NBER WORKING PAPER SERIES

SENSITIVITY TO MISSING DATA ASSUMPTIONS: THEORY AND AN EVALUATION OF THE U.S. WAGE STRUCTURE

Patrick Kline
Andres Santos

Working Paper 15716
http://www.nber.org/papers/w15716

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2010

We are grateful to David Card, Guido Imbens, Justin McCrary, Hal White, and seminar participants at UC Berkeley, UC San Diego, the University of Michigan, USC, Stanford, the 2010 NBER Summer Institute, the 2010 Western Economic Association Summer Meeting, the 2010 Seoul Summer Economics Conference and the 2010 Econometric Society World Congress for useful comments and corrections. We thank Ivan Fernández-Val for assistance in replicating the results of Angrist et al. (2006). A previous version of this paper circulated under the title \Interval Estimation of Potentially Misspecified Quantile Models in the Presence of Missing Data." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Sensitivity to Missing Data Assumptions: Theory and An Evaluation of the U.S. Wage Structure
Patrick Kline and Andres Santos
NBER Working Paper No. 15716
February 2010, Revised February 2011
JEL No. C01,C1,J3

ABSTRACT

This paper develops methods for assessing the sensitivity of empirical conclusions regarding conditional distributions to departures from the missing at random (MAR) assumption. We index the degree of non-ignorable selection governing the missingness process by the maximal Kolmogorov-Smirnov (KS) distance between the distributions of missing and observed outcomes across all values of the covariates. Sharp bounds on minimum mean square approximations to conditional quantiles are derived as a function of the nominal level of selection considered in the sensitivity analysis and a weighted bootstrap procedure is developed for conducting inference. Using these techniques, we conduct an empirical assessment of the sensitivity of observed earnings patterns in U.S. Census data to deviations from the MAR assumption. We find that the well-documented increase in the returns to schooling between 1980 and 1990 is relatively robust to deviations from the missing at random assumption except at the lowest quantiles of the distribution, but that conclusions regarding heterogeneity in returns and changes in the returns function between 1990 and 2000 are very sensitive to departures from ignorability.

Patrick Kline
Department of Economics
UC, Berkeley
508-1 Evans Hall #3880
Berkeley, CA 94720
and NBER
pkline@econ.berkeley.edu

Andres Santos
Department of Economics
9500 Gilman Drive
La Jolla, CA 92093-0508
a2santos@ucsd.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w15716

1 Introduction

Despite major advances in the design and collection of survey and administrative data, missingness remains a pervasive feature of virtually every modern economic dataset. Hirsch and Schumacher (2004), for instance, find that nearly 30% of the earnings observations in the Outgoing Rotation Groups of the Current Population Survey are imputed. Similar allocation rates are present in other major earnings sources such as the March CPS and Decennial Census with the problem growing worse in more recent years.

The dominant framework for dealing with missing data has been to assume that it is "missing at random" (Rubin (1976)) or "ignorable" conditional on observable demographics; an assumption whose popularity owes more to convenience than plausibility. Even in settings where it is reasonable to believe that non-response is approximately ignorable, the extent of missingness in modern economic data suggests that economists ought to assess the sensitivity of their conclusions to small deviations from this assumption.

Previous work on non-ignorable missing data processes has either relied upon parametric models of missingness in conjunction with exclusion restrictions to obtain point identification (Greenlees et al. (1982) and Lillard et al. (1986)) or considered the "worst case" bounds on population moments that result when all assumptions regarding the missingness process are abandoned (Manski (1994, 2003)). Neither approach has garnered much popularity.[1] It is typically quite difficult to find variables which shift the probability of missingness but are uncorrelated with population outcomes. And for most applied problems, the worst case bounds are overly conservative in the sense that they consider missingness processes unlikely to be found in modern datasets.

We propose here an alternative approach for use in settings where one lacks prior knowledge of the missing data mechanism. Rather than ask what can be learned about the parameters of interest given assumptions on the missingness process, we investigate the level of non-ignorable selection necessary to undermine ones' conclusions regarding the conditional distribution of the data obtained under a missing at random (MAR) assumption. We do so by making use of a nonparametric measure of selection – the maximal Kolmogorov-Smirnov (KS) distance between the distributions of missing and observed outcomes across all values of the covariates. The KS distance yields a natural parameterization of deviations from ignorability, with a distance of zero corresponding to MAR and a distance of one encompassing the totally unrestricted missingness processes considered in Manski (1994). Between these extremes lie a continuum of selection mechanisms which may be studied to determine a critical level of selection above which conclusions obtained under an analysis predicated upon MAR may be overturned.

To enable such an analysis, we begin by deriving sharp bounds on the conditional quantile function (CQF) under nominal restrictions on the degree of selection present. We focus on the commonly encountered setting where outcome data are missing and covariates are discrete. In order

[1] See DiNardo et al. (2006) for an applied example comparing these two approaches.

to facilitate the analysis of datasets with many covariates, results are also developed summarizing the conclusions that can be drawn regarding "pseudo-true" parametric approximations to the underlying nonparametric CQF of the sort considered by Chamberlain (1994). When point identification of the CQF fails due to missingness, the identified set of pseudo true parameters consists of all coefficients associated with minimum mean square approximations to functions lying within the CQF bounds.

We obtain sharp bounds on the coordinates of the pseudo true parameter vector and propose computationally simple estimators for them. We show that these estimators converge in distribution to a Gaussian process indexed by the quantile of interest and the level of the nominal restriction on selection and develop a weighted bootstrap procedure for consistently estimating that distribution. This procedure enables inference on the entire pseudo-true quantile process as indexed both by the quantile of interest and the level of the selection bound.

Substantively these methods allow a determination of the critical level of selection for which hypotheses regarding conditional quantiles, parametric approximations to conditional quantiles, or entire conditional distributions cannot be rejected. For example we study the "breakdown" function defined implicitly as the level of selection necessary for conclusions to be overturned at each quantile. The uniform confidence region for this function effectively summarizes the differential sensitivity of the entire conditional distribution to violations of MAR. These techniques substantially extend the recent econometrics literature on sensitivity analysis (Altonji et al. (2005, 2008), Imbens (2003), Rosenbaum and Rubin (1983), Rosenbaum (2002)), most of which has focused on the sensitivity of scalar treatment effect estimates to confounding influences, typically by using assumed parametric models of selection.

Having established our inferential procedures, we turn to an empirical assessment of the sensitivity of heavily studied patterns in the conditional distribution of U.S. wages to deviations from the MAR assumption. We begin by revisiting the results of Angrist et al. (2006) regarding changes across Decennial Censuses in the quantile specific returns to schooling. Weekly earnings information is missing for roughly a quarter of the observations in their study, suggesting the results may be sensitive to small deviations from ignorability. We show that despite extensive missingness in the dependent variable, the well-documented increase in the returns to schooling between 1980 and 1990 is relatively robust to deviations from the missing at random assumption except at the lowest quantiles of the conditional distribution. However, deterioration in the quality of Decennial Census data renders conclusions regarding heterogeneity in returns and changes in the returns function between 1990 and 2000 very sensitive to departures from ignorability at all quantiles. We also show, using a more flexible model studied by Lemieux (2006), that the apparent convexification of the earnings-education profile between 1980 and 2000 is robust to modest deviations from MAR while changes in the wage structure at lower quantiles are more easily obscured by selection.

To gauge the practical relevance of these sensitivity results we analyze a sample of workers from the 1973 Current Population Survey for whom IRS earnings records are available. This sample allows us to observe the earnings of CPS participants who, for one reason or another, failed to provide valid earnings information to the CPS. We show that IRS earnings predict non-response to the CPS within demographic covariate bins, with very high and very low earning individuals most likely to have invalid CPS earnings records. Measuring the degree of selection using our proposed KS metric we find significant deviations from ignorability with patterns of selection that vary substantially across demographic groups. Given recent trends in survey imputation rates, these findings suggest economists' knowledge of the location and shape of conditional earnings distributions in the U.S. may be more tentative than previously supposed.

The remainder of the paper is structured as follows: Section 2 describes our index of selection and our general approach to assessing sensitivity. Section 3 develops our approach to assessing the sensitivity of parametric approximations to conditional quantiles. Section 4 obtains the results necessary for estimation and inference on the bounds provided by restrictions on the selection process. In Section 5 we present our empirical study and briefly conclude in Section 6.

2 Assessing Sensitivity

Consider the random variables (Y, X, D) with joint distribution F, where $Y \in \mathbf{R}$, $X \in \mathbf{R}^l$ and $D \in \{0,1\}$ is a dummy variable that equals one if Y is observable and zero otherwise. Denote the distribution of Y given X and the distribution of Y given X and D respectively as:

$$F_{y|x}(c) \equiv P(Y \leq c|X = x) \qquad F_{y|d,x}(c) \equiv P(Y \leq c|D = d, X = x) \ , \tag{1}$$

where $d \in \{0,1\}$ and further define the probability of Y being observed conditional on X to be:

$$p(x) \equiv P(D = 1|X = x) \ . \tag{2}$$

In conducting a sensitivity analysis the researcher seeks to assess how the identified features of $F_{y|x}(c)$ depend upon alternative assumptions regarding the process generating D. In particular, we will concern ourselves with the sensitivity of conclusions regarding $q(\tau|X)$, the conditional τ-quantile of Y given X, which is often of more direct interest than the distribution function itself. Towards this end, we impose the following assumptions on the data generating process:

Assumption 2.1. *(i) $X \in \mathbf{R}^l$ has finite support $\mathcal{X}$; (ii) $F_{y|d,x}(c)$ is continuous and strictly increasing at all c such that $0 < F_{y|d,x}(c) < 1$; (iii) D equals one if Y is observable and zero otherwise.*

The discrete support requirement in Assumption 2.1(i) simplifies inference as it obviates the need to employ nonparametric estimators of conditional quantiles. While this assumption may be restrictive in some environments, it is still widely applicable as illustrated in our study of quantile specific returns to education in Section 5. It is also important to emphasize that Assumption 2.1(i) is not necessary for our identification results, but only for our discussion of inference. Assumption 2.1(ii) ensures that for any $0 < \tau < 1$, the τ-conditional quantile of Y given X is uniquely defined.

Most previous work on sensitivity analysis (e.g. Rosenbaum and Rubin (1983), Altonji et al.

(2005)) has relied upon parametric models of selection. While potentially appropriate in cases where particular deviations from ignorability are of interest, such approaches risk understating sensitivity by implicitly ruling out a wide class of selection mechanisms. We now develop an alternative approach designed to allow an assessment of sensitivity to arbitrary deviations from ignorability that retains much of the parsimony of parametric methods. Specifically, we propose studying a nonparametric class of selection models indexed by a scalar measure of the deviations from MAR they generate. A sensitivity analysis may then be conducted by considering the conclusions that can be drawn under alternative levels of the selection index, with particular attention devoted to determination of the threshold level of selection necessary to undermine conclusions obtained under ignorability.

Since ignorability occurs when $F_{y|1,x}$ equals $F_{y|0,x}$, it is natural to measure deviations from MAR in terms of the distance between these two distributions. We propose as an index of selection the maximal Kolmogorov-Smirnov (KS) distance between $F_{y|1,x}$ and $F_{y|0,x}$ across all values of the covariates.[2] Thus, for $\mathcal{X}$ the support of X, we define the selection metric:

$$\mathcal{S}(F) \equiv \sup_{x \in \mathcal{X}} \sup_{c \in \mathbf{R}} |F_{y|1,x}(c) - F_{y|0,x}(c)| \ . \tag{3}$$

Note that the missing at random assumption may be equivalently stated as $\mathcal{S}(F) = 0$, while $\mathcal{S}(F) = 1$ corresponds to severe forms of selection where $F_{y|1,x}$ and $F_{y|0,x}$ fail to overlap at some point $x \in \mathcal{X}$. For illustrative purposes, Appendix A provides a numerical example mapping the parameters of a bivariate normal selection model into values of $\mathcal{S}(F)$ and plots of the corresponding observed and missing data CDFs.

Restrictions on $\mathcal{S}(F)$ can be shown to yield sharp tractable bounds on the conditional distribution function $F_{y|x}(\cdot)$ as well as its value at any particular point of evaluation $F_{y|x}(c)$. This facilitates the study of both conditional quantiles $(F^{-1}_{y|x}(c))$ and conditional quantile processes $(F^{-1}_{y|x}(\cdot))$ in families of non-ignorable selection mechanisms indexed by $\mathcal{S}(F)$. By construction, any scalar metric of selection will be less informative than a full description of the selection process. Researchers who suspect particular forms of heterogeneity in the selection mechanism across covariate values may wish to consider separate indices of selection for each point in the support of X or consider a maximum of weighted KS distances in (3). Likewise, if one has in mind particular classes of selection mechanisms, it is possible to consider indices based upon weighted KS distances with weights that vary across points of evaluation c. Though such approaches entail a simple extension of our methods, we do not pursue them here. Our approach is tailored to environments where prior knowledge of the selection mechanism is not available. If substantial prior information is available, or if particular sorts of violations of MAR are of interest, it may be better to work with a semi-parametric or even parametric model of selection.

We note in passing that parametric models of selection often imply stronger restrictions on the relationship between the observed and missing data distributions than is sometimes appreciated.

[2] The Kolmogorov-Smirnov distance between two distributions $H_1(\cdot)$ and $H_2(\cdot)$ is defined as $\sup_{c \in \mathbf{R}} |H_1(c) - H_2(c)|$.

For example, the bivariate normal selection model considered in Appendix A tends to achieve maximal distances between missing and observed distributions at points near the center of the unselected distribution while points in the tails of the distribution exhibit relatively small discrepancies. However, this model cannot accommodate patterns of selection that would lead the CDFs of missing and observed outcomes to cross at some values as might occur if individuals with very high or very low values of Y are most likely to have missing observations – a pattern conjectured to be present in earnings data by Lillard et al. (1986) and corroborated in our later analysis of earnings validation data in Section 5.[3] In cases where crossing occurs, the maximal distance between distributions will tend to be achieved far from the crossing point, often in the tails of the unselected distribution. Since one typically does not know whether or where the CDFs cross, or how this behavior varies across covariate bins, it can be difficult to develop a model of non-response suitable for the study of conditional distributions. For this reason, the consideration of a nonparametric measure of deviations from MAR of the sort indexed by our maximal KS metric $\mathcal{S}(F)$, is likely to be of interest in most settings where missing data are present.

For $q(\tau|X)$ the conditional τ-quantile of Y given X, we examine what can be learned about the conditional quantile function $q(\tau|\cdot)$ under the nominal restriction:

$$\mathcal{S}(F) \leq k \ . \tag{4}$$

Knowledge of a true value of k for which (4) holds is not presumed. Rather, we propose examining the conclusions that may be drawn on the CQF given various candidate values of k. By considering multiple values of k, it is possible to deduce what level of selection is necessary to overturn conclusions of interest obtained under a MAR analysis.

In the absence of additional restrictions, the conditional quantile function ceases to be identified under any deviation from ignorability ($k > 0$). Nonetheless, $q(\tau|\cdot)$ may still be shown to lie within a nominal identified set. This set consists of the values of $q(\tau|\cdot)$ that would be compatible with the distribution of observables were the putative restriction $\mathcal{S}(F) \leq k$ known to hold. We qualify such a set as nominal due to restriction (4) being part of a hypothetical exercise only.

The following Lemma provides a sharp characterization of the nominal identified set:

Lemma 2.1. *Suppose Assumptions 2.1(ii)-(iii) hold, $\mathcal{S}(F) \leq k$ and let $F^{-}_{y|1,x}(c) = F^{-1}_{y|1,x}(c)$ if $0 < c < 1$, $F^{-}_{y|1,x}(c) = -\infty$ if $c \leq 0$ and $F^{-}_{y|1,x}(c) = \infty$ if $c \geq 1$. Defining $(q_L(\tau,k|x), q_U(\tau,k|x))$ by:*

$$q_L(\tau,k|x) \equiv F^{-}_{y|1,x}\Big(\frac{\tau - \min\{\tau + kp(x), 1\}(1-p(x))}{p(x)}\Big)$$
$$q_U(\tau,k|x) \equiv F^{-}_{y|1,x}\Big(\frac{\tau - \max\{\tau - kp(x), 0\}(1-p(x))}{p(x)}\Big) \ ,$$

it follows that the identified set for $q(\tau|\cdot)$ is $\mathcal{C}(\tau,k) \equiv \{\theta : \mathcal{X} \to \mathbf{R} : q_L(\tau,k|\cdot) \leq \theta(\cdot) \leq q_U(\tau,k|\cdot)\}$.

The bounds in Lemma 2.1 are given by quantiles of the conditional distribution of observed

[3] Crossing of CDFs may also occur in two-sided selection models of the sort considered by Neal (2004).

outcomes. The nominal identified set $\mathcal{C}(\tau, k)$ is sharp for $q(\tau|\cdot)$ in that for every function $\theta \in \mathcal{C}(\tau, k)$ there exists a distribution $\tilde{F}$ of (Y, X, D) that matches the distribution of observables, satisfies $\mathcal{S}(\tilde{F}) \le k$ and has conditional τ-quantile function θ. It is interesting to note that $\mathcal{C}(\tau, k)$ provides a smooth parametrization between identification under MAR ($k = 0$) and the bounds derived in Manski (1994) which impose no restrictions on the selection mechanism ($k = 1$). Between these two extremes, however, lie a continuum of identified sets corresponding to families of selection mechanisms yielding different degrees of departure from ignorability.

2.1 Examples

We conclude this section by illustrating through examples how the bound functions (q_L, q_U) may be used to evaluate the sensitivity of conclusions obtained under MAR. For simplicity, we let X be binary so that the conditional τ-quantile function $q(\tau|\cdot)$ takes only two values.

Example 2.1. (Pointwise Conclusions) Suppose interest centers on whether $q(\tau|X = 1)$ equals $q(\tau|X = 0)$ for a specific quantile τ_0. A researcher who finds them to differ under a MAR analysis may easily assess the sensitivity of his conclusion to the presence of selection by employing the functions $(q_L(\tau_0|\cdot), q_U(\tau_0|\cdot))$. Concretely, the minimal amount of selection necessary to overturn the conclusion that the conditional quantiles differ is given by:

$$k_0 \equiv \inf k : q_L(\tau_0, k|X = 1) - q_U(\tau_0, k|X = 0) \le 0 \le q_U(\tau_0, k|X = 1) - q_L(\tau_0, k|X = 0) \ . \qquad (5)$$

That is, k_0 is the minimal level of selection under which the nominal identified sets for $q(\tau_0|X = 0)$ and $q(\tau_0|X = 1)$ contain a common value. ■

Example 2.2. (Distributional Conclusions) A researcher is interested in whether the conditional distribution $F_{y|x=0}$ first order stochastically dominates $F_{y|x=1}$, or equivalently, whether $q(\tau|X = 1) \le q(\tau|X = 0)$ for all $\tau \in (0, 1)$. She finds under MAR that $q(\tau|X = 1) > q(\tau|X = 0)$ at multiple values of τ leading her to conclude that first order stochastic dominance does not hold. Employing the functions (q_L, q_U), she may assess what degree of selection is necessary to cast doubt on this conclusion by examining:

$$k_0 \equiv \inf k : q_L(\tau, k|X = 1) \le q_U(\tau, k|X = 0) \quad \text{for all } \tau \in (0, 1) \ . \qquad (6)$$

Here, k_0 is the smallest level of selection for which an element of the identified set for $q(\cdot|X = 1)$ $(q_L(\cdot, k_0|X = 1))$ is everywhere below an element of the identified set for $q(\cdot|X = 0)$ $(q_U(\cdot, k_0|X = 0))$. Thus, k_0 is the threshold level of selection under which $F_{y|x=0}$ may first order stochastically dominate $F_{y|x=1}$. ■

Example 2.3. (Breakdown Analysis) A more nuanced sensitivity analysis might examine what degree of selection is necessary to undermine the conclusion that $q(\tau|X = 1) \neq q(\tau|X = 0)$ at each specific quantile τ. As in Example 2.1, we can define the quantile specific critical level of selection:

$$\kappa_0(\tau) \equiv \inf k : q_L(\tau, k|X = 1) - q_U(\tau, k|X = 0) \le 0 \le q_U(\tau, k|X = 1) - q_L(\tau, k|X = 0) \ . \qquad (7)$$

By considering $\kappa_0(\tau)$ at different values of τ, we implicitly define a "breakdown" function $\kappa_0(\cdot)$ which reveals the differential sensitivity of the initial conjecture at each quantile $\tau \in (0,1)$. ■

3 Parametric Modeling

Analysis of the conditional τ-quantile function $q(\tau|\cdot)$ and its corresponding nominal identified set $\mathcal{C}(\tau, k)$ can be cumbersome when many covariates are present as the resulting bounds will be of high dimension and difficult to visualize. Moreover, it can be arduous even to state the features of a high dimensional CQF one wishes to examine for sensitivity. It is convenient in such cases to be able to summarize $q(\tau|\cdot)$ using a parametric model. Failure to acknowledge, however, that the model is simply an approximation can easily yield misleading conclusions.

Figure 1 illustrates a case where the nominal identified set $\mathcal{C}(\tau, k)$ possesses an erratic (though perhaps not unusual) shape. The set of linear CQFs obeying the bounds provide a poor description of this set, covering only a small fraction of its area. Were the true CQF known to be linear this reduction in the size of the identified set would be welcome, the benign result of imposing additional identifying information. But in the absence of true prior information these reductions in the size of the identified set are unwarranted – a phenomenon we term "identification by misspecification".

The specter of misspecification leaves the applied researcher with a difficult choice. One can either conduct a fully nonparametric analysis of the nominal identified set, which may be difficult to interpret with many covariates, or work with a parametric set likely to overstate what is known about the CQF. Under identification, this tension is typically resolved by estimating parametric models that possess an interpretation as best approximations to the true CQF and adjusting the corresponding inferential methods accordingly as advocated in Chamberlain (1994) and Angrist et al. (2006). Following Horowitz and Manski (2006), Stoye (2007), and Ponomareva and Tamer (2009), we extend this approach and develop methods for conducting inference on potentially misspecified parametric models under partial identification.

