Econometric Theory, **33**, 2017, 955–979.
doi:10.1017/S0266466616000207

IDENTIFICATION OF PAIRED NONSEPARABLE MEASUREMENT ERROR MODELS

YINGYAO HU AND YUYA SASAKI
Johns Hopkins University

This paper studies the paired nonseparable measurement error models, where two measurements, X and Y, are produced by mutually independent unobservables, U, V, and W, through the system, $X = g(U,V)$ and $Y = h(U,W)$. We propose restrictions to identify the marginal distribution of the common component U and the conditional distributions of X and Y given U. Applying this method to twin panel data, we find the following robust reporting patterns for years of education: (1) 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; (2) sibling reports are accurate whenever the true years of education are 12, 14, 16, and 18, which are typical diploma years.

1. INTRODUCTION

We 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 variables U, V, and W are not. For example, we may think of U as the true years of education in which econometricians are interested but do not observe in the data. Instead, we observe self reports X and sibling reports Y of U. The nonseparable errors V and W are nonadditive factors of self- and sibling-reporting errors, respectively. Throughout the main text, we focus on the case where U is finitely supported—see Section B.5 in the online appendix for a general case.

Under the stated independence condition, we can represent the model (1.1) by the triple $(F_{X|U}, F_{Y|U}, F_U)$ of conditional and marginal distribution functions.[1] For observed variables X and Y, we assume their conditional distributions given

We received helpful comments from the co-editor, Arthur Lewbel, four anonymous referees, and seminar participants at Ohio State University, Greater New York Metropolitan Area Econometrics Colloquium 2012, AMES 2013, and ESEM 2013. We thank Peter C.B. Phillips and Elizabeth Frankfield for their useful editorial input to our paper. The usual disclaimer applies. Address correspondence to Yingyao Hu and Yuya Sasaki, Department of Economics, Johns Hopkins University, Wyman Park Building 544E, 3400 N. Charles St, Baltimore, MD 21218; e-mail: yhu@jhu.edu, sasaki@jhu.edu.

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). Let f_U denote the marginal pmf of U, the distribution of which is assumed to be finitely supported. 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 $(f_{X|U}, f_{Y|U}, f_U)$ is identified from the observed joint distribution $f_{X,Y}$.

Our identification strategy works in the following manner. We order the support of U as $\mathcal{U} = \{u_1, \ldots, u_J\}$ from u_1 to u_J. If $x_1 \in \mathcal{X}$ and $y_1 \in \mathcal{Y}$ are the reports produced only by those individuals with u_1 (i.e., $f_{X|U}(x_1 \mid u_1) > 0$ and $f_{Y|U}(y_1 \mid u_1) > 0$ but $f_{X|U}(x_1 \mid u_j) = f_{Y|U}(y_1 \mid u_j) = 0$ for all $j > 1$), then we use this "support exclusion restriction" to identify $f_{X|U}(\,\cdot \mid u_1) > 0$ with y_1 as a control variable and to identify $f_{Y|U}(\,\cdot \mid u_1) > 0$ with x_1 as a control variable. This high-level restriction of the support exclusion can be rationalized by assumptions on economic behaviors. For example, if we assume that the respondent and his sibling have no incentive to report numbers lower than the true years of education, then we obtain $f_{X|U}(u_1 \mid u_j) = f_{Y|U}(u_1 \mid u_j) = 0$ for all $j > 1$. The identification of $f_{X|U}(\,\cdot \mid u_j) > 0$ and $f_{Y|U}(\,\cdot \mid u_j) > 0$ for $j > 1$ follows inductively by similar arguments based on the principle of mathematical induction.

The model (1.1) of our interest 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; Chen, Hu, and Lewbel, 2009; Carroll, Chen and Hu, 2010; D'Haultfoeuille and Février, 2010; Song, Schennach and White, 2012). One of the most closely related 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 are based only on self and sibling reports, and contains neither a third measurement nor an additional instrument. We thus need to relax the data requirements 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. Another one of the most closely related is Chen, Hu, and Lewbel (2009), where they identify a regression model under misclassification without requiring an additional measurement. This setup parallels our data requirement, and the assumptions that they impose on regression functions imply our assumption. In other words, we provide a generalized assumption for this particular setup, although a direct comparison is difficult due to the different support cardinality assumptions about U.

Related to our model (1.1) is the repeated measurement model with additive errors:

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

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

While we do not consider the case of continuous (U, V, W), a large number of econometric papers use this paired additive model or its variants under the setting of continuous random variables.[2] Furthermore, (1.1) can be used to model nonadditive structural functions. For example, consider a production structure $Y = h(U, W)$, 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 errors $\phi(U, V)$. 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 interested in identifying the structural responses of the produced quantity Y to the true unobserved quantity U of factors, i.e., h or $F_{Y|U}$.[3]

This paper is organized as follows. In Section 2, we derive identification of the triple $(f_{X|U}, f_{Y|U}, f_U)$ representing the model (1.1). Two alternative assumptions tailored to our empirical application are proposed in Sections 2.1 and 2.2. In Section 3, we propose an estimation procedure following the identification approach. We apply our method to twin panel data in Section 4, where we analyze self- and sibling-reporting patterns for years of education. Before presenting our conclusions, we discuss an application to regression models in Section 5.

2. IDENTIFICATION

Let the support $\mathcal{U} = \{u_1, \ldots, u_J\}$ of U satisfy the following restriction.

Restriction 1 (The Basic Identifying Restriction)**.**

(i) $\mathcal{X}(u_j) := support\big(f_{X|U}(\,\cdot\mid u_j)\big) \setminus \cup_{j<k}\, support\big(f_{X|U}(\,\cdot\mid u_k)\big) \neq \emptyset$ *for all* $j \in \{1, \ldots, J\}$,

(ii) $\mathcal{Y}(u_j) := support\big(f_{Y|U}(\,\cdot\mid u_j)\big) \setminus \cup_{j<k}\, support\big(f_{Y|U}(\,\cdot\mid u_k)\big) \neq \emptyset$ *for all* $j \in \{1, \ldots, J\}$.

Examples of sufficient conditions for this restriction will be stated as Assumptions 1 and 2 in Section 2.1 and Assumption 3 in Section 2.2. Furthermore, when we apply our model to regression models, Restriction 1 implies the assumptions that papers in the literature (e.g., Chen, Hu, and Lewbel, 2009) impose on regression functions, although a direct comparison is not possible due to the different support cardinality assumptions about U—see Section 5. This restriction can be considered as a support exclusion restriction, where an element $x_j \in \mathcal{X}(u_j)$ excluded from support$\big(f_{X|U}(\,\cdot\mid u_k)\big)$ for all $k > j$ is used as a control variable to identify $f_{Y|U}(\,\cdot\mid u_j)$. Similarly, an element $y_j \in \mathcal{Y}(u_j)$ excluded from support$\big(f_{Y|U}(\,\cdot\mid u_k)\big)$ for all $k > j$ is used as a control variable to identify $f_{X|U}(\,\cdot\mid u_j)$. In this sense, this restriction is also related to the monotonicity restriction often used in the treatment effects literature. The following two auxiliary lemmas provide main devices to prove the identification of

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

$(f_{X|U}, f_{Y|U}, f_U)$ by Restriction 1 through the principle of mathematical induction on the index set $\{1, \ldots, J\}$.

LEMMA 1. *Suppose that Restriction 1 holds for the model (1.1). If sets* $\mathcal{X}(u_1)$ *and* $\mathcal{Y}(u_1)$ *are known, then* $(f_{X|U}(\cdot \mid u_1), f_{Y|U}(\cdot \mid u_1), f_U(u_1))$ *is identified.*

LEMMA 2. *Suppose that Restriction 1 holds for the model (1.1). Let* $1 \leqslant j$ *and* $j+1 \leqslant J$*. If* $(f_{X|U}(\cdot \mid u_k), f_{Y|U}(\cdot \mid u_k), f_U(u_k))$ *is identified for all* $k \leqslant j$ *and if sets* $\mathcal{X}(u_{j+1})$ *and* $\mathcal{Y}(u_{j+1})$ *are known, then* $(f_{X|U}(\cdot \mid u_{j+1}), f_{Y|U}(\cdot \mid u_{j+1}), f_U(u_{j+1}))$ *is identified.*

See Sections A.1 and A.2 in the appendix for proof of Lemmas 1 and 2, respectively. Lemma 1 serves as the base step, and Lemma 2 serves as the inductive step in the principle of mathematical induction. We illustrate how to use these auxiliary lemmas in the following two subsections under alternative lower-level assumptions.

