(y_0)$$

The variance of the estimator that does not use the covariates is larger and the difference is equal to $V(\mathbb{E}[y_1 - y_0|x])$. This is 0 if y_0, y_1 are not mean dependent on x. In finite samples the variance with covariates may be larger than that without if the relation between the outcomes and the covariates is sufficiently weak, but in large samples the former is always smaller than the latter.[7] Given that estimators that use the covariates are more accurate, one could be tempted to use the HIT estimator to obtain a more precise estimate of the ATE. This note shows that it would be a bad choice. The HIT estimator is not asymptotically normal if the propensity score is constant and therefore inference based on asymptotic normality based, such as the normal confidence interval is fragile.

5.1 Constant Propensity Score

We assume that

$$y_0 = \beta_0 + x'\gamma_0 + \varepsilon_0,$$
$$y_1 = \beta_1 + x'\gamma_1 + \varepsilon_1,$$

[7] Duflo, Glennerster, and Kremer (2008) discuss the use of covariates to reduce the variance of an ATE estimate.

where x is a K-dimensional covariate. We assume that $x \sim N(\mu, I_K)$, $\varepsilon_0 \sim N(0,1)$, and $\varepsilon_1 \sim N(0,1)$ are mutually independent.[8] In order to simplify the analysis, we will assume that our parameter of interest is the mean outcome for the controls

$$E[y_0] = \beta_0 + \mu'\gamma_0.$$

We estimate the "propensity score" $\Lambda(\zeta + x'\alpha)$ by logit MLE. As in HIT we will use regression on the propensity score, i.e., the estimator is based on

$$E(y_0) = E\left[E\left[y_0 \middle| \Lambda(\zeta + x'\alpha)\right]\right] = E\left[E\left[y_0 \middle| x'\alpha\right]\right] \tag{16}$$

We assume that $\alpha_* = 0$, i.e., the population propensity score is constant.

The constant propensity assumption leads to non-differentiability. For this purpose, we note by exploiting normality that

$$\begin{aligned} E\left[y_0 \middle| x'\alpha\right] &= \beta_0 + \mu'\gamma_0 + \frac{\operatorname{Cov}(x'\alpha, y_0)}{\operatorname{Var}(x'\alpha)}(x-\mu)'\alpha \\ &= \beta_0 + \mu'\gamma_0 + \frac{\alpha'\gamma_0}{\alpha'\alpha}(x-\mu)'\alpha \qquad \text{if } \alpha \neq 0, \end{aligned}$$

and

$$E\left[y_0 \middle| x'\alpha\right] = E[y_0] = \beta_0 + \mu'\gamma_0 \qquad \text{if } \alpha = 0.$$

This conditional expectation is non-differentiable at $\alpha = \alpha_* = 0$, which follows immediately from the fact that it is not even continuous at this point. For example, consider a path $\alpha = t \cdot \widetilde{\alpha}$ for some scalar t, then we can see that

$$\frac{\alpha'\gamma_0}{\alpha'\alpha}(x-\mu)'\alpha = \frac{\widetilde{\alpha}'\gamma_0}{\widetilde{\alpha}'\widetilde{\alpha}}(x-\mu)'\widetilde{\alpha}$$

which implies that

$$\lim_{t\to 0} E\left[y_0 \middle| x'\alpha\right] = \beta_0 + \mu'\gamma_0 + \frac{\widetilde{\alpha}'\gamma_0}{\widetilde{\alpha}'\widetilde{\alpha}}(x-\mu)'\widetilde{\alpha} \neq E\left[y_0 \middle| x'\alpha_*\right]$$

in general. This implies that an important regularity condition in Hahn and Ridder (2013) does not hold.

5.2 Non-differentiability and Non-normality

We next discuss the consequences of the non-differentiability. The HIT estimator uses a nonparametric estimator of $E[y_0| \Lambda(\zeta + x'\alpha)]$, but we exploit the built-in linearity (guaranteed by normality) and estimate $E[y_0| x'\alpha]$ (16 shows that that is equivalent) by

$$y_0 \approx \theta_0 + \left(x'\widehat{\alpha}\right)\delta_0$$

[8] The functional form and distributional assumptions simplify the argument but are not necessary for non-differentiability.

by OLS over the subsample $D = 0$. Because we are interested in $E[y_0]$, the parametric version of the HIT estimator would estimate this mean by

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}_0\left(x_i'\widehat{\alpha}\right)\widehat{\delta}_0\right) = \widehat{\theta}_0 + \left(\overline{x}'\widehat{\alpha}\right)\widehat{\delta}_0$$

If we average over the subsample $D = 0$, then by a well-known algebraic property of OLS we have

$$\widehat{\theta}_0 + \left(\overline{x}_0'\widehat{\alpha}\right)\widehat{\delta}_0 = \overline{y}_0$$

where $\overline{x}_0'$, and $\overline{y}_0$ denote the sample means of x and y over the subsample $D = 0$. Therefore, the estimator of $E[y_0]$ is equal to

$$\begin{aligned}\widehat{E}[y_0] &= \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}_0 + \left(x_i'\widehat{\alpha}\right)\widehat{\delta}_0\right) = \widehat{\theta}_0 + \left(\overline{x}'\widehat{\alpha}\right)\widehat{\delta}_0 \\ &= \widehat{\theta}_0 + \left(\overline{x}_0'\widehat{\alpha}\right)\widehat{\delta}_0 + \left(\left(\overline{x}-\overline{x}_0\right)'\widehat{\alpha}\right)\widehat{\delta}_0 \\ &= \overline{y}_0 + \left(\left(\overline{x}-\overline{x}_0\right)'\widehat{\alpha}\right)\widehat{\delta}_0 \qquad (17)\end{aligned}$$

Lemma 7 $\sqrt{n}\left(\left(\overline{x}-\overline{x}_0\right)'\widehat{\alpha}\right)\widehat{\delta}_0 \xrightarrow{d} \frac{(\mathcal{Z}_1'\mathcal{Z}_2)(\gamma_0'\mathcal{Z}_1)}{\mathcal{Z}_1'\mathcal{Z}_1}$.

The estimator $E[y_0]$ is equal to

$$\overline{y}_0 + \left(\left(\overline{x}-\overline{x}_0\right)'\widehat{\alpha}\right)\widehat{\delta}_0$$

by (17). Noting that $\sqrt{n}\left(\overline{y}_0 - E[y_0]\right) \xrightarrow{d} \mathcal{Y}_0$ for some $\mathcal{Y}_0 \sim N\left(0, \mathrm{Var}\left(y_0\right)\right)$, we can see that Lemma 7 above implies that the estimator cannot be asymptotically normal. More formally, we have

$$\sqrt{n}\left(\widehat{E}[y_0] - E[y_0]\right) \to \mathcal{Y}_0 + \frac{\left(\mathcal{Z}_1'\mathcal{Z}_2\right)\left(\gamma_0'\mathcal{Z}_1\right)}{\mathcal{Z}_1'\mathcal{Z}_1},$$

which establishes that the estimator in large samples has a mixture of normals rather than a normal distribution.

Appendix

A Proof of Lemma 1

We are interested in the asymptotics of

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i, \widehat{m}\left(X_i\right)\right) = \frac{1}{n}\sum_{i=1}^{n} h\left(X_i, \widehat{G}\left(X_i; \widehat{\alpha}, \widehat{\alpha}\right)\right)$$

We use the Newey (1994) path-derivative method. Let W denote the vector of observed variables. The corresponding parametric submodel is indexed by the scalar parameter θ and so is the conditional mean of Y given $X = x, V = v$ denoted by $\mu(x, v; \theta)$. The parameter of interest is

$$\beta = \mathbb{E}\left[h\left(X, m\left(X\right)\right)\right] = \mathbb{E}\left[h\left(X, \mathbb{E}_V\left[\mu\left(X, V\right)\right]\right)\right] = \int h\left(x, \left(\int \mu\left(x, v\right) f\left(v\right) dv\right)\right) f\left(x\right) dx \tag{18}$$

where f is generic notation for a PDF. [We can express (18) further as

$$\beta = \int h\left(x, \left(\int \mu\left(x, \varphi(\tilde{x}, \tilde{z}; \alpha_*)\right) f\left(\tilde{x}, \tilde{z}\right) d\tilde{x} d\tilde{z}\right)\right) f\left(x\right) dx$$

For the path-derivative we consider the parameter of interest for a path in the distribution of W indexed by θ and the control variable indexed by α_1, α_2 (with the distinction made for accounting purposes)

$$\beta(\theta, \alpha_1, \alpha_2) = \int h\left(x, \left(\int\int \gamma\left(x, \varphi(\tilde{x}, \tilde{z}; \alpha_1); \alpha_2, \theta\right)\right) f\left(\tilde{x}, \tilde{z}; \theta\right) d\tilde{x} d\tilde{z}\right)\right) f\left(x; \theta\right) dx \tag{19}$$

Because we calculate total derivatives so that we deal with θ and α separately, we also define $\beta(\theta) = \beta(\theta, \alpha_*, \alpha_*)$. The path derivative is a total derivative with respect to $\theta, \alpha_1, \alpha_2$ where we distinguish the two roles of α. In the total derivative we consider e.g. first de derivative with respect to θ in $f(x; \theta)$ and evaluate this derivative at $\theta = 0$ which gives the population densities. This implies that we can initially set $\alpha_1 = \alpha_2 = \alpha_*$ and consider the path derivative for (19) and deal with the derivatives with respect to α_1, α_2 where we set $\theta = 0$. This term by term approach corresponds to the usual procedure in stochastic expansions. In the first part of the derivation the observed variables in W are Y, X, V.

The influence function of the estimator is the function $k(w)$ that satisfies

$$\frac{\partial \beta}{\partial \theta}\left(0\right) = \mathbb{E}[s(W)k(W)]$$

with $s(w) = \frac{\partial \ln f}{\partial \theta}(w; \theta)\Big|_{\theta=0}$. Therefore we we start by totally differentiating $\beta(\theta)$ and evaluating the derivative for $\theta = 0$

$$\begin{aligned} \frac{\partial \beta}{\partial \theta}\left(0\right) = &\int f\left(x\right) s\left(x\right) h\left(x, \left(\int f\left(v\right) \mu\left(x, v\right) dv\right)\right) dx \\ &+ \int f\left(x\right) \frac{\partial h\left(x, m\left(x\right)\right)}{\partial m(x)} \left(\int f\left(v\right) s\left(v\right) \mu\left(x, v\right) dv\right) dx \\ &+ \int f\left(x\right) \frac{\partial h\left(x, m\left(x\right)\right)}{\partial m(x)} \left(\int f\left(v\right) \frac{\partial \mu\left(x, v; 0\right)}{\partial \theta} dv\right) dx \end{aligned} \tag{20}$$

where s is the generic notation for the scores of the corresponding densities.

The first term in (20) is equal to

$$\begin{aligned}\int f(x)\, s(x)\, h(x, m(x))\, dx &= \int f(w)\,(s(w) - s(w|\, x))\, h(x, m(x))\, dw \\ &= \int f(w)\, s(w)\, h(x, m(x))\, dw \\ &- \int\int\int f(x)\, f(y, v|\, x)\, s(y, z|\, x)\, h(x, m(x))\, dxdydv \\ &= \int f(w)\, s(w)\, h(x, m(x))\, dw \\ &= \mathbb{E}\left[s(W) \cdot h(X, m(X))\right]\end{aligned}$$

The third equality is based on the observation

$$\begin{aligned}&\int\int\int f(x)\, f(y, v|\, x)\, s(y, v|\, x)\, h(x, m(x))\, dxdydz \\ &= \int f(x)\, h(x, m(x)) \left(\int f(y, z|\, x)\, s(y, z|\, x)\, dydz\right) dx = 0\end{aligned}$$

by the mean zero property of the score. The same argument can be used in general to write an integral of a function that depends on s subvector of w with respect to the density times the score of that subvector as an integral with respect to the product of the density times score of the distribution of W.

By Newey (1994, Theorem 2.1), we conclude that the contribution of the first term to the influence function is

$$h(X_i, M(X_i)) - \mathbb{E}\left[h(X, M(X))\right] \tag{21}$$

The second term in (20) is equal to

$$\int f(x)\, \tau(x) \left(\int f(v)\, s(v)\, \mu(x, v)\, dv\right) dx = \int f(v)\, s(v) \left(\int f(x)\, \tau(x)\, \mu(x, v)\, dx\right) dv$$

where we write the derivative of h with respect to its argument $m(x)$ as

$$\tau(x) = \frac{\partial h(x, m(x))}{\partial m(x)}$$

to simplify the notation. Using the same argument as for the first term

$$\begin{aligned}\int f(v)\, s(v) \left(\int f(x)\, \tau(x)\, \mu(x, v)\, dx\right) dv &= \int f(w)\, s(w) \left(\int f(x)\, \tau(x)\, \mu(x, v)\, dx\right) dw \\ &= \mathbb{E}\left[s(W) \cdot \mathbb{E}_X\left[\tau(X)\, \mu(X, V)\right]\right]\end{aligned}$$

and by Newey (1994, Theorem 2.1) the contribution is

$$\mathbb{E}_X\left[\tau(X)\, \mu(X, V_i)\right] - \mathbb{E}\left[\tau(X)\, m(X)\right] \tag{22}$$

The third term in (20) is equal to

$$\int f(x)\tau(x)\left(\int f(v)\frac{\partial\mu(x,v;\theta)}{\partial\theta}dv\right)dx = \int\int f(x,v)\frac{f(x)f(v)}{f(x,v)}\tau(x)\frac{\partial\mu(x,v;\theta)}{\partial\theta}dvdx$$
$$= \mathbb{E}\left\{s(W)\cdot\frac{f(X)f(V)}{f(X,V)}\tau(X)(Y-\mu(X,V))\right\}$$

by Newey (1994, p. 1361) with $\delta(x,v) = \frac{f(x)f(v)}{f(x,v)}\tau(x)$. This contributes

$$\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\tau(X_i)(Y_i-\mu(X_i,V_i)) \tag{23}$$

to the asymptotically linear representation.

Finally we consider the derivative with respect to α (or α_1 and α_2 to get the total derivative) of

$$\mathbb{E}\left[h\left(\widetilde{X},\mathbb{E}_{X,Z}\left[\gamma\left(\widetilde{X},\varphi(X,Z;\alpha_1);\alpha_2\right)\right]\right)\right]$$

where $\tilde{X}$ denotes a random variable that is independent of but has the same distribution as X. This is

$$\mathbb{E}\left[\frac{\partial h\left(\widetilde{X},m\left(\widetilde{X}\right)\right)}{\partial m(\widetilde{X})}\mathbb{E}_{X,Z}\left[\frac{\partial\gamma\left(\widetilde{X},\varphi(X,Z;\alpha_*);\alpha_*\right)}{\partial\alpha_1}\right]\right]$$
$$+\mathbb{E}\left[\frac{\partial h\left(\widetilde{X},m\left(\widetilde{X}\right)\right)}{\partial m(\widetilde{X})}\mathbb{E}_{X,Z}\left[\frac{\partial\gamma\left(\widetilde{X},\varphi(X,Z;\alpha_*);\alpha_*\right)}{\partial\alpha_2}\right]\right]$$
$$=\mathbb{E}\left[\tau\left(\widetilde{X}\right)\mathbb{E}_{X,Z}\left[\frac{\partial\mu\left(\widetilde{X},\varphi(X,Z;\alpha_*)\right)}{\partial v}\frac{\partial\varphi(X,Z;\alpha_*)}{\partial\alpha}\right]\right]$$
$$+\mathbb{E}\left[\tau\left(\widetilde{X}\right)\mathbb{E}_V\left[\frac{\partial\gamma\left(\widetilde{X},V;\alpha_*\right)}{\partial\alpha_2}\right]\right]$$

To summarize, our objective is to analyze the derivative

$$\mathbb{E}\left[\tau\left(\widetilde{X}\right)\mathbb{E}_{X,Z}\left[\frac{\partial\mu\left(\widetilde{X},\varphi(X,Z;\alpha_*)\right)}{\partial v}\frac{\partial\varphi(X,Z;\alpha_*)}{\partial\alpha}\right]\right]+\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\tau(X)\frac{\partial\gamma(X,V;\alpha_*)}{\partial\alpha_2}\right]. \tag{24}$$

The first term gives the usual linear approximation coefficient. The second term is accounts for the effect of α on the conditional expectation. It can be handled as in Hahn and Ridder (2013). Because $\gamma(x,\varphi(x,z,\alpha);\alpha)$ is the solution to the projection

$$\min_p \mathbb{E}\left[(Y-p(X,\varphi(X,Z,\alpha);\alpha))^2\right]$$

we have that for all α,

$$\mathbb{E}\left[t(X,\varphi(X,Z,\alpha))(Y-\gamma(X,\varphi(X,Z,\alpha);\alpha))\right]=0$$

for all possible functions t of $X, \varphi(X, Z, \alpha)$. In particular, we have that for all α

$$0 = \mathbb{E}\left[\frac{f(X) f(\varphi(X, Z, \alpha))}{f(X, \varphi(X, Z, \alpha))}\tau(X)(Y - \gamma(X, \varphi(X, Z, \alpha); \alpha))\right]$$

Differentiating with respect to α and evaluating the result at $\alpha = \alpha_*$, we find

$$\begin{aligned}
0 &= \mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{f(X) f(V)}{f(X, V)}\right)\frac{\partial \varphi(X, Z, \alpha_*)}{\partial \alpha}\tau(X)(Y - \mu(X, V))\right] \\
&- \mathbb{E}\left[\frac{f(X) f(V)}{f(X, V)}\tau(X)\frac{\partial \mu(X, V)}{\partial v}\frac{\partial \varphi(X, Z, \alpha_*)}{\partial \alpha}\right] \\
&- \mathbb{E}\left[\frac{f(X) f(V)}{f(X, V)}\tau(X)\frac{\partial \gamma(X, V; \alpha_*)}{\partial \alpha_2}\right]
\end{aligned}$$

with $V = \varphi(X, Z, \alpha_*)$ from which we obtain

$$\begin{aligned}
\mathbb{E}\left[\frac{f(X) f(V)}{f(X, V)}\tau(X)\frac{\partial \gamma(X, V; \alpha_*)}{\partial \alpha_2}\right] &= \mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{f(X) f(V)}{f(X, V)}\right)\tau(X)(Y - \mu(X, V))\frac{\partial \varphi(X, Z, \alpha_*)}{\partial \alpha}\right] \\
&- \mathbb{E}\left[\frac{f(X) f(V)}{f(X, V)}\tau(X)\frac{\partial \mu(X, V)}{\partial v}\frac{\partial \varphi(X, Z, \alpha_*)}{\partial \alpha}\right] \qquad (25)
\end{aligned}$$

Combining (24) and (25), we obtain the adjustment equal to

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_{\tilde{X}}\left[\tau\left(\tilde{X}\right)\frac{\partial \mu\left(\tilde{X}, V\right)}{\partial v}\right]\frac{\partial \varphi(X, Z; \alpha_*)}{\partial \alpha}\right] \\
&+ \mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{f(X) f(V)}{f(X, V)}\right)\tau(X)(Y - \mu(X, V))\frac{\partial \varphi(X, Z, \alpha_*)}{\partial \alpha}\right] \\
&- \mathbb{E}\left[\frac{f(X) f(V)}{f(X, V)}\tau(X)\frac{\partial \mu(X, V)}{\partial v}\frac{\partial \varphi(X, Z, \alpha_*)}{\partial \alpha}\right] \qquad (26)
\end{aligned}$$

times $\sqrt{n}(\widehat{\alpha} - \alpha_*)$. In this expression the expectation is with respect to the joint distribution of Y, X, Z.

B Proof of Lemma 2

Let $\varphi(x, z) = F(x|z)$ and note that

$$\varphi(x, z) = \int 1(\tilde{x} \leq x) f(\tilde{x}|z) d\tilde{x}$$

where $f(x|z)$ denotes the conditional density of X given $Z = z$. We would like to find an analog of Newey (1994, equation 3.9), i.e., we would like to find $k(W)$ such that

$$\frac{\partial}{\partial \theta}\mathbb{E}[h(W, \varphi(X, Z))] = \mathbb{E}[k(W) s(W)]$$

The parameter of interest is

$$\beta = \mathbb{E}[h(W, \varphi(X, Z))]$$

so that we consider

$$\beta(\theta)=\int h\left(w,\varphi(x,z;\theta)\right)f(w;\theta)dw$$

The total derivative evaluated for $\theta=0$ is

$$\frac{\partial\beta}{\partial\theta}(0)=\int h(w,\varphi(x,z))s(w)f(w)dw+\int\frac{\partial h}{\partial v}(w,\varphi(x,z))\frac{\partial\varphi(x,z)}{\partial\theta}f(w)dw$$

where $s(W)$ denotes the score of the distribution of W. For the second term on the right hand side we define

$$\tau(x,z)=\mathbb{E}\left[\frac{\partial h}{\partial v}(W,\varphi(X,Z))\middle| X=x,Z=z\right]$$

so that the second term is

$$\int\int\tau(x,z)\frac{\partial\varphi(x,z)}{\partial\theta}f(x,z)dxdz$$

The path for the control variable is

$$\varphi(x,z;\theta)=\int 1\left(\tilde{x}\leq x\right)f\left(\tilde{x}|\,z;\theta\right)d\tilde{x}$$

so that[9]

$$\frac{\partial\varphi(x,z;0)}{\partial\theta}=\int\int 1\left(\tilde{x}\leq x\right)s(\tilde{x}|z)f(\tilde{x}|z)d\tilde{x}$$

Substitution and changing the order of integration gives

$$\int\int\tau(x,z)\frac{\partial\varphi(x,z)}{\partial\theta}f(x,z)dxdz=\int\int\left(\int\tau\left(x,z\right)1\left(\tilde{x}\leq x\right)f\left(x|\,z\right)dx\right)s\left(\tilde{x}|\,z\right)f\left(\tilde{x}|\,z\right)f\left(z\right)d\tilde{x}dz$$

Define

$$\delta\left(\tilde{x},z\right)=\left(\int\tau\left(x,z\right)1\left(\tilde{x}\leq x\right)f\left(x|\,z\right)dx\right)=\mathbb{E}\left[\tau\left(X,Z\right)1\left(\tilde{x}\leq X\right)|\,Z=z\right] \tag{27}$$

so that (replace $\tilde{x}$ by x)

$$\mathbb{E}\left[\frac{\partial h\left(W,\varphi(X,Z)\right)}{\partial v}\frac{\partial\varphi\left(X,Z\right)}{\partial\theta}\right]=\int\int\delta\left(x,z\right)s\left(x|\,z\right)f\left(x|\,z\right)f\left(z\right)dxdz$$

so that because $s(x|z)=s(x,z)-s(z)$

$$\mathbb{E}\left[\frac{\partial h\left(W,\varphi(X,Z)\right)}{\partial v}\frac{\partial\varphi\left(X,Z\right)}{\partial\theta}\right]=\int\int\delta(x,z)s(x,z)f(x,z)dxdz-\int\left[\int\delta(x,z)f(x|z)dx\right]s(z)f(z)dz$$

We conclude that if we define

$$\overline{\delta}(z)=\mathbb{E}\left[\delta(X,Z)|Z=z\right]$$

we have

$$\mathbb{E}\left[\frac{\partial h\left(W,\varphi(X,Z)\right)}{\partial v}\frac{\partial\varphi\left(X,Z\right)}{\partial\theta}\right]=\mathbb{E}[(\delta(X,Z)-\overline{\delta}(Z))s(W)]$$

[9]If we alternatively define the path as $f(x|z;\theta)=\frac{f(x,z;\theta)}{f(z;\theta)}$ the total derivative is $\int 1(\tilde{x}\leq x)\left(\frac{1}{f(\tilde{x},z)}\frac{\partial f}{\partial\theta}(\tilde{x},z)-\frac{f(\tilde{x}|z)}{f(z)}\frac{\partial f}{\partial\theta}(z)\right)d\tilde{x}$. The expression between parentheses is equal to $(s(\tilde{x},z)-s(z))f(\tilde{x}|z)$ so that because $s(\tilde{x},z)-s(z)=s(\tilde{x}|z)$ we find the same result.

where we use that X, Z are components of the random vector W. Therefore

$$k(w) = h(w, \varphi(x, z)) + \delta(x, z) - \overline{\delta}(z)$$

with

$$\delta(x, z) = \mathbb{E}\left[\frac{\partial h\left(W, F(X|Z)\right)}{\partial v} 1\left(x \leq X\right) \middle| Z = z\right]$$

and

$$\overline{\delta}(z) = \mathbb{E}\left[\frac{\partial h\left(W, F(X|Z)\right)}{\partial v} F(X|Z) \middle| Z = z\right]$$

C Proof of Lemma 3

The result follows from (26) that defines $\xi(W)$ as the function of W such that (26) is equal to $\mathbb{E}\left[\xi(W)\frac{\partial\varphi(X,Z)}{\partial\theta}\right]$ where we replace α by θ in the nonparametric first stage. In Lemma 2 we then replace $\frac{\partial h(W,F(X|Z))}{\partial v}$ by $\xi(W)$.

