Uncovering Hidden Harmony: Latent Binary Quantile Regression and an Application

Yingyao Hu [a], Zhongjian Lin [b], and Ning Neil Yu [c]

[a] *Department of Economics, Johns Hopkins University*
[b] *John Munro Godfrey, Sr. Department of Economics, University of Georgia*
[c] *Institute for Social and Economic Research, Nanjing Audit University*

April 17, 2024

ABSTRACT

This paper develops a method of latent binary quantile regression for settings in which the binary regressand is unobserved and proxied by multiple indicators. We demonstrate how to identify and estimate parameters for conditional quantiles of the hidden outcome, prove the strong consistency of the estimator, and run Monte Carlo experiments to verify its finite-sample performance. Our empirical application attempts to uncover factors affecting the harmony levels within college dormitory rooms. Among other findings, we discover that sleeping schedule discordance damages relationship. Both our approach and findings are applicable for management research and practice.

Keywords: Latent Binary Quantile Regression; Latent Maximum Score Estimator; Measurement Error; Misclassification; Small Group Harmony

I. INTRODUCTION

Entire literatures attempt to operationalize the measurement of vague but meaningful concepts: health, utility, psychological traits, corporate social responsibility, and so on (Bridgman, 1927; Feest, 2005). Inherent vagueness aside, another difficulty lies in the systematic bias of available measurements. For example, a management scientist studying the harmony of small teams may survey team members, but they are likely to underreport disharmony.

This paper proposes a method to study such outcome variables when at least 2 available binary measurements satisfy assumptions to be detailed below. In particular, we generalize *binary quantile regression* and the associated *maximum score estimator* of Manski (1975, 1985) to allow for the measurement error of the binary regressand.

A binary quantile regression model (see Kordas, 2006; Horowitz, 2009; Sherman, 2010, for more formal introductions) works with a latent continuous regressand, and an observed binary regressand which takes the value of 1 or 0 depending on whether the continuous regressand is positive or not: e.g., the difference in utility levels of two options and the associated dichotomous decision. By assumption, observations are independent and identically distributed (assumed throughout the paper), and a quantile of the continuous regressand conditional on regressors is a linear function of them.[1] The classical maximum score estimator consistently recovers the parameters of the function, without making distributional assumptions – in particular, allowing for unknown patterns of heteroskedasticity and nonexistence of error moments.

The estimator and its extensions have found success in empirical analysis of residential mobility (Bartik, Butler, and Liu, 1992), credit card ownership (Bult, 1993), work-trip mode (Horowitz, 1993b), food purchase (Blackburn, Harrison, and Rutström, 1994), brand choice (Briesch, Chintagunta, and Matzkin, 2002), wireless service preferences (Bajari, Fox, and Ryan, 2008), athlete-team matching (Yang, Shi, and Goldfarb, 2009), faculty-office matching (Baccara et al., 2012), FCC spectrum auction (Fox and Bajari, 2013), university-industry collaboration (Mindruta, 2013; Banal-Estañol, Macho-Stadler, and Pérez-Castrillo, 2018), online advertiser-publisher matching (Wu, 2015), bank mergers

[1] When the continuous regressand is observed, this is the standard quantile regression model of Koenker and Bassett (1978). See Koenker and Hallock (2001) for an excellent review.

(Akkus, Cookson, and Hortacsu, 2016), company research alliances (Mindruta, Moeen, and Agarwal, 2016), executive-firm matching (Pan, 2017), sourcing (Fox, 2018), lending relationships (Schwert, 2018), and physician collaboration network (Linde, 2019), among others. All of them involve choice-based binary regressands, and the latent continuous regressands are decision utility.

Our *latent binary quantile regression* allows the binary regressand to be latent. It takes the value of 1 or 0 depending on whether the continuous regressand exceeds a threshold (not necessarily 0). There are at least 2 binary measurements each of which takes the value of 1 whenever the binary regressand is 1, but otherwise may fail to takes the value of 0. In other words, we allow for misclassification but only in one direction. How restrictive this assumption is depends on the application and measurements available, and Remark 2 details reasons why it may hold in many research settings including our application.

Under regularity assumption, we demonstrate how to achieve a closed-form identification of the conditional distribution of the latent binary regressand, and how the parameters of the linear conditional quantile function are *identified up to scale* (Lewbel, 2019). Accordingly, we develop a two-step estimation procedure: nonparametric kernel-based estimation of the conditional distribution of the latent binary regressand, followed by a step that seeks to find a maximizer of the *latent score function*, analogue of the score function used as the criterion function for the maximum score estimator. We call such a maximizer the *latent maximum score estimator*, and prove its strong consistency.

We also show how to utilize more than 2 measurements. The age of big data presents ample opportunities to apply our method: different measurements can come from administrative data, psychological experiments, survey, etc. Monte Carlo experiments showcase how misclassification would cause systematic bias if we were to use the traditional maximum score estimator, and that the latent maximum score estimator addresses the issue and performs as expected.

Our approach is further applied to study the factors determining the harmony level of small groups, an application of interest to management scientists.[2] We surveyed graduating seniors of a Chinese university,[3] where most students stayed in the same

[2]The current editorial statement of the Organizations Department of the journal states that it "welcomes submissions relevant to the internal operations and design of firms and other organizations," and in particular, "those that examine the dynamics of groups and teams."

[3]We agreed not to publicize the name of the university as a precondition for the data collection. Lin,

dormitory rooms for four years, and asked each student whether his/her room had been harmonious for the past four years. The answers serve as multiple measures.

Accordingly, we run latent binary quantile regression to analyze the impact of gender, ethnicity, having siblings, poverty, education in liberal arts, difference in past wake-up time, and room size, on conditional quantiles of group harmony level across a spectrum of quantiles. Among other findings, it is revealed that greater variance in past wake-up time consistently predicts lower group harmony. To the best of our knowledge, the current paper is the first regression analysis of the determinants of group harmony.[4]

Our approach builds upon an influential literature on measurement error, which is too vast to review (an incomplete list of summaries includes Fuller, 1987; Wansbeek and Meijer, 2000; Bound, Brown, and Mathiowetz, 2001; Hausman, 2001; Hyslop and Imbens, 2001; Carroll et al., 2006; Chen, Hong, and Nekipelov, 2011; Schennach, 2016; Hu, 2017). Measurement error in discrete variables are called misclassification and treated by various authors: e.g., Bollinger (1996); Hausman, Abrevaya, and Scott-Morton (1998); Hsiao and Sun (1998); Lewbel (2000); Li, Trivedi, and Guo (2003); Cameron et al. (2004); Lewbel (2007); Molinari (2008); Meyer and Mittag (2017); Ura (2018); Yanagi (2019); Ura (2021). We borrow many insights from that literature, as well as the tradition of taking advantage of multiple measurements to achieve identification (e.g., Hausman et al., 1991; Li and Vuong, 1998; Li, 2002; Li and Hsiao, 2004; Schennach, 2004, 2007; Hu, 2008; Hu and Schennach, 2008; Hu and Shum, 2012; Feng and Hu, 2013; Bonhomme, Jochmans, and Robin, 2016; Gillen, Snowberg, and Yariv, 2019). It should also be noted that the seminal method of Hausman et al. (2021) tackles the problem of classical measurement error in (observed) regressands of quantile regressions (Koenker and Bassett, 1978).

The rest of the paper unfolds as follows. Section II introduces the latent binary quantile regression model. Section III explains an identification strategy. Section IV introduces the latent maximum score estimator and proves its strong consistency. Section V explains how to utilize more than 2 measurements. Section VI conduct Monte Carlo experiments to examine the finite sample performance of our estimator. Section VII describes the

---

Tang, and Yu (2020) study the same population of students.

[4]Such dearth is surprising to us. Social network researchers study enemy links, e.g., via the social balance theory (Heider, 1958), and, in general, how various factors influence social interaction (Jackson, 2008; Goyal, 2012; Borgatti, Everett, and Johnson, 2018; Jackson, 2019).

application of our method to the study of small group harmony. Section VIII concludes and discusses future research.

II. The Model

In the *latent binary quantile regression model*, Y° is a latent continuous outcome variable, e.g., the harmony level of a small group in our empirical exercise. Two observable variables (Y^1, Y^2) are binary *measurements*, e.g., survey answers to the question of whether a group is harmonious. Section V treats cases of more than 2 measurements. Probabilistically, they tend to take the value of 1 when Y° is high, and otherwise tend to take the value of 0. A $K \times 1$ vector X summarizes explanatory variables or *regressors*.

Assumption 1 rules out *severe underreporting*: there is a sufficiently large constant $C \in \mathbb{R}$ such that when Y° is above C, neither measurement takes the value of 0.

Assumption 1. *There exists* $C \in \mathbb{R}$ *such that* $\mathbb{P}(Y^1 = 0 \mid Y^\circ \geq C) = \mathbb{P}(Y^2 = 0 \mid Y^\circ \geq C) = 0$.

Instead of Assumption 1, our method works equally well with a symmetric "*no severe overreporting assumption*," which states that there exists a sufficiently small constant $C' \in \mathbb{R}$ such that when Y° is below C', neither measurement takes the value of 1.

Assumption 1′. *There exists* $C' \in \mathbb{R}$ *such that* $\mathbb{P}(Y^1 = 1 \mid Y^\circ \leq C') = \mathbb{P}(Y^2 = 1 \mid Y^\circ \leq C') = 0$.

