cemmap
centre for microdata methods and practice

Testing for the presence of measurement error

Daniel Wilhelm

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP48/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Testing for the Presence of Measurement Error

Daniel Wilhelm*

September 13, 2019

Abstract

This paper proposes a simple nonparametric test of the hypothesis of no measurement error in explanatory variables and of the hypothesis that measurement error, if there is any, does not distort a given object of interest. We show that, under weak assumptions, both of these hypotheses are *equivalent* to certain restrictions on the joint distribution of an observable outcome and two observable variables that are related to the latent explanatory variable. Existing nonparametric tests for conditional independence can be used to directly test these restrictions without having to solve for the distribution of unobservables. In consequence, the test controls size under weak conditions and possesses power against a large class of nonclassical measurement error models, including many that are not identified. If the test detects measurement error, a multiple hypothesis testing procedure allows the researcher to recover subpopulations that are free from measurement error. Finally, we use the proposed methodology to study the reliability of administrative earnings records in the U.S., finding evidence for the presence of measurement error originating from young individuals with high earnings growth (in absolute terms).

*Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom; E-Mail address: d.wilhelm@ucl.ac.uk. This work was supported in part by the European Research Council (ERC-2014-CoG-646917-ROMIA) and the ESRC Centre for Microdata Methods and Practice at IFS (RES-589-28-0001). I thank Joe Altonji, Richard Blundell, Chris Bollinger, Stéphane Bonhomme, Xavier D'Haultfoeuille, James Heckman, Yuichi Kitamura, Arthur Lewbel, Ulrich Müller, Derek Neal, Whitney Newey, Susanne Schennach, Azeem Shaikh, and Alex Torgovitsky for helpful comments and discussions. Young Jun Lee provided excellent research assistance.

1 Introduction

In empirical research, concerns about measurement error, broadly defined as *any* difference between what we observe and what we would like to observe in the data, are widespread and emerge for a variety of different reasons such as reporting errors on surveys, proxy errors, model misspecification, or the desire to measure imprecisely defined concepts like ability or skills. Failing to account for such errors may lead to significantly distorted conclusions or no distortions at all, depending on the nature of the errors. Perhaps surprisingly, apart from a few prominent exceptions (e.g. Altonji (1986), Card (1996), Erickson and Whited (2000, 2012), Cunha, Heckman, and Schennach (2010), Feng and Hu (2013), Arellano, Blundell, and Bonhomme (2017)), current empirical practice is dominated by informal arguments for the absence of measurement error or that it is of second-order importance. A healthy dose of judgment and critical thinking is, of course, essential for any evaluation of empirical work, but discussions about the importance of measurement difficulties are typically based on little to no information about the origin of measurement errors, how they relate to other variables in the model or how large signal-to-noise ratios might be. Therefore, a formal analysis of the extent of the measurement problem could significantly strengthen the credibility of subsequent findings. Unfortunately, identification of measurement error models relies on strong assumptions and leads to estimators that perform poorly. The credibility of such an analysis that recovers the "true" latent structure is therefore often limited.

In this paper, we propose a fundamentally different and more pragmatic approach to addressing potential measurement problems, which avoids the recovery of the "true" latent structure and leads to simple nonparametric inference for assessing the importance of measurement error. Consider a generic production problem in which an outcome $Y \in \mathbb{R}$ is produced from inputs $X^* \in \mathbb{R}^{d_x}$. We observe a variable $X \in \mathbb{R}^{d_x}$ which we suspect may be an error-contaminated measurement of the inputs X^* and another variable $Z \in \mathbb{R}^{d_z}$ that is related to X^*, perhaps an instrument or a repeated measurement.

We are interested in two types of hypotheses. The first is the hypothesis of no measurement error in X,

$$H_0^{no\,ME}\colon P(X = X^*) = 1, \tag{1}$$

This hypothesis may be of direct economic interest because it can sometimes be interpreted as the absence of certain frictions such as inattention, price misperceptions, or imperfect competition (Chetty (2012)). Testing $H_0^{no\,ME}$ could also be useful as a model specification test. However, in this case, the researcher is likely less interested in whether there is measurement error or not, but rather in whether the measurement error, if there is any, distorts objects of interest such as the production function. We therefore also study a second type of hypothesis,

$$H_0^{func}\colon P(E_P[\Lambda_P(Y, X) \mid X^*, X] = 0) = 1, \tag{2}$$

where the function Λ_P characterizes the equality of the researcher's object of interest and an observable counterpart. For example, H_0^{func} allows us to test whether the observed and true

conditional expectations of $Y|X$ and $Y|X^*$ are the same,

$$P(E_P[Y|X^*] = E_P[Y|X]) = 1, \tag{3}$$

or whether inequality in the distribution of Y among individuals with the same X^*, for example measured by a conditional entropy $E_P[Y \log(Y)|X^*]$, is the same as the observed conditional entropy,

$$P(E_P[Y \log(Y)|X^*] = E_P[Y \log(Y)|X]) = 1. \tag{4}$$

When a hypothesis such as (3) or (4) holds, it does not necessarily imply that there is no measurement error in X but only that, if there is any, it does not affect the relevant functional. In fact, in a finite sample, a test may fail to reject such hypotheses even when measurement error *does* affect the functionals, but the distortion is small relative to sampling noise. In this sense, we can use the test to find out whether measurement error leads to severe enough distortions such that the data can distinguish functionals based on true and mismeasured explanatory variables. Testing (2) for various functions Λ_P then allows us to explore for which type of economic questions measurement error matters and for which ones it does not. For example, measurement error might affect the average level of Y (i.e. (3) does not hold), but not inequality in Y (i.e. (4) holds). In this case, policy questions depending on how a change in X^* affects inequality in the outcome can be answered without accounting for measurement error, but questions related to the effect on average outcomes cannot.

The two null hypotheses $H_0^{no\,ME}$ and H_0^{func} depend on the latent variable X^* and thus cannot directly be tested. Instead of attempting to recover the "true" latent structure and then perform tests based on that, we start by imposing a weak restriction that does not pin down the latent structure but nevertheless can be exploited to test $H_0^{no\,ME}$ and H_0^{func}:

$$Y \perp Z \mid X^*. \tag{5}$$

This assumption postulates that outcomes and the measurement Z are independent conditional on the actual inputs of production, X^*, and is thus a type of exclusion restriction. It implies that Y is produced by X^*, not by the measurement Z. This exclusion restriction is standard in the literature on identification and estimation of measurement error models (Chen, Hong, and Nekipelov (2011), Schennach (2013, 2016), Hu (2017)) and has already been justified in a wide range of empirical applications (see Section 2 for examples).

If the restriction (5) holds, then $H_0^{no\,ME}$ implies that Z must also be independent of the outcome conditional on the observed X,

$$Y \perp Z \mid X. \tag{6}$$

Unlike $H_0^{no\,ME}$, this is a restriction that depends only on observables and can directly be tested using existing tests for conditional independence. $H_0^{no\,ME}$ implies (6) without any other assumptions besides the exclusion restriction (5), so that a valid test of (6) is also a valid test of the null of no measurement error under very general conditions. In particular, all variables

may be either continuous or discrete, and the variable Z may be binary even when X, X^* are continuous. The outcome equation and the equations determining X and Z (the "measurement system") may be fully nonseparable with unobserved heterogeneity of any dimension.

While the observable implication (6) is a fairly immediate consequence of $H_0^{no\,ME}$ and the exclusion restriction, significantly more effort is required to derive conditions under which the two are in fact *equivalent*. This result requires a monotonicity condition in the outcome model such that some moment of Y monotonically varies with X^* and a relevance condition that ensures Z is sufficiently related to X^*. Again, the outcome equation and the measurement system may be fully nonseparable with unobserved heterogeneity of any dimension. In the case of continuous X, X^*, no completeness assumptions are required. The measurement errors are allowed to depend on each other and on the true variable. As a result the test has power against, and thus is able to detect, a wide range of realistic nonclassical measurement error models without requiring that they are identified.

Having established the equivalence result for $H_0^{no\,ME}$, we extend it to the hypothesis H_0^{func} with functions Λ_P that satisfy a certain monotonicity condition.

For the case in which a test of $H_0^{no\,ME}$ or H_0^{func} rejects, we propose a multiple testing procedure that performs the same test on a number of prespecified subpopulations so as to account for multiplicity by controlling the family-wise error. By collecting those null hypotheses that are not rejected the researcher then recovers subpopulations in which there is no measurement error or no distortion of the object of interest, respectively.

Finally, we study the reliability of administrative earnings records in the United States. The proposed test rejects the null of no measurement error on the full dataset as well as on most subsamples we consider, with p-values that are numerically zero or very close to zero. We present some evidence that is consistent with young individuals who experience high earnings growth (in absolute terms) being responsible for the measurement error.

Related Literature In the special case of a binary explanatory variable, Mahajan (2006) provides conditions under which the null of no measurement error is equivalent to the same conditional independence condition that we derive in this paper. His assumptions imply identification of the whole model, which is not necessary for the purpose of testing for the presence of measurement error. We show that the conditions in this paper are implied by Mahajan's. More importantly, our result applies to arbitrary discrete and continuous distributions for the explanatory variable. Especially the latter case requires more general arguments to avoid claims of identification based on strong completeness and independence conditions. In addition, we show how to test whether measurement error affects objects of interest, provide low-level conditions in various measurement error models and a procedure for recovering measurement error-free subpopulations.

In principle, one could construct a test for the presence of measurement error by comparing an estimator of the model that accounts for the possibility of measurement error with one that ignores it, similar in spirit to the work by Durbin (1954), Wu (1973), and Hausman (1978). If the

difference between the two is statistically significant, then one could conclude that this is evidence for the presence of measurement error. However, this strategy would require identification and consistent estimation of the measurement error model, which leads to overly strong assumptions, the necessity of solving ill-posed inverse problems in the continuous variable case, and potentially highly variable estimators. These difficulties can all be avoided by the approach presented in this paper.

There are some existing tests for the presence of measurement error in parametric models that require identification and consistent estimators of the model: Hausman (1978), Chesher (1990), Chesher, Dumangane, and Smith (2002), Hahn and Hausman (2002), and Hu (2008). Related to Hausman (1978), in empirical work it is common to estimate linear regressions by OLS and IV, and then attribute a difference in the two estimates to the presence of measurement error, treating the IV estimate as the consistent and unbiased estimator. Of course, this strategy is valid only if the true relationship of interest is in fact linear, the measurement error is classical, and the model is identified. None of these assumptions are required in the approach proposed in the present paper.

Our proposed testing approach is related to, but different from, some other testing problems such as testing for exogeneity or significance of regressors. In linear models, measurement error and any other unobservable causing endogeneity are absorbed by the error in the outcome equation. This is not the case in nonlinear models, so that testing for the presence of measurement error and testing for other forms of endogeneity are distinct problems. In Section 2.3, we discuss in more detail the differences between the hypotheses considered in this paper and the hypothesis of exogeneity. The connection to testing for significance of regressors is closer because we show that the null of no measurement error is equivalent to a conditional independence condition, which can be tested with existing significance tests such as Gozalo (1993), Fan and Li (1996), Delgado and Gonzalez Manteiga (2001), Mahajan (2006), and Huang, Sun, and White (2016).

Software Implementations We provide R and STATA code implementing the proposed testing procedure at `http://github.com/danielwilhelm/R-ME-test` and `http://github.com/danielwilhelm/STATA-ME-test`.

2 The Null Hypothesis of No Measurement Error

In this section, we focus on testing the null of no measurement error[1] in the explanatory variable:

$$H_0^{no\,ME}\colon P \in \mathbf{P}_0 \tag{7}$$

versus $H_1^{no\,ME}\colon P \in \mathbf{P} \setminus \mathbf{P}_0$, where $\mathbf{P}$ is a "large" set of distributions and

$$\mathbf{P}_0 := \{P \in \mathbf{P}\colon P(X^* = X) = 1\}.$$

[1] Appendix A.1 provides reasons for why we do not consider flipping the null and alternative hypotheses.

The null hypothesis $H_0^{no\,ME}$ cannot be tested directly in the sense that it involves the joint distribution of the unobservable X^* and the observable X. In Sections 2.1 and 2.2, we therefore first provide general conditions on $\mathbf{P}$ under which (7) is *equivalent* to a hypothesis about the joint distribution of the observables (Y, X, Z). Appendix B discusses the assumptions in the context of specific measurement error models and economic applications. Section 2.3 explains the difference between testing $H_0^{no\,ME}$ and testing the null of exogeneity.

2.1 Observable Implication

Probability measures and random vectors in this paper are all defined on the same Borel space $(\Omega, \mathcal{B}(\Omega))$.[2] Let $Y \in \mathbb{R}$ be an outcome variable, and $X^* \in \mathbb{R}^{d_x}$, $X \in \mathbb{R}^{d_x}$, $Z \in \mathbb{R}^{d_z}$ some random vectors.

Assumption 1 (exclusion)**.** $Y \perp Z \mid X^*$ *under* P*.*

This assumption requires Z to be excluded from the outcome model conditional on X^*. Z may be some general indicator of X^*, i.e. a variable that depends on X^* in some way, but does not need to measure the same concept as X^*. The exclusion restriction requires that it affects the outcome Y only through the explanatory variable X^*. In Section B, we show that this assumption can accommodate interpretations of Z as a second measurement of X^* and as an instrument for X^* as special cases.

Assumption 1 is standard in the literature on identification and estimation of measurement error models (Chen, Hong, and Nekipelov (2011), Schennach (2013, 2016), Hu (2017)) and has already been justified in a wide range of empirical applications. Since the assumption is central to all arguments of this paper, we provide a few examples. For the simplicity of discussion, however, we ignore the fact that, in some applications, the exclusion restriction should be modified to conditional mean independence or to the presence of additional conditioning variables as these are straightforward extensions as elaborated on below.

There is a recent and rapidly growing field that studies early childhood interventions by identifying and estimating the production function of skills (e.g. Cunha, Heckman, and Schennach (2010), Heckman, Pinto, and Savelyev (2013), Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2015), Attanasio, Meghir, and Nix (2017)). Depending on the objective of the analysis, Y represents an outcome in adulthood (e.g. income or and indicator for college attendance) or simply a future skill measurement (e.g. a test score). The outcome is produced from a vector of inputs (X^*) such as (non-)cognitive skills, health, and parental investment. These inputs are measured by a vector X, typically a battery of test scores and observable aspects of parental investment, among others. In this context, Z is a set of measurements different from those in X. The authors above justify the exclusion restriction by arguing that the inputs to production are the true inputs X^* rather than the measurements (X, Z). This means that conditional on knowing the true inputs X^*, the measurements should not provide any additional

[2] Appendix E defines this probability space slightly more carefully.

information about outcomes. This argument also applies to other production problems in which inputs are difficult to measure (e.g. Olley and Pakes (1996)).

In the firm-level investment literature (Bond and Van Reenen (2007)), Y represents a firm's investment, X^* the discounted sum of expected future marginal benefits as perceived by the firm's manager ("marginal q"), X the firm's book-to-market ratio ("average q"), and Z could either be average q from a different time period or a variable that does not enter the manager's information set. According to the q-theory of investment (Lucas and Prescott (1971), Mussa (1977)) X^* is the only determinant of investment, so that conditional on X^* the measurements (X, Z) should not provide any additional information about investment. This justifies the exclusion restriction.

In the empirical part (Section 5), Y, X, Z are three measurements of earnings, but Y and (X, Z) come from two different data sources, one from a survey and the other from an administrative dataset. We then argue Assumption 1 holds because (i) the error in Z has a very different origin from the error in Y, at least conditional on X^*, and (ii) the two errors are unlikely to share common components because of the way the sample is selected.

There are many other empirical applications that also impose Assumption 1: for instance, Altonji (1986) studies labor supply, Kane and Rouse (1995) and Kane, Rouse, and Staiger (1999) the returns to education, Card (1996) the effect of unions on the wage structure, Hu, McAdams, and Shum (2013) auctions with unobserved heterogeneity, Feng and Hu (2013) unemployment dynamics, and Arellano, Blundell, and Bonhomme (2017) earnings dynamics.

The following definition introduces the set of distributions $\mathbf{M}$ that satisfy the maintained exclusion restriction.

Definition 1. *Let* $\mathbf{M}$ *be the set of distributions* P *that satisfy Assumption 1.*

For distributions in $\mathbf{M}$ we can state an intuitive observable implication of the null of no measurement error.

Theorem 1. *If* $\mathbf{P} = \mathbf{M}$*, then* (7) *implies*

$$Y \perp Z \mid X \quad \textit{under } P. \tag{8}$$

If condition (8) holds, then the conditional distribution of Y given X and Z does not vary with Z. The result implies that, under the exclusion restriction, a rejection of the hypothesis (8) implies a rejection of the null of no measurement error, $H_0^{no\,ME}$. Therefore, a level-α test of (8) is also a level-α test of $H_0^{no\,ME}$. In fact, (8) is a hypothesis of conditional independence, for which many tests have already been proposed (e.g. Gozalo (1993), Fan and Li (1996), Delgado and Gonzalez Manteiga (2001), Mahajan (2006), Huang, Sun, and White (2016), among many others). Appendix C reports simulation results for the test by Delgado and Gonzalez Manteiga (2001), which we use in the empirical part (Section 5), and confirms that it controls size under the null of no measurement error.[3]

[3]R and STATA implementations are available at `http://github.com/danielwilhelm/R-ME-test` and `http://github.com/danielwilhelm/STATA-ME-test`.

Unlike (8) the null of no measurement error, $H_0^{no\,ME}$, depends on the distribution of unobservable variables. A direct test of $H_0^{no\,ME}$ would therefore require solving for the joint distribution of (X, X^*) from the distribution of observables (Y, X, Z). This is a hard statistical problem, especially in the case of continuous explanatory variables, in the sense that estimators tend to be highly variable and may possess slow, possibly logarithmic, convergence rates (Fan (1991)). Furthermore, the implementation of such a test in the continuous case would involve relatively complicated estimators based on operator inversions. Tests based on (8) avoid all of these complications.

For Theorem 1 to hold, we only need to impose the exclusion restriction and, perhaps surprisingly, no other conditions. Therefore, it is important to emphasize what is not assumed. First of all, distributions in $\mathbf{M}$ allow Y, X^*, X, Z each to be discrete, continuous or mixed, and the supports may be bounded or unbounded. The randomness in the conditional distributions of $Y|X^*$ and $Z|X^*$ may originate from unobserved heterogeneity of unrestricted dimensions and those sources of heterogeneity may interact with X^* in nonlinear, nonseparable ways. In addition, they are allowed to depend on X^* which means we can allow for endogeneity of X^* (see Appendix A.4 for more details). Finally, notice that the observable restriction (8) is testable even though, under the null, the structural relationship between Y and X^* and that between X^* and Z are not identified without further assumptions and normalizations.

Theorem 1 is an intuitive consequence of the exclusion restriction because, under the null, the conditioning variable X^* is equal to X with probability one, so we expect Y to be independent of Z not only conditional on X^* but also conditional on X. Consider the following, slightly more formal, derivation of the observable implication. Under the null, for any two bounded functions a and b,

$$\begin{aligned} E_P[a(Y)b(Z)|X] &\overset{(\star)}{=} E_P[a(Y)b(Z)|X^*] \\ &= E_P[a(Y)|X^*]\, E_P[b(Z)|X^*] \\ &\overset{(\star)}{=} E_P[a(Y)|X]\, E_P[b(Z)|X] \end{aligned}$$

where all equalities hold a.s. and the second equality follows from the exclusion restriction. Since the equality holds for all bounded functions a and b for which the expectations exist, the desired conditional independence then follows. To formalize this argument one needs to establish equality of the sigma algebras on both sides of the equalities marked by $(\star)$. Appendix E contains the details.

Remark 1 (non-directional test)**.** As is common with many other hypothesis tests, a test of $H_0^{no\,ME}$ versus $H_1^{no\,ME}$ is non-directional in the sense that rejecting the null does not lead to a recommendation as to which measurement error model is appropriate for a given sample. This paper provides two possible routes towards a more constructive conclusion in that case. First, Section 3 extends the results of this section to the hypothesis H_0^{func} mentioned in the introduction. In case of a rejection of $H_0^{no\,ME}$, we can therefore further explore whether the measurement error distorts a given object of interest or which type of objects it does and does

not affect. Second, Section 4 proposes a sequential testing procedure that allows us to recover subpopulations that are free from measurement error. □

Remark 2 (additional controls)**.** Theorem 1 readily extends to the presence of additional, correctly measured covariates W that affect the outcome Y. In this case, the exclusion restriction in Assumption 1 is replaced by $Y \perp Z \mid (X^*, W)$ which means that the exclusion holds only after additional conditioning on W. This assumption may be more plausible in economic applications, in which there are variables W that jointly determine Y and Z. In addition, this exclusion restriction allows the distributions of unobservable errors in Y and Z conditional on X^* and W to vary with X^* and W, i.e. individuals with different characteristics W are allowed to draw errors from different distributions. Under the exclusion restriction with additional controls, the null hypothesis (7) then has the observable implication $Y \perp Z \mid (X, W)$. □

Remark 3 (mean independence)**.** One could relax the exclusion restriction to conditional mean independence. Let $\mathbf{M}'$ be the set of distributions P such that there exists a function μ with $E_P[\mu(Y)|X^*, Z] = E_P[\mu(Y)|X^*]$ a.s.. If $\mathbf{P} = \mathbf{M}'$, then the null of no measurement error, (7), implies

$$E_P[\mu(Y)|X, Z] = E_P[\mu(Y)|X] \qquad a.s. \tag{9}$$

which is an observable restriction that can be tested just like (8). At this point, conditional mean independence appears strictly more desirable than the conditional independence condition in Assumption 1, but in the next subsection we see that there is a cost to this generalization when showing equivalence of (7) and (9). See Remark 6. □

Remark 4 (invariance of the null)**.** Let $\tau : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$ be an invertible mapping. Of course, the null hypothesis $H_0^{no\,ME}$ is invariant under invertible transformations, i.e. $\mathbf{P}_0$ is equal to the set $\{P \in \mathbf{P} \colon P(\tau(X^*) = \tau(X)) = 1\}$. Similarly, we can also allow the transformation τ to depend on P, assuming that $\tau(\cdot, P)$ is invertible for every P, and apply it only to X^*, not to X, i.e.

$$P(\tau(X^*, P) = X) = 1,$$

which implies (8) if $\mathbf{P} = \mathbf{M}$.[4] Let $F_{X^*}(\cdot, P)$ and $F_X(\cdot, P)$ be the cumulative distribution functions of X^* and X under P and consider, for example, the case in which $\tau(\cdot, P) = F_X^{-1}(F_{X^*}(\cdot, P), P)$, assuming that the two distribution functions are invertible over their respective supports. Then testing (8) can be interpreted as testing

$$P(F_{X^*}(X^*, P) = F_X(X, P)) = 1,$$

i.e. whether the true rank of X^* is equal to the observed rank of X. □

[4]This can be seen as follows. Let $\widetilde{X}^* := \tau(X^*, P)$. Then there is $\widetilde{P} \in \mathbf{M}$ such that $(Y, \widetilde{X}^*, X, Z) \sim \widetilde{P}$ with $\widetilde{P}(\widetilde{X}^* = X) = 1$ so that (8) holds under $\widetilde{P}$ and thus also under P.