C.1 Alternate Proof of Lemma 3

We use the path-derivative method. The parameter of interest is

$$\beta = \mathbb{E}\left[h\left(X, m\left(X\right)\right)\right] = \mathbb{E}\left[h\left(X, \mathbb{E}_V\left[\mu\left(X, V\right)\right]\right)\right] = \int h\left(x, \left(\int \mu\left(x, v\right) f\left(v\right) dv\right)\right) f\left(x\right) dx$$
$$= \int h\left(\tilde{x}, \left(\int\int \mu\left(\tilde{x}, \varphi\left(x, z\right)\right) f\left(x, z\right) dxdz\right)\right) f\left(\tilde{x}\right) d\tilde{x}$$

where

$$v = \varphi\left(x, z\right) = \int 1\left(\tilde{x} \leq x\right) f\left(\tilde{x}\middle| z\right) d\tilde{x}$$

For the path-derivative we consider the parameter of interest for a path in the distribution of W indexed by θ:

$$\beta\left(\theta\right) = \int h\left(\tilde{x}, \left(\int\int \mu\left(\tilde{x}, \varphi\left(x, z; \theta\right); \theta\right) f\left(x, z; \theta\right) dxdz\right)\right) f\left(\tilde{x}; \theta\right) d\tilde{x}$$

Note that

$$\begin{aligned}
\frac{\partial\beta}{\partial\theta}\left(0\right) &= \int f\left(\tilde{x}\right) s\left(\tilde{x}\right) h\left(\tilde{x}, \left(\int f\left(v\right)\mu\left(\tilde{x}, v\right) dv\right)\right) d\tilde{x} \\
&\quad + \int f\left(\tilde{x}\right) \frac{\partial h\left(\tilde{x}, m\left(\tilde{x}\right)\right)}{\partial m(\tilde{x})} \left(\int f\left(v\right) s\left(v\right) \mu\left(\tilde{x}, v\right) dv\right) d\tilde{x} \\
&\quad + \int f\left(\tilde{x}\right) \frac{\partial h\left(\tilde{x}, m\left(\tilde{x}\right)\right)}{\partial m(\tilde{x})} \left(\int\int f\left(x, z\right) \frac{\partial\mu\left(\tilde{x}, \varphi\left(x, z; 0\right); 0\right)}{\partial\theta} dxdz\right) d\tilde{x} \\
&\quad + \int f\left(\tilde{x}\right) \frac{\partial h\left(\tilde{x}, m\left(\tilde{x}\right)\right)}{\partial m(\tilde{x})} \left(\int\int f\left(x, z\right) \frac{\partial\mu\left(\tilde{x}, v\right)}{\partial v} \frac{\partial\varphi\left(x, z; 0\right)}{\partial\theta} dxdz\right) d\tilde{x}
\end{aligned}$$

where s is the generic notation for the scores of the corresponding densities.

As in the proof of Lemma 1, the first and second terms are

$$\int f(\tilde{x})\, s(\tilde{x})\, h\left(\tilde{x}, \left(\int f(v)\, \mu(\tilde{x}, v)\, dv\right)\right) d\tilde{x} = \int f(x)\, s(x)\, h(x, m(x))\, dx$$
$$= \mathbb{E}\left[s(W) \cdot h(X, m(X))\right] \tag{28}$$

and

$$\int f(\tilde{x}) \frac{\partial h(\tilde{x}, m(\tilde{x}))}{\partial m(\tilde{x})} \left(\int f(v)\, s(v)\, \mu(\tilde{x}, v)\, dv\right) d\tilde{x} = \int f(x)\, \tau(x) \left(\int f(v)\, s(v)\, \mu(x, v)\, dv\right) dx$$
$$= \mathbb{E}\left[s(W) \cdot \mathbb{E}_X\left[\tau(X)\, \mu(X, V)\right]\right] \tag{29}$$

The third term

$$\begin{aligned}
&\int f(\tilde{x}) \frac{\partial h(\tilde{x}, m(\tilde{x}))}{\partial m(\tilde{x})} \left(\int\int f(x, z) \frac{\partial \mu(\tilde{x}, \varphi(x, z; 0); 0)}{\partial \theta} dx dz\right) d\tilde{x} \\
&= \int f(\tilde{x})\, \tau(\tilde{x}) \left(\int f(v) \frac{\partial \mu(\tilde{x}, v; 0)}{\partial \theta} dv\right) d\tilde{x} \\
&= \int\int f(x, v) \frac{f(x)\, f(v)}{f(x, v)} \tau(x) \frac{\partial \mu(x, v; 0)}{\partial \theta} dv dx \\
&= \mathbb{E}\left[\frac{f(X)\, f(V)}{f(X, V)} \tau(X) \frac{\partial \mu(X, V; 0)}{\partial \theta}\right]
\end{aligned}$$

is a little tricky, because the path that we take it is not totally arbitrary; it should only go through the conditional CDF. For a given $\rho(x, \varphi(x, z; \theta))$, we have

$$0 = \int\int \rho(x, \varphi(x, z; \theta)) (y - \mu(x, \varphi(x, z; \theta); \theta))\, f(w; \theta)\, dx dz$$

Differentiating with respect to θ, we obtain

$$\begin{aligned}
0 = &\int \frac{\partial \rho(x, \varphi(x, z; 0))}{\partial v} \frac{\partial \varphi(x, z; 0)}{\partial \theta} (y - \mu(x, \varphi(x, z; 0); 0))\, f(w)\, dw \\
&- \int \rho(x, \varphi(x, z; 0)) \frac{\partial \mu(x, \varphi(x, z; 0); 0)}{\partial v} \frac{\partial \varphi(x, z; 0)}{\partial \theta} f(w)\, dw \\
&- \int \rho(x, \varphi(x, z; 0)) \frac{\partial \mu(x, \varphi(x, z; 0); 0)}{\partial \theta} f(w)\, dw \\
&+ \int \rho(x, \varphi(x, z; 0)) (y - \mu(x, \varphi(x, z; 0)))\, s(w)\, f(w)\, dw
\end{aligned}$$

Letting

$$\rho(x, v) = \frac{f(x)\, f(v)}{f(x, v)} \tau(x)$$

we have the third term equal to

$$
\begin{aligned}
&\mathbb{E}\left[\frac{f(X)\,f(V)}{f(X,V)}\tau(X)\,\frac{\partial\mu(X,V;0)}{\partial\theta}\right] \\
&=\int\frac{\partial}{\partial v}\left(\frac{f(x)\,f(v)}{f(x,v)}\right)\tau(x)\,(y-\mu(x,v))\,\frac{\partial\varphi(x,z;0)}{\partial\theta}f(w)\,dw \\
&-\int\frac{f(x)\,f(v)}{f(x,v)}\tau(x)\,\frac{\partial\mu(x,v)}{\partial v}\frac{\partial\varphi(x,z;0)}{\partial\theta}f(w)\,dw \\
&+\int\frac{f(x)\,f(v)}{f(x,v)}\tau(x)\,(y-\mu(x,v))\,s(w)\,f(w)\,dw
\end{aligned}
$$

Because the fourth term is equal to

$$
\begin{aligned}
&\int f(\tilde{x})\,\frac{\partial h(\tilde{x},m(\tilde{x}))}{\partial m(\tilde{x})}\left(\int\int f(x,z)\,\frac{\partial\mu(\tilde{x},v)}{\partial v}\frac{\partial\varphi(x,z;0)}{\partial\theta}dxdz\right)d\tilde{x} \\
&=\int f(\tilde{x})\,\tau(\tilde{x})\left(\int\int f(x,z)\,\frac{\partial\mu(\tilde{x},v)}{\partial v}\frac{\partial\varphi(x,z;0)}{\partial\theta}dxdz\right)d\tilde{x}
\end{aligned}
$$

we can write the sum of the third and fourth term as

$$
\begin{aligned}
&\int\frac{f(x)\,f(v)}{f(x,v)}\tau(x)\,(y-\mu(x,v))\,s(w)\,f(w)\,dw+\int\xi(w)\,\frac{\partial\varphi(x,z;0)}{\partial\theta}f(w)\,dw \\
&=\mathbb{E}\left[s(W)\cdot\frac{f(X)\,f(V)}{f(X,V)}\tau(X)\,(Y-\mu(X,V))\right] \\
&+\int\xi(w)\,\frac{\partial\varphi(x,z;0)}{\partial\theta}f(w)\,dw
\end{aligned}
\tag{30}
$$

where

$$
\begin{aligned}
\xi(w)&=\int f(\tilde{x})\,\tau(\tilde{x})\,\frac{\partial\mu(\tilde{x},v)}{\partial v}d\tilde{x}+\frac{\partial}{\partial v}\left(\frac{f(x)\,f(v)}{f(x,v)}\right)\tau(x)\,(y-\mu(x,v)) \\
&-\frac{f(x)\,f(v)}{f(x,v)}\tau(x)\,\frac{\partial\mu(x,v)}{\partial v} \\
&=\mathbb{E}_{\tilde{X}}\left[\tau\left(\tilde{X}\right)\frac{\partial\mu\left(\tilde{X},v\right)}{\partial v}\right]+\frac{\partial}{\partial v}\left(\frac{f(x)\,f(v)}{f(x,v)}\right)\tau(x)\,(y-\mu(x,v)) \\
&-\frac{f(x)\,f(v)}{f(x,v)}\tau(x)\,\frac{\partial\mu(x,v)}{\partial v}
\end{aligned}
$$

Using

$$
\frac{\partial\varphi(x,z;0)}{\partial\theta}=\int 1\,(\tilde{x}\leq x)\,s(\tilde{x}|z)f(\tilde{x}|z)d\tilde{x}
$$

which was established in the proof of Lemma 2, we can write

$$
\begin{aligned}
\int \xi(w) \frac{\partial \varphi(x,z;0)}{\partial \theta} f(w)\, dw &= \int\int \tau(x,z) \frac{\partial \varphi(x,z;0)}{\partial \theta} f(x,z)\, dxdz \\
&= \int\int \left(\int \tau(x,z)\, 1(\tilde{x} \leq x)\, f(x|\, z)\, dx \right) s(\tilde{x}|\, z)\, f(\tilde{x}|\, z)\, f(z)\, d\tilde{x}dz \\
&= \int\int \delta(x,z)\, s(x|\, z)\, f(x|\, z)\, f(z)\, dxdz \\
&= \int\int \delta(x,z) s(x,z) f(x,z) dxdz - \int \left[\int \delta(x,z) f(x|z) dx \right] s(z) f(z) dz \\
&= \int\int \left(\delta(x,z) - \overline{\delta}(z) \right) s(x,z) f(x,z) dxdz \\
&= \mathbb{E}\left[s(W) \cdot \left(\delta(X,Z) - \overline{\delta}(Z) \right) \right]
\end{aligned}
$$

where

$$
\begin{aligned}
\tau(X,Z) &= \mathbb{E}[\xi(W)|\, X, Z] \\
\delta(\tilde{x},z) &= \int \tau(x,z)\, 1(\tilde{x} \leq x)\, f(x|\, z)\, dx = \mathbb{E}[\tau(X,Z)\, 1(\tilde{x} \leq X)|\, Z = z] \\
&= \mathbb{E}[\xi(W)\, 1(\tilde{x} \leq X)|\, Z = z] \\
\overline{\delta}(z) &= \int \delta(\tilde{x},z) f(\tilde{x}|z) d\tilde{x} = \int\int \tau(x,z)\, 1(\tilde{x} \leq x)\, f(\tilde{x}|z) f(x|\, z)\, d\tilde{x}dx \\
&= \int \tau(x,z)\, F(x|\, z)\, f(x|\, z)\, d\tilde{x}dx = \mathbb{E}[\tau(X,Z)\, F(X|\, Z)|\, Z = z] \\
&= \mathbb{E}[\xi(W)\, F(X|\, Z)|\, Z = z]
\end{aligned}
$$

It follows that the sum of the third and fourth terms is we can write the sum of the third and fourth term as

$$
\mathbb{E}\left[s(W) \cdot \frac{f(X) f(V)}{f(X,V)} \tau(X) (Y - \mu(X,V)) \right] + \mathbb{E}\left[s(W) \cdot \left(\delta(X,Z) - \overline{\delta}(Z) \right) \right] \tag{31}
$$

Combining (28), (29), and (31), we obtain the conclusion.

D Proof of Lemma 4

This proof follows the method in the proof of Lemma 1. Note that

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial M(X)}{\partial x} \right] &= \int f(x) \frac{\partial}{\partial x} \left(\int f(v)\, \mu(x,v)\, dv \right) dx \\
&= \int f(x) \left(\int f(v) \frac{\partial \mu(x,v)}{\partial x} dv \right) dx \\
&= \int\int f(v)\, f(x) \frac{\partial \mu(x,v)}{\partial x} dxdv
\end{aligned}
$$

and define

$$\beta(\theta,\alpha_1,\alpha_2)=\int f(\tilde{x};\theta)\left(\int\int f(x,z;\theta)\frac{\partial\gamma(\tilde{x},\varphi(x,z,\alpha_1);\alpha_2,\theta)}{\partial x}dxdz\right)d\tilde{x};$$

As before we define $\beta(\theta)=\beta(\theta,\alpha_*,\alpha_*)$ with

$$\beta(\theta)=\int f(\tilde{x};\theta)\left(\int\int f(x,z;\theta)\frac{\partial\mu(\tilde{x},\varphi(x,z,\alpha_*);\theta)}{\partial x}dxdz\right)d\tilde{x}$$

The total derivative of $\beta(\theta)$ at 0 is, if we replace $\tilde{x}$ by x and $\varphi(x,z,\alpha_*)$ by v,

$$\begin{aligned}\frac{\partial\beta(0)}{\partial\theta}&=\int f(x)s(x)\left(\int f(v)\frac{\partial\mu(x,v)}{\partial x}dv\right)dx\\&+\int f(x)\left(\int f(v)s(v)\frac{\partial\mu(x,v)}{\partial x}dv\right)dx\\&-\int\int f(v)\frac{\partial f(x)}{\partial x}\frac{\partial\mu(x,v)}{\partial\theta}dvdx\end{aligned}\tag{32}$$

where we used for the third term on the right-hand side that by partial integration

$$\int\int f(v)f(x)\frac{\partial\mu(x,v)}{\partial x}dxdv=-\int\int f(v)\frac{\partial f(x)}{\partial x}\mu(x,v)\,dxdv$$

if the density of X is 0 at the integration limits, i.e., the boundary of the support. Using the same argument as before we can replace $s(x)f(x)$ by $s(w)f(w)$ so that for the first term

$$\begin{aligned}\int f(x)s(x)\frac{\partial m(x)}{\partial x}dx&=\int f(w)s(w)\frac{\partial m(x)}{\partial x}dw\\&=\mathbb{E}\left[s(W)\cdot\frac{\partial m(X)}{\partial x}\right]\end{aligned}$$

and the contribution to the influence function is

$$\frac{\partial m(X_i)}{\partial x}-\mathbb{E}\left[\frac{\partial m(X)}{\partial x}\right]\tag{33}$$

The second term in (32) is equal to

$$\begin{aligned}\int f(v)s(v)\left(\int f(x)\frac{\partial\mu(x,v)}{\partial x}dx\right)dv&=\int f(w)s(w)\left(\int f(x)\frac{\partial\mu(x,v)}{\partial x}dx\right)dw\\&=\mathbb{E}\left[s(W)\cdot\mathbb{E}_X\left[\frac{\partial\mu(X,V)}{\partial x}\right]\right]\end{aligned}$$

where it is understood that the expectation $\mathbb{E}_X$ is with respect to the marginal distribution of X. This term contributes

$$\mathbb{E}_X\left[\frac{\partial\mu(X,V_i)}{\partial x}\right]-\mathbb{E}\left[\frac{\partial m(X)}{\partial x}\right]\tag{34}$$

to the influence function. The third term in (32) is equal to

$$\int\int f(x,v)\frac{\frac{\partial f(x)}{\partial x}f(v)}{f(x,v)}\frac{\partial\mu(x,v;\theta)}{\partial\theta}dvdx$$

so that by Newey (1994), Proposition 4, the contribution to the influence function is

$$\frac{\frac{\partial f(X_i)}{\partial x} f\left(V_i\right)}{f\left(X_i, V_i\right)}\left(Y_i-\mu\left(X_i, V_i\right)\right) \tag{35}$$

To find the contribution of the first stage we consider

$$\begin{aligned}
\beta\left(0, \alpha_1, \alpha_2\right) &= \int\int f(x, z)\left(\int f(\tilde{x}) \frac{\partial \gamma\left(\tilde{x}, \varphi(x, z, \alpha_1) ; \alpha_2, 0\right)}{\partial x} d \tilde{x}\right) d x d z \\
&= -\int\int f(x, z)\left(\int \frac{\partial f(\tilde{x})}{\partial x} \gamma\left(\tilde{x}, \varphi(x, z, \alpha_1) ; \alpha_2, 0\right) d \tilde{x}\right) d x d z \\
&= -\mathbb{E}_{X, Z}\left[\mathbb{E}_{\widetilde{X}}\left[\frac{\frac{\partial f(\widetilde{X})}{\partial x}}{f(\widetilde{X})} \gamma\left(\widetilde{X}, \varphi(X, Z, \alpha_1) ; \alpha_2, 0\right)\right]\right]
\end{aligned}$$

by partial integration. Hence the total derivative is (we use α instead of α_1)

$$-\mathbb{E}_{X, Z}\left[\mathbb{E}_{\widetilde{X}}\left[\frac{\frac{\partial f(\widetilde{X})}{\partial x}}{f\left(\widetilde{X}\right)} \frac{\partial \mu\left(\widetilde{X}, \varphi\left(X, Z ; \alpha_*\right)\right)}{\partial v} \frac{\partial \varphi\left(X, Z ; \alpha_*\right)}{\partial \alpha}\right]\right]-\mathbb{E}_{X, Z}\left[\mathbb{E}_{\widetilde{X}}\left[\frac{\frac{\partial f(\widetilde{X})}{\partial x}}{f(\widetilde{X})} \frac{\partial \gamma\left(\widetilde{X}, \varphi(X, Z, \alpha_*) ; \alpha_*, 0\right)}{\partial \alpha_2}\right]\right]. \tag{36}$$

The second term is equal to

$$-\mathbb{E}_V\left[\mathbb{E}_{\widetilde{X}}\left[\frac{\frac{\partial f(\widetilde{X})}{\partial x}}{f(\widetilde{X})} \frac{\partial \gamma\left(\widetilde{X}, V ; \alpha_*, 0\right)}{\partial \alpha_2}\right]\right]=-\mathbb{E}_{X, V}\left[\frac{\frac{\partial f(X)}{\partial x} f(V)}{f(X, V)} \frac{\partial \gamma\left(X, V ; \alpha_*, 0\right)}{\partial \alpha_2}\right]$$

Because $\gamma(x, \varphi(x, z, \alpha) ; \alpha)$ is the solution to

$$\min_p \mathbb{E}\left[\left(Y-p\left(X, \varphi\left(X, Z, \alpha\right) ; \alpha\right)\right)^2\right]$$

we have that for all α,

$$\mathbb{E}\left[t\left(X, \varphi\left(X, Z, \alpha\right) ; \alpha\right)\left(Y-\gamma\left(X, \varphi\left(X, Z, \alpha\right) ; \alpha\right)\right)\right]=0$$

for all functions $t\left(X, \varphi\left(X, Z, \alpha\right)\right)$. In particular, we have

$$0=\mathbb{E}\left[\frac{\frac{\partial f(X)}{\partial x} f\left(\varphi\left(X, Z, \alpha\right)\right)}{f\left(X, \varphi\left(X, Z, \alpha\right)\right)}\left(Y-\gamma(X, \varphi(X, Z, \alpha) ; \alpha)\right)\right]$$

Differentiating with respect to α and evaluating the result at $\alpha=\alpha_*$, we find

$$\begin{aligned}
0=\mathbb{E}&\left[\frac{\partial}{\partial v}\left(\frac{\frac{\partial f(X)}{\partial x} f\left(V\right)}{f\left(X, V\right)}\right) \frac{\partial \varphi\left(X, Z, \alpha_*\right)}{\partial \alpha}\left(Y-\mu\left(X, V\right)\right)\right] \\
&-\mathbb{E}\left[\frac{\frac{\partial f(X)}{\partial x} f\left(V\right)}{f\left(X, V\right)} \frac{\partial \mu\left(X, V\right)}{\partial v} \frac{\partial \varphi\left(X, Z, \alpha_*\right)}{\partial \alpha}\right] \\
&-\mathbb{E}\left[\frac{\frac{\partial f(X)}{\partial x} f\left(V\right)}{f\left(X, V\right)} \frac{\partial \gamma\left(X, V ; \alpha_*\right)}{\partial \alpha_2}\right]
\end{aligned}$$

from which we obtain

$$\mathbb{E}\left[\frac{\frac{\partial f(X)}{\partial x}f\left(V\right)}{f\left(X,V\right)}\frac{\partial\gamma\left(X,V;\alpha_*\right)}{\partial\alpha_2}\right]=\mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{\frac{\partial f(X)}{\partial x}f\left(V\right)}{f\left(X,V\right)}\right)\left(Y-\mu\left(X,V\right)\right)\frac{\partial\varphi\left(X,Z,\alpha_*\right)}{\partial\alpha}\right]$$
$$-\mathbb{E}\left[\frac{\frac{\partial f(X)}{\partial x}f\left(V\right)}{f\left(X,V\right)}\frac{\partial\mu\left(X,V\right)}{\partial v}\frac{\partial\varphi\left(X,Z,\alpha_*\right)}{\partial\alpha}\right] \tag{37}$$

Combining (36) and (37), we obtain the adjustment equal to

$$-\mathbb{E}\left[\mathbb{E}_{\tilde{X}}\left[\frac{\frac{\partial f(\tilde{X})}{\partial x}}{f\left(\tilde{X}\right)}\frac{\partial\mu\left(\tilde{X},V\right)}{\partial v}\right]\frac{\partial\varphi\left(X,Z;\alpha_*\right)}{\partial\alpha}\right]-\mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{\frac{\partial f(X)}{\partial x}f\left(V\right)}{f\left(X,V\right)}\right)\left(Y-\mu\left(X,V\right)\right)\frac{\partial\varphi\left(X,Z,\alpha_*\right)}{\partial\alpha}\right]+$$
$$+\mathbb{E}\left[\frac{\frac{\partial f(X)}{\partial x}f\left(V\right)}{f\left(X,V\right)}\frac{\partial\mu\left(X,V\right)}{\partial v}\frac{\partial\varphi\left(X,Z,\alpha_*\right)}{\partial\alpha}\right] \tag{38}$$

times $\sqrt{n}(\widehat{\alpha}-\alpha_*)$.