Remark 1. *The rest of the paper assumes Assumption 1, not Assumption 1′. When the latter holds instead, we can define* $Y^\dagger := -Y^\circ$, $Y^3 := 1 - Y^1$, *and* $Y^4 := 1 - Y^2$, *and recognize that Assumption 1 now holds with* $Y^\dagger$, Y^3, Y^4, *and* $-C'$ *taking the place of* Y°, Y^1, Y^2, *and* C *respectively. The analysis can thus proceed with these new variables. Our approach fails when, with respect to any* $C \in \mathbb{R}$, *one measurement can be overreported and the other can be underreported.*

Remark 2. *It must be emphasized that only one of the two assumptions above needs to hold for our method to apply. There are several important reasons why one might fail but not the other. First, for instance, when a measurement is self-reported, "social desirability biases" may be unidirectional (Bound, Brown, and Mathiowetz, 2001): e.g., there was no clear incentive for a subject in a sufficiently harmonious group to report disharmony, while the opposite is not true due to a general disinclination to report disharmony. Second, incomplete record or imperfect memory may also affect a measurement in one direction but not the other: e.g., a sufficiently harmonious* *group should not have an administrative record of severe conflicts, but "absence of evidence does not mean evidence of absence;" in a study of health, a sufficiently healthy person should not be diagnosed with a severe illness, but the absence of such a diagnosis is not the same as good health. Indeed, the Introduction lists past applications of binary quantile regression where regressands are binary decisions, and it might as well happen that such decisions are recorded inaccurately in a way that satisfies our assumptions.*[5]

Define $Y^{**} := Y^{\circ} - C$. Because C is a constant, Y^{**} is an equally valid outcome variable of interest, serving as our *regressand*. Our model with respect to a fixed quantile $\tau \in (0,1)$ has

$$Q_\tau(Y^{**} \mid X) \;=\; X'\beta_\tau,$$

where β_τ is our primal *estimands*, a fixed $K \times 1$ vector in a parameter space $B_\tau \subset \mathbb{R}^K$; and $Q_\tau(Y^{**} \mid X)$ denotes the τ-quantile of Y^{**} conditional on X.

Remark 3. *Defining $\varepsilon_\tau := Y^{**} - Q_\tau(Y^{**} \mid X)$, we obtain a linear regression form as $Y^{**} = X'\beta_\tau + \varepsilon_\tau$, where the conditional τ-quantile of the error term is $Q_\tau(\varepsilon_\tau \mid X) = 0$. So conditional heteroskedasticity is allowed. Also, β_τ is allowed to vary across quantiles. This is a standard "quantile regression model" (Koenker and Bassett, 1978; Koenker and Hallock, 2001) except that Y^{**} is unobserved.*

Let $\mathbf{1}\{\cdot\}$ be an *indicator function* which takes the value of 1 or 0 depending on whether a statement is true or false, and define

$$Y^{*} := \mathbf{1}\{Y^{\circ} \geq C\} = \mathbf{1}\{Y^{**} \geq 0\}.$$

So Y^* equals 1 if $Y^{**} \geq 0$ and 0 otherwise. We call Y^* a *latent binary regressand*.

Remark 4. *Suppose Y^* is not latent but observed. This becomes a classical "binary quantile regression model" (Horowitz, 2009, Chapter 4). Manski (1975, 1985)*[6] *applies the analogy principle to define the classical "score function" on $\mathbb{R}^K$:*

$$S^0_\tau(b) = \frac{1}{n}\sum_{i=1}^{n}(Y^*_i + \tau - 1)\cdot \mathbf{1}\{X_i^T b \geq 0\}. \tag{1}$$

[5] For example, for work-trip mode (Horowitz, 1993b), a self-report might underreport automobile due to social desirability biases, and an administrative record of parking might also underreport because a worker can park elsewhere.

[6] He works with the case of $\alpha = 0.5$. See Kordas (2006) for the general case.

A classical "maximum score estimator" $\hat{\beta}_\tau^0$ of β_τ maximizes $S_\tau^0(\cdot)$:

$$\hat{\beta}_\tau^0 \in \hat{B}_\tau^0 := \underset{b \in B_\tau:\ |b_1|=1}{\arg\max}\ S_\tau^0(b),$$

where the restriction $|b_1| = 1$ (or an alternative restriction such as fixing the Euclidean norm $\|b\| = 1$) is necessary because β_τ is only identified up to scale. Manski (1985) proves its "strong consistency" under appropriate regularity conditions, i.e.,

$$\sup_{\hat{\beta}_\tau^0 \in \hat{B}_\tau^0} \left\| \hat{\beta}_\tau^0 - \beta_\tau^0 \right\| \overset{\text{a.s.}}{\to} 0.$$

However, neither Y^{**} nor Y^* is observed. The dataset contains n independent draws, $(Y_i^1, Y_i^2, X_i)_{i=1,\ldots,n}$, from the distribution of (Y^1, Y^2, X). Note that we abuse notation: subscripts i and k for X respectively index different observations and regressors.

The next section shows how to identify β_τ.

III. The Identification

Given the model above, we first identify the conditional *probability function* of the latent binary regressand, $\mathbb{P}(Y^* = y \mid X)$, where $y \in \{0, 1\}$.

Denote the support of X by $\operatorname{supp}(X)$. Assumption 2 states that conditional on $Y^\circ < C$ and X taking a value in $\operatorname{supp}(X)$, the probability of both measurements being 0 is nonnegative. This is almost a trivial requirement for Y^1 and Y^2 to be relevant measurements.

Assumption 2. $\mathbb{P}(Y^1 = 0, Y^2 = 0 \mid Y^* = 0, X = x) > 0$ *for all* $x \in \operatorname{supp}(X)$.

We impose a standard condition on the multiple measurements (e.g., see Li and Vuong, 1998; Li, 2002; Mahajan, 2006; Hu, 2008; Hu and Schennach, 2008; Arellano and Bonhomme, 2012; Schennach, 2016; Bonhomme, Jochmans, and Robin, 2017; Hu, 2017; Bonhomme, Lamadon, and Manresa, 2019): they are mutually independent conditional on the latent binary regressand and regressors.

Assumption 3. $Y^1 \perp Y^2 \mid (Y^*, X)$.

Remark 5. *This nontrivial assumption is well accepted in the literature. In our case, there is an extra justification. If we choose any $C' > C$ to take the role of C, Assumptions 1 and 2 are still* *valid. So Assumption 3 only needs to hold for one such C' and $Y^* := \mathbf{1}\{Y^\circ \geq C'\}$ for the latent binary quantile regression framework to apply.*

Assumptions 1 to 3 support a closed-form identification of the conditional distribution of the latent binary regressand.

Proposition 1. *Given Assumptions 1 to 3, there is a closed-form identification of $\mathbb{P}(Y^* = 0 \mid X = x)$ for all $x \in \operatorname{supp}(X)$:*

$$\mathbb{P}(Y^* = 0 \mid X = x) = \frac{\mathbb{P}(Y^1 = 0 \mid X = x) \cdot \mathbb{P}(Y^2 = 0 \mid X = x)}{\mathbb{P}(Y^1 = 0, Y^2 = 0 \mid X = x)}. \tag{2}$$

Proposition 1 allows us to treat $\mathbb{P}(Y^* = 0 \mid X)$ and

$$\mathbb{E}(Y^* \mid X) = \mathbb{P}(Y^* = 1 \mid X) = 1 - \mathbb{P}(Y^* = 0 \mid X)$$

as observables.

We assume an extra rank condition standard for maximum score methods (Horowitz, 2009). Let $X_{-1} := (X_2, \ldots, X_K)$, and similarly for any vector.

Assumption 4. *The dimension of $\operatorname{supp}(X)$ is K. Further, X_1 is continuous, and for almost every x_{-1} in the projection of $\operatorname{supp}(X)$ to the $(K-1)$-dimensional space for X_{-1}, the distribution of X_1 conditional on $X_{-1} = x_{-1}$ has an everywhere positive density.*

Our main identification result states that under assumptions above, β_τ is identified up to scale: with the absolute value of the first component of β_τ normalized to 1, β_τ is exactly identified. This is the case for all maximum score methods we are aware of (Manski, 1985, 1988; Horowitz, 1992, 2009).

Proposition 2. *Given Assumptions 1 to 4 and the normalization $|\beta_{\tau 1}| = 1$, β_τ is identified.*

The next section defines the latent maximum score estimator for β_τ and proves its strong consistency.

IV. The Estimation

The identification above naturally leads to an estimation procedure with two major steps corresponding to the two propositions.

The first major step follows Proposition 1 to estimate the conditional expectation of the latent binary regressand $\mathbb{E}(Y^* \mid X)$ in a nonparametric way. Because it follows standard kernel estimation procedures, we relegate details to Appendix A. The proposed estimator has a strong uniform consistency property under regularity conditions.

Lemma 1. *Given Assumptions 1 to 3 and standard regularity conditions, we have*

$$\sup_{x \in \mathrm{supp}(X)} \left| \hat{\mathbb{E}}(Y^* \mid X = x) - \mathbb{E}(Y^* \mid X = x) \right| \xrightarrow{\text{a.s.}} 0.$$

The second major step follows Proposition 2 to obtain a practical estimator for β_τ that depends on $(Y_i^1, Y_i^2, X_i)_{i=1,\dots,n}$ and not Y_i^*s.

We modify the classical score function in Equation (1) to define the *latent score function* $S_\tau : \mathbb{R}^K \to \mathbb{R}$ such that for all $b \in \mathbb{R}^K$,

$$S_\tau(b) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\mathbb{E}}(Y_i^* \mid X = X_i) + \tau - 1 \right) \cdot \mathbf{1}\{X_i^T b \geq 0\}. \tag{3}$$

In other words, we replace Y_i^* in the classical score function with its estimated conditional expectation.

Accordingly, a *latent maximum score estimator* is defined as

$$\hat{\beta}_\tau \in \hat{B}_\tau \coloneqq \operatorname*{arg\,max}_{b \in B_\tau:\ |b_1|=1} S_\tau(b).$$

Our main theorem concludes that our latent maximum score estimator is strongly consistent for the estimands.

Theorem 1. *Given the conditions of Lemma 1, Assumption 4, the normalization* $|\beta_{\tau 1}| = 1$*, and a compact set* $B \subset \mathbb{R}^K$ *with* $B_\tau \subset B$*, we have*

$$\sup_{\hat{\beta}_\tau \in \hat{B}_\tau} \left\| \hat{\beta}_\tau - \beta_\tau \right\| \xrightarrow{\text{a.s.}} 0.$$

For a class of estimators including the maximum score estimators, Kim and Pollard (1990) establish that, due to the lack of smoothness of the criterion functions, their convergence rates are $\sqrt[3]{n}$, slower than the usual root-n convergence (see Chamberlain, 1986, for an impossibility result for faster rates); and that the limiting distributions are nonstandard. The standard bootstrap inference is invalid for such estimators (see Manski and Thompson, 1986; Abrevaya and Huang, 2005; Léger and MacGibbon, 2006, for more details), but subsampling (Politis and Romano, 1994; Politis, Romano, and Wolf, 1999) is shown to be consistent by Delgado, Rodrıguez-Poo, and Wolf (2001). The logic applies to our latent maximum score estimators (the readers can refer to the papers cited above for details), so this paper adopts the subsampling approach to obtain the confidence intervals of our latent maximum score estimator in the empirical exercises.

