Dummy Endogenous Variables in Nonseparable Models

Edward Vytlacil*

October 11, 2003

Abstract

This paper considers the nonparametric identification and estimation of the average effect of a dummy endogenous variable in nonseparable models. The analysis includes the case of a dummy endogenous variable in a discrete choice model as a special case. This paper establishes conditions under which it is possible to identify and consistently estimate the average effect of the dummy endogenous variable without the use of large support conditions and without relying on parametric functional form or distributional assumptions. A root-N consistent and asymptotically normal estimator is developed for a special case of the model.

JEL Numbers: C50, H43
KEYWORDS: instrumental variables, sample selection models, social program evaluation

1 Introduction

This paper considers dummy endogenous variables in models where the error term is not additively separable from the regressors. The paper shows conditions for identification and estimation of the average effect of the dummy endogenous variable without imposing large support assumptions as are required by "identification-at-infinity" arguments, and without imposing parametric functional form or distributional assumptions.

An important special case of this analysis is to examine the effect of a dummy endogenous variable in a discrete choice model. For example, if a researcher wishes to examine the effect

*Stanford University, Department of Economics. This paper has previously circulated under the title "Common Treatment Effects on Indices of Nonseparable Models: A Semiparametric Identification Analysis." I would like to thank Jaap Abbring, Hide Ichimura, Jim Heckman, John Pepper and Jim Powell for very helpful comments on this paper. I would like to give special thanks to Nese Yildiz for excellent research assistance and helpful comments on the paper, she made an enormous contribution to this paper. I would also especially like to acknowledge my debt to James Heckman for his continued support. I would also like to thank seminar participants at Berkeley, Chicago, Harvard/MIT, NYU, Princeton, Rutgers, Tinbergen Institute, UCLA, UCL, and participants at the Midwest Econometrics Association annual meeting at Chicago and the CEME meeting at Rochester. Financial support from the Joint Center for Poverty Research is gratefully acknowledged. Correspondence: Landau Economics Building, 579 Serra Mall, Stanford CA 94305; Email: vytlacil@stanford.edu; Phone: 650-725-7836; Fax: 650-725-5702.

of a job training program on later employment, he or she might specify a probit equation for employment and include a dummy variable regressor for whether the individual received job training. One might expect that job training is endogenous, in particular, is correlated with the error term in the employment decision rule. In the discrete choice model, the error term is not additively separable from the regressors and thus standard instrumental variable techniques are not valid even if one has a variable that is correlated with job training but not with the error term of the employment equation.[1] Following Heckman (1978), one can impose a system of equations for the joint determination of the endogenous variable (job training) and the outcome variable (later employment). Heckman (1978) imposes joint normality assumptions and develops the maximum likelihood estimator for the resulting model. The model has a form similar to a multivariate probit model, and is referred to as a "multivariate probit model with structural shift" by Heckman (1978).[2]

This raises the question of whether it is possible to identify and consistently estimate the effect of a dummy endogenous variable in nonseparable outcome equations such as discrete choice models without imposing parametric distributional assumptions. One approach is to follow the analysis of Heckman (1990a,b) to use "identification-at-infinity" arguments to identify and estimate the average effect of the dummy endogenous variable on the outcome of interest if large support conditions hold. In particular, this approach assumes that the propensity score has support equal to the full unit interval, where the propensity score is the probability of receiving treatment conditional on observed covariates.[3] The drawbacks of this method is that it requires very strong, large support conditions, and that estimation that directly follows the identification strategy involves estimation on "thin sets" and thus a slow rate of convergence.[4] Angrist (2001) suggests that, if large support conditions do not hold, then the average effect of the dummy endogenous variable in nonseparable models is identified only through distributional assumptions.[5] Angrist (1991, 2001) proposes treating the outcome equation as a linear equation as an approximation,[6] or using instrumental variables to identify the "local average treatment effect."[7]

[1]See, e.g., the discussion in Heckman and Robb, 1985.

[2]A closely related model is the simultaneous probit model of Amemiya (1978), in which a probit model contains a continuous endogenous regressor. Later analysis of this model includes Lee (1981), Rivers and Vuong (1988), and Newey (1986). See Blundell and Powell (2000) for analysis of a semiparametric version of this model. The assumptions and methods used by Blundell and Powell (2000) are not appropriate for the case of a dummy endogenous variable, and likewise the assumptions and methods imposed here are not appropriate for the case of a continuous endogenous variable.

[3]Heckman (1990a,b) assumed that the outcome equation is additively separable in the regressors and the error term, but his analysis extends immediately to the nonseparable case. See also Cameron and Heckman (1998) and Heckman and Taber (1998) for identification-at-infinity arguments in the context of a system of discrete choice equations. Heckman and Vytlacil (1991,2001a) also further develop relevant identification-at-infinity arguments.

[4]See Andrews and Schafgans (1998), Schafgans (2000), and Schafgans and Zinde-Walsh (200) for results for the additively separable model.

[5]Angrist (2000) states one exception to this rule, that average effect will be identified if one imposes conditions such that average effect coincides with the local average treatment effect (LATE). This paper does not impose any such conditions.

[6]See Bhattacharya, et al., 1999, for a related Monte Carlo analysis.

[7]The "local average treatment effect" (LATE) was introduced in Imbens and Angrist, 1994. Under their

Heckman and Vytlacil (2001b) establish that the large support assumption invoked for identification-at-infinity arguments is necessary and sufficient for identification of the average effect of the dummy endogenous variable in the case of a switching regression model if no auxiliary assumptions are imposed. However, the Heckman and Vytlacil (2001b) result is for switching regression models, which leaves open the possibility of avoiding large support assumptions in nonseparable models that are not of the form of switching regression models.

This paper shows that it is possible to identify and estimate the average effect of a dummy endogenous variable in a nonseparable outcome equation (a) without imposing large support conditions, and (b) without relying on parametric distributional or functional form assumptions. This result holds in the a large class of nonseparable models referred to as "generalized regression" models by Han (1987), and includes both threshold crossing models as used in discrete choice analysis and transformation models such as the Box-Cox model and the proportional hazards model with unobserved heterogeneity. A root-N consistent and asymptotically normal estimator is developed for a special case of the model.

Other work that considers endogenous regressors in semiparametric or nonparametric nonseparable models includes Altonji and Matzkin (1997), Altonji and Ichimura (1998), Blundell and Powell (1999), and Imbens and Newey (2001).[8] Blundell and Powell (1999) and Imbens and Newey (2001) consider estimation of the average partial effect of a continuous endogenous regressor in nonseparable models, but their identification strategies are not appropriate for a discrete endogenous regressor as considered in this paper. Altonji and Ichimura (1998) consider estimation of the average derivatives of a general class of nonseparable outcome equations with tobit-type censoring of the outcome, but do not consider the effect of an endogenous binary regressor. Altonji and Matzkin (1997) allow for endogenous regressors in a panel data model with exchangeability. See Blundell and Powell (2000) for a survey of this literature.

---

conditions, instrumental variables will consistently estimate LATE even though it will not consistently estimate the average effect of the variable. See Heckman and Vytlacil (2000) for the relationship between the LATE parameter and other mean treatment parameters including the average treatment effect, and see Vytlacil (2002) for the connection between the assumptions imposed in Imbens and Angrist (1994) and the nonparametric selection model. Another alternative is to use the instrumental variables assumption to bound the average treatment effect. See Robins (1989) and Balke and Pearl (1997) for bounds that exploit a statistical independence version of the instrumental variables assumption, and see Manski (1990, 1994) for bounds that exploit a mean independence version of the instrumental variables assumption. See Manski and Pepper (2000) and Heckman and Vytlacil (2001a) for bounds that combine an instrumental variables assumption with additional restrictions.

[8] Work on nonseparable models with exogenous regressors includes Matzkin (1991, 1992, 1993, 1999). There is also a large literature on identification and estimation of the slope parameters of binary choice models without parametric distributional assumptions and while relaxing the independence of the error terms and the regressors to a weaker condition such as median independence (see, e.g., Manski 1975, 1988). This literature recovers the slope parameters of the binary choice models but not the error distribution, and thus cannot answer questions related to the average effect of one of the regressors on the outcome variable.

2 Model:

For any random variable A, let a denote a realization of A, let F_A denote the distribution of A, and let Supp(A) denote the support of A. Let Y denote the outcome variable of interest and D denote the binary endogenous variable of interest. Following Heckman (1978), consider

$$\begin{aligned} Y^* &= X\beta + \alpha D + \epsilon \\ D^* &= Z\gamma + U \\ Y &= 1[Y^* \geq 0] \\ D &= 1[D^* \geq 0] \end{aligned}$$

where (X, Z) is an observed random vector, (ϵ, U) is an unobserved random vector, $1[\cdot]$ is the indicator function, $(X, Z) \perp\!\!\!\perp (\epsilon, U)$, and (ϵ, U) is normally distributed. Heckman (1978) refers to a model of this form as a multivariate probit model with a structural shift. In this model, the average effect of D on Y given covariates X is $F_\epsilon(X\beta + \alpha) - F_\epsilon(X\beta)$. Heckman (1978) develops the maximum likelihood estimator for the model.

This paper examines the more general model where one does not impose parametric distribution assumptions on the error terms, does not impose linear index assumptions, and is for a more general class of outcome equations that include the above threshold crossing model as a special case. In particular, we assume that Y and D are determined by:

$$Y = g(\nu(X, D), \epsilon) \tag{1}$$
$$D = 1[\vartheta(Z) - U \geq 0] \tag{2}$$

where $(X, Z) \in \Re^{K_X} \times \Re^{K_Z}$ is a random vector of other observed covariates, $(\epsilon, U) \in \Re^2$ are unobserved random variables, $g : \Re^2 \mapsto \Re$, and $\nu(\cdot, \cdot) : \Re^{K_X} \times \{0, 1\} \mapsto \Re$. We are assuming that ϵ is a scalar random variable for simplicity, the analysis can be directly extended to allow ϵ to be a random element.[9] We will assume that (X, Z) is exogenous, in particular, that $(X, Z) \perp\!\!\!\perp (\epsilon, U)$. This system of equations includes the classical case discussed above by taking $\vartheta(Z) = Z\delta$, $\nu(X, D) = X\beta + \alpha D$, $g(t, \epsilon) = 1[t + \epsilon \geq 0]$, and (ϵ, U) distributed joint normal. In the following analysis, the functions ν and g need not be known and no parametric distributional assumption will be imposed on (ϵ, U).

The form of the outcome equation for Y is referred to as a generalized regression model by Han (1987), who considered the estimation of $\nu(\cdot)$ when $\nu(\cdot)$ is known up to a finite dimensional parameter vector and all regressors are exogenous.[10] This form of the outcome equation for Y imposes that (X, D) is weakly separable from ϵ. This weak separability restriction will be critical

[9]See Altonji and Ichimura (1998) for related analysis that allows the error term to be a random element. I would like to thank Hide Ichimura for suggesting this point to me.

[10]See also Matzkin (1991), who considers estimation of $\nu(\cdot)$ when curvature restrictions but no parametric assumptions are imposed on $\nu(\cdot)$, and again all regressors are exogenous. Note that this paper differs from Han (1987) and Matzkin (1991) both by allowing for the dummy endogenous variable and by defining the object of interest to be the average effect of the endogenous variable and not recovery of the ν function.

to the following analysis, and makes the model more restrictive than the Roy-model/switching regression framework considered in Heckman (1990a,b). The purpose of this paper is to exploit this weak separability condition to by-pass the identification-at-infinity arguments for identification and estimation which are required for nonparametric switching regression models.[11] However, the results in this paper will directly extend to the switching regression model of $Y = g(\nu(X, D), \epsilon_D)$ with $\epsilon_D = D\epsilon_1 + (1 - D)\epsilon_0$, if one restricts ϵ_1 and ϵ_0 to have the same distribution conditional on U.

The model for D is a threshold-crossing model. Here, $\vartheta(Z) - U$ is interpreted as net utility to the agent from choosing $D = 1$. Without loss of generality, assume that $U \sim \text{Unif}[0, 1]$ and $\vartheta(z) = P(z)$, where $P(z) = \Pr(D = 1|Z = z)$. $P(Z)$ is sometimes called the "propensity score", following Rosenbaum and Rubin (1983).

I will maintain the following assumptions:

(A-1) The distribution of (U, ϵ) is absolutely continuous with respect to Lebesgue measure on $\Re^2$;

(A-2) (U, ϵ) is independent of (Z, X);

(A-3) $g(\nu(X, 1), \epsilon)$ and $g(\nu(X, 0), \epsilon)$ have finite first moments;

(A-4) $E(g(t, \epsilon)|U = u)$ is strictly increasing in t for a.e. u;

(A-5) There exist sets $\mathcal{S}^1_{X,Z}$ and $\mathcal{S}^0_{X,Z}$ with the following properties, where $I_j = \mathbf{1}[(X, Z) \in \mathcal{S}^j_{X,Z}]$,

- **(A-5-a)** $\Pr[I_j = 1] > 0$, $j = 0, 1$.
- **(A-5-b)** $\Pr[0 < P(Z) < 1|I_j = 1] = 1$
- **(A-5-c)** $P(Z)$ is nondegenerate conditional on $(X, I_j = 1)$, $j = 0, 1$.
- **(A-5-d)** $\text{Supp}\big[(\nu(X, 1), P(Z))\big|I_1 = 1\big] \subseteq \text{Supp}\big[(\nu(X, 0), P(Z))\big]$,
  $\text{Supp}\big[(\nu(X, 0), P(Z))\big|I_0 = 1\big] \subseteq \text{Supp}\big[(\nu(X, 1), P(Z))\big]$.

Assumption (A-1) is a regularity condition imposed to guarantee smoothness of the relevant conditional expectation functions. Assumption (A-2) is a critical independence condition, that the observed covariates (besides for the treatment choice) are independent of the unobserved covariates. Assumption (A-3) is a standard regularity condition required to have the parameter of interest be well defined. We will strengthen (A-3) for estimation.

Assumption (A-4) is a monotonicity condition.[12] It is important to note that (A-4) does not require g to be strictly increasing in t, it does not impose any form of monotonicity of g in ϵ, nor does it impose any form of monotonicity on the ν_1, ν_0 functions. One example of a

[11] Heckman and Vytlacil (2001b) prove that the large support conditions imposed in identification-at-infinity arguments are necessary and sufficient for identification of the average treatment effect in general switching regression models.

[12] The following analysis can be trivially extended to the case where $E(g(t, \epsilon)|U = u)$ is strictly decreasing in t for a.e. u.

model which will satisfy (A-4) is the transformation model, where $g(t_0, \epsilon) = r(t_0 + \epsilon)$, and r is a (possibly unknown) strictly increasing function. This model is referred to as a transformation model, and includes as special cases the Box-Cox model and the proportional hazards model with unobserved heterogeneity. Since r is strictly increasing, condition (A-4) is immediately satisfied. However, (A-4) also allows for cases where g is not strictly monotonic in t. An important special case is the threshold crossing models for a binary outcome variable, where $g(t, \epsilon) = \mathbf{1}(\epsilon \leq t)$ so that $E(g(t, \epsilon)|U = u) = \Pr(\epsilon \leq t|U = u)$. If $\text{Supp}(\epsilon, U) = \Re \times [0, 1]$, then condition (A-4) is immediately satisfied, even though g itself is not strictly increasing.

Let $\mathcal{X}^j = \{x : \exists \ z \ \text{ s.t. } (x, z) \in \mathcal{S}^j_{X,Z}\}$, $j = 0, 1$. The analysis for will be done for $x \in \mathcal{X}^j$. Condition (A-5-a) guarantees that these sets have positive probability. Condition (A-5-b) guarantees there are both treated and untreated individuals with positive probability for (almost every) realization of Z within the set. Assumption (A-5-c) requires an exclusion restriction: there exists a variable that enters the decision rule for D but does not directly determine Y. Assumption (A-5-d) is a support condition, which will be discussed at length later in this paper. As will be shown in this paper, (A-5-d) has an empirical analog and it is possible to empirically determine these sets even though they are defined in terms of the ν function.

Our goal is to identify and consistently estimate the average effect of D on Y. Using counterfactual notation, let

$$Y_d = g(\nu(X, d), \epsilon)$$

denote the outcome that would have been observed had an individual with observable vector X and unobservable ϵ been randomly assigned the "treatment" d. In this case, for any measurable set $\mathcal{A} \subseteq \text{Supp}(X)$, we can define the average outcome if all individuals with observed covariates $X \in \mathcal{A}$ had been randomly assigned the treatment $d = 1$,

$$E(Y_1|X \in \mathcal{A}) = E(g(\nu(X, 1), \epsilon)|X \in \mathcal{A}),$$

and the average outcome if all individuals with observed covariates X had been randomly assigned the treatment $d = 0$,[13]

$$E(Y_0|X \in \mathcal{A}) = E(g(\nu(X, 0), \epsilon)|X \in \mathcal{A}).$$

In this notation, the average effect of $D = 1$ versus $D = 0$ is[14]

$$E(Y_1 - Y_0|X \in \mathcal{A}) = E(g(\nu(X, 1), \epsilon) - g(\nu(X, 0), \epsilon)|X \in \mathcal{A}).$$

Within the treatment effect literature, $E(Y_1 - Y_0|X \in \mathcal{A})$ is referred to as the average treatment effect.[15] Another parameter commonly studied in the treatment effect literature is the effect of

[13]Note that, since X is exogenous, the function $\phi(x, d) \equiv E(Y_d|X = x)$ corresponds to the average structural function as defined by Blundell and Powell (1999). From assumption (A-3), we have that $E(Y_1|X \in \mathcal{A})$ and $E(Y_0|X \in \mathcal{A})$ exist and are finite for every set $\mathcal{A}$ such that $\Pr[X \in A] > 0$.

[14]From assumption (A-3), it follows that $E(Y_1 - Y_0|X \in \mathcal{A})$ exists and is finite for every measurable set $\mathcal{A}$ such that $\Pr[X \in \mathcal{A}] > 0$.

[15]See Heckman and Vytlacil (2000) for a discussion of treatment parameters and the connections among them.

treatment on the treated,[16]

$$E(Y_1 - Y_0|D=1,(X,Z)\in\mathcal{B}) = E(g(\nu(x,1),\epsilon) - g(\nu(x,0),\epsilon)|D=1,(X,Z)\in\mathcal{B}),$$

for any measurable set $\mathcal{B} \subseteq \text{Supp}(X,Z)$. This paper will include identification and estimation results for $E(Y_0|X\in\mathcal{A})$, $E(Y_1|X\in\mathcal{A})$, the average treatment effect conditional on covariates, $E(Y_1 - Y_0|X\in\mathcal{A})$, and the effect of treatment on the treated conditional on covariates, $E(Y_1 - Y_0|(X,Z)\in\mathcal{B}, D=1)$.

3 Identification Analysis

In this section I assume that the distribution of (Y,D,X,Z) is known and consider identification of the average effect of the dummy endogenous variable. In particular, I will show identification conditions given that one knows the following functions over the support of (X,Z),[17]

$$\begin{array}{rcl} \Pr[D=1|Z=z] & = & P(z) \\ E(DY|X=x,Z=z) & = & P(z)E(Y_1|D=1,X=x,Z=z) \\ E((1-D)Y|X=x,Z=z) & = & (1-P(z))E(Y_0|D=0,X=x,Z=z). \end{array} \tag{3}$$

We wish to identify the average effect of D on Y given covariates, $E(Y_1 - Y_0|X=x)$, and thus need to identify $E(Y_1|X=x)$ and $E(Y_0|X=x)$. Using equation 1 and that Z is independent of ϵ conditional on X, we have that Y_1, Y_0 are mean independent of Z conditional on X, $E(Y_j|X) = E(Y_j|X,Z)$, $j=0,1$. Thus, applying the law of iterated expectations, we have that

$$E(Y_1|X=x) = P(z)E(Y_1|D=1,X=x,Z=z) + (1-P(z))E(Y_1|D=0,X=x,Z=z),$$

$$E(Y_0|X=x) = P(z)E(Y_0|D=1,X=x,Z=z) + (1-P(z))E(Y_0|D=0,X=x,Z=z).$$

From equation (3), we identify the first term of the first equation and the second term of the second equation but we do not immediately identify the other terms. My analysis will use the model to identify these terms.

To see how the identification analysis will proceed, note that for any version of the conditional expectations that is consistent with our model of equations (1)-(2) and assumptions (A-1)-(A-4),

$$E(Y_1|X=x,Z=z,D=1) = E(g(\nu(x,1),\epsilon)|U\leq P(z)) \tag{4}$$

[16]From assumption (A-3), we have that $E(Y_1 - Y_0|D=1,(X,Z)\in\mathcal{B})$ exists and is finite for all measurable sets $\mathcal{B}$ such that $\Pr((X,Z)\in\mathcal{B}) > 0$.

[17]Throughout the identification section, a statement that a conditional expectation is identified or known is used a shorthand for the more correct statement that the appropriate equivalence class of conditional expectation functions is known. For example, the statement that the function $P(z) = \Pr[D=1|Z=z]$ is known is a shorthand for the statement that the F_Z-equivalence class, $[P] := \{q \in \mathcal{L}^1 : q = P \text{ a.e. } F_Z\}$, is known. In the estimation section, smoothness conditions will be imposed which will imply that the conditional expectations are unique subject to the smoothness conditions, but no such smoothness conditions are imposed here for identification.

$$E(Y_0|X=x,Z=z,D=0)=E(g(\nu(x,0),\epsilon)|U>P(z)), \tag{5}$$

where we have substituted in the models for D and Y and are using the independence assumption (A-2). The problem is to identify

$$E(Y_0|X=x,Z=z,D=1)=E(g(\nu(x,0),\epsilon)|U\leq P(z)), \tag{6}$$

$$E(Y_1|X=x,Z=z,D=0)=E(g(\nu(x,1),\epsilon)|U>P(z)). \tag{7}$$

The central idea for the identification analysis is that if we can find shifts in X which directly compensate for a shift in D, then we can use information from equation 4 to fill in the missing information for equation 6, and from equation 5 to fill in the missing information for equation 7. In particular, if we identify (x,x_1) and (x_0,x) pairs such that $\nu(x,0)=\nu(x_1,1)$ and $\nu(x,1)=\nu(x_0,0)$, then evaluating equation 4 at x_1 tells us the answer for evaluating equation 6 at x, and evaluating equation 5 at x_0 tells us the answer for evaluating equation 7 at x. Because of selection (D being endogenous), we cannot immediately use the conditional expectations in the data to recover such pairs. However, given our model and assumptions, we can use the variation in the conditional expectations for changes in Z to identify such pairs. Given that equations 6 and 7 are identified by this procedure, then (a version of) $E(Y_0|X=x)$, $E(Y_1|X=x)$ and thus $E(Y_1-Y_0|X=x)$ will be identified if the appropriate support condition holds.

For the identification analysis, it will be convenient to work with expectations conditional on $P(Z)$ instead of conditional on Z. Note that, given our assumptions, we have that any version of the conditional expectations that is consistent with our model of equations (1)-(2) and assumptions (A2) and (A3) will satisfy the following index sufficiency restriction,

$$\begin{aligned} E(DY|X=x,Z=z) &= E(DY|X=x,P(Z)=P(z)), \\ E((1-D)Y|X=x,Z=z) &= E((1-D)Y|X=x,P(Z)=P(z)). \end{aligned} \tag{8}$$

Define

$$\begin{aligned} h_1(p_0,p_1,x) &= \frac{1}{p_1-p_0}\Big[E\big(DY\big|X=x,P(Z)=p_1\big)-E\big(DY\big|X=x,P(Z)=p_0\big)\Big] \\ h_0(p_0,p_1,x) &= -\frac{1}{p_1-p_0}\Big[E\big((1-D)Y\big|X=x,P(Z)=p_1\big)-E\big((1-D)Y\big|X=x,P(Z)=p_0\big)\Big]. \end{aligned}$$

One can easily show that

$$h_1(p_0,p_1,x)-h_0(p_0,p_1,x)=\frac{E(Y|X=x,P(Z)=p_1)-E(Y|X=x,P(Z)=p_0)}{p_1-p_0}.$$

This expression is the probability limit of the Wald IV estimator with $P(Z)$ as the instrument shifting from $P(Z)=p_0$ to $P(Z)=p_1$.[18] h_1 and h_0 individually have the form of the probability

[18] This is the form used by Heckman and Vytlacil (1999, 2001a) for the LATE parameter, building on Imbens and Angrist (1994).

limit of the Wald IV estimator applied to DY and $(1-D)Y$ separately. Evaluating $h_1(p_0,p_1,x_1)-h_0(p_0,p_1,x_0)$ with $x_0 \neq x_1$, the difference has a form similar to the Wald IV estimator but shifting X and the instrument simultaneously. We will use the h_1, h_0 functions to uncover (x_0,x_1) pairs such that $\nu(x_1,1)=\nu(x_0,0)$.

Let sgn(t) denote the sign function, defined as follows:

$$\text{sgn}[t] = \begin{cases} 1 & \text{if } t>0 \\ 0 & \text{if } t=0 \\ -1 & \text{if } t<0. \end{cases}$$

We then have the following Lemma.

Lemma 3.1 *Assume that (D,Y) are generated according to equations (1)-(2). Assume conditions (A-1)-(A-5). Then*

$$sgn[h_1(p_0,p_1,x_1)-h_0(p_0,p_1,x_0)] = sgn[\nu(x_1,1)-\nu(x_0,0)].$$

Proof: See Appendix A.

We thus have that $h_1(p_0,p_1,x_1)-h_0(p_0,p_1,x_0)=0$ implies $\nu(x_1,1)=\nu(x_0,0)$. In other words, if $h_1(p_0,p_1,x_1)=h_0(p_0,p_1,x_0)$, then shifting X from x_0 to x_1 directly compensates for shifting D from 0 to 1. Note that if $h_1(p_0,p_1,x_1)-h_0(p_0,p_1,x_0)=0$ for some (p_0,p_1) evaluation points, then $h_1(p_0,p_1,x_1)-h_0(p_0,p_1,x_0)=0$ for all p_0,p_1 evaluation points. Let

$$\begin{aligned} h_1^{-1}h_0(x_0) &= \{x \in \text{Supp}(X) : \exists\ p_0,p_1 \text{ such that } h_1(p_0,p_1,x)=h_0(p_0,p_1,x_0)\} \\ h_0^{-1}h_1(x_1) &= \{x \in \text{Supp}(X) : \exists\ p_0,p_1 \text{ such that } h_1(p_0,p_1,x_1)=h_0(p_0,p_1,x)\}. \end{aligned} \tag{9}$$

From Lemma 3.1, we have that

$$x \in h_1^{-1}h_0(x_0) \Rightarrow \nu(x,1)=\nu(x_0,0)$$

$$x \in h_0^{-1}h_1(x_1) \Rightarrow \nu(x_1,1)=\nu(x,0).$$

There is a support condition required in order to be able to find such pairs – one needs to find enough variation in X to compensate for a shift in D. Recall that $\mathcal{X}^j = \{x : \exists\ z \text{ s.t. } (x,z) \in \mathcal{S}^j_{X,Z}\}$, $j=0,1$. From assumption (A-5-d), we have that, for any $x \in \mathcal{X}^1$, there is enough variation in X to compensate for a shift from $D=1$ to $D=0$. Likewise, for any $x \in \mathcal{X}^0$, there is enough variation in X to compensate for a shift from $D=0$ to $D=1$. In particular, we have that $h_1^{-1}h_0(x_0)$ is nonempty for $x_0 \in \mathcal{X}^0$, and $h_0^{-1}h_1(x_1)$ is nonempty for $x_1 \in \mathcal{X}^1$. We have the following theorem.

Theorem 3.1 *Assume that (D,Y) are generated according to equations (1)-(2). Assume conditions (A-1)-(A-5). Assume that the distribution of (D,Y,X,Z) is known.*

1. *For any* $\mathcal{A} \subset \mathcal{X}^0$, $E(Y_0|X \in \mathcal{A})$ *is identified and given by*

$$E(Y_0|X \in \mathcal{A}) = \int \left[\int \left(E(DY|X \in h_1^{-1}h_0(x), P = p) + E((1-D)Y|X = x, P = p) \right) dG_{P|X}(p|x) \right] dF_{X|\mathcal{A}}(x)$$

where $F_{X|\mathcal{A}}$ *is the distribution function of* X *conditional on* $X \in \mathcal{A}$*, and* $G_{P|X}$ *is any distribution function that is absolutely continuous with respect to the distribution of* $P(Z)$ *conditional on* X*.*

2. *For any* $\mathcal{A} \subset \mathcal{X}^1$, $E(Y_1|X \in \mathcal{A})$ *is identified and given by*

$$E(Y_1|X \in \mathcal{A}) = \int \left[\int \left(E(DY|X = x, P = p) + E((1-D)Y|X \in h_0^{-1}h_1(x), P = p) \right) dG_{P|X}(p|x) \right] dF_{X|\mathcal{A}}(x)$$

where $F_{X|\mathcal{A}}$ *is the distribution function of* X *conditional on* $X \in \mathcal{A}$*, and* $G_{P|X}$ *is any distribution function that is absolutely continuous with respect to the distribution of* $P(Z)$ *conditional on* X*.*

3. *For any* $\mathcal{A} \subset \mathcal{S}^0_{X,Z}$, $E(Y_1 - Y_0|(X,Z) \in \mathcal{A}, D = 1)$ *is identified and given by*

$$\begin{aligned} &E(Y_1 - Y_0|X \in \mathcal{A}, P(Z) = p, D = 1) \\ &\quad = E(Y|(X,Z) \in \mathcal{A}, D = 1) - \int E(Y|X \in h_1^{-1}h_0(x), Z = z, D = 1) dF_{X,Z|\mathcal{A}}(x,z) \end{aligned}$$

where $F_{X,Z|\mathcal{A}}$ *is the distribution function of* (X,Z) *conditional on* $(X,Z) \in \mathcal{A}$*.*

4. *For any* $\mathcal{A} \in \mathcal{X}^0 \bigcap \mathcal{X}^1$, $E(Y_1 - Y_0|X \in \mathcal{A})$ *is identified and given by*

$$\begin{aligned} E(Y_1 - Y_0|X \in \mathcal{A}) = \int \Bigg[\int \Bigg(&E(DY|X = x, P = p) + E((1-D)Y|X \in h_0^{-1}h_1(x), P = p) \\ &- E(DY|X \in h_1^{-1}h_0(x), P = p) - E((1-D)Y|X = x, P = p) \Bigg) dG_{P|X}(p|x) \Bigg] dF_{X|X \in A}(x) \end{aligned}$$

where $F_{X|\mathcal{A}}$ *is the distribution function of* X *conditional on* $X \in \mathcal{A}$*, and* $G_{P|X}$ *is any distribution function that is absolutely continuous with respect to the distribution of* $P(Z)$ *conditional on* X*.*

Proof: See Appendix A.

