Quantile Structural Treatment Effect: Application to Smoking Wage Penalty and its Determinants

Yu-Chin Hsu

Institute of Economics
Academia Sinica

Kamhon Kan

Institute of Economics
Academia Sinica

Tsung-Chih Lai*

Department of Economics
Feng Chia University

This version: March 1, 2018

* Corresponding author. E-mail: tclai@fcu.edu.tw

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

This paper proposes the quantile structural treatment effect (QSTE) to partial out different values of covariates from the quantile treatment effect which allows us to distinguish between observed and unobserved treatment heterogeneity. We show the QSTE is identified under the unconfoundedness assumption and propose an inverse probability weighted estimator which converges weakly to a Gaussian process at $\sqrt{n}$ rate. A multiplier bootstrap is proposed for uniform confidence bands. Using data from the Panel Study of Income Dynamics, we examine determinants of the wage penalty induced by smoking. Our findings suggest that the smoking wage gap is mainly driven by different levels of observable human capital for males.

JEL Classification: C14, C21, J31

Keywords: Treatment effects, structural functions, smoking wage penalty

1 Introduction

In the causal inference literature, various estimands are proposed to assess treatment effects under the assumption that selection to treatment is based on observable characteristics, e.g., the average treatment effect (ATE, Rosenbaum and Rubin, 1983; Hirano, Imbens, and Ridder, 2003), the quantile treatment effect (QTE, Firpo, 2007), and the inequality treatment effects (Firpo and Pinto, 2016).[1] In this literature, however, the covariates are usually used only to achieve the identification of treatment effect parameters and the researchers do not have a primary interest in how the effects relate to different values of covariates, especially from a distributional point of view.

In this paper, we propose to examine the quantile structural treatment effect (QSTE) to partial out different values of covariates from the QTE in semiparametric additive treatment effect models. That is, we specify the potential outcome as a (possibly nonlinear) parametrized structural function of covariates plus an unobserved error representing individual heterogeneity. The QSTE is the difference between the quantile structural functions (QSFs, Imbens and Newey, 2009) of the potential outcomes, where the QSF is the inversion of the distribution structural function (DSF, Chernozhukov, Fernandez-Val, Newey, Stouli, and Vella, 2017) defined as the distribution function of the potential outcome when covariates are exogenously shifted to a fixed value while keeping unobserved heterogeneity unchanged.

The QSTE allows us to answer a number of counterfactual experiments asking "by how much the treatment effect would change if the covariate was switched from its actual value to its average value?" (Fortin, Lemieux, and Firpo, 2011). This question is of particular interest if one aims to distinguish between observed and unobserved treatment effect heterogeneity. For instance, as will be illustrated in the empirical study, one possible causal explanation for the wage gap between smokers and non-smokers is that the smoking behavior may lead to impaired health and reduced productivity which in turn lower wages (Grafova and Stafford, 2009). On the other hand, the wage penalty can also be attributed to discrimination against smokers. To understand the extent to which observable human capital contributes to the smoking gap (while keeping other unobserved factors such as discrimination fixed), one can compare the QSTE with the QTE to disentangle the wage effects of smoking from different levels of human capital for all individuals.

To identify the object of interest, we first show that the DSF is equivalent to the distribution function of the error with a constant location shift whose magnitude equals to the structural function evaluated at the fixed point. We next show that the parameters of the structural function and the distribution function of the error are both identified under

[1]Please see Imbens and Wooldridge (2009) for a comprehensive review.

the unconfoundedness assumption. For estimation, we employ the weighted nonlinear least squares (WNLS) to estimate the parameters of the structural function and propose an inverse probability weighted (IPW) estimator for the distribution function of the error which is $\sqrt{n}$-consistent and converges weakly to a mean zero Gaussian process. The IPW-type estimators for the DSF, QSF, and QSTE can be constructed accordingly sharing the same asymptotic properties. We also propose a multiplier bootstrap to construct uniform confidence bands for uniform inference. A step-by-step implementation is also demonstrated in Appendix C.

For the empirical application, we revisit the issue of smoking wage penalty and explore its determinants using data from the Panel Study of Income Dynamics (PSID). In particular, we focus on those who started and continued smoking after entering the workforce for the plausibility of unconfoundedness.[2] Despite the well-established average wage differential in favor of non-smokers, this study is the first to investigate the wage effects of smoking beyond the mean by the QTE. Moreover, we examine the QSTE to see whether the wage gap exists after fixing observed heterogeneity. This allows us to examine the determinants directly and yield new insights into the issue. Our findings suggest that the smoking wage gap is mainly driven by different levels of observable human capital for males. For females, we only find a significant wage penalty at the upper tail of the wage distribution.

The QSTE considered in this paper can be viewed as the structural counterpart of the QTE proposed by Firpo (2007) with two substantial differences. First, we focus on the QSTE uniformly over a continuum of quantile indices which includes the pointwise QSTE as a special case. Second, the QSTE can be decomposed into the difference between structural functions and the difference between individual heterogeneity under the assumption of additive separability. This paper is also related to Donald and Hsu (2014) and Donald, Hsu, and Barrett (2012), where the former focus on the distribution and quantile functions of the potential outcomes and the latter focus on the conditional distribution and quantile functions. The DSF and QSF are equivalent to the distribution and quantile functions of the potential outcomes when covariate values are fixed at a point, respectively, and coincide to the conditional distribution and quantile functions under the conditional independence assumption as we will see in Section 2. The idea of structural functions has appeared in the literature (Blundell and Powell, 2003, 2004; Wooldridge, 2005; Imbens and Newey, 2009; Chernozhukov, Fernandez-Val, Newey, Stouli, and Vella, 2017). To the best of our knowledge, however, this paper is the first to incorporate structural functions into the potential outcome framework.[3]

[2] Please see Section 4 for more details.

[3] Note that the QSF is different from the quantile treatment response function proposed in Chernozhukov and Hansen (2005) where the latter is essentially the conditional quantile function of the

The remainder of this paper is organized as follows. In Section 2 we give a formal description of the model and the objects of interest. The identification and estimation results are also provided therein. Section 3 covers the asymptotic properties of the estimators and the multiplier bootstrap for conducting uniform inference. The empirical application is illustrated in Section 4. Section 5 concludes. All proofs are collected in Appendix D.

2 Model, Identification and Estimation

2.1 Model and Objects of Interest

Following the Rubin causal model (Rubin, 1974), let D be the binary treatment indicator such that $D = 1$ if the individual receives treatment and $D = 0$ otherwise. Define Y_1 as the potential outcome for the individual under treatment and Y_0 as that without treatment. Let X be a d_x-dimensional vector of pretreatment characteristics with a compact support $\mathcal{X} \subseteq \mathbb{R}^{d_x}$. What we can observe are X, D, and the actual outcome $Y = D\cdot Y_1 + (1-D)\cdot Y_0$. In this paper, we consider a semiparametric additive model of the potential outcomes

$$\begin{aligned} Y_1 &= m_1(X, \beta_1) + \epsilon_1, \quad E(\epsilon_1|X) = 0, \\ Y_0 &= m_0(X, \beta_0) + \epsilon_0, \quad E(\epsilon_0|X) = 0, \end{aligned} \tag{2.1}$$

where for $d = 0$ and 1, $m_d(x, b_d)$ represents the structural function of Y_d given $X = x$ that is known up to a finite-dimensional parameter vector b_d belonging to a compact support $\mathcal{B}_d \subseteq \mathbb{R}^{d_{b_d}}$. We denote β_d as the true parameter vector such that the conditional mean function $E(Y_d|X = x) = m_d(x, \beta_d)$ for all $x \in \mathcal{X}$. The error term ϵ_d plays the role of unobserved individual heterogeneity and is assumed to be conditional mean zero. Note that (2.1) nests linear regression as a special case when $m_d(X, \beta_d) = X\beta_d$, and is more general in that we allow for different functional forms of m_0 and m_1 as well as different dimensions of β_0 and β_1. One may extend (2.1) to be nonseparable in covariates and errors. The advantage of using separable models as in Brinch, Mogstad, and Wiswall (2017) is to obtain less complicated identification results and the final parameters of interest can be estimated at the parametric rate.[4] For notational simplicity, we hereafter use d for 0 and 1 when the arguments or discussions apply to both cases.

For a specific covariate value x, we follow Chernozhukov, Fernandez-Val, Newey, Stouli and Vella (2017) and Imbens and Newey (2009) to define the DSF and the QSF of the

potential outcomes.