It must be emphasized that there are a few attractive alternatives for inference. For instance, Lee and Pun (2006) and Lee and Yang (2020) demonstrate the validity of using the m-out-of-n bootstrap (Bickel, Götze, and van Zwet, 1997). Dümbgen (1993) and Hong and Li (2020) respectively propose the rescaled bootstrap and the numerical bootstrap. Patra, Seijo, and Sen (2018) provide a model-based smoothed bootstrap procedure. Cattaneo, Jansson, and Nagasawa (2020) achieve consistency of bootstrap-based inference by altering the shape of the criterion functions defining a large class of estimators.

It should be mentioned that, according to Manski (1985, Corollary 1), the maximum score estimator based on a single measurement Y^1 is consistent when misclassification is "random," i.e., when $Y^1 = Y^*$ with probability $p > 0$ and $Y^1 = 1$ or -1 with the same probability $0.5(1 - p) \geq 0$.[7] If this pattern of randomness is justifiable in an empirical setting, there is no need to deploy our method.

Remark 6. *An alternative latent score function appears to utilize more information for estimation: for every observation i with* $\min\{Y_i^1, Y_i^2\} = 0$*, we know* $Y_i^* = 0$ *by Assumption 1, so it seems reasonable to replace* $\hat{\mathbb{E}}(Y_i^* \mid X = X_i)$ *with* $Y_i^* = 0$ *in Equation (3) for such observations. But Appendix C shows why this intuitively appealing approach may not be valid.*

So far we assume the availability of two measurements, but there are plenty of settings where more than two are available. The next section introduces one way to apply the latent binary quantile regression method to those settings.

V. An Extensions to Cases of More Than 2 Measurements

When additional measurements are available, we may utilize them to improve the efficiency of our estimation. For example, assume that two binary variables (Y^3, Y^4) are extra measurements of the latent binary variable Y^*. They satisfy the assumptions

[7] In his model, $Y^* = -1$ is used in place of $Y^* = 0$.

analogous to those for (Y^1, Y^2), i.e., Assumptions 1-3 holds with superscripts 1 and 2 replaced by 3 and 4 respectively. Possibly, (Y^3, Y^4) are available only for a subpopulation with support $\text{supp}^{3,4}(X)$.[8]

Analogously, we may use (Y^3, Y^4) to estimate the conditional expectation of the latent binary regressand, $\hat{\mathbb{E}}^{3,4}(Y^* \mid X = x)$ for any $x \in \text{supp}^{3,4}(X)$. Denote the original estimate generated from (Y^1, Y^2) by $\hat{\mathbb{E}}^{1,2}(Y^* \mid X = x)$. For any $x \in \text{supp}^{3,4}(X)$, we estimate $\hat{\mathbb{E}}(Y^* \mid X = X_i)$ using the average of $\hat{\mathbb{E}}^{1,2}(Y^* \mid X = X_i)$ and $\hat{\mathbb{E}}^{3,4}(Y^* \mid X = X_i)$:

$$\hat{\mathbb{E}}(Y^* \mid X = x) = \begin{cases} \frac{1}{2}\hat{\mathbb{E}}^{1,2}(Y^* \mid X = x) + \frac{1}{2}\hat{\mathbb{E}}^{3,4}(Y^* \mid X = x) & \text{if } x \in \text{supp}^{3,4}(X); \\ \hat{\mathbb{E}}^{1,2}(Y^* \mid X = x) & \text{if } x \in \text{supp}(X) \setminus \text{supp}^{3,4}(X). \end{cases} \tag{4}$$

Analogous to Lemma 1, given the assumptions, we have

$$\sup_{x \in \text{supp}^{3,4}(X)} \left| \hat{\mathbb{E}}^{3,4}(Y^* \mid X = x) - \mathbb{E}(Y^* \mid X = x) \right| \xrightarrow{\text{a.s.}} 0,$$

and thus

$$\sup_{x \in \text{supp}(X)} \left| \hat{\mathbb{E}}(Y^* \mid X = x) - \mathbb{E}(Y^* \mid X = x) \right| \xrightarrow{\text{a.s.}} 0.$$

Theorem 1 still holds for a latent maximum score estimator using the new $\hat{\mathbb{E}}(Y^* \mid X = x)$ from Equation (4). The finite sample performance of the new estimator shall improve thanks to an more accurate estimate of $\mathbb{E}(Y^* \mid X = X_i)$ for those observations associated with (Y_i^3, Y_i^4).

Similarly, extra pairs of measurements such as (Y^5, Y^6), if available and satisfying analogous assumptions, can be utilized to improve the finite sample performance of our latent maximum score estimator in the same way: we only need to update $\hat{\mathbb{E}}(Y^* \mid X = x)$ by taking the average of $\hat{\mathbb{E}}^{1,2}(Y^* \mid X = x)$, $\hat{\mathbb{E}}^{3,4}(Y^* \mid X = x)$, and $\hat{\mathbb{E}}^{5,6}(Y^* \mid X = x)$.

The next section follows the aforementioned procedure to analyze idealized datasets. For our empirical exercises, the first major estimation step relied on the R package "np" (Nonparametric Kernel Methods for Mixed Datatypes), and the second step relied on Gurobi for Matlab in conducting mixed-integer programming. An off-the-shelf personal computer carried out all tasks.

[8] For instance, in our empirical exercise, room size is an explanatory variable, so $\text{supp}^{3,4}(X)$ only covers rooms with at least 4 students.

VI. Monte Carlo Experiments

We run Monte Carlo experiments to assess the finite sample performance of latent binary quantile regression (LBQR), and compare the results with binary quantile regression (BQR) which simply uses each measurement as the regressand.

The latent continuous and binary regressands come from the specification

$$\begin{aligned} Y^{**} &= \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon; \\ Y^{*} &= \mathbf{1}\{Y^{**} \geq 0\}. \end{aligned}$$

Here, the regressors X_1, X_2, and X_3 respectively follow the standard normal distribution, the Bernoulli distribution on $\{-1, 1\}$ with equal probabilities, and the uniform distribution on $[-1, 1]$. The error term ε follows the normal distribution with mean 0 and standard deviation $\frac{1}{1+\|X\|^2}$ conditional on X, so heteroskedasticity is built in. Let $\beta_1 = \beta_2 = \beta_3 = 1$. We have

$$Q_{0.5}(Y^{**} \mid X) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

so this is a binary median regression setup, with the normalization $\beta_1 = 1$ and the estimands being (β_2, β_3).

Two measurements, Y^1 and Y^2, are independently distributed conditional on (Y^*, X) (Assumption 3) such that $\mathbb{P}(Y^1 = 0 \mid Y^* = 1, X) = \mathbb{P}(Y^2 = 0 \mid Y^* = 1, X) = 0$ (Assumption 1), and

$$\mathbb{P}(Y^1 = 1 \mid Y^* = 0, X) = \mathbb{P}(Y^2 = 1 \mid Y^* = 0, X) = \frac{X_2 + X_3}{5} + 0.4.$$

Therefore, potential overreporting probabilities fully cover the interval of $[0, 0.8]$ as well as those estimated probabilities in our empirical application (Section VII).

Given the data generating process above, our experiment adopts a $3 \times 3 \times 3$ factorial design. The first factor is the sample size: the three levels being 250, 500, and 1000. The second factor is the numbers of measurements: the levels being 2, 4, and 8. For each of these nine settings, we draw 200 random samples. For each sample, we calculate two score functions based on Y^1 and Y^2 respectively, and estimate one latent score function through the aforementioned kernel approach (Section V applies when there are more than 2 measurements); then obtain two maximum score estimators of BQR and one latent

Table 1: Summaries of Monte Carlo Experiments

n	Average Bias																	
	BQR of Y^1						BQR of Y^2						LBQR					
	β_2			β_3			β_2			β_3			β_2			β_3		
$\#Y^i$	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8
250	0.336	0.294	0.308	0.393	0.291	0.349	0.312	0.332	0.298	0.351	0.380	0.334	0.207	0.102	0.057	-0.039	-0.103	-0.030
500	0.270	0.266	0.243	0.309	0.279	0.264	0.259	0.321	0.248	0.246	0.354	0.246	0.162	0.125	0.105	-0.027	-0.086	-0.102
1000	0.236	0.247	0.218	0.227	0.228	0.240	0.237	0.247	0.239	0.247	0.281	0.246	0.132	0.094	0.093	0.012	-0.072	-0.076
	Mean Squared Error																	
	BQR of Y^1						BQR of Y^2						LBQR					
	β_2			β_3			β_2			β_3			β_2			β_3		
$\#Y^i$	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8	2	4	8
250	0.644	0.548	0.320	1.264	0.637	0.676	0.424	0.463	0.293	0.717	0.865	0.560	0.208	0.119	0.268	0.297	0.243	0.699
500	0.192	0.152	0.155	0.404	0.285	0.319	0.137	0.434	0.137	0.226	0.736	0.261	0.094	0.060	0.044	0.131	0.145	0.137
1000	0.102	0.096	0.083	0.171	0.141	0.148	0.094	0.095	0.103	0.169	0.168	0.171	0.049	0.026	0.023	0.086	0.053	0.047

Notes: $\#Y^i$ represents the number of measurements.

maximum score estimator of LBQR via maximizing the three criterion functions subject to the normalization $\beta_1 = 1$. In other words, the third factor is the estimation method.

Table 1 summarizes the performance of BQR and LBQR: for each of the 27 experimental conditions, there are 200 estimates of (β_2, β_3), and it reports the average biases and mean squared errors against the benchmark of $(1, 1)$.

As would be predicted by theory, the average biases of BQR estimates (based on measurements Y^1 or Y^2) exhibit no clear signs of trending towards zero: all of them are larger than 0.2, that is, BQR consistently overestimates the effects of X_2 and X_3. This is reasonable: it is assumed that the overreporting probabilities are strictly increasing functions of two regressors, which gives the false impression that they positively influence the latent binary and continuous regressands to a significantly greater extent than the truth. Our LBQR produces uniformly smaller average biases (in absolute values) than BQR, across all sample sizes and all measurement numbers.

