Distributional Effects with Two-Sided Measurement Error: An Application to Intergenerational Income Mobility[*]

Brantly Callaway[†]   Tong Li[‡]   Irina Murtazashvili[§]   Emmanuel S. Tsyawo[¶]

September 20, 2025

Abstract

This paper considers identification and estimation of distributional effect parameters that depend on the joint distribution of an outcome and another variable of interest ("treatment") in a setting with "two-sided" measurement error — that is, where both variables are possibly measured with error. Examples of these parameters in the context of intergenerational income mobility include transition matrices, rank-rank correlations, and the poverty rate of children as a function of their parents' income, among others. Building on recent work on quantile regression (QR) with measurement error in the outcome (particularly, Hausman, Liu, Luo, and Palmer (2021)), we show that, given (i) two linear QR models separately for the outcome and treatment conditional on other observed covariates and (ii) assumptions about the measurement error for each variable, one can recover the joint distribution of the outcome and the treatment. Besides these conditions, our approach does not require an instrument, repeated measurements, or distributional assumptions about the measurement error. Using recent data from the 1997 National Longitudinal Study of Youth, we find that accounting for measurement error notably reduces several estimates of intergenerational mobility parameters.

JEL Codes: C21, J62

Keywords: Measurement Error, Quantile Regression, Continuous Treatment Effects, Intergenerational Income Mobility, Poverty, Inequality

---

[†]John Munro Godfrey, Sr. Department of Economics, University of Georgia, B422 Amos Hall, Athens, GA 30605. Email: brantly.callaway@uga.edu

[‡]Department of Economics, Vanderbilt University, VU Station B #351819, Nashville, TN 37235. Email: tong.li@vanderbilt.edu

[§]Division of Economic and Risk Analysis, U.S. Securities and Exchange Commission, 1617 John F Kennedy Blvd #520, Philadelphia, PA 19103. Email: murtazashviliirina@gmail.com

[¶]Department of Economics, Finance and Legal Studies, Culverhouse College of Business, University of Alabama. Email: estsyawo@gmail.com

1 Introduction

Measurement error is a widespread challenge in empirical work in economics. It is well documented that many important economic variables are often measured with error (e.g., income (Bound and Krueger (1991), Pischke (1995), and Bound, Brown, and Mathiowetz (2001)), consumption (Attanasio, Battistin, and Ichimura (2007)), health status (Bound, Schoenbaum, Stinebrickner, and Waidmann (1999)), government transfers (Meyer, Mok, and Sullivan (2009)), and participation in social programs (Bollinger and David (1997)), among many others). Measurement error can arise due to misreporting (e.g., a survey respondent incorrectly reporting their health insurance coverage) or for other reasons. In the context of our application on intergenerational income mobility, economists have primarily been interested in the effect of parents' *permanent* income on child's *permanent* income. A key empirical challenge in this literature is that, instead of observing permanent incomes, researchers typically observe annual income (or some other shorter-term measure of income). The modern literature on intergenerational income mobility (e.g., Solon (1992) and much subsequent work) considers annual income as an error-ridden measure of permanent income and has developed a number of techniques to deal with measured-with-error versions of permanent incomes, particularly in the context of linear models.[1]

In some settings, it is well-established how to deal with measurement error, or it is at least well-known what the consequences of measurement error are. A leading example is that, in the presence of classical measurement error, a mismeasured right-hand side variable in a linear model leads to attenuation bias, while a mismeasured left-hand side variable does not lead to inconsistency of the parameter estimates in the same setting. In the current paper, we consider more complicated target parameters that are nonlinear functionals of the joint distribution of a continuous outcome and a continuous treatment variable in a setting where both the outcome and the treatment could be mismeasured.[2]

It is useful to give some examples of the types of parameters that we consider in the paper. One example is transition matrices, which summarize transition probabilities from a starting point in the distribution of the treatment variable to an ending point in the distribution of the outcome. In the context of intergenerational mobility, transition matrices provide a way to see how likely children are to move to different parts of the income distribution, conditional on what part of the income distribution their parents were in. Work on transition matrices includes Jantti et al. (2006), Bhattacharya and Mazumder (2011), Black and Devereux (2011), Richey and Rosburg (2018), and Millimet, Li, and Roychowdhury (2020). Another example is Spearman's Rho — the rank-rank correlation of the outcome and the treatment variable. This parameter has received considerable recent attention in the intergenerational mobility literature

---

[1]Much of the literature has focused on estimating Intergenerational Elasticities (IGEs) which are the coefficient on the log of parents' income in a regression of the log of child's income on the log of parents' income. This literature demonstrates that accounting for measurement error leads to notable quantitative differences in estimates of intergenerational mobility. See, for example, Solon (1999), Mazumder (2005), and Haider and Solon (2006), among others.

[2]As a comment on terminology, we refer to the class of parameters that are possibly nonlinear functionals of the joint distribution as *distributional effects*. Slightly abusing terminology, we use the term *treatment* to denote the right-hand side variable of primary interest. Under additional conditions, the distributional effects we consider can have a causal interpretation, but this is not a central issue for our main arguments. We refer to other right-hand side variables as *covariates* and assume that they are not measured with error.

(Chetty, Hendren, Kline, and Saez (2014), Collins and Wanamaker (2017), Chetty and Hendren (2018), and Chetty, Hendren, Jones, and Porter (2020)). Third, another class of parameters are those that are functionals of the entire distribution of the outcome as a function of the treatment; these parameters are related to the literature on continuous treatment effects and include specific examples like the fraction of children whose income is below the poverty line as a function of parents' income (Richey and Rosburg (2016) and Callaway and Huang (2020)). We consider these parameters, along with several more, in substantially more detail in the next section.

What each of the above parameters has in common is that they depend on the joint distribution of the outcome (child's permanent income) and the treatment (parents' permanent income), possibly conditional on covariates. Thus, our main methodological contributions concern identifying joint distributions in the presence of measurement error in both the outcome and the treatment. The main case that we consider in the paper is the one where a researcher has access to a single measurement of the outcome and of the treatment. In this case, we impose two main conditions. First, we impose "classical" measurement error conditions; in particular, we impose that measurement errors are additively separable, independent of the true value of the outcome/treatment and other covariates, and independent of each other.[3] The second main assumption that we make is on auxiliary, preliminary models for the outcome and the treatment conditional on other covariates. In particular, we assume that both the outcome and the treatment are generated by a quantile regression (QR) model that is linear in parameters across all quantiles. Besides these two main conditions, we do not require an instrument, repeated measurements, or distributional assumptions about the measurement errors for identification.

The first part of our identification argument concerns separately recovering the marginal distributions of the outcome and the treatment conditional on other covariates in the presence of measurement error. For this step, our arguments follow from Hausman, Liu, Luo, and Palmer (2021), who show that the structure imposed by the QR model is enough to identify the QR parameters as well as the distribution of the measurement error in the case with classical measurement error in the outcome variable. Applying this argument to both our outcome and treatment implies that the conditional quantiles and conditional distributions of the outcome and treatment are identified.[4] However, identifying these marginal distributions is not enough to identify the distributional effect parameters that we are ultimately interested in, since these parameters depend on the joint distribution of the outcome and the treatment. From Sklar's Theorem (Sklar (1959)), a well-known result in the literature on copulas, we know that joint distributions can

---

[3]In the intergenerational mobility literature, the assumption of classical measurement error is typically considered more plausible for some ages of children and parents (typically mid-career ages). We discuss how to adjust our arguments to the case of "life-cycle" measurement error (as in Haider and Solon (2006)) rather than classical measurement error; see Remark 2 and Appendix SC in the Supplementary Appendix for more details.

[4]This part of our identification strategy can be viewed as semiparametric in the sense that it depends on the two QR models being correctly specified, but it does not place restrictions (besides regularity conditions) on the distribution of the measurement error. This discussion also highlights a connection between this part of our identification strategy and early work on measurement error, where it was common to make distributional assumptions about the measurement error without restricting the distribution of the true outcome/treatment (Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991), and Diggle and Hall (1993)). If we consider the QR model as being a (flexible) distributional assumption for the true outcome/treatment, then the first stage of our identification strategy reverses the variables on which distributional assumptions are placed: for our approach, we effectively make distributional assumptions about the true outcome/treatment without substantively restricting the distribution of the measurement error. A helpful byproduct of semiparametric identification is that, like early work on measurement error, our approach does not require repeated measurements (Li and Vuong (1998)).

be written as the copula (which captures the dependence between two random variables) of the marginal distributions of the outcome and the treatment. We show that, under the conditions above, the copula is also identified. This means that the joint distribution is identified, and hence, all of the distributional effect parameters that we consider are also identified.

For estimation, we propose a three-step approach to estimating the joint distribution of the outcome and the treatment conditional on covariates and then recovering parameters of interest from the joint distribution. In the first step, we estimate the quantile regression parameters and then convert these into estimates of the marginal distributions of the outcome/treatment conditional on covariates. Our estimation approach here is new. We implement an EM-type algorithm that iterates back and forth between making draws of the measurement error (conditional on the values of the parameters) and estimating new values of the parameters. In the second step, we estimate the copula parameters using a simulated maximum likelihood approach that plugs in the first step estimators. To simplify estimation, we assume that the copula is known up to a finite number of parameters (in particular, we consider copulas from an Archimedean family or the Gaussian copula). This step provides an estimate of the joint distribution. Third, we can manipulate the estimated joint distribution into any parameters of interest. We also derive the asymptotic properties of our estimators and show that the empirical bootstrap can be used to conduct inference.

We apply our approach to study intergenerational income mobility using recent data from the 1997 cohort of the National Longitudinal Survey of Youth (NLSY97). This is a rich dataset where we observe much information about families and can link child's income with their parents' income. We employ the 1997 wave of the NLSY97 to obtain information about parents' income and the 2015 wave of the NLSY to obtain the child's income. However, there are some notable challenges. First, we only observe parents' income in a single year. Second, although in principle we observe child's income in every period where they participate in the survey, children are still relatively young, and, in light of much work in the intergenerational mobility literature, it seems most appropriate to only use their income in more recent waves. Our main results indicate that (i) adjusting for measurement error and (ii) using our copula-based approach relative to quantile regression of child's income directly on parents' income and covariates are both important. Accounting for measurement error in child's and parents' income tends to reduce various estimates of intergenerational mobility, resulting in estimates that are more similar to those in the literature where multiple observations of income are available relative to estimates coming directly from the observed data that do not account for measurement error.

Related Work

Our work is related to an extensive literature on measurement error in econometrics (see Hausman (2001) for an overview of the literature). Our paper builds on the literature that recovers the distribution of a variable of interest in the presence of measurement error (important early examples include Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991), and Diggle and Hall (1993)). These papers relied on parametric assumptions about the distribution of the measurement error; Li and Vuong (1998) relax the distributional assumption on the measurement error when there are repeated measurements available.

Notable advances in measurement error problems have also been made for nonlinear models, particularly for the case of the right-hand side (RHS) mismeasured variable (e.g., Hsiao (1989), Li (2000, 2002), Hausman, Newey, Ichimura, and Powell (1991), Hausman, Newey, and Powell (1995), and Schennach (2004, 2008)). Dealing with the left-hand side (LHS) measurement error in a nonlinear context has received much less attention. This is unfortunate since what we know about the linear setting with either RHS or LHS measurement error does not generally carry over to nonlinear models (some examples include Hausman, Abrevaya, and Scott-Morton (1998), Cosslett (2004), Abrevaya and Hausman (1999), and Cavanagh and Sherman (1998)). Furthermore, even less consideration has been given to nonlinear models with *both* RHS and LHS measurement error. Rare exceptions include Lewbel (1996) that studies parametric Engel curves with measurement error on both sides of the equation and De Nadai and Lewbel (2016) that extends Schennach (2007) to allow for both RHS and LHS measurement error in nonparametric models.

Within the measurement error literature, the current paper is most closely related to work on quantile regression with measurement error. The econometric literature for estimating quantile regressions with RHS measurement error includes Schennach (2008), Wei and Carroll (2009), Chesher (2017), and Firpo, Galvao, and Song (2017). Our approach, however, is more closely related to quantile regression with LHS measurement error. Hausman, Liu, Luo, and Palmer (2021) show that the parameters of a linear quantile regression model are identified when there is measurement error in the outcome variable without imposing distributional assumptions on the measurement error or requiring instruments or repeated observations. Our approach builds on this as we use their identification arguments in our first step to show that the marginal distributions of the outcome and the treatment are identified. On the other hand, our estimators of the conditional quantiles are different from those proposed in Hausman, Liu, Luo, and Palmer (2021); we propose an EM-type algorithm while their estimators are based on maximum likelihood.

Finally, there are a few related papers that examine measurement error in the context of nonlinear approaches to understanding intergenerational mobility. Millimet, Li, and Roychowdhury (2020) consider transition matrices, particularly in the case where individuals can be misclassified into the wrong cell. Nybom and Stuhler (2017) and Kitagawa, Nybom, and Stuhler (2018) study rank-rank correlations (i.e., Spearman's Rho) in the presence of measurement error. Nybom and Stuhler (2017) consider measurement error in the ranks of child's income and parents' income rather than in the levels of income. The approach in Kitagawa, Nybom, and Stuhler (2018) builds on the small variance approximations of Chesher (1991) to recover rank-rank correlations when there is measurement error in the levels of child's and parents' income. An, Wang, and Xiao (2020) study a nonparametric version of the IGE in the presence of non-classical measurement error (particularly, life-cycle measurement error).

Organization of the paper

The paper is organized as follows. In Section 2, we introduce a number of parameters that depend on the joint distribution of the outcome and the treatment. Section 3 considers identification of these parameters of interest, allowing for measurement error. In Section 4, we propose

a three-step estimation procedure to estimate the parameters of interest, and, in Section 5, we study the asymptotic properties of the estimators. Our application on intergenerational income mobility using our approach is in Section 6. Final remarks are contained in Section 7.

2   Parameters of Interest

Notation

Let Y^* denote the outcome of interest and T^* denote a continuous treatment variable. We allow both Y^* and T^* to be mismeasured. Also, let X denote a $K \times 1$ vector of covariates that are correctly measured.

All of the parameters that we consider are features of the joint distribution $\mathrm{F}_{Y^*T^*|X}$. Later, our identification arguments will center on identifying this joint distribution. For now, notice that, in the absence of measurement error, this joint distribution is directly identified by the sampling process. In the presence of measurement error, identifying this distribution is non-trivial. We take this issue up in Section 3 below, but for now, we consider a wide variety of parameters that are features of this joint distribution and can, therefore, be obtained if this joint distribution is identified.

We divide the parameters in this section into two categories that we call *Conditional Distribution-type* parameters and *Copula-type* parameters. In the context of intergenerational mobility, the first category includes parameters such as the fraction of children whose income is below the poverty line as a function of parents' income; we can also consider this parameter after adjusting for differences in the distribution of characteristics (e.g., race and education) across different values of parents' income. Another parameter in this class is quantiles of child's income conditional on parents' income and other background characteristics (or after adjusting for differences in other background characteristics). This category also includes parameters from the treatment effects literature (e.g., quantile treatment effects and dose-response functions) as well as counterfactual distributions and distributional policy effects. These types of parameters are of general interest in the case where a researcher is interested in the effect of a continuous treatment on features of the distribution of some outcome.

The copula-type parameters that we consider include transition matrices, measures of upward mobility (i.e., the probability that child's rank in the income distribution is greater than their parents' rank), and rank-rank correlations. Each of these is a nonlinear measure of intergenerational income mobility and has received considerable interest in recent work on intergenerational mobility (though not much attention has been paid to measurement error for these types of parameters).

2.1   Conditional Distribution-type Parameters

The parameters that we consider in this section come from the distribution of Y^* conditional on T^* and X, which is given by

$$\mathrm{F}_{Y^*|T^*X}(y|t,x) = \mathrm{P}(Y^* \le y | T^* = t, X = x) \tag{1}$$

and which will be identified when the joint distribution of (Y^*, T^*) conditional on X is identified. In our application on intergenerational income mobility, this conditional distribution is the distribution of child's income conditional on parents' income and on other characteristics. One parameter of particular interest is given by setting y equal to the poverty line, varying T^* (parents' income), and fixing characteristics X. This gives the fraction of children with income below the poverty line as a function of parents' income and holding other characteristics fixed. Another parameter of interest is the quantiles of Y^* conditional on T^* and X. The quantiles are given by

$$Q_{Y^*|T^*X}(\tau|t,x) = \inf\left\{y : \mathrm{F}_{Y^*|T^*X}(y|t,x) \geq \tau\right\} \tag{2}$$

for $\tau \in (0,1)$. These conditional quantiles are what would be recovered in the absence of measurement error using quantile regression. Quantile regression has been used in the context of intergenerational mobility in Eide and Showalter (1999), Grawe (2004), and Bratberg, Nilsen, and Vaage (2007).[5]

In many applications, it makes sense to integrate out the covariates from the above conditional distributions. Manipulations of the distribution of covariates (while fixing the conditional distribution $\mathrm{F}_{Y^*|T^*X}$) are called counterfactual distributions. Work on counterfactual distributions (without measurement error) includes DiNardo, Fortin, and Lemieux (1996), Machado and Mata (2005), Melly (2005), Chernozhukov, Fernandez-Val, and Melly (2013), Rothe (2010), and Callaway and Huang (2020). There are multiple possibilities here, but the most common one is

$$\mathrm{F}^C_{Y^*|T^*}(y|t) = \int_{\mathcal{X}} \mathrm{F}_{Y^*|T^*X}(y|t,x) \; d\mathrm{F}_X(x) \tag{3}$$

which adjusts the distribution of covariates at every value of T^* to be given by the overall population distribution of covariates.

2.2 Copula-type Parameters

The next set of parameters that we consider depends on the copula of the outcome and the treatment. The copula is the joint distribution of the ranks of random variables. That is, $C_{Y^*T^*}(r,s) = \mathrm{P}\big(\mathrm{F}_{Y^*}(Y^*) \leq r, \mathrm{F}_{T^*}(T^*) \leq s\big)$. Sklar's Theorem, which is one of the most well-known results in the literature on copulas, says that joint distributions can be written as the copula of the marginal distributions, where the copula contains the information about the dependence between the two random variables.[6]

Parameters that depend on the copula of child's income and parents' income are common in the intergenerational mobility literature. Intuitively, one reason why these types of parameters are useful is that they only depend on child and parent ranks in the income distribution and will not be contaminated with information related to how income distributions change across

---

[5]The conditional mean of Y^* given T^* can also be recovered from the conditional distribution or the conditional quantiles. For example, $\mathbb{E}[Y^*|T^*=t, X=x] = \int_0^1 Q_{Y^*|T^*,X}(\tau|t,x)\,d\tau$. In the context of intergenerational mobility, this parameter is a nonparametric/nonlinear version of the IGE. Work on nonlinear IGEs includes Behrman and Taubman (1990), Bratsberg et al. (2007), Björklund, Roine, and Waldenström (2012), Murtazashvili (2012), Murtazashvili, Liu, and Prokhorov (2015), Landersø and Heckman (2017), Callaway and Huang (2019), An, Wang, and Xiao (2020), and Kourtellos, Marr, and Tan (2020).

[6]Joe (1997) and Nelsen (2007) are leading references on copulas.

generations (Chetty, Hendren, Kline, and Saez (2014), Chetty et al. (2014), and Nybom and Stuhler (2017)). Main examples include transition matrices, upward mobility parameters, and rank-rank correlations. We consider each of these in turn next.

Transition Matrices

Transition matrices are one of the main tools to study intergenerational mobility (Jantti et al. (2006), Bhattacharya and Mazumder (2011), Black and Devereux (2011), Richey and Rosburg (2018), and Millimet, Li, and Roychowdhury (2020)). However, not much attention has been paid to transition matrices allowing for measurement error (see Millimet, Li, and Roychowdhury (2020) for an exception, though their arguments are substantially different from ours). Particular cells in transition matrices are given by

$$\theta_{TM}(r_1, r_2, s_1, s_2) = P\big(r_1 \leq F_{Y^*}(Y^*) \leq r_2 | s_1 \leq F_{T^*}(T^*) \leq s_2\big)$$

where (r_1, r_2) and (s_1, s_2) all take values from 0 to 1 (in a typical case, these are chosen so that cells in the transition matrix have cutoffs at the quartiles of the income distribution for children and parents). Notice that transition matrices can be written in terms of the copula of child's income and parents' income. In particular,

$$
\begin{aligned}
\theta_{TM}(r_1, r_2, s_1, s_2) &= \frac{P\big(r_1 \leq F_{Y^*}(Y^*) \leq r_2, s_1 \leq F_{T^*}(T^*) \leq s_2\big)}{P\big(s_1 \leq F_{T^*}(T^*) \leq s_2\big)} \\
&= \frac{C_{Y^*T^*}(r_2, s_2) - C_{Y^*T^*}(r_1, s_2) - C_{Y^*T^*}(r_2, s_1) + C_{Y^*T^*}(r_1, s_1)}{s_2 - s_1}
\end{aligned}
\tag{4}
$$

Upward Mobility

Another parameter of interest in the intergenerational mobility literature is the probability that child's income rank is greater than their parents' income rank. Following Bhattacharya and Mazumder (2011), we define this as a function of parents' income rank,

$$\theta_U(\Delta, s_1, s_2) = P\big(F_{Y^*}(Y^*) > F_{T^*}(T^*) + \Delta | s_1 \leq F_{T^*}(T^*) \leq s_2\big)$$

The leading case here is when $\Delta = 0$ (other values of Δ allow for one to consider the probability of child's income rank exceeding parents' income rank by Δ). In this case, $\theta_U(0, s_1, s_2)$ is the fraction of children who have a higher rank in the income distribution than their parents' rank in the income distribution, conditional on their parents' rank falling in a particular range. Notice that θ_U can be written in terms of the copula of parents' income and child's income. That is,

$$
\begin{aligned}
\theta_U(\Delta, s_1, s_2) &= \frac{P\big(F_{Y^*}(Y^*) > F_{T^*}(T^*) + \Delta, s_1 \leq F_{T^*}(T^*) \leq s_2\big)}{s_2 - s_1} \\
&= \int_{s_1}^{s_2} \int_{\tilde{v}+\Delta}^{1} c_{Y^*T^*}(\tilde{u}, \tilde{v}) \, d\tilde{u} d\tilde{v} \Big/ (s_2 - s_1)
\end{aligned}
$$

where the first equality holds by writing the conditional distribution as the joint distribution divided by the marginal distribution (and because the ranks of parents' income are uniformly

distributed), and the second equality holds because the numerator depends on the joint distribution of the ranks of child's income and parents' income (i.e., the copula).[7]

Rank-Rank Correlations

Finally, we consider the rank-rank correlation, which has received much attention in the intergenerational mobility literature (Chetty, Hendren, Kline, and Saez (2014), Collins and Wanamaker (2017), Chetty and Hendren (2018), and Chetty, Hendren, Jones, and Porter (2020)). This is the correlation between the rank of child's income and the rank of parents' income. The primary advantage of the rank-rank correlation over the more traditional intergenerational elasticity is that the rank-rank correlation does not depend on (changes in) the marginal distributions of income over time, which is often a desirable feature of an intergenerational mobility measure. In the language of the literature on copulas and dependence measures, rank-rank correlations are called Spearman's Rho. The rank-rank correlation is given by

$$\rho_S = 12 \int_0^1 \int_0^1 C_{Y^*T^*}(r, s) \, dr \, ds - 3$$

which, like the other parameters in this section, depends solely on the copula of child's income and parents' income.