[4]As pointed out by Brinch, Mogstad, and Wiswall (2017), the assumption of additive separability between observed and unobserved heterogeneity in treatment effects is implied by the additive separability between the treatment indicator and covariates in linear regression models.

potential outcome Y_d as

$$G_d(x,y) \equiv \int F_{Y_d|X,\epsilon_d}(y|x,e)F_{\epsilon_d}(\mathrm{de}), \qquad q_d(x,\tau) \equiv \inf\{y : G_d(x,y) \geq \tau\},$$

where F denotes the distribution function and $\tau \in [0,1]$. We also define the QSTE $\delta(x,\tau)$ as the difference between the QSFs,

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

Note that the DSF $G_d(x,y)$ can be interpreted as the distribution function of the counterfactual potential outcome when the population covariate values are exogenously switched to a fixed value x, whereas the unobserved individual heterogeneity ϵ_d of the whole population remains unchanged. The QSF $q_d(x,\tau)$ is the corresponding quantile function of $G_d(x,y)$. As mentioned in the introduction, the DSFs and QSFs of the potential outcomes allow us to distinguish the distributional and quantile treatment effects from different values of X. Moreover, the DSFs and QSFs can be used to test for stochastic dominance (Donald and Hsu, 2014) or Lorenz dominance (Barrett, Donald, and Bhattacharya, 2014) for a given x in the treatment effect models.

Under (2.1), the DSF and QSF can be simplified in the following lemma.

Lemma 2.1. *Suppose Y_d is generated according to* (2.1) *for $d = 0, 1$. Then*

$$G_d(x,y) = F_{\epsilon_d}\big(y - m_d(x,\beta_d)\big), \qquad q_d(x,\tau) = m_d(x,\beta_d) + Q_{\epsilon_d}(\tau),$$

where $F_{\epsilon_d}(\cdot)$ and $Q_{\epsilon_d}(\cdot)$ are the distribution and quantile functions of ϵ_d, respectively.

Lemma 2.1 shows that the DSF corresponds to the distribution function of ϵ_d with a horizontal shift of $m_d(x,\beta_d)$ and the QSF corresponds to the quantile function of ϵ_d with a vertical shift of $m_d(x,\beta_d)$. As a result, under (2.1) the QSTE can be expressed as $\delta(x,\tau) = [m_1(x,\beta_1) - m_0(x,\beta_0)] - [Q_{\epsilon_1}(\tau) - Q_{\epsilon_0}(\tau)]$ where the first part is the difference between observable structural functions and the second part is the difference between unobservable quantile functions of ϵ_d. It is also worth noting that the DSF is in general not equal to the conditional distribution function unless ϵ_d is independent of X. To see this, note that

$$\begin{aligned} F_{Y_d|X}(y|x) &= E_{\epsilon_d|X}\big[1\big\{m_d(X,\beta_d) + \epsilon_d \leq y\big\}\big|x\big] = E_{\epsilon_d|X}\big[1\big\{m_d(x,\beta_d) + \epsilon_d \leq y\big\}\big|x\big] \\ &= E_{\epsilon_d}\big[1\big\{m_d(x,\beta_d) + \epsilon_d \leq y\big\}\big] = E_{\epsilon_d}\big[1\big\{\epsilon_d \leq y - m_d(x,\beta_d)\big\}\big] = G_d(x,y), \end{aligned}$$

where $1\{\cdot\}$ denotes the indicator function and the third equality holds only under the independence assumption. In other words, the results regarding the DSF and QSF in this paper can be readily extended to the conditional distribution function and the conditional quantile function if ϵ_d and X are independent.

2.2 Identification

In this section, we discuss the identification of the DSF, QSF, and QSTE under the unconfoundedness assumption introduced by Rosenbaum and Rubin (1983). This assumption requires that treatment assignment be independent of the potential outcomes conditional on the observable covariates. It is also known as the selection on observables, conditional independence, or ignorability in the literature. In addition, the propensity score $p(x)$ which is the conditional probability of receiving treatment given $X = x$ is also assumed to be bounded away from 0 and 1 on $\mathcal{X}$. This requirement is called the common support or overlap condition since it is equivalent to assuming that treated and untreated supports are completely overlapped.

Assumption 2.1. *Suppose that*

(i) $D \perp (Y_0, Y_1)|X$.

(ii) $0 < \underline{p} \leq p(x) = P(D = 1|X = x) \leq \overline{p} < 1$ *for all* $x \in \mathcal{X}$.

Note that by the specification in (2.1), Assumption 2.1(i) is equivalent to $D \perp (\epsilon_0, \epsilon_1)|X$. To identify β_0 and β_1, we impose the following condition.

Assumption 2.2. *Suppose that*

(i) $E\big[Y_d\big|X = x\big] = m_d(x, \beta_d)$ *for some* $\beta_d \in \mathcal{B}_d$ *and for all* $x \in \mathcal{X}$.

(ii) $E\big[m_d(X, \beta_d) - m_d(X, b_d)\big]^2 > 0$ *for all* $b_d \in \mathcal{B}_d$*,* $b_d \neq \beta_d$.

Assumption 2.2(i) requires that the conditional mean of Y_d in (2.1) is correctly specified which is equivalent to assuming $E[\epsilon_d|X = x] = 0$ for all $x \in \mathcal{X}$. Assumption 2.2(ii) requires that β_d is the unique parameter value such that $E[Y_d - m_d(X, \beta_d)|X = x] = 0$ for all $x \in \mathcal{X}$.

Under Assumptions 2.1 and 2.2, we have the following identification results.

Lemma 2.2. *Suppose Assumptions 2.1 and 2.2 hold. Then for* $d = 0, 1$*,* β_d*,* $F_{\epsilon_d}(e)$*,*

$Q_{\epsilon_d}(\tau)$, $G_d(x,y)$, $q_d(x,\tau)$ and $\delta(x,\tau)$ are identified by

$$
\begin{aligned}
\beta_d &= \operatorname*{argmin}_{b_d \in \mathcal{B}_d} E\left[\frac{1\{D=d\}\cdot\big(Y-m_d(X,b_d)\big)^2}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right],\\
F_{\epsilon_d}(e) &= E\left[\frac{1\{D=d\}\cdot 1\big\{Y-m_d(X,\beta_d)\le e\big\}}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right],\\
Q_{\epsilon_d}(\tau) &= \inf\big\{e : F_{\epsilon_d}(e)\ge\tau\big\},\\
G_d(x,y) &= E\left[\frac{1\{D=d\}\cdot 1\big\{Y-m_d(X,\beta_d)\le y-m_d(x,\beta_d)\big\}}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right],\\
q_d(x,\tau) &= m_d(x,\beta_d)+Q_{\epsilon_d}(\tau),\\
\delta(x,\tau) &= q_1(x,\tau)-q_0(x,\tau).
\end{aligned}
$$

2.3 Estimation

Given a random sample $\{D_i, X_i, Y_i\}_{i=1}^n$ of size n, this paper proposes semiparametric estimators for $G_d(x,y)$, $q_d(x,\tau)$ and $\delta(x,\tau)$ that are later shown to converge at the parametric $\sqrt{n}$ rate. We first estimate β_d by the WNLS as in Wooldridge (2010),

$$
\hat\beta_d = \operatorname*{argmin}_{b_d\in\mathcal{B}_d}\sum_{i=1}^n \frac{1\{D_i=d\}\cdot\big(Y_i-m_d(X_i,b_d)\big)^2}{\big[\hat p(X_i)\big]^d\big[1-\hat p(X_i)\big]^{1-d}}, \tag{2.2}
$$

where $\hat p(x)$ is a nonparametric estimator for $p(x)$. Similar to 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.[5] Next, the (normalized) IPW estimators for $F_{\epsilon_d}(e)$ and $Q_{\epsilon_d}(\tau)$ are given by

$$
\begin{aligned}
\hat F_{\epsilon_d}(e) &= \sum_{i=1}^n \frac{1\{D_i=d\}\cdot 1\big\{Y_i-m_d(X_i,\hat\beta_d)\le e\big\}}{\big[\hat p(X_i)\big]^d\big[1-\hat p(X_i)\big]^{1-d}}\Bigg/\sum_{i=1}^n\frac{1\{D_i=d\}}{\big[\hat p(X_i)\big]^d\big[1-\hat p(X_i)\big]^{1-d}}, \\
\hat Q_{\epsilon_d}(\tau) &= \inf\{e:\hat F_{\epsilon_d}(e)\ge\tau\},
\end{aligned} \tag{2.3}
$$

[5] Although the WNLS is not the most efficient estimator among the class of estimators under the conditional moment restriction, the focus of this paper is the estimation and inference for the DSF, QSF and QSTE so we use WNLS here for its simplicity.

which can be used as the building blocks to estimate $G_d(x,y)$, $q_d(x,\tau)$ and $\delta(x,\tau)$:

$$\begin{aligned} \hat{G}_d(x,y) &= \hat{F}_{\epsilon_d}\big(y - m_d(x,\hat{\beta}_d)\big), \quad \hat{q}_d(x,\tau) = m_d(x,\hat{\beta}_d) + \hat{Q}_{\epsilon_d}(\tau), \\ \hat{\delta}(x,\tau) &= \hat{q}_1(x,\tau) - \hat{q}_0(x,\tau). \end{aligned} \tag{2.4}$$

3 Asymptotic Properties and Uniform Inference

This section consists of two parts. First, we investigate the asymptotic properties of all estimators under regularity conditions. Second, a multiplier bootstrap is proposed to approximate the limiting processes for uniform inference. To improve readability, we defer all regularity conditions and corresponding discussions to Appendix A.