2.1. Monotone Support Boundaries

One special instance to satisfy Restriction 1 is the case of monotone support boundaries of $f_{X|U}$ and $f_{Y|U}$, as formally stated below.

Assumption 1 (Monotone Support Boundaries)**.** The supports of X and Y are bounded, and the following two conditions are satisfied.

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

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

We will later characterize this assumption in terms of reporting patterns in Assumption 2. Furthermore, when we apply our baseline model to the regression analysis, this assumption is implied by the standard assumptions (e.g., monotonicity and independence) used in the literature—see Section 5. The most closely related assumptions used in the literature are a variety of monotonicity assumptions used to allow for unique ordering of eigenvalues in the spectral decomposition approaches when an additional measurement is available (e.g., Hu, 2008). While those existing monotonicity assumptions concern the values of conditional densities or the values of conditional expectations, our monotonicity assumption concerns the support boundaries of conditional distributions. Our monotonicity assumption, together with the support cardinality assumption for U, allows us to identify the model without additional measurements.

PROPOSITION 1. *Assumption 1 for (1.1) implies Restriction 1 for (1.1).*

See Section A.3 for a proof. While this proposition only shows that Assumption 1 is sufficient for nonemptiness of $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ for each $j \in \{1, \ldots, J\}$,

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

this assumption also allows these nonempty sets to be identified. We state this argument as the following two auxiliary lemmas.

LEMMA 3. *Suppose that Assumption 1 holds for (1.1). The sets* $\mathcal{X}(u_1)$ *and* $\mathcal{Y}(u_1)$ *are identified.*

LEMMA 4. *Suppose that Assumption 1 holds for (1.1). If* $(f_{X|U}(\cdot \mid u_k), f_{Y|U}(\cdot \mid u_k), f_U(u_k))$ *is known for each* $k < j$*, then the sets* $\mathcal{X}(u_j)$ *and* $\mathcal{Y}(u_j)$ *are identified.*

See Sections A.4 and A.5 in the appendix for proofs of Lemmas 3 and 4, respectively. With the Lemmas 1–4, we can now identify the triple $(f_{X|U}, f_{Y|U}, f_U)$ through the principle of mathematical induction.

THEOREM 1. *If Assumption 1 holds for (1.1), then*

(i) *the sets* $\mathcal{X}(u_j)$ *and* $\mathcal{Y}(u_j)$ *are identified for each* $j \in \{1, \ldots, J\}$*; and*

(ii) $(f_{X|U}(\cdot \mid u_j), f_{Y|U}(\cdot \mid u_j), f_U(u_j))$ *is identified for each* $j \in \{1, \ldots, J\}$.

Proof. First, note that Restriction 1 is satisfied by Proposition 1. We prove the theorem by the principle of mathematical induction on $\{1, \ldots, J\}$. For the base step, the sets $\mathcal{X}(u_1)$ and $\mathcal{Y}(u_1)$ are identified by Lemma 3. Consequently, $(f_{X|U}(\cdot \mid u_1), f_{Y|U}(\cdot \mid u_1), f_U(u_1))$ is identified by Lemma 1. Now, assume inductively that $(f_{X|U}(\cdot \mid u_k), f_{Y|U}(\cdot \mid u_k), f_U(u_k))$ is identified for each $k < j+1$. Then, the sets $\mathcal{X}(u_{j+1})$ and $\mathcal{Y}(u_{j+1})$ are identified by Lemma 4. Consequently, $(f_{X|U}(\cdot \mid u_{j+1}), f_{Y|U}(\cdot \mid u_{j+1}), f_U(u_{j+1}))$ is identified by Lemma 2. ■

In the special case where $\mathcal{X}$ and $\mathcal{Y}$ are exactly the same sets as $\mathcal{U} = \{u_1, \ldots, u_J\}$, the sets $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ for each $j = \{1, \ldots, J\}$ can be constructed easily without relying on Lemmas 3 and 4. Specifically, if $\inf\left(\operatorname{supp}(f_{X|U}(\cdot \mid u_j))\right)$ is increasing in j as in Assumption 1(i), then the equality $\mathcal{U} = \mathcal{X}$ forces $\mathcal{X}(u_j) = \{u_j\}$ for each $j = \{1, \ldots, J\}$. A similar argument applies to the set $\mathcal{Y}(u_j)$ for each $j = \{1, \ldots, J\}$.

Finally, we discuss the main assumption of the current subsection in the context of our application to years of education. What kind of survey reporting pattern rationalizes the monotone support boundaries of Assumption 1? We propose the following reporting pattern as a sufficient condition for Assumption 1, which in turn is sufficient for Restriction 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 honest individuals exist for each actual year u of education. Under the triangular conditional support imposed by part (i), the requirement (ii) of the positive probabilities of zero errors plays a similar role to the matrix invertibility assumption made in the three-measurements literature (e.g., Hu, 2008). At the expense of assuming this specific invertibility, we improve upon this three-measurements literature by reducing the required number of measurements from three to two. This assumption is consistent with the empirical fact that the self-reporting errors for years of education are likely to be negatively correlated with the true years of education (Siegel and Hodge, 1968). In other words, this negative correlation may well arise when people do not under-report their education, as individuals with low education have more room for over-reporting while individuals with high education have little choice but to report truthfully. Section B.4.1 in the online appendix proposes a choice model where Assumption 2 is rationalized by a utility maximization behavior. Part (i) may be restrictive in applications. Thus, Section 2.2 introduces an alternative assumption to relax this restriction. The following proposition claims that Assumption 2 implies Assumption 1.

PROPOSITION 2. *If Assumption 2 holds, then Assumption 1 holds with the ordering of* $\mathcal{U} = \{u_1, \ldots, u_J\}$ *defined by* $j < k$ *if and only if* $u_j < u_k$.

2.2. Alternative Reporting Patterns

Assumption 2, which entirely prohibits under-reporting of years of education, may be restrictive in applications. Certainly, those individuals having just completed diploma-granting years of education, e.g., $U = 12$, 14, 16, and 18,[4] may have no incentive to under-report their education. On the other hand, the remaining individuals, i.e., those with $U = 13$, 15, and 17, may have an incentive to under-report their education by one year due to the stigma of dropping out before gaining a diploma, or simply by rounding numbers to the diploma-granting year for mnemonic reasons. As such, $\min(\text{support}(X \mid U = 12)) = \min(\text{support}(X \mid U = 13)) = 12$ may result and Assumption 1 can thus fail. In light of this possibility, we propose the following alternative assumption in the current subsection.

Assumption 3 (Dropout and Diploma). Let $\mathcal{D} = \{d_1, \ldots, d_L\} \subset \mathcal{U}$ be a set of diploma-granting years, and $\mathcal{D}^c = \mathcal{U} \backslash \mathcal{D}$ be its complement. The following are true.

(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) 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 they have no incentive to voluntarily lie to avoid the stigma of dropout. Part (ii) states that under-reporting individuals report the years of education associated with the highest diploma that they have received, so they can signal that they did not drop out while only minimally suppressing the years. Part (iii) requires the existence of an honest subpopulation. Section B.4.2 in the online appendix proposes a choice model where Assumption 3 is rationalized by a utility maximization behavior. Under this set of assumptions, the general identifying restriction of Section 2 is satisfied as follows

PROPOSITION 3. *If Assumption 3 holds for (1.1), then $\mathcal{U}$ can be written as an indexed set $\mathcal{U} = \{u_1, \ldots, u_J\}$ such that Restriction 1 is satisfied.*

Proof is given in Section A.7 in the appendix. In view of the proof, we can construct the indices $1, \ldots, J$ for an ordering of the set $\mathcal{U}$ by following the rule:

$$\text{If } u_j \in \mathcal{D}^c \text{ and } u_k = \max\{u \in \mathcal{D} \mid u < u_j\}, \text{ then } j < k. \qquad \text{Otherwise, } u_j < u_k \Longleftrightarrow j < k. \tag{2.1}$$

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.

