Recovering the True Distribution with Paired Self Reports

Yingyao Hu*   Yuya Sasaki

March 20, 2014

Abstract

This paper proposes nonparametric identifying restrictions for paired nonseparable measurement error models. Applying this method to twin panel data, we recover the distribution of true years of education using self reports and sibling reports, and find the following robust reporting patterns. Self reports are accurate only when the true years of education are 16 or 18, typically corresponding to advanced university degrees in the US education system. Sibling reports are accurate whenever the true years of education are 12, 14, 16, and 18, that are typical diploma years. Such irregular results would not have been obtained with the classical additive-error models.

Keywords: measurement errors, nonseparable errors, twin panel data, years of education

1 Introduction

Consider the paired nonseparable measurement error model of the following form.

$$\begin{cases} X = g(U,V) \\ Y = h(U,W) \end{cases} \quad \text{where } U,\ V \text{ and } W \text{ are mutually independent.} \tag{1.1}$$

The random variables X and Y are observed by econometricians, but the rest is not. To fix ideas, think of U as the true years of education which econometricians do not observe. Instead

*Johns Hopkins University. We can be reached at yhu@jhu.edu and sasaki@jhu.edu. We received helpful comments from seminar participants at Ohio State University, Greater New York Metropolitan Area Econometrics Colloquium 2012, AMES 2013, and ESEM 2013. The usual disclaimer applies.

we observe self reports X and sibling reports Y of U. The nonseparable errors V and W are non-additive factors of self and sibling reporting errors, respectively.

The general model (1.1) encompasses a variety of important sub-models as follows.

Example 1 (Paired Classical Model)**.** *The paired classical measurement error model is:*

$$\begin{cases} X = U + V \\ \\ Y = U + W \end{cases} \qquad \textit{where } U,\ V \textit{ and } W \textit{ are mutually independent.}$$

Figure 1 (a) illustrates the conditional supports of X and Y given U for this model. △

A large number of econometric papers use this paired additive model or its variants.[1] It is well known that the marginal distributions of the unobserved variables are identified for Example 1 under a locational normalization and regularity, based on deconvolution approaches using multiplicative separability of characteristic functions under the stated independence restriction of the additive latent variables.

The classical model of Example 1 rules out endogeneity in measurement/reporting errors. The following two examples generalize the classical model. Suppose again that U, X, and Y are the true, self-reported, and sibling-reported years of education, respectively.

Example 2 (Paired Heteroskedastic Model)**.** *Heteroskedasticity can be introduced as follows:*

$$\begin{cases} X = U + \phi(U) \cdot V \\ \\ Y = U + \psi(U) \cdot W \end{cases} \qquad \textit{where } U,\ V \textit{ and } W \textit{ are mutually independent.}$$

Functions ϕ and ψ are positive skedastic functions. This case may arise when individuals with more years of education U provide their self-reported years of education X with higher precision, i.e., smaller $\phi(U)$. Figure 1 (b) illustrates this heteroskedastic error model. △

[1]Examples include, but are not limited to, measurement error models (Li and Vuong, 1998; Li, 2002; Schennach, 2004; Song, Schennach and White, 2012), auction models (Li, Perrigne, and Vuong, 2000; Krasnokutskaya, 2011), panel models (Evdokimov, 2010; Arellano and Bonhomme, 2012), and labor economic applications (Cunha, Heckman, and Navarro, 2005; Bonhomme and Robin, 2010; Hansen, Heckman and Mullen, 2004; Kennan and Walker, 2011).

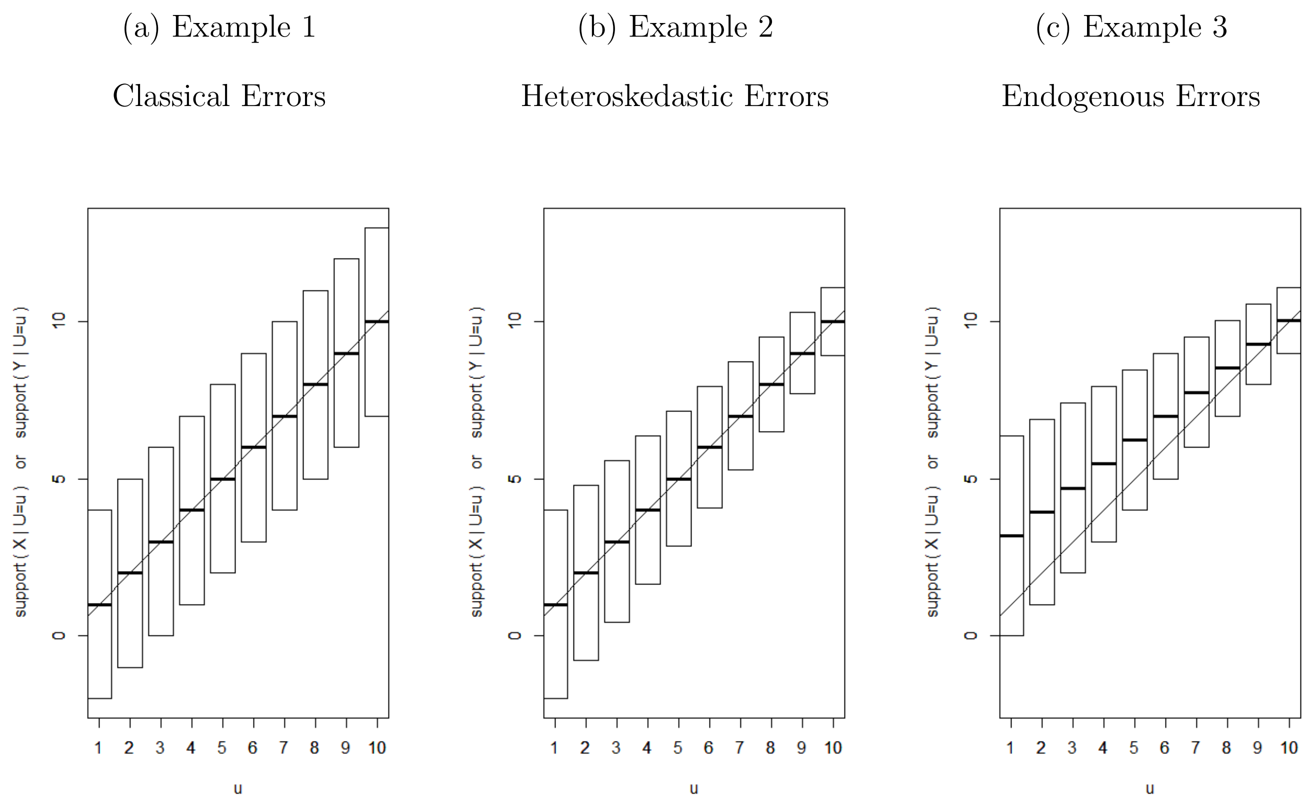


Figure 1: Illustration of the conditional supports of $X \mid U = u$ for (a) paired classical error models, (b) paired heteroskedastic error models, and (c) paired endogenous error models.

Example 3 (Paired Endogenous Error Model)**.** *Measurement errors may be endogenous:*

$$\begin{cases} X = U + \phi(U,V) \\ Y = U + \psi(U,W) \end{cases} \qquad \textit{where } U,\ V \textit{ and } W \textit{ are mutually independent.}$$

Individuals with less years of education U may tend to provide more overstated self reports X, e.g., $E[\phi(u,V)]$ is decreasing in u. Figure 1 (c) illustrates this endogenous error model. $\triangle$

The last two examples describe plausible reporting behaviors which additively separable independent errors of the classical model cannot accommodate. The general model (1.1) that we study in this paper nests these two examples. In addition to these paired measurement error models, the baseline setup (1.1) can also be used to reflect nonparametric structural models with endogenous measurement errors as illustrated in the following example.

Example 4 (Unobserved Factors). *Consider a production function* $Y = h(U, W)$*, where* U *is the quantity of a factor of production and* W *summarizes unobserved technologies. The true quantity* U *is often imperfectly observed with conceivably endogenous measurement error* $\phi(U, V)$*. Let* X *denote an observed proxy of* U*. We hence obtain the following paired structure.*

$$\begin{cases} X = U + \phi(U, V) \\ Y = h(U, W) \end{cases} \quad \text{where } U,\ V \text{ and } W \text{ are mutually independent.}$$

Economists are often interested in identifying the structural responses of the produced quantity Y *to the true unobserved quantity* U *of factors, i.e.,* h *or* $F_{Y|U}$*. For a related example, Kim, Petrin, and Song (2013) propose how to estimate production functions with mis-measured factors.* $\triangle$

Under the stated independence condition, we can represent the model (1.1) by the triple $(F_{X|U}, F_{Y|U}, \mu_U)$ of conditional and marginal distribution functions,[2] where μ_U denotes the probability measure of U, which may be discrete, continuous, or a mixture of them. In the case of classical model (Example 1), identification of this triple $(F_{X|U}, F_{Y|U}, \mu_U)$ leads to identification of the triple (μ_V, μ_W, μ_U) of the marginal distributions of V, W and U. For observed variables X and Y, we assume that their conditional distributions given U are either discrete or continuous. Let $f_{X|U}$ and $f_{Y|U}$ denote the conditional pmf (respectively, pdf) when they are discrete (respectively, continuous). The supports of the marginal distributions of X, Y, and U are denoted by $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{U}$, respectively. We are interested in finding restrictions under which the triple $(F_{X|U}, F_{Y|U}, \mu_U)$ is identified.

Our model (1.1) is related to a number of nonclassical measurement error models considered in the literature (e.g., Chen, Hong and Tamer, 2005; Mahajan, 2006; Lewbel, 2007; Chen, Hong and Tarozzi, 2008; Hu, 2008; Hu and Schennach, 2008; Carroll, Chen and Hu, 2010; D'Haultfoeuille and Février, 2010; Song, Schennach and White, 2012). The most closely related

[2] This representation of otherwise observationally equivalent set of underlying structures follows by normalizing the distributions of V and W. See Matzkin (2003) for necessity of normalizing the error distributions for nonseparable models, and for economically-grounded examples of normalization.

is D'Haultfoeuille and Février (2010), who show nonparametric identification of nonseparable measurement error models using support variations and three or more measurements of the latent variable. Similarly to the approach of D'Haultfoeuille and Février, we use support variations as a source of identification. The empirical data that we use in this paper is based only on self and sibling reports, and contains neither three measurements nor instruments. We thus need to relax data requirement of these existing econometric methods. To this end, we develop alternative identifying restrictions where our model (1.1) requires only two measurements, X and Y, instead of three, and our identification strategy does not rely on instrumental variables.

In the rest of the paper, we describe how to identify the triple. Section 2 introduces a general identifying restriction without any specific economic application in mind. Section 3 follows up by providing primitive sufficient condition in the context of our empirical application.

2 The Basic Identification Result

In this section, we propose a general identification result based on a high-level assumption. The identification strategy proposed in this paper relies on a set structure equipped with order relations. We first introduce the definition of well-orders which we use as the main device in this paper.

Definition 1 (Total Order and Well-Order). *A relation $\preceq$ on $\mathcal{U}$ is a total order if (i) $u \preceq u'$ and $u' \preceq u$ imply $u = u'$, (ii) $u \preceq u'$ and $u' \preceq u''$ imply $u \preceq u''$, and (iii) $u \preceq u'$ or $u' \preceq u$. A total order relation $\preceq$ on $\mathcal{U}$ is a well-order if every nonempty subset of $\mathcal{U}$ has a $\preceq$-least element.*

The Well-Ordering Theorem (equivalent to the Axiom of Choice and to Zorn's Lemma) states that any set $\mathcal{U}$ can be equipped with a well-order. In addition to its guaranteed existence, we require that a well-order $\preceq$ on $\mathcal{U}$ satisfies the following condition in order to derive identification.

Restriction 1 (The Basic Identifying Restriction)**.** *There exists a well-order $\preceq$ on $\mathcal{U}$ such that*

(i) $\mathcal{X}(u) :=$ *support* $\left(f_{X|U}(\ \cdot \mid u)\right) \setminus \cup_{u \prec u'}$ *support* $\left(f_{X|U}(\ \cdot \mid u')\right) \neq \emptyset$ *holds for all* $u \in \mathcal{U}$*,*

(ii) $\mathcal{Y}(u) :=$ *support* $\left(f_{Y|U}(\ \cdot \mid u)\right) \setminus \cup_{u \prec u'}$ *support* $\left(f_{Y|U}(\ \cdot \mid u')\right) \neq \emptyset$ *holds for all* $u \in \mathcal{U}$*,*

(iii) $\{u' \in \mathcal{U} \mid u' \preceq u\}$ *and* $\{u' \in \mathcal{U} \mid u' \prec u\}$ *are* μ_U*-measurable for each* $u \in \mathcal{U}$*, and*

(iv) $\mu_U\left(\{u' \in \mathcal{U} \mid u' \preceq u\} \cap B(u, r)\right) > 0$ *for each* $u \in \mathcal{U}$ *for each* $r > 0$*,*

where $B(u, r)$ *denotes the ball of radius* r *around* u *with respect to the Euclidean metric on* $\mathcal{U}$*.*

This identifying restriction, stated for general applicability, may sound somewhat abstract. In practice, intuitive lower-level assumptions relevant to specific applications are more useful. In the current section without any specific application in mind, we postpone discussing the economic intuition of this high-level assumption. Section 3 discusses how to achieve this abstractly stated general restriction from more primitive restrictions in the context of our empirical application. Economic and behavioral interpretation of Restriction 1 becomes clearer in that section.

In general, neither $\mathcal{X}(u)$ nor $\mathcal{Y}(u)$ is known. For general treatment, the current section proceeds by assuming that they are known. We remark, however, that these sets can be identified by the observed data in many regular cases of (1.1) – see Section 3.1. In Section 3.2, we show an alternative approach to define $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ on the basis of behavioral assumptions. In Section 3.3, we demonstrate a yet alternative method of obtaining $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ through behavioral restrictions and empirical data.

