cemmap

centre for microdata methods and practice

Measurement error and rank correlations

Toru Kitagawa
Martin Nybom
Jan Stuhler

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP28/18

An ESRC Research Centre

Measurement Error and Rank Correlations

Toru Kitagawa[*], Martin Nybom[†] and Jan Stuhler[‡§]

This version: April 2018

Abstract

This paper characterizes and proposes a method to correct for errors-in-variables biases in the estimation of rank correlation coefficients (Spearman's ρ and Kendall's τ). We first investigate a set of sufficient conditions under which measurement errors bias the sample rank correlations toward zero. We then provide a feasible nonparametric bias-corrected estimator based on the technique of small error variance approximation. We assess its performance in simulations and an empirical application, using rich Swedish data to estimate intergenerational rank correlations in income. The method performs well in both cases, lowering the mean squared error by 50-85 percent already in moderately sized samples ($n = 1,000$).

Keywords: Errors-in-variables, Spearman's rank correlation, Kendall's tau, Small variance approximation, Intergenerational mobility.

[*]University College London, Department of Economics, CeMMAP, and Kyoto University, Department of Economics, Email: t.kitagawa@ucl.ac.uk

[†]IFAU Uppsala, SOFI at Stockholm University, and UCLS. Email: martin.nybom@ifau.uu.se

[‡]Universidad Carlos III de Madrid and SOFI. Email: jan.stuhler@uc3m.es

[§]We would like to thank Andrew Chesher for inspiring suggestions and encouragement on this project. We also thank Cristian Bartolucci, Jesper Bagger, Yingyao Hu, Ignacio Monzón, Jeff Racine, and seminar particpants at the Workshop on Intergenerational Mobility in Madrid 2016 and the Conference on Computational and Financial Econometrics 2016 for their valuable comments. Financial support from the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001), the ERC (Starting grant No. 715940), the Ministerio Economía y Competitividad (Spain, MDM 2014-0431 and ECO2014-55858-P) and the Comunidad de Madrid (MadEco-CM S2015/HUM-3444) is gratefully acknowledged.

1 Introduction

Spearman's rank correlation coefficient ρ (Spearman, 1904) and Kendall's τ (Kendall, 1938) are two widely used measures of dependence of two random variables. They are invariant to monotonic transformations of the variables and provide unit-free measurements of their statistical association. While not as pervasive as in other disciplines, rank correlations are used in an increasing variety of empirical literatures in economics, including research on intergenerational mobility (Chetty et al., 2014a), labor market sorting (Hagedorn, Law, and Manovskii 2017), in experimental economics (Dohmen and Falk, 2011), and health economics (Abellan-Perpiñan, Bleichrodt, and Pinto-Prades, 2009).[1]

However, errors-in-variables are ubiquitous in these as in other empirical applications. One common source is error in a reported variable itself. For instance, self-reported data of household expenditure, consumption or income are often contaminated by reporting errors. Measurement error may also arise if a latent variable of interest has to be approximated using observed quantities. For example, when one's lifetime or permanent income is not observed, the measurement of income at a particular age only acts as an imprecise approximation. If a variable refers to an intangible conceptual object, such as cognitive or non-cognitive ability, observable measures can only serve as proxies.

Although rank correlations are frequently estimated using imprecisely measured variables, little is known about how measurement errors influence these estimates, or how to correct for the resulting bias. This gap is notable, as the presence of measurement error is a major and often the prime motivation for the use of rank correlations. As Spearman argues (Spearman 1904, p. 81), "*The chief [advantage of the rank method] is the large reduction of the 'accidental erro*r'", as the conversion into ranks restricts the influence of outliers.[2] Rank correlations have therefore become popular in settings in which errors-in-variables are prevalent, such as the intergenerational mobility literature (e.g. Dahl and DeLeire, 2008; Bhattacharya and Mazumder, 2011; Chetty et al., 2014a; Chetty et al., 2014b; Corak, Lindquist, and Mazumder, 2014; Gregg, Macmillan, and Vittori, 2017; Nybom and Stuhler, 2017; Bratberg et al., 2017; Chen, Ostrovsky, and Piraino, 2017; Pekkarinen, Salvanes, and Sarvimäki, 2017).

This paper analyzes the effects of measurement error on the estimation of Spearman's ρ and Kendall's τ, and makes the following contributions. First, we provide analytical characterizations

[1] According to a keyword search in the database *Scopus*, the share of economics articles that report rank correlations has doubled in the 2010s compared to previous decades, and increased also in comparison to linear regression or correlation coefficients.

[2] Spearman favored the use of rank correlation coefficients (see Part I of Spearman 1904), but developed a method to correct for measurement error for the Pearson correlation instead (Part II of Spearman 1904). Due to their nonlinear nature, it is more difficult to characterize the effect of measurement error on rank correlations. Spearman's *correction for attenuation* is not applicable, as it builds on the assumption that the errors are uncorrelated to true values, while they are always negatively correlated in ranks.

of the errors-in-variables biases, allowing us to investigate whether and when the sample rank correlations identify the signs of the rank correlations of the true measurements. Second, we propose bias-corrected estimation methods for both Spearman's and Kendall's rank correlations that can be applied under classical and certain types of non-classical measurement error. We assess their performances in both simulations and a real empirical example in the context of intergenerational mobility research.

Our bias-correction proposal builds on the idea of small variance approximation introduced by Chesher (1991). The method yields an approximation of the measurement error bias by exploiting a series expansion of the estimand with respect to the measurement error variances. The expansion leads to an accurate approximation when these variances are close to zero. The first-order terms in the expansion, which are functionals of the distribution of the underlying true measurements multiplied by the measurement error variances, can capture the first-order influence of the measurement errors on the parameter estimate of interest. Following the general construction of a feasible bias-corrected estimator in Chesher (1991), we then construct bias-corrected estimators for Spearman's ρ and Kendall's τ by estimating the first-order bias terms nonparametrically and subtracting them from the sample estimates. We assess their performance by means of Monte Carlo studies. Depending on sample size and variance of the measurement error, the bias-corrected estimators reduce the bias by between 50 and 80 percent relative to estimators that ignore the measurement errors. The mean squared error (MSE) improves in samples as small as $n = 100$, and the MSE reductions become substantial at larger error variances or larger sample sizes (up to 85 percent at $n = 1000$).

Applications of the small error variance approximation appear in various contexts in econometrics and applied economics.[3] In the context of measurement error, all existing work invokes a classical errors-in-variables model, i.e. assuming that the errors are independent of the underlying true measurements. One notable feature of the approach when applied to rank correlations is that it can accommodate non-classical measurement errors if their dependence on the true measurements is constrained in some form. Specifically, our bias-correction formulas remain valid in the following two scenarios. First, if the error depends on the true measurement only through its conditional mean and the conditional mean does not alter the population ranking of the true measurements. Second, if the dependence between the errors and the true measurements arises because the measurements are strictly increasing and nonlinear transformations of the object defined as the true measurement plus an additive error. These capabilities of accommodating non-classical errors stem from the invariance property of the rank correlation coefficients to monotonic transformations, and enhance the applicability of the correction method in empirically relevant problems.

[3] They include measurement errors in inequality and social welfare measurements (Chesher and Schluter, 2002), random coefficients in discrete choice (Chesher and Silva, 2002), errors in covariates in program evaluation (Battistin and Chesher, 2014), quantile regressions with mismeasured regressors (Chesher, 2017), and violation of instrument monotonicity in the local average (marginal) treatment effect model (Klein, 2010), to list a few.

Our empirical application concerns the estimation of Spearman's ρ and Kendall's τ of father's and son's lifetime incomes using rich administrative data on life-cycle incomes and parent-child links for Sweden. Since the data set contains annual incomes that span most of the lifecycle of both fathers and their sons we can construct approximate measurements of lifetime income. We then compare our bias-corrected estimates based on snapshots of income at a certain age with estimates of Spearman's ρ and Kendall's τ based on our measures of lifetime income. We find that for sufficiently large samples (for $n = 1000$), the bias-corrected estimator reduces the MSE by between 50 and 60 percent compared to the estimator with no bias correction, while the reduction is less substantial in smaller samples.[4] This finding suggests that our correction method is a useful tool for the estimation of intergenerational income correlations in settings in which incomes are observed only over limited periods, as is typically the case. To our knowledge, previous work has not examined the actual bias-correction performance of small variance approximations in real data, since precise measurements and a bias-free parameter estimate are usually unavailable. In contrast, we can keep the distributions of true measurements and measurement errors at those obtained from the rich data, and compare the MSE performances of our bias-corrected methods to the "oracle" procedure that we would run were the true measurements not available.

Despite the long history of Spearman's ρ and Kendall's τ in statistics, surprisingly little work exists on their relationship with measurement error. Bartolucci, Devicienti, and Monzón (2015) show that measurement error causes an attenuation bias for Kendall's τ in a setup where the unobserved true measurements are functions of a scalar unobserved random variable. Our analysis allows for a more general setting where the unobserved true measurements have a non-degenerate bivariate distribution. Nybom and Stuhler (2017) study how various summary measures of the dependence between parent and child income compare in annual and lifetime incomes, and propose a correction method for Spearman's ρ under the assumption that the measurement error *in ranks* can be well described by a linear projection. While this property is shown to hold approximately in their data, it is not generally known what assumptions on the underlying distributions that can guarantee it. Our analysis here does not rely on assumptions on the error in ranks, but on the actual values underlying those ranks. An, Wang, and Xiao (2017) analyze identification and estimation of an intergenerational mobility function in a setting in which the relation between measurement error and true measures can deviate from the classical errors-in-variables model.

The remainder of the paper is organized as follows. Section 2 lays out the framework of small variance approximation and derives an analytical expression of the first-order bias. It also shows analytical results concerning the attenuation bias and extensions to settings with non-classical errors-in-variables. Section 3 proposes plug-in based bias-corrected estimators for Spearman's ρ and Kendall's τ. To assess the actual performance of the method, we conduct extensive Monte

[4] We restrict our analyses to sample sizes up to $n = 1000$. However, since the variance of the bias-corrected estimator decreases with sample size, the MSE reduction is likely to be even larger for larger sample sizes.

Carlo studies in Section 4. Section 5 illustrates the empirical implementation of our method using Swedish income data, and Section 6 concludes. Proofs omitted from the main text are collected in the Appendix.

2 Measurement Error Bias: Small Error Variance Approximation

Consider continuously distributed bivariate random variables $(X, Y) \in \mathcal{R}^2$ with the joint cdf $F_{X,Y}(\cdot, \cdot)$. Assume that (X, Y) are observable and measure latent random variables $(X^*, Y^*) \in \mathcal{R}^2$ subject to additive measurement errors;

$$\begin{aligned} X &= X^* + \mu_{\epsilon_X} + \sigma_{\epsilon_X}\epsilon_X, \\ Y &= Y^* + \mu_{\epsilon_Y} + \sigma_{\epsilon_Y}\epsilon_Y, \end{aligned} \tag{2.1}$$

where ϵ_X and ϵ_Y are random variables with mean zero, unit marginal variances, and finite third-order moments, μ_{ϵ_X} and μ_{ϵ_Y} are the means of the measurement errors, and $\sigma_{\epsilon_X} \geq 0$ and $\sigma_{\epsilon_Y} \geq 0$ are the standard deviations of the measurement errors. As evident from this representation, we assume the measurement errors, $X - X^*$ and $Y - Y^*$ have finite variances with potentially nonzero means. The baseline setup of our analysis assumes classical measurement errors.

Assumption 2.1. (ϵ_X, ϵ_Y) *are statistically independent of* (X^*, Y^*).

Even though the rank correlations between X and Y and X^* and Y^* are invariant to monotonic transformations, validity of the additive measurement error representation with Assumption 2.1 is not. Admitting (2.1) and imposing Assumption 2.1 for a particular scale of measurement of (X, Y), nonlinear transformations of (X, Y) does not guarantee the existence of the additive representation in the form of (2.1) for the transformed measurements with additive classical measurement errors. As a result, the bias-correction procedure proposed in this paper is *not* generally transformation invariant, and this encourages us to argue for which scale of measurements of (X, Y), the additive error representation of (2.1) and Assumption 2.1 is reasonable in view of economic theory or any available knowledge of the sampling process.

While the classical measurement error assumption can be restrictive in some contexts, we maintain it in our baseline setup. In Section 2.4 below, we relax the classical measurement error assumption and allow the errors to be correlated with X^* and Y^*.

We do not constrain the covariance of the measurement errors, so ϵ_X and ϵ_Y can have nonzero correlation $\sigma_{\epsilon_X\epsilon_Y} \equiv Corr(\epsilon_X, \epsilon_Y) = Cov(\epsilon_X, \epsilon_Y)$. We denote the joint cdf of (X^*, Y^*) by $F_{X^*,Y^*}(\cdot, \cdot)$, and the cdfs of (ϵ_X, ϵ_Y), ϵ_X, and ϵ_Y by $G_{\epsilon_X,\epsilon_Y}(\cdot, \cdot)$, $G_{\epsilon_X}(\cdot)$, and $G_{\epsilon_Y}(\cdot)$, respectively. The next assumption imposes continuity and smoothness of the distribution of (X^*, Y^*).

Assumption 2.2. *(i) $F_{X^*,Y^*}(\cdot,\cdot)$ has bounded probability density with respect to the Lebesgue measure on $\mathcal{R}^2$.*

(ii) The density function of $F_{X^*,Y^*}(\cdot,\cdot)$ is everywhere three-times continuously differentiable.

2.1 Spearman's ρ and Kendall's τ

The Spearman's rank correlation coefficient ρ of bivariate random variables (X,Y) is defined by the correlation of ranks $Corr(F_X(X), F_Y(Y))$. For our purpose of characterizing and correcting the measurement error bias, it is convenient to express ρ in terms of the probabilities of concordance and discordance. To do so, let (X_1,Y_1), (X_2,Y_2), and (X_3,Y_3) be independent copies of $(X,Y) \sim F_{X,Y}$. Following equation (5.1.14) of Nelsen (2006), ρ can be equivalently written as three times the difference between the concordance and discordance probabilities of (X_1,Y_1) and (X_2,Y_3),

$$\begin{aligned}
\rho &= 3\left[\Pr\left((X_1-X_2)(Y_1-Y_3)>0\right)-\Pr\left((X_1-X_2)(Y_1-Y_3)<0\right)\right] \\
&= 6\left[\Pr(X_1>X_2,Y_1>Y_3)+\Pr(X_1<X_2,Y_1<Y_3)\right]-3. \\
&= 12\Pr(X_1>X_2,Y_1>Y_3)-3, \qquad (2.2)
\end{aligned}$$

where the second equality follows by noting that $\Pr(X_1>X_2,Y_1<Y_3)=\frac{1}{2}-\Pr(X_1<X_2,Y_1<Y_3)$ and $\Pr(X_1<X_2,Y_1>Y_3)=\frac{1}{2}-\Pr(X_1>X_2,Y_1>Y_3)$, and the third equality follows by noting that $\Pr(X_1<X_2,Y_1<Y_3)=\Pr(Y_1<Y_3)-\Pr(X_1>X_2)+\Pr(X_1>X_2,Y_1>Y_3)=\Pr(X_1>X_2,Y_1>Y_3)$.

Kendall's τ for continuous bivariate random variables (X,Y) is defined as the difference between the concordance and discordance probabilities of (X_1,Y_1) and (X_2,Y_2),

$$\begin{aligned}
\tau &\equiv \Pr\left((X_1-X_2)(Y_1-Y_2)>0\right)-\Pr\left((X_1-X_2)(Y_1-Y_2)<0\right) \\
&= 4\Pr(X_1>X_2,Y_1>Y_2)-1. \qquad (2.3)
\end{aligned}$$

Let (X_1^*,Y_1^*), (X_2^*,Y_2^*), and (X_3^*,Y_3^*) be independent copies of $(X^*,Y^*) \sim F_{X^*,Y^*}$, and denote the Spearman's and Kendall's rank correlations of (X^*,Y^*) by ρ^* and τ^*, respectively.