Similarly, LBQR produces smaller mean squared errors than BQR across all sample sizes and all measurement numbers except for one instance (for β_3 in the single case of 250 observations and 8 measurements). The mean squared errors from LBQR also show clear signs of trending towards zero as the sample size increases, though at a rate slower than the conventional root-n convergence; this is expected by the theory. Adding measurements also substantially reduces mean square errors, so Section V appears to be a reasonable approach of utilizing more than 2 measurements.

The next section empirically investigates the influencing factors of the harmony levels of small groups, running LBQR with real-life data.

VII. An Application to Group Harmony

VII.A. The Background and Data Description

Group assignment and its effects have been well researched (e.g., Gale and Shapley, 1962; Sacerdote, 2001; Bhattacharya, 2009; Calvó-Armengol, Patacchini, and Zenou, 2009; Epple and Romano, 2011; Sacerdote, 2011; Lin, Tang, and Yu, 2020, among many). But we have yet to find regression analysis on the determinants of group harmony in the literature. This paper attempts to fill the gap. We first describe the data.

3,982 students enrolled in a Chinese university in 2015. Administrators manually assigned them to 955 dormitory rooms, each designed to house 3, 4, or 6 students. The assignment followed the reformative guideline of randomizing and maximizing the diversity of home province, ethnicity, and major within rooms. Most students stayed put for four years: by the end of June 2019, 3,569 still lived in the same dormitory rooms originally assigned to them.[9]

In June 2019, the Commission of Student Affairs of the university, in collaboration with one of the authors, designed an online survey for all 3,933 students who were still with the university. Administrators distributed survey links and requested all students to finish the survey, and also provided administrative records of students' demographic information, room assignments, and other information.[10] Using the dataset, Lin, Tang, and Yu (2020) study the interdependence of roommates' volunteering decisions.

Measurements of group harmony come from the 10th question of the survey: "Have the interactions among all residents of your first dormitory room been harmonious in the past four years?" The response rate was 96.21%. 951 (out of 955) dormitory rooms have at least 2 measurements and no missing variables. Among them, 824 have at least 4 measurements, and only 36 have 6. Based on Section V, 2 or 4 measurements are randomly drawn out of available candidates: *Harmony* ι for $\iota = 1, 2, 3, 4$. Conflicting measurements appeared in 356 rooms out of 951, i.e., 37.43% rooms have reports of both harmony and disharmony. This stands as compelling evidence that misclassification is a valid concern.

For each room, the regressors we consider include *Female* (whether all residents were female), *Minority* (whether at least one resident was non-Han Chinese), *Sibling* (whether at least one resident had at least one sibling)[11], *Poverty* (whether at least one resident received allowances for students from low-income families), *Arts* (whether at least one resident took the liberal-arts type of the college entrance examination instead of the sciences type), and *Room-Size* (the number of original residents). In addition, *Wake-Up* is the standard

[9]49 left the university (as dropouts, transfer students, or army recruits), 183 decided to live off campus, and only 181 changed rooms. We learned that students needed approvals to change rooms or live off campus, but were discouraged from doing so; this is common practice in China. Thanks to this feature, attrition bias is most likely small in our study.

[10]The AEA RCT Registry number is AEARCTR-0004296. The public URL for the trial is http://www.socialscienceregistry.org/trials/4296.

[11]An alternative is the proportion of residents who had at least one sibling, which exacerbates the curse of dimensionality for nonparametric estimations. Similarly, *Poverty* and *Arts* are dummy variables instead of continuous ones.

deviation of the second answer to Question 4: "In high school years, roughly speaking, on average, at what time did you go to bed? At what time did you wake up?"[12] All the regressors are arguably exogenous because they were determined before students arrived on campus and started to interact with roommates. Table 2 reports summary statistics.

Table 2: Summary Statistics

	Mean	Std. Dev.
Female	0.692	0.462
Minority	0.229	0.421
Sibling	0.763	0.425
Poverty	0.492	0.500
Arts	0.766	0.424
Wake-Up	0.504	0.391
Room-Size	4.105	0.692
Harmony 1	0.869	0.338
Harmony 2	0.843	0.364
Harmony 3	0.850	0.358
Harmony 4	0.877	0.328

VII.B. Regression Analysis

In this application, the latent continuous regressand is the harmony level of a group *Harmony*** and the binary one is the harmony indicator *Harmony**.

Our latent binary quantile regression framework makes three key assumptions beyond regularity conditions. First, the observations (corresponding to rooms) are largely independent and identically distributed. It is not our view that each room is isolated from others; however, we consider the assumption a first-order approximation.[13] Second, Assumption 1 rules out underreporting. As Section II shows, we effectively assume that there exists a sufficiently high cutoff C that students in more harmonious rooms

[12]We allowed them to report decimal numbers – e.g., 22.5 means 22:30 p.m.

[13]The principle of mixing students in each room inadvertently reduced interaction between roommates outside the room significantly: they usually took different classes and pursued different career goals (the university also offered few elective courses which are open to students from different majors). As a first approximation, we ignore interactions across different rooms within a dorm.

never reported disharmony, which, in our view, is a reasonable approximation to reality. Third, Assumption 3 requires measurements be independent conditional on *Harmony** and regressors. To this end, we used a red color to make it clear in our instruction: "You must answer these questions independently and truthfully as soon as possible. And do not discuss the details with anybody before the survey ends."

We run latent binary quantile regression to study the impacts of aforementioned regressors. Instead of assuming $|\beta_{\tau 1}| = 1$, we apply the normalization of $\|\beta_\tau\| = 1$, which enables us to sensibly compare effects of all regressors across different quantiles. Table 3 reports the estimation results at the 0.25, 0.5, and 0.75 quantiles. Hypothesis testing is based on subsampling.[14]

Table 3: The Estimation Results

Quantile	0.25	0.5	0.75
Female	-0.051	0.025	-0.086
Minority	0.955***	-0.004	-0.691***
Sibling	0.019	0.027	0.143
Poverty	-0.201**	-0.008	-0.066
Arts	-0.020	-0.955***	-0.686***
Wake-Up	-0.122***	-0.103**	-0.074
Room-Size	0.074**	0.142***	0.078
Intercept	0.154	-0.238	0.093***

Notes: ***Significant at the 99% level; **95%; *90%.

The estimates for the effects of *Wake-Up* are consistently negative, and, at the 0.25 and 0.5 quantiles, statistically significant at the 0.05 level. In other words, sleeping schedule discordance (greater variation in wake-up times) predicts less harmony. For colleges, this means putting students with similar sleeping schedules promotes dormitory harmony. For companies, this might indicate the necessity of more research on how to avoid noise and disturbance in office.

The estimates for the effects of *Poverty* and *Arts* are also consistently negative, half of which are significant at the 0.05 level. For organizations striving for diversity, equity, and

[14] We randomly draw 1,000 subsamples with (block) size, $b = n^{4/5}$ (Politis, Romano, and Wolf, 1999). There is no established method for computing p-values for maximum score estimators.

inclusion, one policy implication is the need for targeted team building mechanisms to promote group harmony. The effects of *Room-Size* are consistently positive and mostly significant, so smaller groups are less harmonious (here, group membership is not self-selected). Less inference can be made regarding other regressors.

VIII. CONCLUDING REMARKS

Heraclitus famously said, "the hidden harmony is better than the obvious." To uncover the hidden harmony of small groups, this paper harnesses the obvious, i.e., the reports from group members regarding whether it is harmonious or not.

Building upon the classical binary quantile regression model (Manski, 1985, 1975), we assume that the binary regressand is latent. Given multiple measurements of the binary regressand and a set of assumptions, parameters of latent binary quantile regression are identified up to scale, and the latent maximum score estimator is strongly consistent. We illustrate its finite sample performance in Monte Carlo experiments, and confirm that it effectively corrects the bias caused by misclassification.

We apply the econometric method to examine how different factors affect group harmony. Using reports from college roommates as multiple measurements, we find that across different quantiles, the standard deviation of all roommates' high-school wake-up time is consistently and negatively correlated with higher conditional quantiles of group harmony level, among other findings.

The method we develop is applicable to a wide variety of research topics, as pointed out in the Introduction. For management research, our method is particularly suitable for studying teamwork: many important aspects of teams are hard to measure, but surveys of team members naturally endow researchers with multiple measurements, which may be used to apply our method.

As our first attempt at addressing the latency of the binary regressand for binary quantile regression, this paper uses the original approach of Manski (1975, 1985) as a natural starting point. For example, there are extensions in the spirit of the methods of Manski (1975), Matzkin (1993), Fox (2007), and Ouyang, Yang, and Zhang (2020), for multinomial choices; the methods of Han (1987), Sherman (1993), Abrevaya (2000), and Krief (2014) for more general regression functions; the methods of Manski (1987), Kyriazidou (1997), Honoré and Kyriazidou (2000), and Thomas (2006) for discrete choice panel data; the methods of Manski and Tamer (2002) and Wan and Xu (2014) for interval data and games of incomplete information; the methods of Moon (2004) and De Jong and Woutersen (2011) for time series; the method of Chen (2010) for censored quantile regression; the matching maximum score estimator of Fox (2010, 2018) for transferable utility matching games; the method of Chen, Lee, and Sung (2014) which accommodates decision under uncertainty; and the method of Chen and Lee (2018) for variable selection.

In addition, it is straightforward to generalize our estimator to a *smoothed* latent maximum score estimator (Horowitz, 1992, 1993a,b, 2002; Kordas, 2006; Chen, Gao, and Li, 2018), by replacing the latent score function for calculating the estimator with a smooth approximation. At the cost of making slightly stronger distributional assumptions, we can obtain faster rates of convergence and asymptotic normality, and use standard bootstrap for inference. We can also generalize, for instance, the Bayesian approach of Benoit and Van den Poel (2012), the local non-linear least squares approach of Blevins and Khan (2013), the local polynomial smoothing approach of Chen and Zhang (2015), and the Laplacian approach of Jun, Pinkse, and Wan (2017), all of which improve estimation in important aspects. It is also meaningful to generalize the assumption of binary reports to allow for ordered discrete choice, given that Likert-type scales are popular (Likert, 1932). These are all left for future research.

APPENDIX

A. The First Step of the Estimation Procedure

Appendix A explains how to estimate $\mathbb{E}(Y^* \mid X)$ using a kernel method.

