Distribution and Quantile Structural Functions in Treatment Effect Models: Application to Smoking Effects on Wages

Yu-Chin Hsu*

Institute of Economics

Academia Sinica

Kamhon Kan†

Institute of Economics

Academia Sinica

Tsung-Chih Lai‡

Department of Economics

National Taiwan University

First version: March 19, 2015

This version: April 19, 2016

* Corresponding author. E-mail: ychsu@econ.sinica.edu.tw. 128 Academia Road, Section 2, Nankang, Taipei, 115 Taiwan. Tel: 886-2-27822791 ext. 322. Fax: 886-2-2785-3946.

† E-mail: kk@econ.sinica.edu.tw. 128 Academia Road, Section 2, Nankang, Taipei, 115 Taiwan.

‡ E-mail: d00323011@ntu.edu.tw. 1 Roosevelt Road, Section 4, Daan, Taipei, 106 Taiwan.

Acknowledgement: We thank Jen-Che Liao, Shakeeb Khan, Chung-Ming Kuan, Tatsushi Oka, and the participants at the 10th International Symposium on Econometric Theory and Applications for helpful suggestions and comments. All errors are our responsibility. Yu-Chin Hsu gratefully acknowledges the research support from Ministry of Science and Technology of Taiwan (MOST103-2628-H-001-001-MY4) and Career Development Award of Academia Sinica, Taiwan.

Abstract

We consider the distribution structural functions (DSFs) and quantile structural functions (QSFs) of the potential outcomes in a separable semiparametric treatment effect model. We show that DSFs and QSFs are identified under the unconfoundedness assumption. We propose estimators for DSFs and QSFs that weakly converge to mean zero Gaussian processes at the parametric rate and a simulation method to approximate these limiting processes. We apply these results to construct uniform confidence bands for the structural quantile treatment effect of smoking on wages and find that smoking does not impose any wage penalty on male workers with low unobserved heterogeneity.

JEL Classification: C14, C31, I19

Keywords: distribution structural function, semiparametric models, smoking, treatment effects, quantile structural function, wages

1 Introduction

In the program evaluation literature, various parameters are proposed to evaluate the impact of a treatment. For example, the average treatment effect (ATE) captures the mean impact of a treatment on the potential outcomes, see e.g., Rosenbaum and Rubin (1983), Heckman, Ichimura, and Todd (1997, 1998), Heckman, Ichimura, Smith, and Todd (1998), Hahn (1998), and Hirano, Imbens, and Ridder (2003). Firpo (2007) focuses on the pointwise quantile treatment effect (QTE) which measures the difference between certain quantiles of the potential outcomes. Donald and Hsu (2014) estimate and draw inference on the entire distributions and quantile functions of the potential outcomes. For a recent review, please see Imbens and Wooldridge (2009) among others.

In this paper, we consider the distribution structural functions (DSFs) and quantile structural functions (QSFs) of the potential outcomes in a separable semiparametric treatment effect model. Analogous to the average structural function (ASF) introduced by Blundell and Powell (2003), the DSF is defined as the distribution function of the *counterfactual* potential outcome when the covariate values of the whole population are exogenously switched to a fixed value, while the unobserved heterogeneity of the whole population remains unchanged. The QSF is the corresponding quantile function of the DSF. The treatment effect model we consider is separable because each of the potential outcomes is defined as a sum of a conditional mean function and an error term. The model is semiparametric because the conditional mean functions are functions of covariates known up to finite-dimensional parameter vectors, whereas the distributions of the error terms are left completely unspecified except that the conditional means of the errors are assumed to be zero almost surely. Note that the error term (which usually represents the unobserved individual heterogeneity) plays a crucial role in accounting for heterogeneous treatment effects.

To identify the DSFs, we first show that the DSFs are equivalent to the distribution functions of corresponding error terms with location shifts, and the corresponding magnitudes of location shifts are the conditional mean functions. We then show that the parameters in the conditional mean functions and the distribution functions of the error terms are identified under the unconfoundedness assumption that requires the treatment assignment to be independent of the potential outcomes conditional on covariates. Therefore the DSFs are identified and so are the QSFs. Based on the identification result, we employ the weighted nonlinear least squares (WNLS) estimator to estimate the parameters in the conditional mean functions. More importantly, we propose inverse probability weighted (IPW) estimators for the distribution functions of the error terms which are shown to be $\sqrt{n}$-consistent and to converge weakly to mean zero Gaussian processes. The estimators for DSFs and QSFs can be constructed accordingly and they are $\sqrt{n}$-consistent and converge weakly to mean zero Gaussian processes. We also propose a simulation approach to approximate these limiting processes and show the validity of the proposed approach.

In the empirical study we investigate the wage effects of smoking by using data from the Panel Study of Income Dynamics (PSID) in 2011. We find that there is a significant mean wage differential for men. Next, we go beyond the mean effect and study the structural quantile treatment effect (SQTE) that is defined as the difference between the QSF under treatment and that without treatment. We apply the two-sided uniform confidence band for the SQTE in this case and find that given a fixed value of observed characteristics, the smoking wage penalty only occurs for male workers with high unobserved heterogeneity. Put it differently, smoking does *not* decrease the wages for male workers with low unobserved heterogeneity. This result is new in the literature and is robust regardless of which fixed value of characteristics is selected. We also investigate the female counterparts. However, the mean wage differential is not significant for women, and the SQTE for female workers is sensitive to the selection of fixed value.

The structural functions are useful tools for an important empirical question raised by Fortin, Lemieux, and Firpo (2011): "by how much the treatment effect would change if [the covariate] were switched from its value ... to its average value?" The concept of structural functions has appeared in the literature, see e.g., Blundell and Powell (2003, 2004), Wooldridge (2005), Imbens and Newey (2009), Kasy (2011), and Chernozhukov, Fernandez-Val, Hahn, and Newey (2013).

This paper is related to Donald and Hsu (2014) where they are interested in the distribution functions and quantile functions of the potential outcomes in a nonparametric model; however, we are interested in the corresponding DSFs and QSFs in a separable semiparametric model. Donald, Hsu and Barrett (2012) focus on the conditional distribution functions and corresponding quantile functions in parametric, semiparametric and nonparametric models under the conditional independence assumption. The DSFs and QSFs are in general different from the conditional distribution functions and conditional quantile functions unless the conditional independence assumption holds as we will see in Section 2.2.

In addition, our method is related to but different from the decomposition method as in Fortin, Lemieux, and Firpo (2011), Rothe (2010), Chernozhukov, Fernandez-Val and Melly (2013) and Hsu, Lai and Lieli (2015) where they consider the counterfactual distribution of an outcome when the distribution of covariates is changed while holding the conditional distribution of the outcome given covariates fixed. Even if the distribution of covariates concentrates on one covariate value, the counterfactual distribution would be the conditional distribution functions instead of the DSF that we consider. Please see the discussion at the end of Section 2.2 for more details.

Remainder of this paper is organized as follows. In Section 2, we give a formal description of the model and the structural functions of interest. We also provide identification and estimation results for the DSFs and QSFs. Section 3 includes the asymptotic properties of the estimators and a simulation method to approximate the limiting processes. In Section 4, we discuss the asymptotics and the uniform confidence bands for the SQTE as well as a step-by-step implementation procedure. The empirical application of smoking on wages is illustrated in Section 5. All proofs are collected in the Appendix.

2 Model, Identification and Estimation of Structural Functions

2.1 Model Framework

Let D be the treatment indicator such that $D = 1$ if the individual receives treatment and $D = 0$ otherwise. Let X be a d_x-dimensional vector of covariates with a compact support $\mathcal{X} \subseteq \mathbb{R}^{d_x}$. Define Y_1 as the potential outcome for the individual under treatment and Y_0 as that without treatment. We observe D, X and $Y = D \cdot Y_1 + (1 - D) \cdot Y_0$. We have a random sample of size n. Assume that the potential outcomes Y_0 and Y_1 satisfy:

$$Y_0 = m_0(X, \beta_0) + \epsilon_0 \quad \text{and} \quad Y_1 = m_1(X, \beta_1) + \epsilon_1, \tag{2.1}$$

where for $j = 0$ and 1, $m_j(x, b_j) : \mathcal{X} \times \mathcal{B}_j \to \mathbb{R}$ represents the conditional mean function of $E[Y_j|X = x]$ that is known up to a finite-dimensional of parameter vector b_j which belongs to a compact subset $\mathcal{B}_j \subseteq \mathbb{R}^{d_{b_j}}$. For $j = 0$ and 1, let β_j denotes the true parameter vector in that $E[Y_j|X = x] = m_j(x, \beta_j)$ for all $x \in \mathcal{X}$. The error terms ϵ_0 and ϵ_1 denote the individual heterogeneity which are unobservable to econometricians.[1]

Remark:

1. The specification in (2.1) plays an important role in our paper as we will see soon. First, in the literature, models without separability are proposed, but we impose separability here so that the DSFs are the unconditional distributions of the errors terms with location shifts. Please see Section 2.2 for details. Second, our results can be extended to cases where conditional mean functions of the potential outcomes are nonparametrically specified. However, we impose the parametric specification on the conditional mean functions so the estimated conditional mean functions converge at the parametric rate instead of nonparametric rate. This is important to obtain estimators of DSFs and QSFs that converge at the parametric rate instead of nonparametric rate. Please see Section 3 for details.

2.2 Distribution and Quantile Structural Functions

For notational simplicity, we simply use j when the arguments or discussions apply to both $j = 0$ and 1 in the rest of the paper. For a specific covariate value x, we define the DSF of Y_j, $G_j(x, y)$, as the distribution function of $m_j(x, \beta_j) + \epsilon_j$. Specifically $G_j(x, y)$ is defined as

$$G_j(x, y) \equiv E_{\epsilon_j}\Big[1\big\{m_j(x, \beta_j) + \epsilon_j \le y\big\}\Big],$$

where the expectation is taken with respect to the *unconditional* distribution of ϵ_j and $1\{\cdot\}$ denotes the indicator function. The DSF can be interpreted as the distribution function of the counterfactual potential outcome when the covariate values of the population are exogenously switched to a fixed value x, while the heterogeneity of the individuals ϵ_j remains unchanged. By the definition of $G_j(x, y)$ and under (2.1), we can see that

$$G_j(x, y) = E_{\epsilon_j}\Big[1\big\{m_j(x, \beta_j) + \epsilon_j \le y\big\}\Big] = E_{\epsilon_j}\Big[1\big\{\epsilon_j \le y - m_j(x, \beta_j)\big\}\Big] = F_{\epsilon_j}\big(y - m_j(x, \beta_j)\big),$$

[1]Note that our theory allows for $m_0(x, b_0)$ and $m_1(x, b_1)$ to be different functions, and β_0 and β_1 to be of different dimensions.

where $F_{\epsilon_j}(\cdot)$ denotes the distribution function of ϵ_j. That is, the DSF, $G_j(x,\cdot)$, is equivalent to $F_{\epsilon_j}(\cdot)$ with a location shift whose magnitude is $m_j(x,\beta_j)$.

By the same logic, the QSF $q_j(x,\tau)$ proposed by Imbens and Newey (2009) can also be defined as the corresponding quantile function of $G_j(x,y)$. That is, for any $\tau \in [0,1]$,

$$q_j(x,\tau) \equiv \inf\big\{y : G_j(x,y) \geq \tau\big\}.$$

Under (2.1), we have $q_j(x,\tau) = m_j(x,\beta_j) + Q_{\epsilon_j}(\tau)$ where $Q_{\epsilon_j}(\tau)$ is the τ-th quantile of ϵ_j.

We summarize the relations between $F_{\epsilon_j}(\cdot)$, $Q_{\epsilon_j}(\cdot)$, $G_j(x,\cdot)$ and $q_j(x,\cdot)$ in the following lemma.

Lemma 2.1. *If Y_j is generated according to* (2.1)*, then*

$$G_j(x,y) = F_{\epsilon_j}\big(y - m_j(x,\beta_j)\big) \quad \textit{and} \quad q_j(x,\tau) = m_j(x,\beta_j) + Q_{\epsilon_j}(\tau).$$

Remarks:

1. Note that the DSF is in general not equal to the conditional distribution function of Y_j given $X = x$, $F_{Y_j|X}(y|x)$; however, they would be the same if ϵ_j is independent of X. That is,

$$\begin{aligned} F_{Y_j|X}(y|x) &= E_{\epsilon_j|X}\Big[1\big\{m_j(X,\beta_j) + \epsilon_j \leq y\big\}\Big|x\Big] = E_{\epsilon_j|X}\Big[1\big\{m_j(x,\beta_j) + \epsilon_j \leq y\big\}\Big|x\Big] \\ &= E_{\epsilon_j}\Big[1\big\{m_j(x,\beta_j) + \epsilon_j \leq y\big\}\Big] = G_j(x,y), \end{aligned}$$

   where the third equality holds (only) under the independence assumption. In this paper we focus on the case where ϵ_j is dependent of X, but when the independence assumption holds, the DSF coincides with the conditional distribution function of Y_j given $X = x$. Similarly, in general the QSF, $q_j(x,\tau)$, is different from the conditional quantile of Y_j given $X = x$ unless ϵ_j is independent of X.