Given a random sample of observable measurements (X,Y), we can consistently estimate Spearman's ρ and Kendall's τ by their sample analogues. However, if the measurements (X,Y) are subject to nondegenerate measurement errors as in (2.1), they are generally biased estimators for ρ^* and τ^*. In Section 2.3, we investigate the conditions under which τ is subject to attenuation bias, and the signs of ρ^* and τ^* are identified. Without strong distributional assumptions on the measurement errors, it is not feasible to nonparametrically identify ρ^* and τ^*. To develop a feasible and widely applicable way to reduce the biases, we consider approximating the bias of ρ by applying the small error variance approximation of Chesher (1991).

2.2 Small Error Variance Approximations for Measurement Error Bias

For $k, l = 1, 2, 3$, $k \neq l$, let $\Delta_{kl}X = X_k - X_l$, $\Delta_{kl}Y = Y_k - Y_l$, $\Delta_{kl}X^* = X_k^* - X_l^*$, $\Delta_{kl}Y^* = Y_k^* - Y_l^*$, $\Delta_{kl}\epsilon_X = \epsilon_{X_k} - \epsilon_{X_l}$, and $\Delta\epsilon_Y = \epsilon_{Y_k} - \epsilon_{Y_l}$. We denote their joint distribution functions by $F_{\Delta_{kl}X, \Delta_{k'l'}Y}$, $F_{\Delta_{kl}X^*, \Delta_{k'l'}Y^*}$ and $G_{\Delta_{kl}\epsilon_X, \Delta_{k'l'}\epsilon_Y}$, respectively. The measurement equations (2.1) and Assumption 2.1 imply that

$$\begin{aligned}
\Pr(X_k > X_l, Y_{k'} > Y_{l'}) &= \Pr(\Delta_{lk}X^* < \sigma_{\epsilon_X}\Delta_{kl}\epsilon_X, \Delta_{l'k'}Y^* < \sigma_{\epsilon_Y}\Delta_{k'l'}\epsilon_Y) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{\Delta_{lk}X^*, \Delta_{l'k'}Y^*}(\sigma_{\epsilon_X}\Delta_{kl}\epsilon_X, \sigma_{\epsilon_Y}\Delta_{k'l'}\epsilon_Y) dG_{\Delta_{kl}\epsilon_X, \Delta_{k'l'}\epsilon_Y}
\end{aligned}$$

where the second line follows as the convolution of $(\Delta_{lk}X^*, \Delta_{l'k'}Y^*)$ with independent random variables $(\Delta_{kl}\epsilon_X, \Delta_{k'l'}\epsilon_Y)$. Viewing $\Pr(X_k > X_l, Y_{k'} > Y_{l'})$ as a function of $(\sigma_{\epsilon_X}, \sigma_{\epsilon_Y})$ and imposing Assumption 2.2, we consider the second-order Taylor expansion of $\Pr(X_k > X_l, Y_{k'} > Y_{l'})$ around $(\sigma_{\epsilon_X}, \sigma_{\epsilon_Y}) = (0, 0)$. Since $\Pr(X_k > X_l, Y_{k'} > Y_{l'})$ at $(\sigma_{\epsilon_X}, \sigma_{\epsilon_Y}) = (0, 0)$ equals $\Pr(X_k^* > X_l^*, Y_{k'}^* > Y_{l'}^*)$, this expansion provides an approximation of the measurement error bias in $\Pr(X_k > X_l, Y_{k'} > Y_{l'})$.

Lemma A1.1 in Appendix shows that this expansion yields

$$\begin{aligned}
\Pr(X_k > X_l, Y_{k'} > Y_{l'}) &= \Pr(X_k^* > X_l^*, Y_k^* > Y_l^*) + \frac{1}{2}f'_{\Delta_{lk}X^*|\Delta_{l'k'}Y^*<0}(0)\sigma_{\epsilon_X^2} + \frac{1}{2}f'_{\Delta_{l'k'}Y^*|\Delta_{lk}X^*<0}(0)\sigma_{\epsilon_Y^2} \\
&\quad + [1_{k=k'} - 1_{k=l'} - 1_{l=k'} + 1_{l=l'}]f_{\Delta_{lk}X^*, \Delta_{l'k'}Y^*}(0,0)\sigma_{\epsilon_X\epsilon_Y}\sigma_{\epsilon_X}\sigma_{\epsilon_Y} \\
&\quad + O((\sigma_{\epsilon_X} + \sigma_{\epsilon_Y})^3).
\end{aligned}$$

where $f_{\Delta_{lk}X^*|\Delta_{l'k'}Y^*\leq 0}(\cdot)$ is the pdf of $\Delta_{lk}X$ conditional on $\Delta_{l'k'}Y^* \leq 0$, $f_{\Delta_{l'k'}Y^*|\Delta_{lk}X^*\leq 0}(\cdot)$ is the pdf of $\Delta_{l'k'}Y^*$ conditional on $\Delta_{lk}X^* \leq 0$, and $f_{\Delta_{lk}X^*, \Delta_{l'k'}Y^*}(\cdot, \cdot)$ is the joint pdf of $(\Delta_{lk}X^*, \Delta_{l'k'}Y^*)$. Plugging this expansion into (2.2) and (2.3), we obtain the small error variance approximations of ρ and τ.

Proposition 2.1. *Under Assumptions 2.1 and 2.2,*

$$\begin{aligned}
\rho &= \rho^* + 6f'_{\Delta_{21}X^*|\Delta_{31}Y^*<0}(0)\sigma_{\epsilon_X}^2 + 6f'_{\Delta_{31}Y^*|\Delta_{21}X^*<0}(0)\sigma_{\epsilon_Y}^2 \\
&\quad + 12f_{\Delta_{21}X^*, \Delta_{31}Y^*}(0,0)\sigma_{\epsilon_X\epsilon_Y}\sigma_{\epsilon_X}\sigma_{\epsilon_Y} + O((\sigma_{\epsilon_X} + \sigma_{\epsilon_Y})^3), \qquad (2.4)
\end{aligned}$$

$$\begin{aligned}
\tau &= \tau^* + 2f'_{\Delta_{21}X^*|\Delta_{21}Y^*\leq 0}(0)\sigma_{\epsilon_X}^2 + 2f'_{\Delta_{21}Y^*|\Delta_{21}X^*\leq 0}(0)\sigma_{\epsilon_Y}^2 \qquad (2.5) \\
&\quad + 8f_{\Delta_{21}X^*, \Delta_{21}Y^*}(0,0)\sigma_{\epsilon_X\epsilon_Y}\sigma_{\epsilon_X}\sigma_{\epsilon_Y} + O\left((\sigma_{\epsilon_X} + \sigma_{\epsilon_Y})^3\right).
\end{aligned}$$

This proposition shows that the first-order measurement error biases of Spearman's ρ and Kendall's τ depend on the density or the density derivatives of the joint or conditional distributions of $(\Delta_{21}X^*, \Delta_{l'k'}Y^*)$ at the origin, where $(l', k') = (3, 1)$ for Spearman's ρ and $(l', k') = (2, 1)$ for Kendall's τ. These expressions of the bias approximations facilitate nonparametric estimation of the biases (given a user's choice of $(\sigma^2_{\epsilon_X}, \sigma^2_{\epsilon_Y}, \sigma_{\epsilon_X \epsilon_Y})$), as we discuss further in Section 3. The validity of these bias approximations does not require any distributional assumptions of the true measurements and errors. Furthermore, these approximation formulae remain valid even if we relax the classical mesurement error assumption to a certain extent (see Section 2.4 below). The quality of approximation, however, crucially depends on the magnitudes of the measurement error biases relative to the variances of X^* and Y^*. We assess the quality of approximations by simulation in Section 4.

2.3 Attenuation Bias and Identifying Signs

It is well known that the Pearson's correlation coefficient of (X, Y) identifies the sign of the correlation coefficient of (X^*, Y^*), but underestimates its magnitude. This is the well-known attenuation bias caused by classical measurement errors. Are the rank correlation coefficients also subject to attenuation bias? In this section, we characterize a set of sufficient conditions that implies an attenuation bias for τ and the sign identification of ρ and τ.

We start with our analysis of Kendall's τ. Suppose Assumption 2.1 holds and ϵ_X is independent of ϵ_Y. Then, τ admits the following representation:

$$\tau = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{y_1^*} b(y_1^*, y_2^*) \left[2 \Pr\left(X_1 > X_2 | y_1^*, y_2^*\right) - 1\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right), \tag{2.6}$$

where $b(y_1^*, y_2^*) = 2\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - 1$. See Appendix A1 for a derivation of (2.6). On the other hand, note that τ^* can be written as

$$\begin{aligned} \tau^* &= 2\left[2\Pr\left(X_1^* > X_2^*, Y_1^* > Y_2^*\right) - \Pr(Y_1^* > Y_2^*)\right] \\ &= 2\int_{-\infty}^{\infty} \int_{-\infty}^{y_1^*} \left[2\Pr\left(X_1^* > X_2^* | y_1^*, y_2^*\right) - 1\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right). \end{aligned} \tag{2.7}$$

A comparison of (2.6) and (2.7) shows that if (i) $0 \le b(y_1^*, y_2^*) \le 1$ and (ii) $\frac{1}{2} \le \Pr\left(X_1 > X_2 | y_1^*, y_2^*\right) \le \Pr\left(X_1^* > X_2^* | y_1^*, y_2^*\right)$ hold for all $y_1^* > y_2^*$, the sign of τ agrees with the sign of τ^* but τ is biased toward zero. The condition (i) always holds under the classical measurement errors, since the facts that $\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) = \Pr(\Delta_{21}\epsilon_Y < \sigma_{\epsilon_Y}^{-1}(y_1^* - y_2^*))$ holds and $\Delta_{21}\epsilon_Y$ has median zero imply $\frac{1}{2} \le \Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) \le 1$ for any $y_1^* \ge y_2^*$. Hence, the condition (ii) is sufficient for τ to be attenuated in case of $\tau^* \ge 0$. Along the same line of argument, if $\Pr\left(X_1^* > X_2^* | y_1^*, y_2^*\right) \le \Pr\left(X_1 > X_2 | y_1^*, y_2^*\right) \le \frac{1}{2}$ for all $y_1^* > y_2^*$, τ has an attenuation bias when $\tau^* \le 0$. We summarize these results in the next proposition.

Proposition 2.2. *Suppose Assumption 2.1 holds and ϵ_X is independent of ϵ_Y.*
(i) If $\frac{1}{2} \le \Pr\left(X_1 > X_2 | y_1^*, y_2^*\right) \le \Pr\left(X_1^* > X_2^* | y_1^*, y_2^*\right)$ holds for every $y_1^* > y_2^*$, then

$$0 \le \tau \le \tau^*.$$

(ii) Symmetrically, if $\Pr(X_1^* > X_2^* | y_1^*, y_2^*) \leq \Pr(X_1 > X_2 | y_1^*, y_2^*) \leq \frac{1}{2}$ *holds for every* $y_1^* > y_2^*$*, then* $\tau^* \leq \tau \leq 0$.

The conditions of Proposition 2.2 (i) or (ii) can be implied if we restrict the joint distribution of (X^*, Y^*) to a class characterized by the next proposition:

Proposition 2.3. *Suppose Assumption 2.1 holds and* ϵ_X *is independent of* ϵ_Y*. If* $\Pr(X^* > x | Y^* = y)$ *is nondecreasing in* y *for all* x *and the distribution of* $(X_1^* - X_2^*)$ *conditional on* $\{Y_1^* = y_1^*, Y_2^* = y_2^*, y_1^* > y_2^*\}$ *is symmetric and unimodal with the nonnegative mode for any* $y_1^* > y_2^*$*, then the condition of Proposition 2.2 (i) holds. Symmetrically, if* $\Pr(X^* > x | Y^* = y)$ *is nonincreasing in* y *for all* x *and the distribution of* $(X_1^* - X_2^*)$ *conditional on* $\{Y_1^* = y_1^*, Y_2^* = y_2^*, y_1^* > y_2^*\}$ *is symmetric and unimodal with nonpositive mode, then the condition of Proposition 2.2 (ii) holds.*

The assumption that $\Pr(X^* > x | Y^* = y)$ is non-decreasing in y for all x corresponds to the concept of *stochastically increasing positive dependence* defined in Lehmann (1966). This is a concept of strong positive dependence that is invariant to monotonic transformations and implies positive rank correlations. The same concept has been referred to as *stochastic monotonicity* in the literature (see, e.g., Lee., Linton, and Whang (2009) and Delgado and Escanciano (2012)). The assumption is likely to hold in the intergenerational context considered in our empirical application, as we discuss below.

Note that the attenuation bias result of Proposition 2.2 does not rely on the small variance approximation, so under the stated conditions, τ is subject to attenuation bias for any magnitude of the error variances. We note that the sign of the measurement error bias predicted by the first-order bias terms in (2.5) is consistent with the exact result of Proposition 2.3 under $\sigma_{\epsilon_X \epsilon_Y} = 0$.

As for Spearman's ρ, a representation analogous to (2.6) becomes more involved, and finding succinct and interpretable sufficient conditions for the attenuation bias appears challenging. We can, however, find a simple sufficient condition for the identification of the sign of ρ^* by combining Proposition 2.2 with the Daniels' inequalities (Daniels (1950)) (see also Theorem 5.1.10 in Nelsen (2006)),

$$\frac{3\tau^* - 1}{2} \leq \rho^* \leq \frac{3\tau^* + 1}{2},$$

which holds for any continuously distributed bivariate random variables (X^*, Y^*). Under the setting of Proposition 2.2 (i), $\tau \leq \tau^*$ leads to $\frac{3\tau - 1}{2} \leq \rho^*$. Hence, if $\tau \geq 1/3$, we can conclude $\rho^* \geq 0$ and the sign of ρ agrees with that of ρ^*. A symmetric result is obtained under the setting of Proposition 2.2 (ii). The next proposition summarizes these results.

Proposition 2.4. *Suppose Assumption 2.1 holds and* ϵ_X *is independent of* ϵ_Y*.*
(i) If $\frac{1}{2} \leq \Pr(X_1 > X_2 | y_1^*, y_2^*) \leq \Pr(X_1^* > X_2^* | y_1^*, y_2^*)$ *hold for every* $y_1^* > y_2^*$ *and* $\tau \geq 1/3$*, then,* $\rho, \rho^* \geq 0$*. (ii) Symmetrically, if* $\Pr(X_1^* > X_2^* | y_1^*, y_2^*) \leq \Pr(X_1 > X_2 | y_1^*, y_2^*) \leq \frac{1}{2}$ *holds for every* $y_1^* > y_2^*$ *and* $\tau \leq -1/3$*, then* $\rho, \rho^* \leq 0$.

2.4 Nonclassical Measurement Errors

The baseline model specification considered above assumes classical measurement errors (Assumption 2.1). This assumption can be restrictive in some contexts (e.g., in the intergenerational mobility example considered in Section 5). Exploiting the invariance property of the rank correlation with respect to monotonic transformations, we can relax the assumption of classical errors to some extent.

Specifically, the small error variance approximation developed above can accommodate nonclassical errors in the following two scenarios specified in Assumptions 2.3 and 2.4 below.

Assumption 2.3. *(Rank preserving measurement errors with mean dependence)*

The conditional means of the measurement errors given their true measurements

$$E\left(X-X^{*}|X^{*}=x\right)=\mu_{\epsilon_{X}}\left(x\right),\ \ E\left(Y-Y^{*}|Y^{*}=y\right)=\mu_{\epsilon_{Y}}\left(y\right)$$

satisfy $\mu'_{\epsilon_X}>-1$ *and* $\mu'_{\epsilon_Y}>-1$ *, and the residuals* $\left(X-X^{*}-\mu_{\epsilon_{X}}\left(X^{*}\right),Y-Y^{*}-\mu_{\epsilon_{Y}}\left(Y^{*}\right)\right)$ *are statistically independent of* (X^{*},Y^{*}) *and have finite variances, i.e., the measurement equations can be represented by*

$$\begin{aligned} X &= X^{*}+\mu_{\epsilon_{X}}\left(X^{*}\right)+\sigma_{\epsilon_{X}}\epsilon_{X}, \\ Y &= Y^{*}+\mu_{\epsilon_{Y}}\left(Y^{*}\right)+\sigma_{\epsilon_{Y}}\epsilon_{Y}, \end{aligned} \tag{2.8}$$

where $(\epsilon_{X},\epsilon_{Y})$ *are the unit variance random variables independent of* (X^{*},Y^{*}).