The most important parts of Restriction 1 are (i) and (ii). Section 3 demonstrates intuitive sufficient conditions for these parts of the restriction in the context of our application. Part (iii) of the restriction requires that the well-order relation $\preceq$ conform with the measure space $(\mathcal{U}, \sigma(U), \mu_U)$. This is trivially satisfied if $\sigma(U)$ is a sub-family of the Borel sigma algebra induced by the $\preceq$-order topology on $\mathcal{U}$. Part (iv) of the restriction is a regularity assumption used in the case of a continuous distribution of U, and particularly used in Lemma 3 shortly. It is trivially satisfied at point masses of U, and it is automatically and globally satisfied when

U follows a discrete distribution.

The main claim of this paper is that the model representation $(F_{X|U}, F_{Y|U}, \mu_U)$ can be identified under Restriction 1. This result is derived by the principle of transfinite induction on the support $\mathcal{U}$ of U. It extends the standard principle of mathematical induction, and is applicable to any well-ordered set $(\mathcal{U}, \preceq)$, be it finite, countably infinite, or uncountably infinite. To use transfinitely inductive arguments, it is important to note that the elements of a well-ordered set are categorized into the following three groups.

Definition 2 (Ordinals)**.** *Let $(\mathcal{U}, \preceq)$ be a well-ordered set. The zero ordinal of $\mathcal{U}$ is the $\preceq$-least element of $\mathcal{U}$. An element of $\mathcal{U}$ is the successor ordinal of $u \in \mathcal{U}$ if it is the $\preceq$-least element of $\{u' \in \mathcal{U} \mid u \prec u'\}$. All the other elements of $\mathcal{U}$ are called limit ordinals.*

For convenience of operating the principle of transfinite induction with minimal notations, we introduce the following restricted probability measures, which is defined under Restriction 1 (iii).

Definition 3 (Restricted Measures)**.** *For each $u \in \mathcal{U}$, we define $\sigma^{\preceq}(U, u) = \{B \cap \{u' \in \mathcal{U} \mid u' \preceq u\} \mid B \in \sigma(U)\}$ and $\sigma^{\prec}(U, u) = \{B \cap \{u' \in \mathcal{U} \mid u' \prec u\} \mid B \in \sigma(U)\}$. Denote the measure μ_U restricted to $\sigma^{\preceq}(U, u)$ and $\sigma^{\prec}(U, u)$ by $\mu_U \mid_{\sigma^{\preceq}(U,u)}$ and $\mu_U \mid_{\sigma^{\prec}(U,u)}$, respectively*

The principle of transfinite induction consists of three steps. In order to claim that a property holds with each point u of $\mathcal{U}$, one shows that 1. the property holds for the zero ordinal; 2. given any ordinal $u \in \mathcal{U}$, the property holds for the $\preceq$-successor ordinal of u whenever it holds for all $u' \preceq u$; and 3. the property holds for any limit ordinal $u \in \mathcal{U}$ whenever it holds for all $u' \prec u$. Thus, we state the following three auxiliary lemmas.

Lemma 1 (Transfinite Induction: Zero Ordinal)**.** *Suppose that Restriction 1 holds for (1.1). Let u_0 be the zero ordinal of $\mathcal{U}$. If sets $\mathcal{X}(u_0)$ and $\mathcal{Y}(u_0)$ are known and $P(u_0) = \mu_U(\{u_0\}) > 0$, then $(f_{X|U}(\,\cdot \mid u_0), f_{Y|U}(\,\cdot \mid u_0), P(u_0))$ is identified.*

Lemma 2 (Transfinite Induction: Successor Ordinal). *Suppose that Restriction 1 holds for (1.1). Assume transfinite-inductively that $\mu_U \mid_{\sigma^{\preceq}(U,u)}$ is known and $(f_{X|U}(\,\cdot \mid u'), f_{Y|U}(\,\cdot \mid u'))$ is known for $[\mu_U]$-a.s. $u' \preceq u$. Let u_+ be the $\preceq$-successor ordinal of u. If sets $\mathcal{X}(u_+)$ and $\mathcal{Y}(u_+)$ are known and $P(u_+) = \mu_U(\{u_+\}) > 0$, then $(f_{X|U}(\,\cdot \mid u_+), f_{Y|U}(\,\cdot \mid u_+), P(u_+))$ is identified.*

Lemma 3 (Transfinite Induction: Limit Ordinal). *Suppose that Restriction 1 holds for (1.1). Assume transfinite-inductively that $\mu_U \mid_{\sigma^{\prec}(U,u)}$ is known and $(f_{X|U}(\,\cdot \mid u'), f_{Y|U}(\,\cdot \mid u'))$ is known for $[\mu_U]$-a.s. $u' \prec u$. If the sets $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ are known, then the following results hold.*
(i) If $P(u) = \mu_U(\{u\}) > 0$, then $(f_{X|U}(\,\cdot \mid u), f_{Y|U}(\,\cdot \mid u), P(u))$ is identified.
(ii) If $\mu_U(\{u\}) = 0$, then $(f_{X|U}(\,\cdot \mid u), f_{Y|U}(\,\cdot \mid u), f_U(u))$ is identified, where f_U is the Radon-Nikodym derivative of the absolutely continuous part of μ_U with respect to the Lebesgue measure.

See the appendix for proofs of Lemmas 1–3. Lemma 1 is analogous to the base step of the classical induction. Similarly, Lemma 2 is analogous to the inductive step of the classical induction. Lemma 3 branches into two cases, depending on whether the limit ordinal u has a positive measure or zero measure. In the former case, the argument is again analogous to the inductive step of the classical induction. In the latter case, the density f_U of the absolutely continuous part of the probability measure μ_U evaluated at u is derived using the Lebesgue Differentiation Theorem based on Restriction 1 (iv). Albeit the technical difference, however, this latter case is the same in terms of the underlying idea as the former case. The former and the latter cases identify the probability of the point mass and the probability density at the location $u \in \mathcal{U}$ of interest, respectively.

With these three lemmas, the transfinite induction yields the following main identification result.

Theorem 1 (Identification). *Suppose that Restriction 1 holds for the model (1.1). If sets $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ are known for $[\mu_U]$-a.s. $u \in \mathcal{U}$, then μ_U is identified and $(f_{X|U}, f_{Y|U})$ is identified $[\mu_U]$-almost surely.*

This theorem assumes that $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ defined in Restriction 1 are known. In practice, we often need to invoke additional assumptions to define or identify these sets for each u. This requirement also demands that the unobserved elements of $\mathcal{U}$ be suitably labeled and well-ordered based on the plausibility of specific empirical problems. As previously mentioned, however, the sets $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ can be identified in many regular cases with natural labeling and natural ordering of the unobserved latent variable. The following section describes some cases in point – we propose several lower-level assumptions that are sufficient for identification of the sets $\mathcal{X}(u)$ and $\mathcal{Y}(u)$, as well as for the high-level assumption of Restriction 1.

The identifying condition of Restriction 1 is in fact satisfied in many regular cases if U is discrete, as shortly discussed in Section 3. These regular cases include those depicted in Figure 1 (a), (b), and (c). Particularly, the classical paired measurement error model of Example 1 necessarily satisfies Restriction 1 if the conditional supports are bounded and U is discrete. Example 1 ceases to satisfy this identifying restriction when U is continuous, and thus our result does not necessarily provide a strict improvement upon the ones based on the deconvolution approach. While our empirical application is concerned about the case of discrete U, however, our basic identification result does not rule out the cases of continuously supported U – see Section A.4 in the appendix for an example. Having stated so, we usually deal with discrete variables in empirical data, and thus the classical case (Example 1 and Figure 1 (a)) is usually encompassed by the basic identifying restrictions of this paper.

3 Examples of Primitive Sufficient Conditions

Our key identification result, Theorem 1, is based on the high-level assumption stated as Restriction 1. While Section 2 presents this abstractly stated restriction for the sake of generality, we can devise primitive assumptions that satisfy Restriction 1. Appropriate sufficient conditions should of course be sought based on plausibility in the contexts of specific applications. In the current section, we propose several examples of economically grounded sufficient conditions, particularly in relation to our empirical analysis of the years of schooling, where (X, Y) are reported years of schooling, U is the true year of schooling, and (V, W) are non-additive factors of reporting errors.

First, as remarked previously, existence of a well-order on $\mathcal{U}$ is not an issue, because the Well-Ordering Theorem (equivalent to the Axiom of Choice and to Zorn's Lemma) guarantees its existence for any set $\mathcal{U}$, be it finite or infinite. Thus, our problem is to let such a well-order satisfy conditions (i)–(iv) of Restriction 1. While the general result in Section 2 uses no such restriction, we assume in the current section that $\mathcal{U}$ is discrete, reflecting the nature of the economic variables, that we use for the empirical analysis in this paper. For finitely supported U, the conditions (iii) and (iv) of Restriction 1 are trivially satisfied, and thus we focus on finding primitive assumptions under which a well-order satisfies conditions (i) and (ii).

3.1 Monotone Support Boundaries – No Under-Reporting

Figure 1 in Section 1 suggests that, when measurement errors have bounded supports, the support boundaries are monotone in many regular cases, i.e., the upper and/or lower bounds of $\text{support}(f_{X|U}(\,\cdot\mid u))$ and $\text{support}(f_{X|U}(\,\cdot\mid u))$ are increasing in u. Notice that this monotonicity holds in the classical-error cases (Figure 1 (a)). Moreover, it is also often consistent with heteroskedastic and endogenous errors (Figure 1 (b) and (c)). This observation motivates the following assumption.

Assumption 1 (Monotone Support Boundaries)**.** *Let $(\mathcal{U}, \prec) = \{u_1, \cdots, u_J\}$ be a finite ordered set with $u_1 \prec \cdots \prec u_J$. Let the supports of X and Y be bounded. The following two conditions are satisfied.*

(i) $\inf\left(\text{supp}(f_{X|U}(\,\cdot\mid u_j))\right)$ *is increasing in j or* $\sup\left(\text{supp}(f_{X|U}(\,\cdot\mid u_j))\right)$ *is decreasing in j.*

(ii) $\inf\left(\text{supp}(f_{Y|U}(\,\cdot\mid u_j))\right)$ *is increasing in j or* $\sup\left(\text{supp}(f_{Y|U}(\,\cdot\mid u_j))\right)$ *is decreasing in j.*

This intuitive assumption of monotone support boundaries alone is sufficient for the general identifying restriction of Section 2, as stated below.

Proposition 1. *If Assumption 1 holds for (1.1), then Restriction 1 is satisfied.*

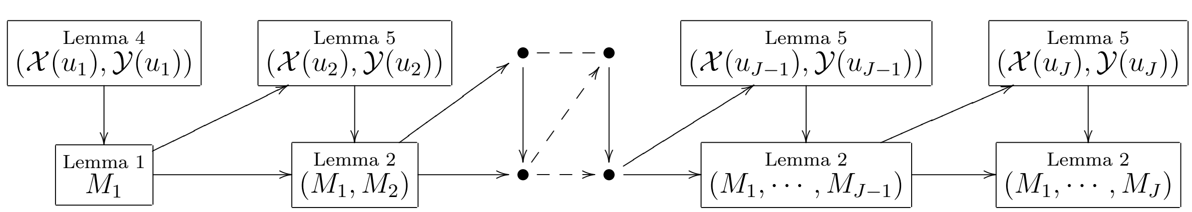
The proposition warrants that Lemmas 1–3 are effective under Assumption 1. Therefore, the model $(F_{X|U}, F_{Y|U}, F_U)$ representing (1.1) is identifiable, provided that we knew $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ for each j. Fortunately, under exactly the same Assumption 1, we too can inductively identify these sets $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ for each j. The following two lemmas establish this auxiliary result.

Lemma 4 (Supports: Zero Ordinal)**.** *Suppose that Assumption 1 holds for (1.1). The sets $\mathcal{X}(u_1)$ and $\mathcal{Y}(u_1)$ are identified.*

Lemma 5 (Supports: Successor Ordinal)**.** *Suppose that Assumption 1 holds for (1.1). Assume inductively that $(f_{X|U}(\ \cdot \mid u_k), f_{Y|U}(\ \cdot \mid u_k), f_U(u_k))$ is known for each $k < j$. The sets $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ are identified.*

Note that, even though we assume that U is finitely distributed in the current section, the supports of X and Y are still allowed to be countably or uncountably infinite. Therefore, the sets $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ identified by the above two lemmas may have infinite cardinalities.

Because we currently restrict the support of latent types to a finite set $\mathcal{U} = \{u_1, \cdots, u_J\}$ with $u_1 \prec \cdots \prec u_J$, the mixture component $M_j = (f_{X|U}(\ \cdot \mid u_j),\ f_{Y|U}(\ \cdot \mid u_j),\ f_U(u_j))$ is sequentially identified for each $j = 1, \cdots, J$ by the principle of classical mathematical induction together with Lemmas 1, 2, 4 and 5. The following diagram illustrates this iterative procedure as a flow chart.



First, the sets $\mathcal{X}(u_1)$ and $\mathcal{Y}(u_1)$ are identified in the base step through Lemma 4. Given $\mathcal{X}(u_1)$

and $\mathcal{Y}(u_1)$, Lemma 1 in turn shows identification of $(f_{X|U}(\,\cdot \mid u_1),\ f_{Y|U}(\,\cdot \mid u_1),\ f_U(u_1))$ in the base step. Subsequently, Lemmas 2 and 5 inductively identify the sequences of sets $\{\mathcal{X}(u_j)\}_j$, and $\{\mathcal{Y}(u_j)\}_j$, and the sequence of the triple $\{(f_{X|U}(\,\cdot \mid u_j),\ f_{Y|U}(\,\cdot \mid u_j),\ f_U(u_j))\}_j$. The last step obtains the model $(f_{X|U}, F_{Y|U}, F_U) = (M_1, \cdots, M_J)$. We summarize this identification result for the case of monotone support boundaries in the following proposition.

Proposition 2. *Suppose that Assumption 1 holds. The model $(F_{X|U}, F_{Y|U}, F_U)$ is identified.*

Next, we propose a behavioral assumption stated in the context of our empirical application that implies the monotone support boundaries of Assumption 1.

Assumption 2 (No Under-Reporting)**.** *The following two conditions are satisfied.*

(i) $Pr(X < U) = Pr(Y < U) = 0$.

(ii) $Pr(X = U \mid U = u) > 0$ and $Pr(Y = U \mid U = u) > 0$ for each $u \in \mathcal{U}$.