2. Our method is different from the decomposition method in Rothe (2010), Fortin, Lemieux, and Firpo (2011), Chernozhukov, Fernandez-Val and Melly (2013) and Hsu, Lai and Lieli (2015) where they consider the counterfactual distribution of an outcome when the distribution of covariates is changed while holding the conditional distribution of the outcome given covariates fixed. To be specific, let $F_{Y|X}(y|x)$, $F_X(x)$ and $F_{\widetilde{X}}(x)$ denote conditional

CDF of Y given $X = x$, the CDF of X and CDF of $\widetilde{X}$, respectively. It is straightforward to see that the unconditional CDF of Y can be given by $F_Y(y) = \int F_{Y|X}(y|x)dF_X(x)$. Let $\widetilde{Y}$ be a counterfactual outcome with $F_{\widetilde{Y}}(y) = \int F_{Y|X}(y|x)dF_{\widetilde{X}}(x)$. Note that $\widetilde{Y}$ can be interpreted as the counterfactual outcome generated by changing the distributions of covariates (from $F_X(x)$ to $F_{\widetilde{X}}(x)$) while holding the conditional distribution of the outcome given covariates ($F_{Y|X}(y|x)$) fixed. Note that in this method, when the distribution of X changes, the overall individual heterogeneity changes as well while in our method, we would like the overall individual heterogeneity to be fixed.

On the other hand, if the density of covariates ($F_{\widetilde{X}}(x)$) concentrates on one point x, $F_{\widetilde{Y}}(y)$ would be the conditional distribution of $F_{Y|X}(y|x)$ instead of the DSF that we consider here and as we discussed above, DSF is in general different from the conditional distribution.

2.3 Identification

In this section, we discuss the identification of the DSFs and QSFs. From Lemma 2.1, it is true that once we show that $F_{\epsilon_j}(\cdot)$ and β_j are identifiable, then DSFs and QSFs are also identified. However as in the traditional treatment effect models, we face the same missing variable problem because we only observe one of the potential outcomes, i.e., $Y = D \cdot Y_1 + (1 - D) \cdot Y_0$ depending on the realization of the treatment status. This further implies that we would only observe one of the error terms even if β_j's were known. To resolve this, we assume that the treatment assignment is unconfounded. The unconfoundedness assumption introduced by Rosenbaum and Rubin (1983) requires that the treatment assignment be independent of the potential outcomes conditional on the observable covariates. Unconfoundedness assumption is also known as selection-on-observables, conditional independence, and ignorability in the literature. The formal definition of the unconfoundedness assumption is as follows.

Assumption 2.2. *Assume that $D \perp (Y_0, Y_1)|X$.*

By the specification of Y_0 and Y_1 in (2.1), Assumption 2.2 is equivalent to $D \perp (\epsilon_0, \epsilon_1)|X$. Let $p(x) = \Pr(D = 1|X = x)$ denote the propensity score, which is the conditional probability of receiving treatment given $X = x$. Following the literature, we assume $p(x)$ to be bounded away from 0 and 1.

Assumption 2.3. *Assume that* $0 < \underline{p} \le p(x) \le \overline{p} < 1$ *on* $\mathcal{X}$.

For the identification of β_0 and β_1, we then assume the following conditional moment restriction:

Assumption 2.4. *Assume that*

(i) $E\big[Y_j \big| X = x\big] = m_j(x, \beta_j)$ *for some* $\beta_j \in \mathcal{B}_j$.

(ii) $E\big[m_j(X, \beta_j) - m_j(X, b_j)\big]^2 > 0$ *for all* $b_j \in \mathcal{B}_j, b_j \neq \beta_j$.

Assumption 2.4(i) requires that conditional mean of Y_j in (2.1) is correctly specified which is equivalent to $E[\epsilon_j|X = x] = 0$ for all $x \in \mathcal{X}$. Assumption 2.4(ii) requires that β_j is the unique parameter value such that $E[Y_j - m_j(X, \beta_j)|X = x] = 0$ for all $x \in \mathcal{X}$.

Under Assumptions 2.2, 2.3 and 2.4, one can show that β_0 and β_1 are identified by

$$\beta_0 = \operatorname*{arg\,min}_{b_0 \in \mathcal{B}_0} E\big[Y_0 - m_0(X, b_0)\big]^2 = \operatorname*{arg\,min}_{b_0 \in \mathcal{B}_0} E\left[\frac{(1-D)}{1-p(X)} \cdot \big(Y - m_0(X, b_0)\big)^2\right],$$
$$\beta_1 = \operatorname*{arg\,min}_{b_1 \in \mathcal{B}_1} E\big[Y_1 - m_1(X, b_1)\big]^2 = \operatorname*{arg\,min}_{b_1 \in \mathcal{B}_1} E\left[\frac{D}{p(X)} \cdot \big(Y - m_1(X, b_1)\big)^2\right].$$

Given that β_0 and β_1 are identified, we can further identify $F_{\epsilon_j}(e)$ and $Q_{\epsilon_j}(\tau)$ in the following lemma.

Lemma 2.5. *Suppose Assumptions 2.2, 2.3 and 2.4 hold.* $F_{\epsilon_0}(e)$ *and* $F_{\epsilon_1}(e)$ *are identified by*

$$F_{\epsilon_0}(e) = E\left[\frac{(1-D) \cdot 1\big\{Y - m_0(X, \beta_0) \le e\big\}}{1 - p(X)}\right] \quad \textit{and}$$
$$F_{\epsilon_1}(e) = E\left[\frac{D \cdot 1\big\{Y - m_1(X, \beta_1) \le e\big\}}{p(X)}\right].$$

Moreover, $Q_{\epsilon_0}(\tau)$ *and* $Q_{\epsilon_1}(\tau)$ *are then identified by*

$$Q_{\epsilon_0}(\tau) = \inf\big\{e : F_{\epsilon_0}(e) \ge \tau\big\} \quad \textit{and} \quad Q_{\epsilon_1}(\tau) = \inf\big\{e : F_{\epsilon_1}(e) \ge \tau\big\}.$$

Given that β_j, $F_{\epsilon_j}(e)$ and $Q_{\epsilon_j}(\tau)$ are identified, $G_j(x, y)$ and $q_j(x, \tau)$ are identified based on Lemma 2.1.

2.4 Estimation

We propose estimators for β_j, $F_{\epsilon_j}(e)$, $Q_{\epsilon_j}(\tau)$, $G_j(x,y)$ and $q_j(x,\tau)$ in this section. First, based on the identification of β_j, we use the WNLS as in Wooldridge (2010) to estimate β_j:

$$\begin{aligned}
\hat{\beta}_0 &= \underset{b_0\in\mathcal{B}_0}{\arg\min}\,\frac{1}{n}\sum_{i=1}^{n}\frac{(1-D_i)}{1-\hat{p}(X_i)}\big(Y_i - m_0(X_i,b_0)\big)^2, \\
\hat{\beta}_1 &= \underset{b_1\in\mathcal{B}_1}{\arg\min}\,\frac{1}{n}\sum_{i=1}^{n}\frac{D_i}{\hat{p}(X_i)}\big(Y_i - m_1(X_i,b_1)\big)^2,
\end{aligned} \tag{2.2}$$

where $\hat{p}(x)$ is a nonparametric estimator for $p(x)$. As in Hirano, Imbens and Ridder (2003), we use the series logit estimator (SLE) to estimate $p(x)$ based on a power series. Other nonparametric estimators can be used, e.g., local polynomial estimator as in Ichimura and Linton (2005) and kernel estimator as in Abrevaya, Hsu and Lieli (2015). The main advantage of SLE over local polynomial and kernel estimators is that the estimated propensity score function based on SLE is automatically bounded away from 0 and 1, meaning that the trimming is not required. Note that the WNLS is not the most efficient estimator among the class of estimators under conditional moment restriction or Assumption 2.4. For example, Donald, Imbens, and Newey (2003) propose the empirical likelihood estimator where the efficiency is achieved as the number of restrictions grows with the sample size. Kitamura, Tripathi, and Ahn (2004) also propose the empirical likelihood-based procedure to obtain efficiency by using kernel smoothing method. Newey (2004) uses a moment tangent set to characterize when the semiparametric efficiency bound can be attained by the GMM. However, the focus of this paper is the estimation and inference on the DSFs and QSFs, so we use WNLS here for its simplicity.

Next, based on Lemma 2.5, the IPW estimators for $F_{\epsilon_0}(e)$ and $F_{\epsilon_1}(e)$ are

$$\begin{aligned}
\widehat{F}_{\epsilon_0}(e) &= \sum_{i=1}^{n}\frac{(1-D_i)\cdot 1\big\{Y_i - m_0(X_i,\hat{\beta}_0)\le e\big\}}{1-\hat{p}(X_i)}\Big/\sum_{i=1}^{n}\frac{1-D_i}{1-\hat{p}(X_i)}, \\
\widehat{F}_{\epsilon_1}(e) &= \sum_{i=1}^{n}\frac{D_i\cdot 1\big\{Y_i - m_1(X_i,\hat{\beta}_1)\le e\big\}}{\hat{p}(X_i)}\Big/\sum_{i=1}^{n}\frac{D_i}{\hat{p}(X_i)},
\end{aligned} \tag{2.3}$$

and the estimators for $Q_{\epsilon_0}(\tau)$ and $Q_{\epsilon_1}(\tau)$ are defined as

$$\widehat{Q}_{\epsilon_0}(\tau) = \inf\big\{e : \widehat{F}_{\epsilon_0}(e)\ge\tau\big\} \quad\text{and}\quad \widehat{Q}_{\epsilon_1}(\tau) = \inf\big\{e : \widehat{F}_{\epsilon_1}(e)\ge\tau\big\}.$$

Finally, we can estimate $G_j(x,y)$ and $q_j(x,\tau)$ according to Lemma 2.1:

$$\widehat{G}_j(x,y) = \widehat{F}_{\epsilon_j}\big(y - m_j(x,\hat{\beta}_j)\big) \quad\text{and}\quad \hat{q}_j(x,\tau) = m_j(x,\hat{\beta}_j) + \widehat{Q}_{\epsilon_j}(\tau). \tag{2.4}$$

3 Asymptotic Properties and the Simulation Method

In this section we discuss the asymptotic properties of estimators $\hat{\beta}_j$, $\widehat{F}_{\epsilon_j}(e)$, $\widehat{Q}_{\epsilon_j}(\tau)$, $\widehat{G}_j(x,y)$ and $\hat{q}_j(x,\tau)$, and then propose a simulation method to approximate the limiting processes. Note that these results are similar to Donald, Hsu and Barrett (2012) and Donald and Hsu (2014) in that to derive our theory, we need to account for to account for the estimation effect of $\hat{\beta}_j$ as in Donald, Hsu and Barrett (2012) and the estimation effect of the inverse probability weighting as in Donald and Hsu (2014) at the same time. We introduce all regularity conditions below.

3.1 Assumptions

Let $\nabla m_j(X,b_j)$ denote the $d_{b_j}\times 1$ gradient of $m_j(X,b_j)$ with respect to b_j and $\nabla^2 m_j(X,b_j)$ denote the $d_{b_j}\times d_{b_j}$ Hessian of $m_j(X,b_j)$. Let $\|\cdot\|$ denote the Euclidean norm of a matrix.

Assumption 3.1. *Assume that*

(i) The support of the d_x-dimensional covariates X is a Cartesian product of compact intervals, $\mathcal{X}=\prod_{i=1}^{d_x}[x_{i\ell},x_{iu}]$; the density of X is bounded away from 0 on $\mathcal{X}$.

(ii) The propensity score $p(x)$ is continuously differentiable of order $s\geq 7d_x$; the SLE of $p(x)$ uses a power series with $K=a\cdot n^{\nu}$ for some $a>0$ and $d_x/\big(4(s-d_x)\big)<\nu<1/9$.

Assumption 3.2. *Assume that*

(i) β_j is in the interior of $\mathcal{B}_j$ which is a compact subset of $\mathbb{R}^{d_{b_j}}$.

(ii) For each $b_j\in\mathcal{B}_j$, $m_j(\cdot,b_j)$ is Borel measurable on $\mathcal{X}$.

(iii) For each $x\in\mathcal{X}$, $m_j(x,\cdot)$ is continuously differentiable of order 2 in $b_j\in\mathcal{B}_j$.

(iv) $E\Big[\sup\limits_{b_j\in\mathcal{B}_j}\big\|\nabla m_j(X,b_j)\big\|^2\Big]<\infty$ and $E\Big[\sup\limits_{b_j\in\mathcal{B}_j}\big\|\nabla^2 m_j(X,b_j)\big\|^2\Big]<\infty$.

(v) $E\big[\nabla m_j(X,\beta_j)\nabla m_j(X,\beta_j)'\big]$ is positive definite.

Assumption 3.3. *Assume that*

(i) ϵ_0 and ϵ_1 have convex and compact supports $[e_{0\ell},e_{0u}]$ and $[e_{1\ell},e_{1u}]$; let $\mathcal{E}=[e_\ell,e_u]$ where $e_\ell=\min\{e_{0\ell},e_{1\ell}\}$ and $e_u=\max\{e_{0u},e_{1u}\}$.

(ii) $F_{\epsilon_0}(e)$ *and* $F_{\epsilon_1}(e)$ *have respectively density functions* $f_{\epsilon_0}(e)$ *and* $f_{\epsilon_1}(e)$ *which are bounded away from 0 and continuously differentiable of order 2 on* $[e_{0\ell}, e_{0u}]$ *and* $[e_{1\ell}, e_{1u}]$.

