Klein, Roger; Shen, Chan

Working Paper

Semiparametric instrumental variable estimation in an endogenous treatment model

Working Paper, No. 2015-11

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Klein, Roger; Shen, Chan (2015) : Semiparametric instrumental variable estimation in an endogenous treatment model, Working Paper, No. 2015-11, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/130736

Semiparametric Instrumental Variable Estimation in an Endogenous Treatment Model[1]

Roger Klein[2] and Chan Shen[3]

Abstract

We propose instrumental variables (IV) estimators for quantile marginal effects and the parameters upon which they depend in the context of a semiparametric outcome model with endogenous discrete treatment variables. We prove identification, consistency and $\sqrt{N}$-asymptotic normality of the estimators. We also show that they are efficient under correct model specification. Further, we show that they are robust to misspecification of the treatment model in that consistency and asymptotic normality continue to hold in this case. In the Monte Carlo study, the estimators perform well over a wide variety of designs covering both correct and incorrect treatment model specifications.

JEL classification: C14, C26.

Keywords: semiparametric estimation, IV, marginal effects, efficiency, robustness.

[1]Correspondence to: Chan Shen, 1400 Pressler St, Houston, TX 77030. Tel: 713-563-4309. Fax: 713-563-0059. Email: cshen@mdanderson.org.

[2]Department of Economics, Rutgers University, New Brunswick, NJ 08901, United States

[3]Departments of Biostatistics and Health Services Research, University of Texas MD Anderson Cancer Center, Houston, TX 77030, United States

1 Introduction

In this paper, we consider a class of semiparametric models in which the outcome depends on a vector of endogenous treatments with unknown treatment effect functions of exogenous variables. For these models, our main objective is to develop robust, efficient, and $\sqrt{N}-$asymptotically normally distributed estimators for quantile marginal effects and the parameters upon which they depend. We refer to these effects as quantile marginal effects because they estimate the expected treatment impact conditioned on the exogenous variables being in a quantile region of interest. We term the estimators as robust as they retain consistency and asymptotic normality when the treatment model is misspecified. When the treatment model is correctly specified, we provide efficient estimators for quantile marginal effects and the parameter upon which they depend.

For the models being considered, the treatments are discrete endogenous variables that flexibly interact with other explanatory variables. Specifically, with Y as an outcome, X a vector of exogenous variables, and $\mathfrak{T}$ a discrete treatment vector, we focus on the class for which the treatment effects are given by $M(\mathcal{W})$, where M is a vector of unknown functions and $\mathcal{W}$ is a vector of indices that depend on X and a parameter vector θ_0. To estimate marginal effect functions, we develop a localized Instrumental Variable (IV) strategy that iteratively employs two sets of instruments for estimating them and the underlying finite dimensional parameter vector. We provide closed form expressions for the lower bound of the asymptotic covariance matrices for the estimators and the corresponding instrument. To obtain these results, trimming out observations is required to deal with small density denominators, and here we retain observations whose exogenous variables lie in a set based on sample quantiles. For the parameter estimates, the quantile trimming set does not need to be consistently estimated to obtain consistent parameter estimates. Accordingly, it is not surprising that such trimming can be taken as known in developing the theory for the index parameter estimator. In contrast, quantile marginal effects are obtained by averaging marginal effects over estimated quantile sets, and consistency of the former does depend on consistency of the latter. Therefore, as one would expect from Newey and McFadden (1994), the covariance matrix for quantile marginal effects depends on that for estimated quantiles in a manner that is formally established in Lemma 3.

Endogeneity is a very important issue in econometric analysis, and there is a large parametric literature on this issue (see e.g. Heckman (1978) and Hausman (1983)). To deal with this problem, the IV estimator is widely employed in the empirical literature for parametric models (e.g. Card (2001)). In part, IV is very appealing because it is robust against misspecification in modeling the instruments. In nonlinear parametric models, Amemiya (1974,1977) developed an optimal instrument which depends on an unknown conditional expectation of the derivative of a residual function. Employing nonparametric expectations while retaining the parametric outcome model structure, Newey (1990) provided a way to implement the optimal instrument and showed that the estimator is consistent, asymptotically normal and efficient. The emphasis in this literature is on estimating a finite dimensional parameter vector. With known parametric response functions and with marginal effects being evaluated at a given value for the explanatory variables (often their average or median), the marginal effects are easy to recover as they are known functions of the parameters.

The literature has continued to progress to fully nonparametric models based on either generalizations of two-stage-least-squares or control approaches. For example, Newey and Powell (2003) developed a two-stage series estimator with a detailed discussion on identification. Imbens and Newey (2009) provided identification results and convergence rates for average structural functions. Das (2005) and Cai et. al. (2005) developed a substitution estimator for a model that is linear in endogenous variables with an associated multiplicative nonparametric impact or marginal effect response functions. Newey, Powell, and Vella (1999) and Imbens and Newey (2009) developed control estimators for a nonparametric triangular system where both the treatment and the outcome variables of interest are continuous. Our paper differs from these in two respects. First, for nonparametric models, the issue of imposing incorrect parametric restrictions does not arise. In contrast, robust estimation is an issue in a semiparametric context and is part of the focus of the present paper. Second, while this literature estimates marginal effects or average structural functions, with the exception of Das (2005) and Cai et. al. (2005), it has not developed distributional results for them.

Turning to semiparametric models with endogeneity, we note that while such models are very important in empirical studies, there is a scarce literature on distributional results for marginal effects in such models. Further, many of the estimators are not robust to treatment misspecification. For example, the estimator in Das (2005) and Cai et. al. (2005) for nonparametric models, can be extended to the semiparametric case. However, because of the structure of the estimator, its semiparametric extension is not robust to misspecification of the treatment model. Semiparametric extremum approaches such as Rothe (2009) could be developed for the model considered here, but would not be robust to treatment misspecification. There is also a substantial literature on control function approaches in semiparametric models (e.g. Blundell and Powell (2004)), but such approaches are not applicable with discrete treatments. GMM estimators (e.g. Ai and Chen (2003)) could be applied to the present model and would have robustness properties. However, while this literature has recently focused on developing estimators for parameters under a large and growing number of moment conditions, here we derive closed form expressions for optimal marginal effect and parameter estimators and the corresponding covariance matrices. The present paper also differs from this literature in developing $\sqrt{N}-$ asymptotics for quantile marginal effects. We derive an explicit form for the asymptotic covariance matrix that accounts for the presence of indicators on estimated quantile sets. Theorem 5 shows that this matrix is minimal under correct treatment specification.

To obtain these results, in Section 2, we describe the model and the estimators for the quantile marginal effect functions and the parameters upon which they depend. As discussed in this section, the estimators may be viewed as iterative versions of global and local IV approaches, where the local formulation is motivated by the approach in Fan and Gibels (1996). Section 3 provides assumptions and definitions needed to obtain large sample results. Section 4 provides the main theorems for both a finite dimensional parameter vector and for quantile marginal effects. In Section 5, we provide a Monte Carlo study with designs in which we vary the form of the marginal effect function. For each of these functions, we consider cases where the treatment model is misspecified, taking the form of omitting relevant variables, and where it is correctly specified. We find that the estimator developed here has good finite sample properties under correct and incorrect treatment specification. Concluding remarks are provided in Section 6. The Appendix contains proofs of main theorems along with required intermediate lemmas, and their proofs.

2 The Model and the Estimator

The model we consider is one with discrete endogenous treatments and a continuous outcome. There are many empirical applications with such a model structure. For example, in health economics, the health outcome (e.g. survival) can be continuous and the treatment could be no treatment, chemotherapy only, radiation therapy only, or both chemotherapy and radiation therapy. The treatment could also be smoking choices: no smoking, smoking 1-5 cigarettes per week, smoking 6-10 cigarettes, etc. There are also many examples in labor economics such as job training program choice with wages being the outcome.

Denote Y_i as the continuous outcome and $\mathcal{W}_i$ is an i.i.d. vector of induces. With $L+1$ possible treatment options, the endogenous treatment variable $\mathfrak{T}_i$ takes on values $0, .., L$ with 0 as the null option. Then, the model that we consider has the following structure:

$$Y_i = g(\mathcal{W}_i, \mathfrak{T}_i) + \varepsilon_i,$$

where ε_i is the error component with conditional expectation zero given the exogenous variables. Write the treatment indicator as $T_{il} = 1\{\mathfrak{T}_i = l\}$, $l = 1, .., L$. Then, because of the discreteness of the treatment, we can rewrite the model as:

$$\begin{aligned} Y_i &= \sum_{l=1}^{L} g(\mathcal{W}_i, l) T_{il} + g(\mathcal{W}_i, 0) \left[1 - \sum_{l=1}^{L} T_{il}\right] + \varepsilon_i \quad (1) \\ &= \sum_{l=1}^{L} [g(\mathcal{W}_i, l) - g(\mathcal{W}_i, 0)] T_{il} + g(\mathcal{W}_i, 0) + \varepsilon_i \\ &\equiv R_i \alpha_0 (\mathcal{W}_i) + \varepsilon_i \quad (2) \end{aligned}$$

where

$$R_i \equiv [T_i, 1], \alpha_0(\mathcal{W}_i) \equiv \begin{pmatrix} [g(\mathcal{W}_i, 1) - g(\mathcal{W}_i, 0)] \\ ... \\ [g(\mathcal{W}_i, L) - g(\mathcal{W}_i, 0)] \\ g(\mathcal{W}_i, 0) \end{pmatrix}. \quad (3)$$

The data set $\{Y_i, \ T_{i1}, .., T_{iL}, \mathcal{W}_i\}_{i=1}^{N}$ is identically and independently distributed (i.i.d.) taking on values in $\mathfrak{X}_Y \times \mathfrak{X}_T \times \mathfrak{X}_W \subset \mathfrak{R}^{1+L+K_W}$ with K_W denoting the dimension of $\mathcal{W}_i$. This model structure is the same as in Das (2005) and Cai et. al. (2005) in the nonparametric case. For this case, they proposed a substitution estimator where the endogenous treatment variables are replaced by estimates of their conditional expectations. This estimator is very attractive for this nonparametric case where by definition there are no incorrect parametric restrictions to consider. However, there are applications where sample sizes are not sufficiently large to obtain reliable estimates of nonparametric models. Semiparametric models have become important alternatives in such cases. If the approaches discussed above are applied in a semiparametric context, the estimator will typically be inconsistent when the semiparametric treatment model is not correctly specified. In contrast, here we develop a semiparametric IV estimator that retains consistency and asymptotic normality when the treatment model is misspecified. Further, we show that the IV estimator is efficient under correct semiparametric specifications. Regardless of whether or not the treatment model is correct, we show that the covariance matrix for our estimator does not depend on the covariance matrix for the estimator of the treatment model.

In the outcome model discussed above, we allow for multiple treatments and therefore we allow for multiple indices in the treatment model. For example, consider a model with treatments T_1 (e.g. chemotherapy) and T_2 (radiation). Suppose that each of these treatments is described by a single index model:

$$T_1 = 1\{I_1 > u_1); T_2 = 1\{I_2 > u_2\},$$

with thresholds u_1 and u_2 independent of the exogenous variables upon which the indices I_1 and I_2 depend. For the model we consider, we allow for the joint treatment $T_3 = \{I_1 > u_1 \text{ and } I_2 > u_2\}$ to be another treatment, which results in a multiple index model. To handle these types of cases, we allow for multiple treatment indices. For the outcome model, for expositional simplicity, we adopt a single index specification, but note that the extension to multiple outcome indices is straightforward.

With the parametric outcome index $V_i(\theta_0)$ defined in (D1) in Section 3 and with v as a fixed value for the index, refer to (2) and let $\Delta_i(v) \equiv R_i\alpha_0(V_i(\theta_0)) - R_i\alpha_0(v)$. Then, we can write the localized model as

$$Y_i = R_i\alpha_0(v) + [\Delta_i(v) + \varepsilon_i].$$

This localization is important, because it enables us to treat the $\alpha - functions$ evaluated at v as estimable parameters.[1] Kernel weighting is employed to enforce this localization by severely downweighting any observations where V_i is not close to v.

If the above functions did not depend on unknown parameter values, we could then proceed to develop a local IV estimator for the above model to recover the unknown functions at all points of interest. With θ_0 having dimension K_1 and taking values in $\Phi_1 \subset \Re^{K_1}$, assume that the outcome index is a known function of $\theta_0 : V_i(\theta_0)$. If we knew θ_0 and hence the true index values and if we only used observations $V_i(\theta_0)$ in a neighborhood of v, then $\Delta_i(v)$ would be close to zero. It would then be natural to employ treatment probabilities as instruments and write localized IV moment conditions as:

$$\hat{Q}' D_N(v, \theta_0)[Y - R\alpha_0(v)] = 0$$

where with $\mathbf{1}$ denoting a vector of ones, $\hat{Q} \equiv \hat{t}\{\hat{P} \quad \mathbf{1}\}$ includes trimming $\hat{t}$ and an $N \times L$ matrix of probabilities $\hat{P}$ as the treatment instruments. The diagonal matrix $D_N(v, \theta_0)$ is given by:

$$D_N(v, \theta_0) \equiv diag\left\{\frac{1}{h}k\left(\frac{v - V_i(\theta_0)}{h}\right)\right\}$$

where the kernel function $k(z)$ is a symmetric density and represents the localization that heavily weights observations close to v. For known θ_0, the above moment conditions generate the estimator $\hat{\alpha}$:

$$\hat{\alpha}(v, \theta_0) = \left[\hat{Q}' D_N(v, \theta_0) R\right]^{-1} \hat{Q}' D_N(v, \theta_0) Y,$$

provided that the inverse on the right-hand-side exists (as established in Theorem 1). In practice, θ_0 is unknown and must be estimated in conjunction with the unknown functions in α. To this end, with functions evaluated at an arbitrary vector of parameter values θ, define:

$$\hat{\alpha}(v, \theta) = \left[\hat{Q}' D_N(v, \theta) R\right]^{-1} \hat{Q}' D_N(v, \theta) Y.$$

[1] Fan and Gibels (2008) analyzed a local polynomial approach in the case where endogeneity is not an issue.

We now estimate the index using a second and different IV strategy. To motivate this strategy, consider the Non-linear Least-Squares (NLS) estimator:

$$\hat{\theta} = \arg\min_{\theta} \sum \left[Y_i - R_i \hat{\alpha}(V_i(\theta), \theta)\right]^2$$

The first-order conditions for the NLS estimator are given as:

$$\frac{1}{N} \sum \left[R_i \nabla_\theta \alpha\right]' \left[Y_i - R_i \hat{\alpha}(V_i(\hat{\theta}), \hat{\theta})\right] = 0.$$

Since this estimator will be inconsistent in general, we require an instrument for $R_i \nabla_\theta \alpha$. The conditions the instrument $\hat{Z}_i$ must satisfy are provided in (D9) and (D10) in Section 3. We also discuss its implementation in the Monte Carlo study, with the instrument being the optimal one as obtained in Theorem 4. Allowing for the possibility that this instrument may (and typically will) be estimated, denote $\hat{Z}$ as the instrument, which we assume converges in probability to Z. Then, the IV estimator for θ_0 is:

$$\hat{\theta} = \arg\min_{\theta} \left[\frac{1}{N} \sum_{i=1}^{N} \hat{Z}_i' \left[Y_i - R_i \hat{\alpha}(V_i(\theta), \theta)\right]\right]^2.$$

In proving properties for the estimator, we show that $\hat{\alpha}(V_i(\theta), \theta)$ converges uniformly in θ to a fixed function $\alpha(V_i(\theta), \theta)$. Further, we show that when evaluated at θ_0, this function is equal to the true marginal effect function.

The constructions of $\hat{\alpha}$ and $\hat{\theta}$ above provide the insight as to why the estimator is consistent even when the treatment model is misspecified. Informally and abstracting away from trimming issues, we use the term misspecified to mean that the treatment instruments are not the expected value of the treatment vector conditioned on the model's exogenous variables. For example, in the Monte Carlo study, we consider a misspecified case where the treatment model is subject to omitted variable bias. To intuitively see why the estimator is robust to such misspecification, it is useful to consider the moment conditions that define the estimators. For the $\theta - moment$ conditions, we employ the uniform convergence property of $\hat{\alpha}(V_i(\theta), \theta)$ to show that:

$$\frac{1}{N} \sum_{i=1}^{N} \hat{Z}_i' \left[Y_i - R_i \hat{\alpha}(V_i(\theta), \theta)\right]$$

is uniformly close to:

$$E\left[\frac{1}{N} \sum_{i=1}^{N} Z_i' \left[Y_i - R_i \alpha(V_j(\theta), \theta)\right]\right].$$

Similarly, for the $\alpha-$moment conditions, it can be shown that they are uniformly close to:

$$E\left[\frac{1}{N} \sum_{i=1}^{N} Q_i' \left[Y_i - R_i \alpha(V_j(\theta), \theta)\right] \frac{1}{h} k\left(\frac{V_i(\theta) - V_j(\theta)}{h}\right)\right], j = 1, .., N.$$

Subject to identification conditions, the estimator will be robust if θ_0 solves the moment conditions when the treatment model is not correct. This is the case, because $\left[Y_i - R_i \alpha(V_i(\theta_0), \theta_0)\right] = \varepsilon_i$

is uncorrelated with Q_i regardless of whether $Q_i = E(R_i|X_{iF})$ where X_{iF} denotes the full set of exogenous variables in the model. Similarly, with Z_i as an appropriately selected vector of exogenous variables, the $\theta-$moment conditions will also be solved at θ_0 when the treatment model is misspecified. In contrast, if we use a substitution estimator with Q_i replacing R_i, then $(R_i - Q_i)\alpha_{0i}$ becomes an error component. When the treatment model is misspecified, it is highly likely that this error component will be related to other included variables in a semiparametric context. We explore this issue further in Monte Carlo experiments below.

It is interesting to note that for $\theta \neq \theta_0$, the $\alpha(V_j(\theta), \theta)$ functions will depend on the instrument Q that is employed. However, at θ_0, $\alpha(V_j(\theta_0), \theta_0) = \alpha_0(V_j(\theta_0))$ does not depend on Q. In other words, at θ_0, we are recovering a treatment effect that depends on X, but does not depend on the instrument.

Having described the estimators for index parameters and $\alpha - functions$, we turn to the estimator for marginal effects. While marginal effects at specific points are implied by the $\alpha - functions$, these functions converge at a rate below the parametric rate of $\sqrt{N}$. Further, as the model is nonlinear, the marginal effects are not constant throughout the index distribution. In this paper, we propose a marginal effect estimator over quantiles that converges at the $\sqrt{N}$ rate as researchers are often interested in marginal effects over different quantiles of the distribution of the exogenous variables.

Let $V_j(\hat{\theta})$ be the estimated index for observation j and $\hat{t}_{qj}$ as an indicator for a particular variable of interest X_{ik} being in an estimated quantile set. We then define a quantile marginal effect as:

$$\hat{M}_{qj} \equiv \sum_{j=1}^{N} \hat{t}_{qj} \hat{\alpha}\left(V_j\left(\hat{\theta}\right), \hat{\theta}\right) / \sum_{j=1}^{N} \hat{t}_{qj}. \tag{4}$$

We will show that this estimator is consistent and asymptotically normally distributed whether or not the treatment model is correctly specified. When this model is correctly specified and the index parameter is based on the efficient instrument given below, we show that the quantile marginal effects estimator will also be efficient. In a parametric problem, this result would directly follow when estimating the marginal effect at a point. In the present context, the marginal effect is random even if evaluated at θ_0. Here, we establish efficiency under correct specification by showing that the distribution of the marginal effect depends on several uncorrelated components, one of which accounts for parameter uncertainty. The efficient estimator for the marginal effect will then be based on an efficient estimator for the parameters.

3 Definitions and Assumptions

In this section, we provide the definitions for our proposed estimators, the corresponding assumptions and also a brief discussion about them.

D1. Outcome Model. Referring to the general model in (1), with θ_0 having dimension K_1 and taking values in $\Phi_1 \subset \Re^{K_1}$, let $\mathcal{W}_i \equiv V_i(\theta_0) = X_{1i} + X_{IIi}\theta_0$ be the outcome index which depends on$X_i \equiv [X_{1i}, \ X_{IIi}]$. a vector of exogenous variables[2] in the outcome equation. With

[2]For expositional purposes, we consider the single index case, but extensions to the multiple index case are straightforward under bias corrections or higher order kernels.

$R_i \equiv [T_i \ \ \mathbf{1}]$ and $\alpha_0(V_i(\theta_0))$ the vectors in (2), the outcome model is defined as:

$$Y_i = R_i\alpha_0\left(V_i(\theta_0)\right) + \varepsilon_i.$$

Letting $\Delta_i(v) \equiv R_i\alpha_0\left(V_i(\theta_0)\right) - R_i\alpha_0(v)$, the localized model at v can be written as:

$$Y_i = R_i\alpha_0(v) + \left[\Delta_i(v) + \varepsilon_i\right]. \tag{5}$$

D2. Treatment Probability and Instrument. Denote $P_i^c \equiv E(T_i|X_{iF})$ as the correct treatment probability, where X_{iF} contains all exogenous variables as specified in (A5). Let $V_T(\gamma)$ be a vector of semiparametric indices for estimating the treatment probability, where γ is a vector of K_2 parameters taking on values in $\Phi_2 \subset \Re^{K_2}$. Denote $\hat{\gamma}$ as the parameter estimator with limiting value γ_*. Then, define the treatment instrument as $P_i \equiv E(T_i|V_{Ti}\left(\gamma_*\right))$, which may not be the same as P_i^c due to misspecification in the treatment model.[3]

D3. Trimming. Let λ be a quantile fraction and W a vector of random variables with λ^{th} population quantile vector $q\left(\lambda\right)$.[4] Let W_i be an i.i.d. sample from the density for W, and $\hat{q}\left(\lambda\right)$ be the λ^{th} sample quantile vector. Then, with all inequalities holding element by element, define population and sample trimming functions that constrain the random variable of interest to a quantile set:

$$\begin{aligned} t_i\left(q\right) \equiv t\left(q(\lambda_1), q(\lambda_2), W_i\right) &\equiv 1\{q\left(\lambda_1\right) < W_i < q\left(\lambda_2\right)\} \\ t_i(\hat{q}) \equiv t\left(\hat{q}(\lambda_1), \hat{q}(\lambda_2), W_i\right) &\equiv 1\{\hat{q}\left(\lambda_1\right) < W_i < \hat{q}\left(\lambda_2\right)\}. \end{aligned}$$

D4. Kernel. Referring to (D1), define $D_N(v,\theta) \equiv diag\left(\frac{1}{h}k\left(\frac{v-V_i(\theta)}{h}\right)\right)$, where $k(\cdot)$ is a symmetric density, $\int z^2k(z)dz$ is bounded, and $h = O(N^{-r})$ with $r = \frac{1}{7.99}$.