Under the derivative restrictions of Assumption 2.3, the random variables $\tilde{X}\equiv X^{*}+\mu_{\epsilon_{X}}\left(X^{*}\right)$ and $\tilde{Y}\equiv Y^{*}+\mu_{\epsilon_{Y}}\left(Y^{*}\right)$ can be viewed as strictly monotonic transformations of X^{*} and Y^{*} so that the rank correlation of $(\tilde{X},\tilde{Y})$ agrees with that of (X^{*},Y^{*}). Hence, under Assumption 2.3, the estimation of the rank correlation of (X^{*},Y^{*}) can be reduced to the estimation of the rank correlation of $(\tilde{X},\tilde{Y})$ in the measurement equations (2.8) with additive classical errors. The small error variance approximations of Proposition 2.1 then apply and provide the first-order bias approximations.

Assumption 2.3 extends the applicability of the bias-corrected estimator to other settings in which classical measurement error assumptions would be too restrictive. An important example is the literature on intergenerational mobility, in which *lifetime* incomes of fathers and sons (X^{*},Y^{*}) are often the variables of interest. Jenkins (1987) and Haider and Solon (2006) note that the relationship between observable (log) annual incomes (X,Y) and lifetime incomes (X^{*},Y^{*}) departs substantially from a classical errors-in-variables model. Haider and Solon show that a better approximation is given by the "generalized" errors-in-variables model, $X=\lambda X^{*}+u_{X}$ with $Cov(X^{*},u_{X})=0$, and Nybom and Stuhler (2016) find that this model captures much (but not all) of the measurement error bias in estimates of the intergenerational income *elasticity* (the slope coefficient in the linear regression of Y^* on X^*). This generalized model is the standard way of addressing life-cycle bias in the literature, and corresponds to a linearized version of the rank-preserving measurement error with mean dependence assumed here, i.e. $\mu_{\epsilon_X}(X^*) = (\lambda - 1)X^*$. The condition $\frac{d}{dx}\mu_{\epsilon_X}(x) > -1$ is satisfied if $\lambda > 0$, as is the case as long as incomes are not measured at very young ages (Böhlmark and Lindquist, 2006). These results suggest that the bias-corrected estimators proposed here can be usefully applied to this literature.

The second scenario where the bias formulas of Proposition 2.1 can straightforwardly accommodate nonclassical errors is specified in the next assumption.

Assumption 2.4. *The measurement equations are strictly increasing transformations of the right-hand sides of equation (2.1), i.e.,*

$$\begin{aligned} X &= g_X(X^* + \mu_{\epsilon_X} + \sigma_{\epsilon_X}\epsilon_X), \\ Y &= g_Y(Y^* + \mu_{\epsilon_Y} + \sigma_{\epsilon_Y}\epsilon_Y), \end{aligned} \tag{2.9}$$

where g_X and g_Y and are strictly increasing functions and (ϵ_X, ϵ_Y) are mean zero random variables with unit variances independent of (X^*, Y^*) as in Assumption 2.1.

The measurement equations specified by (2.9) generally makes the errors $X - X^*$ dependent on X^*. But the rank correlation of (X, Y) and that of $(g_X^{-1}(X), g_Y^{-1}(Y))$ are identical, due to the strict monotonicity of g_X and g_Y (a similar point is made by Bhattacharya and Mazumder 2011 in the context of transition probabilities). As such, the bias approximation formulas of Proposition 2.1 remain valid even without knowledge of g_X and g_Y. The lack of knowledge of these transformations, however, complicates the construction of a bias-corrected estimator as discussed in Section 3.

The invariance arguments of the rank correlations discussed here imply that the attenuation bias results for τ shown in Proposition 2.2 is valid also when the classical errors-in-variables assumption (Assumption 2.1) is weakened to the rank preserving measurement error assumption (Assumption 2.3) as far as the independence between ϵ_X and ϵ_Y is maintained. In addition, Proposition 2.2 can be generalized to a class of measurement equations of the form given by (2.9).

3 Bias-corrected Estimators for Rank Correlations

In the benchmark setting of Assumptions 2.1 and 2.2, the first-order biases shown in Proposition 2.1 depend on several unknown objects: the variances and correlation of the measurement errors, and some distributional features of the latent variables (X^*, Y^*). In the absence of additional assumptions or auxiliary data, they cannot be identified so that it is infeasible to estimate them directly from the observations of (X, Y). To make the estimation of the first-order bias feasible, we assume that the researcher has a credible point estimate $\left(\hat{\sigma}^2_{\epsilon_X}, \hat{\sigma}^2_{\epsilon_Y}, \hat{\sigma}_{\epsilon_X\epsilon_Y}\right)$ for the unknown parameters $\left(\sigma^2_{\epsilon_X}, \sigma^2_{\epsilon_Y}, \sigma_{\epsilon_X\epsilon_Y}\right)$. This point estimate may be based on user's background knowledge on the sampling process or an analysis of auxiliary data. To facilitate the elicitation of $\left(\hat\sigma^2_{\epsilon_X}, \hat\sigma^2_{\epsilon_Y}\right)$, we normalize the variances of X^* and Y^* to be one. That is, $\sigma^2_{\epsilon_X}$ and $\sigma^2_{\epsilon_Y}$ are interpreted as the relative size of the measurement error variances when the variances of the latent variables are normalized to one. With this standardization, the variance of the observable X and Y are equal to $1+\sigma^2_{\epsilon_X}$ and $1+\sigma^2_{\epsilon_Y}$, respectively. In what follows, we assume that the observable measurements (X, Y) in the data are scaled accordingly, i.e., the raw observations of X and Y in the data are each multiplied by $\sqrt{1+\hat\sigma^2_{\epsilon_X}}/\hat\sigma_X$ and $\sqrt{1+\hat\sigma^2_{\epsilon_Y}}/\hat\sigma_Y$, where $\hat\sigma_X$ and $\hat\sigma_Y$ are the sample standard deviations of X and Y in the given data.

Given $\left(\hat\sigma^2_{\epsilon_X}, \hat\sigma^2_{\epsilon_Y}, \hat\sigma_{\epsilon_X\epsilon_Y}\right)$, the unknown quantities in the bias expression are those that depend on the distribution of (X^*, Y^*). Following Chesher (1991), we replace them with their analogues in terms of the distribution of observables (X, Y). This simple replacement does notalter the leading terms in the small error variance approximation, since the incurred additional approximation errors are all in the order smaller than $O((\sigma_{\epsilon_X}+\sigma_{\epsilon_Y})^2)$.

The bias expressions of Proposition 2.1 suggest the following first-order-bias corrected estimator for ρ^* and τ^*,

$$\hat\rho^*_{bc} = \hat\rho - 6\widehat{f'_{\Delta_{21}X|\Delta_{31}Y\le 0}}(0)\hat\sigma^2_{\epsilon_X} - 6\widehat{f'_{\Delta_{31}Y|\Delta_{21}X\le 0}}(0)\hat\sigma^2_{\epsilon_Y} - 12\widehat{f_{\Delta_{21}X,\Delta_{31}Y}}(0,0)\,\hat\sigma_{\epsilon_X}\hat\sigma_{\epsilon_Y}\hat\sigma_{\epsilon_X\epsilon_Y}, \tag{3.1}$$

$$\hat\tau^*_{bc} = \hat\tau - 2\widehat{f'_{\Delta_{21}X|\Delta_{21}Y\le 0}}(0)\hat\sigma^2_{\epsilon_X} - 2\widehat{f'_{\Delta_{21}Y|\Delta_{21}X\le 0}}(0)\hat\sigma^2_{\epsilon_Y} - 8\widehat{f_{\Delta_{21}X,\Delta_{21}Y}}(0,0)\,\hat\sigma_{\epsilon_X}\hat\sigma_{\epsilon_Y}\hat\sigma_{\epsilon_X\epsilon_Y}, \tag{3.2}$$

where $\hat\rho$ and $\hat\tau$ are consistent estimators for the Spearman's and Kendall's rank correlations of (X, Y). $\widehat{f'_{\Delta_{l1}X|\Delta_{l'1}Y\le 0}}(0)$ is an estimator for the density derivative of $\Delta_{l1}X$ given $\Delta_{l'1}Y \le 0$ evaluated at $\Delta_{l1}X = 0$, $\widehat{f'_{\Delta_{l'1}Y|\Delta_{l1}X\le 0}}(0)$ is an estimator for the density derivative of $\Delta_{l'1}Y$ given $\Delta_{l1}X \le 0$ evaluated at $\Delta_{l'1}Y = 0$, and $\widehat{f_{\Delta_{l1}X,\Delta_{l'1}Y}}(0,0)$ is an estimator for the joint density of $(\Delta_{l1}X, \Delta_{l'1}Y)$ evaluated at the origin. In estimation of these quantities, we construct a size n sample of $(X_{1i}, X_{li}, Y_{1i}, Y_{l'i})$ by setting $\{(X_{1i}, Y_{1i}) : i = 1, \ldots, n\}$ at the original sample $\{(X_i, Y_i) : i = 1, \ldots, n\}$ and, for each $l = 2, 3$, generating $\{(X_{li}, Y_{li}) : i = 1, \ldots, n\}$ by randomly permuting $\{(X_{1i}, Y_{1i}) : i = 1, \ldots, n\}$. In the Monte Carlo study and empirical application below, we estimate $f'_{\Delta_{l1}X|\Delta_{l'1}Y\le 0}$ by the kernel density derivative estimator using the subsample $\{i : \Delta_{l'1}Y_i \le 0\}$. Tuning the smoothing parameters is a nontrivial task here given that the ultimate object of interest is the rank correlation coefficients rather than the density derivative of the observables itself. In what follows, we estimate the densities $f_{\Delta_{l1}X|\Delta_{l'1}Y\le 0}$ and $f_{\Delta_{l'1}Y|\Delta_{l1}X\le 0}$ using kernel density estimators with bandwidths chosen by maximum likelihood cross validation (see Hall, Racine, and Li, 2004) with Gaussian kernel adjusted by multiplying $n^{1/5-1/7}$ to the cross-validated bandwidths. We then take their derivatives and evaluate them at the origin. We can estimate $f'_{\Delta_{l'1}Y|\Delta_{l1}X\le 0}$ in a similar manner. For the estimation of the joint density $f_{\Delta_{l1}X,\Delta_{l'1}Y}$, we use a bivariate kernel density estimator with a product kernel.

Replacing the classical errors-in-variables assumption of Assumption 2.1 by Assumption 2.3 with unknown error dependences or Assumption 2.4 with unknown transformations complicates the constructions of the bias-corrected estimators. In the former case, having nonclassical errors affect the choices of $\sigma^2_{\epsilon_X}$ and $\sigma^2_{\epsilon_Y}$, since $\sigma^2_{\epsilon_X}$ corresponds to the conditional variance of X given X^* rather than the marginal variance of $X - X^*$, whose elicitation requires knowledge of their magnitude relative to the variances of $X^* + \mu_{\epsilon_X}(X^*)$ and $Y^* + \mu_{\epsilon_X}(Y^*)$. Hence, it can be more challenging to elicit $\sigma^2_{\epsilon_X}$ and $\sigma^2_{\epsilon_Y}$ when we do not know $\mu_{\epsilon_X}(\cdot)$ or $\mu_{\epsilon_Y}(\cdot)$.

In case of Assumption 2.4, the lack of knowledge of $g_X(\cdot)$ and/or $g_Y(\cdot)$ leads to ambiguity about the unit of (X, Y) to be used for estimating the bias terms. The first-order bias terms expressed in terms of the joint distribution of (X^*, Y^*) are certainly invariant to $g_X(\cdot)$ and/or $g_Y(\cdot)$, but their estimators based on the observations of (X, Y) are not since the standard kernel density estimators are not invariant to nonlinear transformations of (X, Y). Hence, without knowing g_X and/or g_Y, it is hard to argue why a particular unit of measurement is used to construct the bias-corrected rank correlations.

4 Simulation Studies

We examine the approximation performance of the bias-corrected estimators for the Spearman and Kendall rank correlations, starting in this section with Monte Carlo studies in simulated data. Specifically, we draw (X^*, Y^*) from a joint normal distribution with mean zero, variance one, and correlation 0.5, and draw (ϵ_X, ϵ_Y) separately from independent normal distributions (Simulation 1) or mixtures of a point mass at zero and a normal distribution (Simulation 2). The measurement errors have mean zero and variances $\sigma^2_{\epsilon_X} = \sigma^2_{\epsilon_Y} = \sigma^2_{\epsilon}$ that are common across X and Y. We compare the performance of the bias-corrected estimator $\hat{\rho}^*_{bc}$ (or $\hat{\tau}^*_{bc}$ for the Kendall rank correlation) with correct specifications of the error variance terms, the "observed" estimator $\hat{\rho}$ (or $\hat{\tau}$) based on the sample correlation of (X, Y), and the infeasible "oracle" estimator $\hat{\rho}^*$ (or $\hat{\tau}^*$) defined by the sample correlation of (X^*, Y^*). We consider $\sigma^2_{\epsilon} = \{0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5\}$ to examine how the quality of the small error variance approximation varies with the degree of measurement error, and $n = \{100, 300, 1000\}$ to illustrate how it changes with sample size.

4.1 Spearman's ρ

Figure 1 illustrates the performance of the bias-corrected estimator for the Spearman rank correlation as given by equation (3.1) when the true data generating process is jointly normal and (ϵ_X, ϵ_Y) are drawn from independent standard normal distributions. The sub-panels (a) to (c) report the average bias (estimator's mean minus true correlation ρ^*), standard deviation, and the square root of the mean-squared error (MSE) of the bias-corrected, observed and oracle estimators across 1000 repetitions. Given the data generating process, the true Spearman correlation is $\rho^* \cong 0.485$. We construct the bias-corrected estimator under the assumption that the error variance σ_ϵ^2 is known.

Plots in panel (a) show that the oracle estimator $\hat{\rho}^*$ is unbiased. In contrast, the estimator $\hat{\rho}$ based on observed (X, Y) understates the true correlation, by about one third when the errors have half of the variance of the true observations ($\sigma_\epsilon^2 = 0.5$). The bias-corrected estimator $\hat{\rho}_{bc}^*$ reduces this bias, with the extent of bias reduction being inversely related to the error variance: the bias reduction amounts to more than 70 percent at $\sigma_\epsilon^2 = 0.05$ and to about 55 percent at $\sigma_\epsilon^2 = 0.5$. The average bias of the observed estimator $\hat{\rho}$ is stable across sample sizes, while the performance of the bias-corrected estimator $\hat{\rho}_{bc}^*$ improves slightly with sample size. However, these improvements are modest, and the bias-corrected estimator performs well also in small samples: much of the bias reduction is achieved with $n = 100$ or (not shown) fewer observations.

Panel (b) plots the average standard deviations of the three estimators across sample sizes. The oracle and observed estimator have similar standard deviations, while the standard deviation of the bias-corrected estimator is higher. In particular, the standard deviation is increasing in σ_ϵ^2 since the nonparametric bias estimators become more volatile as σ_ϵ^2 becomes larger. However, this difference in the standard deviation remains negligible if the error variance is small, or if the sample size is large.