Example 1

In the US education system, high-school diplomas, associate degrees, bachelor's degrees, and master's degrees are associated with 12, 14, 16, and 18 years of education, respectively. If $\mathcal{D} = \{12, 14, 16, 18\}$ and $\mathcal{D}^c = \{13, 15, 17\}$ for $\mathcal{U} = \{12, 13, 14, 15, 16, 17, 18\}$, then we order $\mathcal{U}$ by $u_1 = 13, u_2 = 12, u_3 = 15, u_4 = 14, u_5 = 17, u_6 = 16$, and $u_7 = 18$ according to (2.1).

Example 2

Associate degrees are less likely to be terminal degrees. If $u = 14$ is removed from $\mathcal{D}$, i.e., $\mathcal{D} = \{12, 16, 18\}$ and $\mathcal{D}^c = \{13, 14, 15, 17\}$, then the rule (2.1) produces the alternative ordering of $\mathcal{U}$ by $u_1 = 13, u_2 = 14, u_3 = 15, u_4 = 12, u_5 = 17, u_6 = 16$, and $u_7 = 18$.

In addition to ensuring their nonemptiness, we can also construct the sets $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ for each u. Specifically, if $\mathcal{X} = \mathcal{Y} = \mathcal{U}$ is the case, then we can see that $\mathcal{X}(u_j) = \mathcal{Y}(u_j) = \{u_j\}$ for each $u_j \in \mathcal{U}$. As the sets $\mathcal{X}(u)$ and $\mathcal{Y}(u)$ are known for each $u \in \mathcal{U}$, we can readily use Lemmas 1 and 2 to identify $(f_{X|U}, f_{Y|U}, f_U)$ through the principle of mathematical induction.

THEOREM 2. *If Assumption 3 holds for (1.1) and a given $\mathcal{D} \subset \mathcal{U}$, then for the rule (2.1):*

(i) *the sets $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ are identified for each $j \in \{1, \ldots, J\}$; and*
(ii) *$(f_{X|U}(\cdot \mid u_j), f_{Y|U}(\cdot \mid u_j), f_U(u_j))$ is identified for each $j \in \{1, \ldots, J\}$.*

Proof. First, note that Restriction 1 is satisfied by Proposition 3. Assumption 3 and the rule (2.1) construct the sets $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ for each $j \in \{1, \ldots, J\}$. Thus, applying the principle of mathematical induction, with Lemma 1 for the base step and Lemma 2 for the inductive step, yields identification of $(f_{X|U}(\cdot \mid u_j), f_{Y|U}(\cdot \mid u_j), f_U(u_j))$ for each $j \in \{1, \ldots, J\}$. ■

2.3. Identifying Formulas

In this section, we display for convenience the inductive identifying formulas obtained in the proofs. For the first element $u_1 \in \mathcal{U}$, the identifying formulas are:

$$f_{X|U}(x \mid u_1) = \frac{f_{XY}(x, y_1)}{f_Y(y_1)} \qquad \text{for all } x \in \mathcal{X}$$

$$f_{Y|U}(y \mid u_1) = \frac{f_{XY}(x_1, y)}{f_X(x_1)} \qquad \text{for all } y \in \mathcal{Y}$$

$$f_U(u_1) = \frac{f_X(x_1) f_Y(y_1)}{f_{XY}(x_1, y_1)}$$

with $x_1 \in \mathcal{X}(u_1)$ and $y_1 \in \mathcal{Y}(u_1)$. Recall that there is an analogy between our identification strategy and the existing control variable approaches. The first one of the above identifying formulas reflects the idea that the excluded variable y_1 plays the role of a control variable for u_1 while varying x. Likewise, the second identifying formula above reflects the idea that the excluded variable x_1 plays the role of a control variable for u_1 while varying y.

For subsequent elements $u_{j+1} \in \mathcal{U}$, the identifying formulas are:

$$f_{X|U}(x \mid u_{j+1}) = \frac{f_{XY}(x, y_{j+1}) - \sum_{k \leqslant j} f_{X|U}(x \mid u_k) f_{Y|U}(y_{j+1} \mid u_k) f_U(u_k)}{f_Y(y_{j+1}) - \sum_{k \leqslant j} f_{Y|U}(y_{j+1} \mid u_k) f_U(u_k)} \qquad \text{for all } x \in \mathcal{X}$$

$$f_{Y|U}(y \mid u_{j+1}) = \frac{f_{XY}(x_{j+1}, y) - \sum_{k \leqslant j} f_{X|U}(x_{j+1} \mid u_k) f_{Y|U}(y \mid u_k) f_U(u_k)}{f_X(x_{j+1}) - \sum_{k \leqslant j} f_{X|U}(x_{j+1} \mid u_k) f_U(u_k)} \qquad \text{for all } y \in \mathcal{Y}$$

$$f_U(u_{j+1}) = \frac{\left[f_X(x_{j+1}) - \sum_{k \leqslant j} f_{X|U}(x_{j+1} \mid u_k) f_U(u_k)\right]\left[f_Y(y_{j+1}) - \sum_{k \leqslant j} f_{Y|U}(y_{j+1} \mid u_k) f_U(u_k)\right]}{f_{XY}(x_{j+1}, y_{j+1}) - \sum_{k \leqslant j} f_{X|U}(x_{j+1} \mid u_k) f_{Y|U}(y_{j+1} \mid u_k) f_U(u_k)}$$

with $x_{j+1} \in \mathcal{X}(u_{j+1})$ and $y_{j+1} \in \mathcal{Y}(u_{j+1})$, where $f_{X|U}(\cdot \mid u_k)$, $f_{Y|U}(\cdot \mid u_k)$ and $f_U(u_k)$ for all $k \leqslant j$ have been inductively identified in previous steps. Furthermore, see Section B.2 in the online appendix for closed-form identifying formulas obtained by successive substitutions of these inductive formulas. As in the previous paragraph, the first identifying formula above reflects the idea that the excluded variable y_{j+1} plays the role of a control variable for u_{j+1} while varying x. Likewise, the second identifying formula above reflects the idea that the excluded variable x_{j+1} plays the role of a control variable for u_{j+1} while varying x.

3. ESTIMATION

The identification results imply that information accumulates as the sample size increases, which leads to consistent estimation of the representing model

$(f_{X|U}, f_{Y|U}, f_U)$. Section 2.3 suggests that the following iterative procedure estimates the representing model. Let f^N_{XY} denote the empirical joint pmf of the observed variables (X, Y), with f^N_X and f^N_Y 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.

$$\hat{f}_{X|U}(x \mid u_1) = \frac{f^N_{XY}(x, y_1)}{f^N_Y(y_1)} \qquad \text{for all } x \in \mathcal{X}$$
$$\hat{f}_{Y|U}(y \mid u_1) = \frac{f^N_{XY}(x_1, y)}{f^N_X(x_1)} \qquad \text{for all } y \in \mathcal{Y}$$
$$\hat{f}_U(u_1) = \frac{f^N_X(x_1) f^N_Y(y_1)}{f^N_{XY}(x_1, y_1)}.$$

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.

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

Because J is finite, we can complete this iterative procedure to eventually obtain an estimate $(\hat{f}_{X|U}, \hat{f}_{Y|U}, \hat{f}_U)$ of the representing model. See Section B.3 in the online appendix for closed-form estimators.

3.1. Asymptotic Properties

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 a smooth transformation of the empirical data F^N_{XY} through the above closed-form arithmetic formulas provided that singular cases are excluded. 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. In the main text, we focus on the case of $j = 1$ for compactness of exposition. Similar arguments continue to apply for higher j—see Section B.3 in the online appendix.

PROPOSITION 4. *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* $\frac{f_{XY}(x,y_1)[f_Y(y_1)-f_{XY}(x,y_1)]}{f_Y(y_1)^3}$.