We focus on linear parametric models and approximations that minimize a known quadratic loss function. For S a known measure on $\mathcal{X}$ and $E_S[g(X)]$ denoting the expectation of $g(X)$ when X is distributed according to S, we define the pseudo true parameter to be:[4]

$$\beta(\tau) \equiv \arg\min_{\gamma \in \mathbf{R}^l} E_S[(q(\tau|X) - X'\gamma)^2] \ . \tag{8}$$

Lack of identification of the conditional quantile function $q(\tau|\cdot)$ due to missing data implies lack of identification of the pseudo true parameter $\beta(\tau)$. We therefore consider the set of pseudo true

[4]The measure S weights the squared deviations in each covariate bin. Its specification is an inherently context-specific task depending entirely upon the researcher's objectives. In Section 4 we weight the deviations by sample size. Other schemes (including equal weighting) may also be of interest in some settings.

parameters which constitute a best approximation to *some* CQF in $\mathcal{C}(\tau, k)$. Formally, we define:

$$\mathcal{P}(\tau, k) \equiv \{\beta \in \mathbf{R}^l : \beta \in \arg\min_{\gamma \in \mathbf{R}^l} E_S[(\theta(X) - X'\gamma)^2] \text{ for some } \theta \in \mathcal{C}(\tau, k)\} . \tag{9}$$

Figure 2 illustrates an element of $\mathcal{P}(\tau, k)$ graphically. While intuitively appealing, the definition of $\mathcal{P}(\tau, k)$ is not necessarily the most convenient for computational purposes. Fortunately, the choice of quadratic loss and the characterization of $\mathcal{C}(\tau, k)$ in Lemma 2.1 imply a tractable alternative representation for $\mathcal{P}(\tau, k)$, which we obtain in the following Lemma.

Lemma 3.1. *If Assumptions 2.1(ii)-(iii), $\mathcal{S}(F) \leq k$ and $E_S[XX']$ is invertible, then it follows that:*

$$\mathcal{P}(\tau, k) = \{\beta \in \mathbf{R}^l : \beta = (E_S[XX'])^{-1} E_S[X\theta(X)] \ \ s.t. \ q_L(\tau, k|x) \leq \theta(x) \leq q_U(\tau, k|x) \ for\ all\ x \in \mathcal{X}\} .$$

Interest often centers on either a particular coordinate of $\beta(\tau)$ or the pseudo-true conditional quantile at a specified value of the covariates. Both these quantities may be expressed as $\lambda'\beta(\tau)$ for some known vector $\lambda \in \mathbf{R}^l$. Using Lemma 3.1 it is straightforward to show that the nominal identified set for parameters of the form $\lambda'\beta(\tau)$ is an interval with endpoints characterized as the solution to linear programming problems.[5]

Corollary 3.1. *Suppose Assumptions 2.1(ii)-(iii), $\mathcal{S}(F) \leq k$, $E_S[XX']$ is invertible and define:*

$$\pi_L(\tau, k) \equiv \inf_{\beta \in \mathcal{P}(\tau,k)} \lambda'\beta = \inf_{\theta} \lambda'(E_S[XX'])^{-1} E_S[X\theta(X)] \ \ s.t. \ q_L(\tau, k|x) \leq \theta(x) \leq q_U(\tau, k|x) \tag{10}$$

$$\pi_U(\tau, k) \equiv \sup_{\beta \in \mathcal{P}(\tau,k)} \lambda'\beta = \sup_{\theta} \lambda'(E_S[XX'])^{-1} E_S[X\theta(X)] \ \ s.t. \ q_L(\tau, k|x) \leq \theta(x) \leq q_U(\tau, k|x) . \tag{11}$$

The nominal identified set for $\lambda'\beta(\tau)$ is then given by the interval $[\pi_L(\tau, k), \pi_U(\tau, k)]$.

Corollary 3.1 provides sharp bounds on the quantile process $\lambda'\beta(\cdot)$ at each point of evaluation τ under the restriction that $\mathcal{S}(F) \leq k$. However, sharpness of the bounds at each point of evaluation does not, in this case, translate into sharp bounds on the entire process. To see this, note that Corollary 3.1 implies $\lambda'\beta(\cdot)$ must belong to the following set:

$$\mathcal{G}(k) \equiv \{g : [0, 1] \to \mathbf{R} : \pi_L(\tau, k) \leq g(\tau) \leq \pi_U(\tau, k) \text{ for all } \tau\} . \tag{12}$$

While the true $\lambda'\beta(\cdot)$ must belong to $\mathcal{G}(k)$, not all functions in $\mathcal{G}(k)$ can be justified as some distribution's pseudo-true process.[6] Therefore, $\mathcal{G}(k)$ does not constitute the nominal identified set for the process $\lambda'\beta(\cdot)$ under the restriction $\mathcal{S}(F) \leq k$. Fortunately, $\pi_L(\cdot, k)$ and $\pi_U(\cdot, k)$ *are* in the identified set over the range of (τ, k) for which the bounds are finite. Thus, the set $\mathcal{G}(k)$, though not sharp, does retain the favorable properties of: (i) sharpness at any point of evaluation τ, (ii) containing the true identified set for the process so that processes not in $\mathcal{G}(k)$ are also known not to be in the identified set; (iii) sharpness of the lower and upper bound functions $\pi_L(\cdot, k)$ and $\pi_U(\cdot, k)$; and (iv) ease of analysis and graphical representation.

[5]Since X has discrete support, we can characterize the function θ by the finite number of values it may take. Because the weighting scheme S is known, so is $\lambda'(E_S[XX'])^{-1}$, and hence the objectives in (10) and (11) are of the form $w'\theta$ where w is a known vector and θ is a finite dimensional vector over which the criterion is optimized.

[6]For example, under our assumptions $\lambda'\beta(\cdot)$ is a continuous function of τ. Hence, any $g \in \mathcal{G}(k)$ that is discontinuous is not in the nominal identified set for $\lambda'\beta(\cdot)$ under the hypothetical that $\mathcal{S}(F) \leq k$.

3.1 Examples

We now revisit Examples 2.1-2.3 from Section 2.1 in order to illustrate how to characterize the sensitivity of conclusions drawn under MAR with parametric models. We keep the simplifying assumption that X is scalar, but no longer assume it is binary and instead consider the model:

$$q(\tau|X) = \alpha(\tau) + X\beta(\tau) \ . \tag{13}$$

Note that when X is binary equation (13) provides a non-parametric model of the CQF, in which case our discussion coincides with that of Section 2.1.

Example 2.1 (cont.) Suppose that an analysis under MAR reveals $\beta(\tau_0) \neq 0$ at a specific quantile τ_0. Employing the functions (π_L, π_U) we may then define the critical level of selection k_0 necessary to cast doubt on this conclusion as:

$$k_0 \equiv \inf k : \pi_L(\tau_0, k) \leq 0 \leq \pi_U(\tau_0, k) \ . \tag{14}$$

That is, under any level of selection $k \geq k_0$ it is no longer possible to conclude that $\beta(\tau_0) \neq 0$. ■

Example 2.2 (cont.) In a parametric analogue of first order stochastic dominance of $F_{y|x}$ over $F_{y|x'}$ for $x < x'$, a researcher examines whether $\beta(\tau) \leq 0$ for all $\tau \in (0, 1)$. Suppose that a MAR analysis reveals that $\beta(\tau) > 0$ for multiple values of τ. The functions (π_L, π_U) enable her to assess what degree of selection is necessary to undermine her conclusions by considering:

$$k_0 \equiv \inf k : \pi_L(\tau, k) \leq 0 \quad \text{for all } \tau \in (0, 1) \ . \tag{15}$$

Note that finding $\pi_L(\tau, k_0) \leq 0$ for all $\tau \in (0, 1)$ does in fact cast doubt on the conclusion that $\beta(\tau) > 0$ for some τ because $\pi_L(\cdot, k_0)$ is itself in the nominal identified set for $\beta(\cdot)$. That is, under a degree of selection k_0, the *process* $\beta(\cdot)$ may equal $\pi_L(\cdot, k_0)$. ■

Example 2.3 (cont.) Generalizing the considerations of Example 2.1, we can examine what degree of selection is necessary to undermine the conclusion that $\beta(\tau) \neq 0$ at each specific τ. In this manner, we obtain a quantile specific critical level of selection:

$$\kappa_0(\tau) \equiv \inf k : \pi_L(\tau, k) \leq 0 \leq \pi_U(\tau, k) \ . \tag{16}$$

As in Section 2.1, the resulting "breakdown" function $\kappa_0(\cdot)$ enables us to characterize the differential sensitivity of the entire conditional distribution to deviations from MAR. ■

4 Estimation and Inference

In what follows we develop methods for conducting sensitivity analysis using sample estimates of $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$. This section is primarily technical and applied readers may wish to skip to the application in Section 5 before studying these methods in detail.

Our strategy for estimating the bounds $\pi_L(\tau,k)$ and $\pi_U(\tau,k)$ consists of first obtaining estimates $\hat{q}_L(\tau,k|x)$ and $\hat{q}_U(\tau,k|x)$ of the conditional quantile bounds and then employing them in place of $q_L(\tau,k|x)$ and $q_U(\tau,k|x)$ in the linear programming problems given in (10) and (11). Thus, an appealing characteristic of our estimator is the reliability and low computational cost involved in solving a linear programming problem – considerations which become particularly salient when implementing a bootstrap procedure for inference.

Recall that the conditional quantile bounds $q_L(\tau,k|x)$ and $q_U(\tau,k|x)$ may be expressed as quantiles of the observed data (see Lemma 2.1). We estimate these bounds using their sample analogues. For the development of our bootstrap procedure, however, it will be useful to consider a representation of these sample estimates as the solution of a general M-estimation problem. Towards this end, we define a family of population criterion functions (as indexed by (τ,b,x)) given by:

$$Q_x(c|\tau,b) \equiv (P(Y\le c, D=1, X=x) + bP(D=0,X=x) - \tau P(X=x))^2 . \tag{17}$$

Under appropriate restrictions on (τ,k), to be shortly specified, $q_L(\tau,k|x)$ and $q_U(\tau,k|x)$ then satisfy:

$$q_L(\tau,k|x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau,\tau+kp(x)) \qquad q_U(\tau,k|x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau,\tau-kp(x)) . \tag{18}$$

Hence, there exists a direct relationship between the bounds $q_L(\tau,k|x)$ and $q_U(\tau,k|x)$ as indexed by (τ,k) and the minimizers of $Q_x(c|\tau,b)$ as indexed by (τ,b).

We therefore employ the sample analogue to $Q_x(c|\tau,b)$ for estimation, which we denote by:

$$Q_{x,n}(c|\tau,b) \equiv \Big(\frac{1}{n}\sum_{i=1}^{n}\{1\{Y_i\le c, X_i=x, D_i=1\} + b1\{D_i=0,X_i=x\} - \tau 1\{X_i=x\}\}\Big)^2 . \tag{19}$$

Exploiting (17), the extremum estimators for the bounds $q_L(\tau,k|x)$ and $q_U(\tau,k|x)$ are then:

$$\hat{q}_L(\tau,k|x) \in \arg\min_{c\in\mathbf{R}} Q_{x,n}(c|\tau,\tau+k\hat{p}(x)) \qquad \hat{q}_U(\tau,k|x) \in \arg\min_{c\in\mathbf{R}} Q_{x,n}(c|\tau,\tau-k\hat{p}(x)) , \tag{20}$$

where $\hat{p}(x) \equiv (\sum_i 1\{D_i=1,X_i=x\})/(\sum_i 1\{X_i=x\})$. Finally, solving the sample analogues to the linear programming problems given in (10) and (11) we obtain the estimators:

$$\hat{\pi}_L(\tau,k) \equiv \inf_{\theta} \lambda'(E_S[XX'])^{-1}E_S[X\theta(X)] \qquad \text{s.t. } \hat{q}_L(\tau,k|x)\le\theta(x)\le\hat{q}_U(\tau,k|x) \tag{21}$$

$$\hat{\pi}_U(\tau,k) \equiv \sup_{\theta} \lambda'(E_S[XX'])^{-1}E_S[X\theta(X)] \qquad \text{s.t. } \hat{q}_L(\tau,k|x)\le\theta(x)\le\hat{q}_U(\tau,k|x) \tag{22}$$

For this approach to prove successful we focus our analysis on choices of (τ,k) for which (18) holds, which is guaranteed by two restrictions. First, we require that (τ,k) be such that the bounds $\pi_L(\tau,k)$ and $\pi_U(\tau,k)$ are finite. Second, we demand that (τ,k) be such that $\mathcal{S}(F)\le k$ proves more informative than the restriction that $F_{y|0,x}$ lie between zero and one. Succinctly, for an arbitrary fixed $\epsilon>0$, we focus on values of (τ,k) that lie in the set:

$$\mathcal{B} \equiv \left\{(\tau,k)\in[0,1]^2 : \begin{array}{ll} \text{(i) } kp(x)(1-p(x))+2\epsilon\le\tau p(x) & \text{(iii) } k\le\tau \\ \text{(ii) } kp(x)(1-p(x))+2\epsilon\le(1-\tau)p(x) & \text{(iv) } k\le 1-\tau \end{array} \ \text{for all } x\in\mathcal{X}\right\}$$

Provided that the conditional probability of missing is bounded away from one and ϵ is small, the set $\mathcal{B}$ is nonempty since it contains the MAR analysis as a special case. In general, however, the set $\mathcal{B}$ imposes that large or small values of τ must be accompanied by small values of k. This simply reflects that the fruitful study of quantiles close to one or zero requires stronger assumptions on the nature of the selection process than the study of, for example, the conditional median.

We introduce the following additional assumption in order to develop our asymptotic theory:

Assumption 4.1. *(i) $\mathcal{B} \neq \emptyset$; (ii) $F_{y|1,x}(c)$ has a continuous bounded derivative $f_{y|1,x}(c)$; (iii) $f_{y|1,x}(c)$ has a continuous bounded derivative $f'_{y|1,x}(c)$; (iv) $E_S[XX']$ is invertible; (v) $f_{y|1,x}(c)$ is bounded away from zero uniformly on all c satisfying $\epsilon \leq F_{y|1,x}(c)p(x) \leq p(x) - \epsilon \ \forall x \in \mathcal{X}$.*

Letting $\hat{\pi}_L$ and $\hat{\pi}_U$ be the functions defined pointwise by (21) and (22), we obtain their asymptotic distribution as elements of $L^\infty(\mathcal{B})$ (the space of bounded functions on $\mathcal{B}$). Such a result is a key step towards constructing confidence intervals for $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$ that are uniform in (τ, k). As we illustrate in Section 4.2, these uniformity results are particularly useful for conducting the sensitivity analyses illustrated in Examples 2.1-2.3.

Theorem 4.1. *If Assumptions 2.1, 4.1 hold and $\{Y_i, X_i, D_i\}_{i=1}^n$ is an i.i.d. sample, then:*

$$\sqrt{n}\begin{pmatrix} \hat{\pi}_L - \pi_L \\ \hat{\pi}_U - \pi_U \end{pmatrix} \xrightarrow{\mathcal{L}} G \ , \tag{23}$$

where G is a Gaussian process on the space $L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$.

We note that since G is a Gaussian process, its marginals $G(\tau, k)$ are simply bivariate normal random variables. For notational convenience, we let $G^{(i)}(\tau, k)$ denote the i^{th} component of the vector $G(\tau, k)$. Thus, $G^{(1)}(\tau, k)$ is the limiting distribution corresponding to the lower bound estimate at the point (τ, k), while $G^{(2)}(\tau, k)$ is the limiting distribution of the upper bound estimate at (τ, k).

4.1 Examples

We now return to the Examples of Section 2.1 and 3.1 and discuss how to conduct inference on the various sensitivity measures introduced there. For simplicity, we assume the relevant critical values are known. In Section 4.2 we develop a bootstrap procedure for their estimation.

Example 2.1 (cont.) Since under any level of selection k larger than k_0 it is also not possible to conclude $\beta(\tau_0) \neq 0$, it is natural to construct a one sided (rather than two sided) confidence interval for k_0. Towards this end, let $r_{1-\alpha}^{(i)}(k)$ be the $1 - \alpha$ quantile of $G^{(i)}(\tau_0, k)$ and define:

$$\hat{k}_0 \equiv \inf k : \hat{\pi}_L(\tau_0, k) - \frac{r_{1-\alpha}^{(1)}(k)}{\sqrt{n}} \leq 0 \leq \hat{\pi}_U(\tau_0, k) + \frac{r_{1-\alpha}^{(2)}(k)}{\sqrt{n}} \ . \tag{24}$$

The confidence interval $[\hat{k}_0, 1]$ then covers k_0 with asymptotic probability at least $1 - \alpha$. ■

Example 2.2 (cont.) Construction of a one sided confidence interval for k_0 in this setting is more challenging as it requires us to employ the uniformity of our estimator in τ. First, let us define:

$$r_{1-\alpha}(k) = \inf r : P\Big(\sup_{\tau \in \mathcal{B}(k)} \frac{G^{(1)}(\tau, k)}{\omega_L(\tau, k)} \leq r \Big) \geq 1 - \alpha \ , \tag{25}$$

where $\mathcal{B}(k) = \{\tau : (\tau, k) \in \mathcal{B}\}$ and ω_L is a positive weight function chosen by the researcher. For every fixed k, we may then construct the following function of τ:

$$\hat{\pi}_L(\cdot, k) - \frac{r_{1-\alpha}(k)}{\sqrt{n}} \omega_L(\cdot, k) \tag{26}$$

which lies below $\pi_L(\cdot, k)$ with asymptotic probability $1 - \alpha$. The function in (26) thus provides a one sided confidence interval for the *process* $\pi_L(\cdot, k)$. The weight function ω_L allows the researcher to account for the fact that the variance of $G^{(1)}(\tau, k)$ may depend heavily on (τ, k). Defining:

$$\hat{k}_0 \equiv \inf k : \sup_{\tau \in \mathcal{B}(k)} \hat{\pi}_L(\tau, k) - \frac{r_{1-\alpha}(k)}{\sqrt{n}} \omega_L(\tau, k) \leq 0 \ , \tag{27}$$

it can then be shown that $[\hat{k}_0, 1]$ covers k_0 with asymptotic probability at least $1 - \alpha$. ■

Example 2.3 (cont.) Employing Theorem 4.1 it is possible to construct a two sided confidence interval for the function $\kappa_0(\cdot)$. Towards this end, we exploit uniformity in τ and k by defining:

$$r_{1-\alpha} \equiv \inf r : P\Big(\sup_{(\tau,k) \in \mathcal{B}} \max\Big\{ \frac{|G^{(1)}(\tau, k)|}{\omega_L(\tau, k)}, \frac{|G^{(2)}(\tau, k)|}{\omega_L(\tau, k)} \Big\} \leq r \Big) \geq 1 - \alpha \ , \tag{28}$$

where as in Example 2.2, ω_L and ω_U are positive weight functions. In addition, we also let:

$$\hat{\kappa}_L(\tau) \equiv \inf k : \hat{\pi}_L(\tau, k) - \frac{r_{1-\alpha}}{\sqrt{n}} \omega_L(\tau, k) \leq 0 \ , \quad \text{and} \quad 0 \leq \hat{\pi}_U(\tau, k) + \frac{r_{1-\alpha}}{\sqrt{n}} \omega_U(\tau, k) \tag{29}$$

$$\hat{\kappa}_U(\tau) \equiv \sup k : \hat{\pi}_L(\tau, k) + \frac{r_{1-\alpha}}{\sqrt{n}} \omega_L(\tau, k) \geq 0 \ , \quad \text{or} \quad 0 \geq \hat{\pi}_U(\tau, k) - \frac{r_{1-\alpha}}{\sqrt{n}} \omega_U(\tau, k) \ . \tag{30}$$

It can then be shown that the functions $(\hat{\kappa}_L(\cdot), \hat{\kappa}_U(\cdot))$ provide a functional confidence interval for $\kappa_0(\cdot)$. That is, $\hat{\kappa}_L(\tau) \leq \kappa_0(\tau) \leq \hat{\kappa}_U(\tau)$ for all τ with asymptotic probability at least $1 - \alpha$. ■

4.2 Bootstrap Critical Values

As illustrated in Examples 2.1-2.3, conducting inference requires use of critical values that depend on the unknown distribution of G, the limiting Gaussian process in Theorem 4.1, and possibly on weight functions ω_L and ω_U (as in (25), (28)). We will allow the weight functions ω_L and ω_U to be unknown, but require the existence of consistent estimators of them:

Assumption 4.2. *(i) $\omega_L(\tau, k) \geq 0$ and $\omega_U(\tau, k) \geq 0$ are continuous and bounded away from zero on $\mathcal{B}$; (ii) There exist estimators $\hat{\omega}_L(\tau, k)$ and $\hat{\omega}_U(\tau, k)$ that are uniformly consistent on $\mathcal{B}$.*

Given (ω_L, ω_U), let G_ω be the Gaussian process on $L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$ that is pointwise defined by:

$$G_\omega(\tau, k) = \begin{pmatrix} G^{(1)}(\tau, k)/\omega_L(\tau, k) \\ G^{(2)}(\tau, k)/\omega_U(\tau, k) \end{pmatrix} . \tag{31}$$

The critical values employed in Examples 2.1-2.3 can be expressed in terms of quantiles of some Lipschitz transformation $L : L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B}) \to \mathbf{R}$ of the random variable G_ω. For instance, in Example 2.2, the relevant critical value, defined in (25), is the $1-\alpha$ quantile of the random variable:

$$L(G_\omega) = \sup_{\tau \in \mathcal{B}(k)} G_\omega^{(1)}(\tau, k) . \tag{32}$$