Remark 1. In Appendix SB in the Supplementary Appendix, we show that, under the assumption of unconfoundedness, many of the parameters discussed above have a causal interpretation. For example, $F_{Y^*|T^*}^C(y|t)$, in Equation (3) above, corresponds to a distributional version of a dose response function, which is a common target parameter in the literature (e.g., Hirano and Imbens (2004), Flores (2007), Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), Galvao and Wang (2015)). Thus, under unconfoundedness, our framework extends work on causal inference with a continuous treatment to settings where the outcome and treatment are measured with error.

3 Identification

The previous section suggests that one could obtain all the parameters of interest from the joint distribution of Y^* and T^* conditional on X. In our setting, this is challenging because Y^* and T^* are mismeasured. We make the following assumptions about the nature of the measurement errors:

Assumption 1 (Measurement Error).
(i) $Y = Y^* + U_{Y^*}$ $\qquad$ and $\qquad$ $T = T^* + U_{T^*}$
(ii) $(U_{Y^*}, U_{T^*}) \perp\!\!\!\perp (Y^*, T^*, X)$
(iii) $U_{Y^*} \perp\!\!\!\perp U_{T^*}$

---

[7]A related parameter to the upward mobility parameter discussed here is the probability that a child's income itself is greater than their parents' income. The conditional-on-covariates version of this parameter is given by $P(Y^* > t | T^* = t, X = x) = 1 - F_{Y^*|T^*X}(t|t,x)$; thus, this parameter can be immediately derived from the conditional distribution in Equation (1), though we note that it should be classified as a conditional distribution-type parameter rather than a copula-type parameter.

Assumption 1 is a "classical" measurement error assumption. Assumption 1(i) says that the researcher observes Y and T which are mismeasured versions of Y^* and T^*. Assumption 1(ii) says that the measurement error (U_{Y^*}, U_{T^*}) is independent of Y^*, T^*, and X. In terms of our application, it says that the joint distribution of the measurement errors does not depend on the true value of child's permanent income, the true value of parents' permanent income, and observed covariates. This type of assumption is very common both in the literature on nonlinear models with measurement error (e.g., Assumption 1(i) and (ii) are the same as the assumptions in Hausman, Liu, Luo, and Palmer (2021) applied to both the outcome conditional on covariates and the treatment conditional on covariates; similar assumptions are also made in Li and Vuong (1998) and Firpo, Galvao, and Song (2017), among many others) as well as in the literature on intergenerational mobility that allows for measurement error (e.g., the main case considered in Kitagawa, Nybom, and Stuhler (2018) involves classical measurement error though they discuss how their arguments can continue to apply under certain violations of classical measurement error).

Assumption 1(iii) says that U_{Y^*} and U_{T^*} are mutually independent. For our application on intergenerational mobility, this assumption says that the measurement error in child's permanent income is independent of the measurement error in parents' permanent income. This sort of assumption will hold as long as shocks to parents' income are independent of shocks to child's income, which seems plausible because parents' income and child's income are typically observed many years apart. Related assumptions that measurement error for child's income and parents' income are uncorrelated are common in the literature (see, for example, Assumption 2 and related discussion in An, Wang, and Xiao (2020)); instead of uncorrelatedness, we require full independence due to focusing on parameters that depend on the entire joint distribution of child's and parents' income.

Remark 2. A leading alternative setup for measurement error in the intergenerational mobility literature is life-cycle measurement error (see Jenkins (1987) and Haider and Solon (2006) as well as Vogel (2006), Black and Devereux (2011), Nybom and Stuhler (2016), and An, Wang, and Xiao (2020) for additional related discussion). Here, a typical setup would impose that

$$Y_{a_Y} = \lambda_{a_Y}^Y Y^* + U_{Y^*, a_Y} \quad \text{and} \quad T_{a_T} = \lambda_{a_T}^T T^* + U_{T^*, a_T}$$

where a_Y and a_T denote particular ages at which child's income and parents' income could be observed. This modification to Assumption 1 allows for things like income at younger ages to be systematically lower than permanent income and for the distribution of the measurement error to change at different ages. This sort of model is also consistent with the common finding that estimates of intergenerational mobility tend to be higher when child's income is observed when they are relatively young. We discuss how to adapt our arguments to the case of life-cycle measurement error in detail in Appendix SC in the Supplementary Appendix. One of the conditions that is often invoked in the life-cycle measurement error literature is that there exist known ages, typically thought to be starting in the early thirties and perhaps lasting until the late forties (Haider and Solon (2006) and Nybom and Stuhler (2016)), when $\lambda_{a_Y}^Y = 1$ and $\lambda_{a_T}^T = 1$. Thus, even in the presence of life-cycle measurement error, one simple way to get back

to the case with classical measurement error is to use observations from prime-age children and parents only — this is the approach that we take in the application.[8]

Remark 3. Another way to relax the classical measurement error assumption would be to allow for the independence conditions in Assumption 1(ii) and Assumption 1(iii) to hold conditional on X (instead of the measurement error also being independent of X). In that case, our arguments exploiting quantile regression do not go through, but it seems likely that one could identify the joint distribution of (Y^*, T^*) conditional on covariates building on existing arguments on recovering distributions of mismeasured variables, particularly in settings where the researcher has access to repeated measurements. In general, many approaches exist for identifying these distributions even in the presence of complicated forms of non-classical measurement error (see Hu (2017) for a recent review). In applications where repeated observations are available, it would seem possible to substitute these types of approaches for quantile regression in the first step of our identification arguments.[9] See Remarks 4 and 5 for additional discussion on the pros and cons of using quantile regression in the first step.

3.1 Identifying Parameters of Interest

Recall that all the parameters of our interest will be identified if the joint distribution of Y^* and T^* conditional on X is identified. Besides the assumptions on measurement error above, the main additional condition that we require is that the quantiles of the outcome conditional on the covariates and the quantiles of the treatment conditional on the covariates are linear in parameters across all quantiles.

Next, we present a series of results to identify this joint distribution. As a first step, recall that by Sklar's Theorem (Sklar (1959)), joint distributions can be written as the copula of their marginals, i.e.,

$$\mathrm{F}_{Y^*T^*|X}(y, t \mid x) = C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y \mid x), \mathrm{F}_{T^*|X}(t \mid x) \mid x\big) \tag{5}$$

where $C_{Y^*T^*|X}$ is the copula of Y^* and T^* conditional on X which captures the dependence between the ranks of Y^* and T^* conditional on X. Equation (5) is helpful because it splits the identification challenge into two parts: (i) identifying the marginal distributions and (ii) identifying the copula. This section provides two results showing that (i) the distributions of the outcome/treatment conditional on covariates are identified, and (ii) the conditional copula in Equation (5) is identified. Taken together, these results show that the joint distribution is identified and, therefore, that all of our parameters of interest are also identified.

---

[8]This strategy might not always be available, but, especially for applications like ours that use NLSY data where there is little variation in child's age, this is a viable approach.

[9]One common strategy in the intergenerational mobility literature is to average repeated measures of annual income into a single measure of permanent income and then subsequently ignore measurement error. Generally, in nonlinear settings (such as ours), approaches using averaging would require consistent estimation of permanent income, which, due to the incidental parameters problem, would require the number of time periods to grow to infinity (see, e.g., Kato, Galvao Jr, and Montes-Rojas (2012) and Fernández-Val and Weidner (2016)). For these arguments to work, it seems likely that a researcher would already have access to an individual's entire annual earnings history. In this case, there is no measurement error issue at all, as one can simply compute permanent income. Even in the context of linear models, averaging a few years of earnings may be of limited use in accounting for measurement error, especially when transitory income shocks can be persistent (Mazumder (2005)).

Before stating these results, we make the following additional assumption.

Assumption 2 (Quantile Regression)**.**
 (i) $Y^* = Q_{Y^*|X}(V_{Y^*} \mid X) = X'\beta_{Y^*}(V_{Y^*})$ with $V_{Y^*}|X \sim U[0,1]$.
 (ii) $T^* = Q_{T^*|X}(V_{T^*} \mid X) = X'\beta_{T^*}(V_{T^*})$ with $V_{T^*}|X \sim U[0,1]$.

Assumption 2 imposes a quantile regression type of model for Y^* and T^* in terms of X. Assumption 2 implies that Y^* and T^* arise from models that are both monotonically increasing in a scalar unobservable as well as being known up to the parameters $\beta_{Y^*}(\cdot)$ and $\beta_{T^*}(\cdot)$, respectively. Next, we introduce some additional assumptions for recovering the distributions of the outcome/treatment conditional on covariates in the presence of measurement error.

Assumption 3 (Additional Assumptions for QR with Measurement Error)**.** *For $j \in \{Y^*, T^*\}$, we make the following assumptions:*

 (i) $\beta_j(\tau)$ is in the space $M[B_1 \times B_2 \times \cdots \times B_K]$ (with K the dimension of X) where the functional space M is the collection of all functions $b = (b_1, \ldots, b_K) : [0,1] \to [B_1 \times \cdots \times B_K]$ with B_k a closed bounded interval of $\mathbb{R}$ for $k = 1, \ldots, K$ such that $x'b(\tau) : [0,1] \to \mathbb{R}$ is monotonically increasing in τ for all $x \in support(X)$. Each component $\beta_{jk}(\cdot)$ is twice continuously differentiable with uniformly bounded derivatives.

 (ii) $\mathbb{E}[XX']$ is nonsingular. There exists at least one covariate X_1, such that conditional on the remaining covariates X_{-1}, $X_1 \mid X_{-1}$ has a continuous distribution with strictly positive density on an open set and the corresponding coefficient $\beta_{j1}(\cdot)$ is strictly monotonic.

 (iii) The measurement error U_j has a continuously differentiable density, mean zero, and satisfies the condition that there exists a constant $C > 0$ such that for all $k > 0, \mathbb{E}[|U_j|^k] < k! \cdot C^k$.

Assumption 3 reproduces Assumptions 1-3 in Hausman, Liu, Luo, and Palmer (2021) with slight changes in notation. Part (i) is mainly a technical condition. The condition that the mapping $\tau \mapsto x'\beta_j(\tau)$ is monotonically increasing is used to define the admissible class of quantile regression models in Proposition 1 below. Without such a restriction, rearrangements of quantile indices would yield observationally equivalent models once measurement error is introduced. Part (ii) is the most substantive part of the assumption. It requires (a) continuous variation in at least one covariate labeled X_1 and (b) a strictly monotone coefficient path for that covariate across quantiles. In the context of our application, a natural candidate for X_1 is age. The strict monotonicity requirement encodes a shape restriction on heterogeneous returns to age across the conditional income distribution; e.g., the marginal effect of age should increase with τ. This sort of condition would hold if those at the higher parts of the conditional income distribution see faster income growth with respect to age than those at lower parts of the conditional distribution. Part (iii) imposes standard regularity conditions on the measurement error (smooth density, mean zero, and rules out thick tails) that are common in the measurement error literature and ensures the deconvolution is well-posed. All of these conditions are likely to be relatively mild in our application.

Hausman, Liu, Luo, and Palmer (2021) show that QR parameters are identified in exactly the sort of setup discussed above. Notice that identification of $\beta_{Y^*}(\cdot)$ and $\beta_{T^*}(\cdot)$ also implies that the distributions of the outcome/treatment conditional on covariates are also identified. The next proposition states this result.

Proposition 1. *Under Assumptions 1 to 3, $\beta_j(\cdot)$ and F_{U_j} are uniquely identified within the class of quantile regression models satisfying Assumptions 1 to 3 for $j \in \{Y^*, T^*\}$.*

The proof of Proposition 1 is provided in Appendix A. With the quantiles identified, one can obtain the marginals simply by inverting the quantiles. The next part of our identification argument shows that the copula of Y^* and T^* conditional on covariates, $C_{Y^*T^*|X}$, is identified.

Theorem 1. *Under Assumptions 1 to 3, and under the assumption that the characteristic functions of U_{Y^*} and U_{T^*} are non-vanishing, $C_{Y^*T^*|X}$ is identified.*

The proof of Theorem 1 is provided in Appendix A. Together, Proposition 1 and Theorem 1 imply that the joint distribution $\mathrm{F}_{Y^*T^*|X}$ is identified — Proposition 1 implies that the marginals are identified (because the conditional quantiles can be inverted) and Theorem 1 implies that the conditional copula is identified. These results also imply that all of our parameters of interest — both the Conditional Distribution-type parameters and the Copula-type parameters — are identified.

Remark 4. The quantile regression structure imposed in Assumption 2 is crucial for being able to identify the conditional distributions of the outcome and the treatment without requiring repeated measurements of the outcome and treatment. It does not require the parameters $\beta_{Y^*}(\cdot)$ and $\beta_{T^*}(\cdot)$ to have a structural/causal interpretation, but it does require that both models of the conditional quantiles are correctly specified. There is an interesting connection between this requirement and early work on identifying the distribution of a variable of interest in the presence of measurement error (e.g., Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991), and Diggle and Hall (1993)); this literature often invoked parametric assumptions on the measurement error (e.g., that the measurement error follows a normal distribution) which enabled recovering the distribution of the random variable of interest without extra requirements such as repeated measurements. Instead of making a parametric assumption about the measurement error, Assumption 2 amounts to a flexible parametric assumption on the variable of interest while leaving the distribution of the measurement error unrestricted up to some regularity conditions.

Remark 5. Assumption 2 is relatively less important for showing the result in Theorem 1; for example, given that the conditional distributions of the outcome and treatment are identified and that the distributions of the measurement errors in the outcome and treatment equation are identified, Theorem 1 does not otherwise rely on Assumption 2. This suggests that our arguments in the second step are likely to go through under alternative first step assumptions. In particular, in applications where a researcher has access to repeated measurements of the outcome and treatment, it might be attractive to impose less structure than is provided by the quantile regression setup in Assumption 2.

Remark 6. A leading alternative to our approach would be to use quantile regression of the outcome conditional on the treatment and covariates. We are unaware of work that allows for measurement error in both the outcome and the treatment in a quantile regression setup. Conditional on handling measurement error, this setup would deliver the Conditional Distribution-type parameters that we consider; however, it would not immediately deliver the Copula-type parameters that we consider.

4 Estimation

This section considers estimating the parameters of interest from Section 2 in the presence of measurement error based on our identification results from Section 3. We make some simplifying assumptions next that make estimation simpler; in general, we have tried to strike a balance between generality and being able to implement our model in a practical way in realistic applications, though noting that one could proceed in a less restrictive way here at the cost of more complicated estimation. In Appendix D, we assess the performance of our estimation strategy using Monte Carlo simulations.

Assumption 4 (Random Sampling). $\{Y_i, T_i, X_i\}_{i=1}^n$ *are i.i.d. draws from the joint distribution* F_{YTX}.

Assumption 5 (Distribution of Measurement Error). $U_{Y^*} \sim \mathrm{F}_{U_{Y^*}}(\cdot | \sigma_{Y^*})$ *and* $U_{T^*} \sim \mathrm{F}_{U_{T^*}}(\cdot | \sigma_{T^*})$ *where* σ_{Y^*} *and* σ_{T^*} *are finite-dimensional parameters.*

Assumption 6 (Copula). *The conditional copula* $C_{Y^*T^*|X}(\cdot, \cdot \mid x)$ *is invariant in* $x \in \mathrm{support}(X)$ *and is denoted by* $C_{Y^*T^*|X}(r, s; \delta)$ *for some finite-dimensional parameter* δ. *It is twice continuously differentiable in* $(r, s) \in (0, 1)^2$, *and its density*

$$c_{Y^*T^*|X}(r, s; \delta) := \frac{\partial^2}{\partial r \, \partial s} C_{Y^*T^*|X}(r, s; \delta)$$

is strictly positive, uniformly bounded in its arguments, and twice continuously differentiable in δ.

Assumption 7 (Splines). *Let* L *denote a (fixed) number of knots and consider the sequence of equally spaced* τ, $0 < \tau_1 < \tau_2 < \cdots < \tau_L < 1$. *For any* $\tau \in [\tau_1, \tau_L]$ *and* $j \in \{Y^*, T^*\}$,

$$\beta_j(\tau) = \beta_j(\tau_l) + (\tau - \tau_l)\frac{\beta_j(\tau_u) - \beta_j(\tau_l)}{\tau_u - \tau_l}$$

where $\tau_l = \sup\{u \in \{\tau_1, \ldots, \tau_L\} : u \leq \tau\}$ *and* $\tau_u = \inf\{u \in \{\tau_1, \ldots, \tau_L\} : u \geq \tau\}$; *i.e.,* τ_l *is the closest smaller value of* τ *for which* $\beta_j(\tau)$ *has been estimated and* τ_u *is the closest larger value of* τ *for which* $\beta_j(\tau)$ *has been estimated. For* $\tau < \tau_1$ *and* $\tau \geq \tau_L$:

$$\beta_{Y^*}(\tau) = \begin{cases} \beta_{Y^*}(\tau_1) + \frac{\log(\tau/\tau_1)}{1-\tau_1}e_1, & \tau < \tau_1 \\ \beta_{Y^*}(\tau_L) - \frac{\log((1-\tau)/(1-\tau_L))}{\tau_L}e_1, & \tau \geq \tau_L \end{cases}$$

where e_1 *denotes the* $K \times 1$ *standard basis vector with 1 as its first element and zeroes everywhere else.*

Assumption 4 says that we have access to a sample of n i.i.d. draws of Y, T, and X. Assumption 5 says that the distribution of the measurement errors U_{Y^*} and U_{T^*} is known up to a finite number of parameters. In practice, we will assume that U_{Y^*} and U_{T^*} are both mixtures of normal distributions. This is similar to what Hausman, Liu, Luo, and Palmer (2021) do in their application, and mixtures of normals are likely to have good approximating properties for

14

continuously distributed measurement error.[10] The first part of Assumption 6 says that the conditional copula does not vary across different values of the covariates (note that this does not imply that it is equal to the unconditional copula). This sort of condition is frequently invoked in the literature on conditional copulas, where it is often referred to as the "simplifying assumption" as it greatly simplifies estimating conditional copulas. Despite being a common assumption, it is a strong assumption (see, for example, the discussion in Derumigny and Fermanian (2017) as well as Veraverbeke, Omelka, and Gijbels (2011) and Gijbels, Veraverbeke, and Omelka (2011)); Nagler (2025) provides a recent systematic discussion of the plausibility of this assumption both in terms of models that rationalize it and in terms of empirical evidence on its plausibility across several different applications. The second part of Assumption 6 says that the conditional copula of Y^* and T^* is known up to a finite number of parameters. In principle, the conditional copula could be nonparametrically estimated (Theorem 1 implies that it is identified, though it does not provide a directly estimable expression). While possible, proceeding this way is likely to be quite complicated in practice (see, for example, Li and Vuong (1998), Li (2002), Veraverbeke, Omelka, and Gijbels (2011), and Firpo, Galvao, and Song (2017)). Instead, we proceed by considering the case where $C_{Y^*T^*|X}$ is specified parametrically. Assumption 7 allows us to recover the process $\beta_j(\cdot)$ from estimating a finite number of $\beta_j(\tau)$, i.e., estimate them over a finite grid of τ and interpolate the process from the particular set of $\beta_j(\tau)$ that we estimate (see Wei and Carroll (2009) and Arellano and Bonhomme (2016) for similar assumptions in related contexts).[11]

4.1 Step 1: Estimating Conditional Quantiles

The first step is to estimate $Q_{Y^*|X}(\tau \mid x) = x'\beta_{Y^*}(\tau)$ and $Q_{T^*|X}(\tau \mid x) = x'\beta_{T^*}(\tau)$. In this section, we outline the estimation procedure for $Q_{Y^*|X}$ and note that the arguments for $Q_{T^*|X}$ hold analogously.

First, notice that if U_{Y^*} were observed, then we could estimate the QR parameters by quantile regression of $(Y - U_{Y^*})$ on X. Next, let $\psi_\tau(w) = \tau - \mathbb{1}\{w < 0\}$ which is the derivative of the check function $\rho_\tau(w) = (\tau - \mathbb{1}\{w < 0\})w$. If Y^* were observed, the first-order condition for estimating the QR parameters is given by

$$0 = \mathbb{E}\big[X\psi_\tau\big(Y^* - X'\beta_{Y^*}(\tau)\big)\big] = \mathbb{E}\big[X\psi_\tau\big(Y - U_{Y^*} - X'\beta_{Y^*}(\tau)\big)\big]$$

This is an infeasible moment condition because Y^* (and U_{Y^*}) is unobserved. However, notice that it can be rewritten as

$$0 = \mathbb{E}\left[X \int_{\mathcal{U}} \psi_\tau\big(Y - u - X'\beta_{Y^*}(\tau)\big) f_{U_{Y^*}|Y,X}(u|Y,X) \ du\right] \tag{6}$$

---

[10]A side-effect of assuming that the distributions of measurement error are known up to a finite number of parameters is that the smoothness of the measurement error—e.g., ordinary-smooth vs. super-smooth measurement error—does not affect the rate of convergence of our estimator or our inference procedure. This is in contrast with Hausman, Liu, Luo, and Palmer (2021), whose main estimator does not specify a parametric model for the measurement error, and, hence, the smoothness of the measurement error affects inference and rates of convergence.

[11]The parameters ρ_S and $\theta_U(\Delta, s_1, s_2)$ depend on integrals over the distributional tails. The second part of Assumption 7 complements the first part by specifying a spline model for the extreme intervals, following Arellano and Bonhomme (2016, Appendix D) and Geraci and Bottai (2007).

Further, notice that from Bayes' Theorem,

$$f_{U_{Y^*}|Y,X}(u|y,x) = \frac{f_{Y|U_{Y^*},X}(y|u,x) f_{U_{Y^*}|X}(u \mid x)}{f_{Y|X}(y \mid x)} \tag{7}$$

and that

$$f_{Y|U_{Y^*},X}(y|u,x) = f_{Y^*|X}(y - u \mid x) \quad \text{and} \quad f_{U_{Y^*}|X}(u \mid x) = f_{U_{Y^*}}(u)$$

which hold by Assumption 1(i)[12] and Assumption 1(ii). Moreover, given $\beta_{Y^*}(\cdot)$, the conditional cdf of Y^* can be recovered as $F_{Y^*|X}(y \mid x) = \int_0^1 \mathbb{1}\{x'\beta_{Y^*}(\tau) \leq y\}\, d\tau$. In estimation, this transformation ensures that the conditional cdf will be increasing in y (Chernozhukov, Fernandez-Val, and Galichon (2010)). Similarly, the conditional density of Y^* can be recovered as $f_{Y^*|X}(y \mid x) = \left[x'\beta_{Y^*}^{\partial}\left(F_{Y^*|X}(y \mid x)\right)\right]^{-1}$ where $\beta_{Y^*}^{\partial}(\tau) := \frac{\partial \beta_{Y^*}(\tau)}{\partial \tau}$ (see, for example, Rothe and Wied (2020)). Equation (6) thus suggests a two-step estimation procedure. In the first step, make draws from $f_{U_{Y^*}|Y,X}$; in the second step, run QR using the observed Y, X, and the simulated measurement error. In practice, we use the following estimation scheme, iterating back and forth between making draws of the measurement error and estimating the parameters using QR and maximum likelihood.

