Instrumental Variable Estimation of Nonlinear Models with Nonclassical Measurement Error Using Control Variates*

Jinyong Hahn† Yingyao Hu‡ Geert Ridder §

March 2008

INCOMPLETE; DO NOT QUOTE WITHOUT PERMISSION FROM THE AUTHORS

Abstract

We consider nonlinear models with an independent variable that is measured with error. The measurement error can be correlated with the true value, i.e. the measurement error is allowed to be nonclassical. We show that we can use a control variate estimator to estimate the parameters of interest. If we are prepared to make an assumption of the joint distribution of the first-stage and measurement errors the estimator is parametric. If we are only willing to specify the marginal distribution of the measurement error (up to a finite dimensional parameter vector), the estimator is semi-parametric. In the semi-parametric case the instrument must be sufficiently powerful, it must have a sufficiently large support. We derive the influence function of the semi-parametric estimator that properly account for the estimation of the control variates in the first stage.

1 Introduction

We study the estimation of models in which an independent variable is measured with error. Until recently, the literature focused on linear regression models and on what is usually called classical measurement error, i.e. measurement error that is independent of the true value of the variable and of other covariates in the model. However, many models that are used in empirical research are nonlinear in the covariates, for instance discrete choice and duration models, and most structural models that are derived from economic theory. Also replication studies in which both the mismeasured and true value of the covariate (and therefore the measurement error) are observed, have shown that the classical measurement error assumptions do not hold in practice. In this paper we assume that we have an instrumental variable that is independent of the measurement error. Recently, Hu and Schennach (2006) have obtained general results on the identification of models with non-classical measurement error by instrumental variables. Their results show that the parameters of nonlinear models can often

*Geert Ridder received financial support for this research through NSF grant SES 0452572.

†Department of Economics, UCLA

‡Department of Economics, Johns Hopkins University

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

be identified using instrumental variables[1]. We suggest relatively simple estimators that are consistent with a parametric rate of convergence and that are asymptotically normal.

We use the control function/variate approach. Recently instrumental variable estimation under weak conditions has been studied as an inverse problem. However, Hahn and Ridder (2007) argue that in many cases formal inverses can be replaced by averages over control variates and that the latter procedure has advantages over the former. This paper is a further illustration of that point in the context of the IV estimation of models with measurement error in the covariates. The key idea behind the use of control variates is that if we condition on them the 'endogenous' independent variable and the model error are independent. Control variate estimators add the first-stage residual to the set of covariates and estimate the resulting model. There are two complications in this procedure. First, we have to recover the model parameters from the relation between the dependent variable, the (endogenous) covariates and the control variate. If we average this relation over the control variate we obtain the average structural function that can be related to the model parameters. Second, identifying the average structural function on (a subset of) the support of the covariates requires assumptions on the strength of the instrument. These issues were first discussed by Imbens and Newey (2003) in the context of nonparametric IV estimation.

It should be noted that the identification of parameters in a model with mismeasured covariates by instrumental variables involves two inversions. The first deals with the correlation between the observed (mismeasured) covariate and the model random error. Conditioning on and averaging over a control variate yields the average structural function instead of a formal inverse. The second inversion deals with the attenuation[2] bias due to the measurement error. In the case of a (nonlinear) regression model the control function is the convolution of the regression function and the measurement error distribution. The control variate is not helpful in performing this inversion. In this paper we resolve this by assuming that the *marginal* distribution of the measurement error is in a parametric family. With this assumption we do not need to perform a deconvolution to recover the model parameters. Instead of this assumption we can e.g. assume that the distribution of the measurement error is symmetric, but that may be less attractive in practice.

The control variate estimator can be implemented in different ways depending on the assumptions that one is willing to make. These assumptions concern: (i) the relation between the mismeasured covariate and other covariates that are correlated with the measurement error and the instrument, and (ii) the relation between the measurement error and the first stage error. In this paper we assume that the first stage is a parametric (nonlinear) regression model with an independent error. This assumption implies that the first stage regression function is correctly specified which ensures that the first stage error is mean independent of the the instrument. We strengthen this to full independence. Hahn and Ridder (2007) note that full independence will hold by construction if the first stage is nonparametric. This complicates the asymptotic analysis of the estimator and we do not consider this possibility here. The estimator is fully parametric if we assume that the conditional distribution of the measurement error given the first stage error is in some parametric family. This assumption is in the spirit of the control function/variate literature. However in applications such an assumption is hard to justify and for that reason we also consider a semi-parametric version of the estimator that does nor require an assumption on the joint distribution of the measurement and first stage errors. In this case the relation between the dependent variable on the one hand and the mismeasured regressor and the control variate on the other is estimated nonparametrically. This nonparametric regression estimator is an approximation of the infeasible nonparametric regression of the dependent variable on the covariate and the first stage error. Because we average over both argument of the

[1]Our assumptions neither imply or are implied by theirs.
[2]In the case of nonclassical measurement error this bias is not necessarily towards 0.

nonparametric regression, be it independently as in a V statistic, the control variate estimator has the parametric rate of convergence. Because the control variate is a residual from the first stage regression we must deal with two problems in the derivation of the influence function. First, the support of the residual does not coincide with the support of the first stage error. We solve this by an extension of the population conditional mean function given the covariate and the first stage error that is sufficiently often continuously differentiable and has the same uniform approximation as the population conditional mean. Second, we must account for the effect of the estimation of the first stage residuals. The estimated residuals have an asymptotically non-negligible effect on both the control variate and the nonparametric regression and we derive the contribution of both to the influence function.

In section 2 we present the model and the assumptions. Section 3 proposes a fully parametric control function estimator that requires the specification (up to a vector of parameters) of the conditional distribution of the measurement error given the first-stage regression error. In section 4 we propose a semiparametric estimator that only requires that the marginal distribution of the measurement error is in a parametric family. The distribution theory of the parametric and semiparametric control function estimators is developed in section 5. Section 6 contains a small simulation study and finally section 7 discusses the control function estimator for nonlinear models with nonseparable errors and for models in which the measurement error is correlated with other covariates that are observed without error.

2 The model and assumptions

We develop the estimator for the nonlinear regression model

$$Y = m\left(X^*; \theta\right) + \eta \tag{1}$$

with m a known function, θ a parameter vector and η an error term. In section 7 we show that our results can be generalized to the nonlinear model with implicit random error

$$Y = m\left(X^*, \eta; \theta\right) \tag{2}$$

that has as special cases most nonlinear econometric models, for instance limited-dependent variable and duration models.

The latent X^* that is assumed to be a continuous variable, is not observed, but instead we observe

$$X = X^* + \varepsilon \tag{3}$$

Hence ε is the measurement error in the true value of the covariate. To keep the exposition simple we initially assume that X^* is the only covariate in the relation. It is easy to allow for additional covariates W that are measured without error and that are not correlated with the measurement error. The measurement error need not have mean 0, nor does it have to be independent of X^*. In other words, we deviate from the classical measurement error assumptions. The second deviation from the classical assumptions, correlation between the measurement error and other covariates, is considered in section 7. We also do not assume that the equation error η and the measurement error ε are independent. They can be correlated if they depend on common unobserved variables.

The relation between Y and the observed X is

$$Y = m(X - \varepsilon; \theta) + \eta$$

where X is correlated with ε and possibly with η as well. If X were independent of ε, η we could estimate θ by nonlinear least squares. This requires the specification of the marginal distribution of the measurement error ε. The estimator proposed below essentially deals with the correlation between X and ε, η, but requires the specification of the marginal distribution of ε. However, any semi-parametric estimator of θ in the model with exogenous X and weaker assumptions on the marginal distribution of the measurement error, can be adapted to the presence of nonclassical measurement error by using a control variate in much the same way as we adapt the nonlinear least squares estimator with parametric marginal measurement error distribution. In section 7 we sketch such a procedure under the weaker assumption that the measurement error has a symmetric distribution.

We observe a third variable Z that is assumed to be independent of the measurement error and the equation error

Assumption 1 (Instrumental Variable)

$$Z \perp \varepsilon, \eta \tag{4}$$

This variable is the instrument that we use to estimate θ. The mismeasured covariate X is related to the instrument Z. To be specific we assume that the relation between X and Z is a possibly nonlinear parametric regression model

Assumption 2 (Parametric First Stage with Independent Error)

$$X = h(Z; \alpha) + V \qquad\qquad V \perp Z \tag{5}$$

To see how restrictive this assumption is we note that if the conditional distribution of X given $Z = z$ has cdf $F(x|z)$, we have $V^* = F(X|Z)$, where V^* has a uniform distribution on $[0,1]$ that by construction is independent of Z. This gives the 'first-stage model'

$$X = F^{-1}(V^*|Z) = h^*(Z, V^*). \tag{6}$$

This 'model' does not impose any restrictions on the relation between X and Z. By projection we obtain

$$X = h(Z) + V, \tag{7}$$

with $h(z) = \mathbb{E}[h^*(Z, V^*)|Z = z]$. By construction $\mathbb{E}[V|Z] = 0$. Again (7) does not impose any restriction on the relation between X and Z. Therefore assumption 2 strengthens mean independence to full independence and restricts $h(z)$ to be in a parametric family. We do not need these assumptions if we are willing to have a nonparametric first-stage model. Because this complicates the asymptotic analysis of our estimators, we leave this to future research.

Assumption 2 implies that, given V, X is a function of Z only and hence by assumption 1 independent of ε, η, i.e.

$$\varepsilon, \eta \perp X \stackrel{d}{=} h(Z; \alpha) + v \mid V = v$$

Therefore, if we condition on V the correlation between the mismeasured covariate X and the measurement error which results in measurement error bias disappears. This is the basic idea behind the use of a control variate.

We can compare our assumptions 1 and 2 to those in the small literature on IV estimation of nonlinear models with measurement error. Schennach (2007) assumes that

$$\begin{aligned} Y &= m(X^*) + \eta \\ X &= X^* + \varepsilon \\ X^* &= Z + U \end{aligned}$$

with

$$X^* \perp \varepsilon \qquad\qquad Z \perp U, \varepsilon$$

Hence $X = Z + V$ with $V = U + \varepsilon \perp Z$ which is the same as our assumption 2. The main difference is that she assumes that the measurement error is classical. She also is more ambitious, because she estimates m nonparametrically. Hu and Schennach (2006) make assumptions on conditional densities

$$\begin{aligned} f(y|x^*, x, z) &= f(y|x^*) \\ f(x|x^*, z) &= f(x|x^*) \\ \mathbb{E}[X|X^* = x^*] &= x^* \end{aligned}$$

In our setup the final assumption is that the mean of the measurement error is 0 (they actually make an assumption on the mode). The main difference with our assumptions is the second one. To see the difference assume joint normality of ε, V so that $\varepsilon = \rho V + \nu$ and

$$\mathbb{E}[X|X^*, Z] = h(Z;\alpha) + \frac{(1-\rho)\sigma_V^2}{(1-\rho)^2\sigma_V^2 + \sigma_\nu^2}(X^* - h(Z;\alpha))$$

which does not depend on Z if and only if $\rho(1-\rho)\sigma_V^2 = \sigma_\nu^2$ which in general will not hold[3]. We conclude that our assumptions are not implied by, nor do they imply the assumptions in the literature.

3 A parametric control variate estimator

Following the traditional approach to estimation with control variates/functions (see e.g. Rivers and Vuong (1988)) we first propose a fully parametric procedure that requires the specification of the conditional distribution of the measurement error given the first-stage regression error.

Assumption 3 (Conditional distribution of ε given V and mean independence of η and V) *The equation error η is mean independent of V. The conditional distribution of ε given $V = v$ is in a parametric family with c.d.f. $G(\varepsilon|v;\beta)$. If the model is a linear regression, only the conditional mean $\mathbb{E}[\varepsilon|V = v] = k(v;\beta)$ has to be in a parametric family.*

If the measurement error ε is assumed to have a (marginal) normal distribution, a natural assumption is that ε given $V = v$ is also normal with conditional mean $\mu(v;\beta_1)$ and conditional variance $\sigma^2(v;\beta_2)$.

The next theorem gives an expression for the conditional mean $\mathbb{E}(Y \mid X = x, V = v)$.