The requirement that $\mathcal{A} \subseteq \mathcal{X}^j$ involves two types of support conditions. One is that there is sufficient variation in $P(Z)$ conditional on X for $X \in \mathcal{A}$. This requires that there be an exclusion restriction, a variable in Z that is not contained in X. The second, less standard type of support condition is that it is possible to find variation in X that compensates for a change from $D = 0$ to $D = 1$. This support condition is likely to fail near the boundaries of the support of X, as illustrated by the following example.

Illustrative Example: To illustrate the conditions of Theorem 1, take the special case of a threshold-crossing model with linear indices. In particular, assume that the true data generating process is:

$$Y = \mathbf{1}(\epsilon \leq X\beta + \delta D),$$
$$D = \mathbf{1}(V \leq Z\gamma)$$

with (ϵ, V) independent of (X, Z), having a distribution which is absolutely continuous with respect to Lebesgue measure on $\Re^2$, and having support $\Re^2$. We can map the equation for D into the form of equation 2 by taking $U = F_V(V)$. We thus have

$$\begin{aligned} E(DY|X = x, P = p) &= \Pr(V \leq F_V^{-1}(p), \epsilon \leq x\beta + \delta), \\ E((1 - D)Y|X = x, P = p) &= \Pr(V > F_V^{-1}(p), \epsilon \leq x), \end{aligned}$$

and thus

$$\begin{aligned} h_1(p_0, p_1, x) &= \Pr(F_V^{-1}(p_0) < V \leq F_V^{-1}(p_1), \epsilon \leq x\beta + \delta), \\ h_0(p_0, p_1, x) &= \Pr(F_V^{-1}(p_0) < V \leq F_V^{-1}(p_1), \epsilon \leq x\beta). \end{aligned}$$

Suppose that (X, Z) has support equal to the cross product of the support of X and the support of Z, $\text{Supp}(X, Z) = \text{Supp}(X) \times \text{Supp}(Z)$. For simplicity, suppose that the support of $X\beta$ is an interval, $\text{Supp}(X\beta) = [t_L, t_U]$. Then

$$\begin{aligned} h_1^{-1}h_0(x_0) &= \{x \in \text{Supp}(X) : (x_0 - x)\beta = \delta\} \\ h_0^{-1}h_1(x_1) &= \{x \in \text{Supp}(X) : (x - x_1)\beta = \delta\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{X}^1 &= \{x \in \text{Supp}(X) : x\beta + \delta \in [t_L, t_U]\} \\ \mathcal{X}^0 &= \{x \in \text{Supp}(X) : x\beta \in [t_L + \delta, t_U + \delta]\}. \end{aligned}$$

Thus, if $\delta \geq 0$,

$$\mathcal{X}^1 \bigcap \mathcal{X}^0 = \{x \in \text{Supp}(X) : x\beta \in [t_L + \delta, t_U - \delta]\}$$

and if $\delta \leq 0$,

$$\mathcal{X}^1 \bigcap \mathcal{X}^0 = \{x \in \text{Supp}(X) : x\beta \in [t_L - \delta, t_U + \delta]\}.$$

In this example, $E(Y_1 - Y_0|X = x)$ is identified for all $x \in \text{Supp}(X)$ if $\text{Supp}(X\beta)$ is unbounded. If the support of $X\beta$ is bounded, then $E(Y_1 - Y_0|X = x)$ is identified for some x values if $t_U - t_L > 2\delta$. It will not be identified for x values such that $x\beta$ is within δ of the limits of the support of $X\beta$.

The conditional expectations $E(Y_0|X = x)$ and $E(Y_1|X = x)$ need not be identified for all $x \in \text{Supp}(X)$, and thus $E(Y_1 - Y_0|X = x)$ need not be identified for all $x \in \text{Supp}(X)$. In the case where these quantities are not identified for a given x value, we can bound these quantities. We now consider bounds on $E(Y_0|X = x)$. Define

$$\mathcal{U}(x_0) = \{x : \exists\ p_1, p_0, p_1 > p_0, \text{ such that } (x, p_0), (x_0, p_0), (x, p_1), (x_0, p_1) \in \text{Supp}(X, P(Z)), \\ \text{and } h_1(p_0, p_1, x) - h_0(p_0, p_1, x_0) \geq 0\}.$$

$$\mathcal{L}(x_0) = \{x : \exists\ p_1, p_0, p_1 > p_0, \text{ such that } (x, p_0), (x_0, p_0), (x, p_1), (x_0, p_1) \in \text{Supp}(X, P(Z)), \\ \text{and } h_1(p_0, p_1, x) - h_0(p_0, p_1, x_0) \leq 0\}.$$

From Lemma 3.1, we have

$$x \in \mathcal{U}(x_0) \Rightarrow \nu(x, 1) \geq \nu(x_0, 0)$$
$$x \in \mathcal{L}(x_0) \Rightarrow \nu(x, 1) \leq \nu(x_0, 0).$$

We now have the following theorem.

Theorem 3.2 *Assume that (D, Y) are generated according to equations (1)-(2). Assume conditions (A-1)-(A-5). Let $G_{P|x,x_0}(p)$ denote any distribution function which is absolutely continuous with respect to the distribution of $P(Z)$ conditional on $X = x$ and with respect to the distribution of $P(Z)$ conditional on $X = x_0$. If $\mathcal{U}(x_0)$ is nonempty, then $B^U(x_0) \geq E(Y_0|X = x_0)$, with*

$$B^U(x_0) = \inf_{x \in \mathcal{U}(x_0)} \left\{ \int \Big[E((1 - D)Y|X = x_0, P(Z) = p) \\ + E(DY|X = x, P(Z) = p) \Big] dG_{P|x,x_0}(p) \right\}.$$

If $\mathcal{L}(x_0)$ is nonempty, then $B^L(x_0) \leq E(Y_0|X = x_0)$, with

$$B^L(x_0) = \sup_{x \in \mathcal{L}(x_0)} \left\{ \int \Big[E((1 - D)Y|X = x_0, P(Z) = p) \\ + E(DY|X = x, P(Z) = p) \Big] dG_{P|x,x_0}(p) \right\}.$$

Proof: See Appendix A.

Note that the value of $\int \left[E((1-D)Y|X = x_0, P(Z) = p) + E(DY|X = x, P(Z) = p) \right] dG_{P|x,x_0}(p)$ and of $\int \left[E((1-D)Y|X = x_0, P(Z) = p) + E(DY|X = x, P(Z) = p) \right] dG_{P|x,x_0}(p)$ does not depend on the choice of $G_{P|x,x_0}$ distribution as long as the distribution is absolutely continuous with respect to both the distribution of $P(Z)$ conditional on $X = x$ and the distribution of $P(Z)$ conditional on $X = x_0$. When there exists $x \in \mathcal{U}(x_0)$ (or in $\mathcal{U}(x_0)$) such that $h_1(p_0, p_1, x) = h_0(p_0, p_1, x_0)$ for some $p_1 > p_0$, then the bounds collapse to point identification. Similar bounds can easily be constructed for $E(Y_1|X = x)$, $E(Y_1 - Y_0|X = x)$, and $E(Y_1 - Y_0|X = x, D = 1)$.

The selection model considered here is a special case of that considered by Heckman and Vytlacil (1999,2001a), so that their bounds on $E(Y_1 - Y_0|X = x)$ immediately apply to the present model if we assume that the outcome variable Y is bounded.[19] The additional assumptions invoked in this analysis beyond what was assumed in Heckman and Vytlacil (1999,2001a) allows us to construct bounds that do not require Y to be bounded and which collapse to point identification without large support conditions. However, while Heckman and Vytlacil (2001b) establishes that the Heckman and Vytlacil (1999,2001a) bounds are sharp given there assumptions, there is no similar result yet for the bounds of Theorem 3.2. Whether the bounds of Theorem 3.2 can be improved upon thus remains a question for future research.

I conclude the section by considering the testable restrictions imposed by the model. The assumption of a selection model imposes testable restrictions. Heckman and Vytlacil (2001a) consider a model which includes the model of the present paper as a special case, and derive two testable restrictions of the model.

Testable Restriction (1): Index sufficiency,

$$\begin{aligned} \Pr(DY \in \mathcal{A}|X = x, Z = z) &= \Pr(DY \in \mathcal{A}|X = x, P(Z) = P(z)), \\ \Pr((1-D)Y \in \mathcal{A}|X = x, Z = z) &= \Pr((1-D)Y \in \mathcal{A}|X = x, P(Z) = P(z)). \end{aligned}$$

Testable Restriction (2): If $\Pr[Y_1 \geq y_x^1|X = x] = 1$, $\Pr[Y_0 \geq y_x^0|X = x] = 1$, then $E[(Y_0 - y_x^0)(1-D) \mid X, P(Z) = p]$ is decreasing in p and $E[(Y_1 - y_x^1)D \mid X, P(Z) = p]$ is increasing in p.

The model of this paper implies additional testable restrictions. Under conditions (A-1)-(A-5), we have

[19] See Heckman and Vytlacil (2001b) for the relationship between the bounds of Heckman and Vytlacil (1999,2001a) and the instrumental variable bounds of Balke and Pearl (1997) and Manski (1990, 1994).

Testable Restriction (3):

$$\left|\int\int\big(h_1(p_0,p_1,x_1)-h_0(p_0,p_1,x_0)\big)dG(p_0|x_0,x_1)dG(p_1|x_0,x_1)\right|$$
$$=\int\int\Big(\Big|h_1(p_0,p_1,x_1)-h_0(p_0,p_1,x_0)\Big|\Big)dG(p_0|x_0,x_1)dG(p_1|x_0,x_1)$$

where $G(\cdot|x_0,x_1)$ is any distribution function that is absolutely continuous with respect to both the distribution of $P(Z)$ conditional on $X=x_1$ and the distribution of $P(Z)$ conditional on $X=x_0$.

Testable Restriction (4): Define $\mathcal{U}(x),\mathcal{L}(x),B^U(x),B^L(x)$ as in the statement of Theorem 3.2. Let $\mathcal{A}$ denote the set of x values such that both $\mathcal{U}(x)$ and $\mathcal{L}(x)$ are nonempty. Then

$$\inf_{x\in\mathcal{A}}|B^U(x)-B^L(x)|\geq 0.$$

Testable Restriction (3) follows directly from Lemma 3.1, while Testable Restriction (4) follows directly from Theorem 3.2.

4 Estimation

For simplicity, the estimation analysis will proceed under the assumption that Z contains a continuous element not contained in X. Recall that the identification analysis of the previous section does not require this assumption, and note that the following estimation strategy can be adapted for the case where Z contains only discrete elements. For ease of exposition, I only consider estimation of $E(Y_0)$. However, estimation of $E(Y_1)$ is completely symmetric, which in turn implies an estimator for the average treatment effect.

Given that Z contains a continuous element, and given smoothness conditions on $P(Z)$ and $E(Y|X,P(Z),D)$ as functions of Z, we can work with the derivative form of the h_1 and h_0 functions. In particular, let

$$\begin{aligned} h_1(x,p) &= \frac{\partial}{\partial p}E\big(DY\big|X=x,P(Z)=p\big) \\ h_0(x,p) &= -\frac{\partial}{\partial p}E\big((1-D)Y\big|X=x,P(Z)=p\big) \end{aligned}$$

and

$$q(t_1,t_2)=E(Y|D=1,h_1(X,P(Z))=t_1,P(Z)=t_2).$$

Define $h_1^{-1}(t_1;t_2)=\{x:h_1(x,t_2)=t_1\}$, $h_0^{-1}(t_1;t_2)=\{x:h_0(x,t_2)=t_1\}$. For a given t_1,t_2, $x_1\in h_1^{-1}(t_1;t_2)$ and $x_0\in h_0^{-1}(t_1;t_2)$ implies that $x_1\in h_1^{-1}h_0(x_0)$ where $h_1^{-1}h_0(\cdot)$ was defined in equation (9). From the identification analysis of the previous section, we have that

$$\begin{aligned} q(t_1,t_2) &= E(Y_1|D=1,X\in h_1^{-1}(t_1;t_2),P(Z)=t_2) \\ &= E(Y_0|D=1,X\in h_0^{-1}(t_1;t_2),P(Z)=t_2) \end{aligned}$$

Assume that the support of $(h_1(X, P(Z)), P(Z))$ contains the support of $(h_0(X, P(Z)), P(Z))$ so that we can evaluate $q(t_1, t_2)$ at all (t_1, t_2) evaluation points in the support of $(h_0(X, P(Z)), P(Z))$. Let $P_i = P(Z_i)$, $h_{ji} = h_j(X_i, P_i)$, and assume that $\{(X_i, Z_i, D_i, Y_i) : i = 1, ..., N\}$ is an i.i.d sample. The identification analysis then suggests the following infeasible estimator of $E(Y_0)$,

$$\hat{\Delta} = \frac{1}{N} \sum_i \Big[(1 - D_i) Y_i + D_i q(h_{0i}, P_i) \Big].$$

Theorem 4.1 *Assume conditions (A-1)-(A-5). Assume that* $\{X_i, Z_i, D_i, Y_i : i = 1, ..., N\}$ *is i.i.d, that* Y_0 *has a positive, finite second moment, and that the support of* $(h_1(X, P(Z)), P(Z))$ *contains the support of* $(h_0(X, P(Z)), P(Z))$*. Then*

$$\sqrt{N} \left(\frac{\hat{\Delta} - \Delta}{\sqrt{V}} \right) \xrightarrow{d} N(0, 1),$$

where

$$\begin{aligned} \Delta &= E(Y_0) \\ V &= Var\Big[E\Big(Y_0 \Big| X, P, D \Big) \Big] + E\Big[(1 - P)\, Var\Big(Y_0 \Big| X, P, D = 0 \Big) \Big] \end{aligned}$$

Proof: From Theorem 3.1, we have that $q(h_{0i}, P_i) = E(Y_0 | D = 1, X = X_i, P(Z) = P_i)$. The theorem then follows from applying the Central Limit Theorem for i.i.d. data with a positive, finite second moment.

The estimator has the form of an imputation based estimator, with the value of Y_0 for those with $D = 1$ being imputed. The form is reminiscent of a matching estimator (see, e.g., Heckman, Ichimura, and Todd, 1998, and Hahn,1998). However, the underlying assumptions of the matching estimator is different from those assumptions imposed here, and the form of the imputation is quite different as a result. If $D_i = 1$, then the matching estimator uses $E(Y_0 | D = 0, X = X_i)$ to impute Y_{0i}. The missing Y_{0i} information for $D_i = 1$ observations is filled in using $Y_{0i'}$ data from $D_{i'} = 0$ observations that have (approximately) the same value of X. In contrast, the estimator proposed here fills in the missing Y_{0i} information for $D_i = 1$ observations using $Y_{1i'}$ information from $D_{i'} = 1$ observations that have different values of X, with the different value of X chosen in a way to compensate for the effect of D. These very different imputation procedures are driven by the difference in the underlying assumptions.

The above estimator would be feasible if the functions $P(\cdot)$, $h_1(\cdot, \cdot)$, $h_0(\cdot, \cdot)$, and $E(Y | D = 1, h_1(X, P(Z)) = \cdot, P(Z) = \cdot)$ were known. They are not known, which suggests using a two step semiparametric estimator where these unknown functions are replaced by consistent, nonparametric estimates. In addition, trimming is needed in practice for two reasons. First, to get uniformly consistent estimates for P, h_0 and h_1 functions, we have to trim out those observations of (X_i, Z_i) for which the value of the density $f_{X,Z}$ is low. Second, we have assumed thus far that the support of $(h_1(X, P(Z)), P(Z))$ contains the support of $(h_0(X_i, P(Z_i)), P(Z_i))$, but this is not a realistic assumption. Thus, we need to trim out those observations for which $f_{h_1,P}$ evaluated at $(h_1(X_i, P(Z_i)), P(Z_i))$ is low. Let the two trimming functions be denoted by $I_{1i} = 1\{f_{X,Z}(X_i, Z_i) \geq q_{01}\}$ and $I_{2i} = 1\{f_{h_1,P}(h_{0i}, P_i) \geq q_{02}\}$, where $q_{01}, q_{02} > 0$. These trimming functions are not known since the corresponding densities are not known, and thus these trimming functions must also be estimated. Thus, consider

$$\tilde{\Delta} = \frac{\frac{1}{N}\sum_i \Big[(1-D_i)Y_i + D_i\hat{q}(\hat{h}_{0i}, \hat{P}_i)\Big]\hat{I}_{1i}\hat{I}_{2i}}{\frac{1}{N}\sum_i \hat{I}_{1i}\hat{I}_{2i}}$$

where $\hat{P}_i = \hat{P}(Z_i)$ with $\hat{P}(\cdot)$ a consistent nonparametric estimator of $P(\cdot)$, and so forth, and $\hat{I}_{1i} = 1\{\hat{f}_{X,Z}(X_i, Z_i) \geq q_{01}\}$, $\hat{I}_{2i} = 1\{\hat{f}_{\hat{h}_1,\hat{P}}(\hat{h}_{0i}, \hat{P}_i) \geq q_{02}\}$. In current work, I am deriving the asymptotic distribution of this estimator when local polynomial regression estimators are used in a first step for these unknown conditional expectations functions. See Appendix B. Note that these estimation results are not finished. The preliminary results are that under regularity conditions

(i) $\hat{P}(z)$ is asymptotically linear with trimming:

$$\big[\hat{P}(z) - P(z)\big]\hat{I}_1(x,z) = \frac{1}{n}\sum_{j=1}^{n} \psi_{nP}(D_i, X_i, Z_i; x, z) + \hat{b}_P(z) + \hat{R}_P(z)$$

where $n^{-1/2}\sum_{i=1}^n \hat{R}_P(X_i, Z_i) = o_p(1)$, $plim_{n\to\infty} n^{-1/2}\sum_{i=1}^n \hat{b}_P(X_i, Z_i) = b_P < \infty$, $E[\psi_{nP}(D_i, X_i, Z_i; X, Z|X = x, Z = z] = 0$.

(ii) $\hat{h}_0(x, \hat{P}(z))$ is asymptotically linear with trimming:

$$\begin{aligned}\Big[\hat{h}_0(x,\hat{P}(z),x) - h_0(x,P(z),x)\Big]\hat{I}_1(x,z) &= N^{-1}\sum_{j=1}^N\Big[\psi_{Nh_0}(-(1-D_j)Y_j, X_j, P(Z_j); x, z) + \tfrac{\partial h_0(x,P(z))}{\partial p}\psi_{nP}(D_j, X_j, Z_j; x, z)\Big] \\ &+ \hat{b}_{\hat{h}_0}(x,z) + \hat{R}_{\hat{h}_0}(x,z)\end{aligned}$$

with $\text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{j=1}^N \hat{b}_{\hat{h}_0}(X_j, Z_j) = b_{h_0} + b_{h_0P} < \infty$, $\text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{j=1}^N \hat{R}_{\hat{h}_0}(X_j, Z_j) = 0$,

(iii) $\frac{1}{\sqrt{N}}\sum_j D_j\big[\hat{q}(\hat{h}_{0j}, \hat{P}_j) - q(h_{0j}, P_j)\big]\hat{I}_{1j}\hat{I}_{2j}$ is asymptotically equivalent to

$$\frac{1}{\sqrt{N}N}\sum_{j=1}^N\sum_{i=1}^N \frac{D_j}{P(Z_j)}\psi_{Nq}(Y_i, h_{1i}, P_i; X_j, Z_j, P_j, h_{0j})I_{2j} + b_q$$

(iv) $\sqrt{N}[\tilde{\Delta} - E(Y_0|A_1 \cap A_2)]$ is asymptotically equivalent to

$$\left[\frac{1}{N}\sum_i I_{1i}I_{2i}\right]^{-1} \times \left(\frac{1}{\sqrt{N}}\sum_i E\left[D_jI_{2j}\psi(D_i, Y_i, X_i, Z_i; X_j, Z_j)\middle| Y_i, D_i, X_i, Z_i\right] + b \right.$$
$$\left. + \frac{1}{\sqrt{N}}\sum_i\Big[(1-D_i)Y_i + D_iq(h_{0i}, P_i) - E(Y_0|A_1\cap A_2)\Big]I_{1i}I_{2i}\right)$$

where

$$A_1 = \{(x,z) \in \text{supp}(X,Z) : f_{X,Z}(x,z) \geq q_{01}\},$$
$$A_2 = \{(x,z) \in \text{supp}(X,Z) : f_{h_1,P}(h_0(x,P(z)),P(z)) \geq q_{02}\},$$
$$b = b_{qP} + b_{qh_0} + b_{qh_0P} + b_q,$$

and

$$\psi(D_i,Y_i,X_i,Z_i;X_j,Z_j) = \frac{\partial q}{\partial P}(h_{0j},P_j)\psi_{NP}(D_i,Y_i,X_i,Z_i;X_j,Z_j) + \frac{\partial q}{\partial h_1}(h_{0j},P_j)\psi_{Nh_0P}(D_i,Y_i,X_i,Z_i;X_j,Z_j) + \frac{1}{P(Z_j)}\psi_{Nq}(Y_i,h_{1i},P_i;X_j,Z_j,P_j,h_{0j}),$$

with

$$\psi_{Nh_0P}(D_j,Y_j,X_j,Z_j;x,z) := \psi_{Nh_0}(-(1-D_j)Y_j,P(Z_j),X_j;P(z),x,z) + \frac{\partial h_0(P(z),x)}{\partial p}\psi_{NP}(D_j,X_j,Z_j;x,z).$$

The main argument and the regularity conditions are presented in Appendix B. Result (i), that $\hat{P}(z)$ is asymptotically linear with trimming under regularity conditions, is proven in Appendix C.1. Result (ii), that $\hat{h}_0(x,\hat{P}(z))$ is asymptotically linear with trimming under regularity conditions, is proven in Appendix C.2. Result (iii), that $\frac{1}{\sqrt{N}}\sum_j D_j\left[\hat{q}(\hat{h}_{0j},\hat{P}_j) - q(h_{0j},P_j)\right]\hat{I}_{1j}\hat{I}_{2j}$ is asymptotically equivalent to a particular sum, is proven in Appendix C.3. Results related to the trimming functions are collected in Appendix C.4. Appendix B combines these results to show result (iv), which in turn establishes the asymptotic distribution theory for my estimator. These results are still preliminary and incomplete.

5 Conclusion

This paper has shown identification and a consistent estimator of the average effect of a dummy endogenous variable in a nonparametric, nonseparable model. While the paper has only considered nonparametric identification and estimation of treatment effects, the results are promising for identification more generally in models with dummy endogenous variables. For example, the results can easily be extended to identification and estimation of the structural parameters of semiparametric models with dummy endogenous variables. As another example, the analysis of this paper can be immediately applied to identify state dependence in panel data models with binary outcomes as long as there is a time-varying continuous regressor and the lagged dependent variables do not have random coefficients associated with them.

References

[1] Abadie, A., 1999, "Semiparametric Estimation of Instrumental Variable Models for Causal Effects," unpublished working paper, Massachusetts Institute of Technology.

[2] Altonji, J. and H. Ichimura, 1998, "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables," unpublished manuscript, Northwestern University and University College London.

[3] Altonji, J. and R. Matzkin, 1998, "Panel Data Estimators for Nonseparable Models with Endogenous Regressors," unpublished manuscript, Northwestern University.

[4] Amemiya, T., 1978, "The Estimation of a Simultaneous Equation Generalized Probit Model," *Econometrica*, **46**, 1193-1205.

[5] Andrews, D., and M. Schafgans, 1998, "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, **65**, 497-518.

[6] Angrist, J., 1991, "Instrumental Variables Estimation of Average Treatment Effects in Econometrics and Epidemiology," NBER Technical Working Paper No. 115.

[7] _____, 2001, "Estimation of Limited-Dependent Variable Models with Binary Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics.*

[8] Balke, A., and Pearl, J., 1997, "Bounds on Treatment Effects From Studies with Imperfect Compliance," *Journal of the American Statistical Association,* **92**, 1171-1176.

[9] Bhattacharya, J., D. McCaffrey, and D. Goldman, 1999, "Estimating Probit Models with Endogenous Covariates," unpublished working paper, RAND.

[10] Blundell, R. and J. Powell, 1999, "Endogeneity in Single Index Models," unpublished working paper, University College London and UC-Berkeley.

[11] Blundell, R. and J. Powell, 2000, "Endogeneity in Nonparametric and Semiparametric Regression Models," unpublished working paper presented at the World Conference of the Econometric Society.

[12] Blundell, R. and R. Smith, 1989, "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, **56**, 37-58.

[13] Das, M., W. Newey, and F. Vella, 2000, "Non-parametric Estimation of the Sample Selection Model," unpublished working paper, Columbia University, Massachusetts Institute of Technology, and Rutgers.

[14] Hahn, J., 1998, "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, **66**, 315-331.

[15] Han, A. K., 1987, "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics,* **35**, 303-316.

[16] Heckman, J., 1978, "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica* , **46**, 931-959.

[17] _____, 1990a, "Varieties of Selection Bias," *American Economic Review*, **80**, 313-318.

[18] _____, 1990b, "Alternative Approaches to the Evaluation of Social Programs," Barcelona Lecture, World Conference of the Econometric Society.

[19] _____, 1997, "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, **32**, 441-462.

[20] Heckman, J., H. Ichimura, and P. Todd, 1998, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies,*, **65**, 261-294.

[21] Heckman, J., and R. Robb, 1985, "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*, (New York: Cambridge University Press), 156-245.

[22] Heckman, J., and E. Vytlacil, 1999, "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences,* **96**, 4730-4734.

[23] _____, 2000, "The Relationship Between Treatment Parameters within a Latent Variable Framework," *Economics Letters,* January 2000, 66(1): 33-39.

[24] _____, 2001a, "Local Instrumental Variables" with J. Heckman, in C. Hsiao, K. Morimune, and J. Powell, eds., *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, (Cambridge: Cambridge University Press), 1-46.

[25] _____, 2001b, "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect," in M. Lechner and F. Pfeiffer, eds., *Econometric Evaluations of Active Labor Market Policies in Europe,* (Heidelberg; New York: Physica-Verlag), 1-23.

[26] Imbens, G., and J. Angrist, 1994, "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, **62**, 467-476.

[27] Imbens, G., and W. Newey, 2001, "Identification and Inference in Triangular Simultaneous Equation Models without Additivity", unpublished working paper, UCLA and MIT.

[28] Manski, C., 1975, "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics,* **3**, 205-28.

[29] ——, (1988), "Identification of Binary Response Models Source," *Journal of the American Statistical Association* 83:403, 729-38

[30] ——, 1990, "Nonparametric Bounds on Treatment Effects," *American Economic Review, Papers and Proceedings,* **80**, 319-323.

[31] ——, 1994, "The Selection Problem," in C. Sims, ed., *Advances in Econometrics: Sixth World Congress*, (Cambridge: Cambridge University Press), 143-170.

[32] Manski, C. and J. Pepper, 2000, "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica,* **68**, 997-1012.

[33] Matzkin, R., 1991, "A Nonparametric Maximum Rank Correlation Estimator," in W. A. Barnett, J. L. Powell, and G. E. Tauchen, eds., *Nonparametric and Semiparametric Methods in Economics and Statistics,* (Cambridge University Press: Cambridge).

[34] Matzkin, R., 1992, "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica* 60:2, 239-70

[35] Matzkin, R., 1993, "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica* 59:5, 1315-27

[36] Newey, W., 1986, "Linear Instrumental Variable Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Journal of Econometrics*, 32, 127-141.

[37] Newey, W., 1994, "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, **62**, 1349-1382.

[38] Rivers, D., and Q. Vuong, 1988, "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics,* **39**, 347-366.

[39] Robins, J., 1989, "The Analysis of Randomized and Non-randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies", in L. Sechrest, H. Freeman and A. Mulley, eds., *Health service Research Methodology: A Focus on AIDS* (U.S. Public Health Service, Washington, DC), 113-159.

[40] Rosenbaum, P., and D. Rubin, 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, **70**, 41-55.

[41] Schafgans, M., 2000, "Finite Sample Properties for the Semiparametric Estimation of the Intercept of a Censored Regression Model," unpublished working paper, London School of Economics.

[42] Schafgans, M., and V. Zinde-Walsh, 2000, "On Intercept Estimation in the Sample Selection Model," unpublished working paper, London School of Economics.

[43] Vytlacil, E., 2002, "Independence, Monotonicity, and Latent Variable Models: An Equivalence Result," forthcoming, *Econometrica*.