3.1 Asymptotic Properties

The following lemma summarizes the asymptotic properties of $\hat{\beta}_d$ and $\hat{F}_{\epsilon_d}(e)$.

Lemma 3.1. *Suppose Assumptions 2.1, 2.2 and A.1–A.4 in Appendix A hold. Let $\hat{\beta}_d$ and $\hat{F}_{\epsilon_d}(e)$ be the estimators defined in* (2.2) *and* (2.3)*, respectively. Then for $W = (X,Y,D)$,*

$$\sqrt{n}(\hat{\beta}_d - \beta_d) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_{\beta_d}(W_i,\beta_d) + o_p(1),$$

where $\psi_{\beta_d}(W,\beta_d) =$

$$E\left[\frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\nabla m_d(X,\beta_d)'}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right]^{-1}\frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\big[Y - m_d(X,\beta_d)\big]}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}.$$

Moreover,

$$\sqrt{n}\big(\hat{F}_{\epsilon_d}(e) - F_{\epsilon_d}(e)\big) \Rightarrow \mathcal{F}_{\epsilon_d}(e),$$

where $\Rightarrow$ denotes the weak convergence and $\mathcal{F}_{\epsilon_d}(e)$ is a mean zero Gaussian process with the covariance kernel being generated by the influence function

$$\begin{aligned} \psi_{\epsilon_d}(W,e) = {} & \frac{1\{D=d\}\cdot 1\big\{Y - m_d(X,\beta_d)\le e\big\}}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}} - F_{\epsilon_d}(e) \\ & + \left\{1 - \frac{1\{D=d\}}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right\} F_{\epsilon_d|X}(e|X) \\ & + f_{\epsilon_d}(e)E\big[\nabla m_d(X,\beta_d)\big]'\psi_{\beta_d}(W,\beta_d), \end{aligned}$$

where $F_{\epsilon_d|X}(e|x)$ is the conditional distribution function of ϵ_d given $X = x$ and $f_{\epsilon_d}(e)$ is the density function of ϵ_d.

The first part of Lemma 3.1 shows that the estimation effect of $\hat{p}(x)$ on $\hat{\beta}_d$ is asymptotically negligible. In other words, the asymptotic variance of $\hat{\beta}_d$ is equal to the variance of $\psi_{\beta_d}(W, \beta_d)$ despite $\hat{p}(x)$ being a first-stage estimator. The second part of Lemma 3.1 gives the influence function representation of $\hat{F}_{\epsilon_d}(e)$. Note that the first term of $\psi_{\epsilon_d}(W, e)$ corresponds to the influence function if the true propensity score $p(x)$ were known, the second term is the contribution of estimating $p(x)$ to the asymptotic process of $\hat{F}_{\epsilon_d}(e)$, and the last term comes from the estimation error of $\hat{\beta}_d$.

By (2.4) and Lemma 3.1, the asymptotic properties of $\hat{G}_d(x, y)$, $\hat{q}_d(x, \tau)$ and $\hat{\delta}(x, \tau)$ are given in the following theorem.

Theorem 3.1. *Suppose Assumptions 2.1, 2.2 and A.1–A.4 in Appendix A hold. Let $\hat{G}_d(x, y)$, $\hat{q}_d(x, \tau)$ and $\hat{\delta}(x, \tau)$ be the estimators defined in (2.4), then*

$$\sqrt{n}\big(\hat{G}_d(x, y) - G_d(x, y)\big) \Rightarrow \mathcal{G}_d(x, y), \qquad \sqrt{n}\big(\hat{q}_d(x, \tau) - q_d(x, \tau)\big) \Rightarrow \mathcal{Q}_d(x, \tau),$$
$$\sqrt{n}\big(\hat{\delta}(x, \tau) - \delta(x, \tau)\big) \Rightarrow \Delta(x, \tau),$$

where $\mathcal{G}_d(x, y)$ is a mean zero Gaussian process with the covariance kernel generated by the influence function

$$\psi_{G_d}(W, x, y) = \psi_{\epsilon_d}\big(W, y - m_d(x, \beta_d)\big) - f_{\epsilon_d}\big(y - m_d(x, \beta_d)\big)\nabla m_d(x, \beta_d)'\psi_{\beta_d}(W, \beta_d),$$

where $\psi_{\epsilon_d}(W, e)$ and $\psi_{\beta_d}(W, \beta_d)$ are defined in Lemma 3.1. Moreover, $\mathcal{Q}_d(x, \tau)$ is also a mean zero Gaussian process with the covariance kernel being generated by the influence function

$$\psi_{Q_d}(W, x, \tau) = -\frac{\psi_{G_d}\big(W, x, q_d(x, \tau)\big)}{g_d\big(x, q_d(x, \tau)\big)},$$

where $g_d(x, y) = f_{\epsilon_d}\big(y - m_d(x, \beta_d)\big)$ is the corresponding density function of $G_d(x, y)$. Finally, $\Delta(x, \tau) = \mathcal{Q}_1(x, \tau) - \mathcal{Q}_0(x, \tau)$.

We have several remarks on Theorem 3.1. First, from (2.4) it is not surprising that the asymptotic of $\hat{G}_d(x, y)$ has a similar influence function representation to that of $\hat{F}_{\epsilon_d}(e)$ in Lemma 3.1. Note, however, that the estimation error attributed to $\hat{\beta}_d$ affects two parts: (i) the distribution function itself depends on $\hat{\beta}_d$ and (ii) the point we evaluate is estimated by $y - m_d(x, \hat{\beta}_d)$. This gives the additional term in $\psi_{G_d}(W, x, y)$. Next, given that the quantile map is Hadamard differentiable, the limiting process of $\hat{q}_d(x, \tau)$ follows from the functional delta method.

3.2 Uniform Inference

We introduce a multiplier bootstrap for uniform inference on the DSF, QSF, and QSTE. Although the asymptotic properties shown above can be used for pointwise inference provided that the influence functions are consistently estimated, conducting uniform inference is not an easy task even if one adopts the nonparametric bootstrap – it may be time-consuming due to the nonparametric estimation of $p(x)$. Hence, we suggest an alternative simulation-based method known as the multiplier bootstrap to approximate the limiting processes $\mathcal{G}_d(x, y)$ and $\mathcal{Q}_d(x, \tau)$. Note that in doing so the estimators for $F_{\epsilon_d|X}(\cdot|x)$ and $f_{\epsilon_d}(\cdot)$ must be uniformly consistent so that the estimation errors will disappear in the limit. We provide such estimators in Appendix B. The explicit forms of the estimated pointwise influence functions $\hat{\psi}_{G_d}(W_i, x, y)$ and $\hat{\psi}_{Q_d}(W_i, x, \tau)$ can also be found in Appendix B.

Given consistently estimated influence functions, let $\{U_1, U_2, \ldots\}$ be i.i.d. pseudo random variables with mean 0 and variance 1 (e.g., standard normal random variables) that are independent of the whole sample. The simulated processes for $\mathcal{G}_d(x, y)$, $\mathcal{Q}_d(x, \tau)$ and $\Delta(x, \tau)$ are given by

$$\begin{aligned} &\mathcal{G}_d^u(x, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \cdot \hat{\psi}_{G_d}(W_i, x, y), \quad \mathcal{Q}_d^u(x, \tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \cdot \hat{\psi}_{Q_d}(W_i, x, \tau), \\ &\Delta^u(x, \tau) = \mathcal{Q}_1^u(x, \tau) - \mathcal{Q}_0^u(x, \tau). \end{aligned} \tag{3.1}$$

Theorem 3.2. *Suppose the assumptions in Theorem 3.1 hold. Also suppose that $\hat{F}_{\epsilon_d|X}(e|x)$ and $\hat{f}_{\epsilon_d}(e)$ are uniformly consistent estimators for $F_{\epsilon_d|X}(e|x)$ and $f_{\epsilon_d}(e)$, respectively. Then*

$$\mathcal{G}_d^u(x, y) \overset{p}{\Rightarrow} \mathcal{G}_d(x, y), \qquad \mathcal{Q}_d^u(x, \tau) \overset{p}{\Rightarrow} \mathcal{Q}_d(x, y), \qquad \Delta^u(x, \tau) \overset{p}{\Rightarrow} \Delta(x, \tau),$$

where $\overset{p}{\Rightarrow}$ denotes the weak convergence conditional on sample path with probability approaching 1.