D5. **Treatment Probabilities.** Denote the i^{th} observation on the L treatment probabilities and the corresponding estimator as:

$$P_i \equiv \left\{ \begin{array}{cccc} P_{i1} & P_{i2} & ... & P_{iL} \end{array} \right\}_{1\times L}; \ \ \hat{P}_i \equiv \left\{ \begin{array}{cccc} \hat{P}_{i1} & \hat{P}_{i2} & ... & \hat{P}_{iL} \end{array} \right\}_{1\times L}.$$

With t_i as any indicator restricting V_T to a fixed compact set :

$$\sup_i t_i \left|\hat{P}_i - P_i\right| = o_p\left(N^{-1/4}\right) \ .$$

D6. Instruments for estimating $\alpha - functions$. Refer to (D2)-(D3) and let X_{CV} be the continuous variables in the outcome index V. Define percentiles: $0 < \lambda_1^* < \lambda_1 < \lambda_2 < \lambda_2^*$, with $\lambda_1 - \lambda_1^*$ and $\lambda_2^* - \lambda_2$ each $o_p(h^c), 0 < c < 1$. Then, define the following trimming functions:

$$\begin{aligned} \hat{t}_{1i}^* &\equiv t\left(\hat{q}(\lambda_1^*), \hat{q}(\lambda_2^*), X_{CV}\right) \text{ and } t_{1i} \equiv t\left(q(\lambda_1), q(\lambda_2), X_{CV}\right), \\ \hat{t}_{2i} &\equiv t\left(q(\lambda_1), q(\lambda_2), V_{iT}\left(\hat{\gamma}\right)\right) \text{ and } t_{2i} \equiv t(q(\lambda_1), q(\lambda_2), V_{iT}\left(\gamma_*\right)), \end{aligned}$$

[3]The treatment model is estimated under index restrictions that may or may not hold. If the treatment model is correctly specified, then $\gamma_* = \gamma_0$, the true parameter vector. However, we allow for misspecification in which case $\gamma_* \neq \gamma_0$ and $E(T_i|V_{Ti}\left(\gamma_*\right)) \neq E(T_i|X_{iF})$.

[4]For example, if $\lambda = .5$ then $q\left(\lambda\right)$ will contain a vector of medians for elements of W.

In establishing convergence rates, the following trimming function will be useful:

$$t_{1i}^{*} \equiv t\left(q\left(\lambda_1^{*}\right), q\left(\lambda_2^{*}\right), X_{CV}\right).$$

The instruments for estimating the $\alpha - functions$ can now be defined as:

$$\begin{array}{cc} \hat{Q}_i \equiv \hat{t}_{2i}\left[\hat{P}_i \;\; \mathbf{1}\right] & Q_i \equiv t_{2i}\left[P_i \;\; \mathbf{1}\right] \\ \hat{Q}_i^{*} \equiv \hat{t}_{1i}^{*}\hat{t}_{2i}\left[\hat{P}_i \;\; \mathbf{1}\right] & \bar{Q}_i^{*} \equiv t_{1i}t_{2i}\left[P_i \;\; \mathbf{1}\right], Q_i^{*} \equiv t_{1i}^{*}t_{2i}\left[P_i \;\; \mathbf{1}\right]. \end{array}$$

D7. Local Projection. Define the local linear projection of T_i on an arbitrary variable F_i as

$$T_i = F_i \times (\pi(v)', c(v)') + e_i,$$

where:

$$(\pi(v)', c(v))' \equiv E\left[F_i'F_i|V_i = v\right]^{-1} E[F_i'T_i|V_i = v].$$

D8. $\boldsymbol{\alpha - functions}$. Referring to (D1), (D4)-(D6), define $\hat{\alpha}(v,\theta)$ and its population counterpart $\alpha(v,\theta)$ as the solutions for the following:

$$\begin{aligned} \frac{\hat{Q}'D_N(v,\theta)R}{N}\hat{\alpha}(v,\theta) &\equiv \sum_i \frac{\hat{Q}_i'D_N(v,\theta)Y_i}{N}, \\ \Omega_j\alpha(v,\theta) &\equiv E\left(Q_j'Y_j|v\right)g_V(v), \;\; \Omega_j \equiv E\left(Q_j'R_j|v\right)g_V(v). \end{aligned}$$

With $\hat{\Delta}_i(v,\theta) \equiv R_i[\hat{\alpha}(V_i(\theta),\theta) - \hat{\alpha}(v,\theta)]$ as a bias adjustment, define a bias corrected estimator $\hat{\alpha}^{*}(v,\theta)$ and its limiting value $\bar{\alpha}(v,\theta)$ as:

$$\begin{aligned} \frac{\hat{Q}^{*\prime}D_N(v,\theta)R}{N}\hat{\alpha}^{*}(v,\theta) &\equiv \sum_i \frac{\hat{Q}_i^{*\prime}D_N(v,\theta)[Y_i - \hat{\Delta}_i(v,\theta)]}{N-1} \\ \bar{\Omega}_j^{*}\bar{\alpha}(v,\theta) &\equiv E\left(\bar{Q}_j^{*\prime}Y_j|v\right)g_V(v), \;\; \bar{\Omega}_j^{*} \equiv E\left(\bar{Q}_j^{*\prime}R_j|v\right)g_V(v). \end{aligned}$$

where $\hat{Q}_j^{*}$ and $\bar{Q}_j^{*}$ are defined in (D6). When $v = V_j$, all averages above are taken over the $N-1$ terms with $i \neq j$. In establishing convergence results, it will also be useful to define: $\Omega_j^{*} \equiv E(Q_j^{*\prime}R_j|v)g_V(v)$, with Q_j^{*} given in (D6).

D9. Instruments for estimating the outcome index. Recalling that K_1 is the dimension of θ, let $\hat{Z}_i^{o}$ $:1 \times K_1$ be the i^{th} observation on the vector of estimated instruments for estimating the outcome parameters and denote its population counterpart as Z_i^{o} , which is a function of X_F. With t_i as any indicator restricting the treatment index V_{Ti} and the outcome variables X_i to a fixed compact set, $\hat{Z}_{ik}^{o}$, $k = 1, ..., K_1$ satisfies

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} t_i\left[\hat{Z}_{ik}^{o} - Z_{ik}^{o}\right]^2 = o_p(1), \tag{6}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} t_i\left[\hat{Z}_{ik}^{o} - Z_{ik}^{o}\right]\varepsilon_i = o_p(1). \tag{7}$$

D10. Index Parameter Estimator. With $\hat{Z}_i^o$ defined in (D9), let $\hat{Z}_i \equiv \hat{t}_{1i}\hat{t}_{2i}\hat{Z}_i^o$ and $Z_i \equiv t_{1i}t_{2i}Z_i^o$. Then:

$$\begin{aligned}\hat{\theta}^* &\equiv \arg\min_{\theta} \hat{G}^*(\theta)' \hat{G}^*(\theta), \\ \hat{G}^*(\theta) &\equiv \frac{1}{N}\sum_{j=1}^{N} \hat{Z}_j' \left[Y_j - R_j \hat{\alpha}^*(V_j(\theta), \theta)\right].\end{aligned}$$

D11. Marginal Effect Trimming. Letting X_C be the model's K_C distinct continuous variables in treatment and outcome equations, define $[\lambda_{1C}, \lambda_{2C}]$ as a region without small denominator issues (e.g. quantile region $[.01, .99]$ for all the variables in X_C). Further, to estimate quantile marginal effects with respect to the variable $X_k \subset X_C$, define $[\lambda_{1k}, \lambda_{2k}]$ as a quantile region of interest (e.g. the second quartile with $[\lambda_{1k}, \lambda_{2k}] = [.25, .50]$). With $[q'(\lambda_1), q'(\lambda_2)] : 2\times K_C$ as the vector of quantiles defined by the intersection of the quantile sets, the marginal effect trimming is defined as:

$$\begin{aligned}t_{qj}(q(\lambda_1), q(\lambda_2), X_{jC}) &\equiv t(q(\lambda_{1C}), q(\lambda_{2C}), X_{jC}) \times t(q(\lambda_{1k}), q(\lambda_{2k}), X_{jk}), \\ \hat{t}_{qj}(q(\lambda_1), q(\lambda_2), X_{jC}) &\equiv t(\hat{q}(\lambda_{1C}), \hat{q}(\lambda_{2C}), X_{jC}) \times t(\hat{q}(\lambda_{1k}), \hat{q}(\lambda_{2k}), X_{jk}).\end{aligned}$$

D12. Quantile Marginal Effects. The population quantile marginal effect and its estimator are given respectively as:

$$\begin{aligned}M_q &\equiv \frac{E\left[t_{qj}\alpha(V_j(\theta_0), \theta_0)\right]}{E\left[t_{qj}\right]}, \\ \hat{M}_q &\equiv \frac{\sum_{j=1}^{N} \hat{t}_{qj}\hat{\alpha}^*(V_j(\hat{\theta}^*), \hat{\theta}^*)}{\sum_{j=1}^{N} \hat{t}_{qj}}.\end{aligned}$$

D13. Bahadur Representation. Referring to (D3) and (D11), let λ_1 and λ_2 be $K_C\times 1$ vectors of lower and upper percentiles, denote $q(\lambda_1)$ and $q(\lambda_2)$ as the corresponding population quantile vectors for the model's K_C continuous variables, X_C . Let $g_{X_{kC}}(\bullet)$ be the marginal density for the k^{th}continuous variable, X_{kC}. For $s : K_C\times 1$ with k^{th} component s_k, denote $g_C(s)$ as the $K_C\times 1$ vector having k^{th} element $g_{X_{kC}}(s_k)$. With $\varnothing$ denoting division by element, let

$$\mathfrak{B}_j \equiv \left[\begin{array}{c}\left[\left(1\left\{X_{jC} \leq q(\lambda_1)\right\} - \lambda_1\right) \varnothing g_C(q(\lambda_1))\right]' \\ \left[\left(1\left\{X_{jC} \geq q(\lambda_2)\right\} - \lambda_2\right) \varnothing g_C(q(\lambda_2))\right]'\end{array}\right]'_{2K_C\times 1}.$$

The Bahadur representaion (see Bahadur (1966) and David (1981)) can now be defined as:

$$\sqrt{N}\left[\hat{q}(\lambda) - q(\lambda)\right] = \sqrt{N}\bar{\mathfrak{B}} + o_p(1), \bar{\mathfrak{B}} \equiv \frac{1}{N}\sum_{j=1}^{N} \mathfrak{B}_j,$$

where $\hat{q}(\lambda)$ is a vector of $\lambda-$sample quantiles.

There are several observations that we would like to make about the above definitions. First, in (D5) and (D9), we provide required convergence conditions on $\hat{P}_i$ and $\hat{Z}_{ik}^o$ under known trimming. In the Appendix we verify that these conditions hold for the particular instruments that we employ. Further, as estimated trimming is employed in practice, we prove that it can be taken as known. Second, the estimation of marginal effect functions and quantile marginal effects are both based on regular kernels with a bias correction. Referring to the local model in (D1), at the true parameter values, the bias corrected model has the form:

$$\begin{aligned} Y_i - \hat{\Delta}_i(v) &= R_i\alpha(v,\theta_0) + \left[\Delta_i(v) - \hat{\Delta}_i(v)\right] + \varepsilon_i, \\ \Delta_i(v) &\equiv R_i\left[\alpha(V_i(\theta_0),\theta_0) - \alpha(v,\theta_0)\right]. \end{aligned}$$

While the kernel weighting is intended to reduce the bias from the localization error $\Delta_i(v)$, the resulting bias in the estimator is not sufficiently small to establish asymptotic normality. For this purpose, we require that the gradient bias in the limiting value of $\sqrt{N}\hat{G}^*(\theta_0)$ in (D10) vanishes. By removing an estimate of this localization bias from the outcome variable Y, we are able to ensure that this is the case. Third, note that in (D12) we have defined a quantile marginal effect estimator. In applied work, we would like to know how these marginal effects vary at different places in the distribution of the X-variables. For example, with age as one of the explanatory variables, we would want to know how the treatment effect varies over different quantiles of the age distribution. The definition above captures this feature by averaging the marginal effect function over quantiles of each explanatory variable of interest.[5] It should be noted that an additional benefit of such averaging is the rate of convergence is increased to the parametric rate.

Finally, the quantile marginal effect estimator in (D12) depends on an indicator for being in an estimated quantile set. The covariance matrix for this estimator will contain components that reflect the estimation uncertainty in having to estimate these quantiles. The Bahadur representation, in conjunction with Lemma 3, play an important role in characterizing this uncertainty.

The following assumptions are needed to obtain the theoretical results for the estimators defined above.

A1. Referring to the outcome and treatment models in (D1)-(D2), observations $\{Y_i, T_{i1}, .., T_{iL}, X_{iF}\}_{i=1}^N$ are i.i.d. and take on values in $\mathfrak{X}_Y \times \mathfrak{X}_T \times \mathfrak{X}_F \subset \mathfrak{R}^{1+L+K_F}$, and $E(\varepsilon_i|X_{iF}) = 0$. For expositional simplicity, we also assume $E(\varepsilon\varepsilon'|X_F) = \sigma_\varepsilon^2 I$.[6] The vector of exogenous variables in the outcome model X_i is a subset of X_{iF}, which is the full set of exogenous variables in the treatment model.

A2. The parameters of the outcome model, θ_0 and those of the treatment model, γ_* take values in the interior of a compact set $\Phi \subset \mathfrak{R}^{K_1+K_2}$.

A3. Referring to (D1), the outcome index $V(\theta_0) = X_1 + X_{II}\theta_0$ contains at least one absolutely continuous exogenous variable.

[5] As discussed following Theorem 5 on quantile marginal effects, we can define quantiles to be support points when the variable of interest is discrete.

[6] For the current estimator, if $E(\varepsilon\varepsilon'|X_F) = \Sigma \neq \sigma_\varepsilon^2 I$, consistency and normality results hold with a modified covariance matrix that is estimable. If the model is appropriately transformed based on an estimate for Σ, then all results apply for the current estimator of the transformed model.

A4. Recalling (D2), there exists a value for the treatment parameter vector γ_*, such that its estimator satisfies:

$$[\hat{\gamma} - \gamma_*] = o_p\left(N^{-1/4}\right).$$

A5. With X_F as the full set of exogenous variables in treatment and outcome models (D1)-(D2) and with R as the treatment vector in (D1), assume that the treatment model is correct in that the treatment instrument satisfies: $[P \;\; \mathbf{1}] = E(R|X_F)$.

A6. Referring to (A2), let $V(\theta) \equiv X_1 + X_{II}\theta$, $v(\theta) \equiv x_1 + x_{II}\theta$, and let $g_{V|Y}(v(\theta)|y)$ be the density for $V(\theta)$ conditioned on $Y = y$. Denote $\nabla_k^p g_{V|Y}$ as the derivative of g with respect to the k^{th} component of θ up to order p, with $\nabla_k^0 g_{V|Y} = g_{V|Y}$. Assume $g_{V|Y} > 0$ on all fixed compact subsets of the support for the index $V(\theta)$ and that $\left|\nabla_k^p g_{V|Y}\right|$ is uniformly bounded for $p = 0, 1, 2, 3, 4$.

A7. With X_{ik} as any one of the exogenous variables in the outcome equation, assume thatthere exists $m > 4$ such that $E\left[(X_{ik}Y_i)^m\right] = O(1)$. For the treatment index V_{Ti}, the outcome variables X_i, and θ in any bounded set, Z_i^o, $\nabla_\theta \alpha(V_i(\theta), \theta)$, $\nabla_\theta^2 \alpha(V_i(\theta), \theta)$, and $\nabla_\theta E\left[Y_i|V_i(\theta)\right]$ are bounded.

A8. Referring to (D6) and (D7), with F_i as either Q_i or $\bar{Q}_i^*$ let $\pi(v)$ be the parameters of the local projection of T_i on F_i. Then, assume that $\pi(v)$ has full rank, $\Sigma \equiv var(F_i'|V_i(\theta) = v)$ is positive definite.

A9. With Z and $\bar{\alpha}$ defined in (D8)-(D10) there is a unique solution at θ_0 to the population moment condition: $E\left\{Z_j'\left[Y_j - R_j\bar{\alpha}(V_j(\theta), \theta)\right]\right\} = 0$. Further, the matrix $H_0 \equiv -E(Z_j' R_j \nabla_\theta \bar{\alpha}(V_j(\theta), \theta)_{\theta=\theta_0})$ is non-singular.

Assumptions (A1)-(A3) and (A6)-(A7) are somewhat standard for estimating semiparametric models. Assumption (A4) provides the conditions for the treatment parameter estimator, whose limiting value need not be the true parameter value, γ_o; while Assumption (A5) defines the correctly specified case for the treatment model estimation. In estimating marginal effect functions and index parameters of the outcome equation, (A5) is needed for efficiency, but not for consistency or asymptotic normality. The conditions in (A8) are needed to identify the $\alpha - function$ at the point v. The assumptions in (A9) are standard for estimating parameters based on moment conditions.

4 Large Sample Results:

We begin with Theorem 1 which provides identification results for the marginal effect functions and the index parameter vector.

Theorem 1 ***Identification.*** *Under (D1) -(D10) and assumptions (A1)-(A3) and (A8)-(A9),* $\bar{\alpha}(v, \theta)$, $\alpha(v, \theta)$, *and* θ_0 *are identified.*

For estimating the marginal effect functions, there are two special cases of interest in which we can simplify the identification conditions. Referring to (D7) , when the treatment model is correct,

$$R_i = Q_i + u_i, E\left(u_i | X_F\right) = 0 \tag{8}$$

and π effectively becomes the identity matrix, satisfying the rank condition. Second, when the treatment model is incorrect and there is a single treatment, π is the local coefficient on the instrument. Similar to the parametric linear case with one endogenous variable, we require this coefficient to be non-zero so that the instrument locally explains R in the presence of the other exogenous variables. Having established identification, Theorems 2 and 3 provide consistency and normality results for the estimator of the outcome equation parameters.

Theorem 2 ***Consistency.*** *Under (D1)-(D10) and assumptions (A1)-(A4) and (A6)-(A9):*

$$\hat{\theta}^* \xrightarrow{p} \theta_0.$$

Theorem 3 ***Normality.*** *Under (D1)-(D10) and assumptions (A1)-(A4) and (A6)-(A9):*

$$\sqrt{N}\left[\hat{\theta}^* - \theta_0\right] \xrightarrow{d} \; W \sim N\left(0, \Upsilon\right), \text{ with } \Upsilon \equiv \sigma_\varepsilon^2 H_0^{-1} E\left[\bar{\mathfrak{R}}_z(V_j\left(\theta_0\right))' \bar{\mathfrak{R}}_z(V_j\left(\theta_0\right))\right] H_0^{-1},$$

where

$$\begin{aligned} \bar{\mathfrak{R}}_z'(V_j\left(\theta_0\right)) &\equiv Z_j' - \left[E(Z_j R_j | V_j(\theta_0))\right]\left[E[\bar{Q}_j^* R_j | V_j(\theta_0)]\right]^{-1} \bar{Q}_j^{*\prime} \\ H_0 &\equiv -E(Z_j' R_j \nabla_\theta \bar{\alpha}(V_j\left(\theta\right), \theta)_{\theta=\theta_0}) \end{aligned}$$

The results in Theorems 2-3 are important for conducting inferences in applications. In obtaining them, we have taken into account several important features for implementing these estimators in practice. First, for technical reasons, it is important to trim observations and in practice the set over which we trim will not be known. Here, we base trimming on estimated sample quantiles and employ results from Pakes and Pollard (1989) to address this issue. Second, we have found that the estimator performs much better under regular than higher order kernels. For the consistency result, regular kernels would suffice without any bias corrections, and the range of permissible windows would be much wider than required in Theorem 3. For the normality result, we employ a bias correction mechanism for which regular kernels are theoretically valid. We note that in stating the normality result above, we have written the gradient component of the covariance matrix in terms of the residual $\bar{\mathfrak{R}}_z'(V_j\left(\theta_0\right))$. This form facilitates the derivation of the optimality result in Theorem 4.

Theorem 4 ***Optimality.*** *Under (D1)-(D10) and assumptions (A1)-(A9), the lower bound on the covariance matrix is given as:*

$$\sigma_\varepsilon^2 \left[E\left(\bar{\mathfrak{R}}_\varphi'(V_j\left(\theta_0\right))\bar{\mathfrak{R}}_\varphi(V_j\left(\theta_0\right))\right)\right]^{-1},$$

where with $\varphi_j \equiv \bar{Q}_j^* \nabla_\theta \alpha(V_j(\theta), \theta_0)_{\theta=\theta_0}$:

$$\begin{aligned} \bar{\mathfrak{R}}_\varphi\left(V_j\left(\theta\right)\right) &= \varphi_j - \bar{Q}_j^{*'} E\left[\bar{Q}_j^{*'} \bar{Q}_j^{*\prime} | V_j\left(\theta_0\right)\right]^{-1} E\left[\bar{Q}_j^{*'} \varphi_j | V_j(\theta_0)\right] \quad (9) \\ &= \bar{Q}_j^* \nabla_\theta \alpha(V_j(\theta), \theta)_{\theta=\theta_0} \quad (10) \end{aligned}$$

The efficiency bound is attained by an instrument with j^{th} *observation:* $\bar{\mathfrak{R}}_\varphi(V_j)$.

To interpret this result, we note that when the $\alpha-functions$ are known parametric functions, the efficiency bound is given in Newey as $E[\varphi\varphi']$ with instrument φ. Notice that φ is the expectation of the treatment conditioned on X_F multiplied by the parametric derivative of the $\alpha-functions$. For the semiparametric model considered here, the functional form of the $\alpha-functions$ are unknown. We write these functions as $\alpha(V_j(\theta),\theta)$ to indicate a dual dependence on θ. Namely, changes in θ will directly change α through the index (the parametric derivative), holding the function fixed. Indirectly, as $\alpha(V_j(\theta),\theta)$ is defined in (D8) as an expectation conditioned on $V_j(\theta)$, changes in θ will change the form of the $\alpha-functions$. The semiparametric derivative, $\bar{\mathfrak{R}}_\varphi$, reflects this dual dependence of the α functions on θ, which accounts for the difference between it and the parametric derivative.