Similarly, in Example 2.3 the appropriate critical value defined in (28) is the $1-\alpha$ quantile of:

$$L(G_\omega) = \sup_{(\tau,k) \in \mathcal{B}} \max\{G_\omega^{(1)}(\tau, k), G_\omega^{(2)}(\tau, k)\} . \tag{33}$$

We therefore conclude by establishing the validity of a weighted bootstrap procedure for consistently estimating the quantiles of random variables of the form $L(G_\omega)$. The bootstrap procedure is similar to the traditional nonparametric bootstrap with the important difference that the random weights on different observations are independent from each other. Specifically, letting $\{W_i\}_{i=1}^n$ be an *i.i.d.* sample from a random variable W, we impose the following:

Assumption 4.3. *(i) W is positive almost surely, independent of (Y, X, D) and satisfies $E[W] = 1$ and $Var(W) = 1$; (ii) The functional $L : L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B}) \to \mathbf{R}$ is Lipschitz continuous.*

A consistent estimator for quantiles of $L(G_\omega)$ may then be obtained through the algorithm:

STEP 1: Generate a random sample of weights $\{W_i\}_{i=1}^n$ satisfying Assumption 4.3(i) and define:

$$\tilde{Q}_{x,n}(c|\tau, b) \equiv \Big(\frac{1}{n}\sum_{i=1}^n W_i\{1\{Y_i \le c, X_i = x, D_i = 1\} + b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\}\}\Big)^2 . \tag{34}$$

Employing $\tilde{Q}_{x,n}(c|\tau, b)$, obtain the following bootstrap estimators for $q_L(\tau, k|x)$ and $q_U(\tau, k|x)$:

$$\tilde{q}_L(\tau, k|x) \in \arg\min_{c \in \mathbf{R}} \tilde{Q}_{x,n}(c|\tau, \tau + k\tilde{p}(x)) \qquad \tilde{q}_U(\tau, k|x) \in \arg\min_{c \in \mathbf{R}} \tilde{Q}_{x,n}(c|\tau, \tau - k\tilde{p}(x)) \tag{35}$$

where $\tilde{p}(x) \equiv (\sum_i W_i 1\{D_i = 1, X_i = x\})/(\sum_i W_i 1\{X_i = x\})$. Note that $\tilde{q}_L(\tau, k|x)$ and $\tilde{q}_U(\tau, k|x)$ are simply the weighted empirical quantiles of the observed data evaluated at a point that depends on the reweighted missingness probability. Note also that if we had used the conventional bootstrap we would run the risk of drawing a sample for which a covariate bin is empty. This is not a concern with the weighted bootstrap as the weights are required to be strictly positive. ■

STEP 2: Using the bootstrap bounds $\tilde{q}_L(\tau, k|x)$ and $\tilde{q}_U(\tau, k|x)$ from Step 1, obtain the estimators:

$$\tilde{\pi}_L(\tau, k) \equiv \inf_\theta \lambda'(E_S[XX'])^{-1}E_S[X\theta(X)] \quad \text{s.t. } \tilde{q}_L(\tau, k|x) \le \theta(x) \le \tilde{q}_U(\tau, k|x) \tag{36}$$

$$\tilde{\pi}_U(\tau, k) \equiv \sup_\theta \lambda'(E_S[XX'])^{-1}E_S[X\theta(X)] \quad \text{s.t. } \tilde{q}_L(\tau, k|x) \le \theta(x) \le \tilde{q}_U(\tau, k|x) . \tag{37}$$

Algorithms for quickly solving linear programming problems of this sort are available in most modern computational packages. The weighted bootstrap process for G_ω is then defined pointwise by:

$$\tilde{G}_\omega(\tau,k) \equiv \sqrt{n}\begin{pmatrix} (\tilde{\pi}_L(\tau,k) - \hat{\pi}_L(\tau,k))/\hat{\omega}_L(\tau,k) \\ (\tilde{\pi}_U(\tau,k) - \hat{\pi}_U(\tau,k))/\hat{\omega}_U(\tau,k) \end{pmatrix} . \tag{38}$$

Step 3: Our estimator for $r_{1-\alpha}$, the $1-\alpha$ quantile of $L(G_\omega)$, is then given by the $1-\alpha$ quantile of $L(\tilde{G}_\omega)$ conditional on the sample $\{Y_i, X_i, D_i\}_{i=1}^n$ (but not $\{W_i\}_{i=1}^n$):

$$\tilde{r}_{1-\alpha} \equiv \inf\left\{r : P\Big(L(\tilde{G}_\omega) \geq r \ \Big| \{Y_i, X_i, D_i\}_{i=1}^n\Big) \geq 1-\alpha\right\} . \tag{39}$$

In applications, $\tilde{r}_{1-\alpha}$ will generally need to be computed through simulation. This can be accomplished by repeating Steps 1 and 2 until the number of bootstrap simulations of $L(\tilde{G}_\omega)$ is large. The estimator $\tilde{r}_{1-\alpha}$ is then well approximated by the empirical $1-\alpha$ quantile of the bootstrap statistic $L(\tilde{G}_\omega)$ across the computed simulations. ■

We conclude our discussion of inference by establishing $\tilde{r}_{1-\alpha}$ is indeed consistent for $r_{1-\alpha}$.

Theorem 4.2. *Let $r_{1-\alpha}$ be the $1-\alpha$ quantile of $L(G_\omega)$. If Assumptions 2.1, 4.1, 4.2, and 4.3 hold, the cdf of $L(G_\omega)$ is strictly increasing and continuous at $r_{1-\alpha}$ and $\{Y_i, X_i, D_i, W_i\}_{i=1}^n$ is i.i.d, then:*

$$\tilde{r}_{1-\alpha} \xrightarrow{p} r_{1-\alpha} \ .$$

5 Evaluating the U.S. Wage Structure

We turn now to an empirical assessment of the sensitivity of observed patterns in the U.S. wage structure to deviations from the MAR assumption. A large literature reviewed by (among others) Autor and Katz (1999), Heckman et al. (2006) and Acemoglu and Autor (2010) finds important changes over time in the conditional distribution of earnings with respect to schooling levels.

We begin our investigation of the sensitivity of these findings to alternative missing data assumptions by revisiting the results of Angrist et al. (2006) regarding changes across Decennial Censuses in the quantile specific returns to schooling. We analyze the 1980, 1990, and 2000 Census samples considered in their study but, to simplify our estimation routine, and to correct small mistakes found in the IPUMS files since the time their extract was created, we use new extracts of the 1% unweighted IPUMS files for each decade rather than their original mix of weighted and unweighted samples.[7] Sample sizes and imputation rates for the weekly earnings variable are given in Table 1.

We estimate linear conditional quantile models for log earnings per week of the form:

$$q(\tau|X,E) = X'\gamma(\tau) + E\beta(\tau) \ , \tag{40}$$

[7] The sample consists of native born black and white men ages 40-49 with six or more years of schooling who worked at least one week in the past year. Rather than dropping observations with allocated earnings we treat them as missing. We also drop 10 observations falling in demographic cells with greater than 66% missing and 1,404 observations falling into demographic cells with less than 20 observations. Use of the original extracts analyzed in Angrist et al. (2006) yields nearly identical results.

where X consists of an intercept, a black dummy, and a quadratic in potential experience, and E represents years of schooling. Our analysis focuses on the quantile specific "returns" to a year of schooling $\beta(\tau)$ though we note that, particularly in the context of quantile regressions, the Mincerian earnings coefficients need not map into any proper economic concept of individual returns (Heckman et al. (2006)).

Figure 3 provides estimates of the pseudo-true returns functions $\beta(\cdot)$ in 1980, 1990, and 2000 that result from assuming the data are missing at random. Uniform confidence regions for these estimates were constructed by applying the methods of Section 3 subject to the restriction that $\mathcal{S}(F) = 0$.[8] In defining our parametric approximation metric we weight bin-specific deviations by sample size (i.e. we choose S equal to empirical measure, see Section 3).

Our MAR results are similar to those found in Figure 2A of Angrist et al. (2006). They suggest that the returns function increased uniformly across quantiles between 1980 and 1990 but exhibited a change in slope in 2000. The change between 1980 and 1990 is consistent with a general economy-wide increase in the return to human capital accumulation as conjectured by Juhn et al. (1993). However the finding of a shape change in the quantile process between 1990 and 2000 represents a form of heteroscedasticity in the conditional earnings distribution with respect to schooling that appears not to have been present in previous decades. This pattern of heteroscedasticity is consistent with nonlinear human capital pricing models of the sort studied in Card and Lemieux (1996) and more nuanced multi-factor views of technical change reviewed in Acemoglu and Autor (2010).

5.1 Sensitivity Analysis

A natural concern is the extent to which some or all of the conclusions regarding the wage structure drawn under a missing at random assumption are compromised by limitations in the quality of Census earnings data. As Table 1 shows, the prevalence of earnings imputations increases steadily across Censuses with roughly a quarter of the observations allocated by 2000.[9] With these levels of missingness, the bounds on quantiles below the 25th percentile and above the 75th are not even finite in the absence of restrictions on the missingness process.

We begin by examining the sensitivity of conclusions regarding changes in the wage structure between 1990 and 2000. Figures 4 shows the 95% uniform confidence regions for the set $\mathcal{G}(k)$, as defined in (12), that result when we allow for a small amount of selection by setting $\mathcal{S}(F) \leq 0.05$. Though it remains clear the returns function increased between 1980 and 1990, we cannot reject the null hypothesis that the quantile process was unchanged from 1990 to 2000. Moreover, there is little evidence of heterogeneity across quantiles in the returns in any of the three Census samples –

[8] In constructing uniform confidence intervals we set $\omega_L(\tau, k) = \omega_U(\tau, k) = \phi(\Phi^{-1}(\tau))^{1/2}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and CDF . These weights are inversely proportional to the square root of the variance of the quantiles of a standard normal. The bootstrap weights $\{W_i\}_{i=1}^n$ were drawn from an exponential distribution.

[9] It is interesting to note that only 7% of the men in our sample report working no weeks in the past year. Hence, at least for this population, assumptions regarding the determinants of non-response appear to be more important for drawing conclusions regarding the wage structure than assumptions regarding non-participation in the labor force.

a straight line can be fit through each sample's confidence region.

To further assess the robustness of our conclusions regarding changes between 1980 and 1990, it is informative to find the level of k necessary to fail to reject the hypothesis that no change in fact occurred between these years under the supposition that $\mathcal{S}(F) \leq k$. Specifically, for $\pi_L^t(\tau, k)$ and $\pi_U^t(\tau, k)$ the lower and upper bounds on the returns coefficients in year t, we aim to obtain a confidence interval for the values of selection k under which:

$$\pi_U^{80}(\tau, k) \geq \pi_L^{90}(\tau, k) \quad \text{for all } \tau \in [0.2, 0.8] \ . \tag{41}$$

As in Example 2.2, we are particularly interested in k_0, the smallest value of k such that (41) holds, as it will hold trivially for all $k \geq k_0$. A search for the smallest value of k such that the 95% uniform confidence intervals for these two decades overlap at all quantiles between 0.2 and 0.8 found this "critical k" to be $\hat{k}_0 = 0.175$. Due to the independence of the samples between 1980 and 1990, the one-sided interval $[\hat{k}_0, 1]$ provides an asymptotic coverage probability for k_0 of at least 90%. The lower end of this confidence interval constitutes a large deviation from MAR (see Appendix A) indicating the evidence is quite strong that the returns process changed between 1980 and 1990. Figure 5 plots the uniform confidence regions corresponding to the hypothetical $\mathcal{S}(F) \leq \hat{k}_0$.

Though severe selection would be necessary for all of the changes between 1980 and 1990 to be spurious, it is clear that changes at some quantiles may be more robust than others. It is interesting then to conduct a more detailed analysis by evaluating the critical level of selection necessary to undermine the conclusion that the returns increased at each quantile. Towards this end, we generalize Example 2.3 and define $\kappa_0(\tau)$ to be the smallest level of k such that:

$$\pi_U^{80}(\tau, k) \geq \pi_L^{90}(\tau, k) \ . \tag{42}$$

The function $\kappa_0(\cdot)$ summarizes the level of robustness of each quantile-specific conclusion. In this manner, the "breakdown" function $\kappa_0(\cdot)$ reveals the differential sensitivity of the entire conditional distribution to violations of the missing at random assumption.

The point estimate for $\kappa_0(\tau)$ is given by the value of k where $\hat{\pi}_U^{80}(\tau, k)$ intersects with $\hat{\pi}_L^{90}(\tau, k)$ (see Figure 6). To obtain a confidence interval for $\kappa_0(\tau)$ that is uniform in τ we first construct 95% uniform two sided confidence intervals in τ and k for the 1980 upper bound $\pi_U^{80}(\tau, k)$ and the 1990 lower bound $\pi_L^{90}(\tau, k)$. Given the independence of the 1980 and 1990 samples, the intersection of the true bounds $\pi_U^{80}(\tau, k)$ and $\pi_L^{90}(\tau, k)$ must lie between the intersection of their corresponding confidence regions with asymptotic probability of at least 90%. Since $\kappa_0(\tau)$ is given by the intersection of $\pi_U^{80}(\tau, k)$ with $\pi_L^{90}(\tau, k)$, a valid lower bound for the confidence region of the function $\kappa_0(\cdot)$ is given by the intersection of the upper envelope for $\pi_U^{80}(\tau, k)$ with the lower envelope for $\pi_L^{90}(\tau, k)$ and a valid upper bound is given by the converse intersection.

Figure 7 illustrates the resulting estimates of the breakdown function $\kappa_0(\cdot)$ and its corresponding confidence region. Unsurprisingly, the most robust results are those for quantiles near the center of the distribution for which very large levels of selection would be necessary to overturn the hypothesis that the returns increased. However the curve is fairly asymmetric with the returns at low quantiles being much more sensitive to deviations from ignorability than those at the upper quantiles. Hence, changes in reporting behavior between 1980 and 1990 pose the greatest threat to hypotheses regarding changes at the bottom quantiles of the earnings distribution.

To conclude our sensitivity analysis we also consider the fitted values that result from the more flexible earnings model of Lemieux (2006) which allows for quadratic effects of education on earnings quantiles.[10] Figure 8 provides bounds on the 10th, 50th, and 90th conditional quantiles of weekly earnings by schooling level in 1980, 1990, and 2000 using our baseline hypothetical restriction $\mathcal{S}(F) \leq 0.05$. Little evidence exists of a change across Censuses in the real earnings of workers at the 10th conditional quantile. At the conditional median, however, the returns to schooling (which appear roughly linear) increased substantially, leading to an increase in inequality across schooling categories. Uneducated workers witnessed wage losses while skilled workers experienced wage gains, though in both cases these changes seem to have occurred entirely during the 1980s. Finally, we also note that, as observed by Lemieux (2006), the returns to schooling appear to have gradually convexified at the upper tail of the weekly earnings distribution with very well educated workers experiencing substantial gains relative to the less educated.

5.2 Estimates of the Degree of Selection in Earnings Data

Our analysis of Census data revealed that the finding of a change in the quantile specific returns to schooling process between 1990 and 2000 is easily undermined by small amounts of selection while changes between 1980 and 1990 (at least above the lower quantiles of the distribution) appear to be relatively robust. Employing a sample where validation data are present, we now turn to an investigation of what levels of selection, as indexed by $\mathcal{S}(F)$, are plausible in U.S. survey data.

In order to estimate $\mathcal{S}(F)$ we first derive an alternative representation of the distance between $F_{y|0,x}$ and $F_{y|1,x}$ which illustrates its dependence on the conditional probability of the outcome being missing. Towards this end, let us define the following conditional probabilities:

$$p_L(x,\tau) \equiv P(D=1|X=x, F_{y|x}(Y) \leq \tau) \tag{43}$$

$$p_U(x,\tau) \equiv P(D=1|X=x, F_{y|x}(Y) > \tau)\ . \tag{44}$$

By applying Bayes' Rule, it is then possible to express the distance between the distribution of missing and non-missing observations at a given quantile as a function of the selection probabilities:[11]

$$|F_{y|1,x}(q(\tau|x)) - F_{y|0,x}(q(\tau|x))| = \frac{\sqrt{(p_L(x,\tau)-p(x))(p_U(x,\tau)-p(x))\tau(1-\tau)}}{p(x)(1-p(x))}\ . \tag{45}$$

Notice that knowledge of the missing probability $P(D=0|X=x, F_{y|x}(Y)=\tau)$ is sufficient to compute by integration all of the quantities in (45) and (by taking the supremum over τ and x)

[10] The model also includes a quartic in potential experience. Our results differ substantively from those of Lemieux both because of differences in sample selection and our focus on weekly (rather than hourly) earnings.

[11] See Appendix B for a detailed derivation of (45).

of $\mathcal{S}(F)$ as well.[12] For this reason, our efforts focus on estimating this function in a dataset with information on the earnings of survey non-respondents.

We work with an extract from the 1973 March Current Population Survey (CPS) for which merged Internal Revenue Service (IRS) earnings data are available. Our sample consists of black and white men between the ages of 25 and 50 with six or more years of schooling who reported working at least one week in the past year and had valid IRS earnings. We drop observations with annual IRS earnings less than \$1,000 or equal to the IRS topcode of \$50,000.

As in our study of the Decennial Census, we take the relevant covariates to be age, years of schooling, and race. However, because our CPS sample is much smaller than our Census sample, we coarsen our covariate categories and drop demographic cells with fewer than 50 observations.[13] This yields an estimation sample of 13,598 observations distributed across 33 demographic cells. Because weeks worked are only measured categorically in this CPS extract we simply take log IRS earnings as our measure of Y and use response to the March CPS annual civilian earnings question as our measure of D. This yields a missingness rate of 8.4%.

We approximate the probability of non-response $P(D=0|X=x, F_{y|x}(Y)=\tau)$ with the following sequence of increasingly flexible logistic models:

$$P(D=0|X=x, F_{y|x}(Y)=\tau) = \Lambda(b_1\tau + b_2\tau^2 + \delta_x) \quad \text{(M1)}$$
$$P(D=0|X=x, F_{y|x}(Y)=\tau) = \Lambda(b_1\tau + b_2\tau^2 + \gamma_1\delta_x\tau + \gamma_2\delta_x\tau^2 + \delta_x) \quad \text{(M2)}$$
$$P(D=0|X=x, F_{y|x}(Y)=\tau) = \Lambda(b_{1,x}\tau + b_{2,x}\tau^2 + \delta_x) \quad \text{(M3)}$$

where $\Lambda(\cdot) = \exp(\cdot)/(1+\exp(\cdot))$ is the Logistic CDF. These models differ primarily in the degree of demographic bin heterogeneity allowed for in the relationship between earnings and the probability of responding to the CPS. Model M1 relies entirely on the nonlinearities in the index function $\Lambda(\cdot)$ to capture heterogeneity across cells in the response profiles. The model M2 allows for additional heterogeneity through the interaction coefficients (γ_1, γ_2) but restricts these interactions to be linear in the cell fixed effect δ_x. Finally, M3, which is equivalent to a cell specific version of M1, places no restrictions across demographic groups on the shape of the response profile.

Maximum likelihood estimates from the three models are presented in Table 2.[14] A comparison of the model log likelihoods reveals that the introduction of the interaction terms (γ_1, γ_2) in Model 2 yields a substantial improvement in fit over the basic separable logit of Model 1 despite the insignificance of the resulting parameter estimates. However, the restrictions of the linearly interacted Model 2 cannot be rejected relative to its fully interacted generalization in Model 3 which appears

[12] Note that $P(D=0, F_{y|x}(Y) \leq \tau | X=x) = \int_0^\tau P(D=0|F_{y|x}(Y)=u, X=x)du$ because $F_{y|x}(Y)$ is uniformly distributed on $[0,1]$ conditional on $X=x$. Thus $p_L(x,\tau) = \int_0^\tau P(D=0|F_{y|x}(Y)=u, X=x)du/\tau$. Likewise $p_U(x,\tau) = \int_\tau^1 P(D=0|F_{y|x}(Y)=u, X=x)du/(1-\tau)$ and $p(x) = \int_0^1 P(D=0|F_{y|x}(Y)=u, X=x)du$.

[13] We use five-year age categories instead of single digit ages and collapse years of schooling into four categories: <12 years of schooling, 12 years of schooling, 13-15 years of schooling, and 16+ years of schooling.

[14] We use the respondent's sample quantile in his demographic cell's distribution of Y as an estimate of $F_{y|x}(Y)$. It can be shown that sampling errors in the estimated quantiles have asymptotically negligible effects on the limiting distribution of the parameter estimates.

to be substantially overfit.

A Wald test of joint significance of the earned income terms (b_1, b_2) in the first model easily rejects the null hypothesis that the data are missing at random. Evidently, missingness follows a U-shaped response pattern with very low and very high wage men least likely to provide valid earnings information – a pattern conjectured (but not directly verified) by Lillard et al. (1986). This pattern is also found in the two more flexible logit models as illustrated in the third panel of the table which provides the average marginal effects of earnings evaluated at three quantiles of the distribution. These average effects are consistently negative at $\tau = 0.2$ and positive at $\tau = 0.8$. It is important to note however that Models 2 and 3 allow for substantial heterogeneity across covariate bins in these marginal effects which in some cases yields response patterns that are monotonic rather than U-shaped.

It is straightforward to estimate the distance between missing and nonmissing earnings distributions in each demographic bin by integrating our estimates of $P(D = 0|X = x, F_{y|x}(Y) = \tau)$ across the relevant quantiles of interest. We implement this integration numerically via one dimensional Simpson quadrature. The bottom panel of Table 2 shows quantiles of the distribution of resulting cell specific KS distance estimates. Model 1 is nearly devoid of heterogeneity in KS distances across demographic bins because of the additive separability implicit in the model. Model 2 yields substantially more heterogeneity with a minimum KS distance of 0.02 and a maximum distance $\mathcal{S}(F)$ of 0.12. Finally, Model 3, which we suspect has been overfit, yields a median KS distance of 0.11 and an enormous maximum KS distance of 0.44.