Assumption 3.1(i) requires that all of the covariates are continuous. It is not restrictive since we can deal with the case where X has both continuous and discrete components by sample splitting as in Donald and Hsu (2014) and please see Remark 5 after Assumption 3.5 in Donald and Hsu (2014) for details. Assumption 3.1(ii) regulates the rate at which additional terms are added to the series approximation to $p(x)$ depending on the dimension of X and the number of derivatives of $p(x)$. Assumption 3.2 consists of standard assumptions for asymptotic properties of the WNLS and please see Wooldridge (2010) for details. Assumption 3.3(i) requires that ϵ_0 and ϵ_1 have compact supports. This is not restrictive in that the theory regarding the estimators for $F_{\epsilon_0}(e)$ and $F_{\epsilon_1}(e)$ remains the same when ϵ_0 and ϵ_1 have supports on the whole real line. If this is the case, we need to assume that $\mathrm{Var}(\epsilon_0) < \infty$ and $\mathrm{Var}(\epsilon_1) < \infty$ (which hold automatically when ϵ_0 and ϵ_1 have compact supports) so that the following result of WNLS would hold. Assumption 3.3(ii) implies that $F_{\epsilon_0}(e)$ and $F_{\epsilon_1}(e)$ are strictly increasing on $[e_{0\ell}, e_{0u}]$ and $[e_{1\ell}, e_{1u}]$, respectively.

3.2 Asymptotic Properties

The following lemma summarizes the asymptotic properties of $\hat{\beta}_j$.

Lemma 3.4. *Suppose Assumptions 2.2, 2.3, 2.4, 3.1 and 3.2 hold. Let* $\hat{\beta}_j$ *be the estimator defined in* (2.2)*. Then*

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\beta_0}(W_i, \beta_0) + o_p(1) \quad \text{and} \quad \sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\beta_1}(W_i, \beta_1) + o_p(1),$$

where $W_i \equiv \{X_i, Y_i, D_i\}$ *and*

$$\psi_{\beta_0}(W, \beta_0) = E\big[\nabla m_0(X, \beta_0) \nabla m_0(X, \beta_0)'\big]^{-1} \frac{(1-D)\cdot \nabla m_0(X, \beta_0)\big[Y - m_0(X, \beta_0)\big]}{1 - p(X)},$$
$$\psi_{\beta_1}(W, \beta_1) = E\big[\nabla m_1(X, \beta_1) \nabla m_1(X, \beta_1)'\big]^{-1} \frac{D \cdot \nabla m_1(X, \beta_1)\big[Y - m_1(X, \beta_1)\big]}{p(X)}.$$

Lemma 3.4 shows that the estimation error of $\hat{p}(x)$ is asymptotically negligible, i.e., the asymptotic variance of $\hat{\beta}_j$ is equal to the variance of $\psi_{\beta_j}(X, Y, D, \beta_j)$ despite the fact that $\hat{p}(x)$ is obtained from the first-stage SLE.

The following lemma summarizes the influence function representation for the IPW estimator for $F_{\epsilon_j}(e)$ as shown in (2.3).

Lemma 3.5. *Suppose Assumptions 2.2, 2.3, 2.4, 3.1, 3.2 and 3.3 hold. Let $\widehat{F}_{\epsilon_j}(e)$ be the estimator defined in* (2.3). *Then*

$$\sqrt{n}\big(\widehat{F}_{\epsilon_0}(e) - F_{\epsilon_0}(e)\big) \Rightarrow \Psi_{\epsilon_0}(e) \quad \textit{and} \quad \sqrt{n}\big(\widehat{F}_{\epsilon_1}(e) - F_{\epsilon_1}(e)\big) \Rightarrow \Psi_{\epsilon_1}(e),$$

where $\Rightarrow$ denotes weak convergence and $\Psi_{\epsilon_0}(e)$ and $\Psi_{\epsilon_1}(e)$ are mean zero Gaussian processes with the covariance kernels being generated by the influence functions

$$\begin{aligned}
\psi_{\epsilon_0}(W,e) &= \left\{\frac{(1-D)\cdot 1\big\{Y - m_0(X,\beta_0) \le e\big\}}{1-p(X)} - F_{\epsilon_0}(e)\right\} + \frac{F_{\epsilon_0|X}(e|X)}{1-p(X)}\big(D - p(X)\big) \\
&\quad + f_{\epsilon_0}(e)E\big[\nabla m_0(X,\beta_0)\big]'\psi_{\beta_0}(W,\beta_0), \\
\psi_{\epsilon_1}(W,e) &= \left\{\frac{D\cdot 1\big\{Y - m_1(X,\beta_1) \le e\big\}}{p(X)} - F_{\epsilon_1}(e)\right\} - \frac{F_{\epsilon_1|X}(e|X)}{p(X)}\big(D - p(X)\big) \\
&\quad + f_{\epsilon_1}(e)E\big[\nabla m_1(X,\beta_1)\big]'\psi_{\beta_1}(W,\beta_1),
\end{aligned}$$

where $F_{\epsilon_j|X}(e|x)$ is the conditional distribution function of ϵ_j given $X = x$, $f_{\epsilon_j}(e)$ is the density function of ϵ_j and $\psi_{\beta_j}(W,\beta_j)$ is defined in Lemma 3.4.

Note that the first term of each influence function is the influence function if we substitute the true propensity score $p(x)$ for $\hat{p}(x)$ and true β_j for $\hat{\beta}_j$. The second term gives the contribution of estimating $p(x)$ to the asymptotic process of $\widehat{F}_{\epsilon_j}(e)$. The last term is the estimation error from the WNLS, $\hat{\beta}_j$.

By (2.4) and Lemma 3.5, the influence function representations for $\widehat{G}_j(x,y)$ and $\hat{q}_j(x,\tau)$ are summarized in the following theorems.

Theorem 3.6. *Suppose the assumptions in Lemma 3.5 hold. Then*

$$\sqrt{n}\big(\widehat{G}_0(x,y) - G_0(x,y)\big) \Rightarrow \Psi_{G_0}(x,y) \quad \textit{and} \quad \sqrt{n}\big(\widehat{G}_1(x,y) - G_1(x,y)\big) \Rightarrow \Psi_{G_1}(x,y),$$

where $\Psi_{G_0}(x,y)$ and $\Psi_{G_1}(x,y)$ are mean zero Gaussian processes with the covariance kernels *being generated by the influence functions*

$$\psi_{G_0}(W,x,y) = \left\{\frac{(1-D)\cdot 1\big\{Y - m_0(X,\beta_0) \le y - m_0(x,\beta_0)\big\}}{1-p(X)} - F_{\epsilon_0}\big(y - m_0(x,\beta_0)\big)\right\}$$
$$+ \frac{F_{\epsilon_0|X}\big(y - m_0(x,\beta_0)\big|X\big)}{1-p(X)}\cdot\big(D - p(X)\big)$$
$$+ f_{\epsilon_0}\big(y - m_0(x,\beta_0)\big)\big\{E\big[\nabla m_0(X,\beta_0)\big] - \nabla m_0(x,\beta_0)\big\}'\psi_{\beta_0}(W,\beta_0),$$
$$\psi_{G_1}(W,x,y) = \left(\frac{D\cdot 1\big\{Y - m_1(X,\beta_1) \le y - m_1(x,\beta_1)\big\}}{p(X)} - F_{\epsilon_1}\big(y - m_1(x,\beta_1)\big)\right)$$
$$- \frac{F_{\epsilon_1|X}\big(y - m_1(x,\beta_1)\big|X\big)}{p(X)}\cdot\big(D - p(X)\big)$$
$$+ f_{\epsilon_1}\big(y - m_1(x,\beta_1)\big)\big\{E\big[\nabla m_1(X,\beta_1)\big] - \nabla m_1(x,\beta_1)\big\}'\psi_{\beta_1}(W,\beta_1),$$

where $\psi_{\beta_0}(W,\beta_0)$ and $\psi_{\beta_1}(W,\beta_1)$ are defined in Lemma 3.4.

Similar to the remark after Lemma 3.5, the first term of each influence functions is the influence function that would obtain if we substitute the true β_j and $p(x)$ for the estimators $\hat{\beta}_j$ and $\hat{p}(x)$. The second term gives the contribution of estimating $p(x)$ to the asymptotic process of $\widehat{G}_j(x,y)$. The last term is again the estimation error from the WNLS, $\hat{\beta}_j$. Note that the estimation error attributed to $\hat{\beta}_j$ affects two parts: (i) the distribution itself depends on $\hat{\beta}_j$ and (ii) the point where we evaluate is estimated by $y - m_j(x,\hat{\beta}_j)$.

Next, given that the quantile map is Hadamard differentiable, the limiting process of $\hat{q}_j(x,\tau)$ follows from Theorem 3.6 and the functional delta method. We summarize the asymptotic properties for $\hat{q}_0(x,\tau)$ and $\hat{q}_1(x,\tau)$ in the following lemma.

Theorem 3.7. *Suppose the assumptions in Theorem 3.6 hold. Then*

$$\sqrt{n}\big(\hat{q}_0(x,\tau) - q_0(x,\tau)\big) \Rightarrow \Psi_{q_0}(x,\tau) \quad \textit{and} \quad \sqrt{n}\big(\hat{q}_1(x,\tau) - q_1(x,\tau)\big) \Rightarrow \Psi_{q_1}(x,\tau),$$

where $\Psi_{q_0}(x,\tau)$ and $\Psi_{q_1}(x,\tau)$ are mean zero Gaussian processes such that

$$\Psi_{q_0}(x,\tau) = -\frac{\Psi_{G_0}\big(x, q_0(x,\tau)\big)}{g_0\big(x, q_0(x,\tau)\big)} \quad \textit{and} \quad \Psi_{q_1}(x,\tau) = -\frac{\Psi_{G_1}\big(x, q_1(x,\tau)\big)}{g_1\big(x, q_1(x,\tau)\big)},$$

where $\Psi_{G_0}(x,\cdot)$ and $\Psi_{G_1}(x,\cdot)$ are defined in Theorem 3.6 and $g_j(x,y) \equiv f_{\epsilon_j}\big(y - m_j(x,\beta_j)\big)$ is the corresponding density function of $G_j(x,y)$.

3.3 Simulation Method

Although the asymptotic properties of $\hat{\beta}_j$, $\widehat{F}_{\epsilon_j}(e)$, $\widehat{Q}_{\epsilon_j}(\tau)$, $\widehat{G}_j(x,y)$ and $\hat{q}_j(x,\tau)$ shown in the previous section can be used for the pointwise inference provided the influence functions are consistently estimated, drawing uniform inference is not an easy task even if the nonparametric bootstrap is adopted because it may be time-consuming since for each replication $p(x)$, β_j have to be estimated. In this section we suggest an alternative simulation-based method which allows us to approximate the limiting processes $\Psi_{G_j}(x,y)$ and $\Psi_{q_j}(x,\tau)$ in Theorems 3.6 and 3.7. The idea is to first construct a simulated process by the estimated pointwise influence function, and then show the simulated process can approximate the true one well by the conditional multiplier central limit theorem. Note that in doing so we require that the estimation errors of $\widehat{F}_{\epsilon_j|X}(\cdot|x)$ and $\hat{f}_{\epsilon_j}(\cdot)$ disappear in the limit, which is true given the estimators are uniformly consistent. We introduce a series-based estimator for $F_{\epsilon_j|X}(\cdot|x)$ and a kernel-based estimator for $f_{\epsilon_j}(\cdot)$ in the Appendix. The estimators play important roles in the simulation-based method. For more details, please see Donald, Hsu, and Barrett (2012), and Donald and Hsu (2014).

The estimated pointwise influence functions $\hat{\psi}_{G_0}(W_i,x,y)$ and $\hat{\psi}_{G_1}(W_i,x,y)$ are defined as

$$\begin{aligned}
&\hat{\psi}_{G_0}(W_i,x,y)\\
&= \left\{\frac{(1-D_i)\cdot 1\{Y_i - m_0(X_i,\hat{\beta}_0) \le y - m_0(x,\hat{\beta}_0)\}}{1-\hat{p}(X_i)} - \widehat{F}_{\epsilon_0}\big(y - m_0(x,\hat{\beta}_0)\big)\right\}\\
&\quad + \frac{\widehat{F}_{\epsilon_0|X}\big(y - m_0(x,\hat{\beta}_0)\big|X_i\big)}{1-\hat{p}(X_i)}\cdot\big(D_i - \hat{p}(X_i)\big)\\
&\quad + \widehat{f}_{\epsilon_0}\big(y - m_0(x,\hat{\beta}_0)\big)\{\nabla_{b_0}\overline{m}_0(X_i,\hat{\beta}_0) - \nabla_{b_0}m_0(x,\hat{\beta}_0)\}'\widehat{\psi}_{\beta_0}(W_i,\hat{\beta}_0),\\
&\hat{\psi}_{G_1}(W_i,x,y)\\
&= \left\{\frac{D_i\cdot 1\{Y_i - m_1(X_i,\hat{\beta}_1) \le y - m_1(x,\hat{\beta}_1)\}}{\hat{p}(X_i)} - \widehat{F}_{\epsilon_1}\big(y - m_1(x,\hat{\beta}_1)\big)\right\}\\
&\quad - \frac{\widehat{F}_{\epsilon_1|X}\big(y - m_1(x,\hat{\beta}_1)\big|X_i\big)}{\hat{p}(X_i)}\cdot\big(D_i - \hat{p}(X_i)\big)\\
&\quad + \widehat{f}_{\epsilon_1}\big(y - m_1(x,\hat{\beta}_1)\big)\{\nabla_{b_1}\overline{m}_1(X_i,\hat{\beta}_1) - \nabla_{b_1}m_1(x,\hat{\beta}_1)\}'\widehat{\psi}_{\beta_1}(W_i,\beta_1),
\end{aligned} \tag{3.1}$$

where

$$\hat{\psi}_{\beta_0}(W_i,\hat{\beta}_0)=\left\{\frac{1}{n}\sum_{i=1}^{n}\nabla m_0(X_i,\hat{\beta}_0)\nabla m_0(X_i,\hat{\beta}_0)'\right\}^{-1}\frac{(1-D_i)\cdot\nabla m_0(X_i,\hat{\beta}_0)\big(Y_i-m_0(X_i,\hat{\beta}_0)\big)}{1-\hat{p}(X_i)},$$

$$\hat{\psi}_{\beta_1}(W_i,\hat{\beta}_1)=\left\{\frac{1}{n}\sum_{i=1}^{n}\nabla m_1(X_i,\hat{\beta}_1)\nabla m_1(X_i,\hat{\beta}_1)'\right\}^{-1}\frac{D_i\cdot\nabla m_1(X_i,\hat{\beta}_1)\big(Y_i-m_1(X_i,\hat{\beta}_1)\big)}{\hat{p}(X_i)},$$

$$\overline{m}_0(X_i,\hat{\beta}_0)=\frac{1}{n}\sum_{i=1}^{n}m_0(X_i,\hat{\beta}_0)\quad\text{and}\quad\overline{m}_1(X_i,\hat{\beta}_1)=\frac{1}{n}\sum_{i=1}^{n}m_1(X_i,\hat{\beta}_1).$$