2.2 Equivalence

This subsection provides conditions under which the restriction (8) on observables is not only implied by but also implies the null hypothesis $H_0^{no\,ME}$, a result that is important for establishing power of a test against alternatives violating $H_0^{no\,ME}$. If Z was independent of all other variables, then the observable implication (8) holds even when the null hypothesis $H_0^{no\,ME}$ is violated. Therefore, it is clear that showing the desired equivalence result requires further restrictions on the set $\mathbf{M}$, denoted by $\mathbf{M}_R \subset \mathbf{M}$. The key challenge lies in restricting $\mathbf{M}_R$ enough such that (8) holds only for those $P \in \mathbf{M}_R$ that also satisfy the null, while at the same time restricting $\mathbf{M}_R$ not too much so the class of distributions remains "large" and, thus, contains measurement error models as general as possible.[5]

Throughout this subsection, we assume X^*, X, Z are scalar ($d_x = d_z = 1$).[6] We first explain the approach of this section in a heuristic fashion. Consider the special case when X^*, X are continuously distributed. Suppose the observable implication (8) holds. Then, for any two values z_1, z_2, we have $P_{Y|X,Z=z_1} = P_{Y|X,Z=z_2}$ which means expectations of Y with respect to these two distributions must also be equal. Let us also strengthen the exclusion restriction to $Y \perp (X, Z) \mid X^*$. Then

$$\int E_P[\mu(Y)|X^*]\; d(P_{X^*|X,Z=z_1} - P_{X^*|X,Z=z_2}) = 0,$$

for any function μ. If $E_P[\mu(Y)|X^* = \cdot]$ is differentiable, then integration by parts yields

$$\int \left(P_{X^*|X=x,Z=z_1}(x^*) - P_{X^*|X=x,Z=z_2}(x^*)\right) \frac{\partial E_P[\mu(Y)|X^* = x^*]}{\partial x^*} dx^* = 0. \tag{10}$$

We want to show that this equation implies the null hypothesis $H_0^{no\,ME}$. On the contrary, assume that this is not the case. To generate a contradiction, we want to ensure that (10) does not hold under the alternative H_1. This is the case, for example, when $E_P[\mu(Y)|X^* = \cdot]$ is monotone (and not constant) and $P_{X^*|X=x,Z=z_2}$ first-order stochastically dominates (FOSD) $P_{X^*|X=x,Z=z_1}$ (and they are not equal) under H_1. The monotonicity assumption makes sure the derivative of the conditional expectation does not change sign (and is nonzero somewhere) and the dominance condition implies that the difference of the conditional distributions is nonnegative (and positive somewhere). In conclusion, the integral in (10) is nonzero under H_1, yielding the desired contradiction, so the null of no measurement error must hold.

While the monotonicity condition is often directly implied by economic theory (see the discussion below), the FOSD condition is an opaque high-level assumption in that it is difficult to see what restrictions it places on any given measurement error model. We therefore first

[5] We do not claim that these maintained assumptions are minimal in any sense, but we argue that they allow for a diverse range of realistic nonclassical measurement error models. In addition, we show that these assumptions are easy to interpret in concrete models. It is therefore our view that generalizing them further (or perhaps even finding a minimal set of such assumptions), though a worthwhile theoretical exercise for future work, is not of first-order importance at this stage.

[6] Appendix A.2 discusses the multivariate extension.

formally state the above result for the general case when X^* is scalar, but not restricted to be continuous, and without assuming differentiability of the conditional expectation, and then turn to the main task of this section, deriving sufficient conditions for the FOSD requirement that can be interpreted in concrete measurement error models.

Assumption 2 (stronger exclusion)**.** $Y \perp (X, Z) \mid X^*$ *under* P

This assumption strengthens the exclusion restriction so that not only Z but also the observed explanatory variable, X, is excluded from the outcome model conditional on X^*. This means that X cannot provide any more information about Y than the true explanatory variable X^* already does ("non-differential measurement error"). The empirical examples discussed in Section 2.1 do in fact impose this stronger assumption rather than the weaker version from that section.

Denote by $\mathcal{S}_A$ the support of a random variable or random vector A under P. To simplify the exposition we omit the dependence of $\mathcal{S}_A$ on P.

Assumption 3 (monotonicity)**.** *There is a measurable function* $\mu\colon \mathcal{S}_Y \to \mathbb{R}$ *and a set* $\mathcal{X}^* \subseteq \mathcal{S}_{X^*}$ *so that* $x^* \mapsto E_P[\mu(Y)|X^* = x^*]$ *exists, is monotone over* $\mathcal{S}_{X^*}$ *and strictly monotone over* $\mathcal{X}^*$*.*

This assumption requires monotonicity of some conditional moment of $Y|X^*$. It holds for example if, for some y, $P(Y \leq y|X^* = \cdot)$ or $E_P[Y|X^* = \cdot]$ is monotone. Such monotonicity assumptions are often directly implied by economic theory, e.g. when $E_P[\mu(Y)|X^* = x^*]$ is a production, cost, or utility function. Examples can be found in Matzkin (1994), Olley and Pakes (1996), Cunha, Heckman, and Schennach (2010), Blundell, Horowitz, and Parey (2012, 2016), Kasy (2014), Wilhelm (2015), Hoderlein, Holzmann, Kasy, and Meister (2016), Chetverikov and Wilhelm (2017), among many others.

Assumption 4 (FOSD)**.** *There exist a constant* $C > 0$ *and sets* $\mathcal{X} \subseteq \mathcal{S}_X$*,* $\mathcal{Z}_1, \mathcal{Z}_2 \subseteq \mathcal{S}_Z$*,* $\mathcal{Z}_1 \cap \mathcal{Z}_2 = \emptyset$*, such that* $P_{X,Z}(\mathcal{X} \times \mathcal{Z}_k) > 0$*,* $k = 1, 2$*, and the following two conditions hold: for any* $(x, z_1, z_2) \in \mathcal{X} \times \mathcal{Z}_1 \times \mathcal{Z}_2$*,*

$$P(X^* \geq x^*|X = x, Z = z_1) \leq P(X^* \geq x^*|X = x, Z = z_2) \qquad \forall x^* \in \mathbb{R} \tag{11}$$

and, for $\mathcal{X}^*$ *the same set as in Assumption 3,*

$$P(X^* \geq x^*|X = x, Z = z_1) \leq P(X^* \geq x^*|X = x, Z = z_2) - C \qquad \forall x^* \in \mathcal{X}^*. \tag{12}$$

Condition (11) is the FOSD requirement introduced in the heuristic explanation above. (12) ensures that the dominance is strict at least for some values of x^* for which $E_P[\mu(Y)|X^* = x^*]$ is strictly monotone. The assumption rules out the case in which Z is independent of X^* and can thus be viewed as a relevance condition.

We now define the set $\mathbf{M}_R$ of distributions that satisfy the maintained Assumptions 2–4 under the alternative and then state the desired equivalence result.

Definition 2. *Let* $\mathbf{M}_R$ *be the set of distributions* $P \in \mathbf{M}$ *that satisfy Assumptions 2–4 when* $P(X = X^*) < 1$.

Theorem 2. *Suppose* $\mathbf{P} = \mathbf{M}_R$. *Then* $P \in \mathbf{P}$ *satisfies* (7) *if, and only if, it satisfies* (8).

Theorem 1 already establishes that a level-α test of (8) is also a level-α test of $H_0^{no\,ME}$. However, it is easy to construct examples of distributions in $\mathbf{M}$ for which $H_0^{no\,ME}$ is violated, but (8) holds. A test of (8) has no power against such violations of $H_0^{no\,ME}$. Theorem 2 shows that, for all distributions in the restricted set $\mathbf{M}_R$, (8) implies $H_0^{no\,ME}$. Therefore, a test that has power against all violations of (8) within $\mathbf{M}_R$ must then also have power against all violations of $H_0^{no\,ME}$ within $\mathbf{M}_R$. In this sense, a consistent test of the observable restriction (8) can recover any measurement error model that is compatible with the restrictions defining $\mathbf{M}_R$.

Except in some models such as binary misclassification (see Appendix B.1), it is not easy to see how the FOSD assumption restricts measurement error processes under the alternative. In the remainder of this section, we therefore discuss sufficient conditions that can fairly easily be interpreted in concrete measurement error models and at the same time are general enough to allow for realistic forms of nonclassical measurement error. Appendix B presents several measurement error models that have been considered in the literature and shows how the assumptions of this section can be verified.

Let ρ_x and ρ_z each denote either Lebesgue on $\mathbb{R}$ or counting measure on a discrete subset of $\mathbb{R}$, but they do not need to be the same. For two elements (or vectors) A and B, we denote by $P_{A|B}$ the regular conditional distribution[7] of A given the sigma algebra generated by B. We use the notation $E_P(A|B)$ for conditional expectations under P of A given the sigma algebra generated by B. The event of no measurement error, $D := \{\omega \in \Omega \colon X(\omega) = X^*(\omega)\}$, subsequently plays a prominent role. Let D^c denote the complement of D. For two sets $G, H \in \mathcal{B}(\Omega)$, we sometimes use the notation

$$P^H(G) := \frac{P(G \cap H)}{P(H)}$$

if $P(H) > 0$. Since this defines just another distribution, conditional distributions of $A|B$ under P^H, denoted by $P^H_{A|B}$, are well-defined.

Any distribution P can be decomposed as a mixture of two distributions, one on the event D of no measurement error and one on the event D^c of some measurement error. These two mixture components are the distributions P^D and $Q := P^{D^c}$, respectively, i.e.

$$P = \lambda P^D + (1 - \lambda) Q \tag{13}$$

with $\lambda := P(D)$ the probability of no measurement error.[8] We can therefore think of the data-generating process in two steps. First, draw a Bernoulli random variable with probability λ for whether there is measurement error or not. If there is, then the data (Y, X^*, X, Z) are drawn from Q, otherwise from P^D.

[7] See chapter 10.2 of Dudley (2002) for a definition. Regular conditional distributions exist because all random vectors are defined on the same Borel space from Appendix E (Theorem 10.2.2 in Dudley (2002)).

[8] When $\lambda = 0$ (or 1), we let P^D (or Q) be an arbitrary distribution.

Sufficient conditions for the FOSD requirement in Assumption 4 need to restrict the set of possible measurement error models that are entertained. In particular, this means placing restrictions on the distribution Q.

Assumption 5 (existence of a density)**. (i)** *The distributions* Q_Z, $P^D_{X^*|Z=z}$, *and* $Q_{X,X^*|Z=z}$ *have densities, bounded by* $C_q < \infty$, *with respect to* ρ_z, ρ_x, *and* $\rho_x \times \rho_x$ *for all* $z \in \mathcal{Z}_1 \cup \mathcal{Z}_2$ *with* $\mathcal{Z}_1, \mathcal{Z}_2 \subseteq \mathcal{S}_Z$, $\mathcal{Z}_1 \cap \mathcal{Z}_2 = \emptyset$. **(ii)** $Q_Z = P^D_Z$.

The existence of a density is a weak assumption, allowing for discrete and continuous distributions for (X, X^*) and Z. One could also allow for mixed continuous-discrete distributions at the expense of more complicated notation and proofs. Notice that the dominating measures ρ_x and ρ_z might be different, thus allowing for (X, X^*) to be continuous and Z to be discrete (or vice versa). We denote densities by lower letters of the corresponding distributions.

Part (ii) is imposed only to simplify the notation in this section and the proofs, but is not necessary for the subsequent results. It requires the marginal distribution of Z to be the same regardless of whether there is measurement error in X.

Assumption 6 (restriction on ME dependence)**.** $X \perp Z \mid X^*$ *under* Q

For the general result in this section, we impose the assumption that the sources of randomness in $X|X^*$ and $Z|X^*$ are independent of each other. If Z is a repeated measurement of X^*, for example, the assumption means that the measurement error in X is independent of that in Z conditional on X^*. Therefore, both measurement errors can be nonclassical and thus depend on each other through X^*, but there cannot be any direct dependence between them. However, with more structure on the measurement system, it is actually possible to allow for direct dependence between the measurement errors (see Appendix B.2).[9]

Assumption 7 (probability of ME)**.** $P(D|X^* = x^*, Z = z) = P(D|X^* = x^*)$ *for all* $x^* \in \mathcal{S}_{X^*}$ *and* $z \in \mathcal{Z}_1 \cup \mathcal{Z}_2$.

This assumption allows for the conditional probability of no measurement error to vary with X^*, but not with Z. This is a common assumption in the literature on identification of models with misclassification in discrete explanatory variables (e.g. Lewbel (2000) and Mahajan (2006)). In the case of continuous explanatory variables, the existing literature typically assumes the probability is equal to 1 for all conditioning values x^* and z (e.g. Chen, Hong, and Nekipelov (2011) and Schennach (2013)), so that the measurement error distribution $X - X^*$ does not have a pointmass at zero.[10] Assumption 7, on the other hand, allows for the probability of no measurement error to freely vary in the interval $[0, 1]$ as a function of x^*. This is important, for example, when X^* is discrete or when X is the response of an individual on a survey. For

[9] In addition, an inspection of the proof of Theorem 3 reveals that we do not actually need conditional independence, but rather only that $q_{X|X^*,Z}(x|x^*, z_1) = q_{X|X^*,Z}(x|x^*, z_2)$ for $z_k \in \mathcal{Z}_k$. If the support of Z has more than two elements, then the two densities can differ for values $(z_1, z_2) \notin \mathcal{Z}_1 \times \mathcal{Z}_2$.

[10] An and Hu (2012) is a noticeable exception.

example, if X^* is true income and X is reported income, we can allow high- and low-income individuals to have different propensities to misreport.

To be able to state the final assumption, which ensures relevance of Z, we introduce the following concept of a single-crossing function.

Definition 3. *A function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, is called single-crossing (from below) if, for any $x', x'' \in A$ with $x' < x''$, $f(x') > 0$ implies $f(x'') \geq 0$.*

This definition allows a single-crossing function to be equal to zero in multiple places, but it can change sign only once and only from negative to positive as the argument of the function increases.

For $c \in \mathbb{R}$ and $z = (z_1, z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2$, define the function $\Delta p_{c,z} : \mathcal{S}_{X^*} \to \mathbb{R}$ as

$$\Delta p_{c,z}(x^*) := p_{X^*|Z}(x^*|z_2) - c\, p_{X^*|Z}(x^*|z_1).$$

Also, let $\mathcal{C}$ be the range of $p_{X|Z}(x|z_2)/p_{X|Z}(x|z_1)$ as (x, z_1, z_2) varies over $\mathcal{X} \times \mathcal{Z}_1 \times \mathcal{Z}_2$.

Assumption 8 (single-crossing)**.** *There are a constant $c_q > 0$, an interval $\mathcal{I} \subset \mathbb{R}$, and a set $\mathcal{X} \subseteq \mathcal{S}_X$ such that $q_{X|X^*}(x|x^*) \geq c_q$ and $P(D^c|X^* = x^*) \geq c_q$ for all $x \in \mathcal{X}$, $x^* \in \mathcal{X}^* := \mathcal{I} \cap \mathcal{S}_{X^*}$, $z_k \in \mathcal{Z}_k$, $k = 1, 2$, and, for all $c \in \mathcal{C}$ and $z \in \mathcal{Z}_1 \times \mathcal{Z}_2$:*

(i) $\Delta p_{c,z}$ is single-crossing,

(ii) $\Delta p_{c,z}(x^*) \geq c_q$ for all $x^* \in \mathcal{X}^*$.

This assumption is the main restriction that ensures Z is sufficiently related to X^*. Part (i) requires that there exist two sets $\mathcal{Z}_1, \mathcal{Z}_2$ so that, for $z \in \mathcal{Z}_1 \times \mathcal{Z}_2$, the function $\Delta p_{1,z}$ is single-crossing. This means the two conditional densities $p_{X^*|Z}(\cdot|z_2)$ and $p_{X^*|Z}(\cdot|z_1)$ cross at most once (they may be equal over an interval intersected with their supports). Part (ii) then simply strengthens the single-crossing requirement to be strict in the sense that the two densities differ over $\mathcal{X}^*$.

First, notice that this assumption is a restriction on the distribution of X^* given Z whereas the FOSD condition in Assumption 4 is a restriction on the distribution of X^* given X and Z. Restrictions on the former can easily be interpreted in concrete measurement error models but restrictions on the latter cannot. For example, if Z is an instrument for X^*, Assumption 8 is a requirement on how changing the value of the instrument from z_1 to z_2 shifts the distribution of X^*. To see this consider the examples in Figure 1. Each panel plots the densities $p_{X^*|Z}(\cdot|z_2)$ and $p_{X^*|Z}(\cdot|z_1)$. In panel (a), both densities are normally distributed in which case an arbitrarily small shift in Z from z_1 to z_2 ensures that the two densities cross only once. This is because the normal density is very smooth. Panel (b) shows a case in which the two densities are mixtures of normals and Z has to shift the conditional distribution further to avoid them crossing at several separated points. Panel (c) is an example of two compactly supported densities whose supports overlap. Even though the two densities do not equal each other over the intersection of their supports, the upper bound of the support of $X^*|Z = z_1$ is the point to the left of which

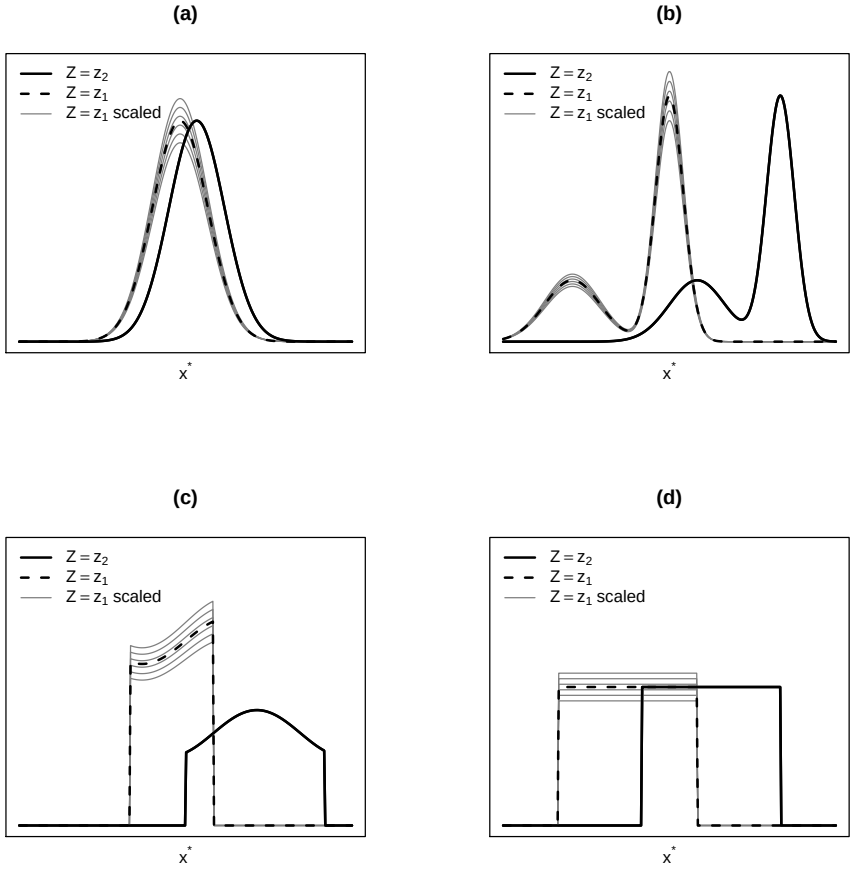

Figure 1: Examples of continuous densities $p_{X^*|Z}$ that satisfy the single-crossing condition. Panel (a) shows the normal case, panel (b) a mixture of normals, panel (c) two compactly supported distributions whose densities do not equal each other on the intersection of their supports, and panel (d) a case with infinitely many intersection points that are not separated.

$p_{X^*|Z}(\cdot|z_1)$ is above $p_{X^*|Z}(\cdot|z_2)$ and vice versa to its right. Therefore, the single-crossing property holds.[11] Finally, panel (d) shows the case of two uniform densities that are equal to each other over an interval, but still satisfy the single-crossing property. Similarly, single-crossing is trivially also guaranteed when the supports of $X^*|Z = z_1$ and $X^*|Z = z_2$ do not intersect. Of course, Z does not need to shift the location of the distribution of $X^*|Z$, but it may arbitrarily affect the shape of the distribution (as, for example, in (c)) as long as the single-crossing property is maintained. The grey lines in the graphs represent the density $p_{X^*|Z}(\cdot|z_1)$ scaled by constants $c \in \mathcal{C}$. In the cases (a) and (b), there are weak conditions on the distribution of $X|X^*$ under which $p_{X|Z}(\cdot|z)$ is continuous over whole $\mathbb{R}$, so that it is always possible to find a value $\bar{x}$ for which $p_{X|Z}(\bar{x}|z_1) = p_{X|Z}(\bar{x}|z_2)$ (e.g. as in Lemma 3). Then letting $\mathcal{X}$ be an arbitrarily

[11] Notice that Assumption 8 requires single-crossing only over the union of the two supports, outside of which the two densities of course equal each other.

small neighborhood of $\bar{x}$, the set $\mathcal{C}$ simply becomes an arbitrarily small neighborhood of 1. Therefore, the single-crossing property in Assumption 1(i) has to hold not only for $p_{X^*|Z}(\cdot|z_2)$ and $p_{X^*|Z}(\cdot|z_1)$, but also for $p_{X^*|Z}(\cdot|z_2)$ and the slightly in- and deflated function $c\,p_{X^*|Z}(\cdot|z_1)$. In the examples (c) and (d), this condition is trivially satisfied. In conclusion, the single-crossing property trades-off the strength of the effect Z has on the distribution of $X^*|Z$ and the complexity of that distributions shape.

It is important to emphasize that we only need to find two sets $\mathcal{Z}_1, \mathcal{Z}_2$ for which single-crossing holds, it does not have to hold for all values z_1, z_2 in the support of Z. If Z is discrete, then $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are just singletons each containing a different support point of Z. If Z is continuous, then the two sets could be arbitrarily small neighborhoods around two values z_1, z_2 in the support of Z.

Under additional assumptions, the single-crossing property for $\Delta p_{c,z}$ translates into single-crossing of the observable difference $p_{X|Z}(\cdot|z_2) - p_{X|Z}(\cdot|z_1)$ and can be nonparametrically tested by testing the implication that the distribution of $X|Z = z_2$ FOSD that of $X|Z = z_1$. See Appendix A.3 for more details.

In some models such as Kotlarski's repeated measurement model (Kotlarski (1967)), it is useful to write the single-crossing condition for the conditional density of $Z|X^*$ rather than of $X^*|Z$. For $c \in \mathbb{R}$ and $z = (z_1, z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2$, define the function $\Delta p'_{c,z} : \mathcal{S}_{X^*} \to \mathbb{R}$ as

$$\Delta p'_{c,z}(x^*) := p_{Z|X^*}(z_2|x^*) - c\,p_{Z|X^*}(z_1|x^*).$$

Also, let $\mathcal{C}'$ be the range of $p_{Z|X}(z_2|x)/p_{Z|X}(z_1|x)$ as (x, z_1, z_2) varies over $\mathcal{X} \times \mathcal{Z}_1 \times \mathcal{Z}_2$.