Figure 9 provides a visual representation of our estimates from Model 2 of the underlying distance functions $|F_{y|1,x}(q(\tau|x)) - F_{y|0,x}(q(\tau|x))|$ in each of the 33 demographic bins in our sample. The outer envelope of these functions corresponds to the quantile specific level of selection considered in the breakdown analysis of Figure 7, while the maximum point on the envelope corresponds to $\mathcal{S}(F)$. Note that while some of the distance functions exhibit an unbroken inverted U shaped pattern others exhibit double or even triple arches. The pattern of multiple arches occurs when the CDFs are estimated to have crossed at some quantile which yields a distance of zero at that point. A quadratic relationship between missingness and earnings can easily yield such patterns. Because of the interactions in Model 2, some cells exhibit effects that are not quadratic and tend to generate CDFs exhibiting first order stochastic dominance. It is interesting to note that the demographic cell obtaining the maximum KS distance of 0.12 corresponds to young (age 25-30), black, high school dropouts for whom more IRS earnings are estimated to monotonically increase the probability of responding to the CPS earnings question. This leads to a distribution of observed earnings which stochastically dominates that of the corresponding unobserved earnings.

Our estimates of selection in Figure 9, when compared to the breakdown function of Figure 7, reinforce our earlier conclusion that most of the apparent changes in wage structure between 1980 and 1990 are robust to plausible violations of MAR but that conclusions regarding lower quantiles could be overturned by selective non-response. Likewise, the apparent emergence of heterogeneity in the returns function in 2000, may easily be justified by selection of the magnitude found in our CPS sample. Though our estimates of selection are fairly sensitive to the manner in which cell specific heterogeneity is modeled, we take the patterns in Table 2 and Figure 9 as suggestive evidence that small, but by no means negligible, deviations from missing at random are likely present in modern earnings data. These deviations may yield complicated discrepancies between observed and missing CDFs about which it is hard to develop strong priors. We leave it to future research to examine these issues more carefully with additional validation datasets. Given however the likely absence of such prior knowledge for most prospective studies, we expect the sensitivity techniques developed in this paper to be quite useful for applied research.

6 Conclusion

We have proposed assessing the sensitivity of estimates of conditional quantile functions with missing outcome data to violations of the MAR assumption by considering the minimum level of selection, as indexed by the maximal KS distance between the distribution of missing and nonmissing outcomes across all covariate values, necessary to overturn conclusions of interest. Inferential methods were developed that account for uncertainty in estimation of the nominal identified set and that acknowledge the potential for model misspecification. We found in an analysis of U.S. Census data that the well-documented increase in the returns to schooling between 1980 and 1990 is relatively robust to alternative assumptions on the missing process, but that conclusions regarding heterogeneity in returns and changes in the returns function between 1990 and 2000 are very sensitive to departures from ignorability.

While we have focused on methods for gauging sensitivity to non-response in cross-sectional datasets, a number of interesting extensions are possible. An obvious (and important) one is an adaptation to environments where missingness arises due to non-participation in the labor force as in Heckman (1974) and Blundell et al. (2007). Additional side restrictions may be appropriate here, particularly in a panel data setting of the sort studied by Johnson et al. (2000) and Neal (2004). Another extension is to treatment effects where, again, researchers may wish to combine our nominal KS restriction with additional restrictions of the sort studied by Lee (2009) or (if outcomes are discrete) those of Shaikh and Vytlacil (2005) and Bhattacharya et al. (2008). Adding restrictions will shrink the nominal identified set and, in general, reduce sensitivity, but will also substantially complicate inference. We leave the development of such methods to future work.

References

Acemoglu, D. and Autor, D. H. (2010). Skills, tasks, and technology: Implications for employment and earnings. In *Handbook of Labor Economics*, vol. 4. Elsevier, Forthcoming.

Altonji, J. G., Elder, T. and Taber, C. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, **113** 151–184.

Altonji, J. G., Elder, T. and Taber, C. (2008). Using selection on observed variables to assess bias from unobservables when evaluating swan-ganz catheterization. *American Economic Review, Paper and Proceedings*, **98** 145–350.

Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, **74** 539–563.

Autor, D. H. and Katz, L. F. (1999). Changes in the wage structure and earning inequality. In *Handbook of Labor Economics*, vol. 3A. North-Holland, 1463–1555.

Bhattacharya, J., Shaikh, A. and Vytlacil, E. (2008). Treatment effect bounds under monotonicity assumptions: An application to swan-ganz catheterization. *American Economic Review, Papers and Proceedings*, **98** 351–356.

Blundell, R., Gosling, A., Ichimura, H. and Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, **75** 323–363.

Card, D. and Lemieux, T. (1996). Wage dispersion, returns to skill, and black-white wage differentials. *Journal of Econometrics*, **74** 319–361.

Chamberlain, G. (1994). Quantile regression, censoring and the structure of wages. In *Advances in Econometrics, Sixth World Congress* (C. Sims, ed.). Elsevier, 171–209.

DiNardo, J., McCrary, J. and Sanbonmatsu, L. (2006). Constructive proposals for dealing with attrition: An empirical example. Working paper, University of Michigan.

Greenlees, J. S., Reece, W. S. and Zieschang, K. D. (1982). Imputations of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, **77** 251–261.

Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, **42** 679–694.

Heckman, J. J., Lochner, L. and Todd, P. (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. In *Handbook of Education Economics* (E. Hanushek and F. Welch, eds.), vol. 1, chap. 7. Elsevier.

Hirsch, B. T. and Schumacher, E. J. (2004). Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics*, **22** 689–722.

Horowitz, J. L. and Manski, C. F. (2006). Identification and estimation of statistical functionals using incomplete data. *Journal of Econometrics*, **132** 445–459.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review, Papers and Proceedings*, **93** 126–132.

Johnson, W., Neal, D. and Kitamura, Y. (2000). Evaluating a simple method for estimating black-white gaps in median wages. *American Economic Review, Papers and Proceedings*, **90** 339–343.

Juhn, C., Murphy, K. M. and Pierce, B. (1993). Wage inequality and the rise in the return to skill. *Journal of Political Economy*, **101** 410–442.

Lee, D. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, **76** 1071–1102.

Lemieux, T. (2006). Postsecondary education and increasing wage inequality. *The American Economic Review*, **96** 195–199.

Lillard, L., Smith, J. P. and Welch, F. (1986). What do we really know about wages? the importance of nonreporting and census imputation. *The Journal of Political Economy*, **94** 489–506.

Manski, C. F. (1994). The selection problem. In *Advances in Econometrics, Sixth World Congress* (C. Sims, ed.). Cambridge University Press, Cambridge, UK, 143–170.

Manski, C. F. (2003). *Partial Identification of Probability Distributions.* Springer-Verlag, New York.

Neal, D. (2004). The measured black-white wage gap is too small. *Journal of Political Economy*, **112** S1–S28.

Ponomareva, M. and Tamer, E. (2009). Misspecification in moment inequality models: Back to moment equalities? Working paper, Northwestern University.

Rosenbaum, P. (2002). *Observational Studies.* Springer, New York.

Rosenbaum, P. and Rubin, D. (1983). Assessing sensitivity to an unobserved binary covarite in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B*, **45** 212–218.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* 581–592.

Shaikh, A. M. and Vytlacil, E. (2005). Threshold crossing models and bounds on treatment effects: A nonparametric analysis. Working paper, University of Chicago.

Stoye, J. (2007). Bounds on generalized linear predictors with incomplete outcome data. *Reliable Computing*, **13** 293–302.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics.* Springer, New York.

Appendix A - The bivariate normal selection model and KS distance

To develop intuition for our metric $\mathcal{S}(F)$ of deviations from missing at random we provide here a mapping between the parameters of a standard bivariate selection model, the resulting CDFs of observed and missing outcomes, and the implied values of $\mathcal{S}(F)$. Using the notation of Section 2, our DGP of interest is:

$$(Y_i, v_i) \sim N(0, \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)) \qquad D_i = 1\{\mu + v_i > 0\} \ . \tag{46}$$

In this model, the parameter ρ indexes the degree of non-ignorable selection in the outcome variable Y_i. We choose $\mu = .6745$ to ensure a missing fraction of 25% which is approximately the degree of missingness found in our analysis of earnings data in the US Census. We computed the distributions of missing and observed outcomes for various values of ρ by simulation, some of which are plotted in Figures A.1 and A.2. The resulting values of $\mathcal{S}(F)$, which correspond to the maximum vertical distance between the observed and missing CDFs across all points of evaluation, are given in the table below:

Table A.1: $\mathcal{S}(F)$ as a function of ρ

ρ	$\mathcal{S}(F)$	ρ	$\mathcal{S}(F)$	ρ	$\mathcal{S}(F)$
0.05	0.0337	0.35	0.2433	0.65	0.4757
0.10	0.0672	0.40	0.2778	0.70	0.5165
0.15	0.1017	0.45	0.3138	0.75	0.5641
0.20	0.1355	0.50	0.3520	0.80	0.6158
0.25	0.1721	0.55	0.3892	0.85	0.6717
0.30	0.2069	0.60	0.4304	0.90	0.7377

Figure A.1: Missing and Observed Outcome CDFs

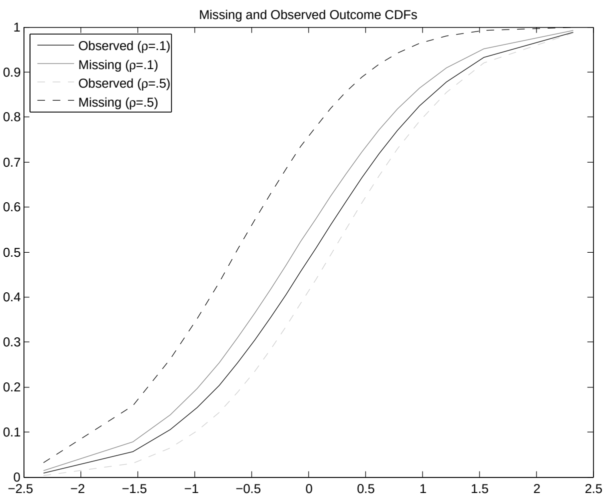
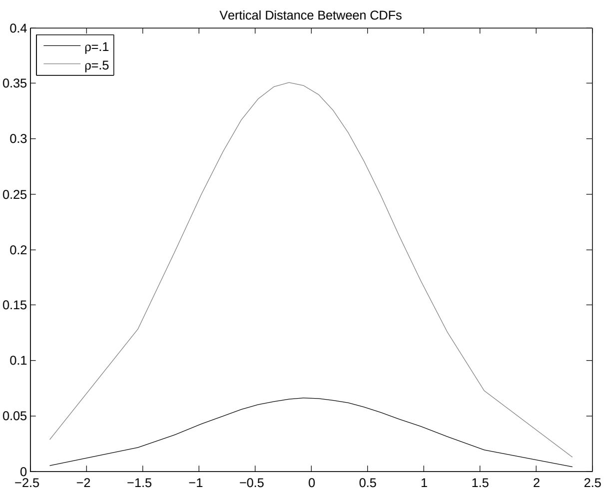


Figure A.2: Vertical Distance Between CDFs



Appendix B - Derivations of Section 5.2

The following Appendix provides a justification for the derivations in Section 5.2, in particular of the representation derived in equation (45). Towards this end, observe first that by Bayes' rule:

$$\begin{aligned} F_{y|1,x}(c) &= \frac{P(D=1|X=x, Y \leq c) \times F_{y|x}(c)}{p(x)} \\ &= \frac{P(D=1|X=x, F_{y|x}(Y) \leq F_{y|x}(c)) \times F_{y|x}(c)}{p(x)} , \end{aligned} \tag{47}$$

where the second equality follows from $F_{y|x}$ being strictly increasing. Evaluating (47) at $c = q(\tau|x)$, employing the definition of $p_L(x, \tau)$ in (43), and noting that $F_{y|x}(q(\tau|x)) = \tau$ yields:

$$F_{y|1,x}(q(\tau|x)) = \frac{p_L(\tau, x) \times \tau}{p(x)} . \tag{48}$$

Moreover, by identical arguments, but working instead with the definition of $p_U(\tau, x)$, we derive:

$$1 - F_{y|1,x}(q(\tau|x)) = \frac{P(D=1|Y > q(\tau|x), X=x) \times (1 - F_{y|1,x}(q(\tau|x)))}{p(x)} = \frac{p_U(\tau, x) \times (1-\tau)}{p(x)} \tag{49}$$

Finally, we note that the same manipulations applied to $F_{y|0,x}$ instead of $F_{y|1,x}$ enable us to obtain:

$$F_{y|0,x}(q(\tau|x)) = \frac{(1 - p_L(\tau, x)) \times \tau}{1 - p(x)} \qquad 1 - F_{y|0,x}(q(\tau|x)) = \frac{(1 - p_U(\tau, x)) \times (1-\tau)}{1 - p(x)} . \tag{50}$$

Hence, we can obtain by direct algebra from the results (47) and (50) that we must have:

$$|F_{y|1,x}(q(\tau|x)) - F_{y|0,x}(q(\tau|x))| = \frac{|p(x) - p_L(x, \tau)| \times \tau}{p(x)(1 - p(x))} . \tag{51}$$

Analogously, exploiting (47) and (50) once again, we can also obtain:

$$\begin{aligned} |F_{y|1,x}(q(\tau|x)) - F_{y|0,x}(q(\tau|x))| &= |(1 - F_{y|1,x}(q(\tau|x))) - (1 - F_{y|0,x}(q(\tau|x)))| \\ &= \frac{|p(x) - p_U(x, \tau)| \times (1-\tau)}{p(x)(1 - p(x))} . \end{aligned} \tag{52}$$

The desired equality in (45) then follows immediately from taking the square root of the product of (51) and (52).

Appendix C - Proof of Results

Lemma 6.1. *Under Assumptions 2.1(ii)-(iii), if $\mathcal{S}(F) \leq k$, then the nominal identified set $\mathcal{C}(\tau, k)$ equals:*

$$\{\theta : \mathcal{X} \to \mathbf{R} : \tau - \min\{\tau + kp(x), 1\} \times \{1 - p(x)\} \leq F_{y|1,x}(\theta(x))p(x) \leq \tau - \max\{\tau - kp(x), 0\} \times \{1 - p(x)\}\}$$

Proof of Lemma 6.1: Letting $KS(F_{y|1,x}, F_{y|0,x}) = \sup_c |F_{y|1,x}(c) - F_{y|0,x}(c)|$, we first observe that:

$$\begin{aligned} KS(F_{y|1,x}, F_{y|0,x}) &= \frac{1}{p(x)} \times \sup_{c \in \mathbf{R}} |F_{y|1,x}(c) \times p(x) + F_{y|0,x}(c) \times \{1 - p(x)\} - F_{y|0,x}(c)| \\ &= \frac{1}{p(x)} \times \sup_{c \in \mathbf{R}} |F_{y|x}(c) - F_{y|0,x}(c)| \ . \end{aligned} \tag{53}$$

Therefore, if $\theta(x) = c_\tau(x)$, then it immediately follows from the hypothesis $\mathcal{S}(F) \leq k$ and result (53) that:

$$\begin{aligned} \tau &= F_{y|1,x}(\theta(x)) \times p(x) + F_{y|0,x}(\theta(x)) \times \{1 - p(x)\} \\ &\leq F_{y|1,x}(\theta(x)) \times p(x) + \min\{F_{y|x}(\theta(x)) + kp(x), 1\} \times \{1 - p(x)\} \\ &= F_{y|1,x}(\theta(x)) \times p(x) + \min\{\tau + kp(x), 1\} \times \{1 - p(x)\} \ . \end{aligned} \tag{54}$$

By identical manipulations, $F_{y|1,x}(\theta(x)) \times p(x) \leq \tau - \max\{\tau - kp(x), 0\} \times \{1 - p(x)\}$ and hence $\theta \in \mathcal{C}(\tau, k)$. To prove the bounds are sharp, let $\theta \in \mathcal{C}(\tau, k)$ and define the function $\kappa : \mathcal{X} \to \mathbf{R}$ by:

$$\kappa(x) \equiv \frac{\tau - F_{y|1,x}(\theta(x)) \times p(x)}{1 - p(x)} \ . \tag{55}$$

Further observe that by virtue of $\theta \in \mathcal{C}(\tau, k)$, the following two inequalities hold uniformly in $x \in \mathcal{X}$:

$$\max\{\tau - kp(x), 0\} \leq \kappa(x) \leq \min\{\tau + kp(x), 1\} \qquad |\kappa(x) - F_{y|1,x}(\theta(x))| \leq k \ . \tag{56}$$

We now aim to construct a distribution for Y conditional on X and Y being missing such that all assumptions are met and in addition θ is the conditional quantile of Y given X. Define:

$$\begin{aligned} \tilde{F}^{+}_{y|0,x}(c) \equiv& 1\{c \geq \theta(x)\} \times \max\{F_{y|1,x}(c),\ \min\{\frac{1}{2}(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x), 1\}\} \\ &+ 1\{c < \theta(x)\} \times \max\{F_{y|1,x}(c),\ 2(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x)\} \\ \tilde{F}^{-}_{y|0,x}(c) \equiv& 1\{c \geq \theta(x)\} \times \min\{F_{y|1,x}(c),\ 2(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x)\} \\ &+ 1\{c < \theta(x)\} \times \min\{F_{y|1,x}(c),\ \max\{\frac{1}{2}(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x), 0\}\} \end{aligned} \tag{57}$$

and let the distribution of Y conditional on X and Y being unobservable be pointwise given by:

$$\tilde{F}_{y|0,x}(c) \equiv 1\{\kappa(x) \geq F_{y|1,x}(\theta(x))\} \times \tilde{F}^{+}_{y|0,x}(c) + 1\{\kappa(x) < F_{y|1,x}(\theta(x))\} \times \tilde{F}^{-}_{y|1,x}(c) \ . \tag{58}$$

Note that $\tilde{F}_{y|0,x}(c)$ is strictly increasing and continuous at all c such that $0 < F_{y|0,x}(c) < 1$ by virtue of $F_{y|1,x}(c)$ being strictly increasing and continuous. Since $\tilde{F}_{y|0,x}$ is bounded between zero and one, we conclude it is a properly defined cdf. Denoting $\tilde{F}_{y|x}(c) = F_{y|1,x}(c) \times p(x) + \tilde{F}_{y|0,x}(c) \times \{1 - p(x)\}$, we obtain:

$$\tilde{F}_{y|x}(\theta(x)) = F_{y|1,x}(\theta(x)) \times p(x) + \tilde{F}_{y|0,x}(\theta(x)) \times \{1 - p(x)\} = F_{y|1,x}(\theta(x)) \times p(x) + \kappa(x) \times \{1 - p(x)\} = \tau \ , \tag{59}$$

so that $\theta(x)$ is the conditional τ-quantile of Y given X. In addition, by construction and (56) we have:

$$\sup_{c\in\mathbf{R}} |\tilde{F}_{y|0,x}(c) - F_{y|1,x}(c)| = |\tilde{F}_{y|0,x}(\theta(x)) - F_{y|1,x}(\theta(x))| \leq k \ , \tag{60}$$

uniformly in $x \in \mathcal{X}$. It follows that $\mathcal{S}(F) \leq k$ and hence conclude the bounds are sharp as claimed. ■

Proof of Lemma 2.1: Follows immediately from Lemma 6.1. ■

Proof of Lemma 3.1: For any $\theta \in \mathcal{C}(\tau, k)$, the first order condition of the norm minimization problem yields $\beta(\tau) = (E_S[X_iX_i'])^{-1}E_S[X_i\theta(X_i)]$. The Lemma then follows from Corollary 2.1. ■

Proof of Corollary 3.1: Since $\mathcal{P}(\tau, k)$ is convex by Lemma 3.1, it follows that the identified set for $\lambda'\beta(\tau)$ is a convex set in $\mathbf{R}$ and hence an interval. The fact that $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$ are the endpoints of such interval follows directly from Lemma 3.1. ■

Lemma 6.2. *Let Assumption 2.1 hold, W_i be independent of (Y_i, X_i, D_i) with $E[W_i] = 1$ and $E[W_i^2] < \infty$ and positive almost surely. If $\{Y_i, X_i, D_i, W_i\}$ is an i.i.d. sample, then the following class is Donsker:*

$$\mathcal{M} \equiv \{m_c : m_c(y, x, d, w) \equiv w1\{y \leq c, d = 1, x = x_0\} - P(Y_i \leq c,\ D_i = 1,\ X_i = x_0),\ c \in \mathbf{R}\} \ .$$

Proof: For any $1 > \epsilon > 0$, by Assumption 2.1(ii) there is an increasing sequence $\{y_0, \ldots, y_{\lceil \frac{4}{\epsilon} \rceil}\}$ such that for $\{[y_{j-1}, y_j]\}_{i=1}^{\lceil \frac{8}{\epsilon} \rceil}$ partitions $\mathbf{R}$ and for any $1 \leq j \leq \lceil \frac{8}{\epsilon} \rceil$ we have $F_{y|1,x}(y_j) - F_{y|1,x}(y_{j-1}) < \epsilon/4$. Let

$$l_j(y, x, d, w) \equiv w1\{y \leq y_{j-1},\ d = 1,\ x = x_0\} - P(Y_i \leq y_j,\ D_i = 1,\ X_i = x_0) \tag{61}$$

$$u_j(y, x, d, w) \equiv w1\{y \leq y_j,\ d = 1,\ x = x_0\} - P(Y_i \leq y_{j-1},\ D_i = 1,\ X_i = x_0) \tag{62}$$

and notice the brackets $\{[l_j, u_j]\}_{j=1}^{\lceil \frac{8}{\epsilon} \rceil}$ form a partition of the class $\mathcal{M}_c$ (since $w \in \mathbf{R}_+$). In addition, note:

$$\begin{aligned} &E[(u_j(Y_i, X_i, D_i, W_i) - l_j(Y_i, X_i, D_i, W_i))^2] \\ &\quad \leq 2E[W_i^2 1\{y_{j-1} \leq Y_i \leq y_j,\ D_i = 1,\ X_i = x_0\}] + 2P^2(y_{j-1} \leq Y_i \leq y_j,\ D_i = 1,\ X_i = x_0) \\ &\quad \leq 4E[W_i^2] \times (F_{y|1,x}(y_j) - F_{y|1,x}(y_{j-1})) \ , \end{aligned} \tag{63}$$

and hence each bracket has norm bounded by $\sqrt{E[W_i^2]\epsilon}$. Therefore, $N_{[\,]}(\epsilon, \mathcal{M}, \|\cdot\|_{L^2}) \leq 16E[W_i^2]/\epsilon^2$, and the Lemma follows by Theorem 2.5.6 in van der Vaart and Wellner (1996). ■

Lemma 6.3. *Let Assumption 2.1 hold, W_i be independent of (Y_i, X_i, D_i) with $E[W_i] = 1$, $E[W_i^2] < \infty$ and positive almost surely. Also let $\mathcal{S} \equiv \{(\tau, b) \in [0, 1]^2 : b\{1 - p(x)\} + \epsilon \leq \tau \leq p(x) + b\{1 - p(x)\} - \epsilon\ \forall x \in \mathcal{X}\}$ for some ϵ satisfying $0 < 2\epsilon < \inf_{x\in\mathcal{X}} p(x)$ and denote the minimizers:*

$$s_0(\tau, b, x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b) \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b) \ . \tag{64}$$

Then $s_0(\tau, b, x)$ is bounded in $(\tau, b, x) \in \mathcal{S} \times \mathcal{X}$ and if $\{Y_i, X_i, D_i, W_i\}$ is i.i.d. then for some $M > 0$:

$$P\Big(\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\hat{s}_0(\tau, b, x)| > M\Big) = o(1) \ .$$

Proof: First note that Assumption 2.1(ii) implies $s_0(\tau, b, x)$ is uniquely defined, while $\hat{s}_0(\tau, b, x)$ may be one of multiple minimizers. By Assumption 2.1(ii) and the definition of $\mathcal{S}$, it follows that the equality:

$$P(Y_i \le s_0(\tau, b, x), D_i = 1 | X_i = x) = \tau - bP(D_i = 0 | X_i = x) \tag{65}$$

implicitly defines $s_0(\tau, b, x)$. Let $\bar{s}(x)$ and $\underline{s}(x)$ be the unique numbers satisfying $F_{y|1,x}(\bar{s}(x)) \times p(x) = p(x) - \epsilon$ and $F_{y|1,x}(\underline{s}(x)) \times p(x) = \epsilon$. By result (65) and the definition of the set $\mathcal{S}$ we then obtain that for all $x \in \mathcal{X}$:

$$-\infty < \underline{s}(x) \le \inf_{(\tau,b)\in\mathcal{S}} s_0(\tau, b, x) \le \sup_{(\tau,b)\in\mathcal{S}} s_0(\tau, b, x) \le \bar{s}(x) < +\infty \ . \tag{66}$$

Hence, we conclude that $\sup_{(\tau,b)\in\mathcal{S}} |s_0(\tau, b, x)| = O(1)$ and the first claim follows by $\mathcal{X}$ being finite.