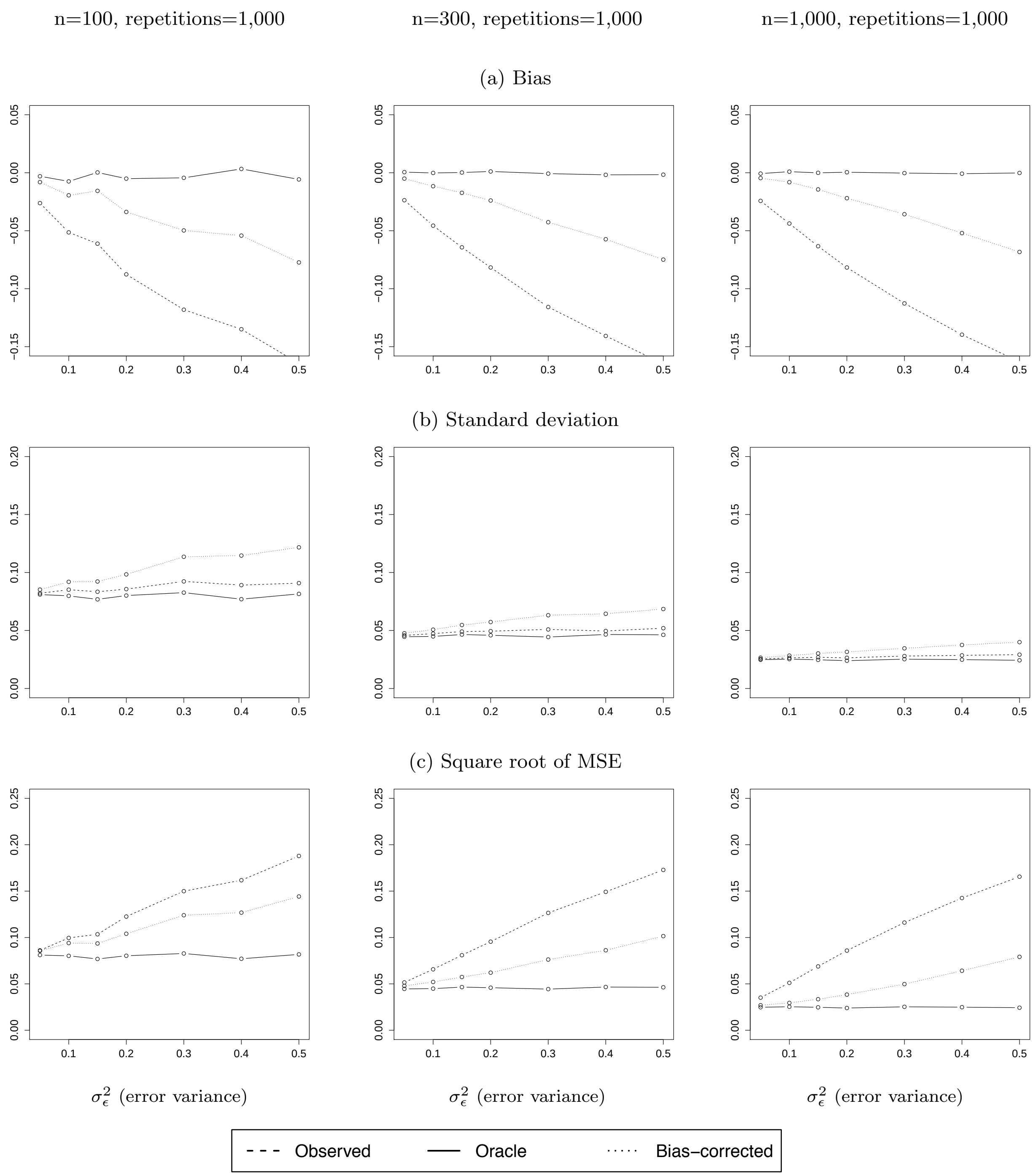
Panel (c) plots the square root of the MSE, which is a function of both the bias and the standard deviation of the estimators. The bias correction procedure leads to a substantial reduction in the MSE: while it increases slightly the standard deviation of the estimator, the bias reduction is more sizable, resulting in a net reduction in the MSE. In percentage terms, this reduction is larger for small error variances and for larger sample sizes. With n=1000 and $\sigma_\epsilon^2 \geq 0.2$, the bias-corrected estimator reduces the MSE by about 80 percent. However, the correction procedure reduces the MSE for any of the three sample sizes and all values of the error variance (the reduction is negligible at $n = 100$ and $\sigma_\epsilon^2 = 0.05$).

The observation that the correction procedure works well under different sample sizes and error variances is important, as it suggests that in settings where the distribution of true measurement and the error distributions are believed to be not far from Gaussian, it can be usefully applied even in small samples and over a wide range of error variances.

To assess how sensitive these results are to the distribution of the measurement errors we implement a second simulation, in which (ϵ_X, ϵ_Y) are instead drawn from a mixture of point mass at zero and independent normal distributions with equal mixture weights. This choice is motivated by the observation that in self-reported survey data, some interviewees do report truthfully and accurately (Bollinger, 1998; An and Hu, 2012).

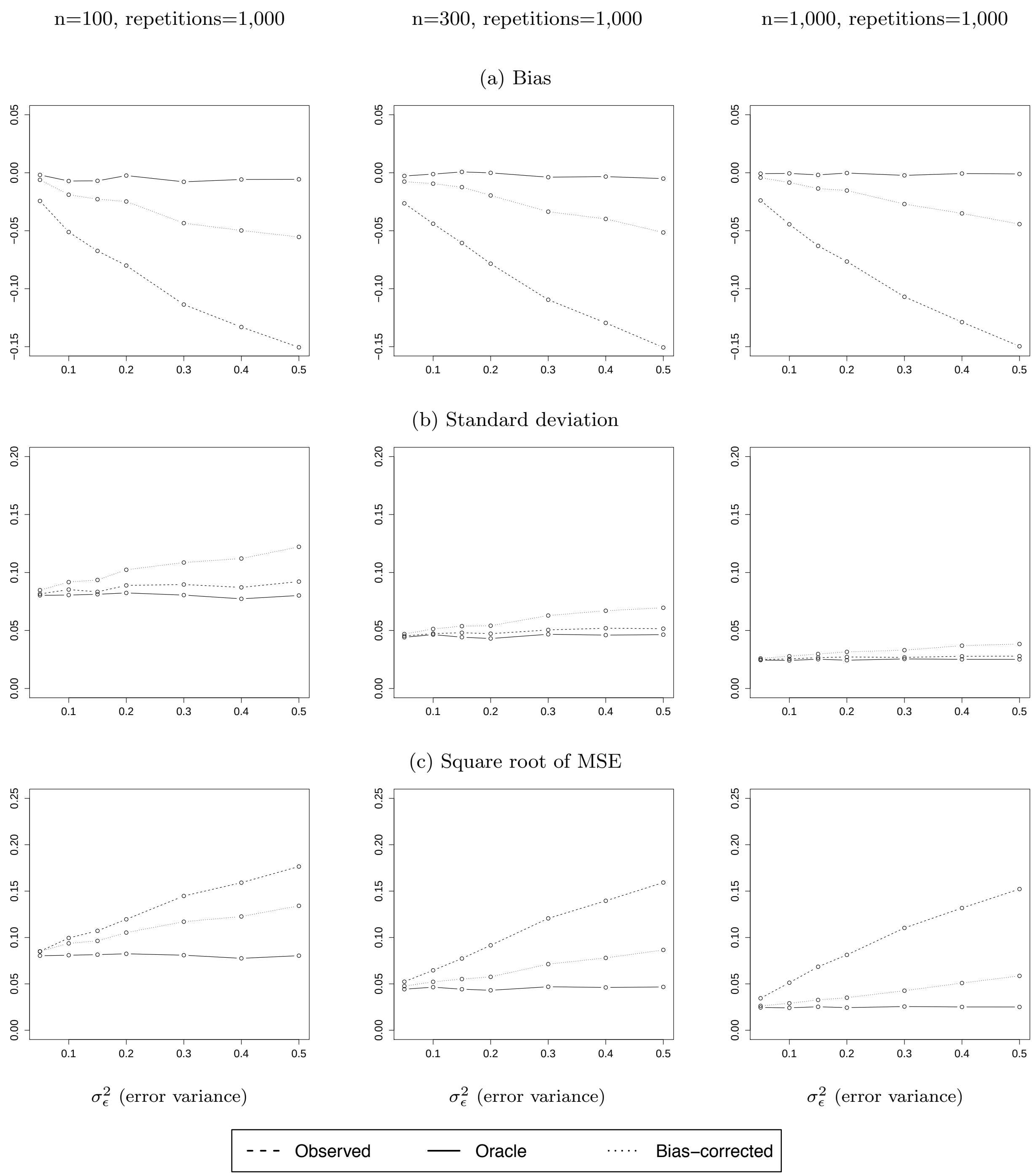
Figure 2 shows the results, again plotting the average bias, average standard deviation, and square root of the MSE of the bias-corrected, observed and oracle estimators in sub-panels (a) to (c). Comparing the results to Figure 1, the observed estimator is slightly less biased, but the bias correction procedure performs equally well. The correction procedure reduces the bias and slightly

Figure 1: Bias-corrected estimator of Spearman's ρ (Simulation 1)



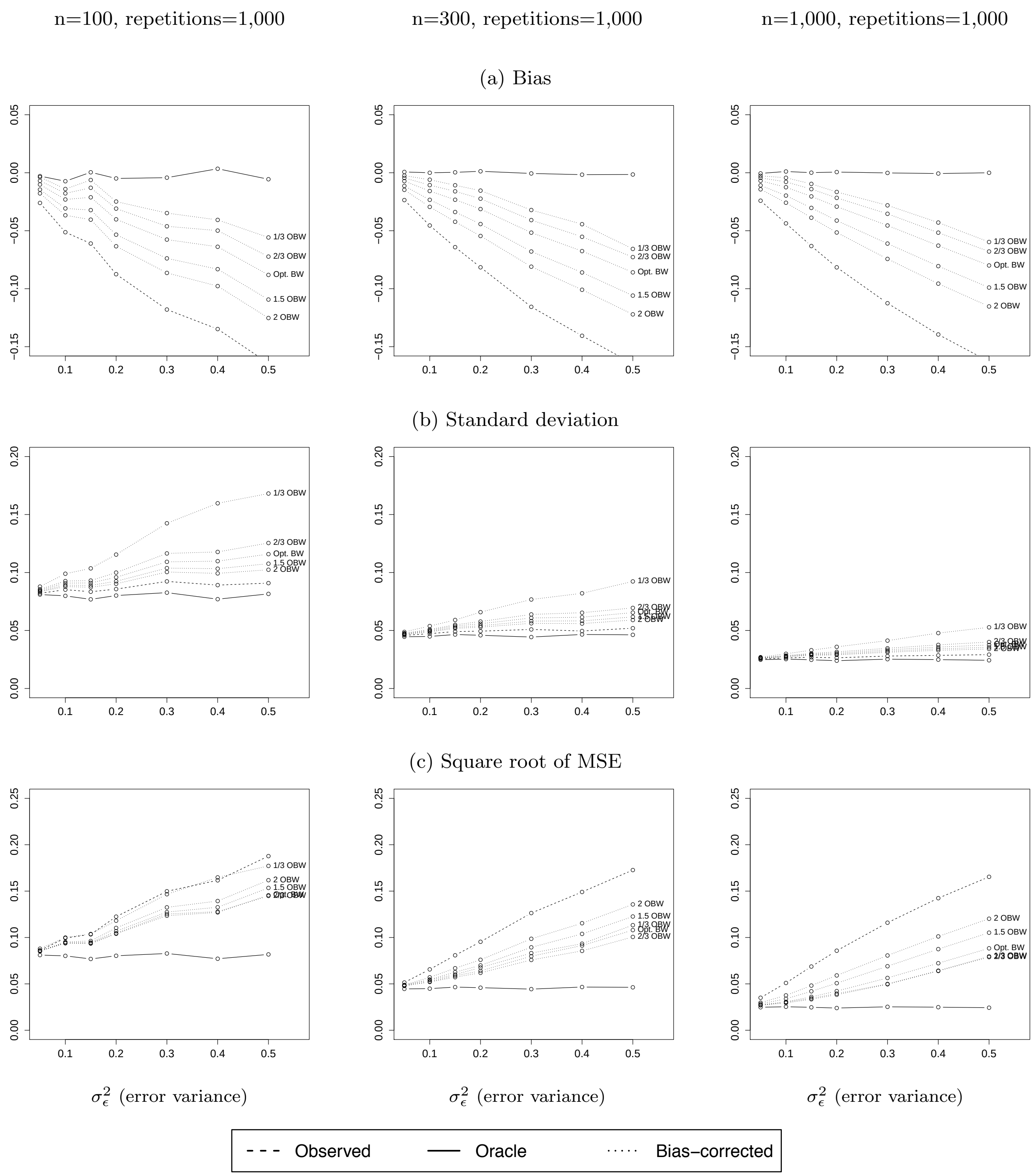
Note: (X^*, Y^*) from the joint normal distribution with mean zero and variance one with (Pearson) correlation 0.5. (ϵ_X, ϵ_Y) drawn from independent standard normal distributions, $\mu_{\epsilon_X} = \mu_{\epsilon_Y} = 0$, and $\sigma^2_{\epsilon_X} = \sigma^2_{\epsilon_Y} = \sigma^2_{\epsilon}$. Bandwidths for the kernel density estimation of f' are chosen by maximum likelihood cross validation as implemented by the "np" package (npudensbw) in the statistical software R, multiplied with the adjustment factor $n^{1/5-1/7}$ for derivative estimation.

Figure 2: Bias-corrected estimator of Spearman's ρ (Simulation 2)



Note: (X^*, Y^*) from the joint normal distribution with mean zero and variance one with (Pearson) correlation 0.5. (ϵ_X, ϵ_Y) are drawn from a mixture of point mass at zero and independent normal distributions with mean zero. Bandwidths for the kernel density estimation of f' are chosen by maximum likelihood cross validation as implemented by the "np" package (npudensbw) in the statistical software R, multiplied with the adjustment factor $n^{1/5-1/7}$ for derivative estimation.

Figure 3: Bandwidth selection for the bias-corrected estimator of Spearman's ρ (Simulation 1)



Note: (X^*, Y^*) from the joint normal distribution with mean zero and variance one with (Pearson) correlation 0.5. (ϵ_X, ϵ_Y) drawn from independent standard normal distributions, $\mu_{\epsilon_X} = \mu_{\epsilon_Y} = 0$, and $\sigma^2_{\epsilon_X} = \sigma^2_{\epsilon_Y} = \sigma^2_\epsilon$. Bandwidths for the kernel density estimation of f' are perturbed by multiplying a factor from the set $\{1/3, 2/3, 1, 1.5, 2\}$ to the bandwidths used in the simulation study of Figure 1.

increases the standard deviation, resulting in a net reduction in the MSE for any sample size and error variance σ_ϵ^2. With $n = 1000$ and sufficiently large error variance, the MSE reduction amounts to up to 85 percent.

Finally, we examine how sensitive these results are to the choice of bandwidth for the kernel density estimation. The choice of bandwidth is not straightforward, and the bandwidths chosen by maximum likelihood cross validation are not guaranteed to be optimal in any formal sense for our particular estimator. In Figure 3 we plot the results from Simulation 1, scaling the chosen cross-validation bandwidths by factors from the set $\{1/3, 2/3, 1, 1.5, 2\}$ to compute bias-corrected estimators that differ only in their choice of bandwidth. The figure illustrates that a smaller bandwidth may lead to a larger bias reduction, but at the cost of a larger standard deviation of $\hat{\rho}^*$ as estimates of the density derivatives become more noisy. The net effect on the MSE is substantial at $n = 100$, but the difference become more negligible at larger sample sizes. Overall, the performance of the bias-corrected estimator appears quite robust to the choice of bandwidth for its underlying kernel density estimation.