For the non-parametric first stage we follow the same proof as in Lemma 3

E Proof of Lemma 5

The parameter for a path indexed by $\theta, \alpha_1, \alpha_2$ is

$$\beta(\theta,\alpha_1,\alpha_2)=\int\int f(x,z;\theta)\frac{\partial\gamma}{\partial x}(x,\varphi(x,z,\alpha_1);\alpha_2,\theta)dxdz$$

and as before $\beta(\theta)=\beta(\theta,\alpha_*,\alpha_*)$. Then

$$\beta(\theta,\alpha_*,\alpha_*)=\int\int f\left(x,v;\theta\right)\frac{\partial\mu(x,v;\theta)}{\partial x}dxdv$$

and therefore,

$$\begin{aligned}\frac{\partial\beta}{\partial\theta}(0)&=\int\int f(x,v)s(x,v)\frac{\partial\mu}{\partial x}(x,v)dxdv+\int\int f(x,v)\frac{\partial}{\partial\theta}\left(\frac{\partial\mu(x,v)}{\partial x}\right)dxdv\\&=\int f(w)s(w)\frac{\partial\mu}{\partial x}(x,v)dv-\int\int\frac{\frac{\partial f(x,v)}{\partial x}}{f(x,v)}\frac{\partial\mu(x,v)}{\partial\theta}f(x,v)dxdv\end{aligned}$$

where in the second term we have used partial integration and the assumption on the density at the boundary of the support. Therefore

$$\frac{\partial\beta}{\partial\theta}(0)=\mathbb{E}\left[s(W)\frac{\partial\mu}{\partial x}(X,V)\right]-\mathbb{E}\left[\frac{\frac{\partial f(X,V)}{\partial x}}{f(X,V)}\frac{\partial\mu(X,V)}{\partial\theta}\right]$$

so that the contribution to the influence function is, using Proposition 4 in Newey (1994)

$$\frac{\partial\mu}{\partial x}(X_i,V_i)-\frac{\frac{\partial f(X_i,V_i)}{\partial x}}{f(X_i,V_i)}(Y_i-\mu(X_i,V_i))$$

Next we consider $\beta(0,\alpha_1,\alpha_2)$ and take the total derivative using

$$\beta(0,\alpha_1,\alpha_*) = \int\int f(x,z)\frac{\partial\mu}{\partial x}(x,\varphi(x,z,\alpha_1))dxdz = -\int\int \frac{\partial f}{\partial x}(x,z)\mu(x,\varphi(x,z,\alpha_1))dxdz$$

so that

$$\frac{\partial\beta}{\partial\alpha_1}(0,\alpha_*,\alpha_*) = -\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,Z)}{f(X,Z)}\frac{\partial\mu}{\partial v}(X,V)\frac{\partial\varphi(X,Z,\alpha_*)}{\partial\alpha}\right]$$

Also using partial integration

$$\beta(0,\alpha_*,\alpha_2) = \mathbb{E}\left[\frac{\partial\gamma}{\partial x}(X,V;\alpha_2)\right] = -\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,V)}{f(X,V)}\gamma(X,V;\alpha_2)\right]$$

so that

$$\frac{\partial\beta}{\partial\alpha_2}(0,\alpha_*,\alpha_*) = -\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,V)}{f(X,V)}\frac{\partial\gamma(X,V;\alpha_*)}{\partial\alpha_2}\right]$$

An analogous projection argument as in Lemma 4 but with

$$t(x,\varphi(x,z,\alpha);\alpha) = \frac{\frac{\partial f}{\partial x}(x,\varphi(x,z,\alpha))}{f(x,\varphi(x,z,\alpha))}$$

so that for all α

$$\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,\varphi(X,Z,\alpha))}{f(X,\varphi(X,Z,\alpha))}(Y-\gamma(X,\varphi(X,Z,\alpha);\alpha)\right] = 0$$

Taking the derivative with respect to α we find

$$\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,V)}{f(X,V)}\frac{\partial\gamma(X,V;\alpha_*)}{\partial\alpha_2}\right] = \mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{\frac{\partial f}{\partial x}(X,V)}{f(X,V)}\right)(Y-\mu(X,V))\frac{\partial\varphi(X,V,\alpha_*)}{\partial\alpha}\right] -$$

$$\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,V)}{f(X,V)}\frac{\partial\mu}{\partial v}(X,V)\frac{\partial\varphi(X,V,\alpha_*)}{\partial\alpha}\right]$$

The contribution of the first stage is therefore

$$-\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,Z)}{f(X,Z)}\frac{\partial\mu}{\partial v}(X,V)\frac{\partial\varphi(X,Z,\alpha_*)}{\partial\alpha}\right] - \mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{\frac{\partial f}{\partial x}(X,V)}{f(X,V)}\right)(Y-\mu(X,V))\frac{\partial\varphi(X,V,\alpha_*)}{\partial\alpha}\right]$$

$$+\mathbb{E}\left[\frac{\frac{\partial f}{\partial x}(X,V)}{f(X,V)}\frac{\partial\mu}{\partial v}(X,V)\frac{\partial\varphi(X,V,\alpha_*)}{\partial\alpha}\right]$$

times $\sqrt{n}(\widehat{\alpha}-\alpha_*)$.

The nonparametric first stage is dealt with as in the proof of Lemma 3.

F Proof of Lemma 6

Define

$$A=\left(\mathbb{E}\left[\frac{\partial r}{\partial\beta}(X;\beta_*)\frac{\partial r}{\partial\beta'}(X;\beta_*)\right]\right)^{-1}$$

and consider

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widehat{m}(X_i)-m(X_i)\right)\frac{\partial r}{\partial\beta}(X_i;\beta_*)$$

We apply Lemma 1 with

$$h(X,\widehat{m}(X))=(\widehat{m}(X)-m(X))\frac{\partial r}{\partial\beta}(X;\beta_*)$$

so that

$$\tau(X)=\frac{\partial r}{\partial\beta}(X;\beta_*)$$

The first term in the influence function of Lemma 1 is 0 and the second term is

$$\mathbb{E}_X\left[\mu(X,V_i)\frac{\partial r}{\partial\beta}(X;\beta_*)\right]-\mathbb{E}\left[m(X)\frac{\partial r}{\partial\beta}(X;\beta_*)\right]$$

The third term is

$$\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i-\mu(X_i,V_i))\frac{\partial r}{\partial\beta}(X_i;\beta_*)$$

and the contribution of the first-stage estimation error

$$\left\{\begin{array}{c}\mathbb{E}\left[\mathbb{E}_{\tilde{X}}\left[\frac{\partial r}{\partial\beta}(\widetilde{X};\beta_*)\frac{\partial\mu(\tilde{X},V)}{\partial v}\right]\frac{\partial\varphi(X,Z;\alpha_*)}{\partial\alpha'}\right]\\+\mathbb{E}\left[\frac{\partial}{\partial v}\left(\frac{f(X)f(V)}{f(X,V)}\right)\frac{\partial r}{\partial\beta}(X;\beta_*)\left(Y-\mu\left(X,V\right)\right)\frac{\partial\varphi(X,Z,\alpha_*)}{\partial\alpha'}\right]\\-\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial r}{\partial\beta}(X;\beta_*)\frac{\partial\mu(X,V)}{\partial v}\frac{\partial\varphi(X,Z,\alpha_*)}{\partial\alpha'}\right]\end{array}\right\}\sqrt{n}(\widehat{\alpha}-\alpha_*)$$

The derivation of the influence function for the nonparametric first stage is as in Lemma 3.

G Proof of Lemma 7

Lemma 8 *Let $p=E\left[D_i\right]$. We have*

$$\left(\begin{array}{c}\sqrt{n}\widehat{\alpha}\\\sqrt{n}(\overline{x}-\overline{x}_0)\end{array}\right)\overset{d}{\to}\left(\begin{array}{c}p^{-\frac{1}{2}}\mathcal{X}_1-(1-p)^{-\frac{1}{2}}\mathcal{X}_0\\\sqrt{p}\mathcal{X}_1-\frac{p}{\sqrt{1-p}}\mathcal{X}_0\end{array}\right)\equiv\left(\begin{array}{c}\mathcal{Z}_1\\\mathcal{Z}_2\end{array}\right)$$

with $\mathcal{X}_0,\mathcal{X}_1$ independent standard normal random vectors of dimension K.

Proof. We have

$$p=\lim_{n\to\infty}\frac{n_1}{n}$$

where n_1 denotes the size of the subsample such that $D_i=1$. By logit MLE and $\alpha_*=0$, we have

$$\sqrt{n}\left[\begin{array}{c}\widehat{\zeta}-\zeta_*\\\widehat{\alpha}\end{array}\right]=\left(p\left(1-p\right)\left[\begin{array}{cc}1&\mu'\\\mu&\mu\mu'+I_K\end{array}\right]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(D_i-p\right)\left[\begin{array}{c}1\\x_i\end{array}\right]+o_p\left(1\right)$$

with under the distributional assumptions above $I_\alpha(0) = p(1-p)I_K$, which implies that

$$\sqrt{n}\widehat{\alpha} = \frac{1}{p\,(1-p)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(D_i - p\right)\left(x_i - \mu\right) + o_p\left(1\right)$$

Therefore

$$\begin{aligned}\sqrt{n}\widehat{\alpha} &= \frac{1}{p(1-p)\sqrt{n}}\left(n_1(1-p)\left(\overline{x}_1 - \mu\right) - n_0 p\left(\overline{x}_0 - \mu\right)\right) + o_p\left(1\right)\\ &= p^{-\frac{1}{2}}\sqrt{n_1}\left(\overline{x}_1 - \mu\right) - (1-p)^{-\frac{1}{2}}\sqrt{n_0}\left(\overline{x}_0 - \mu\right)\end{aligned}$$

Also note that

$$\begin{aligned}\overline{x} - \overline{x}_0 &= \frac{1}{n_0 + n_1}\left(\sum_{D_i=0} x_i + \sum_{D_i=1} x_i\right) - \frac{1}{n_0}\sum_{D_i=0} x_i\\ &= \frac{n_1}{n_0 + n_1}\left(\overline{x}_1 - \overline{x}_0\right)\\ &= \frac{n_1}{n_0 + n_1}\left(\left(\overline{x}_1 - \mu\right) - \left(\overline{x}_0 - \mu\right)\right),\end{aligned}$$

and therefore

$$\frac{n_1}{n_0 + n_1}\frac{\sqrt{n}}{\sqrt{n_0}}\sqrt{n_0}\left(\overline{x}_1 - \mu\right)$$

$$\sqrt{n}\left(\overline{x} - \overline{x}_0\right) = \sqrt{p}\sqrt{n_1}\left(\overline{x}_1 - \mu\right) - \frac{p}{\sqrt{1-p}}\sqrt{n_0}\left(\overline{x}_0 - \mu\right) + o_p\left(1\right).$$

Stacking the expressions and applying the Central Limit Theorem gives the desired result. ■

Proof of Lemma 7. Note that with $\tilde{x}_i = x_i - \overline{x}_0$

$$\begin{aligned}\sqrt{n}\left(\left(\overline{x} - \overline{x}_0\right)'\widehat{\alpha}\right)\widehat{\delta}_0 &= \left(\left(\overline{x} - \overline{x}_0\right)'\widehat{\alpha}\right)\frac{\sum_{D_i=0}\left(\tilde{x}_i'\widehat{\alpha}\right)y_i}{\sum_{D_i=0}\left(\tilde{x}_i'\widehat{\alpha}\right)^2}\\ &= \sqrt{n}\widehat{\alpha}'\left(\overline{x} - \overline{x}_0\right)\frac{\left(\frac{1}{n_0}\sum_{D_i=0}\left(\tilde{x}_i'\widehat{\alpha}\right)y_i\right)}{\frac{1}{n_0}\sum_{D_i=0}\left(\tilde{x}_i'\widehat{\alpha}\right)^2}\\ &= \frac{\sqrt{n}\widehat{\alpha}'\sqrt{n}\left(\overline{x} - \overline{x}_0\right)\left(\frac{1}{n_0}\sum_{D_i=0} y_{0i}\tilde{x}_i\right)'\sqrt{n}\widehat{\alpha}}{\sqrt{n}\widehat{\alpha}'\left(\frac{1}{n_0}\sum_{D_i=0}\tilde{x}_i\tilde{x}_i'\right)\sqrt{n}\widehat{\alpha}}\end{aligned} \tag{39}$$

Now

$$\frac{1}{n_0}\sum_{D_i=0} y_{0i}\tilde{x}_i = \frac{1}{n_0}\sum_{D_i=0}\tilde{x}_i(\beta_0 + x_i'\gamma_0 + \varepsilon_{0i}) = \gamma_0 + o_p(1)$$

Therefore

$$\sqrt{n}\left(\left(\overline{x} - \overline{x}_0\right)'\widehat{\alpha}\right)\widehat{\delta}_0 = \frac{\left(\sqrt{n}\widehat{\alpha}\right)'\left(\sqrt{n}\left(\overline{x} - \overline{x}_0\right)\right)\gamma'\left(\sqrt{n}\widehat{\alpha}\right)}{\left(\sqrt{n}\widehat{\alpha}\right)'\left(\sqrt{n}\widehat{\alpha}\right)} + o_p\left(1\right)$$

and by Lemma 1 and the continuous-mapping theorem

$$\sqrt{n}\left(\left(\overline{x} - \overline{x}_0\right)'\widehat{\alpha}\right)\widehat{\delta}_0 = \frac{\left(\mathcal{Z}_1'\mathcal{Z}_2\right)\left(\gamma_0'\mathcal{Z}_1\right)}{\mathcal{Z}_1'\mathcal{Z}_1} + o_p\left(1\right)$$

■

Derivation of the distribution of the semi-parametric CV estimator by stochastic expansion

H Motivating example

This appendix derives the asymptotic distribution of the control variable estimator by asymptotic expansion. Although the derivation is general, it is helpful to consider as an example a logit model with a single continuous endogenous regressor

$$Y^* = \tau X - \varepsilon$$

where ε has a logistic distribution and X and ε are correlated. If $\tilde{X}$ and ε are independent then

$$\mathbb{E}[Y|\tilde{X} = x] = \frac{e^{\tau x}}{1 + e^{\tau x}} \equiv R(x;\tau) \tag{40}$$

We observe

$$Y = I(Y^* \geq 0)$$

Let Z be an instrument.The general first stage is

$$X = F^{-1}\left(V|\, Z\right)$$

which is more general than a linear first-stage

$$X = \gamma_0 + \gamma_1 Z + V$$

We assume

$$Z \perp V, \varepsilon$$

This assumption is not automatically satisfied in the linear first stage.

The Average Structural Function (ASF) is

$$L\left(x\right) = \int_0^1 \mathbb{E}\left[Y|\, X = x, V = v\right] \mathrm{d}v$$

This requires that V has $[0,1]$ support, i.e., in

$$V = F\left(x|\, Z\right)$$

V takes all values in $[0,1]$. For some x the control variable V may take values in a subset of $[0,1]$, but we rule out such a possibility.

The estimator of the ASF is

$$\hat{L}\left(x\right) = \frac{1}{n}\sum_{j=1}^{n} \hat{\mu}\left(x, \hat{V}_j\right)$$

with $\hat{V}_j$ denoting our estimator of $F\left(X_j\right|Z_j)$, and $\hat{\mu}(x,v)$ the nonparametric series estimator of $\mu\left(x,v\right) = \mathbb{E}\left[Y\right| X = x, V = v]$.

The estimator of the logit model is

$$\min_{\tau} \sum_{i=1}^{n} \left(\hat{L}\left(X_i\right) - R\left(X_i;\tau\right)\right)^2$$

The estimator of τ satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \left(\hat{L}(X_i) - R(X_i;\hat{\tau})\right) \frac{\partial R}{\partial \tau}(X_i;\hat{\tau}) = 0 \tag{41}$$

The equation (41) that defines the estimator is generic and is the starting point of our analysis.

I The first stage

For a matrix A we define the matrix norm[10] $||A|| = \sqrt{\text{tr}(A'A)} = \sqrt{\sum_k \sum_l a_{kl}^2}$. We use $|a|$ for the Euclidean norm in the case that a is a vector or a scalar. We assume that for the control variable

Assumption 3

$$\frac{1}{n}\sum_{j=1}^{n} \left(\hat{V}_j - V_j\right)^2 = O\left(n^{-2\delta}\right)$$

Remark 1 *We discuss sufficient conditions for Assumption 3 in Sections K.2. We find that* $0 < \delta < \frac{1}{2}$ *and that the fastest rate depends on the smoothness of* $F(x|z)$ *and related functions as defined in Assumption 11.*

.

J The asymptotic distribution of the control variable estimator

J.1 Consistency and linearization

Because $L(X) = R(X;\tau_0)$ with $L(x) = \int_0^1 \mu(x,v)\text{d}v$, (41) can be rewritten as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(\hat{L}(X_i) - L(X_i)\right) \frac{\partial R}{\partial \tau}(X_i;\hat{\tau}) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(R(X_i;\hat{\tau}) - R(X_i;\tau_0)\right) \frac{\partial R}{\partial \tau}(X_i;\hat{\tau}) = 0 \tag{42}$$

If α is an l vector we define κ as an l vector of nonnegative integers with $|\kappa| = \sum_{j=1}^{l} \kappa_j$ and $\alpha^\kappa = \prod_{j=1}^{l} \alpha_j^{\kappa_j}$.

[10]If A and B are such that AB is well-defined, then $||AB|| \leq ||A||||B||$. A smaller upper bound is $||AB|| \leq \sqrt{\lambda_{max}(A'A)}||B||$.

Assumption 4 (Structural model) *The structural model has for exogenous $\tilde{X}$ that $\mathbb{E}[Y|\tilde{X}=x] = R(x;\tau_0)$. We assume that for all $\xi > 0$, there is an $\zeta > 0$ such that*

$$\inf_{\tau \in T, |\tau-\tau_0|>\xi} \mathbb{E}\left[(R(X;\tau) - R(X;\tau_0))^2\right] > \zeta$$

with T the parameter space of τ. Also for $d = 0,1,2$ and κ a vector of nonnegative integers of the same dimension as τ

$$\sup_{\tau \in T} \max_{|\kappa|=d} \left| \frac{\partial^d R}{\partial \tau^\kappa}(X;\tau) \right| \leq N_d(X)$$

with $\mathbb{E}[N_d(X)^2] < \infty$ for $d = 0,1$, $\mathbb{E}[N_2(X)] < \infty$ and the matrix

$$\mathbb{E}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial R}{\partial \tau'}(X;\tau_0)\right]$$

is nonsingular. $R(x;\tau)$ is r times continuously differentiable in x.

The next lemma gives conditions for weak consistency and an intermediate linearization result

Lemma 9 *If Assumption 4 holds and $\sup_{x\in\mathcal{X}} |\hat{L}(x) - L(x)| = o_p(1)$ with $\mathcal{X}$ the support of the distribution of X, then $\hat{\tau}$ defined in (41) is weakly consistent for τ_0 and*

$$\sqrt{n}(\hat{\tau} - \tau_0) = \left(\mathbb{E}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial R}{\partial \tau'}(X;\tau_0)\right]\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\hat{L}(X_i) - L(X_i)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0) + o_p(1) \tag{43}$$

Proof. We have

$$\frac{1}{n}\sum_{i=1}^n \left(\hat{L}(X_i) - R(X_i;\tau)\right)^2 = \frac{1}{n}\sum_{i=1}^n \left(\hat{L}(X_i) - L(X_i)\right)^2 + \frac{1}{n}\sum_{i=1}^n (R(X_i;\tau) - R(X_i;\tau_0))^2 -$$

$$\frac{2}{n}\sum_{i=1}^n (\hat{L}(X_i) - L(X_i))\,(R(X_i;\tau) - R(X_i;\tau_0))$$

The final term of this equation is bounded uniformly in τ by

$$\sup_{x\in\mathcal{X}} |\hat{L}(x) - L(x)| \frac{4}{n}\sum_{i=1}^n N_0(X_i) = o_p(1)$$

because by Assumption 4 $\mathbb{E}(N_0(X)) < \infty$. Because $\sup_{x\in\mathcal{X}} |\hat{L}(x) - L(x)| = o_p(1)$

$$\frac{1}{n}\sum_{i=1}^n \left(\hat{L}(X_i) - L(X_i)\right)^2 = o_p(1)$$

Finally because by Assumption 4 $\sup_{\tau\in T} |R(X;\tau)| \leq N_0(X)$ with $\mathbb{E}[N_0(X)^2] < \infty$, we have by the uniform weak law of large numbers

$$\frac{1}{n}\sum_{i=1}^n (R(X_i;\tau) - R(X_i;\tau_0))^2 \xrightarrow{p} \mathbb{E}\left[(R(X;\tau) - R(X;\tau_0)^2\right]$$

uniformly in τ. Combining these results we find

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{L}(X_i)-R(X_i;\tau)\right)^2 \xrightarrow{p} \mathbb{E}\left[(R(X;\tau)-R(X;\tau_0)^2\right]$$

uniformly in τ. By Assumption 4 the conditions in e.g. Van der Vaart (1998), Theorem 5.7 hold, so that $\hat{\tau}$ is weakly consistent for τ_0.

For the linearization first-order Taylor series expansions with $\tilde{\tau}$ and $\overline{\tau}$ intermediate points in (42) result in

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\hat{\tau})\frac{\partial R}{\partial \tau'}(X_i;\tilde{\tau})-\frac{1}{n}\sum_{i=1}^{n}\left(\hat{L}(X_i)-L(X_i)\right)\frac{\partial^2 R}{\partial \tau \partial \tau'}(X_i;\overline{\tau})\right)\sqrt{n}(\hat{\tau}-\tau_0)=$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{L}(X_i)-L(X_i)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)$$

By Assumption 4 and $\mathbb{E}(N_1(X)^2)<\infty$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau)\frac{\partial R}{\partial \tau'}(X_i;\tau) \xrightarrow{p} \mathbb{E}\left[\frac{\partial R}{\partial \tau}(X;\tau)\frac{\partial R}{\partial \tau'}(X;\tau)\right]$$

uniformly for $\tau \in T$. Finally because $\mathbb{E}(N_2(X))<\infty$

$$\sup_{\tau\in T}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{L}(X_i)-L(X_i)\right)\frac{\partial^2 R}{\partial \tau \partial \tau'}(X_i;\tau)\right| \leq \sup_{x\in\mathcal{X}}\left|\hat{L}(x)-L(x)\right|\frac{1}{n}\sum_{i=1}^{n}N_2(X_i)=o_p(1)$$

so that (43) follows. ■

Sufficient conditions for $\sup_{x\in\mathcal{X}}|\hat{L}(x)-L(x)|=o_p(1)$ will be discussed after Lemma 13 is proved.

J.2 Asymptotically linear representation

J.2.1 Decomposition and assumptions

For the rest of the derivation we take, without loss of generality, τ as scalar. We rewrite the sum in (43) as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{L}(X_i)-L(X_i)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)=$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{\mu}_K(X_i,\hat{V}_j)-\tilde{\mu}_K(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)+ \tag{44}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\tilde{\mu}_K(X_i,V_j)-\mu(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)+ \tag{45}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu(X_i,V_j)\frac{\partial R}{\partial \tau}(X_i;\tau_0)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L(X_i)\frac{\partial R}{\partial \tau}(X_i;\tau_0) \tag{46}$$

We write the three expressions (44), (45) and (46) as normalized sample averages. The three expressions each have a contribution to the influence function of our estimator. The contribution of (44) accounts for the estimation of the residuals in the first stage, the contribution of (45) accounts for the variability of the nonparametric regression estimator, and (46) is the pure variance term.

Equation (44) involves the feasible nonparametric regression estimator $\hat{\mu}_K$ of Y on X and $\hat{V}$, and the infeasible nonparametric regression estimator $\tilde{\mu}$ of Y on X and V. To simplify the discussion we define $W = (X\ V)'$ and $\hat{W} = (X\ \hat{V})'$. We use a series estimator. As the basis functions we take a power series and K is the number of basis functions in the series. To include all powers of x and v up to order L, we need to include $K = \frac{1}{2}(L+1)(L+2)$ terms. The resulting basis functions are denoted by the vector $Q_K^*(w) = (x^{\lambda_1} v^{\lambda_2}, \lambda_1 + \lambda_2 \le L)$. We order the basis function by $\lambda_1 + \lambda_2$. We make an assumption on the support of the joint distribution of X, V.

Assumption 5 (Support) *The support of X, V is $\mathcal{W} = \mathcal{X} \times \mathcal{V} = [x_L, x_U] \times [0,1]$. The joint density of X, V is bounded away from 0 on $\mathcal{W}$ and is r times continuously differentiable on its support. Also* $\mathbb{E}\left[\frac{f(X)^2 f(V)^2}{f(X,V)^2}\right] < \infty$.