In order to establish the second claim of the Lemma, we define the functions:

$$R_x(\tau, b) \equiv bP(D_i = 0, X_i = x) - \tau P(X_i = x) \tag{67}$$

$$R_{x,n}(\tau, b) \equiv \frac{1}{n}\sum_{i=1}^{n} W_i(b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\}) \tag{68}$$

as well as the maximizers and minimizers of $R_{x,n}(\tau, b)$ on the set $\mathcal{S}$, which we denote by:

$$(\underline{\tau}_n(x), \underline{b}_n(x)) \in \arg\max_{(\tau,b)\in\mathcal{S}} R_{x,n}(\tau, b) \qquad (\bar{\tau}_n(x), \bar{b}_n(x)) \in \arg\min_{(\tau,b)\in\mathcal{S}} R_{x,n}(\tau, b) \ . \tag{69}$$

Also denote the set of maximizers and minimizers of $\tilde{Q}_{x,n}(c|\tau, b)$ at these particular choices of (τ, b) by:

$$\underline{S}_n(x) \equiv \Big\{\underline{s}_n(x) \in \mathbf{R} : \underline{s}_n(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\underline{\tau}_n(x), \underline{b}_n(x))\Big\} \tag{70}$$

$$\bar{S}_n(x) \equiv \Big\{\bar{s}_n(x) \in \mathbf{R} : \bar{s}_n(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\bar{\tau}_n(x), \bar{b}_n(x))\Big\} \tag{71}$$

From the definition of $\tilde{Q}_{x,n}(c|\tau, b)$, we then obtain from (69), (70) and (71) that for all $x \in \mathcal{X}$:

$$\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) \le \inf_{(\tau,b)\in\mathcal{S}} \hat{s}_0(\tau, b, x) \le \sup_{(\tau,b)\in\mathcal{S}} \hat{s}_0(\tau, b, x) \le \sup_{\bar{s}_n(x)\in\bar{S}_n(x)} \bar{s}_n(x) \ . \tag{72}$$

We establish the second claim of the Lemma, by exploiting (72) and showing that for some $0 < M < \infty$:

$$P\Big(\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) < -M\Big) = o(1) \qquad P\Big(\sup_{\bar{s}_n(x)\in\bar{S}_n(x)} \bar{s}_n(x) > M\Big) = o(1) \ . \tag{73}$$

To prove that $\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x)$ is larger than $-M$ with probability tending to one, note that:

$$|R_{x,n}(\underline{\tau}_n(x), \underline{b}_n(x)) + \epsilon P(X_i = x)| = |R_{x,n}(\underline{\tau}_n(x), \underline{b}_n(x)) - \max_{(\tau,b)\in\mathcal{S}} R_x(\tau, b)| = o_p(1) \ , \tag{74}$$

where the second equality follows from the Theorem of the Maximum and the continuous mapping theorem. Therefore, using the equality $a^2 - b^2 = (a - b)(a + b)$, result (74) and Lemma 6.2, it follows that:

$$\sup_{c\in\mathbf{R}} |\tilde{Q}_{x,n}(c|\underline{\tau}_n(x), \underline{b}_n(x)) - (F_{y|1,x}(c)p(x) - \epsilon)^2 P^2(X_i = x)| = o_p(1) \ . \tag{75}$$

Fix $\delta > 0$ and note that since $F_{y|1,x}(\underline{s}(x))p(x) = \epsilon$ and $\epsilon/p(x) < 1$, Assumption 2.1(ii) implies that:

$$\eta \equiv \inf_{|c-\underline{s}(x)|>\delta} (F_{y|1,x}(c)p(x) - \epsilon)^2 > 0 \ . \tag{76}$$

Therefore, it follows from direct manipulations and the definition of $\underline{S}_n(x)$ in (70) and of $\underline{s}(x)$ that:

$$P\Big(|\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) - \underline{s}(x)| > \delta\Big) \le P\Big(\inf_{|c-\underline{s}(x)|>\delta} \tilde{Q}_{x,n}(c|\underline{\tau}_n(x), \underline{b}_n(x)) \le \tilde{Q}_{x,n}(\underline{s}(x)|\underline{\tau}_n(x), \underline{b}_n(x))\Big)$$
$$\le P\Big(\eta \le \sup_{c\in\mathbf{R}} 2|\tilde{Q}_{x,n}(c|\underline{\tau}_n(x), \underline{b}_n(x)) - (F_{y|1,x}(c)p(x) - \epsilon)^2 P^2(X_i = x)|\Big) \ .$$

We hence conclude from (75) that $\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) \xrightarrow{p} \underline{s}(x)$, which together with (66) implies that $\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x)$ is larger than $-M$ with probability tending to one for some $M > 0$. By similar arguments it can be shown that $\sup_{\bar{s}_n(x)\in\bar{S}_n(x)} \bar{s}_n(x) \xrightarrow{p} \bar{s}(x)$ which together with (66) establishes (73). The second claim of the Lemma then follows from (72), (73) and $\mathcal{X}$ being finite. ■

Lemma 6.4. *Let Assumption 2.1 hold, W_i be independent of (Y_i, X_i, D_i) with $E[W_i] = 1$, $E[W_i^2] < \infty$ and positive almost surely. Also let $\mathcal{S} \equiv \{(\tau, b) \in [0,1]^2 : b\{1 - p(x)\} + \epsilon \le \tau \le p(x) + b\{1 - p(x)\} - \epsilon\ \forall x \in \mathcal{X}\}$ for some ϵ satisfying $0 < 2\epsilon < \inf_{x\in\mathcal{X}} p(x)$ and denote the minimizers:*

$$s_0(\tau, b, x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b) \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b) \ . \tag{77}$$

If $\{Y_i, X_i, D_i, W_i\}$ is an i.i.d. sample, then $\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\hat{s}_0(\tau, b, x) - s_0(\tau, b, x)| = o_p(1)$.

Proof: First define the criterion functions $M : L^\infty(\mathcal{S} \times \mathcal{X}) \to \mathbf{R}$ and $M_n : L^\infty(\mathcal{S} \times \mathcal{X}) \to \mathbf{R}$ by:

$$M(\theta) \equiv \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} Q_x(\theta(\tau, b, x)|\tau, b) \qquad M_n(\theta) \equiv \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \tilde{Q}_{x,n}(\theta(\tau, b, x)|\tau, b) \ . \tag{78}$$

For notational convenience, let $s_0 \equiv s_0(\cdot,\cdot,\cdot)$ and $\hat{s}_0 \equiv \hat{s}_0(\cdot,\cdot,\cdot)$. By Lemma 6.3, $s_0 \in L^\infty(\mathcal{S} \times \mathcal{X})$ while with probability tending to one $\hat{s}_0 \in L^\infty(\mathcal{S} \times \mathcal{X})$. Hence, (77) implies that with probability tending to one:

$$\hat{s}_0 \in \arg\min_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})} M_n(\theta) \qquad s_0 = \arg\min_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})} M(\theta) \ . \tag{79}$$

By Assumption 2.1(ii) and (65), $Q_x(c|\tau, b)$ is strictly convex in a neighborhood of $s_0(\tau, b, x)$. Furthermore, since by (65) and the implicit function theorem $s_0(\tau, b, x)$ is continuous in $(\tau, b) \in \mathcal{S}$ for every $x \in \mathcal{X}$:

$$\inf_{\|\theta - s_0\|_\infty \ge \delta} M(\theta) \ge \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \inf_{|c - s_0(\tau,b,x)|\ge\delta} Q_x(c|\tau, b)$$
$$= \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \min\{Q_x(s_0(\tau, b, x) - \delta|\tau, b), Q_x(s_0(\tau, b, x) + \delta|\tau, b)\} > 0 \ , \tag{80}$$

where the final inequality follows by compactness of $\mathcal{S}$ which together with continuity of $s_0(\tau, b, x)$ implies the inner infimum is attained, while the outer infimum is trivially attained due to $\mathcal{X}$ being finite. Since (80) holds for any $\delta > 0$, s_0 is a well separated minimum of $M(\theta)$ in $L^\infty(\mathcal{S} \times \mathcal{X})$. Next define:

$$G_{x,i}(c) \equiv W_i 1\{Y_i \le c, D_i = 1, X_i = x\} \tag{81}$$

and observe that compactness of $\mathcal{S}$, a regular law of large numbers, Lemma 6.2 and finiteness of $\mathcal{X}$ yields:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \sup_{c\in\mathbf{R}} |\frac{1}{n}\sum_{i=1}^n G_{x,i}(c) + R_{x,n}(\tau, b) - E[G_{x,i}(c)] - R_x(\tau, b)|$$
$$\le \sup_{x\in\mathcal{X}} \sup_{c\in\mathbf{R}} |\frac{1}{n}\sum_{i=1}^n G_{x,i}(c) - E[G_{x,i}(c)]| + \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |R_{x,n}(\tau, b) - R_x(\tau, b)| = o_p(1) \ , \tag{82}$$

where $R_x(\tau,b)$ and $R_{x,n}(\tau,b)$ are as in (67) and (68) respectively. Hence, using (82), the equality $a^2-b^2=(a-b)(a+b)$ and $Q_x(c|\tau,b)$ uniformly bounded in $(c,\tau,b)\in\mathbf{R}\times\mathcal{S}$ due to the compactness of $\mathcal{S}$, we obtain:

$$\sup_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})}|M_n(\theta)-M(\theta)|\le\sup_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})}\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}|\tilde{Q}_{x,n}(\theta(\tau,b,x)|\tau,b)-Q_x(\theta(\tau,b,x)|\tau,b)|$$
$$\le\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\sup_{c\in\mathbf{R}}|\tilde{Q}_{x,n}(c|\tau,b)-Q_x(c|\tau,b)|=o_p(1)\ . \quad (83)$$

The claim of the Lemma then follows from results (79), (80) and (83) together with Corollary 3.2.3 in van der Vaart and Wellner (1996). ■

Lemma 6.5. *Let Assumptions 2.1, 4.1 hold, W_i be independent of (Y_i,X_i,D_i) with $E[W_i]=1$, $E[W_i^2]<\infty$ and positive a.s. Also let $\mathcal{S}\equiv\{(\tau,b)\in[0,1]^2: b\{1-p(x)\}+\epsilon\le\tau\le p(x)+b\{1-p(x)\}-\epsilon\ \forall x\in\mathcal{X}\}$ for some ϵ satisfying $0<2\epsilon<\inf_{x\in\mathcal{X}}p(x)$ and denote the minimizers:*

$$s_0(\tau,b,x)=\arg\min_{c\in\mathbf{R}}Q_x(c|\tau,b)\qquad\qquad \hat{s}_0(\tau,b,x)\in\arg\min_{c\in\mathbf{R}}\tilde{Q}_{x,n}(c|\tau,b)\ . \quad (84)$$

For $G_{x,i}(c)\equiv W_i1\{Y_i\le c,D_i=1,X_i=x\}$ and $R_{x,n}(\tau,b)$ as defined in (68), denote the criterion function:

$$\tilde{Q}^s_{x,n}(c|\tau,b)\equiv\Big(\frac{1}{n}\sum_{i=1}^n\{E[G_{x,i}(c)-G_{x,i}(s_0(\tau,b,x))]+G_{x,i}(s_0(\tau,b,x))\}+R_{x,n}(\tau,b)\Big)^2\ . \quad (85)$$

If $\{Y_i,X_i,D_i,W_i\}$ is an i.i.d. sample, it then follows that:

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\Big|\frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc}\Big|=o_p(n^{-\frac{1}{2}})\ . \quad (86)$$

Proof: We first introduce the criterion function $M_n^s: L^\infty(\mathcal{S}\times\mathcal{X})\to\mathbf{R}$ to be given by:

$$M_n^s(\theta)\equiv\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\tilde{Q}^s_{x,n}(\theta(\tau,b,x)|\tau,b)\ . \quad (87)$$

We aim to characterize and establish the consistency of an approximate minimizer of $M_n^s(\theta)$ on $L^\infty(\mathcal{S}\times\mathcal{X})$. Observe that by Lemma 6.2, compactness of $\mathcal{S}$, finiteness of $\mathcal{X}$ and the law of large numbers:

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}|\frac{1}{n}\sum_{i=1}^n\{G_{x,i}(s_0(\tau,b,x))-E[G_{x,i}(s_0(\tau,b,x))]\}+R_{x,n}(\tau,b)-R_x(\tau,b)|$$
$$\le\sup_{x\in\mathcal{X}}\sup_{c\in\mathbf{R}}|\frac{1}{n}\sum_{i=1}^n\{G_{x,i}(c)-E[G_{x,i}(c)]\}|+\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}|R_{x,n}(\tau,b)-R_x(\tau,b)|=o_p(1)\ , \quad (88)$$

where $R_x(\tau,b)$ is as in (67). Hence, by definition of $\mathcal{S}$ and $R_x(\tau,b)$, with probability tending to one:

$$\frac{\epsilon}{2}P(X_i=x)\le\frac{1}{n}\sum_{i=1}^n\{E[G_{x,i}(s_0(\tau,b,x))]-G_{x,i}(s_0(\tau,b,x))\}-R_{x,n}(\tau,b)$$
$$\le(p(x)-\frac{\epsilon}{2})P(X_i=x)\qquad\qquad \forall(\tau,b,x)\in\mathcal{S}\times\mathcal{X}\ . \quad (89)$$

By Assumption 2.1(ii), whenever (89) holds, we may implicitly define $\hat{s}^s_0(\tau,b,x)$ by the equality:

$$P(Y_i\le\hat{s}^s_0(\tau,b,x),D_i=1,X_i=x)=\frac{1}{n}\sum_{i=1}^n\{E[G_{x,i}(s_0(\tau,b,x))]-G_{x,i}(s_0(\tau,b,x))\}-R_{x,n}(\tau,b)\ , \quad (90)$$

for all $(\tau, b, x) \in \mathcal{S} \times \mathcal{X}$ and set $\hat{s}_0^s(\tau, b, x) = 0$ for all $(\tau, b, x) \in \mathcal{S} \times \mathcal{X}$ whenever (89) does not hold. Thus,

$$\sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} |\tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau, b, x)|\tau, b) - \inf_{c \in \mathbf{R}} \tilde{Q}_{x,n}^s(c|\tau, b)| = o_p(n^{-1}) \ . \tag{91}$$

Let $\hat{s}_0^s \equiv \hat{s}_0^s(\cdot, \cdot, \cdot)$ and note that by construction $\hat{s}_0^s \in L^\infty(\mathcal{S} \times \mathcal{X})$. From (91) we then obtain that:

$$M_n^s(\hat{s}_0^s) \leq \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \inf_{c \in \mathbf{R}} \tilde{Q}_{x,n}^s(c|\tau, b) + o_p(n^{-1}) \leq \inf_{\theta \in L^\infty(\mathcal{S} \times \mathcal{X})} M_n^s(\theta) + o_p(n^{-1}) \ . \tag{92}$$

In order to establish $\|\hat{s}_0^s - s_0\|_\infty = o_p(1)$, let $M(\theta)$ be as in (78) and notice that arguing as in (83) together with result (88) and Lemma 6.2 implies that:

$$\begin{aligned} \sup_{\theta \in L^\infty(\mathcal{S} \times \mathcal{X})} |M_n^s(\theta) - M(\theta)| &\leq \sup_{\theta \in L^\infty(\mathcal{S} \times \mathcal{X})} \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} |\tilde{Q}_{x,n}^s(\theta(\tau, b, x)|\tau, b) - Q_x(\theta(\tau, b, x)|\tau, b)| \\ &\leq \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \sup_{c \in \mathbf{R}} |\tilde{Q}_{x,n}^s(c|\tau, b) - Q_x(c|\tau, b)| = o_p(1) \ . \end{aligned} \tag{93}$$

Hence, by (80), (92), (93) and Corollary 3.2.3 in van der Vaart and Wellner (1996) we obtain:

$$\sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} |\hat{s}_0^s(\tau, b, x) - s_0(\tau, b, x)| = o_p(1) \ . \tag{94}$$

Next, define the random mapping $\Delta_n : L^\infty(\mathcal{S} \times \mathcal{X}) \to L^\infty(\mathcal{S} \times \mathcal{X})$ to be given by:

$$\Delta_n(\theta) \equiv \frac{1}{n} \sum_{i=1}^n \{(G_{x,i}(\theta(\tau, b, x)) - E[G_{x,i}(\theta(\tau, b, x))]) - (G_{x,i}(s_0(\tau, b, x)) - E[G_{x,i}(s_0(\tau, b, x))])\} \ , \tag{95}$$

and observe that Lemma 6.2 and finiteness of $\mathcal{X}$ implies that $\|\Delta_n(\bar{s})\|_\infty = o_p(n^{-\frac{1}{2}})$ for any $\bar{s} \in L^\infty(\mathcal{S} \times \mathcal{X})$ such that $\|\bar{s} - s_0\|_\infty = o_p(1)$. Since $\tilde{Q}_{x,n}(\hat{s}_0(\tau, b, x)|\tau, b) \leq \tilde{Q}_{x,n}(s_0(\tau, b, x)|\tau, b)$ for all $(\tau, b, x) \in \mathcal{S} \times \mathcal{X}$, and by Lemma 6.2 and finiteness of $\mathcal{X}$, $\sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \tilde{Q}_{x,n}(s_0(\tau, b, x)|\tau, b) = O_p(n^{-1})$, we conclude that:

$$\begin{aligned} &\sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \{\tilde{Q}_{x,n}^s(\hat{s}_0(\tau, b, x)|\tau, b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau, b, x)|\tau, b)\} \\ &\quad \leq \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \{\tilde{Q}_{x,n}(\hat{s}_0(\tau, b, x)|\tau, b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau, b, x)|\tau, b)\} + \|\Delta_n^2(\hat{s}_0)\|_\infty + 2\|\Delta_n(\hat{s}_0)\|_\infty \times M_n^{\frac{1}{2}}(\hat{s}_0) \\ &\quad \leq \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \{\tilde{Q}_{x,n}(\hat{s}_0^s(\tau, b, x)|\tau, b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau, b, x)|\tau, b)\} + o_p(n^{-1}) \ , \end{aligned} \tag{96}$$

where $M_n(\theta)$ is as in (78). Furthermore, since by (92) we have $M_n^s(\hat{s}_0^s) \leq M_n^s(s_0) + o_p(n^{-1})$ and by Lemma 6.2 and finiteness of $\mathcal{X}$ we have $M_n^s(s_0) = O_p(n^{-1})$, similar arguments as in (96) imply that:

$$\begin{aligned} &\sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \{\tilde{Q}_{x,n}(\hat{s}_0^s(\tau, b, x)|\tau, b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau, b, x)|\tau, b)\} \\ &\qquad \leq \|\Delta_n(\hat{s}_0^s)\|_\infty^2 + 2\|\Delta_n(\hat{s}_0^s)\|_\infty \times [M_n^s(\hat{s}_0^s)]^{\frac{1}{2}} = o_p(n^{-1}) \ . \end{aligned} \tag{97}$$