Employing the results above in Theorems 2-4, Theorem 5 below provides large sample results for quantile marginal effects. Namely, it provides conditions under which the estimator is consistent, asymptotically distributed as normal, and efficient.

Theorem 5. Consistency, Normality and Optimality for Marginal Effects. *Under (D1)-(D12) and assumptions (A1)-(A4) and (A6)-(A9):*

$$\hat{M}_q \xrightarrow{p} M_q$$

and

$$\sqrt{N}\left[\hat{M}_q - M_q\right] \xrightarrow{d} Z \sim N(0, E\left[\psi_j\psi_j'\right])$$

where $\psi_j \equiv \psi_{1j} + \psi_{2j} + \psi_{3j}$. *To define* ψ_{kj}, *recall the Bahadur representation in (D13), and let* $t_{qj} \equiv t_{qj}(q, X_{jC})$. *Then, with* $\bar{\mathfrak{R}}_z(V_j)$ defined as in Theorem 3 and $\alpha_{oj} \equiv \alpha(V_j(\theta_0),\theta_0)$:

$$\psi_{1j} \equiv \{\bigtriangledown_q E[t_{qj}\alpha_{oj}] \quad - \bigtriangledown_q E(t_{qj}) M_q\}\frac{\mathfrak{B}_j}{E(t_{qj})} - \frac{t_{qj} - E[t_{qj}]}{E[t_{qj}]}M_q + \frac{t_{qj}\alpha_{oj}}{E[t_{qj}]} - M_q$$

$$\psi_{2j} \equiv -\frac{1}{E[t_{qj}]}E\{t_{qj}\bigtriangledown_\theta[\bar{\alpha}(V_j(\theta),\theta))]_{\theta_0}\}H_0^{-1}\bar{\mathfrak{R}}_z(V_j)'\varepsilon_j$$

$$\psi_{3j} \equiv \frac{E(t_{qj}|V_j)}{E[t_{qj}]}\bar{\Omega}^*(V_j)^{-1} g_V(V_j)\bar{Q}_j^{*\prime}\varepsilon_j = \frac{E(t_{qj}|V_j)}{E[t_{qj}]}E\left[\bar{Q}_j^{*\prime}R_j\right]^{-1}\bar{Q}_j^{*\prime}\varepsilon_j.$$

Under (A5), the covariance matrix $E\left[\psi_j\psi_j'\right]$ *is minimal.*

With marginal effects being important in applications with nonlinear models, Theorem 5 makes it possible to construct confidence intervals for those quantile marginal effects of interest. For expositional purposes, the theorem covers the case in which marginal effects are calculated over quantiles for a continuous variable of interest. This theorem also readily applies to the case in which the variable of interest is discrete. In this case, it would be natural to calculate the marginal effect by averaging over observations for which the discrete variable takes on a given support point. Accordingly, in the discrete case, replace the quantile indicator for this variable with an indicator for each support point. With the support points known, this indicator poses no difficulties as it can be taken as known. Because of the additively separable form of the covariance matrix, an efficient instrument for the parameter estimated ensures an efficient estimator for quantile marginal effects.

The form of the covariance matrix also reveals that it depends on the estimation uncertainty in estimating quantile regions. Intuitively, from Newey and McFadden (1994), such uncertainty should enter the covariance matrix because the consistency of quantile marginal effects will in part depend on consistently estimating quantile regions. Lemma 3 provides results needed to incorporate such uncertainty.

5 Monte Carlo

In the Monte Carlo study, we investigate six different designs. They differ in two dimensions: whether the treatment model is misspecified and the form of the marginal effect function. For the odd-numbered designs, the treatment model is correctly specified. The model structure is given as:

$$T = 1\{V_2 \geq u\} \text{ where } V_2 = X_2 + Z; \tag{11}$$
$$Y = V_1 + M(V_1) \times T + \varepsilon \text{ where } V_1 = X_1 + X_2. \tag{12}$$

The error terms u and ε are homoskedastic, correlated, and follow normal distributions, each with expectation 0 and standard deviation 2. The variables X_1 and X_2 follow normal distributions and are correlated with expectation 0 and standard deviation of 2 and 1 respectively. The marginal effect function $M(\cdot)$ is set to be linear for the first and second designs, quadratic for the third and fourth, and exponential for the fifth and sixth. The even-numbered designs differ from the odd-numbered ones in that a variable X_3 is incorrectly omitted from the true treatment model given as:

$$T = 1\{V_2 + X_3 \geq u\} \text{ where } V_2 = X_2 + Z. \tag{13}$$

We generate X_3 so that it is normally distributed and correlated with X_1 and X_2. The sample size in the Monte Carlo is 2000, and we ran 100 replications.

The optimal semiparametric instrument depends on the derivative of the unknown $\alpha-functions$ and the vector of true parameters. To obtain these quantities, we require an initial instrument that does not depend on them. For this purpose, notice in the parametric case that the optimal instrument vector is : $[E(T_j|V_{Tj}), 1]'[\nabla_v \alpha]X_{2j}$. To obtain the initial instrument vector, we set $\nabla_v \alpha$ at the constant vector $[1, 1]'$ so that the preliminary instruments becomes: $\left[\hat{P}_j \quad + 1\right] X_{2j}$. Using this instrument to estimate the $\alpha-$functions and θ_0, we then construct the optimal instrument given in Theorem 4 and re-estimate the model**.** While the estimators based on this two-step IV strategy have desirable large-sample properties, we have found it beneficial in finite samples to perform an additional iteration where we recalculate the optimal instrument based on this new estimator. The IV estimator for θ_0 reported below is based on this final iteration.

The substitution estimator minimizes:

$$\frac{1}{N} \sum_j \left[Y_j - \hat{Q}_j \hat{\alpha}\left(V_j(\theta), \theta\right) \right]^2$$

where the $\hat{\alpha}\left(V_j(\theta), \theta\right)$ function is defined as:

$$\hat{\alpha}\left(V_j(\theta), \theta\right) = \left[\hat{Q}' D_N\left(V_j(\theta), \theta\right) \hat{Q} \right]^{-1} \hat{Q}' D_N\left(V_j(\theta), \theta\right) Y$$

with $\hat{Q}_j \equiv \left[\hat{P}_j \;\; 1\right]$.

In reporting results, recall that the treatment model (11) is correctly specified over odd-numbered designs, and the misspecified model (13) is used in even-numbered designs. The tables below provide the estimation results for the index parameters in both treatment and outcome equations and also estimated marginal effects. We supply not only the median marginal effect estimates but also the quartile marginal effect estimates. The true parameter value and marginal effects are provided in the first column; while mean, median, standard deviation (STDEV), median absolute deviation (MAD), root mean squared error (RMSE) and median absolute error (MAE) are provided in the other columns.

Beginning with the treatment index parameter, it is quite well estimated with a very small RMSE when the treatment model is correct. As expected, under misspecification, the RMSE is much larger, primarily due to the increased bias but also to the higher standard deviation.

Turning to the estimation results for the outcome equation, the outcome index parameter is well estimated by the IV estimator in all designs. In comparison, the substitution estimator generally performs well in correctly specified design except the exponential design where its RMSE is significantly higher than that of the IV estimator (.16 vs. .08). In the misspecified designs, where the substitution estimator is not theoretically consistent, it indeed has a very large RMSE due mainly to a substantial bias.

The results for the quantile marginal effects are more pronounced than those for the outcome index parameter. The IV estimator performs significantly better than the substitution estimator even in the correctly specified designs. The RMSE is around half of the substitution estimator for median marginal effects. In the misspecified designs, the advantage of the IV estimator is even more striking. The RMSE's of the IV estimator for median marginal effects stay low at .33, .55 and .57 for linear, quadratic and exponential designs, while the corresponding RMSE's for the substitution estimator are very high at 2.65, 1.39 and 2.69 respectively. As marginal effect estimation is arguably the main object of interest in empirical studies and treatment models are likely to be misspecified, the robustness of the marginal effect IV estimator is important.

	TRUTH	MEAN	MEDIAN	STDEV	MAD	RMSE	MAE
Design 1							
Treatment Model Parameter Estimator							
	1.00	1.00	1.00	0.08	0.06	0.08	0.06
Outcome Model Parameter IV Estimator							
	1.00	1.04	1.03	0.07	0.05	0.08	0.04
Outcome Model Parameter Substitution Estimator							
	1.00	1.01	1.00	0.07	0.05	0.07	0.05
Marginal Effect IV Estimator							
Median	5.00	5.04	5.01	0.24	0.15	0.25	0.15
1st Quartile	3.20	3.32	3.28	0.32	0.18	0.34	0.20
2nd Quartile	4.49	4.54	4.52	0.30	0.21	0.30	0.21
3rd Quartile	5.51	5.50	5.50	0.30	0.22	0.30	0.21
4th Quartile	6.81	6.79	6.75	0.36	0.21	0.36	0.18
Marginal Effect Substitution Estimator							
Median	5.00	5.42	5.42	0.27	0.17	0.50	0.42
1st Quartile	3.20	3.65	3.61	0.32	0.22	0.56	0.41
2nd Quartile	4.49	4.99	4.89	0.32	0.21	0.52	0.40
3rd Quartile	5.51	5.91	5.89	0.30	0.23	0.49	0.39
4th Quartile	6.81	7.21	7.24	0.37	0.28	0.54	0.44
Design 2							
Treatment Model Parameter Estimator							
	1.00	1.49	1.49	0.21	0.14	0.54	0.49
Outcome Model Parameter IV Estimator							
	1.00	1.06	1.06	0.08	0.05	0.10	0.06
Outcome Model Parameter Substitution Estimator							
	1.00	0.39	0.39	0.07	0.05	0.61	0.61
Marginal Effect IV Estimator							
Median	5.00	4.98	4.96	0.33	0.21	0.33	0.22
1st Quartile	3.20	3.34	3.29	0.59	0.29	0.60	0.31
2nd Quartile	4.49	4.49	4.47	0.48	0.33	0.48	0.33
3rd Quartile	5.51	5.39	5.40	0.44	0.29	0.45	0.30
4th Quartile	6.81	6.70	6.65	0.53	0.39	0.54	0.41
Two-stage Marginal Effect Estimator							
Median	5.00	7.54	7.43	0.77	0.52	2.65	2.43
1st Quartile	3.20	5.67	5.70	0.80	0.57	2.60	2.50
2nd Quartile	4.49	7.04	6.93	0.94	0.63	2.72	2.44
3rd Quartile	5.51	8.07	8.15	0.91	0.59	2.72	2.64
4th Quartile	6.81	9.36	9.37	1.03	0.74	2.75	2.57

	TRUTH	MEAN	MEDIAN	STDEV	MAD	RMSE	MAE
Design 3							
Treatment Model Parameter Estimator							
	1.00	1.00	1.00	0.08	0.06	0.08	0.06
Outcome Model Parameter IV Estimator							
	1.00	1.06	1.06	0.06	0.04	0.08	0.06
Outcome Model Parameter Substitution Estimator							
	1.00	1.01	1.02	0.09	0.06	0.09	0.06
Marginal Effect IV Estimator							
Median	7.36	7.31	7.29	0.22	0.13	0.23	0.14
1st Quartile	8.93	8.64	8.64	0.34	0.21	0.45	0.35
2nd Quartile	5.77	5.83	5.81	0.25	0.18	0.25	0.16
3rd Quartile	5.78	5.87	5.86	0.29	0.20	0.30	0.21
4th Quartile	8.94	8.89	8.89	0.47	0.24	0.47	0.26
Marginal Effect Substitution Estimator							
Median	7.36	7.75	7.73	0.25	0.18	0.47	0.37
1st Quartile	8.93	9.00	8.97	0.43	0.32	0.44	0.32
2nd Quartile	5.77	6.19	6.15	0.27	0.20	0.50	0.38
3rd Quartile	5.78	6.32	6.28	0.29	0.22	0.62	0.51
4th Quartile	8.94	9.49	9.47	0.60	0.38	0.81	0.59
Design 4							
Treatment Model Parameter Estimator							
	1.00	1.49	1.49	0.21	0.14	0.54	0.49
Outcome Model Parameter IV Estimator							
	1.00	1.06	1.08	0.13	0.06	0.14	0.10
Outcome Model Parameter Substitution Estimator							
	1.00	0.54	0.54	0.11	0.07	0.47	0.46
Marginal Effect IV Estimator							
Median	7.36	7.32	7.25	0.55	0.22	0.55	0.28
1st Quartile	8.93	8.58	8.58	0.86	0.45	0.93	0.55
2nd Quartile	5.77	5.82	5.83	0.52	0.37	0.52	0.37
3rd Quartile	5.78	5.86	5.79	0.62	0.34	0.63	0.34
4th Quartile	8.94	9.04	8.77	1.22	0.36	1.22	0.38
Marginal Effect Substitution Estimator							
Median	7.36	8.59	8.51	0.63	0.43	1.39	1.15
1st Quartile	8.93	7.72	7.81	1.11	0.82	1.64	1.12
2nd Quartile	5.77	6.21	6.16	0.64	0.38	0.77	0.53
3rd Quartile	5.78	7.64	7.57	0.89	0.61	2.07	1.79
4th Quartile	8.94	12.79	12.77	1.64	1.15	4.18	3.83

	TRUTH	MEAN	MEDIAN	STDEV	MAD	RMSE	MAE
Design 5							
Treatment Model Parameter Estimator							
	1.00	1.00	1.00	0.08	0.06	0.08	0.06
Outcome Model Parameter IV Estimator							
	1.00	1.07	1.07	0.04	0.03	0.08	0.07
Outcome Model Parameter Substitution Estimator							
	1.00	1.06	1.07	0.15	0.10	0.16	0.10
Marginal Effect IV Estimator							
Median	7.90	7.87	7.86	0.30	0.18	0.30	0.18
1st Quartile	5.23	5.18	5.15	0.29	0.20	0.29	0.19
2nd Quartile	5.77	5.67	5.60	0.27	0.16	0.29	0.21
3rd Quartile	7.14	7.06	7.06	0.33	0.24	0.33	0.23
4th Quartile	13.48	13.55	13.49	0.77	0.46	0.77	0.46
Marginal Effect Substitution Estimator							
Median	7.90	8.41	8.38	0.59	0.35	0.77	0.50
1st Quartile	5.23	5.57	5.57	0.36	0.24	0.49	0.36
2nd Quartile	5.77	6.20	6.16	0.38	0.23	0.57	0.41
3rd Quartile	7.14	7.70	7.67	0.48	0.30	0.73	0.54
4th Quartile	13.48	14.17	14.14	1.43	0.91	1.59	1.02
Design 6							
Treatment Model Parameter Estimator							
	1.00	1.49	1.49	0.21	0.14	0.54	0.49
Outcome Model Parameter IV Estimator							
	1.00	1.10	1.09	0.06	0.03	0.12	0.09
Outcome Model Parameter Substitution Estimator							
	1.00	0.67	0.66	0.18	0.11	0.38	0.35
Marginal Effect IV Estimator							
Median	7.90	7.69	7.66	0.53	0.27	0.57	0.40
1st Quartile	5.23	5.18	5.05	0.74	0.31	0.74	0.34
2nd Quartile	5.77	5.50	5.45	0.54	0.36	0.60	0.43
3rd Quartile	7.14	6.76	6.76	0.53	0.36	0.65	0.45
4th Quartile	13.48	13.31	13.14	1.04	0.56	1.04	0.63
Marginal Effect Substitution Estimator							
Median	7.90	10.19	10.03	1.42	0.76	2.69	2.13
1st Quartile	5.23	6.09	5.99	0.87	0.60	1.23	0.85
2nd Quartile	5.77	6.93	6.78	0.95	0.60	1.49	1.01
3rd Quartile	7.14	8.98	8.98	1.26	0.82	2.23	1.84
4th Quartile	13.48	18.75	18.43	3.38	1.91	6.26	4.99

6 Conclusions

In conclusion, we have proposed IV estimators for the index parameters and quantile marginal effects in an outcome model with endogenous treatments. Since marginal effects are of primary interest in applied work, we have developed the theory for quantile marginal effects as well as the index parameters that enter the marginal effect functions. We have shown that the proposed IV estimators are consistent and asymptotically normally distributed irrespective of whether or not the treatment model is correctly specified. As the model for treatments can be difficult to specify in practice, this robustness property is important in applications. When the treatment model is correct, we establish efficiency for the proposed IV estimators for quantile marginal effects and for the finite dimensional vector of index parameters upon which they depend.

As is evidenced in the Monte Carlo study, the proposed estimators perform very well in finite samples. There is a payoff to an efficient estimator, particularly in the case for marginal effects. Perhaps most importantly for applications where the treatment model may be difficult to specify, the proposed IV estimator is very robust and performs well under misspecification.

References

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

[2] Amemiya, T. (1974): "The Non-linear Two-stage Least Squares Estimator," *Journal of Econometrics*,2, 105-110.

[3] Amemiya, T. (1977): "The Maximum Likelihood and Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955-968.

[4] Amemiya, T. (1985): *Advanced Econometrics*, Harvard University Press, Cambridge.

[5] Berge, C.(1963). *Topological Spaces*, Oliver and Boyd, Edinburgh.

[6] Bahadur, R. R (1963): "A Note on Quantiles in Large Samples," *Annals of Mathematical Statistics*, 37, 577-580.

[7] Bhatacharaya, P. K. (1967): "Estimation of a Probability Density and its Derivatives," *Indian Journal of Statistics,* Series A, 373-383.

[8] Blundell, R. and J.L. Powell (2001): "Endogeneity in Nonparametric and Semiparametric Regression Models," presented at the World;d Congress of the Econometric Society, 2000.

[9] Blundell, R. W. and J.L. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655-679

[10] Cai, Z., M. Das, H. Xiong, and X. Wu (2005): "Functional Coefficient Instrumental Variable Models, " *Journal of Econometrics*, 133, 207-241.

[11] Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric. Problems," *Econometrica*, Vol. 69, No. 5, pp. 1127-1160

[12] David, H. A. (1981): *Order Statistics, 2nd Edition,* Wiley Series in Probability and Mathematical Statistics.

[13] Das, M. (2005): "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.

[14] Fan, J. and I. Gijbels (1996): *Local Polynomial Modeling and Its Applications*, Chapman and Hall, Monographs on Statistics and Applied Probability, 66.

[15] Hausman, J.A. (1983): "Specification and Estimation of Simultaneous Equation Models," Chapter 7 in *Handbook of Econometrics*, V. 1, Elsevier, Z. Griliches and M.D. Intriligator editors, 391-448.

[16] Heckman, J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 931-959

[17] Imbens, G. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, V. 77, No. 5, 1481-1512.

[18] Klein, R. and R. Spady (1993): "An Efficient Semiparametric Estimator for the Binary Response Model," Econometrica. V. 59.

[19] Klein, R., C. Shen, and F. Vella (2014): "Marginal Effects in Semiparametric Selection Models with Binary Outcomes," Journal of Econometrics, V. 185, 62-94.

[20] Klein, R. and C. Shen (2011): "Bias Corrections in Estimating and Testing Single Index Models," Econometric Theory.

[21] Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables ", Journal of Econometrics, 97, 145–177.

[22] Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," Chapter 9 in *Handbook of Econometrics*, V. 4, Elsevier, R. F. Engle and D. L. Mcfadden editors, 2111-2245.

[23] Newey, W. K. (1990): "Efficient Instrumental Variable Estimation of Nonlinear Models, *Econmometrica*, JV. 58, No. 4, 809-817.

[24] Newey, W. K., J.L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Models," *Econometrica*, 67, 565–603.

[25] Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, v. 71, No. 5, 1565-1578.

[26] Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, v.57, No.5, 1027-1057.

[27] Rothe, C. (2009): "Semiparametric Estimation of Binary Response Models with Endogenous Regressors," *Journal of Econometrics*, V. 153, 61-64.

[28] Yang, X. M, .X. Q. Yang, and K. L. Teo, "A matrix trace inequality," *Journal of Mathematical Analysis and Applications*, vol. 263, no. 1, pp. 327–331, 2001.

7 Appendix

7.1 Main Theorems

Proof of Theorem 1. Referring to (D8), $\bar{\Omega}^*(v,\theta)\bar{\alpha}(v,\theta) \equiv E\left[\bar{Q}_j^{*\prime} Y_j | V_j(\theta) = v\right] g_V(v)$ has a unique solution for $\bar{\alpha}(v,\theta)$ if $\bar{\Omega}^*(v,\theta)^{-1}$ exists. From Lemma 6a, $\bar{\Omega}^*(v,\theta)$ is non-singular for all v in a compact set. Therefore $\bar{\alpha}\left(v,\theta\right)$ is identified. A similar argument holds for $\alpha\left(v,\theta\right)$. From Newey and McFadden(1994, Lemma 2.3), identification of θ_0 follows from the assumption (A9) that $E\left\{Z_j'\left[Y_j - R_j\bar{\alpha}\left(V_j\left(\theta\right),\theta\right)\right]\right\} = 0$ is uniquely solved at θ_0.

Proof of Theorem 2. Denoting $\hat{G}_k^*\left(\theta\right)$ as the k^{th} component of $\hat{G}^*\left(\theta\right)$ as in (D10), we establish the following uniform convergence results:

$$\sup_\theta |\hat{G}_k^*\left(\theta\right) - E[Z_{jk}[Y_j - R_j\bar{\alpha}(V_j\left(\theta\right),\theta)]]| \xrightarrow{p} 0. \tag{14}$$

We prove this result by analyzing the following upper bound for the above term.

$$\sup_\theta \left| \frac{1}{N}\sum_{j=1}^N Z_{jk}\left[Y_j - R_j\bar{\alpha}\left(V_j\left(\theta\right),\theta\right)\right] - E\left(Z_{jk}\left[Y_j - R_j\bar{\alpha}\left(V_j\left(\theta\right),\theta\right)\right]\right)\right| \tag{15}$$

$$+\sup_\theta \left| \frac{1}{N}\sum_{j=1}^N \left[\hat{Z}_{jk} - Z_{jk}\right]\left[Y_j - R_j\bar{\alpha}\left(V_j\left(\theta\right),\theta\right)\right]\right| \tag{16}$$

$$+\sup_\theta \left| \frac{1}{N}\sum_{j=1}^N Z_{jk}R_j\left[\hat{\alpha}^*\left(V_j\left(\theta\right),\theta\right) - \bar{\alpha}\left(V_j\left(\theta\right),\theta\right)\right]\right| \tag{17}$$

$$+\sup_\theta \left| \frac{1}{N}\sum_{j=1}^N \left[\hat{Z}_{jk} - Z_{jk}\right] R_j\left[\hat{\alpha}^*\left(V_j\left(\theta\right),\theta\right) - \bar{\alpha}\left(V_j\left(\theta\right),\theta\right)\right]\right|. \tag{18}$$

Turning to the term in (15), it is $o_p(1)$ from a standard uniform convergence result on an i.i.d. sample mean converging uniformly to its expectation (see e.g. Amemiya (1985)). Noting the definition of $\hat{Z}$ in (D10), and the assumption on $\hat{Z}^o$ in (D9), from Lemma 5b with $\hat{W}_i \equiv \hat{Z}_{jk}^o$ and $W_i \equiv Z_{jk}^o$, the second term converges to zero. With $E\left|Z\right|$ bounded from (A7), the third term vanishes from Lemma 5a and Lemma 8d. The final term is $o_p(1)$ from (D9), Lemma 8d, and Lemma 5b.