By Newey (1995, Lemma A.15), this assumption implies that for all K there is a nonsingular matrix A_K such that if we define $\tilde{Q}_K(w) = A_K Q_K^*(w)$, the smallest eigenvalue of $\Omega_K = \mathbb{E}[\tilde{Q}_K(W)\tilde{Q}_K(W)']$ satisfies $\lambda_{min}(\Omega_K) \ge C > 0$ for all K. To simplify some of the argument we choose $Q_K(w) = \Omega_K^{-1/2}\tilde{Q}_K(w)$ so that $\mathbb{E}[Q_K(W)Q_K(W)'] = I_K$.

Define

$$\zeta_d(K) = \max_{|\lambda| \le d} \sup_{w \in \mathcal{W}} \left\| \frac{\partial^d Q_K}{\partial w^\lambda}(w) \right\|.$$

By Newey (1995, Lemma A.15), we have

$$\zeta_d(K) = O\left(K^{2d+1}\right).$$

Assumption 6 (Regression function) *The conditional mean of Y given X, V, $\mu(x,v)$, is twice continuously differentiable as a function of v. There is a vector γ_K such that for constants $C_D, a_D > 0$*

$$\max_{0 \le d \le D} \sup_{w \in \mathcal{W}} \left| \frac{\partial^d \mu}{\partial v^d}(w) - \frac{\partial^d Q_K}{\partial v^d}(w)'\gamma_K \right| \le C_D K^{-a_D}$$

and for $\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X,\tau_0)$ there is a vector δ_K such that for $C_D, a_D > 0$

$$\max_{0 \le d \le D} \sup_{w \in \mathcal{W}} \left| \frac{\partial^d}{\partial v^d}\left(\frac{f(x)f(v)}{f(x,v)}\frac{\partial R}{\partial \tau}(x;\tau_0)\right) - \frac{\partial^d Q_K}{\partial v^d}(w)'\delta_K \right| \le C_D K^{-a_D}$$

If the population regression function $\mu(w)$ is s times continuously differentiable on $\mathcal{W}$ we have by Lorentz (1986) that $a_0 = s/2$ (with 2 the dimension of w) for μ. For $\frac{f(x)f(v)}{f(x,v)}\frac{\partial R}{\partial \tau}(x;\tau_0)$ we have by Assumptions 4 and 5 that $a_0 = r/2$. In the sequel we need approximations up to the second derivative of μ and up to the first derivative of $\frac{f(x)f(v)}{f(x,v)}\frac{\partial R}{\partial \tau}(x;\tau_0)$, i.e., $D = 2$.

J.2.2 Basic properties of the series estimator

The first step in our proof is to derive some basic properties of the feasible and infeasible nonparametric regression estimators. For that purpose define the $K \times K$ matrices

$$\hat{\Omega}_K = \frac{1}{n} \sum_{j=1}^{n} Q_K(X_j, \hat{V}_j) Q_K(X_j, \hat{V}_j)'$$

and

$$\tilde{\Omega}_K = \frac{1}{n} \sum_{j=1}^{n} Q_K(X_j, V_j) Q_K(X_j, V_j)'$$

Lemma 10 *If Assumption 5 holds then*

$$\left\|\tilde{\Omega}_K - I_K\right\| = O_p\left(\zeta_0(K) K^{1/2} n^{-1/2}\right) = O_p\left(K^{3/2} n^{-1/2}\right),$$
$$\left\|\hat{\Omega}_K - \tilde{\Omega}_K\right\| = O_p\left(\zeta_1(K) n^{-\delta}\right) = O_p\left(K^3 n^{-\delta}\right),$$

if $K^3 n^{-\delta} \to 0$. *If that limit holds also*

$$\left\|\hat{\Omega}_K - I_K\right\| = O_p\left(K^3 n^{-\delta}\right).$$

Further

$$|\lambda_{max}(\tilde{\Omega}) - 1| = O_p\left(K^{3/2} n^{-1/2}\right), \qquad |\lambda_{min}(\tilde{\Omega}) - 1| = O_p\left(K^{3/2} n^{-1/2}\right),$$

and

$$|\lambda_{max}(\hat{\Omega}_K) - 1| = O_p\left(K^3 n^{-\delta}\right), \qquad |\lambda_{min}(\hat{\Omega}_K) - 1| = O_p\left(K^3 n^{-\delta}\right).$$

Proof. First, we note that by the argument in Newey (1997, proof of Theorem 1)

$$\left\|\tilde{\Omega}_K - I_k\right\| = O_p\left(\zeta_0(K) K^{1/2} n^{-1/2}\right) = O_p\left(K^{3/2} n^{-1/2}\right)$$

For a square symmetric $K \times K$ matrix A we have that for $k = 1, \ldots, K$

$$\lambda_k(A)^2 = \lambda_k(A^2) \leq \sum_{k=1}^{K} \lambda_k(A^2) = \mathrm{tr}(A^2) = \mathrm{tr}(A'A) = \|A\|^2$$

so that for $k = 1, \ldots, K$

$$|\lambda_k(A)| \leq \|A\|$$

and in particular $\lambda_{min}(A) \leq \|A\|$ and $\lambda_{max}(A) \leq \|A\|$. Also for any square matrix B, the eigenvalues of $B - I$ are equal to those of B minus 1. Therefore by choosing $A = \tilde{\Omega}_K - I_K$ we have

$$\left|\lambda_{max}(\tilde{\Omega}_K) - 1\right| \leq \left\|\tilde{\Omega}_K - I_K\right\| = O_p\left(K^{3/2} n^{-1/2}\right)$$

and the same bound holds for $|\lambda_{min}(\tilde{\Omega}_K) - 1|$.

By Assumption 3, we have $n^{-1}\sum_{i=1}^{n}\left(\hat{V}_i - V_i\right)^2 = O_p\left(n^{-2\delta}\right)$. By the mean value theorem, we have using the uniform bound on the derivative of $Q_K(w)$

$$
\begin{aligned}
\left\|\hat{Q}_K - Q_K\right\| \Big/ n^{1/2} &= \sqrt{n^{-1}\sum_{i=1}^{n}\left|Q_K\left(\hat{W}_i\right) - Q_K\left(W_i\right)\right|^2} \\
&= \sqrt{n^{-1}\sum_{i=1}^{n}\left|\frac{\partial Q_K}{\partial v}\left(\overline{W}_i\right)\left(\hat{V}_i - V_i\right)\right|^2} \\
&\leq \sqrt{\zeta_1\left(K\right)^2 n^{-1}\sum_{i=1}^{n}\left(\hat{V}_i - V_i\right)^2} \\
&= O_p\left(\zeta_1\left(K\right) n^{-\delta}\right),
\end{aligned}
$$

where $Q_K = \left[Q_K\left(W_1\right), \ldots, Q_K\left(W_n\right)\right]'$ and $\hat{Q}_K = \left[Q_K\left(\hat{W}_1\right), \ldots, Q_K\left(\hat{W}_n\right)\right]'$ are $n \times K$ matrices. Note that $Q_K(w)$ is a K vector of basis functions while Q_K is the $n \times K$ matrix of observed values of these basis functions. We have

$$
\begin{aligned}
\left\|\hat{\Omega}_K - \tilde{\Omega}_K\right\| &= \left\|\hat{Q}_K'\hat{Q}_K \Big/ n - Q_K'Q_K/n\right\| \\
&= \left\|\left(\hat{Q}_K - Q_K\right)'\left(\hat{Q}_K - Q_K\right)\Big/ n + Q_K'\left(\hat{Q}_K - Q_K\right)\Big/ n + \left(\hat{Q}_K - Q_K\right)' Q_K \Big/ n\right\| \\
&\leq \left\|\hat{Q}_K - Q_K\right\|^2 \Big/ n + 2\left\|Q_K'\left(\hat{Q}_K - Q_K\right)\Big/ n\right\|
\end{aligned}
$$

For a constant $C > 1$ define

$$
1_{PSD,n} = 1\left(\lambda_{max}(\tilde{\Omega}_K) < C\right),
$$

where we note that if $K^{3/2}n^{-1/2} \to 0$, then

$$
1_{PSD,n} \xrightarrow{p} 1.
$$

Now

$$
\left\|Q_K'\left(\hat{Q}_K - Q_K\right)\Big/ n\right\|^2 = \frac{1}{n^2}\text{tr}\left((\hat{Q}_K - Q_K)'Q_K Q_K'(\hat{Q}_K - Q_K)\right)
$$

and

$$
\begin{aligned}
&\frac{1}{n^2}\text{tr}\left((\hat{Q}_K - Q_K)'Q_K C\tilde{\Omega}_K^{-1}Q_K'(\hat{Q}_K - Q_K)\right) - \frac{1}{n^2}\text{tr}\left((\hat{Q}_K - Q_K)'Q_K Q_K'(\hat{Q}_K - Q_K)\right) \\
&= \frac{1}{n^2}\text{tr}\left((\hat{Q}_K - Q_K)'Q_K(C\tilde{\Omega}_K^{-1} - I_K)Q_K'(\hat{Q}_K - Q_K)\right) \\
&\geq \frac{1}{n^2}\left(\frac{C}{\lambda_{max}(\tilde{\Omega}_K)} - 1\right)\text{tr}\left((\hat{Q}_K - Q_K)'Q_K Q_K'(\hat{Q}_K - Q_K)\right),
\end{aligned}
$$

and the right-hand side is nonnegative if $\lambda_{max}(\tilde{\Omega}_K) < C$. Therefore

$$
1_{PSD,n}\frac{1}{n^2}\text{tr}\left((\hat{Q}_K - Q_K)'Q_K Q_K'(\hat{Q}_K - Q_K)\right) \leq 1_{PSD,n}\frac{1}{n^2}\text{tr}\left((\hat{Q}_K - Q_K)'Q_K C\tilde{\Omega}_K^{-1}Q_K'(\hat{Q}_K - Q_K)\right)
$$

and we have because $Q_K\tilde{\Omega}_K^{-1}Q_K' = nQ_K(Q_K'Q_K)^{-1}Q_K'$ with the matrix $Q_K(Q_K'Q_K)^{-1}Q_K'$ a projection matrix so that its eigenvalues are 0 or 1, that

$$1_{PSD,n}\left\|Q_K'\left(\hat{Q}_K - Q_K\right)\Big/ n\right\|^2 \leq 1_{PSD,n}\frac{C}{n}\text{tr}\left((\hat{Q}_K - Q_K)'(\hat{Q}_K - Q_K)\right)$$

or

$$1_{PSD,n}\left\|Q_K'\left(\hat{Q}_K - Q_K\right)\Big/ n\right\| \leq 1_{PSD,n}\frac{\sqrt{C}}{\sqrt{n}}\|\hat{Q}_K - Q_K\|,$$

Therefore if $K^{3/2}n^{-1/2} \to 0$

$$\left\|Q_K'\left(\hat{Q}_K - Q_K\right)\Big/ n\right\| = O_p\left(\zeta_1\left(K\right)n^{-\delta}\right).$$

We conclude that if $K^{3/2}n^{-1/2} \to 0$

$$\left\|\hat{\Omega}_K - \tilde{\Omega}_K\right\| = O_p\left(\zeta_1\left(K\right)^2 n^{-2\delta}\right) + O_p\left(\zeta_1\left(K\right)n^{-\delta}\right),$$

which is bounded by $O_p\left(\zeta_1\left(K\right)n^{-\delta}\right)$ if $\zeta_1\left(K\right)n^{-\delta} \to 0$. The bound on $\|\hat{\Omega}_K - I_K\|$ follows from the triangle inequality. The bounds on $|\lambda_{max}(\hat{\Omega}_K) - 1|$ and $|\lambda_{min}(\hat{\Omega}_K) - 1|$ are obtained in the same way as for the largest and smallest eigenvalue of $\tilde{\Omega}_K$. ■

We make the following assumption on the conditional variance of Y given $W = w$

Assumption 7 (Variance)

$$\sup_{w\in\mathcal{W}} \text{Var}(Y|W = w) \leq \overline{\sigma}^2 < \infty$$

Define

$$U_j = Y_j - \mu(X_j, V_j)$$

By assumption 7

$$\sup_{w\in\mathcal{W}} \text{Var}(U|W = w) \leq \overline{\sigma}^2 < \infty$$

Lemma 11 *If Assumptions 3, 5 and 7 hold then*

$$\left|\frac{1}{n}\sum_{j=1}^{n} Q_K(W_j)U_j\right| = O_p\left(K^{1/2}n^{-1/2}\right)$$

and

$$\left|\frac{1}{n}\sum_{j=1}^{n} Q_K(\hat{W}_j)U_j\right| = O_p\left(K^3 n^{-\delta}\right)$$

Proof. We have by first-order Taylor series expansions

$$\left|\frac{1}{n}\sum_{j=1}^{n} Q_K(X_j, \hat{V}_j)U_j\right| \leq \left|\frac{1}{n}\sum_{j=1}^{n} Q_K(X_j, V_j)U_j\right| + \left|\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(X_j, \overline{V}_j)\left(\hat{V}_j - V_j\right)U_j\right|$$

Now by Assumption 7

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^{n} Q_K(W_j)U_j\right|^2\right] = \frac{1}{n}\mathbb{E}\left[U^2 Q_K(W)'Q_K(W)\right] =$$

$$\frac{1}{n}\mathbb{E}\left[\text{Var}(Y|W)Q_K(W)'Q_K(W)\right] \le \frac{1}{n}\overline{\sigma}^2 \text{tr}\left(\mathbb{E}\left[Q_K(W)Q_K(W)'\right]\right) = \overline{\sigma}^2\frac{K}{n}$$

so that

$$\left|\frac{1}{n}\sum_{j=1}^{n} Q_K(X_j,V_j)U_j\right| = O_p(K^{1/2}n^{-1/2}).$$

Further if B is the matrix with columns $\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)U_j$ and $\hat{V},V$ the vectors with components $\hat{V}_j,V_j$, then

$$\begin{aligned}\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)\left(\hat{V}_j - V_j\right)U_j\right| &= \left|\frac{1}{n}B(\hat{V}-V)\right| \\ &\le \frac{1}{n}||B||\cdot|\hat{V}-V| \\ &= \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left|\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)U_j\right|^2}\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j-V_j\right)^2}\end{aligned}$$

Because by Assumption 7

$$\mathbb{E}\left[\left|\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)U_j\right|^2\right] = \mathbb{E}\left[\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)'\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)U_j^2\right] \le \overline{\sigma}^2\sup_{w\in\mathcal{W}}\left|\frac{\partial Q_K}{\partial v}(w)\right|^2$$

so that with Assumption 3

$$\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)\left(\hat{V}_j-V_j\right)U_j\right| = O_p\left(\zeta_1\left(K\right)\right)O_p\left(n^{-\delta}\right),$$

we conclude

$$\left|\frac{1}{n}\sum_{j=1}^{n} Q_K(\hat{W}_j)U_j\right| = O_p\left(K^{1/2}n^{-1/2}\right) + O_p\left(K^3 n^{-\delta}\right) = O_p\left(K^3n^{-\delta}\right).$$

■

Define $1_n = 1\left(\lambda_{min}\left(\hat{\Omega}_K\right)\ge 1/2\right)$ with $1(.)$ the indicator of the event between parentheses. By Lemma 10 if $K^3n^{-\delta}\to 0$ then $1_n \xrightarrow{p} 1$. The coefficients in the series estimator of $\mu(w)$ are

$$\hat{\gamma}_K = 1_n\left(\sum_{j=1}^{n} Q_K(\hat{W}_j)Q_K(\hat{W}_j)'\right)^{-1}\sum_{j=1}^{n}Q_K(\hat{W}_j)Y_j$$

We also consider the infeasible OLS estimator of the coefficients in the regression of Y on $Q_K(W)$

$$\tilde{\gamma}_K = \tilde{1}_n \left(\sum_{j=1}^{n} Q_K(W_j) Q_K(W_j)' \right)^{-1} \sum_{j=1}^{n} Q_k(W_j) Y_j$$

with $\tilde{1}_n = 1\left(\lambda_{min}\left(\tilde{\Omega}_K\right) \geq 1/2\right)$ where by Lemma 10 if $K^{3/2}n^{-1/2} \to 0$ then $\tilde{1}_n \xrightarrow{p} 1$.

In the proof of Lemma 13 we will need a bound on $\max_{1\leq j\leq n} |U_j|$ and to obtain this bound we make the following assumption

Assumption 8 (Finite absolute moments) *For some* $m \geq 3$

$$\mathbb{E}\left(|U_j|^m\right) < \infty$$

We have the following bound on $\max_{1\leq j\leq n} |U_j|$

Lemma 12 *If Assumption 8 holds, then*

$$\max_{1\leq j\leq n} |U_j| = O_p\left(n^{1/m}\right)$$

Proof. Because

$$\left(\max_{1\leq j\leq n} |U_j|\right)^m = \max_{1\leq j\leq n} |U_j|^m \leq \sum_{j=1}^{n} |U_j|^m$$

we have for any $C > 0$

$$\Pr\left(\left(\max_{1\leq j\leq n} |U_j|\right)^m \geq nC^m\right) \Pr\left(\max_{1\leq j\leq n} |U_j|^m \geq nC^m\right) \leq \Pr\left(\frac{1}{n}\sum_{j=1}^{n} |U_j|^m \geq nC^m\right) \leq \frac{\mathbb{E}\left(|U_j|^m\right)}{C^m}$$

Therefore $\frac{\max_{1\leq j\leq n}|U_j|}{n^{1/m}}$ is bounded in probability and the conclusion follows. ■

Lemma 13 *If Assumptions 3, 5, 6, 7 and 8 hold, then if* $K^{3/2}n^{-1/2} \to 0$

$$\tilde{1}_n \left|\tilde{\gamma}_K - \gamma_K\right| = O_p\left(K^{1/2}n^{-1/2}\right) + O_p\left(K^{1-a_D}\right)$$

and if in addition $K^3 n^{-\delta} \to 0$,

$$1_n \left|\hat{\gamma}_K - \gamma_K\right| = O_p\left(K^3 n^{-\delta}\right) + O_p\left(K^{1-a_D}\right)$$

and under the same assumptions

$$\begin{aligned} &\left| \hat{\gamma}_K - \tilde{\gamma}_K - \begin{pmatrix} \frac{1}{n}\sum_{j=1}^{n} \frac{\partial Q_K}{\partial v}(W_j) U_j (\hat{V}_j - V_j) \\ -\frac{1}{n}\sum_{j=1}^{n} Q_K(W_j) \frac{\partial \mu}{\partial v}(W_j) \left(\hat{V}_j - V_j\right) \end{pmatrix} \right| \\ &= O_p\left(K^6 n^{-2\delta}\right) + O_p\left(K^5 n^{\frac{1}{m}-2\delta}\right) + O_p\left(K^{1-a_D}\right) \end{aligned}$$

For $\hat{\mu}_K(w) = Q_K(w)'\hat{\gamma}_K$ *and* $\frac{\partial \hat{\mu}_K}{\partial v}(w) = \frac{\partial Q_K}{\partial v}(w)'\hat{\gamma}_K$

$$\sup_{w\in\mathcal{W}} |\hat{\mu}_K(w) - \mu(w)| = O_p\left(K^4 n^{-\delta}\right) + O_p\left(K^{2-a_D}\right)$$

and

$$\sup_{w\in\mathcal{W}} \left|\frac{\partial \hat{\mu}_K}{\partial v}(w) - \frac{\partial \mu}{\partial v}(w)\right| = O_p\left(K^6 n^{-\delta}\right) + O_p\left(K^{4-a_D}\right)$$

and for $\tilde{\mu}_K(w) = Q_K(w)'\tilde{\gamma}_K$ *and* $\frac{\partial \tilde{\mu}_K}{\partial v}(w) = \frac{\partial Q_K}{\partial v}(w)'\tilde{\gamma}_K$

$$\sup_{w\in\mathcal{W}} |\tilde{\mu}_K(w) - \mu(w)| = O_p\left(K^{3/2} n^{-1/2}\right) + O_p\left(K^{2-a_D}\right)$$

and

$$\sup_{w\in\mathcal{W}} \left|\frac{\partial \tilde{\mu}_K}{\partial v}(w) - \frac{\partial \mu}{\partial v}(w)\right| = O_p\left(K^{7/2} n^{-1/2}\right) + O_p\left(K^{4-a_D}\right)$$

Proof. We have

$$\tilde{1}_n(\tilde{\gamma}_K - \gamma_K) = \tilde{1}_n\tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)(Y_j - \mu(W_j)) + \tilde{1}_n\tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)(\mu(W_j) - Q_K(W_j)'\gamma_K)$$

Define $U = (U_1, \ldots, U_n)'$ and $\tilde{U} = (\mu(W_1) - Q_K(W_1)'\gamma_K, \ldots, \mu(W_n) - Q_K(W_n)'\gamma_K)'$. Consider the matrix $4I_K - (\tilde{\Omega}_K^{-1})^2$. This is positive semi-definite if its smallest eigenvalue is nonnegative

$$4 - \frac{1}{\lambda_{min}(\tilde{\Omega}_K)^2} \geq 0$$

and this holds if $\tilde{1}_n = 1$. Then if $K^{3/2}n^{-1/2} \to 0$

$$\begin{aligned}\left|\tilde{1}_n\tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)(Y_j - \mu(W_j))\right|^2 &= \tilde{1}_n\frac{1}{n^2}U'Q_K\left(\tilde{\Omega}_K^{-1}\right)^2 Q_K'U \\ &\leq \tilde{1}_n\frac{4}{n^2}\left|Q_K'U\right|^2 = O_p\left(Kn^{-1}\right).\end{aligned}$$

by Lemma 11. By an analogous argument if Assumption 6 holds and $K^{3/2}n^{-1/2} \to 0$

$$\left|\tilde{1}_n\tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)(\mu(W_j) - Q_K(W_j)'\gamma_K)\right|^2 \leq \tilde{1}_n\frac{4}{n^2}\left|Q_K'\tilde{U}\right|^2 \leq \tilde{1}_n\frac{4}{n^2}||Q_K||^2|\tilde{U}|^2 =$$

$$4\tilde{1}_n\left(\frac{1}{n}\sum_{j=1}^n |Q_K(W_j)|^2\right)\left(\frac{1}{n}\sum_{j=1}^n(\mu(W_j) - Q_K(W_j)'\gamma_K)^2\right) = O_p\left(\zeta_0(K)^2K^{-2a_D}\right)$$

Therefore if $K^{3/2}n^{-1/2} \to 0$

$$\tilde{1}_n|\tilde{\gamma}_K - \gamma_K| = O_p\left(K^{1/2}n^{-1/2}\right) + O_p\left(K^{1-a_D}\right).$$

Next

$$1_n(\hat{\gamma}_K-\gamma_K)=1_n\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)U_j+1_n\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)(\mu(W_j)-Q_K(\hat{W}_j)'\gamma_K)=$$

$$1_n\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)U_j+1_n\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)(\mu(W_j)-\mu(\hat{W}_j))+1_n\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)(\mu(\hat{W}_j)-Q_K(\hat{W}_j)'\gamma_K) \tag{47}$$

We use (47) to derive two results: the rate of convergence of $\hat{\gamma}_K-\gamma_K$ and the rate of convergence of the difference $\hat{\gamma}_K-\tilde{\gamma}_K$. Using a similar argument as above and Lemma 11 we have if $K^3n^{-\delta}\to 0$

$$\left|1_n\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)U_j\right|\le 2\cdot 1_n\left|\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)U_j\right|=O_p\left(K^3n^{-\delta}\right)$$