Theorem 1 *Suppose that assumptions 1, 2, and 3 are satisfied. We have*

$$\mathbb{E}(Y \mid X = x, V = v) = \mathbb{E}(m(x - \varepsilon;\theta) \mid V = v) \tag{8}$$

Proof. We have

$$\mathbb{E}(Y \mid X = x, V = v) = \mathbb{E}(m(X^*;\theta) + \eta \mid X = x, V = v) = \mathbb{E}(m(X^*;\theta) \mid X = x, V = v) + \mathbb{E}(\eta \mid X = x, V = v)$$

By assumptions 1 and 2 $Z \perp \eta, V$ so that from $f(z,\eta,v) = f(z)f(\eta,v)$ follows that $f(z,\eta|v) = f(z)f(\eta|v) = f(z|v)f(\eta|v)$ and therefore $Z \perp \eta | V$. By the same assumptions $X \overset{d}{=} h(Z;\alpha) + v$ given $V = v$. This implies that $X \perp \eta | V$ so that

$$\mathbb{E}(\eta \mid X = x, V = v) = \mathbb{E}(\eta \mid V = v) = 0$$

[3] A special case in which this holds is $\rho = 1, \sigma\nu^2 = 0$ so that $\varepsilon = V$, i.e. the first-stage error and the measurement error coincide. If we were to specify the first stage as $X^* = h(Z;\alpha) + V$ the conditional expectation is independent of Z if and only if $\rho = 0$ so that the measurement error is classical.

where the last equality follows from assumption 3. By assumptions 1 and 2 $Z \perp \varepsilon, V$ so that by the same argument as above $X \perp \varepsilon | V$ and therefore

$$\mathbb{E}\left(m(X^*;\theta) \mid X=x, V=v\right) = \mathbb{E}\left(m(X-\varepsilon;\theta) \mid X=x, V=v\right) = \mathbb{E}\left(m(x-\varepsilon;\theta) \mid X=x, V=v\right) =$$

$$\mathbb{E}\left(m(x-\varepsilon;\theta) \mid V=v\right)$$

■

In the special case of a linear regression

$$\mathbb{E}\left[Y \mid X=x, V=v\right] = \theta_0 + \theta_1 x - \theta_1 \mathbb{E}\left[\varepsilon | V=v\right]$$

so that in case of a linear conditional mean

$$\mathbb{E}\left[Y \mid X=x, V=v\right] = \theta_0 + \theta_1 x - \theta_1 \mu_\varepsilon - \theta_1 \beta v$$

Note that the regression coefficient on X^* is identified. The intercept is identified if we assume that the measurement error has mean 0.

Define

$$R(x,v;\tau) = \int_{\mathcal{E}} m(x-\varepsilon;\theta) g(\varepsilon|v;\beta) \mathrm{d}\varepsilon$$

with $\tau = (\theta' \ \beta')'$. We identify τ from the conditional moment restriction

$$m(y|x,v;\tau) = y - R(x,v;\tau)$$

This suggests following strategy of estimating τ:

1. Obtain the first-stage residuals $\widehat{V}_i = X_i - h\left(Z_i;\widehat{\alpha}\right)$.
2. Do a (nonlinear) regression with dependent variable Y and with regression function $R(x,v;\theta,\beta)$, i.e. solve the minimization problem

$$\min_{\tau} \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - R\left(X_i, \widehat{V}_i;\tau\right)\right)^2$$

This estimator is consistent and its asymptotic variance matrix can be derived from general results for nonlinear models with covariates that depend on preliminary estimates.

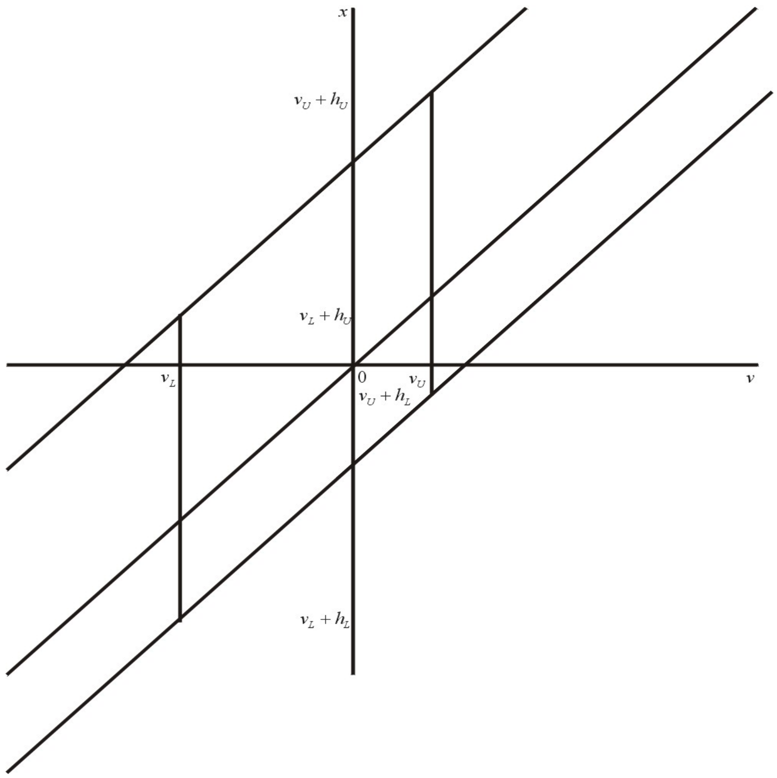
4 The semi-parametric control variate estimator

By adding the control variate to the relation between the dependent variable and the mismeasured covariate we remove the correlation between the mismeasured covariate and the random error of that relation. In a nonlinear model this does not remove all bias since we end up with the convolution of the regression function and the measurement error distribution. For this approach to work we need an assumption on the joint support of X and V that is essentially an assumption on the 'strength' of the instrument Z.

Assumption 4 (Joint support of X and V) *If we denote the support of the joint distribution of X, V by $\mathcal{A}$ and the supports of the marginal distributions of X and V by $\mathcal{X}$ and $\mathcal{V}$, then there is a subset $\mathcal{X}_0$ of $\mathcal{X}$ such that $\mathcal{X}_0 \times \mathcal{V} \subset \mathcal{A}$.*

To see that this is an assumption on the strength of the instrument consider the first-stage model in (5). The set of values taken by the regression function is $\mathcal{H} = \{h(z;\alpha)|z \in \mathcal{Z}\}$ with α the population value of the parameter and $\mathcal{Z}$ the support of the distribution of the instrument. In the sequel we assume that $\mathcal{H} = [z_L, z_U]$, i.e. that it is a closed interval. The support of the distribution of V is $\mathcal{V}$ which again is assumed to be an interval $\mathcal{V} = [v_L, v_U]$ with without loss of generality $v_L \leq 0 \leq v_U$. Because Z and V are independent the support of the joint distribution of $h(Z;\alpha), V$ is $\mathcal{H} \times \mathcal{V}$. The implied support of the joint distribution of V, X is in Figure 1. The set of values of X for which the

Figure 1: *The support of the joint distribution of* V, X



conditional distribution of V given $X = x$ has full support is $\mathcal{X}_0 = [v_U + h_L, v_L + h_U]$. This set is not empty if $h_U - h_L > v_U - v_L$, i.e. the instrument must induce induce sufficient variation in the first-stage regression function. Hence assumption 14 resembles the rank condition in the linear IV case.

Inspection of the proof of Theorem 1 shows that we have

Theorem 2 *If assumptions 1, 2 and 14 are satisfied, then for all* $x \in \mathcal{X}_0$

$$\int_{\mathcal{V}} \mathbb{E}\left(Y \mid X = x, V = v\right) g(v) dv = \mathbb{E}[m(x - \varepsilon;\theta)] \tag{9}$$

with g the marginal density of V *and the expectation over the marginal distribution of the measurement error.*

Proof. In the proof of Theorem 1 we established that

$$\mathbb{E}\left(Y \mid X = x, V = v\right) = \mathbb{E}\left(m(X^*;\theta) + \eta \mid X = x, V = v\right) = \mathbb{E}\left(m(x - \varepsilon;\theta) \mid V = v\right) + \mathbb{E}(\eta|V = v)$$

Averaging over V for values $x \in \mathcal{X}_0$ gives the result by the law of iterated expectations (note $\mathbb{E}(\eta) = 0$). ■

Define for $x \in \mathcal{X}_0$

$$L(x) = \int_{\mathcal{V}} \mathbb{E}\left(Y \mid X = x, V = v\right) g(v) \mathrm{d}v$$

This function is identified from the data, if we add the first stage error to the data. If we define

$$R(x;\tau) = \int_{-\infty}^{\infty} m\left(x - \varepsilon; \theta\right) g\left(\varepsilon; \gamma\right) \mathrm{d}\varepsilon \tag{10}$$

then we identify $\tau = (\theta' \ \gamma')'$ from the conditional moment restriction

$$m(x;\tau) = L(x) - R(x;\tau) = 0 \qquad x \in \mathcal{X}_0$$

In general, $L(x)$ does not identify the regression function $m(x)$ and the measurement error distribution nonparametrically. In this paper do not consider the question what additional restrictions on $m(x)$ and $g(\varepsilon)$ suffice to establish identification. Instead we assume that both the regression function and the distribution of ε are in a parametric family and that the conditional moment restriction uniquely point identifies θ and γ. It is clear that the assumption that the measurement error distribution is in a parametric family is not necessary for identification of θ and in section 7 we briefly discuss an alternative assumption on the measurement error distribution.

In the sequel we assume

Assumption 5 (Marginal distribution of ε) *The marginal distribution of the measurement error ε is in a parametric family $g(\varepsilon;\gamma)$.*

In the case of a linear regression the conditional moment restriction has the same $L(x)$ (which now is a linear function of x which is a remarkable implication of the model) and

$$R(x;\theta,\mu_\varepsilon) = \theta_0 + \theta_1 x - \theta_1 \mu_\varepsilon$$

with μ_ε the average measurement error. In this case we do not need the assumption that the measurement error distribution is in a parametric family. The slope coefficient of the linear regression model is identified. The intercept is identified if $\mu_\varepsilon = 0$.

The moment condition suggests the following strategy of estimating θ, γ with γ the parameters of the marginal distribution of the measurement error:

1. Estimate $\mu(x,v) = \mathbb{E}\left[Y \mid X = x, V = v\right]$ nonparametrically using the sample $Y_i, X_i, \widehat{V}_i = X_i - h\left(Z_i;\widehat{\alpha}\right), i = 1,\ldots,n$. Call the estimator $\hat{\mu}(x,v)$.

2. Estimate the left hand side of (9) by

$$\hat{L}\left(x\right) \equiv \frac{1}{n}\sum_{j=1}^{n} \hat{\mu}(x,\hat{V}_j) \tag{11}$$

   for $x \in \mathcal{X}_0$.

3. Solve the minimization problem

$$\min_{\tau} \frac{1}{n}\sum_{i=1}^{n} 1_{\mathcal{X}_0}(X_i)\left(\hat{L}\left(X_i\right) - R\left(X_i;\tau\right)\right)^2 \tag{12}$$

5 Distribution theory

The first assumption is on the first stage

Assumption 6 (First-stage) *X and Z are related by*

$$X = h(Z;\alpha_0) + V \qquad V \perp Z \qquad \mathbb{E}(V^2) < \infty$$

With $h(Z;\alpha)$ twice continuously differentiable with respect to α in an open neighborhood of α_0 for almost all Z. We assume $\alpha_0 \in A$ with A a compact set and $\sup_{\alpha \in A} |\frac{\partial^k h}{\partial \alpha^k}(Z;\alpha)| \leq M_k(Z)$ with M_k a bounded function of Z for $k = 0,1,2$. For all $\delta > 0$, there is an $\zeta > 0$ such that

$$\sup_{\alpha \in A, |\alpha - \alpha_0| > \delta} \mathbb{E}\left[(h(Z;\alpha) - h(Z;\alpha_0))^2\right] > \zeta \tag{13}$$

Finally, the matrix

$$\mathbb{E}\left[\frac{\partial h}{\partial \alpha} h(Z;\alpha) \frac{\partial h}{\partial \alpha'} h(Z;\alpha_0)\right]$$

is nonsingular.