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

(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*

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

Note that the singularity issue occurs simply when $f_Y(y_1) = 0$, $f_X(x_1) = 0$, and $f_{XY}(x_1,y_1) = 0$ for parts (i), (ii), and (iii), respectively. For higher j, the singularity occurs in more complicated ways. Specifically, for the asymptotic normality of $\sqrt{N}\left(\hat{f}_U(u_2) - f_U(u_2)\right)$, we require the nonsingularity condition $f_{XY}(x_1,y_1)\cdot f_{XY}(x_2,y_2) \neq f_{XY}(x_1,y_2)\cdot f_{XY}(x_2,y_1)$ in addition to $f_{XY}(x_1,y_1) > 0$—see Proposition 5 in Section B.3 in the online appendix for details. Continuing with u_j for higher j for a couple of steps, we can see that nonzero leading minors of the $j \times j$ matrix $[f_{XY}(x_r,y_c)]_{r,c=1}^{j}$ are required for the asymptotic normality for $\sqrt{N}\left(\hat{f}_U(u_j) - f_U(u_j)\right)$. We numerically study the performance of the estimators for parameter values near these singularities in the following section.

3.2. Monte Carlo Simulations

We consider the following simulation design. The observed reports (X,Y) and the latent variable U are supported on the set, $\mathcal{U} = \mathcal{X} = \mathcal{Y} = \{1,2,3\}$, consisting of three ordered elements, $u_1 = 1$, $u_2 = 2$, and $u_3 = 3$. The marginal distribution of U is given by the uniform law $f_U(1) = f_U(2) = f_U(3) = 1/3$. The reporting patterns follow the assumption of no under-reporting (Assumption 2). We vary probabilities $f_{X|U}(1 \mid 1)$, $f_{Y|U}(1 \mid 1)$, $f_{X|U}(2 \mid 2)$, and $f_{Y|U}(2 \mid 2)$ of honest reporting as shown in the first two columns of Table 1 across sets of simulations. In addition, we fix the conditional probabilities, $\Pr(X = 3 \mid U = 1, X > 1) = 0.5$ and $\Pr(Y = 3 \mid U = 1, Y > 1) = 0.5$. These specifications are sufficient to define the joint probability of (U,X,Y). We run Monte Carlo simulations with the sample size of $N = 300$,[5] repeated for 5,000 iterations in each set.

The third and fourth columns in Table 1 show the first two leading principal minors, $D_1 := f_{XY}(1,1)$ and $D_2 := f_{XY}(1,1)f_{XY}(2,2) - f_{XY}(1,2)f_{XY}(2,1)$, of the matrix $[f_{XY}(x_r,y_c)]_{r,c=1}^{3}$. Recall from Section 3.1 (and Section B.3 in the

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

TABLE 1. Simulation results for $N = 300$ with 5,000 Monte Carlo iterations. The first two columns show the data-generating processes. The next two columns show the first and second leading principal minors, $D_1 := f_{XY}(1,1)$ and $D_2 := f_{XY}(1,1)f_{XY}(2,2) - f_{XY}(1,2)f_{XY}(2,1)$, respectively, for each data-generating process. The remaining columns show estimation results, including the bias, the root mean square errors (RMSE), and the 95% coverage probabilities of the estimators that are computed using the asymptotic normality results provided in Propositions 4 and 5

Data-Generating Process				Estimators					
$f_{X\|U}(1 \mid 1)$	$f_{X\|U}(2 \mid 2)$	Minors		$\hat{f}_U(1)$			$\hat{f}_U(2)$		
$f_{Y\|U}(1 \mid 1)$	$f_{Y\|U}(2 \mid 2)$	D_1	D_2	Bias	RMSE	95%	Bias	RMSE	95%
0.800	0.100	0.213	0.001	−0.000	0.029	0.947	6E+11	4E+13	0.733
0.800	0.200	0.213	0.003	0.000	0.028	0.954	0.106	2.226	0.909
0.800	0.400	0.213	0.011	−0.000	0.028	0.955	0.008	0.065	0.946
0.800	0.800	0.213	0.046	−0.000	0.029	0.949	−0.000	0.030	0.946
0.400	0.800	0.053	0.011	0.004	0.047	0.946	−0.003	0.042	0.955
0.200	0.800	0.013	0.003	0.013	0.089	0.940	−0.009	0.061	0.964
0.100	0.800	0.003	0.001	0.026	0.120	0.939	−0.015	0.089	0.952

online appendix) that the singularity of these leading principal minors is associated with problems in the estimators—see Propositions 4 and 5 . In other words, we expect that the estimators behave poorly when these minors take small values. The middle row of the table shows the benchmark setting, $f_{X|U}(1 \mid 1) = f_{Y|U}(1 \mid 1) = f_{X|U}(2 \mid 2) = f_{Y|U}(2 \mid 2) = 0.8$, where the probabilities of honest reporting is high enough. In this case, both leading principal minors, D_1 and D_2, are relatively far away from zero. Consequently, we indeed see that the estimators $\hat{f}_U(1)$ and $\hat{f}_U(2)$ behave fairly well in terms of the biases, the root mean square errors (RMSE), and the 95% coverage rates that are computed based on the asymptotic normality results displayed in Propositions 4 and 5.

As we move down from the middle row in the table, the probabilities, $f_{X|U}(1 \mid 1)$ and $f_{Y|U}(1 \mid 1)$, of honest reporting given $U = 1$ decrease toward zero. Accordingly, both of the two leading principal minors, D_1 and D_2, also decrease toward zero. Notice that the performance of the estimators, $\hat{f}_U(1)$ and $\hat{f}_U(2)$, becomes worse in terms of the biases, the RMSE, and the 95% coverage rates toward the bottom row. This result is consistent with the fact that the singularity of D_1 is ruled out in Proposition 4 for the estimator $\hat{f}_U(1)$, and the fact that the singularity of both D_1 and D_2 is ruled out in Proposition 5 for the estimator $\hat{f}_U(2)$.

On the other hand, as we move up from the middle row in the table, the probabilities, $f_{X|U}(2 \mid 2)$ and $f_{Y|U}(2 \mid 2)$, of honest reporting given $U = 2$ decrease toward zero. Accordingly, the second leading principal minor D_2 also decreases toward zero, but the first leading principal minor D_1 stays constant. Notice that the performance of the estimator $\hat{f}_U(2)$ becomes worse in terms of the biases, the RMSE, and the 95% coverage rates toward the top row, but the performance of the other estimator $\hat{f}_U(1)$ is not affected. This result is consistent with the fact that the singularity of D_2 is *not* ruled out in Proposition 4 for the estimator $\hat{f}_U(1)$, and the fact that the singularity of D_2 is ruled out in Proposition 5 for the estimator $\hat{f}_U(2)$.

4. 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 panels 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. See Miller, Mulvey, and Martin (1995), Behrman and Rosenzweig (1999), and Rouse (1999) for related empirical research.

To correct errors in self-reported education, Ashenfelter and Krueger collected 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

$$\begin{cases} X = U + V \\ Y = U + W \end{cases} \quad \text{where } U,\ V, \text{ and } W \text{ are mutually independent.} \tag{4.1}$$

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 (4.1) 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. 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,

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

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, such as $Y = 14$, 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 nondegenerate otherwise. In other words, the distribution of W is likely to depend on U without any monotonic patterns. For such irregular 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 (4.1) by the paired nonseparable measurement error model (1.1) in our empirical analysis:

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

Sections 2.1 and 2.2 propose sufficient conditions for identifying this model, particularly in the context of the current empirical problem. Recall that Assumption 2 and thus Assumption 1 are consistent with the aforementioned empirical fact that the self-reporting errors for years of education are likely to be negatively correlated with the true years of education (Siegel and Hodge, 1968). For example, this negative correlation may well arise when people do not under-report their education, as individuals with low education have more room for over-reporting while individuals with high education have little choice but to report truthfully. In addition, we propose a couple of choice models in Sections B.4.1 and B.4.2 in the online appendix as theoretical support for Assumptions 2 and 3, respectively. A remaining issue is whether these assumptions are also consistent with other potential reasons for reporting errors, such as poor memory, misunderstanding the question, and recording errors. If reporters with poor memory are to round up (respectively, round down) to the nearest diploma granting year, then the resultant reporting pattern can be consistent with Assumptions 1 and 2 (respectively, Assumption 3). A similar argument applies to the case of recording errors where rounding occurs on the part of the interviewers.

