Interval Estimation of Potentially Misspecified Quantile Models in the Presence of Missing Data

Patrick Kline
Department of Economics
UC Berkeley / NBER
pkline@econ.berkeley.edu

Andres Santos*
Department of Economics
UC San Diego
a2santos@ucsd.edu

January, 2010

Abstract

This paper develops practical methods for relaxing the missing at random assumption when estimating models of conditional quantiles with missing outcome data and discrete covariates. We restrict the degree of non-ignorable selection governing the missingness process by imposing bounds on the Kolmogorov-Smirnov (KS) distance between the distribution of outcomes among missing observations and the overall (unselected) distribution. Two methods are developed for conducting inference in this environment. The first allows us to perform finite sample inference on the identified set and is well suited to tests of model specification. The second enables us to conduct inference on the parameters of potentially misspecified models. To illustrate our techniques, we revisit the results of Angrist, Chernozhukov, and Fernández-Val (2006) regarding changes across Decennial Censuses in the quantile specific returns to schooling.

Keywords: Quantile regression, missing data, sensitivity analysis, misspecification, partial identification.

*We thank David Card, Justin McCrary, Hal White, and seminar participants at UC Berkeley for useful comments and Ivan Fernández-Val for assistance in replicating the results of Angrist, Chernozhukov, and Fernández-Val (2006).

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.

This paper develops practical methods for relaxing the missing at random (MAR) assumption when estimating models of conditional quantiles with missing outcome data and discrete covariates. 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 (Greenless, Reece, and Zieschang (1982) and Lillard, Smith, and Welch (1986)) or considered the worst case bounds on population moments that result when all assumptions about the missingness process are abandoned (Manski (1994, 2003)). Neither approach has garnered much popularity.[1] It is typically quite difficult to find variables that 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 intermediate approach allowing the researcher to assess the sensitivity of his or her conclusions regarding the conditional distribution of the data to modest deviations from MAR of the sort likely to be found in a given economic dataset. In particular we consider the identifying power of restrictions on the Kolmogorov-Smirnov (KS) distance between the distribution of outcomes among missing observations and the overall (unselected) distribution. The level of the bound on the KS metric yields a natural parametrization of deviations from ignorability, with a bound of zero corresponding to MAR and a bound of one yielding the totally unrestricted missingness process considered in Manski (1994). The resulting identified set for the conditional quantile

[1]See DiNardo, McCrary, and Sanbonmatsu (2006) for an applied example comparing these two approaches.

function (CQF) has a simple representation similar to the one obtained in Horowitz and Manski (1995) for processes with contaminated data but known bound on the probability of contamination.

Our strategy of nonparametrically bounding deviations from MAR is similar in spirit to Lee (2009) who imposes nonparametric restrictions on the missingness process designed for use with experiments and to Blundell, Gosling, Ichimura, and Meghir (2007) who develop nonparametric bounds on the distribution of offered wages implied by particular models of labor supply. Our approach is distinguished from these two in that we restrict an outcome of the selection process, which is potentially observable with auxiliary validation data, rather than some feature of the selection process itself, which in many cases is intrinsically unobservable without auxiliary model-based assumptions.[2] As such, our methods are strongly complementary with applied research documenting the extent of nonrandom missingness in particular datasets through use of validation data (e.g. David, Little, Samuhel, and Triest (1986), Grovers and Couper (1998), DiNardo, McCrary, and Sanbonmatsu (2006)).

Two methods are developed for conducting inference on the identified set. The first technique, which is useful for specification testing, satisfies the coverage requirement advocated in Imbens and Manski (2004) in finite samples.[3] The procedure operates by constructing finite sample confidence intervals for population quantiles within each covariate bin and then exploiting conditional independence across bins to obtain a confidence region for the entire identified set. This confidence region can be employed to construct confidence intervals for parametric specifications. For example, if we posit that the conditional quantile function is linear, a finite sample confidence interval is given by all linear specifications that lie within the estimated bounds. Failure to find any such lines represents a rejection of the linear specification.

Our second inferential procedure acknowledges the possibility that parametric models of conditional quantiles may be misspecified. It is convenient in such cases to be able to cover parameters that possess an interpretation as best parametric approximations to the true conditional quantile function as in Chamberlain (1994) and Angrist, Chernozhukov, and Fernández-Val (2006). Follow-

[2]Lee (2009), for example, explores the identifying power of restricting the effect of a randomly assigned treatment on outcome missingness to be one-sided – i.e. for treatment to either encourage or discourage nonresponse among all sample members. In doing so he rules out the existence of a subpopulation of "defiers" who respond to treatment in an opposite manner from the bulk of the population. This assumption, though perhaps quite appropriate for the program he studies is not directly refutable.

[3]The identified set we study could be analyzed using results from the existing literature on moment inequalities; see Chernozhukov, Hong, and Tamer (2007), Andrews and Soares (2007), Romano and Shaikh (2008) and references therein. However, our focus on conditional quantiles allows us to conduct finite sample inference and hence avoid the complications inherent in the asymptotic analysis of partially identified models.

ing Horowitz and Manski (2006), Stoye (2007), and Ponomareva and Tamer (2009), we extend the concept of "pseudo-true" approximating parameters White (1980, 1982) to a partially identified setting. In particular, we characterize the set of parameter values constituting the best approximation to *some* CQF in the identified set.

In the cases we consider, the pseudo-true conditional quantile is given by the projection of the CQF onto the linear subspace spanned by the covariates. Hence, when point identification of the CQF fails due to missingness, the identified set of pseudo true parameters consists of all coefficients associated with the best approximation to a function that lies within the CQF bounds. We obtain sharp bounds on the coordinates of the pseudo true parameter vector and show that they have a representation as the solutions to a pair of linear programming problems.[4] This result suggests simple estimators for these bounds, which we show converge in distribution to a Gaussian process indexed by the quantile of interest and the level of the KS restriction on the missingness process. In addition, we establish the consistency of a weighted bootstrap for estimating the limiting distribution of the process, which enables us to construct uniform confidence intervals. This latter feature is particularly useful for sensitivity analysis as it enables researchers to determine a critical level of the KS bound for which a given hypothesis cannot be rejected.

To empirically assess the likely magnitude of deviations from ignorability in actual data we analyze a sample of workers from the 1973 Current Population Survey for which IRS earnings records are available. This sample allows us to observe the IRS 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 nonresponse to the CPS question within narrow demographic covariate bins, with very high and very low earning individuals most likely to have invalid CPS earnings records. Having rejected the MAR assumption, we develop a method of estimating the appropriate choice of KS bounds on the earnings nonresponse process. We find the data are consistent with nearly random earnings nonresponse within covariate bins, suggesting that a worst case bounds analysis would be vastly over-conservative.

We then proceed to illustrate our inferential methods by revisiting the results of Angrist, Chernozhukov, and Fernández-Val (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, we are able to reject the simple Mincerian model used for the conditional quantiles of log earnings.

[4] See Stoye (2007) for a similar conclusion for approximations to conditional expectations.

Having rejected the Mincer specification as a literal description of the data generating process we proceed to bound the pseudo-true quantile process governing returns to schooling. We conclude 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. 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.

The remainder of the paper is structured as follows: Section 2 characterizes the identified set for the conditional quantile function under the KS bound and Section 3 develops methods for performing finite sample inference on this set. Section 4 contains the results concerning pseudo-true parameters. In Section 5 we present our empirical study and briefly conclude in Section 6.

2 Identified Set

We consider a triplet of random variables (Y_i, X_i, D_i) where $Y_i \in \mathbf{R}$, $X_i \in \mathbf{R}^l$ and $D_i \in \{0,1\}$ is a dummy variable that equals one if Y_i is observable and zero otherwise. Denote the distribution of Y_i given X_i and the distribution of Y_i given X_i and D_i respectively as:

$$F_{y|x}(c) \equiv P(Y_i \leq c|X_i = x) \qquad F_{y|d,x}(c) \equiv P(Y_i \leq c|D_i = d, X_i = x) \ , \tag{1}$$

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

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

It is well known that without additional assumptions $F_{y|x}$ is not identified in the presence of missing data, but can be restricted to lie in an identified set (see Manski (2003)). As a result, it is still possible to derive potentially informative bounds on the conditional quantiles of Y_i given X_i. We explore the nature of these bounds under the following assumptions on (Y_i, X_i, D_i):

Assumption 2.1. *(i) $X_i \in \mathbf{R}^l$ 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_i equals one if Y_i is observable and zero otherwise.*

The discrete support requirement in Assumption 2.1(i) allows us to perform finite sample inference and to avoid the problem of bandwidth selection that accompanies most nonparametric estimators. 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^{th} conditional quantile of Y_i given X_i, denoted $c_\tau(x)$, is the unique number satisfying $F_{y|x}(c_\tau(x)) = \tau$.

The identified set for $c_\tau(x)$ has been studied under different assumptions on the missingness process. Manski (1994), for example, derives the identified set that results from making no assumptions on $F_{y|0,x}$. He notes that since $0 \leq F_{y|0,x}(c_\tau(x)) \leq 1$, the law of total probability implies that $c_\tau(x)$ must satisfy the inequalities:

$$F_{y|1,x}(c_\tau(x)) \times p(x) \leq F_{y|x}(c_\tau(x)) \leq F_{y|1,x}(c_\tau(x)) \times p(x) + \{1 - p(x)\} \ . \tag{3}$$

Under Assumption 2.1(ii), the identified set for $c_\tau(x)$ is given by the set of functions that agree with the observable implications of (3). We denote this identified set by:

$$\mathcal{C}_\tau = \{\theta : \mathcal{X} \to \mathbf{R} \ : \tau - \{1 - p(x)\} \leq F_{y|1,x}(\theta(x)) \times p(x) \leq \tau\} \ . \tag{4}$$

In most cases, however, these worst case bounds will be too conservative, allowing for selection mechanisms which, on prior grounds, are thought to be implausible. Few economists, for example, worry that surveys are so poorly designed (or respondents so strategic) that *all* nonrespondents have earnings above the conditional median of their demographic group.

For applied researchers interest usually centers on considering the effects of relatively mild deviations from missingness at random. It is useful then to devise a metric for comparing selection mechanisms in terms of the deviations they generate from ignorability. We propose use of the following nonparametric metric: the Kolmogorov-Smirnov distance between the distribution of Y_i conditional on X_i, and the distribution of Y_i conditional on X_i and Y_i being missing. More precisely, we examine the set of conclusions that may be drawn under the following additional assumption:

Assumption 2.2. $\sup_{x \in \mathcal{X}} KS(F_{y|x}, F_{y|0,x}) \leq k$*, where* $KS(F_{y|x}, F_{y|0,x}) \equiv \sup_{c \in \mathbf{R}} |F_{y|x}(c) - F_{y|0,x}(c)|$.

Assumption 2.2 is satisfied by a nonparametric family of selection models that yield conditional quantiles in the unselected population close to those in the overall population in a well-defined sense. By imposing this restriction, we depart from the conventional econometric practice of first positing a selection mechanism and then deriving its implications for observed data. For comparison with this approach Appendix A provides a numerical example mapping the parameters of a bivariate normal selection model into values of our KS distance bound k. Our methods are tailored to situations in which little is known about the selection mechanism. If a researcher has prior knowledge of features of the selection mechanism additional restrictions can, in principle, be incorporated into the analysis. We leave such extensions for future work.

Since $c_\tau(x)$ is the conditional quantile of Y_i given X_i, Assumption 2.2 implies that:

$$\max\{\tau - k, 0\} \leq F_{y|x,0}(c_\tau(x)) \leq \min\{\tau + k, 1\} \ , \tag{5}$$

or equivalently, that $c_\tau(x)$ must be between the $\max\{\tau - k, 0\}$ and $\min\{\tau + k, 1\}$ quantile of Y_i conditional on X_i and Y_i being missing.[5] Applying this relationship, it is possible to obtain a sharp characterization of the identified set for $c_\tau(x)$ under Assumptions 2.1(ii)-(iii) and 2.2.

Lemma 2.1. *If Assumptions 2.1(ii)-(iii) and 2.2 hold, then the identified set for $c_\tau(x)$ is:*

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

As Lemma 2.1 shows, imposing bounds on the Kolmogorov-Smirnov distance between $F_{y|0,x}$ and $F_{y|x}$ yields a simple parametrization for $\mathcal{C}_\tau^k$ as a function of k. The set $\mathcal{C}_\tau^k$ is increasing in k, with identification occurring at $k = 0$ (the missing at random case), and $k = 1$ yielding the opposite extreme with $\mathcal{C}_\tau^1 = \mathcal{C}_\tau$ (the worst case bounds). By studying the set of conclusions that may be reached under different choices of k, it is possible to analyze what degree of selection is necessary to overturn conclusions obtained under a missing at random assumption.

We conclude the discussion of the identified set with an alternative representation of $\mathcal{C}_\tau^k$.

Corollary 2.1. *Suppose Assumptions 2.1(ii)-(iii), 2.2 hold and let $F_{y|1,x}^{-}(q) = F_{y|1,x}^{-1}(q)$ if $0 < q < 1$, $F_{y|1,x}^{-}(q) = -\infty$ if $q \leq 0$ and $F_{y|1,x}^{-}(q) = \infty$ if $q \geq 1$. If the bounds $(c_{\tau,L}^k(x), c_{\tau,U}^k(x))$ are given by:*

$$c_{\tau,L}^k(x) \equiv F_{1,x}^{-}\Big(\frac{\tau - \min\{\tau + k, 1\}(1 - p(x))}{p(x)}\Big) \qquad c_{\tau,U}^k(x) \equiv F_{1,x}^{-}\Big(\frac{\tau - \max\{\tau - k, 0\}(1 - p(x))}{p(x)}\Big) \ ,$$

then it follows that $\mathcal{C}_\tau^k = \{\theta : \mathcal{X} \to \mathbf{R} : c_{\tau,L}^k(x) \leq \theta(x) \leq c_{\tau,U}^k(x)\}$.

The characterization of $\mathcal{C}_\tau^k$ as the set of functions satisfying particular pointwise bounds provides a simple geometric interpretation of the identified set. Both the representation of $\mathcal{C}_\tau^k$ obtained in Lemma 2.1 and in Corollary 2.1 will prove useful when studying how to conduct inference.

3 CQF Inference

An appealing feature of quantiles is the ability to perform finite sample inference on them in the presence of an *i.i.d.* sample; see Chernozhukov, Hansen, and Jansson (2009). In this section, we

[5] In a different context, this restriction has also been applied to copulas by Hoderlein and Stoye (2009).

exploit this property to develop a simple computational procedure for conducting inference on $\mathcal{C}_\tau^k$. For a specified desired level of coverage $1-\alpha$, we construct a confidence collection $\hat{\mathcal{C}}_\tau^k$ satisfying:

$$P(\theta \in \hat{\mathcal{C}}_\tau^k \;|\{X_i\}_{i=1}^n) \geq 1-\alpha \ , \tag{6}$$

for all functions θ in the identified set $\mathcal{C}_\tau^k$. The coverage requirement in (6) was originally proposed in Imbens and Manski (2004) and is weaker than the one in Chernozhukov, Hong, and Tamer (2007) or Romano and Shaikh (2010), who explore methods for covering the entire identified set.

By Corollary 2.1, the identified set $\mathcal{C}_\tau^k$ may be characterized as the set of functions $\theta : \mathcal{X} \to \mathbf{R}$ that satisfy certain pointwise bounds on $x \in \mathcal{X}$. These pointwise bounds determine intervals $[c_{\tau,L}^k(x), c_{\tau,U}^k(x)]$ which constitute the identified set for the τ^{th} quantile of Y_i conditional on X_i being equal to x. For computational simplicity, we focus on confidence sets that are of the similar form:

$$\hat{\mathcal{C}}_\tau^k = \{\theta : \mathcal{X} \to \mathbf{R} : \hat{c}_{\tau,L}^k(x) \leq \theta(x) \leq \hat{c}_{\tau,U}^k(x) \ \forall x \in \mathcal{X}\} \ , \tag{7}$$

for appropriate choices of $\hat{c}_{\tau,L}^k(x)$ and $\hat{c}_{\tau,U}^k(x)$. We construct these bounds so that (i) $\hat{c}_{\tau,L}^k(x)$ and $\hat{c}_{\tau,U}^k(x)$ are independent of $\hat{c}_{\tau,L}^k(x')$ and $\hat{c}_{\tau,U}^k(x')$ conditional on $\{X_i\}_{i=1}^n$ for $x \neq x'$; and (ii) $[\hat{c}_{\tau,L}^k(x), \hat{c}_{\tau,U}^k(x)]$ provides coverage of the interval $[c_{\tau,L}^k(x), c_{\tau,U}^k(x)]$ at a prespecified level $1-\alpha_x$ as in Imbens and Manski (2004). Due to independence, the coverage probability of $\hat{\mathcal{C}}_\tau^k$ can in turn be accurately controlled by the individual coverage probabilities $1-\alpha_x$, thus ensuring (6) is satisfied.

3.1 The Bounds $\hat{c}_{\tau,L}^k(x)$ and $\hat{c}_{\tau,U}^k(x)$

We aim to obtain pointwise bounds $\hat{c}_{\tau,L}^k(x)$ and $\hat{c}_{\tau,U}^k(x)$ such that for a specified level of coverage $1-\alpha_x$ and all values c with $c_{\tau,L}^k(x) \leq c \leq c_{\tau,U}^k(x)$, it follows that:

$$P\left(\hat{c}_{\tau,L}^k(x) \leq c \leq \hat{c}_{\tau,U}^k(x) \;|\{X_i\}_{i=1}^n\right) \geq 1-\alpha_x \ . \tag{8}$$

Bounds $\hat{c}_{\tau,L}^k(x)$ and $\hat{c}_{\tau,U}^k(x)$ satisfying (8) can be constructed exploiting the dual characterizations of $\mathcal{C}_\tau^k$ obtained in Lemma 2.1 and Corollary 2.1. In particular, defining the parameter space:

$$\Gamma_{\tau,x}^k \equiv \{(\gamma,\xi) \in [0,1]^2 : \tau - \min\{\tau+k,1\} \times \gamma \leq \xi \leq \tau - \max\{\tau-k,0\} \times \gamma \text{ and } \gamma + \xi \leq 1\} \ , \tag{9}$$

it follows that for any $c \in \mathbf{R}$, the inequalities $c_{\tau,L}^k(x) \leq c \leq c_{\tau,U}^k(x)$ are satisfied, if and only if $(1-p(x), F_{y|1,x}(c)p(x)) \in \Gamma_{\tau,x}^k$. Consequently, bounds $\hat{c}_{\tau,L}^k(x)$ and $\hat{c}_{\tau,U}^k(x)$ satisfying the coverage requirement in (8) may be obtained through test inversion of the hypothesis:

$$H_0 : (1-p(x), F_{y|1,x}(c)p(x)) \in \Gamma_{\tau,x}^k \qquad\qquad H_1 : (1-p(x), F_{y|1,x}(c)p(x)) \notin \Gamma_{\tau,x}^k \ . \tag{10}$$

In turn, because the hypothesis in (10) concerns conditional probabilities and X_i has finite support, the distribution of test statistics conditional on $\{X_i\}_{i=1}^n$ can be readily simulated under the null hypothesis in order to attain finite sample control of size, as required for satisfying (8).

A number of different test statistics may be employed in order to examine the hypothesis in (10). For concreteness, we employ a standard Likelihood Ratio test and denote the relevant test statistic by $T_{n,x}(c)$. The following algorithm outlines how to obtain the bounds $\hat{c}_{\tau,L}(x)$ and $\hat{c}_{\tau,U}(x)$ and we refer the reader to the appendix for computational details.

STEP 1: Obtain the critical value $r(\alpha_x)$ that delivers finite sample control in size by:

$$r(\alpha_x) \equiv \{\inf r : \sup_{(\gamma,\xi)\in\Gamma^k_{\tau,x}} P_{\gamma,\xi}\left(T_{n,x}(c) > r \mid \{X_i\}_{i=1}^n\right) \leq \alpha_x\} \; , \tag{11}$$

where $P_{\gamma,\xi}$ denotes the distribution of $T_{n,x}(c)$ if $(1-p(x), F_{y|1,x}(c)p(x)) = (\gamma, \xi)$. Since the constraint set $\Gamma^k_{\tau,x}$ does not depend on c, the critical value $r(\alpha_x)$ *does not* depend on c either. ■

STEP 2: Letting $t_{n,x}(c)$ denote the actual realization of $T_{n,x}(c)$ in the sample, we obtain:

$$\hat{c}^k_{\tau,L}(x) \equiv \{\inf c : t_{n,x}(c) \leq r(\alpha_x)\} \qquad\qquad \hat{c}^k_{\tau,U}(x) \equiv \{\sup c : t_{n,x}(c) \leq r(\alpha_x)\} \; . \tag{12}$$

The estimated bound $\hat{c}^k_{\tau,L}(x)/\hat{c}^k_{\tau,U}(x)$ may potentially equal negative/positive infinity. In particular, this will occur whenever $t_{n,x}(c)$ evaluated at c equal to the smallest/largest observation of Y_i on the set of the sample for which $X_i = x$, is smaller/larger than $r(\alpha_x)$. ■

As the following Lemma shows, the resulting bounds $\hat{c}^k_{\tau,L}(x)$ and $\hat{c}^k_{\tau,U}(x)$ indeed satisfy (8).

Lemma 3.1. *If $\{Y_i, X_i, D_i\}$ are i.i.d. and Assumptions 2.1 and 2.2 hold, then it follows that:*

$$P\left(\hat{c}^k_{\tau,L}(x) \leq c \leq \hat{c}^k_{\tau,U}(x) | \{X_i\}_{i=1}^n\right) \geq 1 - \alpha_x \; ,$$

for all $c \in \mathbf{R}$ such that $c^k_{\tau,L}(x) \leq c \leq c^k_{\tau,U}(x)$.