Assumption 8′ (single-crossing)**.** *There are a constant $c_q > 0$, an interval $\mathcal{I} \subset \mathbb{R}$, and a set $\mathcal{X} \subseteq \mathcal{S}_X$ such that $q_{X|X^*}(x|x^*)q_{X^*}(x^*) \geq c_q$ and $P(D^c|X^* = x^*) \geq c_q$ for all $x \in \mathcal{X}$, $x^* \in \mathcal{X}^* := \mathcal{I} \cap \mathcal{S}_{X^*}$, $z_k \in \mathcal{Z}_k$, $k = 1, 2$, and, for all $c \in \mathcal{C}'$ and $z \in \mathcal{Z}_1 \times \mathcal{Z}_2$:*

(i) $\Delta p'_{c,z}$ is single-crossing,

(ii) $\Delta p'_{c,z}(x^*) \geq c_q$ for all $x^* \in \mathcal{X}^*$.

Comments similar to those about Assumption 8 apply to Assumption 8′. Kotlarski's model is an important special case in which the single-crossing condition is satisfied under simple primitive conditions. To see this consider the model

$$X = X^* + \eta_X$$
$$Z = X^* + \eta_Z$$

where η_X, η_Z, and X^* are mutually independent and continuously distributed. In this case, the single-crossing condition is trivially satisfied when η_Z is compactly supported. If the support of η_Z is unbounded, then we need to add the requirement that the density of η_Z does not oscillate until infinity. This is a weak requirement and we are not aware of any commonly used distributions that violate it. The details can be found in Appendix B.2. Of course, the single-crossing condition does not require the measurement error to be classical (independent of X^*).

For example, it is easy to check that all nonclassical measurement error examples in Section 4 of Hu and Schennach (2008) satisfy the single-crossing condition.

We now state the second main result of this section, the sufficient conditions for the FOSD requirement of Assumption 4.

Theorem 3 (FOSD under single-crossing)**.** *Any P for which there exist sets $\mathcal{X}^*, \mathcal{X}, \mathcal{Z}_1, \mathcal{Z}_2$ satisfying Assumptions 5–7 and either 8 or 8′ also satisfies Assumption 4.*

Theorems 2 and 3 together imply that all P satisfying Assumptions 2–3, 5–7, and either 8 or 8′ are in $\mathbf{M}_R$. Even though $\mathbf{M}_R$ is a more restrictive set than $\mathbf{M}$, we argue that these assumptions are fairly weak and allow for a wide range of realistic nonclassical measurement error models (i.e. $\mathbf{M}_R$ is "large").

First, Appendix C provides simulation results for the test by Delgado and Gonzalez Manteiga (2001), which we use in the empirical part (Section 5), and confirms that it possesses power against several classical and nonclassical measurement error models that have been analyzed in the literature.[12]

Second, measurement error models in $\mathbf{M}_R$ can be close to the null along two dimensions: the probability of no measurement error, λ, may be close to one and/or the measurement error distribution, $Q_{X|X^*}$, may be close to a pointmass at X^*. In this sense, Theorem 2 and the sufficient conditions in Theorem 3 allow for a rich set of deviations from the null of no measurement error.

Third, a test of (8) can recover all measurement error models in $\mathbf{M}_R$ even though the distribution P of (Y, X^*, X, Z) in $\mathbf{M}_R$ is not identified. Of course, testing for the presence of measurement error is a much more modest goal than identification of the distribution of (Y, X^*, X, Z) and, thus, it is clear that the former can be achieved under weaker assumptions than the latter. Testing $H_0^{no\,ME}$ by testing (8) does not require solving for the distribution of unobservables from the distribution of observables. Theorems 2 and 3 imply that this inversion and thus identification of the true measurement error data-generating process is not necessary for testing $H_0^{no\,ME}$. Instead it suffices to check whether the observed distribution is compatible with the conditional independence in (8). In case of a violation, any distribution $P_{X^*|Y,X,Z}$ of the unobservables conditional on the observables such that $P_{Y,X,Z} \times P_{X^*|Y,X,Z} \in \mathbf{M}_R$ is then a potential candidate for the true data-generating process. The test has power against all of these alternatives, but is not able to point out which one generated the data. There are at least two advantages of this approach. The first advantage is that we avoid some strong assumptions that have been employed for identification of measurement error models. In particular, we

[12] Similarly as in nonparametric specification testing problems, we expect there to be a trade-off between being able to detect high- and low-frequency local alternatives (e.g. Fan and Li (2000), Horowitz and Spokoiny (2001)). For testing conditional independence, we are not aware of any work that has provided a rate-optimal test. In the simulations in Appendix C and the empirical part in Section 5, we therefore use the nonparametric test by Delgado and Gonzalez Manteiga (2001) which can detect low-frequency local alternatives that converge to the null at the fast $n^{-1/2}$-rate.

avoid completeness conditions on the distributions of $X|X^*$ and $X^*|Z$ when X, X^* are continuous. These are restrictive (Santos (2012)), untestable (Canay, Santos, and Shaikh (2013)), and required by even the most general identification results available (Hu and Schennach (2008)). Theorem 3 does not impose completeness on $X|X^*$ and replaces the completeness of $X^*|Z$ by a single-crossing condition. The latter has a simple graphical interpretation, can be tested under an additional assumption on the distribution of $X|X^*$ (Appendix A.3), and is compatible with incomplete distributions $X^*|Z$ (e.g. when X^* is continuous and Z binary). The second advantage is that not solving for the distribution of unobservables circumvents the need for somewhat complicated estimators involving matrix inversions (in the discrete case; see Hu (2008)) or operator inversions (in the continuous case; see Hu and Schennach (2008)). These estimators may be highly variable, especially in the continuous case in which convergence rates may be as slow as a logarithmic function of the sample size. On the other hand, nonparametric estimation of the conditional distribution in (8) is a significantly simpler statistical problem involving standard estimators and faster convergence rates.

To gain some insight into why the theorem holds it is useful to consider the density of $P_{X^*|X,Z}$. By Assumption 7, there exists a function f such that $P(D|X^* = x^*, Z = z) = f(x^*)$, and by Assumption 6 the density can then be written as

$$p_{X^*|X,Z}(x^*|x,z) = \left[\lambda f(x^*)\delta(x-x^*) + (1-f(x^*))q_{X|X^*}(x|x^*)\right] \frac{p_{X^*|Z}(x^*|z)}{p_{X|Z}(x|z)},$$

where δ is a pointmass at zero. While FOSD in the distribution of $X^*|Z = z$ for different values of z does not translate into FOSD in $X^*|X, Z = z$, the single-crossing condition in Assumption 8 *does* carry over from $X^*|Z$ to $X^*|X, Z$. We then use the fact that single-crossing of densities implies FOSD of their cdfs. Formalizing this argument, in particular that strict dominance in $X^*|Z$ implies strict dominance in $X^*|X, Z$, requires some additional effort because of the presence of the pointmass δ. The details are in Appendix E.

Remark 5 (additional controls)**.** Similarly as in Remark 2, it is straightforward to extend Theorem 3 to the case with additional, correctly measured covariates W. □

Remark 6 (mean independence)**.** Similarly as in Remark 3, we can replace the exclusion restriction in Assumption 2 by the conditional mean independence

$$E_P[\mu(Y)|X^*, X, Z] = E_P[\mu(Y)|X^*] \quad \text{a.s.}, \tag{14}$$

where μ is the same function as in Assumption 3. Theorem 2 then continuous to hold with (8) replaced by (9). The relative advantage of Theorem 2 over the equivalence in this remark is that the implementation of a test of (8), unlike (9), does not require knowledge of which conditional moment (the function μ) is monotone in x^*. On the other hand, the result in this remark only requires the weaker conditional mean independence. □

2.3 Relationship to the Hypothesis of Exogeneity

In linear regression models, measurement error in the explanatory variable and any other source of endogeneity manifest themselves in the same way, namely the correlation of the regressor with the regression error. Therefore, standard instrumental variable methods are able to account for both. In nonlinear models, measurement error and other forms of endogeneity require different treatment because the measurement error introduces an additional unobservable in the outcome equation that cannot be subsumed in the outcome equation errors (Schennach (2013)). In particular, testing for exogeneity and testing for the presence of measurement error are two distinct problems as we explain in more detail in the following paragraphs.

Exogeneity in the linear IV model. Testing (8) is related to but different from some specification tests such as those by Durbin (1954), Wu (1973), and Hausman (1978). Specifically, Hausman (1978) considers the linear regression model for Y and X^* with the possibility of classical measurement error in the observed regressor X and the availability of an instrument Z. He shows that testing for the presence of measurement error is equivalent to (I) testing whether OLS and IV estimands are equal, and to (II) testing whether the projection $\widetilde{Z}$ of X onto Z is statistically significant in a regression of Y on X and $\widetilde{Z}$. The strategy in part (I) could, in principle, be extended to our nonparametric problem of testing $H_0^{no\,ME}$ by comparing a standard kernel estimator of $P(Y \leq y|X = x)$ with a measurement error-robust estimator of $P(Y \leq y|X^* = x)$ (e.g. one of those surveyed in Chen, Hong, and Nekipelov (2011) or Schennach (2013)). However, this procedure would require solving for the joint distribution of Y and the latent variable X^*. This is particularly difficult in the case of continuous X^* as estimators are solutions to ill-posed inverse problems and therefore may be very imprecise. Directly testing (8) on the other hand avoids such nonstandard and possibly imprecise estimators. In addition, the approach based on solving for the distribution of unobservables would assume identification of those distributions which in turn requires much stronger assumptions than those in Theorem 2.

Exogeneity in the nonparametric IV Model. Testing (8) is related to but different from testing for exogeneity of a regressor in nonparametric instrumental variable models. Consider the model $Y = g(X) + \varepsilon$ where X is a potentially endogenous regressor and we observe an instrument Z that satisfies $E[\varepsilon|Z] = 0$. The null hypothesis of exogeneity of X is $E[\varepsilon|X] = 0$. It is easy to see that, under a completeness condition on the distribution of $X|Z$, this hypothesis is equivalent to

$$E[E(Y|X)|Z] = E[Y|Z] \tag{15}$$

which can be tested as in Blundell and Horowitz (2007), for example. Compare this hypothesis to the conditional mean version of the observable restriction in Remarks 3 and 6. Under the conditional mean independence (14) with $\mu(y) = y$, (15) is implied by, but does not imply (9). Therefore, (15) is not equivalent to the null of no measurement error and a test of exogeneity such as that in Blundell and Horowitz (2007) is not consistent against all alternatives in $\mathbf{M}_R$

violating $H_0^{no\,ME}$.

Exogeneity in the triangular model. Testing (8) is also related to testing exogeneity in the triangular model

$$\begin{aligned} Y &= \bar{g}(X, \bar{\varepsilon}) \\ X &= \bar{m}(Z, \bar{\eta}) \end{aligned} \tag{16}$$

where $\bar{\varepsilon}$ and $\bar{\eta}$ are unobservables.

Assumption 9. **(i)** *$Z \perp (\bar{\varepsilon}, \bar{\eta})$, where Z, $\bar{\eta}$ and $\bar{\varepsilon}$ are scalar random variables,* **(ii)** *$\bar{g}(X, \cdot)$ is strictly monotone with probability one and X a scalar random random variable,* **(iii)** *$\bar{m}(\cdot, \bar{\eta})$ and $\bar{m}(Z, \cdot)$ are both strictly monotone with probability one,* **(iv)** *$\bar{\eta}$ is continuously distributed with a cdf that is strictly increasing over the support of $\bar{\eta}$.*

This assumption contains the same conditions as Theorem 1 of Imbens and Newey (2009) and two additional assumptions: monotonicity of $\bar{g}(X, \cdot)$ and $\bar{m}(\cdot, \bar{\eta})$, and the condition that $\bar{\varepsilon}$ is scalar.

Lemma 1. *In the triangular model (16), suppose Assumption 9 holds. Then:*

$$\bar{\varepsilon} \perp \bar{\eta} \qquad \Leftrightarrow \qquad Y \perp Z \mid X.$$

This lemma shows that, under Assumption 9, exogeneity in the triangular model is equivalent to conditional independence of Y and Z given X, the same condition as in (8).

In contrast, consider the triangular model with measurement error, which is studied in more detail in Appendix B.3,

$$Y = g(X^*, \varepsilon) \tag{17}$$

$$X^* = m_Z(Z, \eta_Z) \tag{18}$$

where instead of X^* we observe X,

$$X = m_X(X^*, \eta_X) \tag{19}$$

This model can be rewritten as a triangular model of the form (16) under the additional assumption of strict monotonicity of $m_X(\cdot, \eta_X)$ with probability one. Let $m_X^{-1}(\cdot, \eta_X)$ denote the inverse function of $m_X(\cdot, \eta_X)$. Then (17)–(19) implies

$$\begin{aligned} Y &= g(m_X^{-1}(X, \eta_X), \varepsilon) \\ X &= m_X(m_Z(Z, \eta_Z), \eta_X) \end{aligned}$$

which becomes (16) with $\bar{\varepsilon} = (\varepsilon, \eta_X)$, $\bar{\eta} = (\eta_X, \eta_Z)$ and the implied choices for $\bar{g}$ and $\bar{m}$. In this representation as a triangular model, Assumption 9(i) is violated by construction because $\bar{\varepsilon}$ and $\bar{\eta}$ are both of dimension larger than one (unless g, m_X, and m_Z are all linear). Therefore, in the

triangular model, testing for measurement error is not a special case of testing for exogeneity. More generally, the set of maintained assumptions in Lemma 1 neither nest nor are nested by the assumptions defining $\mathbf{M}_R$. In consequence, in the triangular model, the conditional independence $Y \perp Z \mid X$ can be interpreted as the hypothesis of exogeneity or as the absence of measurement error under different sets of assumptions.[13]

3 The Null Hypothesis of Equal Functionals

The economic object of interest is often a functional of the conditional distribution of the outcome variable Y given the explanatory variable X^*. In such cases, we might be more interested in testing whether this functional is distorted by measurement error or not, rather than whether there is any measurement error at all. In this section, we therefore consider the null hypothesis of equal functionals,

$$H_0^{func}\colon P \in \mathbf{P}_0^{func} \tag{20}$$

versus $H_1^{func}\colon P \in \mathbf{P} \setminus \mathbf{P}_0^{func}$ where

$$\mathbf{P}_0^{func} := \{P \in \mathbf{P}\colon P\left(E_P[\Lambda_P(Y,X)|X^*,X] = 0\right) = 1\}$$

and Λ_P is a given function that may depend on P. Importantly, Λ_P takes as arguments only the observable variables Y and X, but neither X^* nor Z. In this section, Y is allowed to be a random vector.

For example, choosing $\Lambda_P(y,x) = y - E_P[Y|X=x]$ and imposing the exclusion in Assumption 2, the null hypothesis becomes a test for whether the true conditional mean of $Y|X^*$ is equal to that of the observed conditional distribution $Y|X$,

$$P\left(E_P[Y|X^*] = E_P[Y|X]\right) = 1. \tag{21}$$

Similarly, we could test equality of any other conditional moment of $Y|X^*$ and $Y|X$, e.g. entropy as a measure of inequality in the conditional distribution:

$$P\left(E_P[Y\log(Y)|X^*] = E_P[Y\log(Y)|X]\right) = 1. \tag{22}$$

To obtain an observable restriction that is equivalent to H_0^{func}, we employ the exclusion restriction from Assumption 2, the FOSD condition from Assumption 4, and the following monotonicity condition that is similar to Assumption 3.

Assumption 10. *There are sets $\mathcal{X}^* \subseteq \mathcal{S}_{X^*}$ and $\mathcal{X} \subseteq \mathcal{S}_X$ so that, for every $x \in \mathcal{X}$, the function $x^* \mapsto E_P[\Lambda_P(Y,X)|X^*=x^*, X=x]$ exists, is monotone over $\mathcal{S}_{X^*}$ and strictly monotone over $\mathcal{X}^*$.*

[13] Notice, however, that the assumptions of Lemma 1, which justify the interpretation of exogeneity are fairly strong, so there might be more attractive approaches to testing exogeneity in the triangular model.

First notice that the null hypothesis is not a contradiction to $x^* \mapsto E_P[\Lambda_P(Y,X)|X^* = x^*, X = x]$ being monotone (and non-constant). The important difference is that the monotonicity holds in x^* *for a fixed value of* x whereas, under the null, both X and X^* vary together so as to yield a zero conditional expectation.

In the conditional mean example of hypothesis (21), the monotonicity requires monotonicity of the true conditional mean $E_P[Y|X^* = \cdot]$. Similarly as in the discussion of Assumption 3, such a condition is often implied by economic theory when the conditional mean represents a production, utility or cost function, for example.

Definition 4. *Let* $\mathbf{M}^{func}$ *be the set of distributions* P *satisfying Assumption 2 and* $\mathbf{M}_R^{func}$ *the set of* $P \in \mathbf{M}^{func}$ *that also satisfy Assumptions 4 and 10, where the set* $\mathcal{X}^*$ *in Assumption 4 is that of Assumption 10.*

Theorem 4. (i) *If* $\mathbf{P} = \mathbf{M}^{func}$*, then* (20) *implies*

$$P(E_P[\Lambda_P(Y,X)|X,Z] = 0) = 1. \tag{23}$$

(ii) *If* $\mathbf{P} = \mathbf{M}_R^{func}$*, then* $P \in \mathbf{P}$ *satisfies* (20) *if, and only if, it satisfies* (23)*.*

The first part of Theorem 4 shows that the null hypothesis H_0^{func}, which depends on unobservables, implies the observable restriction in (23), provided the exclusion holds. The observable restriction can be tested similarly as the conditions in (8) and (9). A level-α test of (23) is thus also a level-α test of H_0^{func}.

The second part of the theorem shows that, under the additional conditions of monotonicity and FOSD, the null hypothesis is equivalent to the observable restriction. The maintained assumptions defining $\mathbf{M}_R^{func}$ are very similar to those in $\mathbf{M}_R$, so the comments on the generality of the equivalence result following Theorem 2 apply here, too.

Failing to reject the hypothesis H_0^{func} does not necessarily imply that there is no measurement error in X but only that, if there is any, it does not affect the functional defining Λ_P. In fact, in a finite sample, a test may fail to reject the hypothesis even when measurement error *does* affect the functional, but the distortion is small relative to sampling noise. In this sense, we can use the test to find out whether measurement error leads to severe enough distortions for the data being able to distinguish functionals based on true and mismeasured explanatory variables. Testing H_0^{func} for various functions Λ_P then allows us to explore for which type of economic questions measurement error matters and for which ones it does not. For example, measurement error might affect the average level of Y (i.e. (21) is violated), but not inequality in Y (i.e. (22) holds). In this case, policy questions depending on how a change in X^* affects inequality in the outcome can be answered without accounting for measurement error, but questions related to the effect on average outcomes cannot.

Remark 7 (additional controls)**.** Similarly as in Remarks 2 and 5, it is straightforward to extend Theorem 4 to the case with additional, correctly measured covariates W. □

Remark 8 (interpretation of the null)**.** The null (21) can be interpreted as follows. For a randomly drawn individual with true and observed characteristics X^* and X, $E[Y|X^*]$ and $E[Y|X]$ represent the (random) values of that individual's true and observed conditional means before X^* and X are realized. The null hypothesis (21) then requires these two conditional means to equal with probability one. It is important to note that this null is not equivalent to the condition

$$E_P[Y|X^* = x] = E_P[Y|X = x] \qquad \text{for all } x. \tag{24}$$

In general, (24) neither implies nor is implied by (21). One difficulty with phrasing the null as in (24), and which the formulation (21) avoids, is that, in general, the supports of X^* and X differ so that it is not clear for which values of x to require equality of the two moments. □

4 Recovering Measurement Error-Free Subpopulations

If we reject the null of no measurement error, it is not necessarily the case that X is mismeasured for all individuals in the population. In particular, there may exist interesting subpopulations that are free of measurement error and allow the researcher to carry out the intended analysis on such a subpopulation without accounting for measurement error. In this section, we describe a multiple testing procedure that recovers measurement error-free subpopulations from a set of pre-specified subpopulations.

Suppose we divide the population into K subpopulations and let $W \in \{1, \ldots, K\}$ be a random variable that indicates subpopulation membership. For example, W could be a function of individuals' observable characteristics such as age, gender, education, and so on. Consider the hypothesis of no measurement error in subpopulation k:

$$H_{0,k}^{no\,ME}\colon P \in \mathbf{P}_{0,k} := \{P \in \mathbf{P}\colon P(X = X^* \mid W = k) = 1\}. \tag{25}$$

Further, define $I(P) \subset \{1, \ldots, K\}$ to be the set of null hypotheses that are true under P, and the family-wise error

$$FWE_P := P\left(\text{reject at least one } H_{0,k}^{no\,ME}\colon k \in I(P)\right),$$

which is the probability that at least one true null hypothesis is rejected. We now describe a testing procedure that asymptotically (in large samples) controls the FWE_P at a nominal level α while rejecting as many false hypotheses as possible.

Testing (25) is different from testing $H_0^{no\,ME}$ in the presence of additional controls (Remarks 2 and 5) because the former hypothesis involves the conditional probability of measurement error and the latter the unconditional probability. However, under very similar conditions, we can show that $H_{0,k}^{no\,ME}$ holds if, and only if,

$$Y \perp Z \mid (X, \{W = k\}) \tag{26}$$

Therefore, every subpopulation-specific null hypothesis $H_{0,k}^{no\,ME}$ can be tested using a test of (26), e.g. any of those listed in Section 2.1.

To control the FWE_P when jointly testing all null hypotheses $H_{0,k}^{noME}$, $k = 1, \ldots, K$, we need to adjust the critical values of the individual tests. Some well-known procedures such the Bonferroni and Holm corrections (see the survey by Romano, Shaikh, and Wolf (2010), for example), are particularly simple to implement, but we now briefly describe how to use the StepM procedure in Romano and Wolf (2005, Algorithm 3.1), which may be considerably more powerful.

The test statistic by Delgado and Gonzalez Manteiga (2001), which we use in the simulations (Appendix C) and the empirical part (Section 5), is based on the Cramér-von Mises estimand

$$\theta_k(P) := E_P[T_k(X, Z, Y)^2]$$

with

$$T_k(x, z, y) := E_P\left[p_{X|W}(X|k)\Big\{\mathbb{1}\{Y \leq y\} - E_P(\mathbb{1}\{Y \leq y\}|W = k, X)\Big\}\iota_{x,z,k}(X, Z, W)\right]$$

and $\iota_{x,z,k}(X, Z, W) := \mathbb{1}\{X \leq x\}\mathbb{1}\{Z \leq z\}\mathbb{1}\{W = k\}$. The null hypothesis $H_{0,k}^{noME}$ is then equivalent to

$$H_{0,k}\colon P \in \{P \in \mathbf{P}\colon \theta_k(P) = 0\}.$$

Written in this form the StepM procedure by Romano and Wolf (2005) can directly be applied. The idea is to create a rectangular joint confidence set for $\theta_1(P), \ldots, \theta_K(P)$ and then reject the k-th hypothesis if the k-th dimension of the confidence set does not contain zero. In the second step, we form a joint confidence set for those $\theta_k(P)$ whose corresponding null hypotheses have not been rejected in the first step. Then we reject all further hypotheses for which zero is not in the relevant dimension of the new confidence set. We then iterate until no further hypotheses are rejected. By Theorem 3.1 in Romano and Wolf (2005), this procedure asymptotically controls the FWE_P at a given nominal level and leads to a consistent test.[14]

As an outcome of the multiple testing procedure we obtain a set of null hypotheses $H_{0,k}^{noME}$ that are not rejected. These indicate the subpopulations for which the test does not find any evidence of measurement error.