Table 2 summarizes the orderings of $\mathcal{U} = \{u_1, \ldots, u_J\}$ obtained through each of the approaches proposed in Sections 2.1 and 2.2. Recall that each of these restrictions uniquely defines $\mathcal{X}(u_j)$ and $\mathcal{Y}(u_j)$ for each j in the current setup provided $\mathcal{U} = \mathcal{X} = \mathcal{Y}$; hence there is no need of pre-estimating them using data. As neither Assumption 2 nor Assumption 3 is empirically testable, we do not want to rely on any one of these particular identifying restrictions. Instead, we estimate

TABLE 2. Summary of identifying restrictions and the implied well-orders

Section	Assumption	Reporting Pattern	Implied Ordering of $\mathcal{U}$
2.1	1, 2	No Under-Reporting	$\{u_1, \ldots, u_7\} = \{12, 13, 14, 15, 16, 17, 18\}$
2.2	3	Stigma against Dropout	$\{u_1, \ldots, u_7\} = \{13, 12, 15, 14, 17, 16, 18\}$

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 consist of an extract from a survey of twins conducted at the 16th annual Twins Days Festival in Twinsburg, Ohio, in 1991. The sample contains 340 twins (680 individuals). Figure 1(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

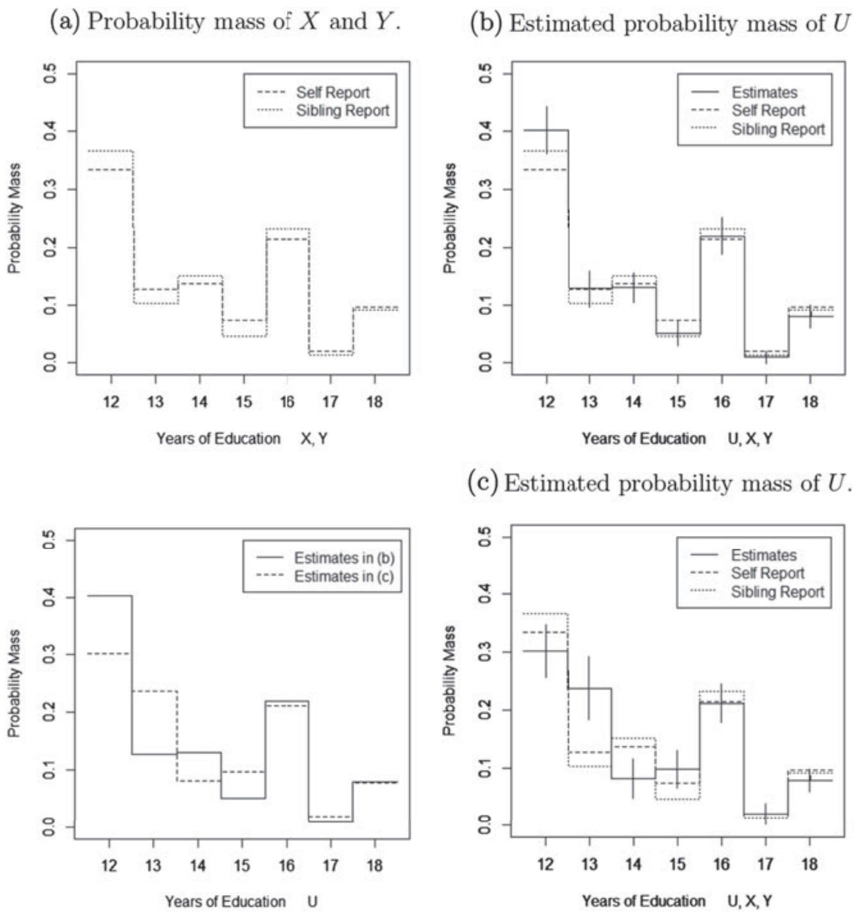

FIGURE 1. (a) Probability masses of self-reported education X and sibling-reported education Y. The remaining two graphs illustrate estimated probability masses of true education U under (b) the assumption of no under-reporting and under (c) the assumption of stigma against dropout without diploma. The vertical lines indicate $\pm 1.96 \times$ estimated standard errors. The bottom left graph overlays the estimates in (c) on top of the estimates in (b) for the purpose of comparison.

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

(12), associate degrees (14), bachelors degrees (16), and masters degrees (18). The sibling report Y particularly stands out at these peaks, which is consistent with the hypothesis 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.

Following the iterative procedure outlined in Section 3, we estimate the distribution F_U of the true years of schooling under each of the alternative restrictions given in Table 2. The two remaining graphs in Figure 1 show the probability masses of the estimated true years of schooling under Figure 1(b) the assumption of no under-reporting and under Figure 1(c) the assumption of stigma against dropout without diploma. The frequencies of self reports are accurate at 16 and 17 years of education robustly across the alternative 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 due only to frequency inflows from other values of U into $X = 18$.

In Figure 2, we compare our estimates to more naive estimates of the cumulative distribution of U, such as using the distribution of X, the distribution of Y, or the distribution of $(X+Y)/2$. The left figure shows that our estimate based on the assumption of no under-reporting is first-order stochastically dominated by all three naive estimates. Likewise, the figure on the right shows that our esti-

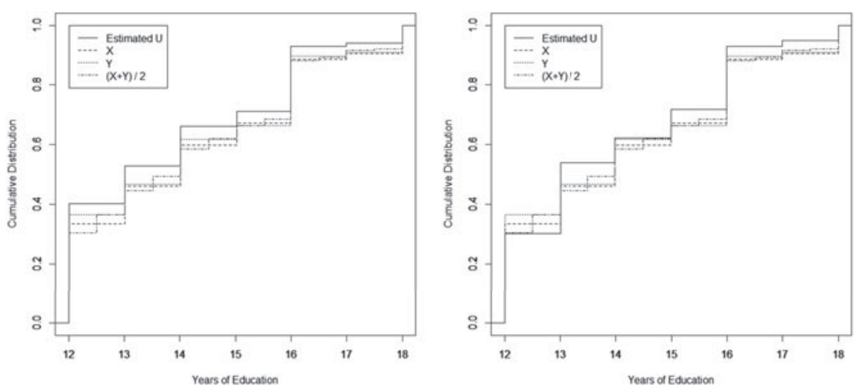

FIGURE 2. Comparisons of our estimates for the distribution of true years of schooling with more naive estimates based on X, Y, and $(X+Y)/2$. The left figure displays our estimate based on the assumption of no under-reporting, and the right figure displays our estimate based on the assumption of stigma against dropout without diploma.

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

mate based on the assumption of stigma against dropping out without a diploma is almost stochastically dominated by all three naive estimates, except at the lowest level $u \in [12, 13)$. These comparisons imply overall left shifts of our estimated distributions relative to the naively estimated distributions.

To assess the actual reporting behaviors (i.e., to reveal who tends 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:

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

These estimated conditional probabilities of correct self and sibling reports are shown in Figure 3. The left and right columns show the results of self reports and sibling reports, respectively. The results displayed in the top and bottom rows are based on the estimates under Figure 3(a) the assumption of no under-reporting and under Figure 3(b) the assumption of stigma against dropout without diploma, respectively. The pattern of self reports are somewhat different across the alternative 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 $U = 13$, who may be characterized as freshman/sophomore dropouts. 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 Figure 3(a) 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. In other words, these robust results are not entirely driven by the assumptions.

By these robust parts of the results across the alternative identifying assumptions, we draw 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 typical diploma-granting years for mnemonic reasons is not overturned. These conclusions are not due to statistical sampling variation, except for the second conclusion about the sibling reports given $U = 17$, for which the long 95% confidence interval extends all the way up to 1.0.