3.2 The Set $\hat{\mathcal{C}}^k_\tau$

The confidence collection $\hat{\mathcal{C}}^k_\tau$, as in (7), may be interpreted to be the *joint* confidence interval implied by the product of the marginal confidence regions $[\hat{c}^k_{\tau,L}(x), \hat{c}^k_{\tau,U}(x)]$ for the intervals $[c^k_{\tau,L}(x), c^k_{\tau,U}(x)]$ at each $x \in \mathcal{X}$. However, since conditional on $\{X_i\}_{i=1}^n$ the estimated bounds $\hat{c}^k_{\tau,L}(x)$ and $\hat{c}^k_{\tau,U}(x)$ are independent across $x \in \mathcal{X}$, the coverage level of the joint confidence intervals can be accurately controlled by that of the marginals. Indeed, as the following Lemma shows, the resulting $\hat{\mathcal{C}}^k_\tau$ satisfies the desired coverage requirement in (6) provided that the coverage requirement for each bound, as in (8), is properly selected.

Lemma 3.2. *If $\{Y_i, D_i, X_i\}_{i=1}^n$ are i.i.d., Assumptions 2.1, 2.2 hold and $\Pi_x(1-\alpha_x) = 1-\alpha$, then:*

$$\inf_{\theta\in\mathcal{C}_\tau^k} P(\theta \in \hat{\mathcal{C}}_\tau^k \ |\{X_i\}_{i=1}^n) \geq 1-\alpha \ .$$

An appealing consequence of $\hat{\mathcal{C}}_\tau^k$ being characterized by a set of pointwise bounds is the computational ease with which it enables us construct confidence intervals for parametric specifications of $c_\tau(x)$. For example, a standard model for $c_\tau(x)$ is to assume linearity, so that:

$$c_\tau(x) = x'\beta_\tau \tag{13}$$

for some $\beta_\tau \in \mathbf{R}^l$. Under the assumption that (13) holds, the relevant identified set may be further restricted to be the intersection of $\mathcal{C}_\tau^k$ with the set of linear functions. Often, of special interest in this context is a particular coordinate of β_τ or the conditional quantile evaluated at a particular value of the covariates. Both these quantities may be expressed as $\lambda'\beta_\tau$ for some known vector $\lambda \in \mathbf{R}^l$. The following Lemma characterizes the identified set for parameters of the form $\lambda'\beta_\tau$:

Lemma 3.3. *Suppose Assumptions 2.1(ii)-(iii) and 2.2 hold, and in addition define:*

$$\phi_{\tau,L}^k \equiv \inf_{\beta\in\mathbf{R}^l} \lambda'\beta \qquad s.t.\ c_{\tau,L}^k(x) \leq x'\beta \leq c_{\tau,U}^k(x) \ \ \forall x \in \mathcal{X} \tag{14}$$

$$\phi_{\tau,U}^k \equiv \sup_{\beta\in\mathbf{R}^l} \lambda'\beta \qquad s.t.\ c_{\tau,L}^k(x) \leq x'\beta \leq c_{\tau,U}^k(x) \ \ \forall x \in \mathcal{X} \tag{15}$$

The identified set for $\lambda'\beta$ is then given by the interval $[\phi_{\tau,L}^k, \phi_{\tau,U}^k]$.

While Lemma 3.3 establishes the identified set for $\lambda'\beta_\tau$ is an interval with endpoints characterized as the solution to linear programs, the structure of $\hat{\mathcal{C}}_\tau^k$ enables us to obtain a confidence region for $\lambda'\beta_\tau$ employing immediate sample analogues. Specifically, defining:

$$\hat{\phi}_{\tau,L}^k \equiv \inf_{\beta\in\mathbf{R}} \lambda'\beta \qquad \text{s.t. } \hat{c}_{\tau,L}^k(x) \leq x'\beta \leq \hat{c}_{\tau,U}^k(x) \ \ \forall x \in \mathcal{X} \tag{16}$$

$$\hat{\phi}_{\tau,U}^k \equiv \sup_{\beta\in\mathbf{R}} \lambda'\beta \qquad \text{s.t. } \hat{c}_{\tau,L}^k(x) \leq x'\beta \leq \hat{c}_{\tau,U}^k(x) \ \ \forall x \in \mathcal{X} \tag{17}$$

we obtain a valid confidence region in $[\hat{\phi}_{\tau,L}^k, \hat{\phi}_{\tau,U}^k]$. It is important to note that the optimization problems in (16) and (17) may be infeasible. This implies there are no linear functions in $\hat{\mathcal{C}}_\tau^k$ and constitutes a rejection of the linear model. For moderate values of k, this is precisely the outcome of our analysis of returns to schooling in Section 5.

The following Corollary establishes the coverage properties of the interval $[\hat{\phi}_{\tau,L}^k, \hat{\phi}_{\tau,U}^k]$.

Corollary 3.1. *If $\{Y_i, X_i, D_i\}_{i=1}^n$ are i.i.d. and Assumptions 2.1-2.2 hold, then it follows that:*

$$P(\hat{\phi}_{\tau,L}^k \leq \phi \leq \hat{\phi}_{\tau,U}^k \ |\{X_i\}_{i=1}^n) \geq 1-\alpha$$

for any $\phi \in \mathbf{R}$ such that $\phi_{\tau,L}^k \leq \phi \leq \phi_{\tau,U}^k$ and $\hat{\mathcal{C}}_\tau^k$ is as in Lemma 3.2.

4 Misspecification

Parametric models can serve as useful devices for summarizing complex multivariate relationships. However, in practice, most parametric models are misspecified. In such situations the methods developed in the previous section may yield misleading results. Figure 1 illustrates a case where the nonparametric identified set for the conditional quantile function 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."

Hence, the methods developed thus far confront the applied researcher with a difficult choice. One can either conduct a fully nonparametric analysis of the 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 often 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 Angrist, Chernozhukov, and Fernández-Val (2006). In this section we extend this approach to the present partially identified setting and develop methods for estimating parametric models that acknowledge the possibility of misspecification.

Our strategy is to develop bounds on parametric approximations to the CQF rather than the CQF itself and then to derive the corresponding sampling theory of those bounds. We focus on linear parametric models and approximations that minimize a known quadratic loss function. For S a known measure on $\mathcal{X}$ and $E_S[g(X_i)]$ denoting expectation of a function $g(x)$ under S, we define the pseudo true parameter to be:[6]

$$\beta^*_\tau \equiv \arg\min_{\beta \in \mathbf{R}^l} E_S[(c_\tau(X_i) - X_i'\beta)^2] \ . \tag{18}$$

Lack of identification of the conditional quantile function $c_\tau(x)$ due to missing data implies lack of identification of the pseudo true parameter β^*_τ. We therefore consider the identified set of pseudo true parameters to be the set of parameter values that constitute a best approximation to *some*

[6]An important task of the applied researcher is to specify the measure S which weights the squared deviations in each x bin. This is an inherently context-specific task depending entirely upon the researcher's objectives. In Section 5 we weight the deviations by sample size which approximates the impact on an average sample member. Other schemes (including equal weighting) may also be of interest in some settings.

conditional quantile function in the identified set $\mathcal{C}^k_\tau$:

$$\mathcal{P}^k_\tau \equiv \{\beta^* \in \mathbf{R}^l : \beta^* = \arg\min_{\beta \in \mathbf{R}^l} E_S[(\theta(X_i) - X_i'\beta)^2] \ \text{ for some } \theta \in \mathcal{C}^k_\tau\} \ . \tag{19}$$

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

Lemma 4.1. *If Assumptions 2.1(ii)-(iii), 2.2 hold and $E_S[X_iX_i']$ is invertible, then it follows that:*

$$\mathcal{P}^k_\tau = \{\beta^* \in \mathbf{R}^l : \beta^* = (E_S[X_iX_i'])^{-1}E_S[X_i\theta(X_i)] \ \ s.t. \ c^k_{\tau,L}(x) \leq \theta(x) \leq c^k_{\tau,U}(x) \ \forall x \in \mathcal{X}\} \ .$$

It follows from Lemma 4.1 that $\mathcal{P}^k_\tau$ is convex, a property that is not immediate from its definition in (19). As in Section 3, we focus on studying the identified set for parameters of the form $\lambda'\beta^*_\tau$ for some known vector $\lambda \in \mathbf{R}^l$. The analysis hence includes both coordinate values of β^*_τ as well as pseudo true quantiles evaluated at a particular choice of covariates as special cases. Using Lemma 4.1 it is possible to show that the identified set for parameters of the form $\lambda'\beta^*_\tau$ is an interval with endpoints characterized as the solution to linear programming problems (see Stoye (2007)).

Corollary 4.1. *Suppose Assumptions 2.1(ii)-(iii), 2.2 hold, $E_S[X_iX_i']$ is invertible and define:*

$$\pi_L(\tau,k) \equiv \inf_{\beta^* \in \mathcal{P}^k_\tau} \lambda'\beta^* = \inf_\theta \lambda'(E_S[X_iX_i'])^{-1}E_S[X_i\theta(X_i)] \ \ s.t. \ c^k_{\tau,L}(x) \leq \theta(x) \leq c^k_{\tau,U}(x) \tag{20}$$

$$\pi_U(\tau,k) \equiv \sup_{\beta^* \in \mathcal{P}^k_\tau} \lambda'\beta^* = \sup_\theta \lambda'(E_S[X_iX_i'])^{-1}E_S[X_i\theta(X_i)] \ \ s.t. \ c^k_{\tau,L}(x) \leq \theta(x) \leq c^k_{\tau,U}(x) \ . \tag{21}$$

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

It is worth pointing out that while the identified set $\mathcal{P}^k_\tau$ is a clear extension of the concept of a pseudo-true parameter to a partially identified context, it is not necessarily the obvious one. Under the assumption that the true conditional quantile function is linear, it is possible to further restrict $\mathcal{C}^k_\tau$ by including only the linear functions that satisfy the pointwise bounds of Corollary 2.1. As a result, letting $(a)_+ \equiv \max\{0,a\}$, we may characterize the identified set under the assumption of linearity as the set of zeroes to the criterion function:

$$E_S[(c^k_{\tau,L}(X_i) - X_i'\beta)^2_+ + (X_i'\beta - c^k_{\tau,U}(X_i))^2_+] \ . \tag{22}$$

Under identification pseudo-true parameters are often characterized as the minimizers of a criterion function with minimum equal to zero under proper specification, but positive minimum otherwise. Hence, while we study $\mathcal{P}_\tau^k$, we could have alternatively defined the relevant set of pseudo true parameters to be the set of minimizers of (22), denoted by:

$$\mathcal{A}_\tau^k \equiv \{\beta^* \in \mathbf{R}^l : \beta^* \in \arg\min_{\beta\in\mathbf{R}^l} E_S[(c_{\tau,L}^k(X_i) - X_i'\beta)_+^2 + (X_i'\beta - c_{\tau,U}^k(X_i))_+^2]\} \ . \tag{23}$$

If the model is identified, then both definitions agree and $\mathcal{P}_\tau^k = \mathcal{A}_\tau^k$. However, important differences arise when the model is partially identified. The set $\mathcal{A}_\tau^k$ may be alternatively expressed as:

$$\mathcal{A}_\tau^k = \{\beta^* \in \mathbf{R}^l : \beta^* \in \arg\min_{\beta\in\mathbf{R}^l}\{\min_{\theta\in\mathcal{C}_\tau^k} E_S[(\theta(X_i) - X_i'\beta)^2]\}\} \ . \tag{24}$$

Therefore, the elements of $\mathcal{A}_\tau^k$ are those that minimize the distance between $\mathcal{C}_\tau^k$ and the set of linear functions, while $\mathcal{P}_\tau^k$ consists of *all* linear models that minimize distance to *some* conditional quantile function in $\mathcal{C}_\tau^k$. Because interest in applied work is in the best approximation to the true conditional function, which may not be in $\mathcal{A}_\tau^k$ but is guaranteed to be in $\mathcal{P}_\tau^k$, we develop inferential procedures for $\mathcal{P}_\tau^k$ rather than $\mathcal{A}_\tau^k$. See Ponomareva and Tamer (2009) for an excellent discussion of this issue in the context of estimating conditional expectations with interval valued dependent variables.

4.1 Bounds Estimation

In what follows we study the asymptotic properties of estimators for $\pi_L(\tau,k)$ and $\pi_U(\tau,k)$ as in (20) and (21). To enable inference on the entire quantile process, we examine the asymptotic distribution of these estimators uniformly in both quantile τ and KS bound k.

For particular choices of τ and k, the bounds on the conditional quantile function may be infinite in turn leading to an unbounded identified set for the pseudo true parameter. We focus our analysis on choices of τ and k for which the bounds $\pi_L(\tau,k)$ and $\pi_U(\tau,k)$ are in fact finite. For this reason, we restrict our analysis to values of τ and k that belong to the set:

$$\mathcal{B} \equiv \{(\tau,k) \in [0,1]^2 : \min\{\tau+k,1\}\times\{1-p(x)\}+\epsilon \leq \tau \leq p(x)+\max\{\tau-k,0\}\times\{1-p(x)\}-\epsilon \ \forall x \in \mathcal{X}\}$$

for some ϵ satisfying $0 < 2\epsilon < \inf_{x\in\mathcal{X}} p(x)$. Provided that the conditional probability of missing is bounded away from one, the set $\mathcal{B}$ is nonempty since it contains the MAR analysis as the special case $k = 0$. 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 $f'_{y|1,x}(x)$; (iv) $E_S[X_i X_i']$ is invertible; (v) $f_{y|1,x}(c)$ is bounded away from zero uniformly on all c satisfying $\epsilon \leq F_{y|1,x}(c)p(x) \leq p(x) - \epsilon \; \forall x \in \mathcal{X}$.*

For any (τ, k), it is possible to construct a confidence interval for $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$ by employing the confidence region $\hat{\mathcal{C}}^k_\tau$. However, Corollary 4.1 implies the bounds $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$ are linear combinations of $c^k_{\tau,L}(x)$ and $c^k_{\tau,U}(x)$ evaluated at different $x \in \mathcal{X}$. An inferential procedure that requires correct inference on *all* bounds $c^k_{\tau,L}(x)$ and $c^k_{\tau,U}(x)$, as $\hat{\mathcal{C}}^k_\tau$ provides, will therefore prove conservative relative to a procedure that focuses directly on the desired linear combinations. For this reason, our strategy for estimating the bounds $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$ instead consists of first obtaining estimates $\breve{c}^k_{\tau,L}(x)$ and $\breve{c}^k_{\tau,U}(x)$ of the conditional quantile bounds and then employing them in place of $c^k_{\tau,L}(x)$ and $c^k_{\tau,U}(x)$ in the linear programming problems given in (20) and (21). We study the estimation of the conditional quantile bounds $c^k_{\tau,L}(x)$ and $c^k_{\tau,U}(x)$ in the context of a general M-estimation problem characterized by the family of population criterion functions:

$$Q_x(c|\tau, b) \equiv \left(P(Y_i \leq c, X_i = x, D_i = 1) + bP(D_i = 0, X_i = x) - \tau P(X_i = x)\right)^2 . \qquad (25)$$

At points $(\tau, k) \in \mathcal{B}$, the bounds $c^k_{\tau,L}(x)$ and $c^k_{\tau,U}(x)$ can be expressed as the unique minimizers of $Q_x(c|\tau, b)$ for a specific choice of parameters (τ, b). In particular, it follows from Lemma 2.1 that:

$$c^k_{\tau,L}(x) = \arg\min_{c \in \mathbf{R}} Q_x(c|\tau, \min\{\tau + k, 1\}) \qquad c^k_{\tau,U}(x) = \arg\min_{c \in \mathbf{R}} Q_x(c|\tau, \max\{\tau - k, 0\}) , \qquad (26)$$

and hence there exists a direct relationship between the bounds $c^k_{\tau,L}(x)$ and $c^k_{\tau,U}(x)$ as indexed by (τ, k) and the minimizers of the criterion function $Q_x(c|\tau, b)$ as indexed by (τ, b).

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

$$Q_{x,n}(c|\tau, b) \equiv \Big(\frac{1}{n}\sum_{i=1}^{n}\{1\{Y_i \leq c, X_i = x, D_i = 1\} + b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\}\}\Big)^2 . \qquad (27)$$

Exploiting (26), the extremum estimators for the bounds $c^k_{\tau,L}(x)$ and $c^k_{\tau,U}(x)$ are then given by:

$$\breve{c}^k_{\tau,L}(x) \in \arg\min_{c \in \mathbf{R}} Q_{x,n}(c|\tau, \min\{\tau + k, 1\}) \qquad \breve{c}^k_{\tau,U}(x) \in \arg\min_{c \in \mathbf{R}} Q_{x,n}(c|\tau, \max\{\tau - k, 0\}) . \qquad (28)$$

By studying the criterion function $Q_{x,n}(c|\tau, b)$ as a stochastic process indexed by (τ, b) it is possible to derive the asymptotic behavior of the estimators $\breve{c}^k_{\tau,L}(x)$ and $\breve{c}^k_{\tau,U}(x)$ uniformly in the parameters $(\tau, k) \in \mathcal{B}$. However, since our primary focus is on the estimation of $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$, we relegate this intermediate result to the appendix.

Employing $\check{c}^k_{\tau,L}(x), \check{c}^k_{\tau,U}(x)$ we can in turn obtain estimates of $\pi_L(\tau,k), \pi_U(\tau,k)$ through the sample analogues to the linear programming problems given in (20) and (21):

$$\hat{\pi}_L(\tau,k) \equiv \inf_{\theta} \lambda'(E_S[X_iX_i'])^{-1}E_S[X_i\theta(X_i)] \quad \text{s.t. } \check{c}^k_{\tau,L}(x) \leq \theta(x) \leq \check{c}^k_{\tau,U}(x) \tag{29}$$

$$\hat{\pi}_U(\tau,k) \equiv \sup_{\theta} \lambda'(E_S[X_iX_i'])^{-1}E_S[X_i\theta(X_i)] \quad \text{s.t. } \check{c}^k_{\tau,L}(x) \leq \theta(x) \leq \check{c}^k_{\tau,U}(x) \ . \tag{30}$$

Under assumptions 2.1, 2.2 and 4.1, the uniform asymptotic behavior of $\check{c}^k_{\tau,L}(x)$ and $\check{c}^k_{\tau,U}(x)$ is inherited by $\hat{\pi}_L(\tau,k)$ and $\hat{\pi}_U(\tau,k)$. The following theorem establishes this point by obtaining the asymptotic distribution of these estimators uniformly both in τ and k.

Theorem 4.1. *If Assumptions 2.1, 2.2, 4.1 hold and $\{Y_i, X_i, D_i\}_{i=1}^n$ is an i.i.d. sample, then:*

$$\sqrt{n}\begin{pmatrix} \hat{\pi}_L(\tau,k) - \pi_L(\tau,k) \\ \hat{\pi}_U(\tau,k) - \pi_U(\tau,k) \end{pmatrix} \xrightarrow{\mathcal{L}} G(\tau,k) \ , \tag{31}$$

where $G(\tau,k)$ is a gaussian process on $L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$.

4.2 Inference

For every fixed k, the functions $\pi_L(\tau,k)$ and $\pi_U(\tau,k)$ constitute the lower and upper envelopes for the identified set of the pseudo-true quantile process $\lambda'\beta^*_\tau$ under a KS bound of k. In this section, we employ Theorem 4.1 to devise an inferential procedure for the set of functions that lie between these envelopes, which we denote by:

$$\mathcal{G} \equiv \{g : \mathcal{B} \to \mathbf{R} : \pi_L(\tau,k) \leq g(\tau,k) \leq \pi_U(\tau,k) \text{ for all } (\tau,k) \in \mathcal{B}\} \ . \tag{32}$$

It is important to note that while $\pi_L(\tau,k)$ and $\pi_U(\tau,k)$ are themselves in the identified set for the pseudo-true quantile process $\lambda'\beta^*_\tau$ for all k, not all functions in $\mathcal{G}$ are. For example, under our assumptions $\lambda'\beta^*_\tau$ must be continuous in τ and hence functions in $\mathcal{G}$ that are discontinuous in τ do not belong in the identified set for the pseudo-true quantile process $\lambda'\beta^*_\tau$. However, the set $\mathcal{G}$ does have the favorable properties of (i) containing the true identified set so that processes not in $\mathcal{G}$ are also known not to be in the identified set; (ii) sharpness of the lower and upper envelopes $\pi_L(\tau,k)$ and $\pi_U(\tau,k)$; (iii) ease of analysis and graphical representation. We leave the development and analysis of sharp bounds on the identified set for pseudo-true quantile processes to future work.[7]

[7] To the best of our knowledge, the only work studying inference on identified sets in function spaces is in the nonparametric instrumental variables literature, where the structure of the problem implies simple characterizations of the identified set (Santos (2007a,b)).

A natural way to construct a confidence region for $\mathcal{G}$ is by "extending" $\hat{\pi}_L(\tau,k)$ and $\hat{\pi}_U(\tau,k)$ by an appropriate amount. We hence consider confidence regions that are of the general form:

$$\hat{\mathcal{G}}(\omega,r) \equiv \{g : \mathcal{B} \to \mathbf{R} : \hat{\pi}_L(\tau,k) - \frac{r}{\sqrt{n}}\omega_L(\tau,k) \leq g(\tau,k) \leq \hat{\pi}_U(\tau,k) + \frac{r}{\sqrt{n}}\omega_U(\tau,k) \; \forall(\tau,k) \in \mathcal{B}\} ,$$

where $\omega_L(\tau,k)$ and $\omega_U(\tau,k)$ are positive weight functions which we assume are known for the present discussion but allow to be estimated in the next section. These weight functions enable us to adjust for settings in which the estimators $\hat{\pi}_L(\tau,k)$ and $\hat{\pi}_U(\tau,k)$ have very different asymptotic variances at different points (τ,k).[8]

The constant r needs to be selected so that $\hat{\mathcal{G}}(\omega,r)$ satisfies the coverage requirement advocated in Imbens and Manski (2004), namely that:

$$\liminf_{n\to\infty} P(g \in \hat{\mathcal{G}}(\omega,r)) \geq 1-\alpha \tag{33}$$

for all functions $g \in \mathcal{G}$ and desired level of confidence $1-\alpha$. In the following lemma, we exploit Theorem 4.1 to characterize the appropriate choices of r that ensure (33) is satisfied:

Lemma 4.2. *Let Assumptions 2.1, 2.2, 4.1 hold and suppose $\omega_L(\tau,k)$ and $\omega_U(\tau,k)$ are both positive, continuous and bounded away from zero on $(\tau,k) \in \mathcal{B}$. For $G(\tau,k)$ as in* (31) *define:*

$$Z \equiv \sup_{(\tau,k)\in\mathcal{B}} \max\left\{\frac{G^{(1)}(\tau,k)}{\omega_L(\tau,k)}, -\frac{G^{(2)}(\tau,k)}{\omega_U(\tau,k)}\right\} , \tag{34}$$

where $G^{(i)}(\tau,k)$ denotes the i^{th} coordinate of $G(\tau,k)$. If $\{Y_i,X_i,D_i\}_{i=1}^n$ is i.i.d. and $r_{1-\alpha}$ is the $1-\alpha$ quantile of Z, then it follows that (33) *is satisfied if and only if $r \geq r_{1-\alpha}$.*

By Lemma 4.2, the smallest choice of r for which the coverage requirement in (33) is satisfied is the $1-\alpha$ quantile of Z, as defined in (34). Interestingly, this choice of r also ensures that with asymptotic probability $1-\alpha$ the *entire* set $\mathcal{G}$ is included in the confidence collection $\hat{\mathcal{G}}(\omega,r)$. Therefore, for the type of confidence collections we consider, the coverage requirement in Imbens and Manski (2004) is equivalent to the one proposed in Chernozhukov, Hong, and Tamer (2007). While in parametric models the former coverage requirement is weaker than the latter, in the present nonparametric context the set $\mathcal{G}$ is so rich that the only way in which (33) can be satisfied is if in fact $\hat{\mathcal{G}}(\omega,r)$ covers $\mathcal{G}$ in its entirety. We formalize these conclusions in the following corollary.