Algorithm 1.

Initialize:

1. Set the counter $l = 0$. Set initial values of the parameters $\{\widehat{\beta}_{Y^*}^{(l)}, \hat{\sigma}_{Y^*}^{(l)}\}$. Set a tolerance level $\varepsilon = o(n^{-1/2})$.

Iterate:

2. Set $l = l + 1$. For $i = 1, \ldots, n$, make S draws from $\widehat{f}_{U_{Y^*}|Y,X}(u|Y_i, X_i) \propto \widehat{f}_{Y^*|X}(Y_i - u \mid X_i; \widehat{\beta}_{Y^*}^{(l-1)}) \widehat{f}_{U_{Y^*}}(u|\hat{\sigma}_{Y^*}^{(l-1)})$.

3. Using the nS pseudo-observations from Step 2 above, estimate $\widehat{\beta}_{Y^*}^{(l)}$ using quantile regression of $Y_i - U_{is}$ as the outcomes and X_i as the covariates. Estimate $\hat{\sigma}_{Y^*}^{(l)}$ using the same nS observations.

4. If $\left\| \left(\widehat{\beta}_{Y^*}^{(l)}, \hat{\sigma}_{Y^*}^{(l)}\right) - \left(\widehat{\beta}_{Y^*}^{(l-1)}, \hat{\sigma}_{Y^*}^{(l-1)}\right) \right\| < \varepsilon$, return $\left(\widehat{\beta}_{Y^*}^{(l)}, \hat{\sigma}_{Y^*}^{(l)}\right)$, else continue iterating.

4.2 Step 2: Estimating the copula parameter

Under Assumption 6, the copula $C_{Y^*T^*|X}$ is known up to a finite number of parameters. In this section, we propose to estimate the copula parameters by simulated maximum likelihood.

Building up to an estimation routine based on maximum likelihood, first notice that the

---

[12]To see the first part, notice that $P(Y \leq y|X, U_{Y^*} = u) = P(Y^* \leq y - u \mid X)$ which holds by Assumption 1(ii) and then the result follows immediately.

observed distribution of mismeasured variables Y and T is given by

$$
\begin{aligned}
\mathrm{P}(Y \le \bar{y}, T \le \bar{t}|X = x) &= \mathrm{P}(Y^* + U_{Y^*} \le \bar{y}, T^* + U_{T^*} \le \bar{t}|X = x) \\
&= \int \mathbb{1}\{y + u \le \bar{y}, t + v \le \bar{t}\} f_{Y^* T^*|X}(y, t \mid x) f_{U_{Y^*}}(u) f_{U_{T^*}}(v) \, dy \, dt \, du \, dv \\
&= \int C_{Y^* T^*|X}\big(\mathrm{F}_{Y^*|X}(\bar{y} - u|x), \mathrm{F}_{T^*|X}(\bar{t} - v|x)\big) f_{U_{Y^*}}(u) f_{U_{T^*}}(v) \, du \, dv
\end{aligned}
\tag{8}
$$

where the second equality holds by Assumption 1, and the third holds (after some manipulation) by Theorem 1 and Assumption 6. Note that Proposition 1 implies that $\mathrm{F}_{Y^*|X}$, $\mathrm{F}_{T^*|X}$, $f_{U_{Y^*}}$, and $f_{U_{T^*}}$ are identified; they are each known up to a finite number of parameters by Assumptions 5 to 7. What remains is estimating the copula parameters.

Equation (8) suggests estimating the copula parameters by maximum likelihood, and it implies

$$
\begin{aligned}
f_{YT|X}(\bar{y}, \bar{t} \mid x) &= \int c_{Y^* T^*|X}\big(\mathrm{F}_{Y^*|X}(\bar{y} - u \mid x), \mathrm{F}_{T^*|X}(\bar{t} - v \mid x); \delta\big) \\
&\qquad \times f_{Y^*|X}(\bar{y} - u \mid x) f_{T^*|X}(\bar{t} - v \mid x) f_{U_{Y^*}}(u) f_{U_{T^*}}(v) \, du \, dv \\
&= \mathbb{E}\Big[c_{Y^* T^*|X}\big(\mathrm{F}_{Y^*|X}(\bar{y} - U_{Y^*} \mid x), \mathrm{F}_{T^*|X}(\bar{t} - U_{T^*} \mid x); \delta\big) f_{Y^*|X}(\bar{y} - U_{Y^*} \mid x) f_{T^*|X}(\bar{t} - U_{T^*} \mid x)\Big]
\end{aligned}
$$

where $c_{Y^* T^*|X}$ is the copula pdf, and the expectation is over the measurement error variables U_{Y^*} and U_{T^*}. The copula parameter can be estimated by maximizing the log-likelihood function. Let

$$
\begin{aligned}
\widehat{L}_i^{(S)}(\delta) &= \sum_{s=1}^{S} c_{Y^* T^*|X}\big(\widehat{\mathrm{F}}_{Y^*|X}(Y_i - U_{Y^*}^{(s)} \mid X_i), \widehat{\mathrm{F}}_{T^*|X}(T_i - U_{T^*}^{(s)} \mid X_i); \delta\big) \\
&\qquad \times \widehat{f}_{Y^*|X}(Y_i - U_{Y^*}^{(s)}|X_i) \widehat{f}_{T^*|X}(T_i - U_{T^*}^{(s)}|X_i)
\end{aligned}
$$

where $U_{Y^*}^{(s)}$ and $U_{T^*}^{(s)}$ are draws from $\widehat{f}_{U_{Y^*} U_{T^*}|Y,X}(\cdot|Y_i, X_i; \widehat{\sigma})$ for $\widehat{\sigma} = (\widehat{\sigma}_{Y^*}, \widehat{\sigma}_{T^*})$. Then, we estimate the copula parameter by

$$
\widehat{\delta} = \arg\max_{\delta} \sum_{i=1}^{n} \log\big(\widehat{L}_i^{(S)}(\delta)\big).
$$

4.3 Step 3: Estimating Parameters of Interest

4.3.1 Conditional Distribution-type Parameters

First, we focus on estimating the Conditional Distribution-type parameters such as $\mathrm{F}_{Y^*|T^*X}$ and $Q_{Y^*|T^*X}$. Bouyé and Salmon (2009) study the relationship between copulas and quantile regression in the bivariate case (see also Chen and Fan (2006)). The bivariate case is the relevant case for the setup in the current paper, as the results for quantile regression and a bivariate copula go through in our case, with everything holding conditional on X. The results of Bouyé

and Salmon ([2009], p. 726) imply that

$$\mathrm{F}_{Y^*|T^*X}(y|t,x) = C_{2|X}\big(\mathrm{F}_{Y^*|X}(y \mid x), \mathrm{F}_{T^*|X}(t \mid x)\big) \tag{9}$$

where $C_{2|X}(u_1, u_2) = \frac{\partial C_{Y^*T^*|X}(u_1,u_2)}{\partial u_2}$ which provides an alternative expression for $\mathrm{F}_{Y^*|T^*X}$ (here we also use Assumption 6 which says that the conditional copula does not vary with x). Under the conditions that $C_{2|X}$ is invertible in its first argument, the quantiles can be directly obtained by

$$Q_{Y^*|T^*X}(\tau|t,x) = Q_{Y^*|X}\left(C_{2;1|X}^{-1}\big(\tau; \mathrm{F}_{T^*|X}(t \mid x)\big) \,\Big|\, x\right) \tag{10}$$

where $C_{2;1|X}^{-1}(\cdot;\cdot)$ is the inverse of $C_{2|X}$ with respect to its first argument.

Example 1 (Archimedean Family of Copulas)**.** An Archimedean copula is given by

$$C(u_1, u_2) = \psi^{-1}\big(\psi(u_1) + \psi(u_2)\big)$$

where ψ is called a generator function (see Nelsen ([2007]) for more details). The results of Bouyé and Salmon ([2009], p. 729) imply that

$$\mathrm{F}_{Y^*|T^*X}(y|t,x) = \frac{\psi'\big(\mathrm{F}_{T^*|X}(t \mid x)\big)}{\psi'\big(\psi^{-1}\big[\psi\big(\mathrm{F}_{T^*|X}(t \mid x)\big) + \psi\big(\mathrm{F}_{Y^*|X}(y \mid x)\big)\big]\big)}$$

and

$$Q_{Y^*|T^*X}(\tau|t,x) = Q_{Y^*|X}\left(\psi^{-1}\left[\psi\left(\psi'^{-1}\left(\frac{1}{\tau}\psi'\big(\mathrm{F}_{T^*|X}(t \mid x)\big)\right)\right) - \psi\big(\mathrm{F}_{T^*|X}(t \mid x)\big)\right] \,\Big|\, x\right)$$

Example 2. The Clayton copula is a particular Archimedean family with $C(u_1, u_2; \delta) = (u_1^{-\delta} + u_2^{-\delta} - 1)^{-1/\delta}$ where we restrict $\delta > 0$; the generator function is given by $\psi(s; \delta) = \delta^{-1}(s^{-\delta} - 1)$, and one can show that

$$\mathrm{F}_{Y^*|T^*X}(y|t,x) = \left[1 + \mathrm{F}_{T^*|X}(t \mid x)^\delta\big(\mathrm{F}_{Y^*|X}(y \mid x)^{-\delta} - 1\big)\right]^{-(1+\delta)/\delta}$$

and

$$Q_{Y^*|T^*X}(\tau|t,x) = Q_{Y^*|X}\left(\left[(\tau^{-\delta/(1+\delta)} - 1)\mathrm{F}_{T^*|X}(t \mid x)^{-\delta} + 1\right]^{-1/\delta} \,\Big|\, x\right)$$

The Clayton copula has the nice property of having a one-to-one correspondence with Kendall's Tau.

Example 3 (Gaussian Copula)**.** The Gaussian copula is given by $C(u_1, u_2; \rho) = \Phi_2\big(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho\big)$ where Φ is the cdf of a standard normal random variable and Φ_2 is the cdf of a pair of random variables that are jointly normally distributed with mean 0, variance 1, and correlation coefficient

ρ. Then,

$$
\mathrm{F}_{Y^*|T^*X}(y|t,x) = \Phi\left(\frac{\Phi^{-1}\big(\mathrm{F}_{Y^*|X}(y\mid x)\big) - \rho\Phi^{-1}\big(\mathrm{F}_{T^*|X}(t\mid x)\big)}{\sqrt{1-\rho^2}}\right)
$$

and

$$
Q_{Y^*|T^*X}(\tau|t,x) = Q_{Y^*|X}\left(\Phi(\rho\Phi^{-1}\big(\mathrm{F}_{T^*|X}(t\mid x)\big) + \sqrt{1-\rho^2}\Phi^{-1}(\tau)\,\Big|\,x\right)
$$

See Bouyé and Salmon (2009) for more discussion of particular parametric families of copulas and quantile regression.

4.3.2 Copula-type Parameters

Next, we consider estimating our Copula-type parameters. One thing to notice is that the copula that we estimated was the *conditional* copula, but the copula parameters that we focus on are functionals of the unconditional copula. Thus, the key step to estimating the Copula-type parameters is to back out the unconditional copula from our estimates of the conditional copula and conditional distributions of Y^* and T^*. In particular, notice that

$$
\widehat{C}_{Y^*T^*}(r,s) := \frac{1}{n}\sum_{i=1}^{n} C_{Y^*T^*|X}\big(\widehat{\mathrm{F}}_{Y^*|X}(\widehat{\mathrm{F}}_{Y^*}^{-1}(r)\mid X_i), \widehat{\mathrm{F}}_{T^*|X}(\widehat{\mathrm{F}}_{T^*}^{-1}(s)\mid X_i); \widehat{\delta}\big)
$$

is an estimator of the unconditional copula. Then, to estimate particular cells in the transition matrix, one can simply plug into Equation (4); i.e.,

$$
\widehat{\theta}_{TM}(r_1,r_2,s_1,s_2) = \frac{\widehat{C}_{Y^*T^*}(s_2,r_2) - \widehat{C}_{Y^*T^*}(s_1,r_2) - \widehat{C}_{Y^*T^*}(s_2,r_1) + \widehat{C}_{Y^*T^*}(s_1,r_1)}{s_2 - s_1}
$$

Similarly, a plug-in estimator via numerical integration of the upward mobility parameter is given by

$$
\widehat{\theta}_U(\Delta, s_1, s_2) := \int_0^1\int_0^1 \frac{\mathbb{1}\{r > s + \Delta\}\,\mathbb{1}\{s_1 \le s \le s_2\}}{(s_2 - s_1)}\frac{\widehat{f}_{Y^*T^*}\big(\widehat{\mathrm{F}}_{Y^*}^{-1}(r),\,\widehat{\mathrm{F}}_{T^*}^{-1}(s)\big)}{\widehat{f}_{Y^*}\big(\widehat{\mathrm{F}}_{Y^*}^{-1}(r)\big)\,\widehat{f}_{T^*}\big(\widehat{\mathrm{F}}_{T^*}^{-1}(s)\big)}\,dr\,ds.
$$

where

$$
\widehat{f}_{Y^*T^*}(y,t) = \frac{1}{n}\sum_{i=1}^{n} c_{Y^*T^*|X}\big(\widehat{\mathrm{F}}_{Y^*|X}(y\mid X_i),\widehat{\mathrm{F}}_{T^*|X}(t\mid X_i); \widehat{\delta}\big)\widehat{f}_{Y^*|X}(y\mid X_i)\widehat{f}_{T^*|X}(t\mid X_i)
$$

is an estimator of the unconditional joint pdf.

5 Inference

In this section, we develop the asymptotic properties of our estimators. We focus on showing asymptotic normality and bootstrap validity of the estimators of our main target parameters. The key ingredient for providing these results is to establish asymptotic normality of our first

step estimators of the QR and measurement error parameters. Algorithm 1 delivers estimates of the QR and measurement error parameters *at convergence.* Thus, the estimated parameters solve the sample first-order conditions up to an asymptotically negligible term, and we can use proof techniques similar to those for establishing the limiting distribution of M-estimators. In particular, after establishing the consistency of our first step estimators, we linearize the score around the population parameters and apply the Central Limit Theorem (CLT). The main technical challenge lies in handling the nuisance conditional density arising from the measurement-error component. Specifically,

$$f_{U_{Y^*}|Y,X}(u \mid y,x;\sigma_{Y^*}) := \frac{\left[x'\beta_{Y^*}^{\partial}\left(\int_0^1 \mathbb{1}\{x'\beta_{Y^*}(\tilde{\tau}) \leq (y-u)\}\, d\tilde{\tau}\right)\right]^{-1} f_{U_{Y^*}}(u;\sigma_{Y^*})}{\int_{\mathcal{U}} f_{Y|U_{Y^*},X}(y \mid \tilde{u},x)\, f_{U_{Y^*}}(\tilde{u};\sigma_{Y^*})\, d\tilde{u}},$$

where this expression follows from Equation (7) after invoking Assumption 5. This expression makes explicit the dependence on σ_{Y^*}, the quantile-slope function $\beta_{Y^*}^{\partial}(\tau) := \partial\beta_{Y^*}(\tau)/\partial\tau$, as well as the entire quantile coefficient process $\{\beta_{Y^*}(\tau) : \tau \in (0,1)\}$. Consequently, conventional asymptotic arguments that treat each index τ separately are not applicable. Note that, for the discussion of results for our first step estimators, we present results for (Y^*, X', U_{Y^*}), but corresponding results apply for (T^*, X', U_{T^*}) using analogous arguments. Finally, given asymptotic normality of our estimator of the QR parameters, we conclude this section by proving asymptotic normality and bootstrap validity for all of our main target parameters.

5.1 Consistency of First Step Estimators of QR and Measurement Error Parameters

To start with, we show the consistency of our first step estimators of the quantile regression parameters and measurement error parameters. This part relies on the identification result in Proposition 1. We make the following assumption.

Assumption 8 (Compact parameter space and uniform interior)**.** *Let* $\mathcal{T} := [\tau_1, \tau_L]$, $0 < \tau_1 < \tau_L < 1$, $\Gamma_\beta \subset \mathbb{R}^K$, *and* $\Gamma_\delta \subset \mathbb{R}^P$ *be compact with nonempty interior, and* $\Gamma_\sigma = [\underline{\sigma}, \overline{\sigma}]$ *with* $0 \leq \underline{\sigma} < \overline{\sigma} < \infty$. *Define* $\Gamma := \Gamma_\beta \times \Gamma_\sigma \times \Gamma_\delta$. *For all* $\tau \in \mathcal{T}$,

$$\left(\beta_{Y^*}(\tau)',\ \sigma_{Y^*}, \delta'\right)' \in \text{int}(\Gamma), \quad and \quad \inf_{\tau \in \mathcal{T}} \text{dist}\left(\left(\beta_{Y^*}(\tau)', \sigma_{Y^*}, \delta'\right)',\ \partial\Gamma\right) \ \geq\ \kappa$$

for some $\kappa > 0$, *where* $\text{dist}(\cdot, \partial\Gamma)$ *is the Euclidean distance to the boundary of* Γ.

Assumption 8 is a standard regularity condition on the parameter space commonly imposed in the analysis of M-estimators. Importantly, the product space Γ does not depend on τ, and we focus on $\mathcal{T} := [\tau_1, \tau_L]$, a compact subset of $(0,1)$.

For the proof of consistency, we first proceed by providing a representation of the population criterion corresponding to Algorithm 1. The population criterion for a fixed quantile index $\tau \in \mathcal{T}$ is given by

$$Q_\tau(\beta, \sigma) = \mathbb{E}\left[\int_{\mathcal{U}} \rho_\tau\left(Y - u - X'\beta\right) f_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma)\, du\right]$$

where $(\beta_{Y^*}(\tau)', \sigma_{Y^*})' := \underset{(\beta',\sigma)' \in \Gamma}{\arg\min} Q_\tau(\beta, \sigma)$.

The next crucial ingredient for the consistency result is uniform convergence in probability of the sample criterion. Define the empirical (Monte–Carlo) conditional density built from the S draws U_{is}, $s = 1, \ldots, S$ for each i by

$$\widehat{f}^{(S)}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) := \frac{1}{S} \sum_{s=1}^{S} \delta_{U_{is}}(u)$$

where $\delta_{U_{is}}(u)$ is the Dirac point mass at u. Consider the following sample criterion:

$$Q_{n,\tau}(\beta, \sigma) = \frac{1}{n} \sum_{i=1}^{n} \left[\int_{\mathcal{U}} \rho_\tau \left(Y_i - u - X_i'\beta \right) \widehat{f}^{(S)}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma) du \right]$$

with the estimator in Algorithm 1 defined by $\left(\widehat{\beta}_{Y^*}(\tau)', \widehat{\sigma}_{Y^*}\right)' := \underset{(\beta',\sigma)' \in \Gamma}{\arg\min} Q_{n,\tau}(\beta, \sigma)$, $\tau \in \{\tau_1, \ldots, \tau_L\}$.

For each $i \in \{1, \ldots, n\}$ and $\sigma \in \Gamma_\sigma$, let $K_{i,\sigma}$ be the transition kernel of the Markov chain $\{U_{i1}, \ldots, U_{iS}\}(\sigma)$ on $\mathcal{U}$, conditional on (Y_i, X_i), with invariant law $\mathcal{F}_{U_{Y^*}|Y,X}(\cdot \mid Y_i, X_i; \sigma)$ where $\mathcal{F}_{U_{Y^*}|Y,X}\big((-\infty, u] \mid Y_i, X_i; \sigma\big) := F_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma)$. The following assumption is imposed on the convergence of the Markov chain uniformly in σ.

Assumption 9 (MCMC Chain). *There exists $q : \mathcal{U} \to [1, \infty)$, $M < \infty$, and $\rho \in (0, 1)$ such that, for all $u \in \mathcal{U}$, $s \geq 1$,*

$$\sup_{\sigma \in \Gamma_\sigma} \big\| K_{i,\sigma}^s(u, \cdot) - \mathcal{F}_{U|Y,X}(\cdot \mid Y_i, X_i; \sigma) \big\|_{\mathrm{TV}} \leq M q(u) \rho^s$$

where $\sup_{\sigma \in \Gamma_\sigma} \mathbb{E}[q(U_{i0}(\sigma)) \mid Y_i, X_i] < \infty$ almost surely (a.s.) and $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm.

Assumption 9 enables the use of a conditional strong law of large numbers, ensuring that the Monte Carlo approximation error is $O_p(S^{-1/2})$. Sufficient conditions for Assumption 9 are provided, for example, by Rosenthal (1995). Notably, Assumption 9 implies that the Markov chain $\{U_{i1}, \ldots, U_{iS}\}(\sigma)$ for each i is β-mixing (see Doukhan (1995, p.89)), which in turn guarantees that the Monte Carlo approximation error is $O_p(S^{-1/2})$; see, e.g., Prakasa Rao (2009, Theorem 10). We also make the following assumption.

Assumption 10 (Dominance). *There exists a positive constant $C < \infty$ such that almost surely (a.s.), (a) $\mathbb{E}[\|X\|^2] + \mathbb{E}[\|Y\|^2] \leq C$; (b) $\sup_{\sigma \in \Gamma_\sigma} f_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma) \leq C$; (c) $\sup_{\sigma \in \Gamma_\sigma} \int_{\mathcal{U}} u^2 f_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma) \leq C$; (d) $f_{Y^*|X}(y \mid X) \in (0, \infty)$ and continuous in $y \in \mathcal{Y}$; (e) $f_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma)$ is continuously differentiable with respect to σ, and $\mathbb{E}\Big[\Big(\int_{\mathcal{U}} \big|u f^\partial_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\big| du\Big)^2\Big] \leq C$.*

Assumption 10 is a standard regularity condition that imposes boundedness and smoothness conditions on the outcomes, covariates, and measurement error.

The following lemma provides a uniform weak law of large numbers result necessary for establishing the consistency of the estimator in Algorithm 1.

21

Lemma 1. *Suppose Assumptions 4, 5 and 8 to 10 hold, then as $n \to \infty$ and $n/S \to 0$,*

$$\sup_{(\tau,\sigma)' \in \mathcal{T} \times \Gamma_\sigma} \left| Q_{n,\tau}\big(\beta(\tau),\sigma\big) - Q_\tau\big(\beta(\tau),\sigma\big) \right| \xrightarrow{p} 0.$$

The proof of Lemma 1 is provided in the Supplementary Appendix. Next, the following theorem provides the consistency result of $\big(\widehat{\beta}_{Y^*}(\tau)', \widehat{\sigma}_{Y^*}\big)$ uniformly in $\mathcal{T}$.

Theorem 2. *Suppose Assumptions 1 to 5 and 7 to 10 hold, then as $n \to \infty$ and $n/S \to 0$,*

$$\sup_{\tau \in \mathcal{T}} \left\| \big(\widehat{\beta}_{Y^*}(\tau)', \widehat{\sigma}_{Y^*}\big)' - \big(\beta_{Y^*}(\tau)', \sigma_{Y^*}\big)' \right\| = o_p(1).$$

The proof of Theorem 2 is provided in Appendix B.