In concluding the empirical application, we remark on some interpretation problems under a possible violation of our identifying assumptions. While the nonseparability condition generalizes the classical measurement error model to a large extent, the assumption that U, V, and W are independent can be still questionable for this application. For example, twins may agree upon a mismea-

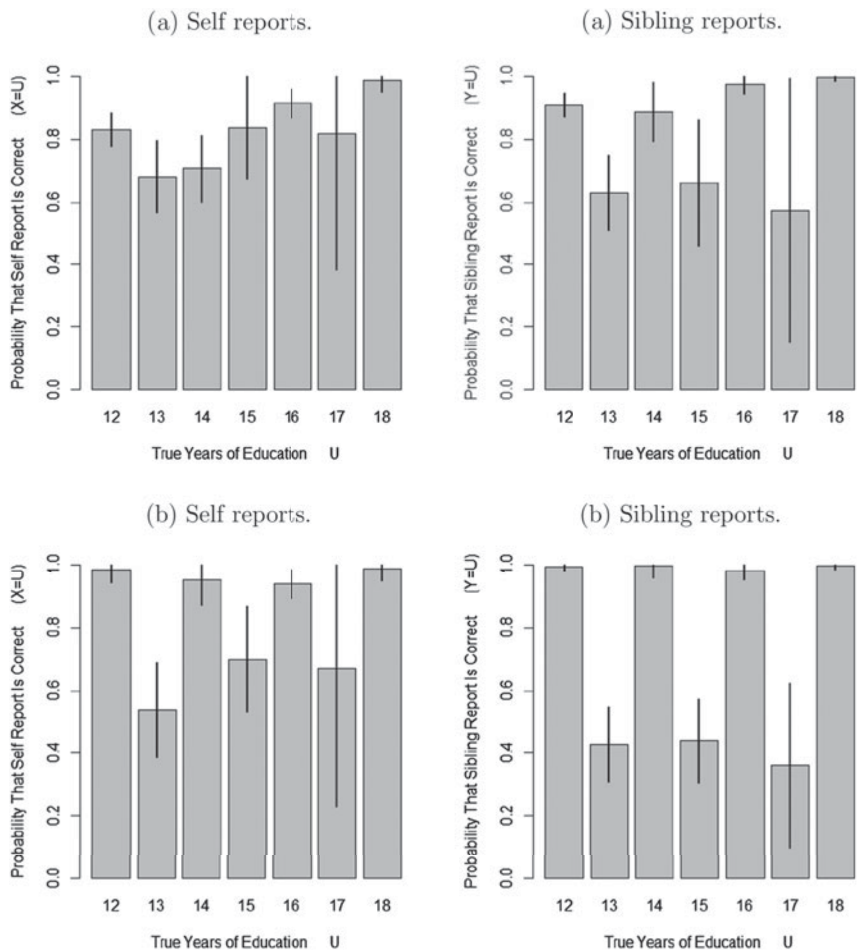

FIGURE 3. Left column: conditional probabilities that self reporting is correct given the truth. Right column: conditional probabilities that sibling reporting is correct given the truth. The conditional probabilities are estimated under (a) the assumption of no under-reporting, and (b) the assumption of stigma against dropping out without a diploma The vertical lines indicate $\pm 1.96 \times$ estimated standard errors.

sure $\psi(U, \eta)$ with the common error η. Suppose that the true structure consists of $X = \tilde{g}(\psi(U, \eta), \tilde{V})$ and $Y = \tilde{h}(\psi(U, \eta), \tilde{W})$. It is observationally equivalent to the structure, $X = g(U, V)$ and $Y = h(U, W)$, where $V = (\eta, \tilde{V})$ and $W = (\eta, \tilde{W})$. In this case, even if U, η, $\tilde{V}$, and $\tilde{W}$ are mutually independent, U, V, and W are not. Our method then identifies the distribution of $\psi(U, \eta)$ instead of U, but it is still better than the aforementioned naive estimates, such as using the distribution of X or the distribution of Y, in the sense that $\psi(U, \eta)$ is at least free of the additional noises $\tilde{V}$ and $\tilde{W}$.

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

5. EXTENSION TO REGRESSION ANALYSIS

Thus far, we have focused on the repeated measurement model (1.1), where U is the unobserved latent variable and (X, Y) are two measurements. In econometrics, we are often interested in structural models and regression models. Before concluding the paper, we discuss how our identification results can be applied to regression analyses.

Suppose that we are interested in the model

$$Y = h(U) + W, \tag{5.1}$$

where the outcome variable Y is observed, but the explanatory variable U is not observable. Instead, we observe a noisy measure X of U, produced by the nonadditive measurement error model:

$$X = g(U, V). \tag{5.2}$$

We assume the explanatory variable U is finitely supported on $\mathcal{U} = \{u_1, \ldots, u_J\}$ with the order $u_1 < \cdots < u_J$ on $\mathbb{R}$. The noisy measure X and the outcome variable Y may be distributed discretely or continuously distributed.

Assumption 4. The following conditions are satisfied for the regression model (5.1) and the measurement model (5.2).

(i) The regression function h is strictly monotone.
(ii) U, V, and W are mutually independent, $\mathbb{E}[W] = 0$, and W is compactly supported.
(iii) $\Pr(X < U) = 0$.
(iv) $\Pr(X = U \mid U = u) > 0$ for each $\mathcal{U}$.

Part (i) is the shape restriction ruling out hump-shaped and wavy regression functions. The function h is assumed to be either strictly increasing or strictly decreasing. This sort of shape restriction is not new in the literature of measurement errors (e.g., Chen, Hu, and Lewbel, 2009). Part (ii) requires the strong exogeneity assumption that the regressor U is statistically independent of the residual W in (5.1), in addition to the standard locational normalization $\mathbb{E}[W] = 0$ and a compact support restriction for W. We require these two parts, (i) and (ii), to guarantee that the conditional distributions $f_{Y|X}(\,\cdot \mid u_j)$ have monotone support boundaries required by Assumption 1 (ii). Parts (iii) and (iv) of Assumption 4 are the same as the assumption of no under-reporting (Assumption 2 for X), which in turn implies Assumption 1(i) by Proposition 2. Therefore, $f_{Y|U}$ is identified under Assumption 4 by Theorem 1. Furthermore, by Assumption 4(ii), this identification of $f_{Y|U}$ in turn implies the identification of the nonparametric regression function h as $h(u_j) = \int y f_{Y|U}(y \mid u_j) dy$ or $h(u_j) = \sum_y y f_{Y|U}(y \mid u_j)$ for each $j \in \{1, \ldots, J\}$. Summarizing these arguments, we obtain the following corollary of Theorem 1.

COROLLARY 1. *If Assumption 4 is satisfied for the regression model (5.1) and the measurement model (5.2), then the regression function h is identified.*

At the cost of invoking Assumption 4 and the finite support $\mathcal{U}$, we can identify the nonparametric regression model using only one measurement. This feature is to be contrasted with the existing identification results for nonparametric regression models that require two measurements (e.g., Li, 2002; Schennach, 2004ab; Hu and Sasaki, 2015) and the identification results for the nonseparable models (e.g., Hu, 2008; Hu and Schennach, 2008; D'Haultfoeuille and Février, 2010). On the other hand, this corollary parallels the result by Chen, Hu, and Lewbel (2009) where they also identify regression models without requiring additional measurements, although direct comparisons are difficult due to the different support cardinality assumptions about U. They require the monotonicity of h as in our Assumption 4(i) as well as the independence and the locational normalization as in our Assumption 4(ii) in order to identify the regression function m without using additional measurements as in our context. Although our baseline restriction (Restriction 1) imposes a strong condition when the model is applied to regression models, it is effectively no stronger than the assumption imposed on the regression models with the same measurement setting in the literature.

6. SUMMARY

This paper proposes nonparametric identifying restrictions for nonseparable paired measurement error models. The general identifying restriction requires that some ordering on the support of unobserved truth entails nonoverlapping conditional support. We provide sufficient conditions for this high-level assumption in the context of our empirical application.

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, which are typical diploma years. Such a nonlinear result would not have been obtained with the traditional methods based on additively separable independent errors.

NOTES

1. This representation of an 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 examples of normalization.

2. Examples include, but are not limited to, measurement error models (Li and Vuong, 1998; Li, 2002; Schennach, 2004ab; 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).

3. For a related example, Kim, Petrin, and Song (2016) propose how to estimate production functions with mismeasured factors.

4. In the United States, 12, 14, 16, and 18 years of education are often, but not necessarily, associated with high-school diplomas, associate degrees, bachelor's degrees, and master's degrees, respectively.