Remark 9. An and Hu (2012) show that if there is a pointmass at zero in the measurement error distribution, then solving for the distribution of unobservables as a function of the distribution of observables is a well-posed inverse problem. In consequence, estimators of the measurement error model possess faster convergence rates compared to those in the ill-posed inverse problems without a pointmass in the measurement error distribution. Since the multiple hypothesis testing procedure described in this section recovers measurement error-free subpopulations, it could be used as a way of providing evidence for An and Hu (2012)'s pointmass condition. □

Remark 10. If the researcher is interested in finding measurement error-free subpopulations among a huge number of subpopulations (large K), then control of the FWE_P may be too

[14] Delgado and Gonzalez Manteiga (2001) derive the limiting distribution of their test statistic and show that a multiplier bootstrap consistently estimates it. Therefore, the conditions of Theorem 3.1 in Romano and Wolf (2005) are satisfied.

demanding in the sense that the resulting multiple hypothesis testing procedure may possess poor power. In that case, one might want to choose a less demanding criterion such as those described in Romano, Shaikh, and Wolf (2008), for example. □

5 The Reliability of Administrative Data

In this section, we test for measurement error in the U.S. Social Security Administration's measure of earnings. While measurement error in survey responses is a wide-spread concern that has occupied a large literature (Bound, Brown, and Mathiowetz (2001)) only recently empirical researchers have emphasized concerns about the reliability of administrative data (e.g. Fitzenberger, Osikominu, and Völter (2006), Kapteyn and Ypma (2007), Abowd and Stinson (2007), Groen (2011), Bollinger, Hirsch, Hokayem, and Ziliak (2018)). First, administrative data is collected for the purpose of administration, which means that priorities in the data collection process may differ from those of a researcher and thus variables may not measure the economic concept a researcher would like them to capture. Second, a researcher typically gains access to only a small number of variables of an administrative dataset which means they have to be matched to other data sources containing, for example, more detailed demographics. Anonymization of the administrative dataset implies that mistakes in the matching procedure are difficult to avoid. Finally, firms may submit inaccurate earnings records to the administration. All these concerns lead to the possibility of observed administrative earnings deviating from true earnings and thus to measurement error. Since the use of administrative data is becoming increasingly popular in economics and researchers tend to view such datasets as free of measurement problems, we believe testing for measurement error in this context is of direct economic interest.

Earnings measures are used, for example, as dependent variable in human capital regressions and appear as explanatory variable in unemployment duration, labor supply, and consumption function models. Especially in the case of labor supply, there is compelling evidence that measurement error in earnings has a first-order effect on empirical findings (e.g. Ashenfelter (1984), Altonji (1986), Abowd and Card (1987)). The existing literature has argued that measurement error in earnings is negatively correlated with true earnings and positively correlated over time (e.g. Bound and Krueger (1991), Bound, Brown, Duncan, and Rodgers (1994), Bound, Brown, and Mathiowetz (2001)). Although these stylized facts about the data-generating process of the measurement error are based on a comparison of survey and administrative earnings data and the assumption that the administrative data are measured without error (the assumption we want to test in this section), they indicate at least that we should not rule out the possibility of such nonclassical measurement error properties a priori. In addition, nonlinearities in labor supply may be important (e.g. Blomquist and Newey (2002), Blomquist, Kumar, Liang, and Newey (2015)). The combination of nonclassical measurement error and a nonlinear model render identification and estimation of an econometric model that formally accounts for the presence of measurement error difficult.

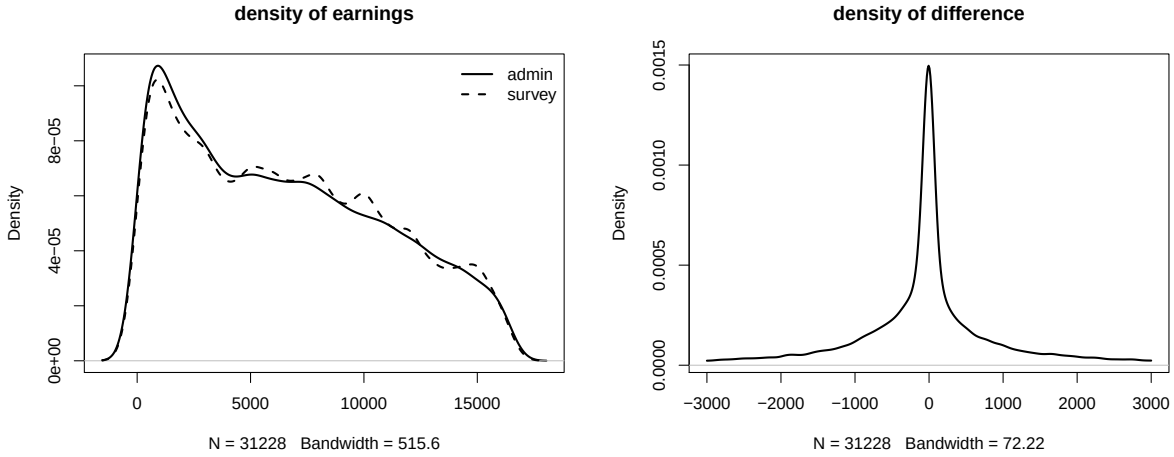

Figure 2: Nonparametric density estimates, using cross-validated bandwidths.

As shown in previous sections, the proposed test of the null of no measurement error can recover a broad class of nonclassical measurement error processes and, in the econometric specification described below, allows for dependence of measurement errors across time as well as dependence of the measurement errors on the true level of earnings.

Data The data come from the 1978 Current Population Survey-Social Security Earnings Records Exact Match File. The dataset contains a single cross section of survey data from the 1978 March CPS file, including annual earnings for 1977 ("survey earnings"), linked with a time series of social security earnings for several years leading up to and including 1977 ("administrative earnings"). More detailed information about the dataset can be found in Bound and Krueger (1991).

The original dataset contains 168,904 individuals. We restrict attention to individuals who work either full- or part-time, have no other income source apart from salary and wages, have positive administrative earnings, are an exact CPS-SER match, and are not top-coded (the threshold was \$16,500 in 1977). The resulting sample is of size 31,228. Table 4 in Appendix D shows summary statistics before and after sample selection. As expected, keeping only full- and part-time workers removes those who work less, thereby shifting the earnings and education distribution upwards. Figure 2 displays nonparametric estimates of the density of survey and administrative earnings in 1977, and of the difference of the two. The distribution of the difference is centered at zero, but possesses considerable probability mass at differences of more than $\pm\$1,000$, indicating that one or both of the earnings measures are contaminated measures of true earnings.

Econometric Specification Let earnings in time period two, E_2^*, depend on earnings in time period one through the following representation:

$$E_2^* = h(E_1^*, U_2)$$

where U_2 is some vector of unobservable shocks to earnings in period two. We observe a measure of earnings in the administrative dataset in both time periods, potentially with error,

$$A_t = m_{At}(E_t^*, \eta_{At}), \qquad t = 1, 2,$$

and we observe another measure of earnings in a survey dataset in period two, potentially with error,

$$S_2 = m_{S2}(E_2^*, \eta_{S2}).$$

In these representations, η_{At} and η_{St} are random vectors of measurement errors in the administrative and survey measures, respectively. Suppose $h(\cdot, U_2)$ is invertible with probability one with inverse function $h^{-1}(\cdot, U_2)$, then the model can be written as

$$\begin{aligned} S_2 &= m_{S2}(E_2^*, \eta_{S2}) \\ A_2 &= m_{A2}(E_2^*, \eta_{A2}) \\ A_1 &= m_{A1}(h^{-1}(E_2^*, U_2), \eta_{A1}) \end{aligned}$$

which fits into the framework of Section 2 with $Y = S_2$, $X^* = E_2^*$, $X = A_2$, $Z = A_1$, with measurement errors in Z and X defined as $\eta_Z = (U_2', \eta_{A1}')'$ and $\eta_X = \eta_{A2}$, respectively. The exclusion restriction in Assumption 1 holds if, under P,

$$\eta_{S2} \perp (\eta_{A1}, U_2) \mid E_2^*. \tag{27}$$

This condition requires the measurement error in the survey to be independent of the lagged measurement error in the administrative data, conditional on true earnings. This means that both measurement errors are allowed to be nonclassical, i.e. depend on the true level of earnings. This is important as existing work has argued for high-earners under-reporting and low-earners over-reporting their earnings (Bound, Brown, and Mathiowetz (2001)). The conditional independence therefore allows the measurement errors from the two datasets to depend on each other through the true earnings, but rules out direct dependence. We now argue why we believe this exclusion restriction to be reasonable. First, since the data collection in surveys and administrative data follow different protocols and are carried out by different individuals at different points in time, the sources of measurement errors are likely to be quite different. Second, our sample excludes workers who reported income other than salary and wages, e.g. tips and other side payments. Therefore, measurement errors in the survey and administrative earnings are unlikely to depend on each other because they both fail to fully include such forms of income. Notice also that the assumption (27) does not restrict the dependence of η_{A1} and η_{A2}, which means the administrative measurement errors may be arbitrarily dependent over time, even conditional on

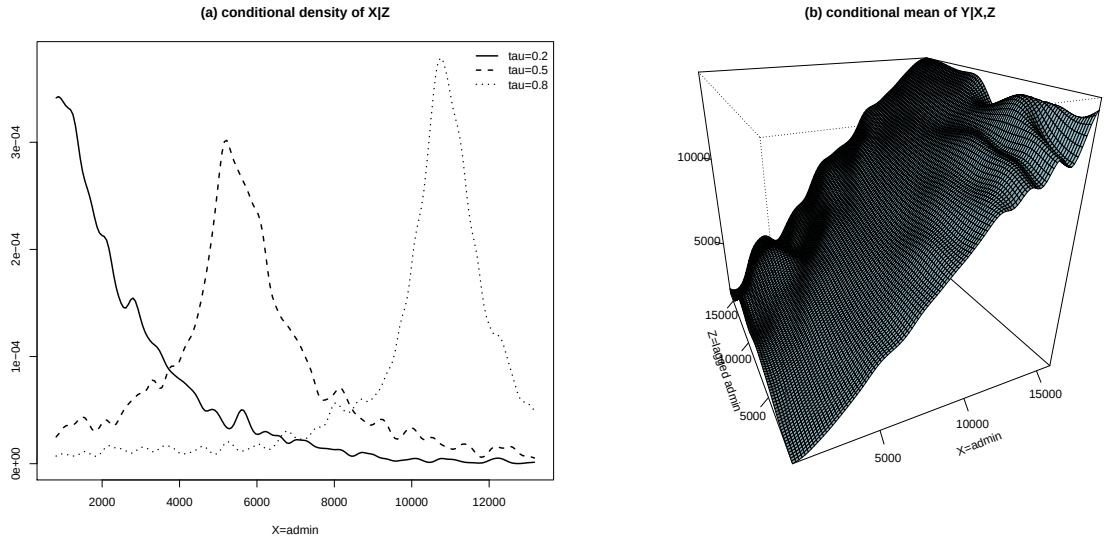

Figure 3: Panel (a) is showing a nonparametric estimate of the conditional density of $X|Z = q$ for q being the τ-quantile of Z and different values of τ. X and Z are administrative earnings in 1977 and 1976, respectively. Panel (b) shows a nonparametric estimate of $E[Y|X, Z]$, where Y is survey earnings in 1977, X is administrative earnings in 1977, and Z is administrative earnings in 1976. All bandwidths are chosen by cross-validation.

true earnings. (27) also requires the measurement error in the survey to be independent of the shock to earnings in the second period, U, conditional on E_2^*. This means the measurement error can depend on the shock, but only through contemporary earnings themselves.

The observable implication under the null of no measurement error and the exclusion restriction (27) then becomes

$$S_2 \perp A_1 \mid A_2. \tag{28}$$

One might be concerned that (27) is violated because high earnings growth could cause an individual to be less certain about the exact value of earnings, leading to measurement error in survey earnings that depends on the growth in true earnings.[15] To investigate this possibility, suppose there is an additional time period $t = 0$ before period $t = 1$ and replace the exclusion restriction in (27) by

$$\eta_{S2} \perp (\eta_{A0}, U_1, U_2) \mid (E_2^*, E_2^* - E_1^*), \tag{29}$$

which allows for survey and administrative measurement errors to depend not only on the level of of true earnings but also on its growth. The null of no measurement error in periods $t = 1, 2$ then yields the observable restriction

$$S_2 \perp A_0 \mid (A_2, A_2 - A_1). \tag{30}$$

See Appendix D for more details. We report results for tests of both restrictions, (28) and (30).

[15] See Appendix D for more details on why this leads to a violation of (27).

	level		growth			
	p-value	**h**	**p-value**	**h_1**	**h_2**	**n**
full sample	0.000	287.6	0.000	519.9	866.8	31,228
earnings in IQR	0.000	251.8	0.015	421.9	758.7	15,614
white males	0.000	127.1	0.000	691.3	1,055.0	12,591
+ singles	0.000	340.5	0.000	348.3	1,173.6	5,043
+ age ≥ 25	0.039	908.4	0.510	896.9	5,077.2	1,738
+ full-time, full-year	0.027	997.0	0.695	897.1	∞	985
+ at least highschool	0.026	827.7	0.813	464.5	2,424.3	877
+ earnings in IQR	0.015	758.0	0.109	168.5	1,892.2	346

Table 1: Test results for the null of no measurement error in 1977 administrative earnings based on (28) ("level") and (30) ("growth"). The table shows the p-value, the bandwidths chosen by cross-validation ("h", "h_1", "h_2"), and the sample size ("n").

Implementation We test the hypothesis of no measurement error as in Remark 3 by testing the observable restriction (9) for $\mu(y) = y$. We use the Cramér-von Mises test statistic in Delgado and Gonzalez Manteiga (2001) with the Epanechnikov kernel and the bandwidth chosen by cross-validation for the regression of Y on X. Critical values are computed by the multiplier bootstrap based on $1,000$ bootstrap samples and multipliers drawn from the two-point distribution in Härdle and Mammen (1993).

Results Panel (a) of Figure 3 provides empirical evidence for the relevance condition. It shows a nonparametric estimate of the conditional density of administrative earnings in 1977 given administrative earnings in 1976 being equal to their 0.2-, 0.5-, or 0.8-quantile. Moving the conditioning variable from its 0.2- to the 0.8-quantile induces a shift in the conditional density that is consistent with the single-crossing condition in Assumption 8 (Appendix A.3).

Panel (b) of Figure 3 shows a nonparametric estimate of the conditional mean of survey earnings given contemporary and lagged administrative earnings. If there was no measurement error in administrative earnings in 1977, then the conditional mean surface should not vary with lagged administrative earnings. The graph shows some variation in that dimension, but it is difficult to judge whether this variation is due to measurement error or just sampling noise.

Table 1 shows the results of the formal test based on the full sample (first row) and various subsamples (other rows). The columns denoted by "level" refer to the test of (28) and those denoted by "growth" to that of (30). The second line restricts the sample to those individuals with earnings in the interquartile range, thereby removing individuals in the extreme quartiles. The third row considers only white males and, in subsequent rows, this sample is then further narrowed by adding more and more restrictions. In the first four samples, the p-values of both the "level"- and the "growth"-test are numerically zero, which means we strongly reject the null of no measurement error. For the remaining subsamples, the "level"-test produces p-values that are positive, but below 4%, so we continue to reject the null at reasonable confidence levels even though the sample size drops to 346. The "growth"-test on the other hand produces large

	small h	large h	Z from 1975	Z from 1974	n
full sample	0.000	0.000	0.000	0.000	31,228
earnings in IQR	0.000	0.000	0.000	0.000	15,614
white males	0.000	0.000	0.000	0.000	12,591
+ singles	0.000	0.000	0.000	0.000	5,043
+ age $\geq$ 25 (*)	0.056	0.033	0.054	0.160	1,738
+ full-time, full-year	0.029	0.020	0.036	0.119	985
+ at least highschool	0.015	0.022	0.030	0.142	877
+ earnings in IQR	0.019	0.003	0.089	0.159	346

Table 2: Sensitivity analysis: testing for presence of ME in administrative earnings 1977; the table shows p-values for four cases: using a bandwidth that is 3/4 times the optimal cross-validated one ("small h"), using a bandwidth that is 5/4 times the optimal cross-validated one ("large h"), using for Z administrative earnings in 1975 ("Z from 1975"), and using for Z administrative earnings in 1975 ("Z from 1974").

p-values for the subsamples in which age is restricted to at least 25 years: it jumps from zero without the age restriction to 51% with the age restriction. Further restrictions to individuals in full-time, full-year work, and with at least a highschool degree lead to even higher p-values. These findings are consistent with the presence of young individuals having high earnings growth (in absolute terms) that leads to measurement error in the survey. Such dependence leads to a violation of the testable restriction (28) and thus to a false rejection of the null of no measurement error in the "level"-test that excludes young individuals. Interestingly, the p-value of the "growth"-test decreases from 81% down to 11% when we impose the additional restriction of earnings lying in the inter-quartile range. This finding is consistent with measurement error in administrative earnings being mainly due to young individuals with high earnings growth (in absolute terms) and earnings level outside the lower or upper quartile.

Sensitivity Analysis – Bandwidths The first two columns of Table 2 show the p-values for the "level"-tests as in Table 1 but using 3/4 times ("small h") and 5/4 times ("large h") the cross-validated bandwidth. The p-values barely change compared to those in Table 1 and remain under or around 5%, providing further evidence for the presence of measurement error in all subsamples.

Sensitivity Analysis – Definition of Z The third and fourth columns of Table 2 report the test results when using as Z administrative earnings in years 1975 and 1974. The p-values in the first four samples remain numerically zero, but increase more sharply as we start restricting the age of individuals, especially when Z is chosen from the 1974 data. In that case, the p-value climbs above 10%. To investigate potential reasons, Figure 4 shows nonparametric estimates of the conditional density of $X|Z$ and of the conditional mean of $Y|X,Z$ estimated on the subsample (*). The conditional mean of $Y|X,Z$ appears flat in the Z-dimension (panel (b)), which is consistent with the absence of measurement error. However, there are two alternative potential reasons explaining the lack of variation in the conditional mean surface. First, the

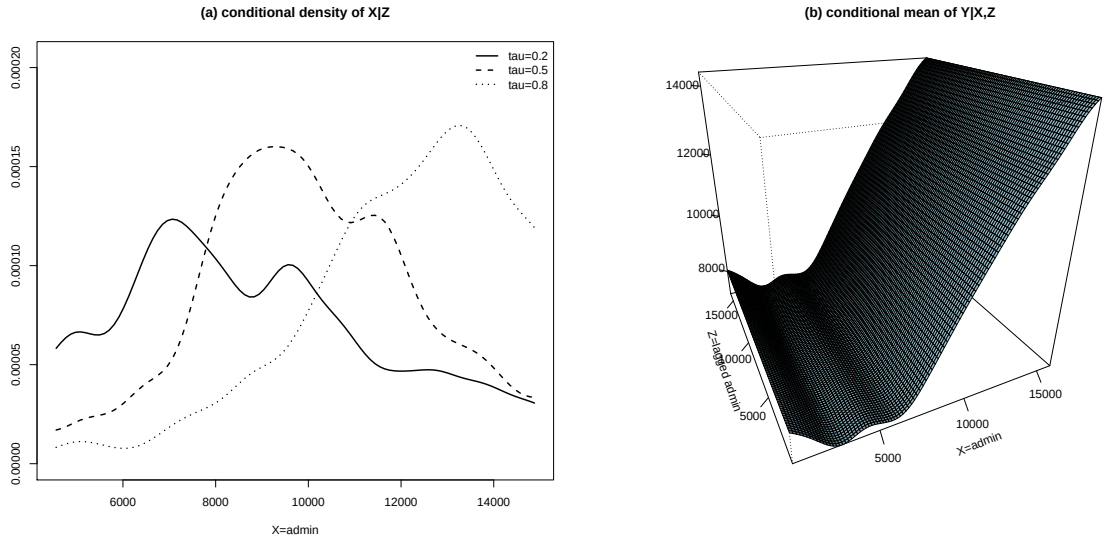

Figure 4: Sensitivity analysis for the restricted sample (*) in Table 2. Panel (a) is showing a nonparametric estimate of the conditional density of $X|Z = q$ for q being the τ-quantile of Z and different values of τ. X and Z are administrative earnings in 1977 and 1974, respectively. Panel (b) shows a nonparametric estimate of $E[Y|X, Z]$, where Y is survey earnings in 1977, X is administrative earnings in 1977, and Z is administrative earnings in 1974. All bandwidths are chosen by cross-validation.

sample size is smaller than in the full sample, which typically lets cross-validation choose a larger bandwidth and therefore leads to more smoothing. This means that the sampling noise in the smaller sample makes it harder to detect nonlinearities. Since the smaller sample size does not lead to such large p-values in Table 1, this is likely only a partial explanation. A second reason is the possibility that the relationship between true earnings X^* and administrative earnings Z from the distant past is weaker than for Z from the less distant past. Figure 4(a) shows that the two conditional distributions for $\tau = 0.2$ and $\tau = 0.8$ still cross only once, which is consistent with the single-crossing condition in Assumption 8 (Appendix A.3), but compared to the corresponding distributions with Z from 1976 there is significantly more probability mass in the intersection of their supports. One might interpret this finding as Z from 1974 possessing a "weaker" relationship to X^* than the Z from 1976. Therefore, it is likely the case that Z from 1974 induces less variation in Y than does Z from 1976, which could manifest itself in the flatter conditional mean surface and does not necessarily imply that there is no measurement error.

On balance, we therefore view the results of the sensitivity analyses as confirming the discovery of measurement error in administrative earnings present in Table 1.

6 Conclusions

In this paper, we propose a simple nonparametric test of the null hypothesis of no measurement error and of the null hypothesis that the measurement error, if there is any, does not distort a given object of interest. We argue that the hypothesis of no measurement error may be

of direct interest to the economist because, in some economic models, it corresponds to the hypothesis of no frictions. On the other hand, one could also test the null of no measurement error for the purpose of model selection, which precedes an intended empirical analysis on the same dataset. In that case, inference in the second step remains valid for fixed data-generating processes because the test of no measurement error is a consistent test. Of course, the second step inference incurs the usual uniformity problems as in Leeb and Pötscher (2005), but it would be possible to develop uniformly valid post-model selection inference similar to the recent developments on selective inference (e.g. Tibshirani, Taylor, Lockhart, and Tibshirani (2016), Tian and Taylor (2017), Tibshirani, Rinaldo, Tibshirani, and Wasserman (2018)). The idea is to derive the joint limiting distribution of the test statistic in the first-step and the estimator used after model selection under sequences of drifting data-generating processes. This is easy, for example, in the misclassification model of Appendix B.1 combined with the measurement error-robust estimator of Mahajan (2006) where we obtain joint normality. Then we can compute the second-step estimator's limiting distribution conditional on either having rejected or failed to reject the null of no measurement error and use this limiting distribution for uniformly valid inference conditional on the outcome of the first-step test.