Part (i) states that individuals do not under-report years of education. Part (ii) states that there exist honest individuals for each actual year u. The former part may be restrictive in some applications, and the next subsection discusses how to relax this restriction. The behavioral restriction in Assumption 2 can be shown to imply the mechanical restriction in Assumption 1, and hence identifies the model by Proposition 2.

Proposition 3. *If Assumption 2 holds, then Assumption 1 holds with the well-order relation $\preceq$ on $\mathcal{U}$ defined by $u \prec u'$ if and only if $u < u'$. In particular, the model $(F_{X|U}, F_{Y|U}, F_U)$ is identified under Assumption 2.*

3.2 Restrictions by Reporting Behaviors – Stigma of Dropout

While Figure 1 illustrates that Assumption 1 is general enough to accommodate a wide class of regular models, this monotone-boundary assumption may be still too artificial to handle idiosyncratic reporting behaviors. To see this, consider the following example.

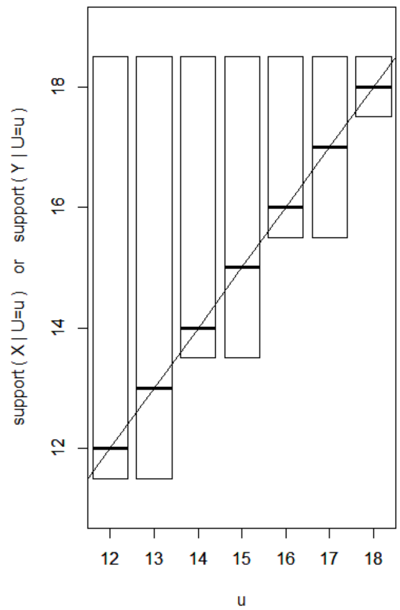


Figure 2: Illustration of the conditional supports of reported years of education when individuals may under-report their education for stigma against college dropout.

Example 5 (Academic Diploma and Stigma against Dropout)**.** *Those individuals having just completed diploma-granting years of education, e.g.,* $U = 12$*,* 14*,* 16*, and* 18*,*[3] *may have no incentive to under-report their education. On the other hand, the rest of individuals, i.e., those with* $U = 13$*,* 15*, and* 17*, may have an incentive to under-report their education by one year for stigma against dropout without diploma, or simply by rounding numbers to the nearest diploma-granting year for mnemonic reasons. As such, min support*$(X \mid U = 12) =$ *min support*$(X \mid U = 13) = 12$ *may result and Assumption 1 can thus fail. Figure 2 illustrate potential supports of* X *or* Y *given* U*, under this hypothetical story.*

This example illustrates a hypothetical yet plausible case where the true reporting behaviors are based on non-monotonic economic incentives, and, as a result, one may find Assumption 1 somewhat strong. In this light, we propose that the following behavioral assumption can be

[3] In the United States, 12, 14, 16, and 18 years of education are often, but not necessarily, associated with high school diploma, Associate degrees, Bachelor's degrees, and Master's degrees, respectively.

substituted for Assumption 1 as an alternative sufficient condition for the general requirement stated as Restriction 1.

Assumption 3 (Stigma against Dropout without Diploma)**.** *Let* $\mathcal{D} = \{d_1, \cdots, d_L\} \subset \mathcal{U}$ *be a set of diploma-granting years, and* $\mathcal{D}^c = \mathcal{U} \backslash \mathcal{D}$ *be its complement.*

(i) $Pr(X \in \mathcal{D}^c \mid U \in \mathcal{D}) = Pr(Y \in \mathcal{D}^c \mid U \in \mathcal{D}) = 0$*.*

(ii) $X < U \Rightarrow X = \max\{d \in \mathcal{D} \mid d \leqslant U\}$ *and* $Y < U \Rightarrow Y = \max\{d \in \mathcal{D} \mid d \leqslant U\}$*.*

(iii) $Pr(X = U \mid U = u) > 0$ *and* $Pr(Y = U \mid U = u) > 0$ *for each* $u \in \mathcal{U}$*.*

For example, $\mathcal{D} = \{12, 14, 16, 18\}$ can be used for common years of education associated with high school diploma, Associate degrees, Bachelor's degrees, and Master's degrees in the US education system. Part (i) of this restriction states that individuals who have actually just completed diploma-granting years of education do not report non-diploma-granting years of education. This restriction is plausible if we assume that they have no incentive to voluntarily lie to suffer from the dropout stigma. Part (ii) states that under-reporting individuals report the years of education associated with the highest diploma that they have actually received, so they can signal they did not drop out while only minimally suppressing the years. Part (iii) requires existence of honest subpopulation. Under this set of assumptions, the general identifying restriction of Section 2 is satisfied as follows

Proposition 4. *If Assumption 3 holds for (1.1), then Restriction 1 is satisfied.*

In the proof of this proposition, the well-order $\preceq$ on $\mathcal{U}$ induced by the following order relation $\prec$ is suggested.

$$\text{If } u \in \mathcal{D}^c \text{ and } u' = \max\{u'' \in \mathcal{D} \mid u'' < u\}, \text{ then } u \prec u'.$$
$$\text{Otherwise, } u < u' \Longleftrightarrow u \prec u'.$$

Intuitively, this definition of ordering states that 1. a non-diploma-granting year should precede the highest lower diploma-granting year; and 2. otherwise lower years should precede higher years. For example, we order the support $\mathcal{U} = \{12, 13, 14, 15, 16, 17, 18\}$ by $13 \prec 12 \prec 15 \prec 14 \prec 17 \prec 16 \prec 18$ according to the rule if $\mathcal{D} = \{12, 14, 16, 18\}$ and $\mathcal{D}^c = \{13, 15, 17\}$. This ordering also well defines the sets $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ for each u. If instead 14 (the typical year for associate degrees) is removed from $\mathcal{D}$, i.e., $\mathcal{D} = \{12, 16, 18\}$ and $\mathcal{D}^c = \{13, 14, 15, 17\}$, then the above rule specifies a well-order on $\mathcal{U}$ by $13 \prec 14 \prec 15 \prec 12 \prec 17 \prec 16 \prec 18$. However, we remark that a well-order satisfying Restriction 1 under Assumption 3 may not be unique in general. The above rule provides a general prescription of how to construct one particular well-order among others.

3.3 Restrictions on Empirical Data and Reporting Behaviors

In practice, one may want to impose the minimal amount of hypothetical restrictions just enough to achieve Restriction 1 given a specific empirical data set. In this section, we demonstrate a case in which behavioral assumptions alone do not satisfy Restriction 1, but these weaker behavioral assumptions can imply Restriction 1 for a specific data structure. In other words, we consider hybrid restrictions by empirical data and reporting behaviors as jointly sufficient conditions for the general identifying restrictions.

We continue to use the illustrative example from the Section 3.2, where the set $\mathcal{U} = \{12, 13, 14, 15, 16, 17, 18\}$ consists of years of education, $\mathcal{D} = \{12, 14, 16, 18\}$ contains diploma-granting years, and $\mathcal{D}^c = \{13, 15, 17\}$ contains of the rest of years. Instead of Assumption 3, we consider the following behavioral assumption.

Assumption 4. *Let $\mathcal{D} = \{d_1, \cdots, d_L\} \subset \mathcal{U}$ be a set of diploma-granting years, and $\mathcal{D}^c = \mathcal{U} \backslash \mathcal{D}$ be its complement.*

(i) $Pr(X = U \mid U = u) > 0$ and $Pr(Y = U \mid U = u) > 0$ for each $u \in \mathcal{U}$.

(ii) $Pr(X = \max\{d \in \mathcal{D} \mid d \leqslant u\} \mid U = u) > 0$ and

$Pr(Y = \max\{d \in \mathcal{D} \mid d \leqslant u\} \mid U = u) > 0$ for each $u \in \mathcal{U}$.

(iii) $Pr(X \in \mathcal{D}^c \mid U \in \mathcal{D}) = Pr(X < U \mid U \in \mathcal{D}) = 0$ and

$$Pr(Y \in \mathcal{D}^c \mid U \in \mathcal{D}) = Pr(Y < U \mid U \in \mathcal{D}) = 0 \text{ for each } u \in \mathcal{U}.$$

(iv) support$(f_{X|U}(\,\cdot \mid u))$ = *support*$(f_{Y|U}(\,\cdot \mid u))$ *for each* $u \in \mathcal{U}$.

Part (i) assumes that there is a positive probability of honestly reporting individuals for each true year u. Part (ii) assumes that there also exists a positive probability of individuals reporting the years of education corresponding to last-received diploma, e.g., freshman dropouts ($U = 13$) report $X = 12$. Part (iii) assumes that individuals who just completed diploma-granting years of education do not falsely report non-diploma-granting years of education or less years of education for self depreciation. Part (iv) assume that the conditional supports of self and sibling reports are the same.

Unlike Assumption 3 in Section 3.2, Assumption 4 alone does not imply Restriction 1 in general. However, we demonstrate that these restrictions may suffice for Restriction 1 given specific empirical data. The following table shows the support of the empirical data f_{XY} which we use for our application in Section 4.

f_{XY}		Y: 12	13	14	15	16	17	18
	12	+	+	+	0	+	+	0
	13	+	+	+	+	+	+	0
	14	+	+	+	+	+	0	0
X	15	+	+	+	+	+	0	+
	16	0	0	+	+	+	+	+
	17	0	+	+	0	+	+	+
	18	0	+	+	0	+	+	+

The entries '+' indicate $f_{XY}(x, y) > 0$, and the entries '0' indicate $f_{XY}(x, y) = 0$.

By the independence condition in the model (1.1), the equality $f_{XY}(x, y) = \sum_{u=12}^{18} f_{X|U}(x \mid u) f_{X|U}(y \mid u) f_U(u)$ must hold for each pair (x, y). Assuming that $f_U(u) \neq 0$ for each $u = 12, \cdots, 18$, the above restrictions together with the zero entries in the above table successively narrow down the supports of the unknown pmfs, $f_{X|U}(\,\cdot \mid u)$ and $f_{Y|U}(\,\cdot \mid u)$. For example, the entry $0 = f_{XY}(16, 13) \geqslant f_{X|U}(16 \mid 16) f_{Y|U}(13 \mid 16) f_U(16)$ of the empirical data and the restriction (i), i.e., $f_{X|U}(16 \mid 16) > 0$, imply that $f_{Y|U}(13 \mid 16) = 0$ must hold. For another example, the entry $0 = f_{XY}(16, 13) \geqslant f_{X|U}(16 \mid 17) f_{Y|U}(13 \mid 17) f_U(17)$ of the empirical data and the restriction (ii), i.e., $f_{X|U}(16 \mid 17) > 0$, imply that $f_{Y|U}(13 \mid 17) = 0$ must hold. Continuing these calculations for every pair (x, y) and for every restriction from (i)–(iv), we can successively eliminate many pairs (x, u) and (y, u) from the supports of $f_{X|U}$ and $f_{Y|U}$, respectively. The following tables summarize restrictions $f_{X|U}(\,\cdot \mid u)$ and $f_{X|U}(\,\cdot \mid u)$ obtained by these operations. The entries '0' indicate $f_{X|U}(x \mid u) = 0$ and $f_{Y|U}(y \mid u) = 0$. The subscripts under zeros indicate which restriction was used to obtain the zeros.

				U			
$f_{X\mid U}$	12	13	14	15	16	17	18
12			$0_{(iii)}$	$0_{(i)}$	$0_{(iii)}$	$0_{(iv)}$	$0_{(i)}$
13	$0_{(iii)}$		$0_{(iii)}$		$0_{(iii)}$	$0_{(iv)}$	$0_{(i)}$
14					$0_{(iii)}$	$0_{(i)}$	$0_{(i)}$
X 15	$0_{(iv)}$	$0_{(iv)}$	$0_{(iii)}$		$0_{(iii)}$	$0_{(i)}$	$0_{(iii)}$
16	$0_{(i)}$	$0_{(i)}$					$0_{(iii)}$
17	$0_{(i)}$	$0_{(ii)}$	$0_{(iii)}$	$0_{(i)}$	$0_{(iii)}$		$0_{(iii)}$
18	$0_{(i)}$	$0_{(ii)}$	$0_{(iv)}$	$0_{(i)}$			

				U			
$f_{Y\mid U}$	12	13	14	15	16	17	18
12			$0_{(iii)}$	$0_{(iv)}$	$0_{(i)}$	$0_{(i)}$	$0_{(i)}$
13	$0_{(iii)}$		$0_{(iii)}$		$0_{(i)}$	$0_{(ii)}$	$0_{(iii)}$
14					$0_{(iii)}$	$0_{(iv)}$	$0_{(iii)}$
Y 15	$0_{(i)}$	$0_{(ii)}$	$0_{(iii)}$		$0_{(iii)}$	$0_{(i)}$	$0_{(i)}$
16	$0_{(iv)}$	$0_{(iv)}$					$0_{(iii)}$
17	$0_{(iv)}$	$0_{(iv)}$	$0_{(i)}$	$0_{(iv)}$	$0_{(iii)}$		$0_{(iii)}$
18	$0_{(i)}$	$0_{(i)}$	$0_{(i)}$	$0_{(ii)}$			

By inspection, we can see from these tables that Restriction 1 is satisfied with a well-order $\preceq$ on $\mathcal{U} = \{12, \cdots, 18\}$ defined by $15 \prec 13 \prec 12 \prec 14 \prec 17 \prec 16 \prec 18$. To see how the high-level conditions of Restriction 1 (i) and (ii) are satisfied, in particular, note that this order relation yields $\mathcal{X}(15) = \mathcal{Y}(15) = \{15\}$, $\mathcal{X}(13) = \mathcal{Y}(13) = \{13\}$, $\mathcal{X}(12) = \mathcal{Y}(12) = \{12\}$, $\mathcal{X}(14) = \mathcal{Y}(14) = \{14\}$, $\mathcal{X}(17) = \mathcal{Y}(17) = \{17\}$, $\mathcal{X}(16) = \mathcal{Y}(16) = \{16\}$, and $\mathcal{X}(18) = \mathcal{Y}(18) = \{18\}$.