A Identification Proofs

Proof. (Lemma 3.1)

Consider the case where $p_1 > p_0$. (the case where $p_0 < p_1$ is symmetric). Consider the numerator of $h_1(p_0, p_1, x_1)$,

$$\begin{aligned} &E(DY|X = x_1, P(Z) = p_1) - E(DY|X = x_1, P(Z) = p_0) \\ &\quad = \int_0^{p_1} E(Y_1|X = x_1, U = u)du - \int_0^{p_0} E(Y_1|X = x_1, U = u)du \\ &\quad = \int_{p_0}^{p_1} E(Y_1|X = x_1, U = u)du \\ &\quad = \int_{p_0}^{p_1} E(g(\nu(x_1, 1), \epsilon)|U = u)du, \end{aligned}$$

where the last equality is using assumption (A-2). Likewise, for the numerator of $h_0(p_0, p_1, x_0)$, we have

$$\begin{aligned} &-\Big[E((1-D)Y|X = x_0, P(Z) = p_1) - E((1-D)Y|X = x_0, P(Z) = p_0)\Big] \\ &\quad = \int_{p_0}^{p_1} E(g(\nu(x_0, 0)\epsilon)|U = u)du. \end{aligned}$$

Thus,

$$h_1(p_0, p_1, x_1) - h_0(p_0, p_1, x_0) = \frac{1}{p_1 - p_0}\int_{p_0}^{p_1} E(g(\nu(x_1, 1), \epsilon) - g(\nu(x_0, 0), \epsilon)|U = u)du.$$

Using assumption (A-4), we have that the sign of this expression will be determined by the sign of $\nu(x_1, 0) - \nu(x_0, 1)$. Q.E.D..

Proof: (Lemma 3.1) Consider assertion (1). By Lemma 3.1, $\nu(\tilde{x}, 1) = \nu(x, 0)$ for any $\tilde{x} \in h_1^{-1}h_0(x)$. Thus,

$$\begin{aligned} &E(DY|X \in h_1^{-1}h_0(x), P(Z) = p) \\ &\quad = E(\mathbf{1}[U \leq P(Z)]g(\nu(X, 1), \epsilon)|X \in h_1^{-1}h_0(x), P(Z) = p) \\ &\quad = \int\left[\int \mathbf{1}[U \leq p]g(\nu(\tilde{x}, 1), \epsilon)dG(\tilde{x}|X \in h_1^{-1}h_0(x), P = p))\right]dF_{\epsilon,U} \\ &\quad = \int\left[\int \mathbf{1}[U \leq p]g(\nu(x, 0), \epsilon)dG(\tilde{x}|X \in h_1^{-1}h_0(x), P(Z) = p)\right]dF_{\epsilon,U} \\ &\quad = \int \mathbf{1}[U \leq p]g(\nu(x, 0), \epsilon)dF_{\epsilon,U} \\ &\quad = E(DY_0|X = x, P = p) \end{aligned}$$

where $G(\tilde{x}|X \in h_1^{-1}h_0(x), P = p)$ is the distribution of X conditional on $X \in h_1^{-1}h_0(x), P = p$, and $F_{\epsilon,U}$ is the distribution of (ϵ, U). The first equality follows from plugging in the model for Y and D given by equations (1) and (2); the second equality follows from assumption (A-2), that $(X, Z) \perp\!\!\!\perp (\epsilon, U)$; and the third equality is using that $\nu(\tilde{x}, 1) = \nu(x, 0)$ for any $\tilde{x} \in h_1^{-1}h_0(x)$ by Lemma 3.1. Thus,

$$
\begin{aligned}
&E(DY|X \in h_1^{-1}h_0(x), P = p) + E((1-D)Y|X = x, P = p) \\
&\qquad = E(DY_0|X = x, P = p) + E((1-D)Y_0|X = x, P = p) \\
&\qquad = E(Y_0|X = x, P = p) \\
&\qquad = E(Y_0|X = x),
\end{aligned}
$$

so that

$$
\begin{aligned}
&\int \Big(E(DY|X \in h_1^{-1}h_0(x), P = p) + E((1-D)Y|X = x, P = p) \Big) dG_{P|X}(p|x) \\
&\qquad = \int E(Y_0|X = x) dG_{P|X}(p|x) \\
&\qquad = E(Y_0|X = x)
\end{aligned}
$$

and the result now follows immediately. Assertions (2) and (3) follow from analogous arguments, and assertion (4) follows from assertions (1) and (2). QED.

Proof: (Lemma 3.2) For any x such that $h_1(p_0, p_1, x) - h_0(p_0, p_1, x_0) \geq 0$, we have $\nu(x, 1) \geq \nu(x_0, 0)$ by Lemma 3.1. Thus, for any $x \in \mathcal{U}(x_0)$,

$$
\begin{aligned}
&E((1-D)Y|X = x_0, P(Z) = p) + E(DY|X = x, P(Z) = p) \\
&= E((1-D)Y_0|X = x_0, P(Z) = p) + E(1[U \leq P(Z)]g(\nu(X, 1), \epsilon)|X = x, P(Z) = p) \\
&= E((1-D)Y_0|X = x_0, P(Z) = p) + E(1[U \leq p]g(\nu(x, 1), \epsilon)) \\
&\geq E((1-D)Y_0|X = x_0, P(Z) = p) + E(1[U \leq p]g(\nu(x_0, 0), \epsilon)) \\
&= E((1-D)Y_0|X = x_0, P(Z) = p) + E(DY_0|X = x_0, P(Z) = p) \\
&= E(Y_0|X = x_0, P(Z) = p) \\
&= E(Y_0|X = x_0)
\end{aligned}
$$

and thus

$$
\inf_{x \in \mathcal{U}(x_0)} \Big\{ E((1-D)Y|X = x_0, P(Z) = p) + E(DY|X = x, P(Z) = p) \Big\} \geq E(Y_0|X = x_0).
$$

The lower bound follows from the analogous argument. QED.

B Estimation Proofs: Main Results (PRELIMINARY AND INCOMPLETE)

$$\begin{aligned}
\sqrt{N}(\tilde{\Delta} - E(Y_0|A_1 \cap A_2)) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{\big[(1-D_i)Y_i + D_i \hat{q}(\hat{h}_{0i}, \hat{P}_i)\big] \hat{I}_{1i}\hat{I}_{2i}}{N^{-1}\sum_{i=1}^{n} \hat{I}_{1i}\hat{I}_{2i}} - E(Y_0|(X,Z) \in A_1 \cap A_2) \right] \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\big[(1-D_i)Y_i + D_i \hat{q}(h_{0i}, P_i) - E(Y_0|(X,Z) \in A_1 \cap A_2)\big] \hat{I}_{1i}\hat{I}_{2i}}{N^{-1}\sum_{i=1}^{n} \hat{I}_{1i}\hat{I}_{2i}}
\end{aligned}$$

where

$$\begin{aligned}
\hat{P}(z) &= E(\widehat{D|Z} = z) \\
P(z) &= E(D|Z = z) \\
\hat{h}_0(x, \hat{P}(z)) &= \frac{\partial}{\partial P} E\big[-(1-D)\widehat{Y|X} = x, P(Z) = \hat{P}(z)\big] \\
h_0(x, P(z)) &= \frac{\partial}{\partial P} E\big[-(1-D)Y|X = x, P(Z) = P(z)\big] \\
h_1(x, P(z)) &= \frac{\partial}{\partial P} E\big[DY|X = x, P(Z) = P(z)\big] \\
\hat{q}(\hat{h}_0(x, \hat{P}(z)), \hat{P}(z)) &= E(Y|D = 1, h_1(X, \widehat{P(Z)) =} \hat{h}_0(x, \hat{P}(z)), P(Z) = \hat{P}(z)) \\
q(h_0(x, P(z)), P(z)) &= E(Y|D = 1, h_1(X, P(Z)) = h_0(x, P(z)), P(Z) = P(z))
\end{aligned}$$

and

$$\begin{aligned}
\hat{I}_{1i} &:= 1\left\{\hat{f}_{X,Z}(X_i, Z_i) \geq q_{01}\right\} \\
I_{1i} &:= 1\left\{f_{X,Z}(X_i, Z_i) \geq q_{01}\right\} \\
\hat{I}_{2i} &:= 1\left\{\hat{f}_{\hat{h}_1(X,\hat{P}(Z)),\hat{P}(Z)}(\hat{h}_0(X_i, \hat{P}(Z_i)), \hat{P}(Z_i)) \geq q_{02}\right\} \\
I_{2i} &:= 1\left\{f_{h_1(X,P(Z)),P(Z)}(h_0(X_i, P(Z_i)), P(Z_i)) \geq q_{02}\right\} \\
A_1 &:= \left\{(x,z) \in \text{supp}(X,Z) : f_{X,Z}(x,z) \geq q_{01}\right\} \\
A_2 &:= \left\{(x,z) \in \text{supp}(X,Z) : f_{h_1(X,P(Z)),P(Z)}(h_0(x, P(z)), P(z)) \geq q_{02}\right\}
\end{aligned}$$

with kernel density estimators used for density estimation and local polynomial regression used for estimation of the conditional expectation functions. To study the asymptotic properties of our estimator, we break it into several pieces and study the behavior of each piece separately. In analyzing the behavior of each piece we will rely on the results stated in Heckman, Ichimura and Todd (1998) extensively. On the other hand, our estimator embodies trimming functions, which rely on estimators for the underlying densities. In the following, we will rely on a theorem stated in Silverman (1978) to argue that the kernel density estimators that the trimming functions are based on approach the true density in the sup norm. Before starting our analysis, let us state the equicontinuity and Hoeffding, Powell, Stock and Stoker lemmas used in Heckman, Ichimura and Todd (1998) and Theorem A of Silverman (1978).

To set up the notation for Silverman's theorem suppose g_N is the kernel estimate for the multivariate density g defined by

$$g_N(x) = \sum_{i=1}^{N} N^{-1} h^{-\tilde{d}} \tilde{K} \left(\frac{X_i - x}{\tilde{h}} \right)$$

The following conditions are used in Silverman's theorems:

(C): (a) $\tilde{K}$ is uniformly continuous (with modulus of continuity $w_{\tilde{K}}$) and of bounded variation $V(\tilde{K})$

(b) $\int |\tilde{K}(x)| \mathrm{d}x < \infty$ and $\tilde{K}(x) \to 0$ as $||x|| \to \infty$

(c) $\int \tilde{K}(x) \mathrm{d}x = 1$

(d) $\int \sqrt{||(x \log ||x||)||} |\mathrm{d}\tilde{K}(x)| < \infty$

Theorem (A): Suppose $\tilde{K}$ satisfies conditions (C) and g is uniformly continuous. Suppose $\tilde{h} \to 0$ and $(N\tilde{h}^{\tilde{d}})^{-1} \log N \to 0$ as $N \to \infty$. Then, defining g_N as above,

$$\sup |g_N - g| \to 0 \;\; a.s. \qquad N \to \infty$$

If $\tilde{K}$ is everywhere differentiable and G_N denotes the empirical measure

$$\tilde{h} \frac{\partial g_N}{\partial x_k}(x) = -\int \tilde{h}^{\tilde{d}} \tilde{K}' \left(\frac{t - x}{\tilde{h}} \right) \mathrm{d}G_N(t)$$

has the same structure as g_n with $\tilde{K}$ replaced by $\tilde{K}'$. Therefore, as long as g has uniformly continuous partial derivatives and conditions (C)(a), (b) and (d) are satisfied by $\tilde{K}'$, and $(N\tilde{h}^{\tilde{d}+1})^{-1} \log N \to 0$ as $N \to \infty$, for each $k \in \{1, ..., \tilde{d}\}$, we have

$$\sup \left| \frac{\partial g_N}{\partial x_k}(x) - \frac{\partial g}{\partial x_k}(x) \right| \to 0 \;\; a.s. \qquad N \to \infty$$

To state the two lemmas from Heckman, Ichimura and Todd (1998) we need to define some notation: For $r = 1$ and 2, let $\mathcal{S}^r$ denote the r-fold product space of $\mathcal{S} \subset \mathbb{R}^d$ and define a class of functions Λ_N over $\mathcal{S}^r$. For any $\lambda_N \in \Lambda_N$, write λ_{N,i_r} as a short hand for either $\lambda_N(s_i)$ or $\lambda_N(s_{i_1}, s_{i_2})$, where $i_1 \neq i_2$. We define $U_N \lambda_N = \sum_{i_r} \lambda_{N,i_r}$, where $\sum_{i_r}$ denotes the summation over all permutations of r elements of $\{s_1, ..., s_N\}$ for $r = 1$ or 2. Then $U_N \lambda_N$ is called a U-process over $\lambda_N \in \Lambda_N$. For $r = 2$, we assume that $\lambda_N(S_i, S_j) = \lambda_N(S_j, S_i)$. Note that a normalizing constant might be included as a part of λ_N. We call a U-process degenerate if all conditional expectations given other elements are 0. When $r = 1$, this condition is defined to mean that $E\lambda_N = 0$.

In the following, we assume that Λ_N is a subset of $\mathcal{L}^2(\mathbb{P}^r)$, the $\mathcal{L}^2$ space defined over $\mathcal{S}^r$ using the product measure of $\mathbb{P}$, $\mathbb{P}^r$. $N_2(\tau, \mathbb{P}, \Lambda_N)$ denotes the $\mathcal{L}^2$ covering number of Λ_N. On the other hand, $||\lambda_N||_2 := \sqrt{\sum_{i_r} E(\lambda_{N,i_r})^2}$.

Equicontinuity Lemma: Let $\{S_i\}_{i=1}^N$ be an iid sequence of random variables generated by $\mathbb{P}$. For a degenerate U-process $\{U_N\lambda_N\}$ over a separable class of functions $\Lambda_N \subset \mathcal{L}^2(\mathbb{P}^r)$ suppose the following assumptions hold:

(i) There exists an $F_N \in \mathcal{L}^2(\mathbb{P}^r)$ such that for any $\lambda_N \in \Lambda_N$, $|\lambda_N| < F_N$ such that $\limsup_{N\to\infty} \sum_{i_r} E(F^2_{N,i_r}) < \infty$;

(ii) For each $\delta > 0$, $\lim_{N\to\infty} \sum_{i_r} E(F^2_{N,i_r} 1\{F_{N,i_r} > \delta\}) = 0$;

(iii) There exists $\alpha(\tau)$ and $\overline{\tau} > 0$ such that for each $0 < \tau \leq \overline{\tau}$, $\sup_{\mathbb{P}} N_2(\tau, \mathbb{P}, \Lambda_n) \leq \alpha(\tau)$ and $\int_0^{\overline{\tau}} [\log \alpha(t)]^{r/2} \mathrm{d}t < \infty$.

Then for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\lim_{N\to\infty} P\left(\sup_{||\lambda_{1N} - \lambda_{2n}||_2 \leq \delta} |U_N(\lambda_{1N} - \lambda_{2N})| > \epsilon \right) = 0$$

Hoeffding, Powell, Stock and Stoker Lemma: Suppose $\{S_i\}_{i=1}^N$ is i.i.d., $U_N\lambda_N = (N(N-1))^{-1} \sum_{i_r} \lambda_N(S_i, S_j)$ where λ_N is symmetric in its arguments, $E[\lambda_N(S_i, S_j)] = 0$, and $\hat{U}_N\lambda_N = N^{-1} \sum_{i=1}^N 2p_N(S_i)$, with $p_N(S_i) = E[\lambda_N(S_i, S_j)|S_i]$. If $E[\lambda_N(S_i, S_j)^2] = o(N)$, then $NE[(U_N\lambda_N - \hat{U}_N\lambda_N)^2] = o(1)$.

Now we are ready to state our assumptions. Suppose $\{\tilde{h}_{N1}\}$, $\tilde{K}_1$, $\{\tilde{h}_{N2}\}$ and $\tilde{K}_2$ denote the bandwidth parameter sequence and kernel function used to estimate $f_{X,Z}$ and $f_{h_1,P}$, respectively. Similarly, let $\{h_{NP}\}$, $\{h_{Nh}\}$ and $\{h_{Nq}\}$ and K^P, K^h and K^q denote the bandwidth sequences and kernel functions used in estimating, $P(Z)$, h_0 (h_1)[20] and q, respectively. We will call a function p-smooth if it is $p+1$ times continuously differentiable and its $p+1^{st}$ derivative is Holder continuous with parameter $0 < a \leq 1$[21].

Assumption B.1 *$\{D_i, Y_i, X_i, Z_i\}$ are i.i.d., (X_i, Z_i) takes values in $\mathbb{R}^{dx} \times \mathbb{R}^{dz} = \mathbb{R}^d$, and $var(Y_i) < \infty$*

Assumptions related to the estimation of $f_{X,Z}$ and $f_{h_1,P}$:

[20] We can use the same kernel function and bandwidth sequence in the estimation of h_0 and h_1.

[21] We use the same definiton as in Heckman, Ichimura and Todd. Namely, we say a function ϱ is Holder continuous at $X = x_0$ with constant $0 < a \leq 1$ if $|\varrho(x,t) - \varrho(x_0,t)| \leq C||x - x_0||^a$ for some $C > 0$ for all x and t in the domain of the function $\varrho(\cdot,\cdot)$. We assume that Holder continuity holds uniformly over t whenever there is an additional argument.

Assumption B.2 *(a)* $f_{X,Z}$ *and* $f_{h_1,P}$ *are both uniformly continuous and have uniformly continuous first derivatives.*

(b) $f_{X,Z}$, $P(Z)$, $h_1(X,P(Z))$, $h_0(X,P(Z))$ *and* $f_{h_1,P}$ *are all* $\tilde{p}_1$*-smooth with* $\tilde{p}_1 > d$.

(c) Let $q_{01} > 0$ *and* $q_{02} > 0$ *be such that*

(i) q_{01} *has a neighborhood* U *such that* $f_{X,Z}(X,Z)$ *has a continuous Lebesgue density that is strictly positive on* U*. Moreover for each* $(x,z) \in f_{X,Z}^{-1}(U)$, $||Df_{X,Z}(x,z)|| > 0$.

(ii) q_{02} *has a neighborhood* V *such that* $f_{h_1,P}(h_1(X,P(Z)),P(Z))$ *has a continuous Lebesgue density that is strictly positive on* V*. Moreover for each* $(x,z) \in f_{X,Z}^{-1}(U)$, $||Df_{h_1,P}(h_0(x,P(z))|| > 0$.

(d) (i) For each $z \in \operatorname{supp}(Z)$ *such that there exists an* $x \in \operatorname{supp}(X)$ *with* $(x,z) \in f_{X,Z}^{1}(U)$, $||DP(z)|| > 0$.

(ii) For each $(x,z) \in f_{X,Z}^{-1}(U)$, $||D_x h_1(x,P(z))|| > 0$*, and* $||D_P h_1(x,P(z))|| > 0$.

(e) $\tilde{K}_1$, $\tilde{K}_1'$, $\tilde{K}_2$ *and* $\tilde{K}_2'$ *satisfy the conditions of Theorem (A). Moreover,* $\tilde{K}_2$ *is Lipschitz.*

(f) (i) $\tilde{h}_{N1} \to 0$, $\frac{\log N}{N\tilde{h}_{N1}^{d+1}} \to 0$.

(ii) $\tilde{h}_{N2} \to 0$, $\frac{\log N}{N\tilde{h}_{N2}^{3}} \to 0$*, and* $N\tilde{h}_{N2}^{12} \to c \in (0,\infty]$.

Assumptions related to the estimation of $E(D|Z)$**:**

Assumption B.3 *(a)* $E(D_i|Z_i = z)$ *is* $\overline{p}_P$*-smooth with* $\overline{p}_P > d_z$*. The point* z *is in the interior of the support of* Z*.*

(b) Bandwidth sequence $\{h_{NP}\}$ *satisfies* $h_{NP} \to 0$, $Nh_{NP}^{d_z}/\log N \to \infty$*, and* $Nh_{NP}^{2\overline{p}_P} \to c_P \in (0,\infty)$.

(c) Kernel function K^P *is symmetric, supported on acompact set, and is Lipschitz continuous. Also it has moments of order* $p_P + 1$ *through* $\overline{p}_P - 1$ *that are equal to 0.*

Assumptions related to the estimation of $E(DY|P(Z),X)$ **and** $E(-(1-D)Y|P(Z),X)$**:**

Assumption B.4 *(a)* $E(DY|P(Z) = p, X = x)$ *and* $E(-(1-D)Y|P(Z) = p, X = x)$ *are both* $\overline{p}_h$*-smooth with* $\overline{p}_h > d_x + 2$*. The point* (p,x) *is in the interior of the support of* $(X,P(Z))$.

(b) $\{h_{Nh}\}$ *satisfies* $Nh_{Nh}^{d+1}/\log N \to \infty$ *and* $Nh_{Nh}^{2(\overline{p}_h-1)} \to c_h < \infty$ *for some* $c_h \geq 0$.

(c) Kernel function $K^h(\cdot)$ *is 1-smooth, symmetric and supported on a compact set. It has moments of order* $p+1$ *through* $\overline{p}_h - 1$ *that are equal to zero.*

Assumptions related to the estimation of $E(Y|D = 1, h_1(X,P(Z)), P(Z)), P(Z))$**:**

Assumption B.5 *(a)* $E(DY|h_1(X,P(Z))=h,P(Z)=p)$ *is* $\overline{p}_q$*-smooth with* $\overline{p}_q>2$*. The point* (h,p) *is in the interior of the support of* $(h_1(X,P(Z)),P(Z))$*.*

(b) $\{h_{Nq}\}$ *satisfies* $Nh_{Nq}^2/\log N\to\infty$ *and* $Nh_{Nq}^{2\overline{p}_q}\to c_q<\infty$ *for some* $c_q\geq 0$*.*

(c) Kernel function $K^q(\cdot)$ *is 1-smooth, symmetric and supported on a compact set. It has moments of order* $p+1$ *through* $\overline{p}_q-1$ *that are equal to zero.*

(d) The function $P(Z)$ *is bounded away from 0.*

We are now ready to study the asymptotic behavior of our estimator. Note that

$$\begin{aligned}\sqrt{N}\big[\tilde{\Delta}-E(Y_0|A_1\cap A_2)\big] &= \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\frac{\left[(1-D_i)Y_i+D_iq(h_{0i},P_i)-E(Y_0|(X,Z)\in A_1\cap A_2)\right]\hat{I}_{1i}\hat{I}_{2i}}{N^{-1}\sum_{i=1}^{n}\hat{I}_{1i}\hat{I}_{2i}}\\ &+ \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\frac{D_i\left[\hat{q}(\hat{h}_{0i},\hat{P}_i)-q(h_{0i},P_i)\right]\hat{I}_{1i}\hat{I}_{2i}}{N^{-1}\sum_{i=1}^{n}\hat{I}_{1i}\hat{I}_{2i}}\end{aligned}\tag{10}$$

We will study the asymptotic behavior of $N^{-1/2}\sum_{i=1}^{N}D_i\big[\hat{q}(\hat{h}_{0i},\hat{P}_i)-q(h_{0i},P_i)\big]\hat{I}_{1i}\hat{I}_{2i}$, $N^{-1/2}\sum_{i=1}^{N}\big[(1-D_i)Y_i+D_iq(h_{0i},P_i)-E(Y_0|(X,Z)\in A_1\cap A_2)\big]\hat{I}_{1i}\hat{I}_{2i}$, and $N^{-1}\sum_{i=1}^{N}\hat{I}_{1i}\hat{I}_{2i}$ separately. An application of the mean value theorem to the first of these terms reveals that the asymptotic behavior of that term is largely determined by the asymptotic behavior of $\hat{P}(z)$, $\hat{h}_0(x,\hat{P}(z))$ and $\hat{q}(h_0(x,P(z)),P(z))$. The asymptotic properties of $\hat{P}(z)$ can be obtained by applying Theorem 3 of HIT. Analyzing the asymptotic behavior of $\hat{h}_0(x,\hat{P}(z))$ requires simple modifications of Theorems 3 and 4 of HIT. The modifications are needed because $h_0(X,P(Z))$ itself is not a conditional expectation, but it is the derivative of one. Heckman, Ichimura and Todd are interested in the first element of the estimated coefficient vector, we are interested in the second element. Analyzing the asymptotic properties of $\hat{q}(h_0(x,P(z)),P(z))$ is also slightly different because this is an estimator for the expectation of Y given $D=1$, $h_1(X,P(Z))$ and $P(Z)$ evaluated at the value the random vector $\big(h_0(X,P(Z)),P(Z)\big)$ takes (and $D=1$). Evaluating this conditional expectation at $\big(h_0(X,P(Z)),P(Z)\big)$ is meaningful only when $\big(h_0(X,P(Z)),P(Z)\big)$ is an element of the support of $\big(h_1(X,P(Z)),P(Z)\big)$. Consequently, we have to use another trimming function to make sure our evaluation point is in the support of $\big(h_1(X,P(Z)),P(Z)\big)$. The details of our trimming function and how these three estimators behave asymptotically are given in Appendix C.

$$\begin{aligned}\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}D_i\big[\hat{q}(\hat{h}_{0i},\hat{P}_i)-q(h_{0i},P_i)\big]\hat{I}_{1i}\hat{I}_{2i} &= \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}D_i\big[\hat{q}(\hat{h}_{0i},\hat{P}_i)-\hat{q}(h_{0i},P_i)\big]\hat{I}_{1i}\hat{I}_{2i}\\ &+ \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}D_i\big[\hat{q}(h_{0i},P_i)-q(h_{0i},P_i)\big]\hat{I}_{1i}\hat{I}_{2i}\end{aligned}\tag{11}$$

To deal with the first of these two terms, we use the Mean Value Theorem.

$$\begin{aligned}\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}D_i\big[\hat{q}(\hat{h}_{0i},\hat{P}_i)-D_i\hat{q}(h_{0i},P_i)\big]\hat{I}_{1i}\hat{I}_{2i} &= \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}D_i\frac{\partial\hat{q}}{\partial h_1}(\tilde{h}_{0i},\tilde{P}_i)\Big(\hat{h}_0(X_i,\hat{P}(Z_i))-h_0(X_i,P(Z_i))\Big)\hat{I}_{1i}\hat{I}_{2i}\\ &+ \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}D_i\frac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i},\tilde{P}_i)\big(\hat{P}(Z_i)-P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i}\end{aligned}\tag{12}$$

where for each i, $(\tilde{h}_{0i}, \tilde{P}_i)$ is between $(h_0(X_i, P(Z_i)), P(Z_i))$ and $(\hat{h}_0(X_i, \hat{P}(Z_i)), \hat{P}(Z_i))$. It is easier to analyze these terms separately first, and then combine the results later. We now proceed as follows. Steps 1 and 2 (sections B.1 and B.2) examine the first and second terms of equation 12, respectively. Steps 3 (section B.3) considers the second term of equation 11. Thus, steps 1 to 3 consider each term of equation 11, and thus analyze the numerator of the second term of equation 10. In step 4 (section B.4), we consider the numerator of the first term of equation 10. In step 5 (section B.5), we consider the numerator of equation 10. The result stated in then text then immediately follows from Slutsky.