5.2 Asymptotic Normality

This section establishes joint asymptotic linearity and normality for the stacked $(LK+1) \times 1$ parameter vector $\big((\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L)', \ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\big)'$, where $\boldsymbol{\beta}_{Y^*}^L := (\beta_{Y^*}(\tau_1)', \ldots, \beta_{Y^*}(\tau_L)')'$. This result provides the foundation for deriving the limiting distributions of the target parameters introduced in Section 2.

Theorem 3. *Suppose Assumptions 1 to 5 and 7 to 10 hold. Further, assume the $L(K+1) \times (LK+1)$ matrix $\mathbb{H}_{Y^*}^L$ in (A8) is full column rank, then as $(n/S) \to 0$, (a) $\Big((\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L)', \ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\Big)'$ has the following asymptotically linear representation:*

$$\sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix} = \big(\mathbb{H}_{Y^*}^{L}{}' \mathbb{H}_{Y^*}^L\big)^{-1} \mathbb{H}_{Y^*}^{L}{}' \sqrt{n} \mathbb{C}_{n,Y^*}^L + o_p(1)$$

and, (b)

$$\sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix} \xrightarrow{d} \mathcal{N}\big(0, \ \Omega_{Y^*}^L\big)$$

where $\mathbb{E}\big[\mathbb{C}_{n,Y^*}^L\big] = 0$ and $\Omega_{Y^*}^L := \big(\mathbb{H}_{Y^*}^{L}{}' \mathbb{H}_{Y^*}^L\big)^{-1} \mathbb{H}_{Y^*}^{L}{}' \Big(\lim_{n \to \infty} n \mathrm{Var}\big[\mathbb{C}_{n,Y^*}^L\big] \Big) \mathbb{H}_{Y^*}^L \big(\mathbb{H}_{Y^*}^{L}{}' \mathbb{H}_{Y^*}^L\big)^{-1}.$

The proof of Theorem 3 is provided in Appendix B. The condition that $(n/S) \to 0$ requires S to go to infinity faster than n. In practice, this can be achieved by setting $S = an^{1+d}$, $a > 0, d > 0$.

Next, we state a corollary to this result that says that our estimator of $F_{Y^*|X}(y|x)$, which is directly estimable given our estimates of the QR parameters and is an important input for estimating our main target parameters, is also asymptotically normal.

Corollary 1. *Suppose Assumptions 1 to 5 and 7 to 10 hold, then*

$$\sqrt{n}\big(\widehat{F}_{Y^*|X}(y \mid x) - F_{Y^*|X}(y \mid x)\big) \xrightarrow{d} \mathcal{N}\big(0, \ \sigma(F_{Y^*|X}(y \mid x))\big),$$

where $\sigma(F_{Y^*|X}(y \mid x))$ is explicitly defined in Appendix B.3.

The proof of Corollary 1 is provided in Appendix B. The next proposition shows that our estimator of the parameters of the conditional copula is also asymptotically normal.

Proposition 2. *Suppose Assumptions 1 to 10 hold, then*

$$\sqrt{n}(\widehat{\delta} - \delta) \xrightarrow{d} \mathcal{N}(0, \Sigma(\delta)),$$

where $\Sigma(\delta)$ is explicitly defined in Appendix B.4.

Finally, our main result in this section is to show that our estimators of the main target parameters discussed above are all asymptotically normal and that the empirical bootstrap can be used to approximate their limiting distribution.

Theorem 4. *Suppose Assumptions 1 to 10 hold, then*

$(a) \quad \sqrt{n}(\widehat{\mathrm{F}}_{Y^*T^*|X}(y, t \mid x) - \mathrm{F}_{Y^*T^*|X}(y, t \mid x)) \xrightarrow{d} \mathcal{N}(0, \sigma(\mathrm{F}_{Y^*T^*|X}(y, t \mid x)));$

$(b) \quad \sqrt{n}(\widehat{\mathrm{F}}_{Y^*|T^*X}(y|t, x) - \mathrm{F}_{Y^*|T^*X}(y|t, x)) \xrightarrow{d} \mathcal{N}(0, \sigma(\mathrm{F}_{Y^*|T^*X}(y|t, x)));$

$(c) \quad \sqrt{n}(\widehat{Q}_{Y^*|T^*X}(\tau \mid t, x) - Q_{Y^*|T^*X}(\tau \mid t, x)) \xrightarrow{d} \mathcal{N}(0, \sigma(Q_{Y^*|T^*X}(\tau \mid t, x)));$

$(d) \quad \sqrt{n}(\widehat{\theta}_{TM}(r_1, r_2, s_1, s_2) - \theta_{TM}(r_1, r_2, s_1, s_2)) \xrightarrow{d} \mathcal{N}(0, \sigma(\theta_{TM}(r_1, r_2, s_1, s_2)));$

$(e) \quad \sqrt{n}(\widehat{\rho}_S - \rho_S) \xrightarrow{d} \mathcal{N}(0, \sigma(\rho_S));$ *and*

$(f) \quad \sqrt{n}(\widehat{\theta}_U(\Delta, s_1, s_2) - \theta_U(\Delta, s_1, s_2)) \xrightarrow{d} \mathcal{N}(0, \sigma(\theta_U(\Delta, s_1, s_2))).$

where each of the asymptotic variances is defined in the proof of Theorem 4. In addition, the empirical bootstrap consistently estimates the limiting distribution of each estimator.

We provide proofs of part (a) of Theorem 4 and bootstrap validity in Appendix B, and we provide the proofs of parts (b)-(f) in Appendix SA in the Supplementary Appendix.

6 Intergenerational Income Mobility allowing for Measurement Error

In this section, we use the methods developed in the previous sections to estimate a variety of distributional effect parameters in an application about intergenerational mobility, allowing for measurement error in both parents' permanent income and child's permanent income, and assess how much allowing for measurement error affects the resulting estimates.

Data

The data that we use comes from the 1997 National Longitudinal Survey of Youth (NLSY97). The NLSY97 represents the 1997 cohort of the National Longitudinal Survey that contains detailed panel data on individuals who were 12-16 years old at the start of 1997. It is a very useful resource for studying recent intergenerational income mobility. Following much work on intergenerational income mobility, we focus on a subset of the data involving fathers and sons. We use the logarithm of son's total labor income in 2014 (which is self-reported in the 2015 survey wave) as the outcome variable. In 2014, survey respondents were between 29 and 33 years old.[13]

---

[13]This is essentially right at the start of the age range where the literature has generally observed that annual incomes are not systematically different from permanent incomes (Haider and Solon (2006) and Nybom and Stuhler (2016)). Thus, for the

Table 1: Summary Statistics

	Q1	Q2	Q3	Q4
Son's Income	46.17	50.82	59.18	66.34
	(33.77)	(34.91)	(42.2)	(46.89)
Father's Income	19.46	43.2	64.22	112.28
	(9.64)	(5.66)	(6.48)	(51.19)
% White	0.53	0.67	0.74	0.84
	(0.50)	(0.47)	(0.44)	(0.37)
% Black	0.14	0.13	0.13	0.06
	(0.35)	(0.34)	(0.33)	(0.24)
% Hispanic	0.32	0.19	0.12	0.09
	(0.47)	(0.40)	(0.33)	(0.28)
Son's Age 1997	14.15	14.16	14.27	14.27
	(1.43)	(1.48)	(1.51)	(1.47)
Father's Age 1997	40.61	40.91	41.97	42.79
	(7.30)	(5.87)	(5.44)	(4.98)
Father's Educ.	11.38	13.03	13.92	15.32
	(3.45)	(2.46)	(2.45)	(2.76)

Notes: The table provides summary statistics by father's income quartiles in each column. Son's income and father's income are reported in thousands of dollars. Standard deviations of each variable are reported in parentheses.

Sources: NLSY97, as described in text.

We use the logarithm of father's total labor income in 1996 (in 2014 dollars) as the treatment variable. We drop all father-son pairs in which either the father or the son has zero or missing earnings.

The main advantages of using the NLSY97 relative to other datasets, such as the Panel Study of Income Dynamics (PSID), are (i) the NLSY97 sample sizes tend to be larger, and there are more individuals who are of a similar age, and (ii) the NLSY97 tends to have more individual level covariates that can be included in our analysis. The main complication is that we do not observe permanent incomes. For fathers, we only observe a single year of income from the first year of the survey. For sons, we only use income reported for the year 2014 and discard information about income in subsequent years. Having one observation of father's and son's incomes fits into the methodological framework considered in the paper.

Besides the income variables for fathers and sons, we also observe race, ethnicity, ages of both fathers and sons, and the father's education (in years). We drop observations that are missing values for any of these. This results in a dataset with 1,096 observations. Summary statistics for our dataset are available in Table 1. Each column provides average values of each variable by quartile of father's income. Like almost all work on intergenerational mobility, we see that son's income is increasing in father's income and that sons from low-income families tend to have higher incomes than their fathers, while sons from high-income families tend to have lower incomes than their fathers. The estimated intergenerational elasticity (IGE) is 0.054

---

results below, we do not make any adjustments related to life-cycle measurement error as discussed in Appendix SC in the Supplementary Appendix.

in our data. This is broadly similar to early studies of intergenerational income mobility that ignored measurement error in income variables. However, the estimated value of the IGE using our data is small relative to more recent IGE estimates that account for measurement error in income in a linear regression framework. In terms of other covariates, the percentage of sons that are white, father's age, and father's education are all increasing with father's income. The percentage of sons who are black and the percentage of sons who are Hispanic both decrease with fathers' income. The son's age in 1997 does not appear to be correlated with father's income.

Implementation Details

In this section, the only covariates we include are son's age and father's age; this specification is common in the intergenerational mobility literature. In Appendix SD in the Supplementary Appendix, we provide additional results that condition on more covariates. For our first step quantile regression estimates (that allow for measurement error), we estimate $\beta_j(\tau)$ over $L = 25$ equally spaced values of τ from 0.02 to 0.98. We initialize Algorithm 1 by setting $\widehat{\beta}_j^{(0)}(\tau)$ equal to the estimates of the QR parameters ignoring measurement error. We specify the measurement error for both son's permanent income and father's permanent income as a mixture of two normal distributions. The parameters for the mixture of normals are each mixture component's mean and variance, and the mixture probability. We initialize the two means to be -1 and 1, both variances to be 1, and the mixture probability to be 0.5. At each iteration, given estimates of $\left(\widehat{\beta}_j^{(l-1)}(\cdot), \hat{\sigma}_j^{(l-1)}\right)$ from the previous iteration, for each observation, we make 100 measurement error draws using the last 100 draws from a Metropolis–Hastings algorithm, after discarding 200 burn-in draws. The proposal distribution is a Gaussian random walk with variance fixed at 1. As discussed in Algorithm 1, we update the QR parameters using quantile regression accounting for the measurement error draws. We update the parameters for the measurement error itself by fitting a two-component mixture of normals on the measurement error draws. For this step, we use the R `mixtools` package (Benaglia, Chauveau, Hunter, and Young (2010)), which implements a standard (and relatively computationally cheap) EM-algorithm for estimating mixture models (see also McLachlan, Lee, and Rathnayake (2019)). Finally, we set the tolerance level ε to be equal to $\left(L(K-1)+3m\right)/40$ where K is the number of regressors and m is the number of mixture components for the measurement error (here $m = 2$ and $K = 3$, which includes an intercept). For estimating the conditional copula in the second step of our estimation procedure, we specify that the conditional copula is a Clayton copula and estimate its parameter. In Appendix SD in the Supplementary Appendix, we provide analogous results to the ones in this section that use a Gaussian copula instead. To estimate the copula parameter, we use 250 simulated measurement error draws in our simulated maximum likelihood procedure. We report standard errors and confidence intervals below that come from using the empirical bootstrap with 100 iterations.

Copula-type Parameters

To start with, we consider copula-type parameters: rank-rank correlations, transition matrices, and upward mobility measures. Although our estimation approach involves estimating quantile regressions conditional on covariates (and accounting for measurement error), the copula-type

Table 2: Transition Matrix

		Father's Income Quartile						Father's Income Quartile			
		1	2	3	4			1	2	3	4
Son's Income Quartile	4	0.126	0.222	0.288	0.364	Son's Income Quartile	4	0.161	0.234	0.255	0.376
		(0.037)	(0.012)	(0.013)	(0.036)			(0.020)	(0.029)	(0.026)	(0.026)
	3	0.165	0.256	0.287	0.291		3	0.197	0.215	0.303	0.266
		(0.029)	(0.004)	(0.014)	(0.014)			(0.026)	(0.033)	(0.032)	(0.027)
	2	0.241	0.278	0.255	0.226		2	0.328	0.230	0.314	0.175
		(0.007)	(0.015)	(0.004)	(0.013)			(0.031)	(0.035)	(0.038)	(0.030)
	1	0.468	0.244	0.170	0.119		1	0.303	0.277	0.175	0.190
		(0.071)	(0.008)	(0.029)	(0.036)			(0.026)	(0.027)	(0.023)	(0.026)
		(a) Measurement Error						(b) Observed			

Notes: The table provides estimates of transition matrices either allowing for measurement error using the techniques developed in the paper (Panel (a)) or coming directly from the observed data (Panel (b)). The columns are organized by quartiles of father's income; e.g., columns labeled "1" use data from fathers whose income is in the first quartile. Similarly, rows are organized by quartiles of son's income. Standard errors are computed using the bootstrap.

Sources: NLSY97, as described in text.

parameters that we consider are unconditional and, therefore, can be directly compared to rank-rank correlations, transition matrices, or upward mobility measures computed directly from the observed data. We report both of these types of results in this section, which provides a straightforward way to compare our estimates that allow for measurement error with estimates that ignore measurement error.

We start with rank-rank correlations. Using the raw data (and ignoring measurement error), our estimate of the rank-rank correlation is 0.208 (s.e.=0.034). By contrast, our estimate of the rank-rank correlation that allows for measurement error is 0.345 (s.e.=0.104). The difference between these estimates is large — our estimates that adjust for measurement error suggest substantially less intergenerational mobility (due to the rank-rank correlations being higher) than is implied by the estimates that use the observed data directly.[14]

The results on transition matrices are presented in Table 2. First, ignoring measurement error (as in Panel (b)), the estimated transition matrix indicates a large degree of intergenerational income mobility. While sons generally are somewhat more likely to stay in the same income quartile as their father, large movements in the income distribution do not appear to be uncommon.

---

[14] Among the distributional effect parameters that we consider, the effect of measurement error on rank-rank correlations has been considered in more depth in the literature relative to other parameters. First, using administrative data from Sweden with a long earnings history, Nybom and Stuhler (2017) provide results (see, in particular, Panel c of their Figure 1) comparing estimates of intergenerational rank-rank correlations for children at different ages to rank-rank correlations using the entire earnings history (and thus accounting for measurement error). At age 30, which is comparable to our estimates reported above that ignore measurement error, they estimate a rank-rank correlation of about 0.18. When they use the full earnings history, which is infeasible for us but comparable to our estimates that account for measurement error, they estimate a rank-rank correlation of about 0.26. This is fairly similar to our estimate, perhaps reflecting somewhat lower intergenerational mobility in the United States than in Sweden. Second, Chetty, Hendren, Kline, and Saez (2014) use administrative data from the United States for parent-child pairs that are 29-32 in 2011 and 2012, which is very similar to the ages and years that we consider. Their main rank-rank correlation estimate is 0.341, which is remarkably close to ours. We note that these estimates being this similar to each other is likely to be partially a coincidence, as (i) our estimates are not especially precise and (ii) their estimates only use two periods of income, which may not be enough for the averaged income to reflect permanent income.

Table 3: Upward Mobility

	Father's Income Quartile			
	1	2	3	4
Measurement Error				
	0.776	0.617	0.430	0.175
	(0.034)	(0.005)	(0.020)	(0.018)
Observed				
	0.830	0.573	0.387	0.167
	(0.022)	(0.029)	(0.026)	(0.024)

Notes: The table provides estimates of upward mobility allowing for measurement error using the techniques developed in the paper (top panel) or coming directly from the observed data (bottom panel). The columns are organized by quartiles of father's income; e.g., columns labeled "1" use data from fathers whose income is in the first quartile. Standard errors are computed using the bootstrap.
*Sources:* NLSY97, as described in text.

For example, for a son whose father was in the lowest quartile of the income distribution, our estimate shows that the probability that the son is in the lowest quartile of the income distribution is 30%, but that the probability that the son moves to the top quartile of the income distribution is 16%. Similarly, for fathers in the top quartile of the income distribution, we estimate the probability that their son is in the top quartile of the income distribution to be 38% and that the probability that their son is in the bottom quartile of the income distribution to be 19%. Allowing for measurement error (as in Panel (a)) indicates substantially less intergenerational mobility. For example, for fathers in the lowest quartile of the permanent income distribution, we estimate that 47% of their sons stay in the lowest quartile of the permanent income distribution and only 13% move to the top quartile of the permanent income distribution. For fathers in the top quartile of the permanent income distribution, we estimate that 36% of their sons stay in the top quartile while only 12% move to the bottom quartile.

Finally, we present upward mobility estimates in Table 3. These are estimates of the fraction of sons whose rank in the income distribution exceeds the rank of their fathers. We present these results by quartile of father's income as well as separately by whether we use our approach that allows for measurement error or just use the observed data directly. Compared to using the observed data directly, the upward mobility estimates allowing for measurement error are lower when the father is in the first quartile of the permanent income distribution and higher when the father is in the second, third, or fourth quartile of the permanent income distribution. As for the other parameters in this section, these results suggest that allowing for measurement error reduces estimates of intergenerational income mobility.

Conditional Distribution-type Results

The next set of results that we present corresponds to the Conditional Distribution-type parameters. In this section, we primarily compare estimates using our approach that allows for

Figure 1: Quantiles of Son's Income as a Function of Father's Income

(a) Measurement Error

(b) QR No Measurement Error
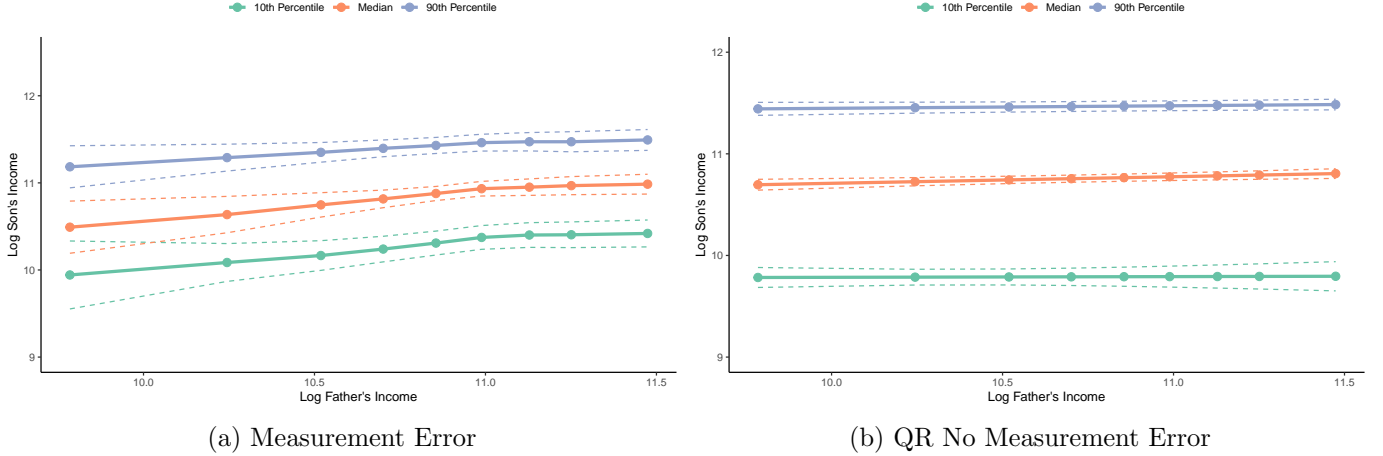
Notes: The figure provides estimates of quantiles of son's income as a function of father's income and conditional on son's age and father's age being equal to their averages in the sample. The estimates in Panel (a) come from the approach suggested in the current paper that allows for measurement error. The estimates in Panel (b) ignore measurement error and come from quantile regression of observed son's income on observed father's income and covariates. The three lines in each panel are estimates of the 10th percentile, median, and 90th percentile of son's income, which are estimated at the 10th, 20th, ..., and 90th percentiles of father's income. Standard errors are computed using the bootstrap.

measurement error to estimates of conditional distribution-type parameters that ignore measurement error and come directly from quantile regression estimates of son's income on father's income, son's age, and father's age.

Figure 1 contains estimates of conditional quantiles (the 10th percentile, median, and 90th percentile) of son's income as a function of father's income, where son's age and father's age (both in 1997) are set at their average values in our sample (14 and 42, respectively).[15] Estimates of these conditional quantiles are reported at the 10th, 20th, ..., and 90th percentiles of observed father's income. These are unequally spaced values of father's income, but they range from about $18,000 to just over $96,000. Panel (a) contains estimates using our approach that allows for measurement error, and Panel (b) contains estimates coming from quantile regression of son's income on father's income and covariates ignoring measurement error.

We start by discussing the estimates that ignore measurement error in Panel (b). By construction, the quantile regression estimates in Panel (b) are linear. Ignoring measurement error, the slopes of the quantile regression estimates are very flat, indicating little effect of father's income on son's income. For interpreting the results, we mainly discuss results conditional on father's income being in the 10th percentile (which is roughly equal to the poverty line) and in the 90th percentile. Using quantile regression and ignoring measurement error, the 10th percentile of son's income conditional on having father's income at the 10th percentile (and conditional on having average values for father's age and son's age) is estimated to be $17,700, the median is estimated to be $44,200, and the 90th percentile is estimated to be $93,200. For sons whose father is in the 90th percentile of the income distribution, the distribution of their income is very similar —

---

[15]We estimate these conditional quantiles by inverting the sample analogue of Equation (9), which ensures monotonicity of the conditional quantiles.