Since the uniform limit of the moment condition is uniquely solved at θ_0 from (A9), it now follows that $\hat{\theta}^* \xrightarrow{p} \theta_0$.

Proof of Theorem 3. Referring to (D10), from a Taylor series expansion:

$$\sqrt{N}(\hat{\theta}^* - \theta_0) = -\left[\nabla_\theta \hat{G}^*(\theta^+)\right]^{-1}\sqrt{N}\hat{G}^*(\theta_0) + o_p\left(1\right).$$

From Lemma 9

$$\nabla_\theta \hat{G}^*(\theta^+) \xrightarrow{p} H_0 = -E\left[Z_j' R_j \nabla_\theta \bar{\alpha}\left(V_j\left(\theta\right), \theta\right)_{\theta=\theta_0}\right].$$

From Lemma 10 with $V_j \equiv V_j(\theta_0)$,

$$\sqrt{N}\hat{G}^*(\theta_0) \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^{N} Z_j'[Y_j - R_j \hat{\alpha}_0^*(V_j)] = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \bar{\mathfrak{R}}_z'\left(V_j\right) \varepsilon_j + o_p\left(1\right),$$

where

$$\bar{\mathfrak{R}}_z'\left(V_j\right) \equiv Z_j' - E(Z_j' R_j | V_j) E\left[\bar{Q}_j^{*'} R_j | V_j\right]^{-1} \bar{Q}_j^{*'}.$$

Substituting for gradient and hessian expressions, we have:

$$\sqrt{N}(\hat{\theta}^* - \theta_0) = -H_0^{-1} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \bar{\mathfrak{R}}_z'\left(V_j\right) \varepsilon_j + o_p\left(1\right). \tag{19}$$

The theorem now directly follows.

Proof of Theorem 4. To prove efficiency, we require a convenient characterization for the covariance matrix under any instrument Z. To this end, it is useful to note that under correct treatment specification in (A5), for any function $f(V)$: $E(f(V)'\bar{Q}^{*'}\bar{\mathfrak{R}}_\varphi) = 0$. To establish this result, recall the definition of $\bar{\mathfrak{R}}_z$. Then, suppressing subscripts and function arguments, for any instrument Z, from (A5)

$$\bar{\mathfrak{R}}_z \equiv Z - \bar{Q}^* E\left[\bar{Q}^{*'} R | V\right]^{-1} E(R' Z | V) = Z - \bar{Q}^* E\left[\bar{Q}^{*'} \bar{Q}^* | V\right]^{-1} E(\bar{Q}^{*'} Z | V), \tag{20}$$

which follows by first taking an expectation conditioned on X_F and then over X_F conditioned on V for each expectation. Employing (20) and (A5):

$$\begin{aligned} E(f(V)'\bar{Q}^{*'}\bar{\mathfrak{R}}_\varphi) &= E\left[f(V)'\bar{Q}^{*'}\varphi - f(V)'\bar{Q}^{*'}\bar{Q}^* E\left[\bar{Q}^{*'}\bar{Q}^* | V\right]^{-1} E(\bar{Q}^{*'}\varphi | V)\right] \\ &= E\left[f(V)'\bar{Q}^{*'}\varphi\right] - E_V E\left[f(V)'\bar{Q}^{*'}\bar{Q}^* E\left[\bar{Q}^{*'}\bar{Q}^* | V\right]^{-1} E(\bar{Q}^{*'}\varphi | V) | V\right] \\ &= E\left[f(V)'\bar{Q}^{*'}\varphi\right] - E\left[f(V)' E[\bar{Q}^{*'}\bar{Q}^* | V] E[\bar{Q}^{*'}\bar{Q}^* | V]^{-1} E(\bar{Q}^{*'}\varphi | V)\right] \\ &= E\left[f(V)'\bar{Q}^{*'}\varphi\right] - E[f(V)' E(\bar{Q}^{*'}\varphi | V)] = 0, \end{aligned} \tag{21}$$

a result that also follows because $\bar{\mathfrak{R}}_\varphi$ is a residual from the local projection of φ on $\bar{Q}^*$. Under (A5) and recalling that Z includes trimming $t_1 t_2$,

$$H_0 = -E\left(Z_j' \bar{Q}_j^* \nabla_\theta \bar{\alpha}\left(V_j\left(\theta\right), \theta\right)_{\theta=\theta_0}\right) \tag{22}$$

Therefore, under Theorem 3 and Lemma 4,

$$H_0 = E\left[Z_j' \bar{\mathfrak{R}}_\varphi(V_j)\right] \tag{23}$$

$$= E(\bar{\mathfrak{R}}_z'(V_j)\bar{\mathfrak{R}}_\varphi(V_j)) + E\left\{[Z_j - \bar{\mathfrak{R}}_z(V_j)]'\bar{\mathfrak{R}}_\varphi\right\} = E(\bar{\mathfrak{R}}_z'(V_j)\bar{\mathfrak{R}}_\varphi(V_j)) \tag{24}$$

where the last equality follows from (21) because $\left[Z_j - \bar{\mathfrak{R}}_z(V_j)\right]$ has the form $\bar{Q}_j^* f(V_j)$. Under assumption (A5), the variance of the estimator based on any instrument Z in Theorem 3 then simplifies to:

$$Cov = \sigma_\varepsilon^2 \left[H_0^{-1}\right] E[\bar{\mathfrak{R}}_z(V_j)'\bar{\mathfrak{R}}_z(V_j)] \left[H_0^{-1}\right]' \tag{25}$$

$$= \sigma_\varepsilon^2 \left[E\left(\bar{\mathfrak{R}}_z'(V_j)\bar{\mathfrak{R}}_\varphi(V_j)\right)\right]^{-1} E[\bar{\mathfrak{R}}_z(V_j)'\bar{\mathfrak{R}}_z(V_j)] \left[E\left(\bar{\mathfrak{R}}_\varphi'(V_j)\bar{\mathfrak{R}}_z(V_j)\right)\right]^{-1} \tag{26}$$

With $Z_j = \bar{\mathfrak{R}}_\varphi(V_j)$ the general form of the hessian in (23) simplifies to:

$$H_0 = -E\left(\bar{\mathfrak{R}}_\varphi'(V_j)\bar{\mathfrak{R}}_\varphi(V_j)\right). \tag{27}$$

Substituting the hessian in (27) into the general form of the covariance matrix Cov in (25) yields:

$$Cov^* = o_\varepsilon^2 \left[E\left(\bar{\mathfrak{R}}_\varphi'(V_j)\bar{\mathfrak{R}}_\varphi(V_j)\right)\right]^{-1}. \tag{28}$$

To show that $Cov^* \leq Cov$, which establishes the efficiency bound, we follow an argument in Newey and McFadden (1994, section 5) and let

$$U \equiv \bar{\mathfrak{R}}_z'(V_j) - E(\bar{\mathfrak{R}}_z'(V_j)\bar{\mathfrak{R}}_\varphi(V_j)) \left[E(\bar{\mathfrak{R}}_\varphi'(V_j)\bar{\mathfrak{R}}_\varphi(V_j))\right]^{-1} \bar{\mathfrak{R}}_\varphi'(V_j).$$

Then, $Cov - Cov^*$ is positive semi-definite because:

$$Cov - Cov^* = \left[E\left(\bar{\mathfrak{R}}_z'(V_j)\bar{\mathfrak{R}}_\varphi(V_j)\right)\right]^{-1} E(UU') \left[E\left(\bar{\mathfrak{R}}_\varphi'(V_j)\bar{\mathfrak{R}}_z(V_j)\right)\right]^{-1}$$

and $E(UU')$ is positive semi-definite. Recalling that Cov^* in (28) was obtained for $Z_j = \bar{\mathfrak{R}}_\varphi(V_j)$, the bound is obtained under this instrument.

Proof of Theorem 5. To prove consistency for the quantile marginal effect estimator, refer to (D11)-(D12) and write:

$$\begin{aligned}
\left[\hat{M}_q - M_q\right] &= \Delta_1 + \Delta_2, \qquad (29)\\
\Delta_1 &= \left(N/\hat{N}_q\right) \frac{1}{N} \sum_{j=1}^{N} \left[t_{qj}\left(\hat{q}, X_{jC}\right) - t_{qj}\left(q, X_{jC}\right)\right] \hat{\alpha}^*(V_j(\hat{\theta}^*), \hat{\theta}^*)\\
\Delta_2 &= \left(N/\hat{N}_q\right) \frac{1}{N} \sum_{i=1}^{N} t_{qj}\left(q, X_{jC}\right) \hat{\alpha}^*(V_j(\hat{\theta}^*), \hat{\theta}^*) - M_q,
\end{aligned}$$

where $\hat{N}_q \equiv \sum_{i=1}^{N} t_{qj}\left(\hat{q}, X_{jC}\right)$. From Lemma 3,

$$\frac{\hat{N}_q - N_q}{N} \equiv \frac{1}{N} \sum_{j=1}^{N} \left[t_{qj}(\hat{q}, X_{jC}) - t_{qj}\left(q, X_{jC}\right)\right] = O_p(N^{-1/2}).$$

Therefore, $\hat{N}_q/N \xrightarrow{p} E\left[t_{qj}\left(q, X_{jC}\right)\right] > 0$. It follows from Lemma 5 and Lemma 8d that $\Delta_1 = o_p(1)$.

For Δ_2, from the convergence of $\left(\hat{N}_q/N\right)$ to $E[t_{qj}\left(q, X_{jC}\right)] \quad > 0$, it suffices to show $\left(\hat{N}_q/N\right)\Delta_2 = o_p(1)$. We have:

$$\left(\hat{N}_q/N\right)\Delta_2 \leq \frac{1}{N}\sum_{i=1}^{N} t_{qj}\left(q, X_{jC}\right)\left|\hat{\alpha}^*(V_j(\hat{\theta}^*), \hat{\theta}^*) - \alpha(V_j(\hat{\theta}^*), \hat{\theta}^*)\right| \tag{30}$$

$$+\frac{1}{N}\sum_{i=1}^{N} t_{qj}\left(q, X_{jC}\right)\left|\alpha(V_j(\hat{\theta}^*), \hat{\theta}^*) - \alpha(V_j(\hat{\theta}^*), \theta_0)\right| \tag{31}$$

$$+\frac{1}{N}\sum_{i=1}^{N} t_{qj}\left(q, X_{jC}\right)\left|\alpha(V_j(\hat{\theta}^*), \theta_0) - \alpha(V_j(\theta_0), \theta_0)\right| \tag{32}$$

$$+\left|\frac{1}{N}\sum_{i=1}^{N} t_{qj}\left(q, X_{jC}\right)\alpha(V_j(\theta_0), \theta_0) - M_q\left(\hat{N}_q/N\right)\right|. \tag{33}$$

From Lemma 8d, the term in (30) is $o_p(1)$. For the term in (31), with θ and v in compact sets and under the assumption that $\alpha\left(v, \theta\right)$ is continuous in both arguments, from Berge (1963), $\sup_v \left|\alpha\left(v, \theta\right) - \alpha\left(v, \theta_0\right)\right|$ is continuous in θ. Therefore, this term is $o_p(1)$ under the continuous function theorem because $\hat{\theta}^* \xrightarrow{p} \theta_0$. For (32), the result follows from a standard Taylor series argument. For the term in (33), the result immediately follows. Therefore, $\hat{M}_q \xrightarrow{p} M_q$.

To establish asymptotic normality, from (29), we study

$$\sqrt{N}\left[\hat{M}_q - M_q\right] = \sqrt{N}\left[\Delta_1 + \Delta_2\right].$$

Referring to (29) , write : $\sqrt{N}\Delta_1 = \sqrt{N}\Delta_{11} + \sqrt{N}\Delta_{12}$, where:

$$\sqrt{N}\Delta_{11} = \left(N/\hat{N}_q\right)\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left[t_{qj}(\hat{q}, X_{jC}) - t_{qj}\left(q, X_{jC}\right)\right]\left[\hat{\alpha}^*(V_j(\hat{\theta}^*), \hat{\theta}^*) - \alpha(V_j(\theta_0), \theta_0)\right] \tag{34}$$

$$\sqrt{N}\Delta_{12} = \left(N/\hat{N}_q\right)\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left[t_{qj}(\hat{q}, X_{jC}) - t_{qj}\left(q, X_{jC}\right)\right]\alpha(V_j(\theta_0), \theta_0)). \tag{35}$$

With $\left(N/\hat{N}_q\right)$ converging to a positive number, $\sqrt{N}\Delta_{11}$ in (34) is $o_p(1)$ from a Taylor series expansion in $\hat{\theta}^*$, Theorem 3, Lemma 5, and Lemma 3. Therefore, recalling (D13), from Lemma 3 and the convergence of $\left(N/\hat{N}_q\right)$ to $1/E\left[t_{qj}\left(q, X_{jC}\right)\right]$:

$$\sqrt{N}\Delta_1 = \frac{1}{E\left[t_{qj}\left(q, X_{jC}\right)\right]}\nabla_q E\left[t_{qj}(q, X_{jC})\alpha(V_j(\theta_0), \theta_0)\right]\sqrt{N}\bar{\mathfrak{B}} + o_p(1). \tag{36}$$

It now remains to characterize $\sqrt{N}\Delta_2$, the second component in (29). With $t_{qj} \equiv t_{qj}(q, X_{jC})$,

write it as the sum of:

$$\sqrt{N}\Delta_{21} : \left(N/\hat{N}_q\right) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} t_{qj} \left[\hat{\alpha}^*(V_j(\hat{\theta}^*), \hat{\theta}^*) - \hat{\alpha}^*(V_j(\theta_0), \theta_0)\right] \tag{37}$$

$$\sqrt{N}\Delta_{22} : \left(N/\hat{N}_q\right) \frac{1}{\sqrt{N}} \sum_{j=1}^{N} t_{\,qj} [\hat{\alpha}^*(V_j(\theta_0), \theta_0) - \alpha(V_j(\theta_0), \theta_0)] \tag{38}$$

$$\sqrt{N}\Delta_{23} : \left\{ \begin{array}{c} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \frac{t_{qj}\alpha(V_j(\theta_0),\theta_0)}{E[t_{qj}]} - M_q + \\ \frac{1}{\sqrt{N}}[\frac{1}{\hat{N}_q/N} - \frac{1}{E[t_{qj}]}] \sum_{j=1}^{N} t_{qj} \;\; \alpha(V_j(\theta_0), \theta_0) \end{array} \right\}. \tag{39}$$

From the convergence of $N/\hat{N}_q$ to $1/E[t_{qj}]$, a Taylor series expansion in $\hat{\theta}^*$, Theorem 3, and Lemma 8d it can be shown that:

$$\sqrt{N}\Delta_{21} = \frac{\frac{1}{N}\sum_{i=1}^{N} t_{qj} \left[\nabla_\theta \bar{\alpha}\left(V_j\left(\theta\right), \theta\right)\right]_{\theta_o}}{E\left[t_{qj}\right]} \sqrt{N}(\hat{\theta}^* - \theta_0) + o_p(1) \tag{40}$$

$$= \frac{E[t_{qj} \left[\nabla_\theta \bar{\alpha}\left(V_j\left(\theta\right), \theta\right)\right]_{\theta_o}]}{E\left[t_{qj}\right]} \sqrt{N}\left(\hat{\theta}^* - \theta_0\right) + o_p(1). \tag{41}$$

For $\sqrt{N}\Delta_{22,}$ from Lemma 11 and the convergence of $N/\hat{N}_q$ to $1/E[t_{qj}]$,

$$\sqrt{N}\Delta_{22} = \frac{1}{E\left[t_{qj}\right]} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E\left(t_{qj}|V_j\right) \bar{\Omega}^*\left(V_j\right)^{-1} g\left(V_j\right) \bar{Q}_j^{*\prime} \varepsilon_j + o_p(1) \tag{42}$$

Turning to the term in (39), it can be written up to $o_p(1)$ as

$$\sqrt{N}\Delta_{23} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{t_{qj}\left(q, X_{jC}\right) \alpha\left(V_j\left(\theta_0\right), \theta_0\right)}{E\left[t_{qj}\left(q, X_{jC}\right)\right]} - M_q - \Delta_q \tag{43}$$

$$\Delta_q \equiv \sqrt{N}\left[\frac{\hat{N}_q/N - E\left[t_{qj}\left(q, X_{jC}\right)\right]}{E\left[t_{qj}\left(q, X_{jC}\right)\right]}\right] M_q. \tag{44}$$

From the convergence of $\left(N/\hat{N}_q\right)$ to $1/E[t_{qj}(q, X_{jC})]$ and Lemma 3, up to $o_p(1)$, we have

$$\Delta_q = (\sqrt{N}\sum_{j=1}^{N}[t_{qj}(\hat{q}, X_{jC}) - t_{qj}(q, X_{jC})] + \sum_{j=1}^{N}[t_{qj}(q, X_{jC}) - Et_{qj}(q, X_{jC})]) \frac{M_q}{E\left[t_{qj}\left(q, X_{jC}\right)\right]}$$

$$= (\nabla_q E t_{qj}(q, X_{jC}) \sqrt{N}\bar{\mathfrak{B}} + \sum_{j=1}^{N}[t_{qj}(q, X_{jC}) - Et_{qj}(q, X_{jC})]) \frac{M_q}{E\left[t_{qj}\left(q, X_{jC}\right)\right]}$$

Substituting the above into the expression for $\sqrt{N}\Delta_{23}$, we have up to $o_p\left(1\right)$:

$$\sqrt{N}\Delta_{23} = \left\{ \begin{array}{c} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \frac{t_{qj}\left(q, X_{jC}\right)\alpha(V_j(\theta_0),\theta_0)}{E\left[t_{qj}\left(q, X_{jC}\right)\right]} - M_q \\ -\nabla_q E\left\{t_{qj}(q, X_{jC})\right\} \sqrt{N}\bar{\mathfrak{B}} \frac{M_q}{E\left[t_{qj}\left(q, X_{jC}\right)\right]} \\ -\frac{1}{\sqrt{N}} \sum_{j=1}^{N} [t_{qj}\left(q, X_{jC}\right) - E(t_{qj}(q, X_{jC}))] \frac{M_q}{E\left[t_{qj}\left(q, X_{jC}\right)\right]}. \end{array} \right\} \tag{45}$$

Combining (36), (41), (42), and (45):

$$\sqrt{N}\left[\hat{M}_q - M_q\right] = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}[\psi_{1j} + \psi_{2j} + \psi_{3j}] + o_p(1), \tag{46}$$

where with $t_{qj} \equiv t_{qj}\left(q, X_{jC/}\right)$, $\alpha_{oj} \equiv \alpha(V_j(\theta_0), \theta_0)$ and with $\left(\hat{\theta}^* - \theta_0\right)$ given from (19) in the proof of Theorem 3:

$$\psi_{1j} \equiv \{\bigtriangledown_q E[t_{qj}\alpha_{oj}] \quad - \bigtriangledown_q E(t_{qj}) M_q\}\frac{\mathfrak{B}_j}{E(t_{qj})} - \frac{t_{qj} - E[t_{qj}]}{E[t_{qj}]}M_q + \frac{t_{qj}\bigtriangledown_\theta \alpha_{oj}}{E[t_{qj}]} - M_q$$

$$\psi_{2j} \equiv -\frac{1}{E[t_{qj}]}E\{t_{qj}\bigtriangledown_\theta [\bar{\alpha}(V_j(\theta), \theta))]_{\theta_o}\}H_0^{-1}\bar{\mathfrak{R}}_z(V_j)'\varepsilon_j,$$

$$\psi_{3j} \equiv \frac{E(t_{qj}|V_j)}{E[t_{qj}]}\bar{\Omega}^*(V_j)^{-1} g_V(V_j)\bar{Q}_j^{*'}\varepsilon_j = \frac{E(t_{qj}|V_j)}{E[t_{qj}]}E\left[\bar{Q}_j^{*'}R_j\right]^{-1}\bar{Q}_j^{*'}\varepsilon_j.$$

The estimator now has a form to which a standard central limit theorem applies, and it remains to examine the form of the covariance matrix. Under correct treatment specification and the optimal instrument in Theorem 4, this matrix will be minimal if the $\psi_k - components$ above are uncorrelated. As the first term is uncorrelated with the other two, we complete the argument by examining the covariance between the last two terms. It suffices to analyze this covariance conditioned on V_j, which is proportional to:

$$C = E\left[\bar{\mathfrak{R}}_z(V_j)\varepsilon_j^2\bar{Q}_j^*|V_j\right] = \sigma_\varepsilon^2 E[\bar{\mathfrak{R}}_z(V_j)'\bar{Q}_j^*|V_j]. \tag{47}$$

From (20) in the Proof of Theorem 4, under (A5):

$$\bar{\mathfrak{R}}_z(V_j) = Z - \bar{Q}^* E\left[\bar{Q}^{*\prime}\bar{Q}^*|V\right]^{-1} E(\bar{Q}^{*\prime}Z|V). \tag{48}$$

Since $\bar{\mathfrak{R}}_z(V_j)$is the residual from the local projection of Z on $\bar{Q}^*$, C is 0, which completes the proof.

7.2 Intermediate Lemmas

Lemma 1 provides uniform bias and variance results; while Lemma 2 provides pointwise and uniform convergence rates for kernel weighted averages. As the proof for lemma 1 is standard (see, for example, Klein and Shen(2010)), it is not presented below, but is available upon request. The uniform results in Lemma 2, which are standard for bounded random variables, are provided here as in this context this bounded assumption will not always hold.