By Lemma 10 $1_n\xrightarrow{p}1$ if $K^3n^{-\delta}\to 0$, so that if this limit holds 1_n can be omitted and we do so in the rest of this proof (and we also omit $\tilde{1}_n$ with the understanding that $K^{3/2}n^{-1/2}\to 0$). By a first-order Taylor series expansion the absolute value of the second term can be expressed as

$$\left|\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)\frac{\partial\mu}{\partial v}(\overline{W}_j)\left(\hat{V}_j-V_j\right)\right|$$

which is bounded, using a the same argument as above, by

$$\left|\frac{2}{n}\sum_{j=1}^n Q_K(\hat{W}_j)\frac{\partial\mu}{\partial v}(\overline{W}_j)\left(\hat{V}_j-V_j\right)\right|=\frac{2}{n}|B(\hat{V}-V)|\le\frac{2}{n}||B|||\hat{V}-V|$$

with B the matrix with columns $Q_K(\hat{W}_j)\frac{\partial\mu}{\partial v}(\overline{W}_j)$. The last expression is equal to

$$2\sqrt{\frac{1}{n}\sum_{j=1}^n|Q_K(\hat{W}_j)|^2\left|\frac{\partial\mu}{\partial v}(\overline{W}_j)\right|^2}\sqrt{\frac{1}{n}\sum_{j=1}^n\left|\hat{V}_j-V_j\right|^2}=\zeta_0\left(K\right)O_p\left(n^{-\delta}\right)=O_p\left(Kn^{-\delta}\right)$$

by Assumption 3, the uniform bound on the basis functions and Assumption 6 by which $\mu(w)$ is continuously differentiable. The third term in (47) is, after increasing the bound as above to eliminate $\hat{\Omega}_K^{-1}$, bounded in the norm by

$$2\sup_{w\in\mathcal{W}}|Q_K(w)|\sup_{w\in\mathcal{W}}\left|\mu(w)-Q_K(w)'\gamma_K\right|=\zeta_0\left(K\right)O_p\left(K^{-a_D}\right)=O_p\left(K^{1-a_D}\right)$$

Therefore if $K^3n^{-\delta}\to 0$

$$1_n\left|\hat{\gamma}_K-\gamma_K\right|=O_p\left(K^3n^{-\delta}\right)+O_p\left(K^{1-a_D}\right)$$

Next we consider $\hat{\gamma}_K-\tilde{\gamma}_K$. We are particularly interested in terms that depend on $\hat{V}_j-V_j$ and for that reason we will consider second-order Taylor series expansions for the first and second term of

(47). A second-order Taylor series expansion gives for the first term of (47)

$$\begin{aligned}
\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)U_j =& \hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)U_j + \hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n \frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j - V_j) + \\
& \frac{1}{2}\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n \frac{\partial^2 Q_K}{\partial v^2}(X_j,\overline{V}_j)U_j(\hat{V}_j - V_j)^2 \\
=& \tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)U_j + \tilde{\Omega}_K^{-1}(\tilde{\Omega}_K - \hat{\Omega}_K)\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)U_j + \\
& \hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n \frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j - V_j) + \frac{1}{2}\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n \frac{\partial^2 Q_K}{\partial v^2}(X_j,\overline{V}_j)U_j(\hat{V}_j - V_j)^2 \\
=& \tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)\left(U_j + (\mu(W_j) - Q_K(W_j)'\gamma_K)\right) \\
& - \tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)(\mu(W_j) - Q_K(W_j)'\gamma_K) + \tilde{\Omega}_K^{-1}(\tilde{\Omega}_K - \hat{\Omega}_K)\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)U_j \qquad (48) \\
& + \hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n \frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j - V_j) + \frac{1}{2}\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n \frac{\partial^2 Q_K}{\partial v^2}(X_j,\overline{V}_j)U_j(\hat{V}_j - V_j)^2
\end{aligned}$$

The first term on the right-hand side is equal to $\tilde{\gamma}_K - \gamma_K$. For the second term of (48) we found above that if $K^{3/2}n^{-1/2} \to 0$

$$\left|\tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)(\mu(W_j) - Q_K(W_j)'\gamma_K)\right| = O_p\left(K^{1-a_D}\right)$$

By an argument that we have used several times above to remove $\tilde{\Omega}_K^{-1}$ and $\hat{\Omega}_K^{-1}$ (as before we omit the indicators that the relevant matrices are positive definite) the third term on the right-hand side of (48) is bounded by (if $K^{3/2}n^{-1/2} \to 0$)

$$\left|\tilde{\Omega}_K^{-1}(\tilde{\Omega}_K - \hat{\Omega}_K)\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(W_j)U_j\right| \leq 2\left|(\tilde{\Omega}_K - \hat{\Omega}_K)\hat{\Omega}_K^{-1}\frac{1}{n}Q_K'U\right| \leq 2\|\tilde{\Omega}_K - \hat{\Omega}_K\|\left|\hat{\Omega}_K^{-1}\frac{1}{n}Q_K'U\right|$$

$$\leq 4\|\tilde{\Omega}_K - \hat{\Omega}_K\|\left|\frac{1}{n}Q_K'U\right| = O_p\left(K^3 n^{-\delta}\right)O_p\left(K^{1/2}n^{-1/2}\right) = O_p\left(K^{7/2}n^{-\delta-1/2}\right)$$

by Lemma 10 and 11. Also for the fifth term on the right-hand side of (48), removing $\hat{\Omega}_K^{-1}$ by increasing the bound as before (and if $K^3n^{-\delta}\to 0$)

$$\begin{aligned}\left|\frac{1}{2}\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n\frac{\partial^2 Q_K}{\partial v^2}(X_j,\overline{V}_j)U_j(\hat{V}_j-V_j)^2\right| &\le \left|\frac{1}{n}\sum_{j=1}^n\frac{\partial^2 Q_K}{\partial v^2}(X_j,\overline{V}_j)U_j(\hat{V}_j-V_j)^2\right|\\ &\le \sup_{w\in\mathcal{W}}\left|\frac{\partial^2 Q_K}{\partial v^2}(w)\right|\max_{1\le j\le n}|U_j|\frac{1}{n}\sum_{j=1}^n(\hat{V}_j-V_j)^2\\ &= \zeta_2(K)O_p(n^{1/m})O_p(n^{-2\delta}) = O_p\left(K^5 n^{\frac{1}{m}-2\delta}\right)\end{aligned}$$

by Assumptions 3 and 8. Combining these results we have that the first term of (47) is (we keep the first and fourth terms, then (if $K^3n^{-\delta}\to 0$ so that also $K^{3/2}n^{-1/2}\to 0$)

$$\begin{aligned}\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)U_j &= (\tilde{\gamma}_K-\gamma_K)+\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j)\\ &\quad + O_p\left(K^{7/2}n^{-\delta-1/2}\right)+O_p\left(K^5n^{\frac{1}{m}-2\delta}\right)+O_p(K^{1-a_D})\end{aligned}$$

We have

$$\frac{K^{7/2}n^{-\delta-1/2}}{K^5 n^{\frac{1}{m}-2\delta}} = \frac{n^{\delta-\frac{1}{2}-\frac{1}{m}}}{K^{\frac{3}{2}}}\to 0,$$

and we may ignore the $O_p\left(K^{7/2}n^{-\delta-1/2}\right)$ term. Note that

$$\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j) =$$

$$\frac{1}{n}\sum_{j=1}^n\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j)+\hat{\Omega}_K^{-1}(I_K-\hat{\Omega}_K)\frac{1}{n}\sum_{j=1}^n\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j)$$

and if Assumptions 3 and 7 hold and $K^3n^{-\delta}\to 0$ then by the argument that we have used before to remove $\hat{\Omega}_K^{-1}$

$$\begin{aligned}\left|\hat{\Omega}_K^{-1}(I_K-\hat{\Omega}_K)\frac{1}{n}\sum_{j=1}^n\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j)\right| &\le 2\left|(I_K-\hat{\Omega}_K)\frac{1}{n}\sum_{j=1}^n\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j)\right|\\ &\le 2\|I_K-\hat{\Omega}_K\|\sqrt{\frac{1}{n}\sum_{j=1}^n\left|\frac{\partial Q_K}{\partial v}(W_j)U_j\right|^2}\sqrt{\frac{1}{n}\sum_{j=1}^n(\hat{V}_j-V_j)^2}\\ &= O_p\left(K^3n^{-\delta}\right)\zeta_1(K)O_p\left(n^{-\delta}\right) = O_p(K^6n^{-2\delta}).\end{aligned}$$

where we also use Lemma 10. Therefore if $K^3n^{-\delta}\to 0$ the first term of (47) is

$$\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^n Q_K(\hat{W}_j)U_j = (\tilde{\gamma}_K-\gamma_K)+\frac{1}{n}\sum_{j=1}^n\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j)+O_p(K^6n^{-2\delta})+O_p\left(K^5n^{\frac{1}{m}-2\delta}\right)+O_p(K^{1-a_D})$$

The second term in (47) is by a second-order Taylor series expansion and using Assumption 3 and 6 to bound the second term below

$$
\begin{aligned}
\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_K(\hat{W}_j)(\mu(W_j)-\mu(\hat{W}_j)) &= -\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_K(\hat{W}_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right) \\
&\quad -\frac{1}{2}\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_K(\hat{W}_j)\frac{\partial^2\mu}{\partial v^2}(X_j,\overline{V}_j)\left(\hat{V}_j-V_j\right)^2 \\
&= -\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_K(\hat{W}_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)+O_p(\zeta_0(K)n^{-2\delta})
\end{aligned}
$$

and

$$
\begin{aligned}
-\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_K(\hat{W}_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right) &= -\frac{1}{n}\sum_{j=1}^{n}Q_K(W_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right) \\
&\quad -\hat{\Omega}_K^{-1}(I_K-\hat{\Omega}_K)\frac{1}{n}\sum_{j=1}^{n}Q_K(W_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right) \\
&\quad -\hat{\Omega}_K^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(X_j,\overline{V}_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)^2 \\
&= -\frac{1}{n}\sum_{j=1}^{n}Q_K(W_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right) \\
&\quad +O_p\left(K^3n^{-\delta}\right)\zeta_0(K)O_p\left(n^{-\delta}\right)+O_p(\zeta_1(K)n^{-2\delta}) \\
&= -\frac{1}{n}\sum_{j=1}^{n}Q_K(W_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)+O_p(K^4n^{-2\delta})
\end{aligned}
$$

Finally, we already showed that the third term in (47) is $O_p\left(K^{1-a_D}\right)$. Collecting results we conclude that, if if $K^3n^{-\delta}\to 0$ (so that also $K^{3/2}n^{-1/2}\to 0$), then

$$
\begin{aligned}
\hat{\gamma}_K-\gamma_K &= (\tilde{\gamma}_K-\gamma_K) \\
&\quad +\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j(\hat{V}_j-V_j) \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}Q_K(W_j)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right) \\
&\quad +O_p\left(K^6n^{-2\delta}\right)+O_p\left(K^5n^{\frac{1}{m}-2\delta}\right)+O_p\left(K^{1-a_D}\right)
\end{aligned}
$$

Next

$$
\sup_{w\in\mathcal{W}}|\hat{\mu}_K(w)-\mu(w)|\leq\sup_{w\in\mathcal{W}}|Q_K(w)'(\hat{\gamma}_K-\gamma_K)|+\sup_{w\in\mathcal{W}}\left|Q_K(w)'\gamma_K-\mu(w)\right|\leq\zeta_0\left(K\right)|\hat{\gamma}_K-\gamma_K|+O\left(K^{-a_D}\right)
$$

and

$$
\sup_{w\in\mathcal{W}}\left|\frac{\partial\hat{\mu}_K}{\partial v}(w)-\frac{\partial\mu}{\partial v}(w)\right|\leq
$$

$$\sup_{w\in\mathcal{W}}\left|\frac{\partial Q_K}{\partial v}(w)'(\hat{\gamma}_K-\gamma_K)\right| + \sup_{w\in\mathcal{W}}\left|\frac{\partial Q_K}{\partial v}(w)'\gamma_K - \frac{\partial \mu}{\partial v}(w)\right| \leq$$

$$\zeta_1(K)\left|\hat{\gamma}_K-\gamma_K\right| + O\left(K^{-a_D}\right)$$

Finally

$$\sup_{w\in\mathcal{W}}|\tilde{\mu}_K(w)-\mu_K(w)| \leq \sup_{w\in\mathcal{W}}|Q_K(w)'(\tilde{\gamma}_K-\gamma_K)| + \sup_{w\in\mathcal{W}}|Q_K(w)'\gamma_K-\mu(w)| \leq$$

$$\sup_{w\in\mathcal{W}}|Q_K(w)||\tilde{\gamma}_K-\gamma_K| + \sup_{w\in\mathcal{W}}|Q_K(w)'\gamma_K-\mu(w)| \leq$$

$$\zeta_0(K)|\tilde{\gamma}_K-\gamma_K| + O\left(K^{-a_D}\right)$$

and

$$\sup_{w\in\mathcal{W}}\left|\frac{\partial \tilde{\mu}_K}{\partial v}(w) - \frac{\partial \mu}{\partial v}(w)\right| \leq$$

$$\sup_{w\in\mathcal{W}}\left|\frac{\partial Q_K}{\partial v}(w)'(\tilde{\gamma}_K-\gamma_K)\right| + \sup_{w\in\mathcal{W}}\left|\frac{\partial Q_K}{\partial v}(w)'\gamma_K - \frac{\partial \mu}{\partial v}(w)\right| \leq$$

$$\zeta_1(K)|\tilde{\gamma}_K-\gamma_K| + O\left(K^{-a_D}\right)$$

■

Corollary 1 *Under the assumptions of Lemma 13 and if* $a_D > 2$ *and* $K^4 n^{-\delta} \to 0$*, we have for* $\hat{L}(x) = \frac{1}{n}\sum_{j=1}^{n}\hat{\mu}_K(x,\hat{V}_j)$ *and* $L(x) = \mathbb{E}[\mu(x,V)]$

$$\sup_{x\in\mathcal{X}}\left|\hat{L}(x)-L(x)\right| = O_p\left(K^4 n^{-\delta}\right) + O_p\left(K^{2-a_D}\right) = o_p(1)$$

Proof.

$$\sup_{x\in\mathcal{X}}\left|\hat{L}(x)-L(x)\right| = \sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\hat{\mu}_K(x,\hat{V}_j)-\mu(x,\hat{V}_j))\right| + \sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\mu(x,\hat{V}_j)-\mu(x,V_j))\right| +$$

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\mu(x,V_j)-\mathbb{E}[\mu(x,V)])\right|$$

By Lemma 13

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\hat{\mu}_K(x,\hat{V}_j)-\mu(x,\hat{V}_j))\right| \leq \sup_{w\in\mathcal{W}}|\hat{\mu}_K(w)-\mu(w)|$$

$$= O_p\left(K^4 n^{-\delta}\right) + O_p\left(k^{2-a_D}\right)$$

By a first-order Taylor expansion and Assumptions 3 and 6

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\mu(x,\hat{V}_j)-\mu(x,V_j))\right| \leq \sup_{x\in\mathcal{X}}\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left|\frac{\partial\mu_K}{\partial v}(x,\overline{V}_j)\right|^2}\sqrt{\frac{1}{n}\sum_{j=1}^{n}(\hat{V}_j-V_j)^2}$$

$$\leq \sup_{w\in\mathcal{W}}\left|\frac{\partial\mu_K}{\partial v}(w)\right|\sqrt{\frac{1}{n}\sum_{j=1}^{n}(\hat{V}_j-V_j)^2} = O_p(n^{-\delta})$$

and finally by Assumption 6 and the uniform law of large numbers

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\mu(x,V_j)-\mathbb{E}[\mu(x,V)])\right| = o_p(1)$$

■

The next step is to consider (44), (45) and (46) and express these as sample averages.

J.3 Expressing (44) as a sample average

We express (44) as

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{\mu}_K(X_i,\hat{V}_j)-\tilde{\mu}_K(X_i,V_j)\right)\frac{\partial R}{\partial\tau}(X_i;\tau_0) =$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial\tau}(X_i;\tau_0)\hat{\gamma}_k'(Q_K(X_i,\hat{V}_j)-Q_K(X_i,V_j))+ \tag{49}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial\tau}(X_i;\tau_0)(\hat{\gamma}_K-\tilde{\gamma}_K)'Q_K(X_i,V_j) \tag{50}$$

Expressing (50) as a sample average will be the main challenge. This will involve use of the V-statistic projection theorem and in addition the transfer of coefficients in a projection from one factor in the expansion to another. The latter is the key step in obtaining the influence function of our estimator.

Before dealing with (50) we first express (49) as a linear expression in $\hat{V}_j - V_j$ in Lemma 14 and 15

Lemma 14 *If Assumptions 3, 4, 5, 6, 7 hold, and $K^3 n^{-\delta}\to 0$, then (49) is equal to*

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial\tau}(X_i;\tau_0)\hat{\gamma}_K'(Q_K(X_i,\hat{V}_j)-Q_K(X_i,V_j))$$

$$=\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial\tau}(X_i;\tau_0)\frac{\partial\mu}{\partial v}(X_i,V_j)\left(\hat{V}_j-V_j\right)$$

$$+O_p\left(K^6 n^{1/2-2\delta}\right)+O_p\left(K^{4-a_D}n^{1/2-\delta}\right)+O_p(n^{1/2-2\delta})$$

Proof. For (49) by a second order Taylor expansion with respect to $\hat{V}_j$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\hat{\gamma}_K'(Q_K(X_i,\hat{V}_j)-Q_K(X_i,V_j))$$

$$=\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\gamma_K'\frac{\partial Q_K}{\partial v}(X_i,V_j)\left(\hat{V}_j-V_j\right) \quad (51)$$

$$+\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_K-\gamma_K)'\frac{\partial Q_K}{\partial v}(X_i,V_j)\left(\hat{V}_j-V_j\right) \quad (52)$$

$$+\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_K-\gamma_K)'\frac{\partial^2 Q_K}{\partial v^2}(X_i,\overline{V}_j)\left(\hat{V}_j-V_j\right)^2 \quad (53)$$

$$+\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(\gamma_K'\frac{\partial^2 Q_K}{\partial v^2}(X_i,\overline{V}_j)-\frac{\partial^2 \mu}{\partial v^2}(X_i,\overline{V}_j)\right)\left(\hat{V}_j-V_j\right)^2 \quad (54)$$

$$+\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial^2 \mu}{\partial v^2}(X_i,\overline{V}_j)\left(\hat{V}_j-V_j\right)^2 \quad (55)$$

Now for (52) by the triangle inequality, Assumption 3 and 4, and Lemma 13

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_K-\gamma_K)'\frac{\partial Q_K}{\partial v}(X_i,V_j)\left(\hat{V}_j-V_j\right)\right| \leq$$

$$\sqrt{n}\sup_{w\in\mathcal{W}}\left|\frac{\partial Q_K}{\partial v}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\right)\left(\frac{1}{n}\sum_{j=1}^{n}\left|\hat{V}_j-V_j\right|\right)|\hat{\gamma}_K-\gamma_K|$$

$$\begin{aligned}&=n^{1/2}\zeta_1(K)O_p\left(n^{-\delta}\right)\left(O_p\left(K^3n^{-\delta}\right)+O_p\left(K^{1-a_D}\right)\right)\\&=O_p\left(K^6n^{1/2-2\delta}\right)+O_p\left(K^{4-a_D}n^{1/2-\delta}\right)\end{aligned}$$

and for (53) by the triangle inequality, Assumption 3 and 4, and Lemma 13

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_K-\gamma_K)'\frac{\partial^2 Q_K}{\partial v^2}(X_i,\overline{V}_j)\left(\hat{V}_j-V_j\right)^2\right| \leq$$

$$\frac{1}{2}\sqrt{n}\sup_{w\in\mathcal{W}}\left|\frac{\partial^2 Q_K}{\partial v^2}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\right)\left(\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j-V_j\right)^2\right)|\hat{\gamma}_K-\gamma_K|$$

$$\begin{aligned}&=O_p\left(n^{1/2}\right)O_p\left(K^5\right)O_p\left(n^{-2\delta}\right)\left(O_p\left(K^3n^{-\delta}\right)+O_p\left(K^{1-a_D}\right)\right)\\&=O_p\left(K^8n^{1/2-3\delta}\right)+O_p\left(K^{6-a_D}n^{1/2-2\delta}\right)\end{aligned}$$

and for (54) by the triangle inequality, Assumption 3, 4 and 6

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(\gamma_K'\frac{\partial^2 Q_K}{\partial v^2}(X_i,\overline{V}_j)-\frac{\partial^2 \mu}{\partial v^2}(X_i,\overline{V}_j)\right)\left(\hat{V}_j-V_j\right)^2\right| \leq$$

$$\frac{1}{2}\sqrt{n}\sup_{w\in\mathcal{W}}\left|\gamma_K'\frac{\partial^2 Q_K}{\partial v^2}(w)-\frac{\partial^2 \mu_K}{\partial v^2}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\right)\left(\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j-V_j\right)^2\right)$$

$$= O_p\left(n^{1/2}\right)O_p(K^{-a_D})O_p\left(n^{-2\delta}\right) = O_p(K^{-a_D}n^{1/2-2\delta})$$

and for (55) by the triangle inequality, Assumption 3 and 4

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial^2 \mu}{\partial v^2}(X_i,\overline{V}_j)\left(\hat{V}_j-V_j\right)^2\right| \leq$$

$$\frac{1}{2}\sqrt{n}\sup_{w\in\mathcal{W}}\left|\frac{\partial^2 \mu}{\partial v^2}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\right)\left(\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j-V_j\right)^2\right) = O_p(n^{1/2-2\delta})$$

Finally for (51) by Assumption 6

$$\begin{aligned}
&\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\gamma_K'\frac{\partial Q_K}{\partial v}(X_i,V_j)\left(\hat{V}_j-V_j\right)\\
&=\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}\left(X_i;\tau_0\right)\frac{\partial \mu}{\partial v}(X_i,V_j)\left(\hat{V}_j-V_j\right)\\
&-\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}\left(X_i;\tau_0\right)\left(\frac{\partial \mu}{\partial v}(X_i,V_j)-\gamma_K'\frac{\partial Q_K}{\partial v}(X_i,V_j)\right)\left(\hat{V}_j-V_j\right)
\end{aligned}$$

with by Assumption 3, 4 and 6

$$\begin{aligned}
&\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(\gamma_K'\frac{\partial Q_K}{\partial v}(X_i,V_j)-\frac{\partial \mu}{\partial v}(X_i,V_j)\right)\left(\hat{V}_j-V_j\right)\right|\\
&\leq\sqrt{n}\sup_{w}\left|\gamma_K'\frac{\partial Q_K}{\partial v}(w)-\frac{\partial \mu}{\partial v}(w)\right|\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\left|\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j-V_j\right)\right|\\
&=O_p\left(K^{-a_D}n^{1/2-\delta}\right)
\end{aligned}$$

Combining the results we have for (49)

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\hat{\gamma}_K'(Q_K(X_i,\hat{V}_j)-Q_K(X_i,V_j)) =$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}\left(X_i;\tau_0\right)\frac{\partial \mu}{\partial v}(X_i,V_j)\left(\hat{V}_j - V_j\right) +$$
$$O_p\left(K^6 n^{1/2-2\delta}\right) + O_p\left(K^{4-a_D} n^{1/2-\delta}\right) + O_p\left(K^8 n^{1/2-3\delta}\right) + O_p\left(K^{6-a_D} n^{1/2-2\delta}\right)$$
$$+ O_p\left(K^{-a_D} n^{1/2-2\delta}\right) + O_p(n^{1/2-2\delta}) + O_p\left(K^{-a_D} n^{1/2-\delta}\right)$$

Because $K^3 n^{-\delta} \to 0$, we have

$$O_p\left(K^8 n^{1/2-3\delta}\right) = K^2 n^{-\delta} O_p\left(K^6 n^{1/2-2\delta}\right) = o_p\left(K^6 n^{1/2-2\delta}\right)$$
$$O_p\left(K^{6-a_D} n^{1/2-2\delta}\right) = K^2 n^{-\delta} O_p\left(K^{4-a_D} n^{1/2-\delta}\right) = o_p\left(K^{4-a_D} n^{1/2-\delta}\right)$$

We also have

$$O_p\left(K^{-a_D} n^{1/2-2\delta}\right) = K^{-4} n^{-\delta} O_p\left(K^{4-a_D} n^{1/2-\delta}\right) = o_p\left(K^{4-a_D} n^{1/2-\delta}\right)$$
$$O_p\left(K^{-a_D} n^{1/2-\delta}\right) = K^{-4} O_p\left(K^{4-a_D} n^{1/2-\delta}\right) = o_p\left(K^{4-a_D} n^{1/2-\delta}\right)$$