A Remarks on the Null of No Measurement Error

A.1 Flipping the Null

In some situations, one might be interested in testing the "flipped" null hypothesis of some measurement error against the alternative hypothesis of no measurement error. A rejection of this null would be direct evidence of the absence of measurement error. Similarly as in Bahadur and Savage (1956) and Romano (2004) we conjecture that the flipped null is untestable because the set of distributions in the alternative lie in the closure of those satisfying the null. Therefore, one would have to test the null that there is measurement error bounded away from the no-measurement error case against the alternative of no measurement error. This specification of the null introduces a host of new difficulties. First, one would have to specify a metric in which to bound the measurement error away from the no-measurement error case (e.g. the variance of measurement error or the support of the measurement error distribution). Second, one would have to specify the level of this bound. The resulting test would then possess more or less power depending on these two user-specific choices and be susceptible to users choosing these to reach a desired testing outcome. Third, in practice, constructing the set of null distributions that satisfy the bound would require solving for (features of) the distribution of unobservables, which in the continuous case amounts to solving ill-posed inverse problems. For this to be possible, one would re-introduce some of the assumptions necessary for identification together with the statistical difficulties that arise from ill-posedness.

All of these complications can be avoided by testing the null of no measurement error against the alternative of some measurement error. In particular, the testing approach proposed in this paper is valid under weak assumptions (much weaker than those necessary for identification), easy to implement in practice, and does not require any user-chosen parameters to specify the null hypothesis. Even if the researcher may be more interested in testing the flipped null, the costs listed above most likely outweigh the benefits. Instead, it may be preferable to use the approach in this paper combined with a test of (8) that is known to possess desirable power properties. Then, one can at least be confident that failing to reject the null of no measurement error is due to the lack of evidence against it in the data, rather than the use of an inferior test (see Footnote 12 on this point).

A.2 Equivalence in the Multivariate Case

The result in Theorem 2 can be extended to the case of a multivariate explanatory variable $X^* = (X_1^*, \ldots, X_{d_x}^*)'$ in at least two different ways. The first approach would be to replace the FOSD condition in (11) by

$$P(X^* \in A | X = x, Z = z_1) \leq P(X^* \in A | X = x, Z = z_2) \tag{31}$$

for every increasing set[16] A and similarly for the condition (12). The monotonicity in Assumption 3 then becomes that the conditional expectation is increasing or decreasing as we increase x^*, where an inequality for a vector x^* means that each element satisfies the inequality. One could also replace (31) by the marginal FOSD condition

$$P(X_d^* \geq x_d^* | X = x, Z = z_1) \leq P(X_d^* \geq x_d^* | X = x, Z = z_2)$$

for each dimension $d = 1, \ldots, d_x$ and then assume that the copula of the vector X^*, conditional on $X = x, Z = z$, is the same for $z \in \mathcal{Z}_1$ and for $z \in \mathcal{Z}_2$ (Scarsini (1988b)).

A second approach is to keep the FOSD conditions in (11) and (12), where again an inequality between two vectors means that the inequality is satisfied element-by-element, and then replace the monotonicity condition in Assumption 3 by a suitable concept of multivariate monotonicity. Two such possibilities are the concepts of Δ-monotonicity by Rüschendorf (1980) and the definition of multivariate risk aversion in Scarsini (1988a).

A.3 Testing the Single-Crossing Condition

The relevance condition in Assumption 8 cannot directly be tested as it is a condition on the joint distribution of the latent variable X^* and the observable Z. However, under an additional restriction on the distribution of $X|X^*$, it is possible to convert this assumption into an observable, testable restriction. It is well-known in the statistics literature (e.g. Karlin and Rubin (1956); Karlin (1968)) that the single-crossing property is preserved under mixing with respect to a log-supermodular density. Specifically, consider the case in which X^*, X are continuously distributed and Assumptions 6–8 hold. Then, $p_{X|X^*,Z} = p_{X|X^*}$ and

$$p_{X|Z}(x|z_2) - p_{X|Z}(x|z_1) = \int p_{X|X^*}(x|x^*) \left[p_{X^*|Z}(x^*|z_2) - p_{X^*|Z}(x^*|z_1) \right] dx^*$$

is single-crossing in x if $p_{X|X^*}(\cdot|\cdot)$ is log-supermodular. This is a condition that can be graphically assessed and nonparametrically tested. For example, the single-crossing property implies that one of the two distributions of $X|Z = z_2$ and $X|Z = z_1$ FOSD the other (see the proof of Theorem 3). There exists a large number of nonparametric tests for this implication (e.g. McFadden (1989), Barrett and Donald (2003); Linton, Maasoumi, and Whang (2005), Lee, Linton, and Whang (2009), Chetverikov, Kim, and Wilhelm (2018) among many others).

A.4 Endogeneity

The exclusion restriction in Assumption 1 allows for endogeneity in the following sense. Consider a representation of the outcome model of the form $Y = g(X^*, \varepsilon)$, where ε is a vector containing all other determinants of the outcome beyond X^*. Assumption 1 is then equivalent to $\varepsilon \perp Z \mid X^*$ which means ε may be a function of X^* and thus endogenous.

[16] A set is increasing if its indicator function is.

However, the exclusion restriction is not necessarily compatible with a triangular structure $X^* = h(Z, U)$ as in Imbens and Newey (2009). Their Theorem 1 requires the unconditional independence condition $\varepsilon \perp Z$ rather than the conditional independence $\varepsilon \perp Z \mid X^*$ required in this paper. Economic justifications for why an instrument Z is excluded from an outcome equation do not recognize this subtle difference and could therefore be used to motivate either assumption. Nevertheless, since triangular models have been studied extensively, it is interesting to see under which conditions they are compatible with our exclusion restriction. First, this is the case when there is no endogeneity, i.e. $U \perp \varepsilon$. Second, suppose, in addition to Y, X, and Z, we observe a second measurement W of X^*. In that case, our exclusion restriction can be imposed on W (i.e. $Y \perp W \mid X^*$) instead of Z. Then, the dependence between ε and U may be left unrestricted, restoring compatibility with endogeneity generated by triangular models such as Imbens and Newey (2009). See Appendix B.3 for details.

B Sufficient Conditions for Specific Measurement Error Models

In this section, we provide sufficient conditions for data-generating processes to satisfy the restrictions in $\mathbf{M}_R$ and discuss the assumptions in the context of economic examples. We start with a binary misclassification model and the continuous repeated measurement framework, more restrictive versions of which have been extensively studied in the measurement error literature. The third is a triangular instrumental variable model. To the best of our knowledge, none of these examples can be identified or consistently estimated using existing results unless further restrictions are imposed. However, testing (7) is possible in each one of them.

B.1 Binary Misclassification Model

Consider the case in which X^*, X, and Z are binary, but the support of Y is unrestricted. There are many economic examples that fit this setup and in which existing studies have raised concerns about measurement error in X. Well-known examples include studies of the effect of union status on wages (Card (1996)) and of the returns to college (Kane and Rouse (1995); Kane, Rouse, and Staiger (1999)).

Assumption B1. $\mathcal{S}_X = \mathcal{S}_{X^*} = \mathcal{S}_Z = \{0, 1\}$ *and* $\mathcal{S}_{X,Z} = \{0, 1\}^2$.

Assumption B2. $0 < P(X^* = 1|Z = 0) < P(X^* = 1|Z = 1) < 1$

Assumption B3. *There is a measurable function* $\mu \colon \mathcal{S}_Y \to \mathbb{R}$ *so that* $E_P[\mu(Y)|X^* = 0] \neq E_P[\mu(Y)|X^* = 1]$.

Assumption B4. *Under* P, $X \perp Z \mid X^*$.

By the definition of the support, each random variable X, X^*, Z has to attain 0 and 1 with positive probability. Assumption B2 rules out the case in which Z perfectly predicts X^* and requires a monotone relationship between X^* and Z in the sense that the larger value of

X^* is more likely to occur when Z also attains the larger value. This is a relevance condition that ensures X^* and Z are sufficiently related and, together with Assumption B4 implies the FOSD condition of Assumption 4. Assumption B3 requires that some moment of Y conditional on X and $X^* = x^*$ varies with x^*. Assumption B4 is the same as Assumption 6 except that the we impose the conditional independence directly on P rather than separately on Q (as in Assumption 6) and the probability of measurement error (as in Assumption 7).

Example 1 (effect of unions on the structure of wages)**.** In Card (1996), the outcome Y is an individual's wage in period t, X^* is union status in period t, X is reported union status in period t, and Z is reported union status in period $t-1$. To simplify the discussion suppose there are no additional covariates in the outcome equation. Assumption B2 requires that if an individual reports to be unionized in period $t-1$, then the individual is more likely to actually be unionized in period t than not. Assumption B3 means that union status has an effect on some conditional moment of wages. Finally, letting Z^* be union status in period $t-1$, Card (1996) assumes that $p_{X,Z|X^*,Z^*} = p_{X|X^*}p_{Z|Z^*}$, which implies Assumption B4. □

Assumption B2 implies

$$P(X^* = 1|Z = 0)P(X^* = 0|Z = 1) < P(X^* = 1|Z = 1)P(X^* = 0|Z = 0).$$

Multiplying both sides by $P(X = 1|X^* = 1)P(X = 1|X^* = 0)$ and dividing by $P(X = 1|Z = 1)P(X = 1|Z = 0)$ then yields

$$\begin{aligned} &P(X^* = 1|X = 1, Z = 0)P(X^* = 0|X = 1, Z = 1) \\ &\qquad < P(X^* = 1|X = 1, Z = 1)P(X^* = 0|X = 1, Z = 0), \end{aligned} \tag{32}$$

which means that $(x^*, z) \mapsto P(X^* = x^*|X = 1, Z = z)$ is a log-supermodular function. This property is well-known to imply the FOSD of Assumption 4.[17] We therefore obtain the following result.

Lemma 2. *Any P satisfying Assumptions 2, B1–B4 is in $\mathbf{M}_R$.*

Since distributions satisfying Assumptions 2, B1–B4 are in $\mathbf{M}_R$, Theorem 2 implies that, for such distributions, (7) is equivalent to (8). Therefore, a test of (8) rejects against and, in turn, is able to recover any binary misclassification model satisfying Assumptions 2, B1–B4.

Remark 11. If, in Lemma 2, Assumption 2 is replaced by mean independence as in Remark 3, then the statement continues to hold. This result is related to Lemma 4 in Mahajan (2006). In the absence of correctly measured covariates in the outcome equation, Mahajan (2006)'s result is a special case of Lemma 2 as he assumes, in addition, that $\mu(y) = y$ and that misclassification is not too severe in the sense that $P(X = 1|X^* = 0) + P(X = 0|X^* = 1) < 1$. □

[17] Simply add $P(X^* = 1|X = 1, Z = 0)P(X^* = 0|X = 1, Z = 1)$ to both sides of (32) and set $\mathcal{X}^* = \mathcal{X} = \mathcal{Z}_2 = \{1\}$, $\mathcal{Z}_1 = \{0\}$.

B.2 Continuous Repeated Measurements

Consider the case of a general outcome model

$$Y = g(X^*, \varepsilon) \tag{33}$$

in which X and Z are repeated measurements of a continuous explanatory variable X^* with additively separable measurement errors:

$$X = X^* + \eta_X \tag{34}$$

$$Z = X^* + \eta_Z \tag{35}$$

We focus the discussion of this section on applications to the study of production functions with latent inputs such as cognitive and noncognitive skills (e.g. Cunha, Heckman, and Schennach (2010), Heckman, Pinto, and Savelyev (2013), Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2015), Attanasio, Meghir, and Nix (2017) among many others),[18] but there are many other examples in which the above model has been used, e.g. for estimation of Engel curves (Hausman, Newey, and Powell (1995)) or in the context of earnings dynamics (Horowitz and Markatou (1996), Bonhomme and Robin (2010), Arellano, Blundell, and Bonhomme (2017)).

Remark 12. In a typical application, X and Z could for instance be two different test scores attempting to measure the same type of skill or test scores from the same test conducted in two different periods. Therefore, the absence of measurement error in one of the two measures does not necessarily imply the absence of measurement error in the other and our hypothesis $H_0^{no\,ME}$, which is only involving one of the two measures, cannot be tested by simply checking whether the variance of $X - Z$ is zero or not. □

First, we directly impose the exclusion and monotonicity assumptions from Section 2.2.

Assumption RM1. $\varepsilon \perp (\eta_X, \eta_Z) \mid X^*$ *under* P.

This is the exclusion restriction of Assumption 2 rephrased in terms of the errors in the model (33)–(35).

Assumption RM2. *The monotonicity condition, Assumption 3, holds.*

The remaining assumptions restrict the measurement system (34)–(35) and, apart from one regularity condition, are implied by the conditions of the well-known model by Kotlarski (1967) and its extension by Evdokimov and White (2012). This is perhaps the most-studied measurement error model in the econometrics literature and, in particular, the standard model for empirical work related to the technology of skill formation[19], so it may be instructive to see how it fits into the framework of this paper.

[18] Some of the studies cited here do in fact consider a generalization of (34)–(35) that includes a coefficient in front of X^* and an intercept. The discussion in this section trivially extends to that case.

[19] Subject to the comment in Footnote 18.

Assumption RM3. **(i)** *P_{X^*,η_X,η_Z} has a continuous density with respect to Lebesgue measure on $\mathbb{R}^3$, is supported over some Cartesian product of intervals (each possibly equal to $\mathbb{R}$), and satisfies (34)–(35).* **(ii)** *There exists $c \in \mathbb{R}$ such that the density $p_{\eta_Z}(\eta)$ is monotone for all $|\eta| > c$.*

The first part assumes that the variables X^*, η_X, η_Z and thus also X and Z are continuous. In consequence, there is no pointmass in the distribution of η_X, so the whole population contains some measurement error in X (the measurement error distribution may, of course, be arbitrarily tightly centered around zero). The second part requires that the density of the measurement error η_Z becomes monotone, at least for large enough values. This condition rules out densities that oscillate until infinity, but allows arbitrarily many oscillations over a bounded subset of $\mathbb{R}$. We are not aware of any common distribution that violates this condition. As a special case all compactly supported distributions trivially satisfy it.

Assumption RM4. *X^*, η_X, η_Z are mutually independent under P.*

Mutual independence of X^*, η_X, η_Z implies classical measurement error in both measurements X and Z. This is a strong assumption, but nevertheless is imposed in all empirical work studying the technology of skill formation that we are aware of.

The Assumptions RM3–RM4 are implied by Kotlarski's model except part (ii) of Assumption RM3. It should be mentioned, however, that identification of Kotlarski's model requires additional assumptions.[20]

Example 2 (technology of skill formation)**.** Most recent papers (e.g. those cited above) in the literature on identification and estimation of the technology of skill formation make very similar assumptions about the measurement error in observed inputs. For concreteness, consider one particular paper by Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2015). In their setup, we could test for measurement error in, say, cognitive skills by letting X, Z, and Y be three measures of cognitive skills. For example, let Y be an assessment by a trained psychologist (e.g. the Bayley test) and X and Z be measurements reported by the mother of the child (e.g. the scores on the Early Children's Behavior Questionnaire and on the Infant Characteristics Questionnaire). Since the mother's reports are collected by an interviewer, at the mother's home on a different day than the psychologist's measurement which took place at a community center, the measurement errors in Y and (X, Z) have very different origins. This justifies the exclusion restriction, Assumption RM1. Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2015) also directly assume that the measurement errors in X and Z are classical and independent of each other (Assumption RM4). The monotonicity in Assumption RM2 follows from the linearity in the measurement system and the assumption that the measurement errors are classical. Finally, they specify the joint distribution of X^*, η_X, η_Z to be normal which implies Assumption RM3 as a special case. □

[20] Evdokimov and White (2012) also require the characteristic function of one of the measurement errors η_X, η_Z to possess only isolated zeros, some restrictions on the characteristic function (and its derivative) of the other measurement error, and measurement error that is mean zero. These assumptions are not imposed here.

Lemma 3. *Any P satisfying Assumptions RM1–RM4 is in $\mathbf{M}_R$.*

Since distributions satisfying Assumptions RM1–RM4 are in $\mathbf{M}_R$, Theorem 2 implies that, for such distributions, (7) is equivalent to (8). Therefore, a test of (8) rejects against and, in turn, is able to recover any Kotlarski-like measurement system satisfying Assumptions RM1–RM4.

As the discussion of Assumptions 8 and 8′ showed, the measurement error models in $\mathbf{M}_R$ do not have to be classical as in Assumption RM4 above. In particular, the measurement errors are allowed to depend on each other through the latent explanatory variable X^*. For example, the nonclassical measurement error models of Section 4 in Hu and Schennach (2008) are of the form

$$X = X^* + \sigma(X^*)U_X$$
$$Z = X^* + \sigma(X^*)U_Z$$

where σ is a monotone function and U_X, U_Z, X^* are mutually independent. Their examples also satisfy the single-crossing condition in Assumption 8′ and lead to distributions in $\mathbf{M}_R$.

We now want show that, in addition, we can even allow for some forms of direct dependence between the measurement errors. To make this point, we introduce an "individual-specific effect" into the measurement errors of the repeated measurement model above:

$$X = X^* + \alpha + U_X \tag{36}$$
$$Z = X^* + \alpha + U_Z \tag{37}$$

where α is an (unobserved) individual-specific effect and U_X and U_Z are mutually independent errors that are also independent of X^* and α. In this model, $\eta_X = \alpha + U_X$ and $\eta_Z = \alpha + U_Z$. The reason for why α is not simply absorbed by X^* is that the exclusion restriction in Assumption RM1 holds conditional on X^* rather than $X^* + \alpha$, reflecting the fact that X^*, rather than $X^* + \alpha$, is the explanatory variable that determines Y. For example, when X and Z are survey responses, the individual-specific effect could represent an individual's ability to correctly remember the answer to the survey question. The way an individual misreports X is then likely similar to the way she misreports Z.

Assumption RM5. **(i)** *P_{X^*,α,U_X,U_Z} has a continuous density with respect to Lebesgue measure on $\mathbb{R}^4$, is supported over some Cartesian product of intervals (each possibly equal to $\mathbb{R}$), and satisfies (36)–(37).* **(ii)** *There exists $c \in \mathbb{R}$ such that the density $p_{U_Z}(u)$ is monotone for all $|u| > c$.*

Assumption RM6. *Under P, U_X, U_Z are independent and jointly independent of (X^*, α).*

Assumption RM7. *$(x^*, v) \mapsto p_{\alpha|X^*}(v - x^*|x^*)$ is log-supermodular*[21].

[21] A function $f : \mathbb{R}^2 \to \mathbb{R}$ is called log-supermodular if for any $x_1, x_2, y_1, y_2 \in \mathbb{R}$ with $x_1 \leq x_2$ and $y_1 \leq y_2$, $f(x_1, y_2)f(x_2, y_1) \leq f(x_1, y_1)f(x_2, y_2)$.

If the density $p_{\alpha|X^*}$ is twice-differentiable, then it is log-supermodular in the above sense if $\partial^2 \log p_{\alpha|X^*}(v - x^*|x^*)/\partial x^* \partial v \geq 0$. In the special case in which α is independent of X^*, i.e. α is a "random effect" according to the language of panel data models, the condition holds for most well-known members of the exponential family (in particular the normal distribution), the uniform distribution, and is preserved by multiplication, convolution, marginalization and other operations on densities (see, e.g., Saumard and Wellner (2014)).

Lemma 4. *Any P satisfying Assumptions RM1–RM2, RM5–RM7 is in $\mathbf{M}_R$.*

Since distributions satisfying Assumptions RM1–RM2, RM5–RM7 are in $\mathbf{M}_R$, Theorem 2 implies that, for such distributions, (7) is equivalent to (8). Therefore, a test of (8) rejects against and, in turn, is able to recover any measurement system (36)–(37) satisfying those assumptions.

B.3 Triangular Instrumental Variable Model

Consider the following triangular structure

$$Y = g(X^*, \varepsilon) \tag{38}$$

$$X^* = m_Z(Z, \eta_Z) \tag{39}$$

where instead of X^* we observe X,

$$X = m_X(X^*, \eta_X) \tag{40}$$

and the unobservables ε, η_X, and η_Z may be random vectors. This model is a generalization of the measurement error model in Schennach (2007) in which Z is an instrumental variable for the explanatory variable X^*. The linear version of this measurement error model, i.e. g, m_X, m_Z are linear in all arguments, is commonly used in applied work. The nonlinear version, however, is not.[22] In particular, we are not aware of any work showing identification of the nonlinear model when X^*, X are continuous, but Z is binary. This case is of practical importance as instruments in empirical work tend to be discrete or binary.

Assumption IV1. *Assumption 5 holds.*

Notice that the instrument Z may be binary even when X^*, X are continuous.

Assumption IV2. *Under P:* **(i)** $\varepsilon \perp (Z, \eta_X, \eta_Z)$ *and* **(ii)** $\eta_X \perp (Z, \eta_Z)$.

Part (i) of this assumption requires the usual exclusion restriction in triangular models, i.e. that the instrument Z is independent of the error ε in the outcome equation. In addition, however, it also requires independence between the latter and the measurement errors. In repeated measurement models (such as those in Section B.2), it is common to assume some

[22] There are somewhat more technical papers showing identification of the nonlinear model when X^*, X, Z are continuous, which include empirical applications, e.g. Song, Schennach, and White (2015), De Nadai and Lewbel (2016), il Kim and Song (2018).

form of independence between the measurement errors and the structural error ε, but in the triangular model (38)–(40) it implies that X^* is independent of ε, thereby ruling out endogeneity of X^*. Just as identification arguments based on instrumental variables in nonlinear models can be used to correct for measurement error by shutting down all other channels of endogeneity (Schennach (2013, 2016)), here we can use the instrument to test for the presence of measurement error (distortions) by shutting down all other channels of endogeneity.

Below we show that endogeneity can be allowed for if, in addition to Y, X, Z, we also observe a second measurement of X^*. Part (ii) implies that the measurement error in X must be classical.

Assumption IV3. (38)–(40) *hold for some measurable functions* g, m_Z, m_X. *For every* $e \in \mathcal{S}_\varepsilon$, $g(\cdot, e)$ *is increasing over* $\mathcal{S}_{X^*}$ *and strictly increasing over some interval* $\mathcal{X}^* \subseteq \mathcal{S}_{X^*}$.

This assumption imposes monotonicity of $g(\cdot, e)$ which is often implied by economic theory if g is a production, cost, or utility function, for example. See the discussion of Assumption 3.

Assumption IV4. *Assumptions 7 and 8 hold with* $\mathcal{X}^*$ *from Assumption IV3.*

This assumption directly imposes the single-crossing condition on $p_{X^*|Z}$ and that the probability of measurement error does not vary with Z conditional on X^*.

Lemma 5. *Any* P *satisfying Assumptions IV1–IV4 is in* $\mathbf{M}_R$.