Given the influence functions are consistently estimated, let $\{U_1,U_2,\ldots\}$ be i.i.d. pseudo random variables with mean 0 and variance 1 which are independent of the whole sample. The simulated processes for $\Psi_{G_0}(x,y)$ and $\Psi_{G_1}(x,y)$ are

$$\Psi^u_{G_0}(x,y)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\cdot\hat{\psi}_{G_0}(W_i,x,y)\quad\text{and}\quad\Psi^u_{G_1}(x,y)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\cdot\hat{\psi}_{G_1}(W_i,x,y).\tag{3.2}$$

Theorem 3.8. *Suppose the assumptions in Theorem 3.6 hold. Also suppose that $\widehat{F}_{\epsilon_j|X}(e|x)$ and $\hat{f}_{\epsilon_j}(e)$ are uniformly consistent estimators for $F_{\epsilon_j|X}(e|x)$ and $f_{\epsilon_j}(e)$. Then*

$$\Psi^u_{G_0}(x,y)\Rightarrow\Psi_{G_0}(x,y)\quad\textit{and}\quad\Psi^u_{G_1}(x,y)\Rightarrow\Psi_{G_1}(x,y),$$

conditional on the sample path with probability 1, which are denoted by $\Psi^u_{G_0}(x,y)\overset{p}{\Rightarrow}\Psi_{G_0}(x,y)$ and $\Psi^u_{G_1}(x,y)\overset{p}{\Rightarrow}\Psi_{G_1}(x,y)$.

Next, we define the simulated processes for $\Psi_{q_0}(x,\tau)$ and $\Psi_{q_1}(x,\tau)$ as

$$\Psi^u_{q_0}(x,\tau)=-\frac{\Psi^u_{G_0}\big(x,\hat{q}_0(x,\tau)\big)}{\hat{g}_0\big(x,\hat{q}_0(x,\tau)\big)}\quad\text{and}\quad\Psi^u_{q_1}(x,\tau)=-\frac{\Psi^u_{G_1}\big(x,\hat{q}_1(x,\tau)\big)}{\hat{g}_1\big(x,\hat{q}_1(x,\tau)\big)},\tag{3.3}$$

where $\Psi^u_{G_j}(x,\cdot)$ is defined in (3.2) and $\hat{g}_j(x,y)=\hat{f}_{\epsilon_j}\big(y-m_j(x,\hat{\beta}_j)\big)$ is the IPW kernel estimator in the Appendix.

Theorem 3.9. *Suppose the assumptions in Theorem 3.8 hold. Then*

$$\Psi^u_{q_0}(x,\tau)\overset{p}{\Rightarrow}\Psi_{q_0}(x,\tau)\quad\textit{and}\quad\Psi^u_{q_1}(x,\tau)\overset{p}{\Rightarrow}\Psi_{q_1}(x,\tau).$$

4 Structural Quantile Treatment Effect

The preceding section provides a way to estimate and draw uniform inferences on the DSFs and QSFs. With these results one can test for the stochastic dominance relation similar to the references in Donald and Hsu (2014) or test for the Lorenz dominance relation as in Barrett, Donald, and Bhattacharya (2014). In this paper, however, we propose to estimate the SQTE which can be treated as the structural version of the QTE in Firpo (2007). The SQTE $\Delta(x,\tau)$ is defined as the difference between QSFs, i.e., for any $\tau \in [0,1]$,

$$\Delta(x,\tau) \equiv q_1(x,\tau) - q_0(x,\tau).$$

The main differences between this paper and Firpo (2007) is that. we focus on $\Delta(x,\tau)$ uniformly over a continuum of quantile indices which of course includes the pointwise SQTE as a special case.

In what follows, we first discuss the asymptotic properties of the estimator for SQTE, $\widehat{\Delta}(x,\tau) = \hat{q}_1(x,\tau) - \hat{q}_0(x,\tau)$, and then provide methods to construct two-sided and one-sided uniform confidence bands for a continuum of SQTE, $\{\Delta(x,\tau) : \tau \in [\delta, 1-\delta]\}$ for some $\delta \in [0,1/2)$. Finally, a step-by-step implementation procedure to conduct the uniform confidence band is also demonstrated.

4.1 Asymptotic Properties

The following corollary summarizes the asymptotics of $\widehat{\Delta}(x,\tau)$.

Corollary 4.1. *Suppose the assumptions in Theorem 3.7 hold. Then*

$$\sqrt{n}\left(\widehat{\Delta}(x,\tau) - \Delta(x,\tau)\right) \Rightarrow \Psi_{q_1}(x,\tau) - \Psi_{q_0}(x,\tau),$$

where $\Psi_{q_1}(x,\tau)$ and $\Psi_{q_0}(x,\tau)$ are defined in Theorem 3.7. Moreover, suppose the assumptions in Theorem 3.9 hold. Then

$$\Psi^u_{q_1}(x,\tau) - \Psi^u_{q_0}(x,\tau) \overset{p}{\Rightarrow} \Psi_{q_1}(x,\tau) - \Psi_{q_0}(x,\tau),$$

where $\Psi^u_{q_1}(x,\tau)$ and $\Psi^u_{q_0}(x,\tau)$ are defined in (3.3).

For a confidence level α and for $[\delta, 1-\delta]$, $\delta \in [0,1/2)$, let $\widehat{C}^{\alpha}_{\text{2-sided},\delta}$ and $\widehat{C}^{\alpha}_{\text{1-sided},\delta}$ denote the bounds respectively for the two-sided and one-sided uniform confidence bands for $\{\Delta(x,\tau) : \tau \in [\delta, 1-\delta]\}$ that satisfy

$$\begin{aligned}
\widehat{C}^{\alpha}_{\text{2-sided},\delta} &= \inf\left\{ e : \Pr\left(\sup_{\tau\in[\delta,1-\delta]} \left|\Psi_{q_1}(x,\tau) - \Psi_{q_0}(x,\tau)\right| \le e \right) \ge \alpha \right\}, \\
\widehat{C}^{\alpha}_{\text{1-sided},\delta} &= \inf\left\{ e : \Pr\left(\sup_{\tau\in[\delta,1-\delta]} \left(\Psi_{q_1}(x,\tau) - \Psi_{q_0}(x,\tau)\right) \le e \right) \ge \alpha \right\}.
\end{aligned} \tag{4.1}$$

That is, $\widehat{C}^{\alpha}_{\text{2-sided},\delta}$ and $\widehat{C}^{\alpha}_{\text{2-sided},\delta}$ are the α-th quantiles of $\left|\Psi_{q_1}(x,\tau)-\Psi_{q_0}(x,\tau)\right|$ and $\Psi_{q_1}(x,\tau)-\Psi_{q_0}(x,\tau)$, respectively. Then the two-sided uniform confidence band for the SQTE is given by

$$\left\{\left(\widehat{\Delta}(x,\tau)-\frac{\widehat{C}^{\alpha}_{\text{2-sided},\delta}}{\sqrt{n}},\quad \widehat{\Delta}(x,\tau)+\frac{\widehat{C}^{\alpha}_{\text{2-sided},\delta}}{\sqrt{n}}\right):\tau\in[\delta,1-\delta]\right\},$$

and the two one-sided uniform confidence bands are

$$\left\{\left(-\infty,\quad \widehat{\Delta}(x,\tau)+\frac{\widehat{C}^{\alpha}_{\text{1-sided},\delta}}{\sqrt{n}}\right):\tau\in[\delta,1-\delta]\right\}\quad\text{and}$$
$$\left\{\left(\widehat{\Delta}(x,\tau)-\frac{\widehat{C}^{\alpha}_{\text{1-sided},\delta}}{\sqrt{n}},\quad \infty\right):\tau\in[\delta,1-\delta]\right\}.$$

The following theorem summarize the asymptotic coverage rates of the uniform confidence bands we just defined.

Theorem 4.2. *Suppose the assumptions in Theorem 3.9 hold. Then for a confidence level α and for $\delta\in[0,1/2)$,*

$$\lim_{n\to\infty}\Pr\left(\Delta(x,\tau)\in\left(\widehat{\Delta}(x,\tau)-\frac{\widehat{C}^{\alpha}_{\textit{2-sided},\delta}}{\sqrt{n}},\quad \widehat{\Delta}(x,\tau)+\frac{\widehat{C}^{\alpha}_{\textit{2-sided},\delta}}{\sqrt{n}}\right)\ \textit{for all}\ \tau\in[\delta,1-\delta]\right)=\alpha,$$
$$\lim_{n\to\infty}\Pr\left(\Delta(x,\tau)\in\left(-\infty,\quad \widehat{\Delta}(x,\tau)+\frac{\widehat{C}^{\alpha}_{\textit{1-sided},\delta}}{\sqrt{n}}\right)\ \textit{for all}\ \tau\in[\delta,1-\delta]\right)=\alpha,$$
$$\lim_{n\to\infty}\Pr\left(\Delta(x,\tau)\in\left(\widehat{\Delta}(x,\tau)-\frac{\widehat{C}^{\alpha}_{\textit{1-sided},\delta}}{\sqrt{n}},\quad \infty\right)\ \textit{for all}\ \tau\in[\delta,1-\delta]\right)=\alpha,$$

where $\widehat{C}^{\alpha}_{\textit{2-sided},\delta}$ and $\widehat{C}^{\alpha}_{\textit{1-sided},\delta}$ are defined in (4.1).

Note that one can also construct the confidence band for the standardized SQTE. Our results extend to this case easily given the availability of the estimators that are uniformly consistent for the standard errors over $\tau\in[\delta,1-\delta]$.

4.2 Implementation Procedure

In this section we provide an implementation procedure for $\widehat{\Delta}(x,\tau)$ as well as the setup of the estimation in the empirical analysis.

1. **Prerequisite:** we have a random sample of $\{Y_i,X_i,D_i\}_{i=1}^n$.

2. Estimate $\hat{p}(X_i)$ via the SLE: Construct a power series of X_i and run a Logit regression of D_i on the power series. Then predict the fitted value as $\hat{p}(X_i)$. In the empirical study we use a quadratic form of the power series which contains 209 linear independent bases. $\hat{p}(X_i)$ is also trimmed to lie in the interval [0.005, 0.995].

3. **Estimate $\hat{\beta}_j$ via the WNLS:** Estimate $\hat{\beta}_j$ as in (2.2). In the empirical study we consider a linear specification of $m_j(X, \beta_j)$. The closed form of $\hat{\beta}_j$ is therefore given by

$$\hat{\beta}_0 = \left\{ \sum_{i=1}^{n} \frac{(1-D_i)\cdot X_i X_i'}{1-\hat{p}(X_i)} \right\}^{-1} \sum_{i=1}^{n} \frac{(1-D_i)\cdot X_i Y_i}{1-\hat{p}(X_i)},$$

$$\hat{\beta}_1 = \left\{ \sum_{i=1}^{n} \frac{D_i \cdot X_i X_i'}{\hat{p}(X_i)} \right\}^{-1} \sum_{i=1}^{n} \frac{D_i \cdot X_i Y_i}{\hat{p}(X_i)}.$$

In addition, the corresponding influence function $\psi_{\beta_j}(W, \beta_j)$ can be estimated by

$$\hat{\psi}_{\beta_0}(W_i, \hat{\beta}_0) = \left\{ \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right\}^{-1} \frac{(1-D_i)\cdot X_i \big(Y_i - X_i \hat{\beta}_0\big)}{1-\hat{p}(X_i)},$$

$$\hat{\psi}_{\beta_1}(W_i, \hat{\beta}_1) = \left\{ \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right\}^{-1} \frac{D_i \cdot X_i \big(Y_i - X_i \hat{\beta}_1\big)}{\hat{p}(X_i)}.$$

4. **Estimate $\widehat{F}_{\epsilon_j}(e)$ via the IPW:** Estimate $\widehat{F}_{\epsilon_j}(e)$ according to (2.3) for each $e \in \mathcal{E}$. In practice we set 1,000 points for e, i.e., $e = -5, -4.99, \dots, 5$.

5. **Estimate $\widehat{G}_j(x, y)$ and $\hat{\psi}_{G_j}(Y_i, X_i, D_i, x, y)$:** First, set a fixed value x which is of the same dimension of X. Next, use (2.4) to estimate $\widehat{G}_j(x, y)$. We again assign 1,000 evenly spaced points in the support of Y_i for y. Finally, estimate $\hat{\psi}_{G_j}(Y_i, X_i, D_i, x, y)$ as in (3.1). Note that we use the same series as in SLE for $\widehat{F}_{\epsilon_j|X}(\cdot|x)$ and adopt the Gaussian kernel with the bandwidth $h = 1.06 \cdot \hat{\sigma}_j n^{-1/5}$ for $\hat{f}_{\epsilon_j}(\cdot)$, where $\hat{\sigma}_j$ is the sample standard deviation of $\hat{\epsilon}_{ij} = Y_i - X_i' \hat{\beta}_j$.