Lemma 1. Let $\mathfrak{S}_i$ be i.i.d. and not depending on θ. Define $v(\theta) \equiv x_1 + x_{II}\theta$, $x \equiv [x_1, x_{II}]$, $V_i(\theta) \equiv X_{1i} + X_{IIi}\theta$. With $\nabla_\theta^0 f \equiv f$, let $\nabla_\theta^d f$ be the dth partial derivative of f w.r.t. θ. Define:

$$\hat{g}_V(v(\theta)) \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}k\left(\frac{v(\theta) - V_i(\theta)}{h}\right)$$

With $g_{v|\mathfrak{S}}(\bullet)$ as the density for $V_i(\theta)$ conditioned on $\mathfrak{S}_i$, assume that $\bigtriangledown^d g_{V|\mathfrak{S}}, d = 0, 1, 2$ are uniformly bounded in x, θ. Further, $E(\mathfrak{S}_i^2)$ is bounded**.** Then, with the sup taken over x and θ in compact sets:

$$\begin{aligned}
&a);\ \ sup E[\mathfrak{S}_i \nabla_\theta^d \tfrac{1}{h} K\left(\tfrac{v(\theta)-V_i(\theta)}{h}\right)] - \nabla_\theta^d [E(\mathfrak{S}_i | V_i(\theta) = v(\theta)) g_V(v)] = O(h^2) \\
&b):\ \ \sup Var \mathfrak{S}_i \nabla_\theta^d \tfrac{1}{h} K\left(\tfrac{v(\theta)-V_i(\theta)}{h}\right) = O\left(h^{-(1+2d)}\right) \\
&c):\ \ sup \nabla_\theta^d E[\hat{g_V}(v(\theta))] = O(h^2) \\
&d):\ \ \sup Var\{\nabla_\theta^d \left[\frac{1}{h} k\left(\frac{v(\theta) - V_i(\theta)}{h}\right)\right]\} = O\left(h^{-(1+2d)}\right).
\end{aligned}$$

Lemma 2. Employing all notation and assumptions in Lemma 1, define:

$$\mathfrak{S}(v(\theta);\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} \mathfrak{S}_i \frac{1}{h} k\left(\frac{v(\theta) - V_i(\theta)}{h}\right);\ \ \bar{S}(v(\theta);\theta) \equiv \frac{\mathfrak{S}(v(\theta);\theta)}{\hat{g}_V(v(\theta))}.$$

For $\delta > 0$ arbitrarily small let $c_p \equiv \min\left[2r, \frac{1-(1+2d)r}{2}\right] - \delta$. Then for $g_V(v) \neq 0$:

$$\begin{aligned}
&a): \left|\nabla_\theta^d [\mathfrak{S}(v(\theta);\theta) - E(\mathfrak{S}_i | V_i(\theta) = v(\theta)) g_V(v(\theta))]\right| = o_p(N^{-c_p}) \\
&b): \left|\nabla_\theta^d [\hat{g}_V(v) - g_V(v)]\right| = o\left(N^{-c_p}\right) \\
&c): \left|\nabla_\theta^d \left[\bar{S}(v) - E(\mathfrak{S}|V = v)\right]\right| = o_p(N^{-c_p}).
\end{aligned}$$

Further assume that $E[|\mathfrak{S}_i|^m] = O(1)$, $E\left[\left|h^{1+d}\nabla_\theta^d K_i\right|^m\right] = O(1)$, and that $E\left[|\mathfrak{S}_i|^m \left|h^{1+d}\nabla_\theta^d K_i\right|^m\right] = O(1)$ for $m > 4$, where $K_i \equiv \frac{1}{h} k\left(\frac{v(\theta)-V_i(\theta)}{h}\right)$. Define:

$$c_u \equiv \begin{cases} \min\left[2r, \frac{1-2(1+d)r}{2}\right] & \text{if } |\mathfrak{S}_i| = O(1) \text{ and } d = 0 \\ \min\left[2r, \frac{1-2(1+d)r}{2} - \frac{1}{m+2}\right] & Otherwise \end{cases}$$

Then, with the sup taken over x and θ in compact sets:

$$\begin{aligned}
&d): \sup\left|\nabla_\theta^d [\mathfrak{S}(v(\theta);\theta) - E(\mathfrak{S}_i | V_i(\theta) = v(\theta)) g_V(v(\theta))]\right| = o_p(N^{-c_u}) \\
&e): \sup\left|\nabla_\theta^d [\hat{g}_V(v) - g_V(v)]\right| = o_p(N^{-c_u}) \\
&f): \sup\left|\nabla_\theta^d \left[\bar{S}(v) - E(\mathfrak{S}|V = v)\right]\right| = o_p(N^{-c_u}).
\end{aligned}$$

Proof. The proofs of a)-c) follow from Lemma 1**.** To prove d), we write the expression in d) as:

$$\begin{aligned}
&\tfrac{1}{N}\textstyle\sum_i \mathfrak{S}_i \nabla_\theta^d K_i - E\left[\mathfrak{S}_i \nabla_\theta^d K_i\right] + \\
&E\left[\mathfrak{S}_i \nabla_\theta^d K_i\right] - \nabla_\theta^d [E(\mathfrak{S}_i | V_i(\theta) = v(\theta)) g_V(v(\theta))].
\end{aligned} \tag{49}$$

It can be shown that the second line in (49) is $O(h^2)$ from a). For the first line in (49)**,** consider first the case where $|\mathfrak{S}_i| = O(1)$ and $d = 0$. Then, since $hK_i = O(1)$ and $h = O(N^{-r})$, it follows from standard results in the literature (see, for example, Bhattacharya(1967) and Klein(1993)) that:

$$\sup_{x,\theta} |\frac{1}{N}\sum_i \mathfrak{S}_i \nabla_\theta^d K_i - E\left[\mathfrak{S}_i \nabla_\theta^d K_i\right]| = O_p\left(N^{-\left(\frac{1-2r}{2}\right)}\right).$$

The proof for d) then follows for this case. Next, consider the case where $d > 0$ and/or $|\mathfrak{S}_i| \neq O(1)$. Define**:**

$$b_i = \begin{cases} 1: & |\mathfrak{S}_i| \left| h^{1+d} \nabla_\theta^d K_i \right| > N^{\frac{1}{m+2}} \\ 0: & Otherwise \end{cases}.$$

Then, write the first line in (49) as the sum of $T_{(1-b)}$ and T_b, where:

$$T_{(1-b)} \equiv \frac{1}{N} \sum_i (1-b_i) \mathfrak{S}_i \nabla_\theta^d K_i - E\left[(1-b_i) \mathfrak{S}_i \nabla_\theta^d K_i\right] \tag{50}$$

$$T_b \equiv \frac{1}{N} \sum_i b_i \mathfrak{S}_i \nabla_\theta^d K_i - E\left[b_i \mathfrak{S}_i \nabla_\theta^d K_i\right]. \tag{51}$$

It then follows (see, for example, Bhattacharya(1967) and Klein(1993)) that uniformly

$$T_{(1-b)} = O_p\left(N^{-(1/2-(1+d)r-\frac{1}{m+2})}\right).$$

Turning to T_b in (51), we separately order each of its two components. For the first, the absolute value of its expectation is bounded above by:

$$\begin{aligned} E\left[b_i |\mathfrak{S}_i \nabla_\theta^d K_i|\right] &\leq E[b_i]^{1/2} E\left[\mathfrak{S}_i^2 \left(\nabla_\theta^d K_i\right)^2\right]^{1/2} \\ &\leq E\left[\mathfrak{S}_i^4\right]^{1/4} E[b_i]^{1/2} E\left[\left(\nabla_\theta^d K_i\right)^4\right]^{1/4}. \end{aligned} \tag{52}$$

By assumption, $E\left[\mathfrak{S}_i^4\right] = O(1)$. For $E[b_i]$:

$$\begin{aligned} E[b_i] &= \Pr\left[|\mathfrak{S}_i| \left|h^{1+d} \nabla_\theta^d K_i\right| > N^{\frac{1}{m+2}}\right] = \Pr\left[|\mathfrak{S}_i|^m \left|h^{1+d} \nabla_\theta^d K_i\right|^m > N^{\frac{m}{m+2}}\right] \\ &\leq N^{-\frac{m}{m+2}} E\left[|\mathfrak{S}_i|^m \left|h^{1+d} \nabla_\theta^d K_i\right|^m\right] = O\left(N^{-\frac{m}{m+2}}\right) \end{aligned} \tag{53}$$

from Markov's inequality and the Lemma's assumptions. Therefore $E[b_i]^{1/2} = O(N^{-\frac{m}{2(m+2)}})$. Since $E[\left(\nabla_\theta^d K_i\right)^4]^{1/4} = O\left(h^{-(1+d)}\right) = O(N^{r(1+d)})$, it follows that:

$$\frac{1}{N} \sum b_i \mathfrak{S}_i \nabla_\theta^d K_i = O_p(N^{-\frac{m}{2(m+2)}+(1+d)r}).$$

Employing the same arguments, $E\left(\left[b_i |\mathfrak{S}_i \nabla_\theta^d K_i(v(\theta), \theta)|\right]\right) = O(N^{-\frac{m}{2(m+2)}+(1+d)r})$. Hence:

$$\frac{1}{N} \sum b_i \mathfrak{S}_i \nabla_\theta^d K_i - E\left(\left[b_i |\mathfrak{S}_i \nabla_\theta^d K_i(v(\theta), \theta)|\right]\right) = O_p\left(N^{-\frac{m}{2(m+2)}+(1+d)r}\right)$$

and the result then follows. The proofs for e) and f) are very similar to that for d) above.

The estimator for quantile marginal effects will depend on a trimming indicator over estimated quantiles. Lemma 3 below provides the results needed to take such quantile uncertainty into account in deriving the asymptotic distribution of the estimator.

Lemma 3. Quantile Estimation Uncertainty. Assuming that W_j is i.i.d. or constant and that $m \equiv E\left[t_{qj}\left(q, X_{jC}\right) W_j\right]$ is bounded and continuously differentiable in q. Then

$$D \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left[t_{qj}\left(\hat{q}, X_{jC}\right) - t_{qj}\left(q, X_{jC}\right)\right] W_j = \left[\nabla_q m\right] \sqrt{N} \bar{\mathfrak{B}} + o_p(1), \tag{54}$$

where t_{qj} is defined in (D11) and $\bar{\mathfrak{B}}$ in (D13).

Proof.

$$\begin{aligned} D &= \left\{ \begin{array}{l} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left[t_{qj}(\hat{q}, X_{jC}) W_j - m(\hat{q})\right] - \\ \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left[t_{qj}(q, X_{jC}) W_j - m(q)\right] \end{array} \right\} \qquad (55) \\ &\quad + \sqrt{N}\left[m(\hat{q}) - m(q)\right]. \qquad (56) \end{aligned}$$

With $\tilde{q}$ an arbitrary value for q, for the term in (55) it suffices to show that:

$$\sup_{\|\tilde{q}-q\|<\delta_n} \left| \begin{array}{l} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left[t_{qj}(\tilde{q}, X_{jC}) W_j - m(\tilde{q})\right] - \\ \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left[t_{qj}(q, X_{jC}) W_j - m(q)\right] \end{array} \right| = o_p(1)$$

for each sequence δ_n tending to zero. From Pakes and Pollard (1989, Lemma 1.17) the above result holds, implying that the term in (55) vanishes in probability. Turning to the term in (56), from a Taylor series expansion:

$$\sqrt{N}[m(\hat{q}) - m(q)] = \nabla_q m\left(q\right) \sqrt{N}(\hat{q} - q).$$

From the Bahadur representation in (D13) :

$$\sqrt{N}\left(\hat{q} - q\right) = \sqrt{N} \bar{\mathfrak{B}} + o_p(1).$$

The lemma follows.

Lemma 4. $\boldsymbol{\alpha - derivative}$. Recalling Theorem 1) and define:

$$\begin{aligned} \bar{\alpha}(V_j\left(\theta\right), \theta) &\equiv E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta)\right]^{-1} E[\bar{Q}_j^{*\prime} Y_j | V_j(\theta)] \\ \alpha_0(V_j(\theta)) &\equiv \bar{\alpha}(V_j\left(\theta\right), \theta_0). \end{aligned}$$

Then, with ∇_θ as the derivative operator w.r.t. θ :

$$\begin{aligned} \nabla_\theta \bar{\alpha}(V_j(\theta), \theta)_{\theta=\theta_0} = \nabla_\theta \alpha_0(V_j(\theta))_{\theta=\theta_0} \\ -E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta_0)\right]^{-1} E[\bar{Q}_j^{*\prime} R_j [\nabla_\theta \alpha_0(V_j(\theta))] | V_j(\theta_0)]_{\theta=\theta_0}. \end{aligned}$$

Proof. Differentiating $\bar{\alpha}(V_j\left(\theta\right), \theta)$, with respect to θ_k, the k^{th} element of θ. We have:

$$\begin{aligned} \nabla_{\theta_k} \bar{\alpha}(V_j\left(\theta\right), \theta)_{\theta=\theta_0} &= \left\{\nabla_{\theta_k} E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta)\right]^{-1}\right\}_{\theta=\theta_0} E[\bar{Q}_j^{*\prime} Y_j | V_j(\theta_0)] \qquad (57) \\ &\quad + E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta_0)\right]^{-1} \nabla_{\theta_k} E[\bar{Q}_j^{*\prime} Y_j | V_j(\theta)]_{\theta=\theta_0}. \qquad (58) \end{aligned}$$

We first note that for any nonsingular matrix $M(\theta)$ and for θ_k an element of θ :

$$\nabla_{\theta_k} M^{-1}(\theta) = -M^{-1}(\theta)[\nabla_{\theta_k} M(\theta)]M^{-1}(\theta), \tag{59}$$

a result that follows because

$$\nabla_{\theta_k}\left[M^{-1}(\theta)M(\theta)\right] = \nabla_{\theta_k}(I) = 0 = \nabla_{\theta_k}\left[M^{-1}(\theta)\right]M(\theta) + M^{-1}(\theta)\nabla_{\theta_k}\left[M^{-1}(\theta)\right].$$

Employing (59), the first component in (57) is given as:

$$\nabla_{\theta_k} E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta)\right]^{-1}_{\theta=\theta_0} = -E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta_0)\right]^{-1}\left\{\nabla_{\theta_k} E[\bar{Q}_j^{*\prime} R_j | V_j(\theta)]\right\}_{\theta=\theta_0} E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta_0)\right]^{-1}.$$

Recalling that $Y_j = R_j\alpha_0(V_j(\theta_0)) + \varepsilon_j$, the second component in (57) is given as:

$$E[\bar{Q}_j^{*\prime} Y_j | V_j(\theta_0)] = E[\bar{Q}_j^{*\prime} R_j \alpha_0(V_j(\theta_0)) | V_j(\theta_0)] = E[\bar{Q}_j^{*\prime} R_j | V_j(\theta_0)]\alpha_0(V_j(\theta_0)).$$

Therefore, the expression in (57) has the characterization:

$$-E\left[\bar{Q}_j^{*\prime} R_j | V_j(\theta_0)\right]^{-1}\left\{\nabla_{\theta_k} E[\bar{Q}_j^{*\prime} R_j | V_j(\theta)]\right\}_{\theta=\theta_0}\alpha_0\left(V_j(\theta_0)\right). \tag{60}$$

Turning to the gradient component in (58), let $\delta(\theta) \equiv V_j(\theta_0) - V_j(\theta)$. Then we may write :

$$\begin{aligned}
\nabla_{\theta_k} E[\bar{Q}_j^{*\prime} Y_j | V_j(\theta)]_{\theta=\theta_0} &= \nabla_{\theta_k} E\left[\bar{Q}_j^{*\prime} R_j[\alpha_0(V_j(\theta_0))] | V_j(\theta))\right]_{\theta=\theta_0} \\
&= \nabla_{\theta_k} E\left[\bar{Q}_j^{*\prime} R_j[\alpha_0(\delta(\theta) + V_j(\theta))] | V_j(\theta))\right]_{\theta=\theta_0} \\
&= G_1 + G_2, \\
G_1 &\equiv \nabla_{\theta_k} E\left[\bar{Q}_j^{*\prime} R_j[\alpha_0(\delta(\theta) + V_j(\theta_0))] | V_j(\theta_0)\right]_{\theta=\theta_0} \\
G_2 &\equiv \nabla_{\theta_k} E\left[\bar{Q}_j^{*\prime} R_j[\alpha_0(\delta(\theta_0) + V_j(\theta))] | V_j(\theta)\right]_{\theta=\theta_0}
\end{aligned}$$

from the chain rule.[7] For the G_1-term, since $\nabla_{\theta_k}\delta(\theta_0) = -\nabla_{\theta_k} V_j(\theta)$:

$$\begin{aligned}
G_1 &= E\left[\bar{Q}_j^{*\prime} R_j[\nabla_{\theta_k}\alpha_0(\delta(\theta) + V_j(\theta_0))] | V_j\left(\theta_0\right)\right]_{\theta=\theta_0} \\
&= -E[\bar{Q}_j^{*\prime} R_j[\nabla_{\theta_k}\alpha_0(V_j(\theta))] | V_j(\theta_0)]_{\theta=\theta_0}.
\end{aligned}$$

For the G_2-term above, since $\delta(\theta_0) = 0$:

$$\begin{aligned}
G_2 &= \nabla_{\theta_k}\left\{E[\bar{Q}_j^{*\prime} R_j \alpha_0(V_j(\theta)) | V_j(\theta)]\right\}_{\theta=\theta_0} \\
&= \nabla_{\theta_k}\left\{E[\bar{Q}_j^{*\prime} R_j | V_j(\theta)]\alpha_0(V_j(\theta))\right\}_{\theta=\theta_0} \\
&= E[\bar{Q}_j^{*\prime} R_j | V_j(\theta_0)]\nabla_{\theta_k}\alpha_0(V_j(\theta))_{\theta=\theta_0} + \\
&\quad \left\{\nabla_{\theta_k} E[\bar{Q}_j^{*\prime} R_j | V_j(\theta)]_{\theta=\theta_0}\right\}\alpha_0(V_j(\theta_0)).
\end{aligned}$$

[7] To the best of our knowledge, Whitney Newey was the first to make this type of argument for semiparametric expectations. It underlies the efficiency result in Klein and Spady(1993) as well as the bias reduction mechanism in Klein and Shen (2011) and Klein, Shen, and Vella (2014).

Combining the G_1 and G_2 terms, the term in (58) can be characterized as:

$$\left[E[\bar{Q}_j^{*\prime}R_j|V_j(\theta_0)\right]^{-1}\left\{\begin{array}{c} -E\left\{\bar{Q}_j^{*\prime}R_j\nabla_{\theta_k}\alpha_0(V_j(\theta))|V_j(\theta_0)\right]\right\}_{\theta=\theta_0}+ \\ E[\bar{Q}_j^{*\prime}R_j|V_j(\theta_0)]\nabla_{\theta_k}\alpha_0(V_j(\theta))_{\theta=\theta_0}+ \\ \left\{\nabla_{\theta_k}E[\bar{Q}_j^{*\prime}R_j|V_j(\theta)]_{\theta=\theta_0}\right\}\alpha_0(V_j(\theta_0))\ . \end{array}\right\}. \tag{61}$$

Finally, adding (60) and (61) together, the form for $\nabla_{\theta_k}\alpha(V_j(\theta);\theta)_{\theta=\theta_0}$ follows.

Lemma 5. Estimated trimming. Let W_i be an i.i.d. vector with estimator $\hat{W}_i$. With $\hat{q}$ and q as vectors of sample and population quantiles, let $t_i(q)$ be a trimming function of the form shown in (D3). Let $\mathfrak{N}_\delta = \{\tilde{q} : \|\tilde{q}-q\| < \delta\}$ where $\delta \to 0$ as $N \to \infty$. Denote $\|A\| \equiv AA'$. With $c_w > 0$, $c_a \geqslant 0$, $c_s > 0$, assume that A_i and W_i are vectors satisfying :

$$\begin{aligned} C1) \ &: \ \frac{1}{N}\sum_{i=1}^{N} t_i(q) Tr\left\|\hat{W}_i - W_i\right\| = o_p(N^{-2c_w}) \\ C2) \ &: \ \frac{1}{N}\sum_{i=1}^{N} t_i(q) Tr\left\|A_i\right\| = O_p(N^{-2c_a}) \\ C3) \ &: \ \sup_{i;\,\tilde{q}\epsilon\mathfrak{N}_\delta} t_i\left(\tilde{q}\right)A_i = o_p\left(N^{-c_s}\right). \\ C4) \ &: \ E\left[\frac{1}{N}\sum_{j=1}^{N} Tr\left\|W_iA_i\right\|\right] = O(1)\ . \end{aligned}$$

Then, for $\varepsilon > 0$ and arbitrarily small and for the indicators employed here, under (C1), (C3) and (C4) :

$$a) : \frac{1}{N}\sum_{i=1}^{N}[t_i(\hat{q})\hat{W}_i - t_i\left(q\right)W_i]A_i = \frac{1}{N}\sum_{i=1}^{N} t_i\left(q\right)(\hat{W}_i - W_i)A_i + o_p(N^{-1/2+\varepsilon})\ .$$

Under (C1)-(C4):

$$b) : \frac{1}{N}\sum_{i=1}^{N}[t_i(\hat{q})\hat{W}_i - t_i\left(q\right)W_i]A_i = o_p(N^{-(c_w+c_a)})\ .$$

$$c) : \frac{1}{N}\sum_{i=1}^{N}[t_i(\hat{q}) - t_i\left(q\right)] = o_p(N^{-1/2+\varepsilon}).$$

Proof. For a), we have:

$$\frac{1}{N}\sum_{i=1}^{N}[t_i(\hat{q})\hat{W}_i - t_i\left(q\right)W_i]A_i = \frac{1}{N}\sum_{i=1}^{N} t_i(q)\left(\hat{W}_i - W_i\right)A_i + \Delta_{wa}$$

$$\begin{aligned} \Delta_{wa} \ &= \ \frac{1}{N}\sum_{i=1}^{N}[t_i(\hat{q}) - t_i\left(q\right)]W_iA_i + & (62) \\ &\quad \frac{1}{N}\sum_{i=1}^{N}[t_i(\hat{q}) - t_i\left(q\right)]s_i(\hat{q})\left(\hat{W}_i - W_i\right)A_i, & (63) \end{aligned}$$

where s_i is an indicator that is one if $t_i(\hat{q})$ or $t_i(q)$ is 1 and is zero otherwise. For the component, of Δ_{wa} in (62), from Cauchy-Schwarz and (C4), the absolute value is bounded above by:[8]

$$\left\{ \frac{1}{N} \sum_{j=1}^{N} [t_i(\hat{q}) - t_i(q)]^2 \right\}^{1/2} \times O_p(1).$$

Employing an argument in Klein and Shen(2010, Theorem 2) and a convergence rate for indicators in Klein(1993),[9] the above term is $o_p(N^{-1/2+\varepsilon})$ for $\varepsilon > 0$ and arbitrarily small. For the absolute value of the second component of Δ_{wa} in (63), from an argument similar to that above, it is bounded by:

$$|\Delta_{wa}| \leq \left\{ \frac{1}{N} \sum_{j=1}^{N} [t_i(\hat{q}) - t_i(q)]^2 \right\}^{1/2} \sup_{\tilde{q} \epsilon \Re_\delta} \left\{ \frac{1}{N} \sum_{j=1}^{N} s_j(\tilde{q}) Tr \left\| \left(\hat{W}_i - W_i \right) A_i \right\| \right\}^{1/2}.$$

As above, the first term is $o_p(N^{-1/2+\varepsilon})$ for $\varepsilon > 0$ and arbitrarily small. The second term converges in probability to 0 from the uniform assumption (C3) on A_i and the pointwise assumption involving $\hat{W}_i$.