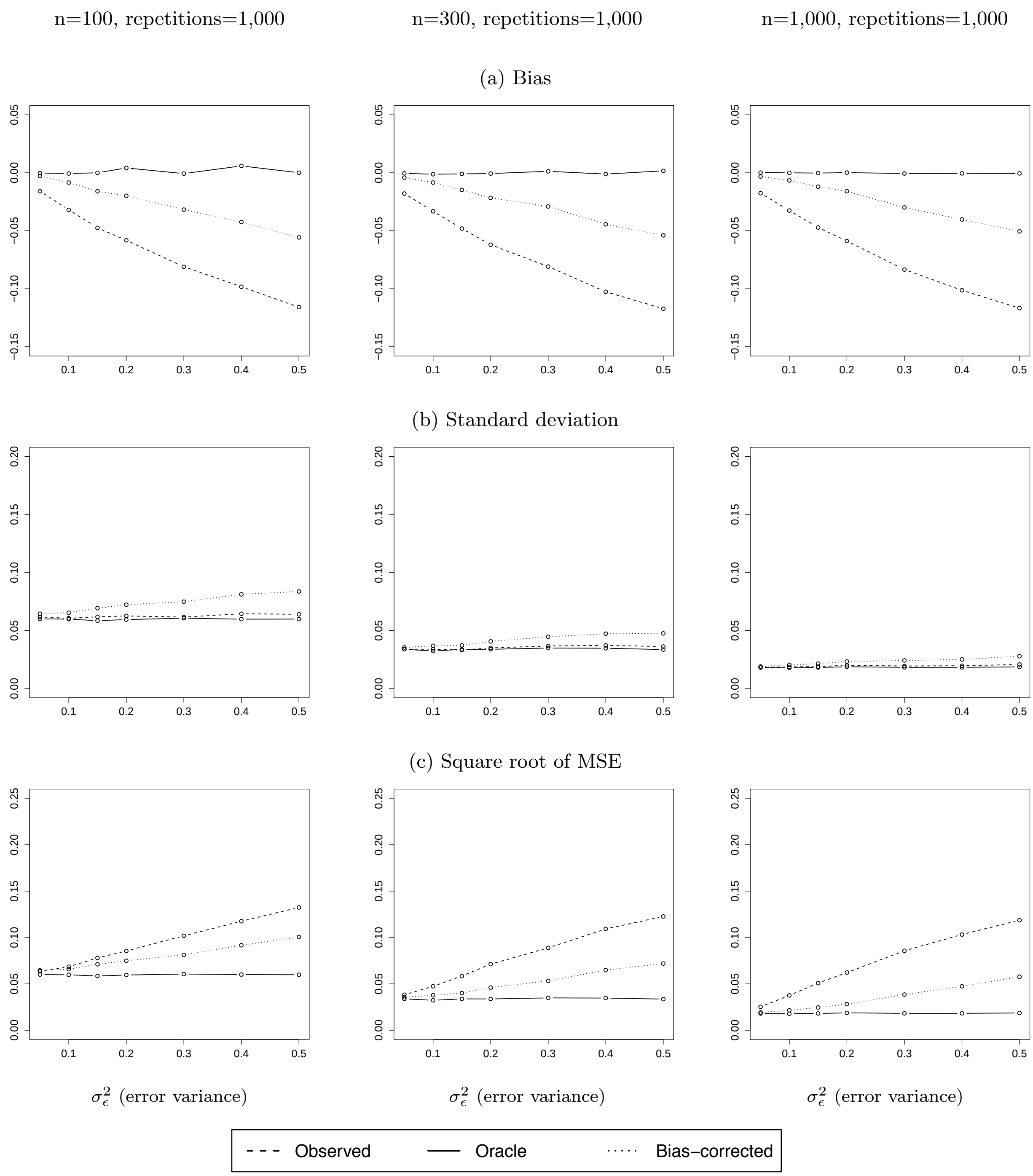
4.2 Kendall's τ

We repeat these exercises for Kendall's rank correlation (τ). We first examine the performance of the bias-corrected estimator as given by equation (3.2) in the setting identical to that of Figure 1. Given the data generating process, the true Kendall correlation is $\tau^* \cong 0.337$.

Panel (a) of Figure 4 plots the average bias of the three estimators. While the bias and the standard deviation in Kendall's $\hat{\tau}$ are both smaller in absolute value than in Spearman's $\hat{\rho}$, they are similar in relative terms. The bias of the observed estimator is again stable across sample sizes, while the performance of the bias-correction improves slightly with sample size. Panel (b) illustrates that the standard deviation of the bias-corrected estimator increases – both in absolute terms and relative to the observed estimator – in the error variance σ_ϵ^2, and decreases in sample size. Overall, we observe again a substantial net reduction in the square root of the MSE, as illustrated in Panel (c). Similarly to the case of Spearman's ρ, the bias correction procedure leads to substantially improved estimates of Kendall's τ, unless both sample and error variance are very small. The gains are again larger for larger error variances and larger sample sizes. At $n = 1000$, the bias correction reduces the bias by up to 75 percent, and the MSE by up to 80 percent, compared to the uncorrected estimator.

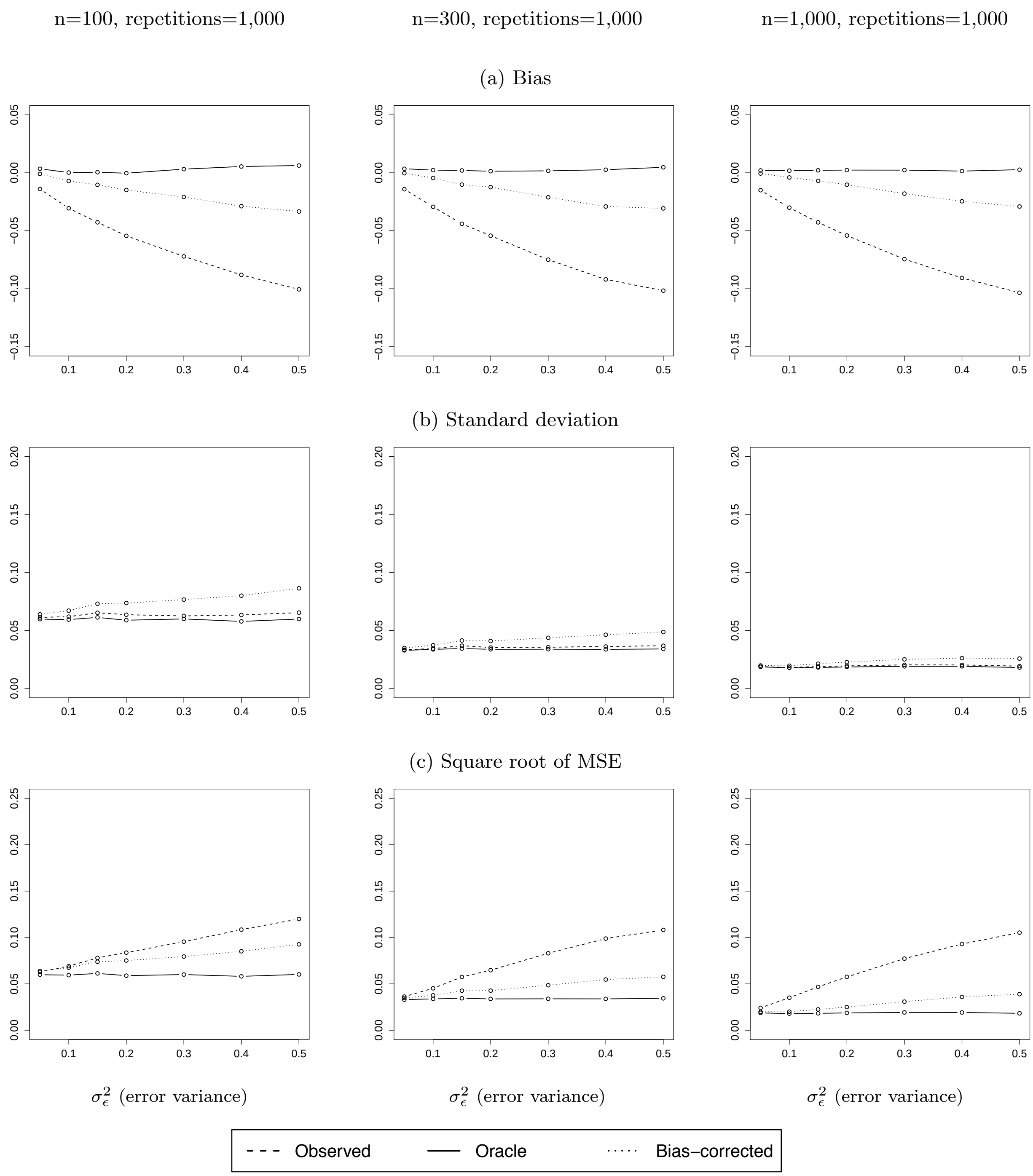
We repeat this exercise for our second simulation, in which (ϵ_X, ϵ_Y) are instead drawn from a mixture of point mass at zero and a normal distribution. Figure 5 shows the results, again plotting the bias, standard deviation, and MSE of the bias-corrected, observed and oracle estimator in panels (a) to (c). Compared to Figure 4, the results are nearly unchanged. The correction procedure reduces the bias, and slightly increases the standard deviation, with a net reduction in the MSE for any of the three sample sizes, and all considered values of the error variance σ_ϵ^2. At $n = 1000$, the

Figure 4: Bias-corrected estimator of Kendall's τ (Simulation 1)



Note: (X^*, Y^*) drawn from the joint normal distribution with mean zero and variance one with (Pearson) correlation 0.5, (ϵ_X, ϵ_Y) drawn from independent normal distributions with mean zero, $\mu_{\epsilon_X} = \mu_{\epsilon_Y} = 0$, and $\sigma^2_{\epsilon_X} = \sigma^2_{\epsilon_Y} = \sigma^2_\epsilon$. Bandwidths for the kernel density estimation of $f'_{\Delta X|\Delta Y \leq 0}$ and $f'_{\Delta Y|\Delta X \leq 0}$ are chosen by maximum likelihood cross validation as implemented by the "np" package (npudensbw) in the statistical software R, multiplied with the adjustment factor $n^{1/5-1/7}$ for derivative estimation.

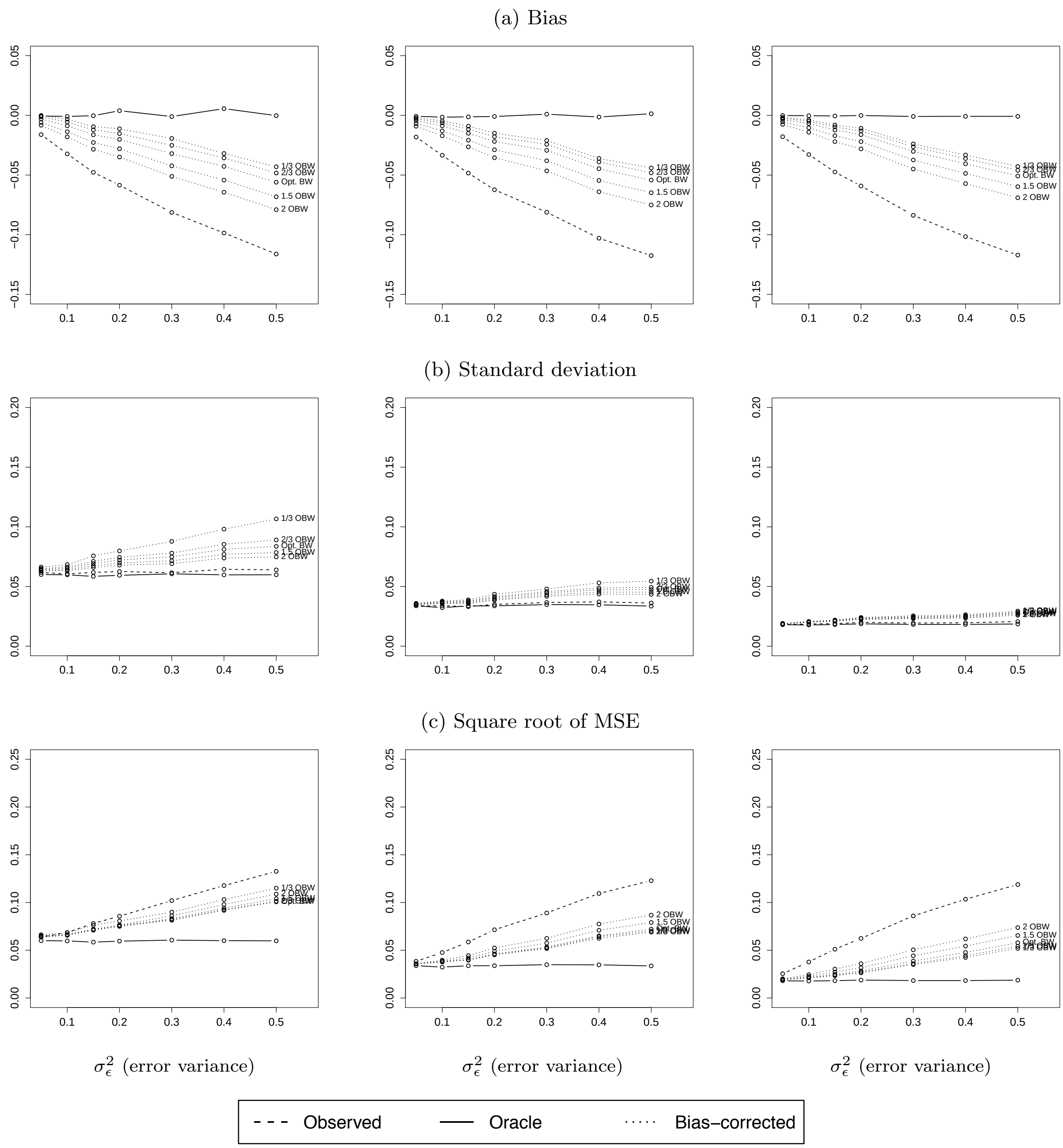
Figure 5: Bias-corrected estimator of Kendall's τ (Simulation 2)



Note: (X^*, Y^*) from the joint normal distribution with mean zero and variance one with (Pearson) correlation 0.5. (ϵ_X, ϵ_Y) are drawn from a mixture of point mass at zero and independent normal distributions with mean zero. Bandwidths for the kernel density estimation of $f'_{\Delta X|\Delta Y\leq 0}$ and $f'_{\Delta Y|\Delta X\leq 0}$ are chosen by maximum likelihood cross validation as implemented by the "np" package (npudensbw) in the statistical software R, multiplied with the adjustment factor $n^{1/5-1/7}$ for derivative estimation.

Figure 6: Bandwidth selection for the bias-corrected estimator of Kendall's τ (Simulation 1)

n=100, repetitions=1,000 | n=300, repetitions=1,000 | n=1,000, repetitions=1,000



Note: (X^*, Y^*) from the joint normal distribution with mean zero and variance one with (Pearson) correlation 0.5. (ϵ_X, ϵ_Y) drawn from independent standard normal distributions, $\mu_{\epsilon_X} = \mu_{\epsilon_Y} = 0$, and $\sigma^2_{\epsilon_X} = \sigma^2_{\epsilon_Y} = \sigma^2_\epsilon$. Bandwidths for the kernel density estimation of $f'_{\Delta X|\Delta Y \le 0}$ and $f'_{\Delta Y|\Delta X \le 0}$ are perturbed by multiplying a factor from the set $\{1/3, 2/3, 1, 1.5, 2\}$ to the bandwidths used in the simulation study of Figure 4.

bias correction reduces the bias by up to 80 percent and the MSE by up to 85 percent compared to the uncorrected estimator.

Finally, we examine how sensitive these results are to the choice of bandwidth for the kernel density estimation. In Figure 6 we plot results from Simulation 1, scaling the cross-validated bandwidth by the factors $\{1/3, 2/3, 1, 1.5, 2\}$. While a smaller bandwidth leads to a larger bias reduction, it also leads to a larger standard deviation, as estimates of the density derivatives, and therefore $\hat{\tau}^*$, become more noisy. The different choices for bandwidth have a somewhat more important effect at $n = 100$ than at larger sample sizes. Overall, the performance of the bias-corrected estimator appears quite robust to the choice of bandwidth for the underlying kernel density estimation.

5 An application to the estimation of intergenerational income mobility

We now examine the performance of the bias-corrected estimator in an empirical application, namely the estimation of intergenerational correlations in income. This literature – often referred to as the literature on intergenerational mobility – aims to measure dependence in lifetime or permanent incomes, but as these are rarely observed, estimations have to be based on short-run incomes instead. However, the traditional log-linear correlations have been found to be severely biased when estimated from short-run income data (Solon 1999, Mazumder 2016), while – consistent with Spearman's conjectures – rank correlations appear more robust (Dahl and DeLeire 2008, Chetty et al. 2014a, Nybom and Stuhler 2017, Chen, Ostrovsky, and Piraino 2017). As a consequence, rank-based measures have become the basis for much of the recent evidence on mobility differentials across time (Chetty et al., 2014b; Pekkarinen, Salvanes, and Sarvimäki, 2017), countries (Corak, Lindquist, and Mazumder, 2014; Bratberg et al., 2017), regions within countries (Chetty et al., 2014a; Chetty and Hendren forthcoming (I); Chetty and Hendren, forthcoming (II)) and groups (Mazumder and Davis 2018). In the absence of formal correction methods, measurement error remains however a central concern.

5.1 Data

We employ an administrative data set based on a 35 percent random sample of the Swedish population. We use as our main sample all males in the random sample that were born 1953-57 and use a multigenerational register to link these individuals to their biological fathers. Our income data come from official tax declaration files and span the years 1968-2007. We focus on a measure of total income, which includes income from all sources, such as labor and capital income, and income-based social transfers. For each son and father we construct one long-run income measure ("true" income) and one potentially mismeasured approximation of long-run income ("observed" income).