This ensures that the first stage estimator defined by

$$\frac{1}{n}\sum_{j}^{n} (X_j - h(Z_j;\hat{\alpha})) \frac{\partial h}{\partial \alpha}(Z_j;\hat{\alpha}) = 0 \tag{14}$$

has the usual asymptotically linear representation

Lemma 1 *If assumption 6 holds, then $\hat{\alpha}$ defined in (14) is weakly consistent for α_0 and*

$$\sqrt{n}(\hat{\alpha} - \alpha_0) = \left(\mathbb{E}\left[\frac{\partial h}{\partial \alpha} h(Z;\alpha_0) \frac{\partial h}{\partial \alpha'} h(Z;\alpha_0)\right]\right)^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j \frac{\partial h}{\partial \alpha} h(Z_j;\alpha_0) + o_p(1)$$

All proofs are in the appendix.

Next we make assumptions on the nonlinear regression model $m(x;\theta)$ and the marginal distribution of the measurement error. We express these as assumptions on $R(x;\tau)$

Assumption 7 (Regression function and marginal distribution of the measurement error) *The nonlinear regression model for Y is*

$$Y = m(X^*;\theta_0) + \eta \qquad \mathbb{E}(\eta | X^*) = 0$$

with $\theta_0 \in \Theta$ and X^* the latent true value of the regressor. For $R(x;\tau)$ defined in (10) we assume that for all $\delta > 0$, there is a $\zeta > 0$ such that

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

Also for $d = 0,1,2$

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

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.

Under some additional assumptions we have that the control variate estimator is weakly consistent and is asymptotically linear in the nonparametric estimator $\hat{L}$.

Lemma 2 *If assumption 7 holds $a_0 > 5/2$[4], and $k = n^\kappa$ with $0 < \kappa < 1/7$, then $\hat{\tau}$ defined in (12) is weakly consistent for τ_0 and*

$$\sqrt{n}(\hat{\tau} - \tau_0) = \left(\mathbb{E} \left[\frac{\partial R}{\partial \tau}(X; \tau_0) \frac{\partial R}{\partial \tau'}(X; \tau_0) \right] \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{L}(X_i) - L(X_i) \right) \frac{\partial R}{\partial \tau}(X_i; \tau_0) + o_p(1) \quad (15)$$

The next step is to study the nonparametric estimator $\hat{L}(x)$ that is a partial average over the nonparametric regression estimator $\hat{\mu}(x, v)$. The analysis would be relatively simple if we could do a nonparametric regression of Y on X, V instead of on $X, \hat{V}$. The nonparametric regression estimator is the series estimator. Define $W = (X \ V)'$ and $\hat{W} = (X \ \hat{V})'$. As the basis functions we take a power series and k is the number of basis function in the series. To include all powers of x and v up to order n, we need to include $k = \frac{1}{2}(n+1)(n+2)$ terms. The resulting k basis functions are denoted by the k vector $P_k(w) = (x^{\lambda_1} w^{\lambda_2}, \lambda_1 + \lambda_2 \leq n)$. 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 8 (Support) *The support of X, V is $\mathcal{W} = \mathcal{X} \times \mathcal{V} = [x_L, x_U] \times [v_L, v_U]$. The joint density of X, V is bounded from 0 on $\mathcal{W}$ and is r times continuously differentiable on its support.*

By Newey (1995) this implies that we can take the basis functions as orthonormal polynomials with respect to the distribution of W.The fact that we estimate the nonparametric regression estimator with first-stage residuals creates two problems. First, the support of W need not be that of $\hat{W}$. To analyze the nonparametric regression on $\hat{W}$ we must extend the definition of $\mu(w)$ to the support of $hatW$. This extension must be sufficiently often (two times in this application) continuously differentiable on the support of $\hat{W}$ and the following assumption on the uniform approximation of $\mu(w)$ must hold on the support of $\hat{W}$.

Assumption 9 (Regression function) *There is a vector γ_k such that for some constants $C_d, a_d, d = 0, 1, \ldots, D$*

$$\sup_{w \in \mathcal{W}} \left| \frac{\partial^d \mu}{\partial v^d}(w) - \frac{\partial Q_k}{\partial v}(w)' \gamma_k \right| \leq C_d k^{-a_d}$$

It is essential that the result holds on the support of $\hat{W}$ with the same coefficients γ_k. We define the following extension

$$\begin{aligned} \mu_k(X, \hat{V}) &= \mu(X, \hat{V}) && v_L \leq \hat{V} \leq v_U \\ &= Q_k(X, \hat{V})'\gamma_k + \mu(X, v_U) - Q_k(X, v_U)'\gamma_k + \sum_{l=1}^{L} \frac{1}{l!} \left(\frac{\partial^l \mu}{\partial v^l}(X, v_U) - \frac{\partial^l Q_k}{\partial v^l}(X, v_U)'\gamma_k \right) (\hat{V} - v_U)^l && \hat{V} > v_U \\ &= Q_k(X, \hat{V})'\gamma_k + \mu(X, v_L) - Q_k(X, v_L)'\gamma_k + \sum_{l=1}^{L} \frac{1}{l!} \left(\frac{\partial^l \mu}{\partial v^l}(X, v_L) - \frac{\partial^l Q_k}{\partial v^l}(X, v_L)'\gamma_k \right) (\hat{V} - v_L)^l && \hat{V} < v_L \end{aligned}$$

The second problem is that the influence function of the control variate estimator has terms that account for the first stage estimation of the residuals. It turns out that the first stage estimation enters

[4] To be defined in assumption 9.

in two ways: we average over $\hat{V}_j$ instead of over V_j and we regress on polynomials in $\hat{V}$ instead of in V. The contribution to the influence function is

$$\left(\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}(W)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\right)\sqrt{n}(\hat{\alpha}-\alpha_0)$$

The matrix is 0, so that there is no contribution, if X and Z are independent.

We also need an assumption on the conditional variance of the dependent variable.

Assumption 10 (Variance)

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

The main result is the asymptotically linear representation of the control variate estimator.

Theorem 3 (Asymptotically linear representation) *If assumptions 6, 7, 8, 9 and 10 hold, then*

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

$$\left\{\left(\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}(W)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\right)\sqrt{n}(\hat{\alpha}-\alpha_0)+\right.$$

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

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

$$O_p(k^{9/2}n^{-1/2})+O_p(k^{7/2-a_0})+O_p(n^{1/2}k^{-a_0})+O_p(n^{1/2}k^{-r/2})+O_p(k^{(4-r)/2})+O_p(n^{1/2}k^{(5-r)/2-a_0})$$

Now take $k=Cn^{\kappa}$. The remainders are negligible if the following inequalities are satisfied simultaneously

$$\kappa<\frac{1}{9}\qquad a_0>\frac{7}{2}\qquad \kappa>\frac{1}{2a_0}\qquad \kappa>\frac{1}{r}\qquad r\geq 5\qquad \kappa>\frac{1}{r-5+2a_0}$$

Therefore

$$a_0>\frac{9}{2}\qquad r\geq 10$$

and

$$\min\left\{\frac{1}{2a_0},\frac{1}{r}\right\}<\kappa<\frac{1}{9}$$

6 A simulation study

We study the finite sample performance of the parametric and semi-parametric control function estimators in a simulation experiment. The model is

$$Y = \theta_0 + \theta_1 X^* + \eta$$

The true and observed value of the covariate are related by

$$X = X^* + \varepsilon$$

and the instrument Z is independent of η, ε. The parametric first stage model is

$$X = \alpha_0 + \alpha_1 Z + V \qquad V \perp Z$$

We take

$$V \sim U[-1,1] \qquad Z \sim U[0,1]$$

This implies that the joint distribution of X, V has bounded support. If $0 \leq x \leq 1$ the conditional support of V is full, i.e. $[-1,1]$. In the second stage only observations with $0 \leq X_i \leq 1$ are used. The variance of V is 1/12. The measurement error is related to the first stage error V by

$$\varepsilon = \kappa V + \nu \qquad \nu \sim N(0, \sigma_\nu^2)$$

Therefore

$$\varepsilon | V \sim N(\kappa V, \sigma_\nu^2)$$

and the marginal distribution of ε is not normal but

$$\mathbb{E}(\varepsilon) = 0 \qquad \text{Var}(\varepsilon) = \frac{\kappa^2}{12} + \sigma_\nu^2$$

The assumption that the mean of the measurement error is 0 ensures that the intercept θ_0 is identified. The covariance of the mismeasured covariate and the measurement error is

$$\text{E}(X\varepsilon) = \frac{\kappa}{12}$$

and its variance is

$$\text{Var}(X) = \frac{\alpha_1^2 + 1}{12}$$

The latent true value is obtained from

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

This implies that the covariance of X^* and the measurement error is

$$\mathbb{E}[X^* \varepsilon] = \frac{\kappa(1-\kappa)}{12} - \sigma_\nu^2$$

and

$$\text{Var}(X^*) = \frac{\alpha_1^2 + (\kappa - 1)^2}{12} + \sigma_\nu^2$$

We choose $\theta_0 = 0, \theta_1 = 1, \alpha_0 = 0, \alpha_1 = 1, \sigma_\eta^2 = 1, \kappa = 5, \sigma_\nu^2 = 1$. This implies that the correlation between X^* and the measurement error is $-.25$. The correlation between the mismeasured covariate X and the error in a regression of Y on X is -.72.

Table 1: Results simulation, $N = 500$, no. of replications is 100

		Par, V	Par, $\hat{V}$	Semi-par, V	Semi-par, $\hat{V}$
θ_0	Ave.	-0.00976097	-0.00936532	-0.03467247	-0.02563894
	Std. Ave.	0.00514093	0.01163065	0.02478237	0.02720358
θ_1	Ave.	0.99986276	0.99052771	1.00767846	1.00168646
	Std. Ave.	0.00304464	0.00730770	0.01594379	0.01783003
κ	Ave.	4.98317958	4.97384453		
	Std. Ave.	0.00546665	0.00872587		

We estimate $\mu(x, v) = \mathbb{E}(Y|X = x, V = v)$ by a nonparametric series estimator. As basis functions we use tensor products of the polynomials $\phi_k(v)$ and $\psi_k(x)$ that are orthonormal with respect to the weight functions on $[-1, 1]$ and $[1, 2]$ respectively. Initially we choose the order of the polynomial $K = 3$. The coefficients of the polynomials are estimated by least squares. Note that in the setup chosen here the population $\mu(x, v)$ is linear in x and v. Therefore the parametric control variate estimator is just the linear regression of Y on the mismeasured covariate and the control variate. We also compare the estimator with the first stage residuals $\hat{V}$ and the first errors V. The results are in Table 1.

7 Implicit error models and additional covariates

We developed the control variate estimator for the nonlinear regression model. Here we show how we can use this estimator in models with an implicit error as in (2). As before we first consider a fully parametric approach that requires the specification of a parametric joint distribution of the implicit error and the measurement error.