Corollary 4.2. *Let Assumptions 2.1, 2.2, 4.1 hold, $\omega_L(\tau,k)$ and $\omega_U(\tau,k)$ be both positive, continuous and bounded away from zero on $(\tau,k) \in \mathcal{B}$. Then $\hat{\mathcal{G}}(\omega,r)$ satisfies* (33) *if and only if:*

$$\liminf_{n\to\infty} P(\mathcal{G} \subseteq \hat{\mathcal{G}}(\omega,r)) \geq 1-\alpha \ . \tag{35}$$

[8] Appropriate choices of $\omega_L(\tau,k)$ and $\omega_U(\tau,k)$ are context specific. Setting them equal to the asymptotic standard deviation of $\sqrt{n}(\hat{\pi}_L(\tau,k) - \pi_L(\tau,k))$ and $\sqrt{n}(\hat{\pi}_U(\tau,k) - \pi_U(\tau,k))$ respectively stabilizes the variance across points, but at the cost of wider intervals around points (τ,k) with high variance (relative to equal weighting).

4.3 Bootstrap Critical Values

The construction of the proposed confidence interval $\hat{\mathcal{G}}(\omega, r)$ is of course infeasible in the absence of suitable estimators for the critical value $r_{1-\alpha}$ and weight functions $\omega_L(\tau, k)$ and $\omega_U(\tau, k)$. In the present section we conclude our discussion on inference by establishing the validity of a bootstrap procedure for the consistent estimation of $r_{1-\alpha}$. The choice of weight functions $\omega_L(\tau, k)$ and $\omega_U(\tau, k)$ is context specific and in some instances they may even be known hence voiding the need for estimation. For this reason, we do not focus on deriving estimators for $\omega_L(\tau, k)$ and $\omega(\tau, k)$ and instead we establish the consistency of a bootstrap procedure under the following high level assumption:

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 $\mathcal{B}$.*

The resampling method we examine is the weighted bootstrap, which has been previously employed in semiparametric M-estimation by Ma and Kosorok (2005) and in conditional moment models by Chen and Pouzo (2008). This procedure is similar to the regular bootstrap except for the important difference that the random weights on different observations are independent from each other. Specifically, let $\{W_i\}_{i=1}^n$ be an *i.i.d.* sample from a random variable W_i satisfying:

Assumption 4.3. *W_i is positive a.s., independent of (Y_i, X_i, D_i) and $E[W_i] = Var(W_i) = 1$.*

The weighted bootstrap estimator for $r_{1-\alpha}$ can then be obtained through the following algorithm:

Step 1: For a generated random sample of weights $\{W_i\}_{i=1}^n$, define the criterion function:

$$\tilde{Q}_{x,n}(c|\tau, b) \equiv \Big(\frac{1}{n}\sum_{i=1}^{n} W_i\{1\{Y_i \leq c, X_i = x, D_i = 1\} + b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\}\}\Big)^2 , \quad (36)$$

which may in turn be employed to obtain the following bootstrap estimators for $c^k_{\tau,L}(x)$ and $c^k_{\tau,L}(x)$:

$$\tilde{c}^k_{\tau,L}(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, \min\{\tau + k, 1\}) \qquad \tilde{c}^k_{\tau,U}(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, \max\{\tau - k, 0\}) . \quad (37)$$

Note that if we were to employ the conventional bootstrap, then we would run the risk of redrawing a sample for which no observations existed with $X_i = x$. This is not a concern under the weighted-bootstrap for which $\tilde{c}^k_{\tau,L}(x)$ and $\tilde{c}^k_{\tau,L}(x)$ are necessarily well defined. ∎

Step 2: Using the bootstrap bounds $\tilde{c}^k_{\tau,L}(x)$ and $\tilde{c}^k_{\tau,U}(x)$ from Step 1, obtain the estimators:

$$\tilde{\pi}_L(\tau, k) \equiv \inf_{\theta} \lambda'(E_S[X_i X_i'])^{-1} E_S[X_i \theta(X_i)] \quad \text{s.t. } \tilde{c}^k_{\tau,L}(x) \leq \theta(x) \leq \tilde{c}^k_{\tau,U}(x) \quad (38)$$

$$\tilde{\pi}_U(\tau, k) \equiv \sup_{\theta} \lambda'(E_S[X_i X_i'])^{-1} E_S[X_i \theta(X_i)] \quad \text{s.t. } \tilde{c}^k_{\tau,L}(x) \leq \theta(x) \leq \tilde{c}^k_{\tau,U}(x) . \quad (39)$$

The weighted bootstrap statistic for the random variable Z, as defined in (34), is then given by:

$$\tilde{Z} \equiv \sup_{(\tau,k)\in\mathcal{B}} \max\left\{\frac{\sqrt{n}(\tilde{\pi}_L(\tau,k)-\hat{\pi}_L(\tau,k))}{\hat{\omega}_L(\tau,k)}, \frac{\sqrt{n}(\hat{\pi}_U(\tau,k)-\tilde{\pi}_U(\tau,k))}{\hat{\omega}_U(\tau,k)}\right\} . \tag{40}$$

Our estimator for $r_{1-\alpha}$ is the $1-\alpha$ quantile of $\tilde{Z}$ conditional on the sample $\{Y_i, X_i, D_i\}_{i=1}^n$:[9]

$$\tilde{r}_{1-\alpha} \equiv \inf\left\{r : P\Big(\tilde{Z} \geq r \,\Big|\{Y_i, X_i, D_i\}_{i=1}^n\Big) \geq 1-\alpha\right\} . \tag{41}$$

Notice that the random variable $\tilde{Z}$ is a function of $\{Y_i, X_i, D_i\}_{i=1}^n$ as well as $\{W_i\}_{i=1}^n$. Hence, the randomness in $\tilde{r}_{1-\alpha}$ is a result of the weights $\{W_i\}_{i=1}^n$ which are not part of the conditioning. ■

In applications, $\tilde{r}_{1-\alpha}$ can be computed through simulations in the same manner as for the regular bootstrap. For this purpose, a random number generator can be employed to obtain samples $\{W_i\}_{i=1}^n$ satisfying Assumption 4.3; for example by generating exponentially distributed random variables with mean one. These samples of random weights $\{W_i\}_{i=1}^n$ can in turn be used to compute a sample of realizations $\tilde{Z}$ conditional on $\{Y_i, X_i, D_i\}_{i=1}^n$ by following Steps 1 and 2. Provided the number of simulations is sufficiently large, the estimator $\tilde{r}_{1-\alpha}$ is then well approximated by the empirical $1-\alpha$ quantile of $\tilde{Z}$ across the computed simulations.

As Theorem 4.2 shows, the estimated critical value $\tilde{r}_{1-\alpha}$ is indeed consistent for $r_{1-\alpha}$.

Theorem 4.2. *If Assumptions 2.1, 2.2, 4.1, 4.2, and 4.3 hold and $\{Y_i, X_i, D_i, W_i\}_{i=1}^n$ is i.i.d., then:*

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

Given the consistency of $\tilde{r}_{1-\alpha}$, established in Theorem 4.2, and of $(\hat{\omega}_L(\tau,k), \hat{\omega}_U(\tau,k))$, imposed in Assumption 4.2, it is straightforward to show that the direct "plug-in" analogue to $\mathcal{G}(\omega, r_{1-\alpha})$ provides the desired coverage. Corollary 4.3 establishes this result.

Corollary 4.3. *If Assumptions 2.1, 2.2, 4.1, 4.2, and 4.3 hold and $\{Y_i, X_i, D_i, W_i\}_{i=1}^n$ is i.i.d:*

$$\liminf_{n\to\infty} P(g \in \hat{\mathcal{G}}(\hat{\omega}, \tilde{r}_{1-\alpha})) \geq 1-\alpha ,$$

for all $g \in \mathcal{G}$, where $\hat{\omega}$ denotes the function $\hat{\omega}(\tau,k) \equiv (\hat{\omega}_L(\tau,k), \hat{\omega}_U(\tau,k))$.

5 Application

To illustrate our techniques we revisit the results of Angrist, Chernozhukov, and Fernández-Val (2006) regarding changes across Decennial Censuses in the quantile specific returns to schooling.

[9]The assumptions imposed on $\tilde{\pi}_L(\tau,k)$ and $\tilde{\pi}_U(\tau,k)$ are not sufficient for implying that $\tilde{Z}$ is a measurable function of $\{W_i\}_{i=1}^n$ for fixed $\{Y_i, X_i, D_i\}_{i=1}^n$. The conditional probability in (41) should be interpreted as an outer probability.

Before doing so, it is useful to empirically assess the likely magnitude of any deviations from ignorable non-response present in similar earnings data in order to inform our choice of k in Assumption 2.2. To that end we begin by examining missingness patterns in the 1973 Exact Match CPS-SSA Earnings File which provides IRS based measures of wage and salary earnings for March CPS respondents with valid Social Security numbers. These data allow us to examine whether IRS earnings predict missingness of CPS earnings within covariate bins.[10]

5.1 Analysis of 1973 CPS-SSA File

In this section we demonstrate how to measure deviations from ignorability using our nonparametric KS metric when validation data are available. We work with an extract from the 1973 CPS 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. To deal with outliers in the IRS data we drop observations with earnings less than 200 or greater than 7500 dollars per week worked. Roughly eight percent of the remaining sample have missing (allocated) CPS based earnings. Like Angrist, Chernozhukov, and Fernández-Val (2006) we take the relevant covariates to be age, years of schooling, and race. Hence our first question is whether conditional on these covariates the data are missing at random.

Table 1: Conditional Fixed Effects Logit Estimates of Missingness in 1973 CPS-SSA sample

log(IRS annual earnings/weeks worked)	-3.34
	(0.86)
(log(IRS annual earnings/weeks worked))^2	0.23
	(0.06)
Log-Likelihood	-3,247.66
$\mathcal{X}^2(2)$	15.22
Number of observations	12,556

Note: Robust Standard Errors in Parentheses

Table 1 presents estimates from a conditional fixed effects logit where all interactions of the demographic covariates have been absorbed as fixed effects and the response probability is modeled as a quadratic in log IRS earned income per week worked (robust standard errors in parenthesis). Evidently, missingness follows a U-shaped response pattern with very low and very high wage men

[10]We focus on IRS earnings rather than Social Security earnings because the latter are severely topcoded.

least likely to provide valid earnings information. This pattern is consistent with the one conjectured in Lillard, Smith, and Welch (1986).

These parametric estimates of the missingness process can be used to construct estimates of the KS distance bound k in Assumption 2.2 given some additional results which we briefly develop here. In the present context we take $F_{y|x}(c)$ and $F_{y|0,x}(c)$ to represent the CDF of IRS based earnings in demographic cell x and the conditional CDF among men with imputed CPS earnings in demographic cell x respectively. The bound k is then given by the maximum across demographic cells x of $KS(F_{y|x}, F_{y|0,x})$.

In order to exploit the parametric structure of our logit model in estimating k, we introduce an alternative representation of $KS(F_{y|x}, F_{y|0,x})$ as a reweighted inverse quantile:[11]

$$KS(F_{y|x}, F_{y|0,x}) = \sup_{c \in \mathbf{R}} \left| \int_{-\infty}^{c} w(y,x) dF(y|X_i = x) \right| , \tag{42}$$

where $w(y,x) = 1 - P(D_i = 0|Y_i = y, X_i = x)/(1 - p(x))$. A byproduct of our conditional logit procedure then is an estimate $\hat{w}(y,x)$ of the weight function $w(y,x)$. The representation in (42) suggests these weights can be used to form an estimate of $KS(F_{y|x}, F_{y|0,x})$ as follows:

$$\widehat{KS}(F_{y|x}, F_{y|0,x}) \equiv \max_{c \in C_n} \left| \frac{1}{n_x} \sum_{i=1}^{n} 1\{Y_i \leq c, X_i = x\} \hat{w}(Y_i, x) \right| , \tag{43}$$

where $n_x \equiv \sum_i 1\{X_i = x\}$ and C_n is a finite set of points in $\mathbf{R}$ that expands with n.[12] In practice we choose C_n to consist of all deciles between 0.1 and 0.9 along with the 0.05 and 0.95 quantiles in order to obtain an estimate of the KS distance in each covariate bin.

To parsimoniously summarize the heterogeneity across bins, we model the cell specific KS distances (and their corresponding estimates) as:

$$KS(F_{y|x}, F_{y|0,x}) = \exp(W_x'\gamma_0) \qquad \widehat{KS}(F_{y|x}, F_{y|0,x}) = \exp(W_x'\gamma_0) + \eta_x \tag{44}$$

where W_x is a vector of six cell level attributes and an intercept and η_x is a random estimation error that vanishes as the number of observations per covariate bin grows large.[13] We estimate the above model via weighted nonlinear least squares with weights proportional to cell size. The fit of the model is quite good with a weighted R^2 of .995.

Letting $\mathcal{W}$ be the set of possible realizations of cell specific covariates W_x, the specification in (44) implies that the KS bound k is equal to the maximum over $W_x \in \mathcal{W}$ of $\exp(W_x'\gamma_0)$. Hence, a

[11] See Lemma 6.7 in the Appendix for a derivation of this equality.

[12] See Lemma 6.8 in the Appendix for a proof of the consistency of $\widehat{KS}(F_{y|x}, F_{y|0,x})$.

[13] The attributes are: years of schooling, years of schooling squared, age, age squared, potential experience, and potential experience squared.

natural estimator of k is given by:

$$\hat{k} \equiv \sup_{W_x \in \mathcal{W}} \exp(W_x' \hat{\gamma}) \ , \tag{45}$$

where $\hat{\gamma}$ is some estimator of γ_0 that is consistent as the number of demographic cells $|\mathcal{X}|$ (the cardinality of $\mathcal{X}$) grows large. In order to conduct inference on k, we derive the asymptotic distribution of $\hat{k}$ under the null of correct model specification and the high level requirement that the estimator $\hat{\gamma}$ be asymptotically normally distributed.

Lemma 5.1. *Suppose* (44) *and* (45) *hold, (i)* $\mathcal{W}$ *is finite and (ii)* $\sqrt{|\mathcal{X}|}(\hat{\gamma} - \gamma_0) \xrightarrow{L} N(0, \Sigma)$ *for some positive definite* Σ*. If* $\mathcal{W}_0 = \{W_x \in \mathcal{W} : k = \exp(W_x'\gamma_0)\}$*, then it follows that:*

$$\sqrt{|\mathcal{X}|}(\hat{k} - k) \xrightarrow{L} k \times \max_{W_x \in \mathcal{W}_0} W_x' Z \ ,$$

where Z *is a normal random variable with mean zero and covariance matrix* Σ*.*

Since $\mathcal{W}_0$ is a subset of $\mathcal{W}$, the asymptotic distribution of $\sqrt{|\mathcal{X}|}(\hat{k}-k)$ is first order stochastically dominated by the distribution of the random variable

$$V \equiv k \times \max_{W_x \in \mathcal{W}} W_x' Z \ , \tag{46}$$

where $Z \sim N(0, \Sigma)$ for Σ the covariance matrix in Lemma 5.1. Hence, for $c_{1-\alpha}$ the $1 - \alpha$ quantile of V, an asymptotically valid one sided confidence interval for k is given by $[\hat{k} - c_{1-\alpha}/\sqrt{|\mathcal{X}|}, \infty)$.

We estimate $c_{1-\alpha}$ via a two-step procedure. First, we use our weighted nonlinear least squares estimate $\hat{\gamma}$ to form an estimate of $\hat{k}$ via (45) which we use in place of k in (46). We then employ a parametric bootstrap procedure, simulating from a normal distribution with mean zero and covariance matrix $\hat{\Sigma}$ to generate draws of $\max_{\mathcal{W}} W_x' Z$, where $\hat{\Sigma}$ is the standard estimate for the asymptotic variance of the NLLS estimate. We take the empirical $1 - \alpha$ quantile of this simulated distribution scaled by $\hat{k}$ as our estimate of $c_{1-\alpha}$.

Implementing these steps, our NLLS based point estimate of $\hat{k}$ is 0.065 and our bootstrap procedure yields a confidence interval with a lower bound value for k of 0.056. Though the CPS earnings data are clearly not missing at random, the deviations from ignorability are relatively minor, suggesting that a worst case bounds analysis would be extremely conservative. Though this conclusion only applies to item nonresponse in the 1973 CPS among men with valid social security numbers, we suspect deviations from MAR of similar magnitude are likely to be present in other settings as well. Additional validation data are necessary to more fully resolve this issue.

5.2 Analysis of Census Data

We turn now to a re-analysis of the 1980, 1990, and 2000 Census samples considered in Angrist, Chernozhukov, and Fernández-Val (2006).[14] To simplify our estimation routine, and to correct small mistakes found in the IPUMS files since the time of their study, we use new extracts of the 1% unweighted IPUMS files for each decade rather than their original mix of weighted and unweighted samples. Sample sizes and imputation rates for the weekly earnings variable are given in Table 2.

Table 2: Fraction of Observations in Census Estimation Sample with Missing Weekly Earnings

Census Year	Number of Observations	Fraction Missing
1980	80,800	19.53%
1990	111,356	23.10%
2000	131,711	27.70%
Overall	323,867	23.67%

Like Angrist, Chernozhukov, and Fernández-Val (2006), we estimate linear conditional quantile models for log earnings per week of the form:

$$c_\tau(x) = X_i'\gamma_\tau + E_i\beta_\tau \ . \tag{47}$$

where X_i consists of an intercept, a black dummy, and a quadratic in potential experience, and E_i represents years of schooling. Our analysis focuses on the quantile specific "returns" to a year of schooling β_τ.[15]

Figure 3 provides estimates of the pseudo-true returns functions 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 4 subject to the restriction that k=0.[16] As in Section 4, our loss metric is chosen to be squared error and we weight bin-specific deviations by sample size (i.e. we choose S equal to empirical measure).

Our results are comparable to those found in Figure 2A of Angrist, Chernozhukov, and Fernández-

[14]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 25 observations falling in demographic cells with greater than 66% missing.

[15]Particularly in the context of quantile regressions, the Mincerian earnings coefficients need not map into any proper economic concept of returns (Heckman, Lochner, and Todd, 2005).

[16]In constructing uniform confidence intervals we use a quantile specific weighting function of the form $\omega(\tau) = \phi(\Phi^{-1}(\tau))^{1/2}$, where $\phi(.)$ and $\Phi(.)$ are the standard normal density and CDF respectively. This can be shown to be inversely proportional to the square root of the variance of the quantiles of a standard normal.

Val (2006).[17] 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 wide array of other evidence on changes in the wage structure (Juhn, Murphy, and Pierce (1993); Autor and Katz (1999); Autor, Katz, and Kearney (2008); Lemieux (2006a)). However, to our knowledge, the finding of a shape change in the quantile process between 1990 and 2000 is new, representing a form of heteroscedasticity in the conditional earnings distribution with respect to schooling that appears not to have been present in previous decades.[18]

A natural concern, however, is the extent to which some or all of these estimated changes in the wage structure are driven by limitations in the quality of Census earnings data. As Table 2 shows the prevalence of earnings imputations increases steadily over the sample period with roughly a quarter of the observations allocated by 2000. Without restrictions on the missingness process quantiles below the 25th percentile and above the 75th are not even bounded. Our question then is how robust the patterns found in Figure 3 are to plausible deviations from the missing at random assumption.

We begin by using our finite sample likelihood ratio procedure described in Section 3 to construct a confidence interval for the β_τ at each τ. Figure 4 plots intervals providing pointwise coverage of the quantile process with the relatively conservative value of $k = 0.10$. For most values of τ the interval is empty meaning the model has been rejected. Only at extreme quantiles do we get nondegenerate intervals. This rejection is remarkable given the degree of missing data present in these samples and our relatively lax restriction on the selection process.[19]

Though the Mincer specification is easily rejected as a literal description of the conditional quantile function, it may still serve as a useful means of summarizing patterns in the data. Hence we now turn to constructing intervals covering the pseudo-true approximating values of β_τ. Figures 5 shows the intervals that result with a value of $k = 0.05$ which we consider reasonable given our earlier analysis of the CPS.

Though the returns function clearly 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

[17]They are not identical because of revisions to the IPUMS files used by Angrist, Chernozhukov, and Fernández-Val (2006) that occured since the time of their analysis and our different choice of loss function for determining the pseudo true approximation.

[18]Lemieux (2006b) finds a similar pattern though his focus is on the nonlinearity of the conditional quantile function with respect to education rather than the implied heteroscedasticity.

[19]Our findings are in line with the conclusion of Heckman, Lochner, and Todd (2006) that the separability restrictions implicit in the Mincer model are easily rejected for the case of the conditional expectation function.

evidence of heterogeneity across quantiles in the returns in any of the three Census samples – a straight line can be fit through each sample's confidence region. To assess the sensitivity of our conclusion that the process changed between 1980 and 1990 to further deviations from MAR we searched for the smallest value of k such that the intervals for these two decades overlap at all quantiles between 0.1 and 0.9. This occurs at $k = 0.135$ which is a fairly large deviation from MAR (see Appendix A). The resulting confidence regions for this critical value of the KS restriction are plotted in Figure 6.