Therefore, by combining the results in (91), (96) and (97), we are able to conclude that:

$$\begin{aligned} &\sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \{\tilde{Q}_{x,n}^s(\hat{s}_0(\tau, b, x)|\tau, b) - \inf_{c \in \mathbf{R}} \tilde{Q}_{x,n}^s(c|\tau, b)\} \\ &\qquad \leq \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} \{\tilde{Q}_{x,n}^s(\hat{s}_0(\tau, b, x)|\tau, b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau, b, x)|\tau, b)\} + o_p(n^{-1}) \leq o_p(n^{-1}) \ . \end{aligned} \tag{98}$$

Let $\epsilon_n \searrow 0$ be such that $\epsilon_n = o_p(n^{-\frac{1}{2}})$ and in addition satisfies:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b) - \inf_{c\in\mathbf{R}} \tilde{Q}^s_{x,n}(c|\tau,b)| = o_p(\epsilon_n^2) \ , \tag{99}$$

which is possible by (98). A Taylor expansion at each $(\tau,b,x)\in\mathcal{S}\times\mathcal{X}$ then implies:

$$\begin{aligned} 0 &\le \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)+\epsilon_n|\tau,b) - \tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)\} + o_p(\epsilon_n^2) \\ &= \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big\{\epsilon_n \times \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + \frac{\epsilon_n^2}{2}\times\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}\Big\} + o_p(\epsilon_n^2) \ , \end{aligned} \tag{100}$$

where $\bar{s}(\tau,b,x)$ is a convex combination of $\hat{s}_0(\tau,b,x)$ and $\hat{s}_0(\tau,b,x)+\epsilon_n$. Since Lemma 6.4 and $\epsilon_n \searrow 0$ imply that $\|\bar{s}-s_0\|_\infty = o_p(1)$, the mean value theorem, $f_{y|1,x}(c)$ being uniformly bounded and (83) yield:

$$\begin{aligned} &\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{1}{n}\sum_{i=1}^n \{E[G_{x,i}(\bar{s}(\tau,b,x)) - G_{x,i}(s_0(\tau,b,x))] + G_{x,i}(s_0(\tau,b,x))\} + R_{x,n}(\tau,b)\Big| \\ &\qquad \le \sup_{c\in\mathbf{R}} f_{y|1,x}(c)p(x)P(X_i=x)\times\|\bar{s}-s_0\|_\infty + M_n^{\frac{1}{2}}(s_0) = o_p(1) \ . \end{aligned} \tag{101}$$

Therefore, exploiting (101), $f'_{y|1,x}(c)$ being uniformly bounded and by direct calculation we conclude:

$$\begin{aligned} &\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2} - 2f^2_{y|1,x}(\bar{s}(\tau,b,x))p^2(x)P^2(X_i=x)\Big| \\ &\qquad \le \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |f'_{y|1,x}(\bar{s}(\tau,b,x))p(x)P(X_i=x)| \times o_p(1) = o_p(1). \end{aligned} \tag{102}$$

Thus, combining results (100) together with (102) and $f_{y|1,x}(c)$ uniformly bounded, we conclude:

$$0 \le \epsilon_n \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + O_p(\epsilon_n^2) \ . \tag{103}$$

In a similar fashion, we note that by exploiting (99) and proceeding as in (100)-(103) we obtain:

$$\begin{aligned} 0 &\le \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)-\epsilon_n|\tau,b) - \tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)\} + o_p(\epsilon_n^2) \\ &\le \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \Big\{-\epsilon_n \times \frac{d\tilde{Q}_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + \frac{\epsilon_n^2}{2}\times\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}\Big\} + o_p(\epsilon_n^2) \\ &\le -\epsilon_n \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \frac{d\tilde{Q}_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + O_p(\epsilon_n^2) \ . \end{aligned} \tag{104}$$

Therefore, since $\epsilon_n = o_p(n^{-\frac{1}{2}})$, we conclude from (103) and (104) that we must have:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} = O_p(\epsilon_n) = o_p(n^{-\frac{1}{2}}) \ . \tag{105}$$

By similar arguments, but reversing the sign of ϵ_n in (100) and (104) it possible to establish that:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} -\frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} = o_p(n^{-\frac{1}{2}}) \ . \tag{106}$$

The claim of the Lemma then follows from (105) and (106). ■

Lemma 6.6. *Let Assumptions 2.1, 4.1 hold, W_i be independent of (Y_i, X_i, D_i) with $E[W_i]=1$, $E[W_i^2]<\infty$*

and positive a.s. Also let $\mathcal{S} \equiv \{(\tau, b) \in [0,1]^2 : b\{1-p(x)\} + \epsilon \leq \tau \leq p(x) + b\{1-p(x)\} - \epsilon \ \forall x \in \mathcal{X}\}$ *for some* ϵ *satisfying* $0 < \epsilon < \inf_{x\in\mathcal{X}} p(x)$ *and denote the minimizers:*

$$s_0(\tau, b, x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b) \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b) \ . \tag{107}$$

If $G_{x,i}(c)$ *is as in* (81), $\inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau, b, x))p(x) > 0$ *and* $\{Y_i, X_i, D_i, W_i\}$ *is i.i.d., then:*

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|(\hat{s}_0(\tau, b, x) - s_0(\tau, b, x)) - \frac{1}{n}\sum_{i=1}^n \frac{G_{x,i}(s_0(\tau, b, x)) + W_i(b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\})}{P(X_i = x)p(x)f_{y|1,x}(s_0(\tau, b, x))}\Big| = o_p(n^{-\frac{1}{2}}) \ . \tag{108}$$

PROOF: For $\tilde{Q}^s_{x,n}(c|\tau, b)$ as in (85), note that the mean value theorem and Lemma 6.5 imply:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|(\hat{s}_0(\tau, b, x) - s_0(\tau, b, x)) \times \frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau, b, x)|\tau, b)}{dc^2} + \frac{d\tilde{Q}^s_{x,n}(s_0(\tau, b, x)|\tau, b)}{dc}\Big| = o_p(n^{-\frac{1}{2}}) \tag{109}$$

for $\bar{s}(\tau, b, x)$ a convex combination of $s_0(\tau, b, x)$ and $\hat{s}_0(\tau, b, x)$. Also note that Lemma 6.2 implies:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{d\tilde{Q}^s_{x,n}(s_0(\tau, b, x)|\tau, b)}{dc}\Big|$$
$$= \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|2f_{y|1,x}(s_0(\tau, b, x))p(x)P(X_i = x) \times \{\frac{1}{n}\sum_{i=1}^n G_{x,i}(s_0(\tau, b, x)) + R_n(\tau, b)\}\Big| = O_p(n^{-\frac{1}{2}}) \ . \tag{110}$$

In addition, by (102), the mean value theorem and $f_{y|1,x}(c)$ being uniformly bounded we conclude that:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{d^2\tilde{Q}_{x,n}(\bar{s}(\tau, b, x)|\tau, b)}{dc^2} - 2f^2_{y|1,x}(s_0(\tau, b, x))p^2(x)P^2(X_i = 1)\Big|$$
$$\lesssim \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |f^2_{y|1,x}(\bar{s}(\tau, b, x)) - f^2_{y|1,x}(s_0(\tau, b, x))| + o_p(1) \lesssim \sup_{c\in\mathbf{R}} |f'_{y|1,x}(c)| \times \|\bar{s} - s_0\|_\infty + o_p(1) \ . \tag{111}$$

Since by assumption $f_{y|1,x}(s_0(\tau, b, x))p(x)$ is bounded away from zero uniformly in $(\tau, b, x) \in \mathcal{S} \times \mathcal{X}$, it follows from (111) and $\|\bar{s} - s_0\|_\infty = o_p(1)$ by Lemma 6.4 that for some $\delta > 0$:

$$\inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \frac{d^2\tilde{Q}_{x,n}(\bar{s}(\tau, b, x)|\tau, b)}{dc^2} > \delta \tag{112}$$

with probability approaching one. Therefore, we conclude from results (109), (110) and (112) that we must have $\|\hat{s}_0 - s_0\|_\infty = O_p(n^{-\frac{1}{2}})$. Hence, by (109) and (111) we conclude that:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|2(\hat{s}_0(\tau, b, x) - s_0(\tau, b, x))f^2_{y|1,x}(s_0(\tau, b, x))p^2(x)P^2(X_i = 1) + \frac{d\tilde{Q}^s_{x,n}(s_0(\tau, b, x)|\tau, b)}{dc}\Big| = o_p(n^{-\frac{1}{2}}) \tag{113}$$

The claim of the Lemma is then established by (110), (112) and (113). ■

Lemma 6.7. *Let Assumptions 2.1, 4.1(ii)-(iii) hold,* W_i *be independent of* (Y_i, X_i, D_i) *with* $E[W_i] = 1$, $E[W_i^2] < \infty$ *and positive a.s. Let* $\mathcal{S} \equiv \{(\tau, b) \in [0,1]^2 : b\{1-p(x)\} + \epsilon \leq \tau \leq p(x) + b\{1-p(x)\} - \epsilon \ \forall x \in \mathcal{X}\}$ *for some* ϵ *satisfying* $0 < 2\epsilon < \inf_{x\in\mathcal{X}} p(x)$ *and for some* $x_0 \in \mathcal{X}$, *denote the minimizers:*

$$s_0(\tau, b, x_0) = \arg\min_{c\in\mathbf{R}} Q_{x_0}(c|\tau, b) \ .$$

If $\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))p(x_0) > 0$ *and* $\{Y_i, X_i, D_i, W_i\}$ *is i.i.d., then the following class is Donsker:*

$$\mathcal{F} \equiv \Big\{ f_{\tau,b}(y,x,d,w) = \frac{w1\{y \le s_0(\tau,b,x_0), d=1, x=x_0\} + bw1\{d=0, x=x_0\} - \tau w1\{x=x_0\}}{P(X_i=x_0)p(x_0)f_{y|1,x}(s_0(\tau,b,x_0))} : (\tau,b)\in\mathcal{S}\Big\}$$

Proof: For $\epsilon > 0$, let $\{B_j\}$ be a collection of closed balls in $\mathbf{R}^2$ with diameter ϵ covering $\mathcal{S}$. Further notice that since $\mathcal{S} \subseteq [0,1]^2$, we may select $\{B_j\}$ so its cardinality is less than $4/\epsilon^2$. On each B_j define:

$$\begin{aligned}
&\underline{\tau}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} \tau && \bar{\tau}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} \tau \\
&\underline{b}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} b && \bar{b}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} b \\
&\underline{s}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} s_0(\tau,b,x_0) && \bar{s}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} s_0(\tau,b,x_0) \\
&\underline{f}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} f_{y|1,x}(s_0(\tau,b,x_0)) && \bar{f}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} f_{y|1,x}(s_0(\tau,b,x_0))\ ,
\end{aligned} \tag{114}$$

where we note that all minimums and maximums are attained due to compactness of $\mathcal{S}\cap B_j$, continuity of $s_0(\tau,b,x_0)$ by (65) and the implicit function theorem and continuity of $f_{y|1,x}(c)$ by assumption 4.1(iii). Next, for $1 \le j \le \#\{B_j\}$ define the functions:

$$l_j(y,x,d,w) \equiv \frac{w1\{y \le \underline{s}_j,\ d=1,\ x=x_0\} + \underline{b}_j w1\{d=0, x=x_0\}}{P(X_i=x_0)p(x_0)\bar{f}_j} - \frac{\bar{\tau}_j w1\{x=x_0\}}{P(X_i=x_0)p(x_0)\underline{f}_j} \tag{115}$$

$$u_j(y,x,d,w) \equiv \frac{w1\{y \le \bar{s}_j,\ d=1,\ x=x_0\} + \bar{b}_j w1\{d=0, x=x_0\}}{P(X_i=x_0)p(x_0)\underline{f}_j} - \frac{\underline{\tau}_j w1\{x=x_0\}}{P(X_i=x_0)p(x_0)\bar{f}_j} \tag{116}$$

and note that the brackets $[l_j, u_j]$ cover the class $\mathcal{F}$. Since $\bar{f}_j^{-1} \le \underline{f}_j^{-1} \le [\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))]^{-1} < \infty$ for all j, there is a finite constant M not depending on j so that $M > 3E[W_i^2]P^{-2}(X_i=x_0)p^{-2}(x_0)\underline{f}_j^{-2}\bar{f}_j^{-2}$ uniformly in j. To bound the norm of the bracket $[l_j, u_j]$ note that for such a constant M it follows that:

$$\begin{aligned}
E[(u_j(Y_i,X_i,D_i,W_i) - l_j(Y_i,X_i,D_i,W_i))^2] &\le M \times (\bar{b}_j\bar{f}_j - \underline{b}_j\underline{f}_j)^2 + M \times (\bar{\tau}_j\bar{f}_j - \underline{\tau}_j\underline{f}_j)^2 \\
&+ M \times E[(1\{Y_i \le \underline{s}_j, D_i=1, X_i=x_0\}\underline{f}_j - 1\{Y_i \le \bar{s}_j, D_i=1, X_i=x_0\}\bar{f}_j)^2]
\end{aligned} \tag{117}$$

Next observe that by the implicit function theorem and result (65) we can conclude that for any $(\tau,b)\in\mathcal{S}$:

$$\frac{ds_0(\tau,b,x_0)}{d\tau} = \frac{1}{f_{y|1,x}(s_0(\tau,b,x_0))} \qquad \frac{ds_0(\tau,b,x_0)}{db} = -\frac{1-p(x_0)}{f_{y|1,x}(s_0(\tau,b,x_0))}\ . \tag{118}$$

Since the minimums and maximums in (114) are attained, it follows that for some $(\tau_1,b_1),(\tau_2,b_2)\in B_j\cap\mathcal{S}$ we have $s_0(\tau_1, b_1.x_0) = \bar{s}_j$ and $s_0(\tau_2,b_2,x_0) = \underline{s}_j$. Hence, the mean value theorem and (118) imply:

$$|\bar{s}_j - \underline{s}_j| = \Big|\frac{1}{f_{y|1,x}(s_0(\tilde{\tau},\tilde{b},x_0))}(\tau_1-\tau_2) + \frac{1-p(x)}{f_{y|1,x}(s_0(\tilde{\tau},\tilde{b},x_0))}(b_1-b_2)\Big| \le \frac{\sqrt{2}\epsilon}{\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))} \tag{119}$$

where $(\tilde{\tau},\tilde{b})$ is between (τ_1,b_1) and (τ_2,b_2) and the final inequality follows by $(\tilde{\tau},\tilde{b})\in\mathcal{S}$ by convexity of $\mathcal{S}$, $(\tau_1,b_1),(\tau_2,b_2)\in B_j\cap\mathcal{S}$ and B_j having diameter ϵ. By similar arguments, and (119) we conclude:

$$|\bar{f}_j - \underline{f}_j| \le \sup_{c\in\mathbf{R}} |f'_{y|1,x}(c)| \times |\bar{s}_j - \underline{s}_j| \le \sup_{c\in\mathbf{R}} |f'_{y|1,x}(c)| \times \frac{\sqrt{2}\epsilon}{\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))}\ . \tag{120}$$

Since $\underline{b}_j \leq \bar{b}_j \leq 1$ due to $\bar{b}_j \in [0,1]$ and $|\bar{b}_j - \underline{b}_j| \leq \epsilon$ by B_j having diameter ϵ, we further obtain that:

$$(\bar{b}_j\bar{f}_j - \underline{b}_j\underline{f}_j)^2 \leq 2\bar{f}_j^2(\bar{b}_j - \underline{b}_j)^2 + 2\underline{b}_j^2(\bar{f}_j - \underline{f}_j)^2 \leq 2\sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times \epsilon^2 + \frac{4\epsilon^2}{\inf_{(\tau,b)\in\mathcal{S}} f^2_{y|1,x}(s_0(\tau,b,x_0))} \ , \quad (121)$$

where in the final inequality we have used result (120). By similar arguments, we also obtain:

$$(\bar{\tau}_j\bar{f}_j - \underline{\tau}_j\underline{f}_j)^2 \leq 2\sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times \epsilon^2 + \frac{4\epsilon^2}{\inf_{(\tau,b)\in\mathcal{S}} f^2_{y|1,x}(s_0(\tau,b,x_0))} \ . \quad (122)$$

Also note that by direct calculation, the mean value theorem and results (119) and (120) it follows that:

$$\begin{aligned} E[(1\{Y_i \leq \underline{s}_j, D_i = 1, X_i = x_0\}\underline{f}_j - 1\{Y_i \leq \bar{s}_j, D_i = 1, X_i = x_0\}\bar{f}_j)^2] \\ \leq 2(\bar{f}_j - \underline{f}_j)^2 + \sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times P(X_i = x_0)p(x_0)(F_{y|1,x}(\bar{s}_j) - F_{y|1,x}(\underline{s}_j)) \\ \leq \frac{4\epsilon^2}{\inf_{(\tau,b)\in\mathcal{S}} f^2_{y|1,x}(s_0(\tau,b,x_0))} + \sup_{c\in\mathbf{R}} f^3_{y|1,x}(c) \times \frac{\sqrt{2}\epsilon}{\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))} \ . \quad (123) \end{aligned}$$

Thus, from (117) and (121)-(122), it follows that for $\epsilon < 1$ and some constant K not depending on j:

$$E[(u_j(Y_i,X_i,D_i,W_i) - l_j(Y_i,X_i,D_i,W_i))^2] \leq K\epsilon \ . \quad (124)$$

Since $\#\{B_j\} \leq 4/\epsilon$, we can therefore conclude that $N_{[\,]}(\epsilon, \mathcal{F}, \|\cdot\|_{L^2}) \leq 4K/\epsilon^2$ and hence by Theorem 2.5.6 in van der Vaart and Wellner (1996) it follows that the class $\mathcal{F}$ is Donsker. ■

Lemma 6.8. *Let Assumptions 2.1, 4.1(ii)-(iii) hold, W_i be independent of (Y_i, X_i, D_i) with $E[W_i] = 1$, $E[W_i^2] < \infty$, positive a.s., $\mathcal{S} \equiv \{(\tau,b) \in [0,1]^2 : b\{1-p(x)\} + \epsilon \leq \tau \leq p(x) + b\{1-p(x)\} \forall x \in \mathcal{X}\}$ and*

$$\tilde{p}(x) \equiv \frac{\sum_{i=1}^n W_i 1\{D_i = 1, X_i = x\}}{\sum_{i=1}^n W_i 1\{X_i = x\}} \qquad p(x) \equiv P(D_i = 1|X_i = x) \qquad s_0(\tau,b,x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau,b) \ .$$

If $\inf_{(\tau,b,x)\in\mathcal{S}\times\mathcal{X}} f_{y|1,x}(s_0(\tau,b,x))p(x) > 0$ and $\{Y_i, X_i, D_i, W_i\}$ is an i.i.d. sample, then for $a \in \{-1, 1\}$:

$$\begin{aligned} s_0(\tau, \tau + ak\tilde{p}(x), x) - s_0(\tau, \tau + akp(x), x) \\ = -\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau,\tau+akp(x),x))P(X=x)} \times \frac{1}{n}\sum_{i=1}^n R(X_i,W_i,x) + o_p(n^{-\frac{1}{2}}) \ , \quad (125) \end{aligned}$$

where $R(W_i, X_i, x) = p(x)\{P(X = x) - W_i 1\{X_i = x\}\} + W_i 1\{D_i = 1, X_i = x\} - P(D = 1, X = x)$ and (125) *holds uniformly in $(\mathcal{B} \times \mathcal{X})$. Moreover, the right hand side of* (125) *is Donsker.*

Proof: First observe that $(\tau, k) \in \mathcal{B}$ implies $(\tau, \tau + akp(x)) \in \mathcal{S}$ for all $x \in \mathcal{X}$, and that with probability tending to one $(\tau, \tau + ak\tilde{p}(x)) \in \mathcal{S}$ for all $(\tau, k) \in \mathcal{B}$. In addition, also note that

$$\tilde{p}(x) - p(x) = \frac{1}{nP(X=x)}\sum_{i=1}^n R(X_i,W_i,x) + o_p(n^{-\frac{1}{2}}) \quad (126)$$

by an application of the Delta method and $\inf_{x\in\mathcal{X}} P(X = x) > 0$ due to X having finite support. Moreover, by the mean value theorem and (118) we obtain for some $\bar{b}(\tau,k)$ between $\tau + ak\tilde{p}(x)$ and $\tau + akp(x)$

$$s_0(\tau,\tau+ak\tilde{p}(x),x) - s_0(\tau,\tau+akp(x),x) = -\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau,\bar{b}(\tau,k),x))}(\tilde{p}(x)-p(x))$$
$$= -\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau,\tau+akp(x),x))}(\tilde{p}(x)-p(x)) + o_p(n^{-\frac{1}{2}}) \quad (127)$$

where the second equality follows from $(\tau,\bar{b}(\tau,k)) \in \mathcal{S}$ for all (τ,k) with probability approaching one by convexity of $\mathcal{S}$, $\inf_{(\tau,b,x)\in\mathcal{S}\times\mathcal{X}} f_{y|1,x}(s_0(\tau,b,x))p(x) > 0$ and $\sup_{(\tau,k)\in\mathcal{B}} |ak(\tilde{p}(x)-p(x))| = o_p(1)$ uniformly in $\mathcal{X}$. The first claim of the Lemma then follows by combining (126) and (127).