4 Estimation Algorithm

The identification result, Theorem 1, implies that information accumulates as the sample size increases, which leads to consistent estimation of the representing model $(F_{X|U}, F_{Y|U}, \mu_U)$. When random variable U is continuously distributed, it is generally difficult to follow the sample counterpart of the proof of identification by the principle of transfinite induction, but we can use a variety of econometric methods available in the literature on nonparametrics. In this paper, we instead focus on the case of discrete variables, because it is relevant to our empirical application. This case allows for an explicit estimation algorithm following the sample counterpart of the inductive identification strategy, and therefore we need not implement numerical optimization on high dimensions.

As in Section 3, consider a finite ordered set $(\mathcal{U}, \preceq)$ with $u_1 \prec \cdots \prec u_J$. The proofs of Lemmas 1 and 2 suggest that the following iterative procedure estimates the representing model. Let f_{XY}^N denote the empirical joint pmf of the observed variables (X, Y), with f_X^N and f_Y^N denoting its marginals. Choosing points $x_1 \in \mathcal{X}(u_1)$ and $y_1 \in \mathcal{Y}(u_1)$, we estimate $(f_{X|U}(\,\cdot \mid u_1), f_{Y|U}(\,\cdot \mid u_1), f_U(u_1))$ by the following formulas.

$$\begin{aligned}
\hat{f}_{X|U}(x \mid u_1) &= \frac{f_{XY}^N(x, y_1)}{f_Y^N(y_1)} && \text{for all } x \in \mathcal{X} \\
\hat{f}_{Y|U}(y \mid u_1) &= \frac{f_{XY}^N(x_1, y)}{f_X^N(x_1)} && \text{for all } y \in \mathcal{Y} \\
\hat{f}_U(u_1) &= \frac{f_X^N(x_1) f_Y^N(y_1)}{f_{XY}^N(x_1, y_1)}
\end{aligned}$$

The fact that u_1 does not appear on the right-hand sides of these formulas may be intuitively understood by noting that x_1 and y_1 serve as control variables for u_1 under the varying support condition.

By the beginning of the j-th step, we have obtained $(\hat{f}_{X|U}(\,\cdot \mid u_k), \hat{f}_{Y|U}(\,\cdot \mid u_k), \hat{f}_U(u_k))$ for all $k < j$. Therefore, choosing $x_j \in \mathcal{X}(u_j)$ and $y_j \in \mathcal{Y}(u_j)$, we estimate $(f_{X|U}(\,\cdot \mid u_j), f_{Y|U}(\,\cdot \mid$

$u_j), f_U(u_j))$ in the j-th step by the following formulas.

$$\begin{aligned}
\hat{f}_{X|U}(x \mid u_j) &= \frac{f^N_{XY}(x, y_j) - \sum_{k=1}^{j-1} \hat{f}_{X|U}(x \mid u_k)\hat{f}_{Y|U}(y_j \mid u_k)\hat{f}_U(u_k)}{f^N_Y(y_j) - \sum_{k=1}^{j-1} \hat{f}_{Y|U}(y_j \mid u_k)\hat{f}_U(u_k)} && \text{for all } x \in \mathcal{X} \\
\hat{f}_{Y|U}(y \mid u_j) &= \frac{f^N_{XY}(x_j, y) - \sum_{k=1}^{j-1} \hat{f}_{X|U}(x_j \mid u_k)\hat{f}_{Y|U}(y \mid u_k)\hat{f}_U(u_k)}{f^N_X(x_j) - \sum_{k=1}^{j-1} \hat{f}_{X|U}(x_j \mid u_k)\hat{f}_U(u_k)} && \text{for all } y \in \mathcal{Y} \\
\hat{f}_U(u_j) &= \frac{\left[f^N_X(x_j) - \sum_{k=1}^{j-1} \hat{f}_{X|U}(x_j \mid u_k)\hat{f}_U(u_k)\right]\left[f^N_Y(y_j) - \sum_{k=1}^{j-1} \hat{f}_{Y|U}(y_j \mid u_k)\hat{f}_U(u_k)\right]}{f^N_{XY}(x_j, y_j) - \sum_{k=1}^{j-1} \hat{f}_{X|U}(x_j \mid u_k)\hat{f}_{Y|U}(y_j \mid u_k)\hat{f}_U(u_k)}
\end{aligned}$$

Because J is finite, we can complete this iterative procedure to eventually obtain the estimated representing model $(\hat{f}_{X|U}, \hat{f}_{Y|U}, \hat{f}_U)$.

Note that the estimator $(\hat{f}_{X|U}(\,\cdot \mid u_j), \hat{f}_{Y|U}(\,\cdot \mid u_j), \hat{f}_U(u_j))_{j=1}^J$ is trivially a smooth transformation of the empirical data F^N_{XY} through the above closed-form arithmetic formulas. Therefore, the standard $\sqrt{N}$-asymptotic normality of this estimator immediately follows from the first-order asymptotics by the weak convergence of the empirical process $\sqrt{N}(F^N_{XY} - F_{XY})$ through the delta method. Although the arguments are standard, we present concrete expressions for asymptotic variances. We focus on the case of $j = 1$ for compactness of exposition. Similar arguments continue to apply for higher j, albeit more writings.

Proposition 5. *Suppose that one of the alternative identifying restrictions is satisfied and that the sample is drawn independently from an identical distribution.*

(i) If $f_Y(y_1) > 0$, then $\sqrt{N}\left(\hat{f}_{X|U}(x \mid u_1) - f_{X|U}(x \mid u_1)\right)$ asymptotically follows the normal distribution with mean zero and variance

$$\frac{f_{XY}(x, y_1)\left[f_Y(y_1) - f_{XY}(x, y_1)\right]}{f_Y(y_1)^3}.$$

(ii) If $f_X(x_1) > 0$, then $\sqrt{N}\left(\hat{f}_{Y|U}(y \mid u_1) - f_{Y|U}(y \mid u_1)\right)$ asymptotically follows the normal distribution with mean zero and variance

$$\frac{f_{XY}(x_1, y)\left[f_X(x_1) - f_{XY}(x_1, y)\right]}{f_X(x_1)^3}.$$

(iii) If $f_{XY}(x_1, y_1) > 0$, then $\sqrt{N}\left(\hat{f}_U(u_1) - f_U(u_1)\right)$ asymptotically follows the normal distribution with mean zero and variance

$$\frac{f_X(x_1)f_Y(y_1)\left[(f_X(x_1) - f_{XY}(x_1, y_1))(f_Y(y_1) - f_{XY}(x_1, y_1)) + f_{XY}(x_1, y_1)(f_{XY}(x_1, y_1) - f_X(x_1)f_Y(y_1))\right]}{f_{XY}(x_1, y_1)^3}.$$

5 The True Distribution of Years of Education

In labor economics and economics of education, isolating unobserved innate abilities from intensities of endogenous treatments, such as years of education, is a great concern for program evaluations. For panel data of monozygotic twins sharing innate abilities as common factors, it is a common practice to assume that within-pair differences in labor outcomes are imputed to differential treatment intensities. Behrman, Taubman, and Wales (1977) use a sample of twin panel to estimate the effects of schooling on labor outcomes. Ashenfelter and Krueger (1994) advance this literature by accounting for potential measurement errors in years of education in addition to controlling for the unobserved heterogeneity. Also see Miller, Mulvey and Martin (1995), Behrman and Rosenzweig (1999), and Rouse (1999) for related empirical research.

In order to correct errors in self-reported education, Ashenfelter and Krueger collect a sample of not only self-reported education, but also sibling-reported education in the 16th Annual Twins Days Festival in Twinsburg, Ohio, in 1991. The paired classical measurement error model assumed by their study can be represented by

$$\begin{cases} X = U + V \\ Y = U + W \end{cases} \quad \text{where } U,\ V \text{ and } W \text{ are mutually independent.} \tag{5.1}$$

See Example 1 in Section 1 for a discussion of this classical error model. The unobserved variable U denotes the true years of education. Econometricians observe the self-reported years of education denoted by X, and the sibling-reported years of education denoted by Y. The exogenous unobserved variables V and W are self-reporting error and sibling reporting error, respectively.

If the additive independent errors in the model (5.1) were indeed true, then existing approaches might be applicable to identify the distribution of true years of education. However, this classical measurement error setup is perhaps too restrictive in the current context for at least two reasons. First, self-reporting errors V are likely to be negatively correlated with U, as reported by Siegel and Hodge (1968). For example, individuals with less U may have upwardly biased errors V due to stigma, whereas individuals with high U may have no such incentive to give biased reports – Figure 1 (c) illustrates this case. In this light, it is more general to assume endogenous self-reporting error via the nonseparable model $X = g(U, V)$, where the self-reporting error defined by $[g(U, V) - U]$ is no longer independent of U by construction. Second, sibling-reporting errors W are likely to be correlated with the true U. For example, siblings may round true U up to the nearest diploma years, like $Y = 14$, 16, and 18, simply due to limited memory. In this case, the reporting errors W may be almost degenerate if true U is already one of the diploma years, whereas W may be non-degenerate otherwise. In other words, the distribution of W is likely to depend on U without any monotonic patterns. For such irregularly endogenous reporting errors, a nonseparable model $Y = h(U, W)$ is probably a more natural description of the true reporting behaviors, where the sibling reporting error defined by $h(U, W) - U$ is no longer independent of U by construction. Therefore, we replace the paired classical measurement error model (5.1) by the paired nonseparable measurement error model (1.1) in our empirical analysis:

$$\begin{cases} X = g(U, V) \\ Y = h(U, W) \end{cases} \qquad \text{where } U,\ V \text{ and } W \text{ are mutually independent.}$$

Sections 3.1, 3.2, and 3.3 propose three primitive sufficient conditions for nonparametric identification of this model, particularly in the context of the current empirical problem. Table 1 summarizes the well-orders obtained through each of the three approaches proposed in Section 3. Because none of Assumptions 2, 3, and 4 is empirically testable, we do not want to stick to any one of these particular identifying restrictions. Instead, we estimate our model under each of these alternative assumptions, and report the results that we obtain robustly across these alternative assumptions.

In our empirical analysis, we use the data of Ashenfelter and Krueger that consists of an extract from a survey of twins conducted at the 16th Annual Twins Days Festival in Twinsburg,

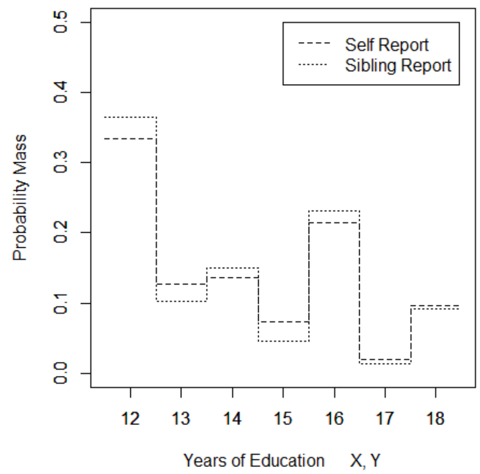
Section	Assumption	Economic Behavior	Implied Well-Order
3.1	2	No Under-Reporting	$12 \prec 13 \prec 14 \prec 15 \prec 16 \prec 17 \prec 18$
3.2	3	Stigma against Dropout	$13 \prec 12 \prec 15 \prec 14 \prec 17 \prec 16 \prec 18$
3.3	4	Stigma against Dropout	$15 \prec 13 \prec 12 \prec 14 \prec 17 \prec 16 \prec 18$

Table 1: Summary of identifying restrictions and the implied well-orders.

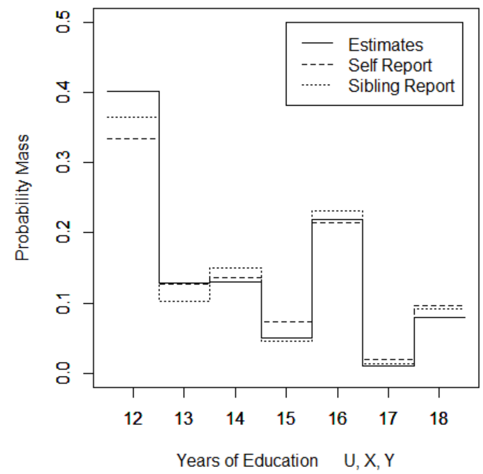
Ohio, in 1991. The sample contains 340 twins, and hence 680 individuals. Figure 3 (a) shows probability masses of self-reported years of education X (dashed lines) and sibling-reported years of education Y (dotted lines). Both X and Y have relative peaks at the diploma years, namely high school graduation (12), associate degrees (14), bachelors degrees (16), and masters degrees (18). The sibling report Y particularly stand out at these peaks, suggesting that sibling reports may perhaps tend to round the true U to near diploma years more evidently than the self reports X. The discrepancy between X and Y suggests that at least one of X and Y is false, and the truth may be neither of them.

Following the iterative procedure outlined in Section 4, we estimate the distribution F_U of the true years of schooling under each of the three restrictions summarized in Table 1. The three remaining graphs in Figure 3 show the probability masses of the estimated true years of schooling under (b) the assumption of no under-reporting, (c) the assumption of stigma against dropout without diploma, and (d) the hybrid restrictions by empirical data and reporting behaviors. The frequencies of self-reports are pretty accurate at 16 and 17 years of education robustly across all the three identifying restrictions. However, this particular result does not imply that individuals with $U = 16$ are honest reporters. There may exist individuals with other values of U who falsely report $X = 16$, i.e., frequency 'inflows' into $X = 16$. These inflows must be compensated for by false reports by individuals with $U = 16$, i.e., frequency 'outflows' from $U = 16$, because $\hat{f}_U(16) \approx f_X^N(16)$ requires conservation of inflowing and outflowing frequencies. By similar arguments, the discrepancy in the frequencies between self reports and the estimated truths at 18 does not imply that individuals with $U = 18$ tend to lie. The difference may be

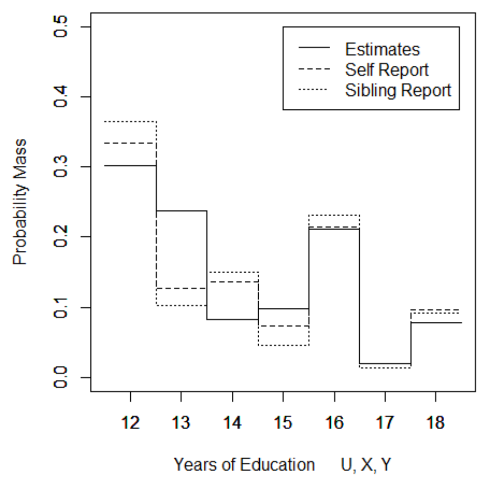
(a) Probability mass of X and Y.