Since distributions satisfying Assumptions IV1–IV4 are in $\mathbf{M}_R$, Theorem 2 implies that, for such distributions, (7) is equivalent to (8). Therefore, a test of (8) rejects against and, in turn, is able to recover the triangular model (38)–(40) with classical measurement error in X. The instrument Z allows us to deal with measurement error in X (in the sense of being able to test for it), but not with any other form of endogeneity in X.

Classical measurement error and no endogeneity in X^* are both undesirable assumptions, but both can be removed when we observe an additional, second measurement of X^*, say W. Consider the triangular structure (38)–(40) augmented by an equation for the second measurement W,

$$W = m_W(X^*, \eta_W). \tag{41}$$

In this case, the second measurement W can play the role of the variable Z in Section 2. Under the assumptions of Theorem 2, the null hypothesis is then equivalent to $Y \perp W \mid X$. The test of this hypothesis ignores the first-stage equation (39), not using the instrument Z, so that η_Z and ε may be arbitrarily dependent, thereby allowing X^* to be endogenous and measurement error in X and W to be nonclassical.

C Simulations

In this section, we report the results of Monte Carlo simulations that explore the finite sample performance of a test for the null of no measurement error, $H_0^{no\,ME}$. We consider the following

outcome equation

$$Y = X^{*2} + \frac{1}{2}X^* + N(0, \sigma_\varepsilon^2)$$

with different models for the measurement system:

Model I (two measurements with independent and homoskedastic ME):

$$X = X^* + D \cdot N(0, \sigma_{ME}^2)$$
$$Z = X^* + N(0, 0.3^2)$$

with $\sigma_\varepsilon = 0.5$.

Model II (two measurements with conditionally independent and heteroskedastic ME):

$$X = X^* + D \cdot N(0, \sigma_{ME}^2)e^{-|X^*-0.5|}$$
$$Z = X^* + N(0, 0.3^2)$$

with $\sigma_\varepsilon = 0.5$.

Model III (two measurements with dependent and heteroskedastic ME):

$$X = X^* + D \cdot N(0, \sigma_{ME}^2)e^{-|X^*-0.5|}$$
$$Z = X^* + N(0, 0.3^2)e^{-|X^*-0.5|}$$

with $\sigma_\varepsilon = 0.5$.

Model IV (nonlinear relationship between X^* and Z):

$$X = X^* + D \cdot N(0, \sigma_{ME}^2)$$
$$Z = -(X^* - 1)^2 + N(0, 0.2^2)$$

with $\sigma_\varepsilon = 0.2$.

In all four models, $X^* \sim U[0, 1]$ and the random variable D is Bernoulli$(1 - \lambda)$, where $1 - \lambda$ is the probability of measurement error occurring in X. $1 - \lambda = 0$ means there is no measurement error in X, which represents the null hypothesis. To generate alternatives, we increase $1 - \lambda$ on a grid up to one. We also vary the standard deviation of the measurement error in X, σ_{ME}, in $\{0.2, 0.5, 1\}$. Therefore, alternatives get closer to the null as we decrease $1 - \lambda$ and/or σ_{ME}. We also vary the sample size $n \in \{200, 500\}$, but all models are simulated on $1,000$ Monte Carlo samples.

We present results for four different tests. They all test the condition (9) for $\mu(y) = y$, i.e. whether the conditional mean of Y depends on Z or not. First, we run a linear regression of Y on X and Z and then employ the standard t-test for the two-sided hypothesis that the coefficient in front of Z is equal to zero. The other three tests are the Cramér-von Mises test proposed in Delgado and Gonzalez Manteiga (2001) using three different bandwidths: $h = 0.2n^{-1/3}$ ("DM")

n	test	Model I	Model II	Model III	Model IV
200	t	0.043	0.051	0.043	0.418
	DM-	0.084	0.084	0.081	0.078
	DM	0.058	0.073	0.068	0.062
	DM+	0.049	0.069	0.060	0.100
500	t	0.067	0.052	0.035	0.798
	DM-	0.076	0.066	0.069	0.074
	DM	0.067	0.053	0.064	0.073
	DM+	0.078	0.056	0.060	0.076

Table 3: Null rejection probabilities

which is the rule-of-thumb bandwidth proposed in Delgado and Gonzalez Manteiga (2001), a smaller bandwidth $h = 0.1n^{-1/3}$ ("DM-") and a larger bandwidth $h = 0.5n^{-1/3}$ ("DM+"). The DM tests use the Epanechnikov kernel and the critical values are computed using 100 bootstrap replications. The nominal level is 5% for all tests.

Table 3 shows the null rejection probabilities of the tests and Figures 5–8 the corresponding power curves. Overall the test by Delgado and Gonzalez Manteiga (2001) controls size well and possesses power against all alternatives. The t-test is more powerful in some scenarios, which is not surprising since it is based on a parametric estimator that converges at a faster rate than the nonparametric estimator in the test by Delgado and Gonzalez Manteiga (2001). However, this test does of course fail to control size when the linearity of the outcome equation is violated (see Figure 8).

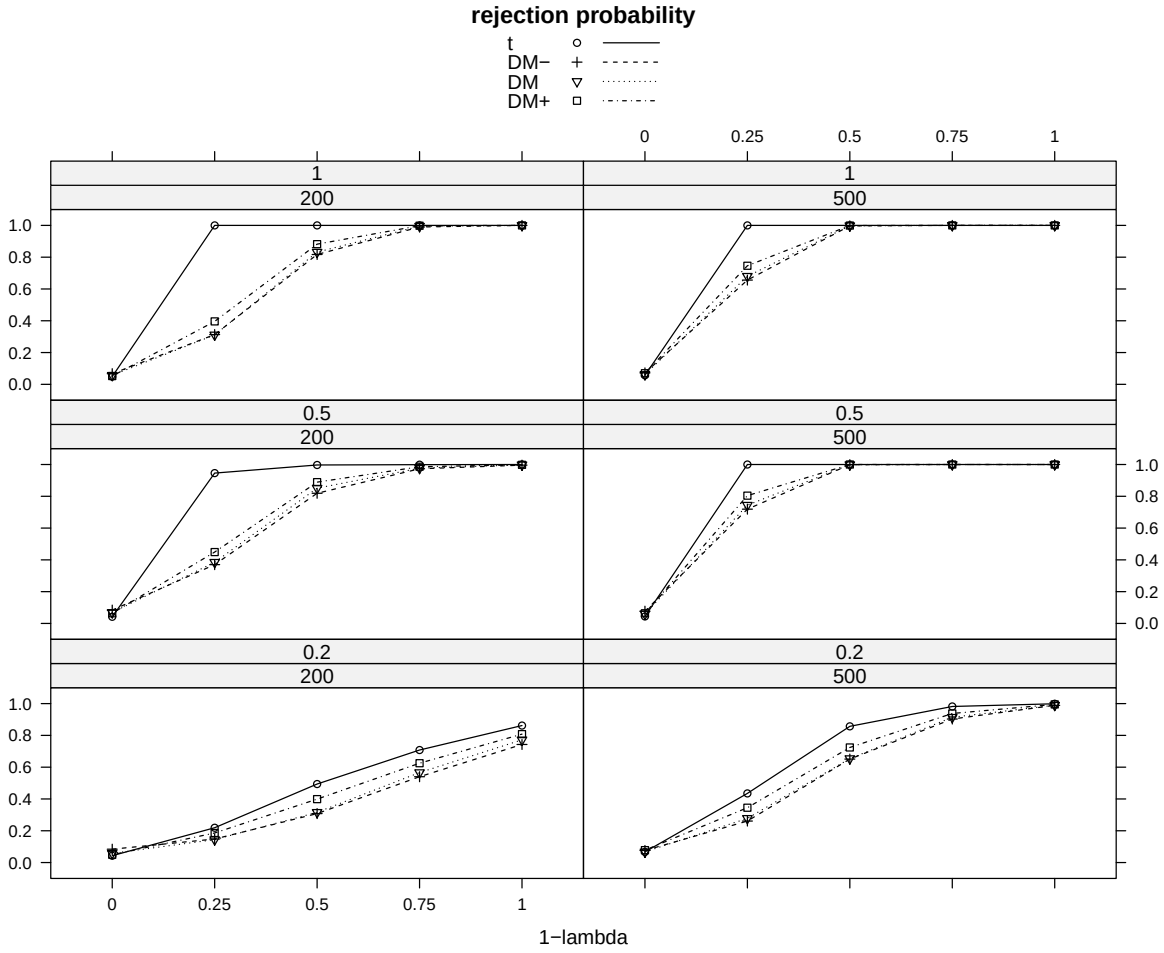

Figure 5: (Model I): rejection probabilities for $n = 200$ (left column) and $n = 500$ (right column) as well as $\sigma^2_{ME} = 1, 0.5, 0.2$ (top to bottom).

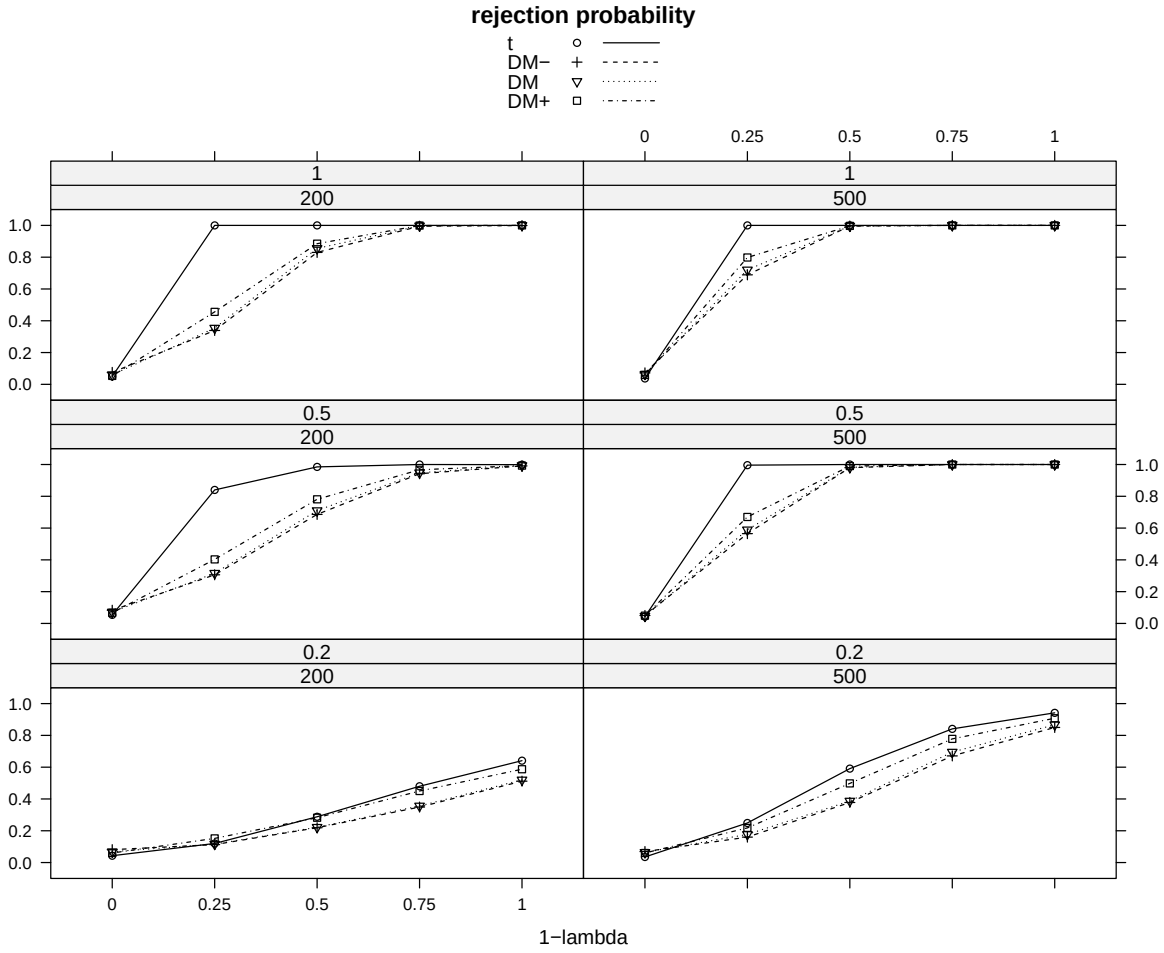

Figure 6: (Model II): rejection probabilities for $n = 200$ (left column) and $n = 500$ (right column) as well as $\sigma^2_{ME} = 1, 0.5, 0.2$ (top to bottom).

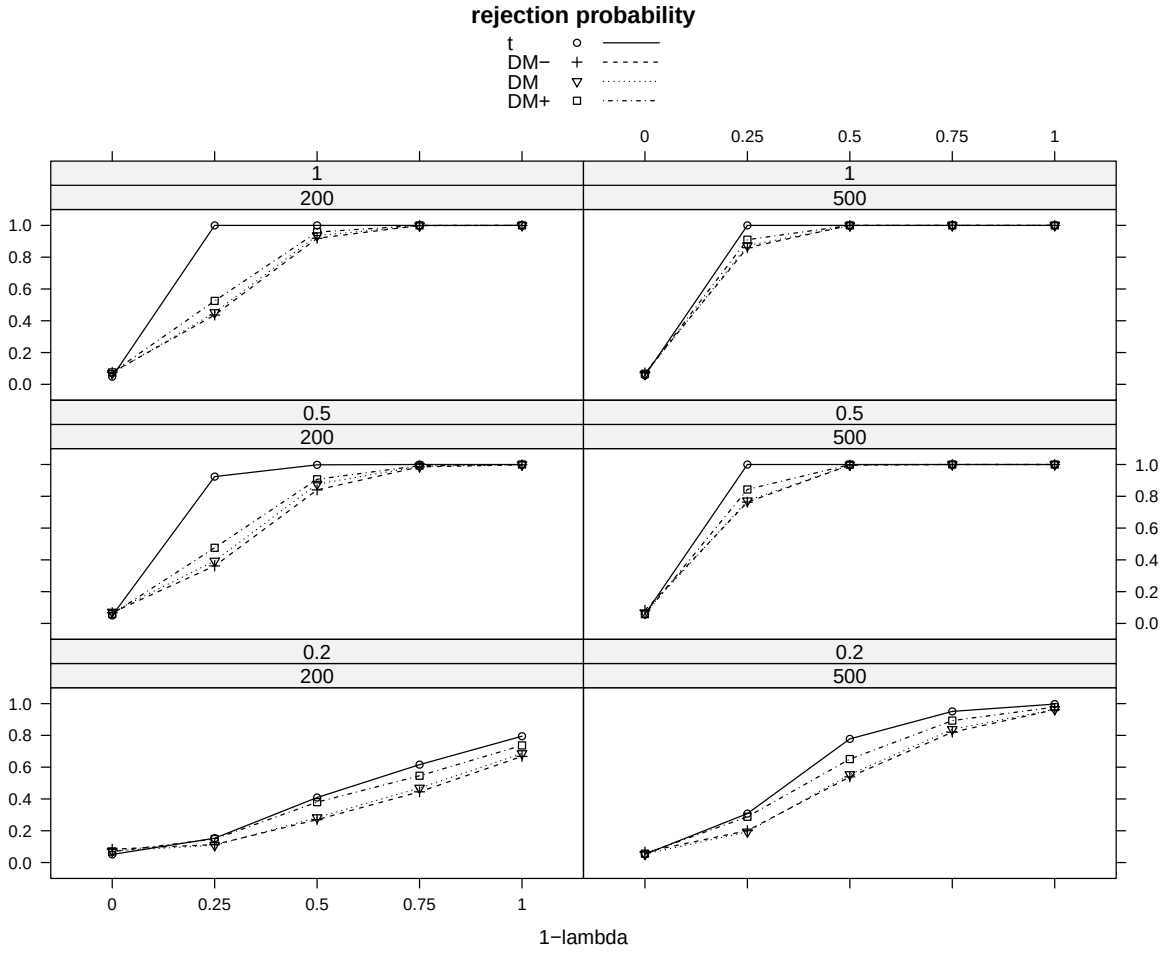

Figure 7: (Model III): rejection probabilities for $n = 200$ (left column) and $n = 500$ (right column) as well as $\sigma^2_{ME} = 1, 0.5, 0.2$ (top to bottom).

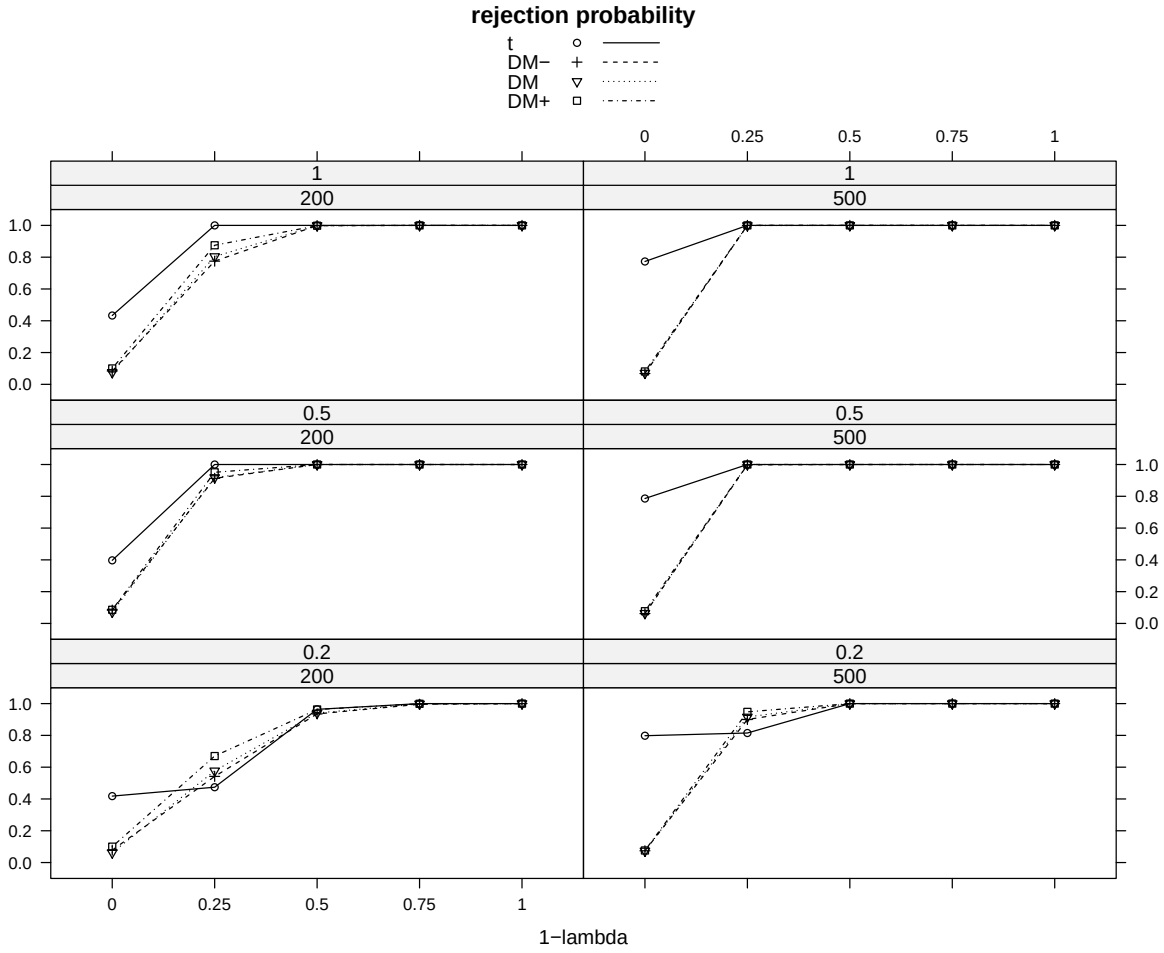

Figure 8: (Model IV): rejection probabilities for $n = 200$ (left column) and $n = 500$ (right column) as well as $\sigma^2_{ME} = 1, 0.5, 0.2$ (top to bottom).

	before		after	
	mean	stand. dev.	mean	stand. dev.
survey earnings 1977	4,195.6	7,143.8	6,567.7	4,508.4
– fraction topcoded	0.1	0.4	0	0
administrative earnings 1977	4,703.6	5,867.8	6,422.6	4,539.6
– fraction topcoded	0.6	0.5	0	0
administrative earnings 1976	4,308.5	5,443.7	5,595.2	4,486.0
white	0.8	0.4	0.9	0.3
age	32.5	21.9	35.4	14.8
education	9.7	6.0	13.1	2.7
married	0.4	0.5	0.6	0.5
sample size	168,904		31,378	

Table 4: Summary statistics before and after sample selection.

D More Details on the Empirical Application

If the survey measurement error η_{S2} depends not only on the level of true earnings E_2^*, but also on its growth, $E_2^* - E_1^*$, then (27) is violated. To see this suppose there is no measurement error in administrative earnings, i.e. $A_t = E_t^*$ for all t, and that η_{S2} depends on $E_2^* - E_1^*$ conditional on E_2^*. Then, η_{S2} depends on $A_2 - A_1$ conditional on A_2, which means S_2 depends on A_1 conditional on A_2. However, this is not possible under the exclusion restriction (27).

Under the exclusion restriction (29), which allows for the dependence of the survey measurement error on both the level and growth of earnings, we can test the modified null that there is no measurement error in both periods, i.e. $P(A_1 = E_1^* \text{ and } A_2 = E_2^*) = 1$ by testing the observable implication

$$S_2 \perp A_0 \mid (A_2, A_2 - A_1). \tag{42}$$

This restriction can be tested in a similar fashion as the other conditional independence conditions described in the main text.

E Proofs

We now introduce the probability space and some notation that is used in the proofs. We consider distributions P on the measurable space $(\Omega, \mathcal{B}(\Omega))$, where $\Omega \subseteq \mathbb{R}^d$, $d := 1 + 2d_x + d_z$, $d_x \geq 1$, $d_z \geq 1$, $\Omega = \Omega_y \times \Omega_x \times \Omega_x \times \Omega_z$, $\Omega_y \subseteq \mathbb{R}$, $\Omega_x \subseteq \mathbb{R}^{d_x}$, $\Omega_z \subseteq \mathbb{R}^{d_z}$, and $\mathcal{B}(\Omega)$ is the Borel-sigma algebra of subsets of Ω. We then define the random vector (Y, X^*, X, Z) to contain the coordinate mappings of $\omega \in \Omega$.

For a collection $\mathcal{G}$ of subsets of Ω, we denote by $\sigma(\mathcal{G})$ the sub-sigma algebra of $\mathcal{B}(\Omega)$ generated by $\mathcal{G}$. For random variables A and B, we denote by $\sigma(A)$ and $\sigma(A, B)$ the sub-sigma algebras of $\mathcal{B}(\Omega)$ generated by A and A, B, respectively. For a collection of sets $\mathcal{G}$ and another set H, we let $\mathcal{G} \cap H := \{G \cap H \colon G \in \mathcal{G}\}$. The abbreviation "a.s." always refers to "P-a.s.". When an equality holds almost surely with respect to a different measure, we will make this explicit. For a set $B \in \mathcal{B}(\Omega)$, we use the abbreviation $\mathbb{1}_B$ to denote the indicator function for the set B, so expressions like $E_P[\mathbb{1}_B]$, for example, make sense (this is equal to $P(B)$).