Without loss of generality, assume that all regressors are continuous.[15] Following Rosenblatt (1956) and Li and Racine (2007, Chapter 1), let $\mathcal{K}_0(\cdot)$ be a valid univariate

[15] When regressors are mixed (remember that Assumption 4 dictates the existence of at least one continuous regressor), to estimate joint distributions, Li and Racine (2007) examine the frequency-based approach in Chapter 3 and the smooth approach in Chapter 4. Our paper follows the latter (see Aitchison and Aitken, 1976; Li and Racine, 2004; Racine and Li, 2004; Li and Racine, 2007, 2008, for discussions of the merits of different approaches).

kernel function such as the standard normal kernel with $\mathcal{K}_0(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}$ for any $v \in \mathbb{R}$. We estimate $\mu(x)$, the joint *probability density function* of X valued at any $x \in \mathbb{R}^K$, as

$$\hat{\mu}(x) := \frac{1}{nh_1 \dots h_K} \sum_{i=1}^{n} \mathcal{K}\Big(\frac{X_i - x}{h}\Big),$$

where $h = (h_1, \dots, h_K) > 0$ collects nuisance bandwidth parameters selected through cross-validation, and $\mathcal{K}(\cdot)$ is a "product kernel function" satisfying $\mathcal{K}\Big(\frac{X_i - x}{h}\Big) := \prod_{k=1}^{K} \mathcal{K}_0\Big(\frac{X_{ik} - x_k}{h_k}\Big)$. Silverman (1978) and Giné and Guillou (2002) provide standard regularity conditions[16] for the strong uniform consistency of the kernel estimate, that is, for

$$\sup_{x \in \mathbb{R}^K} \Big|\hat{\mu}(x) - \mu(x)\Big| \overset{\text{a.s.}}{\to} 0,$$

(as $n \to \infty$, where n is suppressed from $\hat{\mu}_n$ and other expressions for simplicity).

Then, based on the smooth approach to nonparametric estimation in the presence of discrete variables, we estimate $f(y^1, y^2, x)$, the joint density of (Y^1, Y^2, X) valued at (y^1, y^2, x), for any $y^1, y^2 \in \{0, 1\}$ and $x \in \mathbb{R}^K$, as

$$\hat{f}(y^1, y^2, x) := \frac{1}{nh'_1 \dots h'_K} \sum_{i=1}^{n} \mathcal{K}\Big(\frac{X_i - x}{h'}\Big) \cdot \lambda_1^{\mathbf{1}\{Y_i^1 \neq y^1\}} \cdot (1 - \lambda_1)^{\mathbf{1}\{Y_i^1 = y^1\}} \cdot \lambda_2^{\mathbf{1}\{Y_i^2 \neq y^2\}} \cdot (1 - \lambda_2)^{\mathbf{1}\{Y_i^2 = y^2\}},$$

where $h' = (h'_1, \dots, h'_K) > 0$ collects bandwidth parameters, and $\lambda_1, \lambda_2 \in (0, 1)$ are also nuisance parameters which approach 0 as n goes to infinity. Similarly, under standard regularity conditions (Li and Ouyang, 2005; Mason and Swanepoel, 2015), we have

$$\sup_{y^1, y^2 \in \{0,1\};\, x \in \mathbb{R}^K} \Big|\hat{f}(y^1, y^2, x) - f(y^1, y^2, x)\Big| \overset{\text{a.s.}}{\to} 0.$$

Since $\mathbb{P}(Y^1 = y^1, Y^2 = y^2 \mid X = x) = \frac{f(y^1, y^2, x)}{\mu(x)}$ for all $x \in \operatorname{supp}(X)$, we accordingly estimate the conditional joint probability function of the two measurements as

$$\hat{\mathbb{P}}(Y^1 = y^1, Y^2 = y^2 \mid X = x) := \frac{\hat{f}(y^1, y^2, x)}{\hat{\mu}(x)},$$

for any $y^1, y^2 \in \{0, 1\}$ and $x \in \operatorname{supp}(X)$ (Hardle, Janssen, and Serfling, 1988; Li and Racine, 2007, Chapter 5). In practice, $\operatorname{supp}(X)$ may be unknown; and we only calculate $\hat{\mathbb{P}}(Y^1 = y^1, Y^2 = y^2 \mid X = X_i)$ for $i = 1, \dots, n$ and any $y^1, y^2 \in \{0, 1\}$, because these are all we need for the latent maximum score estimator. The strong uniform consistency of $\hat{\mu}(x)$ and

[16] The conditions are technical and precisely stated in those papers. To avoid distraction from our main messages, we refer to the cited papers.

$\hat{f}(y^1, y^2, x)$ implies that under standard regularity conditions,

$$\sup_{y^1,y^2\in\{0,1\};\, x\in\mathrm{supp}(X)} \left|\hat{\mathbb{P}}(Y^1 = y^1, Y^2 = y^2 \mid X = x) - \mathbb{P}(Y^1 = y^1, Y^2 = y^2 \mid X = x)\right| \xrightarrow{\text{a.s.}} 0.$$

Note that the conditional joint probability function of the two measurements can also be estimated in a parametric way via standard bivariate Probit or Logit regressions, when the parametric distributional assumptions on the corresponding error terms are justifiable.

Finally, we obtain $\hat{\mathbb{P}}(Y^* = 0 \mid X = x)$, the estimator for the conditional probability of the latent binary regressand Y^* taking the value of 0, based on the closed-form identification in Proposition 1. Specifically, on the right-hand side of Equation (2), the first half and second half of the denominator can be respectively expressed as

$$\begin{aligned}
\mathbb{P}(Y^1 = 0 \mid X = x) &= \mathbb{P}(Y^1 = 0, Y^2 = 1 \mid X = x) + \mathbb{P}(Y^1 = 0, Y^2 = 0 \mid X = x);\\
\mathbb{P}(Y^2 = 0 \mid X = x) &= \mathbb{P}(Y^1 = 1, Y^2 = 0 \mid X = x) + \mathbb{P}(Y^1 = 0, Y^2 = 0 \mid X = x).
\end{aligned}$$

So the formula can be expressed in terms of $\mathbb{P}(Y^1 = y^1, Y^2 = y^2 \mid X = x)$ with $y^1, y^2 \in \{0, 1\}$. To obtain $\hat{\mathbb{P}}(Y^* = 0 \mid X = x)$, we only need to replace such $\mathbb{P}(Y^1 = y^1, Y^2 = y^2 \mid X = x)$ in the formula with $\hat{\mathbb{P}}(Y^1 = y^1, Y^2 = y^2 \mid X = x)$. Then the estimator for the conditional expectation of the latent binary regressand is $\hat{\mathbb{E}}(Y^* \mid X = x) := 1 - \hat{\mathbb{P}}(Y^* = 0 \mid X = x)$.

B. Main Proofs

Proof of Proposition 1. According to our model setup, two measurements (Y^1, Y^2) and all regressors X are observable. So we know that the joint conditional probability function, $\mathbb{P}(Y^1 = y^1, Y^2 = y^2 \mid X = x)$, where $y^1, y^2 \in \{0, 1\}$ and $x \in \mathrm{supp}(X)$, is identified from the observables. For simplicity, we suppress "conditional on $X = x$ for all $x \in \mathrm{supp}(X)$" in what follows.

We have for all $y^1, y^2 \in \{0, 1\}$,

$$\begin{aligned}
\mathbb{P}(Y^1 = y^1, Y^2 = y^2) &= \sum_{y\in\{0,1\}} \mathbb{P}(Y^1 = y^1, Y^2 = y^2 \mid Y^* = y) \cdot \mathbb{P}(Y^* = y)\\
&= \sum_{y\in\{0,1\}} \mathbb{P}(Y^1 = y^1 \mid Y^* = y) \cdot \mathbb{P}(Y^2 = y^2 \mid Y^* = y) \cdot \mathbb{P}(Y^* = y),
\end{aligned}$$

where the first equality follows from the law of total probability and the second from

Assumption 3. Assumption 1 tells us that $\mathbb{P}(Y^j = 0 \mid Y^* = 1) = 0$ for all $j \in \{0, 1\}$. Hence,

$$\mathbb{P}(Y^1 = 1, Y^2 = 0) = \mathbb{P}(Y^1 = 1 \mid Y^* = 0) \cdot \mathbb{P}(Y^2 = 0 \mid Y^* = 0) \cdot \mathbb{P}(Y^* = 0); \tag{5}$$

$$\mathbb{P}(Y^1 = 0, Y^2 = 1) = \mathbb{P}(Y^1 = 0 \mid Y^* = 0) \cdot \mathbb{P}(Y^2 = 1 \mid Y^* = 0) \cdot \mathbb{P}(Y^* = 0); \tag{6}$$

$$\mathbb{P}(Y^1 = 0, Y^2 = 0) = \mathbb{P}(Y^1 = 0 \mid Y^* = 0) \cdot \mathbb{P}(Y^2 = 0 \mid Y^* = 0) \cdot \mathbb{P}(Y^* = 0). \tag{7}$$

By Bayes' theorem, Assumption 1 further tells us that for all $j \in \{0, 1\}$,

$$\mathbb{P}(Y^* = 1 \mid Y^j = 0) = \frac{\mathbb{P}(Y^j = 0 \mid Y^* = 1) \cdot \mathbb{P}(Y^* = 1)}{\mathbb{P}(Y^j = 0)} = 0. \tag{8}$$

Equation (5) gives us

$$\begin{aligned} \mathbb{P}(Y^1 = 1 \mid Y^* = 0) &= \frac{\mathbb{P}(Y^1 = 1, Y^2 = 0)}{\mathbb{P}(Y^2 = 0 \mid Y^* = 0) \cdot \mathbb{P}(Y^* = 0)} \\ &= \frac{\mathbb{P}(Y^1 = 1, Y^2 = 0)}{\mathbb{P}(Y^* = 0 \mid Y^2 = 0) \cdot \mathbb{P}(Y^2 = 0)} \\ &= \frac{\mathbb{P}(Y^1 = 1, Y^2 = 0)}{\mathbb{P}(Y^2 = 0)}, \end{aligned} \tag{9}$$

where the second equality follows from Bayes' theorem, and the third from Equation (8). Note that Assumption 2 ensures positive denominators. Equation (6) implies an equality symmetric to Equation (9):

$$\mathbb{P}(Y^2 = 1 \mid Y^* = 0) = \frac{\mathbb{P}(Y^1 = 0, Y^2 = 1)}{\mathbb{P}(Y^1 = 0)}. \tag{10}$$