6. **Simulation method:** Draw i.i.d. pseudo random variables $\{U_i\}_{i=1}^{n}$ with mean 0 and variance 1 for B times, say $B = 1,000$. For each replication $b = 1, \dots, B$, calculate the simulated process $\Psi^u_{q_j,b}(x, \tau)$ according to (3.3) and store the maximum absolute difference between $\Psi^u_{q_0,b}(x, \tau)$ and $\Psi^u_{q_1,b}(x, \tau)$ among τ, i.e., $M_b \equiv \max_\tau |\Psi^u_{q_1,b}(x, \tau) - \Psi^u_{q_0,b}(x, \tau)|$. We let $\delta = 0.1$ so that $\tau \in [0.1, 0.9]$. Rank M_b in an ascending order such that $M_{(1)} \leq \dots \leq M_{(B)}$ and then define $M_{(\lfloor \alpha B \rfloor)}$ as the two-sided uniform bound $C^{\alpha}_{\text{2-sided},\delta}$, where $\lfloor c \rfloor$

is the floor function returning the largest integer not greater than c. The α two-sided uniform confidence band for the $\Delta(x, \tau)$ is therefore given by

$$\left[\widehat{\Delta}(x,\tau) - \frac{C^{\alpha}_{\text{2-sided},\delta}}{\sqrt{n}}, \quad \widehat{\Delta}(x,\tau) + \frac{C^{\alpha}_{\text{2-sided},\delta}}{\sqrt{n}}\right].$$

We depict the estimated SQTE and its 90% two-sided uniform confidence band in the empirical analysis.

5 Wage Effects of Smoking

In this section we revisit the wage effects of smoking (see Levine, Gustafson, and Velenchik (1997), van Ours (2004), Auld (2005), and Grafova and Stafford (2009) among others) to illustrate the proposed methods. This study evaluates the smoking wage penalty beyond the mean. Similar to previous findings, we find a significant negative effect of smoking on mean wage for men. In addition, keeping observed characteristics fixed, smoking imposes a wage penalty for men only in the higher quantiles of their unobserved heterogeneity. Unobserved heterogeneity could be interpreted as an individual's ability. Under this interpretation our findings suggest that individuals with better ability suffer more from smoking. By contrast, smoking does not depress the wages of low ability individuals. This is new to the literature and our results are robust to different fixed values of the observed characteristics. Finally, we also consider the female counterparts. In contrast to the results of men, the mean wage differential is insignificant for women and the corresponding SQTE is sensitive to the selection of fixed value of observed characteristics. This may be attributed to the diversity of women's returns to observed characteristics, which overwhelms the unobserved heterogeneity.

5.1 Literature Review

There are several possible channels through which smoking is related to wages (see, for example, Levine, Gustafson, and Velenchik (1997), and Grafova and Stafford (2009)). They can be broadly classified into the causal and non-causal ones. The harmful effect of smoking on health, which reduces productivity, and discrimination against smokers are the main causal channels proposed in the literature. On the other hand, the relationship could be non-causal with smoking being simply an indicator of high rate of time preference that leads to less investment in human capital, and thus lowering wages.

To examine the wage effects of smoking and the potential mechanism of the relationship, Levine, Gustafson and Velenchick (1997) employ data from the National Longitudinal Survey of Youth to find a 4–8% wage penalty for smokers using OLS and difference-in-differences (by time and siblings) approaches. However, without accounting for possible heterogeneous smoking effects the causal and non-causal explanations cannot be teased out. In recognition of the heterogeneity of smokers, Grafova and Stafford (2009) reconstruct a retrospective sample using three waves of the Panel Study of Income Dynamics (PSID). This retrospective sample allows them to obtain heterogeneous effects conditional on personal smoking history. The finding suggests a 8–12% wage penalty for persistent smokers versus never smokers or former smokers.

Moreover, van Ours (2004) and Auld (2005), using Dutch and Canadian data respectively, take into account the simultaneous effects of smoking and alcohol use. After controlling for unobserved heterogeneity, van Ours (2004) finds that male drinkers earn about 10% more than non-drinkers and smokers earn about 10% less than non-smokers, while the wages for females are not affected by either smoking or drinking behavior. In addition, he finds that there is substantial interaction effect between smoking and drinking. Without controlling for drinking, the harmful effect of smoking is underestimated. However, in his analysis a relatively small set of covariates (age and educational level only) are controlled such that the estimates are likely to be confounded by unobserved characteristics.

On the other hand, Auld (2005) estimates a simultaneous equation system, consisting of smoking and drinking and income by a method of maximum simulated likelihood. Without accounting for unobserved heterogeneity, the results suggest that smokers earn 8% less than non-smokers. However, the system results show that the smoking penalty is 24%, indicating that the unobserved heterogeneity is highly correlated with smoking behavior.

In conclusion, to examine the wage effects of smoking one should be aware that (i) the smoking effect is heterogeneous, (ii) the unobserved factors are crucial, and (iii) smoking and drinking behavior must be considered simultaneously. Heeding these observations from previous findings, we specify a model of potential outcomes for average hourly wage, i.e.,

$$Y_0 = X'\beta_0 + \epsilon_0 \quad \text{and} \quad Y_1 = X'\beta_1 + \epsilon_1. \tag{5.1}$$

Although the model does not allow the returns to observed characteristics β_j's to vary with individuals, it captures the heterogeneous effects by the unobserved heterogeneity ϵ_j. In the next section we describe our data and discuss the selection of covariates X in detail.

5.2 Data

We use data from the 2011 wave of the PSID. Began in 1968, the PSID is the longest running household panel survey collecting rich information on respondents' economic and demographic status. The PSID collects data on smoking behavior in the 1986 wave and every wave since the 1999 one. The sample size increases from 4,800 in the first wave to over 8,900 in the 2011 wave. Similar to Grafova and Stafford (2009), we restrict our attention to household heads between the ages of 25 and 60 who were employed full-time (i.e., working at least 1,500 hours annually). This yields a sample of 3,127 men and 1,061 women.

In the empirical analysis, we control for individuals' education, job tenure, years of work experience, age, occupation (whether white-collar or not) geographic location (whether residing in the south or not), union status, health status (self-reported), marital status, spousal smoking status and history (i.e., smoking currently and ever smoked), religious participation (i.e., frequency of attendance at religious services), and state cigarette prices. State cigarette prices are obtained from the Tax Burden on Tobacco (Orzechowski and Walker, 2012). Moreover, to control for the interaction effect of drinking, we also include drinking status (whether or not a drinker and drinking intensity, i.e., number of alcoholic drinks consumed daily) in the wage equation. The outcome variable is the log of average hourly wage rate. All individual characteristics are summarized in Table 1.

Note that in general the unconfoundedness assumption which is the key assumption in our model is not testable without extra conditions as in Huber (2013), Donald, Hsu and Lieli (2014) and de Luna and Johansson (2014). In these three paper, an assumption in common is that there is a valid instrument for the treatment assignment. However, unfortunately in our dataset, we can not find such instrument. As a result, we would assume that unconfoundedness assumption holds after controlling a set of covariates following van Ours (2004).

5.3 Main Results

Our main results are two-fold. We first report the average wage effects of smoking via various estimation approaches, and then discuss the SQTE. Table 2 shows the estimates of ATEs via OLS, IPW, and weighted least squares (WLS), respectively. We see that the mean wage penalty of smoking for men is around 6–12% which is in line with previous findings. In addition, we also report the estimates of returns to observed characteristics obtained by OLS and WLS in Table 3. It should come as no surprise that the OLS and WLS coefficient estimates are similarly, and the details for the estimation can be found in the implementation procedure described above.

The second result is summarized in Figure 1. To begin with, Figure 1(a) shows the histogram of estimated propensity score which is bounded away from 0 and 1. Next, the quantile functions of the observed wages are depicted in Figure 1(b) for smokers (thick line) and non-smokers (thin line), along with the mean wages (dash lines). It is clear that non-smokers have higher hourly wages than smokers at all quantiles. Even after controlling for self-selectivity of smoking behavior by comparing the potential wages, the same pattern appears in Figure 1(c). However, the differences between potential wages are substantial in two aspects. It may either come from the difference in characteristics (i.e., explained component) or the differences in the returns to observed characteristics and unobserved heterogeneity (i.e., unexplained component). We focus on the latter. In fact, the WLS coefficient estimates in Table 3 represent the returns to observed characteristics and the potential unobserved heterogeneity is illustrated by Figure 1(d).

From Figure 1(d) we can now find a clue about the striking result because the penalty is now reversed in the lower quantiles. Put it differently, aside from the impacts owing to observed characteristics, individuals with lower unobserved heterogeneity will surprisingly gain from smoking. This implies that individuals with lower unobserved ability gain more from the beneficial effects of smoking – smokers are less likely to develop Parkinson's and Alzheimer's diseases (Fratiglioni and Wang, 2000), smoking may also increase mental concentration (Stough, Mangan, Bates, and Pellett, 1994), and smokers have a lower incidence of some inflammatory and neurodegenerative diseases (Sopori, 2002) – as well as for individuals with higher unobserved ability the harmful effects of smoking outweigh the beneficial ones.

It is not satisfying to merely compare the potential unobserved heterogeneity, however, since it does not account for the returns to observed characteristics. To complete our task we propose to estimate the SQTE. We first introduce a set of fixed values for the observed characteristics (see Table 4 for details). The combinations of the fixed values are set in a way that we examine the effect of smoking for individuals with typical observed characteristics. For example, a Type I male is white, not residing in the South, college educated, white-collared, in "very good" self-rated health status, an alcohol drinker, and married with a non-smoking spouse. A Type II male is non-white, residing in the South, a high school diploma or GED holder, in "very good" self-rated health status, an alcohol drinker, married with a non-smoking spouse. Note that for other continuous variables we assign the fixed values to the sample means in the original Type I and II subgroups.

After pinning down the fixed values, the corresponding QSFs are presented in Figures 1(e) and 1(f). More importantly, the SQTEs are delineated in Figure 1(g) for Type I manipulation and Figure 1(h) for Type II manipulation with 90% uniform confidence bands respectively. It is obvious from these figures that given characteristics fixed, the wage differentials occur at high quantiles only. At the lower end of the SQTE there is no significant wage difference between smokers and non-smokers.

To conclude our finding, the results suggests that there is a significant wage penalty for smokers at the mean for men. Furthermore, keeping observed characteristics fixed, smoking depresses the wages for men with high unobserved heterogeneity only, whereas for men with low unobserved heterogeneity smoking does not have any effect. This result is robust to fixing the observed characteristics at different values. These results suggest that men with lower unobserved ability gain more from the beneficial effect of smoking, while for those with a higher unobserved ability the harmful effects outweigh the beneficial ones.

5.4 Female Case

Most papers in this literature focus on the wage effect of smoking for men only, except for van Ours (2004) who shows that there is no average wage penalty for women. It may be interesting to see whether there is wage penalty of smoking at different quantiles for women. In this section we perform the same analysis for female household heads of the PSID whose descriptive statistics can be found in Table 5. Note that we drop the variables related to spousal smoking status since the PSID only records spousal smoking behavior for male household heads only.

Table 6 reports various average wage effects of smoking which are around 10–11%. However, not all estimates are statistically significant at conventional levels. Table 7 contains the estimates of the returns to observed characteristics by the OLS and WLS. Unlike the results for men, the coefficient estimates vary dramatically between smokers and non-smokers, yielding very different pictures in the end of the analysis. Figure 2 is the female counterpart of Figure 1. The most prominent difference is that the wage effect of smoking is now sensitive to the choice of fixed values for the observed characteristics. While Figure 2(g) shows that the smoker/non-smoker wage gap is positive for Type I manipulation, the gap becomes negative for Type II manipulation as can be seen from Figure 2(h). The sensitivity of the smoker/non-smoker wage gap is a direct consequence of the huge differences in women's returns to observed characteristics between smokers and non-smokers.

6 Conclusion

This paper identifies and estimates the DSFs and QSFs in a separable semiparametric treatment effect model under the unconfoundedness assumption. We propose IPW the DSF and QSF estimators which are $\sqrt{n}$-consistent and converge weakly to mean zero Gaussian processes. We also propose a simulation approach to approximate the limiting processes and uniform confidence bands for the SQTE.

In the empirical study we revisit the wage effects of smoking. We find a significant mean wage difference between male smokers and non-smokers which is in line with previous findings. Keeping characteristics fixed, there is no wage penalty for male smokers who have low unobserved heterogeneity according to the uniform confidence band for the SQTE. This finding is robust to different fixed values of observed characteristics. However, in the female case the smoker/non-smoker wage gap in terms of SQTE is sensitive to the choice of fixed values.

APPENDIX

A Estimators for the Simulation Method

A.1 Series-Based Estimator for $F_{\epsilon_j|X}(e|x)$

We construct estimator for $F_{\epsilon_j|X}(e|x)$ similar to that in Donald and Hsu (2014). Note that the estimator must satisfy the regularity coditions: (i) bounded between 0 and 1, (ii) monotonically increasing in e for any given x, and (iii) converging in probability to $F_{\epsilon_j}(e|x)$ uniformly in both arguments e and x. The estimator plays an important role in the simulation-based method.