We next show how to construct one-sided and two-sided uniform confidence bands for the QSTE. For a confidence level α and for $[\eta, 1 - \eta]$ where $\eta \in [0, 1/2)$, let $\hat{C}_{\alpha,\eta}^{\text{1-sided}}$ and $\hat{C}_{\alpha,\eta}^{\text{2-sided}}$ be the one- and two-sided critical values for $\{\delta(x, \tau) : \tau \in [\eta, 1 - \eta]\}$ that satisfy

$$\begin{aligned} \hat{C}_{\alpha,\eta}^{\text{1-sided}} &= \inf_{y \in \mathbb{R}} \left\{ \Pr\left(\sup_{\tau \in [\eta, 1-\eta]} \frac{\Delta^u(x, \tau)}{\hat{\sigma}(x, \tau)} \le y \right) \ge \alpha \right\}, \\ \hat{C}_{\alpha,\eta}^{\text{2-sided}} &= \inf_{y \in \mathbb{R}} \left\{ \Pr\left(\sup_{\tau \in [\eta, 1-\eta]} \frac{\left|\Delta^u(x, \tau)\right|}{\hat{\sigma}(x, \tau)} \le y \right) \ge \alpha \right\}, \end{aligned}$$

where

$$\hat{\sigma}(x,\tau)=\left\{\frac{1}{n}\sum_{i=1}^{n}\Big[\hat{\psi}_{Q_1}(W_i,x,\tau)-\hat{\psi}_{Q_0}(W_i,x,\tau)\Big]^2\right\}^{1/2}$$

is the estimated standard error of the QSTE. That is, $\hat{C}^{\text{1-sided}}_{\alpha,\eta}$ and $\hat{C}^{\text{2-sided}}_{\alpha,\eta}$ are the α-th quantiles of $\Delta^u(x,\tau)/\hat{\sigma}(x,\tau)$ and $\left|\Delta^u(x,\tau)\right|/\hat{\sigma}(x,\tau)$, respectively. The lower and upper one-sided $(1-\alpha)$ uniform confidence bands for $\tau\in[\eta,1-\eta]$ are then given by

$$\left(\hat{\delta}(x,\tau)-\hat{C}^{\text{1-sided}}_{\alpha,\eta}\frac{\hat{\sigma}(x,\tau)}{\sqrt{n}},\ \infty\right)\quad\text{and}\quad\left(-\infty,\ \hat{\delta}(x,\tau)+\hat{C}^{\text{1-sided}}_{\alpha,\eta}\frac{\hat{\sigma}(x,\tau)}{\sqrt{n}}\right),\tag{3.2}$$

and the two-sided $(1-\alpha)$ uniform confidence band for the SQTE is

$$\left(\hat{\delta}(x,\tau)-\hat{C}^{\text{2-sided}}_{\alpha,\eta}\frac{\hat{\sigma}(x,\tau)}{\sqrt{n}},\ \hat{\delta}(x,\tau)+\hat{C}^{\text{2-sided}}_{\alpha,\eta}\frac{\hat{\sigma}(x,\tau)}{\sqrt{n}}\right).\tag{3.3}$$

4 Empirical Study

4.1 Literature Review

As an empirical application, we explore the possible nature of the negative wage effects of smoking. It is well established in the literature that smokers on average earn 4–24% less than non-smokers (see, e.g., Levine, Gustafson, and Velenchik, 1997; van Ours, 2004; Auld, 2005; Grafova and Stafford, 2009). According to Grafova and Stafford (2009), possible causal explanations can be broadly classified into observable and unobservable ones: For the former, the adverse effects of smoking on health and productivity are the main causes for low wages. For the latter, discrimination against smokers could be one of the unobserved factors that deteriorate labor market performance for smokers. Nevertheless, there are also non-causal explanations suggesting that the smoking behavior is not the cause, but rather an indicator of preferences that are associated with fewer investments in human capital and thus lower wages.

To examine these potential explanations, Levine, Gustafson, and Velenchick (1997) employ US data to find a wage differential of 4–8% using OLS and difference-in-differences. However, the investigation on the potential explanations cannot be teased out without accounting for heterogeneous effects. In recognition of the heterogeneity of smokers, Grafova and Stafford (2009) construct a retrospective sample by smoking history and find negative wage effects of 8–12% for persistent smokers versus never or former smokers. Their findings suggest that the wage penalty may be driven by a non-causal explanation rather than by smoking per se, indicating the presence of self-selection into smoking. Taking self-selection into account, van Ours (2004) and Auld (2005), using Dutch and Canadian data respectively, consider the instrumental variable estimation. While a 10% wage gap is found in van Ours (2004), Auld (2005) reports that smoking reduces wages by 8–24% before and after correcting for endogeneity.

To sum up, three issues should be addressed in exploring determinants of the wage gap induced by smoking: (i) heterogeneous effects; (ii) observed and unobserved factors; (iii) self-selection problem (validity of the unconfoundedness assumption). Heeding these observations, we first consider the QTE to uncover heterogeneous effects of smoking on wages. Next, we utilize the QSTE to fix observable human capital measures such as education and occupation at the same values for all individuals. The QTE and QSTE allow us to study the mechanism of smoking wage gap: If the wage penalty is indeed driven by different levels of observable human capital, the QTE would be significant while the QSTE would be insignificant since the human capital measures are now fixed.

Before describing the data in detail, it is important to note that in this study we focus on the comparison between late smokers (those who started and continued smoking after entering the workforce) and non-smokers (those who did not smoke at the time of interview). That is, we exclude individuals who started smoking before entering the labor market since early smoking may be confounded with observable characteristics such as education (Zhao, Konishi and Glewwe, 2012) and occupation (Viscusi and Hersch, 2001). The plausibility of unconfoundedness will also be discussed in the next section.

4.2 Data

The data we use are from the 2015 wave of the Panel Study of Income Dynamics (PSID). Started in 1968, the PSID is the longest-running household panel survey collecting rich information on the economic and demographic status of a household and its head (the respondent). The sample size increases from 6,800 in the first wave to over 9,000 in the 2015 wave. In particular, the PSID collects data on a respondent's smoking behavior in the 1986 wave and every wave since the 1999's. This panel structure enables us to track smoking history of the respondents. For example, in the 2015 wave about 41% (2,713 out of 6,588) of household heads reported that they had ever smoked cigarettes. Among the ever smokers, 1,399 of them started smoking after entering the workforce and 1,214 of them continued smoking at the time of interview.

The outcome variable is the log of average hourly wage rate. The covariates include race, education (highest grade completed), occupation (white-collar or not), union status, and the state cigarette price. We set the white-collar dummy equal to one if the respondent's occupation code belonged to management, professional, and related categories as classified by the 2000 Census of Population and Housing. For the state cigarette price, we match the average cost per pack of cigarettes from the Tax Burden on Tobacco (Orzechowski and Walker, 2017) to PSID respondents' state of residence and for the year when they started smoking for smokers or the year they were 18 for non-smokers.[6] Similar to Grafova and Stafford (2009), we restrict the male sample to household heads between the ages of 20 and 60 who worked at least 1,000 hours a year, and the female sample to household heads between the ages of 20 and 65 who worked at least 1,000 hours annually. All individual characteristics are summarized in Table 1.

To claim the validity of the unconfoundedness assumption, we utilize the test proposed by Donald, Hsu, and Lieli (2014) and follow van Ours (2004) to use the ever smoking status as a binary instrument for smoking that satisfies one-sided non-compliance. As can be seen in Table 2, a number of implementations of the SLE are considered to estimate propensity scores. Specifically, we start with a constant model and then add linear, interaction, and quadratic terms in the power series.[7] It follows that the unconfoundedness assumption cannot be rejected at 10% significance level for most cases except for the constant model where no covariate is included. Put differently, these results suggest that (i) the smoking behavior is not randomly assigned and (ii) the selection bias can be explained by observable characteristics. We interpret this as a statistical evidence to rule out the selection on unobservables for our empirical application.

4.3 Main Results

First, we estimate the ATE of smoking on wages based on the linear specification of the potential outcome

$$Y_d = X\beta_d + \epsilon_d.$$

Table 3 compares the WNLS estimate with estimates from the regression adjustment (RA), inverse probability weighted (IPW), and inverse probability weighted regression adjustment (IPWRA) methods.[8] It can be seen that a significant wage differential of 13–24% is found for males which is consistent with the literature albeit we focus on late smokers with a shorter smoking history. Table 3 also shows that smoking does no harm (at least on average) to females' wages similar to the findings in van Ours (2004).

Secondly, the QTE results are summarized in Figure 1. Following Donald and Hsu (2014), we estimate the quantile functions of the potential wages (and the corresponding

[6]The cigarette price is deflated by the CPI-U (2015-based).

[7]The 'pure quadratic' model includes constant, linear, and quadratic terms as the power series.

[8]The specifications we use are linear, logit, and linear-logit models for RA, IPW, and IPWRA, respectively. For WNLS, we use constant, linear, interaction and quadratic terms as the power series to estimate the propensity score by the SLE. The estimated propensity score is trimmed to the interval [0.005, 0.995] following Abrevaya, Hsu, and Lieli (2015).