The proof for b) readily follows from a), an extension of Cauchy Schwarz and the point-wise convergence assumptions. The proof for c) employs the same argument as that in a), making similar use of the s_j indicator .

Lemma 6. Let $v \epsilon \mathfrak{C}$, a compact subset of the support for V and $\theta \epsilon \Phi_1$ a compact set. Refer to (D6) and (D8). Set the window parameter as in (D4) and take the sup over v and θ. Then, under (A8):

$$a) \bar{\Omega}^*(v,\theta), \Omega(v,\theta) \text{ are non-singular for all } v \epsilon \mathfrak{C}, \theta \epsilon \Phi_1 \tag{64}$$

$$b) \sup \begin{bmatrix} |\hat{\Omega} - \Omega| \\ |\hat{\Omega}^* - \bar{\Omega}^*| \\ |\hat{\Omega}^* - \Omega^*| \end{bmatrix} = \begin{bmatrix} o_p\left(N^{-1/4}\right) \\ o_p\left(N^{-\delta}\right) \\ o_p\left(N^{-1/4}\right) \end{bmatrix} \tag{65}$$

$$c) \sup \begin{bmatrix} |\hat{\Omega}^{-1} - \Omega^{-1}| \\ |\hat{\Omega}^{*-1} - \bar{\Omega}^{*-1}| \\ |\hat{\Omega}^{*-1} - \Omega^{*-1}| \end{bmatrix} = \begin{bmatrix} o_p\left(N^{-1/4}\right) \\ o_p\left(N^{-\delta}\right) \\ o_p\left(N^{-1/4}\right) \end{bmatrix} \tag{66}$$

Proof. To show that $\bar{\Omega}^*$ in a) is non-singular, from (D6)-(D8), locally project T_i on $\bar{Q}_i^*$ to obtain:

$$T_i = \bar{Q}_i^* (\pi'(v), c'(v)) + e_i, E\left[e_i \bar{Q}_i^* | V_i = v \right] = 0.$$

[8] Letting $B_i \equiv W_i A_i$ have k^{th} element B_{ik}, from Cauchy-Schwarz

$$\sum_{i=1}^{N} [t_i(\hat{q}) - t_i(q)] B_{ik} \leq [\frac{1}{N} \sum_{i=1}^{N} [t_i(\hat{q}) - t_i(q)]^2]^{1/2} \left[\frac{1}{N} \sum_{i=1}^{N} B_{ik}^2 \right]^{1/2},$$

where $\frac{1}{N} \sum_{i=1}^{N} B_{ik} \leq \frac{1}{N} \sum_{i=1}^{N} Tr[B_i B_i']$.

[9] The rate is obtained by bounding $[t_i(\hat{q}) - t_i(q)]^2$ from above by a smooth function and then employing a Taylor series expansion.

It then follows that:

$$E[\bar{Q}_i^{*\prime} R_i | V_i = v] = E(\bar{Q}_i^{*\prime} \bar{Q}_i^{*} | V_i = v)] \begin{bmatrix} \pi & 0 \\ c' & 1 \end{bmatrix},$$

If $\Sigma \equiv var(t_i P_i' | V_i(\theta) = v)$ is positive definite for all v and θ in a compact sets, then $E(\bar{Q}_i^{*\prime} \bar{Q}_i^{*} | V_i = v)$ will have full rank. It then follows that if π has full rank and $g_V(v)$ is non-zero,

$$\bar{\Omega}^*(v, \theta) \equiv E(\bar{Q}_i^{*\prime} R_i | V_i = v) g_V(v)$$

will have full rank, completing the proof for a). The proof for $\Omega(v, \theta)$ is identical.

Next we show b): For the first and third parts, let $\hat{\Omega}_\alpha(v, \theta)$ refer to either $\hat{\Omega}(v, \theta)$ or $\hat{\Omega}^*$, $\Omega_\alpha(v, \theta)$ to either Ω or Ω^*, and Q_α to either Q or Q^*. Then,

$$\left| \hat{\Omega}_\alpha(v, \theta) - \Omega_\alpha(v, \theta) \right| \leq \left| \frac{Q_\alpha' D_N(v, \theta) R}{N} - \Omega_\alpha(v, \theta) \right| + \left| \frac{\left[\hat{Q}_\alpha - Q_\alpha \right]' D_N(v, \theta) R}{N} \right|.$$

The *sup* of the first component on the right-hand-side is $o_p(N^{-1/4})$ from Lemma 2d. From (D5), Lemma 5, and Lemma 2d, the second component is also $o_p(N^{-1/4})$. The second part of b) follows from the convergence of t_1^* to t_1 in (D6) and the definition of $\bar{\Omega}^*$ in (D8).

For the first part of c), let

$$(v^*, \theta^*) \equiv \arg\sup_{v, \theta} \left\| \hat{\Omega}(v, \theta)^{-1} - \Omega(v, \theta)^{-1} \right\|.$$

Notice that from a)-b):

$$\begin{aligned} \det\left(\hat{\Omega}(v^*, \theta^*) \right) - \det\left(\Omega(v^*, \theta^*) \right) &= o_p(1), \\ \det\left(\Omega(v^*, \theta^*) \right) &\geqslant \inf_{v, \theta} \det\left(\Omega(v, \theta) \right) > 0. \end{aligned}$$

Therefore, it can be shown that $\hat{\Omega}(v, \theta)^{-1} - \Omega(v, \theta)^{-1}$ uniformly converges to 0. To obtain the required rate, write the expression in c) as:

$$\hat{\Omega}(v^*, \theta^*)^{-1} - \Omega(v^*, \theta^*)^{-1} = \hat{\Omega}(v^*, \theta^*)^{-1} [\Omega(v^*, \theta^*) - \hat{\Omega}(v^*, \theta^*)] \Omega(v^*, \theta^*)^{-1}. \tag{67}$$

With the first and third terms being $O_p(1)$, the result follows from the convergence rate for the middle term established in b). The arguments for the other two parts of c) are similar with the slower convergence rate in the second part being due to the slower convergence rate of $||\bar{\Omega}^* - \bar{\Omega}^*|$ in b).

Lemma 7. Recalling (D1) and (D8), define $V_i \equiv V_i(\theta_0)$. Let α_0, $\hat{\alpha}_0^*$, and $\hat{\alpha}_0$ be matrices with i^{th} rows given as $\alpha_{0i} \equiv \alpha(V_i, \theta_0)$, $\hat{\alpha}_{0i}^* \equiv \hat{\alpha}_0^*(V_i, \theta_0)$, and $\hat{\alpha}_{0i} \equiv \hat{\alpha}(V_i, \theta_0)$ respectively. With $l = 1, ..., L+1$, let $\Delta_l(V_i)$ be the l^{th} element from $\Delta(V_i)$ corresponding to L treatments and a default option, and $\hat{\Delta}_l(V_i)$ the l^{th} element from $\hat{\Delta}(V_i, \theta_0)$ in (D8). Assume that t_i is an indicator

constraining V_i to a compact set. Then, recalling that $\|A\| \equiv AA'$:

$$a)\frac{1}{N}\sum_{i=1}^{N} t_i Tr\,\|\hat{\alpha}_{0i} - \alpha_{0i}\| = o_p(N^{-1/2}) \tag{68}$$

$$b)\frac{1}{N}\sum_{i=1}^{N} t_i Tr\,\left\|\left[\Delta(V_i) - \hat{\Delta}(V_i)\right] D_N(V_i\,(\theta_0)\,,\theta_0)\right\| /(N-1) = o_p\left(N^{-1/2}\right) \tag{69}$$

$$c)\frac{1}{N}\sum_{i=1}^{N} t_i Tr\,\|\hat{\alpha}^*_{0i} - \alpha_{0i}\| = o_p(N^{-1/2}). \tag{70}$$

Proof. For a), let $D_N\,(V_i) \equiv D_N(V_i(\theta_0),\,\theta_{\,0})$ and define:

$$\tilde{\alpha}_{0i} \equiv \frac{1}{N}\Omega\,(V_i)^{-1}\,Q' D_N\,(V_i)\,Y = \alpha_{0i} + \frac{1}{N}\Omega\,(V_i)^{-1}\,Q' D_N\,(\varepsilon + \Delta\,(V_i))\,. \tag{71}$$

We prove a) by showing:

$$a1) : \Delta_1 \equiv \frac{1}{N}\sum_{i-1}^{N} t_{\,i}\,Tr\,\|\tilde{\alpha}_{0i} - \alpha_{0i}\| = o_p\left(N^{-1/2}\right) \tag{72}$$

$$a2) : \Delta_2 \equiv \frac{1}{N}\sum_{i=1}^{N} t_i Tr\|\hat{\alpha}_{0i} - \alpha_{0i}\| - \frac{1}{N}\sum_{i=1}^{N} t_i Tr\|\tilde{\alpha}_{0i} - \alpha_{0i}\|| = o_p(N^{-1/2}) \tag{73}$$

$$\tag{74}$$

For a1),

$$E\,(\Delta_1) = E[t_{\,i}\,Tr\|\tilde{\alpha}_{0i} - \alpha_{0i}\|] = E[t_{\,i}\,Tr\,\|\tilde{\alpha}_{0i} - E\,(\tilde{\alpha}_{0i})\,\|] + E\,[t_{\,i}\,Tr\,\|E\,(\tilde{\alpha}_{0i}) - \alpha_{0i}\|] \tag{75}$$

From Lemma 2, it can be shown that last (squared bias) component is uniformly $O(h^4)$ while the remaining (variance) component is uniformly $O\big(\frac{1}{Nh}\big)$. With $h = 1/7.99$, it follows that a1) holds. Turning to a2):

$$\Delta_2 = \frac{1}{N}\sum_{i=1}^{N} t_i Tr\,\|(\hat{\alpha}_{0i} - \tilde{\alpha}_{0i}) + (\tilde{\alpha}_{0i} - \alpha_{0i})\,\| - \frac{1}{N}\sum_{i=1}^{N} t_i Tr\|\tilde{\alpha}_{0i} - \alpha_{0i}\| \tag{76}$$

and hence

$$\Delta_2 = \frac{1}{N}\sum_{i=1}^{N} t_i\,[\hat{\alpha}_{0i} - \tilde{\alpha}_{0i}]'\,[\hat{\alpha}_{0i} - \tilde{\alpha}_{0i}] + 2\frac{1}{N}\sum_{i=1}^{N} t_i\,[\hat{\alpha}_{0i} - \tilde{\alpha}_{0i}]'\,[\tilde{\alpha}_{0i} - \alpha_{0i}]. \tag{77}$$

For the first component in (77), recalling the definition of $\hat{\alpha}$ in (D8):

$$t_i(\hat{\alpha}_{0i} - \tilde{\alpha}_{0i}) = t_i \left\{ \begin{array}{c} \Omega\,(V_i)^{-1}\left(\hat{Q} - Q\right)' + \left[\hat{\Omega}\,(V_i)^{-1} - \Omega\,(V_i)^{-1}\right] \\ + \left(\hat{\Omega}\,(V_i)^{-1} - \Omega\,(V_i)^{-1}\right)\left(\hat{Q} - Q\right)' \end{array} \right\} A_N, \tag{78}$$

where $A_N \equiv D_N(V_i)[\epsilon + \Delta(V_i)]/N$. From (D5) and Lemmas 5-6, $t_i\Omega(V_i)^{-1}\left(\hat{Q} - Q\right)$ is uniformly $o_p\left(N^{-1/4}\right)$ in i and from Lemma 6, $t_i(\hat{\Omega}(V_i)^{-1} - \Omega(V_i)^{-1})$ is also uniformly $o_p\left(N^{-1/4}\right)$. Accordingly, it can be shown that

$$\sup_i |t_i(\hat{\alpha}_{0i} - \tilde{\alpha}_{0i})| = o_p\left(N^{-1/4}\right). \tag{79}$$

It now follows that the first term of Δ_2 in (77) is $o_p\left(N^{-1/2}\right)$. From Cauchy-Schwarz, the absolute value of the second term of Δ_2 is bounded from above by:

$$2\sqrt{\frac{1}{N}\sum_{i=1}^{N} t_i\left[\hat{\alpha}_{0i} - \tilde{\alpha}'_{0i}\right]'\left[\hat{\alpha}_{0i} - \tilde{\alpha}_{0i}\right]}\sqrt{\frac{1}{N}\sum_{i=1}^{N} t_i[\tilde{\alpha}_{0i} - \alpha_{0i}]'\left[\tilde{\alpha}_{0i} - \alpha_{0i}\right]} \tag{80}$$

The first component of the above product is $o_p\left(N^{-1/4}\right)$ from (79) above. The second term in the above product is $o_p\left(N^{-1/4}\right)$ from a1). The result in a) now follows.

To prove b), recall from (D1) and (D8) that:

$$\Delta_j(V_i) = R_j\left[\alpha_{0j} - \alpha_{0i}\right]; \hat{\Delta}_j(V_i) = R_j[\hat{\alpha}_{0j} - \hat{\alpha}_{0i}] \implies \tag{81}$$

$$\Delta_j(V_i) - \hat{\Delta}_j(V_i) = R_j\left\{[\hat{\alpha}_{0i} - \alpha_{0i}] - [\hat{\alpha}_{0j} - \alpha_{0j}]\right\}. \tag{82}$$

Employing the definition of $\tilde{\alpha}_{0i}$ in the proof of a), define

$$\tilde{\Delta}_j(V_i) \equiv R_j\left[\tilde{\alpha}_{0i} - \alpha_{0i}\right] \implies \Delta_j(V_i) - \tilde{\Delta}_j(V_i) = R_j\left\{[\tilde{\alpha}_{0i} - \alpha_{0i})] - [\tilde{\alpha}_{0j} - \alpha_{0j}]\right\}. \tag{83}$$

Under an argument similar to that in a),

$$\frac{1}{N}\sum_{i=1}^{N} t_i\{Tr\|\left[\Delta(V_i) - \hat{\Delta}(V_i)\right]D_N(V_i)\|/(N-1) - Tr\left\|\left[\Delta(V_i) - \tilde{\Delta}(V_i)\right]D_N(V_i)\right\|/(N-1)\} \tag{84}$$

is $o_p\left(N^{-1/2}\right)$. Therefore, to prove (b), we study:

$$E\{t_i\frac{1}{h^2}\left[\Delta(V_i) - \tilde{\Delta}(V_i)\right]' D_N^2(V_i)\left[\Delta(V_i) - \tilde{\Delta}(V_i)\right]/(N-1)\} \tag{85}$$

Write this term as the sum of:

$$T_1 \equiv \frac{1}{N-1}\sum_{r\neq i}^{N} E\left\{t_i[\tilde{\Delta}_r(V_i) - E\left(\tilde{\Delta}_r(V_i)\ |V_i, V_r\right)]^2 K_{ir}^2\right\} \tag{86}$$

$$T_2 \equiv \frac{1}{N-1}\sum_{r\neq i}^{N} E\left\{t_i\left[E(\tilde{\Delta}_r(V_i)\ |V_i, V_r) - \Delta_r(V_i)\right]^2 K_{ir}^2\right\}. \tag{87}$$

where $K_{ir} \equiv K[(V_i - V_r)/h]$. The$T_1$ term is $o_p(N^{-1/2})$ because K^2 is bounded and the variance term:

$$\frac{1}{h^2}E\left\{t_i\left[\tilde{\Delta}_r(V_i) - E\left(\tilde{\Delta}_r(V_i)|V_i\right)\right]'\left[\bar{\Delta}_r(V_i) - E\left(\tilde{\Delta}_r(V_i)|V_i\right)\right]\right\}$$

has order $\frac{1}{h^2}O(\left(\frac{1}{Nh}\right)) = o_p(N^{-1/2})$ uniformly in r for $h = O(N^{-1/7.99})$. Turning to T_2, it can be shown that

$$E[E(\tilde{\Delta}_r(V_i)|V_i, V_r) - \Delta_r(V_i)|V_i, V_r] = h^2[B(V_i) - B(V_r)],$$

where B is a bounded function. Therefore, we need to evaluate:

$$h^3 E\{t_i E[B(V_i) - B(V_r)]' \left[K^2\left[\frac{V_i - V_r}{h}\right]/h\right][B(V_i) - B(V_r)]|V_i]\}\ .$$

The order of this term is further reduced with the kernel restricting V_r to be close to V_i . In particular, letting $Z = (V_r - V_i)/h$, it can be shown that the inner expectation is uniformly $O(h^4)$. Therefore, with $h = N^{-1/7.99}$, the above term is $o(N^{-1/2})$ which establishes part b) of the lemma.

To establish c), refer to (D8), recall the definition of $\tilde{\Delta}(V_i)$ above, and define:

$$\begin{aligned}\tilde{\alpha}^*_{0i} &\equiv \frac{\Omega_i^{*-1} Q^{*\prime} D_N(V_i)[Y - \tilde{\Delta}(V_i)]}{N} \\ &= \alpha_{0i} + \Omega_i^{*-1} Q^{*\prime} D_N(V_i)[\varepsilon + \Delta(V_i) - \tilde{\Delta}(V_i)]\end{aligned} \tag{88}$$

Then, similar to a), the proof follows by showing

$$c1): \Delta_1^* \equiv \frac{1}{N}\sum_{i-1}^{N} t_i\, Tr\, \|\tilde{\alpha}^*_{0i} - \alpha_{0i}\| = o_p\left(N^{-1/2}\right) \tag{89}$$

$$c2): \Delta_2^* \equiv \frac{1}{N}\sum_{i=1}^{N} t_i Tr\, \|\hat{\alpha}^*_{0i} - \alpha_{0i}\| - \frac{1}{N}\sum_{i=1}^{N} t_i Tr\|\tilde{\alpha}^*_{0i} - \alpha_{0i}\|| = o_p(N^{-1/2}). \tag{90}$$

The proof for c1) is very similar to that for a1). Employing b), the proof for c2) is very similar to that for a2).

Lemma 8 provides convergence rates on derivatives, which are needed to establish asymptotic normality.

Lemma 8. Define $V_i(\theta) \equiv X_{1i} + X_{IIi}\theta$, where X_{1i}, X_{IIi}, and θ are in compact sets.[10] Let $\nabla^d_{\theta_k} f(\theta)$ be the d^{th} partial derivative of the function f with respect to θ_k, the k^{th} element of θ where $\nabla^0_{\theta_k} f(\theta) \equiv f(\theta)$. Let δ_d be a small positive number. Then, referring to (D6) and (D8):

$$\begin{aligned} a) &: \sup_{i,\theta}\left|\bigtriangledown_{\theta_k}\left[\hat{\Omega}(V_i(\theta),\theta)^{-1} - \Omega(V_i(\theta),\theta)^{-1}\right]\right| = o_p(N^{-\delta_d}) \\ b) &: \sup_{i,\theta}|\bigtriangledown_{\theta_k}\left[\hat{\Omega}^*(V_i(\theta),\theta)^{-1} - \Omega^*(V_i(\theta),\theta)^{-1}\right]| = o_p(N^{-\delta_d}) \end{aligned}$$

For $d = 0, 1$

$$\begin{aligned} c) &: \sup_{i,\theta}\left|\nabla^d_\theta[\hat{\alpha}[V_i(\theta),\theta] - \alpha(V_i(\theta),\theta)]\right| = o_p(N^{-\delta_d}) \\ d) &: \sup_{i,\theta}\left|\nabla^d_\theta[\hat{\alpha}^*[V_i(\theta),\theta] - \bar{\alpha}(V_i(\theta),\theta)]\right| = o_p(N^{-\delta_d}) \end{aligned}$$

[10] In applications of this lemma, trimming ensures that the index variables are in compact sets.

Proof. For a), employ the derivative of a matrix inverse in (59) and write the difference in derivatives in a) as:

$$\Omega(V_i(\theta),\theta)^{-1}\left[\nabla_{\theta_k}\Omega(V_i(\theta),\theta)\right]\Omega(V_i(\theta),\theta)^{-1}-\hat{\Omega}(V_i(\theta),\theta)^{-1}\left[\nabla_{\theta_k}\hat{\Omega}(V_i(\theta),\theta)\right]\hat{\Omega}(V_i(\theta),\theta)^{-1}$$

From Theorem 1 and Lemma 6,

$$\left|\hat{\Omega}(V_i(\theta),\theta)^{-1}-\Omega(V_i(\theta),\theta)^{-1}\right|$$

is uniformly $o_p(N^{-\delta_d})$. Therefore, it suffices to show that

$$\left|\nabla_{\theta_k}\hat{\Omega}(V_i(\theta),\theta)-\nabla_{\theta_k}\Omega(V_i(\theta),\theta)\right|$$

is uniformly $o_p(N^{-\delta_d})$. An upper bound on this expression is given as:

$$\sup_{i,\theta}\left|\nabla_{\theta_k}\frac{Q'D_N(V_i(\theta),\theta)R}{N}-\nabla_{\theta_k}\Omega(V_i(\theta),\theta)\right| \tag{91}$$

$$+\sup_{i,\theta}\left|\nabla_{\theta_k}\left[\frac{\left[\hat{Q}-Q\right]'D_N(V_i(\theta),\theta)R}{N}\right]\right|. \tag{92}$$

From Lemma 2a with $r=1/7.99$, the component in (91) is $o_p(N^{-\delta_d})$. Based on Lemma 5 and (D5), it can be shown that the component in (92) is $o_p(N^{-\delta_d})$, which completes the proof of a).