Let $\widetilde{F}_{\epsilon_0|X}(e|x)$ and $\widetilde{F}_{\epsilon_1|X}(e|x)$ be the series estimators for $F_{\epsilon_0|X}(e|x)$ and $F_{\epsilon_1|X}(e|x)$,

$$\widetilde{F}_{\epsilon_0|X}(e|x) = \left\{\sum_{i=1}^{n} \frac{(1-D_i)\cdot 1\{Y_i - m_0(X_i,\hat{\beta}_0) \le e\}}{1-\hat{p}(X_i)} R^K(X_i)\right\}' \left\{\sum_{i=1}^{n} R^K(X_i)R^K(X_i)'\right\}^{-1} R^K(x),$$

$$\widetilde{F}_{\epsilon_1|X}(e|x) = \left\{\sum_{i=1}^{n} \frac{D_i\cdot 1\{Y_i - m_1(X_i,\hat{\beta}_1) \le e\}}{\hat{p}(X_i)} R^K(X_i)\right\}' \left\{\sum_{i=1}^{n} R^K(X_i)R^K(X_i)'\right\}^{-1} R^K(x),$$

where $R^K(x)$ is the same power series used in SLE estimator. $\widetilde{F}_{\epsilon_j|X}(e|x)$ is a step function in e with jumps at $Y_i - m_j(X_i,\hat{\beta}_j)$'s for any given x. However, although $\widetilde{F}_{\epsilon_j|X}(e|x)$ converges in probability to $F_{\epsilon_j|X}(e|x)$ uniformly in both arguments e and x, it is neither necessarily bounded between 0 and 1 nor necessarily monotonically increasing in e for any given x. Nevertheless, we can construct estimator for $F_{\epsilon_j|X}(e|x)$ satisfying all requirements based on $\widetilde{F}_{\epsilon_j|X}(e|x)$.

Let $\varepsilon_i \equiv Y_i - m_j(X_i,\hat{\beta}_j)$ for brevity. Without loss of generality, we assume that $\mathcal{E} = [0,\bar{e}]$ and there are no ties between ε_i's. We add $\varepsilon_{(0)} = 0$ and $\varepsilon_{(n+1)} = \bar{e}$. Let $\varepsilon_{(i)}$ denote the i-th smallest element among the ε_i's so that we have $0 = \varepsilon_{(0)} < \varepsilon_{(1)} < \cdots < \varepsilon_{(n)} < \varepsilon_{(n+1)} = \bar{e}$. We define $\widehat{F}_{\epsilon_j|X}(e|x)$ by induction. Define $\widehat{F}_{\epsilon_j|X}(e|x) = \widetilde{F}_{\epsilon_j|X}(e|x) = 0$ for $\varepsilon_{(0)} \le e < \varepsilon_{(1)}$ and $\widehat{F}_{\epsilon_j|X}(\varepsilon_{(n+1)}|x) = 1$. Suppose $\widehat{F}_{\epsilon_j|X}(e|x) = 0$ is defined for $\varepsilon_{(0)} \le e < \varepsilon_{(i)}$, then we define for $\varepsilon_{(i)} \le e < \varepsilon_{(i+1)}$,

$$\begin{aligned}\widehat{F}_{\epsilon_j|X}(e|x) = {} & \widehat{F}_{\epsilon_j|X}(\varepsilon_{(i-1)}|x)\cdot 1\{0 \le \widetilde{F}_{\epsilon_j|X}(\varepsilon_{(i)}|x) \le \widehat{F}_{\epsilon_j|X}(\varepsilon_{(i-1)}|x)\} \\ & + \widetilde{F}_{\epsilon_j|X}(\varepsilon_{(i)}|x)\cdot 1\{\widehat{F}_{\epsilon_j|X}(\varepsilon_{(i-1)}|x) < \widetilde{F}_{\epsilon_j|X}(\varepsilon_{(i)}|x) \le 1\} \\ & + 1\{\widetilde{F}_{\epsilon_j|X}(\varepsilon_{(i)}|x) > 1\}.\end{aligned}$$

The idea is that if $\widetilde{F}_{\epsilon_j|X}(e|x)$ jumps down at $\varepsilon_{(i)}$, then we set $\widehat{F}_{\epsilon_j|X}(e|x) = \widehat{F}_{\epsilon_j}(\varepsilon_{(i-1)}|x)$ for $\varepsilon_{(i)} \le e < \varepsilon_{(i+1)}$. At the same time, we trim $\widetilde{F}_{\epsilon_j|X}(e|x)$ between 0 and 1 by defining $\widehat{F}_{\epsilon_j|X}(e|x) = 0$ when $\widetilde{F}_{\epsilon_j|X}(e|x) < 0$ and defining $\widehat{F}_{\epsilon_j|X}(e|x) = 1$ when $\widetilde{F}_{\epsilon_j|X}(e|x) > 1$. The properties of $\widehat{F}_{\epsilon_0|X}(e|x)$ and $\widehat{F}_{\epsilon_1|X}(e|x)$ are summarized in the following lemma.

Lemma A.1. *Suppose Assumptions 2.2, 2.3, 2.4, 3.1, 3.2 and 3.3 hold. Then for any given x, $\widehat{F}_{\epsilon_0|X}(e|x)$ and $\widehat{F}_{\epsilon_1|X}(e|x)$ are bounded between 0 and 1 and monotonically increasing in e, and*

$$\sup_{e\in\mathcal{E},x\in\mathcal{X}}\left|\widehat{F}_{\epsilon_0|X}(e|x)-F_{\epsilon_0|X}(e|x)\right|=o_p(1)\quad \textit{and}\quad \sup_{e\in\mathcal{E},x\in\mathcal{X}}\left|\widehat{F}_{\epsilon_1|X}(e|x)-F_{\epsilon_1|X}(e|x)\right|=o_p(1).$$

Lemma A.1 follows from $\sup_{e\in\mathcal{E},x\in\mathcal{X}}\left|\widetilde{F}_{\epsilon_j|X}(e|x)-F_{\epsilon_j|X}(e|x)\right|=o_p(1)$ and $\sup_{x\in\mathcal{X}}\left|\widehat{F}_{\epsilon_j|X}(e|x)-F_{\epsilon_j|X}(e|x)\right|\leq \sup_{x\in\mathcal{X}}\left|\widetilde{F}_{\epsilon_j|X}(e|x)-F_{\epsilon_j|X}(e|x)\right|=o_p(1)$ for all $x\in\mathcal{X}$. Note that the compactness of $\mathcal{X}$ is needed to obtain the uniform result. One can also use the kernel estimator to estimate $F_{\epsilon_j|X}(e|x)$ instead of the series estimator. The simulation result remains the same provided the kernel estimator has the properties in Lemma A.1.

A.2 Kernel-Based Estimator for $f_{\epsilon_j}(e)$

Besides $F_{\epsilon_j|X}(e|x)$, we need to estimate $f_{\epsilon_j}(e)$ before proposing the simulation method to approximate the process of $\Psi_{G_j}(x,y)$. We introduce an IPW kernel estimator for $f_{\epsilon_j}(e)$ which is consistent uniformly over $\mathcal{E}$. Let h denote a bandwidth which depends on sample size n and $\mathcal{K}(u)$ a kernel function. For $e\in[e_\ell+h,e_u-h]$, define $\tilde{f}_{\epsilon_0}(e)$ and $\tilde{f}_{\epsilon_1}(e)$ as

$$\begin{aligned}\tilde{f}_{\epsilon_0}(e)&=\frac{1}{nh}\sum_{i=1}^{n}\frac{1-D_i}{1-\hat{p}(X_i)}\mathcal{K}\left(\frac{Y_i-m_0(X_i,\hat{\beta}_0)-e}{h}\right),\\ \tilde{f}_{\epsilon_1}(e)&=\frac{1}{nh}\sum_{i=1}^{n}\frac{D_i}{\hat{p}(X_i)}\mathcal{K}\left(\frac{Y_i-m_1(X_i,\hat{\beta}_1)-e}{h}\right).\end{aligned}$$

The estimators for $f_{\epsilon_0}(e)$ and $f_{\epsilon_1}(e)$ are defined as

$$\hat{f}_{\epsilon_0}(e)=\begin{cases}\tilde{f}_{\epsilon_0}(e_\ell+h) & \text{if } e\in[e_\ell,e_\ell+h)\\ \tilde{f}_{\epsilon_0}(e) & \text{if } e\in[e_\ell+h,e_u-h]\\ \tilde{f}_{\epsilon_0}(e_u-h) & \text{if } e\in(e_u-h,e_u]\end{cases}\quad\text{and}\quad \hat{f}_{\epsilon_1}(e)=\begin{cases}\tilde{f}_{\epsilon_1}(e_\ell+h) & \text{if } e\in[e_\ell,e_\ell+h)\\ \tilde{f}_{\epsilon_1}(e) & \text{if } e\in[e_\ell+h,e_u-h]\\ \tilde{f}_{\epsilon_1}(e_u-h) & \text{if } e\in(e_u-h,e_u]\end{cases}.$$

The reason to use $\hat{f}_{\epsilon_j}(e)$ instead of $\tilde{f}_{\epsilon_j}(e)$ is because $\tilde{f}_{\epsilon_j}(e)$ is in general inconsistent around the boundary point e_ℓ. Therefore, we modify $\tilde{f}_{\epsilon_j}(e)$ around the boundary point to obtain uniformly consistent estimator for $f_{\epsilon_j}(e)$. This method is also used in Donald, Hsu, and Barrett (2012) and Donald and Hsu (2014). We make the following assumptions on $\mathcal{K}(\cdot)$ and h.

Assumption A.2. *Assume that*

(i) The kernel function $\mathcal{K}(\cdot)$ is nonnegative, symmetric around 0, continuous differentiable of order 1 and has support $[-1,1]$.

(ii) The bandwidth h satisfies that $h\to 0, nh^4\to\infty$ and $nh/\log n\to\infty$ when $n\to\infty$.

Lemma A.3. *Suppose Assumptions 2.2, 2.3, 2.4, 3.1, 3.2, 3.3 and A.2 hold. Then*

$$\sup_{e\in\mathcal{E}}\left|\hat{f}_{\epsilon_0}(e)-f_{\epsilon_0}(e)\right|=o_p(1)\quad \textit{and}\quad \sup_{e\in\mathcal{E}}\left|\hat{f}_{\epsilon_1}(e)-f_{\epsilon_1}(e)\right|=o_p(1).$$

B Proofs

Proof of Lemma 2.1

For the DSF,

$$G_j(x,y)=E\Big[1\big\{m_j(x,\beta_j)+\epsilon_j\le y\big\}\Big]=E\Big[1\big\{\epsilon_j\le y-m_j(x,\beta_j)\big\}\Big]=F_{\epsilon_j}\big(y-m_j(x,\beta_j)\big).$$

The QSF follows by

$$\begin{aligned}Q_j(x,\tau)&=\inf\big\{y:G_j(x,y)\ge\tau\big\}=\inf\big\{y:F_{\epsilon_j}\big(y-m_j(x,\beta_j)\big)\ge\tau\big\}\\&=m_j(x,\beta_j)+\inf\big\{y-m_j(x,\beta_j):F_{\epsilon_j}\big(y-m_j(x,\beta_j)\big)\ge\tau\big\}=m_j(x,\beta_j)+Q_{\epsilon_j}(\tau).\end{aligned}$$

□

Proof of Lemma 2.5

We show the case of $F_{\epsilon_1}(\cdot)$ and the argument for $F_{\epsilon_0}(\cdot)$ is similar. By law of iterated expectations,

$$\begin{aligned}&E\left[\frac{D\cdot 1\big\{Y-m_1(X,\beta_1)\le e\big\}}{p(X)}\right]\\&=E\left[E\left[\frac{D\cdot 1\big\{Y-m_1(X,\beta_1)\le e\big\}}{p(X)}\middle|X=x\right]\right]\\&=E\left[\frac{1}{p(x)}E\Big[D\cdot 1\big\{Y-m_1(X,\beta_1)\le e\big\}\Big|X=x,D=1\Big]\Pr(D=1|X=x)\right]\\&=E\Big[E\Big[1\big\{Y_1-m_1(X,\beta_1)\le e\big\}\Big|X=x,D=1\Big]\Big]\\&=E\Big[E\Big[1\big\{Y_1-m_1(X,\beta_1)\le e\big\}\Big|X=x\Big]\Big]\\&=E\Big[1\big\{Y-m_1(X,\beta_1)\le e\big\}\Big]=F_{\epsilon_1}(e).\end{aligned}$$

The second equality holds by expanding the conditional expectation according to D. The third equality comes from $\Pr(D=1|X=x)=p(x)$ and $Y=Y(1)$ when $D=1$. By Assumption 2.2, the fourth equality holds. Then use law of iterated expectations again the fifth equality holds. Since $F_{\epsilon_j}(\cdot)$ is identifiable, $Q_{\epsilon_j}(\cdot)$ is also identifiable by definition. □

Proof of Lemma 3.4

We show the case of $\hat{\beta}_1$ and the argument for $\hat{\beta}_0$ is similar. By a mean-value expansion about β_1 in the first-order condition for $\hat{\beta}_1$, it can be shown that

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = -E\Big[H\big(\beta_1, p(X)\big)\Big]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s\big(\beta_1, \hat{p}(X_i)\big)\right] + o_p(1),$$

where the score and Hessian are defined as

$$\begin{aligned} s\big(\beta_1, p(X)\big) &= -\frac{D \cdot \nabla m_1(X,\beta_1)\big[Y - m_1(X,\beta_1)\big]}{p(X)}, \\ H\big(\beta_1, p(X)\big) &= -\frac{D \cdot \nabla^2 m_1(X,\beta_1)\big[Y - m_1(X,\beta_1)\big]}{p(X)} + \frac{D \cdot \nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'}{p(X)}. \end{aligned}$$