B.1 Step 1:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i)\big(\hat{P}(Z_i) - P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i\big[\tfrac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i) - \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\big]\big(\hat{P}(Z_i) - P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\big(\hat{P}(Z_i) - P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i}$$

By Appendix C.1, we know that

$$\big[\hat{P}(z)) - P(z)\big]\hat{I}_1(x,z) = N^{-1}\sum_{i=1}^{N}\psi_{NP}(D_i, Z_i; x, z) + \hat{b}_P(x,z) + \hat{R}_P(x,z)$$

where $E[\psi_{NP}(D_i, Z_i; X, Z)|X = x, Z = z] = 0$, $\text{plim}_{N\to\infty} N^{-1/2}\sum_{i=1}^{N}\hat{b}_P(X_i, Z_i) = b_P < \infty$, and $\text{plim}_{N\to\infty} N^{-1/2}\sum_{i=1}^{N}\hat{R}_P(X_i, Z_i) = 0$. Substituting all these in yields

$$\begin{aligned}
&\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} D_i\left[\tfrac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i) - \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\right]\big(\hat{P}(Z_i) - P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i} \\
&= \tfrac{1}{N\sqrt{N}}\textstyle\sum_{i=1}^{N} D_i\left[\tfrac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i) - \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\right]\sum_{j=1}^{N}\psi_{NP}(D_j, Z_j; X_i, Z_i)\hat{I}_{2i} \\
&\quad + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} D_i\left[\tfrac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i) - \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\right]\hat{b}_P(X_i, Z_i)\hat{I}_{2i} \\
&\quad + \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} D_i\left[\tfrac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i) - \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\right]\hat{R}_P(X_i, Z_i)\hat{I}_{2i}
\end{aligned}$$

Let $\overline{A}_1 := \{(x,z) \in \text{supp}(X,Z) : \hat{f}(x,z) \geq q_{01} - \epsilon_{f1}\}$. Then by Appendices C.1 and C.2, we know that $\hat{P}(z)$ is uniformly consistent for $P(z)$, and $\hat{h}_0(\hat{P}(z), x)$ is uniformly consistent for $h_0(P(z), x))$ on $\overline{A}_1$. Then applying theorem 4 of Heckman, Ichimura and Todd to $\hat{q}$ for the set of observations for which $D_i = 1$, we know that $\frac{\partial\hat{q}}{\partial P}(h,p)$ is uniformly consistent for $\frac{\partial q}{\partial P}(h,p)$ on $A_1 \cap A_2$[22]. Then using the equicontinuity lemma we can show that the probability limit of each of these terms is

[22]Note that $A_1 \cap A_2 \subset \overline{A}_1$

0, so that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial \hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i)\big(\hat{P}(Z_i) - P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i} =^{AE} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\big(\hat{P}(Z_i) - P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i}$$

The latter term in turn equals

$$\begin{aligned} &\tfrac{1}{N\sqrt{N}}\textstyle\sum_{i=1}^{N} D_i \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\sum_{j=1}^{N} \psi_{NP}(D_j, Z_j; X_i, Z_i)\hat{I}_{2i} \\ +&\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} D_i \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\hat{b}_P(X_i, Z_i)\hat{I}_{2i} + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\hat{R}_P(X_i, Z_i)\hat{I}_{2i} \end{aligned}$$

Using continuity of $\frac{\partial q}{\partial P}(h_{0i}, P_i)$, compactness of $A_1 \cap A_2$, and the explicit form of $\hat{b}_P$, and $\hat{R}_P$, we can show that

$$b_{qP} := \text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\hat{b}_P(X_i, Z_i)\hat{I}_{2i} < \infty$$

and

$$\text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\hat{R}_P(X_i, Z_i)\hat{I}_{2i} = 0$$

On the other hand, using the equicontinuity lemma once more, we can show that

$$\frac{1}{N\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\sum_{j=1}^{N} \psi_{NP}(D_j, Z_j; X_i, Z_i)\left(\hat{I}_{2i} - I_{2i}\right) = o_P(1)$$

Combining these results, we conclude that

$$\begin{aligned} &\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial \hat{q}}{\partial P}(\tilde{h}_{0i}, \tilde{P}_i)\big(\hat{P}(Z_i) - P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i} \\ &\qquad =^{AE} \frac{1}{N\sqrt{N}}\sum_{i=1}^{N}\sum_{j=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\psi_{NP}(D_j, Z_j; X_i, Z_i)I_{2i} + Vb_{qP}. \end{aligned}$$

Next, we focus on

$$\frac{1}{N\sqrt{N}}\sum_{i=1}^{N}\sum_{j=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\psi_{NP}(D_j, Z_j; X_i, Z_i)I_{2i}$$

This term can be broken into

$$\frac{1}{N\sqrt{N}}\sum_{i=1}^{N} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\psi_{NP}(D_i, Z_i; X_i, Z_i)I_{2i} + \frac{1}{N\sqrt{N}}\sum_{i=1}^{N}\sum_{j\neq i} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\psi_{NP}(D_j, Z_j; X_i, Z_i)I_{2i}$$

We can apply a strong law of large numbers to the first of these terms.

$$E\left\{D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)\psi_{NP}(D_i, Z_i; X_i, Z_i)I_{2i}\right\} = E\left\{\frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}E\Big[D_i\psi_{NP}(D_i, Z_i; X_i, Z_i)|X_i, Z_i\Big]\right\}$$

$$E\Big[D_i\psi_{NP}(D_i, Z_i; X_i, Z_i)|X_i, Z_i\Big] = I_1(X_i, Z_i)e_1[M_{pN}^p(Z_i)]^{-1}e_1' h_{NP}^{-d_z}K^p(0)E\Big[D_i\varepsilon_i^P|Z_i\Big]$$

where $\varepsilon_i^P = D_i - E(D_i|Z_i)$, so that $E(D_i\varepsilon_i^P|Z_i) = P(Z_i)\big(1 - P(Z_i)\big)$. So unfortunately, we don't have $E\left\{D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_2(X_i, Z_i)\psi_{NP}(D_i, Z_i; X_i, Z_i)\right\} = 0$. But $E\left\{D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_2(X_i, Z_i)\psi_{NP}(D_i, Z_i; X_i, Z_i)\right\} < \infty$, implies that [23]

$$\begin{gathered}N^{-3/2}\textstyle\sum_{i=1}^N D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_i, Z_i; X_i, Z_i) = \\ \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{1}{N}\left[D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_i, Z_i; X_i, Z_i) - E\left(D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_i, Z_i; Z_i)\right)\right] \\ + \frac{1}{N\sqrt{N}}E\left(D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_i, Z_i; X_i, Z_i)\right) = o_p(1)\end{gathered}$$

Next, we deal with the sum corresponding to different indices:

$$\begin{gathered}N^{-3/2}\textstyle\sum_{i=1}^N \sum_{j\neq i} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_j, Z_j; X_i, Z_i) \\ = N^{-3/2}\sum_{i=1}^N \sum_{j\neq i}\left(\frac{1}{2}D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_j, Z_j; X_i, Z_i) + \frac{1}{2}D_j \frac{\partial q}{\partial P}(h_{0j}, P_j)I_{2j}\psi_{NP}(D_i, Z_i; X_j, Z_j)\right)\end{gathered}$$

$$\begin{gathered}E\left[D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_j, Z_j; X_i, Z_i)\right] = \\ E\left[D_i \frac{\partial q}{\partial P}\big(h_0(X_i, P(Z_i)), P(Z_i)\big)I_2(X_i, Z_i)e_1[M_{pN}^P(Z_i)]^{-1}\varepsilon_j^P\left[\left(\frac{Z_j - Z_i}{h_{NP}}\right)^{Q_p}\right]' h_{NP}^{-d_z}K^P\left(\frac{Z_j - Z_i}{h_{NP}}\right)\right] \\ = E_{X_i, Z_i, D_i, Z_j}\left[D_i \frac{\partial q}{\partial P}\Big(h_0(X_i, P(Z_i)), P(Z_i)\Big)I_2(X_i, Z_i)e_1[M_{pN}^P(Z_i)]^{-1}E\Big(\varepsilon_j^P|Z_j\Big) \quad \frac{Z_j - Z_i}{h_{NP}} \quad {}^{Q_p} \quad {}' h_{NP}^{-d_z}K^P \quad \frac{Z_j - Z_i}{h_{NP}} \quad \right] = 0\end{gathered}$$

Similarly, $E\left[D_j \frac{\partial q}{\partial P}(h_{0j}, P_j)I_2(X_j, Z_j)\psi_{NP}(D_i, Z_i; X_j, Z_j)\right] = 0$. On the other hand,

$$\begin{gathered}N^{-3/2}\textstyle\sum_{i=1}^N \sum_{j\neq i} \frac{1}{2}\left(D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_j, Z_j; X_i, Z_i) + D_j \frac{\partial q}{\partial P}(h_{0j}, P_j)I_{2j}\psi_{NP}(D_i, Z_i; X_j, Z_j)\right) \\ \frac{1}{\sqrt{N}(N-1)}\sum_{i=1}^N \sum_{j\neq i} \frac{N-1}{2N}\left(D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_j, Z_j; X_i, Z_i) + D_j \frac{\partial q}{\partial P}(h_{0j}, P_j)I_{2j}\psi_{NP}(D_i, Z_i; X_j, Z_j)\right)\end{gathered}$$

Since $\lim_{N\to\infty} \frac{N-1}{N} = 1$, the asympototic behavior of this last object is the same as the asymptotic behavior of

$$\frac{1}{\sqrt{N}(N-1)}\sum_{i=1}^N \sum_{j\neq i} \frac{1}{2}\left(D_i \frac{\partial q}{\partial P}(h_{0i}, P_i)I_{2i}\psi_{NP}(D_j, Z_j; X_i, Z_i) + D_j \frac{\partial q}{\partial P}(h_{0j}, P_j)I_{2j}\psi_{NP}(D_i, Z_i; X_j, Z_j)\right)$$

[23] Note that $Nh_{NP}^{2d_z} \to \infty$.

Let

$$\zeta_N(D_i, X_i, Z_i, D_j, D_j, X_j, Z_j) = \frac{1}{2} D_i \frac{\partial q}{\partial P}(h_{0i}, P_i) I_{2i} \psi_{NP}(D_j, Z_j; X_i, Z_i) + \frac{1}{2} D_j \frac{\partial q}{\partial P}(h_{0j}, P_j) I_{2j} \psi_{NP}(D_i, Z_i; X_j, Z_j)$$

Then by Hoeffding, Powell, Stock and Stoker lemma, if $E\left(\zeta_N(D_i, X_i, Z_i, D_j, X_j, Z_j)\right)^2 = o(N)$, then

$$NE\Big[\Big(\tfrac{1}{N(N-1)}\sum_{i=1}^N \sum_{j\neq i} \zeta_N(D_i,X_i,Z_i,D_j,X_j,Z_j) - \tfrac{1}{N}\sum_{i=1}^N 2E[\zeta_N(D_i,X_i,Z_i,D_j,X_j,Z_j)|X_i,Z_i]\Big)^2\Big] = o(1)$$

$$E \quad \tfrac{1}{\sqrt{N}(N-1)}\sum_{i=1}^N \sum_{j\neq i} \zeta_N(D_i,X_i,Z_i,D_j,X_j,Z_j) - \tfrac{1}{\sqrt{N}}\sum_{i=1}^N 2E[\zeta_N(D_i,X_i,Z_i,D_j,X_j,Z_j)|X_i,Z_i] \quad ^2 = o(1)$$

$$\tfrac{1}{\sqrt{N}(N-1)}\sum_{i=1}^N \sum_{j\neq i} \tfrac{1}{2}\left(D_i \tfrac{\partial q}{\partial P}(h_{0i}, P_i) I_{2i}\psi_{NP}(D_j, Z_j; X_i, Z_i) + D_j \tfrac{\partial q}{\partial P}(h_{0j}, P_j) I_{2j}\psi_{NP}(D_i, Z_i; X_j, Z_j)\right) \to^P \tfrac{1}{\sqrt{N}}\sum_{i=1}^N 2E[\zeta_N(D_i, X_i, Z_i, D_j, X_j, Z_j)|D_i, X_i, Z_i]$$

$$2E[\zeta_N(D_i, X_i, Z_i, D_j, X_j, Z_j)|D_i, X_i, Z_i] = D_i \frac{\partial q}{\partial P}(h_{0i}, P_i) I_{2i} E[\psi_{NP}(D_j, Z_j; X_i, Z_i)|D_i, X_i, Z_i] + E\left[D_j \frac{\partial q}{\partial P}(h_{0j}, P_j) I_{2j}\psi_{NP}(D_i, Z_i; X_j, Z_j)|D_i, X_i, Z_i\right]$$

$$E[\psi_{NP}(D_j,Z_j;X_i,Z_i)|D_i,X_i,Z_i] = \tfrac{1}{h_{NP}^{d_z}} I_1(X_i,Z_i)[M_{pN}^q(Z_i)]^{-1} E \quad \tfrac{Z_j-Z_i}{h_{NP}} \quad ^{Q_p} \quad ' K^q \quad \tfrac{Z_j-Z_i}{h_{NP}} \quad \varepsilon_j^P|D_i,X_i,Z_i$$

$$= \tfrac{1}{h_{NP}^{d_z}} I_1(X_i, Z_i)[M_{pN}^q(Z_i)]^{-1} E\left(E\left\{\left[\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)^{Q_p}\right]' K^q\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)\varepsilon_j^P|D_i, X_i, Z_i, Z_j\right\}|D_i, X_i, Z_i\right) = 0$$

Therefore,

$$2E[\zeta_N(D_i, X_i, Z_i, D_j, X_j, Z_j)|D_i, X_i, Z_i] = E\left[D_j \frac{\partial q}{\partial P}(h_{0j}, P_j) I_{2j}\psi_{NP}(D_i, Z_i; X_j, Z_j)|D_i, X_i, Z_i\right]$$

Next, we investigate if can we get $E\zeta(D_i, X_i, Z_i, D_j, X_j, Z_j)^2 = o(N)$ from our basic assumptions? This requires

$$E\left\{D_i\left(\tfrac{\partial q}{\partial P}(h_{0i}, P_i)\right)^2 I_{2i}^2 I_{1i}^2 (\varepsilon_j^P)^2 \left(e_1[M_{pN}^q(Z_i)]^{-1}\left[\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)^{Q_p}\right]'\right)^2 h_{NP}^{-2d_z} K^q\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)^2\right\} = o(N)$$

and

$$E\left\{D_i D_j \tfrac{\partial q}{\partial P}(h_{0i}, P_i)\tfrac{\partial q}{\partial P}(h_{0j}, P_j) I_{2i} I_{2j}\psi_{NP}(D_j, Z_j; X_i, Z_i)\psi_{NP}(D_i, Z_i; X_j, Z_j)\right\} = o(N)$$

By Cauchy-Schwarz inequality, the first one implies the second. Thus, we only need to make sure the first one holds. Now, to make sure that $\hat{P}(z)$ is an asymptotically linear estimator for $P(z)$, we had to assume that for $\overline{p}_P > d_z$, $Nh_{NP}^{2\overline{p}} \to c < \infty$. This means that $Nh_{NP}^{2d_z} \to \infty$, and $\frac{1}{Nh_{NP}^{2d_z}} \to 0$, as long as $c > 0$. With $c > 0$, the required condition will hold if, for each N,

$$E\left\{D_i\left(\frac{\partial q}{\partial P}(h_{0i},P_i)\right)^2 I_{1i}I_{2i}(\varepsilon_j^P)^2\left(e_1[M_{pN}^q(Z_i)]^{-1}\left[\left(\frac{Z_j-Z_i}{h_{NP}}\right)^{Q_p}\right]'\right)^2\left(K^q\left(\frac{Z_j-Z_i}{h_{NP}}\right)\right)^2\right\} < \infty$$

On the other hand,

$$\begin{aligned} &E\left\{D_i\left(\tfrac{\partial q}{\partial P}(h_{0i},P_i)\right)^2 I_{1i}I_{2i}(\varepsilon_j^P)^2\left(e_1[M_{pN}(Z_i)]^{-1}\left[\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)^{Q_p}\right]'\right)^2\left(K^q\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)\right)^2\right\} \\ =\ &E\left\{P(Z_i)P(Z_j)\big(1-P(Z_j)\big)I_{1i}I_{2i}\left(\tfrac{\partial q}{\partial P}(h_{0i},P_i)\right)^2\left(e_1[M_{pN}(Z_i)]^{-1}\left[\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)^{Q_p}\right]'\right)^2\left(K^q\left(\tfrac{Z_j-Z_i}{h_{NP}}\right)\right)^2\right\} \end{aligned}$$

Therefore, the desired condition will hold, as long as, for each N,

$$E\left\{\left(\frac{\partial q}{\partial P}(h_{0i},P_i)\right)^2\left(e_1[M_{pN}(Z_i)]^{-1}I_{1i}I_{2i}\left[\left(\frac{Z_j-Z_i}{h_{NP}}\right)^{Q_p}\right]'\right)^2\left(K^q\left(\frac{Z_j-Z_i}{h_{NP}}\right)\right)^2\right\} < \infty$$

For sufficiently large N this is true, because the kernel function is 0 outside a compact set, $\frac{\partial q}{\partial P}$ and K are continuous functions and M_p is nonsingular. Thus,

$$\begin{aligned} &\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N D_i\frac{\partial\hat{q}}{\partial P}(\tilde{h}_{0i},\tilde{P}_i)\big(\hat{P}(Z_i)-P(Z_i)\big)\hat{I}_{1i}\hat{I}_{2i} =^{AE} \\ &N^{-1/2}\textstyle\sum_{i=1}^N E\left[D_j\frac{\partial q}{\partial P}(h_{0j},P_j)I_{2j}\psi_{NP}(D_i,Z_i;X_j,Z_j)|D_i,Z_i,X_i\right]+b_{qP} \\ &= N^{-1/2}\textstyle\sum_{i=1}^N E\left[D_j\frac{\partial q}{\partial P}(h_{0j},P_j)I_{2j}\psi_{NP}(D_i,Z_i;X_j,Z_j)|Y_i,D_i,Z_i,X_i\right]+b_{qP} \end{aligned}$$

B.2 Step 2:

$$\begin{aligned} &\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N D_i\frac{\partial\hat{q}}{\partial h_1}(\tilde{h}_{0i},\tilde{P}_i)\big(\hat{h}_0(X_i,\hat{P}(Z_i))-h_0(X_i,P(Z_i))\big)\hat{I}_{1i}\hat{I}_{2i} = \\ &\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N D_i\left[\frac{\partial\hat{q}}{\partial h_1}(\tilde{h}_{0i},\tilde{P}_i)-\frac{\partial q}{\partial h_1}(h_{0i},P_i)\right]\left[\hat{h}_0(X_i,\hat{P}(Z_i))-h_0(X_i,P(Z_i))\right]\hat{I}_{1i}\hat{I}_{2i} \\ &+\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N D_i\frac{\partial h_1}{\partial P}(h_{0i},P_i)\left[\hat{h}_0(X_i,\hat{P}(Z_i))-h_0(X_i,P(Z_i))\right]\hat{I}_{1i}\hat{I}_{2i} \end{aligned}$$

By Appendix C.2, we know that

$$\left[\hat{h}_0(X_i,\hat{P}(Z_i))-h_0(X_i,P(Z_i))\right]\hat{I}_1(x,z) = N^{-1}\sum_{i=1}^N \psi_{Nh_0P}(D_i,Y_i,X_i,Z_i;x,z)+\hat{b}_{\hat{h}_0}(x,z)+\hat{R}_{\hat{h}_0}(x,z)$$

where $E[\psi_{Nh_0P}(D_i, Y_i, X_i, Z_i; X, Z)|X = x, Z = z] = 0$,
$\text{plim}_{N\to\infty} N^{-1/2}\sum_{i=1}^N \hat{b}_{\hat{h}_0}(X_i, Z_i) = b_{h_0P} < \infty$, and $\text{plim}_{N\to\infty} N^{-1/2}\sum_{i=1}^N \hat{R}_{\hat{h}_0}(X_i, Z_i) = 0$. Then using arguments similar to those in step 1, we can show that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial \hat{q}}{\partial h_1}(\tilde{h}_{0i}, \tilde{P}_i)\big[\hat{h}_0(X_i, \hat{P}(Z_i)) - h_0(X_i, P(Z_i))\big]\hat{I}_{1i}\hat{I}_{2i} =^{AE}$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\big[\hat{h}_0(X_i, \hat{P}(Z_i)) - h_0(X_i, P(Z_i))\big]\hat{I}_{1i}\hat{I}_{2i}$$

The latter term in turn equals

$$\frac{1}{N\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\sum_{j=1}^N \psi_{Nh_0P}(D_j, Y_j, X_j, Z_j; X_i, Z_i)\hat{I}_{2i}$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\hat{b}_{\hat{h}_0}(X_i, Z_i)\hat{I}_{2i} + \frac{1}{\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\hat{R}_{\hat{h}_0}(X_i, Z_i)\hat{I}_{2i}$$

Using continuity of $\frac{\partial q}{\partial h_1}(h_{0i}, P_i)$, compactness of $A_1 \cap A_2$, and the explicit form of $\hat{b}_{\hat{h}_0}$, and $\hat{R}_{\hat{h}_0}$, we can show that

$$b_{qh_0P} := \text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\hat{b}_{\hat{h}_0}(X_i, Z_i)\hat{I}_{2i} < \infty$$

and

$$\text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\hat{R}_{\hat{h}_0}(X_i, Z_i)\hat{I}_{2i} = 0$$

On the other hand, using the equicontinuity lemma once more, we can show that

$$\frac{1}{N\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\sum_{j=1}^N \psi_{Nh_0P}(D_j, Y_j, X_j, Z_j; X_i, Z_i)\left(\hat{I}_{2i} - I_{2i}\right) = o_P(1)$$

Combining these results, we conclude that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N D_i \frac{\partial \hat{q}}{\partial h_1}(\tilde{h}_{0i}, \tilde{P}_i)\big(\hat{h}_0(X_i, \hat{P}(Z_i)) - h_0(X_i, P(Z_i))\big)\hat{I}_{1i}\hat{I}_{2i} =^{AE}$$
$$\frac{1}{N\sqrt{N}}\sum_{i=1}^N\sum_{j=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\psi_{Nh_0P}(D_j, Y_j, X_j, Z_j; X_i, Z_i)I_{2i} + b_{qh_0P}$$

Next we focus on

$$N^{-3/2}\sum_{i=1}^N D_i \frac{\partial q}{\partial h_1}(h_{0i}, P_i)\left[\psi_{Nh_0}\big(-(1-D_i)Y_i, P(Z_i), X_i; P(Z_i), X_i\big) + \frac{\partial h_0(P(Z_i), X_i)}{\partial p}\psi_{nP}(D_i, Z_i; Z_i)\right]$$

Since

$$E[D_i\varepsilon_i^{h_0}|P(Z_i), X_i] = E\big[-D_i(1-D_i)Y_i + D_iE\big((1-D_i)Y_i|P(Z_i), X_i\big)|P(Z_i), X_i\big]$$
$$= E\big((1-D_i)Y_i|P(Z_i), X_i\big)P(Z_i)$$

$$E \quad D_i \frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}\psi_{Nh_0}\Big(-(1-D_i)Y_i,P(Z_i),X_i;P(Z_i),X_i\Big) \; = \frac{K(0)}{h_{Nh_0}^{d_X+2}} E\Big\{ \frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}e_2[M_{pN}^h(P_i,X_i)]^{-1}e_1' E[D_i\varepsilon_i|P(Z_i),X_i] \Big\}$$

$$= \frac{K(0)}{h_{Nh_0}^{d_X+2}} E\left\{ \frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}e_2[M_{pN}^h(P_i,X_i)]^{-1}e_1' P(Z_i)(1-D_i)Y_i \right\}$$

Using $E\left\{\frac{\partial q}{\partial h_1}(h_{0i},P_i)e_2[M_{pN}^h(P_i,X_i)]^{-1}e_1'P(Z_i)(1-D_i)Y_i\right\} < \infty$, and $Nh_{Nh_0}^{2d_X+4} \to \infty$, we get that $N^{-3/2}\sum_{i=1}^N \; D_i\frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}\psi_{Nh_0}\big(-(1-D_i)Y_i,P(Z_i),X_i;P(Z_i),X_i\big) = o_p(1)$. On the other hand, $E[D_i\varepsilon_i^P|Z_i,X_i]=E\Big[D_i-D_iE\Big(D_i|Z_i\Big)|Z_i,X_i\Big]=P(Z_i)\Big(1-P(Z_i)\Big)$, so that

$$E\Big\{ \; D_i\frac{\partial q}{\partial h_1}(h_{0i},P_i)\frac{\partial h_0(P(Z_i),X_i)}{\partial p}I_{2i}\psi_{NP}(D_i,Z_i;Z_i)\Big\}=\frac{K(0)}{h_{NP}^{d_z}}E \quad \frac{\partial q}{\partial h_1}(h_{0i},P_i)\frac{\partial h_0(P(Z_i),X_i)}{\partial p}I_{2i}e_1[M_{pN}^P(Z_i)]^{-1}e_1'P(Z_i)\Big(1-P(Z_i)\Big)$$

Combining $E\left\{\frac{\partial q}{\partial h_1}(h_{0i},P_i)\frac{\partial h_0(P(Z_i),X_i)}{\partial p}I_{2i}e_1[M_{pN}^P(Z_i)]^{-1}e_1'P(Z_i)\big(1-P(Z_i)\big)\right\} < \infty$ and $Nh_{NP}^{2d_z} \to \infty$, we also conclude that $N^{-3/2}\sum_{i=1}^N \; D_i\frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}\frac{\partial h_0(P(Z_i),X_i)}{\partial p}\psi_{NP}(D_i,Z_i;Z_i) = o_p(1)$. Next, we look at the terms with different indices:

$$N^{-3/2}\sum_{i=1}^N\sum_{j\neq i}D_i\frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}\psi_{Nh_0}\Big(-(1-D_j)Y_j,P(Z_j),X_j;P(Z_i),X_i\Big)=N^{-3/2}\sum_{i=1}^N\sum_{j\neq i}\tilde{g}_{1N}(\tilde{S}_i,\tilde{S}_j)$$

where

$$\tilde{g}_{1N}(\tilde{S}_i,\tilde{S}_j)= \; \frac{D_i}{2}\frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}\psi_{Nh_0}\Big(-(1-D_j)Y_j,P(Z_j),X_j;P(Z_i),X_i\Big)+\frac{D_j}{2}\frac{\partial q}{\partial h_1}(h_{0j},P_j)I_{2j}\psi_{Nh_0}\Big(-(1-D_i)Y_i,P(Z_i),X_i;P(Z_j),X_j\Big)$$

and $\tilde{S}_i = (D_i,Y_i,Z_i,X_i)$. Using the definition of ε^{h_0}, iterated law of expectations and the independence of observations from one another, one could show that the expectation of each term in this sum is 0. Moreover, since $M_p^h(P_i,X_i)$ is nonsingular, $\partial q/\partial h_1$ is continuous, K is 0 outside a compact set, and $var(Y) < \infty$, and $Nh_{Nh_0}^{2(d_X+2)} \to \infty$, $E\big(\tilde{g}_{1N}(\tilde{S}_i,\tilde{S}_j)^2\big) = o(N)$. Therefore by the Hoeffding, Powell, Stock and Stoker lemma,

$$N^{-3/2}\sum_{i=1}^N\sum_{j\neq i}D_i\frac{\partial q}{\partial h_1}(h_{0i},P_i)I_{2i}\psi_{Nh_0}\big(-(1-D_j)Y_j,P(Z_j),X_j;P(Z_i),X_i\big) =^{AE}$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^N E\left[D_j\frac{\partial q}{\partial h_1}(h_{0j},P_j)I_{2j}\psi_{Nh_0}\big(-(1-D_i)Y_i,P(Z_i),X_i;P(Z_j),X_j\big)|D_i,Y_i,P(Z_i),X_i\right]$$

Similarly,

$$N^{-3/2}\sum_{i=1}^N\sum_{j\neq i}D_iI_{2i}\frac{\partial q}{\partial h_1}(h_{0i},P_i)\frac{\partial h_0}{\partial P}(P(Z_i),X_i)\psi_{NP}\big(D_j,Z_j,Z_i\big) =^{AE}$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^N E\left[D_jI_{2j}\frac{\partial q}{\partial h_1}(h_{0j},P_j)\frac{\partial h_0}{\partial P}(P(Z_j),X_j)\psi_{NP}\big(D_i,Z_i,Z_j\big)|D_i,Z_i,X_i\right]$$

Therefore,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N D_i\frac{\partial\hat{q}}{\partial h_1}(\tilde{h}_{0i},\tilde{P}_i)\left[\hat{h}_0(X_i,\hat{P}(Z_i))-h_0(X_i,P(Z_i))\right]\hat{I}_{1i}\hat{I}_{2i} =^{AE}$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^N E\left[D_j\frac{\partial q}{\partial h_1}(h_{0j},P_j)I_{2j}\psi_{Nh_0}\big(-(1-D_i)Y_i,P(Z_i),X_i;X_j,P(Z_j)\big)|Y_i,D_i,X_i,Z_i\right]+$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^N E\left[D_j\frac{\partial q}{\partial h_1}(h_{0j},P_j)I_{2j}\frac{\partial h_0}{\partial P}(P(Z_j),X_j)\psi_{NP}\big(D_i,Z_i;X_j,Z_j\big)|Y_i,D_i,X_i,Z_i\right]+b_{qh_0P}$$

B.3 Step 3:

By Appendix C.3, we know that

$$\tfrac{1}{\sqrt{N}} \textstyle\sum_{i=1}^{N} D_i \big(\hat{q}(h_{0i}, P_i) - q(h_{0i}, P_i)\big) \hat{I}_{1i} \hat{I}_{2i} =^{AE}$$

$$\tfrac{1}{\sqrt{N}N} \textstyle\sum_{i=1}^{N} \sum_{j=1}^{N} \tfrac{D_i}{P(Z_i)} [M_{pN}^q(h_{0i}, P_i)]^{-1} \left[\left(\tfrac{(h_{1j},P_j)-(h_{0i},P_i)}{h_{Nq}} \right)^{Q_p} \right]' \tfrac{1}{h_{Nq}^2} K^q \left(\tfrac{(h_{1j},P_j)-(h_{0i},P_i)}{h_{Nq}} \right) \varepsilon_j^q I_{1i} I_{2i} + b_q$$

where $\varepsilon_j^q = D_j Y_j - E\big[D_j Y_j | h_1\big(X_j, P(Z_j)\big), P(Z_j)\big]$

As in the previous two steps, we first focus on

$$\tfrac{1}{\sqrt{N}N} \textstyle\sum_{j=1}^{N} \tfrac{D_j}{P(Z_j)} [M_{pN}^q(h_{0j}, P_j)]^{-1} \left[\left(\tfrac{(h_{1j},P_j)-(h_{0j},P_j)}{h_{Nq}} \right)^{Q_p} \right]' \tfrac{1}{h_{Nq}^2} K^q \left(\tfrac{(h_{1j},P_j)-(h_{0j},P_j)}{h_{Nq}} \right) \varepsilon_j^q I_{1j} I_{2j}$$

$$\begin{aligned} &\text{plim} \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \tfrac{D_i}{P(Z_i)} e_1 [M_{pN}^q(h_{0i}, P_i)]^{-1} \left[\left(\tfrac{(h_{1i},P_i)-(h_{0i},P_i)}{h_{Nq}} \right)^{Q_p} \right]' \tfrac{1}{h_{Nq}^2} K^q \left(\tfrac{(h_{1i},P_i)-(h_{0i},P_i)}{h_{Nq}} \right) \varepsilon_i^q I_{1i} I_{2i} \\ &= E \left[\tfrac{D_i}{P(Z_i)} e_1 [M_{pN}^q(h_{0i}, P_i)]^{-1} \left[\left(\tfrac{(h_{1i},P_i)-(h_{0i},P_i)}{h_{Nq}} \right)^{Q_p} \right]' \tfrac{1}{h_{Nq}^2} K^q \left(\tfrac{(h_{1i},P_i)-(h_{0i},P_i)}{h_{Nq}} \right) \varepsilon_i^q I_{1i} I_{2i} \right] < \infty \end{aligned}$$

and $N h_{Nq}^4 \to \infty$. Therefore,

$$\text{plim} \tfrac{1}{N\sqrt{N} h_{Nq}^2} \textstyle\sum_{i=1}^{N} \tfrac{D_i}{P(Z_i)} e_1 [M_{pN}(h_{0i}, P_i)]^{-1} \left[\left(\tfrac{(h_{1i},P_i)-(h_{0i},P_i)}{h_{Nq}} \right)^{Q_p} \right]' K \left(\tfrac{(h_{1i},P_i)-(h_{0i},P_i)}{h_{Nq}} \right) \varepsilon_i^q I_{1i} I_{2i} = 0$$

Next, we focus on the sum containing different indices. Using the assumption that the observations are independent, the definition of ε^q and the law of iterated expectations, we first observe that for $i \neq j$,

$$E \left\{ \frac{D_i}{P(Z_i)} e_1 [M_{pN}^q(h_{0i}, P_i)]^{-1} \left[\left(\frac{(h_{1j}, P_j) - (h_{0i}, P_i)}{h_{Nq}} \right)^{Q_p} \right]' \frac{1}{h_{Nq}^2} K^q \left(\frac{(h_{1j}, P_j) - (h_{0i}, P_i)}{h_{Nq}} \right) \varepsilon_j^q I_{1i} I_{2i} \right\} = 0$$