As in the main text let ρ_x and ρ_z each denote either Lebesgue on $\mathbb{R}$ or counting measure on a discrete subset of $\mathbb{R}$, but they do not need to be the same. For two elements (or vectors) A and B, we denote by $P_{A|B}$ the regular conditional distribution[23] of A given the sigma algebra generated by B. We use the notation $E_P(A|B)$ for conditional expectations under P of A given the sigma algebra generated by B. The event of no measurement error, $D := \{\omega \in \Omega \colon X(\omega) = X^*(\omega)\}$, subsequently plays a prominent role. Let D^c denote the complement of D. For two sets $G, H \in \mathcal{B}(\Omega)$, we sometimes use the notation

$$P^H(G) := \frac{P(G \cap H)}{P(H)}$$

if $P(H) > 0$. Since this defines just another distribution, conditional distributions of $A|B$ under P^H, denoted by $P^H_{A|B}$, are well-defined.

E.1 Proofs for Sections 2 and 3

Proof of Theorem 1. Let $P \in \mathbf{M}$ and a, b be any real-valued bounded functions such that the subsequent expectations exist. Then, under H_0, we have $P = P^D$ so that

$$\begin{aligned}
E_P[a(Y)b(Z) \mid X] &= E_{P^D}[a(Y)b(Z) \mid X] \\
&= E_{P^D}[a(Y)b(Z) \mid X^*] \\
&= E_P[a(Y)b(Z) \mid X^*] \\
&= E_P[a(Y)|X^*]\ E_P[b(Z)|X^*] \\
&= E_{P^D}[a(Y)|X^*]\ E_{P^D}[b(Z)|X^*] \\
&= E_{P^D}[a(Y)|X]\ E_{P^D}[b(Z)|X] \\
&= E_P[a(Y)|X]\ E_P[b(Z)|X]
\end{aligned}$$

where all equalities hold P-a.s..[24] The second and second-to-last equalities follow from Lemma 8. The claim is therefore established. Q.E.D.

Before proving Theorem 2 we present the following lemma which is a modification of arguments in Lehmann (1955).

[23] See chapter 10.2 of Dudley (2002) for a definition. Regular conditional distributions exist because all random vectors are defined on the same Borel space from Appendix E (Theorem 10.2.2 in Dudley (2002)).

[24] The second and second-to-last equalities hold P^D-a.s. according to Lemma 8, but since $P(D) = 1$, they also hold P-a.s..

Lemma 6. *Let X be a random variable, P_1,P_2 two probability measures, and $\phi : \mathbb{R} \to \mathbb{R}$ a monotone function. Assume there exist a set $\mathcal{X} \subseteq \mathbb{R}$ and a constant $C > 0$ such that the following conditions hold: (i) ϕ is strictly monotone over $\mathcal{X}$, and (ii) P_1, P_2 satisfy*

$$P_1(X \geq x) \leq P_2(X \geq x) \qquad \forall x \in \mathbb{R}$$

and

$$P_1(X \geq x) \leq P_2(X \geq x) - C \qquad \forall x \in \mathcal{X}.$$

Then

$$E_{P_1}\phi(X) \neq E_{P_2}\phi(X),$$

provided the expectations exist.

Proof. Without loss of generality assume ϕ is increasing and $\phi \geq 0$. Approximate it by the sequence of simple functions defined as $\phi_n(x) := \sum_{i=1}^{N} b_i \mathbb{1}\{x \in S_i^{(n)}\}$, where, for $N := n2^n + 1$,

$$b_i := \begin{cases} \frac{i-1}{2^n}, & i = 1, \ldots, N-1 \\ n, & i = N \end{cases}$$

and

$$S_i^{(n)} := \begin{cases} \left\{x \in \mathbb{R} \colon \frac{i-1}{2^n} \leq \phi(x) < \frac{i}{2^n}\right\}, & i = 1, \ldots, N-1 \\ \left\{x \in \mathbb{R} \colon \phi(x) > n\right\}, & i = N \end{cases}.$$

Similarly as in Lehmann (1955), we can then show that

$$E_{P_2}\phi_n(X) = \sum_{i=1}^{N} a_i P_2\left(\phi(X) \geq \frac{i-1}{2^n}\right)$$

where

$$a_i := \left\{ \begin{array}{cc} b_i - b_{i-1}, & i = 2, \ldots, N \\ 0, & i = 1 \end{array} \right\} = \left\{ \begin{array}{cc} \frac{1}{2^n}, & i = 2, \ldots, N \\ 0, & i = 1 \end{array} \right\}.$$

Let $I_n := \{i \in \{1, \ldots, N\} \colon (i-1)/2^n \in \phi(\mathcal{X})\}$ where $\phi(\mathcal{X})$ is the image of $\mathcal{X}$ under ϕ. Notice that the number of elements of I_n grows at the rate 2^n as $n \to \infty$, so there is a positive constant $\bar{C}$ such that $|I_n| \geq \bar{C}2^n$. Therefore:

$$\begin{aligned} E_{P_2}\phi_n(X) &= \sum_{i=1}^{N} a_i P_2\left(\phi(X) \geq \frac{i-1}{2^n}\right) \\ &= \sum_{i \in I_n} a_i P_2\left(\phi(X) \geq \frac{i-1}{2^n}\right) + \sum_{i \notin I_n} a_i P_2\left(\phi(X) \geq \frac{i-1}{2^n}\right) \\ &\geq \sum_{i \in I_n} a_i \left[P_1\left(X \geq \phi^{-1}\left(\frac{i-1}{2^n}\right)\right) + C\right] + \sum_{i \notin I_n} a_i P_1\left(\phi(X) \geq \frac{i-1}{2^n}\right) \\ &= \sum_{i=1}^{N} a_i P_1\left(\phi(X) \geq \frac{i-1}{2^n}\right) + C\frac{|I_n|}{2^n} \\ &\geq E_{P_1}\phi_n(X) + C\bar{C}. \end{aligned}$$

where ϕ^{-1} denotes the inverse of ϕ on $\mathcal{X}$. Since $\{\phi_n\}_{n\geq 1}$ is an increasing sequence, monotone convergence then implies the desired result. Q.E.D.

Proof of Theorem 2. Sufficiency follows from Theorem 1 because $\mathbf{M}_R \subset \mathbf{M}$. To show necessity assume that $P \in \mathbf{M}_R$ satisfies (8), but not (7). (8) implies that $E_P[\mu(Y)|X,Z] = E_P[\mu(Y)|X]$ a.s.. On the other hand, by Lemma 6, for any $(x, z_1, x_2) \in \mathcal{X} \times \mathcal{Z}_1 \times \mathcal{Z}_2$, we have

$$\begin{aligned} E_P[\mu(Y)|X=x, Z=z_2] &= E_P[\phi(X^*)|X=x, Z=z_2] \\ &\neq E_P[\phi(X^*)|X=x, Z=z_1] = E_P[\mu(Y)|X=x, Z=z_1], \end{aligned} \tag{43}$$

where $\phi(x^*) := E_P[\mu(Y)|X^* = x^*]$. The two equalities follow from Assumption 2. Since $P(\mathcal{X} \times \mathcal{Z}_k) > 0$, $k = 1, 2$, we have found a contradiction, thus establishing necessity. Q.E.D.

Proof of Theorem 3. Consider P that satisfies the assumptions of the theorem with Assumption 8 and $\lambda := P(X = X^*) < 1$. Notice that

$$\begin{aligned} P(X \in A|X^* \in B, Z \in C) &= P(X \in A|X^* \in B, Z \in C, D)P(D|X^* \in B, Z \in C) \\ &\quad + P(X \in A|X^* \in B, Z \in C, D^c)P(D^c|X^* \in B, Z \in C) \\ &= P^D(X^* \in A \cap B|Z \in C)P(D|X^* \in B, Z \in C) \\ &\quad + Q(X \in A|X^* \in B, Z \in C)(1 - P(D|X^* \in B, Z \in C)) \end{aligned}$$

for any sets $A \in \mathcal{B}(\Omega_x)$, $C \in \mathcal{B}(\Omega_z)$, for which the denominator is nonzero. By Assumption 7, there is a function $f : \mathcal{S}_{X^*} \to [0, 1]$ such that $P(D|X^* = x^*, Z = z) = f(x^*)$. Then by the usual limiting argument choosing a suitable sequence of conditioning sets and Assumption 6,

$$p_{X|X^*,Z}(x|x^*, z) = f(x^*)\delta(x - x^*) + (1 - f(x^*))q_{X|X^*}(x|x^*)$$

where δ denotes a pointmass at zero. For any $a > 0$, define $d_{a,x}(x^*) := \mathbb{1}\{x^* = x\}$ if the dominating measure ρ_x in Assumption 5 is counting measure and $d_{a,x}(x^*) := a\phi(a(x^* - x))$ if ρ_x is Lebesgue measure, where $\phi(\cdot)$ denotes the standard normal pdf. Let $P_a(x^*|x, z) := \int_{x^*}^{\infty} p_a(u|x, z)d\rho_x(u)$ with

$$p_a(x^*|x, z) := \frac{p_a(x|x^*)p_{X^*|Z}(x^*|z)}{p_{X|Z}(x|z)}$$

and

$$p_a(x|x^*) := f(x^*)d_{a,x}(x^*) + (1 - f(x^*))q_{X|X^*}(x|x^*).$$

Then

$$\lim_{a\to 0} P_a(x^*|x, z) = P(X^* \geq x^*|X = x, Z = z) \tag{44}$$

holds trivially when ρ_x is counting measure, but also in the case of Lebesgue measure because $\{d_{a,x}(\cdot), a \geq 0\}$ is a delta sequence (e.g. Kanwal (1983)) and because $p_{X^*|Z}(x^*|z)$ is bounded (implied by Assumption 5(i)).

Fix any $z = (z_1, z_2)' \in \mathcal{Z}_1 \times \mathcal{Z}_2$, $x \in \mathcal{X}$, and $a > 0$. Let $\bar{c} := p_{X|Z}(x|z_2)/p_{X|Z}(x|z_1)$ and $\Delta p_a(x^*) := p_a(x^*|x, z_2) - p_a(x^*|x, z_1)$. Now, suppose that, for some $x_1^* \in \mathcal{S}_{X^*}$,

$$\Delta p_a(x_1^*) = \frac{p_a(x|x_1^*)}{p_{X|Z}(x|z_2)} \left[p_{X^*|Z}(x_1^*|z_2) - \bar{c} p_{X^*|Z}(x_1^*|z_1)\right] = \frac{p_a(x|x_1^*)}{p_{X|Z}(x|z_2)} \Delta p_{\bar{c},z}(x_1^*) > 0.$$

Therefore, $\Delta p_{\bar{c},z}(x_1^*) > 0$ which, by Assumption 8(i), implies $\Delta p_{\bar{c},z}(x_2^*) \geq 0$ and thus $\Delta p_a(x_2^*) \geq 0$ for any $x_2^* > x_1^*$. This means Δp_a is also single-crossing. Define $x_0 := \inf\{x^* \colon \Delta p_a(x^*) > 0\}$. Then, it is clear that $P_a(x^*|x, z_2) \geq P_a(x^*|x, z_1)$ for all $x^* \geq x_0$. Suppose there exists a value $y < x_0$ such that $P_a(y|x, z_2) < P_a(y|x, z_1)$. Then

$$0 > P_a(y|x, z_2) - P_a(y|x, z_1) = \int_{-\infty}^{y} (p_a(u|x, z_1) - p_a(u|x, z_2)) d\rho_x(u)$$

which is a contradiction to $p_a(u|x, z_1) - p_a(u|x, z_2) \geq 0$ for all $u < x_0$. Therefore, we have shown that

$$P_a(x^*|x, z_2) \geq P_a(x^*|x, z_1) \qquad \forall x^* \in \mathbb{R}. \tag{45}$$

We now show that this inequality is strict for some x^*. To this end, notice that by parts (i) and (ii) of Assumption 5,

$$p_{X|Z}(x|z) = \lambda p_{X|Z}^D(x|z) + (1-\lambda) q_{X|Z}(x|z) = \lambda p_{X^*|Z}^D(x|z) + (1-\lambda) q_{X|Z}(x|z) \leq C_q.$$

Also, by the bounds in Assumption 8, $p_a(x|x^*) \geq (1 - f(x^*)) q_{X|X^*}(x|x^*) \geq c_q^2$ for all $x^* \in \mathcal{X}^*$. Thus, part (ii) of Assumption 8 implies $\Delta p_a(x^*) \geq c_q^3/C_q > 0$ for all $x^* \in \mathcal{X}^*$. If ρ_x is Lebesgue measure, partition the set $\mathcal{X}^*$ into two sets $\mathcal{X}_1^*$ and $\mathcal{X}_2^*$ (i.e. $\mathcal{X}_1^* \cap \mathcal{X}_2^* = \emptyset$ and $\mathcal{X}_1^* \cup \mathcal{X}_2^* = \mathcal{X}^*$) such that $\rho_x(\mathcal{X}_k^*) > 0$, $k = 1, 2$, and $x_1^* < x_2^*$ for all $x_1^* \in \mathcal{X}_1^*$, $x_2^* \in \mathcal{X}_2^*$. If ρ_x is counting measure, then simply let $\mathcal{X}_1^* = \mathcal{X}_2^* = \mathcal{X}^*$. Then, for all $x^* \in \mathcal{X}_1^*$,

$$\begin{aligned} P_a(x^*|x, z_2) - P_a(x^*|x, z_1) &= \int_{x^*}^{\infty} (p_a(u|x, z_2) - p_a(u|x, z_1)) d\rho_x(u) \\ &\geq \int_{\mathcal{X}_2^*} (p_a(u|x, z_2) - p_a(u|x, z_1)) d\rho_x(u) \\ &\geq \frac{c_q^3}{C_q} \rho_x(\mathcal{X}_2^*) =: \widetilde{C} > 0 \end{aligned} \tag{46}$$

Since $\widetilde{C}, \mathcal{X}_1^*, \mathcal{X}_2^*$ are independent of a, x, z_1, z_2, the convergence in (44) together with (45)–(46) implies

$$P(X^* \geq x^*|X = x, Z = z_2) \geq P(X^* \geq x^*|X = x, Z = z_1) \qquad \forall x^* \in \mathbb{R}$$

$$P(X^* \geq x^*|X = x, Z = z_2) \geq P(X^* \geq x^*|X = x, Z = z_1) + \widetilde{C} \qquad \forall x^* \in \mathcal{X}_1^*.$$

which are the conditions in Assumption 4. Finally, consider the case in which Assumption 8′ holds. Then, by Assumption 8′(i) and letting $\bar{c}' := p_{Z|X}(z_2|x)/p_{Z|X}(z_1|x)$,

$$\Delta p_a(x_1^*) = \frac{p_a(x|x_1^*) p_{X^*}(x_1^*)}{p_{Z|X}(z_2|x) p_X(x)} \Delta p'_{\bar{c},z}(x_1^*) > 0.$$

again implies $\Delta p_a(x_2^*)$ for any $x_2^* > x_1^*$, so the single-crossing property of Δp_a continues to hold and (45) follows. The bound $p_{Z|X}(z|x)p_X(x) = p_{X|Z}(x|z)p_Z(z) \leq C_q^2$ follows from Assumption 5 and $p_a(x|x^*)p_{X^*}(x^*) \geq (1 - f(x^*))q_{X|X^*}(x|x^*)(1-\lambda)q_{X^*}(x^*) \geq (1-\lambda)c_q^2$ so that, by Assumption 8′(ii), we have $\Delta p_a(x^*) \geq (1-\lambda)c_q^3 C_q^{-2} > 0$ for all $x^* \in \mathcal{X}^*$. The argument then procedes similarly as above to show (46), thus concluding the proof. Q.E.D.

Proof of Lemma 1. First notice, that by Theorem 1 of Imbens and Newey (2009),

$$\bar{\varepsilon} \perp X \mid \bar{\eta}. \tag{47}$$

Second, let $\bar{g}^{-1}(x,\cdot)$, $\bar{m}_\eta^{-1}(z,\cdot)$ and $\bar{m}_z^{-1}(\cdot,\bar{\eta})$ denote the inverse functions of $\bar{g}(x,\cdot)$, $\bar{m}(z,\cdot)$, and $\bar{m}(\cdot,\bar{\eta})$, respectively. Then:

$$\begin{aligned}
P(Y \leq y \mid X = x, Z = z) &= P(\bar{\varepsilon} \leq \bar{g}^{-1}(x,y) \mid X = x, Z = z) \\
&= P(\bar{\varepsilon} \leq \bar{g}^{-1}(x,y) \mid X = x, \bar{m}_z^{-1}(X,\bar{\eta}) = z) \\
&= P(\bar{\varepsilon} \leq \bar{g}^{-1}(x,y) \mid X = x, \bar{m}_z^{-1}(x,\bar{\eta}) = z) \\
&= P(\bar{\varepsilon} \leq \bar{g}^{-1}(x,y) \mid X = x, x = \bar{m}(z,\bar{\eta})) \\
&= P(\bar{\varepsilon} \leq \bar{g}^{-1}(x,y) \mid X = x, \bar{\eta} = \bar{m}_\eta^{-1}(z,x))
\end{aligned}$$

where all equalities follow from the strict monotonicity conditions in Assumption 9, guaranteeing that all relevant conditioning sigma algebras are equal. Similarly,

$$P(Y \leq y \mid X = x) = P(\bar{\varepsilon} \leq \bar{g}^{-1}(x,y) \mid X = x)$$

so that

$$Y \perp Z \mid X \qquad \Leftrightarrow \qquad \bar{\varepsilon} \perp \bar{\eta} \mid X.$$

Now, the "⇒" direction of the claim follows because $\bar{\varepsilon} \perp \bar{\eta}$ and $Z \perp (\bar{\varepsilon}, \bar{\eta})$ imply $\bar{\varepsilon} \perp (X, \bar{\eta})$ which in turn implies $\bar{\varepsilon} \perp \bar{\eta} \mid X$. The "⇐" direction follows because $\bar{\varepsilon} \perp \bar{\eta} \mid X$ and (47) imply $\bar{\varepsilon} \perp \bar{\eta}$. Q.E.D.

Proof of Theorem 4. Sufficiency follows immediately because, by the exclusion restriction, the null implies

$$\begin{aligned}
E_P[\Lambda_P(Y,X)|X,Z] &= E_P\left[E_P\left(\Lambda_P(Y,X)|X^*,X,Z\right) \mid X,Z\right] \\
&= E_P\left[E_P\left(\Lambda_P(Y,X)|X^*,X\right) \mid X,Z\right] \\
&= 0
\end{aligned}$$

Necessity is proven similarly as in Theorem 2. Assume that $P \in \mathbf{M}_R$ satisfies (8), but not (7). For any $x \in \mathcal{X}$, let $\phi_x(x^*) := E_P[\Lambda_P(Y,X)|X^* = x^*, X = x]$, then by Lemma 6, for any $(x, z_1, z_2) \in \mathcal{X} \times \mathcal{Z}_1 \times \mathcal{Z}_2$, we have

$$\begin{aligned}
E_P[\Lambda_P(Y,X)|X = x, Z = z_2] &= E_P[\phi_x(X^*)|X = x, Z = z_2] \\
&\neq E_P[\phi_x(X^*)|X = x, Z = z_1] = E_P[\Lambda_P(Y,X)|X = x, Z = z_1].
\end{aligned} \tag{48}$$

The two equalities follow from Assumption 2. Since $P(\mathcal{X} \times \mathcal{Z}_k) > 0$, $k = 1, 2$, we have found a contradiction, thus establishing necessity. Q.E.D.

E.2 Proofs for Appendix B

Proof of Lemma 3. By Theorem 3 it suffices to show that Assumptions 5–7 and 8′ are satisfied. Part (ii) of Assumption 5 and Assumption 7 trivially hold because η_X has a continuous distribution under P and because of (34), so that $P(X = X^*) = 0$. An inspection of the proof of Theorem 3 reveals that the assumption of boundedness of the densities in Assumption 5 is not needed when $P(X = X^*) = 0$, but only that $p_Z(z)$ is bounded over $\mathcal{Z}_1 \cup \mathcal{Z}_2$. This is the case for continuous distributions supported over intervals as long as we choose $\mathcal{Z}_k$ to be bounded subsets of the support of Z, bounded away from the boundaries of the support (which we will further below). Assumption 6 holds because of Assumption RM4, so we only need to show that the single-crossing condition, Assumption 8′, holds.

First, consider the case in which the support of η_Z is equal to $\mathbb{R}$. Notice that the lower bounds on the densities are satisfied for any choice of bounded intervals $\mathcal{X}^*, \mathcal{X}, \mathcal{Z}_1, \mathcal{Z}_2$ of positive length. By the monotonicity condition in Assumption RM3, $x^* \mapsto p_{Z|X^*}(\bar{z}_2|x^*) = p_{\eta_Z}(\bar{z}_2 - x^*)$ and $x^* \mapsto p_{Z|X^*}(\bar{z}_1|x^*) = p_{\eta_Z}(\bar{z}_1 - x^*)$ intersect only once as long as $\bar{z}_1$ and $\bar{z}_2$ are far enough apart. In particular, let $c_0 \geq c$ be such that $p_{\eta_Z}(\eta) < \min_{u\in(-c,c)} p_{\eta_Z}(u)$ for all $\eta \geq c_0$. Then, they intersect only once if $\bar{z}_2 - \bar{z}_1 > 2c_0$.

By Assumptions RM3 and RM4, for any value $x \in \mathcal{S}_X$, the function $z \mapsto p_{Z,X}(z, x) = \int p_{\eta_Z}(z - x^*)p_{\eta_X}(x - x^*)p_{X^*}(x^*)dx^*$ is continuous, supported on whole $\mathbb{R}$, and must have tails that go to zero as $|z| \to \infty$. As a consequence, for any $\bar{x} \in \mathcal{S}_X$, there exist values $\bar{z}_1, \bar{z}_2 \in \mathbb{R}$, $\bar{z}_2 - \bar{z}_1 > 2c_0$, such that $p_{Z,X}(\bar{z}_2, \bar{x}) = p_{Z,X}(\bar{z}_1, \bar{x})$ and thus also $p_{Z|X}(\bar{z}_2|\bar{x}) = p_{Z|X}(\bar{z}_1|\bar{x})$. By continuity of the densities, Assumption 8′ then holds with $\mathcal{X} = N_\delta(\bar{x})$, $\mathcal{Z}_k = N_\delta(\bar{z}_k)$, $k = 1, 2$, and $\mathcal{X}^* = N_\delta(\bar{x}^*)$, where $\bar{x}^* \in \mathbb{R}$ is large enough and $\delta > 0$ is chosen suitably small.