Similar to the proof of Lemma 2.5, we can show that $E\big[s\big(\beta_1, p(X)\big)\big]$ and the first term of $E\big[H\big(\beta_1, p(X)\big)\big]$ are zero under Assumption 2.4(i). Therefore,

$$\begin{aligned} E\Big[H\big(\beta_1, p(X)\big)\Big] &= E\left[\frac{D \cdot \nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'}{p(X)}\right] \\ &= E\left[E\left[\frac{D \cdot \nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'}{p(X)}\middle| X = x\right]\right] \\ &= E\left[\frac{1}{p(x)}E\big[D \cdot \nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'\big| X = x, D = 1\big]\Pr(D = 1|X = x)\right] \\ &= E\left[E\big[\nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'\big| X = x, D = 1\big]\right] \\ &= E\left[E\big[\nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'\big| X = x\big]\right] = E\left[\nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'\right]. \end{aligned}$$

Next, by replacing Y_i's with $-\nabla m_1(X_i,\beta_1)\big[Y_i - m_1(X_i,\beta_1)\big]$'s in the addendum of Hirano, Imbens, and Ridder (2003), it is true that

$$\begin{aligned} \Bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s\big(\beta_1, \hat{p}(X_i)\big) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{ &-\frac{D_i \cdot \nabla m_1(X_i,\beta_1)\big[Y_i - m_1(X_i,\beta_1)\big]}{p(X_i)} \\ &+ \frac{\nabla m_1(X_i,\beta_1)E\big[Y_1 - m_1(X,\beta_1)\big|X = X_i\big]}{p(X_i)}\big(D_i - p(X_i)\big)\Bigg\}\Bigg| = o_p(1), \end{aligned}$$

where the last term of left-hand side is actually zero by Assumption 2.4(i). Thus,

$$\begin{aligned} &\left|\sqrt{n}(\hat{\beta}_1 - \beta_1) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{E\Big[H\big(\beta_1, p(X)\big)\Big]^{-1}\frac{D_i \cdot \nabla m_1(X_i,\beta_1)\big[Y_i - m_1(X_i,\beta_1)\big]}{p(X_i)}\right\}\right| \\ &= \left|\sqrt{n}(\hat{\beta}_1 - \beta_1) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{E\big[\nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'\big]^{-1}\frac{D_i \cdot \nabla m_1(X_i,\beta_1)\big[Y_i - m_1(X_i,\beta_1)\big]}{p(X_i)}\right\}\right| \\ &= o_p(1). \end{aligned}$$

That is, $\hat{\beta}_1$ is asymptotically linear with the influence function

$$\psi_{\beta_1}(x, y, D, \beta_1) = E\big[\nabla m_1(X,\beta_1)\nabla m_1(X,\beta_1)'\big]^{-1}\frac{D \cdot \nabla m_1(X,\beta_1)\big[Y - m_1(X,\beta_1)\big]}{p(X)}. \qquad \square$$

Proof of Lemma 3.5

The proof is a combination of Theorem 3.3 of Donald, Hsu, and Barrett (2012), and Theorem 3.6 of Donald and Hsu (2014) so we omit it. □

Proof of Theorem 3.6

Given Lemma 2.1 and Lemma 3.5, the proof is similar to Theorem 3.3 of Donald, Hsu and Barrett (2012). □

Proof of Theorem 3.7

Given that the quantile map is Hadamard differentiable, the functional delta method applies to the result. A similar proof can be found in Theorem 3.8 of Donald and Hsu (2014). □

Proof of Theorem 3.8

The proof is similar to Theorem 4.2 of Donald and Hsu (2014). □

Proof of Theorem 3.9

The proof is similar to Theorem 4.5 of Donald and Hsu (2014). □

Proof of Corollary 4.1

The result follows directly from Theorems 3.7 and 3.9. □

Proof of Theorem 4.2

The argument is standard in the literature, see, e.g., Bickel and Krieger (1989) or Rothe (2010). □

Proof of Lemma A.1

The proof is similar to Lemma 4.1 of Donald and Hsu (2014). □

Proof of Lemma A.3

The proof is similar to Lemma 4.4 of Donald and Hsu (2014). □

References

Abrevaya, J., Y.-C. Hsu and R. P. Lieli (2015): "Estimating Conditional Average Treatment Effects," *Journal of Business and Economic Statistics*, **33**, 485–505.

Auld, M. C. (2005): "Smoking, Drinking, and Income," *Journal of Human Resources*, **40**, 505–518.

Barrett, G. F., S. G. Donald and D. Bhattacharya (2014): "Consistent Nonparametric Tests for Lorenz Dominance," *Journal of Business and Economic Statistics*, **32**, 1–13.

Bickel, P. J. and A. M. Krieger (1989): "Confidence Bands for a Distribution Function Using the Bootstrap," *Journal of the American Statistical Association*, **84**, 95–100.

Blundell, R. and J. L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," *Advances in Economics and Econometrics*, **2**, 312–357.

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

Chernozhukov, V., I. Fernández-Val, J. Hahn and W. K. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, **81**, 535–580.

Chernozhukov, V., I. Fernández-Val and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, **81**, 2205–2268.

Xavier de Luna, X. and P. Johansson (2014): "Testing for the Unconfoundedness Assumption Using an Instrumental Assumption," *Journal of Causal Inference*, **2**, 187–199.

Donald, S. G. and Y.-C. Hsu (2014): "Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models," *Journal of Econometrics*, **178**, 383–397.

Donald, S. G., Y.-C. Hsu and G. F. Barrett (2012): "Incorporating Covariates in the Measurement of Welfare and Inequality: Methods and Applications," *The Econometrics Journal*, **15**, C1–C30.

Donald, S. G., Y.-C. Hsu and R. P. Lieli (2014): "Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT," *Journal of Business and Economic Statistics*, **32**, 395–415.

Donald, S. G., G. W. Imbens and W. K. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, **117**, 55–93.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, **75**, 259–276.

Fortin, N., T. Lemieux and S. Firpo (2011): "Decomposition Methods in Economics," in D. Card and O. Ashenfelter (Eds.), *Handbook of Labor Economics*, **4**, 1–102. Elsevier.

Fratiglioni, L. and H. X. Wang (2000): "Smoking and Parkinson's and Alzheimer's Disease: Review of the Epidemiological Studies," *Behavioural Brain Research*, **113**, 117–120.

Grafova, I. B. and F. P. Stafford (2009): "The Wage Effects of Personal Smoking History," *Industrial and Labor Relations Review*, **62**, 381–393.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, **66**, 315–331.

Heckman, J., H. Ichimura, J. Smith and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, **66**, 1017–1098.

Heckman, J., H. Ichimura and P. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, **64**, 605–654.

Heckman, J., H. Ichimura and P. Todd (1998): "Matching as an Econometric Evaluations Estimator," *Review of Economic Studies*, **65**, 261–294.

Hirano, K., G. Imbens and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, **71**, 1161–1189.

Hsu, Y.-C., T.-C. Lai and R. P. Lieli (2015): "Forecasting Treatment Effects," Working Paper.

Huber, M. (2013): "A Simple Test for the Ignorability of Non-compliance in Experiments," *Economics Letters*, **120**, 389–391.

Ichimura, H. and O. Linton (2005): "Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators," in D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models*, 149–170. Cambridge University Press.

Imbens, G. W. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*, **77**, 1481–1512.

Imbens, G. W. and J. W. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, **47**, 5–86.

Kasy, M. (2011): "Identification in Triangular Systems Using Control Functions," *Econometric Theory*, **27**, 663–71.

Kitamura, Y., G. Tripathi and H. Ahn (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, **72**, 1667–1714.

Levine, P. B., T. A. Gustafson and A. D. Velenchik (1997): "More Bad News for Smokers? The Effects of Cigarette Smoking on Wages," *Industrial and Labor Relations Review*, **50**, 493–509.

Newey, W. K. (2004): "Efficient Semiparametric Estimation via Moment Restrictions," *Econometrica*, **72**, 1877–1897.

Orzechowski, W. and R. Walker (2012): *The Tax Burden on Tobacco: Historical compilation.* **47**.

Rosenbaum, P. and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, **70**, 41–55.

Rothe, C. (2010): "Nonparametric Estimation of Distributional Policy Effect," *Journal of Econometrics*, **155**, 56–70.

Sopori, M. (2002): "Effects of Cigarette Smoke on the Immune System," *Nature Reviews Immunology*, **2**, 372–377.

Stough, C., G. Mangan, T. Bates and O. Pellett (1994): "Smoking and Raven IQ," *Psychopharmacology,* **116**, 382–384.

US Census Bureau (2003): *2000 census of population and housing*, Bureau of Census.

van Ours, J. C. (2004): "A Pint a Day Raises a Man's Pay; But Smoking Blows that Gain Away," *Journal of health economics*, **23**, 863–886.

Wooldridge, J. M. (2005): "Unobserved Heterogeneity and Estimation of Average Partial Effects," in D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models*, 27–55. Cambridge University Press.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, 2nd Edition. The MIT Press.

Table 1: General Characteristics, by Smoking Status (Male)[a]

	Smokers			Non-Smokers		
	Mean	Min	Max	Mean	Min	Max
Average Hourly Wage	19.36	0.06	200.00	29.91	0.05	723.46
Education – College Degree	0.17	0	1	0.46	0	1
Education – Some College, No College Degree	0.20	0	1	0.19	0	1
Education – High School Diploma or GED	0.47	0	1	0.29	0	1
No. Years Experience with Current Employer	6.71	0	41	8.02	0	42
No. Years Worked for Money Since 18 Years Old	10.97	0	57	10.52	0	41
White-Collar Job[b]	0.16	0	1	0.38	0	1
Job Covered by Union Contract	0.11	0	1	0.15	0	1
Health – Excellent	0.13	0	1	0.26	0	1
Health – Very Good	0.38	0	1	0.41	0	1
Health – Good	0.37	0	1	0.26	0	1
Drinker	0.79	0	1	0.74	0	1
No. Alcoholic Drinks Consumed Daily	3.04	0	25	1.88	0	24
Non-White Race	0.35	0	1	0.30	0	1
South Residence	0.41	0	1	0.40	0	1
Age	39.82	25	60	41.40	25	60
Marital Status	0.56	0	1	0.76	0	1
Wife – Smoker	0.26	0	1	0.06	0	1
Wife – Former Smoker	0.12	0	1	0.17	0	1
Frequency of Attendance at Religious Services	2.00	0	50	3.02	0	80
State Cigarette Price[c]	5.63	4.30	9.85	5.79	4.30	9.85
Sample Size		596			2,531	

[a] The sample includes male household heads between 25 and 60 years old, working at least 1,500 hours a year with positive average hourly wage.

[b] white-collar dummy equals one if individual's occupation code falls into management, professional, and related categories classified by 2000 Census of Population and Housing.

[c] We match the weighted average cigarette price per package (generic brands included) from the Tax Burden on Tobacco (Orzechowski and Walker, 2012) to PSID respondents by their state of residence.

Table 2: Average Smoking Effect on Wages (Male)

	OLS		IPW		WLS	
	Smokers	Non-Smokers	Smokers	Non-Smokers	Smokers	Non-Smokers
ATE	-0.061^{*}		-0.105^{*}		-0.120^{***}	
	(0.035)		(0.063)		(0.032)	
Unconditional Mean	2.976	3.037	2.936	3.042	2.921	3.040
	(0.033)	(0.016)	(0.062)	(0.015)	(0.030)	(0.015)

* Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level. Robust standard errors in parentheses. Power series in SLE includes constant, linear, interaction, and squared terms.

Table 3: OLS and WLS Estimates, by Smoking Status (Male)

	OLS		WLS	
	Smokers	Non-Smokers	Smokers	Non-Smokers
Education – College Degree	0.436***	0.478***	0.446***	0.437***
	(0.087)	(0.063)	(0.110)	(0.063)
Education – Some College	0.329***	0.315***	0.377***	0.281***
	(0.072)	(0.062)	(0.086)	(0.062)
Education – High School or GED	0.141**	0.158***	0.093	0.113*
	(0.068)	(0.061)	(0.076)	(0.061)
Tenure	0.045***	0.040***	0.055***	0.039***
	(0.008)	(0.005)	(0.012)	(0.005)
Tenure Squared	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Experience	−0.011	−0.011*	−0.021*	−0.013**
	(0.008)	(0.006)	(0.011)	(0.006)
Experience Squared	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
White-Collar Job	0.326***	0.285***	0.291***	0.272***
	(0.071)	(0.035)	(0.101)	(0.037)
Union	0.254***	0.086***	0.137*	0.090***
	(0.062)	(0.031)	(0.077)	(0.033)
Health – Excellent	0.042	0.182***	−0.136	0.187***
	(0.111)	(0.064)	(0.112)	(0.062)
Health – Very Good	0.174*	0.128**	0.198*	0.126**
	(0.091)	(0.061)	(0.107)	(0.060)
Health – Good	0.149*	0.018	0.149	0.025
	(0.090)	(0.062)	(0.097)	(0.060)
Drinker	0.332***	0.174***	0.152**	0.143***
	(0.073)	(0.036)	(0.075)	(0.036)
Drinking Intensity	−0.019*	−0.012*	0.002	−0.005
	(0.010)	(0.006)	(0.012)	(0.007)
Race	−0.246***	−0.165***	−0.180***	−0.181***

Table 3: OLS and WLS Estimates, by Smoking Status (Continued)

	OLS		WLS	
	Smokers	Non-Smokers	Smokers	Non-Smokers
	(0.053)	(0.030)	(0.066)	(0.032)
Region	0.043	0.045	0.015	0.042
	(0.047)	(0.031)	(0.069)	(0.032)
Age	0.024	0.057***	0.033	0.055***
	(0.024)	(0.013)	(0.031)	(0.013)
Age Squared	0.000	−0.001***	0.000	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Marital Status	0.148***	0.157***	0.096	0.147***
	(0.049)	(0.033)	(0.059)	(0.034)
Wife – Smoker	−0.045	−0.081	−0.020	−0.095
	(0.064)	(0.060)	(0.073)	(0.061)
Wife – Former Smoker	−0.020	0.042	−0.093	0.037
	(0.068)	(0.037)	(0.090)	(0.037)
Religious Participation	0.006**	0.000	0.004***	0.000
	(0.003)	(0.002)	(0.002)	(0.002)
Cigarette Price	0.020	0.041***	−0.013	0.044***
	(0.023)	(0.014)	(0.027)	(0.013)
Constant	1.368***	0.627**	1.468**	0.732***
	(0.452)	(0.271)	(0.590)	(0.272)
Sample Size	596	2,531	596	2,531

* Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level. Robust standard errors in parentheses. Power series in SLE includes constant, linear, interaction, and squared terms. Estimated propensity score is trimmed by 0.005 from the top and bottom.