We define $S_i := (D_i, Y_i, X_i, Z_i)$, and

$$\begin{aligned} g_N(S_i, S_j) := &\tfrac{D_i}{2h_{Nq}^2 P(Z_i)} e_1 [M_{pN}^q(h_{0i}, P_i)]^{-1} \left[\left(\tfrac{(h_{1j},P_j)-(h_{0i},P_i)}{h_{Nq}} \right)^{Q_p} \right]' K^q \left(\tfrac{(h_{1j},P_j)-(h_{0i},P_i)}{h_{Nq}} \right) \varepsilon_j^q I_{1i} I_{2i} \\ &+ \tfrac{D_j}{2h_{Nq}^2 P(Z_j)} e_1 [M_{pN}^q(h_{0j}, P_j)]^{-1} \left[\left(\tfrac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}} \right)^{Q_p} \right]' K^q \left(\tfrac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}} \right) \varepsilon_i^q I_{1j} I_{2j} \end{aligned}$$

If $E\big[g_N(S_i, S_j)^2\big] = o(N)$, then by the Hoeffding, Powell, Stock and Stoker lemma

$$\left[\frac{1}{\sqrt{N}(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} g_N(S_i, S_j) \right] =^{AE} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} 2E[g_N(S_i, S_j) | S_i]$$

We know that $Nh_{Nq}^{2\overline{p}} \to c$ for some constant c, and that $\overline{p} > 2$. Thus, $Nh_{Nq}^4 \to \infty$. Combining this with the nonsingularity of $M_p^q(h_{0i}, P_i)$, the fact that $P(Z_i)$ is almost surely bounded away from 0, $var(Y) < \infty$, and that the kernel function is zero outside a compact set, we get $E\big[g_N(S_i, S_j)^2\big] = o(N)$. On the other hand,

$$E[g_N(S_i,S_j)|S_i]=\tfrac{1}{2}E\left\{\tfrac{D_j}{P(Z_j)}e_1[M_{pN}^q(h_{0j},P_j)]^{-1}\left[\quad \tfrac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}} \quad ^{Q_p}\right]'\tfrac{1}{h_{Nq}^2}K^q \quad \tfrac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}} \quad \varepsilon_i^q I_{1j}I_{2j}|D_i,Y_i,X_i,Z_i\right\}$$

This equality follows from the fact that $E\big[\varepsilon_j^q|h_{1j}, P_j, D_i, Y_i, X_i, Z_i\big] = E\big[\varepsilon_j^q|h_{1j}, P_j\big] = 0$. Thus,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} D_i\big(\hat{q}(h_{0i}, P_i) - q(h_{0i}, P_i)\big) =^{AE} b_q + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} E\bigg\{\frac{D_j}{P(Z_j)}e_1[M_{pN}^q(h_{0j}, P_j)]^{-1}I_{1j}I_{2j}$$
$$\times\left[\left(\frac{(h_{1i}, P_i) - (h_{0j}, P_j)}{h_{Nq}}\right)^{Q_p}\right]'\frac{\varepsilon_i^q}{h_{Nq}^2}K^q\left(\frac{(h_{1i}, P_i) - (h_{0j}, P_j)}{h_{Nq}}\right)|D_i, Y_i, X_i, Z_i\bigg\}$$

B.4 Step 4:

Here we study the numerator of the first term of equation (10).

$$\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\Big[(1 - D_i)Y_i + D_i q(h_{0i}, P_i) - E(Y_0|A_1 \cap A_2)\Big]\hat{I}_{1i}\hat{I}_{2i}$$

Let $A := A_1 \cap A_2$ and

$$\delta_A(X_i, Z_i) := (1 - D_i)Y_i + D_i q(h_{0i}, P_i) - E(Y_0|A_1 \cap A_2)$$

For $\tilde{I}_1 \in \mathcal{I}_1$, $\tilde{I}_2 \in \mathcal{I}_2$ such that $\tilde{I}_{1i} \neq I_{1i}$ or $\tilde{I}_{2i} \neq I_{2i}$, $E[\delta_A(X_i, Z_i)\tilde{I}_{1i}\tilde{I}_{2i}] \neq 0$. But with probability one $\hat{I}_{1i}\hat{I}_{2i}$ equals

$$I_{1i}\hat{I}_{2i} + [\hat{\sigma}_1(X_i, Z_i)]^{-1}\tilde{J}_-\left(\tfrac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}_1(X_i,Z_i)}\right)1\{\hat{f}(X_i, Z_i) > f_{X,Z}(X_i, Z_i)\}\big[\hat{f}(X_i, Z_i) - f_{X,Z}(X_i, Z_i)\big]\hat{I}_{2i}$$
$$+[\hat{\sigma}_1(X_i, Z_i)]^{-1}\tilde{J}_+\left(\tfrac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}_1(X_i,Z_i)}\right)1\{\hat{f}(X_i, Z_i)) < f_{X,Z}(X_i, Z_i)\}\big[\hat{f}(X_i, Z_i) - f_{X,Z}(X_i, Z_i)\big]\hat{I}_{2i}$$

where $\tilde{J}_-(u) = 1\{-1 \leq u < 0\}$, $\tilde{J}_+(u) = 1\{0 \leq u < 1\}$, and $\hat{\sigma}_1(X_i, Z_i) := \Big|\hat{f}(X_i, Z_i) - f_{X,Z}(X_i, Z_i)\Big|$. Similarly, for $f \in \mathcal{H}_1$, define, $\tilde{\sigma}_1(X_i, Z_i) := |f(X_i, Z_i) - f_{X,Z}(X_i, Z_i)|$, $\tilde{L}_{1i} = 1\{f(X_i, Z_i) > f_{X,Z}(X_i, Z_i)\}$. Then for $\tilde{I}_2 \in \mathcal{I}_2$,

$$N^{-3/2}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_A(X_i, Z_i)\tilde{I}_{2i}\tilde{L}_{1i}[\tilde{\sigma}_1(X_i, Z_i)]^{-1}\tilde{J}_-\left(\tfrac{f_{X,Z}(X_i,Z_i)-q_{01}}{\tilde{\sigma}_1(X_i,Z_i)}\right)$$
$$\times\left(\tfrac{1}{\tilde{h}_{N1}^d}\tilde{K}_1\left(\tfrac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}}\right) - E\left[\tfrac{1}{\tilde{h}_{N1}^d}\tilde{K}_1\left(\tfrac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}}\right)|X_i, Z_i\right]\right)$$
$$+N^{-3/2}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_A(X_i,Z_i)\tilde{I}_{2i}\tilde{L}_{1i}[\tilde{\sigma}_1(X_i,Z_i)]^{-1}\tilde{J}_- \quad \tfrac{f_{X,Z}(X_i,Z_i)-q_{01}}{\tilde{\sigma}_1(X_i,Z_i)} \quad E \quad \tfrac{1}{\tilde{h}_{N1}^d}\tilde{K}_1 \quad \tfrac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}} \quad |X_i,Z_i \;\; -f_{X,Z}(X_i,Z_i)$$

The first of these is an order one degenrate U-process which satisfies the conditions of the equicontinuity lemma. Therefore the first term is $o_p(1)$ for each element of the family of the functions we consider. As for the second term, HIT claim that they can control the bias arising from that by $\overline{p}_1$-smoothness of the density $f_{X,Z}$. I don't necessaruily understand their comment. But using the rates of convergence in Silverman's article we can deal with this term as well. On the other hand, the analysis of the term involving $\tilde{J}_+$ is symmetric. The last step in this section is to repeat these arguments for

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta_A(X_i,Z_i)I_{1i}[\hat{I}_{2i}-I_{2i}]$$

B.5 Step 5:

$$\frac{1}{N}\sum_{i=1}^{N}\hat{I}_{1i}\hat{I}_{2i}=\frac{1}{N}\sum_{i=1}^{N}I_{1i}I_{2i}+\frac{1}{N}\sum_{i=1}^{N}\hat{I}_{1i}[\hat{I}_{2i}-I_{2i}]+\frac{1}{N}\sum_{i=1}^{N}I_{2i}[\hat{I}_{1i}-I_{1i}]$$

By the law of large numbers, the first term on the right hand side converges to $P(A_1\cap A_2)$. $N^{-1}|\sum_{i=1}^{N}\hat{I}_{1i}[\hat{I}_{2i}-I_{2i}]|\le N^{-1}\sum_{i=1}^{N}|\hat{I}_{2i}-I_{2i}|$. Our trimming assumptions guarantee that $E|\hat{I}_{2i}-I_{2i}|$ approaches 0 as N tends to infinity. Therefore, for each fixed $\kappa>0$,

$$P\left(N^{-1}\left|\sum_{i=1}^{N}\hat{I}_{1i}[\hat{I}_{2i}-I_{2i}]\right|>\kappa\right)\le P\left(N^{-1}\sum_{i=1}^{N}|\hat{I}_{2i}-I_{2i}|>\kappa\right)\le\frac{E|\hat{I}_{2i}-I_{2i}|}{\kappa}\to 0$$

Similarly, we can show that the last term is $o_p(1)$.

C Additional Estimation Results

(PRELIMINARY AND INCOMPLETE)

C.1 Asymptotic linearity of $\hat{P}(Z)$:

Note $P(Z)=E(D=1|Z)$ and thus fits into the form in HIT without any change. So under the assumptions of Theorem 3 of HIT, with $Y=D$, $X=Z$, and $d_z=dim(Z)$, for any $0\le p<\overline{p}_z$ the local polynomial regression estimation will yield

$$\left[\hat{P}(z)-P(z)\right]\hat{I}(z)=\frac{1}{n}\sum_{j=1}^{n}\psi_{nP}(Z_i,D_i;z)+\hat{b}_P(z)+\hat{R}_P(z)$$

where $n^{-1/2}\sum_{i=1}^{n}\hat{R}_P(Z_i)=o_p(1)$, $plim_{n\to\infty}n^{-1/2}\sum_{i=1}^{n}\hat{b}_P(Z_i)=b_P<\infty$, $E[\psi_{nP}(Z_i,D_i;Z|Z=z]=0$.

$$\psi_{nP}(Z_i,D_i;z)=\tfrac{1}{h_n^{d_z}}e_1[M_{p,n}(z)]^{-1}\left[\left(\tfrac{Z_i-z}{h_n}\right)^{Q_p}\right]'K\left(\tfrac{Z_i-z}{h_n}\right)\varepsilon_i^D I(z)$$

$$\hat{b}_P(z)=h_n^{\overline{p}_z}e_1[M_{pn}(z)]^{-1}\hat{I}(z)\sum_{s=p+1}^{\overline{p}_z}\left[\int u^{Q(0)}\cdot u^{Q(s)}P^{(s)}(z)'\cdot u^{Q(\overline{p}_z-s)}K(u)\mathrm{d}u,\right.$$
$$\left....,\int u^{Q(p)}\cdot u^{Q(s)}P^{(s)}(z)'\cdot u^{Q(\overline{p}_z-s)}K(u)\mathrm{d}u\right]f_Z^{(\overline{p}_z-s)}(z)'$$

where $P^{(s)}$ denotes the s-th order derivative of P, $\hat{I}(z)=1\{\hat{f}_Z(z)\geq q_{0Z}\}$ and $I(z)=1\{f_Z(z)\geq q_{0Z}\}$. And if $p=\overline{p}$, the estimator has the same form, but $\hat{b}_P(z)=o(h_n^{\overline{p}})$.

C.1.1 Asymptotic Linearity of $\hat{P}(Z)$ with different trimming:

The above formulation involves trimming based on the estimated density of Z. This is fine for this stage. But when we later estimate h_0 and h_1 we need to make sure that the joint density of X and Z is bounded away from 0. If in the first stage, our trimming function is $1\{\hat{f}_Z(z)\geq q_{01}\}$, this only guarantees that the true density of Z evaluated at z is above q_{01} with probability approaching to 1. If we have an (X,Z) value, say (x,z) such that $1\{\hat{f}_Z(z)\geq q_{01}\}=1$, and hence $f_Z(z)\geq q_{01}$, it is still possible that $f_{X,Z}(x,z)=0$. Therefore, even though, $E(D|X)=E(D|Z)$ I will trim based on the estimated density of (X,Z). If we change certain assumptions required for Theorem 3 of HIT everything goes through. The required changes are as follows:

1) For the estimated density of (X,Z) to converge uniformly in probability to the true density of (X,Z) we need $f_{X,Z}$ to be uniformly continuous.

2) For Lemma 4 to go through for $f_{X,Z}$ we need to assume that $f_{X,Z}$ is $\tilde{p}$-smooth, with $\tilde{p}>\dim(X,Z)$, and that $f_{X,Z}$ has a continuous Lebesgue density f_f in a neighborhood of q_{01} with $f_f(q_{01})>0$. So all this means that Assumption 4 in HIT has to be stated as: Trimming has to be $\tilde{p}$-nice on S with $\tilde{p}>\dim(X,Z)$.

Given these changes, everything goes through, and we get

$$\left[\hat{P}(z)-P(z)\right]\hat{I}_1(x,z)=\frac{1}{n}\sum_{j=1}^{n}\psi_{nP}(X_i,Z_i,D_i;x,z)+\hat{b}_P(z)+\hat{R}_P(z)$$

where $n^{-1/2}\sum_{i=1}^{n}\hat{R}_P(X_i,Z_i)=o_p(1)$, $plim_{n\to\infty}n^{-1/2}\sum_{i=1}^{n}\hat{b}_P(X_i,Z_i)=b_P<\infty$, $E[\psi_{nP}(X_i,Z_i,D_i;X,Z|X=x,Z=z]=0$.

$$\psi_{nP}(X_i,Z_i,D_i;x,z)=\tfrac{1}{h_n^{d_z}}e_1[M_{p,n}(z)]^{-1}\left[\left(\tfrac{Z_i-z}{h_n}\right)^{Q_p}\right]'K\left(\tfrac{Z_i-z}{h_n}\right)\varepsilon_i^D I_1(x,z)$$

$$\hat{b}_P(x,z) = h_n^{\overline{p}_z} e_1[M_{pn}(z)]^{-1}\hat{I}_1(x,z)\sum_{s=p+1}^{\overline{p}_z}\Big[\int u^{Q(0)}\cdot u^{Q(s)}P^{(s)}(z)'\cdot u^{Q(\overline{p}_z-s)}K(u)\mathrm{d}u,$$

$$\ldots,\int u^{Q(p)}\cdot u^{Q(s)}P^{(s)}(z)'\cdot u^{Q(\overline{p}_z-s)}K(u)\mathrm{d}u\Big] f_Z^{(\overline{p}_z-s)}(z)'$$

where $P^{(s)}$ denotes the s-th order derivative of P, $\hat{I}_1(x,z) = 1\{\hat{f}_{X,Z}(x,z)\geq q_0\}$ and $I_1(x,z) = 1\{f_{X,Z}(x,z)\geq q_0\}$. And if $p=\overline{p}$, the estimator has the same form, but $\hat{b}_P(z) = o(h_n^{\overline{p}})$.

Remark: We are estimating $E(D|Z)$. To compute and control the bias we have to assume that this function and the marginal density of Z is sufficiently smooth, in fact "the order" of smoothness has to be greater than the dimension of Z. But for the trimming based on (X,Z) to work, we need to assume that the joint distribution of (X,Z) is smooth with order of smoothness greater than the dimension of (X,Z) which is larger than the dimension of Z. Keep in mind that the latter assumption implies the former.

On the other hand, when we do the local polynomial regression estimation of h_0 and h_1, we are going to assume that $(X,P(Z))$ has a smooth Lebesgue density. This implies that $\text{Prob}(\{(x,z): P(z)=a\})=0$ for each $a\in[0,1]$. If we define $E := \{z\in \text{supp}(Z) : \exists(x,z)\in\text{supp}(X,Z)\}$ then for almost every $z\in E$, $|P'(z)|>0$. Using this, continuity[24] of P', continuity of $f_{X,Z}$ and the Lebesgue Differentiation Theorem, we could show that $f_{X,Z}(x_0,z_0)>0 \Rightarrow f_{X,P(Z)}(x_0,P(z_0))>0$. So based on this, we don't need to trim again in the second stage, i.e. the estimation of h_0 and h_1.

C.2 Estimating $h_0(x,P(z))$

$$\begin{aligned}
h_1(x,p) &= \frac{\partial}{\partial p}E(DY|X=x,P(Z)=p)\\
h_0(x,p) &= -\frac{\partial}{\partial p}E((1-D)Y|X=x,P(Z)=p)\\
h_1^{-1}h_0(x_0) &= \{x|\exists p, h_1(x,p)=h_0(x,p)\}\\
P(z) &= E(D|Z=z)\\
q(t_1,t_2) &= E(Y|D=1,h_1(X,P(Z))=t_1,P(Z)=t_2)
\end{aligned}$$

This appendix has two goals: first to show that the local polynomial regression estimator of h_0 is asymptotically linear with trimming; second, to show that its derivative with respect to p is uniformly consistent for the derivative of h_0 with respect to p.
To show that local polynomial regression estimator of h_0 is asymptotically linear with trimming, we follow arguments similar to those in the proof of theorem 3 of Heckman, Ichimura and Todd.

[24]When we do local polynomial regression estimation of $E(D|Z)$ we need to assume this anyway.

Write $Y = m + \varepsilon = X_{\overline{p}}(x_0)\beta^*_{\overline{p}}(x_0) + r_{\overline{p}}(X, x_0) + \varepsilon$, where $\varepsilon = Y - E(Y|X)$. In our case $-(1-D)Y$ will play the role of Y in HIT, and the vector $(P(Z), X)$ will play the role of X in HIT. In the first part of this section, we will use $\tilde{Y}$ and S to denote $(1-D)Y$ and $(P(Z), X)$. Just as HIT, we will consider the case, where p, the order of the polynomial terms included, is less than the underlying smoothness, $\overline{p}$, of the regression function. To do that partition, $S_{\overline{p}}(s_0) = [S_p(s_0), \overline{S}_{\overline{p}}(s_0)]$ and $\beta^*_{\overline{p}}(s_0) = [\beta^*_p(s_0)', \overline{\beta}^*_{\overline{p}}(s_0)']'$. Then,

$$\begin{aligned}[\hat{\beta}_p(s_0) - \beta^*_p(s_0)]\hat{I}_{10} &= H[\hat{M}_{pn}(s_0)]^{-1}n^{-1}H'S_p(s_0)'W(s_0)\varepsilon\hat{I}_{10} \\ &+ H[\hat{M}_{pn}(s_0)]^{-1}n^{-1}H'S_p(s_0)'W(s_0)\overline{S}_{\overline{p}}(s_0)\overline{\beta}^*_{\overline{p}}(s_0)\hat{I}_{10} \\ &+ H[\hat{M}_{pn}(s_0)]^{-1}n^{-1}H'S_p(s_0)'W(s_0)r_{\overline{p}}(s_0)\hat{I}_{10}\end{aligned}$$

where $\hat{I}_{10} = 1\{(x,z) : \hat{f}_{X,Z}(x_0, z_0) \geq q_{01}\}$, with $s_0 = (P(z_0), x_0)$. We need to show that $e_2[\hat{\beta}_p(s_0) - \beta^*_p(s_0)]\hat{I}_{10}$ is asymptotically linear.

C.2.1 First Step

As our first step, we would like to claim that

$$e_2H[\hat{M}_{pn}(s_0)]^{-1}n^{-1}H'S_p(s_0)'W(s_0)\varepsilon\hat{I}_{10} = e_2H[M_{pn}(s_0)]^{-1}n^{-1}H'S_p(s_0)'W(s_0)\varepsilon I_0 + \hat{R}_1(s_0)$$

where $e_2 = (0, 1, 0, ..., 0)$ and $1/\sqrt{n}\sum_{i=1}^n \hat{R}_1(S_i, X_i, Z_i) = o_p(1)$ Note that $e_2H = \frac{1}{h_n}e_2$. Let $\gamma_{n0}(S_j) = e_2[M_{pn}(S_j)]^{-1}$, $\hat{\gamma}_n(S_j) = e_2[\hat{M}_{pn}(S_j)]^{-1}$, $\overline{A}_1 = \{(x,z) : f_{X,Z}(x,z) \geq q_{01} - \epsilon_f > 0\}$, and

$$\Gamma_n = \{\gamma_n(x) | \sup_{x \in \overline{A}_1} |\gamma_n(x) - e_2[M_{pn}(x)]^{-1}| \leq \epsilon_\gamma\}$$

$$\mathcal{H}_1 = \{f : \sup_{(x,z)\in \text{supp}(X,Z)} |f(x,z) - f_{X,Z}(x,z)| \leq \epsilon_f\}$$

$$\mathcal{I}_1 = \{I((x,z) \in \tilde{A}) : \tilde{A} = \{(x,z) : f(x,z) \geq q_{01}\} \text{ for some } f \in \mathcal{H}_1\}$$

$$\mathcal{G}_{1n} = \left\{ g_n : g_n(\varepsilon_i, S_i, S_j, X_j, Z_j) = n^{-3/2}\gamma_n(S_j)\left(\frac{1}{h_n}\right)^{d+1}\left[\left(\frac{S_i - S_j}{h_n}\right)^{Q_p}\right]^T \varepsilon_i K\left(\frac{S_i - S_j}{h_n}\right)\tilde{I}_{1j} \right\}$$

Explanation: $e_2H = 1/(h_n)e_2$, so this is why we have $(1/(h_n)^{d+1})$ as opposed to $(1/(h_n)^d)$

Also let g_{n0} be the same as g_n except with γ_n replaced by γ_{n0}, and $\tilde{I}_{1j}$ replaced by I_{1j}. And define $\hat{g}_n$ similarly with $\hat{\gamma}_n$ and $\hat{I}_j$ replacing γ_n and $\tilde{I}_{1j}$, respectively. With this new notation $1/\sqrt{n}\sum_i \hat{R}_1(S_j, X_j, Z_j) = \sum_i\sum_j[\hat{g}_n(\varepsilon_i, S_i, S_j, X_j, Z_j) - g_{n0}(\varepsilon_i, S_i, S_j, X_j, Z_j)]$. To show that this sum is $o_p(1)$, we first need to show that $\sum_j\sum_i g_n(\varepsilon_i, S_i, S_j, X_j, Z_j)$ is equicontinuous over $\mathcal{G}_{1n}$ in a neighborhood of $g_{n0}(\varepsilon_i, S_i, S_j, X_j, Z_j)$ and that with probability approaching to 1, $\hat{g}_n(\varepsilon_i, S_i, S_j, X_j, Z_j)$ lies within the neighborhood over which equicontinuity is established. For the first step, we try using the third lemma. To apply that lemma, we need to have a degenerate U-process, and $\sum_i\sum_j g_n(\varepsilon_i, S_i, S_j, X_j, Z_j)$ is not degenerate.

First, we split the $\sum_i \sum_j g_n(\varepsilon_i, S_i, S_j, X_j, Z_j)$ process into two parts: $\sum_i \sum_j g_n(\varepsilon_i, S_i, S_j, X_j, Z_j) = \sum_i \sum_{j\neq i} g_n(\varepsilon_i, S_i, S_j, X_j, Z_j) + \sum_i g_n(\epsilon_i, S_i, S_i, X_i, Z_i)$. The latter process is symmetric. To see that it is also degenerate, we observe that $\left(\frac{S_i - S_i}{h_n}\right)^{Q_p}$ is a row vector whose first component equals 1 and all other components equal 0.

$$
\begin{aligned}
g_n(\varepsilon_i, S_i, S_i, X_i, Z_i) &= n^{-3/2}\gamma_n(S_i) e_1^T \varepsilon_i \left(\frac{1}{h_n}\right)^{d+1} K(0)\tilde{I}_{1i} \\
E[g_n(\varepsilon_i, S_i, S_i, X_i, Z_i)] &= n^{-3/2} E[\gamma_n(S_i)\tilde{I}_{1i} e_1^T E(\varepsilon_i | S_i, X_i, Z_i)] \left(\frac{1}{h_n}\right)^{d+1} K(0)
\end{aligned}
$$

$$
\begin{gathered}
E(\varepsilon_i | S_i, X_i, Z_i) = E(\varepsilon_i | P(Z_i), X_i, Z_i) = E\big[(1-D) g(\nu(X,D),\epsilon) \\
-E\big((1-D) g(\nu(X,D),\epsilon) | X, P(Z)\big) | X, Z, P(Z)\big] = E\big[(1-D) g(\nu(X,D),\epsilon) | X, Z, P(Z)\big] \\
-E\big((1-D) g(\nu(X,D),\epsilon) | X, P(Z)\big)
\end{gathered}
$$

The last equality holds because $E\big((1-D) g(\nu(X,D),\epsilon) | X, P(Z)\big)$ is measurable with respect to $\sigma(X, Z)$.

$$
\begin{gathered}
E\big[(1-D) g(\nu(X,D),\epsilon) | X, Z, P(Z)\big] = P(D=0|X,Z,P(Z)) E[g(\nu(X,0),\epsilon) | D=0, X, Z, P(Z)] \\
= P(U > P(Z) | X, Z, P(Z)) E[g(\nu(X,0),\epsilon) | D=0, X, Z, P(Z)] \\
= (1-P(Z)) E[g(\nu(X,0),\epsilon) | D=0, X, P(Z)]
\end{gathered}
$$

The independence of ϵ from Z was used in writing the last equality. On the other hand,

$$
\begin{gathered}
E\big((1-D) g(\nu(X,D),\epsilon) | X, P(Z)\big) = P(D=0|X,P(Z)) E[g(\nu(X,0),\epsilon) | D=0, X, P(Z)] \\
= (1-P(Z)) E[g(\nu(X,0),\epsilon) | D=0, X, P(Z)]
\end{gathered}
$$

Therefore, both $E(\varepsilon_i | S_i, X_i, Z_i)$ and $E[g_n(\varepsilon_i, S_i, S_i, X_i, Z_i)]$ are 0.

Next, define $g_n^0 := \frac{g_n(\varepsilon_i, S_i, S_j, X_j, Z_j) + g_n(\varepsilon_j, S_j, S_i, X_i, Z_i)}{2}$, $L_i := (\varepsilon_i, X_i, Z_i, P(Z_i))$, $\phi_n(L_i) := E[g_n^0(L_i, l) | L_i]$ $= E[g_n^0(l, L_i) | L_i]$, and $\tilde{g}_n^0(L_i, L_j) = g_n^0(L_i, L_j) - \phi_n(L_i) - \phi_n(L_j)$ as in HIT, so that $\sum_i \sum_{j\neq i} g_n^0(L_i, L_j) = \sum_i \sum_{j\neq i} \tilde{g}_n^0(L_i, L_j) + \sum_{i=1}^n 2(n-1)\phi_n(L_i)$. To show equicontinuity of our original process we need to show that that each of the processes $\sum_{i=1}^n g_n(\epsilon_i, S_i, S_i, X_i, Z_i)$, $\sum_i \sum_{j\neq i} \tilde{g}_n^0(L_i, L_j)$ and $\sum_{i=1}^n 2(n-1)\phi_n(L_i)$ are degenerate. We already verified that the first of these is degenerate. Next we show that the latter two are degenerate. Observe that

$$
\begin{aligned}
\phi_n(L_i) &= \frac{1}{2}\left(\frac{1}{h_n}\right)^{d+1} n^{-3/2} \epsilon_i E\left[\gamma_n(S_j)\tilde{I}(X_j, Z_j)\left(\left(\frac{S_i - S_j}{h_n}\right)^{Q_p}\right)^T K\left(\frac{S_i - S_j}{h_n}\right) | \varepsilon_i, X_i, Z_i, P(Z_i)\right] \\
&= \frac{1}{2}\left(\frac{1}{h_n}\right)^{d+1} n^{-3/2} \epsilon_i E\left[\gamma_n(S_j)\tilde{I}(X_j, Z_j)\left(\left(\frac{S_i - S_j}{h_n}\right)^{Q_p}\right)^T K\left(\frac{S_i - S_j}{h_n}\right) | \varepsilon_i, X_i, Z_i\right]
\end{aligned}
$$

To see that ϕ_n and $\tilde{g}_n^0$ are degenerate, first note that the conditional expectation term that appears in $\phi_n(L_i)$ can be thought of some function, say $\varphi(\varepsilon_i, X_i, Z_i)$. Then

$$\begin{aligned} E(\phi_n(L_i)|L_j) &= \frac{1}{2}\left(\frac{1}{h_n}\right)^{d+1} n^{-3/2} E\left(\varepsilon_i \varphi(\varepsilon_i, X_i, Z_i)|\varepsilon_j, X_j, Z_j\right) \\ &= \frac{1}{2}\left(\frac{1}{h_n}\right)^{d+1} n^{-3/2} E(\varepsilon_i \varphi(\varepsilon_i, X_i, Z_i)) = 0 = E(\phi_n(L_i)) \end{aligned}$$

Thus, all the processes are degenerate, and lemma 3 is applicable to each of them. Lemma 3 looks at a degenerate U-process over a separable class of functions $\Psi \subset \mathcal{L}_2$ and concludes that as long as 3 conditions hold, for each $\eta > 0$, there exists a $\delta > 0$ such that $\lim_{n\to\infty} P\big(\sup\{|U_n(\psi_{1n} - \psi_{2n})| > \eta : ||\psi_{1n} - \psi_{2n}||_2 \leq \delta\}\big) = 0$.

HIT assume that the limit of the $M_{pn}(s)$ matrix is positive definite. Therefore, $M_{pn}(s)$ matrix is positive definite, and hence invertible, when n is large. Based on this argument they say that the norm of γ_n will be finite for each $\gamma_n \in \Gamma_n$. Convinced with this argument I started verifying the three conditions of the lemma for each process.