Figure 2: Poverty Rate

(a) Measurement Error

(b) QR No Measurement Error
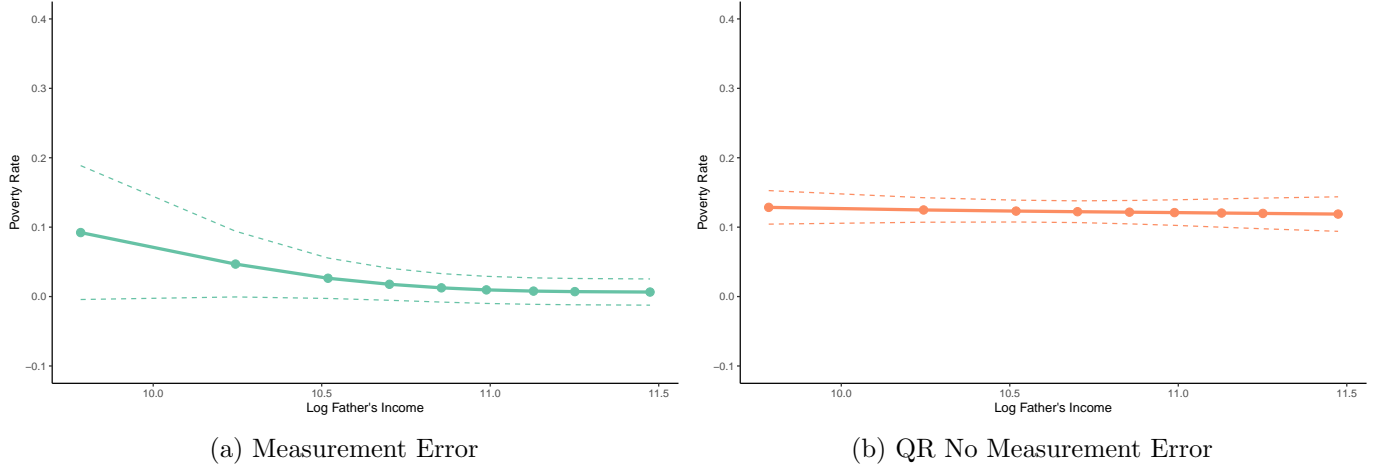
Notes: The figure provides estimates of son's poverty rate as a function of father's income and conditional on son's age and father's age being equal to their averages in the sample. The estimates in Panel (a) come from the approach suggested in the current paper that allows for measurement error. The estimates in Panel (b) ignore measurement error and come from inverting quantile regressions of observed son's income on observed father's income and covariates. The poverty line is set at (the logarithm of) $20,000. Standard errors are computed using the bootstrap.

we estimate that the 10th percentile (conditional on having average values for father's age and son's age) is $17,900, the median is $49,300, and the 90th percentile is $97,200.[16]

Moving to results that allow for measurement error (these are in Panel (a) of Figure 1), there are notable differences. The slopes of the estimates are steeper, which indicates a stronger relationship between father's and son's permanent income. For sons whose father's permanent income is in the 10th percentile, allowing for measurement error substantially reduces the median and 90th percentile of their permanent income distribution — the median is now estimated to be $36,000 and the 90th percentile to be $72,000. The other major difference is in the lower part of the permanent income distribution for sons whose fathers had high permanent incomes. There, we estimate that the 10th percentile of permanent income is $33,500 and the median permanent income is $59,000 — these are both notably higher than our earlier quantile regression estimates that ignored measurement error.

Next, we consider estimates of the fraction of sons whose income is below the poverty line (which we set at $20,000) as a function of father's income and conditional on father's and son's ages being set at their average in our sample. Using quantile regression (and ignoring measurement error), we estimate that 12.9% of sons whose father's income is in the 10th percentile have income that is below the poverty line. For sons whose father's income was in the 90th percentile, we estimate that 11.9% have income that is below the poverty line. Allowing for measurement error again has a large effect and indicates substantially less intergenerational mobility. For sons whose father's permanent income is in the 10th percentile, we estimate that 9.2% have perma-

---

[16]For $\tau \in \{0.1, 0.5, 0.9\}$, from our quantile regression estimates of the log of son's income on the log of father's income and father's and son's age, the estimated effect of father's income is small (ranging from 0.005 to 0.067) but positive across all values of τ. The estimated effect is only statistically significant for $\tau = 0.5$. This immediately translates to the small estimated effects of father's income in Panel (b) of Figure 1.

nent income below the poverty line, but for sons whose father's permanent income is in the 90th percentile, we estimate that only 0.6% have permanent income below the poverty line.

Our results in this section indicate that accounting for measurement error generally reduces estimates of various ways to summarize intergenerational income mobility, which is in line with the literature in many different contexts (e.g., Solon (1992), Mazumder (2005), Haider and Solon (2006), Nybom and Stuhler (2016), and An, Wang, and Xiao (2020)). These results especially matter in the tails of the distribution where the sensitivity to measurement error is higher, which is also in line with the literature (e.g., Bhattacharya and Mazumder (2011) and Nybom and Stuhler (2017)).

7   Conclusion

In this paper, we developed a new method to obtain the joint distribution of an outcome and a continuous treatment conditional on covariates when the outcome and the treatment are potentially measured with error. Our main innovation and departure from the existing literature was to model both the outcome and the treatment as a function of the covariates; that is, changing the problem from measurement error on the left and right to two cases of measurement error on the left. Then, we showed that the copula of the outcome and treatment was identified in this setup, which, in turn, implies that a large number of distributional effect parameters of interest are identified.

We applied our approach to study intergenerational income mobility. We use this as a leading example for our approach because researchers studying intergenerational income mobility typically only have noisy measurements of an individual's permanent income. In addition, we used recent data from the NLSY where only one observation of child's and parents' income was available, which means that many existing approaches to accounting for measurement error are not applicable. We found that accounting for measurement error led to substantially lower estimates of a variety of measures of intergenerational mobility relative to implementing common alternative approaches while ignoring measurement error.

References

Abrevaya, Jason and Jerry Hausman (1999). "Semiparametric estimation with mismeasured dependent variables: An application to duration models for unemployment spells". *Annales D'Economie et de Statistique* 55/56, pp. 243–275.

An, Yonghong, Le Wang, and Ruli Xiao (2020). "A nonparametric nonclassical measurement error approach to estimating intergenerational mobility elasticities". *Journal of Business & Economic Statistics*, pp. 1–17.

Arellano, Manuel and Stéphane Bonhomme (2016). "Nonlinear panel data estimation via quantile regressions". *The Econometrics Journal* 19.3.

Attanasio, Orazio, Erich Battistin, and Hidehiko Ichimura (2007). "What really happened to consumption inequality in the US?" *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*. Ed. by Berndt, Ernst and Charles Hulten. National Bureau of Economic Research, pp. 515–543.

Behrman, J.R. and P. Taubman (1990). "The intergenerational correlation between children's adult earnings and their parents income: Results from the Michigan Panel Survey of Income Dynamics". *Review of Income and Wealth* 36.2, pp. 115–127.

Benaglia, Tatiana, Didier Chauveau, David R Hunter, and Derek S Young (2010). "mixtools: an R package for analyzing mixture models". *Journal of Statistical Software* 32, pp. 1–29.

Bhattacharya, Debopam and Bhashkar Mazumder (2011). "A nonparametric analysis of black–white differences in intergenerational income mobility in the United States". *Quantitative Economics* 2.3, pp. 335–379.

Björklund, Anders, Jesper Roine, and Daniel Waldenström (2012). "Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity?" *Journal of Public Economics* 96.5, pp. 474–484.

Black, Sandra E and Paul Devereux (2011). "Recent developments in intergenerational mobility". *Handbook of Labor Economics*. Ed. by Ashenfelter, Orley and David Card. Vol. 4.B. Amsterdam: North-Holland. Chap. 16, pp. 1487–1541.

Bollinger, Christopher R and Martin H David (1997). "Modeling discrete choice with response error: Food stamp participation". *Journal of the American Statistical Association* 92.439, pp. 827–835.

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

Bound, John and Alan B Krueger (1991). "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of Labor Economics* 9.1, pp. 1–24.

Bound, John, Michael Schoenbaum, Todd R Stinebrickner, and Timothy Waidmann (1999). "The dynamic effects of health on the labor force transitions of older workers". *Labour economics* 6.2, pp. 179–202.

Bouyé, Eric and Mark Salmon (2009). "Dynamic copula quantile regressions and tail area dynamic dependence in Forex markets". *The European Journal of Finance* 15.7-8, pp. 721–750.

Bratberg, Espen, Oivind Anti Nilsen, and Kjell Vaage (2007). "Trends in intergenerational mobility across offspring's earnings distrbution in Norway". *Industrial Relations* 46.1, pp. 112–129.

Bratsberg, Bernt et al. (2007). "Nonlinearities in intergenerational earnings mobility: Consequences for cross-country comparisons". *The Economic Journal* 117.519, pp. C72–C92.

Callaway, Brantly and Weige Huang (2019). "Local intergenerational elasticities". *Economics Bulletin* 39.2, pp. 919–928.

— (2020). "Distributional effects of a continuous treatment with an application on intergenerational mobility". *Oxford Bulletin of Economics and Statistics* 82.4, pp. 808–842.

Carroll, Raymond J and Peter Hall (1988). "Optimal rates of convergence for deconvolving a density". *Journal of the American Statistical Association* 83.404, pp. 1184–1186.

Cavanagh, Christopher and Robert P. Sherman (1998). "Rank estimators for monotonic index models". *Journal of Econometrics* 84.2, pp. 351–381.

Chen, Xiaohong and Yanqin Fan (2006). "Estimation of copula-based semiparametric time series models". *Journal of Econometrics* 130.2, pp. 307–335.

Chernozhukov, Victor, Ivan Fernandez-Val, and Alfred Galichon (2010). "Quantile and probability curves without crossing". *Econometrica* 78.3, pp. 1093–1125.

Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly (2013). "Inference on counterfactual distributions". *Econometrica* 81.6, pp. 2205–2268.

Chesher, Andrew (1991). "The effect of measurement error". *Biometrika* 78.3, pp. 451–462.

— (2017). "Understanding the effect of measurement error on quantile regressions". *Journal of Econometrics* 200.2, pp. 223–237.

Chetty, Raj and Nathaniel Hendren (2018). "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects". *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter (2020). "Race and economic opportunity in the United States: An intergenerational perspective". *The Quarterly Journal of Economics* 135.2, pp. 711–783.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014). "Where is the land of opportunity? The geography of intergenerational mobility in the United States". *The Quarterly Journal of Economics* 129.4, pp. 1553–1623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner (2014). "Is the United States still a land of opportunity? Recent trends in intergenerational mobility". *American Economic Review* 104.5, pp. 141–47.

Collins, William J and Marianne H Wanamaker (2017). "Up from slavery? African American intergenerational economic mobility since 1880". Working Paper.

Cosslett, S.R. (2004). "Efficient semiparametric estimation of censored and truncated regressions via a smoothed self-consistency equation". *Econometrica* 72.4, pp. 1277–1293.

De Nadai, Michele and Arthur Lewbel (2016). "Nonparametric errors in variables models with measurement errors on both sides of the equation". *Journal of Econometrics* 191, pp. 19–32.

Derumigny, Alexis and Jean-David Fermanian (2017). "About tests of the "simplifying" assumption for conditional copulas". *Dependence Modeling* 5.1, pp. 154–197.

Diggle, Peter J and Peter Hall (1993). "A Fourier approach to nonparametric deconvolution of a density estimate". *Journal of the Royal statistical society: series B (Methodological)* 55.2, pp. 523–531.

DiNardo, John, Nicole M Fortin, and Thomas Lemieux (1996). "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach". *Econometrica* 64.5, pp. 1001–1044.

Doukhan, Paul (1995). "Mixing". *Mixing: Properties and Examples*. Springer, pp. 15–23.

Eide, Eric R and Mark H Showalter (1999). "Factors affecting the transmission of earnings across generations: A quantile regression approach". *Journal of Human Resources*, pp. 253–267.

Fan, Jianqing (1991). "On the optimal rates of convergence for nonparametric deconvolution problems". *The Annals of Statistics*, pp. 1257–1272.

Fernández-Val, Iván and Martin Weidner (2016). "Individual and time effects in nonlinear panel models with large N, T". *Journal of Econometrics* 192.1, pp. 291–312.

Firpo, Sergio, Antonio F Galvao, and Suyong Song (2017). "Measurement errors in quantile regression models". *Journal of Econometrics* 198.1, pp. 146–164.

Flores, Carlos A (2007). "Estimation of dose-response functions and optimal doses with a continuous treatment". Working Paper.

Flores, Carlos A, Alfonso Flores-Lagunes, Arturo Gonzalez, and Todd C Neumann (2012). "Estimating the effects of length of exposure to instruction in a training program: The case of Job Corps". *Review of Economics and Statistics* 94.1, pp. 153–171.

Galvao, Antonio F and Liang Wang (2015). "Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment". *Journal of the American Statistical Association* 110.512, pp. 1528–1542.

Geraci, Marco and Matteo Bottai (2007). "Quantile regression for longitudinal data using the asymmetric Laplace distribution". *Biostatistics* 8.1, pp. 140–154.

Gijbels, Irene, Noel Veraverbeke, and Marel Omelka (2011). "Conditional copulas, association measures and their applications". *Computational Statistics & Data Analysis* 55.5, pp. 1919–1932.

Grawe, Nathan D (2004). "Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints". *Journal of Human Resources* 39.3, pp. 813–827.

Haider, Steven and Gary Solon (2006). "Life-cycle variation in the association between current and lifetime earnings". *American Economic Review* 96.4, pp. 1308–1320.

Hansen, Bruce (2022). *Econometrics*. Princeton University Press.

Hausman, Jerry (2001). "Mismeasured variables in econometric analysis: Problems from the right and problems from the left". *Journal of Economic Perspectives* 15.4, pp. 57–68.

Hausman, Jerry, Jason Abrevaya, and F.M. Scott-Morton (1998). "Misclassification of the dependent variable in a discrete-response setting". *Journal of Econometrics* 87.2, pp. 239–269.

Hausman, Jerry, Haoyang Liu, Ye Luo, and Christopher Palmer (2021). "Errors in the dependent variable of quantile regression models". *Econometrica* 89.2, pp. 849–873.

Hausman, Jerry A, Whitney K Newey, Hidehiko Ichimura, and James L Powell (1991). "Identification and estimation of polynomial errors-in-variables models". *Journal of Econometrics* 50.3, pp. 273–295.

Hausman, Jerry A, Whitney K Newey, and James L Powell (1995). "Nonlinear errors in variables estimation of some Engel curves". *Journal of Econometrics* 65.1, pp. 205–233.

Hirano, Keisuke and Guido W Imbens (2004). "The propensity score with continuous treatments". *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* 226164, pp. 73–84.

Hsiao, Cheng (1989). "Consistent estimation for some nonlinear errors-in-variables models". *Journal of econometrics* 41.1, pp. 159–185.

Hu, Yingyao (2017). "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics". *Journal of Econometrics* 200.2, pp. 154–168.

Jantti, Markus et al. (2006). "American exceptionalism in a new light: A comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States". Working Paper.

Jenkins, Stephen (1987). "Snapshots versus movies: 'Lifecycle biases' and the estimation of intergenerational earnings inheritance". *European Economic Review* 31.5, pp. 1149–1158.

Joe, Harry (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

Kato, Kengo, Antonio F Galvao Jr, and Gabriel V Montes-Rojas (2012). "Asymptotics for panel quantile regression models with individual effects". *Journal of Econometrics* 170.1, pp. 76–91.

Kitagawa, Toru, Martin Nybom, and Jan Stuhler (2018). "Measurement error and rank correlations". Working Paper.

Kourtellos, Andros, Christa Marr, and Chih Ming Tan (2020). "Local intergenerational mobility". *European Economic Review* 126, p. 103460.

Landersø, Rasmus and James J Heckman (2017). "The Scandinavian fantasy: The sources of intergenerational mobility in Denmark and the U.S." *The Scandinavian Journal of Economics* 119.1, pp. 178–230.

Lewbel, Arthur (1996). "Demand estimation with expenditure measurement errors on the left and right hand side". *Review of Economics and Statistics* 78.4, pp. 718–725.

Li, Tong (2000). "Estimation of nonlinear errors-in-variables models: A simulated minimum distance estimator". *Statistics and Probability Letters* 47.3, pp. 243–248.

— (2002). "Robust and consistent estimation of nonlinear errors-in-variables models". *Journal of Econometrics* 110.1, pp. 1–26.

Li, Tong and Quang Vuong (1998). "Nonparametric estimation of the measurement error model using multiple indicators". *Journal of Multivariate Analysis* 65.2, pp. 139–165.

Machado, José AF and José Mata (2005). "Counterfactual decomposition of changes in wage distributions using quantile regression". *Journal of Applied Econometrics* 20.4, pp. 445–465.

Mazumder, Bhashkar (2005). "Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data". *Review of Economics and Statistics* 87.2, pp. 235–255.

McLachlan, Geoffrey J, Sharon X Lee, and Suren I Rathnayake (2019). "Finite mixture models". *Annual Review of Statistics and its Application* 6, pp. 355–378.

Melly, Blaise (2005). "Decomposition of differences in distribution using quantile regression". *Labour Economics* 12.4, pp. 577–590.

Meyer, Bruce D, Wallace KC Mok, and James X Sullivan (2009). "The under-reporting of transfers in household surveys: Its nature and consequences". Working Paper.

Millimet, Daniel L, Hao Li, and Punarjit Roychowdhury (2020). "Partial identification of economic mobility: With an application to the United States". *Journal of Business & Economic Statistics* 38.4, pp. 732–753.

Murtazashvili, Irina (2012). "An alternative measure of intergenerational income mobility based on a random coefficient model". *Journal of Applied Econometrics* 27, pp. 1161–1173.

Murtazashvili, Irina, Di Liu, and Artem Prokhorov (2015). "Two-sample nonparametric estimation of intergenerational income mobility in the United States and Sweden". *Canadian Journal of Economics* 48.3, pp. 1733–1761.

Nagler, Thomas (2025). "Simplified vine copula models: State of science and affairs". *Risk Sciences*, p. 100022.

Nelsen, Roger (2007). *An Introduction to Copulas*. 2nd ed. Springer.

Nybom, Martin and Jan Stuhler (2016). "Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation". *Journal of Human Resources* 51.1, pp. 239–268.

— (2017). "Biases in standard measures of intergenerational income dependence". *Journal of Human Resources* 52.3, pp. 800–825.

Pischke, Jörn-Steffen (1995). "Measurement error and earnings dynamics: Some estimates from the PSID validation study". *Journal of Business & Economic Statistics* 13.3, pp. 305–314.

Præstgaard, Jens and Jon A Wellner (1993). "Exchangeably weighted bootstraps of the general empirical process". *The Annals of Probability*, pp. 2053–2086.

Prakasa Rao, Bhagavatula LS (2009). "Conditional independence, conditional mixing and conditional association". *Annals of the Institute of Statistical Mathematics* 61.2, pp. 441–460.

Richey, Jeremiah and Alicia Rosburg (2016). "Understanding intergenerational economic mobility by decomposing joint distributions". Working Paper.

— (2018). "Decomposing economic mobility transition matrices". *Journal of Applied Econometrics* 33.1, pp. 91–108.

Rosenthal, Jeffrey S (1995). "Minorization conditions and convergence rates for Markov chain Monte Carlo". *Journal of the American Statistical Association* 90.430, pp. 558–566.

Rothe, Christoph (2010). "Nonparametric estimation of distributional policy effects". *Journal of Econometrics* 155.1, pp. 56–70.

Rothe, Christoph and Dominik Wied (2020). "Estimating derivatives of function-valued parameters in a class of moment condition models". *Journal of Econometrics* 217.1, pp. 1–19.

Schennach, S. M. (2004). "Estimation of nonlinear models with measurement error". *Econometrica* 72.1, pp. 33–75.

— (2007). "Instrumental variable estimation of nonlinear errors-in-variables models". *Econometrica* 75.1, pp. 201–239.

— (2008). "Quantile regression with mismeasured covariates". *Econometric Theory* 24.4, pp. 1010–1043.

Sklar, Abe (1959). *Fonctions de répartition à n dimensions et leurs marges*. Publications de L Institut de Statistique de L Universite de Paris.

Solon, Gary (1992). "Intergenerational income mobility in the United States". *American Economic Review*, pp. 393–408.

— (1999). "Intergenerational mobility in the labor market". *Handbook of Labor Economics* 3, pp. 1761–1800.

Stefanski, Leonard A and Raymond J Carroll (1990). "Deconvolving kernel density estimators". *Statistics* 21.2, pp. 169–184.

van der Vaart, Aad W and Jon A Wellner (1996). *Weak Convergence and Empirical Processes*. Springer Science & Business Media.

Veraverbeke, Noel, Marek Omelka, and Irene Gijbels (2011). "Estimation of a conditional copula and association measures". *Scandinavian Journal of Statistics* 38.4, pp. 766–780.

Vogel, Thorsten (2006). "Reassessing intergenerational mobility in Germany and the United States: the impact of differences in lifecycle earnings patterns". Working Paper.

Wei, Ying and Raymond J Carroll (2009). "Quantile regression with measurement error". *Journal of the American Statistical Association* 104.487, pp. 1129–1143.

A Proofs of Results in Section 3

Proof of Proposition 1

Proof. The result holds immediately from Hausman, Liu, Luo, and Palmer (2021, Theorem 1). □

Proof of Theorem 1

Proof. Let $\phi_X(t_1, t_2) = \mathbb{E}\big[\exp(it_1 Y + it_2 T) \mid X\big]$ denote the characteristic function of the observed (Y, T) conditional on X. Let $\phi_X^*(t_1, t_2) = \mathbb{E}\big[\exp(it_1 Y^* + it_2 T^*) \mid X\big]$ denote the characteristic function of (Y^*, T^*) conditional on X and let $\phi_{U_{Y^*}}$ and $\phi_{U_{T^*}}$ denote the characteristic functions of U_{Y^*} and U_{T^*}, respectively. Then,

$$\phi_X(t_1, t_2) = \phi_X^*(t_1, t_2)\phi_{U_{Y^*}}(t_1)\phi_{U_{T^*}}(t_2)$$

which holds by Assumption 1. Further, $\phi_{U_{Y^*}}$ and $\phi_{U_{T^*}}$ are identified due to the result in Proposition 1. This implies

$$\phi_X^*(t_1, t_2) = \frac{\phi_X(t_1, t_2)}{\phi_{U_{Y^*}}(t_1)\phi_{U_{T^*}}(t_2)}$$

which holds because the characteristic functions of U_{Y^*} and U_{T^*} are non-vanishing. This proves that the joint distribution of $(Y^*, T^*) \mid X$ is identified, which implies that the copula is identified. □

B Proofs of Results in Section 5

B.1 Proof of Theorem 2

Proof. Under the conditions of Proposition 1, namely Assumptions 1 to 3 and Lemma 1, the proof of this part follows from a straightforward modification of existing results, e.g., Hansen (2022, Proof of Theorem 22.1). The details are therefore omitted. □

B.2 Proof of Theorem 3

Proof. **Part (a):**

The $(k+1) \times 1$ population first-order conditions with respect to $\big(\beta_{Y^*}(\tau)', \sigma_{Y^*}\big)'$ are given by

$$\mathbb{E}\left[X \cdot \int_{\mathcal{U}} \psi_\tau \big(Y - u - X'\beta_{Y^*}(\tau)\big) f_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma_{Y^*})\, du\right] = 0 \tag{A1}$$

$$\mathbb{E}\left[\int_{\mathcal{U}} \rho_\tau \big(Y - u - X'\beta_{Y^*}(\tau)\big) \cdot f^\partial_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma_{Y^*})\, du\right] = 0 \tag{A2}$$

where $f^\partial_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma_{Y^*}) := \frac{\partial}{\partial \sigma_{Y^*}} f_{U_{Y^*}|Y,X}(u \mid Y, X; \sigma_{Y^*})$. (A1) can be alternatively ex-

pressed as

$$
\begin{aligned}
0 =& \mathbb{E}\left[X \int_{\mathcal{U}} \psi_\tau\left(Y - u - X'\beta_{Y^*}(\tau)\right) f_{U_{Y^*}|Y,X}(u \mid Y, X)\, du\right] \\
=& (\tau - 1)\mathbb{E}[X] + \mathbb{E}\left[X \int_{\mathcal{U}} \mathbb{1}\{u \leq Y - X'\beta_{Y^*}(\tau)\} f_{U_{Y^*}|Y,X}(u \mid Y, X)\, du\right] \\
=& (\tau - 1)\mathbb{E}[X] + \mathbb{E}\left[X \cdot \mathrm{F}_{U_{Y^*}|Y,X}\left(Y - X'\beta_{Y^*}(\tau) \mid Y, X\right)\right].
\end{aligned} \tag{A1'}
$$

The following decomposition of the sample moment condition corresponding to (A1), for $\tau \in \mathcal{T}$ satisfies

$$
\begin{aligned}
o_p(n^{-1/2}) =& \frac{1}{nS} \sum_{i=1}^n \sum_{s=1}^S \left[X_i \psi_\tau\left(Y_i - U_{is} - X_i'\widehat{\beta}_{Y^*}(\tau)\right)\right] \tag{A3} \\
=:& \frac{1}{n} \sum_{i=1}^n X_i \int_{\mathcal{U}} \psi_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right) \widehat{f}^{(S)}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*})\, du \\
=& \frac{1}{n} \sum_{i=1}^n X_i \underbrace{\left\{\int_{\mathcal{U}}\left[\psi_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right)\left(\widehat{f}^{(S)}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) - \widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*})\right)\right] du\right\}}_{R_{i,MC} = O_p(S^{-1/2})} \\
& + \frac{1}{n} \sum_{i=1}^n X_i \underbrace{\int_{\mathcal{U}}\left[\psi_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right)\left(\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) - \widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\right) du\right]}_{B_{1n}} \\
& + \frac{1}{n} \sum_{i=1}^n X_i \underbrace{\int_{\mathcal{U}}\left[\psi_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right)\underbrace{\left(\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) - f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\right)}_{R_{i,MC}} du\right]}_{B_{2n}} \\
& + \frac{1}{n} \sum_{i=1}^n X_i \underbrace{\int_{\mathcal{U}}\left[\psi_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right) f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\, du\right]}_{B_{3n}}
\end{aligned}
$$

where $\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) := \dfrac{\left[X_i'\widehat{\beta}^\partial_{Y^*}\left(\widehat{\tau}_y(Y_i - u)\right)\right]^{-1}}{\int_{\mathcal{U}} f_{Y|U_{Y^*},X}(Y_i \mid \tilde{u}, X_i) f_{U_{Y^*}|X}(\tilde{u}; \widehat{\sigma}_{Y^*})\, d\tilde{u}} f_{U_{Y^*}}(u; \widehat{\sigma}_{Y^*})$ and $\widehat{\tau}_y(Y_i -$

$u) := \int_0^1 \mathbb{1}\{X_i'\widehat{\beta}_{Y^*}(\tilde{\tau}) \leq (Y_i - u)\}\, d\tilde{\tau}$ thanks to Assumption 7. The normalizing property of the MCMC algorithm ensures $\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma)$ is a proper pdf, i.e., integrates to one, uniformly in Γ_σ.