Finally, observe that the right hand side of (125) is trivially Donsker since $R(X_i,W_i,x)$ does not depend on (k,τ) and the function $(1-p(x))ka/(f_{y|1,x}(s_0(\tau,\tau+akp(x),x))P(X=x))$ is uniformly continuous on $(\tau,k)\in\mathcal{B}$ due to $\inf_{(\tau,b,x)\in\mathcal{S}\times\mathcal{X}} f_{y|1,x}(s_0(\tau,b,x))p(x) > 0$. ■

Proof of Theorem 4.1: Throughout the proof we exploit Lemmas 6.6 and 6.7 applied with $W_i = 1$ with probability one, so that $\tilde{Q}_{x,n}(c|\tau,b) = Q_{x,n}(c|\tau,b)$ for all (τ,b) in $\mathcal{S}$, where

$$\mathcal{S} \equiv \{(\tau,b)\in[0,1]^2 : b\{1-p(x)\}+\epsilon \le \tau \le p(x)+b\{1-p(x)\}-\epsilon \ \forall x\in\mathcal{X}\} . \quad (128)$$

Also notice that for every $(\tau,k)\in\mathcal{B}$ and all $x\in\mathcal{X}$, the points $(\tau,\tau+kp(x)),(\tau,\tau-kp(x))\in\mathcal{S}$, while with probability approaching one $(\tau,\tau+k\hat{p}(x))$ and $(\tau,\tau-k\hat{p}(x))$ also belong to $\mathcal{S}$. Therefore for $s_0(\tau,b,x)$ and $\hat{s}_0(\tau,b,x)$ as defined in (107) we obtain from Lemmas 6.6 and 6.7, applied with $W_i = 1$ a.s., that:

$$|(\hat{s}_0(\tau,\tau+akp(x),x)-s_0(\tau,\tau+akp(x),x))-(\hat{s}_0(\tau,\tau+ak\hat{p}(x),x)-s_0(\tau,\tau+ak\hat{p}(x),x))| = o_p(n^{-\frac{1}{2}}) \quad (129)$$

uniformly in $(\tau,k,x)\in\mathcal{B}\times\mathcal{X}$ and $a\in\{-1,1\}$. Moreover, by Lemma 6.8 applied with $W_i = 1$ a.s.

$$s_0(\tau,\tau+ak\hat{p}(x),x) - s_0(\tau,\tau+akp(x),x)$$
$$= -\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau,\tau+akp(x),x))P(X=x)} \times \frac{1}{n}\sum_{i=1}^{n} R(X_i,x) + o_p(n^{-\frac{1}{2}}) , \quad (130)$$

where $R(X_i,x) = p(x)\{P(X=x)-1\{X_i=x\}\} + 1\{D_i=1,X_i=x\} - P(D=1,X=x)$ again uniformly in $(\tau,k,x)\in\mathcal{B}\times\mathcal{X}$. Also observe that since $(\tau,\tau+k\hat{p}(x))$ and $(\tau,\tau-k\hat{p}(x))$ belong to $\mathcal{S}$ with probability approaching one, we obtain uniformly in $(\tau,k,x)\in\mathcal{B}\times\mathcal{X}$ that:

$$\begin{aligned} q_L(\tau,k|x) &= s_0(\tau,\tau+kp(x),x) & q_U(\tau,k|x) &= s_0(\tau,\tau-kp(x),x) \\ \hat{q}_L(\tau,k|x) &= \hat{s}_0(\tau,\tau+k\hat{p}(x),x)+o_p(n^{-\frac{1}{2}}) & \hat{q}_U(\tau,k|x) &= \hat{s}_0(\tau,\tau-k\hat{p}(x),x)+o_p(n^{-\frac{1}{2}}) . \end{aligned} \quad (131)$$

Therefore, combining results (129)-(131) and exploiting Lemmas 6.6, 6.7 and 6.8 and the sum of Donsker classes being Donsker we conclude that for J a Gaussian process on $L^\infty(\mathcal{B}\times\mathcal{X})\times L^\infty(\mathcal{B}\times\mathcal{X})$:

$$\sqrt{n}C_n \xrightarrow{L} J \qquad C_n(\tau,k,x) \equiv \begin{pmatrix} \hat{q}_L(\tau,k|x)-q_L(\tau,k|x) \\ \hat{q}_U(\tau,k|x)-q_U(\tau,k|x) \end{pmatrix} . \quad (132)$$

Next, observe that since X has finite support, we may denote $\mathcal{X} = \{x_1,\ldots,x_{|\mathcal{X}|}\}$ and define the matrix

$B = (P(X_i = x_1)x_1, \ldots, P(X_i = x_{|\mathcal{X}|})x_{|\mathcal{X}|})$ as well as the vector of weights:

$$w \equiv \lambda' \, (E_S[X_i X_i'])^{-1} B \ . \tag{133}$$

Since w is also a function on $\mathcal{X}$, we refer to its coordinates by $w(x)$. Solving the linear programming problems in (10) and (11), it is then possible to obtain the closed form solution:

$$\begin{aligned}\pi_L(\tau,k) &= \sum_{x\in\mathcal{X}} \{1\{w(x) \geq 0\}w(x)q_L(\tau,k|x) + 1\{w(x) \leq 0\}w(x)q_U(\tau,k|x)\} \\ \pi_U(\tau,k) &= \sum_{x\in\mathcal{X}} \{1\{w(x) \geq 0\}w(x)q_U(\tau,k|x) + 1\{w(x) \leq 0\}w(x)q_L(\tau,k|x)\}\end{aligned} \tag{134}$$

with a similar representation holding for $(\hat{\pi}_L(\tau,k), \hat{\pi}_U(\tau,k))$ but with $(\hat{q}_L(\tau,k|x), \hat{q}_U(\tau,k|x))$ in place of $(q_L(\tau,k|x), q_U(\tau,k|x))$. We hence define the linear map $K : L^\infty(\mathcal{B}\times\mathcal{X}) \times L^\infty(\mathcal{B}\times\mathcal{X}) \to L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$, to be given by:

$$K(\theta)(\tau,k) \equiv \begin{pmatrix} \sum_{x\in\mathcal{X}}\{1\{w(x)\geq 0\}w(x)\theta^{(1)}(\tau,k,x) + 1\{w(x)\leq 0\}w(x)\theta^{(2)}(\tau,k,x)\} \\ \sum_{x\in\mathcal{X}}\{1\{w(x)\geq 0\}w(x)\theta^{(2)}(\tau,k,x) + 1\{w(x)\leq 0\}w(x)\theta^{(1)}(\tau,k,x)\} \end{pmatrix} \tag{135}$$

where for any $\theta \in L^\infty(\mathcal{X}\times\mathcal{B}) \times L^\infty(\mathcal{B}\times\mathcal{X})$, $\theta^{(i)}(\tau,k,x)$ denotes the i^{th} coordinate of the two dimensional vector $\theta(\tau,k,x)$. It then follows from (132), (134) and (135) that:

$$\sqrt{n}\begin{pmatrix} \hat{\pi}_L - \pi_L \\ \hat{\pi}_U - \pi_U \end{pmatrix} = \sqrt{n}K(C_n) \ . \tag{136}$$

Moreover, employing the norm $\|\cdot\|_\infty + \|\cdot\|_\infty$ on the product spaces $L^\infty(\mathcal{B}\times\mathcal{X}) \times L^\infty(\mathcal{B}\times\mathcal{X})$ and $L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$, we can then obtain by direct calculation that for any $\theta \in L^\infty(\mathcal{B}\times\mathcal{X}) \times L^\infty(\mathcal{B}\times\mathcal{X})$:

$$\|K(\theta)\|_\infty \leq 2\sum_{x\in\mathcal{X}} |w(x)| \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\theta(\tau,b,x)| = 2\sum_{x\in\mathcal{X}} |w(x)| \times \|\theta\|_\infty \ , \tag{137}$$

which implies the linear map K is continuous. Therefore, the theorem is established by (132), (136), the linearity of K and the continuous mapping theorem. ■

Proof of Theorem 4.2: For a metric space $\mathbb{D}$, let $BL_c(\mathbb{D})$ denote the set of real valued bounded Lipschitz functions with supremum norm and Lipschitz constant less than or equal to c. We first aim to show that:

$$\sup_{h\in BL_1(\mathbf{R})} |E[h(L(\tilde{G}_\omega))|\mathcal{Z}_n] - E[h(L(G_\omega))]| = o_p(1) \ , \tag{138}$$

where $\mathcal{Z}_n = \{Y_i, X_i, D_i\}_{i=1}^n$ and $E[h(\tilde{Z})|\mathcal{Z}_n]$ denotes outer expectation over $\{W_i\}_{i=1}^n$ with $\mathcal{Z}_n$ fixed. Let

$$\hat{s}_0(\tau,b,x) \in \arg\min_{c\in\mathbf{R}} Q_{x,n}(c|\tau,b) \qquad \tilde{s}_0(\tau,b,x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau,b) \qquad s_0(\tau,b,x) \in \arg\min_{c\in\mathbf{R}} Q_x(c|\tau,b) \ . \tag{139}$$

Also note that with probability approaching one the points $(\tau, \tau + ak\tilde{p}(x)) \in \mathcal{S}$ for all $(\tau,k,x) \in \mathcal{B}\times\mathcal{X}$

and $a \in \{-1, 1\}$ for $\mathcal{S}$ as in (128). Hence, arguing as in (129) and (130) we obtain:

$$\tilde{q}_L(\tau, k|x) - \hat{q}_L(\tau, k|x) = \tilde{s}_0(\tau, \tau + kp(x), x) - \hat{s}_0(\tau, \tau + kp(x), x) - \frac{(1-p(x))k}{f_{y|1,x}(s_0(\tau, \tau + kp(x), x))P(X = x)} \times \frac{1}{n}\sum_{i=1}^{n} \Delta R(X_i, W_i, x) + o_p(n^{-\frac{1}{2}}) \quad (140)$$

$$\tilde{q}_U(\tau, k|x) - \hat{q}_U(\tau, k|x) = \tilde{s}_0(\tau, \tau - kp(x), x) - \hat{s}_0(\tau, \tau - kp(x), x) + \frac{(1-p(x))k}{f_{y|1,x}(s_0(\tau, \tau - kp(x), x))P(X = x)} \times \frac{1}{n}\sum_{i=1}^{n} \Delta R(X_i, W_i, x) + o_p(n^{-\frac{1}{2}}) \quad (141)$$

where $\Delta R(X_i, W_i, x) = (1 - W_i)(1\{X_i = x\}p(x) - 1\{D_i = 1, X_i = x\})$ and both statements hold uniformly in $(\tau, k, x) \in \mathcal{B} \times \mathcal{X}$. Also note that for the operator K as defined in (135), we have:

$$\sqrt{n}\begin{pmatrix} \tilde{\pi}_L - \hat{\pi}_L \\ \tilde{\pi}_U - \hat{\pi}_U \end{pmatrix} = \sqrt{n}K(\tilde{C}_n) \qquad \tilde{C}_n(\tau, k, x) \equiv \begin{pmatrix} \tilde{q}_L(\tau, k|x) - \hat{q}_L(\tau, k|x) \\ \tilde{q}_U(\tau, k|x) - \hat{q}_U(\tau, k|x) \end{pmatrix} . \quad (142)$$

By Lemmas 6.6, 6.7 and 6.8, results (140) and (141) and Theorem 2.9.2 in van der Vaart and Wellner (1996), the process $\sqrt{n}\tilde{C}_n$ converges unconditionally to a tight Gaussian process on $L^\infty(\mathcal{B} \times \mathcal{X})$. Hence, by the continuous mapping theorem, $\sqrt{n}K(\tilde{C}_n)$ is asymptotically tight. Define,

$$\bar{G}_\omega \equiv \sqrt{n}\begin{pmatrix} (\tilde{\pi}_L - \hat{\pi}_L)/\omega_L \\ (\tilde{\pi}_U - \hat{\pi}_U)/\omega_U \end{pmatrix} , \quad (143)$$

and notice that $\omega_L(\tau, k)$ and $\omega_U(\tau, k)$ being bounded away from zero, $\hat{\omega}_L(\tau, k)$ and $\hat{\omega}_U(\tau, k)$ being uniformly consistent by Assumption 4.2(ii) and $\sqrt{n}K(\tilde{C}_n)$ being asymptotically tight imply that:

$$|L(\tilde{G}_\omega) - L(\bar{G}_\omega)| \leq \sup_{(\tau,k)\in\mathcal{B}} M_0\Big|\frac{\sqrt{n}(\hat{\pi}_L(\tau, k) - \tilde{\pi}_L(\tau, k))}{\hat{\omega}_L(\tau, k)} - \frac{\sqrt{n}(\hat{\pi}_L(\tau, k) - \tilde{\pi}_L(\tau, k))}{\omega_L(\tau, k)}\Big| + \sup_{(\tau,k)\in\mathcal{B}} M_0\Big|\frac{\sqrt{n}(\hat{\pi}_U(\tau, k) - \tilde{\pi}_U(\tau, k))}{\hat{\omega}_U(\tau, k)} - \frac{\sqrt{n}(\hat{\pi}_U(\tau, k) - \tilde{\pi}_U(\tau, k))}{\omega_U(\tau, k)}\Big| = o_p(1) , \quad (144)$$

for some constant M_0 due to L being Lipschitz. By definition of BL_1, all $h \in BL_1$ have Lipschitz constant less than or equal to 1 and are also bounded by 1. Hence, for any $\eta > 0$ Markov's inequality implies:

$$P\Big(\sup_{h\in BL_1(\mathbf{R})} |E[h(L(\tilde{G}_\omega))|\mathcal{Z}_n] - E[h(L(\bar{G}_\omega))|\mathcal{Z}_n]| > \eta\Big) \leq P\Big(2P(|L(\tilde{G}_\omega) - L(\bar{G}_\omega)| > \frac{\eta}{2}|\mathcal{Z}_n) + \frac{\eta}{2}P(|L(\tilde{G}_\omega) - L(\bar{G}_\omega)| \leq \frac{\eta}{2}|\mathcal{Z}_n) > \eta\Big) \leq \frac{4}{\eta}E\Big[E\Big[1\Big\{|L(\tilde{G}_\omega) - L(\bar{G}_\omega)| > \frac{\eta}{2}\Big\}|\mathcal{Z}_n\Big]\Big] . \quad (145)$$

Therefore, by (144), (145) and Lemma 1.2.6 in van der Vaart and Wellner (1996), we obtain:

$$P\Big(\sup_{h\in BL_1(\mathbf{R})} |E[h(L(\tilde{G}_\omega))|\mathcal{Z}_n] - E[h(L(\bar{G}_\omega))|\mathcal{Z}_n]| > \eta\Big) \leq \frac{4}{\eta}P\Big(|L(\tilde{G}_\omega) - L(\bar{G}_\omega)| > \frac{\eta}{2}\Big) = o(1) . \quad (146)$$

Next, let $\stackrel{L}{=}$ stands for "equal in law" and notice that for J the Gaussian process in (132):

$$L(G_\omega) \stackrel{L}{=} T \circ K(J) \qquad L(\bar{G}_\omega) = \sqrt{n}L \circ K(\tilde{C}_n) , \quad (147)$$

due to the continuous mapping theorem and (142). For $w(x)$ as defined in (130) and $C_0 \equiv 2\sum_{x\in\mathcal{X}}|w(x)|$, it follows from linearity of K and (135), that K is Lipschitz with Lipschitz constant C_0. Therefore, for any $h \in BL_1(\mathbf{R})$, result (147) implies that $h \circ L \circ K \in BL_{C_0M_0}(L^\infty(\mathcal{B}\times\mathcal{X}))$ for some $M_0 > 0$ and hence

$$\sup_{h\in BL_1(\mathbf{R})} |E[h(L(\bar{G}_\omega))|\mathcal{Z}_n] - E[h(L(G_\omega))]| \leq \sup_{h\in BL_{C_0M_0}(L^\infty(\mathcal{B}\times\mathcal{X}))} |E[h(\bar{G}_\omega)|\mathcal{Z}_n] - E[h(J)]| = o_p(1) \ , \quad (148)$$

where the final equality follows from (140), (141), (147), arguing as in (145)-(146) and Lemmas 6.7, 6.8 and Theorem 2.9.6 in van der Vaart and Wellner (1996). Hence, (146) and (148) establish (138).

Next, we aim to exploit (138) to show that for all $t \in \mathbf{R}$ at which the cdf of $L(G_\omega)$ is continuous:

$$|P(L(\tilde{G}_\omega) \leq t|\mathcal{Z}_n) - P(L(G_\omega) \leq t)| = o_p(1) \ . \quad (149)$$

Towards this end, for every $\lambda > 0$, and t at which the cdf of $L(G_\omega)$ is continuous define the functions:

$$h^U_{\lambda,t}(u) = 1 - 1\{u > t\}\min\{\lambda(u-t), 1\} \qquad h^L_{\lambda,t}(u) = 1\{u < t\}\min\{\lambda(t-u), 1\} \ . \quad (150)$$

Notice that by construction, $h^L_{\lambda,t}(u) \leq 1\{u \leq t\} \leq h^U_{\lambda,t}(u)$ for all $u \in \mathbf{R}$, the functions $h^L_{\lambda,t}(u)$ and $h^U_{\lambda,t}(u)$ are both bounded by one and they are both Lipschitz with Lipschitz constant λ. Also by direct calculation:

$$0 \leq E[h^U_{\lambda,t}(L(G_\omega)) - h^L_{\lambda,t}(L(G_\omega))] \leq P(t - \lambda^{-1} \leq L(G_\omega) \leq t + \lambda^{-1}) \ . \quad (151)$$

Therefore, exploiting that $h^L_{\lambda,t}, h^U_{\lambda,t} \in BL_\lambda(\mathbf{R})$ and that $h \in BL_\lambda(\mathbf{R})$ implies $\lambda^{-1}h \in BL_1(\mathbf{R})$, we obtain:

$$\begin{aligned}|P(L(\tilde{G}_\omega) \leq\, & t|\mathcal{Z}_n) - P(L(G_\omega) \leq t)| \\ & \leq |E[h^L_{\lambda,t}(L(\tilde{G}_\omega))|\mathcal{Z}_n] - E[h^U_{\lambda,t}(L(G_\omega))]| + |E[h^U_{\lambda,t}(L(\tilde{G}_\omega))|\mathcal{Z}_n] - E[h^L_{\lambda,t}(L(G_\omega))]| \\ & \leq 2\sup_{h\in BL_\lambda(\mathbf{R})} |E[h(L(\tilde{G}_\omega))|\mathcal{Z}_n] - E[h(L(G_\omega))]| + 2P(t-\lambda^{-1} \leq L(G_\omega) \leq t+\lambda^{-1}) \\ & = 2\lambda \sup_{h\in BL_1(\mathbf{R})} |E[h(L(\tilde{G}_\omega))|\mathcal{Z}_n] - E[h(L(G_\omega))]| + 2P(t-\lambda^{-1} \leq L(G_\omega) \leq t+\lambda^{-1}) \ . \quad (152)\end{aligned}$$

If t is a continuity point of the cdf of $L(G_\omega)$, then (149) follows from (138) and (151).

Finally, note that since the cdf of $L(G_\omega)$ is strictly increasing and continuous at $r_{1-\alpha}$, we obtain that:

$$P(L(G_\omega) \leq r_{1-\alpha} - \epsilon) < 1 - \alpha < P(L(G_\omega) \leq r_{1-\alpha} + \epsilon) \quad (153)$$

$\forall \epsilon > 0$. Define the event $A_n \equiv \{P(L(\tilde{G}_\omega) \leq r_{1-\alpha} - \epsilon|\mathcal{Z}_n) < 1 - \alpha < P(L(\tilde{G}_\omega) \leq r_{1-\alpha} + \epsilon|\mathcal{Z}_n)\}$ and notice

$$P(|\tilde{r}_{1-\alpha} - r_{1-\alpha}| \leq \epsilon) \geq P(A_n) \to 1 \ , \quad (154)$$

where the inequality follows by definition of $\tilde{r}_{1-\alpha}$ and the second result is implied by (149) and (153). ■

Table 1: Fraction of Observations in Census Estimation Sample with Missing Weekly Earnings

Census Year	Number of Observations	Fraction Missing
1980	80,128	19.49%
1990	111,070	23.09%
2000	131,265	27.70%
Overall	322,463	23.66%

Table 2: Logit Estimates of $P(D = 0|X = x, F_{y|x}(Y) = \tau)$ in 1973 CPS-IRS Sample

	Model 1	Model 2	Model 3
b_1	-1.30	7.25	
	(0.42)	(15.76)	
b_2	1.33	-0.72	
	(0.41)	(9.75)	
γ_1		3.92	
		(7.25)	
γ_2		-0.92	
		(4.50)	
Log-Likelihood	-3,865.69	-3861.78	-3825.17
Parameters	35	37	99
Number of observations	13,598	13,598	13,598
Demographic Cells	33	33	33
$E[\frac{\partial P(D=0\|X=x,F_{y\|x}(Y)=\tau)}{\partial \tau}\|_{\tau=0.2}]$	-0.06	-0.06	-0.04
$E[\frac{\partial P(D=0\|X=x,F_{y\|x}(Y)=\tau)}{\partial \tau}\|_{\tau=0.5}]$	0.00	0.00	0.00
$E[\frac{\partial P(D=0\|X=x,F_{y\|x}(Y)=\tau)}{\partial \tau}\|_{\tau=0.8}]$	0.06	0.06	-0.05
Min KS Distance	0.03	0.02	0.02
Median KS Distance	0.03	0.06	0.11
Max KS Distance ($\mathcal{S}(F)$)	0.03	0.12	0.44

Note: Asymptotic standard errors in parentheses.