Let $I_{1i}^* = 1\{f_{X,Z}(X_i, Z_i) \geq q_{01} - \epsilon_f\}$. Then $|g_n(\varepsilon_i, S_i, S_i, X_i, Z_i)| \leq n^{-3/2} C |e_1^T| |\varepsilon_i| K(0) I_{1i}^*$, and

$$\sum_{i=1}^{n} E\left[n^{-3} C^2 \varepsilon_i^2 \left(\frac{1}{h_n}\right)^{2(d+1)} K(0)^2 I_{1i}^*\right] \leq C^2 K(0)^2 E(\varepsilon^2)\left(\frac{1}{n h_n^{(d+1)}}\right)^2 < \infty$$

This shows that condition (i) of the equicontinuity lemma holds for the $\sum_i g_n(\varepsilon_i, S_i, S_i, X_i, Z_i)$ process if $nh_n^{d+1} \to \infty$. Condition (ii) holds under the same assumption by the dominated convergence theorem.

Next, we recall that $K(\cdot)$ is zero outside a compact set, so that when $\left|\left|\frac{S_i - S_j}{h_n}\right|\right|$ is "too large" $K\left(\frac{S_i - S_j}{h_n}\right) = 0$. This implies that there exist C_1, C_2 such that any element of $\left[\left(\frac{S_i - S_j}{h_n}\right)^{Q_p}\right]^T K\left(\frac{S_i - S_j}{h_n}\right)$ is bounded by $C_1 \left(\frac{1}{h_n}\right)^d 1\{||S_i - S_j|| \leq C_2 h_n\}$. Then

$$|g_n(\varepsilon_i, S_i, S_j, X_j, Z_j)| \leq n^{-3/2} C C_1 1\{||S_i - S_j|| \leq C_2 h_n\} \left(\frac{1}{h_n}\right)^{d+1} |\varepsilon_i| I_1(X_j, Z_j)^*$$

Thus, as long as $nh_n^{(d+1)} \to \infty$, conditions (i) and (ii) are satisfied for the process $\sum_i \sum_{j \neq i} g_n(\varepsilon_i, S_i, S_j, X_j, Z_j)$ as well.

$$|2n\phi_n(\varepsilon_i, X_i, Z_i)| \leq 2n^{-1/2} C |\varepsilon_i|$$

$$\sum_{i=1}^{n} 4n^{-1} C^2 E\left(\varepsilon_i^2\right) = 4C^2 E\left(\varepsilon_i^2\right) < \infty$$

Thus, the first condition of Lemma 3 of HIT holds for the $2n\phi$ process. On the other hand, since $E\left(\varepsilon_i^2\right) < \infty$, $\varepsilon_i^2 1\{|\varepsilon_i| > \sqrt{n}\frac{\delta}{2C}\} \to 0$ as $n \to \infty$, almost everywhere. Moreover, $\varepsilon_i^2 1\{|\varepsilon_i| > \sqrt{n}\frac{\delta}{2C}\} \leq \varepsilon_i^2$. Therefore, we could apply the Dominated Convergence Theorem to get that

$$\forall \delta > 0, \ \lim_{n\to\infty} \sum_{i=1}^{n} 4n^{-1} C^2 E\left(\varepsilon_i^2 1\left\{|\varepsilon_i| > \sqrt{n}\frac{\delta}{2C}\right\}\right) = 0$$

Now we move on to verifying condition (iii) of Lemma 3 of HIT for our three processes. First, let g_n^1 and g_n^2 be any two elements of $\mathcal{G}_{1n}$. Then

$$\begin{aligned} |g_n^1(\varepsilon_i, S_i, S_j, X_j, Z_j) - g_n^2(\varepsilon_i, S_i, S_j, X_j, Z_j)| &\leq n^{-3/2} C_1 1\{||S_i - S_j|| \leq C_2 h_n\} \left(\frac{1}{h_n}\right)^{d+1} |\varepsilon_i| \times \\ &\quad \left(|I_{1j}^1 - I_{1j}^2| \, ||\gamma_n^1(S_j)|| + |\gamma_n^1(S_j) - \gamma_n^2(S_j)| I_{1j}^2\right) \\ &\leq n^{-3/2} C_1 1\{||S_i - S_j|| \leq C_2 h_n\} \left(\frac{1}{h_n}\right)^{d+1} |\varepsilon_i| \times \\ &\quad \left(|I_{1j}^1 - I_{1j}^2| \, ||\gamma_n^1(S_j)|| + |\gamma_n^1(S_j) - \gamma_n^2(S_j)| I_{1j}^*\right) \end{aligned}$$

Similarly,

$$|2n\phi^1(Z_i) - 2n\phi^2(Z_i)| = \left(\frac{1}{h_n}\right)^{d+1} n^{-1/2} |\epsilon_i|$$

$$\left| E\left[\left(\tilde{I}_{1j}^1(\gamma_n^1(X_j) - \gamma_n^2(X_j)) + (\tilde{I}_{1j}^1 - \tilde{I}_{1j}^2)\gamma_n^2(X_j)\right)\left(\left(\frac{X_i - X_j}{h_n}\right)^{Q_p}\right)^T K\left(\frac{X_i - X_j}{h_n}\right) | X_i\right]\right|$$

Our function families are simpler - i.e. we don't have to worry about $\mathcal{A}$. Other than that the only difference between HIT and this, is that the exponent of h_n is $d+1$ instead of d in our case. They first verify that condition (iii) of lemma 3 is satisfied by Γ_n, $\mathcal{A}$ and $\mathcal{I}_1$. Then they say since $nh_n^d \to \infty$ the condition is also satisfied for $\{g_n(\epsilon_i, X_i, X_j)\}$, $\{g_n(\epsilon_i, X_i, X_i)\}$ and $\{2n\phi(Z)\}$ families. Our Γ_n and $\mathcal{I}_1$ is the same as theirs, and we don't need to worry about $\mathcal{A}$. Therefore, if we change assumptions 2 and 3 of HIT to $\overline{p} > d+1$ and $nh_n^{d+1}/\log n \to \infty$, we will be fine. The arguments so far showed equicontinuity of the process $\sum_{i=1}^n \sum_{j=1}^n g_n(\epsilon_i, S_i, S_j, X_j, Z_j)$ over $\mathcal{G}_{1n}$ in a neighborhood of $g_{n0}(\epsilon_i, S_i, S_j X_j, Z_j)$. Next we need to argue that $\hat{g}_n(\epsilon_i, S_i, S_j, X_j, Z_j)$ lies in that neighborhood. Given the modified assumption 3, this will be true if $\sup_x ||\hat{M}_{pn}(s) - M_{pn}(s)|| \to 0$ where $\lim_{n\to\infty} \inf_s \det\left(M_{pn}(s)\right) > 0$. Lemmas 5 and 6 of HIT take care of that.

C.2.2 Second Step:

Next, we move on to the term that will contain the bias:

$$\begin{aligned} & e_2 H[\hat{M}_{pn}(s_0)]^{-1} n^{-1} H^T S_p^T(s_0) W(s_0) \overline{S}_{\overline{p}}(s_0) \overline{\beta}_{\overline{p}}^*(s_0) \hat{I}_{10} \\ = \ & e_2[\hat{M}_{pn}(s_0)]^{-1} \hat{I}_{10} \sum_{k=p+1}^{\overline{p}} n^{-1} \left(\frac{1}{h_n}\right)^{d+1} \sum_{i=1}^{\overline{p}} \left[\left(\frac{S_i - s_0}{h_n}\right)^{Q_p}\right]^T (S_i - s_0)^{Q(k)} [m^{(k)}(s_0)]^T K\left(\frac{S_i - s_0}{h_n}\right) \end{aligned}$$

We add and subtract

$$e_2[\hat{M}_{pn}(s_0)]^{-1}\sum_{k=p+1}^{\overline{p}}\frac{1}{h_n^{d+1}}E\left\{\left[\left(\frac{S_i-S_j}{h_n}\right)^{Q_p}\right]^T(S_i-S_j)^{Q(k)}K\left(\frac{S_i-S_j}{h_n}\right)|S_j=s_0\right\}[m^{(k)}(s_0)]^T\hat{I}_{10}$$

This gives us three terms. But the difference of the two terms is handled in the same way as in lemma 2. In particular, we take γ_n, Γ_n, and $\mathcal{I}_1$ as before and define

$$g_n(S_i,S_j,X_j,Z_j)=n^{-3/2}\gamma_n(S_j)\left(\frac{1}{h_n}\right)^{d+1}\left[\left(\frac{S_i-S_j}{h_n}\right)^{Q_p}\right]^T(S_i-S_j)^{Q(k)}K\left(\frac{S_i-S_j}{h_n}\right)[m^{(k)}(S_j)]^T\tilde{I}_{1j}$$
$$-n^{-3/2}\gamma_n(S_j)\left(\frac{1}{h_n}\right)^{d+1}E\left[\left[\left(\frac{S_i-S_j}{h_n}\right)^{Q_p}\right]^T(S_i-S_j)^{Q(k)}K\left(\frac{S_i-S_j}{h_n}\right)|S_j\right][m^{(k)}(S_j)]^T\tilde{I}_{1j}$$

Let$\hat{g}_n(S_i,S_j,X_j,Z_j)$ and $g_{n0}(S_i,S_j,X_j,Z_j)$ be defined in the same way as before. Moreover, let $\mathcal{G}_{2n}:=\{g_n(S_i,S_j,X_j,Z_j)|\gamma_n(S_j)\in\Gamma_n,\ \tilde{I}_{1j}\in\mathcal{I}_1\}$. Then going through the same steps as in lemma 2 we can show that $1/\sqrt{n}\sum_{j=1}^n\hat{R}_{21}(S_j,X_j,Z_j)=\sum_{i=1}^n\sum_{j=1}^n[\hat{g}_n(S_i,S_j,X_j,Z_j)-g_{n0}(S_i,S_j,X_j,Z_j)]=o_p(1)$.
Then we deal with the term

$$e_2[\hat{M}_{pn}(s_0)]^{-1}\sum_{k=p+1}^{\overline{p}}\frac{1}{h_n^{d+1}}E\left\{\left[\left(\frac{S_i-S_j}{h_n}\right)^{Q_p}\right]^T(S_i-S_j)^{Q(k)}K\left(\frac{S_i-S_j}{h_n}\right)|S_j=s_0\right\}[m^{(k)}(s_0)]^T\hat{I}_{10}$$

which in turn equals

$$e_2\big([\hat{M}_{pn}(s_0)]^{-1}-[M_p(s_0)]^{-1}\big)\hat{I}_{10}$$
$$\cdot\textstyle\sum_{k=p+1}^{\overline{p}}\frac{1}{h_n^{d+1}}E\left\{\left[\left(\frac{S_i-S_j}{h_n}\right)^{Q_p}\right]^T(S_i-S_j)^{Q(k)}K\left(\frac{S_i-S_j}{h_n}\right)|S_j=s_0\right\}[m^{(k)}(s_0)]^T$$
$$+e_2[M_p(s_0)]^{-1}\hat{I}_{10}\textstyle\sum_{k=p+1}^{\overline{p}}\frac{1}{h_n^{d+1}}E\left\{\left[\left(\frac{S_i-S_j}{h_n}\right)^{Q_p}\right]^T(S_i-S_j)^{Q(k)}K\left(\frac{S_i-S_j}{h_n}\right)|S_j=s_0\right\}[m^{(k)}(s_0)]^T$$

The first expression can be treated in the same way as in lemma 2. The last expression equals

$$h_n^{\overline{p}-1}e_2[M_p(s_0)]^{-1}\times$$
$$\textstyle\sum_{k=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(k)}m^{(k)}(s_0)'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u,...,\int u^{Q(p)}\cdot u^{Q(k)}m^{(k)}(s_0)'\cdot u^{Q(\overline{p}-k)}K(u)\mathrm{d}u\right]$$
$$\times f^{(\overline{p}-k)}(s_0)'\hat{I}_{10}$$

We need

$$plim_{n\to\infty}1/\sqrt{n}\textstyle\sum_{i=1}^n h_n^{\overline{p}-1}e_2[M_p(S_i)]^{-1}\times$$
$$\textstyle\sum_{k=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(k)}m^{(k)}(S_i)'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u,...,\int u^{Q(p)}\cdot u^{Q(k)}m^{(k)}(S_i)'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u\right]$$
$$\times f^{(\overline{p}-k)}(S_i)'\hat{I}(X_i,Z_i)=b<\infty$$

All the terms involving (S_i, X_i, Z_i) are bounded with probability 1. Thus, if $nh_n^{2(\overline{p}-1)} \to a < \infty$ then we are OK. So I have to modify assumption 3.

C.2.3 Third Step:

Here we focus on the $e_2 H[\hat{M}_{pn}(s_0)]^{-1} n^{-1} H^T S_p^T(s_0) W(s_0) r_{\overline{p}}(s_0) \hat{I}_{10}$ term. Since $e_2 H = 1/h_n e_2$, this term equals

$$\frac{1}{h_n} e_2[\hat{M}_{pn}(s_0)]^{-1} n^{-1} H^T S_p^T(s_0) W(s_0) r_{\overline{p}}(s_0) \hat{I}_{10}$$

In the proof of lemma 8, HIT show that $n^{-1} H^T X_p^T(s_0) W(s_0) r_{\overline{p}}(s_0) = o_p(h_n^{\overline{p}+1})$. On the other hand, since $M_{pn}(s_0)$ converges to a positive definite matrix for each s_0 for which $f(s_0) > 0$[25], and since $\sup_s ||\hat{M}_{pn}(s) - M_{pn}(s)|| \to 0$, $\sup_s [\hat{M}_{pn}(s)]^{-1}$ will be finite for large n. Combining these arguments we get that

$$e_2 H[\hat{M}_{pn}(s_0)]^{-1} n^{-1} H^T S_p^T(s_0) W(s_0) r_{\overline{p}}(s_0) \hat{I}_{10} = o_p(h_n^{\overline{p}})$$

Under these assumptions given in the Appendix B,

$$\left[\hat{h}_0(p,x) - h_0(p,x)\right] \hat{I}_1(x,z) = \frac{1}{N} \sum_{j=1}^{N} \psi_{N h_0}(P(Z_i), X_i, D_i Y_i; x, z) + \hat{b}_{h_0}(p,x,z) + \hat{R}_{h_0}(p,x,z)$$

where $N^{-1/2} \sum_{i=1}^{N} \hat{R}_{h_0}(P(Z_i, X_i, Z_i) = o_p(1)$, $\text{plim}_{N\to\infty} N^{-1/2} \sum_{i=1}^{N} \hat{b}_{h_0}(P(Z_i), X_i, Z_i) = b_{h_0} < \infty$, and $E[\psi_{N h_0}(P(Z_i), X_i, D_i Y_i; P(Z_i), X_i, Z_i | P(Z_i) = p, X_i = x, Z_i = x] = 0$, $p = P(z)$. For the case when $\hat{h}_0$ is local polynomial regression estimator of h_0 of order $0 \le p < \overline{p}$:

$$\psi_{N h_0}(S_i, -(1-D_i)Y_i; p, x, z) = \tfrac{1}{h_N^{d+1}} e_2[M_{p,N}(s)]^{-1} \left[\left(\tfrac{S_i - s}{h_N}\right)^{Q_p}\right]^T K\left(\tfrac{S_i - s}{h_N}\right) \varepsilon_i^{h_0} I_1(x,z)$$

$$\begin{aligned}\hat{b}_{h_0}(p,x,z) = h_N^{\overline{p}-1} e_2[M_{p,N}(s)]^{-1} \hat{I}_1(x,z) \sum_{k=p+1}^{\overline{p}} \Bigg[&\int u^{Q(0)} \cdot u^{Q(k)} m^{(k)}(s)' \cdot u^{Q(\overline{p}-k)} K(u) \mathrm{d}u, \\ &..., \int u^{Q(p)} \cdot u^{Q(k)} m^{(k)}(s)' \cdot u^{Q(\overline{p}-k)} K(u) \mathrm{d}u \Bigg] f^{(\overline{p}-k)}(s)'\end{aligned}$$

with $S := (P(Z), X)$, $\varepsilon_i^{h_0} = -(1-D_i)Y_i - E[-(1-D_i)Y_i | P(Z_i), X_i]$ and $d = dim(S)$.

C.2.4 Asymptotic linearity of $\hat{h}_0(\hat{P}(z), x)$:

To show this, we need to use Lemma 1 of HIT. Recall that

[25] The previous section argues that,under our assumptions, $f_{X,Z}(x_0, z_0) > 0$ implies that $f_{X,P(Z)}(x_0, P(z_0)) > 0$.

Lemma C.1 *(HIT) Suppose that:*

1. *Both* $\hat{P}(z)$ *and* $\hat{g}(p,t)$ *are asymptotically linear with trimming where*

$$\big[\hat{P}(z)-P(z)\big]I\big((x,z)\in\hat{A}_1\big)=n^{-1}\sum_{j=1}^{n}\psi_{nP}(D_j,Z_j;x,z)+\hat{b}_P(x,z)+\hat{R}_P(x,z)$$

$$\big[\hat{g}(p,t)-g(p,t)\big]I\big((x,z)\in\hat{A}_1\big)=n^{-1}\sum_{j=1}^{n}\psi_{ng}(Y_j,T_j,P(Z_j);p,t,z)+\hat{b}_g(p,t,z)+\hat{R}_g(p,t,z);$$

2. $\partial\hat{g}(p,t)/\partial p$ *and* $\hat{P}(z)$ *are uniformly consistent and converge to* $\partial g(p,t)/\partial p$ *and* $P(z)$*, respectively and* $\partial g(p,t)/\partial p$ *is continuous;*

3. $\text{plim}_{n\to\infty}n^{-1/2}\sum_{i=1}^{n}\hat{b}_g(P(Z_i),T_i,Z_i)=b_g$ *and* $\text{plim}_{n\to\infty}n^{-1/2}\sum_{i=1}^{n}\frac{\partial g(P(Z_i),T_i)}{\partial p}\hat{b}_P(P(Z_i),T_i,Z_i)=b_{gP}$*;*

4. $\text{plim}_{n\to\infty}n^{-1/2}\sum_{i=1}^{n}\left[\frac{\partial\hat{g}(\overline{P}_{T_i}(Z_i),T_i)}{\partial p}-\frac{\partial g(P(Z_i),T_i)}{\partial p}\right]\hat{R}_P(P(Z_i),T_i,Z_i)=0$*, and* $\text{plim}_{n\to\infty}n^{-1/2}\sum_{i=1}^{n}\left[\frac{\partial\hat{g}(\overline{P}_{T_i}(Z_i),T_i)}{\partial p}-\frac{\partial g(P(Z_i),T_i)}{\partial p}\right]\hat{b}_P(P(Z_i),T_i,Z_i)=0$*;*

5. $\text{plim}_{n\to\infty}n^{-3/2}\sum_{i=1}^{n}\left[\frac{\partial\hat{g}(\overline{P}_{T_i}(Z_i),T_i)}{\partial p}-\frac{\partial g(P(Z_i),T_i)}{\partial p}\right]\psi_{nP}(D_j,Z_j;T_i,Z_i)=0.$

then $\hat{g}(\hat{P}(z),t)$ *is also asymptotically linear with trimming where*

$$\begin{aligned}\big[\hat{g}(\hat{P}(z),t)-g(P(z),t)\big]I\big((x,z)\in\hat{A}_1\big) &= n^{-1}\sum_{j=1}^{n}\big[\psi_{ng}(Y_j,T_j,P(Z_j),Z_j;P(z),t,z)\\ &+ \partial g(t,P(z))/\partial p\ \cdot\psi_{nP}(D_j,Z_j,X_j;x,z)\big]\\ &+ \hat{b}^g(x,z)+\hat{R}^g(x,z),\end{aligned}$$

and $\text{plim}_{n\to\infty}\sum_{i=1}^{n}\hat{b}^g(X_i,Z_i)=b_g+b_{gP}$.

In our case, $g(p,x)=\frac{\partial}{\partial p}E\big(-(1-D)Y|P(Z)=p,X=x\big)$. The verification of the conditions for Lemma 1 of HIT for the case where g itself is the derivative of some conditional expectation with respect to one of the conditioning variables is not really different from what HIT have. The only potential difference is in the proof of theorem 4, but even there, their argument holds for the entire $\nabla\hat{\beta}$ vector, not just the first component.
All these arguments show that

$$\begin{aligned}\Big[\hat{h}_0(\hat{P}(z),x)-h_0(P(z),x)\Big]I\Big((x,z)\in\hat{A}_1\Big) &= N^{-1}\sum_{j=1}^{N}\Big[\psi_{Nh_0}(-(1-D_j)Y_j,P(Z_j),X_j;P(z),x,z)+\frac{\partial h_0(P(z),x)}{\partial p}\psi_{nP}(D_j,Z_j;x,z)\Big]\\ &+ \hat{b}_{\hat{h}_0}(x,z)+\hat{R}_{\hat{h}_0}(x,z)\end{aligned}$$

with $\text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\hat{b}_{\hat{h}_0}(X_j,Z_j) = b_{h_0} + b_{h_0P} < \infty$, $\text{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\hat{R}_{\hat{h}_0}(X_j,Z_j) = 0$, and $\psi_{Nh_0}(-(1-D_j)Y_j, P(Z_j), X_j; P(z), x, z) = \varepsilon_j^{h_0} e_2[M_{pN}(s)]^{-1}\left[\left(\frac{S_j-s}{h_{Nh_0}}\right)^{Q_p}\right]' \frac{1}{h_{Nh_0}^{d+1}} K\left(\frac{S_j-s}{h_{Nh_0}}\right) I_1(x,z)$, with $S = (P(Z), X)$, and $\varepsilon_j^{h_0} = -(1-D_j)Y_j - E\big[-(1-D_j)Y_j|S_j\big]$; and $\psi_{NP}(D_j, Z_j; x, z) = \varepsilon_j^P e_1[M_{pN}(z)]^{-1}\left[\left(\frac{Z_j-z}{h_{NP}}\right)^{Q_p}\right]' \frac{1}{h_{NP}^{d_z}} K\left(\frac{Z_j-z}{h_{NP}}\right) I_1(x,z)$, $\varepsilon_j^P = D_j - E[D_j|Z_j]$. Define

$$\psi_{Nh_0P}(D_j, Y_j, X_j, Z_j; x, z) := \psi_{Nh_0}(-(1-D_j)Y_j, P(Z_j), X_j; P(z), x, z) + \frac{\partial h_0(P(z), x)}{\partial p}\psi_{NP}(D_j, Z_j; x, z)$$

C.3 Estimating $q(h_0(x, P(z)), P(z))$

We need to estimate $E(Y|D=1, h_1(X, P(Z)), P(Z))$. But

$$E(Y|D=1, h_1(X,P(Z)), P(Z)) = \frac{E(DY|h_1(X,P(Z)),P(Z))}{P(D=1|h_1(X,P(Z)),P(Z))} = \frac{E(DY|h_1(X,P(Z)),P(Z))}{P(Z)}$$

We could use local polynomial regression to estimate $E(DY|h_1(X,P(Z)),P(Z))$. Therefore the analysis here is very similar to the proof of their theorem 3. The only difference is that we evaluate this estimator at the value of the random vector $(h_0(X_i,P(Z_i)),P(Z_i))$, which is different from the random vector we condition on. As long as the support of $h_0(X_i,P(Z_i))$ is contained in the support of $h_1(X_i,P(Z_i))$ this is well defined. To simplify the following expressions, define $T_{1i} := (h_1(X_i,P(Z_i)),P(Z_i))$, and $T_{0i} := (h_0(X_i,P(Z_i)),P(Z_i))$. Let t_1 and t_0 denote a value in the interior of the support of T_1 and T_0, respectively. And let p denote that point in the interior of the support of $P(Z)$ that corresponds to t_0. Note that here, $d=2$.

Let $\hat{I}_{1i} := 1\{\hat{f}_{X,Z}(X_i,Z_i) \geq q_{01}\}$, $I_{1i} := 1\{f_{X,Z}(X_i,Z_i) \geq q_{01}\}$, $\hat{I}_{2i} := 1\{\hat{f}_{\hat{h}_1,\hat{P}}(\hat{h}_0(X_i,\hat{P}(Z_i)),\hat{P}(Z_i)) \geq q_{02}\}$, and $I_{2i} := 1\{f_{h_1,P}(h_0(X_i,P(Z_i)),P(Z_i)) \geq q_{02}\}$. Our goal is to derive the asymptotic distribution of

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[D_i\big(\hat{q}(h_{0i},P_i) - q(h_{0i},P_i)\big)\right]\hat{I}_1(X_i,Z_i)\hat{I}_2(X_i,Z_i)$$

Let $W(t_0) := h_{Nq}^{-2} diag\left(K\left(\frac{T_{11}-t_0}{h_{Nq}}\right), \ldots, K\left(\frac{T_{1N}-t_0}{h_{Nq}}\right)\right)$, $\varepsilon_i^q := D_iY_i - E(D_iY_i|T_{1i})$.

$$T_p(t_0) := \begin{pmatrix} (T_{11}-t_0)^{Q_p} \\ \vdots \\ (T_{1N}-t_0)^{Q_p} \end{pmatrix}$$

Now

$$\begin{aligned}\hat{q}(t_0) - q(t_0) &= \frac{1}{p} e_1[\hat{M}_{pN}(t_0)]^{-1} N^{-1} H' T_p'(t_0) W(t_0) \varepsilon^q \\ &+ \frac{1}{p} e_1[\hat{M}_{pN}(t_0)]^{-1} N^{-1} H' T_p(t_0)' W(t_0) \overline{T}_{\overline{p}}(t_0) \overline{\beta}^*_{\overline{p}}(t_0) \\ &+ \frac{1}{p} e_1[\hat{M}_{pN}(t_0)]^{-1} N^{-1} H' T_p'(t_0) W(t_0) r_{\overline{p}}(t_0)\end{aligned}$$

Therefore,

$$\begin{aligned}&\tfrac{1}{\sqrt{N}} \textstyle\sum_{j=1}^N D_j [\hat{q}(T_{0j}) - q(T_{0j})] \hat{I}_{1j} \hat{I}_{2j} = \tfrac{1}{\sqrt{N}} \textstyle\sum_{j=1}^N \tfrac{D_j}{P(Z_j)} e_1[\hat{M}_{pN}(T_{0j})]^{-1} N^{-1} H' T_p'(T_{0j}) W(T_{0j}) \varepsilon^q \hat{I}_{1j} \hat{I}_{2j} \\ &+ \tfrac{1}{\sqrt{N}} \textstyle\sum_{j=1}^N \tfrac{D_j}{P(Z_j)} e_1[\hat{M}_{pN}(T_{0j})]^{-1} N^{-1} H' T_p(T_{0j})'(T_{0j}) W(T_{0j}) \overline{T}_{\overline{p}}(T_{0j}) \overline{\beta}^*_{\overline{p}}(T_{0j}) \hat{I}_{1j} \hat{I}_{2j} \\ &+ \tfrac{1}{\sqrt{N}} \textstyle\sum_{j=1}^N \tfrac{D_j}{P(Z_j)} e_1[\hat{M}_{pN}(T_{0j})]^{-1} N^{-1} H' T_p'(T_{0j}) W(T_{0j}) r_{\overline{p}}(T_{0j}) \hat{I}_{1j} \hat{I}_{2j}\end{aligned}$$

C.3.1 First Term:

Let's start with the first term. Add and subtract

$$\frac{1}{\sqrt{N}} \sum_{j=1}^N \frac{D_j}{P(Z_j)} e_1[M_{pN}(T_{0j})]^{-1} N^{-1} H' T_p'(T_{0j}) W(T_{0j}) \varepsilon^q I_{1j} I_{2j}$$

We will argue that

$$\begin{aligned}&\frac{1}{N\sqrt{N}} \sum_{j=1}^N \sum_{i=1}^{N_1} \frac{D_j}{P(Z_j)} e_1[\hat{M}_{pN}(T_{0j})]^{-1} \left[\left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right)^{Q_p} \right]' h_{Nq}^{-2} \varepsilon_i^q K\left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right) \hat{I}_{1j} \hat{I}_{2j} \\ -&\frac{1}{N\sqrt{N}} \sum_{j=1}^N \sum_{i=1}^{N_1} \frac{D_j}{P(Z_j)} e_1[M_{pN}(T_{0j})]^{-1} \left[\left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right)^{Q_p} \right]' h_{Nq}^{-2} \varepsilon_i^q K\left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right) I_{1j} I_{2j} = o_p(1)\end{aligned}$$

Now we will go through arguments as in lemmas 2 through 6 of HIT, and try to show that this sum is $o_p(1)$. To do that define $\gamma_{N0}(T_{0j}) = e_1[M_{pN}(T_{0j})]^{-1}$, $\hat{\gamma}_N(T_{0j}) = e_1[\hat{M}_{pN}(T_{0j})]^{-1}$, $\overline{A}_1 = \{(x,z) : f_{X,Z}(x,z) \geq q_{01} - \epsilon_f > 0\}$, $A_1 := \{(x,z) : f_{X,Z}(x,z) \geq q_{01}\}$, $A_2 := \{(x,z) : f_{h_1(X,P(Z)),P(Z)}(h_0(x,P(z)),P(z)) \geq q_{02}\}$ and

$$\Gamma_N = \{\gamma_N(x) | \sup_{x \in \overline{A}_1} |\gamma_N(x) - e_1[M_{pN}(x)]^{-1}| \leq \epsilon_\gamma\}$$

$$\begin{aligned}\mathcal{G}_{1N} := \Bigg\{ g_N : &\, g_N(\varepsilon_i^q, T_{1i}; D_j, T_{0j}, X_j, Z_j) \\ &= N^{-3/2} \gamma_N(T_{0j}) \frac{D_j}{P(Z_j)} \left[\left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right)^{Q_p} \right]' h_{Nq}^{-2} \varepsilon_i^q K\left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right) \tilde{I}_{1j} \tilde{I}_{2j} \Bigg\}\end{aligned}$$

$$g_{N0}(\varepsilon_i^q, T_{1i}; D_j, T_{0j}, X_j, Z_j) = N^{-3/2}\gamma_{N0}(T_{0j})\frac{D_j}{P(Z_j)}\left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right) I_{1j}I_{2j}$$

Let $\mathcal{H}_1$, $\mathcal{I}_1$, $\mathcal{H}_2$ and $\mathcal{I}_2$ be as in Appendix B. We are going to try to show that the process $\sum_{i=1}^N \sum_{j=1}^N g_N(\varepsilon_i^q, T_{1i}; D_j, T_{0j}, X_j, Z_j)$ is equicontinuous over $\mathcal{G}_{1n}$ in a neighborhood of $g_{N0}(\varepsilon_i^q, T_{1i}; D_j, T_{0j}, X_j, Z_j)$, and that $\hat{g}_N(\varepsilon_i^q, T_{1i}; D_j, T_{0j}, X_j, Z_j)$ lies in the neighborhood over which we establish equicontinuity with probability approaching to 1.