Assumption 11 (Conditional Distribution of (ε, η) given V) *The equation and measurement errors are independent given V, the equation error is independent of V and the conditional distribution of ε given $V = v$ is in a parametric family. Therefore*

$$\varepsilon, \eta | V = v \overset{d}{=} G(\varepsilon | v; \beta) H(\eta; \gamma) \tag{16}$$

Theorem 4 *Suppose that Conditions 1, 5, and 11 are satisfied. Letting $A(y, x^*; \theta) \equiv \{\eta \mid m(x^*, \eta; \theta) \leq y\}$, we have for the model with implicit error*

$$F(y \mid X = x, V = v) = \mathbb{E}\left[\Pr(\eta \in A(y, x - \varepsilon; \theta) | V = v, \varepsilon) \mid V = v\right] = \mathbb{E}\left[P_\eta(\eta \in A(y, x - \varepsilon; \theta)) \mid V = v\right] \tag{17}$$

Proof. By equations (2) and (3) we have

$$\begin{aligned} F(y \mid X = x, V = v) &= \Pr(m(X^*, \eta; \theta) \leq y \mid X = x, V = v) \\ &= \Pr(m(X - \varepsilon, \eta; \theta) \leq y \mid X = x, V = v) \\ &= \Pr(m(x - \varepsilon, \eta; \theta) \leq y \mid X = x, V = v) \end{aligned}$$

Because Z and V are independent, we have that conditional on $V = v$, $X = h(Z; \alpha) + v$ is a function of Z and hence by assumption 1

$$F(y \mid X = x, V = v) = \Pr(m(x - \varepsilon, \eta; \theta) \leq y \mid V = v)$$

Finally by assumption 11 we have

$$F(y \mid X = x, V = v) = \mathbb{E}\left[\Pr\left(\eta \in A(y, x - \varepsilon; \theta) \middle| V = v, \varepsilon\right) \mid V = v\right] = \mathbb{E}\left[P_\eta\left(\eta \in A(y, x - \varepsilon; \theta)\right) \mid V = v\right]$$

where the expectation is over the conditional distribution of the measurement error ε given V. ■

The set $A(y, x^*; \theta) = \{\eta \mid m(x^*, \eta; \theta) \leq y\}$ takes different forms depending on m. For a probability model with $m(x^*, \eta; \theta) = 1\{x^*\theta + \eta > 0\}$, we have e.g. $A(0, x^*; \theta) = \{\eta \mid \eta \leq -x^*\theta\}$. Hence for a probit for $y = 0$, we have

$$\Pr\left(\eta \in A(y, x - \varepsilon; \theta) \middle| V = v, \varepsilon\right) = \Pr\left(\eta \leq -(x - \varepsilon)\theta \middle| V = v, \varepsilon\right)$$

and hence,

$$F(0 \mid X = x, V = v) = \int_{-\infty}^{\infty} \left[1 - \Phi\left((x - \varepsilon)\theta\right)\right] \mathrm{d}G(\varepsilon \mid V = v)$$

This can be simplified further if e.g. the conditional distribution of ε given $V = v$ is normal with mean $\mu_\varepsilon + \rho(v - \mathbb{E}[V])$ and variance τ^2 so that

$$F(0 \mid X = x, V = v) = 1 - \Phi\left(\frac{\theta x - \theta\left(\mu_\varepsilon + \rho(v - \mathbb{E}[V])\right)}{\sqrt{1 + \theta^2\tau^2}}\right)$$

so that in this case θ is identified up to scale.

Theorem 5 *Suppose that assumptions 1 and 2 are satisfied and ε and η are independent. With $A(y, x^*; \theta)$ as in Theorem 4 we have for the model with implicit error*

$$\int_{\mathcal{V}} F(y \mid X = x, V = v)\, g(v) dv = \mathbb{E}\left[P_\eta\left(\eta \in A(y, x - \varepsilon; \theta)\right)\right] \tag{18}$$

with g is the marginal density of V.

Proof. In the proof of Theorem 4 we established that

$$F(y \mid X = x, V = v) = \mathbb{E}\left[P_\eta\left(\eta \in A(y, x - \varepsilon; \theta)\right) \mid V = v\right]$$

The result follows if we average over V. In the case of a regression model

$$\mathbb{E}\left(m(X^*; \theta) \mid X = x, V = v\right) = \mathbb{E}\left(m(x - \varepsilon; \theta) \mid V = v\right) + \mathbb{E}(\eta | V = v)$$

Again averaging over V gives the result if $\mathbb{E}(\eta) = 0$.

■

For the probit example, if the measurement error has a marginal distribution that is normal with mean μ_ε and variance τ^2

$$\mathbb{E}\left[P_\eta(\eta \in A(y, x - \varepsilon; \theta))\right] = \Phi\left(\frac{\theta x - \theta\mu_\varepsilon}{\sqrt{1 + \theta^2\tau^2}}\right)$$

References

Hahn, J., and G. Ridder (2007): "Conditional Moment Restrictions and Triangular Simultaneous Equations," Discussion paper, Working Paper, IEPR, University of Southern California.

Hu, Y., and S. Schennach (2006): "Identification and estimation of nonclassical nonlinear errors-in-variables models with continuous distributions using instruments," Discussion paper, CeMMAP working papers CWP17/06, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Imbens, G., and W. Newey (2003): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," Working Paper, Department of Economics, Berkeley.

Newey, W. (1995): *Advances in Econometrics and Qualitative Economics: Essays in Honor of C.R.Rao*chap. Convergence rates for series estimators, pp. 254–275. Basil-Blackwell.

Rivers, D., and Q. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39, 347–366.

Schennach, S. (2007): "Instrumental variable estimation of nonlinear errors-in-variables models," *Econometrica*, 75, 201–39.

Appendix

A The estimator

Let $\tau = (\theta'\beta')'$ be the parameter vector and define

$$R(x;\tau) = \int_{\mathcal{E}} m(x-\varepsilon;\theta)g(\varepsilon;\beta)\mathrm{d}\varepsilon \tag{19}$$

Further define

$$\hat{L}(x) = \frac{1}{n}\sum_{j=1}^{n}\hat{\mu}(x,\hat{V}_j) \tag{20}$$

with $\hat{\mu}(x,v)$ the nonparametric series estimator of $\mu(x,v) = \mathbb{E}(Y|X=x,V=v)$ defined below. The first stage residual $\hat{V}_j$ is

$$\hat{V}_j = X_j - h(Z_j;\hat{\alpha})$$

where the Nonlinear Least Squares (NLS) estimator $\hat{\alpha}$ satisfies

$$\frac{1}{n}\sum_{j}^{n}(X_j - h(Z_j;\hat{\alpha}))\frac{\partial h}{\partial \alpha}(Z_j;\hat{\alpha}) = 0 \tag{21}$$

The estimator of τ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{L}(X_i) - R(X_i;\hat{\tau})\right)\frac{\partial R}{\partial \tau}(X_i;\hat{\tau}) = 0 \tag{22}$$

B The first stage estimator

For a matrix D we define the matrix norm $|D| = \sqrt{\mathrm{tr}(D'D)}$. We assume

Assumption 12 (First-stage) *X and Z are related by*

$$X = h(Z;\alpha_0) + V \qquad V \perp Z \qquad \mathbb{E}(V^2) < \infty$$

With $h(Z;\alpha)$ twice continuously differentiable with respect to α in an open neighborhood of α_0 for almost all Z. We assume $\alpha_0 \in A$ with A a compact set and $\sup_{\alpha\in A}|\frac{\partial^k h}{\partial \alpha^k}(Z;\alpha)| \le M_k(Z)$ with M_k a bounded function of Z for $k = 0,1,2$. For all $\delta > 0$, there is an $\zeta > 0$ such that

$$\sup_{\alpha\in A,|\alpha-\alpha_0|>\delta} \mathbb{E}\left[(h(Z;\alpha) - h(Z;\alpha_0))^2\right] > \zeta \tag{23}$$

Finally, the matrix

$$\mathbb{E}\left[\frac{\partial h}{\partial \alpha}h(Z;\alpha)\frac{\partial h}{\partial \alpha'}h(Z;\alpha_0)\right]$$

is nonsingular.

We have

Lemma 3 *If assumption 12 holds, then $\hat{\alpha}$ defined in (21) is weakly consistent for α_0 and*

$$\sqrt{n}(\hat{\alpha}-\alpha_0) = \left(\mathbb{E}\left[\frac{\partial h}{\partial \alpha}h(Z;\alpha_0)\frac{\partial h}{\partial \alpha'}h(Z;\alpha_0)\right]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}V_j\frac{\partial h}{\partial \alpha}h(Z_j;\alpha_0)+o_p(1)$$

Proof. Because $\sup_{\alpha\in A}|h(Z;\alpha)| \leq M_0(Z)$ with $\mathbb{E}[M_0(Z)] < \infty$, we have that

$$\frac{1}{n}\sum_{j=1}^{n}(X_j - h(Z_j;\alpha))^2 \xrightarrow{p} \mathbb{E}(V^2) + \mathbb{E}\left[(h(Z;\alpha)-h(Z;\alpha_0)^2\right]$$

uniformly on A. Hence by the usual argument for M-estimators (see e.g. Van der Vaart ()), (23) implies that $\hat{\alpha} \xrightarrow{p} \alpha_0$. A first-order Taylor expansion of (21) around α_0 gives

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}V_j\frac{\partial h}{\partial \alpha}h(Z_j;\alpha_0) - \left(\frac{1}{n}\sum_{j}^{n}\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\frac{\partial h}{\partial \alpha'}(Z_j;\overline{\alpha}) - \frac{1}{n}\sum_{j}^{n}(X_j - h(Z_j;\overline{\alpha}))\frac{\partial^2 h}{\partial\alpha\partial\alpha'}(Z_j;\overline{\alpha})\right)\sqrt{n}(\hat{\alpha}-\alpha_0) = 0 \tag{24}$$

Because

$$\frac{1}{n}\sum_{j}^{n}\frac{\partial h}{\partial \alpha}(Z_j;\alpha)\frac{\partial h}{\partial \alpha'}(Z_j;\alpha) \xrightarrow{p} \mathbb{E}\left[\frac{\partial h}{\partial \alpha}h(Z;\alpha)\frac{\partial h}{\partial \alpha'}h(Z;\alpha)\right]$$

uniformly in A,

$$\frac{1}{n}\sum_{j}^{n}\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\frac{\partial h}{\partial \alpha'}(Z_j;\overline{\alpha}) \xrightarrow{p} \mathbb{E}\left[\frac{\partial h}{\partial \alpha}h(Z;\alpha_0)\frac{\partial h}{\partial \alpha'}h(Z;\alpha_0)\right]$$

Further

$$\frac{1}{n}\sum_{j}^{n}(X_j - h(Z_j;\overline{\alpha}))\frac{\partial^2 h}{\partial\alpha\partial\alpha'}(Z_j;\overline{\alpha}) = \frac{1}{n}\sum_{j}^{n}V_j\frac{\partial^2 h}{\partial\alpha\partial\alpha'}(Z_j;\overline{\alpha})$$
$$+\frac{1}{n}\sum_{j}^{n}(h(Z_j;\alpha_0) - h(Z_j;\overline{\alpha}))\frac{\partial^2 h}{\partial\alpha\partial\alpha'}(Z_j;\overline{\alpha})$$

Because

$$\sup_{\alpha\in A}\left|V\frac{\partial^2 h}{\partial\alpha\partial\alpha'}(Z;\alpha)\right| \leq |V|M_2(Z)$$

where $\mathbb{E}[|V|M_2(Z)] < \infty$, the first term on the right hand side converges to $\mathbb{E}\left[V\frac{\partial^2 h}{\partial\alpha\partial\alpha'}(Z;\alpha_0)\right] = 0$. Because

$$\sup_{\alpha\in A}|(h(Z_j;\alpha_0)-h(Z_j;\alpha))|\left|\frac{\partial^2 h}{\partial\alpha\partial\alpha'}(Z_j;\alpha)\right| \leq 2M_0(Z)M_2(Z)$$

and by Cauchy-Schwartz $\mathbb{E}[M_0(Z)M_2(Z)] < \infty$, the second term on the right hand side converges in probability to 0. Equation (24) can be solved for $\sqrt{n}(\hat{\alpha}-\alpha_0)$ if the matrix between parentheses H is nonsingular for which event we use the indicator I_{NS}. If the matrix is singular we set $\sqrt{n}(\hat{\alpha}-\alpha_0)$ arbitrarily equal to 0, so that

$$\sqrt{n}(\hat{\alpha}-\alpha_0) = I_{NS}H^{-1}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}V_j\frac{\partial h}{\partial \alpha}h(Z_j;\alpha_0)\right)$$

Because $I_{NS} \xrightarrow{p} 1$ and the inverse is continuous at a nonsingular matrix, the result follows by the continuous mapping and Slutsky theorems. ■

C The asymptotic distribution of the control variate estimator

C.1 Linearization

Because $L(X) = R(X;\tau_0)$ (22) can be rewritten as

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

We assume

Assumption 13 (Regression function and marginal distribution of the measurement error) *The nonlinear regression model for Y is*

$$Y = m(X^*;\theta_0)+\eta \qquad \mathbb{E}(\eta|X^*)=0$$

with $\theta_0 \in \Theta$ and X^* the latent true value of the regressor. For $R(x;\tau)$ defined in (19) we assume that for all $\delta > 0$, there is a $\zeta > 0$ such that