Finally, we show the results of estimating the more flexible earnings model of Lemieux (2006b) which allows for quadratic effects of education on earnings quantiles.[20] Figure 7 provides bounds on the 10th, 50th, and 90th conditional quantiles of weekly earnings by schooling level in 1980, 1990, and 2000 using our baseline KS restriction of $k = 0.05$. 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, as noted by Lemieux (2006b) the returns to schooling appear to have convexified at the upper tail of the weekly earnings distribution with very well educated workers experiencing substantial gains relative to the less educated.

6 Conclusion

We have proposed a nonparametric restriction on the degree of non-ignorable selection governing non-response and developed inferential procedures for parametric quantile models subject to this restriction and the possibility of misspecification. Our examination of missingness patterns in merged CPS-SSA data led us to reject the MAR assumption but to conclude that the degree of non-random selection in earnings data is likely small. Revisiting the empirical analysis in Angrist, Chernozhukov, and Fernández-Val (2006) we found 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.

[20]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.

Appendix A - The bivariate normal selection model and KS distance

To develop intuition for our nonparametric KS metric of deviations from missing at random, we provide here a mapping between some values of a standard bivariate selection model and the implied KS distance. Using the notation of Section 2, our DGP of interest is:

$$(Y_i, v_i) \sim N(0, \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)) \qquad D_i = 1\{\mu + v_i > 0\} \ . \tag{48}$$

In this model, the parameter ρ indexes the degree of non-ignorable selection in the outcome variable Y_i. We choose μ 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 KS distance between the distribution of missing outcomes and the unselected distribution for various values of ρ by simulation. The results are given in the following table:

Table 3: $KS(F_y, F_{y|0})$ as a function of ρ

ρ	KS
0.05	0.0261
0.10	0.0515
0.15	0.0771
0.20	0.1028
0.25	0.1287
0.30	0.1548

ρ	KS
0.35	0.1815
0.40	0.2085
0.45	0.2363
0.50	0.2648
0.55	0.2939
0.60	0.3239

ρ	KS
0.65	0.3554
0.70	0.3887
0.75	0.4239
0.80	0.4621
0.85	0.5043
0.90	0.5529

Appendix B - Computing $\hat{c}^k_{\tau,L}(x)$ and $\hat{c}^k_{\tau,U}(x)$

In this Appendix we discuss the details of the inference procedure outlined in Section 3.1. For notational convenience, define the variables:

$$V_{x,i} \equiv 1\{D_i = 0, X_i = x\} \qquad W_{x,i}(c) \equiv 1\{Y_i \le c,\ D_i = 1, X_i = x\} \ . \tag{49}$$

Letting $\gamma_x \equiv 1 - p(x)$, $\xi_x(c) \equiv F_{y|1,x}(c)p(x)$ it then follows that the log-likelihood of the subsample of $\{V_{x,i}, W_{x,i}\}_{i=1}^n$ for which $X_i = x$, conditional on $\{X_i\}_{i=1}^n$ and $(\gamma_x, \xi_x(c)) = (\gamma, \xi)$, is given by:

$$L_{n,x}(\gamma,\xi) \equiv \log\gamma \times \sum_{i=1}^n V_{x,i} + \log\xi \times \sum_{i=1}^n W_{x,i}(c) + \log(1-\gamma-\xi) \times \sum_{i=1}^n (1\{X_i = x\} - V_{x,i} - W_{x,i}(c)) \ . \tag{50}$$

Therefore, the relevant likelihood ratio test statistic for the hypothesis in (10) is given by:

$$T_{n,x}(c) \equiv \sup_{(\gamma,\xi)} L_{n,x}(\gamma,\xi) - \sup_{(\gamma,\xi)\in\Gamma^k_{\tau,x}} L_{n,x}(\gamma,\xi) \ , \tag{51}$$

for $\Gamma^k_{\tau,x}$ as defined in (9). Also let $n_x \equiv \sum_i 1\{X_i = x\}$, $\bar{V}_x = \frac{1}{n_x}\sum_i V_{x,i}$, $\bar{W}_x(c) \equiv \frac{1}{n_x}\sum_i W_{x,i}(c)$ and

$$B_1 \equiv -(\bar{V}_x + \bar{W}_x(c))(\tau - k)(1-\tau) - \tau(1-\tau+k)(1-\bar{W}_x(c))$$
$$B_2 \equiv -(\bar{V}_x + \bar{W}_x(c))(\tau + k)(1-\tau) - \tau(1-\tau-k)(1-\bar{W}_x(c)) \ .$$

Furthermore, letting $(\hat{\gamma}^c_x, \hat{\xi}^c_x)$ be the maximizers of $L_{n,x}(\gamma,\xi)$ on $\Gamma^k_{\tau,x}$, it is possible to show that:

$$\begin{array}{ll} \hat{\gamma}^c_x = \frac{-B_1 - \sqrt{B_1^2 - 4\tau(\tau-k)(1-\tau+k)(1-\tau)\bar{V}_x}}{2(\tau-k)(1-\tau+k)} & \text{if } \tau - k > 0 \text{ and } \bar{W}_x(c) > \tau - (\tau-k)\bar{V}_x \\ \hat{\gamma}^c_x = \frac{(1-\tau)\bar{V}_x}{1-\bar{W}_x(c)} & \text{if } \tau - k \leq 0 \text{ and } \bar{W}_x(c) > \tau \\ \hat{\gamma}^c_x = \bar{V}_x & \text{if } \tau - \min\{\tau+k,1\}\bar{V}_x \leq \bar{W}_x(c) \leq \tau - \max\{\tau-k,0\}\bar{V}_x \\ \hat{\gamma}^c_x = \frac{\tau\bar{V}_x}{\bar{V}_x + \bar{W}_x(c)} & \text{if } \tau + k \geq 1 \text{ and } \bar{W}_x(c) < \tau - \bar{V}_x \\ \hat{\gamma}^c_x = \frac{-B_2 - \sqrt{B_2^2 - 4\tau(\tau+k)(1-\tau-k)(1-\tau)\bar{V}_x}}{2(\tau+k)(1-\tau-k)} & \text{if } \tau + k < 1 \text{ and } \bar{W}_x(c) < \tau - (\tau+k)\bar{V}_x \ , \end{array} \quad (52)$$

while the closed form solution for $\hat{\xi}^c_x$ is in turn given by:

$$\begin{array}{ll} \hat{\xi}^c_x = \tau - \max\{\tau-k,0\}\hat{\gamma}^c_x & \text{if } \bar{W}_x(c) > \tau - \max\{\tau-k,0\}\bar{V}_x \\ \hat{\xi}^c_x = \bar{W}_x(c) & \text{if } \tau - \min\{\tau+k,1\}\bar{V}_x \leq \bar{W}_x(c) \leq \tau - \max\{\tau-k,0\}\bar{V}_x \\ \hat{\xi}^c_x = \tau - \min\{\tau+k,1\}\hat{\gamma}^c_x & \text{if } \bar{W}_x(c) < \tau - \min\{\tau+k,1\}\bar{V}_x \ . \end{array} \quad (53)$$

Since the unconstrained maximizers of $L_{n,x}(\gamma,\xi)$, denoted $(\hat{\gamma}^u_x, \hat{\xi}^u_x)$, are given by $(\hat{\gamma}^u_x, \hat{\xi}^u_x) = (\bar{V}_x, \bar{W}_x(c))$, it is straightforward to compute a closed form solution for $T_{n,x}(c)$ from (51), (52) and (53).

In order to compute $r(\alpha_x)$ as in (11), let $q_{\gamma,\xi}(1-\alpha_x)$ denote the $1-\alpha_x$ quantile of $T_{n,x}(c)$ when $(\gamma_x, \xi_x(c)) = (\gamma, \xi)$, and observe that through direct manipulations:

$$\begin{aligned} r(\alpha_x) &= \{\inf r : \sup_{(\gamma,\xi)\in\Gamma^k_{\tau,x}} \{1 - P_{\gamma,\xi}(T_{n,x}(c) \leq r | \{X_i\}_{i=1}^n)\} \leq \alpha_x\} \\ &= \{\inf r : \inf_{(\gamma,\xi)\in\Gamma^k_{\tau,x}} P_{\gamma,\xi}(T_{n,x}(c) \leq r | \{X_i\}_{i=1}^n) \geq 1 - \alpha_x\} \\ &= \sup_{(\gamma,\xi)\in\Gamma^k_{\tau,x}} q_{\gamma,\xi}(1-\alpha_x) \ . \end{aligned} \quad (54)$$

However, for every fixed $(\gamma,\xi) \in \Gamma^k_{\tau,x}$, $q_{\gamma,\xi}(1-\alpha_x)$ may be computed either analytically or by simulation by exploiting the closed form solution for $T_{n,x}(c)$ and noting that the random variable $(\sum_i V_{x,i}, \sum_i W_{x,i}(c))$ is multinomially distributed with parameters $(\gamma, \xi, \sum_i 1\{X_i = x\})$. In our application in Section 5, we computed $q_{\gamma,\xi}(1-\alpha_x)$ through simulation and obtained $r(\alpha_x)$ by numerically maximizing $q_{\gamma,\xi}(1-\alpha_x)$ on $\Gamma^k_{x,\tau}$. Having found $r(\alpha_x)$, it is then possible to compute $\hat{c}^k_{\tau,L}(x)$ and $\hat{c}^k_{\tau,U}(x)$ from (12).

Appendix C - Proof of Results

Proof of Lemma 2.1: If $\theta(x) = c_\tau(x)$, then it immediately follows from Assumption 2.2 that:

$$\begin{aligned} \tau &= F_{y|1,x}(\theta(x)) \times p(x) + F_{y|0,x}(\theta(x)) \times \{1 - p(x)\} \\ &\leq F_{y|1,x}(\theta(x)) \times p(x) + \min\{F_{y|x}(\theta(x)) + k, 1\} \times \{1 - p(x)\} \\ &= F_{y|1,x}(\theta(x)) \times p(x) + \min\{\tau + k, 1\} \times \{1 - p(x)\} \ . \end{aligned} \quad (55)$$

By identical manipulations, it follows that $F_{y|1,x}(\theta(x)) \times p(x) \leq \tau - \max\{\tau - k, 0\} \times \{1 - p(x)\}$ and hence we conclude $\theta \in \mathcal{C}_\tau^k$. To prove the bounds are sharp, first observe that:

$$\begin{aligned}\sup_{x\in\mathcal{X}} KS(F_{y|x}, F_{y|0,x}) &= \sup_{x\in\mathcal{X}}\sup_{c\in\mathbf{R}} |F_{y|0,x}(c) \times \{1 - p(x)\} + F_{y|1,x}(c) \times p(x) - F_{y|0,x}(c)| \\ &= \sup_{x\in\mathcal{X}}\sup_{c\in\mathbf{R}} |F_{y|0,x}(c) - F_{y|1,x}(c)| \times p(x) \ . \end{aligned} \tag{56}$$

Therefore, we obtain from (56) that Assumption 2.2 is satisfied if and only if:

$$\sup_{x\in\mathcal{X}}\{KS(F_{y|1,x}, F_{y|0,x}) \times p(x)\} \leq k \ . \tag{57}$$

Next, let $\theta \in \mathcal{C}_\tau^k$ and define the function $\kappa : \mathcal{X} \to \mathbf{R}$ by:

$$\kappa(x) \equiv \frac{\tau - F_{y|1,x}(\theta(x)) \times p(x)}{1 - p(x)} \ . \tag{58}$$

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

$$\max\{\tau - k, 0\} \leq \kappa(x) \leq \min\{\tau + k, 1\} \qquad\qquad |\kappa(x) - F_{y|1,x}(\theta(x))| \leq \frac{k}{p(x)} \ . \tag{59}$$

We now aim to construct a distribution for Y_i conditional on X_i and Y_i being missing such that all assumptions are met and in addition $\theta(x)$ is the conditional quantile of Y_i given X_i. Define:

$$\begin{aligned}\tilde{F}^+_{y|0,x}(c) \equiv& 1\{c \geq \theta(x)\} \times \max\{F_{y|1,x}(c),\ \min\{\frac{1}{2}(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x), 1\}\} \\ &+ 1\{c < \theta(x)\} \times \max\{F_{y|1,x}(c),\ 2(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x)\} \\ \tilde{F}^-_{y|0,x}(c) \equiv& 1\{c \geq \theta(x)\} \times \min\{F_{y|1,x}(c),\ 2(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x)\} \\ &+ 1\{c < \theta(x)\} \times \min\{F_{y|1,x}(c),\ \max\{\frac{1}{2}(F_{y|1,x}(c) - F_{y|1,x}(\theta(x))) + \kappa(x), 0\}\}\end{aligned} \tag{60}$$

and let the distribution of Y_i conditional on X_i and Y_i being unobservable be given by:

$$\tilde{F}_{y|0,x}(c) \equiv 1\{\kappa(x) \geq F_{y|1,x}(\theta(x))\} \times \tilde{F}^+_{y|0,x}(c) + 1\{\kappa(x) < F_{y|1,x}(\theta(x))\} \times \tilde{F}^-_{y|1,x}(c) \ . \tag{61}$$

Note that $\tilde{F}_{y|0,x}(c)$ is strictly increasing and continuous for 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:

$$\tilde{F}_{y|x}(\theta(x)) = F_{y|1,x}(\theta(x)) \times p(x) + \tilde{F}_{y|0,x}(\theta(x)) \times \{1 - p(x)\} = F_{y|1,x}(\theta(x)) \times p(x) + \kappa(x) \times \{1 - p(x)\} = \tau \ , \tag{62}$$

so that $\theta(x)$ is the conditional τ^{th} quantile of Y_i given X_i. In addition, by construction and (59) we have:

$$\sup_{c\in\mathbf{R}} |\tilde{F}_{y|0,x}(c) - F_{y|1,x}(c)| = |\tilde{F}_{y|0,x}(\theta(x)) - F_{y|1,x}(\theta(x))| \leq \frac{k}{p(x)} \ , \tag{63}$$

uniformly in $x \in \mathcal{X}$. Therefore, from (57) and (63) it follows that $KS(\tilde{F}_{y|x}, \tilde{F}_{y|0,x}) \leq k$ for all $x \in \mathcal{X}$, and hence $\tilde{F}_{y|x}$ and $\tilde{F}_{y|0,x}$ satisfy Assumptions 2.1 and 2.2 and are such that $\theta(x) = c_\tau(x)$. We hence conclude the bounds are sharp and the Lemma follows. ■

Proof of Corollary 2.1: Follows immediately form Lemma 2.1 and the definition of $F^{-}_{y|1,x}(q)$. ■

Proof of Lemma 3.1: As in Appendix B, let $(\gamma_x, \xi_x(c)) \equiv (1 - p(x), F_{y|1,x}(c)p(x))$ and note that since $c \notin [\hat{c}^k_{\tau,L}(x), \hat{c}^k_{\tau,U}(x)]$ implies $T_{n,x}(c) > r(\alpha_x)$, it follows that for any c such that $c^k_{\tau,L}(x) \le c \le c^k_{\tau,U}(x)$:

$$\begin{aligned} P_{\gamma_x,\xi_x(c)}\left(c \notin [\hat{c}_{\tau,L}(x), \hat{c}_{\tau,U}(x)] | \{X_i\}_{i=1}^n\right) &\le P_{\gamma_x,\xi_x(c)}\left(T_{n,x}(c) > r(\alpha_x) | \{X_i\}_{i=1}^n\right) \\ &\le \sup_{(\gamma,\xi)\in\Gamma^k_x} P_{\gamma,\xi}\left(T_{n,x}(c) > r(\alpha_x) | \{X_i\}_{i=1}^n\right) \\ &\le \alpha_x \end{aligned} \tag{64}$$

where the second inequality follows by $c \in [c^k_{\tau,L}(x), c^k_{\tau,U}(x)]$ if and only if $(\gamma_x, \xi_x(c)) \in \Gamma^k_x$ and the final inequality follows by definition of $r(\alpha_x)$. ■

Proof of Lemma 3.2: As in Appendix B, let $(\gamma_x, \xi_x(c)) \equiv (1 - p(x), F_{y|1,x}(c)p(x))$. Since $\theta \in \mathcal{C}^k_\tau$ if and only if $c^k_{\tau,L}(x) \le \theta(x) \le c^k_{\tau,U}(x)$ for all $x \in \mathcal{X}$, it follows that:

$$\begin{aligned} P\left(\theta \in \hat{\mathcal{C}}^k_\tau | \{X_i\}_{i=1}^n\right) &= P\left(\hat{c}^k_{\tau,L}(x) \le \theta(x) \le \hat{c}^k_{\tau,U}(x)\ \forall x \in \mathcal{X} | \{X_i\}_{i=1}^n\right) \\ &= \prod_{x\in\mathcal{X}} P\left(\hat{c}^k_{\tau,L}(x) \le \theta(x) \le \hat{c}^k_{\tau,U}(x) | \{X_i\}_{i=1}^n\right) \\ &\ge \prod_{x\in\mathcal{X}} (1 - \alpha_x) \end{aligned} \tag{65}$$

where for the second equality we have used that $(\hat{c}^k_{\tau,L}(x), \hat{c}^k_{\tau,U}(x))$ are independent across $x \in \mathcal{X}$ conditional on $\{X_i\}_{i=1}^n$, and the inequality follows from Lemma 3.1. The claim of the Lemma is then implied by (65) and $\prod_{x\in\mathcal{X}}(1 - \alpha_x) = 1 - \alpha$ by hypothesis. ■

Proof of Lemma 3.3: Observe that $f \in \mathbf{R}$ belongs to the identified set for $\lambda'\beta_\tau$ if and only if there exists a $\beta \in \mathbf{R}^l$ such that $\lambda'\beta = f$ and $c^k_{\tau,L}(x) \le x'\beta \le c^k_{\tau,U}(x)$ for all $x \in \mathcal{X}$. It follows that the identified set is convex, and therefore an interval in $\mathbf{R}$. All that remains is to determine the lower and upper ends of the interval, which are by construction given by $\phi^k_{\tau,L}$ and $\phi^k_{\tau,U}$ respectively. ■

Proof of Corollary 3.1: By Lemma 3.3, it follows that $f \in [\phi^k_{\tau,L}, \phi^k_{\tau,L}]$ if and only if there exists a $\tilde{\beta}_\tau \in \mathcal{C}^k_\tau$ such that $\lambda'\tilde{\beta}_\tau = f$. Hence, by Lemma 3.2, we are then able to conclude:

$$\begin{aligned} 1 - \alpha &\le P(x'\tilde{\beta}_\tau \in \hat{\mathcal{C}}^k_\tau\ | \{X_i\}_{i=1}^n) \\ &\le P(\hat{\phi}^k_{\tau,L} \le f \le \hat{\phi}^k_{\tau,L}\ | \{X_i\}_{i=1}^n)\ , \end{aligned} \tag{66}$$

where the second inequality follows by definition of $\hat{\phi}^k_{\tau,L}$ and $\hat{\phi}^k_{\tau,U}$. ■

Proof of Lemma 4.1: For any $\theta \in \mathcal{C}^k_\tau$, the first order condition of the norm minimization problem yields $\beta^*_\tau = (E_S[X_i X_i'])^{-1} E_S[X_i \theta(X_i)]$. The Lemma then follows from Corollary 2.1. ■

Proof of Corollary 4.1: Since $\mathcal{P}^k_\tau$ is convex by Lemma 4.1, 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 4.1. ■

Lemma 6.1. *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:*

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

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 $1 \le j \le \lceil \frac{8}{\epsilon} \rceil$ we have $F_{y|1,x}(y_j) - F_{y|1,x}(y_{j-1}) < \epsilon/4$. Let

$$l_j(y, x, d, w) \equiv w1\{y \le y_{j-1},\ d = 1,\ x = x_0\} - P(Y_i \le y_j,\ D_i = 1,\ X_i = x_0) \tag{67}$$

$$u_j(y, x, d, w) \equiv w1\{y \le y_j,\ d = 1,\ x = x_0\} - P(Y_i \le y_{j-1},\ D_i = 1,\ X_i = x_0) \tag{68}$$

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:

$$\begin{aligned} &E[(u_j(Y_i, X_i, D_i, W_i) - l_j(Y_i, X_i, D_i, W_i))^2] \\ &\quad \le 2E[W_i^2 1\{y_{j-1} \le Y_i \le y_j,\ D_i = 1,\ X_i = x_0\}] + 2P^2(y_{j-1} \le Y_i \le y_j,\ D_i = 1,\ X_i = x_0) \\ &\quad \le 4E[W_i^2] \times (F_{y|1,x}(y_j) - F_{y|1,x}(y_{j-1})) \ , \end{aligned} \tag{69}$$

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

Lemma 6.2. *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:*

$$s_0(\tau, b, x) = \arg\min_{c \in \mathbf{R}} Q_x(c|\tau, b) \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c \in \mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b) \ . \tag{70}$$

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:

$$P(Y_i \le s_0(\tau, b, x), D_i = 1 | X_i = x) = \tau - bP(D_i = 0 | X_i = x) \tag{71}$$

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 (71) and the definition of the set $\mathcal{S}$ we then obtain that for all $x \in \mathcal{X}$:

$$-\infty < \underline{s}(x) \le \inf_{(\tau,b) \in \mathcal{S}} s_0(\tau, b, x) \le \sup_{(\tau,b) \in \mathcal{S}} s_0(\tau, b, x) \le \bar{s}(x) < +\infty \ . \tag{72}$$

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:

$$R_x(\tau, b) \equiv bP(D_i = 0, X_i = x) - \tau P(X_i = x) \tag{73}$$

$$R_{x,n}(\tau, b) \equiv \frac{1}{n}\sum_{i=1}^{n} W_i(b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\}) \tag{74}$$

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

$$(\underline{\tau}_n(x), \underline{b}_n(x)) \in \arg\max_{(\tau,b)\in\mathcal{S}} R_{x,n}(\tau, b) \qquad (\bar{\tau}_n(x), \bar{b}_n(x)) \in \arg\min_{(\tau,b)\in\mathcal{S}} R_{x,n}(\tau, b) \ . \tag{75}$$