We will first check if the associated processes are degenerate:

$$E\big[g_N(\varepsilon_i^q, T_{1i}; D_i, T_{0i})\big] = N^{-3/2}E\left\{\gamma_N(T_{0i})\frac{h_{Nq}^{-2}}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' K\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}E[D_i\varepsilon_i^q|X_i, Z_i]\right\}$$

with $\varepsilon_i^q = D_iY_i - E(D_iY_i|h_1(X_i, P(Z_i)), P(Z_i))$.

$$\begin{aligned}
E\big[D_i\varepsilon_i^q|X_i, Z_i\big] &= E[D_iY_i|X_i, Z_i] - E\big[D_iE(D_iY_i|h_1(X_i, P(Z_i)), P(Z_i))|X_i, Z_i\big] \\
&= E[D_iY_i|X_i, Z_i] - E(D_i|X_i, Z_i)E(D_iY_i|h_1(X_i, P(Z_i)), P(Z_i)) \\
&= E[D_iY_i|X_i, Z_i] - P(Z_i)E(D_iY_i|h_1(X_i, P(Z_i)), P(Z_i))
\end{aligned}$$

This does not have zero expectation. Therefore the associated U-process is not degenerate. But we can remedy this by adding and subtracting the expectation of g_N from it. Let

$$\begin{aligned}
\tilde{g}_N(\varepsilon_i^q, T_{1i}; D_i, T_{0i}) &= N^{-3/2}\gamma_N(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right) \\
&- E\left\{N^{-3/2}\gamma_N(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\right\}
\end{aligned}$$

This process is degenerate, and satisfies all the conditions of the lemma 3 of HIT. It is thus equicontinuous. But this is only one piece of the $\sum_{i=1}^N g_N(\varepsilon_i^q, T_{1i}; D_i, T_{0i})$ process. The other piece is

$$\begin{aligned}
&\sum_{i=1}^N E\left\{N^{-3/2}\gamma_N(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}\right\} \\
&= N^{-1/2}E\left\{\gamma_N(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}\right\}
\end{aligned}$$

We have to make sure that the limit of this is 0. We know that any element of $\left[\ \frac{T_{1i}-T_{0i}}{h_{Nq}}\ ^{Q_p}\right]' K_h(T_{1i} - T_{0i})$ is bounded by $C_1 h_{Nq}^{-2} I\{||T_{1i} - T_{0i}|| \leq C_2 h_{Nq}\}$ for some finite C_1 and C_2. On the other hand, $|D_i| \leq 1$, $|\tilde{I}_{1i}\tilde{I}_{2i}| \leq 1$ $E|\varepsilon_i^q| < \infty$, and $P(Z_i)$ is almost surely bounded away from 0. Combining

these facts with $Nh_{Nq}^4 \to \infty$, we get that the desired limit is in fact 0. Now $\sum_{i=1}^N \big(g_N(i) - E(g_N)\big)$ is a degenerate U-process, which satisfies the conditions of the equicontinuity lemma.
Next, we focus on the part containing different indices. Let $S_i := (\varepsilon_i^q, D_i, T_{1i}, T_{0i}, X_i, Z_i)$. Define

$$\begin{aligned} g_N^0(S_i, S_j) &= \frac{1}{2} N^{-3/2} \gamma_N(T_{0j}) \frac{D_j}{P(Z_j)} \left[\left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right)^{Q_p} \right]' h_{Nq}^{-2} \varepsilon_i^q K \left(\frac{T_{1i} - T_{0j}}{h_{Nq}} \right) \tilde{I}_{1j} \tilde{I}_{2j} \\ &+ \frac{1}{2} N^{-3/2} \gamma_N(T_{0i}) \frac{D_i}{P(Z_i)} \left[\left(\frac{T_{1j} - T_{0i}}{h_{Nq}} \right)^{Q_p} \right]' h_{Nq}^{-2} \varepsilon_j^q K \left(\frac{T_{1j} - T_{0i}}{h_{Nq}} \right) \tilde{I}_{1i} \tilde{I}_{2i} \end{aligned}$$

Define $\phi_N(S_i) = E[g_N^0(S_i, S_j)|S_i]$. Then,

$$\phi_N(S_i) = E[g_N^0(S_i, S_j)|S_i] = \tfrac{1}{2} N^{-3/2} h_{Nq}^{-2} \times$$

$$E\Big\{\gamma_N(T_{0j})\tfrac{D_j}{P(Z_j)}\Big[\ \tfrac{T_{1i}-T_{0j}}{h_{Nq}}\ ^{Q_p}\Big]'\varepsilon_i^q K\ \tfrac{T_{1i}-T_{0j}}{h_{Nq}}\ \tilde{I}_{1j}\tilde{I}_{2j}+\gamma_{N_1}(T_{0i})\tfrac{D_i}{P(Z_i)}\varepsilon_j^q\Big[\ \tfrac{T_{1j}-T_{0i}}{h_{Nq}}\ ^{Q_p}\Big]'K\ \tfrac{T_{1j}-T_{0i}}{h_{Nq}}\ \tilde{I}_{1i}\tilde{I}_{2i}|\varepsilon_i^q,D_i,T_{1i},T_{0i},X_i,Z_i\Big\}$$

Note that

$$\begin{aligned} & E\left\{\gamma_N(T_{0i})\tfrac{D_i}{P(Z_i)}\varepsilon_j^q\left[\left(\tfrac{T_{1j}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' K\left(\tfrac{T_{1j}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}|\varepsilon_i^q, D_i, T_{1i}, T_{0i}, X_i, Z_i\right\} \\ &= \gamma_N(T_{0i})\tfrac{D_i}{P(Z_i)}\tilde{I}_{1i}\tilde{I}_{2i}E\Big\{E\ \ \varepsilon_j^q\Big[\ \tfrac{T_{1j}-T_{0i}}{h_{Nq}}\ ^{Q_p}\Big]'K\ \tfrac{T_{1j}-T_{0i}}{h_{Nq}}\ |\varepsilon_i^q,D_i,T_{1i},T_{0i},X_i,Z_i,T_{1j}\Big)|\varepsilon_i^q,D_i,T_{1i},T_{0i},X_i,Z_i\Big\} \\ &= \gamma_N(T_{0i})\tfrac{D_i}{P(Z_i)}\tilde{I}_{1i}\tilde{I}_{2i}E\Big\{\Big[\ \tfrac{T_{1j}-T_{0i}}{h_{Nq}}\ ^{Q_p}\Big]'K\ \tfrac{T_{1j}-T_{0i}}{h_{Nq}}\ E\big(\varepsilon_j^q|\varepsilon_i^q,D_i,T_{1i},T_{0i},X_i,Z_i,T_{1j}\big)|\varepsilon_i^q,D_i,T_{1i},T_{0i},X_i,Z_i\Big\} \\ &= \gamma_N(T_{0i})\tfrac{D_i}{P(Z_i)}\tilde{I}_{1i}\tilde{I}_{2i}E\left\{\left[\left(\tfrac{T_{1j}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' K\left(\tfrac{T_{1j}-T_{0i}}{h_{Nq}}\right)E\left(\varepsilon_j^q|T_{1j}\right)|\varepsilon_i^q, D_i, T_{1i}, T_{0i}, X_i, Z_i\right\} = 0 \end{aligned}$$

Therefore,

$$\begin{aligned} \phi_N(S_i) &= \tfrac{1}{2}N^{-3/2}h_{Nq}^{-2}E\Big\{\gamma_N(T_{0j})\tfrac{D_j}{P(Z_j)}\Big[\ \tfrac{T_{1i}-T_{0j}}{h_{Nq}}\ ^{Q_p}\Big]'\varepsilon_i^q K\ \tfrac{T_{1i}-T_{0j}}{h_{Nq}}\tilde{I}_{1j}\tilde{I}_{2j}\ |\varepsilon_i^q,D_i,T_{1i},T_{0i},X_i,Z_i\Big\} \\ &= \frac{1}{2}N^{-3/2}h_{Nq}^{-2}\varepsilon_i^q E\left\{\gamma_{N_1}(T_{0j})\frac{D_j}{P(Z_j)}\left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' K\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\tilde{I}_{1j}\tilde{I}_{2j}|T_{1i}\right\} \end{aligned}$$

This is of the form $\varphi_N(T_{1i})\varepsilon_i^q$, and $E\big(\varphi_N(T_{1i})\varepsilon_i^q\big) = E\Big[\varphi_N(T_{1i})E(\varepsilon_i^q|T_{1i})\Big] = 0$. Thus we can define $\tilde{g}_N^0(S_i, S_j) := g_N^0(S_i, S_j) - \phi_N(S_i) - \phi_N(S_j)$. The process $\sum_i \sum_{j \neq i} \tilde{g}_N^0(S_i, S_j)$ is a degenerate U-process of order two. On the other hand, the above calculations show that $\sum_i 2(N-1)\phi_N(S_i)$ is a degenerate order one process. Since $|D_i| \leq 1$ and $P(Z_i)$ is bounded away from 0, and $\tilde{I}_{1i}\tilde{I}_{2i} \leq \tilde{I}_{1i} \leq I_{1i}^* := 1\{f_{X,Z}(X_i, Z_i) \geq q_{01} - \epsilon_{f1}\}$ the same steps as on p. 287 of HIT prove that each of these processes satisfies the first two conditions of the equicontinuity lemma. For the third condition, take any $g_N^{(1)}, g_N^{(2)} \in \mathcal{G}_{1N}$.

$$\begin{aligned}
|g_N^{(1)} - g_N^{(2)}| &= \left| N^{-3/2} \tfrac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right| \\
&\times \left| \gamma_N^{(1)}(T_{0j}) \left[\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(1)} \tilde{I}_{2j}^{(1)} - \gamma_N^{(2)}(T_{0j}) \tfrac{D_j}{P(Z_j)} \left[\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(2)} \tilde{I}_{2j}^{(2)} \right| \\
&\leq \left| N^{-3/2} \tfrac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right| \left| \left(\gamma_N^{(1)}(T_{0j}) - \gamma_N^{(2)}(T_{0j})\right) \left[\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(1)} \tilde{I}_{2j}^{(1)} \right| \\
&+ \left| N^{-3/2} \tfrac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right| \left| \gamma_N^{(2)}(T_{0j}) \left[\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{2j}^{(1)} \right| \left| \tilde{I}_{1j}^{(1)} - \tilde{I}_{1j}^{(2)} \right| \\
&+ \left| N^{-3/2} \tfrac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right| \left| \gamma_N^{(2)}(T_{0j}) \left[\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(2)} \right| \left| \tilde{I}_{2j}^{(1)} - \tilde{I}_{2j}^{(2)} \right|
\end{aligned}$$

Therefore, the third condition of the equicontinuity lemma will hold, if each of the families $\Gamma_N, \mathcal{I}_1, \mathcal{I}_2$ satisfy it. To verify that these families satisfy that condition, we will use lemma 4 of HIT. For Γ_N to satisfy this condition, we need that both $E[DY|h_1(X,P(Z)),P(Z)]$ and $f_{h_1(X,P(Z)),P(Z)}$ are both $\overline{p}_2$ smooth where $\overline{p}_2 + \alpha_2 > \dim(X,P(Z))$ with α_2 equal to the smaller of the Holder continuity constant of these functions. At the same time, if $f_{h_1(X,P(Z)),P(Z)}$ satisfies this smoothness condition, and if the first derivative of $f_{h_1(X,P(Z)),P(Z)}$ is uniformly continuous and the Lebesgue density $f_{f_{h_1,P}}$ of $f_{h_1(X,P(Z)),P(Z)}$ is continuous in a neighborhood of q_{02} with $f_{f_{h_1,P}}(q_{02}) > 0$, the third condition of the equicontinuity lemma is satisfied for $\mathcal{I}_2$. Similarly, if $f_{X,Z}$ is $\overline{p}_1$ smooth where $\overline{p}_1 + \alpha_1 > \dim(X,Z)$ with α_1 equal to the Holder continuity constant, if the first derivative of $f_{X,Z}$ is uniformly continuous and the Lebesgue density $f_{f_{X,Z}}$ of $f_{X,Z}$ is continuous in a neighborhood of q_{01} with $f_{f_{X,Z}}(q_{01}) > 0$, the third condition of the equicontinuity lemma is satisfied for $\mathcal{I}_1$.

Combining all these results, we conclude that the process $\sum_{j=1}^N \sum_{i=1}^N g_N(\varepsilon_i^q, T_{1i}, T_{0j})$ is equicontinuous over $\mathcal{G}_{1N}$ in a neighborhood of $g_{N0}(\varepsilon_i^q, T_{1i}, D_j, T_{0j}, X_j, Z_j)$.

Lemma 5 and 6 of HIT can be used to show that $\sup_{(x,z)\in A_1\cap A_2} ||\hat{M}_{pN}(h_0(x,P(z)),P(z)) - M_{pN}(h_0(x,P(z)),P(z))|| \to 0$. This result combined with the arguments at the beginning shows that $\hat{g}_N(\varepsilon_i^q, T_{1i}, T_{0j})$ lies in the neighborhood of $g_{N0}(\varepsilon_i^q, T_{1i}, T_{0j}, X_j, Z_j)$ over which equicontinuity was shown.

C.3.2 Second Term:

Next, we look at

$$\frac{1}{\sqrt{N}} \sum_{j=1}^N D_j e_1 [\hat{M}_{pN}(T_{0j})]^{-1} N^{-1} H' T_p(T_{0j})'(T_{0j}) W(T_{0j}) \overline{T}_{\overline{p}}(T_{0j}) \overline{\beta}_{\overline{p}}^*(T_{0j}) \hat{I}_{1j} \hat{I}_{2j}$$

Fix the evaluation point (d_0, x_0, z_0) such that $(x_0, z_0) \in A_1 \cap A_2$. Let $P_0 = P(z_0)$, $t_0 = (h_0(x_0, P(z_0)), P(z_0)$. Then each term in this sum equals:

$$\begin{aligned} & e_1 H[\hat{M}_{pN}(t_0)]^{-1} N^{-1} \frac{d_0}{P_0} H' T_p'(t_0) W(t_0) \overline{T}_{\overline{p}}(t_0) \overline{\beta}^*_{\overline{p}}(t_0) \hat{I}_{10} \hat{I}_{20} \\ = & e_1[\hat{M}_{pN}(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} N^{-1} \frac{1}{h_{Nq}^2} \sum_{i=1}^{\overline{p}} \frac{d_0}{P_0} \left[\left(\frac{T_{1i} - t_0}{h_{Nq}} \right)^{Q_p} \right]' (T_{1i} - t_0)^{Q(s)} [m^{(s)}(t_0)]' K \left(\frac{T_{1i} - t_0}{h_{Nq}} \right) \hat{I}_{10} \hat{I}_{20} \end{aligned}$$

We add and subtract

$$e_1[\hat{M}_{pN}(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} \frac{1}{h_{Nq}^2} E\left\{ \frac{D_j}{P(Z_j)} \left[\frac{T_{1i}-T_{0j}}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-T_{0j})^{Q(s)} K \frac{T_{1i}-T_{0j}}{h_{Nq}} |T_{0j}=t_0, X_j=x_0, Z_j=z_0 \right\} [m^{(s)}(t_0)]' \hat{I}_{10} \hat{I}_{20}$$

This gives us three terms. But the difference of the two terms is handled in the same way as in lemma 2. In particular, we take γ_n and Γ_n as before and define

$$\begin{aligned} g_N(T_{1i}, T_{0j}, D_j, X_j, Z_j) = N^{-3/2} \gamma_N(T_{0j}) h_{Nq}^{-2} \\ \times \frac{D_j}{P(Z_j)} \left[\frac{T_{1i}-T_{0j}}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-T_{0j})^{Q(s)} K \frac{T_{1i}-T_{0j}}{h_{Nq}} - E\left[\frac{D_j}{P(Z_j)} \left[\frac{T_{1i}-T_{0j}}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-T_{0j})^{Q(s)} K \frac{T_{1i}-T_{0j}}{h_{Nq}} \Big| T_{0j}, X_j, Z_j \right] \Big) \\ \times [m^{(s)}(T_{0j})]' \tilde{I}_{1j} \tilde{I}_{2j} \end{aligned}$$

Let$\hat{g}_N(T_{1i}, T_{0j}, X_j, Z_j)$ and $g_{N0}(T_{1i}, T_{0j}, X_j, Z_j)$ be defined in the same way as before. Moreover, let $\mathcal{G}_{2N} := \{g_n(T_{1i}, T_{0j}, X_j, Z_j) | \gamma_N(T_{0j}) \in \Gamma_N\}$. Then going through the same steps as in lemma 2 we can show that $1/\sqrt{N} \sum_{j=1}^N \hat{R}_{21}(T_{0j}, D_j) = \sum_{i=1}^N \sum_{j=1}^N [\hat{g}_N(T_{1i}, T_{0j}, D_j, X_j, Z_j) - g_{N0}(T_{1i}, T_{0j}, D_j, X_j, Z_j)] = o_p(1)$.
Then we deal with the term

$$e_1[\hat{M}_{pN}(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} \frac{1}{h_{Nq}^2} E\left\{ \frac{D_j}{P(Z_j)} \left[\frac{T_{1i}-T_{0j}}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-T_{0j})^{Q(s)} K \frac{T_{1i}-T_{0j}}{h_{Nq}} |T_{0j}=t_0, X_j=x_0, Z_j=z_0 \right\} [m^{(s)}(t_0)]' \hat{I}_{10} \hat{I}_{20}$$

which in turn equals

$$\begin{aligned} e_1\big([\hat{M}_{pN}(t_0)]^{-1} - [M_p(t_0)]^{-1}\big) \\ \cdot \sum_{s=p+1}^{\overline{p}} \frac{1}{h_{Nq}^2} E\left\{ \frac{D_j}{P(Z_j)} \left[\frac{T_{1i}-T_{0j}}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-T_{0j})^{Q(s)} K \frac{T_{1i}-T_{0j}}{h_{Nq}} |T_{0j}=t_0, X_j=x_0, Z_j=z_0 \right\} [m^{(s)}(t_0)]' \hat{I}_{10} \hat{I}_{20} \\ + e_1[M_p(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} \frac{1}{h_{Nq}^2} E\left\{ \frac{D_j}{P(Z_j)} \left[\frac{T_{1i}-T_{0j}}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-T_{0j})^{Q(s)} K \frac{T_{1i}-T_{0j}}{h_{Nq}} |T_{0j}=t_0, X_j=x_0, Z_j=z_0 \right\} [m^{(s)}(t_0)]' \hat{I}_{10} \hat{I}_{20} \end{aligned}$$

The first expression can be treated in the same way as in lemma 2. If $t_0 = (h_0(x_0, P(z_0)), P(z_0))$, the last expression equals

$$\begin{aligned} & e_1[M_p(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} \frac{1}{h_{Nq}^2 P(z_0)} E\left\{ D_j \left[\frac{T_{1i}-t_0}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-t_0)^{Q(s)} K \frac{T_{1i}-t_0}{h_{Nq}} |T_{0j}=t_0, X_j=x_0, Z_j=z_0 \right\} [m^{(s)}(t_0)]' \hat{I}_{10} \hat{I}_{20} \\ & = e_1[M_p(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} \frac{1}{h_{Nq}^2 P(z_0)} E(D_j | Z_j=z_0) E\left\{ \left[\frac{T_{1i}-t_0}{h_{Nq}} ^{Q_p} \right]' (T_{1i}-t_0)^{Q(s)} K \frac{T_{1i}-t_0}{h_{Nq}} \right\} [m^{(s)}(t_0)]' \\ & = h_{Nq}^{\overline{p}} e_1[M_p(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} \Big[\int u^{Q(0)} \cdot u^{Q(s)} m^{(s)}(t_0)' \cdot u^{Q(\overline{p}-1)} K(u) \mathrm{d}u, \ldots, \int u^{Q(p)} \cdot u^{Q(s)} m^{(s)}(t_0)' \cdot u^{Q(\overline{p}-s)} K(u) \mathrm{d}u \Big] f^{(\overline{p}-s)}(t_0)' \end{aligned}$$

We need

$$\text{plim}_{N\to\infty} 1/\sqrt{N}\sum_{i=1}^{N} h_{Nq}^{\overline{p}} e_1[M_p(T_{0i})]^{-1}\times$$
$$\sum_{s=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(s)} m^{(s)}(T_{0i})'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u,...,\int u^{Q(p)}\cdot u^{Q(s)} m^{(s)}(T_{0i})'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u\right]$$
$$\times f^{(\overline{p}-s)}(T_{0i})' = b_q < \infty$$

All the terms involving $T_{0i} = (h_0(X_i, P(Z_i)), P(Z_i))$ are bounded with probability 1. Thus, if $Nh_{Nq}^{2\overline{p}} \to c < \infty$ then we are OK.

C.3.3 Third Term:

We claim that under our assumptions, for each evaluation point (d_0, x_0, z_0) such that $(x_0, z_0) \in A_1 \cap A_2$,

$$e_1[\hat{M}_{pN}(t_0)]^{-1} N^{-1}\frac{d_0}{P(z_0)} H' T_p'(t_0) W(t_0) r_{\overline{p}+1}(t_0)\hat{I}_{10}\hat{I}_{20} = o_p(h_{Nq}^{\overline{p}+1})$$

But, as in lemma 8 of HIT,

$$\begin{aligned}
& N^{-1}||\frac{d_0}{P(z_0)} H' T_p'(t_0) W(t_0) r_{\overline{p}+1}(t_0)\hat{I}_{10}\hat{I}_{20}|| \leq N^{-1} h_{Nq}^{(\overline{p}+1)} \\
\times \quad & \left|\left|\sum_{i\in I_1}\frac{d_0}{P(z_0)}\left[\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q_p}\right]'\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q(\overline{p}+1)}[m^{(\overline{p}+1)}(t_i) - m^{(\overline{p}+1)}(t_0)]\frac{1}{h_{Nq}^2}K\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)\right|\right| \\
\leq \quad & N^{-1} o(h_{Nq}^{\overline{p}+1})\sum_{i=1}^{N}\left|\left|\left[\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q_p}\right]'\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q(\overline{p}+1)}\frac{1}{h_{Nq}^2}K\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)\right|\right| = o_p(h_{Nq}^{\overline{p}+1})
\end{aligned}$$

By lemma 5 of HIT, for any t_0 such that $f_{h_1(X,P(Z)),P(Z)}(t_0) > 0$, for sufficiently large N, $\hat{M}_{pN}(t_0)$ will be nonsingular. Therefore, every element of the matrix $[\hat{M}_{pN}(t_0)]^{-1}$ has finite norm.

C.3.4 Conclusion:

$$\tfrac{1}{\sqrt{N}}\sum_{j=1}^{N} D_j\big(\hat{q}(h_{0j}, P_j) - q(h_{0j}, P_j)\big)\hat{I}_{1j}\hat{I}_{2j} =^{AE}$$
$$\tfrac{1}{\sqrt{N}N}\sum_{j=1}^{N}\sum_{i=1}^{N}\tfrac{D_j}{P(Z_j)}[M_{pN}(h_{0j}, P_j)]^{-1}\left[\left(\tfrac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}}\right)^{Q_p}\right]'\tfrac{1}{h_{Nq}^2}K\left(\tfrac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}}\right)\varepsilon_i^q I_{1j} I_{2j} + b_q$$

where $\varepsilon_i^q = D_i Y_i - E\big[D_i Y_i | h_1\big(X_i, P(Z_i)\big), P(Z_i)\big]$

C.4 Issues in Trimming

The first part of this appendix shows that the family of functions that the trimming function, which is based on the values the the estimated density $\hat{f}(x,z)$, belongs to satisfies the conditions of the equcicontinuity lemma. The second part of the appendix shows the same thing for the trimming function which is based on, $\hat{f}_{\hat{h}_1,\hat{P}}$, the kernel density estimator of $f_{h_1,P}$. Let S be a random variable, whose values we observe. Also let

$$\begin{aligned}
\mathcal{H} &:= \{f : \sup_{s\in \text{supp}(X)} |f(s) - f_S(s)| \le \epsilon_f,\ f\ \ has\ \ smoothness\ \ q > d, \inf_{s\in\overline{A}} ||Df(s)|| \ge \underline{\theta}\} \\
\mathcal{I} &:= \{I((s)\in \tilde{A}) : \tilde{A} = \{s : f(s) \ge q_0\}\ for\ some\ f \in \mathcal{H}\} \\
\overline{A} &:= \{s : f_S(s) \ge q_0 - \epsilon_f\} \\
A &:= \{s : f_S(s) \ge q_0\}
\end{aligned}$$

where $\underline{\theta} > 0$. First we observe that under the assumptions of Silverman's Theorem A on f_S, the kernel function and the bandwidth sequence used to estimate this density function

$$\begin{aligned}
\sup_{s\in\text{supp}(S)} \left|\hat{f}(s) - f(s)\right| &\to 0 \quad a.s. \\
\sup_{s\in\text{supp}(S)} \left|\frac{\partial \hat{f}}{\partial s_j}(s) - \frac{\partial f}{\partial s_j}(s)\right| &\to 0 \quad a.s. \quad for\ j \in \{1,...,d\}
\end{aligned}$$

Using this result, we can claim that $1_{\overline{A}}(s)$ is an envelope for $\mathcal{I}$. On the other hand, we can also prove that if the third condition of the equicontinuity lemma holds for $\mathcal{H}$, it also holds for $\mathcal{I}$. For this purpose, observe that for any function in $\mathcal{H}$, and for any probability measure that is absolutely continuous with respect to Lebesgue measure,

$$P(\{x : f(x) = q_0\}) = 0$$

Then by the dominated convergence theorem,

$$\lim_{\delta\downarrow 0} P(\{x : f(x) = q_0\} \oplus B_\delta(0)) = 0$$

where $A \oplus B := \{a+b : a \in A,\ b \in B\}$, and $B_\delta(0)$ denotes the ball around 0 with radius δ. Next, I claim that for any $f, g \in \mathcal{H}$, such that $\sup_{x\in\overline{A}} |f(x) - g(x)| < \eta$, and for $\delta = \eta/\underline{\theta}$,

$$P(\{f \ge q_0 > g\}) \le P(\{x : f(x) = q_0\} \oplus B_\delta(0))$$

To see this, consider any $s \in \overline{A} \setminus (\{x : f(x) = q_0\} \oplus B_\delta(0))$. Then for each $u \in \{x : f(x) = q_0\}$, $d(u,s) \ge \delta$. If $f(s) \le q_0$, there is nothing to prove. Otherwise, pick some $u \in \{x : f(x) = q_0\}$. Using the mean value theorem, we know that

$$|f(s) - f(u)| = |f(s) - q_0| = ||Df(\tilde{u})|| \cdot ||s-u|| \ge \underline{\theta}||s-u|| > \underline{\theta}\delta = \eta$$

Since $|f(s) - g(s)| < \eta$, this implies that $g(s) \geq q_0$, i.e. $s \in \overline{A} \setminus \{x : f(x) \geq q_0 > g\}$, equivalently that $\{x : f(x) \geq q_0 > g(x)\} \subseteq (\{x : f(x) = q_0\} \oplus B_\delta(0))$. The last step is to note that by choosing η appropriately we could make sure that $P(\{x : f(x) \geq q_0 > g(x)\}) + P(\{x : g(x) \geq q_0 > f(x)\})$ arbitrarily small.