5. We chose this sample size because it is close to the size ($N = 340$) of the empirical data that we use for our empirical application in Section 4.

REFERENCES

Arellano, M. & S. Bonhomme (2012) Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies* 79, 987–1020.

Ashenfelter, O. & A.B. Krueger (1994) Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review* 84, 1157–1173.

Behrman, J.R., P. Taubman, & T.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 (ed.), *Kinometrics: The Determinants of Socioeconomic Success within and between-Families*, pp. 35–96. North-Holland.

Behrman, J.R. & M.R. Rosenzweig (1999) "Ability" biases in schooling returns and twins: a test and new estimates. *Economics of Education Review* 18, 159–167.

Bonhomme, S., & J. Robin (2010) Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies* 77, 491–533.

Carroll, R.J., X. Chen, & Y. Hu (2010) Identification and inference in nonlinear models using two samples with nonclassical measurement errors. *Journal of Nonparametric Statistics* 22, 379–399.

Chen, X., H. Hong, & E. Tamer (2005) Measurement error models with auxiliary data. *Review of Economic Studies* 72, 343–366.

Chen, X., H. Hong, & Tarozzi, A. (2008) Semiparametric efficiency in GMM models with auxiliary data. *Annals of Statistics* 36, 808–843.

Chen, X., Y. Hu, & A. Lewbel (2009) Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information. *Statistica Sinica* 19, 949–968.

Cunha, F., J. Heckman, & S. Navarro (2005) Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers* 57, 191–261.

D'Haultfoeuille, X. & P. Février (2010) Identification of mixture models using support variations. Working Paper, CREST.

Evdokimov, K. (2010) Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working Paper, Princeton University.

Hansen, K.T., J.J. Heckman, & K.J. Mullen (2004) The effect of schooling and ability on achievement test scores. *Journal of Econometrics* 121, 39–98.

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

Hu, Y. & Y. Sasaki (2015) Closed-form estimation of nonparametric models with non-classical measurement errors. *Journal of Econometrics* 185, 392–408.

Hu, Y. & S.M. Schennach (2008) Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76, 195–216.

Kennan, J. & J.R. Walker (2011) The effect of expected income on individual migration decisions. *Econometrica* 79, 211–251.

Kim, K., A. Petrin, & S. Song (2016) Estimating production functions with control functions when capital is measured with error. *Journal of Econometrics* 190, 267–279.

Krasnokutskaya, E. (2011) Identification and estimation of auction models with unobserved heterogeneity. *Review of Economic Studies* 78, 293–327.

Lewbel, A. (2007) Estimation of average treatment effects with misclassification. *Econometrica* 75, 537–551.

Li, T. (2002) Robust and consistent estimation of nonlinear errors-in-variables models. *Journal of Econometrics* 110, 1–26.
Li, T., I. Perrigne, & Q. Vuong (2000) Conditionally independent private information in OCS wildcat auctions. *Journal of Econometrics* 98, 129–161.
Li, T. & Q. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis* 65, 139–165.
Mahajan, A. (2006) Identification and estimation of single index models with misclassified regressor. *Econometrica* 74, 631–665.
Matzkin, R.L. (2003) Nonparametric estimation of nonadditive random functions. *Econometrica* 71, 1339–1375.
Miller, P., C. Mulvey, & N. Martin (1995) What do twins studies reveal about the economic returns to education? A comparison of Australian and U.S. findings. *American Economic Review* 85, 586–599.
Rouse, C.E. (1999) Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review* 18, 149–157.
Schennach, S. (2004a) Estimation of nonlinear models with measurement error. *Econometrica* 72, 33–75.
Schennach, S. (2004b) Nonparametric regression in the presence of measurement errors. *Econometric Theory* 20, 1046–1093.
Siegel, Paul M. & Robert W.H. (1968) A causal approach to the study of measurement error. In H.M. Blalock and A.B. Blalock (eds.), *Methodology in Social Research*, pp. 28–59. McGraw-Hill.
Song, S., S.M. Schennach, & H. White (2012) Estimating nonseparable models with mismeasured endogenous variables. Working Paper, University of Wisconsin Milwaukee.

APPENDIX A

A.1. Proof of Lemma 1

Proof. Note that $f_U(u_1) > 0$ because $u_1 \in \mathcal{U}$. By Restriction 1, we can choose $x^* \in \mathcal{X}(u_1)$ and $y^* \in \mathcal{Y}(u_1)$. Because $f_{X|U}(x^* \mid u_j) = 0$ for all $j > 1$ by the choice of x^*, we have $f_{XY}(x^*, y) = \sum_{j=1}^{J} f_{X|U}(x^* \mid u_j) f_{Y|U}(y \mid u_j) f_U(u_j) = f_{X|U}(x^* \mid u_1) f_{Y|U}(y \mid u_1) f_U(u_1)$ for all $y \in \mathcal{Y}$ by the independence assumption of the model (1.1). Similarly, $f_{XY}(x, y^*) = f_{X|U}(x \mid u_1) f_{Y|U}(y^* \mid u_1) f_U(u_1)$ holds for all $x \in \mathcal{X}$. In particular, $f_{XY}(x^*, y^*) = f_{X|U}(x^* \mid u_1) f_{Y|U}(y^* \mid u_1) f_U(u_1)$. Moreover, $f_X(x^*) = f_{X|U}(x^* \mid u_1) f_U(u_1)$ and $f_Y(y^*) = f_{Y|U}(y^* \mid u_1) f_U(u_1)$ follow. Using all these equalities, we get

$$f_{X|U}(x \mid u_1) = \frac{f_{X|U}(x \mid u_1) f_{Y|U}(y^* \mid u_1) f_U(u_1)}{f_{Y|U}(y^* \mid u_1) f_U(u_1)} = \frac{f_{XY}(x, y^*)}{f_Y(y^*)} \qquad \text{for all } x \in \mathcal{X}$$

$$f_{Y|U}(y \mid u_1) = \frac{f_{X|U}(x^* \mid u_1) f_{Y|U}(y \mid u_1) f_U(u_1)}{f_{X|U}(x^* \mid u_1) f_U(u_1)} = \frac{f_{XY}(x^*, y)}{f_X(x^*)} \qquad \text{for all } y \in \mathcal{Y}$$

$$f_U(u_1) = \frac{f_{X|U}(x^* \mid u_1) f_{Y|U}(y^* \mid u_1) f_U(u_1)^2}{f_{X|U}(x^* \mid u_1) f_{Y|U}(y^* \mid u_1) f_U(u_1)} = \frac{f_X(x^*) f_Y(y^*)}{f_{XY}(x^*, y^*)}.$$

Note that the right-hand sides of these equalities consist of the observed data f_{XY}. Therefore, $(f_{X|U}(\cdot \mid u_1), f_{Y|U}(\cdot \mid u_1), f_U(u_1))$ is identified. ∎

A.2. Proof of Lemma 2

Proof. Note that $f_U(u_{j+1}) > 0$ because $u_{j+1} \in \mathcal{U}$. By Restriction 1, we can choose $x^* \in \mathcal{X}(u_{j+1})$ and $y^* \in \mathcal{Y}(u_{j+1})$. Because $f_{X|U}(x^* \mid u_k) = 0$ for all $k > j+1$ by the choice of x^*, we have $f_{XY}(x^*, y) = \sum_{k=1}^{J} f_{X|U}(x^* \mid u_k) f_{Y|U}(y \mid u_k) f_U(u_k) = \sum_{k\leqslant j} f_{X|U}(x^* \mid u_k) f_{Y|U}(y \mid u_k) f_U(u_k) + f_{X|U}(x^* \mid u_{j+1}) f_{Y|U}(y \mid u_{j+1}) f_U(u_{j+1})$ for all $y \in \mathcal{Y}$. Similarly, $f_{XY}(x, y^*) = \sum_{k\leqslant j} f_{X|U}(x \mid u_k) f_{Y|U}(y^* \mid u_k) f_U(u_k) + f_{X|U}(x \mid u_{j+1}) f_{Y|U}(y^* \mid u_{j+1}) f_U(u_{j+1})$ holds for all $x \in \mathcal{X}$. In particular, $f_{XY}(x^*, y^*) = \sum_{k\leqslant j} f_{X|U}(x^* \mid u_k) f_{Y|U}(y^* \mid u_k) f_U(u_k) + f_{X|U}(x^* \mid u_{j+1}) f_{Y|U}(y^* \mid u_{j+1}) f_U(u_{j+1})$. Moreover, $f_X(x^*) = \sum_{k\leqslant j} f_{X|U}(x^* \mid u_k) f_U(u_k) + f_{X|U}(x^* \mid u_{j+1}) f_U(u_{j+1})$ and $f_Y(y^*) = \sum_{k\leqslant j} f_{Y|U}(y^* \mid u_k) f_U(u_k) + f_{Y|U}(y^* \mid u_{j+1}) f_U(u_{j+1})$ follow. Using all these equalities, we get