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

$$\underline{S}_n(x) \equiv \Big\{\underline{s}_n(x) \in \mathbf{R} : \underline{s}_n(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\underline{\tau}_n(x), \underline{b}_n(x))\Big\} \tag{76}$$

$$\bar{S}_n(x) \equiv \Big\{\bar{s}_n(x) \in \mathbf{R} : \bar{s}_n(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\bar{\tau}_n(x), \bar{b}_n(x))\Big\} \tag{77}$$

From the definition of $\tilde{Q}_{x,n}(c|\tau, b)$, we then obtain from (75), (76) and (77) that for all $x \in \mathcal{X}$:

$$\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) \le \inf_{(\tau,b)\in\mathcal{S}} \hat{s}_0(\tau, b, x) \le \sup_{(\tau,b)\in\mathcal{S}} \hat{s}_0(\tau, b, x) \le \sup_{\bar{s}_n(x)\in\bar{S}_n(x)} \bar{s}_n(x) \ . \tag{78}$$

We establish the second claim of the Lemma, by exploiting (78) and showing that for some $0 < M < \infty$:

$$P\Big(\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) < -M\Big) = o(1) \qquad P\Big(\sup_{\bar{s}_n(x)\in\bar{S}_n(x)} \bar{s}_n(x) > M\Big) = o(1) \ . \tag{79}$$

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:

$$|R_{x,n}(\underline{\tau}_n(x), \underline{b}_n(x)) + \epsilon P(X_i = x)| = |R_{x,n}(\underline{\tau}_n(x), \underline{b}_n(x)) - \max_{(\tau,b)\in\mathcal{S}} R_x(\tau, b)| = o_p(1) \ , \tag{80}$$

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 (80) and Lemma 6.1, it follows that:

$$\sup_{c\in\mathbf{R}} |\tilde{Q}_{x,n}(c|\underline{\tau}_n(x), \underline{b}_n(x)) - (F_{y|1,x}(c)p(x) - \epsilon)^2 P^2(X_i = x)| = o_p(1) \ . \tag{81}$$

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:

$$\eta \equiv \inf_{|c-\underline{s}(x)|>\delta} (F_{y|1,x}(c)p(x) - \epsilon)^2 > 0 \ . \tag{82}$$

Therefore, it follows from direct manipulations and the definition of $\underline{S}_n(x)$ in (76) and of $\underline{s}(x)$ that:

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

We hence conclude from (81) that $\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) \xrightarrow{p} \underline{s}(x)$, which together with (72) 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 (72) establishes (79). The second claim of the Lemma then follows from (78), (79) and $\mathcal{X}$ being finite. ■

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:*

$$s_0(\tau, b, x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b) \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b) \ . \tag{83}$$

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:

$$M(\theta) \equiv \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} Q_x(\theta(\tau, b, x)|\tau, b) \qquad M_n(\theta) \equiv \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \tilde{Q}_{x,n}(\theta(\tau, b, x)|\tau, b) \ . \tag{84}$$

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 6.2, $s_0 \in L^\infty(\mathcal{S}\times\mathcal{X})$ while with probability tending to one $\hat{s}_0 \in L^\infty(\mathcal{S}\times\mathcal{X})$. Hence, (83) implies that with probability tending to one:

$$\hat{s}_0 \in \arg\min_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})} M_n(\theta) \qquad s_0 = \arg\min_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})} M(\theta) \ . \tag{85}$$

By Assumption 2.1(ii) and (71), $Q_x(c|\tau, b)$ is strictly convex in a neighborhood of $s_0(\tau, b, x)$. Furthermore, since by (71) and the implicit function theorem $s_0(\tau, b, x)$ is continuous in $(\tau, b) \in \mathcal{S}$ for every $x \in \mathcal{X}$:

$$\begin{aligned}\inf_{\|\theta - s_0\|_\infty \geq \delta} M(\theta) &\geq \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \inf_{|c - s_0(\tau,b,x)|\geq\delta} Q_x(c|\tau, b) \\ &= \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \min\{Q_x(s_0(\tau, b, x) - \delta|\tau, b), Q_x(s_0(\tau, b, x) + \delta|\tau, b) > 0 \ ,\end{aligned} \tag{86}$$

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 (86) 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:

$$G_{x,i}(c) \equiv W_i 1\{Y_i \leq c, D_i = 1, X_i = x\} \tag{87}$$

and observe that compactness of $\mathcal{S}$, a regular law of large numbers, Lemma 6.1 and finiteness of $\mathcal{X}$ yields:

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

where $R_x(\tau, b)$ and $R_{x,n}(\tau, b)$ are as in (73) and (74) respectively. Hence, using (88), 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:

$$\begin{aligned}\sup_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})} |M_n(\theta) - M(\theta)| &\leq \sup_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})} \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\tilde{Q}_{x,n}(\theta(\tau, b, x)|\tau, b) - Q_x(\theta(\tau, b, x)|\tau, b)| \\ &\leq \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \sup_{c\in\mathbf{R}} |\tilde{Q}_{x,n}(c|\tau, b) - Q_x(c|\tau, b)| = o_p(1) \ .\end{aligned} \tag{89}$$

The claim of the Lemma then follows from results (85), (86) and (89) together with Corollary 3.2.3 in van der Vaart and Wellner (1996). ■

Lemma 6.4. *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 $0 < 2\epsilon < \inf_{x\in\mathcal{X}} p(x)$ and denote the minimizers:*

$$s_0(\tau, b, x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b) \qquad\qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b) \ . \tag{90}$$

For $G_{x,i}(c) \equiv W_i 1\{Y_i \leq c, D_i = 1, X_i = x\}$ and $R_{x,n}(\tau, b)$ as defined in (74), denote the criterion function:

$$\tilde{Q}^s_{x,n}(c|\tau, b) \equiv \Big(\frac{1}{n}\sum_{i=1}^n \{E[G_{x,i}(c) - G_{x,i}(s_0(\tau, b, x))] + G_{x,i}(s_0(\tau, b, x))\} + R_{x,n}(\tau, b)\Big)^2 \ . \tag{91}$$

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

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau, b, x)|\tau, b)}{dc}\Big| = o_p(n^{-\frac{1}{2}}) \ . \tag{92}$$

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

$$M_n^s(\theta) \equiv \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \tilde{Q}^s_{x,n}(\theta(\tau, b, x)|\tau, b) \ . \tag{93}$$

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 6.1, compactness of $\mathcal{S}$, finiteness of $\mathcal{X}$ and the law of large numbers:

$$\begin{aligned}\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} &|\frac{1}{n}\sum_{i=1}^n \{G_{x,i}(s_0(\tau, b, x)) - E[G_{x,i}(s_0(\tau, b, x))]\} + R_{x,n}(\tau, b) - R_x(\tau, b)| \\ &\leq \sup_{x\in\mathcal{X}} \sup_{c\in\mathbf{R}} |\frac{1}{n}\sum_{i=1}^n \{G_{x,i}(c) - E[G_{x,i}(c)]\}| + \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |R_{x,n}(\tau, b) - R_x(\tau, b)| = o_p(1) \ ,\end{aligned} \tag{94}$$

where $R_x(\tau, b)$ is as in (73). Hence, by definition of $\mathcal{S}$ and $R_x(\tau, b)$, with probability tending to one:

$$\begin{aligned}\frac{\epsilon}{2}P(X_i = x) &\leq \frac{1}{n}\sum_{i=1}^n \{E[G_{x,i}(s_0(\tau, b, x))] - G_{x,i}(s_0(\tau, b, x))\} - R_{x,n}(\tau, b) \\ &\leq (p(x) - \frac{\epsilon}{2})P(X_i = x) \qquad\qquad \forall(\tau, b, x) \in \mathcal{S}\times\mathcal{X} \ .\end{aligned} \tag{95}$$

By Assumption 2.1(ii), whenever (95) holds, we may implicitly define $\hat{s}_0^s(\tau, b, x)$ by the equality:

$$P(Y_i \leq \hat{s}^s_0(\tau, b, x), D_i = 1, X_i = x) = \frac{1}{n}\sum_{i=1}^n \{E[G_{x,i}(s_0(\tau, b, x))] - G_{x,i}(s_0(\tau, b, x))\} - R_{x,n}(\tau, b) \ , \tag{96}$$

for all $(\tau, b, x) \in \mathcal{S}\times\mathcal{X}$ and set $\hat{s}^s_0(\tau, b, x) = 0$ for all $(\tau, b, x) \in \mathcal{S}\times\mathcal{X}$ whenever (95) does not hold. Thus,

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\tilde{Q}^s_{x,n}(\hat{s}^s_0(\tau, b, x)|\tau, b) - \inf_{c\in\mathbf{R}} \tilde{Q}^s_{x,n}(c|\tau, b)| = o_p(n^{-1}) \ . \tag{97}$$

Let $\hat{s}^s_0 \equiv \hat{s}^s_0(\cdot,\cdot,\cdot)$ and note that by construction $\hat{s}^s_0 \in L^\infty(\mathcal{S}\times\mathcal{X})$. From (97) we then obtain that:

$$M_n^s(\hat{s}^s_0) \leq \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \inf_{c\in\mathbf{R}} \tilde{Q}^s_{x,n}(c|\tau, b) + o_p(n^{-1}) \leq \inf_{\theta\in L^\infty(\mathcal{S}\times\mathcal{X})} M_n^s(\theta) + o_p(n^{-1}) \ . \tag{98}$$

In order to establish $\|\hat{s}_0^s - s_0\|_\infty = o_p(1)$, let $M(\theta)$ be as in (84) and notice that arguing as in (89) together with result (94) and Lemma 6.1 implies that:

$$\sup_{\theta \in L^\infty(\mathcal{S}\times\mathcal{X})} |M_n^s(\theta) - M(\theta)| \leq \sup_{\theta \in L^\infty(\mathcal{S}\times\mathcal{X})} \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\tilde{Q}_{x,n}^s(\theta(\tau,b,x)|\tau,b) - Q_x(\theta(\tau,b,x)|\tau,b)|$$

$$\leq \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \sup_{c\in\mathbf{R}} |\tilde{Q}_{x,n}^s(c|\tau,b) - Q_x(c|\tau,b)| = o_p(1) \ . \quad (99)$$

Hence, by (86), (98), (99) and Corollary 3.2.3 in van der Vaart and Wellner (1996) we obtain:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\hat{s}_0^s(\tau,b,x) - s_0(\tau,b,x)| = o_p(1) \ . \quad (100)$$

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:

$$\Delta_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \{(G_{x,i}(\theta(\tau,b,x)) - E[G_{x,i}(\theta(\tau,b,x))]) - (G_{x,i}(s_0(\tau,b,x)) - E[G_{x,i}(s_0(\tau,b,x))])\} \ , \quad (101)$$

and observe that Lemma 6.1 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 6.1 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:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}_{x,n}^s(\hat{s}_0(\tau,b,x)|\tau,b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau,b,x)|\tau,b)\}$$

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

$$\leq \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}_{x,n}(\hat{s}_0^s(\tau,b,x)|\tau,b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau,b,x)|\tau,b)\} + o_p(n^{-1}) \ , \quad (102)$$

where $M_n(\theta)$ is as in (84). Furthermore, since by (98) we have $M_n^s(\hat{s}_0^s) \leq M_n^s(s_0) + o_p(n^{-1})$ and by Lemma 6.1 and finiteness of $\mathcal{X}$ we have $M_n^s(s_0) = O_p(n^{-1})$, similar arguments as in (102) imply that:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}_{x,n}(\hat{s}_0^s(\tau,b,x)|\tau,b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau,b,x)|\tau,b)\}$$

$$\leq \|\Delta_n(\hat{s}_0^s)\|_\infty^2 + 2\|\Delta_n(\hat{s}_0^s)\|_\infty \times [M_n^s(\hat{s}_0^s)]^{\frac{1}{2}} = o_p(n^{-1}) \ . \quad (103)$$

Therefore, by combining the results in (97), (102) and (103), we are able to conclude that:

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

$$\leq \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}_{x,n}^s(\hat{s}_0(\tau,b,x)|\tau,b) - \tilde{Q}_{x,n}^s(\hat{s}_0^s(\tau,b,x)|\tau,b)\} + o_p(n^{-1}) \leq o_p(n^{-1}) \ . \quad (104)$$

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

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |\tilde{Q}_{x,n}^s(\hat{s}_0(\tau,b,x)|\tau,b) - \inf_{c\in\mathbf{R}} \tilde{Q}_{x,n}^s(c|\tau,b)| = o_p(\epsilon_n^2) \ , \quad (105)$$

which is possible by (104). A Taylor expansion at each $(\tau, b, x) \in \mathcal{S} \times \mathcal{X}$ then implies:

$$\begin{aligned} 0 &\leq \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)+\epsilon_n|\tau,b) - \tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)\} + o_p(\epsilon_n^2) \\ &= \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in S} \Big\{\epsilon_n \times \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + \frac{\epsilon_n^2}{2} \times \frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}\Big\} + o_p(\epsilon_n^2) \ , \end{aligned} \tag{106}$$

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 6.3 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 (89) yield:

$$\begin{aligned} &\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{1}{n}\sum_{i=1}^n\{E[G_{x,i}(\bar{s}(\tau,b,x)) - G_{x,i}(s_0(\tau,b,x))] + G_{x,i}(s_0(\tau,b,x))\} + R_{x,n}(\tau,b)\}\Big| \\ &\qquad\qquad \leq \sup_{c\in\mathbf{R}} f_{y|1,x}(c)p(x)P(X_i = x) \times \|\bar{s} - s_0\|_\infty + M_n^{\frac{1}{2}}(s_0) = o_p(1) \ . \end{aligned} \tag{107}$$

Therefore, exploiting (107), $f'_{y|1,x}(c)$ being uniformly bounded and by direct calculation we conclude:

$$\begin{aligned} &\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2} - 2f^2_{y|1,x}(\bar{s}(\tau,b,x))p^2(x)P^2(X_i = x)\Big| \\ &\qquad\qquad \leq \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} |f'_{y|1,x}(\bar{s}(\tau,b,x))p(x)P(X_i = x)| \times o_p(1) = o_p(1). \end{aligned} \tag{108}$$

Thus, combining results (106) together with (108) and $f_{y|1,x}(c)$ uniformly bounded, we conclude:

$$0 \leq \epsilon_n \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + O_p(\epsilon_n^2) \ . \tag{109}$$

In a similar fashion, we note that by exploiting (105) and proceeding as in (106)-(109) we obtain:

$$\begin{aligned} 0 &\leq \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \{\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)-\epsilon_n|\tau,b) - \tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)\} + o_p(\epsilon_n^2) \\ &\leq \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}} \Big\{-\epsilon_n \times \frac{d\tilde{Q}_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + \frac{\epsilon_n^2}{2} \times \frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}\Big\} + o_p(\epsilon_n^2) \\ &\leq -\epsilon_n \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \frac{d\tilde{Q}_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + O_p(\epsilon_n^2) \ . \end{aligned} \tag{110}$$

Therefore, since $\epsilon_n = o_p(n^{-\frac{1}{2}})$, we conclude from (109) and (110) that we must have:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} = O_p(\epsilon_n) = o_p(n^{-\frac{1}{2}}) \ . \tag{111}$$

By similar arguments, but reversing the sign of ϵ_n in (106) and (110) it possible to establish that:

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}} -\frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} = o_p(n^{-\frac{1}{2}}) \ . \tag{112}$$

The claim of the Lemma then follows from (111) and (112). ■

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

$$s_0(\tau, b, x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b) \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b) \ . \tag{113}$$

If $G_{x,i}(c)$ is as in (87), $\inf_{x\in\mathcal{X}}\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x))p(x) > 0$ *and* $\{Y_i, X_i, D_i, W_i\}$ *is i.i.d., then:*

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\Big|(\hat{s}_0(\tau,b,x)-s_0(\tau,b,x)) - \frac{1}{n}\sum_{i=1}^n \frac{G_{x,i}(s_0(\tau,b,x)) + W_i(b1\{D_i=0,X_i=x\}-\tau 1\{X_i=x\})}{P(X_i=x)p(x)f_{y|1,x}(s_0(\tau,b,x))}\Big| = o_p(n^{-\frac{1}{2}}) \ . \quad (114)$$

PROOF: For $\tilde{Q}^s_{x,n}(c|\tau,b)$ as in (91), note that the mean value theorem and Lemma 6.4 imply:

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\Big|(\hat{s}_0(\tau,b,x)-s_0(\tau,b,x))\times\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}+\frac{d\tilde{Q}^s_{x,n}(s_0(\tau,b,x)|\tau,b)}{dc}\Big| = o_p(n^{-\frac{1}{2}}) \quad (115)$$

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 6.1 implies:

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\Big|\frac{d\tilde{Q}^s_{x,n}(s_0(\tau,b,x)|\tau,b)}{dc}\Big|$$
$$= \sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\Big|2f_{y|1,x}(s_0(\tau,b,x))p(x)P(X_i=x)\times\{\frac{1}{n}\sum_{i=1}^n G_{x,i}(s_0(\tau,b,x))+R_n(\tau,b)\}\Big| = O_p(n^{-\frac{1}{2}}) \ . \quad (116)$$

In addition, by (108), the mean value theorem and $f_{y|1,x}(c)$ being uniformly bounded we conclude that:

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\Big|\frac{d^2\tilde{Q}_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2} - 2f^2_{y|1,x}(s_0(\tau,b,x))p^2(x)P^2(X_i=1)\Big|$$
$$\lesssim \sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}|f^2_{y|1,x}(\bar{s}(\tau,b,x)) - f^2_{y|1,x}(s_0(\tau,b,x))| + o_p(1) \lesssim \sup_{c\in\mathbf{R}}|f'_{y|1,x}(c)|\times\|\bar{s}-s_0\|_\infty + o_p(1) \ . \quad (117)$$

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 (117) and $\|\bar{s}-s_0\|_\infty = o_p(1)$ by Lemma 6.3 that for some $\delta>0$:

$$\inf_{x\in\mathcal{X}}\inf_{(\tau,b)\in\mathcal{S}}\frac{d^2\tilde{Q}_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2} > \delta \quad (118)$$

with probability approaching one. Therefore, we conclude from results (115), (116) and (118) that we must have $\|\hat{s}_0 - s_0\|_\infty = O_p(n^{-\frac{1}{2}})$. Hence, by (115) and (117) we conclude that:

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}}\Big|2(\hat{s}_0(\tau,b,x)-s_0(\tau,b,x))f^2_{y|1,x}(s_0(\tau,b,x))p^2(x)P^2(X_i=1)+\frac{d\tilde{Q}^s_{x,n}(s_0(\tau,b,x)|\tau,b)}{dc}\Big| = o_p(n^{-\frac{1}{2}}) \quad (119)$$

The claim of the Lemma is then established by (116), (118) and (119). ■

Lemma 6.6. *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 $\mathcal{S}\equiv\{(\tau,b)\in[0,1]^2 : b\{1-p(x)\}+\epsilon\le\tau\le p(x)+b\{1-p(x)\}-\epsilon \ \forall x\in\mathcal{X}\}$ for some ϵ satisfying $0<2\epsilon<\inf_{x\in\mathcal{X}}p(x)$ and 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:

$$\begin{aligned} &\underline{\tau}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} \tau && \bar{\tau}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} \tau \\ &\underline{b}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} b && \bar{b}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} b \\ &\underline{s}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} s_0(\tau,b,x_0) && \bar{s}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} s_0(\tau,b,x_0) \\ &\underline{f}_j = \min\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} f_{y|1,x}(s_0(\tau,b,x_0)) && \bar{f}_j = \max\nolimits_{(\tau,b)\in\mathcal{S}\cap B_j} f_{y|1,x}(s_0(\tau,b,x_0))\ , \end{aligned} \tag{120}$$

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 (71) and the implicit function theorem and continuity of $f_{y|1,x}(c)$ by assumption 4.1(iii). Next, for $1 \leq j \leq \#\{B_j\}$ define the functions:

$$l_j(y,x,d,w) \equiv \frac{w1\{y \leq \underline{s}_j,\ d=1,\ x=x_0\} + \underline{b}_j w1\{d=0, x=x_0\}}{P(X_i=x_0)p(x_0)\bar{f}_j} - \frac{\bar{\tau}_j w1\{x=x_0\}}{P(X_i=x_0)p(x_0)\underline{f}_j} \tag{121}$$

$$u_j(y,x,d,w) \equiv \frac{w1\{y \leq \bar{s}_j,\ d=1,\ x=x_0\} + \bar{b}_j w1\{d=0, x=x_0\}}{P(X_i=x_0)p(x_0)\underline{f}_j} - \frac{\underline{\tau}_j w1\{x=x_0\}}{P(X_i=x_0)p(x_0)\bar{f}_j} \tag{122}$$

and note that the brackets $[l_j, u_j]$ cover the class $\mathcal{F}$. Since $\bar{f}_j^{-1} \leq \underline{f}_j^{-1} \leq [\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))]^{-1} < \infty$ 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:

$$\begin{aligned} E[(u_j(Y_i,X_i,D_i,W_i) - l_j(Y_i,X_i,D_i,W_i))^2] &\leq M \times (\bar{b}_j\bar{f}_j - \underline{b}_j\underline{f}_j)^2 + M \times (\bar{\tau}_j\bar{f}_j - \underline{\tau}_j\underline{f}_j)^2 \\ &+ M \times E[(1\{Y_i \leq \underline{s}_j, D_i=1, X_i=x_0\}\underline{f}_j - 1\{Y_i \leq \bar{s}_j, D_i=1, X_i=x_0\}\bar{f}_j)^2] \end{aligned} \tag{123}$$

Next observe that by the implicit function theorem and result (71) we can conclude that for any $(\tau,b) \in \mathcal{S}$:

$$\frac{ds_0(\tau,b,x_0)}{d\tau} = \frac{1}{f_{y|1,x}(s_0(\tau,b,x_0))} \qquad \frac{ds_0(\tau,b,x_0)}{db} = -\frac{1-p(x_0)}{f_{y|1,x}(s_0(\tau,b,x_0))}\ . \tag{124}$$

Since the minimums and maximums in (120) 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 (124) imply:

$$|\bar{s}_j - \underline{s}_j| = \Big|\frac{1}{f_{y|1,x}(s_0(\tilde{\tau},\tilde{b},x_0))}(\tau_1-\tau_2) + \frac{1-p(x)}{f_{y|1,x}(s_0(\tilde{\tau},\tilde{b},x_0))}(b_1-b_2)\Big| \leq \frac{\sqrt{2}\epsilon}{\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))} \tag{125}$$