Figure 1: Linear Conditional Quantile Functions (Shaded Region) as a Subset of the Identified Set

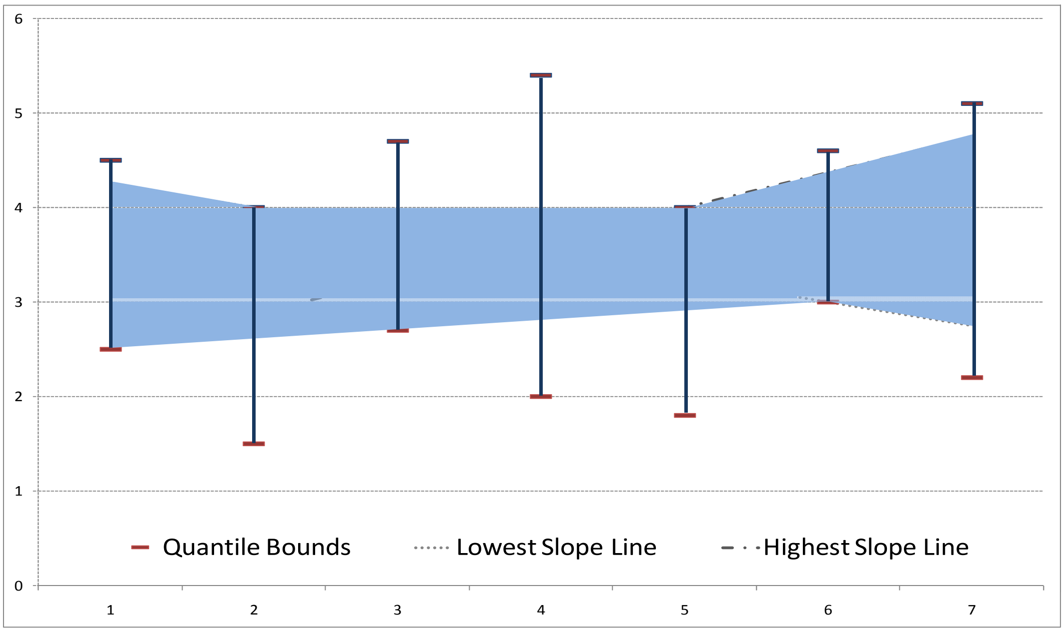


Figure 2: Conditional Quantile and its Pseudo-True Approximation

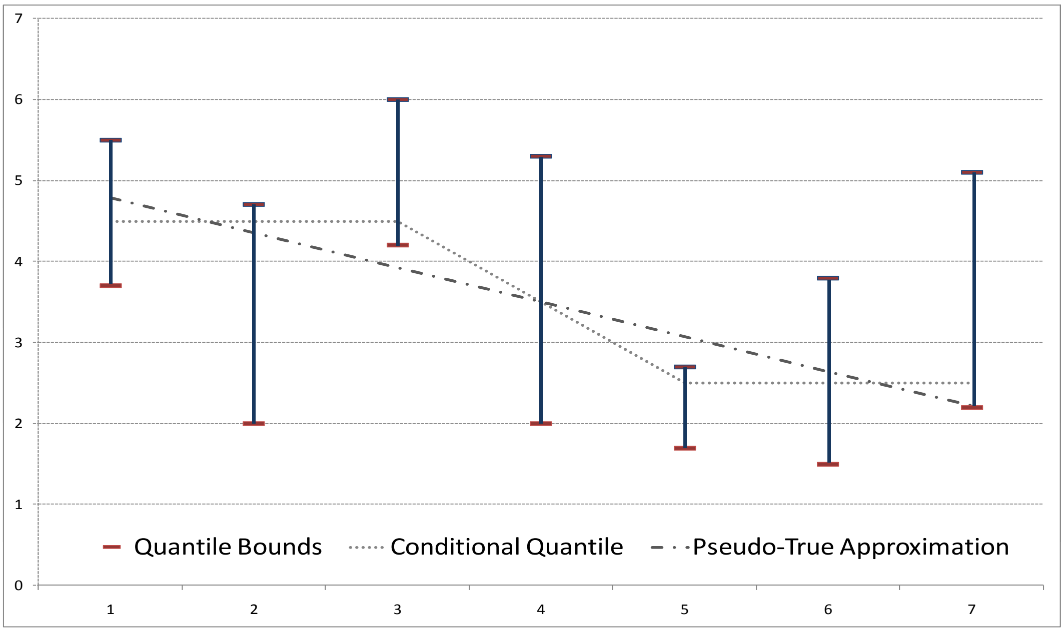


Figure 3: Confidence Intervals Under Missing at Random Assumption

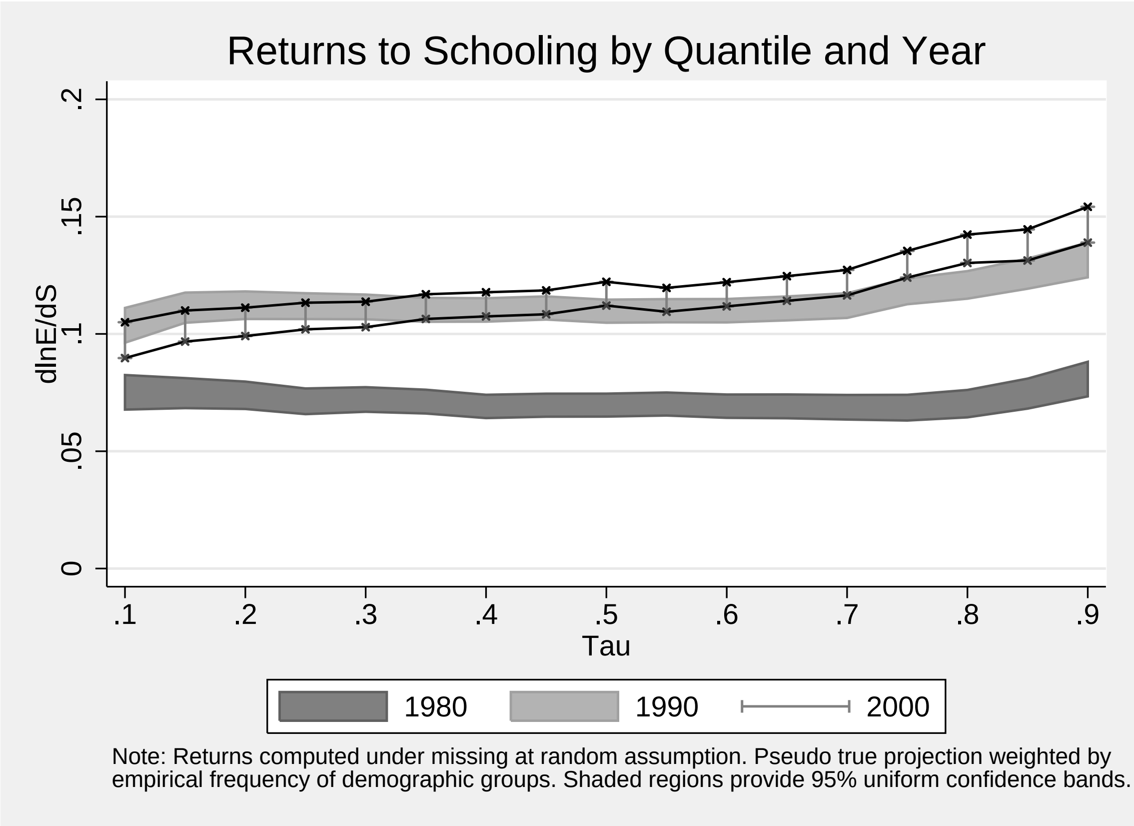


Figure 4: Confidence Intervals Under $\mathcal{S}(F) \leq 0.05$

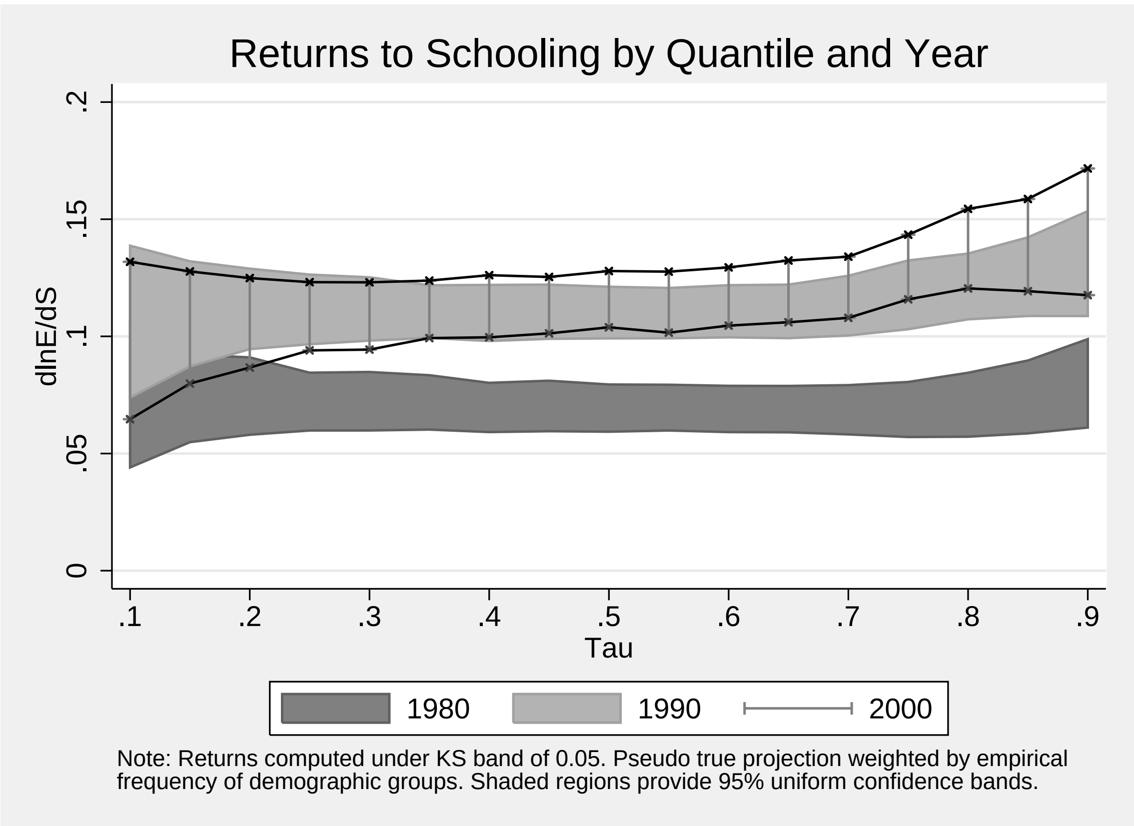


Figure 5: Confidence Intervals Under $\mathcal{S}(F) \leq 0.175$ (1980 vs. 1990)

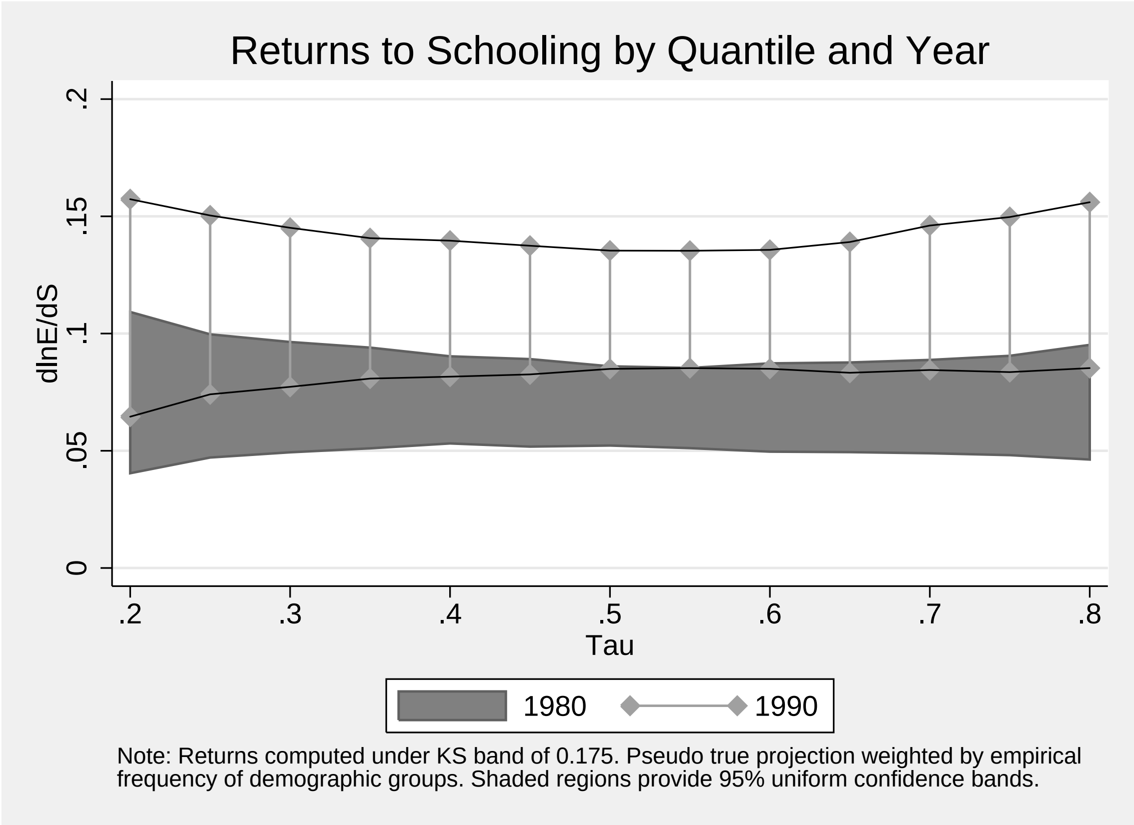


Figure 6: Intersection of 1980 upper envelope and 1990 lower envelope

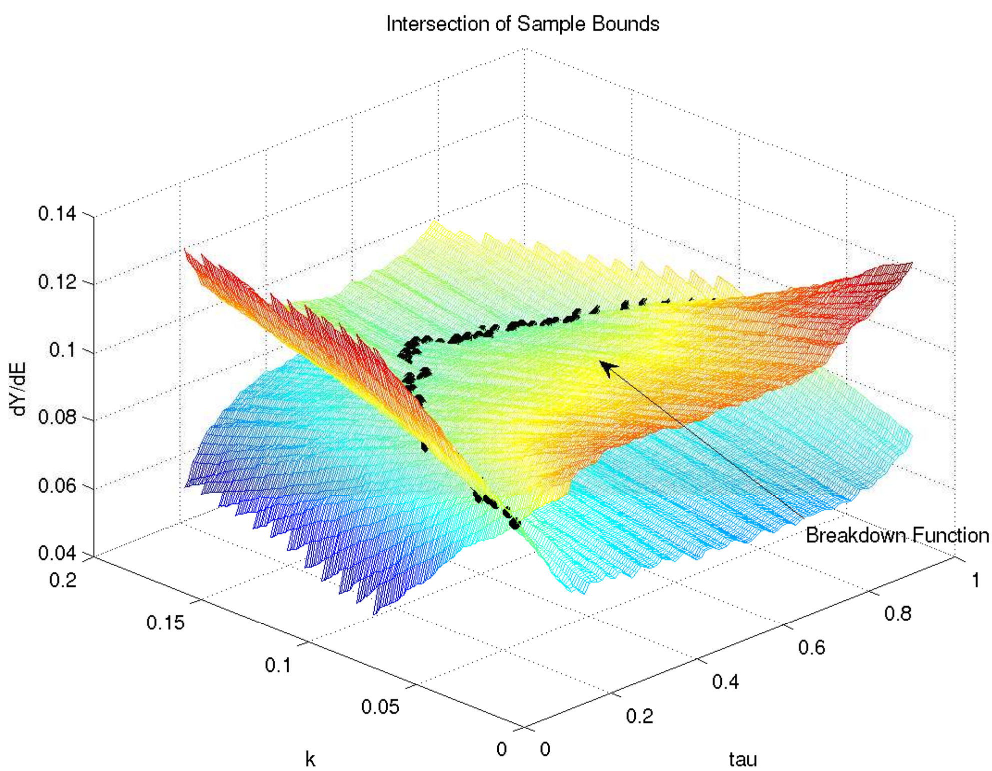


Figure 7: Breakdown Analysis

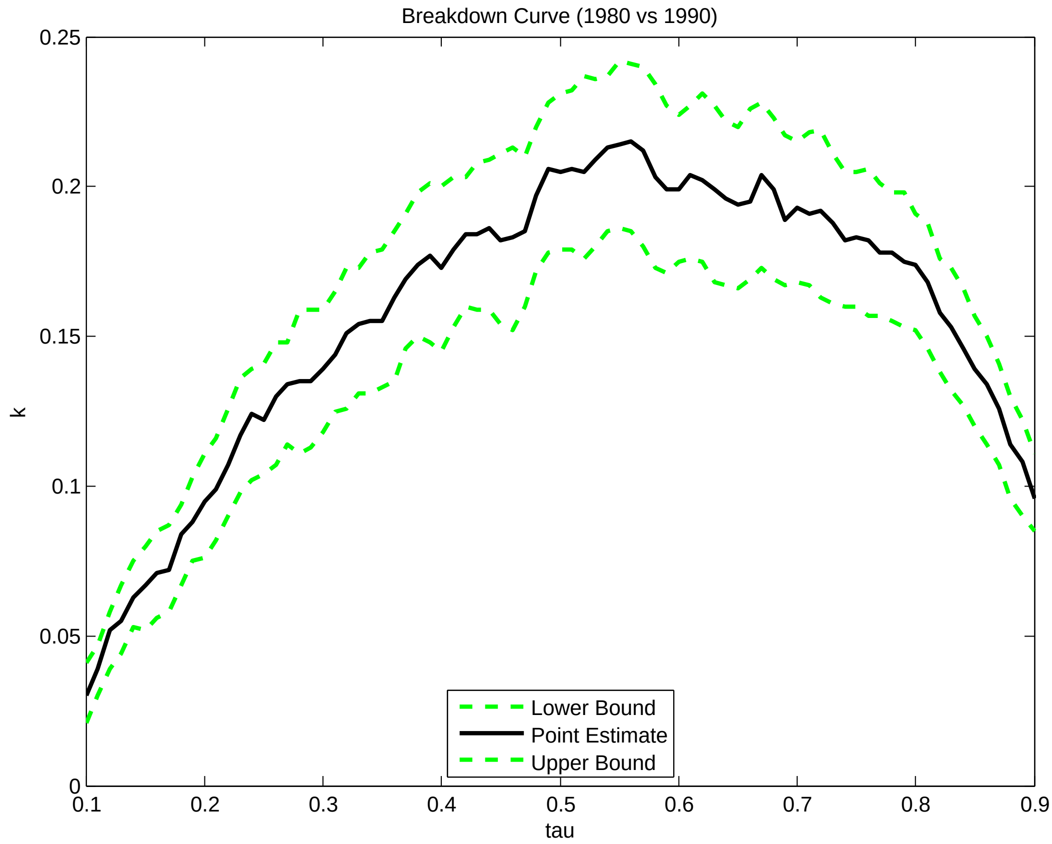


Figure 8: Confidence Intervals for Fitted Values Under $\mathcal{S}(F) \leq 0.05$

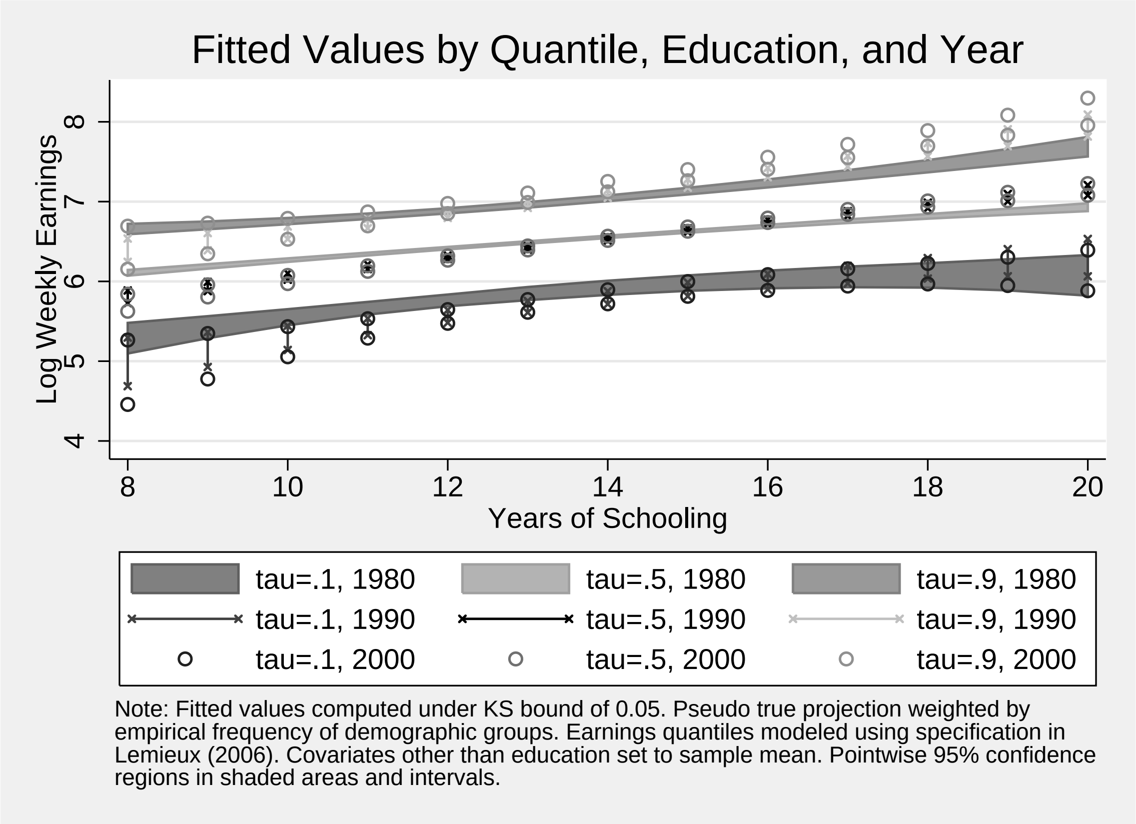


Figure 9: Logit Based Estimates of Selection in 1973 CPS-IRS File