So we can derive from Equation (7) that

$$\begin{aligned} \mathbb{P}(Y^* = 0) &= \frac{\mathbb{P}(Y^1 = 0, Y^2 = 0)}{\mathbb{P}(Y^1 = 0 \mid Y^* = 0) \cdot \mathbb{P}(Y^2 = 0 \mid Y^* = 0)} \\ &= \frac{\mathbb{P}(Y^1 = 0, Y^2 = 0)}{(1 - \mathbb{P}(Y^1 = 1 \mid Y^* = 0)) \cdot (1 - \mathbb{P}(Y^2 = 1 \mid Y^* = 0))} \\ &= \frac{\mathbb{P}(Y^1 = 0, Y^2 = 0) \cdot \mathbb{P}(Y^1 = 0) \cdot \mathbb{P}(Y^2 = 0)}{(\mathbb{P}(Y^2 = 0) - \mathbb{P}(Y^1 = 1, Y^2 = 0)) \cdot (\mathbb{P}(Y^1 = 0) - \mathbb{P}(Y^1 = 0, Y^2 = 1))} \\ &= \frac{\mathbb{P}(Y^2 = 0) \cdot \mathbb{P}(Y^1 = 0)}{\mathbb{P}(Y^1 = 0, Y^2 = 0)}, \end{aligned}$$

where the third equality follows from Equations (9) and (10). This gives us the closed-form identification. $\square$

Proof of Proposition 2. By Proposition 1, $\mathbb{E}(Y^* \mid X) = \mathbb{P}(Y^* = 1 \mid X)$ is identified. According to the model setup,

$$\mathbb{P}(Y^* = 1 \mid X = x) = \mathbb{P}(\varepsilon_\tau \geq -X'\beta_\tau \mid X = x) = 1 - \mathbb{P}(\varepsilon_\tau < -X'\beta_\tau \mid X = x).$$

Since $Q_\tau(\varepsilon_\tau \mid X) = 0$, we further have

$$\mathbb{P}(Y^* = 1 \mid X = x) \gtreqless 1 - \tau \Longleftrightarrow X'\beta_\tau \gtreqless 0.$$

Without loss of generality, let $\beta_{\tau 1} = 1$, and $b := (1, b_{-1}) \neq \beta_\tau$ be a false parameter value vector that also satisfies the normalization, and

$$T_1(b) := \left\{x \in \mathbb{R}^K : x'\beta_\tau < 0 \leq x'b\right\};$$
$$T_2(b) := \left\{x \in \mathbb{R}^K : x'b < 0 \leq x'\beta_\tau\right\}.$$

If $\mathbb{P}(X \in T_1(b)) > 0$, then b is observationally distinguishable from β_τ: we can find a subset of $\operatorname{supp}(X)$ with positive probability such that β_τ entails $\mathbb{P}(Y^* = 1 \mid X = x) < 1 - \tau$ and b entails $\mathbb{P}(Y^* = 1 \mid X = x) \geq 1 - \tau$. Similarly, if $\mathbb{P}(X \in T_2(b)) > 0$, b is observationally distinguishable from β_τ. Thus, β_τ is identified if

$$\mathbb{P}(X \in T_1(b) \cup T_2(b)) > 0,$$

for all possible $b \neq \beta_\tau$.

But for any $b \neq \beta_\tau$, since $b_1 = \beta_{\tau 1} = 1$,

$$T_1(b) = \left\{x \in \mathbb{R} : -x'_{-1}b_{-1} \leq x_1 < -x'_{-1}\beta_{\tau,-1}\right\};$$
$$T_2(b) = \left\{x \in \mathbb{R} : -x'_{-1}\beta_{\tau,-1} \leq x_1 < -x'_{-1}b_{-1}\right\}.$$

By the second part of Assumption 4, $\mathbb{P}(X \in T_1(b)) > 0$ as long as $-x'_{-1}b_{-1} < -x'_{-1}\beta_{\tau,-1}$; and $\mathbb{P}(X \in T_2(b)) > 0$ as long as $-x'_{-1}\beta_{\tau,-1} < -x'_{-1}b_{-1}$. Thus β_τ is identified if

$$\mathbb{P}(X'_{-1}b_{-1} = X'_{-1}\beta_{\tau,-1}) < 1,$$

which holds because the first half of Assumption 4 rules out an exact linear relation among the components of X. $\square$

Proof of Theorem 1. For any $b \in B$, let the *population score function* be defined as the expectation of the classical score function

$$\bar{S}^0_\tau(b) := \mathbb{E}(S^0_\tau(b)) = \mathbb{E}\Big[(Y^* + \tau - 1) \cdot \mathbf{1}\{X^T b \geq 0\}\Big] = \mathbb{E}\Big[(\mathbb{E}(Y^* \mid X) + \tau - 1) \cdot \mathbf{1}\{X^T b \geq 0\}\Big], \quad (11)$$

where the first equality follows from the definition in Equation (1) and the second follows from the law of iterative expectation. We can mechanically follow Manski (1985) to demonstrate that $\bar{S}_\tau^0(\cdot)$ attains its maximum at β_τ given the restriction of $|b_1| = 1$ (his Lemma 3) and that $\bar{S}_\tau^0(\cdot)$ is continuous (his Lemma 5).

Define the *hypothetical latent score function* on B:

$$S_\tau^H(b) = \frac{1}{n}\sum_{i=1}^{n}\Big(\mathbb{E}(Y^* \mid X = X_i) + \tau - 1\Big)\cdot \mathbf{1}\{X_i^T b \geq 0\}.$$

The gap between the latent score function and the hypothetical one converges to 0 almost surely, uniformly over $b \in B$:

$$\begin{aligned}
\sup_{b\in B:\,|b_1|=1}\Big|S_\tau(b) - S_\tau^H(b)\Big| &= \sup_{b\in B:\,|b_1|=1}\left|\frac{1}{n}\sum_{i=1}^{n}\Big(\hat{\mathbb{E}}(Y^* \mid X = X_i) - \mathbb{E}(Y^* \mid X = X_i)\Big)\cdot \mathbf{1}\{X_i^T b \geq 0\}\right| \\
&\leq \sup_{b\in B:\,|b_1|=1}\frac{1}{n}\sum_{i=1}^{n}\Big|\hat{\mathbb{E}}(Y^* \mid X = X_i) - \mathbb{E}(Y^* \mid X = X_i)\Big|\cdot \mathbf{1}\{X_i^T b \geq 0\} \\
&\leq \sup_{b\in B:\,|b_1|=1}\frac{1}{n}\sum_{i=1}^{n}\sup_{x\in\operatorname{supp}(X)}\Big|\hat{\mathbb{E}}(Y^* \mid X = x) - \mathbb{E}(Y^* \mid X = x)\Big|\cdot 1 \\
&= \sup_{x\in\operatorname{supp}(X)}\Big|\hat{\mathbb{E}}(Y^* \mid X = x) - \mathbb{E}(Y^* \mid X = x)\Big| \\
&\overset{\text{a.s.}}{\to} 0,
\end{aligned}$$

where the first line follows from the definitions and the last line follows from Lemma 1.

Now we show that the hypothetical latent score function converges to the population score function almost surely, uniformly over $b \in B$. A family of functions indexed by $b \in \mathbb{R}^K$,

$$\mathcal{F} := \Big\{f_b : \mathbb{R}^K \to \mathbb{R}\,\Big|\, b \in \mathbb{R}^K \text{ and } f_b(x) = x^T b \text{ for all } x \in \mathbb{R}^K\Big\},$$

forms a finite-dimensional vector space, so by Lemma 2.6.15 in Van Der Vaart and Wellner (1996), it is a VC-subgraph class of functions.[17] By their Lemma 2.6.18 (viii) and (vi), another family derived from $\mathcal{F}$,

$$\mathcal{F}' := \Big\{f_b : \mathbb{R}^K \to \mathbb{R}\,\Big|\, b \in \mathbb{R}^K \text{ and } f_b(x) = (\mathbb{E}(Y^* \mid X = x) + \tau - 1)\cdot \mathbf{1}\{x^T b \geq 0\} \text{ for all } x \in \mathbb{R}^K\Big\},$$

is VC-subgraph. By their Theorem 2.6.7 (see the comments below), $\mathcal{F}'$ satisfies the uniform

[17] VC stands for Vapnik-Červonenkis.

entropy condition. Hence, according to their Theorem 2.5.2, $\mathcal{F}'$ is a Donsker class, and thus a Glivenko-Cantelli class (Van Der Vaart and Wellner, 1996, Page 82). We thus have

$$\sup_{b\in B:\,|b_1|=1} |S_\tau^H(b) - \bar{S}_\tau^0(b)| \overset{\text{a.s.}}{\to} 0.$$

Given the two uniform strong convergence results above, we know that the latent score function converges to the population score function almost surely, uniformly over $b \in B$:

$$\sup_{b\in B:\,|b_1|=1} \left|S_\tau(b) - \bar{S}_\tau^0(b)\right| \overset{\text{a.s.}}{\to} 0.$$

Analogous to the proof of strong consistency by Manski (1985), we have established all three major conditions for applying Theorem 2 in Manski (1983). The strong consistency of $\hat{\beta}_\tau$ follows. □

C. An Alternative Latent Score Function

Formally, Remark 6 proposes an *alternative latent score function* $\tilde{S}_\tau : \mathbb{R}^K \to \mathbb{R}$ such that for all $b \in \mathbb{R}^K$,

$$\tilde{S}_\tau(b) = \frac{1}{n}\sum_{i=1}^{n}\Big(\min\{Y_i^1, Y_i^2\} \cdot \hat{\mathbb{E}}(Y_i^* \mid X = X_i) + \tau - 1\Big) \cdot 1\{X_i^T b \geq 0\}.$$

The difference between the values of the latent score function and the alternative one at each $b \in \mathbb{R}^K$ is

$$S_\tau(b) - \tilde{S}_\tau(b) \; = \frac{1}{n}\sum_{i=1}^{n}\Big(1 - \min\{Y_i^1, Y_i^2\}\Big) \cdot \hat{\mathbb{E}}(Y_i^* \mid X = X_i) \cdot \mathbf{1}\{X_i^T b \geq 0\},$$

which in general converges to a strictly positive number, not zero. The proof of the strong consistency of the latent maximum score estimator (Theorem 1) relies on the fact that the latent score function converges to the population score function of Equation (11) almost surely, uniformly over $b \in B$. The alternative latent score function thus cannot converges to the population score function almost surely, uniformly over $b \in B$.