Thanks to Assumption 9 and that $|\psi_\tau(\cdot)| \leq 1$, the Monte Carlo approximation error $R_{i,MC} = O_p(S^{-1/2}) = o_p(n^{-1/2})$ for each $i \in \{1, \ldots, n\}$ since $n/S = o(1)$ following the argument used in the proof of Lemma 1 analogously.

Thanks to the differentiability of $f_{U_{Y^*}}(u; \sigma)$ with respect to σ (Assumption 10(e)), it follows from the Mean-Value Theorem (MVT) that

$$
\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) - \widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) = \widehat{f}^\partial_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \bar{\sigma}_{Y^*})(\widehat{\sigma}_{Y^*} - \sigma_{Y^*})
$$

for some $\bar{\sigma}_{Y^*}$ that satisfies $|\bar{\sigma}_{Y^*} - \sigma_{Y^*}| \leq |\hat{\sigma}_{Y^*} - \sigma_{Y^*}|$. It follows from the above that

$$B_{1n} = \frac{1}{n}\sum_{i=1}^{n} X_i \int_{\mathcal{U}} \left[\psi_\tau \left(Y_i - u - X_i'\hat{\beta}_{Y^*}(\tau) \right) \hat{f}^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \bar{\sigma}_{Y^*})\, du \right] (\hat{\sigma}_{Y^*} - \sigma_{Y^*}).$$

From Lemma 3, the Monte Carlo approximation error has the following representation:

$$R_{i,MC} := \left(\hat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) - f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) \right)$$

$$= \sum_{\ell=1}^{L} \hat{\mathbb{R}}_{\ell, f_{U_{Y^*}|Y,X}}(u, Y_i, X_i; \sigma_{Y^*})' \left(\hat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \right).$$

Thus,

$$B_{2n} = \sum_{\ell=1}^{L} \underbrace{\frac{1}{n}\sum_{i=1}^{n} \left\{ \int_{\mathcal{U}} \left[\psi_\tau(Y_i - u - X_i'\hat{\beta}_{Y^*}(\tau)) X_i \hat{\mathbb{R}}_{\ell, f_{U_{Y^*}|Y,X}}(u, Y_i, X_i; \sigma_{Y^*})'\, du \right] \right\}}_{\mathbb{R}_{\beta,\ell,n}(\tau)} \left(\hat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \right)$$

$$=: \sum_{\ell=1}^{L} \mathbb{R}_{\beta,\ell,n}(\tau) \left(\hat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \right).$$

In addition to Assumption 7, $f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma)$ is continuous in u, thus the resulting CDF $F_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})$ is continuously differentiable in u. By the MVT, $F_{U_{Y^*}|Y,X}(Y_i - X_i'\hat{\beta}_{Y^*}(\tau) \mid Y_i, X_i) = F_{U_{Y^*}|Y,X}(Y_i - X_i'\beta_{Y^*}(\tau) \mid Y_i, X_i) - f_{U_{Y^*}|Y,X}(Y_i - X_i'\bar{\beta}_{Y^*}(\tau) \mid Y_i, X_i) X_i'(\hat{\beta}_{Y^*}(\tau) - \beta_{Y^*}(\tau))$. It follows from the representation in (A1') that

$$B_{3n} = \frac{1}{n}\sum_{i=1}^{n} X_i \int_{\mathcal{U}} \left[\psi_\tau \left(Y_i - u - X_i'\hat{\beta}_{Y^*}(\tau) \right) f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\, du \right]$$

$$= (\tau - 1)\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} X_i \int_{\mathcal{U}} \left[\mathbb{1}\{u \leq Y_i - X_i'\hat{\beta}_{Y^*}(\tau)\} f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\, du \right]$$

$$= (\tau - 1)\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} X_i F_{U_{Y^*}|Y,X}(Y_i - X_i'\hat{\beta}_{Y^*}(\tau) \mid Y_i, X_i)$$

$$= (\tau - 1)\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} X_i F_{U_{Y^*}|Y,X}(Y_i - X_i'\beta_{Y^*}(\tau) \mid Y_i, X_i)$$

$$- \left\{ \frac{1}{n}\sum_{i=1}^{n} X_i X_i' f_{U_{Y^*}|Y,X}(Y_i - X_i'\bar{\beta}_{Y^*}(\tau) \mid Y_i, X_i) \right\} \left(\hat{\beta}_{Y^*}(\tau) - \beta_{Y^*}(\tau) \right).$$

Putting terms together,

$$
o_p(n^{-1/2})
$$

$$
= -\underbrace{\left\{ \frac{1}{n}\sum_{i=1}^{n} X_i \widehat{f}_{U_{Y^*}|Y,X}(Y_i - X_i'\bar{\beta}_{Y^*}(\tau) \mid Y_i, X_i) X_i' \right\}}_{\mathbb{A}_{\beta,n}(\tau)} \left(\widehat{\beta}_{Y^*}(\tau) - \beta_{Y^*}(\tau)\right)
$$

$$
+ \sum_{\ell=1}^{L} \mathbb{R}_{\beta,\ell,n}(\tau)\left(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell)\right)
$$

$$
+ \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i \int_{\mathcal{U}} \left[\psi_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right) \widehat{f}^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \bar{\sigma}_{Y^*})\, du \right](\widehat{\sigma}_{Y^*} - \sigma_{Y^*})}_{\mathbb{B}_{\beta,n}(\tau)}
$$

$$
+ \underbrace{(\tau - 1)\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n} X_i \mathrm{F}_{U_{Y^*}|Y,X}(Y_i - X_i'\beta_{Y^*}(\tau) \mid Y_i, X_i)}_{\mathbb{C}_{\beta,n}(\tau)}
$$

$$
=: \mathbb{A}_{\beta,n}(\tau)\left(\widehat{\beta}_{Y^*}(\tau) - \beta_{Y^*}(\tau)\right) + \sum_{\ell=1}^{L} \mathbb{R}_{\beta,\ell,n}(\tau)\left(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell)\right) + \mathbb{B}_{\beta,n}(\tau)(\widehat{\sigma}_{Y^*} - \sigma_{Y^*}) + \mathbb{C}_{\beta,n}(\tau).
$$
(A4)

Recalling that $\rho_\tau(w)$ is differentiable in w, the following decomposition of the sample moment corresponding to (A2), for $\tau \in (0,1)$ holds:

$$
o_p(n^{-1/2}) + O_p(S^{-1/2})
$$

$$
= \frac{1}{n}\sum_{i=1}^{n} \left[\int_{\mathcal{U}} \rho_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right) \widehat{f}^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*})du \right] \tag{A5}
$$

$$
= \frac{1}{n}\sum_{i=1}^{n} \int_{\mathcal{U}} \left(\rho_\tau\left(Y_i - u - X_i'\widehat{\beta}_{Y^*}(\tau)\right) - \rho_\tau\left(Y_i - u - X_i'\beta_{Y^*}(\tau)\right) \right) \widehat{f}^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*})\, du
$$

$$
+ \frac{1}{n}\sum_{i=1}^{n} \int_{\mathcal{U}} \rho_\tau\left(Y_i - u - X_i'\beta_{Y^*}(\tau)\right) \left(\widehat{f}^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) - \widehat{f}^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) \right) du
$$

$$
+ \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \sigma} \int_{\mathcal{U}} \rho_\tau\left(Y_i - u - X_i'\beta_{Y^*}(\tau)\right) \underbrace{\left(\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) - f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) \right)}_{R_{i,MC}} du
$$

$$
+ \frac{1}{n}\sum_{i=1}^{n} \int_{\mathcal{U}} \rho_\tau\left(Y_i - u - X_i'\beta_{Y^*}(\tau)\right) f^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\, du
$$

$$
=: -\underbrace{\left\{ \frac{1}{n}\sum_{i=1}^{n} \int_{\mathcal{U}} \left[\psi_\tau\left(Y_i - u - X_i'\bar{\beta}_{Y^*}(\tau)\right) \widehat{f}^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*})\, du \right] X_i' \right\}}_{\mathbb{A}_{\sigma,n}(\tau)} \left(\widehat{\beta}_{Y^*}(\tau) - \beta_{Y^*}(\tau)\right)
$$

$$
+ \underbrace{\left\{ \frac{1}{n}\sum_{i=1}^{n} \int_{\mathcal{U}} \rho_\tau\left(Y_i - u - X_i'\beta_{Y^*}(\tau)\right) \widehat{f}^{\partial\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \bar{\sigma}_{Y^*})\, du \right\}}_{\mathbb{B}_{\sigma,n}(\tau)} (\widehat{\sigma}_{Y^*} - \sigma_{Y^*})
$$

38

$$+ \sum_{\ell=1}^{L} \mathbb{R}_{\sigma,\ell,n}(\tau) \big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \big)$$

$$\underbrace{+ \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{U}} \rho_\tau \big(Y_i - u - X_i' \beta_{Y^*}(\tau) \big) f^{\partial}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) \, du}_{\mathbb{C}_{\sigma,n}(\tau)}$$

$$=: \mathbb{A}_{\sigma,n}(\tau) \big(\widehat{\beta}_{Y^*}(\tau) - \beta_{Y^*}(\tau) \big) + \sum_{\ell=1}^{L} \mathbb{R}_{\sigma,\ell,n}(\tau) \big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \big) + \mathbb{B}_{\sigma,n}(\tau)(\widehat{\sigma}_{Y^*} - \sigma_{Y^*}) + \mathbb{C}_{\sigma,n}(\tau).$$

$$(\text{A6})$$

where under the conditions of Lemma 3,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \sigma} \int_{\mathcal{U}} \rho_\tau \big(Y_i - u - X_i' \beta_{Y^*}(\tau) \big) \Big(\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) - f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) \Big) \, du$$

$$= \sum_{\ell=1}^{L} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \Big\{ \int_{\mathcal{U}} \Big[\rho_\tau \big(Y_i - u - X_i' \beta_{Y^*}(\tau) \big) \frac{\partial}{\partial \sigma} \widehat{\mathbb{R}}_{\ell, f_{U_{Y^*}|Y,X}}(u, Y_i, X_i; \sigma_{Y^*})' \, du \Big] \Big\}}_{\mathbb{R}_{\sigma,\ell,n}(\tau)} \times \big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \big)$$

$$=: \sum_{\ell=1}^{L} \mathbb{R}_{\sigma,\ell,n}(\tau) \big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \big).$$

Define

$$\mathbb{A}_n(\tau) := \underbrace{\begin{bmatrix} \mathbb{A}_{\beta,n}(\tau) \\ \mathbb{A}_{\sigma,n}(\tau) \end{bmatrix}}_{(K+1) \times K}, \ \mathbb{R}_{\ell,n}(\tau) := \underbrace{\begin{bmatrix} \mathbb{R}_{\beta,\ell,n}(\tau) \\ \mathbb{R}_{\sigma,\ell,n}(\tau) \end{bmatrix}}_{(K+1) \times K}, \ \mathbb{B}_n(\tau) := \underbrace{\begin{bmatrix} \mathbb{B}_{\beta,n}(\tau) \\ \mathbb{B}_{\sigma,n}(\tau) \end{bmatrix}}_{(K+1) \times 1}, \ \text{and} \ \mathbb{C}_n(\tau) := - \underbrace{\begin{bmatrix} \mathbb{C}_{\beta,n}(\tau) \\ \mathbb{C}_{\sigma,n}(\tau) \end{bmatrix}}_{(K+1) \times 1}.$$

Then, stacking both expanded moments (A4) and (A6) vertically obtains

$$\mathbb{A}_n(\tau) \big(\widehat{\beta}_{Y^*}(\tau) - \beta_{Y^*}(\tau) \big) + \sum_{\ell=1}^{L} \mathbb{R}_{\ell,n}(\tau) \big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \big) + \mathbb{B}_n(\tau)(\widehat{\sigma}_{Y^*} - \sigma_{Y^*}) = \mathbb{C}_n(\tau) + o_p(n^{-1/2}),$$

which, evaluated at each $\tau = \tau_{\ell'} \in \{\tau_1, \ldots, \tau_L\}$ (from Assumption 7) gives,

$$\big(\mathbb{A}_n(\tau_{\ell'}) + \mathbb{R}_{\ell',n}(\tau_{\ell'}) \big) \big(\widehat{\beta}_{Y^*}(\tau_{\ell'}) - \beta_{Y^*}(\tau_{\ell'}) \big)$$

$$+ \sum_{\ell \neq \ell'}^{L} \mathbb{R}_{\ell,n}(\tau_{\ell'}) \big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell) \big) + \mathbb{B}_n(\tau_{\ell'})(\widehat{\sigma}_{Y^*} - \sigma_{Y^*}) = \mathbb{C}_n(\tau_{\ell'}) + o_p(n^{-1/2}).$$

For example, the above evaluated at $\ell' = 1$ has the matrix form

$$
\underbrace{\left[\mathbb{A}_n(\tau_1) + \mathbb{R}_{1,n}(\tau_1) \ \middle| \ \mathbb{R}_{2,n}(\tau_1) \ \middle| \ \cdots \ \middle| \ \mathbb{R}_{L,n}(\tau_1) \ \middle| \ \mathbb{B}_n(\tau_1) \right]}_{(K+1) \times (LK+1)} \underbrace{\begin{bmatrix} \widehat{\beta}_{Y^*}(\tau_1) - \beta_{Y^*}(\tau_1) \\ \widehat{\beta}_{Y^*}(\tau_2) - \beta_{Y^*}(\tau_2) \\ \vdots \\ \widehat{\beta}_{Y^*}(\tau_L) - \beta_{Y^*}(\tau_L) \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix}}_{(LK+1) \times 1} = \underbrace{\mathbb{C}_n(\tau_1)}_{(K+1) \times 1} + o_p(n^{-1/2}).
$$

Next, stacking vertically across $\tau \in \{\tau_1, \ldots, \tau_L\}$,

$$
\underbrace{\begin{bmatrix} \mathbb{A}_n(\tau_1) + \mathbb{R}_{1,n}(\tau_1) & \mathbb{R}_{2,n}(\tau_1) & \cdots & \mathbb{R}_{L,n}(\tau_1) & \mathbb{B}_n(\tau_1) \\ \mathbb{R}_{1,n}(\tau_2) & \mathbb{A}_n(\tau_2) + \mathbb{R}_{2,n}(\tau_2) & \cdots & \mathbb{R}_{L,n}(\tau_2) & \mathbb{B}_n(\tau_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbb{R}_{1,n}(\tau_L) & \mathbb{R}_{2,n}(\tau_L) & \cdots & \mathbb{A}_n(\tau_L) + \mathbb{R}_{L,n}(\tau_L) & \mathbb{B}_n(\tau_L) \end{bmatrix}}_{L(K+1) \times (LK+1)} \underbrace{\begin{bmatrix} \widehat{\beta}_{Y^*}(\tau_1) - \beta_{Y^*}(\tau_1) \\ \widehat{\beta}_{Y^*}(\tau_2) - \beta_{Y^*}(\tau_2) \\ \vdots \\ \widehat{\beta}_{Y^*}(\tau_L) - \beta_{Y^*}(\tau_L) \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix}}_{(LK+1) \times 1} = \underbrace{\begin{bmatrix} \mathbb{C}_n(\tau_1) \\ \mathbb{C}_n(\tau_2) \\ \vdots \\ \mathbb{C}_n(\tau_L) \end{bmatrix}}_{L(K+1) \times 1} + o_p(n^{-1/2}).
$$

$$(A7)$$

Define

$$
\mathbb{H}_{Y^*}^L := \plim_{n \to \infty} \begin{bmatrix} \mathbb{A}_n(\tau_1) + \mathbb{R}_{1,n}(\tau_1) & \mathbb{R}_{2,n}(\tau_1) & \cdots & \mathbb{R}_{L,n}(\tau_1) & \mathbb{B}_n(\tau_1) \\ \mathbb{R}_{1,n}(\tau_2) & \mathbb{A}_n(\tau_2) + \mathbb{R}_{2,n}(\tau_2) & \cdots & \mathbb{R}_{L,n}(\tau_2) & \mathbb{B}_n(\tau_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbb{R}_{1,n}(\tau_L) & \mathbb{R}_{2,n}(\tau_L) & \cdots & \mathbb{A}_n(\tau_L) + \mathbb{R}_{L,n}(\tau_L) & \mathbb{B}_n(\tau_L) \end{bmatrix}, \quad (A8)
$$

$$
\left(\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \right) := \begin{bmatrix} \widehat{\beta}_{Y^*}(\tau_1) - \beta_{Y^*}(\tau_1) \\ \widehat{\beta}_{Y^*}(\tau_2) - \beta_{Y^*}(\tau_2) \\ \vdots \\ \widehat{\beta}_{Y^*}(\tau_L) - \beta_{Y^*}(\tau_L) \end{bmatrix}, \quad \text{and} \quad \mathbb{C}_{n,Y^*}^L := \begin{bmatrix} \mathbb{C}_n(\tau_1) \\ \mathbb{C}_n(\tau_2) \\ \vdots \\ \mathbb{C}_n(\tau_L) \end{bmatrix}.
$$

Then under the full column rank condition on $\mathbb{H}_{Y^*}^L$ in the theorem, $\mathbb{H}_{Y^*}^{L}{}'\mathbb{H}_{Y^*}^L$ is invertible. In addition to the conditions of Theorem 2, one obtains the asymptotically linear representation

$$
\sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix} = \left(\mathbb{H}_{Y^*}^{L}{}'\mathbb{H}_{Y^*}^L \right)^{-1} \mathbb{H}_{Y^*}^{L}{}' \sqrt{n} \mathbb{C}_{n,Y^*}^L + o_p(1)
$$

Part (b):

By (A1') and (A2),

$$
\mathbb{C}_n(\tau) = \frac{1}{n} \sum_{i=1}^n \left[\begin{array}{l} X_i\big((\tau - 1) + \mathrm{F}_{U_{Y^*}|Y,X}(Y_i - X_i'\beta_{Y^*}(\tau) \mid Y_i, X_i)\big) \\ \int_{\mathcal{U}} \rho_\tau (Y_i - u - X_i'\beta_{Y^*}(\tau)) f_{U_{Y^*}|Y,X}^{\partial}(u \mid Y_i, X_i; \sigma_{Y^*}) \, du \end{array} \right] =: \frac{1}{n} \sum_{i=1}^n \mathbb{C}_{in}(\tau)
$$

is mean-zero for any $\tau \in (0,1)$. This implies $\mathbb{E}[\mathbb{C}_{n,Y^*}^L] = 0$. Further, under Assumption 4 and

Assumption [10](a),

$$\mathbb{E}\big[\big\|X_i\big((\tau-1)+\mathrm{F}_{U_{Y^*}|Y,X}(Y_i-X_i'\beta_{Y^*}(\tau)\mid Y_i,X_i)\big)\big\|^2\big]\leq\mathbb{E}[\|X\|^2]\leq C.$$

Define $\mathcal{E}_{i\tau}^*:=Y_i-X_i'\beta_{Y^*}(\tau)$. Note that $\rho_\tau(w)\leq|w|$ and $\rho_\tau(w)$ is 1-Lipschitz. By the triangle inequality,

$$\left|\int_{\mathcal{U}}\rho_\tau\left(Y_i-u-X_i'\beta_{Y^*}(\tau)\right)f_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\,du\right|$$
$$=\left|\int_{\mathcal{U}}\left(\rho_\tau\left(\mathcal{E}_{i\tau}^*-u\right)-\rho_\tau(\mathcal{E}_{i\tau}^*)+\rho_\tau(\mathcal{E}_{i\tau}^*)\right)f_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\,du\right|$$
$$\leq\left|\int_{\mathcal{U}}\left(\rho_\tau\left(\mathcal{E}_{i\tau}^*-u\right)-\rho_\tau(\mathcal{E}_{i\tau}^*)\right)\cdot f_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\,du\right|+\rho_\tau(\mathcal{E}_{i\tau}^*)\left|\int_{\mathcal{U}}f_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\,du\right|$$
$$\leq\int_{\mathcal{U}}|u|\cdot\left|f_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\right|du=\int_{\mathcal{U}}\left|uf_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\right|du$$

since $\int_{\mathcal{U}}f_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\,du=\frac{\partial}{\partial\sigma}\int_{\mathcal{U}}f_{U_{Y^*}|Y,X}(u\mid Y_i,X_i;\sigma_{Y^*})\,du=\frac{\partial}{\partial\sigma}1=0$ holds under Assumption [10](b) by the dominance convergence theorem. From the foregoing and Assumption [10](e),

$$\mathbb{E}\left[\left|\int_{\mathcal{U}}\rho_\tau\left(Y_i-u-X_i'\beta_{Y^*}(\tau)\right)f_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\,du\right|^2\right]\leq\mathbb{E}\left[\left(\int_{\mathcal{U}}\left|uf_{U_{Y^*}|Y,X}^\partial(u\mid Y_i,X_i;\sigma_{Y^*})\right|du\right)^2\right]$$
$$\leq C.$$