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 (125) we conclude:

$$|\bar{f}_j - \underline{f}_j| \leq \sup_{c\in\mathbf{R}} |f'_{y|1,x}(c)| \times |\bar{s}_j - \underline{s}_j| \leq \sup_{c\in\mathbf{R}} |f'_{y|1,x}(c)| \times \frac{\sqrt{2}\epsilon}{\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))}\ . \tag{126}$$

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:

$$(\bar{b}_j\bar{f}_j - \underline{b}_j\underline{f}_j)^2 \leq 2\bar{f}_j^2(\bar{b}_j - \underline{b}_j)^2 + 2\underline{b}_j^2(\bar{f}_j - \underline{f}_j)^2 \leq 2\sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times \epsilon^2 + \frac{4\epsilon^2}{\inf_{(\tau,b)\in\mathcal{S}} f^2_{y|1,x}(s_0(\tau,b,x_0))}\ , \tag{127}$$

where in the final inequality we have used result (126). By similar arguments, we also obtain:

$$(\bar{\tau}_j\bar{f}_j - \underline{\tau}_j\underline{f}_j)^2 \leq 2\sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times \epsilon^2 + \frac{4\epsilon^2}{\inf_{(\tau,b)\in\mathcal{S}} f^2_{y|1,x}(s_0(\tau,b,x_0))}\ . \tag{128}$$

Also note that by direct calculation, the mean value theorem and results (125) and (126) it follows that:

$$
\begin{aligned}
E[(1\{Y_i \le \underline{s}_j, D_i = 1, X_i = x_0\}\underline{f}_j - 1\{Y_i \le \bar{s}_j, D_i = 1, X_i = x_0\}\bar{f}_j)^2] \\
&\le 2(\bar{f}_j - \underline{f}_j)^2 + \sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times P(X_i = x_0)p(x_0)(F_{y|1,x}(\bar{s}_j) - F_{y|1,x}(\underline{s}_j)) \\
&\le \frac{4\epsilon^2}{\inf_{(\tau,b)\in\mathcal{S}} f^2_{y|1,x}(s_0(\tau,b,x_0))} + \sup_{c\in\mathbf{R}} f^3_{y|1,x}(c) \times \frac{\sqrt{2}\epsilon}{\inf_{(\tau,b)\in\mathcal{S}} f_{y|1,x}(s_0(\tau,b,x_0))} . \quad (129)
\end{aligned}
$$

Thus, from (123) and (127)-(128), it follows that for $\epsilon < 1$ and some constant K not depending on j:

$$
E[(u_j(Y_i, X_i, D_i, W_i) - l_j(Y_i, X_i, D_i, W_i))^2] \le K\epsilon . \quad (130)
$$

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

PROOF OF THEOREM 4.1: Throughout the proof we exploit Lemmas 6.5 and 6.6 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

$$
\mathcal{S} \equiv \{(\tau, b) \in [0, 1]^2 : b\{1 - p(x)\} + \epsilon \le \tau \le p(x) + b\{1 - p(x)\} - \epsilon \ \forall x \in \mathcal{X}\} . \quad (131)
$$

Also notice that for every $(\tau, k) \in \mathcal{B}$, the points $(\tau, \min\{\tau + k, 1\}), (\tau, \max\{\tau - k, 0\}) \in \mathcal{S}$ and therefore for $s_0(\tau, b, x)$ and $\hat{s}_0(\tau, b, x)$ as defined in (113) we additionally conclude that:

$$
\begin{aligned}
c^k_{\tau,L}(x) &= s_0(\tau, \min\{\tau + k, 1\}, x) \qquad & c^k_{\tau,U}(x) &= s_0(\tau, \max\{\tau - k, 0\}, x) \\
\check{c}^k_{\tau,L}(x) &= \hat{s}_0(\tau, \min\{\tau + k, 1\}, x) \qquad & \check{c}^k_{\tau,U}(x) &= \hat{s}_0(\tau, \max\{\tau - k, 0\}, x)
\end{aligned} \quad (132)
$$

Next, observe that since X_i 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:

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

Since w is also a function on $\mathcal{X}$, we refer to its coordinates by $w(x)$. Solving the linear programming problems in (20) and (21), it is then possible to obtain the closed form solution:

$$
\begin{aligned}
\pi_L(\tau, k) &= \sum_{x\in\mathcal{X}} \{1\{w(x) \ge 0\}w(x)c^k_{\tau,L}(x) + 1\{w(x) \le 0\}w(x)c^k_{\tau,U}(x)\} \\
\pi_U(\tau, k) &= \sum_{x\in\mathcal{X}} \{1\{w(x) \ge 0\}w(x)c^k_{\tau,U}(x) + 1\{w(x) \le 0\}w(x)c^k_{\tau,L}(x)\}
\end{aligned} \quad (134)
$$

with a similar representation holding for $(\hat{\pi}_L(\tau, k), \hat{\pi}_U(\tau, k))$ but with $(\check{c}^k_{\tau,L}, \check{c}^k_{\tau,U})$ in place of $(c^k_{\tau,L}, c^k_{\tau,U})$. We hence define the linear map $K : L^\infty(\mathcal{S} \times \mathcal{X}) \to L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$, with $K(\theta)(\tau, k)$ to be given by:

$$
\begin{pmatrix}
\sum_{x\in\mathcal{X}} \{1\{w(x) \ge 0\}w(x)\theta(\tau, \min\{\tau + k, 1\}, x) + 1\{w(x) \le 0\}w(x)\theta(\tau, \max\{\tau - k, 0\})\} \\
\sum_{x\in\mathcal{X}} \{1\{w(x) \ge 0\}w(x)\theta(\tau, \max\{\tau - k, 0\}, x) + 1\{w(x) \le 0\}w(x)\theta(\tau, \min\{\tau + k, 1\})\}
\end{pmatrix} \quad (135)
$$

for any $\theta \in L^\infty(\mathcal{S} \times \mathcal{X})$. It then follows from (132), (134) and the linearity of K that:

$$
\sqrt{n}\begin{pmatrix} \hat{\pi}_L - \pi_L \\ \hat{\pi}_U - \pi_U \end{pmatrix} = K(\sqrt{n}(\hat{s}_0 - s_0)) . \quad (136)
$$

We establish the Theorem by exploiting (136). Assumptions 2.1, 2.2 and 4.1 imply the conditions for Lemmas 6.5 and 6.6 are satisfied. Since $\mathcal{X}$ has finite cardinality and the finite sum of Donsker clases is Donsker, we can then conclude from Lemmas 6.5 and 6.6 that:

$$\sqrt{n}(\hat{s}_0(\tau, b, x) - s_0(\tau, b, x)) \xrightarrow{L} J(\tau, b, x) \tag{137}$$

for $J(\tau, b, x)$ Gaussian process on $L^\infty(\mathcal{S} \times \mathcal{X})$. Employing the norm $\|\cdot\|_\infty + \|\cdot\|_\infty$ on $L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$, we can then obtain by direct calculation that for any $\theta \in L^\infty(\mathcal{S} \times \mathcal{X})$:

$$\|K(\theta)\|_\infty \le 2 \sum_{x \in \mathcal{X}} |w(x)| \times \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}} |\theta(\tau, b, x)| = 2 \sum_{x \in \mathcal{X}} |w(x)| \times \|\theta\|_\infty \ , \tag{138}$$

which implies the linear map K is continuous. Therefore, the theorem is established by (136), (137), the continuous mapping theorem and linear maps of Gaussian processes also being Gaussian processes. ■

Proof of Lemma 4.2: We first show that (33) is satisfied if $r \ge r_{1-\alpha}$. To this end, note that if $g \in \mathcal{G}$, then $\pi_L(\tau, k) \le g(\tau, k) \le \pi_U(\tau, k)$ for all $(\tau, k) \in \mathcal{B}$, and hence we obtain:

$$\begin{aligned} \liminf_{n\to\infty} P(g \in \hat{\mathcal{G}}(\omega, r)) &= \liminf_{n\to\infty} P(\hat{\pi}_L(\tau, k) - \frac{r}{\sqrt{n}}\omega_L(\tau, k) \le g(\tau, k) \le \hat{\pi}_U(\tau, k) + \frac{r}{\sqrt{n}}\omega_U(\tau, k) \ \forall(\tau, k) \in \mathcal{B}) \\ &\ge \liminf_{n\to\infty} P(\max\{\frac{\sqrt{n}(\hat{\pi}_L(\tau, k) - \pi_L(\tau, k))}{\omega_L(\tau, k)}, \frac{\sqrt{n}(\pi_U(\tau, k) - \hat{\pi}_U(\tau, k))}{\omega_U(\tau, k)}\} \le r \ \forall(\tau, k) \in \mathcal{B}) \\ &= P(Z \ge r) \ , \end{aligned} \tag{139}$$

where the final inequality follows from Theorem 4.1, the continuous mapping theorem, the Portmanteau lemma and Z being continuously distributed due to Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998). Since for all $r \ge r_{1-\alpha}$ we have $P(Z \ge r) \ge 1 - \alpha$, it follows that (33) holds for all $r \ge r_{1-\alpha}$.

We next show that (33) fails to hold if $r < r_{1-\alpha}$. Since $r_{1-\alpha}$ is the $1 - \alpha$ quantile of Z, it follows that:

$$P(Z \le r) \le 1 - \alpha - \eta \ , \tag{140}$$

for some $\eta > 0$. Next, observe that since continuous functions on compact sets are uniformly continuous and both $\omega_L(\tau, k)$ and $\omega_U(\tau, k)$ are bounded away from zero on $\mathcal{B}$, it follows that $(\omega_L(\tau, k))^{-1}$ and $(\omega_U(\tau, k))^{-1}$ are uniformly continuous on $\mathcal{B}$. Therefore, defining the processes:

$$\tilde{G}_{n,L}(\tau, k) \equiv \frac{\sqrt{n}(\hat{\pi}_L(\tau, k) - \pi_L(\tau, k))}{\omega_L(\tau, k)} \qquad \tilde{G}_{n,U}(\tau, k) \equiv \frac{\sqrt{n}(\hat{\pi}_U(\tau, k) - \pi_U(\tau, k))}{\omega_U(\tau, k)} \ , \tag{141}$$

it follows that $\tilde{G}_{n,L}(\tau, k)$ and $\tilde{G}_{n,U}(\tau, k)$ are asymptotically uniformly equicontinuous in probability by virtue of $\sqrt{n}(\hat{\pi}_L(\tau, k) - \pi_L(\tau, k), \hat{\pi}_U(\tau, k) - \pi_U(\tau, k))$ being asymptotically uniformly equicontinuous in probability and tight as a result of Theorem 4.1. Hence, for any $\epsilon > 0$ there exits a $\delta > 0$ such that:

$$\limsup_{n\to\infty} P\Big(\sup_{\|(\tau,k)-(\tau',k')\|<\delta} \max\{|\tilde{G}_{n,L}(\tau, k) - \tilde{G}_{n,L}(\tau', k')|, |\tilde{G}_{n,U}(\tau, k) - \tilde{G}_{n,U}(\tau', k')|\} > \epsilon\Big) < \frac{\eta}{2} \ . \tag{142}$$

In particular, since Z is continuously distributed, by (140) we may select ϵ so that the following holds:

$$P(Z \le r + \epsilon) < 1 - \alpha - \frac{\eta}{2} \ . \tag{143}$$

For some odd integer $K > \delta^{-1}$, partition $[0,1]^2$ into squares with sides of length $1/K$, and denote this partition by $\{A_j\}_{j=1}^{K^2}$ where the squares are enumerated from the upper left corner proceeding down the rows. Since K is odd, all even squares share sides only with odd squares and vice-versa. Define:

$$g_K(\tau,k) \equiv \sum_{j=1}^{K^2}\{1\{(\tau,k)\in A_j,\ j \text{ odd }\}\pi_L(\tau,k) + 1\{(\tau,k)\in A_j,\ j \text{ even }\}\pi_U(\tau,k)\} \tag{144}$$

for all $(\tau,k)\in[0,1]^2$ and notice that since the $\{A_j\}_{j=1}^{K^2}$ are nonoverlapping and form a partition of $[0,1]^2$, by construction we have that the restriction of g_K to $\mathcal{B}$ is an element of $\mathcal{G}$.

Since the inequalities defining $\mathcal{B}$ become more stringent for smaller values of $p(x)$ it follows that:

$$\mathcal{B} = \{(\tau,k)\in[0,1]^2 : \min\{\tau+k,1\}\times\{1-\underline{p}\}+\epsilon \le \tau \le \underline{p} + \max\{\tau-k,0\}\times\{1-\underline{p}\}-\epsilon\}\ , \tag{145}$$

where $\underline{p}\equiv \inf_{x\in\mathcal{X}} p(x)$. Furthermore, notice that (τ,k) satisfies the lower constraint if and only if $(1-\tau,k)$ satisfies the upper constraint. As a result, if we define the sets:

$$B_1 = \{(\tau,k)\in[0,1]^2 : \tau\le k,\ \tau\le \underline{p}-\epsilon\} \qquad B_2 = \{(\tau,k)\in[0,1]^2 : \tau\ge k,\ \tau\le\underline{p}+(\tau-k)(1-\underline{p})-\epsilon\}\ ,$$

it follows that $\mathcal{B} = (B_1\cup B_2)\cap(\mathcal{R}(B_1)\cup\mathcal{R}(B_2))$ where $\mathcal{R}(B)$ denotes the reflection of B along the line $\tau = 0.5$. Letting $H = \{(\tau,k)\in[0,1]^2 : \tau\ge 1/2\}$ we may therefore decompose $\mathcal{B}$ into the sets:

$$\mathcal{B} = (B_1\cap H)\cup\mathcal{R}(B_1\cap H)\cup(B_2\cap H)\cup\mathcal{R}(B_2\cap H)\ . \tag{146}$$

Hence $\mathcal{B}$ is the union of four convex sets, none of which is a singleton (though some may be empty). Letting K be such that $\sqrt{2}/K$ is smaller than the diameter of the sets in (146) that are not empty, it follows by convexity that each such set intersects with at least two adjacent squares in $\{A_j\}_{j=1}^{K}$. Thus, for every $(\tau,k)\in\mathcal{B}$, either $g_K(\tau,k)=\pi_L(\tau,k)$ or there is a $(\tau',k')\in\mathcal{B}$ such that $g_K(\tau',k')=\pi_L(\tau',k')$ and $\|(\tau,k)-(\tau',k')\|<\delta$ (and similarly for $\pi_U(\tau,k)$). Since $\pi_L(\tau,k)\le g_K(\tau,k)\le\pi_U(\tau,k)$ for all $(\tau,k)\in\mathcal{B}$,

$$\begin{aligned} 0 \le \sup_{(\tau,k)\in\mathcal{B}}\max\{\tilde{G}_{n,L}(\tau,k),-\tilde{G}_{n,U}(\tau,k)\} - \sup_{(\tau,k)\in\mathcal{B}}\max\{\frac{\sqrt{n}(\hat{\pi}_L(\tau,k)-g_K(\tau,k))}{\omega_L(\tau,k)},\frac{\sqrt{n}(g_K(\tau,k)-\hat{\pi}_U(\tau,k))}{\omega_U(\tau,k)}\} \\ \le \sup_{\|(\tau,k)-(\tau',k')\|<\delta}\max\{|\tilde{G}_{n,L}(\tau,k)-\tilde{G}_{n,L}(\tau',k')|,|\tilde{G}_{n,U}(\tau,k)-\tilde{G}_{n,U}(\tau',k')|\}\ . \end{aligned} \tag{147}$$

Therefore, by employing the results in (142), (143) and (147) together with the continuous mapping theorem, Z being continuously distributed and the Portmanteau lemma we are able to conclude that:

$$\begin{aligned} \liminf_{n\to\infty} P(g_K\in\hat{\mathcal{G}}(\omega,r)) = \liminf_{n\to\infty} P\Big(\sup_{(\tau,k)\in\mathcal{B}}\max\{\frac{\sqrt{n}(\hat{\pi}_L(\tau,k)-g_K(\tau,k))}{\omega_L(\tau,k)},\frac{\sqrt{n}(g_K(\tau,k)-\hat{\pi}_U(\tau,k))}{\omega_U(\tau,k)}\}\le r\Big) \\ \le \liminf_{n\to\infty} P\Big(\sup_{(\tau,k)\in\mathcal{B}}\max\{\tilde{G}_{n,L}(\tau,k),-\tilde{G}_{n,U}(\tau,k)\}\le r+\epsilon\Big)+\frac{\eta}{2} < 1-\alpha\ . \end{aligned} \tag{148}$$

It follows that (33) is not satisfied for $g_K\in\mathcal{G}$, which establishes the lemma. ■

Proof of Corollary 4.2: To see that (35) implies (33) is satisfied, notice that for every $g\in\mathcal{G}$ we have:

$$\liminf_{n\to\infty} P(g\in\hat{\mathcal{G}}(\omega,r)) \ge \liminf_{n\to\infty} P(\mathcal{G}\subseteq\hat{\mathcal{G}}(\omega,r)) \ge 1-\alpha\ . \tag{149}$$

On the other hand, if (33) is satisfied, then by Lemma 4.2 $r \geq r_{1-\alpha}$ and arguing as in (139) we obtain:

$$\liminf_{n\to\infty} P(\mathcal{G} \subseteq \hat{\mathcal{G}}(\omega, r))$$
$$= \liminf_{n\to\infty} P\Big(\sup_{(\tau,k)\in\mathcal{B}} \max\{\frac{\sqrt{n}(\hat{\pi}_L(\tau,k) - \pi_L(\tau,k))}{\omega_L(\tau,k)}, \frac{\sqrt{n}(\pi_U(\tau,k) - \hat{\pi}_U(\tau,k))}{\omega_U(\tau,k)}\} \leq r\Big) \geq 1-\alpha \ . \quad (150)$$

The claim of the corollary then follows from (149) and (150). ■

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:

$$\sup_{h\in BL_1(\mathbf{R})} |E[h(\tilde{Z})|\mathcal{Z}_n] - E[h(Z)]| = o_p(1) \ , \quad (151)$$

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

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

and notice that Lemma 6.5 applied to $W_i = 1$ with probability one, implies that:

$$\hat{s}_0(\tau, b, x) - s_0(\tau, b, x)$$
$$= \frac{1}{n}\sum_{i=1}^n \frac{1\{Y_i \leq s_0(\tau, b, x), D_i = 1, X_i = x\} + b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\}}{P(X_i = x)p(x)f_{y|1,x}(s_0(\tau, b, x))} + o_p(n^{-\frac{1}{2}}) \quad (152)$$

uniformly in $(\tau, b, x) \in \mathcal{S} \times \mathcal{X}$, for $\mathcal{S}$ as defined in (131). Furthermore, also by Lemma 6.5 we obtain:

$$\tilde{s}_0(\tau, b, x) - s_0(\tau, b, x)$$
$$= \frac{1}{n}\sum_{i=1}^n \frac{W_i(1\{Y_i \leq s_0(\tau, b, x), D_i = 1, X_i = x\} + b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\})}{P(X_i = x)p(x)f_{y|1,x}(s_0(\tau, b, x))} + o_p(n^{-\frac{1}{2}}) \quad (153)$$

uniformly in $(\tau, b, x) \in \mathcal{S}\times\mathcal{X}$. Letting K be the continuous linear operator defined in (135), we then obtain

$$\sqrt{n}\begin{pmatrix} \tilde{\pi}_L - \hat{\pi}_L \\ \tilde{\pi}_U - \hat{\pi}_U \end{pmatrix} = K(\sqrt{n}(\tilde{s}_0 - \hat{s}_0)) \ , \quad (154)$$

for $\tilde{s}_0 = \tilde{s}_0(\cdot,\cdot,\cdot)$ and $\hat{s}_0 = s_0(\cdot,\cdot,\cdot)$. By Lemma 6.6, results (152) and (153) and Theorem 2.9.2 in van der Vaart and Wellner (1996), the process $\sqrt{n}(\tilde{s}_0 - \hat{s}_0)$ converges to a tight Gaussian process on $L^\infty(\mathcal{S}\times\mathcal{X})$. Hence, by the continuous mapping theorem, the process $K(\sqrt{n}(\tilde{s}_0 - \hat{s}_0))$ is asymptotically tight. Define,

$$Z_0 \equiv \sup_{(\tau,k)\in\mathcal{B}} \max\Big\{\frac{\sqrt{n}(\tilde{\pi}_L(\tau,k) - \hat{\pi}_L(\tau,k))}{\omega_L(\tau,k)}, \frac{\sqrt{n}(\hat{\pi}_U(\tau,k) - \tilde{\pi}_U(\tau,k))}{\omega_U(\tau,k)}\Big\} \ , \quad (155)$$

and notice that $\omega_L(\tau, b)$ and $\omega_U(\tau, b)$ being bounded away from zero, $\hat{\omega}_L(\tau, b)$ and $\hat{\omega}_U(\tau, b)$ being uniformly consistent by Assumption 4.2(ii) and tightness of $K(\sqrt{n}(\tilde{s}_0 - \hat{s}_0))$ imply that:

$$|\tilde{Z} - Z_0| \leq \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{\sqrt{n}(\tilde{\pi}_L(\tau,k) - \hat{\pi}_L(\tau,k))}{\hat{\omega}_L(\tau,k)} - \frac{\sqrt{n}(\tilde{\pi}_L(\tau,k) - \hat{\pi}_L(\tau,k))}{\omega_L(\tau,k)}\Big|$$
$$+ \sup_{(\tau,b)\in\mathcal{S}} \Big|\frac{\sqrt{n}(\tilde{\pi}_U(\tau,k) - \hat{\pi}_U(\tau,k))}{\hat{\omega}_U(\tau,k)} - \frac{\sqrt{n}(\tilde{\pi}_U(\tau,k) - \hat{\pi}_U(\tau,k))}{\omega_U(\tau,k)}\Big| = o_p(1) \ . \quad (156)$$

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. Therefore, for any $\eta > 0$ it follows that:

$$P\Big(\sup_{h\in BL_1(\mathbf{R})} |E[h(\tilde{Z})|\mathcal{Z}_n] - E[h(Z_0)|\mathcal{Z}_n]| > \eta\Big) \leq P\Big(2P(|\tilde{Z} - Z_0| > \frac{\eta}{2}|\mathcal{Z}_n) + \frac{\eta}{2}P(|\tilde{Z} - Z_0| \leq \frac{\eta}{2}|\mathcal{Z}_n) > \eta\Big)$$
$$\leq P\Big(P\Big(|\tilde{Z} - Z_0| > \frac{\eta}{2}|\mathcal{Z}_n\Big) > \frac{\eta}{4}\Big) \leq \frac{4}{\eta}E\Big[E\Big[1\Big\{|\tilde{Z} - Z_0| > \frac{\eta}{2}\Big\}|\mathcal{Z}_n\Big]\Big] \leq \frac{4}{\eta}P\Big(|\tilde{Z} - Z_0| > \frac{\eta}{2}\Big) = o(1) \quad (157)$$

where the third and fourth inequality hold by Markov's inequality and Fubini's theorem for outer expectations (cf Lemma 1.2.6 in van der Vaart and Wellner (1996)), while the final result follows by (156).