REFERENCES

Abrevaya, Jason. 2000. "Rank estimation of a generalized fixed-effects regression model." *Journal of Econometrics*, 95 (1): 1–23.

Abrevaya, Jason and Jian Huang. 2005. "On the bootstrap of the maximum score estimator." *Econometrica*, 73 (4): 1175–1204.

Aitchison, John and Colin G. G. Aitken. 1976. "Multivariate binary discrimination by the kernel method." *Biometrika*, 63 (3): 413–420.

Akkus, Oktay, J. Anthony Cookson, and Ali Hortacsu. 2016. "The determinants of bank mergers: A revealed preference analysis." *Management Science*, 62 (8): 2241–2258.

Arellano, Manuel and Stéphane Bonhomme. 2012. "Identifying distributional characteristics in random coefficients panel data models." *Review of Economic Studies*, 79 (3): 987–1020.

Baccara, Mariagiovanna, Ayşe İmrohoroğlu, Alistair J. Wilson, and Leeat Yariv. 2012. "A field study on matching with network externalities." *American Economic Review*, 102 (5): 1773–1804.

Bajari, Patrick, Jeremy T. Fox, and Stephen P. Ryan. 2008. "Evaluating wireless carrier consolidation using semiparametric demand estimation." *Quantitative Marketing and Economics*, 6 (4): 299.

Banal-Estañol, Albert, Inés Macho-Stadler, and David Pérez-Castrillo. 2018. "Endogenous matching in university-industry collaboration: Theory and empirical evidence from the United Kingdom." *Management Science*, 64 (4): 1591–1608.

Bartik, Timothy J., John Scott Butler, and Jin-Tan Liu. 1992. "Maximum score estimates of the determinants of residential mobility: Implications for the value of residential attachment and neighborhood amenities." *Journal of Urban Economics*, 32 (2): 233–256.

Benoit, Dries F. and Dirk Van den Poel. 2012. "Binary quantile regression: A Bayesian approach based on the asymmetric Laplace distribution." *Journal of Applied Econometrics*, 27 (7): 1174–1188.

Bhattacharya, Debopam. 2009. "Inferring optimal peer assignment from experimental data." *Journal of the American Statistical Association*, 104 (486): 486–500.

Bickel, Peter J., Friedrich Götze, and Willem R. van Zwet. 1997. "Resampling fewer than n observations: Gains, losses, and remedies for losses." *Statistica Sinica*, 7: 1–31.

Blackburn, McKinley, Glenn W. Harrison, and E. Elisabet Rutström. 1994. "Statistical bias functions and informative hypothetical surveys." *American Journal of Agricultural Economics*, 76 (5): 1084–1088.

Blevins, Jason R. and Shakeeb Khan. 2013. "Local NLLS estimation of semi-parametric binary choice models." *Econometrics Journal*, 16 (2): 135–160.

Bollinger, Christopher R. 1996. "Bounding mean regressions when a binary regressor is mismeasured." *Journal of Econometrics*, 73 (2): 387–399.

Bonhomme, Stéphane, Koen Jochmans, and Jean-Marc Robin. 2016. "Non-parametric estimation of finite mixtures from repeated measurements." *Journal of the Royal Statistical Society Series B.*, 78 (1): 211–229.

———. 2017. "Nonparametric estimation of non-exchangeable latent-variable models." *Journal of Econometrics*, 201 (2): 237–248.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A distributional framework for matched employer employee data." *Econometrica*, 87 (3): 699–739.

Borgatti, Stephen P., Martin G. Everett, and Jeffrey C. Johnson. 2018. *Analyzing Social Networks*. SAGE Publications.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement error in survey data." In *Handbook of Econometrics*, vol. 5. Elsevier, 3705–3843.

Bridgman, Percy W. 1927. *The logic of modern physics*. The logic of modern physics. Macmillan.

Briesch, Richard A., Pradeep K. Chintagunta, and Rosa L. Matzkin. 2002. "Semiparametric estimation of brand choice behavior." *Journal of the American Statistical Association*, 97 (460): 973–982.

Bult, Jan Roelf. 1993. "Semiparametric versus parametric classification models: An application to direct marketing." *Journal of Marketing Research*, 30 (3): 380–390.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2009. "Peer effects and social networks in education." *Review of Economic Studies*, 76 (4): 1239–1267.

Cameron, A. Colin, Tong Li, Pravin K. Trivedi, and David M. Zimmer. 2004. "Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts." *Econometrics Journal*, 7 (2): 566–584.

Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. CRC Press.

Cattaneo, Matias D, Michael Jansson, and Kenichi Nagasawa. 2020. "Bootstrap-Based Inference for Cube Root Asymptotics." *Econometrica*, 88 (5): 2203–2219.

Chamberlain, Gary. 1986. "Asymptotic efficiency in semi-parametric models with censoring." *Journal of Econometrics*, 32 (2): 189–218.

Chen, Le-Yu and Sokbae Lee. 2018. "Best subset binary prediction." *Journal of Econometrics*, 206 (1): 39–56.

Chen, Le-Yu, Sokbae Lee, and Myung Jae Sung. 2014. "Maximum score estimation with nonparametrically generated regressors." *Econometrics Journal*, 17 (3): 271–300.

Chen, Songnian. 2010. "An integrated maximum score estimator for a generalized censored quantile regression model." *Journal of Econometrics*, 155 (1): 90–98.

Chen, Songnian and Hanghui Zhang. 2015. "Binary quantile regression with local polynomial smoothing." *Journal of Econometrics*, 189 (1): 24–40.

Chen, Xiaohong, Han Hong, and Denis Nekipelov. 2011. "Nonlinear models of measurement errors." *Journal of Economic Literature*, 49 (4): 901–937.

Chen, Xirong, Wenzheng Gao, and Zheng Li. 2018. "A data-driven bandwidth selection method for the smoothed maximum score estimator." *Economics Letters*, 170: 24–26.

De Jong, Robert M. and Tiemen Woutersen. 2011. "Dynamic time series binary choice." *Econometric Theory*, 27 (4): 673–702.

Delgado, Miguel A., Juan M. Rodrıguez-Poo, and Michael Wolf. 2001. "Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator." *Economics Letters*, 73 (2): 241–250.

Dümbgen, Lutz. 1993. "On nondifferentiable functions and the bootstrap." *Probability Theory and Related Fields*, 95 (1): 125–140.

Epple, Dennis and Richard E. Romano. 2011. "Peer effects in education: A survey of the theory and evidence." In *Handbook of Social Economics*, vol. 1. Elsevier, 1053–1163.

Feest, Uljana. 2005. "Operationism in psychology: What the debate is about, what the debate should be about." *Journal of the History of the Behavioral Sciences*, 41 (2): 131–149. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jhbs.20079`.

Feng, Shuaizhang and Yingyao Hu. 2013. "Misclassification errors and the underestimation of the US unemployment rate." *American Economic Review*, 103 (2): 1054–70.

Fox, Jeremy T. 2007. "Semiparametric estimation of multinomial discrete-choice models using a subset of choices." *RAND Journal of Economics*, 38 (4): 1002–1019.

———. 2010. "Identification in matching games." *Quantitative Economics*, 1 (2): 203–254.

———. 2018. "Estimating matching games with transfers." *Quantitative Economics*, 9 (1): 1–38.

Fox, Jeremy T. and Patrick Bajari. 2013. "Measuring the efficiency of an FCC spectrum auction." *American Economic Journal: Microeconomics*, 5 (1): 100–146.

Fuller, Wayne A. 1987. *Measurement Error Models*. Wiley Series in Probability and Statistics. Wiley.

Gale, David and Lloyd S Shapley. 1962. "College admissions and the stability of marriage." *American Mathematical Monthly*, 69 (1): 9–15.

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with measurement error: Techniques with applications to the Caltech Cohort Study." *Journal of Political Economy*, 127 (4): 1826–1863.

Giné, Evarist and Armelle Guillou. 2002. "Rates of strong uniform consistency for multivariate kernel density estimators." *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 38 (6): 907–921.

Goyal, Sanjeev. 2012. *Connections: An Introduction to the Economics of Networks*. Princeton University Press.

Han, Aaron K. 1987. "Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator." *Journal of Econometrics*, 35 (2-3): 303–316.

Hardle, W., Paul Janssen, and Robert Serfling. 1988. "Strong uniform consistency rates for estimators of conditional functionals." *Annals of Statistics*, 16 (4): 1428–1449.

Hausman, Jerry, Haoyang Liu, Ye Luo, and Christopher Palmer. 2021. "Errors in the dependent variable of quantile regression models." *Econometrica*, 89 (2): 849–873.

Hausman, Jerry A. 2001. "Mismeasured variables in econometric analysis: Problems from the right and problems from the left." *Journal of Economic Perspectives*, 15 (4): 57–67.

Hausman, Jerry A., Jason Abrevaya, and Fiona M. Scott-Morton. 1998. "Misclassification of the dependent variable in a discrete-response setting." *Journal of Econometrics*, 87 (2): 239–269.

Hausman, Jerry A., Whitney K. Newey, Hidehiko Ichimura, and James L. Powell. 1991. "Identification and estimation of polynomial errors-in-variables models." *Journal of Econometrics*, 50 (3): 273–295.

Heider, Fritz. 1958. *The psychology of interpersonal relations*. Wiley Press.

Hong, Han and Jessie Li. 2020. "The numerical bootstrap." *Annals of Statistics*, 48 (1): 397–412.

Honoré, Bo E. and Ekaterini Kyriazidou. 2000. "Panel data discrete choice models with lagged dependent variables." *Econometrica*, 68 (4): 839–874.

Horowitz, Joel L. 1992. "A smoothed maximum score estimator for the binary response model." *Econometrica*, 60 (3): 505–531.

———. 1993a. "Optimal rates of convergence of parameter estimators in the binary response model with weak distributional assumptions." *Econometric Theory*, 9 (1): 1–18.

———. 1993b. "Semiparametric estimation of a work-trip mode choice model." *Journal of Econometrics*, 58 (1-2): 49–70.

———. 2002. "Bootstrap critical values for tests based on the smoothed maximum score estimator." *Journal of Econometrics*, 111 (2): 141–167.

———. 2009. *Semiparametric and Nonparametric Methods in Econometrics*. Springer.