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

Also for $d = 0, 1, 2$

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

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 4 *If assumption 13 holds $a_0 > 5/2$, and $k = n^\kappa$ with $0 < \kappa < 1/7$, then $\hat{\tau}$ defined in (22) is weakly consistent for τ_0 and*

$$\sqrt{n}(\hat{\tau}-\tau_0)=\left(\mathbb{E}\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial R}{\partial \tau'}(X;\tau_0)\right]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{L}(X_i)-L(X_i)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)+o_p(1) \qquad (25)$$

Proof. We have

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

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

The final term is uniformly in τ bounded by

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

with $\mathcal{X}$ the support of the distribution of X and

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

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

with μ_k defined below (as is the other undefined notation). By lemma 7 below

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\hat{\mu}(x,\hat{V}_j)-\mu_k(x,\hat{V}_j))\right| \leq \sup_{w\in\widehat{\mathcal{W}}}|\hat{\mu}(w)-\mu_k(w)| = O_p(k^{7/2}n^{-1/2})+O_p(k^{5/2-a_0})$$

and by a first-order Taylor expansion and assumption 12

$$\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^{n}(\mu_k(x,\hat{V}_j)-\mu(x,V_j))\right| \leq \sup_{w\in\widehat{\mathcal{W}}}\left|\frac{\partial\mu_k}{\partial v}(w)\right|\frac{1}{n}\sum_{j=1}^{n}M_1(Z_j)|\hat{\alpha}-\alpha_0| = O_p(n^{-1/2})$$

and finally because $\frac{\partial\mu}{\partial x}(x,v)$ is continuous and hence bounded on $\mathcal{W}$

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

We conclude that because by assumption 13 $\sup_{\tau\in T}|R(X;\tau)|\leq N_0(X)$ with $\mathbb{E}[N_0(X)^2]<\infty$

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

uniformly in τ. Because the limit has a well-separated minimum in τ_0 weak consistency of $\hat{\tau}$ follows.

We have

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

so that by first-order Taylor series expansions with $\tilde{\tau}$ and $\overline{\tau}$ intermediate points

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

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

By assumption 13 and Newey (1991), corollary 3.1

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

uniformly for $\tau\in T$ and

$$\sup_{\tau\in T}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{L}(X_i)-L(X_i)\right)\frac{\partial^2 R}{\partial\tau\partial\tau'}(X_i;\tau)\right| \leq \sup_{x\in\mathcal{X}}|\hat{L}(x)-L(x)|\frac{1}{n}\sum_{i=1}^{n}N_2(X_i) = O_p(k^{7/2}n^{-1/2})+O_p(k^{5/2-a_0})$$

so that the conclusion follows. ■

C.2 Asymptotic linear representation

C.2.1 Decomposition and assumptions

For the rest of the proof we take, without loss of generality, α and τ as scalar. We consider for $\tau \in T$

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

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{\mu}(X_i,\hat{V}_j)-\tilde{\mu}(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)+ \tag{26}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\tilde{\mu}(X_i,V_j)-\mu(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0)+ \tag{27}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu(X_i,V_j)\frac{\partial R}{\partial \tau}(X_i;\tau_0)-\frac{1}{n}\sum_{i=1}^{n}L(X_i)\frac{\partial R}{\partial \tau}(X_i;\tau_0) \tag{28}$$

with $\tilde{\mu}$ the infeasible nonparametric regression estimator defined below. We write the three expressions (26), (27) and (28) as normalized sample averages. The three expressions each have a contribution to the influence function of our estimator. The contribution of (26) accounts for the estimation of the residuals in the first stage, the contribution of (27) accounts for the variability of the nonparametric regression estimator, and the contribution of (28) is the pure variance term. By assumption 13 the representation will hold uniformly in $\tau \in T$ and hence we can substitute $\hat{\tau}$ for τ.

Equation (26) involves the feasible nonparametric regression estimator $\hat{\mu}$ 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 function in the series. To include all powers of x and v up to order n, we need to include $k=\frac{1}{2}(n+1)(n+2)$ terms. The resulting k basis functions are denoted by the k vector $P_k(w)=(x^{\lambda_1}w^{\lambda_2},\lambda_1+\lambda_2\leq n)$. 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 14 (Support) *The support of X, V is $\mathcal{W}=\mathcal{X}\times\mathcal{V}=[x_L,x_U]\times[v_L,v_U]$. The joint density of X, V is bounded from 0 on $\mathcal{W}$ and is r times continuously differentiable on its support.*

By lemma A.15 in Newey (1995) this assumption implies that for all k there is a nonsingular matrix A_k such that if we define $\tilde{Q}_k(w)=A_kP_k(w)$, the smallest eigenvalue of $\Omega_k=\mathbb{E}[\tilde{Q}_k(W)\tilde{Q}_k(W)']$ satisfies $\lambda_{min}(\Omega_k)\geq C>0$ for all k. To simplify some of the argument we choose $Q_k(w)=\Omega_k^{-1/2}\tilde{Q}_k(w)$ so that $\mathbb{E}[Q_k(W)Q_k(W)']=I_k$. The same lemma in Newey (1995) gives the following bounds on the vector of basis functions for $d=0,1,\ldots$

$$\sup_{w\in\mathcal{W}}\left|\frac{\partial^d Q_k}{\partial w^d}(w)\right|=O(k^{d+1})\qquad \sup_{w\in\widehat{\mathcal{W}}}\left|\frac{\partial^d Q_k}{\partial w^d}(w)\right|=O(k^{d+1})$$

with $\widehat{\mathcal{W}}$ the support of $\hat{W}$ that by assumption 12 is also bounded so that Newey's lemma applies.

In (26) the nonparametric regression estimator $\hat{\mu}(X_i,\hat{V}_j)$ estimates $\mu(X_i,\hat{V}_j)$ with $\mu(x,v)=\mathbb{E}[Y|X=x,V=v]$. However the target is not well-defined if $\hat{V}$ is not in the support of V. If $\mu(x,v)=Q_k(x,v)'\gamma_k$

for some k, then we can extend the definition from the support of V to that of $\hat{V}$. In general we define

$$\begin{aligned}\mu_k(X,\hat{V}) &= \mu(X,\hat{V}) && v_L \leq \hat{V} \leq v_U \\ &= Q_k(X,\hat{V})'\gamma_k + \mu(X,v_U) - Q_k(X,v_U)'\gamma_k + \sum_{l=1}^{L}\frac{1}{l!}\left(\frac{\partial^l\mu}{\partial v^l}(X,v_U) - \frac{\partial^l Q_k}{\partial v^l}(X,v_U)'\gamma_k\right)(\hat{V}-v_U)^l && \hat{V} > v_U \\ &= Q_k(X,\hat{V})'\gamma_k + \mu(X,v_L) - Q_k(X,v_L)'\gamma_k + \sum_{l=1}^{L}\frac{1}{l!}\left(\frac{\partial^l\mu}{\partial v^l}(X,v_L) - \frac{\partial^l Q_k}{\partial v^l}(X,v_L)'\gamma_k\right)(\hat{V}-v_L)^l && \hat{V} < v_L\end{aligned}$$

This extension L times continuously differentiable on the support of $\hat{V}$ and in particular in v_L and v_U. In the sequel we take $L = 2$ so that we can do second-order Taylor expansions. The vector γ_k is given in the next assumption.

Assumption 15 (Regression function) *There is a vector γ_k such that for some constants $C_d, a_d, d = 0, 1, \ldots, D$*

$$\sup_{w\in\mathcal{W}}\left|\frac{\partial^d\mu}{\partial v^d}(w) - \frac{\partial Q_k}{\partial v}(w)'\gamma_k\right| \leq C_d k^{-a_d}$$

If the population regression function $\mu(w)$ is s times continuously differentiable on $\mathcal{W}$ we have by Lorentz (1986), chapter 6, theorem 8 that $a_0 = s/2$. In the sequel we use $D = 2$, so that we assume the existence of uniform approximations of μ and its first and second derivative.

We have that if we denote the support of $\hat{W}$ by $\widehat{\mathcal{W}}$

$$\sup_{w\in\widehat{\mathcal{W}}}\left|\frac{\partial^d\mu_k}{\partial v^d}(w) - \frac{\partial Q_k}{\partial v}(w)'\gamma_k\right| \leq C_d k^{-a_d}$$

Therefore $\mu_k(x,v)$ and its derivatives with respect to v can be approximated on $\widehat{\mathcal{W}}$ with a polynomial that has the same coefficients γ_k as in assumption 15, i.e. as in the approximation of $\mu(x,v)$ (and its derivatives with respect to v) on $\mathcal{W}$.

C.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 5 *If assumption 14 holds then*

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

and

$$|\hat{\Omega}_k - I_k| = O_p\left(k^{3}n^{-1/2}\right)$$

Proof. By a first order Taylor expansion and the triangle inequality

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

$$\left| \frac{1}{n} \sum_{j=1}^{n} \frac{\partial Q_k}{\partial v}(X_j, \overline{V}_j) \frac{\partial Q_k}{\partial v}(X_j, \overline{V}_j)'(\hat{V}_j - V_j)^2 \right| \leq \left| \frac{1}{n} \sum_{j=1}^{n} \frac{\partial Q_k}{\partial v}(X_j, \overline{V}_j) \frac{\partial Q_k}{\partial v}(X_j, \overline{V}_j)' \left(\frac{\partial h}{\partial \alpha}(Z_j; \overline{\alpha}) \right)^2 \right| (\hat{\alpha} - \alpha_0)^2 +$$

$$\left(\left| \frac{1}{n} \sum_{j=1}^{n} \frac{\partial Q_k}{\partial v}(X_j, \overline{V}_j) Q_k(X_j, V_j)' \frac{\partial h}{\partial \alpha}(Z_j; \overline{\alpha}) \right| + \left| \frac{1}{n} \sum_{j=1}^{n} Q_k(X_j, V_j) \frac{\partial Q_k}{\partial v}(X_j, \overline{V}_j)' \frac{\partial h}{\partial \alpha}(Z_j; \overline{\alpha}) \right| \right) |\hat{\alpha} - \alpha_0|$$

By assumptions 12 and 14 this expression is bounded by

$$\left(\frac{1}{n} \sum_{j=1}^{n} M_1(Z_j)^2 \right) \sup_{w \in \widehat{\mathcal{W}}} \left| \frac{\partial Q_k}{\partial v}(w) \right|^2 |\hat{\alpha} - \alpha_0|^2 +$$

$$\sup_{w \in \widehat{\mathcal{W}}} \left| \frac{\partial Q_k}{\partial v}(w) \right| \left(\frac{1}{n} \sum_{j=1}^{n} M_1(Z_j) |Q_k(W_j)| \right) |\hat{\alpha} - \alpha_0|$$

By Cauchy-Schwartz and assumptions 12 and 14

$$\frac{1}{n} \sum_{j=1}^{n} M_1(Z_j) |Q_k(W_j)| \leq \sqrt{\frac{1}{n} \sum_{j=1}^{n} M_1(Z_j)^2} \sqrt{\frac{1}{n} \sum_{j=1}^{n} |Q_k(X_j, V_j)|^2} = O_p(k)$$

so that the expression above is

$$O_p(k^4 n^{-1}) + O_p(k^3 n^{-1/2}) = O_p(k^3 n^{-1/2})$$

Next, as in Newey (1997), because $\mathbb{E}[Q_k(W_j) Q_k(W_j)'] = I_k$

$$\mathbb{E}\left[|\tilde{\Omega}_k - I_k|^2 \right] = \mathbb{E}\left[\text{tr}\left((\tilde{\Omega}_k - I_k)^2 \right) \right] = \text{tr}\left(\mathbb{E}\left[\left(\frac{1}{n} \sum_{j=1}^{n} [Q_k(W_j) Q_k(W_j)' - I_k] \right)^2 \right] \right) =$$

$$\frac{1}{n} \text{tr}\left(\mathbb{E}\left[\left(Q_k(W_j) Q_k(W_j)' - I_k \right)^2 \right] \right) = \frac{1}{n} \text{tr}\left(\mathbb{E}\left[Q_k(W_j) Q_k(W_j)' Q_k(W_j) Q_k(W_j)' \right] - I_k \right) \leq$$

$$\frac{1}{n} \mathbb{E}\left[Q_k(W_j)' Q_k(W_j) \text{tr}(Q_k(W_j) Q_k(W_j)') \right] \leq \frac{1}{n} \sup_{w \in \mathcal{W}} |Q_k(w)|^2 \text{tr}(I_k) = O(k^3 n^{-1})$$

so that by $\mathbb{E}\left[|\tilde{\Omega}_k - I_k| \right] \leq \sqrt{\mathbb{E}\left[|\tilde{\Omega}_k - I_k|^2 \right]}$ and the Markov inequality

$$|\tilde{\Omega}_k - I_k| = O_p(k^{3/2} n^{-1/2})$$

The conclusion now follows from the triangle inequality. ■

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

Assumption 16 (Variance)