For $\theta \in L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$, let $\theta^{(i)}(\tau, k)$ denote its i^{th} coordinate evaluated at (τ, k), and define:

$$T(\theta) \equiv \sup_{(\tau,k)\in\mathcal{B}} \max\Big\{\frac{\theta^{(1)}(\tau, k)}{\omega_L(\tau, k)}, -\frac{\theta^{(2)}(\tau, k)}{\omega_U(\tau, k)}\Big\} . \quad (158)$$

Notice that the mapping $T : L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B}) \to \mathbf{R}$ satisfies for every $\theta_1, \theta_2 \in L^\infty(\mathcal{B}) \times L^\infty(\mathcal{B})$:

$$|T(\theta_1) - T(\theta_2)| \leq M_0\{\|\theta_1^{(1)} - \theta_2^{(1)}\|_\infty + \|\theta_1^{(2)} - \theta_2^{(2)}\|_\infty\} \quad (159)$$

for $M_0 \equiv \inf_{(\tau,k)\in\mathcal{B}} \max\{(\omega_L(\tau, k))^{-1}, (\omega_U(\tau, k))^{-1}\}$. Equipping $L^\infty(\mathcal{B})\times L^\infty(\mathcal{B})$ with the norm $\|\cdot\|_\infty+\|\cdot\|_\infty$, it follows that T is Lipschitz with Lipschitz constant M_0. In addition, note that:

$$Z_0 = T(K(\sqrt{n}(\tilde{s}_0 - \hat{s}_0))) \qquad Z \stackrel{L}{=} T(K(J)) , \quad (160)$$

where $\stackrel{L}{=}$ stands for "equal in law", $J(\tau, b, x)$ is the Gaussian process in (137) and for the second result we have used the continuous mapping theorem. Also define the process:

$$L_n(\tau, b, x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(W_i - 1)(1\{Y_i \leq s_0(\tau, b, x), D_i = 1, X_i = x\} + b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\})}{P(X_i = x)p(x)f_{y|1,x}(s_0(\tau, b, x))}$$

For $w(x)$ as defined in (133) and $C_0 \equiv 2\sum_{x\in\mathcal{X}} |w(x)|$, it follows from linearity of K and (138), that K is Lipschitz with Lipschitz constant C_0. Therefore, for any $h \in BL_1(\mathbf{R})$, result (159) implies that $h \circ T \circ K \in BL_{C_0M_0}(L^\infty(\mathcal{B} \times \mathcal{X}))$. Thus, by (160) it follows that:

$$\sup_{h\in BL_1(\mathbf{R})} |E[h(Z_0)|\mathcal{Z}_n] - E[h(T(K(L_n)))|\mathcal{Z}_n]|$$
$$\leq \sup_{h\in BL_{C_0M_0}(L^\infty(\mathcal{S}\times\mathcal{X}))} |E[h(\sqrt{n}(\tilde{s}_0 - \hat{s}_0)|\mathcal{Z}_n] - E[h(L_n)|\mathcal{Z}_n]| = o_p(1) , \quad (161)$$

where the final result follows from (152), (153) and arguing as in (157). Exploiting (160) again:

$$\sup_{h\in BL_1(\mathbf{R})} |E[h(T(K(L_n)))|\mathcal{Z}_n] - E[h(Z)]| \leq \sup_{h\in BL_{C_0M_0}(L^\infty(\mathcal{S}\times\mathcal{X}))} |E[h(L_n)|\mathcal{Z}_n] - E[h(J)]|$$
$$= C_0M_0 \times \sup_{h\in BL_1(L^\infty(\mathcal{S}\times\mathcal{X}))} |E[h(L_n)|\mathcal{Z}_n] - E[h(J)]| = o_p(1) , \quad (162)$$

where the final result holds by $J(\tau, b, x)$ being the limit in law of the right hand side of (152) and Theorem 2.9.6 in van der Vaart and Wellner (1996). Hence, (157), (161) and (162) establish (151).

Next, we aim to exploit (151) to show that for all $t \in \mathbf{R}$ that are continuity points of the cdf of Z:

$$|P(\tilde{Z} \leq t|\mathcal{Z}_n) - P(Z \leq t)| = o_p(1) \ . \tag{163}$$

Towards this end, for every $\lambda > 0$, and t a continuity point of the cdf of Z define the functions:

$$h^U_{\lambda,t}(x) = 1 - 1\{x > t\}\min\{\lambda(x-t), 1\} \qquad h^L_{\lambda,t}(x) = 1\{x < t\}\min\{\lambda(t-x), 1\} \ . \tag{164}$$

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

$$0 \leq E[h^U_{\lambda,t}(Z) - h^L_{\lambda,t}(Z)] \leq P(t - \lambda^{-1} \leq Z \leq t + \lambda^{-1}) \ . \tag{165}$$

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:

$$\begin{aligned} |P(\tilde{Z} \leq t|\mathcal{Z}_n) - P(Z \leq t)| &\leq |E[h^L_{\lambda,t}(\tilde{Z})|\mathcal{Z}_n] - E[h^U_{\lambda,t}(Z)]| + |E[h^U_{\lambda,t}(\tilde{Z})|\mathcal{Z}_n] - E[h^L_{\lambda,t}(Z)]| \\ &\leq 2 \sup_{h \in BL_\lambda(\mathbf{R})} |E[h(\tilde{Z})|\mathcal{Z}_n] - E[h(Z)]| + 2P(t - \lambda^{-1} \leq Z \leq t + \lambda^{-1}) \\ &= 2\lambda \sup_{h \in BL_1(\mathbf{R})} |E[h(\tilde{Z})|\mathcal{Z}_n] - E[h(Z)]| + 2P(t - \lambda^{-1} \leq Z \leq t + \lambda^{-1}) \ . \end{aligned} \tag{166}$$

If t is a continuity point of the cdf of Z, then (163) follows from (151) and (166).

To conclude the proof, note that Z is continuously distributed with strictly increasing cdf as a result of Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998). Therefore, for every $\epsilon > 0$ we must have:

$$P(Z \leq r_{1-\alpha} - \epsilon) < 1 - \alpha < P(Z \leq r_{1-\alpha} + \epsilon) \ . \tag{167}$$

Define the event $A_n \equiv \{P(\tilde{Z} \leq r_{1-\alpha} - \epsilon|\mathcal{Z}_n) < 1 - \alpha < P(\tilde{Z} \leq r_{1-\alpha} + \epsilon|\mathcal{Z}_n)\}$ and notice that

$$P(|\tilde{r}_{1-\alpha} - r_{1-\alpha}| \leq \epsilon) \geq P(A_n) \to 1 \ , \tag{168}$$

where the inequality follows by definition of $\tilde{r}_{1-\alpha}$ and the second result is implied by (163) and (167). ■

Proof of Corollary 4.3: In order to establish the Corollary, first define the random variable:

$$Z_n \equiv \sup_{(\tau,k) \in \mathcal{B}} \max\left\{\frac{\sqrt{n}(\hat{\pi}_L(\tau,k) - \pi_L(\tau,k))}{\omega_L(\tau,k)}, \frac{\sqrt{n}(\pi_U(\tau,k) - \hat{\pi}_U(\tau,k))}{\omega_U(\tau,k)}\right\} \tag{169}$$

and notice that since $\sqrt{n}(\hat{\pi}_L - \pi_L, \hat{\pi}_U - \pi_U)$ is asymptotically tight by Theorem 4.1, Assumption 4.2 implies:

$$\sup_{(\tau,k) \in \mathcal{B}} \max\left\{\frac{\sqrt{n}(\hat{\pi}_L(\tau,k) - \pi_L(\tau,k))}{\hat{\omega}_L(\tau,k)}, \frac{\sqrt{n}(\pi_U(\tau,k) - \hat{\pi}_U(\tau,k))}{\hat{\omega}_U(\tau,k)}\right\} = Z_n + o_p(1) \ , \tag{170}$$

by arguing as in (156). For T as defined in (158) and $\hat{\pi} - \pi = (\hat{\pi}_L - \pi_L, \hat{\pi}_U - \pi_U)$ we then obtain:

$$Z_n = T(\sqrt{n}(\hat{\pi} - \pi)) \xrightarrow{L} Z \tag{171}$$

by continuity of T, as shown in (159), Theorem 4.1 and the continuous mapping theorem. Hence,

$$\begin{aligned}
\liminf_{n\to\infty} & P(g \in \hat{\mathcal{G}}(\hat{\omega}, \tilde{r}_{1-\alpha})) \\
& \geq \liminf_{n\to\infty} P(\mathcal{G} \subseteq \hat{\mathcal{G}}(\hat{\omega}, \tilde{r}_{1-\alpha})) \\
& = \liminf_{n\to\infty} P\Big(\sup_{(\tau,k)\in\mathcal{B}} \max\Big\{ \frac{\sqrt{n}(\hat{\pi}_L(\tau,k) - \pi_L(\tau,k))}{\hat{\omega}_L(\tau,k)}, \frac{\sqrt{n}(\pi_U(\tau,k) - \hat{\pi}_U(\tau,k))}{\hat{\omega}_U(\tau,k)} \Big\} \leq \tilde{r}_{1-\alpha} \Big) \\
& \geq 1-\alpha \qquad (172)
\end{aligned}$$

where the first inequality hold for any $g \in \mathcal{G}$ and the second is a result of (170), (171), Theorem 4.2, Slutsky's and the Portmanteau Theorem and Z being continuously distributed as a result of Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998). ■

Lemma 6.7. *Suppose Assumption 2.1(ii)-(iii) and 4.1(ii) hold. Then for the weight function* $\omega(y,x) \equiv 1 - P(D_i = 0|Y_i = y, X_i = x)/(1-p(x))$*, it follows that:*

$$KS(F_{y|x}, F_{y|0,x}) = \sup_{c\in\mathbf{R}} \Big| \int_{-\infty}^{c} w(y,x) dF(y|X_i = x) \Big| \ .$$

Proof: Applying Baye's Rule and the definition of $w(y,x)$ we immediately obtain that:

$$\begin{aligned}
P(Y \leq c|X_i = x) &- P(Y \leq c|X_i = x, D = 0) \\
&= \int_{-\infty}^{c} f(y|X_i = x)dy - \int_{-\infty}^{c} f(y|X_i = x, D = 0)dy \\
&= \int_{-\infty}^{c} f(y|X_i = x)dy - \int_{-\infty}^{c} \frac{P(D = 0|Y = y, X_i = x)}{P(D = 0|X_i = x)} f(y|X_i = x)dy \\
&= \int_{-\infty}^{c} w(y,x) f(y|X_i = x)dy \ . \qquad (173)
\end{aligned}$$

Therefore, the claim of the Lemma follows by taking the supremum over c of the absolute value of (173).■

Lemma 6.8. *If Assumptions 2.1(ii)-(iii), 4.1(ii) hold,* $\{Y_i, X_i\}_{i=1}^n$ *be i.i.d.,* $\sup_y |\hat{w}(y,x_0) - w(y,x_0)| = o_p(1)$ *and* $\sup_{c\in\mathbf{R}} \inf_{c'\in C_n} |F_{y|x_0}(c) - F_{y|x_0}(c')| = o(1)$*, then* $\widehat{KS}(F_{y|x_0}, F_{y|0,x_0}) \xrightarrow{p} KS(F_{y|x_0}, F_{y|0,x_0})$.

Proof: In order to establish the Lemma, we first notice that since by assumption $\sup_y |w(y,x_0) - \hat{w}(y,x_0)| = o_p(1)$ and $n_{x_0}^{-1} \sum_i 1\{Y_i \leq c, X_i = x_0\} \leq 1$ we have that:

$$\begin{aligned}
\Big| \max_{c\in C_n} \Big| \frac{1}{n_{x_0}} \sum_{i=1}^{n} 1\{Y_i \leq c, X_i = x_0\} \hat{w}(Y_i, x_0) \Big| &- \max_{c\in C_n} \Big| \frac{1}{n_{x_0}} \sum_{i=1}^{n} 1\{Y_i \leq c, X_i = x_0\} w(Y_i, x_0) \Big| \Big| \\
&\leq \sup_{y\in\mathbf{R}} |\hat{w}(y,x_0) - w(y,x_0)| \times \sup_{c\in\mathbf{R}} \frac{1}{n_{x_0}} \sum_{i=1}^{n} 1\{Y_i \leq c, X_i = x_0\} = o_p(1) \ . \quad (174)
\end{aligned}$$

Similarly, observe that since $|w(y,x_0)| \leq 1 + (1-p(x_0))^{-1}$, using $n^{-1} \sum_i 1\{Y_i \leq c, X_i = x_0\} \leq 1$, we obtain:

$$\begin{aligned}
\Big| \max_{c\in C_n} \Big| \frac{1}{n_{x_0}} \sum_{i=1}^{n} 1\{Y_i \leq c, X_i = x_0\} w(Y_i, x_0) \Big| &- \max_{c\in C_n} \Big| \frac{1}{nP(X_i = x_0)} \sum_{i=1}^{n} 1\{Y_i \leq c, X_i = x_0\} w(Y_i, x_0) \Big| \Big| \\
&\leq \Big| \frac{n}{n_{x_0}} - \frac{1}{P(X_i = x_0)} \Big| \times \sup_{c\in\mathbf{R}} \frac{1}{n} \sum_{i=1}^{n} 1\{Y_i \leq c, X_i = x_0\} |w(Y_i, x_0)| = o_p(1) \ . \quad (175)
\end{aligned}$$

Let $\mathcal{M} = \{m_c : \mathbf{R}^{1+l} \to \mathbf{R} : m_c(y,x) = 1\{y \leq c, x = x_0\}w(y,x_0) - E[1\{y \leq c, x = x_0\}w(y,x_0)], c \in \mathbf{R}\}$. For $M > (1+(1-p(x_0))^{-1})^2$, and any $1 > \epsilon > 0$, under Assumption 2.1(ii), there exists an increasing sequence $\{y_0, \ldots, y_{\lceil \frac{16M}{\epsilon} \rceil}\}$ such that the intervals $\{[y_{j-1}, y_j]\}_{j=1}^{\lceil \frac{16M}{\epsilon} \rceil}$ partition $\mathbf{R}$ and in addition for all $1 \leq j \leq \lceil \frac{16M}{\epsilon} \rceil$ we have $F_{y|x_0}(y_j) - F_{y|x_0}(y_{j-1}) \leq \frac{\epsilon}{16M}$. Define the functions:

$$
\begin{aligned}
l_j(y,x) &= 1\{y \leq y_{j-1}, x = x_0\}w(y,x_0) - 1\{y_{j-1} \leq y \leq y_j, x = x_0\}|w(y,x_0)| \\
u_j(y,x) &= 1\{y \leq y_{j-1}, x = x_0\}w(y,x_0) + 1\{y_{j-1} \leq y \leq y_j, x = x_0\}|w(y,x_0)| \ , \qquad (176)
\end{aligned}
$$

and $\tilde{l}_j(y,x) = l_j(y,x) - E[u_j(Y_i,X_i)]$ as well as $\tilde{u}_j(y,x) = u_j(y,x) - E[l_j(Y_i,X_i)]$. Further observe that the brackets $\{[\tilde{l}_j, \tilde{u}_j]\}_{j=1}^{\lceil \frac{16M}{\epsilon} \rceil}$ partition $\mathcal{M}$ since for every $y_{j-1} \leq c \leq y_j$ we have $\tilde{l}_j(y,x) \leq m_c(y,x) \leq \tilde{u}_j(y,x)$. Arguing as in (69), it can be shown that $E[(\tilde{l}_j(Y_i,X_i) - \tilde{u}_j(Y_i,X_i))^2] \leq \epsilon$. Hence, $N_{[\,]}(\epsilon, \mathcal{M}, \|\cdot\|_{L^2}) \leq \epsilon^2/16M$. Theorem 2.5.6 in van der Vaart and Wellner (1996) then implies the class is Donsker, which establishes:

$$
\begin{aligned}
&\Big| \max_{c \in C_n} \frac{1}{nP(X_i = x_0)} \sum_{i=1}^n 1\{Y_i \leq c, X_i = x_0\}w(y,x_0)\Big| - \max_{c \in C_n} \Big|E[1\{Y_i \leq c, X_i = x_0\}w(Y_i,x_0)|X_i = x_0]\Big|\Big| \\
&\leq \frac{1}{P(X_i = x_0)} \sup_{c \in \mathbf{R}} \Big|\frac{1}{n}\sum_{i=1}^n 1\{Y_i \leq c, X_i = x_0\}w(Y_i,x_0) - E[1\{Y_i \leq c, X_i = x_0\}w(Y_i,x_0)]\Big| = o_p(1) \ . \qquad (177)
\end{aligned}
$$

Employing $|w(y,x_0)| \leq 1 + (1-p(x_0))^{-1}$, we then obtain by $\sup_{c \in \mathbf{R}} \inf_{c' \in C_n} |F_{y|x_0}(c) - F_{y|x_0}(c')| = o(1)$:

$$
\begin{aligned}
&\Big| \max_{c \in C_n} \Big|E[1\{Y_i \leq c, X_i = x_0\}w(Y,x_0)]\Big| - \sup_{c \in \mathbf{R}} \Big|E[1\{Y \leq c, X_i = x_0\}w(Y,x_0)]\Big|\Big| \\
&\qquad \leq (1 + (1-p(x_0)))^{-1} \times \sup_{c \in \mathbf{R}} \inf_{c' \in C_n} |P(Y_i \leq c, X_i = x_0) - P(Y_i \leq c', X_i = x_0)| = o(1) \ . \qquad (178)
\end{aligned}
$$

The claim of the Lemma is then established by (174), (175), (177) and (178). ■

Proof of Lemma 5.1: Since $\hat{\gamma} \xrightarrow{p} \gamma_0$ and $\mathcal{W}$ is finite, it follows that with probability tending to one $\hat{k} = \exp(W_x'\hat{\gamma})$ for some $W_x \in \mathcal{W}_0$. In addition, notice that since $W_x'\gamma_0 = \tilde{W}_x'\gamma_0$ for all $W_x, \tilde{W}_x \in \mathcal{W}_0$,

$$
\begin{aligned}
&P(W_x\hat{\gamma} = \tilde{W}_x\hat{\gamma}, \text{ for some } W_x, \tilde{W}_x \in \mathcal{W}_0) \\
&\qquad\qquad = P(\sqrt{|\mathcal{X}|}(W_x - \tilde{W}_x)'(\hat{\gamma} - \gamma_0) = 0, \text{ for some } W_x, \tilde{W}_x \in \mathcal{W}_0) = o(1) \ , \qquad (179)
\end{aligned}
$$

where the final result follows by $\mathcal{W}$ being finite, $\sqrt{|\mathcal{X}|}(\hat{\gamma} - \gamma_0) \xrightarrow{L} N(0,\Sigma)$ with Σ positive definite and the Portmanteau Lemma. Therefore, exploiting (179), $\exp(W_x'\gamma_0) = \exp(\tilde{W}_x'\gamma_0)$ for all $W_x, \tilde{W}_x \in \mathcal{W}_0$ and applying the delta method we are able to conclude that:

$$
\begin{aligned}
\sqrt{|\mathcal{X}|}(\hat{k} - k) &= \sqrt{|\mathcal{X}|}\{\max_{W_x \in \mathcal{W}_0} \exp(W_x'\hat{\gamma}) - k\} + o_p(1) \\
&= \sqrt{|\mathcal{X}|} \sum_{W_x \in \mathcal{W}_0} 1\{W_x'\hat{\gamma} \geq \tilde{W}_x'\hat{\gamma}, \ \forall \tilde{W}_x \in \mathcal{W}_0\}(\exp(W_x'\hat{\gamma}) - \exp(W_x'\gamma_0)) + o_p(1) \\
&= \sqrt{|\mathcal{X}|} \sum_{W_x \in \mathcal{W}_0} 1\{W_x'(\hat{\gamma} - \gamma_0) \geq \tilde{W}_x'(\hat{\gamma} - \gamma_0), \ \forall \tilde{W}_x \in \mathcal{W}_0\} \exp(W_x'\gamma_0) W_x'(\hat{\gamma} - \gamma_0) + o_p(1) \\
&= k \times \max_{W_x \in \mathcal{W}_0} W_x'\sqrt{|\mathcal{X}|}(\hat{\gamma} - \gamma) + o_p(1) \ . \qquad (180)
\end{aligned}
$$

The Lemma then follows by $\sqrt{|\mathcal{X}|}(\hat{\gamma} - \gamma_0) \xrightarrow{L} N(0,\Sigma)$ and the continuous mapping theorem. ■

References

Andrews, D. W. K., and G. Soares (2007): "Inference for Parameters defined by Moment Inequalities using Generalized Moment Selection," Working paper, Yale University.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74(2), 539–563.

Autor, D. H., and L. F. Katz (1999): "Changes in the Wage Structure and Earning Inequality," in *Handbook of Labor Economics*, vol. 3A, pp. 1463–1555. North-Holland.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90(2), 300–323.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75(2), 323–363.

Chamberlain, G. (1994): "Quantile Regression, Censoring and the Structure of Wages," in *Advances in Econometrics, Sixth World Congress*, ed. by C. Sims, pp. 171–209. Elsevier.

Chen, X., and D. Pouzo (2008): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Moments," Working paper, Yale University.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Finite Sample Inference for Quantile Regression Models," *Journal of Econometrics*, 152, 93–103.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and confidence regions for parameter sets in econometric models," *Econometrica*, 75(5), 1243–1284.