To decrease the amount of non-classical measurement error we focus on mid-career incomes (see Haider and Solon, 2006; Nybom and Stuhler, 2017; and our discussion in Section 2.4). To this end, we drop observations with fathers born before 1927. For the sons we approximate true income as (the log of) the average of annual incomes between ages 30-50, and observed income as (log) annual income at age 40. For the fathers, who we can only observe from age 40, we approximate true income as the (log) average of annual incomes between ages 40-50, and observed income as (log) annual income at age 45. We keep only those fathers and sons who had annual incomes of at least 20,000 SEK (approximately 2,500 USD) in each of the years used to compute their long-run average income.

Despite these restrictions, our data do not perfectly satisfy the classical errors-in-variables assumptions. Hence, remaining correlations between errors and true values, as well as differences in the higher-order moments of their distributions (e.g. skewness and kurtosis), may lead to deviations from the theoretical and simulation results above. We present summary statistics in Table A1 and a correlation matrix in Table A2 in the Appendix. Figure 7 plots histograms of the marginal distribution of (demeaned) log lifetime income and measurement errors. Our full sample consists of 36,135 father-son pairs. The estimates of Spearman's and Kendall's rank correlations using "true" incomes are $\rho^* = 0.26$ and $\tau^* = 0.18$, while the corresponding (biased) estimates based on observed annual incomes are $\rho = 0.21$ and $\tau = 0.14$. We treat this random sample as the true underlying population and assess the performances of the bias-corrected estimators based on repeatedly drawn random subsamples of size n varying over {100,300,1000}.

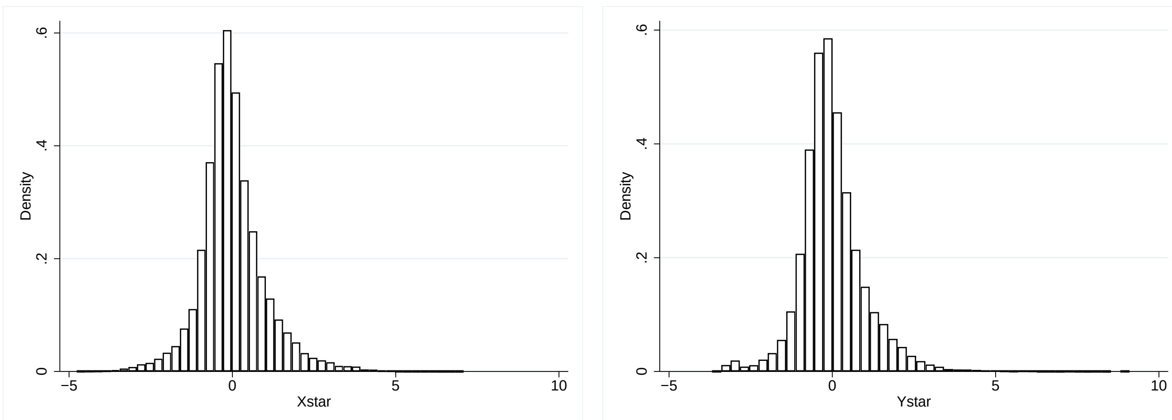
5.2 Spearman's ρ

Figure 8 reports the results for the bias-corrected estimator for Spearman's ρ for each subsample size, $n = \{100, 300, 1000\}$. Panels (a) to (c) report the average bias (estimator minus the true correlation ρ^*), standard deviation, and square root of the MSE of the bias-corrected, observed and oracle estimators across 500 repetitions. The x-axis indicates different specifications for the error variance $\sigma^2_{\epsilon_X}$ in our bias correction procedure, while the actual value $\sigma^2_{\epsilon_X} = 0.259$ (relative to the variance of X^* being standardized to one) is indicated by a vertical line in all graphs. The corresponding error variance $\sigma^2_{\epsilon_Y}$ is scaled proportionally. A comparison along the x-axis is therefore informative about the sensitivity of the bias correction to misspecification of the underlying error variance.

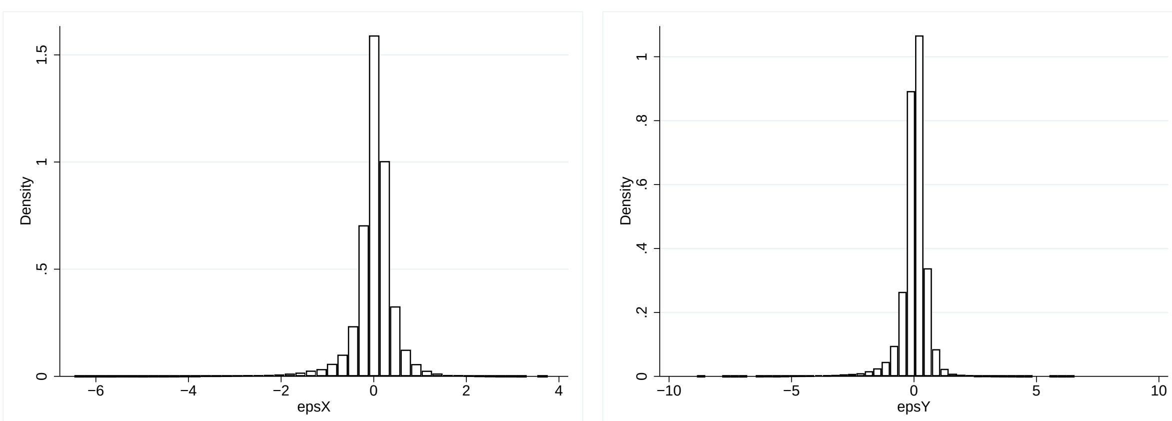
As in the simulations in Section 4, panel (a) shows that the application of the small error variance approximation can greatly reduce the bias in estimates of Spearman's ρ. In contrast to the simulations, at the true value $\sigma^2_{\epsilon_X} = 0.259$ the bias is slightly over- instead of under-corrected, demonstrating that in real-world applications the correction may not always be conservative. To avoid such over-correction, it may thus be advisable in similar applications to use estimates of $\sigma^2_{\epsilon_X}$ and $\sigma^2_{\epsilon_Y}$ that are believed to be slightly smaller than their true values. Panel (b) unsurprisingly

Figure 7: Marginal distributions

(a) Father's and son's log lifetime incomes



(b) Father's and son's measurement errors



shows that the bias correction leads to a modest increase in the standard deviation of the estimator, in particular for smaller sample sizes. Panel (c) demonstrates that the combined effect of the decrease in bias and the increase in standard deviation is a small net increase of the square-root of the MSE at small sample size $n = 100$. However, for larger sample size ($n = 1000$) this combined effect instead results in a substantial net reduction of roughly 60 percent as compared to the sample estimators without the bias correction. Figure 8 also suggests that the over-correction of the biases when choosing excessively large σ_ϵ^2 can increase the MSE compared to the uncorrected estimators.

In this application, the bias correction is therefore beneficial in small samples only if the bias and not the sampling variance of the estimator is the main concern. This pattern is in contrast to our Monte Carlo simulations, in which application of the bias correction led to a net reduction of the MSE even for small sample sizes. A possible explanation is that more observations are in the tails of the distribution in the empirical data compared to a normal distribution, in which f' are only imprecisely estimated in smaller samples. However, the bias correction does lead to a substantial MSE reduction at $n = 1000$ (see Figure 8) or larger sample sizes (not shown here), at which its application becomes advisable also when the MSE is the primary concern.[5]

5.3 Kendall's τ

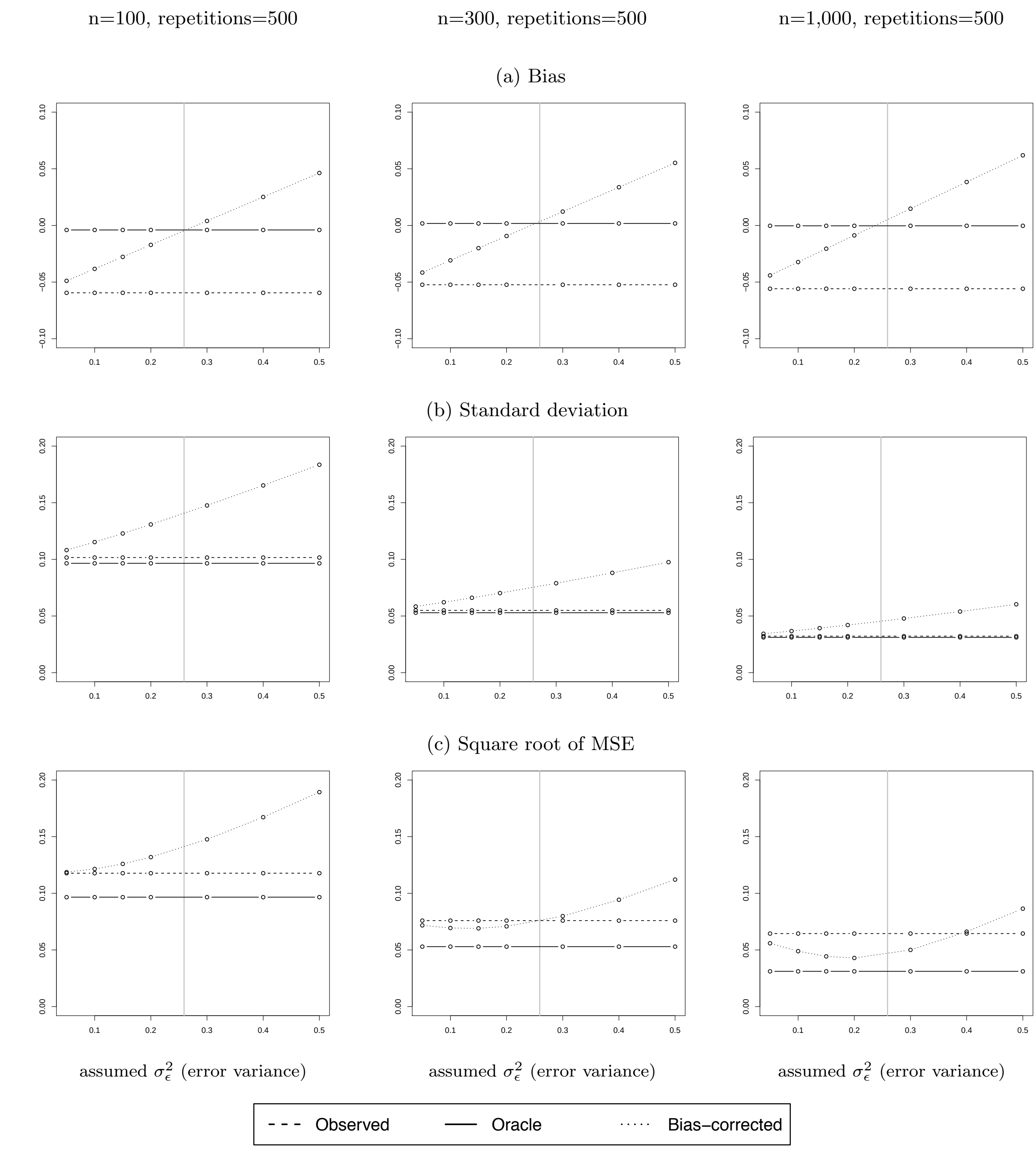
Figure 9 reports the corresponding results for the bias-corrected estimator of Kendall's τ, which are generally similar to those for Spearman's ρ. Panel (a) shows that the uncorrected estimator is attenuated towards zero. This result is in line with Proposition 2.3, as stochastic monotonicity – in this context, that the conditional distribution function of child income given parent income is increasing in parent income – is likely to hold in intergenerational relationships. Dardanoni, Fiorini, and Forcina (2012) test 149 tables of intergenerational *class* mobility covering 35 countries, and find that stochastic monotonicity can be rejected in only four cases. Delgado and Escanciano (2012) propose a distribution-free test of stochastic monotonicity, and fail to reject the null hypothesis of stochastic monotonicity in the association between child and parent income in the Panel Study of Income Dynamics. And Nybom and Stuhler (2017) plot the joint density of parent and child long-run incomes in a sample drawn from the same Swedish register data as we use here, which suggests that stochastic monotonicity holds also in the Swedish population.[6]

The correction procedure again reduces the average bias substantially. In this case, the remaining bias after implementation of the correction procedure is in fact close to zero. As with the Spearman rank correlation, the bias is marginally over-corrected for $n = 1000$. While the error

[5]Tentative evidence (not shown here) suggests that the non-zero correlation between measurement errors and true values has very little effect on the bias correction. However, the extent of (over-)correction depends on chosen bandwidths, and further work on how to optimally choose bandwidths in this context would be useful.

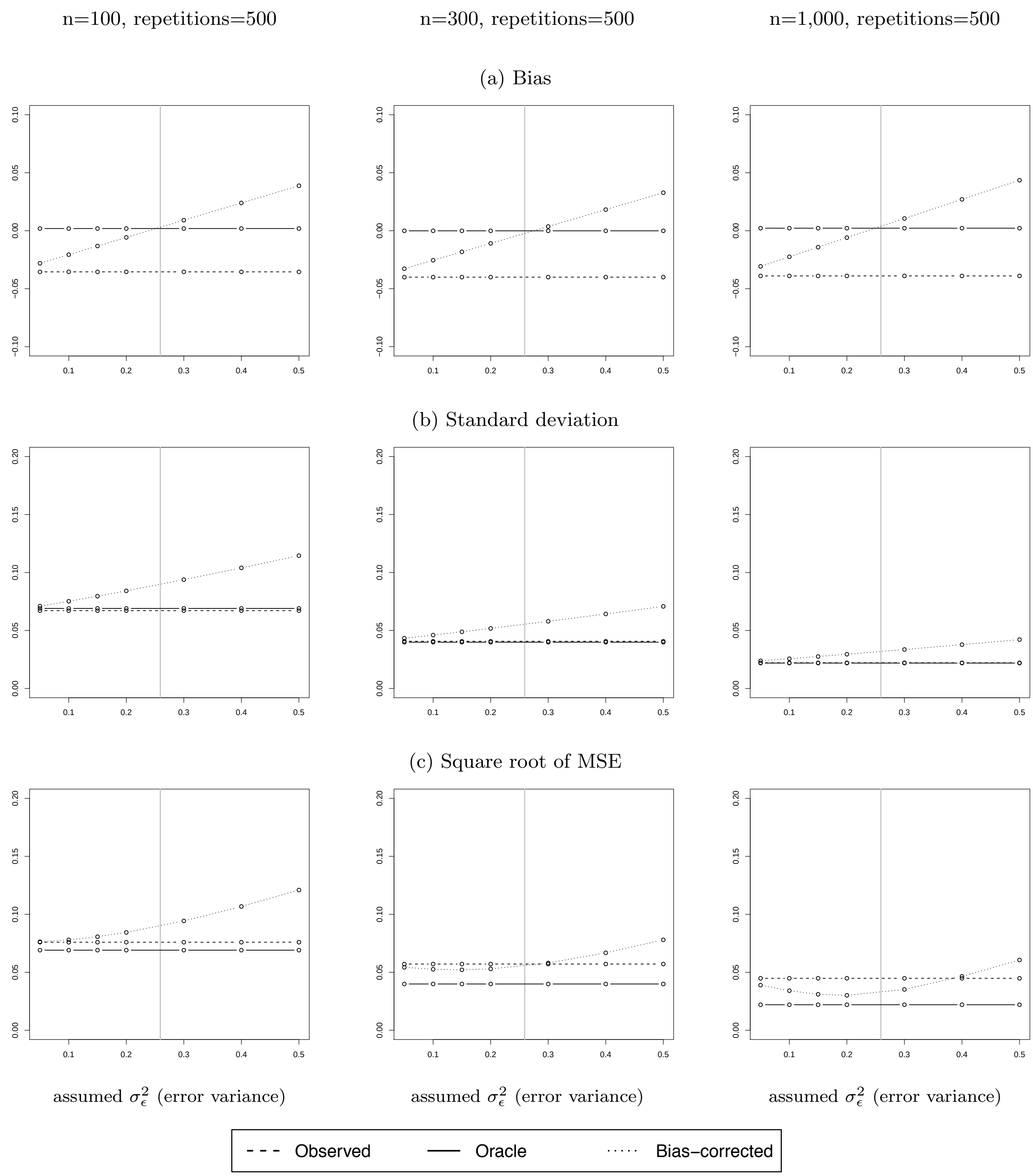
[6]However, stochastic monotonicity would fail to hold if child incomes are measured very early, as children from high-income parents are more likely to enter higher education, and therefore less likely to work full-time in their early 20s.