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

Define

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

By assumption 16

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

Lemma 6 *If assumptions 14 and 16 hold then*

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

and

$$\left|\frac{1}{n}\sum_{j=1}^{n} Q_k(\hat{W}_j)U_j\right| = O_p(k^{2}n^{-1/2})$$

Proof. We have by first-order Taylor series expansions

$$\left|\frac{1}{n}\sum_{j=1}^{n} Q_k(X_j,\hat{V}_j)U_j\right| \leq \left|\frac{1}{n}\sum_{j=1}^{n} Q_k(X_j,V_j)U_j\right| + \left|\frac{1}{n}\sum_{j=1}^{n} \frac{\partial Q_k}{\partial v}(X_j,\overline{V}_j)\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})U_j\right| |\hat{\alpha}-\alpha_0|$$

Now

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

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

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

$$\left|\frac{1}{n}\sum_{j=1}^{n} \frac{\partial Q_k}{\partial v}(X_j,\overline{V}_j)\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})U_j\right| \leq \sup_{w\in\widehat{\mathcal{W}}}\left|\frac{\partial Q_k}{\partial v}(w)\right| \frac{1}{n}\sum_{j=1}^{n} M_1(Z_j)|U_j| = O_p(k^2)$$

because by Cauchy-Schwartz $\mathbb{E}[M_1(Z_j)|U_j|] \leq \sqrt{\mathbb{E}[U^2]}\sqrt{\mathbb{E}[M_1(Z_j)^2]} < \infty$, so that

$$\left|\frac{1}{n}\sum_{j=1}^{n} \frac{\partial Q_k}{\partial v}(X_j,\overline{V}_j)\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})U_j\right| |\hat{\alpha}-\alpha_0| = O_p(k^2 n^{-1/2})$$

■

Define $1_n = 1\left(\lambda_{min}\left(\hat{\Omega}_k\right) > 1/2\right)$ with $1(.)$ the indicator of the event between parentheses. The coefficients in the series estimator 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 $Q_k(X, V)$

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

with $\tilde{1}_n = 1\left(\lambda_{min}\left(\tilde{\Omega}_k\right) > 1/2\right)$

Lemma 7 *If assumptions 14, 16 , and 15 hold, then*

$$\tilde{1}_n |\tilde{\gamma}_k - \gamma_k| = O_p(k n^{-1/2}) + O_p(k^{3/2 - a_0})$$

and

$$1_n |\hat{\gamma}_k - \gamma_k| = O_p(k^{5/2} n^{-1/2}) + O_p(k^{3/2 - a_0})$$

If in addition $k^3 n^{-1/2} \to 0$, then

$$\left| \hat{\gamma}_k - \tilde{\gamma}_k - (\hat{\alpha} - \alpha_0) \frac{1}{n} \sum_{j=1}^n Q_k(W_j) \frac{\partial \mu}{\partial v}(W_j) \frac{\partial h}{\partial \alpha}(Z_j; \alpha_0) \right| = O_p(k^{9/2} n^{-1}) + O(k^{3/2 - a_0})$$

For $\hat{\mu}(w) = Q_k(w)'\hat{\gamma}_k$ and $\frac{\partial \hat{\mu}}{\partial v}(w) = \frac{\partial Q_k}{\partial v}(w)'\hat{\gamma}_k$

$$\sup_{w \in \widehat{\mathcal{W}}} |\hat{\mu}(w) - \mu_k(w)| = O_p(k^{7/2} n^{-1/2}) + O_p(k^{5/2 - a_0})$$

and

$$\sup_{w \in \widehat{\mathcal{W}}} |\frac{\partial \hat{\mu}}{\partial v}(w) - \frac{\partial \mu_k}{\partial v}(w)| = O_p(k^{9/2} n^{-1/2}) + O_p(k^{7/2 - a_0})$$

and for $\tilde{\mu}(w) = Q_k(w)'\tilde{\gamma}_k$ and $\frac{\partial \tilde{\mu}}{\partial v}(w) = \frac{\partial Q_k}{\partial v}(w)'\tilde{\gamma}_k$

$$\sup_{w \in \widehat{\mathcal{W}}} |\tilde{\mu}(w) - \mu_k(w)| = O_p(k^2 n^{-1/2}) + O_p(k^{5/2 - a_0})$$

and

$$\sup_{w \in \widehat{\mathcal{W}}} |\frac{\partial \tilde{\mu}}{\partial v}(w) - \frac{\partial \mu_k}{\partial v}(w)| = O_p(k^3 n^{-1/2}) + O_p(k^{7/2 - a_0})$$

Proof. We have

$$\tilde{1}_n(\tilde{\gamma}_k - \gamma_k) = \tilde{1}_n \tilde{\Omega}_k^{-1} \frac{1}{n} \sum_{j=1}^n Q_k(W_j)(Y_j - \mu(W_j)) + \tilde{1}_n \tilde{\Omega}_k^{-1} \frac{1}{n} \sum_{j=1}^n Q_k(W_j)(\mu(W_j) - Q_k(W_j)'\gamma_k)$$

so that because for any square, symmetric and positive definite matrix of order k D we have $\sqrt{k}\lambda_{min}(D) \leq |D| \leq \sqrt{k}\lambda_{max}(D)$

$$\tilde{1}_n|\tilde{\gamma}_k-\gamma_k| \leq \tilde{1}_n|\tilde{\Omega}_k^{-1}| \left|\frac{1}{n}\sum_{j=1}^n Q_k(W_j)U_j\right| + \tilde{1}_n|\tilde{\Omega}_k^{-1}| \sup_{w\in\mathcal{W}} |Q_k(w)| \sup_{w\in\mathcal{W}} |\mu(w)-Q_k(w)'\gamma_k| = O_p(kn^{-1/2})+O_p(k^{3/2-a_0})$$

Note that if $kn^{-1/2} \to 0$ we can omit $\tilde{1}_n$.

Next

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

$$1_n\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^n Q_k(\hat{W}_j)U_j + 1_n\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^n Q_k(\hat{W}_j)(\mu(W_j)-\mu_k(\hat{W}_j)) + 1_n\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^n Q_k(\hat{W}_j)(\mu_k(\hat{W}_j)-Q_k(\hat{W}_j)'\gamma_k) \tag{29}$$

We use (29) to derive two results: the rate of convergence of $\hat{\gamma}_k - \gamma_k$ and the rate of convergence of the difference $\hat{\gamma}_k - \tilde{\gamma}_k$. By lemma 6 and assumption 14, the first term in (29) is $O_p(k^{5/2}n^{-1/2})$. By a first-order Taylor series expansion the second term can be expressed as

$$(\hat{\alpha} - \alpha_0)1_n\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^n Q_k(\hat{W}_j)\frac{\partial \mu_k}{\partial v}(\overline{W}_j)\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})$$

which is bounded by

$$2|\hat{\alpha} - \alpha_0|\sqrt{k} \sup_{w\in\widehat{\mathcal{W}}} |Q_k(w)| \sup_{w\in\widehat{\mathcal{W}}} \left|\frac{\partial \mu_k}{\partial v}(w)\right| \frac{1}{n}\sum_{j=1}^n M_1(Z_j) = O_p(k^{3/2}n^{-1/2})$$

because $\mu_k(w)$ is continuously differentiable. The third term in (29) is bounded by

$$2\sqrt{k} \sup_{w\in\widehat{\mathcal{W}}} |Q_k(w)| \sup_{w\in\widehat{\mathcal{W}}} \left|\mu_k(w) - Q_k(w)'\gamma_k\right| = O(k^{3/2-a_0})$$

Therefore

$$1_n|\hat{\gamma}_k - \gamma_k| = O_p(k^{5/2}n^{-1/2}) + O(k^{3/2-a_0})$$

Next we consider $\hat{\gamma}_k - \tilde{\gamma}_k$. If $k^3n^{-1/2} \to 0$ then we can omit 1_n and $\tilde{1}_n$ in the expressions. A second-order Taylor expansion gives for the first term of (29)

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

$$\frac{1}{2}\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^n \frac{\partial^2 Q_k}{\partial v^2}(X_j,\overline{V}_j))U_j(\hat{V}_j - V_j)^2 =$$

$$\tilde{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^n Q_k(W_j)U_j + \hat{\Omega}_k^{-1}\tilde{\Omega}_k^{-1}(\tilde{\Omega}_k - \hat{\Omega}_k)\frac{1}{n}\sum_{j=1}^n Q_k(W_j)U_j +$$

$$\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_k}{\partial v}(W_j))U_j(\hat{V}_j-V_j)+\frac{1}{2}\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2 Q_k}{\partial v^2}(X_j,\overline{V}_j))U_j(\hat{V}_j-V_j)^2=$$

$$\tilde{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)U_j+\tilde{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)(\mu(W_j)-Q_k(W_j)'\gamma_k)$$

$$-\tilde{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)(\mu(W_j)-Q_k(W_j)'\gamma_k)+\hat{\Omega}_k^{-1}\tilde{\Omega}_k^{-1}(\tilde{\Omega}_k-\hat{\Omega}_k)\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)U_j-$$

$$(\hat{\alpha}-\alpha_0)\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_k}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)U_j-\frac{(\hat{\alpha}-\alpha_0)^2}{2}\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_k}{\partial v}(W_j))U_j\frac{\partial^2 h}{\partial \alpha^2}(Z_j;\overline{\alpha})+$$

$$\frac{(\hat{\alpha}-\alpha_0)^2}{2}\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2 Q_k}{\partial v^2}(X_j,\overline{V}_j))U_j\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2=$$

$$(\tilde{\gamma}_k-\gamma_k)+O(k^{3/2-a_0})+O_p(k^{9/2}n^{-1})+O_p(k^{5/2}n^{-1})+O_p(k^{3/2}n^{-1})+O_p(k^{7/2}n^{-1})$$

by lemmas 5, 6 and assumptions 12 and 14 and the fact that

$$\mathbb{E}\left[\frac{\partial Q_k}{\partial v}(W)\frac{\partial Q_k}{\partial v}(W)'\left(\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right)^2U^2\right]\leq C\left(\sup_{w\in\mathcal{W}}\left|\frac{\partial Q_k}{\partial v}(w)\right|\right)^2\mathbb{E}[U^2]=O(k^4)$$

The second term in (29) is, if $k^3n^{-1/2}\to 0$

$$(\hat{\alpha}-\alpha_0)\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(\hat{W}_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)+\frac{(\hat{\alpha}-\alpha_0)^2}{2}\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(\hat{W}_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial^2 h}{\partial \alpha^2}(Z_j;\overline{\alpha})-$$

$$\frac{(\hat{\alpha}-\alpha_0)^2}{2}\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(\hat{W}_j)\frac{\partial^2 \mu_k}{\partial v^2}(X_j,\overline{V}_j))\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2=$$

$$(\hat{\alpha}-\alpha_0)\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(\hat{W}_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)+O_p(k^{3/2}n^{-1})+O_p(k^{3/2}n^{-1})$$

and

$$(\hat{\alpha}-\alpha_0)\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}Q_k(\hat{W}_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)=(\hat{\alpha}-\alpha_0)\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)-$$

$$(\hat{\alpha}-\alpha_0)\hat{\Omega}_k^{-1}(\hat{\Omega}_k-I_k)\frac{1}{n}\sum_{j=1}^{n}Q_k(\hat{W}_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)-$$

$$(\hat{\alpha}-\alpha_0)^2\hat{\Omega}_k^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial Q_k}{\partial v}(X_j,\overline{V}_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})=$$

$$(\hat{\alpha}-\alpha_0)\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)+O_p(k^{9/2}n^{-1})+O_p(k^{5/2}n^{-1})$$

Finally, the third term in (29) is, if $k^3 n^{-1/2} \to 0$

$$\left| 1_n \hat{\Omega}_k^{-1} \frac{1}{n} \sum_{j=1}^{n} Q_k(\hat{W}_j)(\mu_k(\hat{W}_j) - Q_k(\hat{W}_j)'\gamma_k) \right| \leq 2\sqrt{k} \sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)| \sup_{w \in \widehat{\mathcal{W}}} |\mu_k(w) - Q_k(w)'\gamma_k| = O(k^{3/2-a_0})$$

Collecting results we conclude that, if if $k^3 n^{-1/2} \to 0$

$$\hat{\gamma}_k - \gamma_k = (\tilde{\gamma}_k - \gamma_k) + (\hat{\alpha} - \alpha_0) \frac{1}{n} \sum_{j=1}^{n} Q_k(W_j) \frac{\partial \mu}{\partial v}(W_j) \frac{\partial h}{\partial \alpha}(Z_j; \alpha_0) + O_p(k^{9/2} n^{-1}) + O(k^{3/2-a_0})$$

Note that the remainder terms are of order n^{-1} (or a pure bias term) while the main terms are of order $n^{-1/2}$.