QTE) by the IPW using the estimated propensity score. As shown in Figure 1(a), the male wage gap between potential smokers and non-smokers is uniformly narrower than the actual one for all quantiles. On the other hand, Figure 1(b) reveals that the smoking wage penalty for females only exists in the middle-upper quantiles. These findings are reinforced by the 90% uniform confidence bands depicted in Figures 1(c) and 1(d). In Figure 1(c), we show that male smokers earn significantly less than non-smokers at most quantiles after correcting for selection bias. For females, we only find a significant wage penalty at the upper tail of the wage distribution, as displayed in Figure 1(d).

Our final set of results, which is the main contribution of this study, is to investigate potential explanations for the smoking wage differential via the QSTE in Figure 2. Note that the exposition of the fixed values for observable characteristics can be found in Table 4. Specifically, we assign the fixed values for binary variables to the most frequently occurring types of individuals according to the data; for continuous variables, the fixed values are assigned to the group means of the most typical subgroup.[9] Figures 2(a) and 2(b) depict the QSFs for males and females, respectively. It follows that the uniform confidence bands for potential smokers are much wider than those for potential non-smokers, reflecting the fact that the observations of smokers is relatively small in our sample. Figures 2(c) displays the QSTE with 90% uniform confidence band as well as the QTE estimate for the male sample. Surprisingly, there is no significant wage differential at the two ends of the unobserved heterogeneity distribution once observable characteristics are fixed. In other words, the results demonstrated in Figures 1(c) and 2(c) provide supportive evidence that the wage gap induced by smoking is mainly driven by different levels of observable human capital for males. This result is new to the literature and is robust to different sets of fixed values for observable characteristics.

5 Conclusion

This paper provides the identification, estimation, and inference for the QSTE in semi-parametric additive treatment effect models. The QSTE allows us to distinguish the QTE from different values of observable characteristics while keeping the unobserved factors unchanged. By comparing the QTE and QSTE, we investigate potential causal explanations for the smoking wage penalty. Our findings suggest that the smoking wage gap for male workers at the two ends of individual heterogeneity distribution can be attributed to different levels of observable characteristics. However, no significant wage differential is found for females in the lower and middle quantiles.

[9] We also try different sets of fixed values for continuous variables such as the overall means or the group means from the k-mean clustering for $k = 2$ (the fixed values for binary variables are adjusted accordingly). The results are similar and available upon request.

APPENDIX

A Regularity Conditions

We impose the following regularity conditions for the asymptotic results in this paper. Let $\nabla m_d(X, b_d)$ denote the $d_{b_d} \times 1$ gradient of $m_d(X, b_d)$ with respect to b_d and $\nabla^2 m_d(X, b_d)$ denote the $d_{b_d} \times d_{b_d}$ Hessian of $m_d(X, b_d)$. Let $\|\cdot\|$ denote the Euclidean norm of a matrix.

Assumption A.1 (Support of X)**.**

(i) The support of X is a Cartesian product of compact intervals, $\mathcal{X} = \prod_{i=1}^{d_x} [x_{i\ell}, x_{iu}]$.

(ii) The density function of X is bounded away from 0 on $\mathcal{X}$.

Assumption A.2 (Structural Functions)**.**

(i) β_d is in the interior of $\mathcal{B}_d$ which is a compact subset of $\mathbb{R}^{d_{b_d}}$.

(ii) For each $b_d \in \mathcal{B}_d$, $m_d(\cdot, b_d)$ is Borel measurable on $\mathcal{X}$.

(iii) For each $x \in \mathcal{X}$, $m_d(x, \cdot)$ is continuously differentiable of order 2 in $b_d \in \mathcal{B}_d$.

(iv) $E\Big[\sup_{b_d \in \mathcal{B}_d} \big\|\nabla m_d(X, b_d)\big\|^2\Big] < \infty$ *and* $E\Big[\sup_{b_d \in \mathcal{B}_d} \big\|\nabla^2 m_d(X, b_d)\big\|^2\Big] < \infty$.

(v) $E\big[\nabla m_d(X, \beta_d) \nabla m_d(X, \beta_d)'\big]$ *is positive definite.*

Assumption A.3 (Unobservables)**.**

(i) ϵ_d has a compact support $[e_{d\ell}, e_{du}]$ and denote $\mathcal{E} = [e_\ell, e_u]$ where $e_\ell = \min\{e_{0\ell}, e_{1\ell}\}$ and $e_u = \max\{e_{0u}, e_{1u}\}$.

(ii) The density function $f_{\epsilon_d}(e)$ is bounded away from 0 and continuously differentiable of order 2 on $[e_{d\ell}, e_{du}]$.

Assumption A.4 (SLE)**.**

(i) $p(x)$ is continuously differentiable of order $s \geq 7d_x$.

(ii) The SLE $\hat{p}(x)$ uses a power series with order $a \cdot n^\nu$ for some $a > 0$ and $d_x/4(s - d_x) < \nu < 1/9$.

Assumption A.2 consists of standard assumptions for asymptotic properties of the WNLS, see Wooldridge (2010) for details. Assumption A.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. However, if

$\mathrm{Var}(\epsilon_0) = \mathrm{Var}(\epsilon_1) = \infty$, we can still estimate $Q_{\epsilon_0}(\tau)$ and $Q_{\epsilon_1}(\tau)$ for $\tau \in [t, 1-t]$, $0 < t < 1/2$ on which the density functions of ϵ_0 and ϵ_1 are bounded away from 0. Assumption A.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. Assumption A.4(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). Assumption A.4(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 $p(x)$.

B Estimated Influence Functions

The plug-in estimators for the influence functions in Theorem 3.1 are given by

$$
\begin{aligned}
\hat{\psi}_{G_d}(W_i, x, y) = {} & \frac{1\{D_i = d\} \cdot 1\{Y_i - m_d(X_i, \hat{\beta}_d) \le y - m_d(x, \hat{\beta}_d)\}}{[\hat{p}(X_i)]^d [1 - \hat{p}(X_i)]^{1-d}} - \hat{F}_{\epsilon_d}(y - m_d(x, \hat{\beta}_d)) \\
& + \left\{ 1 - \frac{1\{D_i = d\}}{[\hat{p}(X_i)]^d [1 - \hat{p}(X_i)]^{1-d}} \right\} \hat{F}_{\epsilon_d|X}(y - m_d(x, \hat{\beta}_d) | X_i) \\
& + \hat{f}_{\epsilon_d}(y - m_d(x, \hat{\beta}_d)) \left[\nabla_{b_d} \overline{m}_d(X_i, \hat{\beta}_d) - \nabla_{b_d} m_d(x, \hat{\beta}_d) \right]' \hat{\psi}_{\beta_d}(W_i, \hat{\beta}_d), \\
\hat{\psi}_{Q_d}(W_i, x, \tau) = {} & -\frac{\hat{\psi}_{G_d}(W_i, x, \hat{q}_d(x, \tau))}{\hat{g}_d(x, \hat{q}_d(x, \tau))},
\end{aligned}
$$

where $\hat{F}_{\epsilon_d|X}(\cdot|x)$ and $\hat{f}_{\epsilon_d}(\cdot)$ are uniformly consistent estimators given below,

$$
\hat{\psi}_{\beta_d}(W_i, \hat{\beta}_d) =
$$

$$
\left[\frac{1}{n} \sum_{i=1}^{n} \frac{1\{D_i = d\} \cdot \nabla m_d(X_i, \hat{\beta}_d) \nabla m_d(X_i, \hat{\beta}_d)'}{[\hat{p}(X_i)]^d [1 - \hat{p}(X_i)]^{1-d}} \right]^{-1} \frac{1\{D_i = d\} \cdot \nabla m_d(X_i, \hat{\beta}_d)(Y_i - m_d(X_i, \hat{\beta}_d))}{[\hat{p}(X_i)]^d [1 - \hat{p}(X_i)]^{1-d}},
$$

and $\overline{m}_d(X_i, \hat{\beta}_d) = n^{-1} \sum_{i=1}^{n} m_d(X_i, \hat{\beta}_d)$.

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

We construct a series-based estimator for $F_{\epsilon_d|X}(e|x)$ similar to that in Donald and Hsu (2014). Note that the estimator must satisfy the following requirements: (i) bounded between 0 and 1; (ii) monotonically increasing in e for any given x; (iii) converging in probability to $F_{\epsilon_d}(e|x)$ uniformly in both e and x. The estimator plays an important role in the multiplier bootstrap method.