Figure 8: Bias-corrected estimator of Spearman's ρ of father's and son's lifetime incomes



Note: (X^*, Y^*) are drawn from our Swedish intergenerational income data standardized to mean zero and variance one, with $\sigma^2_{\epsilon_X} = 0.259$ and $\sigma^2_{\epsilon_Y} = 0.399$. Bandwidths for the kernel density estimation of f' are chosen by maximum likelihood cross validation as implemented by the "np" package (npudensbw) in the statistical software R, multiplied with the adjustment factor $n^{1/5-1/7}$ for derivative estimation.

Figure 9: Bias-corrected estimator of Kendall's τ of father's and son's lifetime incomes



Note: (X^*, Y^*) are drawn from our Swedish intergenerational income data standardized to mean zero and variance one, with $\sigma^2_{\epsilon_X} = 0.259$ and $\sigma^2_{\epsilon_Y} = 0.399$. Bandwidths for the kernel density estimation of $f'_{\Delta X|\Delta Y \leq 0}$ and $f'_{\Delta Y|\Delta X \leq 0}$ are chosen by maximum likelihood cross validation as implemented by the "np" package (npudensbw) in the statistical software R, multiplied with the adjustment factor $n^{1/5-1/7}$ for derivative estimation.

variance in our full sample equals $\sigma^2_{\epsilon_X} = 0.259$, the bias is minimized when assuming a slightly lower value for $\sigma^2_{\epsilon_X}$. In line with both the simulations and the Spearman rank correlation, panel (b) illustrates that the bias correction increases the standard deviation of the estimator, in particular for larger assumed error variances and smaller sample sizes. However, panel (c) again demonstrates that the combined effect of the bias reduction and the increased standard deviation is a net decrease in the square root of the MSE at $n = 1000$. But as for the Spearman rank correlation, there is a slight increase in the square root of the MSE at small sample size ($n = 100$) and when assuming excessively large values of σ^2_ϵ.

In contrast to the simulations, we thus find that in small samples the bias-corrected estimator for Kendall's τ is advantagous only if the bias itself is the main concern. If MSE reduction is the main yardstick then the bias correction makes little difference (on average) for $n = 300$ but leads to a sizable improvement at $n = 1000$, for a wide range of assumed values of σ^2_ϵ. At even larger sample sizes (not shown here), the MSE is reduced further.

6 Conclusion

In this paper, we analyzed the effects of measurement error on the estimation of Spearman's ρ and Kendall's τ. Due to the nonlinear nature of these rank correlation coefficients, analytical characterization of the effect of measurement errors is not as simple as for that of the Pearson correlation or regression coefficients. Using the approach of small error variance approximation first proposed by Chesher (1991), we derive the first-order bias terms proportional to the variances and covariances of the measurement errors. We also provide a simple sufficient condition for when measurement error leads to attenuation biases, which also leads to identification of the signs of the rank correlations. A notable feature of our analysis is that thanks to the transformation invariance of the rank correlations, our method can also accommodate some forms of non-classical measurement errors. This contrasts our analysis with previous applications of the small error variance approximation where classical measurement error type assumptions are commonly assumed.

We construct bias-corrected estimators for Spearman's ρ and Kendall's τ by estimating the first-order bias terms nonparametrically and subtracting them from the sample estimates. We find that the bias-corrected estimators can improve the mean squared errors (MSE) relative to the estimator that ignore the measurement errors. The MSE gains are substantial for a wide range of measurement error variances, even when those variances are as large as half of the variances of the true measurements. Our empirical application concerns the estimation of rank correlations of parent and child lifetime incomes, using administrative data on life-cycle incomes for Sweden. We find that the bias-corrected estimator improves the MSE by between 50 and 60 percent compared to the estimator with no bias correction (for $n = 1000$). For small sample sizes the MSE reduction is smaller, while for larger sample sizes it is larger, mainly due to the reduction of the variance of the non-parametric bias estimators.

Despite that Spearman's ρ and Kendall's τ have a long history in statistics and are also becoming increasingly common in applied economics, surprisingly little work exists on their relationship with measurement error. We hope that this new line of work will spark further research on the relationship between rank-based measures of statistical dependence and measurement error and equip applied researchers with appropriate tools and correction methods.

References

Abellan-Perpiñan, Jose-Maria, Han Bleichrodt, and Jose-Luis Pinto-Prades. 2009. "The predictive validity of prospect theory versus expected utility in health utility measurement." *Journal of Health Economics* 28 (6):1039–1047. URL http://EconPapers.repec.org/RePEc:eee:jhecon:v:28:y:2009:i:6:p:1039-1047.

An, Yonghong and Yingyao Hu. 2012. "Well-posedness of measurement error models for self-reported data." *Journal of Econometrics* 168 (2):259–269. URL https://ideas.repec.org/a/eee/econom/v168y2012i2p259-269.html.

An, Yonghong, Le Wang, and Ruli Xiao. 2017. "Your American Dream is Not Mine! A New Approach to Estimating Intergenerational Mobility Elasticities." Tech. rep., mimeo.

Bartolucci, Cristian, Francesco Devicienti, and Ignacio Monzón. 2015. "Identifying Sorting in Practice." IZA Discussion Papers 9411, Institute for the Study of Labor (IZA). URL https://ideas.repec.org/p/iza/izadps/dp9411.html.

Battistin, E. and A. Chesher. 2014. "Treatment Effect Estimation with Covariate Measurement Error." *Journal of Econometrics* 178:707–715.

Bhattacharya, Debopam and Bhashkar Mazumder. 2011. "A nonparametric analysis of black–white differences in intergenerational income mobility in the United States." *Quantitative Economics* 2 (3):335–379. URL http://ideas.repec.org/a/ecm/quante/v2y2011i3p335-379.html.

Böhlmark, A. and M.J. Lindquist. 2006. "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden." *Journal of Labor Economics* 24 (4):879–896.

Bollinger, Christopher R. 1998. "Measurement Error in the Current Population Survey: A Nonparametric Look." *Journal of Labor Economics* 16 (3):576–594. URL http://www.jstor.org/stable/10.1086/209899.

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage. 2017. "A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US." *The Scandinavian Journal of Economics* 119 (1):72–101. URL http://dx.doi.org/10.1111/sjoe.12197.

Chen, Wen-Hao, Yuri Ostrovsky, and Patrizio Piraino. 2017. "Lifecycle variation, errors-in-variables bias and nonlinearities in intergenerational income transmission: new evidence from Canada." *Labour Economics* 44 (C):1–12. URL https://ideas.repec.org/a/eee/labeco/v44y2017icp1-12.html.

Chesher, Andrew. 1991. “The Effect of Measurement Error.” *Biometrika* 78 (3):451–462.

———. 2017. “Understanding the Effect of Measurement Error on Quantile Regressions.” *cemmap working paper* (19/17).

Chesher, Andrew and Christian Schluter. 2002. “Welfare Measurement and Measurement Error.” *Review of Economic Studies* 69 (2):357–378.

Chesher, Andrew and J.M.C. Santos Silva. 2002. “Taste Variation in Discrete Choice Models.” *Review of Economic Studies* 69:147–168.

Chetty, Raj and Nathaniel Hendren. forthcoming (I). “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects.” *Quarterly Journal of Economics* .

———. forthcoming (II). “The Impact of Neighborhoods on Intergenerational Mobility II: County-Level Estimates.” *Quarterly Journal of Economics* .

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014a. “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *Quarterly Journal of Economics* 129 (4):1553–1623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014b. “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility.” *American Economics Review: Papers and Proceedings* 104 (5):141–47. URL http://EconPapers.repec.org/RePEc:nbr:nberwo:19844.

Corak, Miles, Matthew J. Lindquist, and Bhashkar Mazumder. 2014. “A Comparison of Upward and Downward Intergenerational Mobility in Canada, Sweden and the United States.” *Labour Economics* URL http://www.sciencedirect.com/science/article/pii/S0927537114000530.

Dahl, M. and T. DeLeire. 2008. “The Association Between Children’s Earnings And Fathers’ Lifetime Earnings: Estimates Using Administrative Data.” Discussion Paper No. 1342-08, Institute for Research on Poverty, University of Wisconsin-Madison.

Daniels, H. E. 1950. “Rank Correlation and Population Models.” *Journal of the Royal Statistical Society, Series B* 12 (2):171–191.

Dardanoni, Valentino, Mario Fiorini, and Antonio Forcina. 2012. “Stochastic monotonicity in intergenerational mobility tables.” *Journal of Applied Econometrics* 27 (1):85–107. URL https://ideas.repec.org/a/wly/japmet/v27y2012i1p85-107.html.

Delgado, Miguel A. and Juan Carlos Escanciano. 2012. “Distribution-free tests of stochastic monotonicity.” *Journal of Econometrics* 170 (1):68 – 75. URL http://www.sciencedirect.com/science/article/pii/S0304407612000723.

Dohmen, Thomas and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101 (2):556–590. URL https://ideas.repec.org/a/aea/aecrev/v101y2011i2p556-90.html.

Gregg, Paul, Lindsey Macmillan, and Claudia Vittori. 2017. "Moving Towards Estimating Sons' Lifetime Intergenerational Economic Mobility in the UK." *Oxford Bulletin of Economics and Statistics* 79 (1):79–100. URL https://ideas.repec.org/a/bla/obuest/v79y2017i1p79-100.html.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2017. "Identifying Equilibrium Models of Labor Market Sorting." *Econometrica* 85:29–65. URL https://ideas.repec.org/a/wly/emetrp/v85y2017ip29-65.html.

Haider, Steven and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96 (4):1308–1320.

Hall, Peter, Jeff Racine, and Qi Li. 2004. "Cross-Validation and the Estimation of Conditional Probability Densities." *Journal of the American Statistical Association* 99:1015–1026. URL http://ideas.repec.org/a/bes/jnlasa/v99y2004p1015-1026.html.

Jenkins, S. 1987. "Snapshots versus Movies: 'Lifecycle biases' and the Estimation of Intergenerational Earnings Inheritance." *European Economic Review* 31 (5):1149–1158.

Kendall, M.G. 1938. "A New Measure of Rank Correlation." *Biometrika* 30:81–93.

Klein, Tobias J. 2010. "Heterogeneous Treatment Effects: Instrumental Variables without Monotonicity?" *Journal of Econometrics* 155 (2):99–116.

Lee., Sokbae, Oliver Linton, and Yoon-Jae Whang. 2009. "Testing for Stochastic Monotonicity." *Econometrica* 77 (2):585–602.

Lehmann, E.L. 1966. "Some Concepts of Dependence." *The Annals of Mathematical Statistics* 37:1137–1153.

Mazumder, Bhashkar. 2016. *Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitations of Tax Data.* , *Book Series: Research in Labor Economics*, vol. 43, chap. 3. Emerald Group Publishing Limited, 83–129. URL http://www.emeraldinsight.com/doi/abs/10.1108/S0147-912120160000043012.

Mazumder, Bhashkar and Jonathan Davis. 2018. "Racial and Ethnic Differences in the Geography of Intergenerational Mobility." Tech. rep., mimeo.

Nelsen, Roger B. 2006. *An Introduction to Copulas.* Springer, 2nd ed.

Nybom, Martin and Jan Stuhler. 2016. “Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation.” *Journal of Human Resources* 51 (1).

———. 2017. “Biases in Standard Measures of Intergenerational Income Dependence.” *Journal of Human Resources* 52 (3):800–825. URL https://ideas.repec.org/a/uwp/jhriss/v52y2017i3p800-825.html.

Pekkarinen, Tuomas, Kjell G. Salvanes, and Matti Sarvimäki. 2017. “The Evolution of Social Mobility: Norway during the Twentieth Century.” *The Scandinavian Journal of Economics* 119 (1):5–33. URL http://dx.doi.org/10.1111/sjoe.12205.

Solon, Gary. 1999. “Intergenerational Mobility in the Labor Market.” In *Handbook of Labor Economics*, vol. 3A, edited by O. Ashenfelter and D. Card, chap. 29. Elsevier, 1761–1800.

Spearman, C. 1904. “The Proof and Measurement of Association between Two Things.” *American Journal of Psychology* 15:72–101.

Appendix

A1 Proofs

Proposition 2.1 follows as a corollary of the next lemma.

Lemma A1.1. *Under Assumptions 2.1 and 2.2,*

$$
\begin{aligned}
\Pr(X_k > X_l, Y_{k'} > Y_{l'}) \quad = \quad & \Pr(X_k^* > X_l^*, Y_k^* > Y_l^*) + \frac{1}{2} f'_{\Delta_{lk} X^* | \Delta_{l'k'} Y^* < 0}(0) \sigma_{\epsilon_X^2} + \frac{1}{2} f'_{\Delta_{l'k'} Y^* | \Delta_{lk} X^* < 0}(0) \sigma_{\epsilon_Y^2} \\
& + [1_{k=k'} - 1_{k=l'} - 1_{l=k'} + 1_{l=l'}] f_{\Delta_{lk} X^*, \Delta_{l'k'} Y^*}(0,0) \sigma_{\epsilon_X \epsilon_Y} \sigma_{\epsilon_X} \sigma_{\epsilon_Y} \\
& + O((\sigma_{\epsilon_X} + \sigma_{\epsilon_Y})^3).
\end{aligned}
$$

Proof. The second-order Taylor expansion of $\Pr(X_k > X_l, Y_{k'} > Y_{l'})$ at $(\sigma_{\epsilon_X}, \sigma_{\epsilon_Y}) = (0,0)$ is

$$
\begin{aligned}
& \Pr(X_k > X_l, Y_{k'} > Y_{l'}) \\
= \quad & \Pr(X_k^* > X_l^*, Y_{k'}^* > Y_{l'}^*) + \left. \frac{\partial \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_X}} \right|_{0,0} \sigma_{\epsilon_X} + \left. \frac{\partial \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_Y}} \right|_{0,0} \sigma_{\epsilon_Y} \\
& + \frac{1}{2} \left. \frac{\partial^2 \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_X}^2} \right|_{0,0} \sigma_{\epsilon_X}^2 + \frac{1}{2} \left. \frac{\partial^2 \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_Y}^2} \right|_{0,0} \sigma_{\epsilon_Y}^2, \\
& + \left. \frac{\partial^2 \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_X} \partial \sigma_{\epsilon_Y}} \right|_{0,0} \sigma_{\epsilon_X} \sigma_{\epsilon_Y} + O\left((\sigma_{\epsilon_X} + \sigma_{\epsilon_Y})^3 \right).
\end{aligned}
$$

We compute each derivative term. First,

$$
\begin{aligned}
\left. \frac{\partial \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_X}} \right|_{0,0} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left. \frac{\partial}{\partial a} F_{\Delta_{lk} X^*, \Delta_{l'k'} Y^*}(a,b) \right|_{a=b=0} \Delta_{kl} \epsilon_X dG_{\Delta_{kl}\epsilon_X, \Delta_{k'l'}\epsilon_Y} \\
&= \left. \frac{\partial}{\partial a} F_{\Delta_{lk} X^*, \Delta_{l'k'} Y^*}(a,b) \right|_{a=b=0} \cdot E(\Delta_{kl}\epsilon_X) = 0,
\end{aligned}
$$

as $E(\Delta_{kl}\epsilon_X) = 0$. Similarly, we can show $\left. \frac{\partial \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_Y}} \right|_{0,0} = 0$.