Therefore, we obtain

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\hat{\gamma}_K'(Q_K(X_i,\hat{V}_j) - Q_K(X_i,V_j))$$
$$= \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}\left(X_i;\tau_0\right)\frac{\partial \mu}{\partial v}(X_i,V_j)\left(\hat{V}_j - V_j\right)$$
$$+ O_p\left(K^6 n^{1/2-2\delta}\right) + O_p\left(K^{4-a_D} n^{1/2-\delta}\right) + O_p(n^{1/2-2\delta})$$

■

Lemma 15 *If Assumptions 3, 4, 5, 6, and 7 hold and* $K^3 n^{-\delta} \to 0$, *then (49) is equal to*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right]\left(\hat{V}_j - V_j\right)$$
$$+ O_p\left(K^6 n^{1/2-2\delta}\right) + O_p\left(K^{4-a_D} n^{1/2-\delta}\right) + O_p(n^{1/2-2\delta})$$

Proof. Consider

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}\left(X_i,V_j\right) - \mathbb{E}_X\left[\frac{\partial R}{\partial \tau}\left(X;\tau_0\right)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right]\right)\right)\left(\hat{V}_j - V_j\right) \tag{56}$$

By Cauchy-Schwartz

$$\left|\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}\left(X_i,V_j\right) - \mathbb{E}_X\left[\frac{\partial R}{\partial \tau}\left(X;\tau_0\right)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right]\right)\right)\left(\hat{V}_j - V_j\right)\right)\right|^2$$
$$\leq \sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}\left(X_i,V_j\right) - \mathbb{E}_X\left[\frac{\partial R}{\partial \tau}\left(X;\tau_0\right)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right]\right)\right)^2 \cdot \frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j - V_j\right)^2 \tag{57}$$

Because by Assumption 4 and 6

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}\left(X_i,V_j\right)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}\left(X;\tau_0\right)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right]\right)\right)^2\right]\\
&=\mathbb{E}\left[\operatorname{Var}_X\left(\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\middle| V_j\right)\right]\\
&\leq \mathbb{E}\left[\mathbb{E}_X\left[\left(\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right)^2\middle| V_j\right]\right]\\
&\leq \sup_{w\in\mathcal{W}}\left(\frac{\partial \mu}{\partial v}\left(w\right)\right)^2\mathbb{E}_X\left[N_1(X)^2\right]<\infty
\end{aligned}$$

we have by the Markov inequality that

$$\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}\left(X_i,V_j\right)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}\left(X;\tau_0\right)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right]\right)\right)^2=O_p(1)$$

so that (57) by Assumption 3 is $O_p(n^{-2\delta})$ and therefore

$$\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}\left(X_i,V_j\right)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}\left(X;\tau_0\right)\frac{\partial \mu}{\partial v}\left(X,V_j\right)\right]\right)\right)\left(\hat{V}_j-V_j\right)\right|=O_p\left(n^{-\delta}\right)$$

Because

$$O_p\left(n^{-\delta}\right)=n^{-1/2+\delta}O_p\left(n^{1/2-2\delta}\right)=o_p\left(n^{1/2-2\delta}\right)$$

we get the desired conclusion. ■

For (50) we find after substitution from Lemma 13

$$\begin{aligned}
&\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)'(\hat{\gamma}_K-\tilde{\gamma}_K)\\
&=-\sqrt{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)'\cdot\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right) &(58)\\
&+\sqrt{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)'\cdot\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j-V_j\right) &(59)\\
&+O_p\left(K^7n^{1/2-2\delta}\right)+O_p\left(K^6n^{1/2+1/m-2\delta}\right)+O_p\left(K^{2-a_D}n^{1/2}\right)
\end{aligned}$$

Lemma 16 *If Assumptions 3, 4, 5, 6, 7 hold and $K^3n^{-\delta}\to 0$, then (58) is equal to*

$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)+O_p\left(K^{7/2}n^{-\delta}\right)+O_p\left(K^{2-a_D}n^{1/2-\delta}\right)$$

Proof. We write (58) as

$$-\sqrt{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)'\cdot\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)=$$

$$-\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_K(W_j)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)+$$

$$-\sqrt{n}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)'-\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\cdot$$

$$\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)$$

Now the second term on the right-hand side is bounded by

$$\sqrt{n}\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)-\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)\right|\cdot$$

$$\left|\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right|\leq \tag{60}$$

$$\sqrt{n}\left(\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right]\right|\right.$$

$$\left.+\left|\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)-\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]\right|\right)\cdot$$

$$\left|\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right|$$

where we have used that

$$\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right]=\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]$$

For the double sum in the first term of the upper bound define

$$d(x,v)=\frac{\partial R}{\partial \tau}(x;\tau_0)Q_K(x,v)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right]$$

and note that if $i\neq j$

$$\mathbb{E}[d(X_i,V_j)]=0$$

Consider

$$\mathbb{E}\left[\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}d(X_i,V_j)\right|^2\right]=\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}[d(X_i,V_j)'d(X_k,V_l)] \tag{61}$$

To bound this expectation we partition the values taken by the indices i, j, k, l in subsets where none, one, two, three or all four are equal. If the indices are such that $i \neq j \neq k \neq l$ (there are $n(n-1)(n-2)(n-3)$ such terms), then $\mathbb{E}[d(X_i, V_j)d(X_k, V_l)] = 0$. If $i = k, j \neq i, k \neq l, j \neq l$ (with $n(n-1)(n-2)$ terms) and $j = l, i \neq j, k \neq l, i \neq k$ (also $n(n-1)(n-2)$ terms) then by Assumption 4

$$\mathbb{E}[d(X_i, V_j)' d(X_i, V_l)] = \mathbb{E}_X \left[\left(\frac{\partial R}{\partial \tau}(X; \tau_0) \right)^2 \mathbb{E}_V \left[Q_K(X, V) \right]' \mathbb{E}_V \left[Q_K(X, V) \right] \right] -$$

$$\left(\mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] \right)' \left(\mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] \right) \leq$$

$$\mathbb{E}_X \left[\left(\frac{\partial R}{\partial \tau}(X; \tau_0) \right)^2 \mathbb{E}_V \left[|Q_K(X, V)| \right]' \mathbb{E}_V \left[|Q_K(X, V)| \right] \right] \leq \left(\sup_{w \in \mathcal{W}} |Q_K(w)| \right)^2 \mathbb{E}\left[N_1(X)^2 \right] = \zeta_0(K)^2 = O(K^2)$$

and

$$\mathbb{E}_V \left[\mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) Q_K(X, V) \right]' \mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) Q_K(X, V) \right] \right] -$$

$$\left(\mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] \right)' \left(\mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] \right) \leq$$

$$\mathbb{E}_V \left[\mathbb{E}_X \left[\left| \frac{\partial R}{\partial \tau}(X; \tau_0) \right| |Q_K(X, V)| \right]' \mathbb{E}_X \left[\left| \frac{\partial R}{\partial \tau}(X; \tau_0) \right| |Q_K(X, V)| \right] \right]$$

$$\leq \left(\sup_{w \in \mathcal{W}} |Q_K(w)| \right)^2 \left(\mathbb{E}\left[N_1(X) \right] \right)^2 = \zeta_0(K)^2 = O(K^2)$$

respectively. The contribution of these terms to the expected value (61) is therefore $O(K^2/n)$. If $i = k, j = l, j \neq i, l \neq k$ (with $n(n-1)$ terms), then

$$\mathbb{E}[d(X_i, V_j)' d(X_i, V_j)] = \mathbb{E}_X \left[\left(\frac{\partial R}{\partial \tau}(X; \tau_0) \right)^2 \mathbb{E}_V \left[Q_K(X, V)' Q_K(X, V) \right] \right] -$$

$$\left(\mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] \right)' \left(\mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] \right) \leq$$

$$\mathbb{E}_X \left[\left(\frac{\partial R}{\partial \tau}(X; \tau_0) \right)^2 \mathbb{E}_V \left[Q_K(X, V)' Q_K(X, V) \right] \right] \leq \left(\sup_{w \in \mathcal{W}} |Q_K(w)| \right)^2 \mathbb{E}\left[N_1(X)^2 \right] = \zeta_0(K)^2 = O(K^2)$$

so that the contribution of these terms to the expected value (61) is $O(K^2/n^2)$. If $j = i, k = i, l \neq i$ (with $n(n-1)$ terms), then

$$\mathbb{E}[d(X_i, V_i)' d(X_i, V_l)] =$$

$$\mathbb{E}_{X,V} \left[\left(\frac{\partial R}{\partial \tau}(X; \tau_0) Q_K(X, V) - \mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] \right)' \cdot \right.$$

$$\left. \left(\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] - \mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right) \right] \right] =$$

$$\mathbb{E}_{X,V} \left[\left(\frac{\partial R}{\partial \tau}(X; \tau_0) \right)^2 Q_K(X, V)' \mathbb{E}_V [Q_K(X, V)] \right] - \mathbb{E}_{X,V} \left[\frac{\partial R}{\partial \tau}(X; \tau_0) Q_K(X, V) \right]' \mathbb{E}_X \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \mathbb{E}_V [Q_K(X, V)] \right] =$$

$$\mathbb{E}_{X,V}\left[\left(\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)-\mathbb{E}_{X,V}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]\right)'\cdot\right.$$
$$\left.\left(\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right)\right]\right]$$

By Cauchy-Schwartz the absolute value of the final expression is bounded by

$$\sqrt{\mathbb{E}_{X,V}\left[\left|\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)-\mathbb{E}_{X,V}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]\right|^2\right]}\cdot$$

$$\sqrt{\mathbb{E}_{X}\left[\left|\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]-\mathbb{E}_{X}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right|^2\right]\right]}\leq$$

$$\sqrt{\mathbb{E}_{X,V}\left[\left|\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right|^2\right]}\sqrt{\mathbb{E}_{X}\left[\left|\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right|^2\right]}\leq\left(\sup_{w\in\mathcal{W}}|Q_K(w)|\right)^2\mathbb{E}\left[N_1(X)^2\right]$$

where we use Assumption 4. We conclude that

$$\mathbb{E}[d(X_i,V_i)'d(X_i,V_l)]=O(K^2)$$

and the same argument shows that for $j=i,k\neq i,l=i$ (with $n(n-1)$ terms) $\mathbb{E}[d(X_i,V_i)'d(X_k,V_i)]=O(K^2)$, that for $j\neq i,k=i,l=i$ (with $n(n-1)$ terms) $\mathbb{E}[d(X_i,V_j)'d(X_i,V_i)]=O(K^2)$, and that for $i\neq j,k=j,l=j$ (with $n(n-1)$ terms) $\mathbb{E}[d(X_i,V_j)'d(X_j,V_j)]=O(K^2)$. The contribution of these terms to the expectation in (61) is therefore $O(K^2/n^2)$. Finally if $i=j=k=l$

$$\mathbb{E}[d(X_i,V_i)'d(X_i,V_i)]=\mathbb{E}_{X,V}\left[\left(\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right]\right)'\cdot\right.$$
$$\left.\left(\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right]\right)\right]\leq$$

$$\mathbb{E}[d(X_i,V_i)'d(X_i,V_i)]=\mathbb{E}_{X,V}\left[\left(\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)-\mathbb{E}_{X,V}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]\right)'\cdot\right.$$
$$\left.\left(\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)-\mathbb{E}_{X,V}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]\right)\right]$$

and by Cauchy-Schwartz and Assumption 4 we find that the bound is $O(K^2)$ so that the contribution to the expected value is $O(K^2/n^3)$.

We conclude that the expectation in (61) is $O(K^2/n)$. Together with

$$\left|\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial\mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right|\leq\sup_{w\in\mathcal{W}}|Q_K(w)|\sup_{w\in\mathcal{W}}\left|\frac{\partial\mu}{\partial v}(w)\right|\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j-V_j\right)^2}=O_p(Kn^{-\delta})\tag{62}$$

by Assumption 3 and 6 this implies that the upper bound on (60) is

$$\sqrt{n}\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V[Q_K(X,V)]\right]\right|\cdot$$

$$\left|\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right|=O_p\left(K^2n^{-\delta}\right)$$

Also by Assumption 4 and Assumption 5 that ensures that $f(x,v)$ is bounded from 0

$$\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)-\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]\right|\cdot$$

$$\left|\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right|=O_p\left(K^2n^{-\delta}\right)$$

because

$$\mathbb{E}\left[\frac{f(X)^2f(V)^2}{f(X,V)^2}\left(\frac{\partial R}{\partial \tau}(X;\tau_0\right)^2|Q_K(X,V)|^2\right]=O(K^2) \tag{63}$$

Combining the results we conclude that (60) is $O_p\left(K^2n^{-\delta}\right)$. Therefore (58) is equal to

$$-\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)+O_p\left(K^2n^{-\delta}\right)$$

Now

$$-\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)=$$

$$-\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\tilde{\Omega}_K^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)+$$

$$-\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\tilde{\Omega}_K^{-1}\right)\left(\tilde{\Omega}_K-I_K\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)=$$

$$-\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\tilde{\Omega}_K^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)+O_p\left(K^{7/2}n^{-\delta}\right)$$

because to bound the third displayed equation we increase the bound on the first factor as in the proof of Lemma 13 to remove $\tilde{\Omega}_K^{-1}$ and use the bound on the second moment in (63), use the bound of Lemma 10 in the second factor and also the bound in (62) for the third factor to obtain

$$\sqrt{n}\left|\tilde{\Omega}_K^{-1}\frac{1}{n}\sum_{i=1}^{n}Q_K(X_i,V_i)\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\left|\tilde{\Omega}_K-I_K\right|\left|\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right|\leq$$

$$2\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}Q_K(X_i,V_i)\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\left|\tilde{\Omega}_K-I_K\right|\left|\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right|=O_p\left(K^{7/2}n^{-\delta}\right)$$

Now note that

$$\tilde{\delta}_K=\tilde{\Omega}_K^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)\right)$$

is the regression estimator of $\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)$ on $Q_K(X,V)$. Therefore

$$-\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\tilde{\Omega}_K^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)=$$

$$-\sqrt{n}\left(\tilde{\delta}_K'\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)=-\sqrt{n}\left(\delta_K'\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)-$$

$$\sqrt{n}\left((\tilde{\delta}_K-\delta_K)'\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)=-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta_K'Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)+$$

$$O_p\left(K^{3/2}n^{-\delta}\right)+O_p\left(K^{2-a_D}n^{1/2-\delta}\right)$$

because

$$\left|\sqrt{n}(\tilde{\delta}_K-\delta_K)'\left(\frac{1}{n}\sum_{j=1}^{n}Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)\right|=\sqrt{n}\left(O_p(K^{1/2}n^{-1/2})+O_p(K^{1-a_D})\right)O_p(Kn^{-\delta})$$

by Assumption 3 and 6 and Lemma 13. Finally we obtain

$$-\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta_K'Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)=$$

$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)+$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)-\delta_K'Q_K(W_j)\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\right)$$

where the final term is $O_p\left(K^{-a_D}n^{1/2-\delta}\right)$ by Assumption 3 and 6. Combining all remainders we find that (58) is equal to

$$-\frac{1}{n}\sum_{j=1}^{n}\delta_K'Q_K\left(W_j\right)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)+O_p\left(K^{2}n^{-\delta}\right)+O_p\left(K^{7/2}n^{-\delta}\right)+O_p\left(K^{3/2}n^{-\delta}\right)+O_p\left(K^{2-a_D}n^{1/2-\delta}\right)+$$

$$O_p\left(K^{-a_D}n^{1/2-\delta}\right)$$

Therefore the final result is that keeping the largest remainders (58) is equal to

$$
\begin{aligned}
&-\frac{1}{n}\sum_{j=1}^{n}\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\frac{\partial \mu}{\partial v}(W_j)\left(\hat{V}_j-V_j\right)\\
&+O_p\left(K^{7/2}n^{-\delta}\right)+O_p\left(K^{2-a_D}n^{1/2-\delta}\right)
\end{aligned}
$$

■

Lemma 17 *If Assumptions 3, 4, 5, 6, 7 hold and $K^3 n^{-\delta}\to 0$, then (59) is equal to*

$$
\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\right)U_j\left(\hat{V}_j-V_j\right)+O_p\left(K^{11/2}n^{-\delta}\right)+O_p\left(K^{4-a_D}n^{1/2-\delta}\right)
$$

Proof. For (59)

$$
\sqrt{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)'\cdot\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j-V_j\right)=
$$

$$
\begin{aligned}
&\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j-V_j\right)\right)\\
&+\sqrt{n}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)'-\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\cdot\\
&\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j-V_j\right)\right)
\end{aligned}
$$

Now the third displayed expression is bounded by

$$
\begin{aligned}
&\sqrt{n}\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)-\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)\right|\cdot\\
&\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j-V_j\right)\right|\leq \qquad (64)\\
&\sqrt{n}\left(\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_j)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_V\left[Q_K(X,V)\right]\right]\right|\right.\\
&\left.+\left|\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)-\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)Q_K(X,V)\right]\right|\right)\cdot\\
&\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j-V_j\right)\right|
\end{aligned}
$$

This is equal to the bound in (60) except that the second factor in the upper bound is now

$$\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right| \leq \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left|\frac{\partial Q_K}{\partial v}(W_j)\right|^2 U_j^2}\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j - V_j\right)^2} = O_p\left(\zeta_1(K)n^{-\delta}\right) = O_p\left(K^3 n^{-\delta}\right)$$

by Cauchy-Schwartz and Assumptions 3 and 7. We conclude that (64) is $O_p\left(K^4 n^{-\delta}\right)$, so that (59) is equal to

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) + O_p\left(K^4 n^{-\delta}\right)$$

Now consider

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) =$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\tilde{\Omega}_K^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) +$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\tilde{\Omega}_K^{-1}\right)\left(\tilde{\Omega}_K - I_K\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) =$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\tilde{\Omega}_K^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) + O_p\left(K^{11/2}n^{-\delta}\right)$$

where the bound on the third displayed expression is obtained as in the previous lemma with the only difference that the third factor in the expression is now $O_p\left(K^3 n^{-\delta}\right)$ instead of $O_p\left(K n^{-\delta}\right)$. Denote by $\tilde{\delta}_K$ the same unfeasible regression estimator as in the previous lemma. Then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_K(X_i,V_i)'\right)\tilde{\Omega}_K^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) =$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\tilde{\delta}_K'\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\delta_K'\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) +$$

$$\sqrt{n}\left((\tilde{\delta}_K - \delta_K)'\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) =$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\delta_K'\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j - V_j\right)\right) + O_p\left(K^{7/2}n^{-\delta}\right) + O_p\left(K^{4-a_D}n^{1/2-\delta}\right)$$

using the same argument as in the previous lemma with the noted change that implies that the bound in the previous lemma is multiplied by K^2. Finally

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\delta_K'\frac{\partial Q_K}{\partial v}(W_j)U_j\left(\hat{V}_j-V_j\right)\right)=$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\right)U_j\left(\hat{V}_j-V_j\right)\right)+$$

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\right)-\delta_K'\frac{\partial Q_K}{\partial v}(W_j)\right)U_j\left(\hat{V}_j-V_j\right)\right)$$

where the final term is bounded by

$$\left|\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\right)-\delta_K'\frac{\partial Q_K}{\partial v}(W_j)\right)U_j\left(\hat{V}_j-V_j\right)\right)\right|\leq$$

$$\sqrt{n}\sup_{w\in\mathcal{W}}\left|\frac{\partial}{\partial v}\left(\frac{f(x)f(v)}{f(x,v)}\frac{\partial R}{\partial \tau}(x;\tau_0)\right)-\delta_K'\frac{\partial Q_K}{\partial v}(w)\right|\sqrt{\frac{1}{n}\sum_{j=1}^{n}U_j^2}\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{V}_j-V_j\right)^2}=O_p\left(K^{-a_D}n^{1/2-\delta}\right)$$

Therefore the final result is that (59) is equal to (if we keep the remainders of the largest order)

$$\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\right)U_j\left(\hat{V}_j-V_j\right)\right)$$
$$+O_p\left(K^{11/2}n^{-\delta}\right)+O_p\left(K^{4-a_D}n^{1/2-\delta}\right)$$

■

Combining this with the result on (50) we have

Lemma 18 *If Assumptions 3, 4, 5, 6, 7 hold and* $K^3n^{-\delta}\to 0$ *, then by Lemmas 15,16, 17 and equations (58) and (59)*

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{\mu}_K(X_i,\hat{V}_j)-\tilde{\mu}_K(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau)=$$

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\begin{array}{c}\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}(X,V_j)\right]-\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\frac{\partial \mu}{\partial v}(W_j)\\+\frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\right)U_j\end{array}\right)\left(\hat{V}_j-V_j\right)$$
$$+O_p\left(K^{11/2}n^{-\delta}\right)+O_p\left(K^7n^{1/2-2\delta}\right)+O_p(K^6n^{1/2+1/m-2\delta})+O_p\left(K^{2-a_D}n^{1/2}\right)$$

If $\delta\leq 1/2$ *then the remainder is* $O_p\left(K^7n^{1/2-2\delta}\right)+O_p(K^6n^{1/2+1/m-2\delta})+O_p\left(K^{2-a_D}n^{1/2}\right)$.