Next

$$\sup_{w \in \widehat{\mathcal{W}}} |\hat{\mu}(w) - \mu_k(w)| \leq \sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)'(\hat{\gamma}_k - \gamma_k)| + \sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)'\gamma_k - \mu_k(w)| \leq$$

$$\sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)||\hat{\gamma}_k - \gamma_k| + \sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)'\gamma_k - \mu_k(w)| =$$

$$O_p(k^{7/2} n^{-1/2}) + O_p(k^{5/2-a_0}) + O_p(k^{-a_0}) = O_p(k^{7/2} n^{-1/2}) + O_p(k^{5/2-a_0})$$

and

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

$$\sup_{w \in \widehat{\mathcal{W}}} \left| \frac{\partial Q_k}{\partial v}(w)'(\hat{\gamma}_k - \gamma_k) \right| + \sup_{w \in \widehat{\mathcal{W}}} \left| \frac{\partial Q_k}{\partial v}(w)'\gamma_k - \frac{\partial \mu_k}{\partial v}(w) \right| =$$

$$O_p(k^{9/2} n^{-1/2}) + O_p(k^{7/2-a_0}) + O_p(k^{-a_1}) = O_p(k^{9/2} n^{-1/2}) + O_p(k^{7/2-a_0})$$

Finally

$$\sup_{w \in \widehat{\mathcal{W}}} |\tilde{\mu}(w) - \mu_k(w)| \leq \sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)'(\tilde{\gamma}_k - \gamma_k)| + \sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)'\gamma_k - \mu_k(w)| \leq$$

$$\sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)||\tilde{\gamma}_k - \gamma_k| + \sup_{w \in \widehat{\mathcal{W}}} |Q_k(w)'\gamma_k - \mu_k(w)| =$$

$$O_p(k^2 n^{-1/2}) + O_p(k^{5/2-a_0}) + O_p(k^{-a_0}) = O_p(k^2 n^{-1/2}) + O_p(k^{5/2-a_0})$$

and

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

$$\sup_{w \in \widehat{\mathcal{W}}} \left| \frac{\partial Q_k}{\partial v}(w)'(\tilde{\gamma}_k - \gamma_k) \right| + \sup_{w \in \widehat{\mathcal{W}}} \left| \frac{\partial Q_k}{\partial v}(w)'\gamma_k - \frac{\partial \mu_k}{\partial v}(w) \right| =$$

$$O_p(k^3 n^{-1/2}) + O_p(k^{7/2-a_0}) + O_p(k^{-a_1}) = O_p(k^3 n^{-1/2}) + O_p(k^{7/2-a_0})$$

■

The next step is to consider (26), (27) and (28).

C.3 Expressing (26) as a sample average

We express (26) as

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

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\hat{\gamma}_k'(Q_k(X_i,\hat{V}_j)-Q_k(X_i,V_j))+ \tag{30}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_k-\tilde{\gamma}_k)'Q_k(X_i,V_j) \tag{31}$$

For (30) by a second order Taylor expansion with respect to $\hat{V}_j$ followed by a second order Taylor series expansion of the term in $\hat{V}_j - V_j$ and substitution of the first order Taylor series expansion $\hat{V}_j - V_j = -\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha}_j)$ in $(\hat{V}_j-V_j)^2$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\hat{\gamma}_k'(Q_k(X_i,\hat{V}_j)-Q_k(X_i,V_j))=-\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\gamma_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)(\hat{\alpha}-\alpha_0)-$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_k-\gamma_k)'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)(\hat{\alpha}-\alpha_0)-$$

$$\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\hat{\gamma}_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial^2 h}{\partial \alpha^2}(Z_j;\overline{\alpha})(\hat{\alpha}-\alpha_0)^2+$$

$$\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_k-\gamma_k)'\frac{\partial^2 Q_k}{\partial v^2}(X_i,\overline{V}_j)\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2(\hat{\alpha}-\alpha_0)^2$$

$$\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(\gamma_k'\frac{\partial^2 Q_k}{\partial v^2}(X_i,\overline{V}_j)-\frac{\partial^2 \mu_k}{\partial v^2}(X_i,\overline{V}_j)\right)\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2(\hat{\alpha}-\alpha_0)^2+$$

$$\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial^2 \mu_k}{\partial v^2}(X_i,\overline{V}_j)\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2(\hat{\alpha}-\alpha_0)^2$$

Now

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_k-\gamma_k)'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)(\hat{\alpha}-\alpha_0)\right|\leq$$

$$\sqrt{n}|\hat{\alpha}-\alpha_0|\sup_{w\in\mathcal{W}}\left|\frac{\partial Q_k}{\partial v}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}M_1(Z_j)\right)|\hat{\gamma}_k-\gamma_k|=O_p(k^{9/2}n^{-1/2})+O(k^{7/2-a_0})$$

and

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\hat{\gamma}_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial^2 h}{\partial \alpha^2}(Z_j;\overline{\alpha})(\hat{\alpha}-\alpha_0)^2\right|\leq$$

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_k-\gamma_k)'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial^2 h}{\partial \alpha^2}(Z_j;\overline{\alpha})(\hat{\alpha}-\alpha_0)^2\right|+$$

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(\gamma_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)-\frac{\partial \mu}{\partial v}(X_i,V_j)\right)\frac{\partial^2 h}{\partial \alpha^2}(Z_j;\overline{\alpha})(\hat{\alpha}-\alpha_0)^2\right|+$$

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}(X_i,V_j)\frac{\partial^2 h}{\partial \alpha^2}(Z_j;\overline{\alpha})(\hat{\alpha}-\alpha_0)^2\right|\leq$$

$$\frac{1}{2\sqrt{n}}n|\hat{\alpha}-\alpha_0|^2\sup_{w\in\mathcal{W}}\left|\frac{\partial Q_k}{\partial v}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}M_2(Z_j)\right)|\hat{\gamma}_k-\gamma_k|+$$

$$\frac{1}{2\sqrt{n}}n|\hat{\alpha}-\alpha_0|^2\sup_{w\in\mathcal{W}}\left|\gamma_k'\frac{\partial Q_k}{\partial v}(w)-\frac{\partial \mu}{\partial v}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}M_2(Z_j)\right)+$$

$$\frac{1}{2\sqrt{n}}n|\hat{\alpha}-\alpha_0|^2\sup_{w\in\mathcal{W}}\left|\frac{\partial \mu}{\partial v}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}M_2(Z_j)\right)=$$

$$O_p(k^{9/2}n^{-1})+O(k^{7/2-a_0}n^{-1/2})+O_p(n^{-1/2}k^{-a_1})+O_p(n^{-1/2})$$

and

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)(\hat{\gamma}_k-\gamma_k)'\frac{\partial^2 Q_k}{\partial v^2}(X_i,\overline{V}_j)\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2(\hat{\alpha}-\alpha_0)^2\right|\leq$$

$$\frac{1}{2\sqrt{n}}n|\hat{\alpha}-\alpha_0|^2\sup_{w\in\mathcal{W}}\left|\frac{\partial^2 Q_k}{\partial v^2}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}M_1(Z_j)^2\right)|\hat{\gamma}_k-\gamma_k|=O_p(k^{11/2}n^{-1})+O(k^{9/2-a_0}n^{-1/2})$$

and

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(\gamma_k'\frac{\partial^2 Q_k}{\partial v^2}(X_i,\overline{V}_j)-\frac{\partial^2 \mu_k}{\partial v^2}(X_i,\overline{V}_j)\right)\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2(\hat{\alpha}-\alpha_0)^2\right|\leq$$

$$\frac{1}{2\sqrt{n}}n|\hat{\alpha}-\alpha_0|^2\sup_{w\in\mathcal{W}}\left|\gamma_k'\frac{\partial^2 Q_k}{\partial v^2}(w)-\frac{\partial^2 \mu_k}{\partial v^2}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}M_1(Z_j)^2\right)=O_p(n^{-1/2}k^{-a_2})$$

and

$$\left|\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial^2 \mu_k}{\partial v^2}(X_i,\overline{V}_j)\left(\frac{\partial h}{\partial \alpha}(Z_j;\overline{\alpha})\right)^2(\hat{\alpha}-\alpha_0)^2\right|\leq$$

$$\frac{1}{2\sqrt{n}}n|\hat{\alpha}-\alpha_0|^2\sup_{w\in\mathcal{W}}\left|\frac{\partial^2 \mu_k}{\partial v^2}(w)\right|\left(\frac{1}{n}\sum_{i=1}^{n}N_1(X_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}M_1(Z_j)^2\right)=O_p(n^{-1/2})$$

Finally

$$-\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)\gamma_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)(\hat{\alpha}-\alpha_0)=$$

$$-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\sqrt{n}(\hat{\alpha}-\alpha_0)-$$

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\gamma_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\right)\sqrt{n}(\hat{\alpha}-\alpha_0)$$

with

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\gamma_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\right)=$$

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\frac{\partial R}{\partial \tau}(X_i;\tau_0)\left(\gamma_k'\frac{\partial Q_k}{\partial v}(X_i,V_j)-\frac{\partial \mu}{\partial v}(X_i,V_j)\right)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)+$$

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\frac{\partial R}{\partial \tau}(X_i;\tau_0)\frac{\partial \mu}{\partial v}(X_i,V_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\right)$$

and these terms are $O_p(k^{-a_1})$ and $O_p(n^{-1/2})$ respectively.

Combining the results we have for (30)

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

$$-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\sqrt{n}(\hat{\alpha}-\alpha_0)+O_p(k^{9/2}n^{-1/2})+O_p(k^{7/2-a_0})$$

where we only show the order of the largest remainders.

For (31) we find after substitution from lemma 7

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

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_k(X_i,V_j)'\cdot\frac{1}{n}\sum_{j=1}^n Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\sqrt{n}(\hat{\alpha}-\alpha_0)+O_p(k^{11/2}n^{-1})+O_p(k^{5/2-a_0})=$$

$$\left(\frac{1}{n}\sum_{i=1}^n\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_k(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^n Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)+$$

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

$$\left(\frac{1}{n}\sum_{j=1}^n Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)+O_p(k^{11/2}n^{-1})+O_p(k^{5/2-a_0})$$

Now

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

$$\left|\frac{1}{n}\sum_{j=1}^n Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right|\sqrt{n}|\hat{\alpha}-\alpha_0|\leq \tag{32}$$

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

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

$$\left|\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right|\sqrt{n}|\hat{\alpha}-\alpha_0|$$

By the V-statistic projection theorem

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

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

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

The remainder may grow with k but at a slower rate than the projection itself. We have

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

and

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

Together with

$$\left|\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right|\leq\sup_{w\in\mathcal{W}}|Q_k(w)|\sup_{w\in\mathcal{W}}\left|\frac{\partial \mu}{\partial v}(w)\right|\frac{1}{n}\sum_{j=1}^{n}M_1(Z_j)=O_p(k)$$

this implies that (32) is $O_p(k^2n^{-1/2})$.