Hsiao, Cheng and Bao-Hong Sun. 1998. "Modeling survey response bias – with an analysis of the demand for an advanced electronic device." *Journal of Econometrics*, 89 (1): 15–39.

Hu, Yingyao. 2008. "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution." *Journal of Econometrics*, 144 (1): 27–61.

———. 2017. "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics." *Journal of Econometrics*, 200 (2): 154–168.

Hu, Yingyao and Susanne M. Schennach. 2008. "Instrumental variable treatment of nonclassical measurement error models." *Econometrica*, 76 (1): 195–216.

Hu, Yingyao and Matthew Shum. 2012. "Nonparametric identification of dynamic models with unobserved state variables." *Journal of Econometrics*, 171 (1): 32–44.

Hyslop, Dean R. and Guido W. Imbens. 2001. "Bias from classical and other forms of measurement error." *Journal of Business & Economic Statistics*, 19 (4): 475–481.

Jackson, Matthew O. 2008. *Social and Economic Networks*. Princeton University Press.

———. 2019. *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*. Pantheon Books: New York.

Jun, Sung Jae, Joris Pinkse, and Yuanyuan Wan. 2017. "Integrated score estimation." *Econometric Theory*, 33 (6): 1418.

Kim, Jeankyung and David Pollard. 1990. "Cube root asymptotics." *Annals of Statistics*, 18 (1): 191–219.

Koenker, Roger and Gilbert Bassett, Jr. 1978. "Regression quantiles." *Econometrica*, 46 (1): 33–50.

Koenker, Roger and Kevin F. Hallock. 2001. "Quantile regression." *Journal of Economic Perspectives*, 15 (4): 143–156.

Kordas, Gregory. 2006. "Smoothed binary regression quantiles." *Journal of Applied Econometrics*, 21 (3): 387–407.

Krief, Jerome M. 2014. "An integrated kernel-weighted smoothed maximum score estimator for the partially linear binary response model." *Econometric Theory*, 30 (3): 647–675.

Kyriazidou, Ekaterini. 1997. "Estimation of a panel data sample selection model." *Econometrica*, 65 (6): 1335–1364.

Lee, Stephen M. S. and M. C. Pun. 2006. "On *m* out of *n* bootstrapping for nonstandard M-estimation with nuisance parameters." *Journal of the American Statistical Association*, 101 (475): 1185–1197.

Lee, Stephen M. S. and Puyudi Yang. 2020. "Bootstrap confidence regions based on M-estimators under nonstandard conditions." *Annals of Statistics*, 48 (1): 274–299.

Léger, Christian and Brenda MacGibbon. 2006. "On the bootstrap in cube root asymptotics." *Canadian Journal of Statistics*, 34 (1): 29–44.

Lewbel, Arthur. 2000. "Identification of the binary choice model with misclassification." *Econometric Theory*, 16 (4): 603–609.

———. 2007. "Estimation of average treatment effects with misclassification." *Econometrica*, 75 (2): 537–551.

———. 2019. "The identification zoo: Meanings of identification in econometrics." *Journal of Economic Literature*, 57 (4): 835–903.

Li, Qi and Desheng Ouyang. 2005. "Uniform convergence rate of kernel estimation with mixed categorical and continuous data." *Economics Letters*, 86 (2): 291–296.

Li, Qi and Jeffrey S. Racine. 2004. "Cross-validated local linear nonparametric regression." *Statistica Sinica*: 485–512.

———. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

———. 2008. "Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data." *Journal of Business & Economic Statistics*, 26 (4): 423–434.

Li, Tong. 2002. "Robust and consistent estimation of nonlinear errors-in-variables models." *Journal of Econometrics*, 110 (1): 1–26.

Li, Tong and Cheng Hsiao. 2004. "Robust estimation of generalized linear models with measurement errors." *Journal of Econometrics*, 118 (1): 51–65.

Li, Tong, Pravin K. Trivedi, and Jiequn Guo. 2003. "Modeling response bias in count:

A structural approach with an application to the national crime victimization survey data." *Sociological Methods & Research*, 31 (4): 514–544.

Li, Tong and Quang Vuong. 1998. "Nonparametric estimation of the measurement error model using multiple indicators." *Journal of Multivariate Analysis*, 65 (2): 139–165.

Likert, Rensis. 1932. "A technique for the measurement of attitudes." *Archives of Psychology*, 22 (140): 5–55.

Lin, Zhongjian, Xun Tang, and Ning Neil Yu. 2020. "Uncovering heterogeneous social effects in binary choices." *Journal of Econometrics, forthcoming*.

Linde, Sebastian. 2019. "The formation of physician patient sharing networks in Medicare: Exploring the effect of hospital affiliation." *Health Economics*, 28 (12): 1435–1448.

Mahajan, Aprajit. 2006. "Identification and estimation of regression models with misclassification." *Econometrica*, 74 (3): 631–665.

Manski, Charles F. 1975. "Maximum score estimation of the stochastic utility model of choice." *Journal of Econometrics*, 3 (3): 205–228.

———. 1983. "Closest empirical distribution estimation." *Econometrica*, 51 (2): 305–319.

———. 1985. "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator." *Journal of Econometrics*, 27 (3): 313–333.

———. 1987. "Semiparametric analysis of random effects linear models from binary panel data." *Econometrica*, 55 (2): 357–362.

———. 1988. "Identification of binary response models." *Journal of the American Statistical Association*, 83 (403): 729–738.

Manski, Charles F. and Elie Tamer. 2002. "Inference on regressions with interval data on a regressor or outcome." *Econometrica*, 70 (2): 519–546.

Manski, Charles F. and T. Scott Thompson. 1986. "Operational characteristics of maximum score estimation." *Journal of Econometrics*, 32 (1): 85–108.

Mason, David M. and Jan W. H. Swanepoel. 2015. "Uniform in bandwidth consistency of kernel estimators of the density of mixed data." *Electronic Journal of Statistics*, 9 (1): 1518–1539.

Matzkin, Rosa L. 1993. "Nonparametric identification and estimation of polychotomous choice models." *Journal of Econometrics*, 58 (1-2): 137–168.

Meyer, Bruce D. and Nikolas Mittag. 2017. "Misclassification in binary choice models." *Journal of Econometrics*, 200 (2): 295–311.

Mindruta, Denisa. 2013. "Value creation in university-firm research collaborations: A matching approach." *Strategic Management Journal*, 34 (6): 644–665.

Mindruta, Denisa, Mahka Moeen, and Rajshree Agarwal. 2016. "A two-sided matching approach for partner selection and assessing complementarities in partners' attributes in inter-firm alliances." *Strategic Management Journal*, 37 (1): 206–231.

Molinari, Francesca. 2008. "Partial identification of probability distributions with misclassified data." *Journal of Econometrics*, 144 (1): 81–117.

Moon, Hyungsik Roger. 2004. "Maximum score estimation of a nonstationary binary choice model." *Journal of Econometrics*, 122 (2): 385–403.

Ouyang, Fu, Thomas Tao Yang, and Hanghui Zhang. 2020. "Semiparametric identification and estimation of discrete choice models for bundles." *Economics Letters*, 193: 109321.

Pan, Yihui. 2017. "The determinants and impact of executive-firm matches." *Management Science*, 63 (1): 185–200.

Patra, Rohit Kumar, Emilio Seijo, and Bodhisattva Sen. 2018. "A consistent bootstrap procedure for the maximum score estimator." *Journal of Econometrics*, 205 (2): 488–507.

Politis, Dimitris N. and Joseph P. Romano. 1994. "Large sample confidence regions based on subsamples under minimal assumptions." *Annals of Statistics*: 2031–2050.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf. 1999. *Subsampling*. Springer.

Racine, Jeffrey S. and Qi Li. 2004. "Nonparametric estimation of regression functions with both categorical and continuous data." *Journal of Econometrics*, 119 (1): 99–130.

Rosenblatt, Murray. 1956. "Remarks on some nonparametric estimates of a density function." *Annals of Mathematical Statistics*, 27 (3): 832–837.

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *Quarterly Journal of Economics*, 116 (2): 681–704.

———. 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" *Handbook of the Economics of Education*, 3 (3): 249–277.

Schennach, Susanne M. 2004. "Estimation of nonlinear models with measurement error." *Econometrica*, 72 (1): 33–75.

———. 2007. "Instrumental variable estimation of nonlinear errors-in-variables models." *Econometrica*, 75 (1): 201–239.

———. 2016. "Recent advances in the measurement error literature." *Annual Review of Economics*, 8 (1): 341–377.

Schwert, Michael. 2018. "Bank capital and lending relationships." *Journal of Finance*, 73 (2): 787–830.

Sherman, Robert P. 1993. “The limiting distribution of the maximum rank correlation estimator.” *Econometrica*, 61 (1): 123–137.

———. 2010. “Maximum score methods.” In *Microeconometrics*. Springer, 122–128.

Silverman, Bernard W. 1978. “Weak and strong uniform consistency of the kernel estimate of a density and its derivatives.” *Annals of Statistics*: 177–184.

Thomas, Alban. 2006. “Consistent estimation of binary-choice panel data models with heterogeneous linear trends.” *Econometrics Journal*, 9 (2): 177–195.

Ura, Takuya. 2018. “Heterogeneous treatment effects with mismeasured endogenous treatment.” *Quantitative Economics*, 9 (3): 1335–1370.

———. 2021. “Instrumental variable quantile regression with misclassification.” *Econometric Theory*, 37 (1): 169–204.

Van Der Vaart, Aad W. and Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Wan, Yuanyuan and Haiqing Xu. 2014. “Semiparametric identification of binary decision games of incomplete information with correlated private signals.” *Journal of Econometrics*, 182 (2): 235–246.

Wansbeek, Tom and Erik Meijer. 2000. *Measurement Error and Latent Variables in Econometrics*. Advanced Textbooks in Economics. Elsevier.

Wu, Chunhua. 2015. “Matching value and market design in online advertising networks: An empirical analysis.” *Marketing Science*, 34 (6): 906–921.

Yanagi, Takahide. 2019. “Inference on local average treatment effects for misclassified treatment.” *Econometric Reviews*, 38 (8): 938–960.

Yang, Yupin, Mengze Shi, and Avi Goldfarb. 2009. “Estimating the value of brand alliances in professional team sports.” *Marketing Science*, 28 (6): 1095–1111.