J.4 Expressing (45) as a sample average

We rewrite (45) as

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\tilde{\mu}_K(X_i,V_j)-\mu(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)=$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(Q_K(X_i,V_j)'\gamma_K-\mu(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)+ \tag{65}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(Q_K(X_i,V_j)-Q_K(X_i,V_i)\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\right)'(\tilde{\gamma}_K-\gamma_K)\frac{\partial R}{\partial \tau}(X_i;\tau_0)+ \tag{66}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i-\mu(X_i,V_i))\frac{\partial R}{\partial \tau}(X_i;\tau_0)+ \tag{67}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(\mu(X_i,V_i)-Q_K(X_i,V_i)'\gamma_K)\frac{\partial R}{\partial \tau}(X_i;\tau_0)- \tag{68}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i-Q_K(X_i,V_i)'\tilde{\gamma}_K)\frac{\partial R}{\partial \tau}(X_i;\tau_0) \tag{69}$$

with (67) the main term. The remainder (65) is bounded by

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(Q_K(X_i,V_j)'\gamma_K-\mu(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|$$
$$\leq \sqrt{n}\sup_{w\in\mathcal{W}}\left|Q_K(w)'\gamma_K-\mu(w)\right|\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)=O_p\left(n^{1/2}K^{-a_D}\right)$$

The remainder (66) is by Assumptions 4 and 6 bounded by

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(Q_K(X_i,V_j)-Q_K(X_i,V_i)\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\right)\right||\tilde{\gamma}_K-\gamma_K|$$
$$=O_p(K)\left(O_p(K^{1/2}n^{-1/2})+O_p(K^{1-a_D})\right)$$
$$=O_p\left(K^{3/2}n^{-1/2}\right)+O_p\left(K^{2-a_D}\right)$$

using the bound $O_p(K)$ on the first factor derived in Lemma 16 (see equation (60)) and we use the bound on the second factor in Lemma 13. By Assumptions 4 and 6, (68) is bounded by

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(\mu(X_i,V_i)-Q_K(X_i,V_i)'\gamma_K)\frac{\partial R}{\partial \tau}(X_i;\tau_0)\right|\leq$$

$$\sqrt{n}\sup_{w\in\mathcal{W}}|\mu(w)-Q_K(w)'\gamma_K|\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)=O_p(n^{1/2}K^{-a_D})$$

To bound (69) we note that the residual $Y_i - Q_K(X_i,V_i)'\tilde{\gamma}_K$ is uncorrelated with $Q_K(X_i,V_i)$. Therefore (69) is equal to (where δ_K is chosen as in Assumption 6)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - Q_K(X_i,V_i)'\tilde{\gamma}_K)\left(\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0) - Q_K(X_i,V_i)'\delta_K\right) =$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i - (\tilde{\mu}_K(W_i) - \mu_K(W_i)))\left(\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0) - Q_K(X_i,V_i)'\delta_K\right)$$

By Assumption 6 and 7 and Lemma 13 the final expression is bounded by

$$\sqrt{n}\sup_{w\in\mathcal{W}}\left|\frac{f(x)f(v)}{f(x,v)}\frac{\partial R}{\partial \tau}(x;\tau_0) - Q_K(w)'\delta_K\right|\left(\frac{1}{n}\sum_{i=1}^{n}|U_i| + \frac{1}{n}\sum_{i=1}^{n}|\tilde{\mu}_K(W_i) - \mu(W_i)|\right)$$

$$= \sqrt{n}O_p\left(K^{-a_D}\right)\left(O_p(1) + O_p(K^{3/2}n^{-1/2}) + O_p(K^{2-a_D})\right) = O_p\left(K^{-a_D}n^{1/2}\right) + O_p\left(K^{3/2-a_D}\right) + O_p\left(K^{2-2a_D}n^{1/2}\right)$$

Lemma 19 *If Assumptions 4, 5, 6 and 7 hold, then*

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\tilde{\mu}_K(X_i,V_j) - \mu(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i - \mu(X_i,V_i))\frac{\partial R}{\partial \tau}(X_i;\tau_0) +$$

$$O_p\left(K^{-a_D}n^{1/2}\right) + O_p\left(K^{3/2}n^{-1/2}\right) + O_p\left(K^{2-a_D}\right) + O_p\left(K^{2-2a_D}n^{1/2}\right)$$

J.5 Expressing (46) as a sample average

Upon substitution of $R(X_i;\tau_0)$ for $L(X_i)$ we obtain for (46)

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu(X_i,V_j)\frac{\partial R}{\partial \tau}(X_i;\tau_0) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}R(X_i;\tau_0)\frac{\partial R}{\partial \tau}(X_i;\tau_0)$$

Because

$$\mathbb{E}\left[R(X;\tau_0)\frac{\partial R}{\partial \tau}(X;\tau_0)\right] = \mathbb{E}_V\left[\mathbb{E}_X\left[\mu(X,V)\frac{\partial R}{\partial \tau}(X;\tau_0)\right]\right]$$

we can express (46) as

$$\left(\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu(X_i,V_j)\frac{\partial R}{\partial \tau}(X_i;\tau_0) - \mathbb{E}_V\left[\mathbb{E}_X\left[\mu(X,V)\frac{\partial R}{\partial \tau}(X;\tau_0)\right]\right]\right) -$$

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}R(X_i;\tau_0)\frac{\partial R}{\partial \tau}(X_i;\tau_0) - \mathbb{E}\left[R(X;\tau_0)\frac{\partial R}{\partial \tau}(X;\tau_0)\right]\right)$$

The first term is a two-sample generalized U statistic with projection

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}R(X_i;\tau_0)\frac{\partial R}{\partial \tau}(X_i;\tau_0) - \mathbb{E}\left[R(X;\tau_0)\frac{\partial R}{\partial \tau}(X;\tau_0)\right] +$$

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbb{E}_X\left[\mu(X,V_j)\frac{\partial R}{\partial \tau}(X;\tau_0)\right] - \mathbb{E}_V\left[\mathbb{E}_X\left[\mu(X,V)\frac{\partial R}{\partial \tau}(X;\tau_0)\right]\right] + o_p(1)$$

where we use $\mathbb{E}_V[\mu(x,V)] = R(x;\tau_0)$ so that

Lemma 20 *Under Assumptions 4 and 6*

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu(X_i,V_j)\frac{\partial R}{\partial\tau}(X_i;\tau_0)-\frac{1}{n}\sum_{i=1}^{n}L(X_i)\frac{\partial R}{\partial\tau}(X_i;\tau_0)=$$

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbb{E}_X\left[\mu(X,V_j)\frac{\partial R}{\partial\tau}(X;\tau_0)\right]-\mathbb{E}_V\left[\mathbb{E}_X\left[\mu(X,V)\frac{\partial R}{\partial\tau}(X;\tau_0)\right]\right]+o_p(1)$$

J.6 The asymptotic distribution

The results in Lemmas 18, 19, and 20 give

Theorem 9 (Asymptotically linear representation) *If $a_D > 2$, $\frac{1}{4} < \delta < \frac{1}{2}$, and we choose $K = n^b$ such that*

$$\frac{1}{2(a_D-2)} < b < \frac{2}{7}\delta - \frac{1}{14}$$

where δ and a_D satisfy

$$\frac{1}{a_D-2} < \frac{4}{7}\delta - \frac{1}{7}$$

If assumptions 3, 4, 5, 6 and 7 hold, then

$$\sqrt{n}(\hat{\tau}-\tau_0) = \left(\mathbb{E}\left[\frac{\partial R}{\partial\tau}(X;\tau_0)\frac{\partial R}{\partial\tau'}(X;\tau_0)\right]\right)^{-1}\cdot$$

$$\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\begin{pmatrix}\mathbb{E}_X\left[\frac{\partial R}{\partial\tau}(X;\tau_0)\frac{\partial\mu}{\partial v}(X,V_j)\right]-\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial\tau}(X_j;\tau_0)\frac{\partial\mu}{\partial v}(W_j)\\ +\frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial\tau}(X_j;\tau_0)\right)U_j\end{pmatrix}\left(\hat{V}_j-V_j\right)\right.$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i-\mu(X_i,V_i))\frac{\partial R}{\partial\tau}(X_i;\tau_0)+$$

$$\left.\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\mathbb{E}_X\left[\mu(X,V_j)\frac{\partial R}{\partial\tau}(X;\tau_0)\right]-\mathbb{E}_V\left[\mathbb{E}_X\left[\mu(X,V)\frac{\partial R}{\partial\tau}(X;\tau_0)\right]\right]\right)\right\}+o_p(1)$$

Proof. The relevant remainder terms are

$$O_p\left(K^{11/2}n^{-\delta}\right)+O_p\left(K^7n^{1/2-2\delta}\right)+O_p(K^6n^{1/2+1/m-2\delta})+O_p\left(K^{2-a_D}n^{1/2}\right)+O_p\left(K^{3/2}n^{-1/2}\right)$$

First $a_D > 2$ is required to make the third remainder vanish. Second, the relevant range for δ is between $\frac{1}{4}+\frac{1}{2m}$ and $\frac{1}{2}$. The lower bound derives from the third remainder and is larger by $1/m$ than the well-known rate restriction in Newey (1994). For $K = n^b$ we have the inequalities $b < \frac{2}{11}\delta$, $b < \min\left\{\frac{2}{7}\delta - \frac{1}{14}, \frac{1}{3}\delta - \frac{1}{12} - \frac{1}{6m}\right\}$ and $b < \frac{1}{3}$ with only the second inequality relevant on the relevant range of δ. Finally we have $b > \frac{1}{2(a_D-2)}$. Also the upper bound on b ensures that $K^3n^{-\delta} \to 0$. To ensure that the lower bound on b is not larger than the upper bound we require that

$$\frac{1}{a_D-2} < \frac{4}{7}\delta - \frac{1}{7}$$

■

K Asymptotics for the control variable V_j

K.1 The asymptotic distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^n S_j\left(\tilde{V}_j - V_j\right)$

Theorem 9 shows that the asymptotic distribution of $\hat{\tau}$ is a constant matrix times $\frac{1}{\sqrt{n}}\sum_{i=1}^n S_j\left(\hat{V}_j - V_j\right)$ with

$$\begin{aligned}S_j &= \mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}(X,V_j)\right] - \frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\frac{\partial \mu}{\partial v}(X_j,V_j)\\ &+ \frac{\partial}{\partial v}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\right)U_j\end{aligned}$$

We will derive an asymptotically linear representation of the distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^n S_j\left(\tilde{V}_j - V_j\right)$ where

$$\begin{aligned}\tilde{V}_j &= p_L(Z_j)'\,\tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^n p_L(Z_i)\,1(X_i \le X_j)\right)\\ \tilde{\Psi}_L &= \frac{1}{n}\sum_{j=1}^n p_L(Z_j)\,p_L(Z_j)'\end{aligned}$$

under the implicit assumption that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n S_j\left(\hat{V}_j - V_j\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n S_j\left(\tilde{V}_j - V_j\right) + o_p(1). \tag{70}$$

where $\hat{V}_j$ is a transformation of the series estimator $\tilde{V}_j$ that is in the $[0,1]$ interval. For simplicity of notation, we will assume that S is a scalar although the derivation below can be easily modified to the case that S is a vector.

Define

$$\tilde{\alpha}_L(x) = \tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^n p_L(Z_i)\,1(X_i \le x)\right)$$

then

$$\tilde{V}_j = p_L(Z_j)'\,\tilde{\alpha}_L(X_j)$$

is the series estimator of $F(X_j|Z_j)$ with the L vector of basis function $P_L(Z)$. We make a support assumption as in Assumption 5, i.e.

Assumption 10 *The support of the distribution of Z is a rectangle.*

Under this assumption we have as before

$$\zeta_d(L) = \max_{|\lambda|\le d}\sup_z \left\|\frac{\partial^d p_L}{\partial z^d}(z)\right\| = O\left(L^{2d+1}\right).$$

As before we assume the bound on $\zeta_d(L)$ and do not mention Assumption 10 again. Without loss of generality, we assume $\mathbb{E}\left[p_L(Z_j)\, p_L(Z_j)'\right] = I_L$ as we did before. The implication is that

$$\left\|\tilde{\Psi}_L - I_L\right\| = O_p\left(\zeta_0(L)\, L^{1/2} n^{-1/2}\right) = O_p\left(L^{3/2} n^{-1/2}\right)$$

by Lemma 10. As we did in Assumption 6 we bound the approximation error. Following Imbens and Newey (2002, Assumption 5.1), we assume that

Assumption 11 *There exists a vector $\alpha_L(x)$ and constants $C_D, b_D > 0$ such that*

$$\sup_{x,z}\left|F(x|\, z) - p_L(z)'\, \alpha_L(x)\right| \le C_D L^{-b_D} \tag{71}$$

In addition there exist vectors $\beta_L(x), \beta_L^*(z_1)$ such that

$$\sup_{x,z}\left|\chi(x,z) - p_L(z)'\, \zeta_L(x)\right| \le O\left(L^{-b_D}\right)$$

$$\sup_{\tilde{z},z}\left|\chi^*(\tilde{z},z) - p_L(z)'\, \zeta_L^*(\tilde{z})\right| \le O\left(L^{-b_D}\right)$$

with χ and χ^* defined below.

As in the proof of Lemma 13 we have

$$\tilde{\alpha}_L(x) - \alpha_L(x) = \tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^n p_L(Z_i)\left(1(X_i \le x) - p_L(Z_i)'\, \alpha_L(x)\right)\right)$$

where we omit the indicator $\tilde{1}_n$ of the event that $\tilde{\Psi}_L$ is positive definite under the assumption that $L^{3/2}n^{-1/2} \to 0$. In the rest of this proof we assume that this limit holds, so that we can assume that $\tilde{\Psi}_L$ is positive definite.

We now write

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n S_j\left(\tilde{V}_j - V_j\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^n S_j p_L(Z_j)'\, \tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^n p_L(Z_i)\left(1(X_i \le X_j) - F(X_j|\, Z_i)\right)\right) \tag{72}$$

$$+ \frac{1}{\sqrt{n}}\sum_{j=1}^n S_j p_L(Z_j)'\, \tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^n p_L(Z_i)\left(F(X_j|\, Z_i) - p_L(Z_i)'\, \alpha_L(X_j)\right)\right) \tag{73}$$

$$+ \frac{1}{\sqrt{n}}\sum_{j=1}^n S_j\left(p_L(Z_j)'\, \alpha_L(X_j) - F(X_j|\, Z_j)\right). \tag{74}$$

The first term (72) is the main term and we bound the other terms. First the second term (73) is bounded by, using that if $L^{3/2}n^{-1/2} \to 0$ then $4I_L - (\tilde{\Psi}_L^{-1})^2$ is positive definite,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \left|p_L(Z_j)'\, \tilde{\Psi}_L^{-1} p_L(Z_j)\right|\left(\frac{1}{n}\sum_{i=1}^n |p_L(Z_i)|\left|\left(F(X_j|\, Z_i) - p_L(Z_i)'\, \alpha_L(X_j)\right|\right)\right)|S_j|$$

$$\le 2\zeta_0(L)^2 O(L^{-b_D}) \cdot \frac{1}{\sqrt{n}}\sum_{j=1}^n |S_j|$$

$$= O_p\left(n^{1/2} L^{2-b_D}\right).$$

by Assumption 11 and because Assumptions 4 and 6 ensure that $\mathbb{E}(|S|^2) < \infty$. The last term (74) is

$$O\left(n^{1/2}L^{-b_D}\right).$$

by Assumption 11 and because $\mathbb{E}(|S|^2) < \infty$. Next, we consider the main term in (72)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} p_L(Z_i)' \tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\left(1\left(X_i \le X_j\right) - F\left(X_j|\, Z_i\right)\right)S_j\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} p_L(Z_i)' \tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\left(\begin{array}{c} 1\left(X_i \le X_j\right)S_j - \chi\left(X_i, Z_j\right) \\ -F\left(X_j|\, Z_i\right)S_j + \chi^*\left(Z_i, Z_j\right)\end{array}\right)\right) \tag{75}$$

$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} p_L(Z_i)' \tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\left(\chi\left(X_i, Z_j\right) - \chi^*\left(Z_i, Z_j\right)\right)\right), \tag{76}$$

where we define

$$\chi\left(x_i, z_j\right) = E\left[1\left(x_i \le X\right)S|\, Z = z_j\right]$$
$$\chi^*\left(z_i, z_j\right) = E\left[F\left(X|\, z_i\right)S|\, Z = z_j\right]$$

In the first line the order of summation is interchanged. Therefore in (75)

$$\tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\, 1\left(X_i \le X_j\right)S_j\right)$$

is the coefficient estimator in the nonparametric regression of $1\left(X_i \le X_j\right)S_j$ on Z_j with regression function $\mathbb{E}_{X,S}\left[1\left(X_i \le X\right)S|Z_j\right] = \chi(X_i, Z_j)$. In the same way

$$\tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\, F\left(X_j|\, Z_i\right)S_j\right)$$

is the coefficient estimator in the nonparametric regression of $F\left(X_j|\, Z_i\right)S_j$ on Z_j with regression function $\mathbb{E}\left[F\left(X|\, Z_i\right)S|Z_j\right] = \chi^*(Z_i, Z_j)$. The terms in parentheses in (75)

$$\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\left(1\left(X_i \le X_j\right)S_j - \chi\left(X_i, Z_j\right)\right)$$

and

$$\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\left(F\left(X_j|\, Z_i\right)S_j - \chi^*\left(Z_i, Z_j\right)\right)$$

are therefore covariances of the regressors $P_L(Z_j)$ and errors. We have for (75)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} P_L(Z_i)'\tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)\left(1\left(X_i \le X_j\right)S_j - \chi(X_i, Z_j) - F\left(X_j|\, Z_i\right)S_j + \chi^*(Z_i, Z_j)\right)\right) =$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right)\left(1\left(X_i \le X_j\right)S_j - \chi(X_i,Z_j) - F\left(X_j\right|Z_i\right)S_j + \chi^*(Z_i,Z_j)) =$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n} P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_i\right)\left(1\left(X_i \le X_i\right)S_i - \chi(X_i,Z_i) - F\left(X_i\right|Z_i\right)S_i + \chi^*(Z_i,Z_i)) + \tag{77}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right)\left(1\left(X_i \le X_j\right)S_j - \chi(X_i,Z_j) - F\left(X_j\right|Z_i\right)S_j + \chi^*(Z_i,Z_j)) \tag{78}$$

For (77)

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^{n} P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_i\right)\left(1\left(X_i \le X_i\right)S_i - \chi(X_i,Z_i) - F\left(X_i\right|Z_i\right)S_i + \chi^*(Z_i,Z_i))\right| \le$$

$$4\sup_{z\in\mathcal{Z}}|P_L(z)|^2 n^{-1/2}\frac{1}{n}\sum_{i=1}^{n}(2\mathbb{E}(|S_i|) + 2\mathbb{E}(|S||Z_i)) = \zeta_0(L)^2 n^{-1/2}O_p(1) = O_p(L^2 n^{-1/2})$$

For the double sum in (78) we condition on $Z_1,\ldots,Z_n$ so that we can consider $P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right)$ as constants. Define $h(W_i,W_j) = P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right)\left(1\left(X_i \le X_j\right)S_j - \chi(X_i,Z_j)\right) - F\left(X_j\right|Z_i\left)S_j + \chi^*(Z_i,Z_j)$. Then

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right)\left(1\left(X_i \le X_j\right)S_j - \chi(X_i,Z_j) - F\left(X_j\right|Z_i\right)S_j + \chi^*(Z_i,Z_j)) =$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} h(W_i,W_j) = \frac{n-1}{2\sqrt{n}}\frac{1}{\binom{n}{2}}\sum_{i<j}(h(W_i,W_j) + h(W_j,W_i))$$

Define

$$U_n = \frac{1}{\binom{n}{2}}\sum_{i<j}(h(W_i,W_j) + h(W_j,W_i))$$

We will find a bound on $\mathbb{V}[U_n|Z_1,\ldots,Z_n]$. The statistic U_n is a U-statistic with kernel $\overline{h}(W_i,W_j) = h(W_i,W_j)+h(W_j,W_i)$. However the conditioning implies that $\overline{h}(W_i,W_j)|Z_1,\ldots,Z_n$ are not identically distributed. However we will show that the argument that gives the Hoeffding (1948) formula for the variance of a U-statistic still applies (conditionally). As a first step we compute some expectations

$$\mathbb{E}_{W_i,W_j}[1(X_i \le X_j)S_j|Z_1,\ldots,Z_n] = \mathbb{E}_{X_j,S_j}[F(X_j|Z_i)S_j|Z_i,Z_j] = \chi^*(Z_i,Z_j) \tag{79}$$

$$\mathbb{E}_{W_i,W_j}[\chi(X_i,Z_j)|Z_1,\ldots,Z_n] = \mathbb{E}_{X_i}\left[\mathbb{E}_{X,S}[1(X_i \le X)S|Z=Z_j]|Z_1,\ldots,Z_n\right] = \chi^*(Z_i,Z_j) \tag{80}$$

$$\mathbb{E}_{W_i,W_j}[F(X_j|Z_i)S_j|Z_1,\ldots,Z_n] = \mathbb{E}_{X_j,S_j}[F(X_j|Z_i)S_j|Z_j] = \chi^*(Z_i,Z_j) \tag{81}$$

$$\mathbb{E}_{W_i,W_j}[\chi^*(Z_i,Z_j)|Z_1,\ldots,Z_n] = \chi^*(Z_i,Z_j) \tag{82}$$

$$\tag{83}$$

and in the same way

$$\mathbb{E}_{W_i,W_j}[1(X_j \leq X_i)S_i|Z_1,\ldots,Z_n] = \chi^*(Z_j,Z_i) \tag{84}$$
$$\mathbb{E}_{W_i,W_j}[\chi(X_j,Z_i)|Z_1,\ldots,Z_n] = \chi^*(Z_j,Z_i) \tag{85}$$
$$\mathbb{E}_{W_i,W_j}[F(X_i|Z_j)S_i|Z_1,\ldots,Z_n] = \chi^*(Z_j,Z_i) \tag{86}$$
$$\mathbb{E}_{W_i,W_j}[\chi^*(Z_j,Z_i)|Z_1,\ldots,Z_n] = \chi^*(Z_j,Z_i) \tag{87}$$
$$\tag{88}$$

where in (80) we interchange the order of the expectations which is allowed by the Fubini Theorem. This implies first that

$$\mathbb{E}[\overline{h}(W_i,W_j)|Z_1,\ldots,Z_n] = 0$$

so that

$$\mathbb{V}(U_n) = \mathbb{E}\left[\mathbb{E}[U_n^2|Z_1,\ldots,Z_n]\right]$$

Now

$$U_n^2 = \left[\frac{1}{\binom{n}{2}}\sum_{i<j}\overline{h}(W_i,W_j)\right]^2 = \frac{1}{\binom{n}{2}^2}\sum_{i<j}\sum_{k<l}\overline{h}(W_i,W_j)\overline{h}(W_k,W_l)$$

Consider $\mathbb{E}[\overline{h}(W_i,W_j)\overline{h}(W_k,W_l)|Z_1,\ldots,Z_n]$. If the indices i,j,k,l are all different then

$$\mathbb{E}[\overline{h}(W_i,W_j)\overline{h}(W_k,W_l)|Z_1,\ldots,Z_n] = 0$$

If $i = k$ and $j \neq l$ then

$$\mathbb{E}[\overline{h}(W_i,W_j)\overline{h}(W_i,W_l)|Z_1,\ldots,Z_n] = \mathbb{E}[h_1(W_i,Z_1,\ldots,Z_n)^2|Z_1,\ldots,Z_n]$$

with

$$\begin{aligned}\overline{h}_1(w_i,Z_1,\ldots Z_n) \equiv \mathbb{E}[\overline{h}(w_i,W_j)|Z_1,\ldots,Z_n] = P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L(Z_j)\,\cdot\\
[\mathbb{E}[1(x_i \leq X_j)S_j|Z_1,\ldots,Z_n] - \mathbb{E}[1(x_i \leq X)S|Z=Z_j]\\
-\mathbb{E}[F(X_j|Z_i)S_j|Z_1,\ldots,Z_n] + \mathbb{E}[F(X|Z_i)S|Z_i,Z=Z_j]\\
+\mathbb{E}[1(X_j \leq x_i)s_i|Z_1,\ldots,Z_n] - \mathbb{E}\left[\mathbb{E}[1(X_j \leq X)S|Z=Z_i]|Z_1,\ldots,Z_n\right]\\
-\mathbb{E}[F(x_i|Z_j)s_i|Z_i] + \mathbb{E}[F(X|Z_j)S|Z=Z_i]]\end{aligned}$$

The second and third line of the above equation are obviously equal to 0. The fourth and fifth lines are also 0 because

$$\mathbb{E}[1(X_j \leq x_i)s_i|Z_1,\ldots,Z_n] = \mathbb{E}[F(x_i|Z_j)s_i|Z_i]$$

and

$$\mathbb{E}[F(X|Z_j)S|Z_i] = \mathbb{E}\left[\mathbb{E}[1(X_j \leq X)S|Z_i]|Z_1,\ldots,Z_n\right]$$

Because $h_1(w_i, Z_1, \ldots, Z_n) \equiv 0$ for all $w_i, Z_1, \ldots, Z_n$, we conclude that

$$\mathbb{E}[\overline{h}(W_i, W_j)\overline{h}(W_i, W_l)|Z_1, \ldots, Z_n] = 0 \tag{89}$$

Next if If $i = l$ and $j \neq k$ we have

$$\mathbb{E}[\overline{h}(W_i, W_j)\overline{h}(W_k, W_i)|Z_1, \ldots Z_n] = \mathbb{E}[\overline{h}(W_i, W_j)\overline{h}(W_i, W_k)|Z_1, \ldots Z_n] = 0$$

by symmetry of $\overline{h}$ and the expectation in (89). In the same way symmetry implies that for $i \neq k$ and $i \neq l$

$$\mathbb{E}[\overline{h}(W_i, W_j)\overline{h}(W_k, W_j)|Z_1, \ldots Z_n] = \mathbb{E}[\overline{h}(W_i, W_j)\overline{h}(W_j, W_l)|Z_1, \ldots Z_n] = 0$$

respectively. The conclusion is that because only $\binom{n}{2}$ terms in the conditional expectation of U_n^2 given $Z_1, \ldots, Z_n$ are nonzero we have

$$\mathbb{E}\left[\mathbb{E}[U_n^2|Z_1, \ldots, Z_n]\right] = \frac{1}{\binom{n}{2}}\mathbb{E}\left[\mathbb{E}[\overline{h}(W_i, W_j)^2|Z_1, \ldots Z_n]\right]$$

Now

$$\begin{aligned}\overline{h}(W_i, W_j)^2 &= \left(P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right)\right)^2 \cdot \\ &[1\left(X_i \leq X_j\right)S_j - \chi(X_i, Z_j) - F\left(X_j \middle| Z_i\right)S_j + \chi^*(Z_i, Z_j) + \\ &1\left(X_j \leq X_i\right)S_i - \chi(X_j, Z_i) - F\left(X_i \middle| Z_j\right)S_i + \chi^*(Z_i, Z_j)]^2 \\ &\leq 4\zeta_0(L)^4 M\end{aligned}$$

because by Cauchy-Schwartz

$$\left(P_L(Z_i)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right)\right)^2 \leq p_L\left(Z_i\right)'\tilde{\Psi}_L^{-1}p_L\left(Z_i\right)p_L\left(Z_j\right)'\tilde{\Psi}_L^{-1}p_L\left(Z_j\right) \leq 4\zeta_0(L)^4 M$$

if $L^{3/2}n^{-1/2} \to 0$ and M is the sum of the terms between square brackets and $\mathbb{E}(M) = \mathbb{E}[\mathbb{E}[M|Z_1, \ldots, Z_n]] < \infty$ by Assumptions 4 and 6. Therefore we obtain the bound $\mathbb{V}(U_n) = O(L^4/n^2)$ so that (78) is $O_p(n^{-1/2}L^2)$.