Next we turn to our trimming problem. We have to employ two trimming functions. The first function is needed to guarantee that the estimator $\hat{P(z)}$ is uniformly consistent for $E(D|Z)$. The second trimming function is needed because we need to have a uniformly consistent estimate for $E(DY|h_1(X, P(Z)), P(Z))$ evaluated at the value $(h_0(X < P(Z)), P(Z))$ takes. For this purpose we define $d = \dim(\mathrm{supp(X, Z)})$. We will use the arguments given above with $S = (X, Z)$ to argue that our first trimming function satisfies the conditions of the equicontinuity lemma. We define

$$\begin{aligned} \overline{A}_1 &= \{(x, z) : f_{X,Z}(x, z) \geq q_{01} - \epsilon_{f1}\} \\ \overline{A}_q &= \{(x, z) : q_{01} + \epsilon_{f1} \geq f_{X,Z}(x, z) \geq q_{01} - \epsilon_{f1}\} \\ \overline{B}_z &= \{z \in \mathrm{supp}(Z) : (x, z) \in \overline{A}_1, \text{for some } x \in \mathrm{supp}(X)\} \end{aligned}$$

and[26]

$$\begin{aligned} \mathcal{H}_1 &= \{f : \sup_{(x,z)\in \mathrm{supp}(X,Z)} |f(x, z) - f_{X,Z}(x, z)| \leq \epsilon_{f1},\ f\ has\ smoothness\ q > d,\ \inf_{(x,z)\in\overline{A}_q} ||Df(x,z)|| \geq \underline{\theta}_1\} \\ \mathcal{I}_1 &= \{I((x, z) \in \tilde{A}_1) : \tilde{A}_1 = \{(x, z) : f(x, z) \geq q_{01}\} \text{ for some } f \in \mathcal{H}_1\} \\ \Psi_P &= \{g : \sup_{z\in\overline{B}_z} |g(z) - P(z)| \leq \epsilon_P,\ g\ has\ smoothness\ q > d,\ \inf_{z\in\overline{B}_{zq}} ||Dg(z)|| \geq \underline{\theta}_P\} \\ &\cap \{g : \sup_{z\in\overline{B}_z} |g(z) - P(z)| = o_P(\tilde{h}^3_{N2})\} \\ \Psi_h &= \{\varphi : \sup_{\tilde{P}\in\Psi_P} \sup_{(x,z)\in\overline{A}_1} |\varphi(x, \tilde{P}(z)) - h_0(x, P(z))| \leq \epsilon_h,\ \varphi\ has\ smoothness\ q > d\} \\ &\cap \{\varphi : \sup_{\tilde{P}\in\Psi_P} \sup_{(x,z)\in\overline{A}_1} |\varphi(x, \tilde{P}(z)) - h_0(x, P(z))| = o_P(\tilde{h}^3_{N2})\} \\ &\cap \{\varphi : \inf\{||D_x\varphi(x, \tilde{P}(z))|| : (x, z) \in \overline{A}_q, \tilde{P} \in \Psi_P\} \geq \underline{\theta}_{hx}\} \\ &\cap \{\varphi : \inf\{||D_P\varphi(x, \tilde{P}(z))|| : (x, z) \in \overline{A}_q, \tilde{P} \in \Psi_P\} \geq \underline{\theta}_{hP}\} \end{aligned}$$

$$\begin{aligned} \mathcal{H}_2 &= \{f : \sup_{(\varphi,\tilde{P})\in\Psi_h\times\Psi_P} \sup_{(x,z)\in\overline{A}_1} |f(\varphi(x, \tilde{P}(z)), \tilde{P}(z)) - f_{h_1(X,P(Z)),P(Z)}(h_0(x, P(z)), P(z))| \leq \epsilon_{f2}\} \\ &\cap \{f : f\ \ has\ \ smoothness\ \ q > d,\ \inf_{(x,z,\tilde{P},\varphi)\in\overline{A}_q\times\Psi_P\times\Psi_h} ||Df(\varphi(x, \tilde{P}(z)), \tilde{P}(z))|| \geq \underline{\theta}_2\} \\ \mathcal{I}_2 &= \{I((x, z) \in \tilde{A}_2) : \tilde{A}_2 = \{(x, z) \in \overline{A}_1 : f(\varphi(x, \tilde{P}(z)), \tilde{P}(z)) \geq q_{02}\} \text{ for some } {\scriptstyle f\in\mathcal{H}_2,\varphi\in\Psi_h,\tilde{P}\in\Psi_P}\} \end{aligned}$$

We would like to show that $\hat{I}_1 = 1\{\hat{f}_{X,Z}(x, z) \geq q_{01}\} \in \mathcal{I}_1$ and $\hat{I}_2 = 1\{\hat{f}_{\hat{h}_1,\hat{P}}(\hat{h}_0(x, \hat{P}(z)), \hat{P}(z)) \geq q_{02}\} \in \mathcal{I}_2$ for sufficiently large N with probability approaching to 1. The first of these follows

[26]In these definitions all the θ's are strictly greater than 0, and $\tilde{h}_{N2}$ denotes the smoothing parameter that is used in the trimmed kernel density estimation of $f_{h_1,P}$.

from the analysis above. To investigate the second one, consider $1\{\hat{f}_{\hat{h}_1,\hat{P}}(\hat{h}_0(x,\hat{P}(z)),\hat{P}(z)) \geq q_{02}\}1(\overline{A}_1)$ [27]. For $(x,z) \in \overline{A}_1$, $\hat{P}$ and $h_0(x,\hat{P}(z))$ are uniformly consistent for $P(z)$ and $h_0(x,P(z))$, respectively. So we only need to show that $\hat{f}_{\hat{h}_1,\hat{P}}$ is uniformly consistent for $f_{h_1,P}$. To be able to guarantee this, we need $\hat{h}_1(X_i,\hat{P}(Z_i))$ and $\hat{P}(Z_i)$ to be uniformly consistent for $h_1(X_i,P(Z_i))$ and $P(Z_i)$. However, this occurs only when the density of (X,Z) is bounded away from 0. Therefore, we have to trim out those observations at which $f_{X,Z}$ is very small. Let $\tilde{K}_2$ be a Lipschitz function[28] and define

$$\hat{f}_{\hat{h}_1,\hat{P}}\big(\varphi(x,\tilde{P}(z)),\tilde{P}(z)\big) := \frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^{N}\tilde{K}_2\left(\frac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i))-(\varphi(x,\tilde{P}(z)),\tilde{P}(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i,Z_i)$$

$$\hat{f}_{h_1,P}\big(\varphi(x,\tilde{P}(z)),\tilde{P}(z)\big) := \frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^{N}\tilde{K}_2\left(\frac{(h_1(X_i,P(Z_i)),P(Z_i))-(\varphi(x,\tilde{P}(z)),\tilde{P}(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i,Z_i)$$

Consider

$$\begin{aligned}|\hat{f}_{\hat{h}_1,\hat{P}}(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) - f_{h_1,P}(h_0(x,P(z)),P(z))| \leq \\ \left|\hat{f}_{\hat{h}_1,\hat{P}}(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) - \hat{f}_{\hat{h}_1,\hat{P}}(h_0(x,P(z)),P(z))\right| \\ +\left|\hat{f}_{\hat{h}_1,\hat{P}}(h_0(x,P(z)),P(z)) - \hat{f}_{h_1,P}(h_0(x,P(z)),P(z))\right| \\ +\left|\hat{f}_{h_1,P}(h_0(x,P(z)),P(z)) - f_{h_1,P}(h_0(x,P(z)),P(z))\right|\end{aligned} \tag{13}$$

We will first deal with the first term.

$$\begin{aligned}|\hat{f}_{\hat{h}_1,\hat{P}}(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) - \hat{f}_{\hat{h}_1,\hat{P}}(h_0(x,P(z)),P(z))| \leq \\ \tfrac{1}{N\tilde{h}_{N2}^2}\textstyle\sum_{i=1}^N\left|\tilde{K}_2\left(\tfrac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i))-(\varphi(x,\tilde{P}(z)),\tilde{P}(z))}{\tilde{h}_{N2}}\right) - \tilde{K}_2\left(\tfrac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i))-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)\right|\hat{I}_1(X_i,Z_i) \\ \leq \tfrac{M}{\tilde{h}_{N2}^3}\left[\left|\varphi(x,\tilde{P}(z)) - h_0(x,P(z))\right| + \left|\tilde{P}(z)-P(z)\right|\right]\end{aligned}$$

We know that on $\overline{A}_1$, both $\varphi(x,\tilde{P}(z))$, and $\tilde{P}(z)$ are uniformly consistent. Moreover, $|\varphi(x,\tilde{P}(z)) - h_0(x,P(z))|$ and $|\tilde{P}(z)-P(z)|$ are both $o_p(\tilde{h}_{N2}^3)$ on $\overline{A}_1$. On the other hand,

$$\begin{aligned}\left|\hat{f}_{\hat{h}_1,\hat{P}}(h_0(x,P(z)),P(z)) - \hat{f}_{h_1,P}(h_0(x,P(z)),P(z))\right| = \\ \tfrac{1}{N\tilde{h}_{N2}^2}\textstyle\sum_{i=1}^N \ \tilde{K}_2 \ \tfrac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i))-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}} \ -\tilde{K}_2 \ \tfrac{(h_1(X_i,P(Z_i)),P(Z_i))-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}} \ |\hat{I}_1(X_i,Z_i) \\ \leq \left|\tfrac{M_1}{N\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N\left[\hat{h}_1(X_i,\hat{P}(Z_i)) - h_1(X_i,P(Z_i))\right]\hat{I}_1(X_i,Z_i)\right| \\ +\left|\tfrac{M_1}{N\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N\left[\hat{P}(Z_i)-P(Z_i)\right]\hat{I}_1(X_i,Z_i)\right|\end{aligned}$$

[27] Note that the first trimming function $\hat{I}_1$ would eventually eliminate observations which lie outside of $\overline{A}_1$ with probability approaching to 1. So in terms of the second trimming function, we only need to worry about (x,z) values in $\overline{A}_1$.

[28] Later, we may impose other conditions on this kernel function.

Using the results of Appendix C.1:

$$\begin{aligned}&\left|\tfrac{M_1}{N\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N\left[\hat{P}(Z_i)-P(Z_i)\right]\hat{I}_1(X_i,Z_i)\right|\\ \leq&\left|\tfrac{M_1}{N^2\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N\sum_{j=1}^N\psi_{NP}(D_j,X_j,Z_j;X_i,Z_i)\right|+\left|\tfrac{M_1}{N\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N\hat{b}_P(X_i,Z_i)\right|\\ &+\left|\tfrac{M_1}{N\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N\hat{R}_P(X_i,Z_i)\right|\end{aligned}$$

Let's deal with the first part first. Split the first term into two sums: one containing the terms where i and j are the same, and the other, where they are different:

$$\left|\tfrac{M_1}{N^2\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N\psi_{NP}(X_i,Z_i,D_i;X_i,Z_i)\right|=$$
$$\left|\tfrac{M_1}{N^2\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^N e_1[M_{pN}^P(Z_i)]^{-1}e_1'h_{NP}^{-d_z}K^P(0)\varepsilon_i^P\right|=\left|\tfrac{1}{N}\textstyle\sum_{i=1}^N\tfrac{M_1}{N\tilde{h}_{N2}^3h_{NP}^{d_z}}e_1[M_{pN}^P(Z_i)]^{-1}e_1'K^P(0)\varepsilon_i^P\right|$$

We will apply a strong law of large numbers:

Theorem C.1 ***(Chebyshev)*** *Let $S_1,S_2,...$ be uncorrelated with means $\mu_1,\mu_2,...$ and variances $\sigma_1^2,\sigma_2^2,....$ If $\sum_{i=1}^N\sigma_i^2=o(N^2)$ as $N\rightarrow\infty$ then*

$$\frac{1}{N}\sum_{i=1}^N S_i-\frac{1}{N}\sum_{i=1}^N\mu_i\rightarrow^P 0$$

To apply this theorem we need to check the expectation and the variance of the i^{th} term:

$$\begin{aligned}&E\left[e_1[M_{pN}^P(Z_i)]^{-1}e_1'K^P(0)I_1(X_i,Z_i)\varepsilon_i^P\right]=\\ &E\left[e_1[M_{pN}^P(Z_i)]^{-1}e_1'K^P(0)I_1(X_i,Z_i)E(\varepsilon_i^P|X_i,Z_i)\right]=0\end{aligned}$$

We also need to verify that the required variance condition holds.

$$\begin{aligned}&\lim_{N\rightarrow\infty}\frac{1}{N^2}\sum_{i=1}^N E\left[\frac{M_1^2}{N^2\tilde{h}_{N2}^6h_{NP}^{2d_z}}\left(e_1[M_{pN}^P(Z_i)]^{-1}e_1'\right)^2\left(K^P(0)\right)^2I_1(X_i,Z_i)\left(\varepsilon_i^P\right)^2\right]\\ &=\lim_{N\rightarrow\infty}\frac{M_1}{N^3\tilde{h}_{N2}^6h_{NP}^{2d_z}}E\left[\left(e_1[M_{pN}^P(Z_i)]^{-1}e_1'\right)^2\left(K^P(0)\right)^2I_1(X_i,Z_i)\left(\varepsilon_i^P\right)^2\right]\end{aligned}$$

$Nh_N^{2d_z}\rightarrow\infty$. As long as $N\tilde{h}_{N2}^3$ does not converge to 0, or does not converge to 0 too fast, the variance condition needed to apply the theorem holds and we have

$$plim_{N\rightarrow\infty}\frac{1}{N}\sum_{i=1}^N\frac{M_1}{N\tilde{h}_{N2}^3h_{NP}^{d_z}}e_1[M_{pN}^P(Z_i)]^{-1}e_1'K^P(0)I_1(X_i,Z_i)\varepsilon_i^P=0$$

Then by the continuous mapping theorem

$$plim_{N\to\infty}\left|\frac{M_1}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N \psi_{NP}(X_i,Z_i,D_i;X_i,Z_i)\right| =$$
$$\left|plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \frac{M_1}{N\tilde{h}_{N2}^3 h_{NP}^{d_z}} e_1[M_{pN}^P(Z_i)]^{-1}e_1'K^P(0)I_1(X_i,Z_i)\varepsilon_i^P\right| = 0$$

Next, we deal with

$$\left|\frac{M_1}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N\sum_{j\neq i}^N \psi_{NP}(X_j,Z_j,D_j;X_i,Z_i)\right|$$

For this term, we will appeal to the Hoeffding, Powell, Stock and Stoker lemma. Defining

$$\zeta(D_i,Y_i,X_i,Z_i,D_j,Y_j,X_j,Z_j) = \frac{1}{2h_{NP}^{d_z}}e_1[M_{pN}^P(Z_i)]^{-1}\left[\left(\frac{Z_j-Z_i}{h_{NP}}\right)\right]' K^P\left(\frac{Z_j-Z_i}{h_{NP}}\right)I_1(X_i,Z_i)\varepsilon_j^P$$
$$+\frac{1}{2h_{NP}^{d_z}}e_1[M_{pN}^P(Z_j)]^{-1}\left[\left(\frac{Z_i-Z_j}{h_{NP}}\right)\right]' K^P\left(\frac{Z_i-Z_j}{h_{NP}}\right)I_1(X_j,Z_j)\varepsilon_i^P$$

The arguments in Appendix C.1 show that

$$E\big[\zeta(D_i,Y_i,X_i,Z_i,D_j,Y_j,X_j,Z_j)\big] = 0$$
$$E\big[(\zeta(D_i,Y_i,X_i,Z_i,D_j,Y_j,X_j,Z_j))^2\big] = o(N)$$

Therefore, using the Hoeffding, Powell, Stock and Stoker lemma,

$$plim_{N\to\infty}\frac{M_1}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N\sum_{j\neq i}^N \psi_{NP}(X_j,Z_j,D_j;X_i,Z_i) =$$
$$plim_{N\to\infty}\frac{M_1}{\sqrt{N\tilde{h}_{N2}^6}}plim_{N\to\infty}N^{-1/2}\sum_{i=1}^N E\ \ e_1[M_{pN}^P(Z_j)]^{-1}I_1(X_j,Z_j)\Big[\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ \Big]' K^P\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ h_{NP}^{-d_z}\varepsilon_i^P|D_i,X_i,Z_i$$
$$=plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\frac{M_1}{h_{NP}^{d_z}\tilde{h}_{N2}^3}E\ \ e_1[M_{pN}^P(Z_j)]^{-1}I_1(X_j,Z_j)\Big[\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ \Big]' K^P\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ \varepsilon_i^P|D_i,X_i,Z_i$$

We can now apply the same law of large numbers. Each $\mu_i = 0$. Therefore $1/N\sum_{i=1}^N \mu_i = 0$. We still have to verify that $\sum_{i=1}^N \sigma_i^2 = o(N^2)$.

$$\frac{1}{N^2}\sum_{i=1}^N\frac{M_1^2}{\tilde{h}_{N2}^6}E\left\{\ \ E\left[e_1[M_{pN}^P(Z_j)]^{-1}I_1(X_j,Z_j)\Big[\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ \Big]'\frac{1}{h_{NP}^{d_z}}K^P\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ \varepsilon_i^P|D_i,X_i,Z_i\right]\Big)^2\right\}$$

By Jensen's inequality

$$\left(E\left[e_1[M_{pN}^P(Z_j)]^{-1}I_1(X_j,Z_j)\left[\left(\frac{Z_i-Z_j}{h_{NP}}\right)\right]'\frac{1}{h_{NP}^{d_z}}K^P\left(\frac{Z_i-Z_j}{h_{NP}}\right)\varepsilon_i^P|D_i,X_i,Z_i\right]\right)^2$$
$$\leq E\left\{I_1(X_j,Z_j)\ \ e_1[M_{pN}^P(Z_j)]^{-1}\Big[\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ \Big]'\ \ ^2\frac{1}{h_{NP}^{2d_z}}\ \ K^P\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ ^2\left(\varepsilon_i^P\right)^2|D_i,X_i,Z_i\right\}$$
$$=\frac{\left(\varepsilon_i^P\right)^2}{h_{NP}^{d_z}}E\left\{I_1(X_j,Z_j)\ \ e_1[M_{pN}^P(Z_j)]^{-1}\Big[\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ \Big]'\ \ ^2\frac{1}{h_{NP}^{d_z}}\ \ K^P\ \ \frac{Z_i-Z_j}{h_{NP}}\ \ ^2|D_i,X_i,Z_i\right\}$$

On A_1, $e_1[M_{pN}^P(Z_j)]^{-1}$ and the density $f_{X,Z}$ are bounded. Moreover, the kernel function K^P is assumed to have compact support, and hence for some positive $\tilde{C}$

$$E\left\{I_1(X_j,Z_j)\ \ e_1[M_{pN}^P(Z_j)]^{-1}\Big[\ \frac{Z_i-Z_j}{h_{NP}}\ \Big]'\ {}^2\ \frac{1}{h_{NP}^{d_z}}\ K^P\ \frac{Z_i-Z_j}{h_{NP}}\ \ {}^2|D_i,X_i,Z_i\right\}\leq\tilde{C}$$

Since, we also have $\sigma_P^2 := E(\varepsilon_i^P)^2 < \infty$,

$$\frac{1}{N^2}\sum_{i=1}^N \frac{M_1^2}{\tilde{h}_{N2}^6}E\left\{\ \ E\left[e_1[M_{pN}^P(Z_j)]^{-1}I_1(X_j,Z_j)\Big[\ \frac{Z_i-Z_j}{h_{NP}}\ \Big]'\frac{1}{h_{NP}^{d_z}}K^P\ \frac{Z_i-Z_j}{h_{NP}}\ \ \varepsilon_i^P|D_i,Y_i,X_i,Z_i\right]\Big)^2\right\}$$
$$\leq \frac{1}{N^2}\sum_{i=1}^N \frac{M_1^2}{h_{NP}^{d_z}\tilde{h}_{N2}^6}\sigma_P^2\tilde{C} = \frac{M_2}{Nh_{NP}^{d_z}\tilde{h}_{N2}^6}$$

We assumed that $Nh_{NP}^{2d_z} \to \infty$. Then if $\sqrt{N}\ \tilde{h}_{N2}^6$ does not go to 0, or if it does not go to 0 too fast, then the product of $\sqrt{N}h_{NP}^{d_z}$ and $\sqrt{N}\ \tilde{h}_{N2}^6$ will still go to ∞[29]. Next, we deal with

$$\left|\frac{M_1}{N\tilde{h}_{N2}^3}\sum_{i=1}^N \hat{b}_P(X_i,Z_i)\right|$$

From Appendix C.1, we know that $plim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{b}_P(X_i,Z_i) = b_P < \infty$. Then if $\lim_{N\to\infty}\sqrt{N}\tilde{h}_{N2}^3$ $= \infty$, this term too will be converging to 0 uniformly in probability by the continuous mapping theorem. Finally, let us look at

$$\left|\frac{M_1}{N\tilde{h}_{N2}^3}\sum_{i=1}^N \hat{R}_P(X_i,Z_i)\right| = \left|\frac{M_1}{\sqrt{N}\tilde{h}_{N2}^3}\frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{R}_P(X_i,Z_i)\right|$$

We know that $\frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{R}_P(X_i,Z_i) = o_p(1)$. This, continuous mapping theorem, and our previous assumption that $\lim_{N\to\infty}\sqrt{N}\tilde{h}_{N2}^3 = \infty$ jointly imply that this last term also goes to 0 uniformly in probability.

Using Appendix C.2, we can write

$$\left|\frac{M_1}{N\tilde{h}_{N2}^3}\sum_{i=1}^N \left[\hat{h}_1(X_i,\hat{P}(Z_i)) - h_1(X_i,P(Z_i))\right]\hat{I}_1(X_i,Z_i)\right| \leq$$
$$\left|\frac{M_1}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N\sum_{j=1}^N \psi_{N\hat{h}_1}(D_j,Y_j,X_j,Z_j;X_i,Z_i)\right| + \left|\frac{M_1}{N\tilde{h}_{N2}^3}\sum_{i=1}^N \hat{b}_{\hat{h}_1}(X_i,Z_i)\right|$$
$$+\left|\frac{M_1}{N\tilde{h}_{N2}^3}\sum_{i=1}^N \hat{R}_{\hat{h}_1}(X_i,Z_i)\right|$$

Again, we know that $plim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{b}_{\hat{h}_1}(X_i,Z_i) = b_{h_1} + b_{h_1P} < \infty$. So again, if $\sqrt{N}\tilde{h}_{N2}^3 \to \infty$, by continuous mapping theorem, the middle term goes to 0 in probability. Similarly, we know $plim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N \hat{R}_{\hat{h}_1}(X_i,Z_i) = 0$. Thus, the same condition guarantees that the last sum converges to 0 in probability. As for the first sum, again we can split it into two pieces. One

[29]We could for example, choose $\tilde{h}_{N2} = h_{NP}^{d_z/6}$.

piece contains the terms where the two indices equal, the other piece contains the terms where the indices are different:

$$\left|\frac{M_1}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N \psi_{N\hat{h}_1}(D_i,Y_i,X_i,Z_i;X_i,Z_i)\right| =$$
$$\left|\frac{M_1}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N \left(\psi_{Nh_1}(D_i,Y_i,X_i,Z_i;X_i,Z_i)+\frac{\partial h_1}{\partial P}(X_i,P(Z_i))\psi_{NP}(D_i,X_i,Z_i;X_i,Z_i)\right)\right|$$

Each term has 0 expectation. To verify that the sum of the variances is $o(N^2)$, by Cauchy-Schwarz inequality it suffices to verify that

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N E\big[\psi_{Nh_1}(D_i,Y_i,X_i,Z_i;X_i,Z_i)^2\big] =$$
$$\lim_{N\to\infty}\frac{M_1}{N^3\tilde{h}_{N2}^6 h_{Nh_1}^{2(d_x+2)}}E\left[\left(e_1[M_{pN}^{h_1}(X_i,P(Z_i))]^{-1}e_1'\right)^2\left(K^{h_1}(0)\right)^2 I_1(X_i,Z_i)\left(\varepsilon_i^{h_1}\right)^2\right]=0$$

and

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N E\big[\psi_{NP}(D_i,X_i,Z_i;X_i,Z_i)^2\big] =$$
$$\lim_{N\to\infty}\frac{M_1}{N^3\tilde{h}_{N2}^6 h_{NP}^{2d_z}}E\left[\left(e_1[M_{pN}^{P}(Z_i)]^{-1}e_1'\right)^2\left(K^{P}(0)\right)^2 I_1(X_i,Z_i)\left(\varepsilon_i^{P}\right)^2\right]=0$$

The first one is true because the term inside the parentheses is bounded, $N\tilde{h}_{N2}^6\to\infty$ and $Nh_{Nh_1}^{2(d_x+2)}\to\infty$. The second one is true because $Nh_{NP}^{2d_z}\to\infty$, and again because the term inside the parentheses is bounded, and $N\tilde{h}_{N2}^6\to\infty$. So the sum of terms with $i=j$ converges to 0 in probability. For the other sum, we again use Hoeffding, Powell, Stock and Stoker lemma. By arguments in Appendix B, we know that

$$plim_{N\to\infty}\frac{M_1}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N\sum_{j\neq i}^N \psi_{N\hat{h}_1}(D_j,Y_j,X_j,Z_j;X_i,Z_i) =$$
$$= plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\frac{M_1}{\tilde{h}_{N2}^3}E\left[\psi_{N\hat{h}_1}(D_i,Y_i,X_i,Z_i;D_j,Y_j,X_j,Z_j)|D_i,Y_i,X_i,Z_i\right]$$

Then we apply the Chebyshev's theorem one last time. Again, the expectation of i^{th} term is 0. And given that we have already assumed $Nh_{NP}^{2d_z}\to\infty$, $Nh_{Nh_1}^{2(d_x+1)}\to\infty$ and $N\tilde{h}_{N2}^{12}$ does not go to 0, the variance condition is satisfied. Therefore, this sum converges to 0 in probability as well.

Our last step is to show that

$$\left|\hat{f}_{h_1,P}(h_0(x,P(z)),P(z)) - f_{h_1,P}(h_0(x,P(z)),P(z))\right|$$

converges to 0 uniformly in probability.

$$\left|\hat{f}_{h_1,P}(h_0(x,P(z)),P(z)) - f_{h_1,P}(h_0(x,P(z)),P(z))\right| =$$
$$\left|\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^N \tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i,Z_i) - f_{h_1,P}(h_0(x,P(z)),P(z))\right|$$

Since $E\left[\frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)\right]$ exists[30], for each $\eta > 0$, we can pick a large number, T, so that

$$P\left(\frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right) > T\right) < \frac{\eta}{2}$$

Let G denote the complement of the set in the above expression. Then by Markov's inequality, for each $\alpha > 0$,

$$P\left(\frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)|\hat{I}_{1i}-I_{1i}| > \alpha\right) \leq$$
$$\frac{\eta}{2} + P\left(\left\{\frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)|\hat{I}_{1i}-I_{1i}| > \alpha\right\}\cap G\right) \leq$$
$$\frac{\eta}{2} + \frac{T}{\alpha}E(|\hat{I}_{1i}-I_{1i}|)$$

By our previous assumptions, $E(|\hat{I}_{1i}-I_{1i}|)$ approaches 0. Therefore, by choosing N sufficiently large, that expectation can be made arbitrarily small, in particular, smaller than $\frac{\alpha\eta}{2T}$. But all these arguments only show the convergence in probability for each point $(x,z) \in A_1$. I could try using the equicontinuity lemma (and that's what I thought I was doing before). But the U-process is not degenerate. I will try another trick:

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i,Z_i)=$$
$$\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^N \tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)I_1(X_i,Z_i)+\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^N \tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)$$
$$\times[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^- \quad \frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)} \quad \left[\hat{f}(X_i,Z_i)-f_{X,Z}(X_i,Z_i)\right]1\{\hat{f}(X_i,Z_i)>f(X_i,Z_i)\}$$
$$+\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^N \tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^+ \quad \frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)} \quad \left[\hat{f}(X_i,Z_i)-f_{X,Z}(X_i,Z_i)\right]$$
$$\times 1\{\hat{f}(X_i,Z_i)\leq f(X_i,Z_i)\}$$

My goal is to show that each of the last two terms is uniformly $o_p(1)$. Let's focus on the first of those two. That term equals

$$\sum_{i=1}^N\sum_{j=1}^N \frac{1}{N^2\tilde{h}_{N2}^2\tilde{h}_{N1}^d}\tilde{K}_2\left(\frac{\left(h_1(X_i,P(Z_i)),P(Z_i)\right)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^- \quad \frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)} \quad 1\{\hat{f}(X_i,Z_i)>f(X_i,Z_i)\}$$
$$\times \quad \tilde{K}_1 \quad \frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}} \quad -E \quad \tilde{K}_1 \quad \frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}} \quad |X_i,Z_i \tag{14}$$

[30]We assume $\tilde{K}_2 \geq 0$.

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{N^2\tilde{h}_{N2}^2\tilde{h}_{N1}^d}\tilde{K}_2\left(\frac{\big(h_1(X_i,P(Z_i)),P(Z_i)\big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^- \quad \frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)} \quad 1\{\hat{f}(X_i,Z_i)>f(X_i,Z_i)\}$$

$$\times \; E \quad \tilde{K}_1 \quad \frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}} \quad |X_i,Z_i \quad -f_{X,Z}(X_i,Z_i) \tag{15}$$

Using the equicontinuity lemma we will show that (14) is $o_p(1)$. For this purpose, for $g \in \mathcal{H}_1$, define $\tilde{\sigma}(X_i, Z_i) = |g(X_i, Z_i) - f_{X,Z}(X_i, Z_i)|$, $\tilde{L}_i = 1\{g(X_i, Z_i) > f_{X,Z}(X_i, Z_i)\}$. Then

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{N^2\tilde{h}_{N2}^2\tilde{h}_{N1}^d}\tilde{K}_2\left(\frac{\big(h_1(X_i,P(Z_i)),P(Z_i)\big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\tilde{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^- \quad \frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\tilde{\sigma}(X_i,Z_i)} \quad \tilde{L}_i$$

$$\times \; \tilde{K}_1 \quad \frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}} \quad -E \quad \tilde{K}_1 \quad \frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}} \quad |X_i,Z_i \quad -f_{X,Z}(X_i,Z_i)$$

is a degenerate U-process of order one which satisfies the conditions of the equicontinuity lemma. Finally, (15) is $o_p(\tilde{h}_{N1})$ by the smoothness of $f_{X,Z}$. On the other hand, by using the same tricks, we can also show that the symmetric term (i.e. the term involving $\tilde{J}_+$) is also uniformly $o_p(1)$. As a result, $\hat{f}_{h_1,P}(h_0(x, P(z)), P(z))$ converges in probability uniformly to

$$\tilde{f}_{h_1,P}(h_0(x,P(z)),P(z)) = \frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^{N}\tilde{K}_2\left(\frac{\big(h_1(X_i,P(Z_i)),P(Z_i)\big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)I_1(X_i,Z_i)$$

Given our assumptions on $\tilde{K}_2$ we can use a strong law of large numbers to show that this converges to

$$E\left[\frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\big(h_1(X_i,P(Z_i)),P(Z_i)\big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)1(A_1)\right]$$

Now the set A_1 is closed, but we can find a sequence of open sets that are all contained in A_1. Moreover the limit of this sequence of open sets will be A_1. Using change of variables theorem by breaking the set A_1 into disjoint regions where P and h_1 have non-zero derivatives, if necessary, and then using Silverman's theorem we have the desired result.