Considering both bounds, $\mathbb{E}\big[\big\|\mathbb{C}_{in}(\tau)\big\|^2\big]\leq 2C<\infty$ whence $\mathbb{E}\big[\big\|\mathbb{C}_{iY^*}^L\big\|^2\big]\leq 2CL<\infty$ with

$$\mathbb{C}_{iY^*}^L:=\begin{bmatrix}\mathbb{C}_{in}(\tau_1)\\\mathbb{C}_{in}(\tau_2)\\\vdots\\\mathbb{C}_{in}(\tau_L)\end{bmatrix}.$$

The conclusion follows from the Multivariate Lindeberg-Lévy Central Limit Theorem (CLT) and the Continuous Mapping Theorem (CMT). $\qquad\square$

B.3 Proof of Corollary [1]

Proof. The estimator of $\mathrm{F}_{Y^*|X}(y\mid x)$ is given by $\widehat{\mathrm{F}}_{Y^*|X}(y\mid x)=\int_0^1\mathbb{1}\{x'\widehat{\beta}_{Y^*}(\tilde\tau)\leq y\}\,d\tilde\tau$. Thanks to the strict monotonicity of the map $\tau\mapsto x'\beta_{Y^*}(\tau)$ and its sample counterpart analogously, there exist unique values $\tau_y\in(0,1)$ and $\hat\tau_y$ such that $x'\beta_{Y^*}(\tau_y)=y$ and $x'\widehat{\beta}_{Y^*}(\hat\tau_y)=y$. Consider the function $\hat{h}(\tau):=y-x'\widehat{\beta}_{Y^*}(\tau)$. By Assumption [7] and the MVT,

$$\begin{aligned}0&=\hat{h}(\hat\tau_y)\\&=\hat{h}(\tau_y)+\hat{h}^\partial(\bar\tau_y)(\hat\tau_y-\tau_y)\\&=y-x'\widehat{\beta}_{Y^*}(\tau_y)-x'\widehat{\beta}_{Y^*}^\partial(\bar\tau_y)(\hat\tau_y-\tau_y)\\&=-x'\big(\widehat{\beta}_{Y^*}(\tau_y)-\beta_{Y^*}(\tau_y)\big)-x'\widehat{\beta}_{Y^*}^\partial(\bar\tau_y)(\hat\tau_y-\tau_y).\end{aligned}$$

Notice that under Assumption 2, $\tau_y = F_{Y^*|X}(y \mid x)$ and $\hat{\tau}_y = \widehat{F}_{Y^*|X}(y \mid x)$ by the definition of the quantile function and its estimator. Thus by Assumption 10(d), the Continuous Mapping Theorem (CMT) and Theorem 2,

$$
\sqrt{n}\big(\widehat{F}_{Y^*|X}(y \mid x) - F_{Y^*|X}(y \mid x)\big) = -\frac{1}{x'\widehat{\beta}^{\partial}_{Y^*}(\bar{\tau}_y)} x'\sqrt{n}\big(\widehat{\beta}_{Y^*}(\tau_y) - \beta_{Y^*}(\tau_y)\big)
$$

$$
= -f_{Y^*|X}(y \mid x)x'\sqrt{n}\big(\widehat{\beta}_{Y^*}(\tau_y) - \beta_{Y^*}(\tau_y)\big) + o_p(1).
$$

It follows from the foregoing, Assumption 7, and (A10) that

$$
\sqrt{n}\big(\widehat{F}_{Y^*|X}(y \mid x) - F_{Y^*|X}(y \mid x)\big) = -f_{Y^*|X}(y \mid x) \sum_{\ell=1}^{L} \omega_{\ell,\beta}(\tau_y)x'\sqrt{n}\big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell)\big) + o_p(1)
$$

$$
= \underbrace{-f_{Y^*|X}(y \mid x)\big[\omega_{1,\beta}(\tau_y)x', \ldots, \omega_{L,\beta}(\tau_y)x', 0\big]}_{\mathbb{M}^L_{FY}(y,x)'} \sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}}^L_{Y^*} - \boldsymbol{\beta}^L_{Y^*} \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\end{bmatrix} + o_p(1).
$$

The conclusion follows from the CMT and Theorem 3 with $\sigma(F_{Y^*|X}(y \mid x)) := \mathbb{M}^L_{FY}(y,x)'\Omega^L_{Y^*}\mathbb{M}^L_{FY}(y,x)$.

$\square$

B.4 Proof of Proposition 2

Proof. Consider the simulated maximum likelihood estimator

$$
\widehat{\delta} = \arg\max_{\delta} \sum_{i=1}^{n} \log\big(\widehat{L}_i^{(S)}(\delta)\big)
$$

where

$$
\widehat{L}_i^{(S)}(\delta) = \int_{\mathcal{U}\times\mathcal{U}} \Big[c_{Y^*T^*|X}\big(\widehat{F}_{Y^*|X}(Y_i - u \mid X_i), \widehat{F}_{T^*|X}(T_i - v \mid X_i); \delta\big) \times \widehat{f}_{Y^*|X}(Y_i - u|X_i)\widehat{f}_{T^*|X}(T_i - v|X_i)
$$

$$
\times \widehat{f}^{(S)}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*})\widehat{f}^{(S)}_{U_{T^*}|Y,X}(v \mid Y_i, X_i; \widehat{\sigma}_{T^*})\Big] du dv.
$$

Similarly, define

$$
\widehat{L}_i(\delta) := \int_{\mathcal{U}\times\mathcal{U}} \Big[c_{Y^*T^*|X}\big(\widehat{F}_{Y^*|X}(Y_i - u \mid X_i), \widehat{F}_{T^*|X}(T_i - v \mid X_i); \delta\big) \times \widehat{f}_{Y^*|X}(Y_i - u|X_i)\widehat{f}_{T^*|X}(T_i - v|X_i)
$$

$$
\times \widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*})\widehat{f}_{U_{T^*}|Y,X}(v \mid Y_i, X_i; \widehat{\sigma}_{T^*})\Big] du dv \quad \text{and}
$$

$$
L_i(\delta) := \int_{\mathcal{U}\times\mathcal{U}} \Big[c_{Y^*T^*|X}\big(F_{Y^*|X}(Y_i - u \mid X_i), F_{T^*|X}(T_i - v \mid X_i); \delta\big) \times f_{Y^*|X}(Y_i - u|X_i)f_{T^*|X}(T_i - v|X_i)
$$

$$
\times f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})f_{U_{T^*}|Y,X}(v \mid Y_i, X_i; \sigma_{T^*})\Big] du dv.
$$

Let

$$
\widehat{\mathcal{S}}_i^{(S)}(\delta) := \nabla_\delta \log \widehat{L}_i^{(S)}(\delta) = \frac{\nabla_\delta \widehat{L}_i^{(S)}(\delta)}{\widehat{L}_i^{(S)}(\delta)}, \qquad \mathcal{H}_i(\delta) := \nabla_\delta^2 \log L_i(\delta) = \frac{\nabla_\delta^2 L_i(\delta)}{L_i(\delta)} - \frac{\nabla_\delta L_i(\delta)\,\nabla_\delta L_i(\delta)'}{L_i(\delta)^2},
$$

$$\nabla_\delta L_i(\delta) = \iint c_\delta(\cdot; \delta) \, f_{Y^*|X} \, f_{T^*|X} \, f_{U_{Y^*}|Y,X} \, f_{U_{T^*}|Y,X} \, du \, dv, \quad \text{and}$$

$$\nabla_\delta^2 L_i(\delta) = \iint c_{\delta\delta}(\cdot; \delta) \, f_{Y^*|X} \, f_{T^*|X} \, f_{U_{Y^*}|Y,X} \, f_{U_{T^*}|Y,X} \, du \, dv$$

where $c_\delta(\cdot, \cdot; \delta) := \frac{\partial}{\partial \delta} c_{Y^*T^*|X}(\cdot, \cdot; \delta)$ and $c_{\delta\delta}(\cdot, \cdot; \delta) := \frac{\partial^2}{\partial\delta\partial\delta'} c_{Y^*T^*|X}(\cdot, \cdot; \delta)$.

Granted the differentiability of $c_{Y^*T^*|X}(\cdot, \cdot; \delta)$ (Assumption 6), and that $\widehat\delta$ is the maximizer, the following first-order condition holds:

$$
\begin{aligned}
0 &= \sum_{i=1}^n \widehat{\mathcal{S}}_i^{(S)}(\widehat\delta) \\
&= \sum_{i=1}^n \widehat{\mathcal{S}}_i(\widehat\delta) + o_p(n^{-1/2}) \\
&= \sum_{i=1}^n \mathcal{S}_i(\widehat\delta) + \sum_{i=1}^n \left(\widehat{\mathcal{S}}_i(\widehat\delta) - \mathcal{S}_i(\widehat\delta) \right) + o_p(n^{-1/2}) \\
&=: \sum_{i=1}^n \mathcal{S}_i(\widehat\delta) + \sum_{i=1}^n \Delta_i(\widehat\delta; \widehat\cdot - \cdot) + o_p(n^{-1/2})
\end{aligned}
$$

where $\mathcal{S}_i(\delta) := \nabla_\delta \log L_i(\delta)$. The second equality holds under Assumptions 6, 9 and 10 and the condition $n/S = o(1)$.

Under Assumption 6, the copula density function is twice continuously differentiable with respect to δ. Plugging in a first–order expansion of $\mathcal{S}_i(\widehat\delta)$ around δ into the first-order condition then gives

$$o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^n \mathcal{S}_i(\delta) + \left(\frac{1}{n}\sum_{i=1}^n \mathcal{H}_i(\delta)\right)(\widehat\delta - \delta) + \frac{1}{n}\sum_{i=1}^n \Delta_i(\widehat\delta; \widehat\cdot - \cdot)$$

where $\mathbb{E}[\mathcal{S}_i(\delta)] = 0$. Therefore,

$$\sqrt{n}(\widehat\delta - \delta) = -\left[\frac{1}{n}\sum_{i=1}^n \mathcal{H}_i(\delta)\right]^{-1} \left\{ \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{S}_i(\delta) + \frac{1}{\sqrt{n}}\sum_{i=1}^n \Delta_i(\widehat\delta; \widehat\cdot - \cdot) \right\} + o_p(1) \qquad \text{(A9)}$$

where $\Delta_i(\widehat{\delta}; \widehat{\cdot} - \cdot)$ has the following weighted sum representation:

$$
\begin{aligned}
\Delta_i(\widehat{\delta}; \widehat{\cdot} - \cdot) :&= \widehat{\mathcal{S}}_i(\widehat{\delta}) - \mathcal{S}_i(\widehat{\delta}) \\
&= \nabla_\delta\big(\log \widehat{L}_i(\widehat{\delta}) - \log L_i(\widehat{\delta})\big) \\
&= \widehat{L}_i(\widehat{\delta})^{-1}\nabla_\delta\big(\widehat{L}_i(\widehat{\delta}) - L_i(\widehat{\delta})\big) - \big(\widehat{L}_i(\widehat{\delta})L_i(\widehat{\delta})\big)^{-1}\nabla_\delta L_i(\widehat{\delta})\big(\widehat{L}_i(\widehat{\delta}) - L_i(\widehat{\delta})\big) \\
&=: \int\left(\int W_i^{(F_Y)}(u,v;\widehat{\delta})\,dv\right)\big(\widehat{F}_{Y^*|X} - F_{Y^*|X}\big)(Y_i - u \mid X_i)\,du \\
&\quad + \int\left(\int W_i^{(f_Y)}(u,v;\widehat{\delta})\,dv\right)\big(\widehat{f}_{Y^*|X} - f_{Y^*|X}\big)(Y_i - u \mid X_i)\,du \\
&\quad + \int\left(\int W_i^{(f_{U_Y})}(u,v;\widehat{\delta})\,dv\right)\big(\widehat{f}_{U_{Y^*}|Y,X} - f_{U_{Y^*}|Y,X}\big)(u \mid Y_i, X_i)\,du \\
&\quad + \int\left(\int W_i^{(F_T)}(u,v;\widehat{\delta})\,du\right)\big(\widehat{F}_{T^*|X} - F_{T^*|X}\big)(T_i - v \mid X_i)\,dv \\
&\quad + \int\left(\int W_i^{(f_T)}(u,v;\widehat{\delta})\,du\right)\big(\widehat{f}_{T^*|X} - f_{T^*|X}\big)(T_i - v \mid X_i)\,dv \\
&\quad + \int\left(\int W_i^{(f_{U_T})}(u,v;\widehat{\delta})\,du\right)\big(\widehat{f}_{U_{T^*}|Y,X} - f_{U_{T^*}|Y,X}\big)(v \mid Y_i, X_i)\,dv.
\end{aligned}
$$

First, from the proof of Corollary 1,

$$
\sqrt{n}\big(\widehat{F}_{Y^*|X} - F_{Y^*|X}\big)(Y_i - u \mid X_i) = \mathbb{M}_{FY}^L(Y_i - u, X_i)'\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\end{bmatrix} + o_p(1).
$$

Second, from Lemma 3(a) and the conditions of Theorem 2,

$$
\begin{aligned}
\sqrt{n}\big(\widehat{f}_{Y^*|X} - f_{Y^*|X}\big)(Y_i - u \mid X_i) &= \sum_{\ell=1}^{L}\widehat{\mathbb{R}}_{\ell, f_{Y^*|X}}(Y_i - u, X_i)'\,\sqrt{n}\big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell)\big) \\
&= \begin{bmatrix}\widehat{\mathbb{R}}_{1, f_{Y^*|X}}(Y_i - u, X_i)' & \cdots & \widehat{\mathbb{R}}_{L, f_{Y^*|X}}(Y_i - u, X_i)' & 0\end{bmatrix}\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\end{bmatrix} \\
&=: \mathbb{M}_{f_{Y^*|X}}^L(Y_i - u, X_i)\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\end{bmatrix} + o_p(1).
\end{aligned}
$$

Lastly, under Assumption 10(e), the conditions of Lemma 3(b), and the conditions of Theorem 2,

$$
\begin{aligned}
&\sqrt{n}\big(\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) - f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\big) \\
&= \sqrt{n}\big(\widehat{f}_{U_{Y^*}|Y,X} - f_{U_{Y^*}|Y,X}\big)(u \mid Y_i, X_i; \sigma_{Y^*}) + \sqrt{n}\big(\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \widehat{\sigma}_{Y^*}) - \widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\big) \\
&= \sum_{\ell=1}^{L}\widehat{\mathbb{R}}_{\ell, f_{U_{Y^*}|Y,X}}(u, Y_i, X_i; \sigma_{Y^*})'\,\sqrt{n}\big(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell)\big) + \widehat{f}_{U_{Y^*}|Y,X}^{\partial}(u \mid Y_i, X_i; \bar{\sigma}_{Y^*})\sqrt{n}(\widehat{\sigma}_{Y^*} - \sigma_{Y^*}) \\
&= \begin{bmatrix}\widehat{\mathbb{R}}_{1, f_{U_{Y^*}|Y,X}}(u, Y_i, X_i; \sigma_{Y^*})', \ldots, \widehat{\mathbb{R}}_{L, f_{U_{Y^*}|Y,X}}(u, Y_i, X_i; \sigma_{Y^*})' & \widehat{f}_{U_{Y^*}|Y,X}^{\partial}(u \mid Y_i, X_i; \bar{\sigma}_{Y^*})\end{bmatrix}\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\end{bmatrix} \\
&=: \mathbb{M}_{f_{U_{Y^*}|Y,X}}^L(u; Y_i, X_i)'\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*}\end{bmatrix} + o_p(1).
\end{aligned}
$$

Putting together the foregoing, and applying arguments with respect to Y^* analogously for T^*,

$$
\begin{aligned}
\sqrt{n}\Delta_i(\widehat{\delta};\widehat{\cdot}-\cdot) = & \left\{ \int \left[\left(\int W_i^{(F_Y)}(u,v;\widehat{\delta})\,dv \right) \mathbb{M}_{FY}^L(Y_i-u,X_i)' \right. \right. \\
& + \left(\int W_i^{(f_Y)}(u,v;\widehat{\delta})\,dv \right) \mathbb{M}_{f_{Y^*|X}}^L(Y_i-u,X_i)' \\
& \left. + \left(\int W_i^{(f_{U_Y})}(u,v;\widehat{\delta})\,dv \right) \mathbb{M}_{f_{U_{Y^*|Y,X}}}^L(u;Y_i,X_i)' \right] du \right\} \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix} \\
& + \left\{ \int \left[\left(\int W_i^{(F_T)}(u,v;\widehat{\delta})\,du \right) \mathbb{M}_{FT}^L(T_i-v,X_i)' \right. \right. \\
& + \left(\int W_i^{(f_T)}(u,v;\widehat{\delta})\,du \right) \mathbb{M}_{f_{T^*|X}}^L(T_i-v,X_i)' \\
& \left. + \left(\int W_i^{(f_{U_T})}(u,v;\widehat{\delta})\,du \right) \mathbb{M}_{f_{U_{T^*|T,X}}}^L(v;T_i,X_i)' \right] dv \right\} \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{T^*}^L - \boldsymbol{\beta}_{T^*}^L \\ \widehat{\sigma}_{T^*} - \sigma_{T^*} \end{bmatrix} + o_p(1) \\
=: & \ \mathbb{M}_{i\Delta Y}^L(Y_i,X_i;\widehat{\delta})' \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix} + \mathbb{M}_{i\Delta T}^L(Y_i,X_i;\widehat{\delta})' \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{T^*}^L - \boldsymbol{\beta}_{T^*}^L \\ \widehat{\sigma}_{T^*} - \sigma_{T^*} \end{bmatrix} + o_p(1).
\end{aligned}
$$

Plugging the above into (A9) and applying the CMT,

$$
\begin{aligned}
\sqrt{n}(\widehat{\delta}-\delta) = & -\mathcal{H}_\delta^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{S}_i(\delta) - \mathcal{H}_\delta^{-1} \frac{1}{n} \sum_{i=1}^n \mathbb{M}_{i\Delta Y}^L(Y_i,X_i;\widehat{\delta})' \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix} \\
& - \mathcal{H}_\delta^{-1} \frac{1}{n} \sum_{i=1}^n \mathbb{M}_{i\Delta T}^L(Y_i,X_i;\widehat{\delta})' \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{T^*}^L - \boldsymbol{\beta}_{T^*}^L \\ \widehat{\sigma}_{T^*} - \sigma_{T^*} \end{bmatrix} + o_p(1) \\
= & -\mathcal{H}_\delta^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{S}_i(\delta) - \mathcal{H}_\delta^{-1} \mathbb{M}_{\Delta Y}^{L\prime} \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{Y^*}^L - \boldsymbol{\beta}_{Y^*}^L \\ \widehat{\sigma}_{Y^*} - \sigma_{Y^*} \end{bmatrix} - \mathcal{H}_\delta^{-1} \mathbb{M}_{\Delta T}^{L\prime} \sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{T^*}^L - \boldsymbol{\beta}_{T^*}^L \\ \widehat{\sigma}_{T^*} - \sigma_{T^*} \end{bmatrix} + o_p(1) \\
=: & \ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i^\delta + o_p(1) \xrightarrow{d} \mathcal{N}(0,\Sigma(\delta))
\end{aligned}
$$

where $\Sigma(\delta) = \mathbb{E}\left[\psi^\delta \psi^{\delta\prime}\right]$, $\mathcal{H}_\delta := \underset{n\to\infty}{\operatorname{plim}} \frac{1}{n} \sum_{i=1}^n \mathcal{H}_i(\delta)$ and $\mathbb{M}_{\Delta j}^L := \underset{n\to\infty}{\operatorname{plim}} \frac{1}{n} \sum_{i=1}^n \mathbb{M}_{i\Delta j}^L(Y_i,X_i;\widehat{\delta})$, $j \in \{Y,T\}$. Since a linear combination of asymptotically linear quantities is asymptotically linear, the concluding part follows from the CMT, Theorem 3, in addition to the condition $\mathbb{E}[\|\mathcal{S}(\delta)\|^2] < \infty$ which is guaranteed under Assumptions 6 and 10. $\qquad\square$

B.5   Proof of Theorem 4

We provide the proofs of the first result in Theorem 4 and of bootstrap validity in this subsection. The proofs of the remaining parts are provided in the Supplementary Appendix.

B.5.1   Proof of Theorem 4(a)

The following lemma provides a proof of part (a) of Theorem 4.