The final step is to derive an asymptotically linear representation of (76). We have by Assumption

11

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}p_L\left(Z_i\right)'\tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}p_L\left(Z_j\right)\chi\left(X_i,Z_j\right)\right)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}p_L\left(Z_i\right)'\tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}p_L\left(Z_j\right)\left(\chi\left(X_i,Z_j\right)-p_L\left(Z_j\right)'\beta_L\left(X_i\right)\right)\right)\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}p_L\left(Z_i\right)'\tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}p_L\left(Z_j\right)p_L\left(Z_j\right)'\right)\beta_L\left(X_i\right)\\
&=O_p\left(L^{2-b_D}n^{1/2}\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}p_L\left(Z_i\right)'\beta_L\left(X_i\right)\\
&=O_p\left(L^{2-b_D}n^{1/2}\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi\left(X_i,Z_i\right)\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\chi\left(X_i,Z_i\right)-p_L\left(Z_i\right)'\beta_L\left(X_i\right)\right)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi\left(X_i,Z_i\right)+O_p\left(L^{2-b_D}n^{1/2}\right)+O_p\left(L^{-b_D}n^{1/2}\right)
\end{aligned}
$$

so that

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}p_L\left(Z_i\right)'\tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}p_L\left(Z_j\right)\chi\left(X_i,Z_j\right)\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi\left(X_i,Z_i\right)+O_p\left(L^{2-b_D}n^{1/2}\right)
$$

Likewise, we obtain

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}p_L\left(Z_i\right)'\tilde{\Psi}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}p_L\left(Z_j\right)\chi^*\left(Z_i,Z_j\right)\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi^*\left(Z_i,Z_i\right)+O_p\left(L^{2-b_D}n^{1/2}\right)
$$

Combining the bounds on (73), (74), (76) with the asymptotically linear representation of (76) we obtain, only keeping the largest remainders ($L^{3/2}n^{-1/2}\to 0$) the following lemma

Lemma 21 *If Assumptions 4, 6, 10 and 11 hold and* $L^{3/2}n^{-1/2}\to 0$ *then*

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{j=1}^{n}S_j\left(\tilde{V}_j-V_j\right)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\chi\left(X_i,Z_i\right)-\chi^*\left(Z_i,Z_i\right)\right)\\
&+O_p\left(L^2n^{-1/2}\right)+O_p\left(n^{1/2}L^{2-b_D}\right)
\end{aligned}
$$

K.2 Rate of convergence of $\frac{1}{n}\sum_{j=1}^{n}(\tilde{V}_j - V_j)^2$

We have

$$\begin{aligned}
\tilde{V}_j - V_j &= p_L(Z_j)' \tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} p_L(Z_i)\, 1(X_i \leq X_j)\right) - F(X_j|Z_j) = p_L(Z_j)' \tilde{\alpha}_L(X_j) - F(X_j|Z_j) \\
&= P_L(Z_j)'(\tilde{\alpha}_L(X_j) - \alpha_L(X_j)) - (F(X_j|Z_j) - P_L(Z_j)'\alpha_L(X_j)) \\
&= P_L(Z_j)'\tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} P_L(Z_i)[1(X_i \leq X_j) - P_L(Z_i)'\alpha_L(X_j)]\right) - (F(X_j|Z_j) - P_L(Z_j)'\alpha_L(X_j)) \\
&= P_L(Z_j)'\tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} P_L(Z_i)[1(X_i \leq X_j) - F(X_j|Z_i)]\right) - (F(X_j|Z_j) - P_L(Z_j)'\alpha_L(X_j)) + \\
&\quad P_L(Z_j)'\tilde{\Psi}_L^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} P_L(Z_i)[F(X_j|Z_i) - P_L(Z_i)'\alpha_L(X_j)]\right)
\end{aligned}$$

Let $\tilde{V}$ and V be the n vectors with components $\tilde{V}_j$ and V_j respectively and let P_L be the $n \times L$ matrix with rows $P_L(Z_j)'$. Further let e_j be the n vector of errors $1(X_i \leq X_j) - F(X_j|Z_i)$, u_{1L} be the n vector of approximation residuals $F(X_j|Z_j) - P_L(Z_j)'\alpha_L(X_j)$ and u_{2Lj} be the n vector of approximation residuals $F(X_j|Z_i) - P_L(Z_i)'\alpha_L(X_j)$. Then

$$\tilde{V}_j - V_j = P_L(Z_j)'(P_L'P_L)^{-1}P_L'e_j - u_{1L,j} + P_L(Z_j)'(P_L'P_L)^{-1}P_L'u_{2Lj} \tag{90}$$

so that

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}(\tilde{V}_j - V_j)^2 =& \frac{1}{n}\sum_{j=1}^{n} P_L(Z_j)'(P_L'P_L)^{-1}P_L'e_je_j'P_L(P_L'P_L)^{-1}P_L(Z_j) && (91)\\
&+ \frac{1}{n}\sum_{j=1}^{n} u_{1L,j}^2 + \frac{1}{n}\sum_{j=1}^{n}\left(P_L(Z_j)'(P_L'P_L)^{-1}P_L'u_{2Lj}\right)^2 && (92)\\
&- \frac{2}{n}\sum_{j=1}^{n} P_L(Z_j)'(P_L'P_L)^{-1}P_L'e_ju_{1L,j} && (93)\\
&+ \frac{2}{n}\sum_{j=1}^{n} P_L(Z_j)'(P_L'P_L)^{-1}P_L'e_ju_{2Lj}'P_L(P_L'P_L)^{-1}P_L(Z_j) && (94)\\
&- \frac{2}{n}\sum_{j=1}^{n} P_L(Z_j)'(P_L'P_L)^{-1}P_L'u_{2Lj}u_{1L,j} && (95)
\end{aligned}$$

We have for (91)

$$\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} P_L(Z_j)'(P_L'P_L)^{-1}P_L'e_je_j'P_L(P_L'P_L)^{-1}P_L(Z_j)\middle| Z_1,\ldots,Z_n\right) =$$

$$\frac{1}{n}\sum_{j=1}^{n} P_L(Z_j)'(P_L'P_L)^{-1}P_L'\mathbb{E}\left(e_je_j'|Z_1,\ldots,Z_n\right) P_L(P_L'P_L)^{-1}P_L(Z_j)$$

with for $i \neq i'$ and $i, i' \neq j$

$$\mathbb{E}[e_{ij}e_{i'j}|Z_1,\ldots,Z_n] = \mathbb{E}_{X_j}\left[\mathbb{E}_{X_i,X_{i'}}[(1(X_i \leq X_j) - F(X_j|Z_i))(1(X_{i'} \leq X_j) - F(X_j|Z_{i'}))|X_j, Z_1,\ldots,Z_n]\right] = 0$$

and for $i \neq j$

$$\mathbb{E}[e_{ij}^2|Z_1,\ldots,Z_n] = \mathbb{E}_{X_j}\left[\mathbb{E}_{X_i}[(1(X_i \leq X_j) - F(X_j|Z_i))^2|X_j, Z_1,\ldots,Z_n]\right] =$$

$$\mathbb{E}[(1 - F(X_j|Z_i))F(X_j|Z_i)|Z_1,\ldots,Z_n] \leq 1$$

and because

$$\mathbb{E}[e_{jj}^2|Z_1,\ldots,Z_n] = \mathbb{E}_{X_j}[(1 - F(X_j|Z_j))^2|Z_1,\ldots,Z_n] \leq 1$$

the inequality also holds if $i = j$. Therefore

$$\sup_{Z_1,\ldots,Z_n} \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^n P_L(Z_j)'(P_L'P_L)^{-1}P_L'e_je_j'P_L(P_L'P_L)^{-1}P_L(Z_j)\middle| Z_1,\ldots,Z_n\right) \leq$$

$$\sup_{Z_1,\ldots,Z_n} \frac{1}{n}\sum_{j=1}^n P_L(Z_j)'(P_L'P_L)^{-1}P_L(Z_j) = \sup_{Z_1,\ldots,Z_n} \frac{\operatorname{tr}\left(P_L'(P_L'P_L)^{-1}P_L\right)}{n} = \frac{L}{n}$$

so that by the Markov inequality (91) has bound

$$\frac{1}{n}\sum_{j=1}^n P_L(Z_j)'(P_L'P_L)^{-1}P_L'e_je_j'P_L(P_L'P_L)^{-1}P_L(Z_j) = O_p(Ln^{-1}) \tag{96}$$

For (92) by Assumption 11

$$\frac{1}{n}\sum_{j=1}^n u_{1L,j}^2 = O_p\left(L^{-2b_D}\right)$$

and by Assumption 11

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^n \left(p_L(Z_j)'(P_L'P_L)^{-1}P_L'u_{2Lj}\right)^2 &\leq \frac{1}{n}\sum_{j=1}^n \left|p_L(Z_j)'(P_L'P_L)^{-1}P_L'\right|^2 |u_{2Lj}|^2 \\
&\leq nO(L^{-2b_D}) \cdot \frac{1}{n}\sum_{j=1}^n \left|p_L(Z_j)'(P_L'P_L)^{-1}P_L'\right|^2 \\
&= O(L^{-2b_D})\sum_{j=1}^n p_L(Z_j)'(P_L'P_L)^{-1}P_L'P_L(P_L'P_L)^{-1}p_L(Z_j) \\
&= O(L^{-2b_D})\sum_{j=1}^n p_L(Z_j)'(P_L'P_L)^{-1}p_L(Z_j) \\
&= O(L^{-2b_D})\operatorname{tr}\left(P_L(P_L'P_L)^{-1}P_L'\right) \\
&= O_p(L^{1-2b_D})
\end{aligned}$$

For (93)

$$\begin{aligned}
\left|\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)'(P_L'P_L)^{-1}P_L'e_j u_{1L,j}\right| &\leq \frac{1}{n}\sum_{j=1}^{n}\left|p_L(Z_j)'(P_L'P_L)^{-1}P_L'e_j u_{1L,j}\right| \\
&\leq \sqrt{n}\cdot O(L^{-b_D})\cdot \frac{1}{n}\sum_{j=1}^{n}\left|p_L(Z_j)'(P_L'P_L)^{-1}P_L'\right| \\
&\leq \sqrt{n}\cdot O(L^{-b_D})\cdot \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left|p_L(Z_j)'(P_L'P_L)^{-1}P_L'\right|^2} \\
&= \sqrt{n}\cdot O(L^{-b_D})\cdot \sqrt{\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)'(P_L'P_L)^{-1}p_L(Z_j)} \\
&= \sqrt{n}\cdot O(L^{-b_D})\cdot \sqrt{\frac{1}{n}\operatorname{tr}\left(P_L(P_L'P_L)^{-1}P_L'\right)} \\
&= \sqrt{n}\cdot O(L^{-b_D})\cdot \sqrt{\frac{L}{n}} \\
&= O_p(L^{1/2-b_D})
\end{aligned}$$

For (94)

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)'(P_L'P_L)^{-1}P_L'e_j u_{2Lj}' P_L(P_L'P_L)^{-1}p_L(Z_j)\right| \\
&\leq O(L^{-b_D})\sum_{j=1}^{n} p_L(Z_j)'(P_L'P_L)^{-1}P_L'P_L(P_L'P_L)^{-1}p_L(Z_j) \\
&= O(L^{-b_D})\sum_{j=1}^{n} p_L(Z_j)'(P_L'P_L)^{-1}p_L(Z_j) \\
&= O(L^{-b_D})\operatorname{tr}\left(P_L(P_L'P_L)^{-1}P_L'\right) \\
&= O_p(L^{1-b_D})
\end{aligned}$$

Finally for (95)

$$\begin{aligned}
\left|\frac{1}{n}\sum_{j=1}^{n} p_L(Z_j)'(P_L'P_L)^{-1}P_L'u_{2Lj}u_{1L,j}\right| &\leq \sqrt{n}O(L^{-2b_D})\cdot \frac{1}{n}\sum_{j=1}^{n}\left|p_L(Z_j)'(P_L'P_L)^{-1}P_L'\right| \\
&\leq \sqrt{n}O(L^{-2b_D})\cdot \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left|p_L(Z_j)'(P_L'P_L)^{-1}P_L'\right|^2} \\
&= \sqrt{n}O(L^{-2b_D})\cdot \sqrt{\frac{1}{n}\operatorname{tr}\left(P_L(P_L'P_L)^{-1}P_L'\right)} \\
&= O_p(L^{1/2-2b_D})
\end{aligned}$$

Combining these results we find keeping the remainders of the highest order

Lemma 22 *If Assumptions 10 and 11 hold then*

$$\frac{1}{n}\sum_{j=1}^{n}(\tilde{V}_j - V_j)^2 = O_p(Ln^{-1}) + O_p(L^{1-b_D})$$

so that if $L = n^\kappa$

$$O_p(Ln^{-1}) + O_p(L^{1-b_D}) = O_p(n^{-2\delta})$$

with for $0 < \kappa < 1$

$$\delta = -\frac{1}{2}\max\{\kappa - 1, \kappa(1-b_D)\}$$

so that according to these bounds the fastest rate of convergence is for $b_D > 1$

$$\delta = \frac{b_D - 1}{2b_D}$$

where the rate is parametric if $b_D \to \infty$ *.*

K.3 Use of smoothed control variables

The control variables $\tilde{V}_j$ are not restricted to the $[0, 1]$ interval. Therefore as in Imbens and Newey (2002, p. 20) we will consider 'smoothed' control variables $\hat{V}_j$ that are always in the unit interval

$$\hat{V} = \begin{cases} 1 & \text{if } \tilde{V} > 1 + \xi_n \\ 1 - \frac{\left(1-\tilde{V}+\xi_n\right)^2}{4\xi_n} & \text{if } 1 - \xi_n < \tilde{V} \leq 1 + \xi_n \\ \tilde{V} & \text{if } \xi_n \leq \tilde{V} \leq 1 - \xi_n \\ \frac{\left(\tilde{V}+\xi_n\right)^2}{4\xi_n} & \text{if } -\xi_n \leq \tilde{V} < \xi_n \\ 0 & \text{if } \tilde{V} < -\xi_n \end{cases}$$

In particular we consider (70), which is predicated on the assumption that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_j\left(\tilde{V}_j - \hat{V}_j\right) = o_p(1), \tag{97}$$

which we establish here.

The following two lemmas show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|S_j|\frac{1}{4\xi_n}\left(\tilde{V}_j - \xi_n\right)^2 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right) = o_p(1) \tag{98}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|S_j|\left|\tilde{V}_j\right| 1\left(\tilde{V}_j \leq -\xi_n\right) = o_p(1) \tag{99}$$

Because

$$\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} S_j\left(\tilde{V}_j - \hat{V}_j\right)\right| \leq \frac{1}{\sqrt{n}}\sum_{j=1}^{n}|S_j|\left|\tilde{V}_j - 1\right| 1\left(\tilde{V}_j \geq 1+\xi_n\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}|S_j|\frac{1}{4\xi_n}\left(1-\tilde{V}_j-\xi_n\right)^2 1\left(1-\xi_n \leq \tilde{V}_j \leq 1+\xi_n\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}|S_j|\frac{1}{4\xi_n}\left(\tilde{V}_j-\xi_n\right)^2 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}|S_j|\left|\tilde{V}_j\right| 1\left(\tilde{V}_j \leq -\xi_n\right),$$

it implies that (97) follows from (98) and (99), noting that for the first two term we replace $\tilde{V}_j$ by $1-\tilde{V}_j$ in the lemmas that follow.

Lemma 23 *Let ξ_n be such that that $\hat{\sigma}_n = o_p(\xi_n)$, where $\hat{\sigma}_n = \sup_{1\leq j\leq n}\left|\tilde{V}_j - V_j\right|$.Then (98) holds.*

Proof. Because

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}|S_j|\left|\tilde{V}_j\right| 1\left(\tilde{V}_j \leq -\xi_n\right) \leq \sqrt{\frac{1}{n}\sum_{j=1}^{n}S_j^2\tilde{V}_j^2}\left(\sum_{j=1}^{n}1\left(\tilde{V}_j \leq -\xi_n\right)\right)$$

it suffices to show

$$\sum_{j=1}^{n}1\left(\tilde{V}_j \leq -\xi_n\right) = o_p(1).$$

We have

$$\sum_{j=1}^{n}1\left(\tilde{V}_j \leq -\xi_n\right) = \sum_{j=1}^{n}1\left(\tilde{V}_j - V_j \leq -\xi_n - V_j\right)$$
$$\leq \sum_{j=1}^{n}1\left(\tilde{V}_j - V_j \leq -\xi_n\right)$$
$$\leq \sum_{j=1}^{n}1\left(\left|\tilde{V}_j - V_j\right| \geq \xi_n\right)$$
$$\leq n\cdot 1\left(\hat{\sigma}_n \geq \xi_n\right)$$

Therefore, for any given $\varepsilon > 0$

$$\Pr\left[\sum_{j=1}^{n}1\left(\tilde{V}_j \leq -\xi_n\right) \geq \varepsilon\right] \leq \Pr\left[n\cdot 1\left(\hat{\sigma}_n \geq \xi_n\right) \geq \varepsilon\right]$$
$$= \Pr\left[\hat{\sigma}_n \geq \xi_n\right]$$

It follows that if $\hat{\sigma}_n = o_p(\xi_n)$

$$\Pr[\hat{\sigma}_n \geq \xi_n] \to 0$$

and therefore

$$\sum_{i=1}^{n} 1\left(\tilde{V}_j \leq -\xi_n\right) = o_p(1)$$

■ We will derive the rate of convergence of $\sup_{1\leq j\leq n}\left|\tilde{V}_j - V_j\right|$ after the next lemma.

Lemma 24 *Suppose that $E\left[S_j^p\right] < \infty$ for some $p > 1$. Suppose that $\xi_n = n^{-a}$ for some $a > 0$ such that*

$$a > \max\left(\frac{q}{2(q+1)}, \frac{q-1}{2q+1}\right)$$

where q satisfies $\frac{1}{p} + \frac{1}{q} = 1$. Let ξ_n be such that $\hat{\sigma}_n = o_p(\xi_n)$. Then (99) holds.

Proof. Note that if $-\xi_n \leq \tilde{V}_j \leq \xi_n$

$$\frac{\left(\tilde{V}_j - \xi_n\right)^2}{4\xi_n} \leq \xi_n$$

By the Hölder inequality

$$\begin{aligned}
\xi_n \frac{1}{\sqrt{n}} \sum_{j=1}^{n} |S_j| \, 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right) &= \sqrt{n}\xi_n \frac{1}{n} \sum_{j=1}^{n} |S_j| \, 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right) \\
&\leq \sqrt{n}\xi_n \left(\frac{1}{n}\sum_{j=1}^{n} S_j^p\right)^{\frac{1}{p}} \left(\frac{1}{n}\sum_{i=1}^{n} 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right)^q\right)^{\frac{1}{q}} \\
&= \left(\frac{1}{n}\sum_{j=1}^{n} S_j^p\right)^{\frac{1}{p}} \left(n^{\frac{q}{2}-1}\xi_n^q \sum_{j=1}^{n} 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right)\right)^{\frac{1}{q}}
\end{aligned}$$

with $\frac{1}{p} + \frac{1}{q} = 1$. Because

$$\begin{aligned}
\sum_{j=1}^{n} 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right) &= \sum_{j=1}^{n} 1\left(-\xi_n \leq \tilde{V}_j - V_j + V_j \leq \xi_n\right) \\
&= \sum_{j=1}^{n} 1\left(-\xi_n - \left(\tilde{V}_j - V_j\right) \leq V_j \leq \xi_n - \left(\tilde{V}_j - V_j\right)\right) \\
&\leq \sum_{j=1}^{n} 1\left(-\xi_n - \hat{\sigma}_n \leq V_j \leq \xi_n + \hat{\sigma}_n\right) \\
&= \sum_{j=1}^{n} 1\left(0 \leq V_j \leq \xi_n + \hat{\sigma}_n\right)
\end{aligned}$$

Consider for some $C > 1$

$$\sum_{j=1}^{n} 1(\hat{\sigma}_n \leq (C-1)\xi_n) 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right) \leq \sum_{j=1}^{n} 1\left(0 \leq V_j \leq C\xi_n\right)$$

Therefore

$$1(\hat{\sigma}_n \leq (C-1)\xi_n) n^{\frac{q}{2}-1}\xi_n^q \sum_{j=1}^{n} 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right) \leq n^{\frac{q}{2}-1}\xi_n^q \sum_{j=1}^{n} 1\left(0 \leq V_j \leq C\xi_n\right)$$

but the right-hand side has mean equal to

$$n^{\frac{q}{2}-1}\xi_n^q \cdot n \cdot C\xi_n = O\left(n^{\frac{q}{2}}\xi_n^{q+1}\right) = O\left(n^{\frac{q}{2}-aq-a}\right) = o\left(1\right)$$

and variance

$$n^{q-2}\xi_n^{2q} \cdot n \cdot C\xi_n \left(1 - C\xi_n\right) = O\left(n^{q-1}\xi_n^{2q+1}\right) = O\left(n^{q-1-2aq-a}\right) = o\left(1\right)$$

so by Chebyshev, we have, because we can omit the indicator on the left-hand side, that

$$\xi_n \frac{1}{\sqrt{n}} \sum_{j=1}^{n} |S_j| \, 1\left(-\xi_n \leq \tilde{V}_j \leq \xi_n\right) = o_p\left(1\right)$$

The above requirement can be rewritten

$$(q+1)\, a > \frac{q}{2}$$
$$(2q+1)\, a > q-1$$

or

$$a > \max\left(\frac{q}{2\left(q+1\right)}, \frac{q-1}{2q+1}\right)$$

■

Remark 2 *To derive a bound on* $\sup_{1\leq j\leq n}\left|\tilde{V}_j - V_j\right|$ *we note that*

$$\sup_{1\leq j\leq n}\left|\tilde{V}_j - V_j\right| = \sup_{1\leq j\leq n}\left|P_L(Z_j)'\tilde{\alpha}_L(X_j) - F(X_j|Z_j)\right| \leq \sup_{x,z}\left|P_L(z)'\tilde{\alpha}_L(x) - F_(x|z)\right|$$

Because the same proof as for the uniform bound in Theorem 1 of Newey (1997) holds we find that $\hat{\sigma}_n = O_p(L^{3/2}n^{-1/2}) + O_p(L^{1-b_D})$. *Note that as* $p \to \infty$, *we have* $q \to 1$, *and as such*

$$\max\left(\frac{q}{2\left(q+1\right)}, \frac{q-1}{2q+1}\right) \to \frac{1}{4}$$

Therefore, if S_j *has moments of arbitrarily large order, then we can take*

$$\xi_n = n^{-\left(\frac{1}{4}+\varepsilon\right)}$$

for some small $\varepsilon > 0$. *Therefore if* $L = n^{\kappa}$ *we require that*

$$\max\left\{\frac{3}{2}\kappa - \frac{1}{2}, \kappa(1-b_D)\right\} < -\left(\frac{1}{4}+\varepsilon\right)$$

References

Hahn, J., and G. Ridder (2011): "Conditional moment restrictions and triangular simultaneous equations," *Review of Economics and Statistics*, 93, 683–689.

——— (2013): "Asymptotic variance of semiparametric estimators with generated regressors," *Econometrica*, 81, 315–340.

Hahn, J., and G. Ridder (2014): "Instrumental variable estimation of nonlinear models with nonclassical measurement error using control variables," INET working paper.

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

Matzkin, R. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1339–1375.

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

Rivers, D., and Q. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39, 347–366.

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics.* John Wiley&Sons, New York.