David, M., R. J. A. Little, M. E. Samuhel, and R. K. Triest (1986): "Alternative Methods for CPS Income Imputation," *Journal of the American Statistical Association*, 81(393), 29–41.

Davydov, Y. A., M. A. Lifshits, and N. V. Smorodina (1998): *Local Properties of Distribuions of Stochastic Functionals.* American Mathematical Society, Providence.

DiNardo, J., J. McCrary, and L. Sanbonmatsu (2006): "Constructive Proposals for Dealing with Attrition: An Empirical Example," Working paper, University of Michigan.

Greenless, J. S., W. S. Reece, and K. D. Zieschang (1982): "Imputations of Missing Values when the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association*, 77(378), 251–261.

GROVERS, R. M., AND M. P. COUPER (1998): *Nonresponse in Household Interview Surveys*. John Wiley and Sons, New York.

HECKMAN, J. J., L. LOCHNER, AND P. TODD (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of Education Economics*, ed. by E. Hanushek, and F. Welch, vol. 1, chap. 7. Elsevier.

HIRSCH, B. T., AND E. J. SCHUMACHER (2004): "Match Bias in Wage Gap Estimates Due to Earnings Imputation," *Journal of Labor Economics*, 22(3), 689–722.

HODERLEIN, S., AND J. STOYE (2009): "Revealed Preferences in a Heterogeneous Population," Working paper, Brown University.

HOROWITZ, J. L., AND C. F. MANSKI (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63(2), 281–302.

——— (2006): "Identification and Estimation of Statistical Functionals Using Incomplete Data," *Journal of Econometrics*, 132, 445–459.

IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," 72, 1845–1857.

JUHN, C., K. M. MURPHY, AND B. PIERCE (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: Estiamting Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76(3), 1071–1102.

LEMIEUX, T. (2006a): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data or Rising Demand for Skill?," *The American Economic Review*, 96(3), 461–198.

——— (2006b): "Postsecondary Education and Increasing Wage Inequality," *The American Economic Review*, 96(2), 195–199.

LILLARD, L., J. P. SMITH, AND F. WELCH (1986): "What do We Really Know about Wages? The Importance of Nonreporting and Census Imputation," *The Journal of Political Economy*, 94(3), 489–506.

MA, S., AND M. R. KOSOROK (2005): "Robust Semiparametric M-estimation and the Weighted Bootstrap," *Journal of Multivariate Analysis*, 96, 190–217.

MANSKI, C. F. (1994): "The Selection Problem," in *Advances in Econometrics, Sixth World Congress*, ed. by C. Sims, pp. 143–170. Cambridge University Press, Cambridge, UK.

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

Ponomareva, M., and E. Tamer (2009): "Misspecification in Moment Inequality Models: Back to Moment Equalities?," Working paper, Northwestern University.

Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference – Special Issue in Honor of Ted Anderson*.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, (forthcoming).

Rubin, D. B. (1976): "Inference and Missing Data," *Biometrika*, (3), 581–592.

Santos, A. (2007a): "Inference in Nonparametric Instrumental Variables with Partial Identification," Working paper, University of California at San Diego.

——— (2007b): "Instrumental Variable Methods for Recovering Continuous Linear Functionals," Working paper, University of California at San Diego.

Stoye, J. (2007): "Bounds on Generalized Linear Predictors with Incomplete Outcome Data," *Reliable Computing*, 13, 293–302.

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

White, H. (1980): "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21, 149–170.

——— (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

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

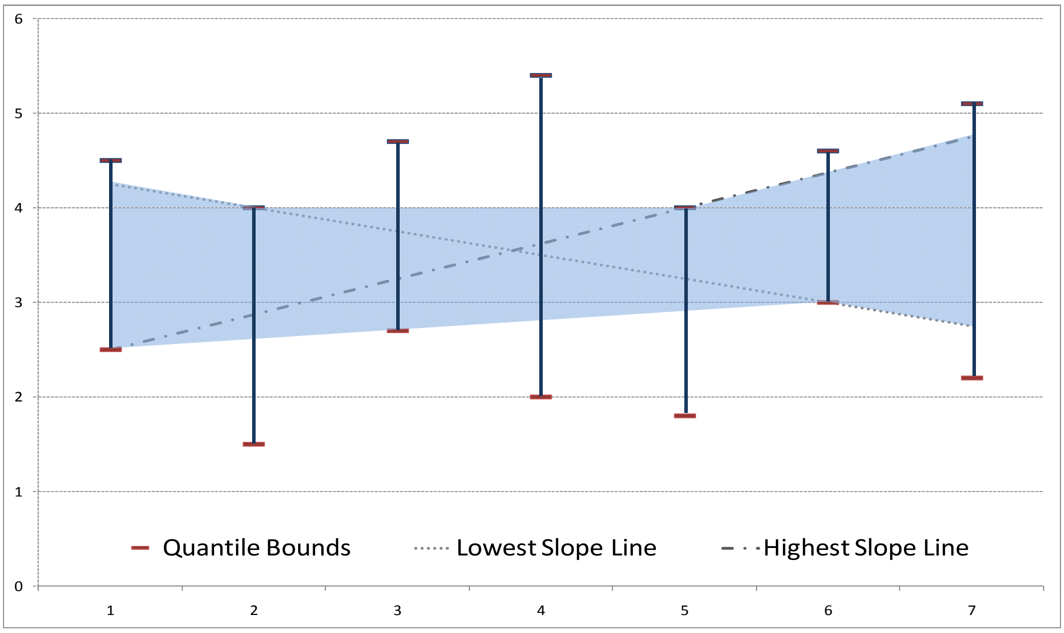


Figure 2: Conditional Quantile and its Pseudo-True Approximation

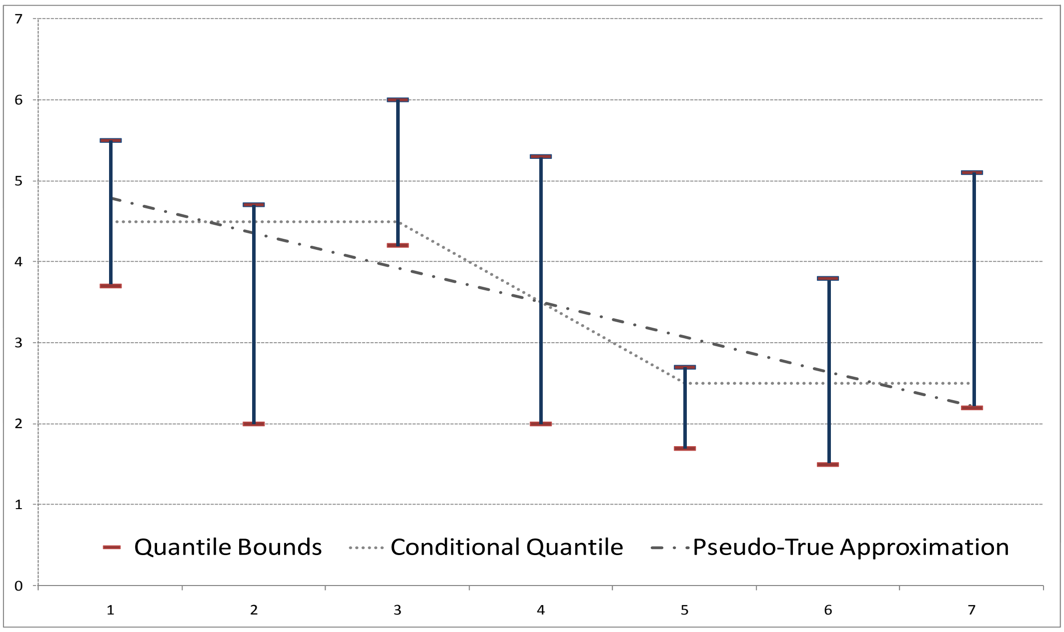
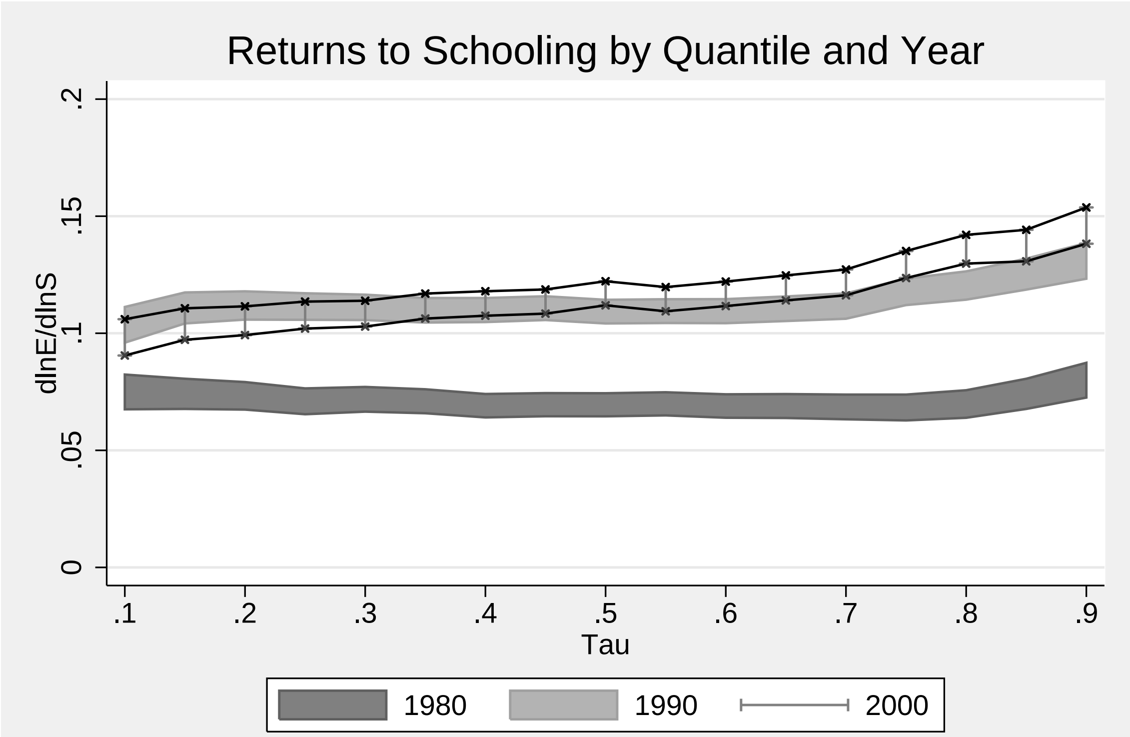
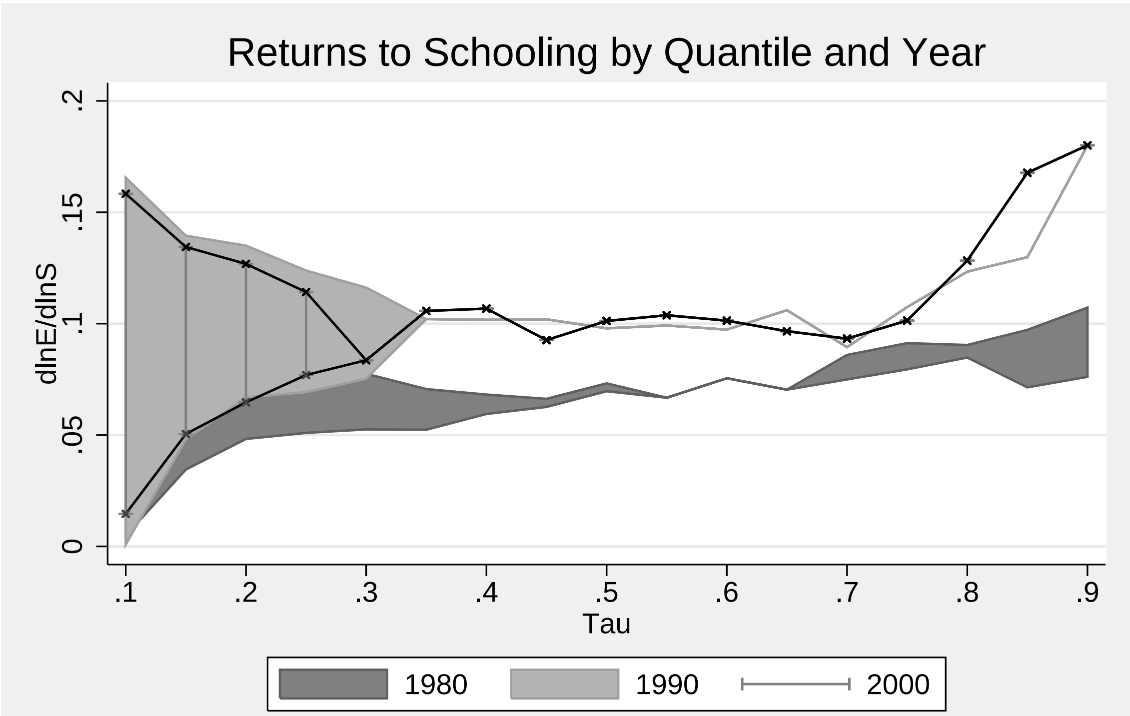


Figure 3: Quantile Regression Estimates Assuming Missing at Random



Note: Returns computed under missing at random assumption. Pseudo true projection weighted by empirical frequency of demographic groups. Shaded regions provide 95% uniform confidence bands.

Figure 4: Finite Sample Confidence Interval for Linear Specification



Note: Returns computed under KS band of 0.10. Shaded regions provide pointwise 95% confidence intervals generated via inversion of finite sample likelihood ratio test statistic. See Appendix B for details. Empty intervals constitute rejection of model at given quantile.

Figure 5: Pseudo True Confidence Intervals

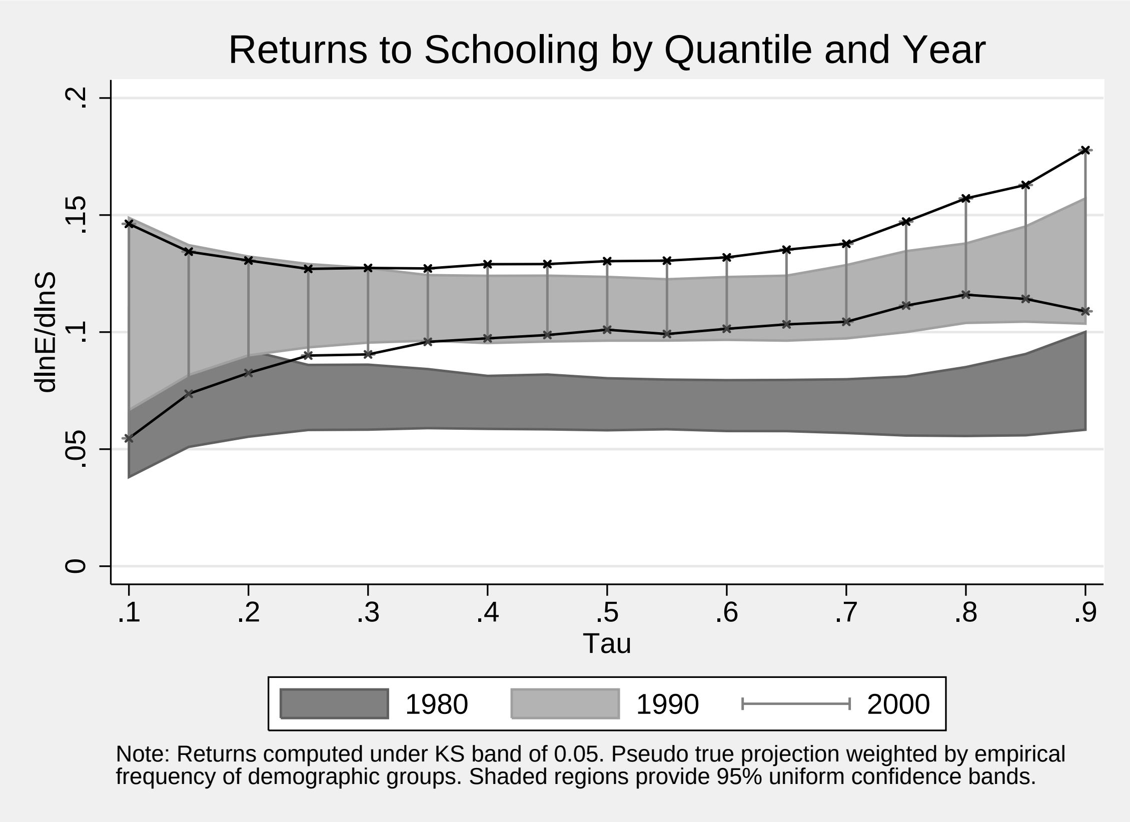


Figure 6: Pseudo True Confidence Intervals, Sensitivity Analysis (1980 vs. 1990)

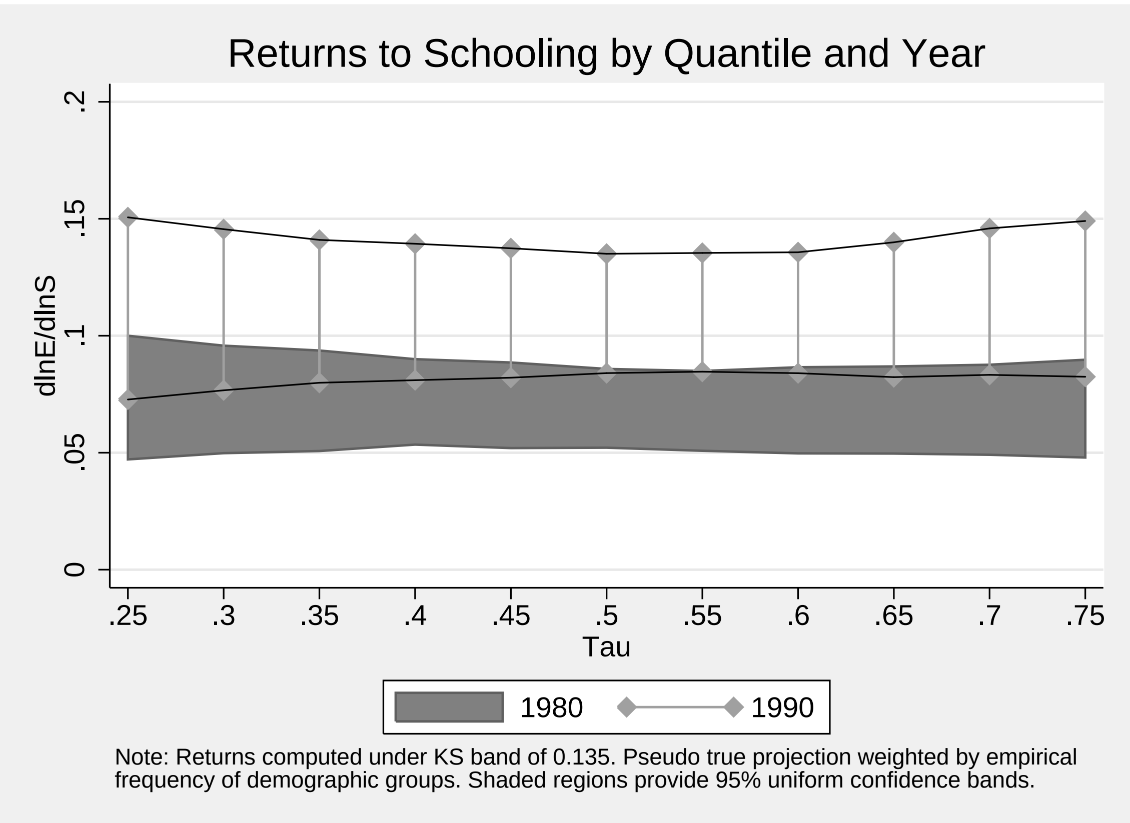
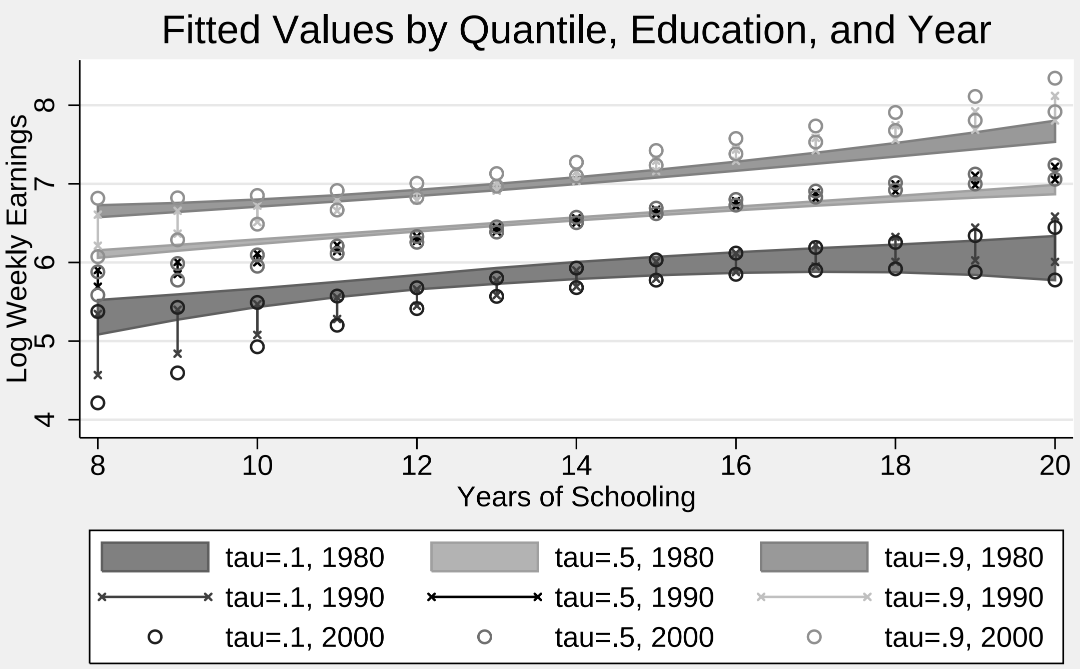


Figure 7: Pseudo True Confidence Intervals, Fitted Values



Note: Fitted values computed under KS bound of 0.05. Pseudo true projection weighted by empirical frequency of demographic groups. Earnings quantiles modeled using specification in Lemieux (2006). Covariates other than education set to sample mean. Pointwise 95% confidence regions in shaded areas and intervals.