(b) Estimated probability mass of U



(c) Estimated probability mass of U.



(d) Estimated probability mass of U.

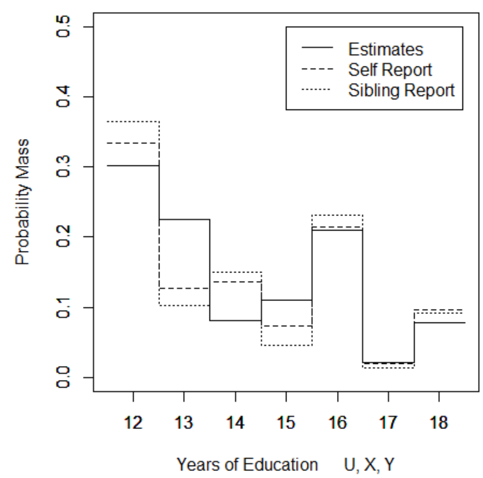


Figure 3: (a) Probability masses of self-reported education X and sibling-reported education Y. The remaining three graphs illustrate estimated probability masses of true education U under (b) the assumption of no under-reporting, (c) the assumption of stigma against dropout without diploma, and (d) the hybrid restrictions by empirical data and reporting behaviors.

due only to frequency inflows from other values of U into $X = 18$.

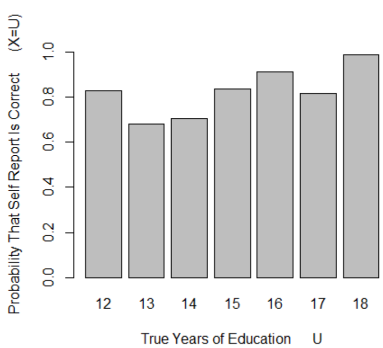
In order to assess the actual reporting behaviors to tell who tend to report correctly or falsely, we can use the estimated pmfs $(\hat{f}_{X|U}, \hat{f}_{Y|U}, \hat{f}_U)$ to compute the conditional probabilities of correct reports given the truth as follows:

$$\begin{aligned} \widehat{\Pr}(\text{Self report is correct} \mid U = u) &= \hat{f}_{X|U}(u \mid u) \\ \widehat{\Pr}(\text{Sibling report is correct} \mid U = u) &= \hat{f}_{Y|U}(u \mid u) \end{aligned}$$

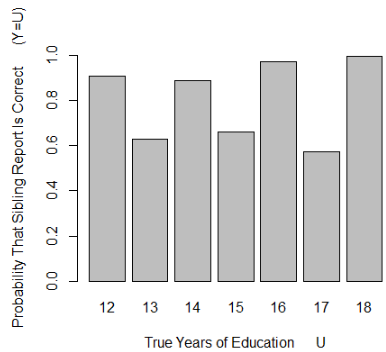
These estimated conditional probabilities of correct self and sibling reports are shown in Figure 4. The left and right columns show the results of self reports and sibling reports, respectively. The results displayed in the top, middle, and bottom rows are based on the estimates under (b) the assumption of no under-reporting, (c) the assumption of stigma against dropout without diploma, and (d) the hybrid restrictions by empirical data and reporting behaviors, respectively. The pattern of self reports are somewhat different across the three specifications, but the left column robustly shows that the self reports tend to be accurate whenever the true years of education are $U = 16$ or 18, corresponding to Bachelor's and Master's degrees in the US education system, while they are robustly inaccurate when the true years of education are $U = 13$ which may be characterized as freshman/sophomore dropout. On the other hand, the right column robustly shows that the accuracy of sibling reports stands out at every even number, $U = 12$, 14, 16, and 18, corresponding to the typical diploma years, while sibling reports are robustly inaccurate whenever the truth is an odd number. Note that the estimation method used to obtain the results in the top row (b) does not rely on direct assumptions associated with a distinction between diploma years and other years, but the results show that the peaks of correct reporting probabilities occur exactly at diploma years.

By these robust parts of the results across the three alternative identifying assumptions, we make the following conclusion. First, the hypothesis that self reports are accurate when the true years of education correspond to the typical years granting high-level diplomas is not overturned. Second, the hypothesis that sibling reports tend to round the true numbers to

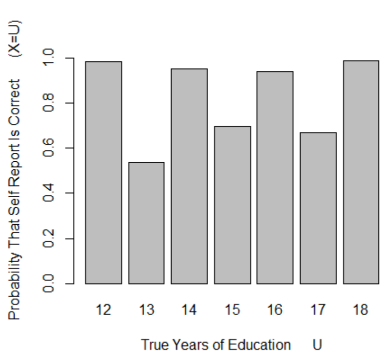
(b) Self reports.



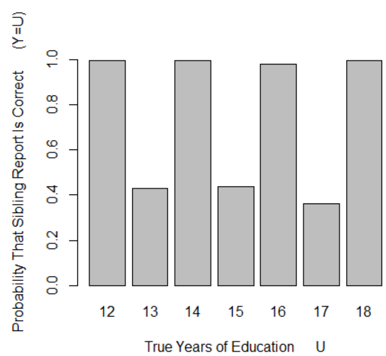
(b) Sibling reports.



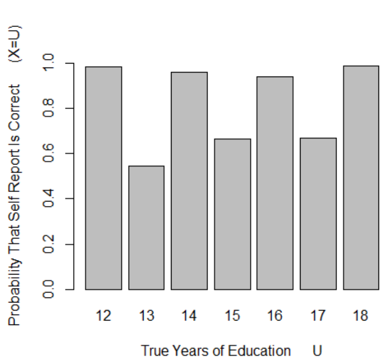
(c) Self reports.



(c) Sibling reports.



(d) Self reports.



(d) Sibling reports.

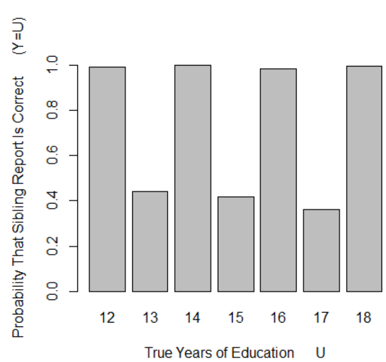


Figure 4: Left column: conditional probabilities that self report is correct given the truth. Right column: conditional probabilities that sibling report is correct given the truth. The conditional probabilities are provided by the estimates of $f_{X|U}$ and $f_{Y|U}$ under (b) the assumption of no under-reporting, (c) the assumption of stigma against dropout without diploma, and (d) the hybrid restrictions by empirical data and reporting behaviors.

typical diploma-granting years for mnemonic reasons is not overturned.

6 Summary

This paper proposes nonparametric identifying restrictions for nonseparable paired measurement error models, which encompass models with heteroskedastic measurement errors, endogenous measurement errors, and a variety of other nonlinear interactions between common factors and errors. The general identifying restriction requires that some well-order on the support of unobserved truth entails non-overlapping conditional supports. Sufficient conditions for this high-level assumption used for the general identification result should be sought in the contexts of specific applications.

Focusing on our empirical application, we propose several primitive sufficient conditions for the general identifying restriction. Applying the method to the twin panel data of Ashenfelter and Krueger (1994) containing self-reported and sibling-reported years of education, we attempt to recover the distribution of true years of education as well as the behavioral patterns of self reports and sibling reports. Across alternative identifying restrictions, we obtain the following robust patterns. Self reports are accurate if the true years of education are 16 or 18, typically corresponding to advanced university degrees. On the other hand, sibling reports are accurate when the true years of education are 12, 14, 16, or 18, that are typical diploma years. Such a nonlinear result would not have been obtained with the traditional methods based on additively separable independent errors.

A Mathematical Appendix

A.1 Proof of Lemma 1

Proof. Assume that $P(u_0) > 0$. By Restriction 1, we can choose $x^* \in \mathcal{X}(u_0)$ and $y^* \in \mathcal{Y}(u_0)$. Because $f_{X|U}(x^* \mid u) = 0$ for all $u \in \mathcal{U} \backslash \{u_0\}$ by the choice of x^*, we have $f_{XY}(x^*, y) = \int_{u \in \mathcal{U}} f_{X|U}(x^* \mid u) f_{Y|U}(y \mid u) d\mu_U(u) = f_{X|U}(x^* \mid u_0) f_{Y|U}(y \mid u_0) P(u_0)$ for all $y \in \mathcal{Y}$ by the independence assumption of the model (1.1). Similarly, $f_{XY}(x, y^*) = f_{X|U}(x \mid u_0) f_{Y|U}(y^* \mid u_0) P(u_0)$ holds for all $x \in \mathcal{X}$. In particular, $f_{XY}(x^*, y^*) = f_{X|U}(x^* \mid u_0) f_{Y|U}(y^* \mid u_0) P(u_0)$. Moreover, $f_X(x^*) = f_{X|U}(x^* \mid u_0) P(u_0)$ and $f_Y(y^*) = f_{Y|U}(y^* \mid u_0) P(u_0)$ follow. Using all these equalities, we get

$$\begin{aligned}
f_{X|U}(x \mid u_0) &= \frac{f_{X|U}(x \mid u_0) f_{Y|U}(y^* \mid u_0) P(u_0)}{f_{Y|U}(y^* \mid u_0) P(u_0)} = \frac{f_{XY}(x, y^*)}{f_Y(y^*)} && \text{for all } x \in \mathcal{X} \\
f_{Y|U}(y \mid u_0) &= \frac{f_{X|U}(x^* \mid u_0) f_{Y|U}(y \mid u_0) P(u_0)}{f_{X|U}(x^* \mid u_0) P(u_0)} = \frac{f_{XY}(x^*, y)}{f_X(x^*)} && \text{for all } y \in \mathcal{Y} \\
P(u_0) &= \frac{f_{X|U}(x^* \mid u_0) f_{Y|U}(y^* \mid u_0) P(u_0)^2}{f_{X|U}(x^* \mid u_0) f_{Y|U}(y^* \mid u_0) P(u_0)} = \frac{f_X(x^*) f_Y(y^*)}{f_{XY}(x^*, y^*)}
\end{aligned}$$

Note that the right-hand sides of these equalities consist of the observed data f_{XY}. Therefore, $(f_{X|U}(\,\cdot \mid u_0), f_{Y|U}(\,\cdot \mid u_0), P(u_0))$ is identified. $\square$

A.2 Proof of Lemma 2

Proof. Assume that $P(u_+) > 0$. By Restriction 1, we can choose $x^* \in \mathcal{X}(u_+)$ and $y^* \in \mathcal{Y}(u_+)$. Because $f_{X|U}(x^* \mid u') = 0$ for all $u' \succ u_+$ by the choice of x^*, we have $f_{XY}(x^*, y) = \int_{u' \in \mathcal{U}} f_{X|U}(x^* \mid u') f_{Y|U}(y \mid u') d\mu_U(u') = \int_{u' \preceq u} f_{X|U}(x^* \mid u') f_{Y|U}(y \mid u') d\mu_U(u') + f_{X|U}(x^* \mid u_+) f_{Y|U}(y \mid u_+) P(u_+)$ for all $y \in \mathcal{Y}$. Similarly, $f_{XY}(x, y^*) = \int_{u' \preceq u} f_{X|U}(x \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u') + f_{X|U}(x \mid u_+) f_{Y|U}(y^* \mid u_+) P(u_+)$ holds for all $x \in \mathcal{X}$. In particular, $f_{XY}(x^*, y^*) = \int_{u' \preceq u} f_{X|U}(x^* \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u') + f_{X|U}(x^* \mid u_+) f_{Y|U}(y^* \mid u_+) P(u_+)$. Moreover, $f_X(x^*) = \int_{u' \preceq u} f_{X|U}(x^* \mid u') d\mu_U(u') + f_{X|U}(x^* \mid u_+) P(u_+)$ and $f_Y(y^*) = \int_{u' \preceq u} f_{Y|U}(y^* \mid u') d\mu_U(u') +$

$f_{Y|U}(y^* \mid u_+)P(u_+)$ follow. Using all these equalities, we get

$$\begin{aligned}
f_{X|U}(x \mid u_+) &= \frac{f_{X|U}(x \mid u_+)f_{Y|U}(y^* \mid u_+)P(u_+)}{f_{Y|U}(y^* \mid u_+)P(u_+)} \\
&= \frac{f_{XY}(x, y^*) - \int_{u' \preceq u} f_{X|U}(x \mid u')f_{Y|U}(y^* \mid u')d\mu_U(u')}{f_Y(y^*) - \int_{u' \preceq u} f_{Y|U}(y^* \mid u')d\mu_U(u')} \qquad \text{for all } x \in \mathcal{X} \\
f_{Y|U}(y \mid u_+) &= \frac{f_{X|U}(x^* \mid u_+)f_{Y|U}(y \mid u_+)P(u_+)}{f_{X|U}(x^* \mid u_+)P(u_+)} \\
&= \frac{f_{XY}(x^*, y) - \int_{u' \preceq u} f_{X|U}(x^* \mid u')f_{Y|U}(y \mid u')d\mu_U(u')}{f_X(x^*) - \int_{u' \preceq u} f_{X|U}(x^* \mid u')d\mu_U(u')} \qquad \text{for all } y \in \mathcal{Y} \\
P(u_+) &= \frac{f_{X|U}(x^* \mid u_+)f_{Y|U}(y^* \mid u_+)P(u_+)^2}{f_{X|U}(x^* \mid u_+)f_{Y|U}(y^* \mid u_+)P(u_+)} \\
&= \frac{\left[f_X(x^*) - \int_{u' \preceq u} f_{X|U}(x^* \mid u')d\mu_U(u')\right]\left[f_Y(y^*) - \int_{u' \preceq u} f_{Y|U}(y^* \mid u')d\mu_U(u')\right]}{f_{XY}(x^*, y^*) - \int_{u' \preceq u} f_{X|U}(x^* \mid u')f_{Y|U}(y^* \mid u')d\mu_U(u')}
\end{aligned}$$