$$f_{X|U}(x \mid u_{j+1}) = \frac{f_{XY}(x, y^*) - \sum_{k\leqslant j} f_{X|U}(x \mid u_k) f_{Y|U}(y^* \mid u_k) f_U(u_k)}{f_Y(y^*) - \sum_{k\leqslant j} f_{Y|U}(y^* \mid u_k) f_U(u_k)} \qquad \text{for all } x \in \mathcal{X}$$

$$f_{Y|U}(y \mid u_{j+1}) = \frac{f_{XY}(x^*, y) - \sum_{k\leqslant j} f_{X|U}(x^* \mid u_k) f_{Y|U}(y \mid u_k) f_U(u_k)}{f_X(x^*) - \sum_{k\leqslant j} f_{X|U}(x^* \mid u_k) f_U(u_k)} \qquad \text{for all } y \in \mathcal{Y}$$

$$f_U(u_{j+1}) = \frac{\left[f_X(x^*) - \sum_{k\leqslant j} f_{X|U}(x^* \mid u_k) f_U(u_k)\right]\left[f_Y(y^*) - \sum_{k\leqslant j} f_{Y|U}(y^* \mid u_k) f_U(u_k)\right]}{f_{XY}(x^*, y^*) - \sum_{k\leqslant j} f_{X|U}(x^* \mid u_k) f_{Y|U}(y^* \mid u_k) f_U(u_k)}.$$

Note that the right-hand sides of these equalities consist of the observed data f_{XY}, or are assumed in the statement of the lemma to be known. Therefore, $(f_{X|U}(\cdot \mid u_{j+1}), f_{Y|U}(\cdot \mid u_{j+1}), f_U(u_{j+1}))$ is identified. ■

A.3. Proof of Proposition 1

Proof. To show that Restriction 1(i) is satisfied, assume without loss of generality that $\inf \operatorname{supp}\left(f_{X|U}(\cdot \mid u_j)\right)$ is increasing in j as in Assumption 1(i). Similar arguments follow in the other case. Let $u_j, u_{j+1} \in \mathcal{U}$. Let $s_j = \inf \operatorname{supp}\left(f_{X|U}(\cdot \mid u_j)\right)$ and $s_{j+1} = \inf \operatorname{supp}\left(f_{X|U}(\cdot \mid u_{j+1})\right)$, where $s_j < s_{j+1}$ holds by Assumption 1(i). By the definition of s_j as the infimum of the set $\operatorname{supp}\left(f_{X|U}(\cdot \mid u_j)\right)$, 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 $\operatorname{supp}\left(f_{X|U}(\cdot \mid u_k)\right)$ 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. ■

A.4. Proof of Lemma 3

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 $j = 1, \ldots, J$. 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 $j = 2, \ldots, J$ by Assumption 1. Therefore, $f_{XY}(x, y) = \sum_j f_{X|U}(x \mid$

https://doi.org/10.1017/S0266466616000207 Published online by Cambridge University Press

$u_j) \cdot f_U(u_j) \cdot f_{Y|U}(y \mid u_j) = f_{X|U}(x \mid u_1) \cdot f_U(u_1) \cdot f_{Y|U}(y \mid u_1)$ 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 $s > s_2$.

Therefore, it 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 $s_2 = \inf\{s \in \mathcal{X} | \exists\, x_1, x_2 \in [s_1, s) \cap \mathcal{X}$ and $y_1, y_2 \in \mathcal{Y}$ s.t. $f_{XY}(x_1, y_1) f_{XY}(x_2, y_2) \neq f_{XY}(x_1, y_2) f_{XY}(x_2, y_1)\}$. Similar lines of argument show that t_2 can be characterized by $t_2 = \inf\{t \in \mathcal{Y} | \exists\, x_1, x_2 \in \mathcal{X}$ and $y_1, y_2 \in [t_1, t) \cap \mathcal{Y}$ s.t. $f_{XY}(x_1, y_1) f_{XY}(x_2, y_2) \neq f_{XY}(x_1, y_2) f_{XY}(x_2, y_1)\}$. 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.5. Proof of Lemma 4

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 $k = 1, \ldots, J$. Assume without loss of generality that $\inf\left(f_{X|U}(\,\cdot \mid u_k)\right)$ and $\inf\left(f_{Y|U}(\,\cdot \mid u_k)\right)$ 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 $k = j+1, \ldots, J$ 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 $s_{j+1} = \inf\{s \in \mathcal{X} \mid \exists\ x_1, x_2 \in [s_j, s) \cap \mathcal{X}$ and $y_1, y_2 \in \mathcal{Y}$ s.t. $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)] \cdot [f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \neq [f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \cdot [f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]\}$. Similarly, t_{j+1} can be characterized by $t_{j+1} = \inf\{t \in \mathcal{Y} \mid \exists\, x_1, x_2 \in \mathcal{X}$ and $y_1, y_2 \in [t_j, t) \cap \mathcal{Y}$ s.t. $[f_{XY}(x_1, y_1) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)] \cdot [f_{XY}(x_2, y_2) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \neq [f_{XY}(x_1, y_2) - \sum_{k<j} f_{X|U}(x_1 \mid u_k) f_U(u_k) f_{Y|U}(y_2 \mid u_k)] \cdot [f_{XY}(x_2, y_1) - \sum_{k<j} f_{X|U}(x_2 \mid u_k) f_U(u_k) f_{Y|U}(y_1 \mid u_k)]\}$. 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 $\mathcal{X}_j = \{x \in \mathcal{X} \mid f_X(x) - \sum_{k<j} f_{XU}(x,u_k) > 0\}$ and $\mathcal{Y}_j = \{y \in \mathcal{Y} \mid f_Y(y) - \sum_{k<j} f_{YU}(x,u_k) > 0\}$. 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.6. Proof of Proposition 2

Proof. By Assumption 2(i), $\inf \text{supp}(f_{X|U}(\cdot \mid u)) \geqslant u$ and $\inf \text{supp}(f_{Y|U}(\cdot \mid u)) \geqslant u$ for each $u \in \mathcal{U}$. On the other hand, by Assumption 2(ii), $\inf \text{supp}(f_{X|U}(\cdot \mid u)) \leqslant u$ and $\inf \text{supp}(f_{Y|U}(\cdot \mid u)) \leqslant u$ for each $u \in \mathcal{U}$. Therefore, $\inf \text{supp}(f_{X|U}(\cdot \mid u)) = \inf \text{supp}(f_{Y|U}(\cdot \mid u)) = u$ for each $u \in \mathcal{U}$. It then follows that $\inf \text{supp}(f_{X|U}(\cdot \mid u_j))$ and $\inf \text{supp}(f_{Y|U}(\cdot \mid u_j))$ are increasing in j with the ordering on $\mathcal{U}$ defined by $j < k$ if and only if $u_j < u_k$. ■

A.7. Proof of Proposition 3

Proof. For ease of writing, we define a relation $\prec$ on $\mathcal{U}$ in the following manner: If $u \in \mathcal{D}^c$ and $u' = \max\{u'' \in \mathcal{D} \mid u'' < u\}$, then $u \prec u'$; Otherwise, $u < u' \iff u \prec u'$. The induced relation $\preceq$ can be shown to be a linear order on $\mathcal{U}$, so it can construct the indexed set $\mathcal{U} = \{u_1, \ldots, u_J\}$ such that $j \leqslant k$ if and only if $u_j \preceq j_k$.

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 $u \in \mathcal{U}$. ■