Let $\tilde{F}_{\epsilon_d|X}(e|x)$ be the series estimator for $F_{\epsilon_d|X}(e|x)$,

$$
\tilde{F}_{\epsilon_d|X}(e|x) = \left\{ \sum_{i=1}^{n} \frac{1\{D_i = d\} \cdot 1\{Y_i - m_d(X_i, \hat{\beta}_d) \le e\}}{[\hat{p}(X_i)]^d [1 - \hat{p}(X_i)]^{1-d}} R^K(X_i) \right\}' \left\{ \sum_{i=1}^{n} R^K(X_i) R^K(X_i)' \right\}^{-1} R^K(x),
$$

where $R^K(x)$ is the same power series used in the SLE estimator for $p(x)$. $\tilde{F}_{\epsilon_d|X}(e|x)$ is a step function in e with jumps at $Y_i - m_d(X_i, \hat{\beta}_d)$'s for any given x, and will converge in probability to $F_{\epsilon_d|X}(e|x)$ in both arguments e and x. However, $\tilde{F}_{\epsilon_d|X}(e|x)$ is not necessarily bounded between 0 and 1 nor monotonically increasing in e for any given x. Hence we introduce a modified version as follows.

Let $\varepsilon_i = Y_i - m_d(X_i, \hat{\beta}_d)$. Without loss of generality 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 $\hat{F}_{\epsilon_d|X}(e|x)$ by induction: First define $\hat{F}_{\epsilon_d|X}(e|x) = \tilde{F}_{\epsilon_d|X}(e|x) = 0$ for $\varepsilon_{(0)} \le e < \varepsilon_{(1)}$ and

$\hat{F}_{\epsilon_d|X}(\varepsilon_{(n+1)}|x) = 1$. Next, suppose $\hat{F}_{\epsilon_d|X}(e|x) = 0$ is already defined for $\varepsilon_{(0)} \leq e < \varepsilon_{(i)}$, we then define for $\varepsilon_{(i)} \leq e < \varepsilon_{(i+1)}$,

$$\begin{aligned}\hat{F}_{\epsilon_d|X}(e|x) = \hat{F}_{\epsilon_d|X}(\varepsilon_{(i-1)}|x) \cdot 1\big\{0 \leq \tilde{F}_{\epsilon_d|X}(\varepsilon_{(i)}|x) \leq \hat{F}_{\epsilon_d|X}(\varepsilon_{(i-1)}|x)\big\} \\ + \tilde{F}_{\epsilon_d|X}(\varepsilon_{(i)}|x) \cdot 1\big\{\hat{F}_{\epsilon_d|X}(\varepsilon_{(i-1)}|x) < \tilde{F}_{\epsilon_d|X}(\varepsilon_{(i)}|x) \leq 1\big\} + 1\big\{\tilde{F}_{\epsilon_d|X}(\varepsilon_{(i)}|x) > 1\big\}.\end{aligned}$$

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

Lemma B.1. *Suppose Assumptions 2.1, 2.2, A.1–A.4 hold. Then for any given x, $\hat{F}_{\epsilon_d|X}(e|x)$ is bounded between 0 and 1, monotonically increasing in e, and*

$$\sup_{e\in\mathcal{E},x\in\mathcal{X}} \left|\hat{F}_{\epsilon_d|X}(e|x) - F_{\epsilon_d|X}(e|x)\right| = o_p(1).$$

Lemma B.1 follows from $\sup_{e\in\mathcal{E},x\in\mathcal{X}} |\tilde{F}_{\epsilon_d|X}(e|x) - F_{\epsilon_d|X}(e|x)| = o_p(1)$ and $\sup_{x\in\mathcal{X}} |\hat{F}_{\epsilon_d|X}(e|x) - F_{\epsilon_d|X}(e|x)| \leq \sup_{x\in\mathcal{X}} |\tilde{F}_{\epsilon_d|X}(e|x) - F_{\epsilon_d|X}(e|x)| = o_p(1)$ for all $x \in \mathcal{X}$. Note that the compactness of $\mathcal{X}$ in Assumption A.1 is needed to obtain the uniform result. One can also use the kernel estimator to estimate $F_{\epsilon_d|X}(e|x)$ instead of the series estimator. The multiplier bootstrap remains valid provided the kernel estimator has the properties in Lemma B.1.

B.2 Kernel-Based Estimator for $f_{\epsilon_d}(e)$

In addition to $F_{\epsilon_d|X}(e|x)$, we need to estimate $f_{\epsilon_d}(e)$ before approximating the limiting process $\mathcal{G}_d(x, y)$. We introduce the IPW kernel estimator for $f_{\epsilon_d}(e)$ which is uniformly consistent over $\mathcal{E}$. Let $h = h_n$ 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 $\tilde{f}_{\epsilon_d}(e)$ as

$$\tilde{f}_{\epsilon_d}(e) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1\{D_i = d\}}{\big[\hat{p}(X_i)\big]^d \big[1 - \hat{p}(X_i)\big]^{1-d}} \mathcal{K}\left(\frac{Y_i - m_d(X_i, \hat{\beta}_d) - e}{h}\right).$$

The estimator for $f_{\epsilon_d}(e)$ is given by

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

The reason to use $\hat{f}_{\epsilon_d}(e)$ instead of $\tilde{f}_{\epsilon_d}(e)$ is because $\tilde{f}_{\epsilon_d}(e)$ is in general inconsistent around the boundary point e_ℓ. Therefore, we modify $\tilde{f}_{\epsilon_d}(e)$ around the boundary point to obtain uniform consistency. This method is also used in Donald, Hsu, and Barrett (2012) and Donald and Hsu

(2014). We make the following assumptions on $\mathcal{K}$ and h.

Assumption B.1 (Kernel).

(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 B.2. *Suppose Assumptions 2.1, 2.2, A.1–A.4 and B.1 hold. Then*

$$\sup_{e \in \mathcal{E}} \left| \hat{f}_{\epsilon_d}(e) - f_{\epsilon_d}(e) \right| = o_p(1).$$

C Implementation of Uniform Confidence Bands

Here we provide a step-by-step implementation of uniform confidence bands for $\hat{\delta}(x,\tau)$.

1. Given a power series of X_i satisfying Assumption A.4(ii), define the SLE estimator $\hat{p}(X_i)$ as the fitted value of the logit regression of D_i on the power series.

2. The WNLS estimator $\hat{\beta}_d$ can be constructed as in (2.2). Specifically, if one imposes the linear specification of $m_d(X,\beta_d)$, the closed-form expression of $\hat{\beta}_d$ is given by

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

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

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

3. Plug $\hat{p}(X_i)$ and $\hat{\beta}_d$ in (2.3). Given a fixed value x which is of the same dimension of X, the DSF, QSF, and QSTE can also be estimated according to (2.4).

4. The influence functions in Theorem 3.1 can be estimated given $\hat{F}_{\epsilon_d|X}(\cdot|x)$ and $\hat{f}_{\epsilon_d}(\cdot)$ in Appendix B. Note that we use the same power series as in step 1 for $\hat{F}_{\epsilon_d|X}(\cdot|x)$ and adopt the Gaussian kernel with bandwidth $h = 1.06 \cdot \hat{\sigma}_d n^{-1/5}$ for $\hat{f}_{\epsilon_d}(\cdot)$, where $\hat{\sigma}_d$ is the sample standard deviation of $\hat{\epsilon}_{di} = Y_i - X_i\hat{\beta}_d$.

5. 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 $b = 1,\ldots,B$, calculate the simulated process $\Delta_b^u(x,\tau)$ according to (3.1) and store the maximum (absolute) value of $\Delta_b^u(x,\tau)/\hat{\sigma}(x,\tau)$ among $\tau \in [\eta, 1-\eta]$ for the one-sided (two-sided) case, where $\hat{\sigma}(x,\tau)$ is the estimated standard error of the QSTE. That is, let $M_b = \max_{\tau\in[\eta,1-\eta]} \Delta_b^u(x,\tau)/\hat{\sigma}(x,\tau)$ for the one-sided case and $M_b = \max_{\tau\in[\eta,1-\eta]} |\Delta_b^u(x,\tau)|/\hat{\sigma}(x,\tau)$ for the two-sided case. In practice, one can let U_i obey the standard normal distribution and let $\eta = 0.1$ so that $\tau \in [0.1, 0.9]$.

6. Rank M_b in an ascending order such that $M_{(1)} \leq \ldots \leq M_{(B)}$ and then define $M_{(\lfloor \alpha B \rfloor)}$ as the critical value $\hat{C}_{\alpha,\eta}$, where $\lfloor c \rfloor$ is the floor function returning the largest integer not greater than c. The uniform confidence bands are given by (3.2) and (3.3).