Now, consider the case in which the support of η_Z is bounded, say equal to $(\underline{\eta}_Z, \overline{\eta}_Z)$. For some $\delta > 0$ and some value x_0^* in the interior of the support of X^*, let $\bar{z}_2 = x_0^* + \overline{\eta}_Z + \delta$ and $\bar{z}_1 = x_0^* + \underline{\eta}_Z - \delta$. Then $p_{\eta_Z}(\bar{z}_1 - x^*) > 0$ if $x^* < x_0^* - \delta$, zero otherwise, and $p_{\eta_Z}(\bar{z}_2 - x^*) > 0$ if $x^* > x_0^* + \delta$, zero otherwise. Therefore,

$$p_{\eta_Z}(\bar{z}_2 - x^*) - p_{\eta_Z}(\bar{z}_1 - x^*) \begin{cases} < 0, & x^* < x_0^* - \delta \\ = 0, & x_0^* - \delta \leq x^* \leq x_0^* + \delta \\ > 0, & x^* > x_0^* + \delta \end{cases}$$

and thus also, for any $C > 0$,

$$p_{Z|X^*}(\bar{z}_2|x^*) - Cp_{Z|X^*}(\bar{z}_1|x^*) = p_{\eta_Z}(\bar{z}_2 - x^*) - Cp_{\eta_Z}(\bar{z}_1 - x^*) \begin{cases} < 0, & x^* < x_0^* \\ \geq 0, & x^* \geq x_0^* \end{cases}. \tag{49}$$

Varying $\delta > 0$ in the above argument shows that (49) holds also for $\bar{z}_k$ replaced by any value z_k in a small enough neighborhood $\mathcal{Z}_k = N_{\delta_z}(\bar{z}_k)$, $\delta_z > 0$. Similarly, we can find a neighborhood $\mathcal{X}^* \subset \mathcal{S}_{X^*}$ of values that are larger than $x_0^* + \delta$, $\delta > 0$, so that there is a constant $c_q > 0$ with $p_{\eta_Z}(z_2 - x^*) \geq c_q$ for all $z_2 \in \mathcal{Z}_2$ and $x^* \in \mathcal{X}^*$. Therefore, both parts (i) and (ii) of Assumption 8′ hold. It remains to show that the lower bounds on the densities are satisfied. Let $\bar{x}$ be in the interior of the support of X such that $p_{X|X^*}(\bar{x}|x^*) = p_{\eta_X}(\bar{x} - x^*) > 0$ and $p_{X^*}(x^*) > 0$ for all

$x^* \in \mathcal{X}^*$. Then by continuity of the densities, we can also choose the neighborhoods $\mathcal{X}, \mathcal{X}^*$ small enough so there is a constant $c_q > 0$ with $q_{X|X^*}(\bar{x}|x^*)q_{X^*}(x^*) = p_{X|X^*}(\bar{x}|x^*)p_{X^*}(x^*) \geq c_q$ for all $x^* \in \mathcal{X}^*$ and $x \in \mathcal{X}$.

A straightforward modification of the argument yields the conclusion in the case that one of $\underline{\eta}_Z, \overline{\eta}_Z$ is not finite. Q.E.D.

Proof of Lemma 4. By the same argument as in the proof of Lemma 3, $p_{U_Z}(z_2 - v) - cp_{U_Z}(z_1 - v)$ as a function of v is single-crossing and there is a constant $c_q > 0$ such that $p_{U_Z}(z_2 - v) - cp_{U_Z}(z_1 - v) \geq c_q$ for all v in some neighborhood. These statements hold for all z_k in some neighborhood $\mathcal{Z}_k$ and all c in $\mathcal{C}$, where the definition of $\mathcal{C}$ relies on a suitably chosen $\mathcal{X}$. All subsequent statements are understood to hold for z_k, v, x, and c in such sets and we omit those qualifications. Now, notice that

$$\begin{aligned}
&p_{X,Z|X^*}(x, z_2|x^*) - cp_{X,Z|X^*}(x, z_1|x^*) \\
&\quad = \int \left[p_{X,Z|X^*,\alpha}(x, z_2|x^*, \alpha) - cp_{X,Z|X^*,\alpha}(x, z_1|x^*, \alpha)\right] p_{\alpha|X^*}(\alpha|x^*)d\alpha \\
&\quad = \int p_{U_X}(x - x^* - \alpha)\left[p_{U_Z}(z_2 - x^* - \alpha) - cp_{U_Z}(z_1 - x^* - \alpha)\right] p_{\alpha|X^*}(\alpha|x^*)d\alpha \\
&\quad = \int p_{U_X}(x - v)\left[p_{U_Z}(z_2 - v) - cp_{U_Z}(z_1 - v)\right] p_{\alpha|X^*}(v - x^*|x^*)dv
\end{aligned}$$

By Assumption RM7 and single-crossing of $v \mapsto p_{U_X}(x-v)\left[p_{U_Z}(z_2 - v) - cp_{U_Z}(z_1 - v)\right]$, Lemma A1 in Athey (2002) implies that $x^* \mapsto p_{X,Z|X^*}(x, z_2|x^*) - cp_{X,Z|X^*}(x, z_1|x^*)$ is also single-crossing and there is a constant $c_q' > 0$ such that $p_{X,Z|X^*}(x, z_2|x^*) - cp_{X,Z|X^*}(x, z_1|x^*) \geq c_q'$ for x^* in some neighborhood.

Similarly as in the proof of Lemma 3 we then show that

$$\begin{aligned}
&p_{X^*|X,Z}(x^*|x, z_2) - p_{X^*|X,Z}(x^*|x, z_1) \\
&\qquad = \frac{p_{X^*}(x^*)}{p_{X,Z}(x, z_2)}\left[p_{X,Z|X^*}(x, z_2|x^*) - \frac{p_{X,Z}(x, z_2)}{p_{X,Z}(x, z_1)}p_{X,Z|X^*}(x, z_2|x^*)\right]
\end{aligned}$$

as a function of x^* is also single-crossing. The proof then follows that of Theorem 3. Q.E.D.

Proof of Lemma 5. Set $\mathcal{Z}_1 = \{0\}$, $\mathcal{Z}_2 = \{1\}$. Assumption IV2(i) implies $\varepsilon \perp (Z, \eta_X, X^*)$, so that $\varepsilon \perp (Z, \eta_X) \mid X^*$, which then implies Assumption 2. By Assumptions IV2(i) and IV3, $x^* \mapsto E_P[Y|X^* = x^*] = E_P[g(x^*, \varepsilon)]$ is increasing over $\mathcal{S}_{X^*}$ and strictly increasing over $\mathcal{X}^*$ (Lemma 9), so that Assumption 3 holds. Assumption IV1 directly implies Assumption 5. Assumption IV4 requires Assumptions 8 and 7 to hold. Assumption IV2(ii) implies that $\eta_X \perp (Z, X^*)$ and thus $\eta_X \perp Z \mid X^*$, so that Assumption 6 holds. Theorem 3 then implies the desired claim. Q.E.D.

F Auxiliary Results

In this section, we provide auxiliary results that are used in the proofs of the main results. Some of the proofs are fairly straightforward, but for the lack of a good reference and for completeness, we provide statements and proofs here.

Lemma 7. *Let $\mathcal{A}$ be a collection of subsets of Ω and $H \in \mathcal{B}(\Omega)$ some set. Then:*

$$\sigma(\mathcal{A} \cap H) = \sigma(\mathcal{A}) \cap H.$$

Proof. First, consider the direction "$\subseteq$". It is easy to see that $\sigma(\mathcal{A}) \cap H$ is also a sigma algebra.[25] Furthermore, it contains $\mathcal{A} \cap H$ and thus must also contain $\sigma(\mathcal{A} \cap H)$.

Now, consider the direction "$\supseteq$". Let $\mathcal{C} := \{C \subseteq \Omega \colon C \cap H \in \sigma(\mathcal{A}) \cap H\}$. Clearly, $\mathcal{A} \subseteq \mathcal{C}$ and $\mathcal{C}$ is a sigma algebra by a similar argument as in Footnote 25. Therefore, $\sigma(\mathcal{A})$ must also be contained in $\mathcal{C}$, which means for all $A \in \sigma(\mathcal{A})$, $A \cap H \in \sigma(\mathcal{A} \cap H)$ and thus $\sigma(\mathcal{A}) \cap H \subseteq \sigma(\mathcal{A} \cap H)$. Q.E.D.

Lemma 8. *For any probability measure P on $(\Omega, \mathcal{B}(\Omega))$ and any integrable f, we have*

$$E_{P^D}[f|\sigma(X, X^*)] = E_{P^D}[f|\sigma(X)] = E_{P^D}[f|\sigma(X^*)] \qquad P^D\text{–a.s.}.$$

Proof. For the first equality, we want to show that $E_{P^D}[f|\sigma(X)]$ is a version of the conditional expectation $E_{P^D}[f|\sigma(X^*, X)]$. This means, we need to establish that $E_{P^D}[f|\sigma(X)]$ satisfies the following two conditions:

(i) $E_{P^D}[f|\sigma(X)]$ is $\sigma(X^*, X)$-measurable

(ii) for all $G \in \sigma(X^*, X)$:
$$\int_G E_{P^D}[f|\sigma(X)]dP^D = \int_G f dP^D.$$

(i) trivially holds because by definition $E_{P^D}[f|\sigma(X)]$ is $\sigma(X)$-measurable and thus also measurable with respect to the larger sigma algebra $\sigma(X^*, X)$. Now, we want to show (ii). Let $\mathcal{G}_X$ be the collection containing all sets of the form $\{\omega \in \Omega \colon X_j(\omega) \leq x\}$ for $x \in \mathbb{R}$ and $j = 1, \ldots, d_x$. Similarly, let $\mathcal{G}_{XX^*}$ be the collection containing the sets in $\mathcal{G}_X$ and all those of the form $\{\omega \in \Omega \colon X_j^*(\omega) \leq x\}$ for $x \in \mathbb{R}$ and $j = 1, \ldots, d_x$. Then $\sigma(\mathcal{G}_X) = \sigma(X)$ and $\sigma(\mathcal{G}_{XX^*}) = \sigma(X^*, X)$ and $\mathcal{G}_X \cap D = \mathcal{G}_{XX^*} \cap D$. Therefore,

$$\sigma(X^*, X) \cap D = \sigma(\mathcal{G}_{XX^*}) \cap D = \sigma(\mathcal{G}_{XX^*} \cap D) = \sigma(\mathcal{G}_X \cap D) = \sigma(\mathcal{G}_X) \cap D = \sigma(X) \cap D$$

where the second and fourth equalities follow from Lemma 7. By definition $E_{P^D}[f|\sigma(X)]$ satisfies

$$\int_G E_{P^D}[f|\sigma(X)]dP^D = \int_G f dP^D \qquad \forall G \in \sigma(X).$$

Therefore, we also have

$$\begin{aligned}
& \int_G E_{P^D}[f|\sigma(X)]dP = \int_G f dP \qquad \forall G \in \sigma(X) \cap D \\
\Rightarrow \quad & \int_G E_{P^D}[f|\sigma(X)]dP = \int_G f dP \qquad \forall G \in \sigma(X^*, X) \cap D \\
\Rightarrow \quad & \int_G E_{P^D}[f|\sigma(X)]dP^D = \int_G f dP^D \qquad \forall G \in \sigma(X^*, X)
\end{aligned}$$

[25] obviously, (i) $H \in \sigma(\mathcal{A}) \cap H$ and (ii) $A \in \sigma(\mathcal{A}) \cap H$ implies $H \setminus A \in \sigma(\mathcal{A}) \cap H$. (iii) $A_1, A_2, \in \sigma(\mathcal{A}) \cap H$ also implies $\cup_{i \geq 1} A_i \in \sigma(\mathcal{A}) \cap H$ as follows: if $\omega \in \cup_{i \geq 1} A_i$, then there must be some j such that $\omega \in A_j \in \sigma(\mathcal{A}) \cap H$.

This establishes (ii) and thus $E_{P^D}[f|\sigma(X^*, X)] = E_{P^D}[f|\sigma(X)]$ P^D–a.s. as desired. Given the above argument, the proof for the second equality is obvious. Q.E.D.

Lemma 9. *Let $(A, \mathcal{A}, \mu)$ be a measure space with $\mu(A) > 0$ and f a measurable, positive function. Then $\int_A f d\mu > 0$.*

Proof. First of all, $\int_A f d\mu$ exists because f is positive. Suppose to the contrary that $\int_A f d\mu = 0$. Let $A_n := \{a \in A \colon f(a) \geq 1/n\}$ so that A is equal to the union of all A_n, $n \in \mathbb{N}$. Then

$$0 = \int_A f d\mu \geq \int_{A_n} f d\mu \geq \mu(A_n)/n.$$

By the continuity of measure, we then have

$$\mu(A) = \mu\left(\cup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n) = 0,$$

which is a contradiction, so the claim follows. Q.E.D.

References

Abowd, J. M., and D. Card (1987): "Intertemporal Labor Supply and Long-Term Employment Contracts," *The American Economic Review*, 77(1), 50–68.

Abowd, J. M., and M. H. Stinson (2007): "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data," Discussion paper.

Altonji, J. G. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *The Journal of Political Economy*, 94(3), S176–S215.

An, Y., and Y. Hu (2012): "Well-posedness of measurement error models for self-reported data," *Journal of Econometrics*, 168(2), 259–269.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," *Econometrica*, 85(3), 693–734.

Ashenfelter, O. (1984): "Macroeconomic analyses and microeconomic analyses of labor supply," *Carnegie-Rochester Conference Series on Public Policy*, 21, 117 – 156.

Athey, S. (2002): "Monotone Comparative Statics under Uncertainty," *The Quarterly Journal of Economics*, 117(1), 187–223.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2015): "Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia," Working Paper 20965, NBER.

Attanasio, O., C. Meghir, and E. Nix (2017): "Human Capital Development and Parental Investment in India," Discussion paper.

Bahadur, R. R., and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *The Annals of Mathematical Statistics*, 27(4), 1115–1122.

Barrett, G. F., and S. G. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71(1), 71–104.

Blomquist, S., A. Kumar, C.-Y. Liang, and W. K. Newey (2015): "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income," Discussion paper.

Blomquist, S., and W. Newey (2002): "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica*, 70(6), 2455–2480.

Blundell, R., J. Horowitz, and M. Parey (2016): "Nonparametric Estimation of a Nonseparable Demand Function under the Slutsky Inequality Restriction," *Review of Economics and Statistics*, forthcoming.

Blundell, R., and J. L. Horowitz (2007): "A Non-Parametric Test of Exogeneity," *Review of Economic Studies*, 74, 1035–1058.

Blundell, R., J. L. Horowitz, and M. Parey (2012): "Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation," *Quantitative Economics*, 3(1), 29–51.

Bollinger, C. R., B. T. Hirsch, C. M. Hokayem, and J. P. Ziliak (2018): "The Good, The Bad and The Ugly: Measurement Error, Non-response and Administrative Mismatch in the CPS," Discussion paper.

Bond, S., and J. Van Reenen (2007): "Microeconometric Models of Investment and Employment," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. VI, chap. 65, pp. 4417–4498. Elsevier.

Bonhomme, S., and J.-M. Robin (2010): "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *The Review of Economic Studies*, 77, 491–533.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994): "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12(3), 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. V, pp. 3705–3843. Elsevier Science B.V.

Bound, J., and A. B. Krueger (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics*, 9(1), 1–24.

Canay, I. A., A. Santos, and A. M. Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models With Endogeneity," *Econometrica*, 81(6), 2535–2559.

Card, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64(4), 957–979.

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49(4), 901–937.

Chesher, A. (1990): "On Assessing the Effect and Detecting the Presence of Measurement Error," Discussion paper, University of Bristol.

Chesher, A., M. Dumangane, and R. J. Smith (2002): "Duration response measurement error," *Journal of Econometrics*, 111(2), 169 – 194.

Chetty, R. (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80(3), 969–1018.

Chetverikov, D., D. Kim, and D. Wilhelm (2018): "An Adaptive Test of Stochastic Monotonicity," Working Paper CWP24/18, CeMMAP.

Chetverikov, D., and D. Wilhelm (2017): "Nonparametric Instrumental Variable Estimation Under Monotonicity," *Econometrica*, 85(4), 1303–1320.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

De Nadai, M., and A. Lewbel (2016): "Nonparametric errors in variables models with measurement errors on both sides of the equation," *Journal of Econometrics*, 191(1), 19–32.

Delgado, M. A., and W. Gonzalez Manteiga (2001): "Significance Testing in Nonparametric Regression Based on the Bootstrap," *The Annals of Statistics*, 29(5), 1469–1507.

Dudley, R. M. (2002): *Real Analysis and Probability*. Cambridge University Press, Cambridge.

Durbin, J. (1954): "Errors in Variables," *Review of the International Statistical Institute*, 22(1/3), 23–32.

Erickson, T., and T. M. Whited (2000): "Measurement Error and the Relationship between Investment and q," *Journal of Political Economy*, 108(5), 1027–1057.

——— (2012): "Treating Measurement Error in Tobin's q," *Review of Financial Studies*, 25(4), 1286–1329.

Evdokimov, K., and H. White (2012): "Some Extensions of a Lemma of Kotlarski," *Econometric Theory*, 28, 925–932.

Fan, J. (1991): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19(3), 1257–1272.

Fan, Y., and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64(4), 865–890.

——— (2000): "Consistent Model Specification Tests: Kernel-Based Tests Versus Bierens' ICM Tests," *Econometric Theory*, 16(6), 1016–1041.

Feng, S., and Y. Hu (2013): "Misclassification Errors and the Underestimation of the US Unemployment Rate," *American Economic Review*, 103(2), 1054–70.

Fitzenberger, B., A. Osikominu, and R. Völter (2006): "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," *Schmollers Jahrbuch (Zeitschrift für Wirtschafts- und Sozialwissenschaften / Journal of the Applied Social Sciences)*, 126(3), 405–436.

Gozalo, P. L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, 9(3), 451–477.

Groen, J. A. (2011): "Sources of Error in Survey and Administrative Data: The Importance of Reporting Procedures," Discussion paper.

Hahn, J., and J. Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70(1), 163–189.

Härdle, W., and E. Mammen (1993): "Comparing Nonparametric Versus Parametric Regression Fits," *The Annals of Statistics*, 21(4), 1926–1947.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–1271.

Hausman, J. A., W. K. Newey, and J. L. Powell (1995): "Nonlinear errors in variables Estimation of some Engel curves," *Journal of Econometrics*, 65(1), 205–233.

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103(6), 2052–2086.

Hoderlein, S., H. Holzmann, M. Kasy, and A. Meister (2016): "Erratum Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment," *The Review of Economic Studies*, forthcoming.

Horowitz, J. L., and M. Markatou (1996): "Semiparametric Estimation of Regression Models for Panel Data," *Review of Economic Studies*, 63(145-168).

Horowitz, J. L., and V. G. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69(3), 599–631.

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

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

Hu, Y., D. McAdams, and M. Shum (2013): "Identification of first-price auctions with non-separable unobserved heterogeneity," *Journal of Econometrics*, 174(2), 186–193.

Hu, Y., and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

Huang, M., Y. Sun, and H. White (2016): "A Flexible Nonparametric Test for Conditional Independence," *Econometric Theory*, 32(6), 1434–1482.

il Kim, K., and S. Song (2018): "Estimation of Semiparametric Models with Mismeasured Endogenous Regressors Using Control Variables," Discussion paper.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

Kane, T. J., and C. E. Rouse (1995): "Labor-Market Returns to Two- and Four-Year College," *The American Economic Review*, 85(3), 600–614.

Kane, T. J., C. E. Rouse, and D. Staiger (1999): "Estimating Returns to Schooling When Schooling is Misreported," Working Paper 7235, NBER.

Kanwal, R. P. (1983): *Generalized Functions: Theory and Technique*. Academic Press, New York.

Kapteyn, A., and J. Y. Ypma (2007): "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data," *Journal of Labor Economics*, 25(3), 513–551.

Karlin, S. (1968): *Total Positivity: Volume I*. Stanford University Press.

Karlin, S., and H. Rubin (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *The Annals of Mathematical Statistics*, 27(2), 272–299.

Kasy, M. (2014): "Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment," *The Review of Economic Studies*, 81(4), 1614.

Kotlarski, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20(1), 69–76.

Lee, S., O. Linton, and Y.-J. Whang (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77(2), 585–602.

Leeb, H., and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21, 21–59.

Lehmann, E. L. (1955): "Ordered Families of Distributions," *The Annals of Mathematical Statistics*, 26(3), 399–419.

Lewbel, A. (2000): "Identification of the Binary Choice Model With Misclassification," *Econometric Theory*, 16, 603–609.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, 72(3), 735–765.

Lucas, Jr., R. E., and E. C. Prescott (1971): "Investment Under Uncertainty," *Econometrica*, 39(5), 659–681.

Mahajan, A. (2006): "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74(3), 631–665.

Matzkin, R. L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, pp. 2523–2558. Elsevier Science B.V.

McFadden, D. (1989): "Testing for Stochastic Dominance," in *Studies in the Economics of Uncertainty*, ed. by T. B. Fomby, and T. K. Seo, pp. 113–134, New York, NY. Springer New York.

Mussa, M. (1977): "External and Internal Adjustment Costs and the Theory of Aggregate and Firm Investment," *Economica*, 44(174), 163–178.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

Romano, J. P. (2004): "On Non-parametric Testing, the Uniform Behaviour of the t-test, and Related Problems," *Scandinavian Journal of Statistics*, 31(4), 567–584.

Romano, J. P., A. M. Shaikh, and M. Wolf (2008): "Formalized Data Snooping Based on Generalized Error Rates," *Econometric Theory*, 24(2), 404–447.

——— (2010): "Hypothesis Testing in Econometrics," *Annual Review of Economics*, 2(1), 75–104.

Romano, J. P., and M. Wolf (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73(4), 1237–1282.

Rüschendorf, L. (1980): "Inequalities for the expectation of Δ-monotone functions," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 54(3), 341–349.

Santos, A. (2012): "Inference in Nonparametric Instrumental Variables With Partial Identification," *Econometrica*, 80(1), 213–275.

Saumard, A., and J. A. Wellner (2014): "Log-concavity and strong log-concavity: A review," *Statistics Surveys*, 8, 45–114.

Scarsini, M. (1988a): "Dominance Conditions for Multivariate Utility Functions," *Management Science*, 34(4), 454–460.

——— (1988b): "Multivariate stochastic dominance with fixed dependence structure," *Operations Research Letters*, 7(5), 237 – 240.

Schennach, S. M. (2007): "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 75(1), 201–239.

——— (2013): "Measurement Error in Nonlinear Models — A Review," in *Advances in Economics and Econometrics*, vol. 3. Cambridge University Press.

——— (2016): "Recent Advances in the Measurement Error Literature," *Annual Review of Economics*, 8, 341–377.

Song, S., S. M. Schennach, and H. White (2015): "Estimating nonseparable models with mismeasured endogenous variables," *Quantitative Economics*, 6(3), 749–794.

Tian, X., and J. Taylor (2017): "Asymptotics of Selective Inference," *Scandinavian Journal of Statistics*, 44(2), 480–499.

Tibshirani, R. J., A. Rinaldo, R. Tibshirani, and L. Wasserman (2018): "Uniform asymptotic inference and the bootstrap after model selection," *Annals of Statistics*, 46(3), 1255–1287.

Tibshirani, R. J., J. Taylor, R. Lockhart, and R. Tibshirani (2016): "Exact Post-Selection Inference for Sequential Regression Procedures," *Journal of the American Statistical Association*, 111(514), 600–620.

Wilhelm, D. (2015): "Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error," Working Paper CWP34/15, cemmap.

Wu, D.-M. (1973): "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrica*, 41(4), 733–750.