As the proof for b) is identical to that for a), we turn to c). As the argument for d = 0 is similar to but less complicated than that for d = 1, we provide the argument for d = 1. For this case, the difference in estimated and true derivatives is given as:

$$\nabla_{\theta_k}\left\{\hat{\Omega}(V_i(\theta),\theta)^{-1}\frac{\hat{Q}'D_N(V_i(\theta),\theta)Y}{N}-\Omega(V_i(\theta),\theta)^{-1}E\left[Q_i'Y_i|V_i(\theta)\right]g_V(V_i(\theta))\right\}$$

$$=\Omega(V_i(\theta),\theta)^{-1}\nabla_{\theta_k}[\frac{\hat{Q}'D_N(V_i(\theta),\theta)Y}{N}-E[Q_i'Y_i|V_i(\theta)]g_V(V_i(\theta))] \tag{93}$$

$$+\nabla_{\theta_k}\Omega(V_i(\theta),\theta)^{-1}\left[\frac{\hat{Q}'D_N(V_i(\theta),\theta)Y}{N}-E\left[Q_i'Y_i|V_i(\theta)\right]g_V(V_i(\theta))\right] \tag{94}$$

$$+\left[\hat{\Omega}(V_i(\theta),\theta)^{-1}-\Omega(V_i(\theta),\theta)^{-1}\right]\nabla_{\theta_k}\left\{\frac{\hat{Q}'D_N(V_i(\theta),\theta)Y}{N}\right\} \tag{95}$$

$$+\nabla_{\theta_k}\left[\hat{\Omega}(V_i(\theta),\theta)^{-1}-\Omega(V_i(\theta),\theta)^{-1}\right]\frac{\hat{Q}'D_N(V_i(\theta),\theta)Y}{N} \tag{96}$$

For the term (93), $|\Omega(V_i(\theta),\theta)^{-1}|\leq\sup_{v,\theta}|\Omega(v,\theta)^{-1}|$ is bounded, hence we study

$$\nabla_{\theta_k}\left[\frac{\hat{Q}'D_N(V_i(\theta),\theta)Y}{N}-E\left[Q_i'Y_i|V_i(\theta)\right]g_V(V_i(\theta))\right]$$

$$=\nabla_{\theta_k}\left[\frac{Q'D_N(V_i(\theta),\theta^*)Y}{N}-E[Q_i'Y_i|V_i(\theta)]g_V(V_i(\theta))\right] \tag{97}$$

$$+\left[\frac{\left[\hat{Q}-Q\right]'\nabla_{\theta_k}D_N(V_i(\theta),\theta)Y}{N}\right]. \tag{98}$$

The absolute value of the term in (97) is bounded above by:

$$\sup_{i,\theta}\left|\nabla_{\theta_k}\left[\frac{Q'D_N(V_i(\theta),\theta)Y}{N}-E\left[Q_i'Y_i|V_i(\theta)\right]g_V(V_i(\theta))\right]\right|,$$

which is $o_p(N^{-\delta_d})$ from Lemma 2d) and (A6). For the term in (98), the absolute value of a typical term is bounded above by:

$$\sup_i t_{2i}|\hat{P}_{li}-P_{li}|\sup_{j,\theta}\left|\sum_{j=1}^{N}\frac{Y_j\nabla_{\theta_k}K_j(V_i(\theta),\theta)}{N}\right|$$

The first component is $o_p(N^{-1/4})$ from (D5). The second component is $O_p(1)$ under Lemma 2d and (A6). Consequently, the term in (98) is $o_p(N^{-\delta_d})$.

Turning to the term in (94) the absolute value of its second component is $o_p(N^{-\delta_d})$ from an argument virtually identical to that for (93). For the first component:

$$\begin{aligned}\left|\nabla_{\theta_k}\Omega(V_i(\theta),\theta)^{-1}\right| &= \left|\Omega(V_i(\theta),\theta)^{-1}\left[\nabla_{\theta_k}\Omega(V_i(\theta),\theta)\right]\Omega(V_i(\theta),\theta)^{-1}\right| \\ &\leq \left\{\begin{array}{c}\sup_{i,\theta}|\Omega(V_i(\theta),\theta)^{-1}|\times \\ \sup_{i,\theta}|\nabla_{\theta_k}\Omega(V_i(\theta),\theta)|\times \\ \sup_{i,\theta}|\Omega(V_i(\theta),\theta)^{-1}|\end{array}\right\}=O_p(1).\end{aligned}$$

For the term in (95), the absolute value of its first component is $o_p(N^{-\delta_d})$ from Lemma 6c). For the second component we must analyze:

$$\left\{\frac{\left[\hat{Q}-Q\right]'\nabla_{\theta_k}D_N(v^*,\theta^*)Y}{N}\right\}+\left\{\frac{Q'\nabla_{\theta_k}D_N(v^*,\theta^*)Y}{N}\right\} \tag{99}$$

Employing arguments similar to those for (93), the absolute value of the elements in the first term is $o_p(N^{-\delta_d})$, while those in the second are $O_p(1)$ from (A6) and Lemma 2d.

For the term in (96), its first component is $o_p(N^{-\delta_d})$ from a), while the second component, which is similar to that in (99), is $O_p(1)$. The proof for d) is very similar to that for c).

To establish asymptotic normality for the estimator of the index parameter vector, the next lemma shows that the estimated Hessian matrix converges to a fixed matrix.

Lemma 9. The Hessian. Recalling the definition of H_0 in Theorem 3, for any $\theta^+ \xrightarrow{p} \theta_0$:

$$\nabla_\theta\hat{G}^*(\theta^+)\xrightarrow{p}H_0$$

Proof. To prove the result, recall the definition of $\hat{\alpha}^*$ and $\bar{\alpha}$ in (D8) and let

$$\nabla_{\theta_{k_1}}\hat{G}^*_{k_2}(\theta)\equiv-\frac{1}{N}\sum_{i=1}^{N}\hat{Z}_{ik_2}R_i\nabla_{\theta_{k_1}}\hat{\alpha}^*(V_j(\theta),\theta)$$

$$H_{k_1k_2}(\theta)\equiv E[-\frac{1}{N}\sum_{i=1}^{N}Z_{ik_2}R_i\nabla_{\theta_{k_1}}\bar{\alpha}(V_j(\theta),\theta)],$$

where $\nabla_{\theta_{k_1}} \hat{G^*_{k2}}(\theta)$ is the derivative of k_2^{th} moment w.r.t. θ_{k_1}; while $H_{k_1 k_2}(\theta)$ is the fixed matrix to which it is converging. Then since $\theta^+ \xrightarrow{p} \theta_0$, we must show:

$$a) \sup_\theta \left| \nabla_{\theta_{k_1}} \hat{G}^*_{k_2}(\theta) + [\frac{1}{N} \sum_{i=1}^N Z_{ik_2} R_i \nabla_{\theta_{k_1}} \bar{\alpha}(V_j(\theta), \theta)] \right| \xrightarrow{p} 0$$

$$b) \sup_\theta |\sum_{i=1}^N Z_{ik_2} R_i \nabla_{\theta_{k_1}} \bar{\alpha}(V_j(\theta), \theta) + H_{k_1 k_2}(\theta)| \quad \xrightarrow{p} 0.$$

For a):

$$\begin{aligned} & [-\frac{1}{N} \sum_{i=1}^N Z_{ik_2} R_i \nabla_{\theta_{k_1}} \bar{\alpha}(V_j(\theta), \theta)] - \nabla_{\theta_{k_1}} \hat{G}^*_{k_2}(\theta) \\ = \; & \frac{1}{N} \sum_{i=1}^N Z_{ik_2} R_i \nabla_{\theta_{k_1}} [\hat{\alpha}^*(V_i(\theta), \theta) - \bar{\alpha}(V_i(\theta), \theta)] \qquad (100) \\ & + \frac{1}{N} \sum_{i=1}^N \left[\hat{Z}_{ik_2} - Z_{ik_2} \right] R_i \nabla_{\theta_{k_1}} \bar{\alpha}(V_i(\theta), \theta) \qquad (101) \\ & + \frac{1}{N} \sum_{i=1}^N \left[\hat{Z}_{ik_2} - Z_{ik_2} \right] R_i \nabla_{\theta_{k_1}} [\hat{\alpha}^*(V_i(\theta), \theta) - \bar{\alpha}(V_i(\theta), \theta)]. \qquad (102) \end{aligned}$$

The absolute value of the term in (100) is bounded above by:

$$\sup_\theta \left| \frac{1}{N} \sum_{i=1}^N Z_{ik_2} R_i \right| \sup_{i,\theta} |\nabla_{\theta_{k_1}} [\hat{\alpha}^*(V_i(\theta), \theta) - \bar{\alpha}(v, \theta)$$

From Lemma 8, the first component is $o_p(1)$. The second component is $O_p(1)$ under (A7), which completes the argument for (100).

Turning to the term in (101), from Lemma 5a, trimming can be taken as known. Then, since θ only enters through $\nabla_{\theta_{k_1}} \bar{\alpha}(V_i(\theta), \theta)$, the resulting term is then $o_p(1)$ from Cauchy-Schwarz.

Using arguments similar to those above, the term in (102) converges to 0 faster than the other two. For b), with $t_i \nabla_{\theta_{k_1}} \bar{\alpha}(V_i(\theta), \theta)$ bounded from (A7), uniform convergence holds from Amemiya(1985, Thm. 4.2.1).

Lemma 10. The Gradient. With $V_j \equiv V_j(\theta_0)$ and $\alpha_0(V_j) \equiv \alpha(V_j, \theta_0)$, recall that $Y_j = R_j \alpha_0(V_j) + \varepsilon_j$ and denote $\hat{\alpha}^*_0(V_j) \equiv \hat{\alpha}^*(V_j, \theta_0)$. With $\hat{G^*}(\theta_0)$ as the gradient and with

$$\bar{\mathfrak{R}}'_z(V_j) \equiv Z'_j - E(Z'_j R_j | V_j) E\left[\bar{Q}^{*\prime}_j R_j | V_j \right]^{-1} \bar{Q}^{*\prime}_j, \qquad (103)$$

it follows that:

$$\sqrt{N} \hat{G^*}(\theta_0) \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^N Z'_j [Y_j - R_j \hat{\alpha}^*_0(V_j)] = \frac{1}{\sqrt{N}} \sum_{j=1}^N \bar{\mathfrak{R}}'_z(V_j) \varepsilon_j + o_p(1). \qquad (104)$$

Proof. Recalling that $\varepsilon_j = Y_j - R_j\alpha_o(V_j)$, write $\sqrt{N}\hat{G}^*(\theta_0)$ in (104) as.

$$\sqrt{N}\hat{G}^*(\theta_0) = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\hat{Z}_j'\varepsilon_j - \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\hat{Z}_j'R_j[\hat{\alpha}_0^*(V_j) - \alpha_0(V_j)] \tag{105}$$

$$= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}Z_j'\varepsilon_j - \frac{1}{\sqrt{N}}\sum_{j=1}^{N}Z_j'R_j[\hat{\alpha}_0^*(V_j) - \alpha_0(V_j)] \tag{106}$$

$$+\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left[\hat{Z}_j - Z_j\right]'\varepsilon_j - \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left[\hat{Z}_j - Z_j\right]'R_j[\hat{\alpha}_0^*(V_j) - \alpha_0(V_j)]. \tag{107}$$

The first term in (106) has the required form. For the second term in (106), from Lemma 11

$$\frac{1}{\sqrt{N}}\sum_{j=1}^{N}Z_j'R_j[\hat{\alpha}^*(V_j) - \alpha_0(V_j)] = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}E[Z_j'R_j|V_j]E(\bar{Q}_j^{*\prime}R_j|V_j)^{-1}\bar{Q}_j^{*\prime}\varepsilon_j + o_p(1). \tag{108}$$

Therefore, the proof will follow if the terms in (107) vanish in probability. For the first term, the result follows under (D9) because estimated trimming can be taken as known under Lemma 5. For the second term in (107), it vanishes from Lemma 5, Lemma 7c and (D9).

For both the index estimator (as illustrated in the proof of Lemma 10 above) and marginal effects, we need a result for a weighted difference between bias corrected estimator, $\hat{\alpha}^*(V_j)$ and the true function, $\alpha_0(V_j)$. Lemma 11 provides the key result, with required intermediate results provided in Lemmas 12-13.

Lemma 11. Letting S_j^* be i.i.d. or constant and define $S_j \equiv t_j S_j^*$, where t_j is any trimming function of the form in (D3) that constrains $V_j(\theta_0)$ to a compact set and assume S_j is finite.[11] Referring to (D8), let $V_j \equiv V_j(\theta_0)$ and $D_N(V_j) \equiv D_N(V_j(\theta_0), \theta_0)$. Then, with $\bar{Q}_j{}^*, \bar{\Omega}_j^*$ defined in (D6) and (D8), let:

$$\begin{aligned} U &\equiv \frac{1}{N}\sum_{j=1}^{N}S_j[\hat{\alpha}^*(V_j) - \alpha_0(V_j)] \\ U^* &\equiv \frac{1}{N}\sum_{j=1}^{N}E\left(\bar{S}_j|V_j\right)\Omega^*(V_j)^{-1}g(V_j)\bar{Q}_j^*\varepsilon_j \equiv \frac{1}{N}\sum_{j=1}^{N}E(S_j|V_j)E(\bar{Q}_j^{*\prime}R|V_j)^{-1}\bar{Q}_j^{*\prime}\varepsilon_j \end{aligned}$$

Then:

$$\sqrt{N}\left[U - U^*\right] = o_p(1).$$

Proof. To prove this result, with Q^* defined in (D6), we will first show:

$$\sqrt{N}U = \sqrt{N}[U_1 + U_2] + o_p(1), U_1 \equiv \frac{1}{N(N-1)}\sum_{j=1}^{N}\Omega^*(V_j)^{-1}Q^{*\prime}D_N(V_j)\varepsilon \tag{109}$$

$$U_2 = \frac{1}{N(N-1)(N-2)}\sum_{j=1}^{N}\sum_{i\neq j}\sum_{k\neq i,j}\rho_{jik}, \tag{110}$$

[11] The proof is simplified under this assumption, but only requires $E(S_{jk}|V_j)$ to be finite.

where

$$\rho_{jik} \equiv \left[S_j \Omega^*(V_j)^{-1} Q_i^{*\prime} K_i(V_j)\right] R_i \left[\begin{array}{c} \Omega(V_i)^{-1} Q_k' K_k(V_i)\left[\varepsilon_k + \Delta_k(V_i)\right] - \\ \Omega(V_j)^{-1} Q_k' K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right] \end{array}\right].$$

The proof will then follow by showing : $\sqrt{N}U_1 = \sqrt{N}U^* + o_p(1)$ and $\sqrt{N}U_2 = o_p(1)$.

Beginning with the decomposition of $\sqrt{N}U$ into $\sqrt{N}[U_1 + U_2]$ above, note that:

$$\begin{aligned} \hat{\alpha}^*(V_j) &\equiv \hat{\Omega}^*(V_j)^{-1} \frac{\hat{Q}^{*\prime} D_N(V_j)\left[Y - \hat{\Delta}(V_j)\right]}{N-1} \\ &= \alpha_0(V_j) + \hat{\Omega}^*(V_j)^{-1} \frac{\hat{Q}^{*\prime} D_N(V_j)\left[\varepsilon + \Delta(V_j) - \hat{\Delta}(V_j)\right]}{N-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} \sqrt{N}U &= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} S_j \hat{\Omega}^*(V_j)^{-1} \frac{\hat{Q}^{*\prime} D_N(V_j)\left[\varepsilon + \Delta(V_j) - \hat{\Delta}(V_j)\right]}{N-1} \\ &= \sqrt{N}U_1 : \frac{1}{\sqrt{N}} \sum_{j=1}^{N} S_j \Omega^*(V_j)^{-1} \left[\frac{Q^{*\prime} D_N(V_j)\varepsilon}{N-1}\right] + \\ &\quad T_1 : \frac{1}{\sqrt{N}} \sum_{j=1}^{N} S_j \Omega^*(V_j)^{-1} \left[\frac{Q^{*\prime} D_N(V_j)\left[\Delta(V_j) - \hat{\Delta}(V_j)\right]}{N-1}\right] + \\ &\quad T_2 : \frac{1}{\sqrt{N}} \sum_{j=1}^{N} S_j \left[\hat{\Omega}^*(V_j)^{-1} - \Omega^*(V_j)^{-1}\right] \frac{Q^{*\prime} D_N(V_j)\left[\varepsilon + \Delta(V_j) - \hat{\Delta}(V_j)\right]}{N-1} + \\ &\quad T_3 : \frac{1}{\sqrt{N}} \sum_{j=1}^{N} S_j \left[\hat{\Omega}^*(V_j)^{-1} - \Omega^*(V_j)^{-1}\right] \frac{\left[\hat{Q}^* - Q^*\right]' D_N(V_j)\left[\varepsilon + \Delta(V_j) - \hat{\Delta}(V_j)\right]}{N-1} + \\ &\quad T_4 : \frac{1}{\sqrt{N}} \sum_{j=1}^{N} S_j \left[\Omega^*(V_j)^{-1}\right] \frac{\left[\hat{Q}^* - Q^*\right]' D_N(V_j)\left[\varepsilon + \Delta(V_j) - \hat{\Delta}(V_j)\right].}{N-1} \end{aligned}$$

We begin by showing that terms T_2, T_3 and T_4 vanish in probability. For T_2, recall that Q^* contains a trimming function and define:

$$\tilde{T_2} \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^{N} S_j [\hat{\Omega}^*(V_j)^{-1} - \Omega^*(V_j)^{-1}] \frac{Q^{*\prime} D_N(V_j)\left[\varepsilon + \Delta(V_j) - \tilde{\Delta}(V_j)\right]}{N-1}, \tag{111}$$

where from (83) and (71) in the proof of Lemma 7a,

$$\Delta_j(V_i) - \tilde{\Delta}_j(V_i) = R_j\left\{[\tilde{\alpha}_{0i} - \alpha_{0i})] - [\tilde{\alpha}_{0j} - \alpha_{0j}]\right\}, \tag{112}$$

$$\tilde{\alpha}_{0i} = \alpha_{0i} + \frac{1}{N} \Omega(V_i)^{-1} Q' D_N(\varepsilon + \Delta(V_i)). \tag{113}$$

From Lemma 6c, $\sup_j |\hat{\Omega}^*(V_j)^{-1} - \Omega^*(V_j)^{-1}| = o_p\left(N^{-1/4}\right)$. From (79) in the proof of Lemma 7a, $\sup t_i|\hat{\alpha_{0i}} - \tilde{\alpha}_{0i}|$ is $o_p\left(N^{-1/4}\right)$. Therefore, $\left|T_2 - \tilde{T}_2\right| = o_p\left(N^{-1/2}\right)$. It then suffices to consider $\tilde{T}_2$. With $\tilde{\alpha}^*_{0i} - \alpha_{0i} = \Omega_i^{*-1}Q^{*\prime}D_N(V_i)\left[\varepsilon + \Delta\left(V_i\right) - \tilde{\Delta}(V_i)\right]$ in (88) of the proof of Lemma 7c:

$$\begin{aligned}\left|\tilde{T}_2\right| &= \frac{1}{\sqrt{N}}|\sum_{j=1j}^{N}[\hat{\Omega}^*(V_j)^{-1} - \Omega^*(V_j)^{-1}]\Omega^*(V_j)[\tilde{\alpha}^*(V_j) - \alpha_0(V_j)]| \\ &\leq \sup_i |[\hat{\Omega}^*(V_j)^{-1} - \Omega^*(V_j)^{-1}]\Omega^*(V_j)|\frac{1}{N}\sum_{j=1}^{N}|\tilde{\alpha}^*(V_j) - \alpha_0(V_j)| \qquad (114)\end{aligned}$$

From Lemma 6c, the first component in (114) is $o_p\left(N^{-1/4}\right)$. Employing arguments similar, but simpler than, those in the proof of Lemma 7c, it can be shown that the second component is $o_p\left(N^{-1/4}\right)$, which completes the argument for T_2.

For T_4, as above,

$$\left|T_4 - \tilde{T}_4\right| = o_p\left(N^{-1/2}\right)$$

$$\tilde{T}_4 = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} S_j\left[\Omega^*(V_j)^{-1}\right]\left[\hat{Q}^* - Q^*\right]' D_N(V_j)\left[\varepsilon + \Delta(V_j) - \tilde{\Delta}(V_j)\right].$$

It can be shown that the $\varepsilon - component$ of $\tilde{T}_4$ converges to 0 as a degenerate U-statistic. The analysis for the remaining components is similar to that above as $\hat{Q}_i^{\ *} - Q_i^*$ is uniformly $o_p\left(N^{-1/4}\right)$ in i from (D5) and Lemma 5.