Note that the right-hand sides of these equalities consist of the observed data f_{XY}, or are transfinite-inductively assumed to be known. Therefore, $(f_{X|U}(\,\cdot \mid u_+), f_{Y|U}(\,\cdot \mid u_+), P(u_+))$ is identified. □

A.3 Proof of Lemma 3

Proof. (i) Let $P(u) > 0$. By Restriction 1, we can choose $x^* \in \mathcal{X}(u)$ and $y^* \in \mathcal{Y}(u)$. Because $f_{X|U}(x^* \mid u') = 0$ for all $u' \succ u$ by the choice of x^*, we have $f_{XY}(x^*, y) = \int_{u' \prec u} f_{X|U}(x^* \mid u')f_{Y|U}(y \mid u')d\mu_U(u') + f_{X|U}(x^* \mid u)f_{Y|U}(y \mid u)P(u)$ for all $y \in \mathcal{Y}$. Similarly, $f_{XY}(x, y^*) = \int_{u' \prec u} f_{X|U}(x \mid u')f_{Y|U}(y^* \mid u')d\mu_U(u') + f_{X|U}(x \mid u)f_{Y|U}(y^* \mid u)P(u)$ holds for all $x \in \mathcal{X}$. In particular, $f_{XY}(x^*, y^*) = \int_{u' \prec u} f_{X|U}(x^* \mid u')f_{Y|U}(y^* \mid u')d\mu_U(u') + f_{X|U}(x^* \mid u)f_{Y|U}(y^* \mid u)P(u)$. Moreover, $f_X(x^*) = \int_{u' \prec u} f_{X|U}(x^* \mid u')d\mu_U(u') + f_{X|U}(x^* \mid u)P(u)$ and $f_Y(y^*) = \int_{u' \prec u} f_{Y|U}(y^* \mid$

$u')d\mu_U(u') + f_{Y|U}(y^* \mid u)P(u)$ follow. Using all these equalities, we get

$$
\begin{aligned}
f_{X|U}(x \mid u) &= \frac{f_{X|U}(x \mid u) f_{Y|U}(y^* \mid u) P(u)}{f_{Y|U}(y^* \mid u) P(u)} \\
&= \frac{f_{XY}(x, y^*) - \int_{u' \prec u} f_{X|U}(x \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u')}{f_Y(y^*) - \int_{u' \prec u} f_{Y|U}(y^* \mid u') d\mu_U(u')} \qquad \text{for all } x \in \mathcal{X} \\
f_{Y|U}(y \mid u) &= \frac{f_{X|U}(x^* \mid u) f_{Y|U}(y \mid u) P(u)}{f_{X|U}(x^* \mid u) P(u)} \\
&= \frac{f_{XY}(x^*, y) - \int_{u' \prec u} f_{X|U}(x^* \mid u') f_{Y|U}(y \mid u') d\mu_U(u')}{f_X(x^*) - \int_{u' \prec u} f_{X|U}(x^* \mid u') d\mu_U(u')} \qquad \text{for all } y \in \mathcal{Y} \\
P(u) &= \frac{f_{X|U}(x^* \mid u) f_{Y|U}(y^* \mid u) P(u)^2}{f_{X|U}(x^* \mid u) f_{Y|U}(y^* \mid u) P(u)} \\
&= \frac{\left[f_X(x^*) - \int_{u' \prec u} f_{X|U}(x^* \mid u') d\mu_U(u')\right] \left[f_Y(y^*) - \int_{u' \prec u} f_{Y|U}(y^* \mid u') d\mu_U(u')\right]}{f_{XY}(x^*, y^*) - \int_{u' \prec u} f_{X|U}(x^* \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u')}
\end{aligned}
$$

Note that the right-hand sides of these equalities consist of the observed data f_{XY}, or are transfinite-inductively assumed to be known. Therefore, $(f_{X|U}(\,\cdot \mid u), f_{Y|U}(\,\cdot \mid u), P(u))$ is identified.

(ii) Let $P(u) = 0$. By Restriction 1, we can choose $x^* \in \mathcal{X}(u)$ and $y^* \in \mathcal{Y}(u)$. Define the set $E_r(u) = \{u' \in \mathcal{U} \mid u' \preceq u\} \cap B(u, r)$ which is μ_U-measurable by Restriction 1 (iii). Moreover, we have $\mu_U(E_r(u)) > 0$ for each $r > 0$ by Restriction 1 (iv). Let $E_r^c(u) = \{u' \in \mathcal{U} \mid u' \preceq u\} \backslash B(u, r)$, which is not exactly the complement of $E_r(u)$, but is also μ_U-measurable by Restriction 1 (iii). By the choice of x^*, we have $f_{XY}(x^*, y) = \int_{E_r^c(u)} f_{X|U}(x^* \mid u') f_{Y|U}(y \mid u') d\mu_U(u') + \int_{E_r(u)} f_{X|U}(x^* \mid u') f_{Y|U}(y \mid u') d\mu_U(u')$ for all $y \in \mathcal{Y}$. Similarly, $f_{XY}(x, y^*) = \int_{E_r^c(u)} f_{X|U}(x \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u') + \int_{E_r(u)} f_{X|U}(x \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u')$ holds for all $x \in \mathcal{X}$. In particular, $f_{XY}(x^*, y^*) = \int_{E_r^c(u)} f_{X|U}(x^* \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u') + \int_{E_r(u)} f_{X|U}(x^* \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u')$. Moreover, $f_X(x^*) = \int_{E_r^c(u)} f_{X|U}(x^* \mid u') d\mu_U(u') + \int_{E_r(u)} f_{X|U}(x^* \mid u') d\mu_U(u')$ and $f_Y(y^*) = \int_{E_r^c(u)} f_{Y|U}(y^* \mid u') d\mu_U(u') + \int_{E_r(u)} f_{Y|U}(y^* \mid u') d\mu_U(u')$ follow. Using all these equalities together with the Lebesgue Differentiation Theorem, we get

$$
\begin{aligned}
f_{X|U}(x \mid u) &= \frac{f_{X|U}(x \mid u) f_{Y|U}(y^* \mid u) f_U(u)}{f_{Y|U}(y^* \mid u) f_U(u)} \\
&= \lim_{r \to 0} \frac{f_{XY}(x, y^*) - \int_{E_r^c(u)} f_{X|U}(x \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u')}{f_Y(y^*) - \int_{E_r^c(u)} f_{Y|U}(y^* \mid u') d\mu_U(u')} \qquad \text{for all } x \in \mathcal{X} \\
f_{Y|U}(y \mid u) &= \frac{f_{X|U}(x^* \mid u) f_{Y|U}(y \mid u) f_U(u)}{f_{X|U}(x^* \mid u) f_U(u)} \\
&= \lim_{r \to 0} \frac{f_{XY}(x^*, y) - \int_{E_r^c(u)} f_{X|U}(x^* \mid u') f_{Y|U}(y \mid u') d\mu_U(u')}{f_X(x^*) - \int_{E_r^c(u)} f_{X|U}(x^* \mid u') d\mu_U(u')} \qquad \text{for all } y \in \mathcal{Y} \\
f_U(u) &= \frac{f_{X|U}(x^* \mid u) f_{Y|U}(y^* \mid u) f_U(u)^2}{f_{X|U}(x^* \mid u) f_{Y|U}(y^* \mid u) f_U(u)} \\
&= \lim_{r \to 0} \frac{\left[f_X(x^*) - \int_{E_r^c(u)} f_{X|U}(x^* \mid u') d\mu_U(u')\right] \left[f_Y(y^*) - \int_{E_r^c(u)} f_{Y|U}(y^* \mid u') d\mu_U(u')\right]}{\mu_U(E_r(u)) \left[f_{XY}(x^*, y^*) - \int_{E_r^c(u)} f_{X|U}(x^* \mid u') f_{Y|U}(y^* \mid u') d\mu_U(u')\right]}
\end{aligned}
$$

Note that the right-hand sides of these equalities consist of the observed data f_{XY}, or are transfinite-inductively assumed to be known. Therefore, $(f_{X|U}(\cdot \mid u), f_{Y|U}(\cdot \mid u), f_U(u))$ is identified. □

A.4 An Example of Continuous U

In this section, we show an example of the model with continuously distributed U satisfying Restriction 1. Let $\mathcal{U} = [0, 1]$, and let $\mathcal{C}$ be any well-ordered Cantor set on $(\mathbb{R}, <)$. Because $\mathcal{C}$ is a Cantor set on $\mathbb{R}$, there exists a bijection β from $\mathcal{C}$ to $\mathcal{U}$. Define $\prec$ on $\mathcal{U}$ by $\beta(c) \prec \beta(c')$ if and only if $c < c'$. Because $(\mathcal{C}, <)$ is well-ordered, so is $(\mathcal{U}, \prec)$ by construction. Define any strictly increasing functions $\theta_x : \mathcal{C} \to \mathbb{R}$ and $\theta_y : \mathcal{C} \to \mathbb{R}$. Construct the conditional distributions of X given U and Y given U so that $\inf \text{support}(f_{X|U}(\cdot \mid u)) = \theta_x \circ \beta^{-1}(u)$ and $\inf \text{support}(f_{Y|U}(\cdot \mid u)) = \theta_y \circ \beta^{-1}(u)$ are satisfied for each $u \in \mathcal{U}$. Since $\mathcal{C}$ is a Cantor set on $\mathbb{R}$, we then have $\inf \text{support}(f_{X|U}(\cdot \mid u)) = \theta_x \circ \beta^{-1}(u) < \inf \{\theta_x \circ \beta^{-1}(u') \mid \beta^{-1}(u) < \beta^{-1}(u')\} = \inf \{\theta_x \circ \beta^{-1}(u') \mid u \prec u'\} = \inf \left\{\inf \text{support}(f_{X|U}(\cdot \mid u')) \mid u \prec u'\right\}$ so that Restriction 1 (i) is satisfied. Similarly, we get $\inf \text{support}(f_{Y|U}(\cdot \mid u)) < \inf \left\{\inf \text{support}(f_{Y|U}(\cdot \mid u')) \mid u \prec u'\right\}$ and Restriction 1 (ii) is satisfied. Let $\mathcal{T}_<$ and $\mathcal{T}_\prec$ be the $<$-order topology and $\preceq$-order topology on $\mathcal{U}$, respectively. If the probability measure μ_U is defined on the sigma algebra generated by $\mathcal{T}_< \cup \mathcal{T}_\prec$, then Restriction 1 (iii) is satisfied. Lastly, let $\mu_U(B) = m(B)$ for every $B \in \sigma(\mathcal{T}_<)$, and let $\mu_U(B) = 0$ for every $B \in \sigma(\mathcal{T}_< \cup \mathcal{T}_\prec)$ such that there exist $B' \in \sigma(\mathcal{T}_<)$ with $B \subset B'$ and $m(B') = 0$, where m is the Lebesgue measure. This construction well-defines μ_U on $\sigma(\mathcal{T}_< \cup \mathcal{T}_\prec)$, and moreover satisfies Restriction 1 (iv).

A.5 Proof of Proposition 1

Proof. First, note that $\preceq$ is induced by the well-ordered set $(\{1, \cdots, J\}, \leqslant)$, where $\leqslant$ is a well-order by the Well-Ordering Principle. Therefore, the induced order $\preceq$ is also a well-order. To show that Restriction 1 (i) is satisfied, assume without loss of generality that $\inf\big(f_{X|U}(\,\cdot \mid u_j)\big)$ is increasing in j as in Assumption 1 (i). Similar arguments follow in the other case. Let $u_j \in \mathcal{U}$ and u_{j+1} be the $\preceq$-successor ordinal of u_j. Let $s_j = \inf\big(f_{X|U}(\,\cdot \mid u_j)\big)$ and $s_{j+1} = \inf\big(f_{X|U}(\,\cdot \mid u_{j+1})\big)$, where $s_j < s_{j+1}$ holds by Assumption 1 (i). By the definition of s_j as the infimum of the set $\big(f_{X|U}(\,\cdot \mid u_j)\big)$, there exists x_j such that $s_j \leqslant x_j < s_{j+1}$. By Assumption 1 (i), x_j is not an element of $\big(f_{X|U}(\,\cdot \mid u_k)\big)$ for all $k > j$. Therefore, Restriction 1 (i) is satisfied. Similar lines of argument show that Assumption 1 (ii) implies that Restriction 1 (ii) is satisfied. With discrete $\mathcal{U}$, we have $\{u' \in \mathcal{U} \mid u' \preceq u\} \in \sigma(U)$ and $\{u' \in \mathcal{U} \mid u' \prec u\} \in \sigma(U)$ for each $u \in \mathcal{U}$. Finally, because $u \in \mathcal{U}$ if and only if $u \in \text{support}(\mu_U)$ by the definition of $\mathcal{U}$ and U is discrete, we have $0 < \mu_U(\{u\}) \leqslant \mu_U(\{u' \in \mathcal{U} \mid u' \leqslant u\} \cap B(u, r))$ holds for each $u \in \mathcal{U}$ for each $r > 0$, due to the monotonicity of the probability measure μ_U. □

A.6 Proof of Proposition 3

Proof. By Assumption 2 (i), $\inf(f_{X|U}(\cdot \mid u) \geqslant u$ and $\inf(f_{Y|U}(\cdot \mid u) \geqslant u$ for each $u \in \mathcal{U}$. On the other hand, by Assumption 2 (ii), $\inf(f_{X|U}(\cdot \mid u) \leqslant u$ and $\inf(f_{Y|U}(\cdot \mid u) \leqslant u$ for each $u \in \mathcal{U}$. Therefore, $\inf(f_{X|U}(\cdot \mid u) = \inf(f_{Y|U}(\cdot \mid u) = u$ for each $u \in \mathcal{U}$. It then follows that $\inf(f_{X|U}(\cdot \mid u)$ and $\inf(f_{Y|U}(\cdot \mid u)$ are increasing in u with respect to the ordering $\preceq$ on $\mathcal{U}$

defined by $u \prec u'$ if and only if $u < u'$. The second result follows by Proposition 2. $\square$

A.7 Proof of Lemma 4

Proof. Assume without loss of generality that $\inf\left(f_{X|U}(\,\cdot \mid u_j)\right)$ and $\inf\left(f_{Y|U}(\,\cdot \mid u_j)\right)$ are increasing in j as in Assumption 1 (i) and (ii). Similar arguments follow in the other cases. We use the short-hand notations $s_j = \inf\{x \in \mathcal{X} \mid f_{X|U}(x \mid u_j) > 0\}$ and $t_j = \inf\{y \in \mathcal{Y} \mid f_{Y|U}(y \mid u_j) > 0\}$ for each $j = 1, \cdots, J$. Notice that $s_1 = \inf \mathcal{X}$ and $t_1 = \inf \mathcal{Y}$ under the current assumption. To prove the lemma, we want to find s_2 and t_2. To this end, we claim that the equality $f_{XY}(x_1, y_1) f_{XY}(x_2, y_2) = f_{XY}(x_1, y_2) f_{XY}(x_2, y_1)$ holds for all $x_1, x_2 \in [s_1, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$ if and only if $s \leqslant s_2$.