Therefore (31) is

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

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_k(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)+$$

$$O_p(k^2n^{-1/2})+O_p(k^{5/2-a_0})$$

Because by assumptions 15 and 14 the function $\frac{f(x)f(v)}{f(x,v)}\frac{\partial R}{\partial \tau}(x;\tau_0)$ is r times continuously differentiable we have that for some vector δ_k

$$\sup_{w\in\mathcal{W}}\left|\frac{f(x)f(v)}{f(x,v)}\frac{\partial R}{\partial \tau}(x;\tau_0)-\delta_k' Q_k(x,v)\right|\leq Ck^{-r/2} \tag{33}$$

Hence

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_k(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)=$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\delta_k' Q_k(X_i,V_i)Q_k(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)+$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial \tau}(X_i;\tau_0)-\delta_k' Q_k(X_i,V_i)\right)Q_k(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)$$

where the final term is $O_p(k^{(4-r)/2})$. The final step is to transfer δ_k from the first to the second sum

$$\left(\frac{1}{n}\sum_{i=1}^{n}\delta_k' Q_k(X_i,V_i)Q_k(X_i,V_i)'\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)=$$

$$\left(\frac{1}{n}\sum_{j=1}^{n}\delta_k' Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)+\delta_k'\left(\tilde{\Omega}_k-I_k\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)$$

with because $|AB|=|BA|\leq|A||B|$ if both AB and BA are well-defined

$$\left|\delta_k'\left(\tilde{\Omega}_k-I_k\right)\left(\frac{1}{n}\sum_{j=1}^{n}Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)\right|\leq$$

$$\left|\tilde{\Omega}_k-I_k\right|\left|\frac{1}{n}\sum_{j=1}^{n}\delta_k' Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right|\sqrt{n}|\hat{\alpha}-\alpha_0|\leq$$

$$\left|\tilde{\Omega}_k-I_k\right|\left|\frac{1}{n}\sum_{j=1}^{n}\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right|\sqrt{n}|\hat{\alpha}-\alpha_0|+$$

$$\left|\tilde{\Omega}_k-I_k\right|\left|\frac{1}{n}\sum_{j=1}^{n}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)-\delta_k' Q_k(W_j)\right)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right|\sqrt{n}|\hat{\alpha}-\alpha_0|\leq$$

$$\left|\tilde{\Omega}_k-I_k\right|\sup_{w\in\mathcal{W}}\left|\frac{\partial \mu}{\partial v}(w)\right|\left(\frac{1}{n}\sum_{j=1}^{n}\frac{f(X_j)f(V_j)}{f(X_j,V_j)}N_1(X_j)M_1(Z_j)+\sup_{w\in\mathcal{W}}\left|\frac{f(x)f(v)}{f(w)}\frac{\partial R}{\partial \tau}(x;\tau_0)-\delta_k' Q_k(w)\right|\frac{1}{n}\sum_{j=1}^{n}M_1(Z_j)\right)\cdot$$

$$\sqrt{n}|\hat{\alpha}-\alpha_0|=O_p(n^{-1/2}k^{3/2})+O_p(n^{-1/2}k^{(3-r)/2})$$

because $\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}N_1(X)M_1(Z)\right] = \mathbb{E}[M_1(Z)]\mathbb{E}[N_1(X)] < \infty$, so that for (31)

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

$$\left(\frac{1}{n}\sum_{j=1}^{n}\delta_k' Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0) + O_p(n^{-1/2}k^2) + O_p(k^{5/2-a_0}) + O_p(k^{(4-r)/2})$$

Finally we obtain

$$\left(\frac{1}{n}\sum_{j=1}^{n}\delta_k' Q_k(W_j)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0) =$$

$$\left(\frac{1}{n}\sum_{j=1}^{n}\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0) -$$

$$\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{f(X_j)f(V_j)}{f(X_j,V_j)}\frac{\partial R}{\partial \tau}(X_j;\tau_0) - \delta_k' Q_k(W_j)\right)\frac{\partial \mu}{\partial v}(W_j)\frac{\partial h}{\partial \alpha}(Z_j;\alpha_0)\right)\sqrt{n}(\hat{\alpha}-\alpha_0)$$

where the final term is $O_p(k^{-r/2})$ by (33).

Therefore the final result is that (31)

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial R}{\partial \tau}(X_i;\tau_0)Q_k(X_i,V_j)'(\hat{\gamma}_k-\tilde{\gamma}_k) = \mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}(W)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\sqrt{n}(\hat{\alpha}-\alpha_0) +$$

$$O_p(n^{-1/2}k^2) + O_p(k^{5/2-a_0}) + O_p(k^{(4-r)/2})$$

Combining this with the result on (31) we have

Lemma 8 *If assumptions 12, 13, 14, 15 and 16 hold, then*

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

$$\left(\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}(W)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right] - \mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\right)\sqrt{n}(\hat{\alpha}-\alpha_0) +$$

$$O_p(k^{9/2}n^{-1/2}) + O_p(k^{7/2-a_0}) + O_p(k^{(4-r)/2})$$

C.4 Expressing (27) as a sample average

We decompose (27) as

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

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(Q_k(X_i,V_j)'\gamma_k-\mu(X_i,V_j)\right)\frac{\partial R}{\partial \tau}(X_i;\tau_0) + \tag{34}$$

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n\left(Q_k(X_i,V_j)-Q_k(X_i,V_i)\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\right)'(\tilde{\gamma}_k-\gamma_k)\frac{\partial R}{\partial\tau}(X_i;\tau_0)+ \tag{35}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i-\mu(X_i,V_i))\frac{\partial R}{\partial\tau}(X_i;\tau_0)+ \tag{36}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(\mu(X_i,V_i)-Q_k(X_i,V_i)'\gamma_k)\frac{\partial R}{\partial\tau}(X_i;\tau_0)- \tag{37}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{f(X_i)f(V_i)}{f(X_i,V_i)}(Y_i-Q_k(X_i,V_i)'\tilde{\gamma}_k)\frac{\partial R}{\partial\tau}(X_i;\tau_0) \tag{38}$$

with (36) the main term.

The bias remainder (34) is bounded by

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n\left(Q_k(X_i,V_j)'\gamma_k-\mu(X_i,V_j)\right)\frac{\partial R}{\partial\tau}(X_i;\tau_0)\right|\le\sqrt{n}\sup_{w\in\mathcal{W}}\left|Q_k(w)'\gamma_k-\mu(w)\right|\frac{1}{n}\sum_{i=1}^nN_1(X_i)=O_p(n^{1/2}k^{-a_0})$$

The remainder (35) is bounded by

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n\frac{\partial R}{\partial\tau}(X_i;\tau_0)\left(Q_k(X_i,V_j)-Q_k(X_i,V_i)\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\right)\right||\tilde{\gamma}_k-\gamma_k|$$

The first factor is the norm of a V-statistic that is equal to the projection and a term that is of smaller order

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

We have

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mathbb{E}\left[\frac{\partial R}{\partial\tau}(X_i;\tau_0)Q_k(X_i,V)\right]-\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial\tau}(X_i;\tau_0)Q_k(X_i,V_i)\right)\right|\right]\le$$
$$\sqrt{\mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mathbb{E}\left[\frac{\partial R}{\partial\tau}(X_i;\tau_0)Q_k(X_i,V)\right]-\frac{f(X_i)f(V_i)}{f(X_i,V_i)}\frac{\partial R}{\partial\tau}(X_i;\tau_0)Q_k(X_i,V_i)\right)\right|^2\right]}\le$$
$$\sqrt{\mathbb{E}\left[\frac{f(X)^2f(V)^2}{f(X,V)^2}\left(\frac{\partial R}{\partial\tau}(X;\tau_0)\right)^2|Q_k(X,V)|^2\right]}\le C\sqrt{\left(\sup_{w\in\mathcal{W}}|Q_k(w)|\right)^2\mathbb{E}\left[N_1(X)^2\right]}=O(k)$$

and

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

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

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

Hence by the Markov inequality the first factor is $O_p(k)$. Combining this with lemma 7 the term (35) is $O_p(k^2 n^{-1/2}) + O_p(k^{5/2-a_0})$.

By assumptions 13, 14 and 15 (37) is bounded by

$$C\sqrt{n}\sup_{w\in\mathcal{W}}|\mu(w)-Q_k(w)'\gamma_k|\frac{1}{n}\sum_{i=1}^{n}N_1(X_i) = O_p(n^{1/2}k^{-a_0}) \tag{39}$$

To bound (38) we observe that the OLS residual $Y_i - Q_k(X_i,V_i)'\tilde{\gamma}_k$ is uncorrelated with $Q_k(X_i,V_i)$. Therefore for the vector δ_k in (33) the remainder (38) is equal to

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

By (33) we have that (38) is bounded by

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

Because assumption 16

$$\mathbb{E}[|Y-\mu(W)|] \leq \sqrt{\mathbb{E}_W\left[\mathbb{E}[Y^2|W]\right]} \leq \overline{\sigma} < \infty$$

we have by the Markov inequality that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|Y_i-\mu(W_i)| = O_p(n^{1/2})$$

By lemma 7

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\tilde{\mu}(W_i)-\mu(X_i,V_i)| = O_p(k^2) + O_p(n^{1/2}k^{5/2-a_0})$$

Therefore (38) is $O_p(n^{1/2}k^{-r/2}) + O_p(k^{(4-r)/2}) + O_p(n^{1/2}k^{(5-r)/2-a_0})$.

We conclude

Lemma 9 *If assumptions 13, 14, 15 and 16 hold, then*

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

$$O_p(n^{-1/2}k^2) + O_p(k^{5/2-a_0}) + O_p(n^{1/2}k^{-a_0}) + O_p(n^{1/2}k^{-r/2}) + O_p(k^{(4-r)/2}) + O_p(n^{1/2}k^{(5-r)/2-a_0})$$

C.5 Expressing (28) as a sample average

Upon substitution of $R(X_i;\tau_0)$ for $L(X_i)$ we obtain for (28)

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

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 (28) 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 V-statistic with projection

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

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

so that

Lemma 10 *Under assumption 13*

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

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

C.6 The asymptotic distribution

The results in the lemmas give

Theorem 6 (Asymptotically linear representation) *If assumptions 12, 13, 14, 15 and 16 hold, then*

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

$$\left\{\left(\mathbb{E}\left[\frac{f(X)f(V)}{f(X,V)}\frac{\partial R}{\partial \tau}(X;\tau_0)\frac{\partial \mu}{\partial v}(W)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]-\mathbb{E}_X\left[\frac{\partial R}{\partial \tau}(X;\tau_0)\mathbb{E}_{V,Z}\left[\frac{\partial \mu}{\partial v}(X,V)\frac{\partial h}{\partial \alpha}(Z;\alpha_0)\right]\right]\right)\sqrt{n}(\hat{\alpha}-\alpha_0)+\right.$$

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

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

$$O_p(k^{9/2}n^{-1/2})+O_p(k^{7/2-a_0})+O_p(n^{1/2}k^{-a_0})+O_p(n^{1/2}k^{-r/2})+O_p(k^{(4-r)/2})+O_p(n^{1/2}k^{(5-r)/2-a_0})$$

Now take $k=Cn^{\kappa}$. The remainders are negligible if the following inequalities are satisfied simultaneously

$$\kappa<\frac{1}{9}\qquad a_0>\frac{7}{2}\qquad \kappa>\frac{1}{2a_0}\qquad \kappa>\frac{1}{r}\qquad r\geq 5\qquad \kappa>\frac{1}{r-5+2a_0}$$

Therefore

$$a_0>\frac{9}{2}\qquad\qquad r\geq 10$$

and

$$\min\left\{\frac{1}{2a_0},\frac{1}{r}\right\}<\kappa<\frac{1}{9}$$