D Proofs

Proof of Lemma 2.1

For the DSF,

$$G_d(x,y) = E_{\epsilon_d}\big[1\{m_d(x,\beta_d)+\epsilon_d \le y\}\big] = E_{\epsilon_d}\big[1\{\epsilon_d \le y - m_d(x,\beta_d)\}\big] = F_{\epsilon_d}(y - m_d(x,\beta_d)),$$

where the expectation is taken with respect to the unconditional distribution of ϵ_d. The QSF follows by

$$\begin{aligned} Q_d(x,\tau) &= \inf\big\{y : G_d(x,y) \ge \tau\big\} = \inf\big\{y : F_{\epsilon_d}\big(y - m_d(x,\beta_d)\big) \ge \tau\big\} \\ &= m_d(x,\beta_d) + \inf\big\{y - m_d(x,\beta_d) : F_{\epsilon_d}\big(y - m_d(x,\beta_d)\big) \ge \tau\big\} = m_d(x,\beta_d) + Q_{\epsilon_d}(\tau). \end{aligned}$$

□

Proof of Lemma 2.2

We show the identification of β_d and $F_{\epsilon_d}(e)$ which is sufficient for other objects of interest. For β_d, from Assumption 2.2(ii) we know that β_d uniquely solves the population problem

$$\min_{b_d \in \mathcal{B}_d} E\big[Y_d - m_d(X,b_d)\big]^2.$$

By law of iterated expectations,

$$\begin{aligned} & E\left[\frac{1\{D=d\}\cdot\big(Y - m_d(X,b_d)\big)^2}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right] \\ &= E\left[E\left[\frac{1\{D=d\}\cdot\big(Y - m_d(X,b_d)\big)^2}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\middle| X\right]\right] \\ &= E\left[\frac{1}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}} E\left[1\{D=d\}\cdot\big(Y - m_d(X,b_d)\big)^2\big| X, D=d\right]\Pr(D=d|X)\right] \\ &= E\left[E\left[\big(Y_d - m_d(X,b_d)\big)^2\big| X, D=d\right]\right] \\ &= E\left[E\left[\big(Y_d - m_d(X,b_d)\big)^2\big| X\right]\right] \\ &= E\left[\big(Y_d - m_d(X,b_d)\big)^2\right], \end{aligned}$$

where the second equality holds by expanding the conditional expectation according to D, the third equality comes from $\Pr(D=d|X) = \big[p(X)\big]^d\big[1-p(X)\big]^{1-d}$ and $Y = Y_d$ when $D = d$. By Assumption 2.1(i), the fourth equality holds. Then by law of iterated expectations again the fifth equality holds. Since D, X, Y are all observable, β_d is identified. For $F_{\epsilon_d}(e)$, simply replace $(Y - m_d(X,b_d))^2$ above with $1\{Y - m_d(X,\beta_d) \le e\}$ and the result follows. Since β_d and $F_{\epsilon_d}(e)$ are identified, $Q_{\epsilon_d}(\tau)$, $G_d(x,y)$, $q_d(x,\tau)$, and $\delta(x,\tau)$ are also identified according to Lemma 2.1. □

Proof of Lemma 3.1

We first show the asymptotic of $\hat{\beta}_d$. The proof for $\hat{F}_{\epsilon_d}(e)$ case 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. By a mean-value expansion about β_d in the first-order condition for $\hat{\beta}_d$, it is true that

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

where the score $s(\beta_d, p(X))$ and Hessian $H(\beta_d, p(X))$ are defined as

$$\begin{aligned}
s\big(\beta_d, p(X)\big) &= -\frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\big[Y - m_d(X,\beta_d)\big]}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}},\\
H\big(\beta_d, p(X)\big) &= -\frac{1\{D=d\}\cdot\nabla^2 m_d(X,\beta_d)\big[Y - m_d(X,\beta_d)\big]}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}} + \frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\nabla m_d(X,\beta_d)'}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}.
\end{aligned}$$

Similar to the proof of Lemma 2.2, one can show that $E\big[s\big(\beta_d, p(X)\big)\big] = 0$ and

$$E\big[H\big(\beta_d, p(X)\big)\big] = E\left[\frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\nabla m_d(X,\beta_d)'}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right].$$

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

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

where the last term is actually zero by Assumption 2.2(i). Thus,

$$\begin{aligned}
&\left|\sqrt{n}(\hat{\beta}_d - \beta_d) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} E\big[H\big(\beta_d, p(X)\big)\big]^{-1} s\big(\beta_d, \hat{p}(X_i)\big)\right|\\
&= \Bigg|\sqrt{n}(\hat{\beta}_d - \beta_d) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg\{E\left[\frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\nabla m_d(X,\beta_d)'}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right]^{-1}\cdot\\
&\qquad\frac{1\{D_i=d\}\cdot\nabla m_d(X_i,\beta_d)\big[Y_i - m_d(X_i,\beta_d)\big]}{\big[p(X_i)\big]^d\big[1-p(X_i)\big]^{1-d}}\Bigg\}\Bigg| = o_p(1).
\end{aligned}$$

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

$$E\left[\frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\nabla m_d(X,\beta_d)'}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}\right]^{-1}\frac{1\{D=d\}\cdot\nabla m_d(X,\beta_d)\big[Y - m_d(X,\beta_d)\big]}{\big[p(X)\big]^d\big[1-p(X)\big]^{1-d}}. \qquad \square$$

Proof of Theorem 3.1

Given Lemma 2.1 and Lemma 3.1, the influence function representation and the asymptotic properties of $\hat{G}_d(x, y)$ is similar to Theorem 3.3 of Donald, Hsu, and Barrett (2012) so is omitted. For the $\hat{q}_d(x, \tau)$ and $\hat{\delta}(x, \tau)$ cases, one can directly apply the functional delta method for the results given that the quantile map is Hadamard differentiable. A similar proof can be found in Theorem 3.8 of Donald and Hsu (2014). $\square$

Proof of Theorem 3.2

The proof is a combination of Theorem 4.2 and Theorem 4.5 of Donald and Hsu (2014) so we omit it. $\square$

Proof of Lemma B.1

The proof is similar to Lemma 4.1 of Donald and Hsu (2014) so is omitted. $\square$

Proof of Lemma B.2

The proof is similar to Lemma 4.4 of Donald and Hsu (2014) so is omitted. $\square$

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.

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.

Brinch, C. N., M. Mogstad and M. Wiswall (2017): "Beyond LATE with a Discrete Instrument," *Journal of Political Economy*, **125**, 985–1039.

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.

Chernozhukov, V., I. Fernández-Val, W. Newey, S. Stouli and F. Vella (2017): "Semiparametric Estimation of Structural Functions in Nonseparable Triangular Models," *Working Paper.*

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, **73**, 245–261.

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.

Firpo, S. and C. Pinto (2016): “Identification and Estimation of Distributional Impacts of Interventions using Changes in Inequality Measures,” *Journal of Applied Econometrics*, **31**, 457–486.

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.

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.

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. (2010): “Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009),” *Journal of Economic Literature*, **48**, 399–423.

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.

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.

Orzechowski, W. and R. Walker (2017): *The Tax Burden on Tobacco: Volume 51, 1970-2016.*

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.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, **66**, 688.

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.

Viscusi, W. K. and J. Hersch (2001): "Cigarette Smokers as Job Risk Takers," *Review of Economics and Statistics*, **83**, 269–280.

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.

Zhao, M., Y. Konishi and P. Glewwe (2012): "Does Smoking Affect Schooling? Evidence from Teenagers in Rural China," *Journal of Health Economics*, **31**, 584–598.

Table 1: Descriptive Statistics

	Males		Females	
	Smokers	Non-Smokers	Smokers	Non-Smokers
Average Hourly Wage	16.705	29.486	13.086	17.994
	(10.677)	(32.186)	(8.292)	(12.744)
Non-White	0.469	0.355	0.682	0.644
	(0.500)	(0.479)	(0.467)	(0.479)
Education	11.880	13.979	12.330	13.858
	(1.998)	(2.339)	(1.957)	(2.208)
White-Collar	0.093	0.356	0.112	0.336
	(0.291)	(0.479)	(0.316)	(0.472)
Union	0.137	0.140	0.112	0.145
	(0.344)	(0.347)	(0.316)	(0.352)
Cigarette Price	3.831	3.319	3.601	3.449
	(1.459)	(1.310)	(1.439)	(1.410)
Ever Smoke	1.000	0.141	1.000	0.132
	(0.000)	(0.348)	(0.000)	(0.339)
Sample Size	409	2610	179	1248

Standard deviations are in parentheses.

Table 2: Unconfoundedness Test

	Size	Covariates	Specifications	p-value
Males	3019	Non-White, Education, White-Collar, Union, Cigarette Price	Constant	0.000
			Linear	0.108
			Interaction	0.331
			Pure Quadratic	0.163
			Quadratic	0.109
Females	1372	Non-White, Education, White-Collar, Union, Cigarette Price	Constant	0.001
			Linear	0.115
			Interaction	0.708
			Pure Quadratic	0.477
			Quadratic	0.647

Propensity scores are trimmed to lie in the interval [0.005, 0.995].