Suppose that $s \leqslant s_2$ holds. For all $x \in [s_1, s) \cap \mathcal{X}$, $f_{X|U}(x \mid u_1) > 0$ but $f_{X|U}(x \mid u_j) = 0$ for all $j = 2, \cdots, J$ by Assumption 1. Therefore, $f_{XY}(x, y) = \sum_j f_{X|U}(x \mid u_j) \cdot f_U(u_j) \cdot f_{Y|U}(y \mid u_j) = f_{X|U}(x \mid u_1) \cdot f_U(u_1) \cdot f_{Y|U}(y \mid u_1)$ for all $x \in [s_1, s) \cap \mathcal{X}$ and all $y \in \mathcal{Y}$. It follows that $f_{XY}(x_1, y_1) f_{XY}(x_2, y_2) = f_{X|U}(x_1 \mid u_1) f_{X|U}(x_2 \mid u_1) f_{Y|U}(y_1 \mid u_1) f_{Y|U}(y_2 \mid u_1) f_U(u_1)^2 = f_{XY}(x_1, y_2) f_{XY}(x_2, y_1)$ holds for all $x_1, x_2 \in [s_1, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$.

Conversely, suppose that $s > s_2$ holds. By definition of s_2 as the infimum of the set $\{x \in \mathcal{X} \mid f_{X|U}(x \mid u_2) > 0\}$, there exists $x_1 \in [s_2, s) \cap \mathcal{X}$ such that $f_{X|U}(x_1 \mid u_2) > 0$. Because of Assumption 1, we can choose such x_1 so that $x_1 < s_3$. Let $x_2 \in [s_1, s_2) \cap \mathcal{X} \subset [s_1, s) \cap \mathcal{X}$, $y_1 \in \{y \in \mathcal{Y} \mid f_{Y|U}(y \mid u_2) > 0\} \subset \mathcal{Y}$, and $y_2 \in [t_1, t_2) \cap \mathcal{Y} \subset \mathcal{Y}$. Note that $f_{XY}(x_1, y_1) f_{XY}(x_2, y_2) = f_{X|U}(x_1 \mid u_1) f_{X|U}(x_2 \mid u_1) f_{Y|U}(y_1 \mid u_1) f_{Y|U}(y_2 \mid u_1) f_U(u_1)^2 + f_{X|U}(x_1 \mid u_2) f_{X|U}(x_2 \mid u_1) f_{Y|U}(y_1 \mid u_2) f_{Y|U}(y_2 \mid u_1) f_U(u_1) f_U(u_2) \neq f_{X|U}(x_1 \mid u_1) f_{X|U}(x_2 \mid u_1) f_{Y|U}(y_1 \mid u_1) f_{Y|U}(y_2 \mid u_1) f_U(u_1)^2$ because $f_{X|U}(x_1 \mid u_2) f_U(u_2) f_{Y|U}(y_1 \mid u_2) \neq 0$ for our choice of x_1 and y_1 as well as $f_{X|U}(x_2 \mid u_1) f_U(u_1) f_{Y|U}(y_2 \mid u_1) \neq 0$ for our choice of x_2 and y_2. On the other hand, $f_{XY}(x_1, y_2) f_{XY}(x_2, y_1) = f_{X|U}(x_1 \mid u_1) f_{X|U}(x_2 \mid u_1) f_{Y|U}(y_1 \mid u_1) f_{Y|U}(y_2 \mid u_1) f_U(u_1)^2$ for our choice of x_2 and y_2 under Assumption 1. It follows that $f_{XY}(x_1, y_1) f_{XY}(x_2, y_2) \neq f_{XY}(x_1, y_2) f_{XY}(x_2, y_1)$ for these $x_1, x_2 \in [s_1, s) \cap \mathcal{X}$ and $y_1, y_2 \in \mathcal{Y}$. This shows that the equality

$f_{XY}(x_1, y_1)f_{XY}(x_2, y_2) = f_{XY}(x_1, y_2)f_{XY}(x_2, y_1)$ need not hold for all $x_1, x_2 \in [s_1, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$ when $s > s_2$.

It therefore follows that the equality $f_{XY}(x_1, y_1)f_{XY}(x_2, y_2) = f_{XY}(x_1, y_2)f_{XY}(x_2, y_1)$ holds for all $x_1, x_2 \in [s_1, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$ if and only if $s \leqslant s_2$. This implies that s_2 can be characterized by

$$\begin{aligned} s_2 \;=\; & \inf\{s \in \mathcal{X} \,|\, \exists\; x_1, x_2 \in [s_1, s) \cap \mathcal{X} \text{ and } y_1, y_2 \in \mathcal{Y} \;\text{ s.t.} \\ & f_{XY}(x_1, y_1)f_{XY}(x_2, y_2) \neq f_{XY}(x_1, y_2)f_{XY}(x_2, y_1)\} \end{aligned}$$

Similar lines of argument show that t_2 can be characterized by

$$\begin{aligned} t_2 \;=\; & \inf\{t \in \mathcal{Y} \,|\, \exists\; x_1, x_2 \in \mathcal{X} \text{ and } y_1, y_2 \in [t_1, t) \cap \mathcal{Y} \;\text{ s.t.} \\ & f_{XY}(x_1, y_1)f_{XY}(x_2, y_2) \neq f_{XY}(x_1, y_2)f_{XY}(x_2, y_1)\}. \end{aligned}$$

Notice that every component in the right-hand sides of these equalities can be directly identified by the observed data f_{XY}. Hence, under Assumption 1, we identify $\mathcal{X}(u_1)$ and $\mathcal{Y}(u_1)$ by $[s_1, s_2) \cap \mathcal{X}$ and $[t_1, t_2) \cap \mathcal{Y}$, respectively. □

A.8 Proof of Lemma 5

Proof. We use the short-hand notations $s_k = \inf\{x \in \mathcal{X} \mid f_{X|U}(x \mid u_k) > 0\}$ and $t_k = \inf\{y \in \mathcal{Y} \mid f_{Y|U}(y \mid u_k) > 0\}$ for each $k = 1, \cdots, J$. Assume without loss of generality that $\inf\big(f_{X|U}(\,\cdot \mid u_k)\big)$ and $\inf\big(f_{Y|U}(\,\cdot \mid u_k)\big)$ are increasing in k as in Assumption 1 (i) and (ii). In this case, s_j and t_j are known by the inductive assumption. Similar arguments follow in the other cases. To prove the lemma, we want to find s_{j+1} and t_{j+1}. To this end, we claim that the equality $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k)f_U(u_k)f_{Y|U}(y_1 \mid u_k)][f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k)f_U(u_k)f_{Y|U}(y_2 \mid u_k)] = f_{X|U}(x_1 \mid u_j)f_{X|U}(x_2 \mid u_j)f_{Y|U}(y_1 \mid u_j)f_{Y|U}(y_2 \mid u_j)f_U(u_j)^2 = [f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k)f_U(u_k)f_{Y|U}(y_2 \mid u_k)][f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k)f_U(u_k)f_{Y|U}(y_1 \mid u_k)]$ holds for all $x_1, x_2 \in [s_j, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$ if and only if $s \leqslant s_{j+1}$.

Suppose that $s \leqslant s_{j+1}$ holds. For all $x \in [s_j, s) \cap \mathcal{X}$, $f_{X|U}(x \mid u_k) = 0$ for all $k = j+1, \cdots, J$ by Assumption 1. Therefore, $f_{XY}(x, y) = \sum_k f_{X|U}(x \mid u_k) \cdot f_U(u_k) \cdot f_{Y|U}(y \mid u_k) = \sum_{k \leqslant j} f_{X|U}(x \mid$

$u_k) \cdot f_U(u_k) \cdot f_{Y|U}(y \mid u_k)$ for all $x \in [s_j, s) \cap \mathcal{X}$ and all $y \in \mathcal{Y}$. It follows that $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)][f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] = f_{X|U}(x_1 \mid u_j) f_{X|U}(x_2 \mid u_j) f_{Y|U}(y_1 \mid u_j) f_{Y|U}(y_2 \mid u_j) f_U(u_j)^2 = [f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)][f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]$ holds for all $x_1, x_2 \in [s_j, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$.

Conversely, suppose that $s > s_{j+1}$ holds. By definition of s_{j+1} as the infimum of the set $\{x \in \mathcal{X} \mid f_{X|U}(x \mid u_{j+1}) > 0\}$, there exists $x_1 \in [s_{j+1}, s) \cap \mathcal{X}$ such that $f_{X|U}(x_1 \mid u_{j+1}) > 0$. Because of Assumption 1, we can choose such x_1 so that $x_1 < s_{j+2}$. Let $y_1 \in \{y \in \mathcal{Y} \mid f_{Y|U}(y \mid u_{j+1}) > 0\} \subset \mathcal{Y}$. Also let $x_2 \in [s_j, s_{j+1}) \cap \mathcal{X} \subset [s_j, s_{j+1}) \cap \mathcal{X}$ and $y_2 \in [t_j, t_{j+1}) \cap \mathcal{Y} \subset \mathcal{Y}$ be such that $f_{X|U}(x_2 \mid u_j) > 0$ and $f_{Y|U}(y_2 \mid u_j) > 0$, where such x_2 and y_2 are guaranteed to exist by the definitions of s_j and t_j as the infima of the sets $\{x \in \mathcal{X} \mid f_{X|U}(x \mid u_j) > 0\}$ and $\{y \in \mathcal{Y} \mid f_{Y|U}(y \mid u_j) > 0\}$, respectively. Given these choices, note that $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)][f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] = f_{X|U}(x_1 \mid u_j) f_{X|U}(x_2 \mid u_j) f_{Y|U}(y_1 \mid u_j) f_{Y|U}(y_2 \mid u_j) f_U(u_j)^2 + f_{X|U}(x_1 \mid u_{j+1}) f_{X|U}(x_2 \mid u_j) f_{Y|U}(y_1 \mid u_{j+1}) f_{Y|U}(y_2 \mid u_j) f_U(u_j) f_U(u_{j+1}) \neq f_{X|U}(x_1 \mid u_j) f_{X|U}(x_2 \mid u_j) f_{Y|U}(y_1 \mid u_j) f_{Y|U}(y_2 \mid u_j) f_U(u_j)^2$ because $f_{X|U}(x_1 \mid u_{j+1}) f_U(u_{j+1}) f_{Y|U}(y_1 \mid u_{j+1}) \neq 0$ for our choice of x_1 and y_1, as well as $f_{X|U}(x_2 \mid u_j) f_U(u_j) f_{Y|U}(y_2 \mid u_j) \neq 0$ for our choice of x_2 and y_2. On the other hand, $[f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)][f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)] = f_{X|U}(x_1 \mid u_j) f_{X|U}(x_2 \mid u_j) f_{Y|U}(y_1 \mid u_j) f_{Y|U}(y_2 \mid u_j) f_U(u_j)^2$ for our choice of x_2 and y_2 under Assumption 1. As a consequence, we have $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)][f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \neq [f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)][f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]$ for these $x_1, x_2 \in [s_j, s) \cap \mathcal{X}$ and $y_1, y_2 \in \mathcal{Y}$. This shows that $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)][f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] = [f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)][f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]$ need not hold for all $x_1, x_2 \in [s_j, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$

when $s > s_{j+1}$.

It therefore follows that the equality $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)][f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] = f_{X|U}(x_1 \mid u_j) f_{X|U}(x_2 \mid u_j) f_{Y|U}(y_1 \mid u_j) f_{Y|U}(y_2 \mid u_j) f_U(u_j)^2 = [f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)][f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]$ holds for all $x_1, x_2 \in [s_j, s) \cap \mathcal{X}$ and all $y_1, y_2 \in \mathcal{Y}$ if and only if $s \leqslant s_{j+1}$. This implies that s_{j+1} can be characterized by

$$
\begin{aligned}
s_{j+1} = \inf\{s \in \mathcal{X} \mid \exists\, x_1, x_2 \in [s_j, s) \cap \mathcal{X} \text{ and } y_1, y_2 \in \mathcal{Y} \text{ s.t.} \\
[f_{XY}(x_1, y_1) - \textstyle\sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)] \times \\
[f_{XY}(x_2, y_2) - \textstyle\sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \neq \\
[f_{XY}(x_1, y_2) - \textstyle\sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \times \\
[f_{XY}(x_2, y_1) - \textstyle\sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]\}
\end{aligned}
$$

Similar lines of argument show that t_{j+1} can be characterized by

$$
\begin{aligned}
t_{j+1} = \inf\{t \in \mathcal{Y} \mid \exists\, x_1, x_2 \in \mathcal{X} \text{ and } y_1, y_2 \in [t_j, t) \cap \mathcal{Y} \text{ s.t.} \\
[f_{XY}(x_1, y_1) - \textstyle\sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)] \times \\
[f_{XY}(x_2, y_2) - \textstyle\sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \neq \\
[f_{XY}(x_1, y_2) - \textstyle\sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \times \\
[f_{XY}(x_2, y_1) - \textstyle\sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]\}
\end{aligned}
$$

Notice that every component in the right-hand sides of these equalities can be directly identified by the observed data f_{XY} or known by the inductive assumption. Therefore, s_{j+1} and t_{j+1} are identified, and we have $\mathcal{X}(u_j) \subset [s_j, s_{j+1})$ and $\mathcal{Y}(u_j) \subset [t_j, t_{j+1})$.