Lemma 2. *Suppose Assumptions [1] to [10] hold, then* $\sqrt{n}\big(\widehat{\mathrm{F}}_{Y^*T^*|X}(y,t\mid x) - \mathrm{F}_{Y^*T^*|X}(y,t\mid x)\big) \xrightarrow{d} \mathcal{N}\big(0,\,\sigma(\mathrm{F}_{Y^*T^*|X}(y,t\mid x))\big).$

Proof. Recall $\mathrm{F}_{Y^*T^*|X}(y,t\mid x) = C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \mathrm{F}_{T^*|X}(t\mid x)\mid x\big) = C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \mathrm{F}_{T^*|X}(t\mid x); \delta\big)$ and the estimator is given by $\widehat{\mathrm{F}}_{Y^*T^*|X}(y,t\mid x) = C_{Y^*T^*|X}\big(\widehat{\mathrm{F}}_{Y^*|X}(y\mid x), \widehat{\mathrm{F}}_{T^*|X}(t\mid x); \widehat{\delta}\big)$. Consider the following decomposition:

$$
\begin{aligned}
&\sqrt{n}\big(\widehat{\mathrm{F}}_{Y^*T^*|X}(y,t\mid x) - \mathrm{F}_{Y^*T^*|X}(y,t\mid x)\big) \\
&= \sqrt{n}\Big(C_{Y^*T^*|X}\big(\widehat{\mathrm{F}}_{Y^*|X}(y\mid x), \widehat{\mathrm{F}}_{T^*|X}(t\mid x); \widehat{\delta}\big) - C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \mathrm{F}_{T^*|X}(t\mid x); \delta\big)\Big) \\
&= \sqrt{n}\Big(C_{Y^*T^*|X}\big(\widehat{\mathrm{F}}_{Y^*|X}(y\mid x), \widehat{\mathrm{F}}_{T^*|X}(t\mid x); \widehat{\delta}\big) - C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \widehat{\mathrm{F}}_{T^*|X}(t\mid x); \widehat{\delta}\big)\Big) \\
&\quad + \sqrt{n}\Big(C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \widehat{\mathrm{F}}_{T^*|X}(t\mid x); \widehat{\delta}\big) - C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \mathrm{F}_{T^*|X}(t\mid x); \widehat{\delta}\big)\Big) \\
&\quad + \sqrt{n}\Big(C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \mathrm{F}_{T^*|X}(t\mid x); \widehat{\delta}\big) - C_{Y^*T^*|X}\big(\mathrm{F}_{Y^*|X}(y\mid x), \mathrm{F}_{T^*|X}(t\mid x); \delta\big)\Big).
\end{aligned}
$$

Let $\tau_y := \mathrm{F}_{Y^*|X}(y\mid x)$ and $\tau_t := \mathrm{F}_{T^*|X}(t\mid x)$. Since (1) $C_{Y^*T^*|X}(\tau_y, \tau_t; \delta)$ is continuously differentiable in $(\tau_y, \tau_t, \delta')' \in \mathcal{T}^2 \times \Gamma_\delta$, with partials $C_1 = \partial C/\partial u$, $C_2 = \partial C/\partial v$, and C_δ (Assumption [6]), it follows from the above decomposition and the CMT that

$$
\begin{aligned}
&\sqrt{n}\big(\widehat{\mathrm{F}}_{Y^*T^*|X}(y,t\mid x) - \mathrm{F}_{Y^*T^*|X}(y,t\mid x)\big) \\
&= C_1(\tau_y, \tau_t; \delta)\sqrt{n}\big(\widehat{\mathrm{F}}_{Y^*|X}(y\mid x) - \mathrm{F}_{Y^*|X}(y\mid x)\big) + C_2(\tau_y, \tau_t; \delta)\sqrt{n}\big(\widehat{\mathrm{F}}_{T^*|X}(t\mid x) - \mathrm{F}_{T^*|X}(t\mid x)\big) \\
&\quad + C_\delta(\tau_y, \tau_t; \delta))'\,\sqrt{n}\big(\widehat{\delta} - \delta\big)\; +\; o_p(1) \\
&=: \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i^{(a)} + o_p(1) \xrightarrow{d} \mathcal{N}(0, \sigma(F_{Y^*T^*|X}(y,t|x)))
\end{aligned}
$$

where $\sigma(F_{Y^*T^*|X}(y,t|x)) = \mathbb{E}\big[(\psi^{(a)})^2\big]$. Since a linear combination of asymptotically linear quantities is asymptotically linear, the conclusion follows from Corollary [1] (applied analogously to T), Proposition [2], and the CMT. □

B.5.2 Validity of the Bootstrap

Given our previous results establishing the limiting distributions of the QR parameters, measurement error parameters, copula parameters, and target parameters, the validity of the bootstrap follows from standard arguments that closely parallel the arguments provided above. For brevity, we only sketch these arguments here. To start with, we consider the validity of the bootstrap for the QR parameters and measurement error parameters. The bootstrap version of these estimators is defined by replacing the score equations (Equations [A3] and [A5]) with their weighted counterparts, where each observation is multiplied by bootstrap weights w_i drawn independently of the data with mean one and variance one (for the empirical bootstrap, these are multinomial weights). Then, the validity of the bootstrap holds by linearizing the weighted score equations around the original estimates (which follows using an analogous argument to the one provided above), and then employing the bootstrap CLT (Præstgaard and Wellner (1993)). Finally, the validity of the bootstrap for the target parameters in Theorem [4] follows using similar arguments

to the ones provided in the proof of Theorem 4 in conjunction with the bootstrap delta method (van der Vaart and Wellner (1996)).

C Supporting Lemmas

Lemma 3. *Under Assumption 7,*

$$(a) \ \left(\widehat{f}_{Y^*|X}(y \mid x) - f_{Y^*|X}(y \mid x)\right) = \sum_{\ell=1}^{L} \widehat{\mathbb{R}}_{\ell, f_{Y^*|X}}(y, x)' \left(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell)\right) \ and$$

$$(b) \ \left(\widehat{f}_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*}) - f_{U_{Y^*}|Y,X}(u \mid Y_i, X_i; \sigma_{Y^*})\right) = \sum_{\ell=1}^{L} \widehat{\mathbb{R}}_{\ell, f_{U_{Y^*}|Y,X}}(u, Y_i, X_i; \sigma_{Y^*})' \left(\widehat{\beta}_{Y^*}(\tau_\ell) - \beta_{Y^*}(\tau_\ell)\right)$$

for some random vectors $\widehat{\mathbb{R}}_{\ell, f_{Y^*|X}}(y, x)$ *and* $\widehat{\mathbb{R}}_{\ell, f_{U_{Y^*}|Y,X}}(u, \sigma_{Y^*}, Y_i, X_i)$, $\ell \in \{1, \ldots, L\}$.

Proof. **Part (a):**

Recall $f_{Y^*|X}(y \mid x) = \left[x'\beta_{Y^*}^{\partial}\left(F_{Y^*|X}(y \mid x)\right)\right]^{-1}$ with the estimator $\widehat{f}_{Y^*|X}(y \mid x) = \left[x'\widehat{\beta}_{Y^*}^{\partial}\left(\widehat{F}_{Y^*|X}(y \mid x)\right)\right]^{-1}$. Let $\tau_y := F_{Y^*|X}(y \mid x)$ and $\widehat{\tau}_y := \widehat{F}_{Y^*|X}(y \mid x)$, then

$$\begin{aligned}
\left(\widehat{f}_{Y^*|X}(y \mid x) - f_{Y^*|X}(y \mid x)\right) &= \left[x'\widehat{\beta}_{Y^*}^{\partial}\left(\widehat{\tau}_y\right)\right]^{-1} - \left[x'\beta_{Y^*}^{\partial}\left(\tau_y\right)\right]^{-1} \\
&= -\left[\widehat{\beta}_{Y^*}^{\partial}\left(\widehat{\tau}_y\right)'xx'\beta_{Y^*}^{\partial}\left(\tau_y\right)\right]^{-1} \\
&\quad \times x'\left(\widehat{\beta}_{Y^*}^{\partial}\left(\widehat{\tau}_y\right) - \beta_{Y^*}^{\partial}\left(\tau_y\right)\right).
\end{aligned}$$

Thanks to Assumption 7, the process $\beta_{Y^*}(\cdot)$ is piecewise linear. Thus, on each open interval between knots, $\beta_{Y^*}(\cdot)$ is linear and the second derivative is zero. At each knot $\tau_\ell, \ell \in \{1, \ldots, L\}$, the first derivative has a jump and the second derivative does not exist (unless the slopes match there). Since L is finite, the set of knots has Lebesgue measure zero and thus, $\beta_{Y^*}^{\partial\partial}(\tau) := \frac{\partial^2 \beta_{Y^*}(\tau)}{\partial \tau^2} = 0$ for Lebesgue-almost everywhere $\tau \in \mathcal{T}$. Further, under Assumption 7, the indices in the set are evenly spaced, i.e., $\tau_\ell = \tau_1 + (\ell - 1)\epsilon$, $\epsilon := \tau_2 - \tau_1$, and

$$\begin{aligned}
\beta_{Y^*}(\tau) &= \sum_{\ell=1}^{L-1} \mathbb{1}\{\tau \in (\tau_\ell, \tau_{\ell+1})\}\left(\beta_{Y^*}(\tau_\ell) + \frac{\tau - \tau_\ell}{\epsilon}\left(\beta_{Y^*}(\tau_{\ell+1}) - \beta_{Y^*}(\tau_\ell)\right)\right) + \sum_{\ell=1}^{L} \mathbb{1}\{\tau = \tau_\ell\}\beta_{Y^*}(\tau_\ell) \\
&= \sum_{\ell=1}^{L}\left[\mathbb{1}\{\ell \leq L-1\}\,\mathbb{1}\{\tau \in [\tau_\ell, \tau_{\ell+1})\}\left(1 - \tfrac{\tau - \tau_\ell}{\epsilon}\right) + \mathbb{1}\{\ell \geq 2\}\,\mathbb{1}\{\tau \in (\tau_{\ell-1}, \tau_\ell)\}\tfrac{\tau - \tau_{\ell-1}}{\epsilon}\right]\beta_{Y^*}(\tau_\ell) \\
&=: \sum_{\ell=1}^{L} \omega_{\ell,\beta}(\tau)\,\beta_{Y^*}(\tau_\ell). \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (A10)
\end{aligned}$$

Hence

$$\frac{\partial}{\partial \tau}\beta_{Y^*}(\tau) = \sum_{\ell=1}^{L} \omega_{\ell,\beta}^{\partial}(\tau)\beta_{Y^*}(\tau_\ell)$$

where $\omega_{\ell,\beta}^{\partial}(\tau) := \epsilon^{-1}\left(\mathbb{1}\{\ell \geq 2\}\,\mathbb{1}\{\tau \in [\tau_{\ell-1}, \tau_\ell)\} - \mathbb{1}\{\ell \leq L-1\}\,\mathbb{1}\{\tau \in (\tau_\ell, \tau_{\ell+1})\}\right)$.

Thus,

$$
\begin{aligned}
\widehat{\beta}_{Y*}^{\partial}(\widehat{\tau}_y) - \beta_{Y*}^{\partial}(\tau_y) &= \left(\widehat{\beta}_{Y*}^{\partial}(\widehat{\tau}_y) - \beta_{Y*}^{\partial}(\widehat{\tau}_y)\right) + \left(\beta_{Y*}^{\partial}(\widehat{\tau}_y) - \beta_{Y*}^{\partial}(\tau_y)\right) \\
&= \frac{\partial}{\partial\tau}\left(\widehat{\beta}_{Y*}(\widehat{\tau}_y) - \beta_{Y*}(\widehat{\tau}_y)\right) + \beta_{Y*}^{\partial\partial}(\bar{\tau}_y)(\widehat{\tau}_y - \tau_y) \\
&= \frac{\partial}{\partial\tau}\left(\widehat{\beta}_{Y*}(\widehat{\tau}_y) - \beta_{Y*}(\widehat{\tau}_y)\right) \\
&= \sum_{\ell=1}^{L} \omega_{\ell,\beta}^{\partial}(\widehat{\tau}_y)\left(\widehat{\beta}_{Y*}(\tau_\ell) - \beta_{Y*}(\tau_\ell)\right).
\end{aligned}
$$

Putting terms together,

$$
\begin{aligned}
\left(\widehat{f}_{Y*|X}(y\mid x) - f_{Y*|X}(y\mid x)\right) &= -\left[\widehat{\beta}_{Y*}^{\partial}(\widehat{\tau}_y)' xx' \beta_{Y*}^{\partial}(\tau_y)\right]^{-1} x' \sum_{\ell=1}^{L} \omega_{\ell,\beta}^{\partial}(\widehat{\tau}_y)\left(\widehat{\beta}_{Y*}(\tau_\ell) - \beta_{Y*}(\tau_\ell)\right) \\
&=: \sum_{\ell=1}^{L} \widehat{\mathbb{R}}_{\ell, f_{Y*|X}}(y, x)' \left(\widehat{\beta}_{Y*}(\tau_\ell) - \beta_{Y*}(\tau_\ell)\right). \quad\quad\quad (A11)
\end{aligned}
$$

Part (b):

From part (a) above,

$$
\begin{aligned}
&\left(\widehat{f}_{U_{Y*}|Y,X}(u\mid Y_i, X_i; \sigma_{Y*}) - f_{U_{Y*}|Y,X}(u\mid Y_i, X_i; \sigma_{Y*})\right) \\
&= \frac{f_{U_{Y*}}(u; \sigma_{Y*})}{\int_{\mathcal{U}} f_{Y|U_{Y*},X}(Y_i\mid\tilde{u}, X_i)\, f_{U_{Y*}|X}(\tilde{u}; \sigma_{Y*})\, d\tilde{u}} \left(\widehat{f}_{Y*|X}(Y_i\mid X_i) - f_{Y*|X}(Y_i\mid X_i)\right) \\
&= \frac{f_{U_{Y*}}(u; \sigma_{Y*})}{\int_{\mathcal{U}} f_{Y|U_{Y*},X}(Y_i\mid\tilde{u}, X_i)\, f_{U_{Y*}|X}(\tilde{u}; \sigma_{Y*})\, d\tilde{u}} \times \sum_{\ell=1}^{L} \widehat{\mathbb{R}}_{\ell, f_{Y*|X}}(Y_i, X_i)' \left(\widehat{\beta}_{Y*}(\tau_\ell) - \beta_{Y*}(\tau_\ell)\right) \\
&=: \sum_{\ell=1}^{L} \widehat{\mathbb{R}}_{\ell, f_{U_{Y*}|Y,X}}(u, Y_i, X_i; \sigma_{Y*})' \left(\widehat{\beta}_{Y*}(\tau_\ell) - \beta_{Y*}(\tau_\ell)\right).
\end{aligned}
$$

$\square$

D   Monte Carlo Simulations

In this section, we provide some Monte Carlo simulations to demonstrate the performance of our proposed approach. For the simulations below, we generate the outcome and the treatment according to

$$
\begin{aligned}
Y_i &= \beta_{0Y*}(V_{iY*}) + \beta_{1Y*}(V_{iY*})X_i + U_{iY*} \\
T_i &= \beta_{0T*}(V_{iT*}) + \beta_{1T*}(V_{iT*})X_i + U_{iT*}
\end{aligned}
$$

where, for $\tau \in [0, 1]$, we set

$$
\begin{aligned}
\beta_{0Y*}(\tau) &= 1 + 3\tau - \tau^2 \\
\beta_{1Y*}(\tau) &= \exp(\tau)
\end{aligned}
$$

$$\beta_{0T^*}(\tau) = 1 + \sqrt{\tau}$$
$$\beta_{1T^*}(\tau) = 0.1 \exp(\tau)$$

We set V_{Y^*} and V_{T^*} to be draws from a parametric copula $C_{V_{Y^*} V_{T^*}|X}$ which we vary across simulations. In particular, we vary this copula between a Clayton copula with parameter $\delta = 1.5$ and a Gaussian copula with parameter $\rho = 0.5$.

U_{Y^*} and U_{T^*} are $i.i.d.$ draws from $\mathcal{N}(0, \sigma_j^2)$ for $j \in \{Y^*, T^*\}$, and we vary σ_j^2 across simulations. Larger values of σ_j^2 indicate more measurement error, and, in particular, we consider the cases where σ_j^2 is equal to 1, 0.25, 0.01, and 0, which roughly correspond to a relatively large amount of measurement error, a relatively small amount of measurement error, a very small amount of measurement error, and no measurement error at all. Finally, we set $X \sim U[0, 1]$.

Table 4 contains the results from the Monte Carlo simulations. The first set of results compares the performance of our first step estimators (which account for measurement error in the outcome) with quantile regression that ignores measurement error (these results are in the columns labeled "QR Coef."). For this part, we estimate the quantile regression parameters for the outcome across a grid of 10 equally spaced values of τ from 0.02 to 0.98. The numbers in those columns correspond to an aggregated version of root mean squared error — in each Monte Carlo simulation, we compute the squared error of the estimate of $\beta_{0Y^*}(\tau)$ and $\beta_{1Y^*}(\tau)$, average these across all 20 parameter estimates, and then calculate the square root of the mean of this quantity across Monte Carlo simulations. For the first step quantile regression estimates, we find that our approach performs substantially better than ignoring measurement error in the case when the measurement error is large, each approach performs similarly in the case with small measurement error, and ignoring measurement error tends to perform better when the amount of measurement error is very small or non-existent. Additionally, both approaches perform moderately better with larger sample sizes. And the results are qualitatively similar for both the Clayton copula and the Gaussian copula that we consider.

Table 4: Monte Carlo Simulations

ME s.d.	QR Coef.		Cop. Param		Transition Mat.			Pov. Rate		
	ME	No ME	ME	No ME	ME	No ME	Obs.	ME	No ME	QR No ME
$n = 250$, **Clayton Copula** ($\delta = 1.5$)										
1	0.463	0.637	0.930	1.358	0.098	0.139	0.147	0.036	0.154	0.151
0.5	0.364	0.335	0.554	1.126	0.058	0.107	0.117	0.023	0.078	0.078
0.1	0.361	0.191	1.263	0.329	0.038	0.033	0.050	0.033	0.013	0.011
0	0.368	0.195	1.402	0.212	0.039	0.023	0.045	0.033	0.011	0.015
$n = 1000$, **Clayton Copula** ($\delta = 1.5$)										
1	0.344	0.546	0.866	1.345	0.086	0.136	0.141	0.026	0.154	0.149
0.5	0.288	0.235	0.404	1.097	0.048	0.103	0.109	0.025	0.077	0.077
0.1	0.331	0.099	1.316	0.276	0.033	0.028	0.032	0.033	0.009	0.007
0	0.335	0.097	1.578	0.102	0.037	0.017	0.025	0.033	0.005	0.013
$n = 250$, **Gaussian Copula** ($\rho = 0.5$)										
1	0.464	0.647	0.369	0.419	0.083	0.100	0.111	0.029	0.118	0.104
0.5	0.361	0.333	0.195	0.298	0.045	0.074	0.087	0.054	0.057	0.041
0.1	0.364	0.200	0.343	0.062	0.075	0.022	0.048	0.081	0.017	0.028
0	0.364	0.192	0.358	0.050	0.086	0.021	0.047	0.081	0.017	0.034
$n = 1000$, **Gaussian Copula** ($\rho = 0.5$)										
1	0.341	0.543	0.283	0.413	0.067	0.099	0.103	0.017	0.119	0.102
0.5	0.290	0.234	0.177	0.283	0.036	0.071	0.078	0.061	0.057	0.038
0.1	0.332	0.100	0.374	0.036	0.080	0.016	0.027	0.082	0.008	0.026
0	0.335	0.097	0.378	0.025	0.091	0.016	0.026	0.082	0.009	0.031

Notes: The table reports root mean squared error for (i) estimates of quantile regression parameters (columns labeled "QR Coef"), (ii) estimates of the conditional copula parameter (columns labeled "Cop. Param"), (iii) estimates of the transition matrix (columns labeled "Transition Mat."), and (iv) estimates of the poverty rate (columns labeled "Pov. Rate"). The column labeled "ME s.d." refers to the standard deviation of the measurement error for Y and T (which are set to be equal to each other). Columns labeled "ME" use the approach suggested in the paper; columns labeled "No ME" ignore measurement error but otherwise follow the same estimation strategy consisting of first step quantile regressions followed by estimating the conditional copula; the column labeled "Obs." directly uses the observed data to calculate the unconditional transition matrix, and the column labeled "QR No ME" ignores measurement error and inverts quantile regression estimates into estimates of the poverty rate. The rows differ by (i) the amount of measurement error, (ii) the number of observations, and (iii) the particular conditional copula used. The results are based on 1000 Monte Carlo simulations.

The second set of results is for estimates of the copula parameter (these results are in the columns labeled "Cop. Param"). As before, our approach outperforms ignoring measurement error in the case of a large amount of measurement error and underperforms ignoring measurement error in the cases with a very small amount of measurement error or non-existent measurement error. However, in the middle case (recall, this is the case with a small but non-negligible amount of measurement error), our approach tends to perform substantially better than ignoring measurement error across simulations. In particular, when measurement error is ignored, estimates of the conditional copula parameter tend to be downward biased for the dependence between Y^* and T^* due to measurement error.

The next set of results is for estimates of unconditional transition matrices (these results are in the columns labeled "Transition Mat.") and the poverty rate conditional on a particular value of T^*. For the transition matrix, we report an aggregated version of the root mean squared error based on computing the average squared error across cells in the transition matrix for a particular simulation and then taking the square root of the average of these across simulations. For the transition matrix, we also report estimates of the root mean squared error of just taking the raw data and computing a transition matrix directly. For the poverty rate, we set the "poverty line" to be $\beta_{0Y^*}(0.1) + \beta_{1Y^*}(0.1)0.5$ (where 0.5 is the median of X since it is distributed $U[0,1]$), and we compute an estimate of the poverty rate conditional on $T^* = \beta_{0T^*}(0.5) + \beta_{1T^*}(0.5)0.5$. For the poverty rate, the column labeled "QR No ME" ignores measurement error and does not use our estimation procedure; rather, the estimates in that column come from running a quantile regression of Y on T and X, inverting the quantile regression estimates to obtain an estimate of $F_{Y|T,X}$, averaging out the covariates from this distribution, and then computing an estimate of the probability of being below the poverty line. As for the copula parameter, our approach tends to perform better than ignoring measurement error for estimating transition matrices and poverty rates, except in cases where the measurement error is extremely small. This can be understood in light of the results on estimating the QR coefficients and copula parameters themselves. In the middle case with a small but non-negligible amount of measurement error, both approaches tend to perform about equally for estimating the QR parameters, but our approach tends to perform better for estimating the copula parameter. Since the transition matrix and poverty rate parameters depend on both of these, our approach tends to perform somewhat better in this case.

The simulations above had measurement error drawn from a normal distribution, which is super-smooth. As we discussed in the main text, due to Assumption 5, the speed of convergence of our estimators is not affected by the smoothness of the measurement error. However, the finite sample performance of our estimators could still depend on the smoothness of the measurement error. We explore this in the simulations reported in Table 5, where we continue to use the same DGP as above except draw the measurement error from a Laplace distribution, which is ordinary-smooth. Note that the rate of convergence of nonparametric estimators is typically faster for ordinary-smooth measurement error than for super-smooth measurement error. We fix the standard deviation of the measurement error to be 0.5, so the results are most comparable to the results from Table 4 with the same standard deviation of the measurement error. The main takeaways from these results are that, relative to the case with Gaussian measurement error, our

Table 5: Monte Carlo Simulations with Laplace Measurement Error

ME s.d.	QR Coef.		Cop. Param		Transition Mat.			Pov. Rate		
	ME	No ME	ME	No ME	ME	No ME	Obs.	ME	No ME	QR No ME
$n = 250$, **Clayton Copula** ($\delta = 1.5$)										
0.5	0.266	0.329	0.626	1.118	0.047	0.106	0.117	0.020	0.078	0.079
$n = 1000$, **Clayton Copula** ($\delta = 1.5$)										
0.5	0.177	0.238	0.431	1.099	0.037	0.103	0.109	0.019	0.077	0.077
$n = 250$, **Gaussian Copula** ($\rho = 0.5$)										
0.5	0.269	0.327	0.180	0.301	0.043	0.074	0.089	0.045	0.057	0.041
$n = 1000$, **Gaussian Copula** ($\rho = 0.5$)										
0.5	0.175	0.235	0.141	0.284	0.030	0.071	0.078	0.044	0.057	0.037

Notes: The table reports root mean squared error for (i) estimates of quantile regression parameters (columns labeled "QR Coef"), (ii) estimates of the conditional copula parameter (columns labeled "Cop. Param"), (iii) estimates of the transition matrix (columns labeled "Transition Mat."), and (iv) estimates of the poverty rate (columns labeled "Pov. Rate"). The results in this table differ from those in Table 4 by having measurement error drawn from a Laplace distribution rather than a Gaussian distribution. The column labeled "ME s.d." refers to the standard deviation of the measurement error for Y and T (which are set to be equal to each other). Columns labeled "ME" use the approach suggested in the paper; columns labeled "No ME" ignore measurement error but otherwise follow the same estimation strategy consisting of first step quantile regressions followed by estimating the conditional copula; the column labeled "Obs." directly uses the observed data to calculate the unconditional transition matrix, and the column labeled "QR No ME" ignores measurement error and inverts quantile regression estimates into estimates of the poverty rate. The rows differ by (i) the amount of measurement error, (ii) the number of observations, and (iii) the particular conditional copula used. The results are based on 1000 Monte Carlo simulations.

estimator (i) tends to perform better with Laplace measurement error for a fixed sample size, and (ii) tends to improve more when moving from a smaller sample size to a larger sample size with Laplace measurement error. This suggests at least somewhat better finite sample performance of our estimators with ordinary-smooth measurement error relative to super-smooth measurement error.