Table 3: ATEs

		Methods			
	Size	RA	IPW	IPWRA	WNLS
Males	3019	-0.140^{***}	-0.246^{***}	-0.147^{**}	-0.132^{**}
		(0.043)	(0.039)	(0.046)	(0.049)
Females	1372	-0.078	-0.066	-0.040	-0.030
		(0.055)	(0.063)	(0.052)	(0.049)

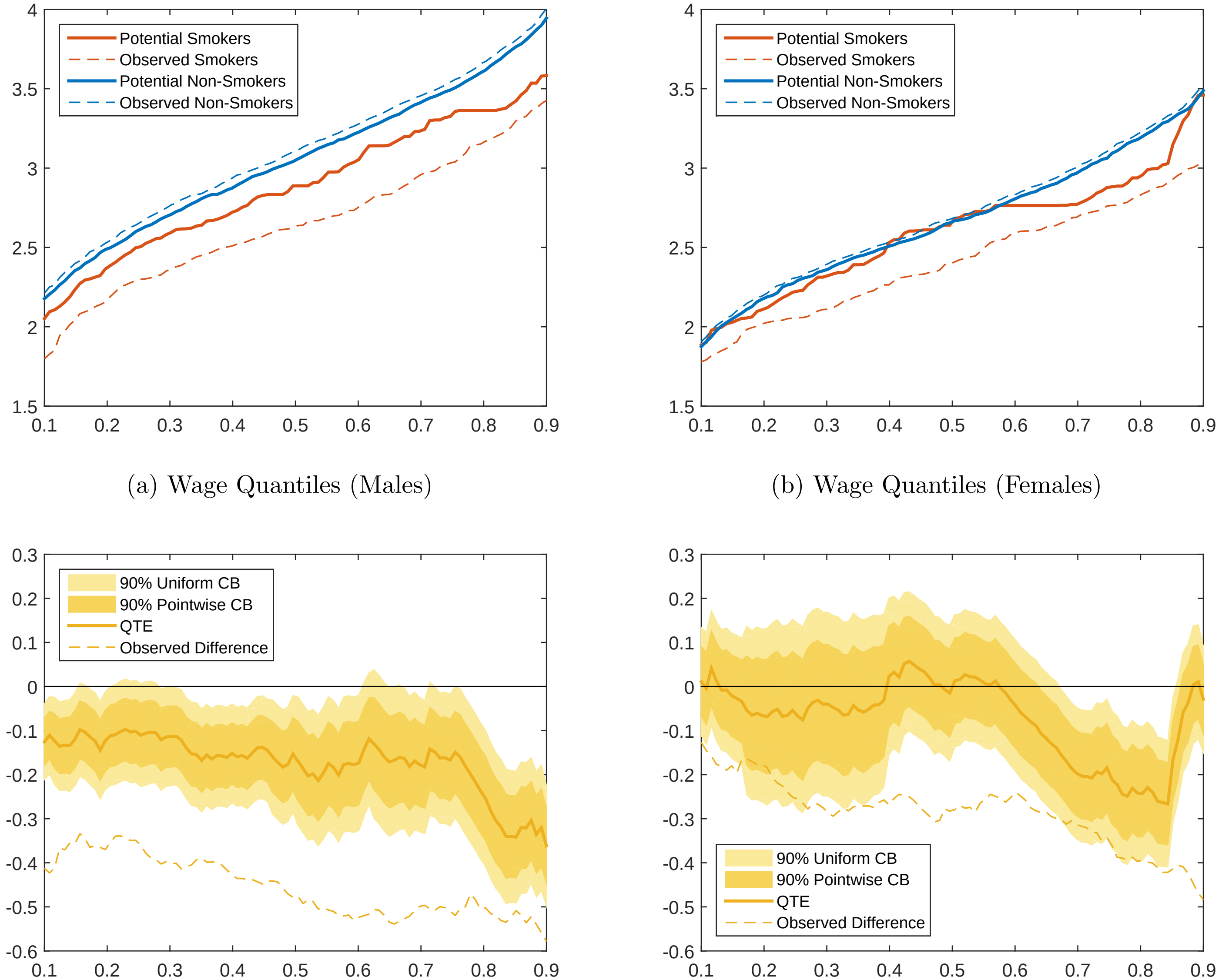
* Significant at 5%, ** Significant at 1%, *** Significant at 0.1%. Robust standard errors are in parentheses.

Table 4: Fixed Values

	Males	Females
Non-White	✗	✓
White-Collar	✗	✗
Union	✗	✗
Education	12.958	12.858
Cigarette Price	3.429	3.45
Number of the Most Typical Individuals	971	592
Percentage (%)	32.16	41.49

For binary variables, the fixed values are assigned to the most frequently occurring types of individuals according to the data. For continuous variables, the fixed values are assigned to the group means of the most typical subgroup.

Figure 1: Potential Wages and QTEs



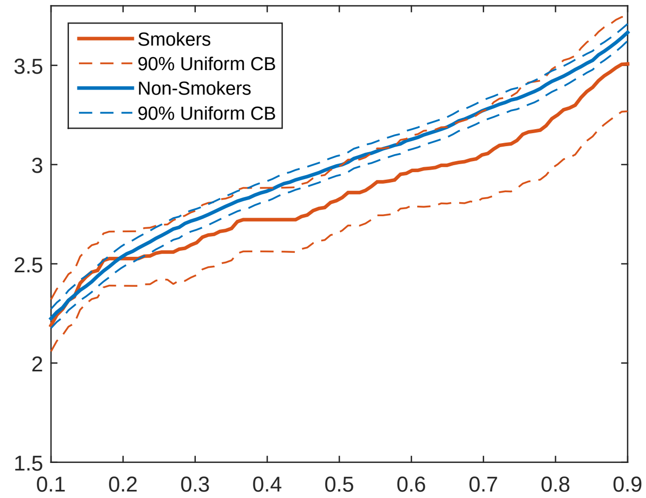
(a) Wage Quantiles (Males)

(b) Wage Quantiles (Females)

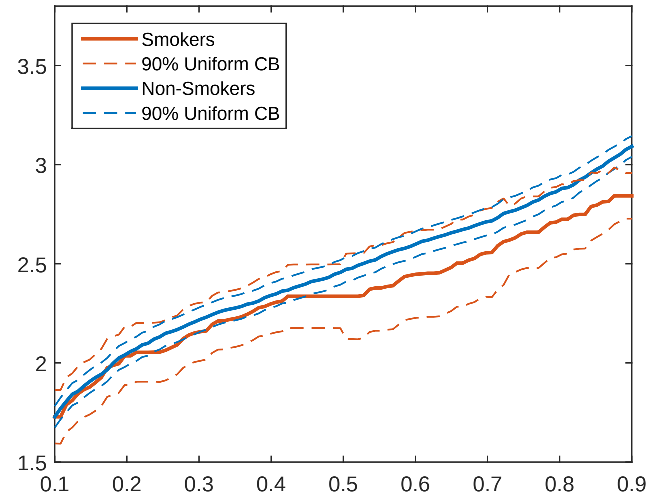
(c) QTE (Males)

(d) QTE (Females)

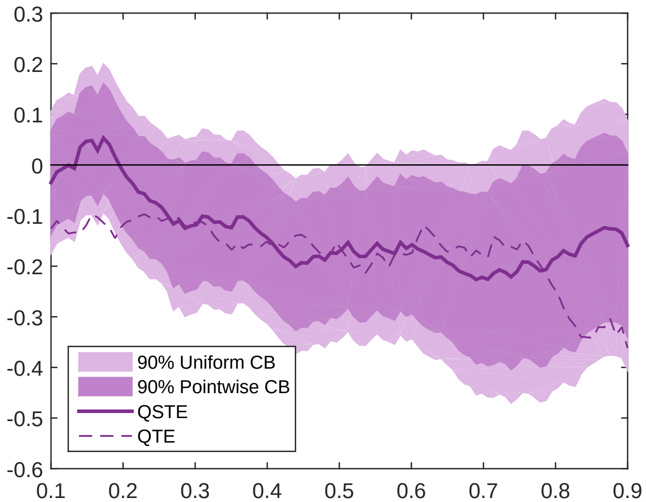
Figure 2: QSFs and QSTEs



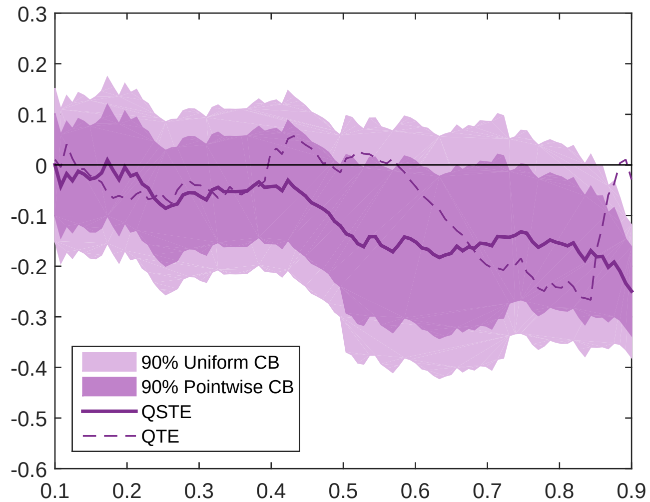
(a) QSFs (Males)



(b) QSFs (Females)



(c) QSTE (Males)



(d) QSTE (Females)