Turning to T_1, notice that

$$\begin{aligned}\Delta_i(V_j) - \hat{\Delta}_i(V_j) &= R_i\left\{[\alpha_0(V_i) - \alpha_0(V_j)] - [\hat{\alpha}_0(V_i) - \hat{\alpha}_0(V_j)]\right\}, i \neq j \\ &= R_i\left\{[\hat{\alpha}_0(V_j) - \alpha_0(V_j)] - [\hat{\alpha}_0(V_i) - \alpha_0(V_i)]\right\} \\ &= R_i\begin{bmatrix} \hat{\Omega}(V_j)^{-1}\sum_{k\neq j}\frac{\hat{Q}_k' K_k(V_j)[\varepsilon_k + \Delta_k(V_j)]}{N-1} - \\ \hat{\Omega}(V_i)^{-1}\sum_{k\neq i}\frac{\hat{Q}_k' K_k(V_i)[\varepsilon_k + \Delta_k(V_i)]}{N-1} \end{bmatrix}.\end{aligned}$$

Therefore,

$$T_1 = \frac{1}{\sqrt{N}(N-1)}\sum_{j=1}^{N} S_j\Omega^*(V_j)^{-1}\sum_{i\neq j} Q_i^{*\prime}K_i(V_j)R_i\left\{\begin{array}{c} \hat{\Omega}(V_j)^{-1}\sum_{k\neq j}\frac{\hat{Q}_k' K_k(V_j)[\varepsilon_k + \Delta_k(V_j)]}{N-1} \\ -\hat{\Omega}(V_i)^{-1}\sum_{k\neq i}\frac{\hat{Q}_k' K_k(V_i)[\varepsilon_k + \Delta_k(V_i)]}{N-1} \end{array}\right\}. \qquad (115)$$

To simplify notation and shorten expressions, let:

$$A_j \equiv S_j\Omega^*(V_j)^{-1}, B_{ij} \equiv Q_i^{*\prime}K_i(V_j)R_i, C_{kj} \equiv \Omega(V_j)^{-1}Q_k'K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right].$$

Then, employing arguments similar to those for analyzing T_2, we may replace $\hat{\Omega}(V_r)^{-1}$with $\Omega(V_r)^{-1}$ for $r = j, i$, $\hat{Q}_k^{*\prime}$ with $Q_k^{*\prime}$, and write the expression in (115) as:

$$T_1 = \frac{1}{\sqrt{N}\,(N-1)} \sum_{j=1}^{N} A_j \sum_{i \neq j} B_{ij} \sum_{k \neq i,j} \frac{[C_{kj} - C_{ki}]}{N-1} + o_p(1) \tag{116}$$

$$= \frac{1}{\sqrt{N}\,(N-1)} \sum_{j=1}^{N} A_j \sum_{i \neq j} B_{ij} \left\{ \left(\sum_{k \neq i,j} \frac{C_{kj} - C_{ki}}{N-2} \right) + \frac{C_{ij} - C_{ii}}{N-2} \right\} \left(\frac{N-2}{N-1} \right) + o_p(1). \tag{117}$$

Since $A_j = O(1)$ and $B_{ij} = O(1/h)$, up to an additive term that is $o_p(1)$, we have

$$T_1 = \frac{1}{\sqrt{N}\,(N-1)} \sum_{j=1}^{N} A_j \sum_{i \neq j} B_{ij} \sum_{k \neq i,j} \frac{[C_{kj} - C_{ki}]}{N-2} + O\left(\frac{N^{-5/2}}{h} \right) \sum_{j=1}^{N} \sum_{i \neq j}^{N} [C_{ij} - C_{ii}]\,.$$

$$= \frac{1}{\sqrt{N}\,(N-1)} \sum_{j=1}^{N} A_j \sum_{i \neq j} B_{ij} \sum_{k \neq i,j} \frac{[C_{kj} - C_{ki}]}{N-2} + o_p(1).$$

From above, as $T_1 = U_2 + o_p(1)$, we have now established that:

$$\sqrt{N}U = \sqrt{N}U_1 + \sqrt{N}U_2 + o_p(1), \quad U_1 \equiv \frac{1}{\sqrt{N}(N-1)} \sum_{j=1}^{N} S_j \Omega^*(V_j)^{-1} Q^{*\prime} D_N(V_j) \varepsilon \tag{118}$$

$$U_2 \equiv \frac{1}{(N-1)\,(N-2)} \sum_{j=1}^{N} \sum_{i \neq j} \sum_{k \neq i,j} \rho_{jik} + o_p(1) \tag{119}$$

From Lemma 12, up to an additive term that is $o_p\,(1)$:

$$\sqrt{N}U_1 = \frac{1}{N} \sum_{j=1}^{N} E\,(S_j|V_j)\,\bar{\Omega}^*(V_j)^{-1} g(V_j) \bar{Q}_j^{*\prime} \varepsilon_j \equiv \frac{1}{N} \sum_{j=1}^{N} E\,(S_j|V_j)\,E(\bar{Q}_j^{*\prime} R|V_j)^{-1} \bar{Q}_j^{*\prime} \varepsilon_j$$

and from Lemma 13, $\sqrt{N}U_2 = o_p(1)$. Lemma 11 then follows.

Lemma 12 Referring to the proof of Lemma 11:

$$\sqrt{N}U_1 \equiv \frac{1}{\sqrt{N}(N-1)} \sum_{j=1}^{N} S_j \Omega^*(V_j)^{-1} Q^{*\prime} D_N(V_j) \varepsilon = \sqrt{N}U_1^*$$

$$U_1^* \equiv \frac{1}{N} \sum_{j=1}^{N} E\,(S_j|V_j)\,\bar{\Omega}^*(V_j)^{-1} g(V_j) \bar{Q}_j^{*\prime} \varepsilon_j \equiv \frac{1}{N} \sum_{j=1}^{N} E\,(S_j|V_j)\,E(\bar{Q}_j^{*\prime} R|V_j)^{-1} \bar{Q}_j^{*\prime} \varepsilon_j.$$

Proof. Defining $\rho_{ji} \equiv S_j \Omega^*(V_j)^{-1} Q_i^{*\prime} K_i(V_j)\varepsilon_i$, we have

$$\begin{aligned}\sqrt{N}U_1 &= \frac{1}{\sqrt{N}(N-1)} \sum_{j=1}^{N}\sum_{i\neq j}^{N} S_j \Omega^*(V_j)^{-1} Q_i^{*\prime} K_i(V_j)\varepsilon_i \\ &= \sqrt{N}\left(\begin{array}{c} N \\ 2 \end{array}\right)^{-1} \sum_{j=1}^{N}\sum_{i>j}^{N} \left[\rho_{ij} + \rho_{ji}\right]/2.\end{aligned}$$

With *obs* j denoting observation j, as the above expression is a centered U-statistic, we have:

$$\begin{aligned}\sqrt{N}U_1 &= \frac{2}{\sqrt{N}} \sum_{j=1}^{N} E\left[\rho_{ij} + \rho_{ji} | obs \;\; j\right]/2 + o_p(1) \\ &= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E\left[\rho_{ij} | obs \;\; j\right] + o_p(1) \\ &= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E\left[S_i \Omega^*(V_i)^{-1} K_j(V_i) | obs \;\; j\right] Q_j^{*\prime}\varepsilon_j + o_p(1) \\ &= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E\left[E\left(S_i|V_i\right) \Omega^*(V_i)^{-1} K_j(V_i) | obs \;\; j\right] Q_j^{*\prime}\varepsilon_j + o_p(1) \\ &= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E\left(S_j|V_j\right) \Omega^*(V_j)^{-1} g(V_j) Q_j^{*\prime}\varepsilon_j + o_p(1) \\ &= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E\left(S_j|V_j\right) E[Q_j^{*'} R_j|V_j]^{-1} Q_j^{*'}\varepsilon_j + o_p(1)\end{aligned}$$

from a standard U-statistic projection and (D8). To complete the proof, we need to show that we can replace Q^* with $\bar{Q}^*$. To this end, let $\delta_j \equiv E[Q_j^{*'} R_j|V_j]^{-1} Q_j^{*'} - E[\bar{Q}_j^{*'} R_j|V_j]^{-1} \bar{Q}_j^{*'}$ and define

$$D \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E\left[S_j|V_j\right] \delta_j \varepsilon_j. \tag{120}$$

With $E(D) = 0$, the proof will follow if $Var(D) \to 0$. We have:

$$Var(D) = E\left\{E\left[S_j|V_j\right] \delta_j \delta_j' E\left[S_j|V_j\right]'\right\} \sigma_\varepsilon^2. \tag{121}$$

Since $E[S_j|V_j]$ is bounded, it suffices to consider: $E\left(\delta_j \delta_j'\right)$. Noting that Δ_j is a difference in terms, one depending from (D6) on t_1^* and the other on t_1 converging to t_1^*, from the same arguments used to establish Lemma 5, it can be shown that $E\left[\delta_j \delta_j'\right] \to 0$, which completes the proof.

Lemma 13. A Degenerate U-Statistic Result. Referring to the proof of Lemma 11:

$$\sqrt{N}U_2 \equiv o_p(1).$$

Proof. With

$$\rho_{jik} \equiv \left[S_j\Omega^*(V_j)^{-1}Q_i^{*\prime}K_i(V_j))\right] * R_i \left[\begin{array}{c} \Omega(V_i)^{-1}Q_k'K_k(V_i)\left[\varepsilon_k + \Delta_k(V_i)\right] - \\ \Omega(V_j)^{-1}Q_k'K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right] \end{array} \right], \tag{122}$$

we have:

$$\sqrt{N}U_2 \equiv \frac{1}{\sqrt{N}\,(N-1)\,(N-2)} \sum_{j=1}^{N} \sum_{i\neq j} \sum_{k\neq i,j} \rho_{jik} \tag{123}$$

Letting

$$\rho^*_{jik} = \left[\rho_{jik} + \rho_{ijk} + \rho_{ikj} + \rho_{kij} + \rho_{jki} + \rho_{kji}\right]/3! \tag{124}$$

we can rewrite (123) as:

$$\sqrt{N}U_2 = \left(\begin{array}{c} N \\ 3 \end{array} \right)^{-1} \sum_{j=1}^{N} \sum_{i>j}^{N} \sum_{k>i>j}^{N} \rho^*_{jik},$$

To see that this result holds, notice that

$$\begin{aligned} \sum_{j=1}^{N} \sum_{i\neq j}^{N} \sum_{k\neq i,j}^{N} \rho_{jik} \;=\; & \sum_{j} \sum_{i>j} \sum_{k>i>j} \rho_{jik} + \sum_{j} \sum_{i>j} \sum_{i>k>j} \rho_{jik} \\ & + \sum_{j} \sum_{i>j} \sum_{i>j>k} \rho_{jik} + \sum_{j} \sum_{i<j} \sum_{k<i<j} \rho_{jik} \\ & + \sum_{j} \sum_{i<j} \sum_{i<k<j} \rho_{jik} + \sum_{j} \sum_{i<j} \sum_{i<j<k} \rho_{jik} \end{aligned}$$

The first term has the required form. For the second term:

$$\sum_{j} \sum_{k>j} \sum_{i>k>j} \rho_{jik} = \sum_{j} \sum_{i>j} \sum_{k>i>j} \rho_{jki},$$

which has the required form. The analysis for other terms is similar. Recalling the definitions of ρ_{jik} and ρ^*_{jik} in (122) and (124), the remainder of the proof consists of showing the following:

$$\begin{array}{lcl} a) & : & E[\rho_{jik}|obs\;\; j] = o(N^{-1/2}),\ \text{uniformly in } V_j \\ b) & : & E\left(\rho^*_{jik}\right) = o(N^{-1/2}) \\ c) & : & E\left(\rho^*_{jik}|obs\;\; j\right) = o(h^2),\ \text{uniformly in } V_j \\ d) & : & \sqrt{N}U_2 = o_p(1). \end{array}$$

For a), we may write $E\left[\rho_{jik}|obs\;\; j\right]$ as: $E_1 - E_2$, where

$$E_1 = \quad E \left\{ \begin{array}{c} S_j\Omega^*(V_j)^{-1}E\left[Q_i^{*\prime}R_i|V_i\right]K_i(V_j)\Omega(V_i)^{-1}\times \\ E\left[Q_k'K_k(V_i)\left[\varepsilon_k + \Delta_k(V_i)\right]\;\;|V_i\right]\,|\;\; obs\;\; j \end{array} \right\} \tag{125}$$

$$E_2 = \quad E \left\{ \begin{array}{c} \left[S_j\Omega^*(V_j)^{-1}Q_i^{*\prime}R_i\right]K_i(V_j)\Omega(V_j)^{-1}\times \\ Q_k'K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right]\,|\;\; obs\;\; j \end{array} \right\}. \tag{126}$$

For the $E_1 - term$ in (125), we require the second interior expectation:

$$\begin{aligned} E\left[Q_k' K_k(V_i)\left[\varepsilon_k + \Delta_k(V_i)\right] \;|V_i\right] &= E\left[Q_k' K_k(V_i)\Delta_k(V_i) \;|V_i\right] \\ &= E\left[Q_k' R_k K_k(V_i)[\alpha_0(V_k(\theta_0)) - \alpha_0(V_i(\theta_0))] \;|V_i\right] \\ &= EE\left[Q_k' R_k K_k(V_i)[\alpha_0(V_k(\theta_0)) - \alpha_0(V_i(\theta_0))] \;|V_i, V_k\right] \\ &= E\left[E\left(Q_k' R_k | V_k\right) K_k(V_i)[\alpha_0(V_k(\theta_0)) - \alpha_0(V_i(\theta_0))] \;|V_i\right] \\ &= \int \frac{\Omega(V_k)}{g(V_k)} \frac{1}{h} K\left(\frac{V_i - V_k}{h}\right)[\alpha_0(V_k(\theta_0)) - \alpha_0(V_i(\theta_0))]\, g(V_k) dV_k \\ &= \int \Omega(V_i + hz) K(z)\left[\alpha_0(V_i + hz) - \alpha_0(V_i(\theta_0))\right] dz \\ &= h^2 \nabla_i^2 \Omega(V_i)\alpha_0(V_i) + O(h^4), \end{aligned} \tag{127}$$

where ∇_i^2 denotes a second derivative taken with respect to V_i. Substituting this expression into (125) yields:

$$\begin{aligned} E_1 = \;& ES_j \Omega^*(V_j)^{-1} E\left[Q_i^{*\prime} R_i | V_i\right] K_i(V_j)\Omega(V_i)^{-1}[h^2 \nabla_i^2[\Omega(V_i)\alpha_0(V_i)] + O(h^4)] |\; obs \;\; j\} \\ = \;& E\left\{S_j \Omega^*(V_j)^{-1} \frac{\Omega^*(V_i)}{g(V_i)} K_i(V_j)\Omega(V_i)^{-1}[h^2 \nabla_i^2[\Omega(V_i)\alpha_0(V_i)]] + O(h^4)] |\; obs \;\; j\right\} \\ = \;& S_j \Omega^*(V_j)^{-1} E\left\{\frac{\Omega^*(V_i)}{g(V_i)} K_i(V_j)\Omega(V_i)^{-1}[h^2 \nabla_i^2[\Omega(V_i)\alpha_0(V_i)]] + O(h^4)] |\; obs \;\; j\right\} \\ = \;& S_j \Omega^*(V_j)^{-1}\left\{\Omega^*(V_j)\Omega(V_j)^{-1}[h^2 \nabla_j^2[\Omega(V_j)\alpha_0(V_j)]] + O(h^4)]\right\} \\ = \;& h^2 S_j \Omega(V_j)^{-1}[\nabla_j^2[\Omega(V_j)\alpha_0(V_j)]] + O(h^4), \end{aligned}$$

Turning to the $E_2 - term$ in (126)

$$\begin{aligned} E_2 &= E\left\{\left[S_j \Omega^*(V_j)^{-1} Q_i^{*\prime} R_i\right] K_i(V_j)\Omega(V_j)^{-1} Q_k' K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right] |\; obs \;\; j\right\} \\ &= E\left\{\left[S_j \Omega^*(V_j)^{-1} E\left[Q_i^{*\prime} R_i | V_i\right]\right] K_i(V_j)\Omega(V_j)^{-1} Q_k' K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right] |\; obs \;\; j\right\} \\ &= E\left\{S_j \Omega^*(V_j)^{-1} \frac{\Omega^*(V_i)}{g(V_i)} K_i(V_j)\Omega(V_j)^{-1} | obs \;\; j\right\} E\left\{Q_k' K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right] |\; obs \;\; j\right\}. \end{aligned}$$

In a derivation similar to that in (127).

$$E\left\{Q_k' K_k(V_j)\left[\varepsilon_k + \Delta_k(V_j)\right] |\; obs \;\; j\right\} = [h^2 \nabla_j^2[\Omega(V_j)\alpha_0(V_j))] + O(h^4)].$$

Substituting this expression into E_2 above:

$$\begin{aligned} E_2 &= E\left\{S_j \Omega^*(V_j)^{-1} \frac{\Omega^*(V_i)}{g(V_i)} K_i(V_j)\Omega(V_j)^{-1} | obs \;\; j\right\}\left[h^2 \nabla_j^2\left[\Omega(V_j)\alpha_0(V_j)\right] + O(h^4)\right] \\ &= E\left\{S_j \Omega^*(V_j)^{-1} \frac{\Omega^*(V_i)}{g(V_i)} K_i(V_j)\Omega(V_j)^{-1}[h^2 \nabla_j^2[\Omega(V_j)\alpha_0(V_j)] + O(h^4)] | obs \;\; j\right\} \\ &= S_j \Omega^*(V_j)^{-1}\left[\int \Omega^*(V_i) K_i(V_j) dV_i\right] \Omega(V_j)^{-1}[h^2 \nabla_j^2\left[\Omega(V_j)\alpha_0(V_j)\right] + O(h^4) \\ &= S_j \Omega^*(V_j)^{-1}\left[\Omega^*(V_j) + O(h_2^2)\right] \Omega(V_j)^{-1}[h^2 \nabla_j^2\left[\Omega(V_j)\alpha_0(V_j)\right] + O(h^4)] \\ &= h^2 S_j \Omega(V_j)^{-1}[\nabla_j^2[\Omega(V_j)\alpha_0(V_j)]] + O(h^4). \end{aligned}$$

The proof now follows as $E_1 - E_2 = O(h^4) = o(N^{-1/2})$ with $h = O\left(N^{-1/7.99}\right)$.

For b) the proof follows immediately from (a) and the fact that all terms in ρ^*_{jik} have the same unconditional expectation. For (c), it suffices to analyze the components involving $\Delta's$. With the analysis being very similar to that for E_1 or E_2, the result follows.

For d), from b), $\text{E}\left(\sqrt{N}\hat{U}\right) = o(1)$. For the variance, it can be shown that

$$Var\left(\sqrt{N}\hat{U}\right) = Var(E[\rho^*_{jik}|obs\ \ j]) = o\left(1\right),$$

from which the result now follows.

The results in Lemma 14 hold under regular kernels when the treatment depends on a single index. In the multiple index case, we would require a bias reduction strategy (e.g. higher order kernels) to obtain the required results. Below, for expositional purposes, we provide the argument for the single index case.

Lemma 14. Convergence Rates for Instruments. With the treatment instrument given as $\hat{P}_j$, let the preliminary instrument for estimating the index parameters in the outcome equation be

$$[\hat{P}_j + 1]X_{2j} \tag{128}$$

and denote the resulting parameter estimate as $\hat{\theta}_p$. Define the second stage instrument as

$$[\hat{P}_j \ \ 1]'[\nabla_\theta \hat{\alpha}^*(V_j\left(\hat{\theta}_p\right), \hat{\theta}_p)] \tag{129}$$

Then, denoting Z^o_{jk} as the k^{th} element of either (128) or (129), the instruments satisfy the conditions in (D5) and (D9) in that

$$\begin{aligned}
a) \ &: \ \sup_j \ \left|t_j\left[\hat{P}_j\left(\hat{\gamma}\right) - P_j\left(\gamma_*\right)\right]\right| = o_p(N^{-1/4}), \\
b) \ &: \ \frac{1}{\sqrt{N}}\sum_{j=1}^N t_j\left[\hat{Z}^o_{jk} - Z^o_{jk}\right]^2 = o_p(1), \\
c) \ &: \ \frac{1}{\sqrt{N}}\sum_{j=1}^N t_j\left[\hat{Z}^o_{jk} - Z^o_{jk}\right]\varepsilon_j = o_p(1).
\end{aligned}$$

Proof. To establish a) , from a Taylor series expansion in $\hat{\gamma}$, (A4), and the uniform convergence of probability derivatives (Lemma 2), it suffices to consider:

$$\left|t_i\left[\hat{P}_i\left(\gamma_*\right) - P_i\left(\gamma_*\right)\right]\right|.$$

From Lemma 2, with a window of $h_1 = O(N^{-1/5})$, the difference is uniformly $o_p(N^{-1/4})$, which completes the argument.

For b), we first consider the preliminary instrument in (128). With trimming constraining X_{2j} to a compact set, the result follows from a). Turning to the second stage instrument in (129), it suffices to show that:

$$\Delta_1(\hat{\theta}_p) \equiv \frac{1}{\sqrt{N}}\sum_{i=1}^N Tr\left\|\nabla_\theta \hat{\alpha}^*\left(V_j\left(\hat{\theta}_p\right), \hat{\theta}_p\right) - \nabla_\theta \alpha\left(V_j\left(\theta_0\right), \theta_0\right)\right\| = o_p(1),$$

where $\|A\| \equiv AA'$. From a Taylor series expansion,

$$\Delta_1(\hat{\theta}_p) = \Delta_1(\theta_0) + \frac{\partial \Delta_1(\theta_0)}{\partial \theta}(\hat{\theta}_p - \theta_0) + \frac{1}{2}(\hat{\theta}_p - \theta_0)' \frac{\partial^2 \Delta_1(\theta^+)}{\partial \theta \partial \theta'}(\hat{\theta}_p - \theta_0). \tag{130}$$

From the pointwise results in Lemma 2, the first term in (130) is $o_p(1)$. For the second and third terms, it can be shown that $|\nabla^2_\theta(\hat{\alpha}^* - \alpha_0)| = o_p(N^{-\delta_1}), \delta_1 > 0$ and that $|\nabla^3_\theta(\hat{\alpha}^* - \alpha_0)| = O_p(N^{\frac{1}{2}-\delta_2})$, $\delta_2 > 0$. Therefore, since $|\hat{\theta}_p - \theta_0| = O_p(N^{-\frac{1}{2}})$ from Theorem 3, the remaining terms of the expansion in (130) are $o_p(N^{-\frac{1}{2}})$.

For c), beginning with the preliminary instrument, it suffices to consider:

$$\Delta_2 \equiv \frac{1}{\sqrt{N}} \sum_j t_j \left[\hat{P}_{lj} - P_{lj}\right] X_{2j}\varepsilon_j$$

We first reduce the problem to studying a linear combination of estimated functions. To this end, write

$$\hat{P}_{lj} = \frac{\frac{1}{N-1}\sum_{i \neq j} \frac{1}{h_1} T_{li} K\left(\frac{V_{jT} - V_{iT}}{h_1}\right)}{\frac{1}{N-1}\sum_{i \neq j} \frac{1}{h_1} \left(\frac{V_{jT} - V_{iT}}{h_1}\right)} \equiv \frac{\hat{f}_j}{\hat{g}_j}$$

Then, with Δ_2^* defined as

$$\Delta_2^* \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^{N} t_j \left[\hat{P}_{lj} - P_{lj}\right] \frac{\hat{g}_j}{g_j} X_{2j}\varepsilon_j. \tag{131}$$

Employing Cauchy-Schwarz, it readily follows that $|\Delta_2 - \Delta_2^*| = o_p(1)$:

$$\begin{aligned} |\Delta_2 - \Delta_2^*| &= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} t_j \left[\hat{P}_{lj} - P_{lj}\right] \left[\frac{\hat{g}_j - g_j}{g_j}\right] X_{2j}\varepsilon_j \\ &\leq \sqrt{N} \sqrt{\frac{1}{N}\sum_{j=1}^{N} t_j \left[\hat{P}_{lj} - P_{lj}\right]^2} \sqrt{\frac{1}{N}\sum_{j=1}^{N} t_j \left[\frac{\hat{g}_j - g_j}{g_j}\right]^2 X_{2j}^2 \varepsilon_j^2} \end{aligned}$$

The proof will now follow if $\Delta_2^* = o_p(1)$. With ε in Δ_2^* ensuring that the bias in Δ_2^* is $o(1)$, it is possible to employ a U-statistic argument to show that $\Delta_2^* = o_p(1)$. The argument for the second stage instrument is not provided as it is very similar to that above with more steps required to reduce the problem to one of studying a linear combination of estimated functions.