To further pin down $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$, it remains to find the subsets of $[s_j, s_{j+1})$ and $[t_j, t_{j+1})$ on which $f_{X|U}(\,\cdot \mid u_j) > 0$ and $f_{Y|U}(\,\cdot \mid u_j) > 0$, respectively. Consider the sets

$$
\begin{aligned}
\mathcal{X}_j &= \left\{ x \in \mathcal{X} \mid f_X(x) - \sum_{k<j} f_{X|U}(x \mid u_k) f_U(u_k) > 0 \right\} \qquad \text{and} \\
\mathcal{Y}_j &= \left\{ y \in \mathcal{Y} \mid f_Y(y) - \sum_{k<j} f_{Y|U}(x \mid u_k) f_U(u_k) > 0 \right\}
\end{aligned}
$$

Note that every component in the right-hand sides of these equalities can be directly identified by the observed data f_{XY} or known by the inductive assumption. Therefore, these sets $\mathcal{X}_j$ and $\mathcal{Y}_j$ are identified. We now claim that $\mathcal{X}(u_j) = \mathcal{X}_j \cap [s_j, s_{j+1})$, where the right-hand side is identified. First, $\mathcal{X}(u_j) \subset [s_j, s_{j+1})$ was already claimed. Furthermore, if $x \in \mathcal{X}(u_j) \cap [s_j, s_{j+1})$, then $f_X(x) - \sum_{k<j} f_{X|U}(x \mid u_k) f_U(u_k) = f_{X|U}(x \mid u_j) > 0$ so that $x \in \mathcal{X}_j$ holds. Conversely, let $x \in \mathcal{X}_j \cap [s_j, s_{j+1})$. Then, we have $x \in \mathcal{X}_j \cap [s_j, s_{j+1}) \subset \mathcal{X}_j \backslash [s_{j+1}, \infty) \subset \text{support}(f_{X|U}(\,\cdot \mid u_k)) \backslash \cup_{j<k} \text{support}(f_{X|U}(\,\cdot \mid u_k))$, thus showing that $x \in \mathcal{X}(u_j)$. Similarly, we can show $\mathcal{Y}(u_j) = \mathcal{X}_j \cap [t_j, t_{j+1})$ where the right-hand side is identified. □

A.9 Proof of Proposition 4

Proof. We define an order relation $\prec$ on $\mathcal{U}$ in the following manner.

$$\text{If } u \in \mathcal{D}^c \text{ and } u' = \max\{u'' \in \mathcal{D} \mid u'' < u\}, \text{ then } u \prec u'.$$

$$\text{Otherwise, } u < u' \iff u \prec u'.$$

The induced order relation $\preceq$ can be shown to be a well-order on $\mathcal{U}$.

Let $u \in \mathcal{D}$. If $\{u' \in \mathcal{D} \mid u' > u\} = \emptyset$, then there exist no element $u' \in \mathcal{U}$ for which $u \prec u'$ holds due to our definition of $\prec$, and thus $u \notin \text{support}(f_{X|U}(\,\cdot \mid u'))$ trivially holds for this u'. Next, assume that $\{u' \in \mathcal{D} \mid u' > u\} \neq \emptyset$, and let $u_+ = \min\{u' \in \mathcal{D} \mid u' > u\}$. If $u \prec u'$, then we have $u_+ \leqslant u'$ by our definition of $\prec$. Assumption 3 (ii) then implies $u \notin \text{support}(f_{X|U}(\,\cdot \mid u'))$. Therefore, Restriction 1 (i) follows for this u by Assumption 3 (iii).

Let $u \in \mathcal{D}^c$. If $u \prec u'$, then we have $u' = \max\{u'' \in \mathcal{D} \mid u'' < u\}$ or $u < u'$ by our definition of $\prec$. If the former is the case, then Assumption 3 (i) implies $u \notin \text{support}(f_{X|U}(\,\cdot \mid u'))$. If the latter is the case, then Assumption 3 (i) and (ii) together imply $u \notin \text{support}(f_{X|U}(\,\cdot \mid u'))$. Therefore, Restriction 1 (i) follows in both cases for this u by Assumption 3 (iii).

The above two paragraphs show that Restriction 1 (i) is satisfied by Assumption 3. Similar lines of argument show that Restriction 1 (ii) is satisfied. Because U is discrete, $\{u' \in \mathcal{U} \mid$

$u' \preceq u\} \in \sigma(U)$ and $\{u' \in \mathcal{U} \mid u' \prec u\} \in \sigma(U)$, so Restriction 1 (iii) trivially holds. Finally, Restriction 1 (iii) follows from the monotonicity of probability measures, i.e., $\mu_U(\{u' \in \mathcal{U} \mid u' \preceq u\} \cap B(u,r)) \geqslant \mu_U(\{u\}) > 0$ for all $u \in \mathcal{U}$. □

A.10 Proof of Proposition 5

Proof. Consider the estimators, T_{xu_1}, T_{yu_1}, and T_{u_1}, of $f_{X|U}(x \mid u_1)$, $f_{Y|U}(y \mid u_1)$, and $f_U(u_1)$, defined by

$$\begin{aligned}
\hat{f}_{X|U}(x \mid u_1) &= T_{xu_1}\left(f_X^N, f_Y^N, f_{XY}^N\right) = \frac{f_{XY}^N(x, y_1)}{f_Y^N(y_1)} \\
\hat{f}_{Y|U}(y \mid u_1) &= T_{yu_1}\left(f_X^N, f_Y^N, f_{XY}^N\right) = \frac{f_{XY}^N(x_1, y)}{f_X^N(x_1)} \\
\hat{f}_U(u_1) &= T_{u_1}\left(f_X^N, f_Y^N, f_{XY}^N\right) = \frac{f_X^N(x_1) f_Y^N(y_1)}{f_{XY}^N(x_1, y_1)},
\end{aligned}$$

respectively. With some abuse of notations, the derivatives of these estimators are written by

$$\begin{aligned}
\nabla T_{xu_1}\left(f_X^N, f_Y^N, f_{XY}^N\right) &= \begin{bmatrix} 0 & -\frac{f_{XY}^N(x, y_1)}{f_Y^N(y_1)^2} & \frac{1}{f_Y^N(y_1)} \end{bmatrix}^T \\
\nabla T_{yu_1}\left(f_X^N, f_Y^N, f_{XY}^N\right) &= \begin{bmatrix} -\frac{f_{XY}^N(x_1, y)}{f_X^N(x_1)^2} & 0 & \frac{1}{f_X^N(x_1)} \end{bmatrix}^T \\
\nabla T_{u_1}\left(f_X^N, f_Y^N, f_{XY}^N\right) &= \begin{bmatrix} \frac{f_Y^N(y_1)}{f_{XY}^N(x_1, y_1)} & \frac{f_X^N(x_1)}{f_{XY}^N(x_1, y_1)} & -\frac{f_X^N(x_1) f_Y^N(y_1)}{f_{XY}^N(x_1, y_1)^2} \end{bmatrix}^T
\end{aligned}$$

On the other hand, we have the asymptotic normality

$$\sqrt{N}\begin{pmatrix} f_X^N(x) - f_X(x) \\ f_Y^N(y) - f_Y(y) \\ f_{XY}^N(x,y) - f_{XY}(x,y) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} f_X(x)[1 - f_X(x)] & f_{XY}(x,y) - f_X(x) f_Y(y) & f_{XY}(x,y)[1 - f_X(x)] \\ & f_Y(y)[1 - f_Y(y)] & f_{XY}(x,y)[1 - f_Y(y)] \\ & & f_{XY}(x,y)[1 - f_{XY}(x,y)] \end{bmatrix}\right)$$

for each (x, y) under i.i.d. sampling. (Note that the bounded $(2+\delta)$-th moment is automatically satisfied in this context.) Let $V(x, y)$ denote the variance matrix in the above asymptotic normal distribution. Applying the delta method yields

$$\sqrt{N}\left(\hat{f}_{X|U}(x \mid u_1) - f_{X|U}(x \mid u_1)\right) \xrightarrow{d} N(0, V_{xu_1})$$

where some calculations give the simplified asymptotic variance formula

$$V_{xu_1} = \nabla T_{xu_1}^T \cdot V(x, y_1) \cdot \nabla T_{xu_1} = \frac{f_{XY}(x, y_1)\left[f_Y(y_1) - f_{XY}(x, y_1)\right]}{f_Y(y_1)^3}.$$

By symmetric arguments, we obtain

$$\sqrt{N}\left(\hat{f}_{Y|U}(y \mid u_1) - f_{Y|U}(y \mid u_1)\right) \xrightarrow{d} N(0, V_{yu_1})$$

where asymptotic variance is given by

$$V_{yu_1} \quad = \quad \nabla T_{yu_1}^T \cdot V(x_1, y) \cdot \nabla T_{yu_1} \quad = \quad \frac{f_{XY}(x_1, y)\left[f_X(x_1) - f_{XY}(x_1, y)\right]}{f_X(x_1)^3}.$$

Lastly, by similar lines of arguments, we can derive

$$\sqrt{N}\left(\hat{f}_U(u_1) - f_U(u_1)\right) \xrightarrow{d} N(0, V_{u_1})$$

where asymptotic variance is given by $V_{u_1} = \nabla T_{u_1}^T \cdot V(x_1, y_1) \cdot \nabla T_{u_1} =$
$\frac{f_X(x_1) f_Y(y_1)\left[(f_X(x_1) - f_{XY}(x_1, y_1))(f_Y(y_1) - f_{XY}(x_1, y_1)) + f_{XY}(x_1, y_1)(f_{XY}(x_1, y_1) - f_X(x_1) f_Y(y_1))\right]}{f_{XY}(x_1, y_1)^3}$

□

References

Arellano, Manuel, and Stephane Bonhomme (2012) "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *Review of Economic Studies,* Vol. 79, pp. 987–1020.

Ashenfelter, Orley, and Alan B. Krueger (1994) "Estimates of the Economic Returns to Schooling from a New Sample of Twins," *American Economic Review,* Vol. 84, pp. 1157–1173.

Behrman, Jere, Paul Taubman, and Terence J. Wales (1977) "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success," in P. Taubman, eds., *Kinometrics: The Determinants of Socioeconomic Success within and between-Families.* Amsterdam: North-Holland, pp. 35–96.

Behrman, Jere R. and Mark R. Rosenzweig (1999) ""Ability" Biases in Schooling Returns and Twins: a Test and New Estimates," *Economics of Education Review,* Vol. 18, pp. 159-167.

Bonhomme, Stephane, and Jean-Mark Robin (2010) "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *Review of Economic Studies,* Vol. 77, pp. 491–533.

Carroll, Raymond J, Xiaohong Chen, and Yingyao Hu (2010) "Identification and Inference in Nonlinear Models Using Two Samples with Nonclassical Measurement Errors," *Journal of Nonparametric Statistics,* Vol. 22, pp. 379–399.

Chen, Xiaohong, Han Hong, and Elie Tamer (2005) "Measurement Error Models with Auxiliary Data," *Review of Economic Studies,* Vol. 72, pp. 343–366.

Chen, Xiaohong, Han Hong, and Tarozzi, Alessandro (2008) "Semiparametric Efficiency in GMM Models with Auxiliary Data," *Annals of Statistics,* Vol. 36, pp. 808–843.

Cunha, Flavio, James Heckman, and Salvador Navarro (2005) "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers,* Vol. 57, pp. 191–261.

D'Haultfoeuille, Xavier and Philippe Février (2010) "Identification of Mixture Models Using Support Variations," Working Paper, CREST.

Evdokimov, Kirill (2010) "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Working Paper, Princeton University.

Hansen, Karsten T., James J. Heckman, and Kathleen J. Mullen (2004) "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics,* Vol. 121, pp. 39–98.

Hu, Yingyao (2008) "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics,* Vol. 144, pp. 27-61.

Hu, Yingyao, and Susanne M. Schennach (2008) "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica,* Vol. 76, pp. 195-216.

Kennan, John, and James R. Walker (2011) "The Effect of Expected Income on Individual Migration Decisions," *Econometrica,* Vol. 79, pp. 211–251.

Kim, Kyoo il, Amil Petrin, and Suyong Song (2013) "Estimating Production Functions When Capital Input Is Measured With Error," Working Paper.

Krasnokutskaya, Elena (2011) "Identification and Estimation of Auction Models with Unobserved Heterogeneity," *Review of Economic Studies,* Vol. 78, pp. 293–327.

Lewbel, Arthur (2007) "Estimation of Average Treatment Effects with Misclassification," *Econometrica,* Vol. 75, pp. 537–551.

Li, Tong (2002) "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models," *Journal of Econometrics,* Vol. 110, pp. 1–26.

Li, Tong, Isabelle Perrigne, and Quang Vuong (2000) "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics,* Vol. 98, pp. 129–161.

Li, Tong, and Quang Vuong (1998) "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis,* Vol. 65, pp. 139–165.

Mahajan, Aprajit (2006) "Identification and Estimation of Single Index Models with Misclassified Regressor," *Econometrica,* Vol. 74, pp. 631–665.

Matzkin, Rosa L. (2003) "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica,* Vol. 71, pp. 1339-1375.

Miller, Paul, Charles Mulvey and Nick Martin (1995) "What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison of Australian and U.S. Findings," American Economic Review, Vol. 85, pp. 586–599.

Rouse, Cecilia E. (1999) "Further Estimates of the Economic Return to Schooling from a New Sample of Twins," *Economics of Education Review,* Vol. 18, pp. 149-157.

Schennach, Susanne (2004) “Estimation of Nonlinear Models with Measurement Error,” *Econometrica,* Vol. 72, pp. 33–75.

Song, Suyong, Susanne M. Schennach, and Halbert White (2012) “Estimating Nonseparable Models with Mismeasured Endogenous Variables,” Working Paper, University of Wisconsin Milwaukee.

Siegel, Paul M., and Robert W. Hodge (1968), “A Causal Approach to the Study of Measurement Error,” in: H M Blalock and A B Blalock, eds., *Methodology in Social Research,* New York: McGraw-Hill, pp. 28–59.