Table 4: Characteristics for Type I and II Manipulation, by Sex[a,b]

	Male		Female	
	I	II	I	II
Education – College Degree	✓		✓	
Education – Some College, No College Degree				
Education – High School Diploma or GED		✓		✓
White-Collar Job	✓	✗	✓	✗
Job Covered by Union Contract	✗	✗	✗	✗
Health – Excellent				
Health – Very Good	✓	✓	✓	✓
Health – Good				
Drinker	✓	✓	✓	✓
Non-White Race	✗	✓	✗	✓
South Residence	✗	✓	✗	✓
Marital Status	✓	✓	✗	✗
Wife – Smoker	✗	✗	—	—
Wife – Former Smoker	✗	✗	—	—
Observations[c]	89	20	43	18

[a] Type I and II are the first two most typical observations according to the data.

[b] The fixed values for other continuous variables are assigned to the sample means in original Type I and II subgroups.

[c] If each characteristic is evenly distributed, we would have approximately 2 observations for male and 1 observation for female.

Table 5: General Characteristics, by Smoking Status (Female)[a]

	Smokers			Non-Smokers		
	Mean	Min	Max	Mean	Min	Max
Average Hourly Wage	14.75	1.31	52.86	19.16	2.33	225
Education – College Degree	0.23	0	1	0.44	0	1
Education – Some College, No College Degree	0.26	0	1	0.24	0	1
Education – High School Diploma or GED	0.41	0	1	0.26	0	1
No. Years Experience with Current Employer	5.85	0	30	7.61	0	40
No. Years Worked for Money Since 18 Years Old	12.33	0	40	12.875	0	41
White-Collar Job[b]	0.22	0	1	0.40	0	1
Job Covered by Union Contract	0.11	0	1	0.14	0	1
Health – Excellent	0.12	0	1	0.17	0	1
Health – Very Good	0.39	0	1	0.38	0	1
Health – Good	0.32	0	1	0.35	0	1
Drinker	0.75	0	1	0.64	0	1
No. Alcoholic Drinks Consumed Daily	2.04	0	15	1.29	0	12
Non-White Race	0.47	0	1	0.61	0	1
South Residence	0.46	0	1	0.53	0	1
Age	39.50	25	60	40.33	25	60
Marital Status	0.02	0	1	0.01	0	1
Frequency of Attendance at Religious Services	3.63	0	52	4.64	0	60
State Cigarette Price[c]	5.51	4.30	9.85	5.59	4.30	9.85
Sample Size		213			848	

[a] The sample includes female household heads between 25 and 60 years old, working at least 1,500 hours a year with positive average hourly wage.

[b] white-collar dummy equals one if individual's occupation code falls into management, professional, and related categories classified by 2000 Census of Population and Housing.

[c] We match the weighted average cigarette price per package (generic brands included) from the Tax Burden on Tobacco (Orzechowski and Walker, 2012) to PSID respondents by their state of residence.

Table 6: Average Smoking Effect on Wages (Female)

	OLS		IPW		WLS	
	Smokers	Non-Smokers	Smokers	Non-Smokers	Smokers	Non-Smokers
ATE	-0.113^{**}		-0.095		-0.108^{*}	
	(0.044)		(0.104)		(0.059)	
Unconditional Mean	2.635	2.748	2.642	2.737	2.636	2.744
	(0.043)	(0.020)	(0.107)	(0.026)	(0.058)	(0.019)

* Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level. Robust standard errors in parentheses. Power series in SLE includes constant, linear, interaction, and squared terms.

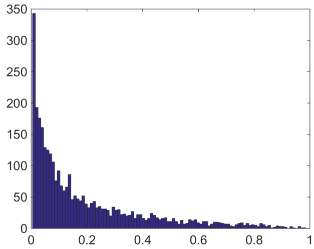
Table 7: OLS and WLS Estimates, by Smoking Status (Female)

	OLS		WLS	
	Smokers	Non-Smokers	Smokers	Non-Smokers
Education – College Degree	0.341**	0.540***	0.194	0.566***
	(0.136)	(0.075)	(0.193)	(0.066)
Education – Some College	0.166	0.304***	−0.008	0.322***
	(0.114)	(0.076)	(0.170)	(0.067)
Education – High School or GED	0.045	0.175**	−0.276*	0.201***
	(0.108)	(0.073)	(0.147)	(0.065)
Tenure	0.067***	0.033***	0.055***	0.035***
	(0.014)	(0.007)	(0.015)	(0.007)
Tenure Squared	−0.002***	−0.001**	−0.002***	−0.001***
	(0.001)	(0.000)	(0.000)	(0.000)
Experience	−0.018	−0.007	−0.037*	−0.006
	(0.015)	(0.007)	(0.022)	(0.007)
Experience Squared	0.001**	0.000	0.001**	0.000
	(0.000)	(0.000)	(0.001)	(0.000)
White-Collar Job	0.398***	0.219***	0.354***	0.213***
	(0.102)	(0.038)	(0.115)	(0.039)
Union	0.184**	0.119***	0.269***	0.131***
	(0.093)	(0.045)	(0.070)	(0.043)
Health – Excellent	0.115	0.189***	0.270*	0.147**
	(0.148)	(0.068)	(0.158)	(0.063)
Health – Very Good	0.177	0.141**	0.316*	0.085
	(0.125)	(0.061)	(0.164)	(0.056)
Health – Good	0.031	0.070	0.146	0.030
	(0.111)	(0.063)	(0.127)	(0.058)
Drinker	0.089	0.195***	0.131	0.204***
	(0.109)	(0.047)	(0.130)	(0.046)
Drinking Intensity	−0.011	−0.020	−0.003	−0.027**
	(0.015)	(0.015)	(0.025)	(0.012)
Race	−0.171*	−0.087**	−0.210**	−0.075*

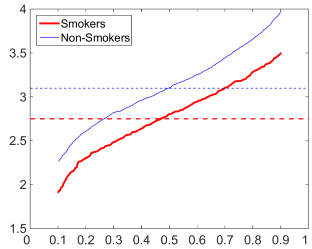
Table 7: OLS and WLS Estimates, by Smoking Status (Continued)

	OLS		WLS	
	Smokers	Non-Smokers	Smokers	Non-Smokers
	(0.089)	(0.037)	(0.105)	(0.040)
Region	−0.075	0.082**	0.088	0.099***
	(0.091)	(0.036)	(0.109)	(0.038)
Age	0.034	0.054***	0.061	0.048***
	(0.035)	(0.016)	(0.048)	(0.016)
Age Squared	0.000	−0.001***	−0.001	−0.001***
	(0.000)	(0.000)	(0.001)	(0.000)
Marital Status	0.046	0.159**	−0.091	0.160***
	(0.201)	(0.080)	(0.096)	(0.054)
Religious Participation	0.005*	0.000	0.008***	−0.001
	(0.003)	(0.002)	(0.002)	(0.002)
Cigarette Price	0.063**	0.070***	0.090***	0.076***
	(0.031)	(0.017)	(0.032)	(0.017)
Constant	1.009	0.532	0.579	0.607*
	(0.661)	(0.337)	(0.867)	(0.332)
Sample Size	213	848	213	848

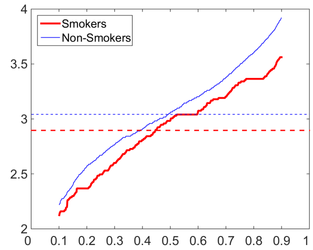
* Significant at 10% level. ** Significant at 5% level. *** Significant at 1% level. Robust standard errors in parentheses. Power series in SLE includes constant, linear, interaction, and squared terms. Estimated propensity score is trimmed by 0.005 from the top and bottom.



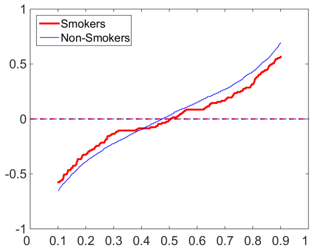
(a) Propensity Score



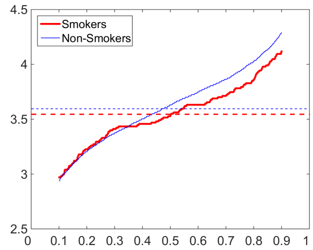
(b) Observed Wages



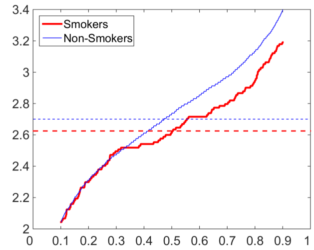
(c) Potential Wages



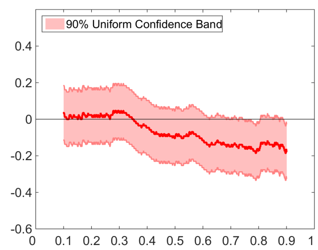
(d) Unobservables



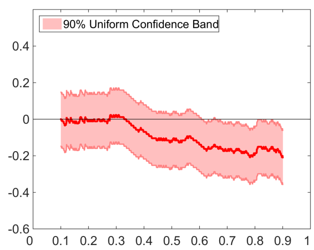
(e) QSFs for Type I



(f) QSFs for Type II

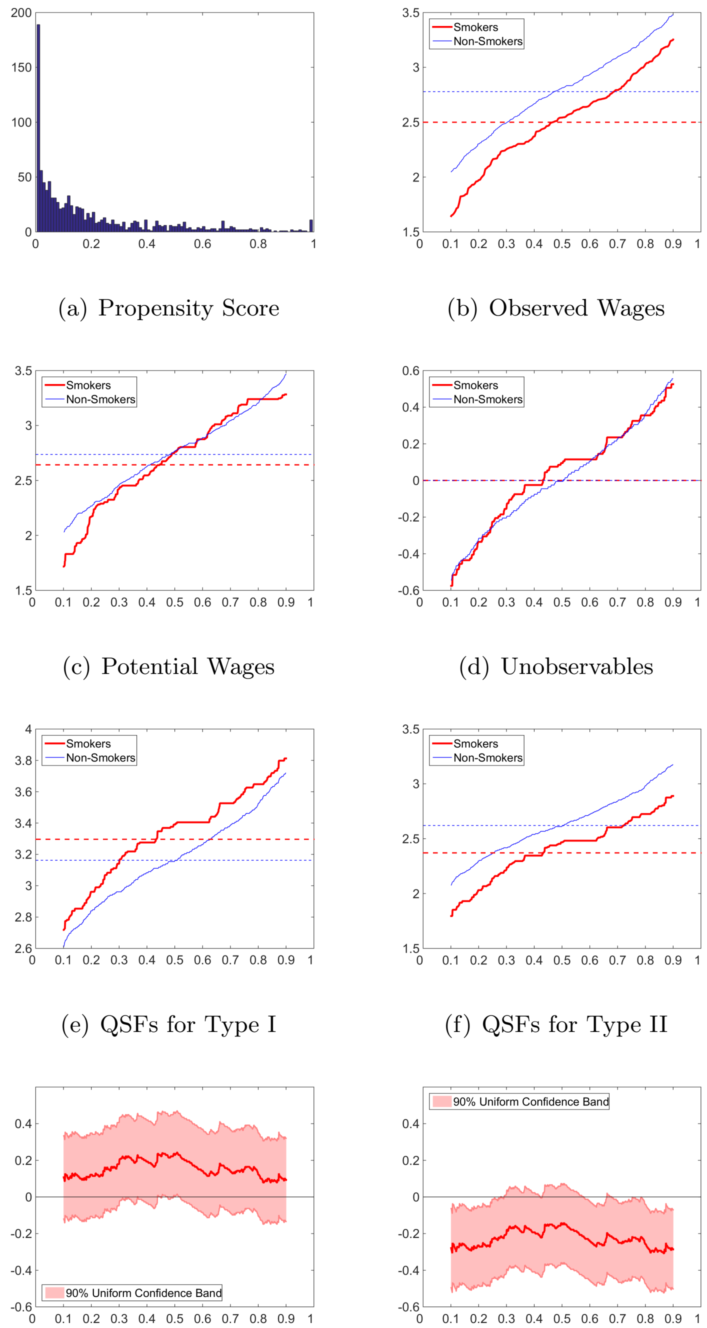


(g) SQTE for Type I



(h) SQTE for Type II

Figure 1: Propensity Score; Quantile Functions of Observed Wages, Potential Wages and Unobservables; QSFs; SQTEs for Males



(a) Propensity Score (b) Observed Wages

(c) Potential Wages (d) Unobservables

(e) QSFs for Type I (f) QSFs for Type II

(g) SQTE for Type I (h) SQTE for Type II

Figure 2: Propensity Score; Quantile Functions of Observed Wages, Potential Wages and Unobservables; QSFs; SQTEs for Females