As for the second derivatives,

$$
\begin{aligned}
\left. \frac{\partial^2 \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_X}^2} \right|_{0,0} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left. \frac{\partial^2}{\partial a^2} F_{\Delta_{lk} X^*, \Delta_{l'k'} Y^*}(a,b) \right|_{a=b=0} (\Delta_{kl}\epsilon_X)^2 \, dG_{\Delta_{kl}\epsilon_X, \Delta_{k'l'}\epsilon_Y} \\
&= 2 \left. \frac{\partial}{\partial a} f_{\Delta_{lk} X^* | \Delta_{l'k'} Y^* \le 0}(a) \right|_{a=0} F_{\Delta_{l'k'} Y^*}(0) \\
&= \left. \frac{\partial}{\partial a} f_{\Delta_{lk} X^* | \Delta_{l'k'} Y^* \le 0}(a) \right|_{a=0},
\end{aligned}
$$

where the second line follows from $Var(\Delta_{kl}\epsilon_X) = 2$ and

$$
F_{\Delta_{lk} X^*, \Delta_{l'k'} Y^*}(a,b) = \Pr(\Delta_{lk} X^* \le a | \Delta_{l'k'} Y^* \le b) F_{\Delta_{l'k'} Y^*}(b),
$$

and the third line follows from $F_{\Delta_{l'k'}Y^*}(0) = \frac{1}{2}$. Similarly, it holds

$$\left.\frac{\partial^2 \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma^2_{\epsilon_X}}\right|_{0,0} = \left.\frac{\partial}{\partial b} f_{\Delta_{l'k'}Y^*|\Delta_{lk}X^* \leq 0}(b)\right|_{b=0}.$$

For the cross derivative term, we have

$$\begin{aligned}\left.\frac{\partial^2 \Pr(X_k > X_l, Y_{k'} > Y_{l'})}{\partial \sigma_{\epsilon_Y} \partial \sigma_{\epsilon_X}}\right|_{0,0} &= f_{\Delta_{lk}X^* \Delta_{l'k'}Y^*}(0,0) \cdot Cov(\Delta_{kl}\epsilon_X, \Delta_{k'l'}\epsilon_Y) \\ &= [1_{k=k'} - 1_{k=l'} - 1_{l=k'} + 1_{l=l'}]\sigma_{\epsilon_X \epsilon_Y} f_{\Delta_{lk}X^*, \Delta_{l'k'}Y^*}(0,0).\end{aligned}$$

With all these combined, we obtain the conclusion of this lemma.

□

Proof. [Derivation of equation (2.6)] Note

$$\begin{aligned}
& \frac{1}{2}\tau \\
= & \left[\Pr(X_1 > X_2, Y_1 > Y_2) - \Pr(X_1 < X_2, Y_1 > Y_2)\right] \\
= & \left[\begin{array}{l} \Pr(X_1 > X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*) + \Pr(X_1 > X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* < Y_2^*) \\ + \Pr(X_1 > X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*) + \Pr(X_1 > X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* < Y_2^*) \end{array}\right] \\
- & \left[\begin{array}{l} \Pr(X_1 < X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*) + \Pr(X_1 < X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* < Y_2^*) \\ + \Pr(X_1 < X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*) + \Pr(X_1 < X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* < Y_2^*) \end{array}\right] \\
= & \left[\begin{array}{l} \Pr(X_1 > X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*) + \Pr(X_1 < X_2, Y_1 < Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*) \\ - \Pr(X_1 < X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*) - \Pr(X_1 > X_2, Y_1 < Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*) \end{array}\right] \qquad \text{(A1)} \\
+ & \left[\begin{array}{l} \Pr(X_1 > X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*) + \Pr(X_1 < X_2, Y_1 < Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*) \\ - \Pr(X_1 > X_2, Y_1 < Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*) - \Pr(X_1 < X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*) \end{array}\right]
\end{aligned}$$

where the second equality follows from the law of total probability, and the third line follows from the exchangeability of (X_1, X_1^*, Y_1, Y_1^*) and (X_2, X_2^*, Y_2, Y_2^*), i.e., we can permute the subscripts of the random variables in $\Pr(\cdot)$. The first term in the first squared bracketed term in equation (A1) can be written as

$$
\begin{aligned}
& \Pr\left(X_1 > X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_1^*}\int_{-\infty}^{y_1^*} \Pr\left(X_1 > X_2, Y_1 > Y_2 | x_1^*, x_2^*, y_1^*, y_2^*\right) dF_{X^*,Y^*}\left(x_2^*, y_2^*\right) dF_{X^*,Y^*}\left(x_1^*, y_1^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_1^*}\int_{-\infty}^{y_1^*} \Pr\left(X_1 > X_2 | x_1^*, x_2^*, y_1^*, y_2^*\right) \Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) dF_{X^*|Y^*}\left(x_2^* | y_2^*\right) dF_{Y^*}\left(y_2^*\right) \\
& \cdot dF_{X^*|Y^*}\left(x_1^* | y_1^*\right) dF_{Y^*}\left(y_1^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{x_1^*} \Pr\left(X_1 > X_2 | x_1^*, x_2^*, y_1^*, y_2^*\right) dF_{X^*|Y^*}\left(x_2^* | y_2^*\right) dF_{X^*|Y^*}\left(x_1^* | y_1^*\right)\right] \\
& \cdot \Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \Pr\left(X_1 > X_2, X_1^* > X_2^* | y_1^*, y_2^*\right) \Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right),
\end{aligned}
$$

where the second equality follows by the conditional independence $(X_1, X_2) \perp (Y_1, Y_2) \,|\, (X_1^*, X_2^*, Y_1^*, Y_2^*)$ and $(Y_1 \perp Y_2) \perp (X_1^*, X_2^*) \,|\, (Y_1^*, Y_2^*)$ implied from the independence of the classical measurement errors $\epsilon_X \perp \epsilon_Y$, the third equality follows from interchanging the order of integrations. Similarly, for each of the terms in the first squared brackets, we can write

$$
\begin{aligned}
& \Pr\left(X_1 < X_2, Y_1 < Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \Pr\left(X_1 < X_2, X_1^* > X_2^* | y_1^*, y_2^*\right) \Pr\left(Y_1 < Y_2 | y_1^*, y_2^*\right) dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right)
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left(X_1 < X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \Pr\left(X_1 < X_2, X_1^* > X_2^* | y_1^*, y_2^*\right) \Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right),
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left(X_1 > X_2, Y_1 < Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \Pr\left(X_1 > X_2, X_1^* > X_2^* | y_1^*, y_2^*\right) \Pr\left(Y_1 < Y_2 | y_1^*, y_2^*\right) dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right).
\end{aligned}
$$

Hence, the first square-bracketed terms are reduced to

$$
\begin{aligned}
& \left[\begin{array}{c} \Pr\left(X_1 > X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) + \Pr\left(X_1 < X_2, Y_1 < Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) \\ - \Pr\left(X_1 < X_2, Y_1 > Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) - \Pr\left(X_1 > X_2, Y_1 < Y_2, X_1^* > X_2^*, Y_1^* > Y_2^*\right) \end{array}\right] \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \Pr\left(X_1 > X_2, X_1^* > X_2^* | y_1^*, y_2^*\right)\left[\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - \Pr\left(Y_1 < Y_2 | y_1^*, y_2^*\right)\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right) \\
- & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \Pr\left(X_1 < X_2, X_1^* > X_2^* | y_1^*, y_2^*\right)\left[\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - \Pr\left(Y_1 < Y_2 | y_1^*, y_2^*\right)\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right) \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*}\left[\Pr\left(X_1 > X_2, X_1^* > X_2^* | y_1^*, y_2^*\right) - \Pr\left(X_1 < X_2, X_1^* > X_2^* | y_1^*, y_2^*\right)\right] \qquad \text{(A2)} \\
& \cdot \left[2\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - 1\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right).
\end{aligned}
$$

By repeating a similar algebra, the second square-bracketed terms in (A1) can be written as

$$\begin{aligned}
&\left[\begin{array}{c} \Pr\left(X_1 > X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*\right) + \Pr\left(X_1 < X_2, Y_1 < Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*\right) \\ - \Pr\left(X_1 > X_2, Y_1 < Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*\right) - \Pr\left(X_1 < X_2, Y_1 > Y_2, X_1^* < X_2^*, Y_1^* > Y_2^*\right) \end{array}\right] \\
= \;& \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \left[\Pr\left(X_1 > X_2, X_1^* < X_2^* | y_1^*, y_2^*\right) - \Pr\left(X_1 < X_2, X_1^* < X_2^* | y_1^*, y_2^*\right)\right] \\
& \cdot \left[2\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - 1\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right).
\end{aligned} \tag{A3}$$

Summing up (A2) and (A3) leads to

$$\begin{aligned}
& \frac{1}{2}\tau \\
= \;& \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \left[\Pr\left(X_1 > X_2 | y_1^*, y_2^*\right) - \Pr\left(X_1 < X_2 | y_1^*, y_2^*\right)\right]\left[2\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - 1\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right) \\
= \;& \int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} \left[2\Pr\left(X_1 > X_2 | y_1^*, y_2^*\right) - 1\right]\left[2\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - 1\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right).
\end{aligned}$$

Hence, with $b(y_1^*, y_2^*) = 2\Pr\left(Y_1 > Y_2 | y_1^*, y_2^*\right) - 1$, we obtain

$$\tau = 2\int_{-\infty}^{\infty}\int_{-\infty}^{y_1^*} b(y_1^*, y_2^*)\left[2\Pr\left(X_1 > X_2 | y_1^*, y_2^*\right) - 1\right] dF_{Y^*}\left(y_2^*\right) dF_{Y^*}\left(y_1^*\right).$$

□

Proof of Proposition 2.3. We prove the claim for the positive τ^* case only, since a symmetric argument applies to the negative τ^* case.

First, we shall show that if (X^*, Y^*) has stochastically increasing positive dependence, $\Pr(X_1^* > X_2^* | y_1^*, y_2^*) \geq 1/2$ and $\Pr(X_1 > X_2 | y_1^*, y_2^*) \geq 1/2$ hold for $y_1^* \geq y_2^*$. For $\delta > 0$, it holds

$$\begin{aligned}
& \Pr(X_1^* > X_2^* | Y_1^* = y^* + \delta, Y_2^* = y^*) \\
= \;& \textstyle\int_{-\infty}^{\infty} \Pr(X_1^* > x | Y_1^* = y^* + \delta, Y_2^* = y^*, X_2^* = x) dF_{X_2^*|Y_2^*}(x|y^*) \\
= \;& \textstyle\int_{-\infty}^{\infty} \Pr(X_1^* > x | Y_1^* = y^* + \delta) dF_{X_2^*|Y_2^*}(x|y^*) \\
\geq \;& \textstyle\int_{-\infty}^{\infty} \Pr(X_1^* > x | Y_1^* = y^*) dF_{X_2^*|Y_2^*}(x|y^*) \\
= \;& \textstyle\int_{-\infty}^{\infty} \Pr(X_1^* > x | Y_1^* = Y_2^* = y^*, X_2^* = x) dF_{X_2^*|Y_2^*}(x|y^*) \\
= \;& \Pr(X_1^* > X_2^* | Y_1^* = Y_2^* = y^*), \qquad = 1/2
\end{aligned}$$

where the third line follows by the stochastically increasing dependence and the last equality follows since (X_1^*, Y_1^*) and (X_2^*, Y_2^*) are independently and identically distributed.

To show $\Pr(X_1 > X_2|y_1^*, y_2^*) \geq 1/2$, let $G_{\Delta_{21}\epsilon_X}(\cdot)$ be the distribution function of $\Delta_{21}\epsilon_X$. Consider

$$
\begin{aligned}
&\Pr(X_1 > X_2|y_1^*, y_2^*)\\
=&\int_{-\infty}^{\infty} \Pr(X_1^* > X_2^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)dG_{\Delta_{21}\epsilon_X}\\
=&\int_{-\infty}^{0} \Pr(X_1^* > X_2^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)dG_{\Delta_{21}\epsilon_X}\\
&+\int_{0}^{\infty} \Pr(X_1^* > X_2^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)dG_{\Delta_{21}\epsilon_X}\\
=&\frac{1}{2} - \int_{-\infty}^{0} \Pr(X_2^* \geq X_1^* - \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)dG_{\Delta_{21}\epsilon_X}\\
&+\int_{0}^{\infty} \Pr(X_1^* > X_2^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)dG_{\Delta_{21}\epsilon_X}\\
=&\frac{1}{2} - \int_{0}^{\infty} \Pr(X_2^* \geq X_1^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)dG_{\Delta_{21}\epsilon_X}\\
&+\int_{0}^{\infty} \Pr(X_1^* > X_2^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)dG_{\Delta_{21}\epsilon_X},
\end{aligned}
$$

where we exploit the symmetry of the distribution of $\Delta_{21}\epsilon_X$ in the last line. When $y_1^* = y_2^*$, the two integrals in the last expression cancel out, so $\Pr(X_1 > X_2|Y_1^* = Y_2^* = y^*) = 1/2$ holds. For $\delta > 0$, set $y_1^* = y^* + \delta > y_2^* = y^*$ and rewrite the previous equation as

$$
\begin{aligned}
&\frac{1}{2} + \int_{0}^{\infty} \left[\Pr(X_1^* > X_2^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X) - \Pr(X_2^* \geq X_1^* + \Delta_{21}\epsilon_X|y_1^*, y_2^*, \Delta_{21}\epsilon_X)\right] dG_{\Delta_{21}\epsilon_X}\\
=&\frac{1}{2} + \int_{0}^{\infty} \left[\begin{array}{c} \int_{-\infty}^{\infty} \Pr(X_1^* > x + \Delta_{21}\epsilon_X|Y_1^* = y^* + \delta)dF_{X^*|Y^*}(x|y^*)\\ - \int_{-\infty}^{\infty} \Pr(X_2^* \geq x + \Delta_{21}\epsilon_X|Y_2^* = y^*)dF_{X^*|Y^*}(x|y^* + \delta) \end{array}\right] dG_{\Delta_{21}\epsilon_X}\\
\geq&\frac{1}{2} + \int_{0}^{\infty} \left[\begin{array}{c} \int_{-\infty}^{\infty} \Pr(X^* > x + \Delta_{21}\epsilon_X|Y^* = y^*)dF_{X^*|Y^*}(x|y^*)\\ - \int_{-\infty}^{\infty} \Pr(X^* \geq x + \Delta_{21}\epsilon_X|Y^* = y^*)dF_{X^*|Y^*}(x|y^* + \delta) \end{array}\right] dG_{\Delta_{21}\epsilon_X},
\end{aligned}
$$

where the inequality follows by the assumption of stochastic increasing dependence. Since $\Pr(X^* > x + \Delta_{21}\epsilon_X|Y^* = y^*)$ is non-increasing in x and $F_{X^*|Y^*}(\cdot|y^*)$ is first-order stochastically dominated by $F_{X^*|Y^*}(\cdot|y^* + \delta)$, we have

$$
\int_{-\infty}^{\infty} \Pr(X^* > x+\Delta_{21}\epsilon_X|Y^* = y^*)dF_{X^*|Y^*}(x|y^*) \geq \int_{-\infty}^{\infty} \Pr(X^* \geq x+\Delta_{21}\epsilon_X|Y^* = y^*)dF_{X^*|Y^*}(x|y^*+\delta).
$$

Hence, $\Pr(X_1 > X_2|y_1^*, y_2^*) \geq 1/2$.

Next, we show that the symmetric and unimodality with the nonnegative mode of the conditional distribution of $(X_1^* - X_2^*)$ given $\{Y_1^* = y_1^*, Y_2^* = y_2^*, y_1^* > y_2^*\}$ implies $\Pr(X_1^* > X_2^*|y_1^*, y_2^*) \geq \Pr(X_1 > X_2|y_1^*, y_2^*)$ for all $y_1^* > y_2^*$. Let the conditional cdf and pdf of $\Delta_{12}X^* = X_1^* - X_2^*$ given $\{Y_1^* = y_1^*, Y_2^* = y_2^*\}$ be denoted by $F_{\Delta_{12}X^*}(\cdot|y_1^*, y_2^*)$ and $f_{\Delta_{12}X^*}(\cdot|y_1^*, y_2^*)$, respectively. We denote by $g_{\Delta_{21}\epsilon_X}(\cdot)$ the Lebesgue density of $\Delta_{21}\epsilon_X$. Let $m \geq 0$ be the mode of the conditional distribution of $\Delta_{12}X^*$ given $\{Y_1^* = y_1^*, Y_2^* = y_2^*, y_1^* \geq y_2^*\}$. Consider

$$\begin{aligned}
\Pr(X_1 > X_2 | y_1^*, y_2^*) &= \int_{-\infty}^{\infty} [1 - F_{\Delta_{12}X^*}(\Delta_{21}\epsilon_X | y_1^*, y_2^*)] \, dG_{\Delta_{21}\epsilon_X} \\
&= \int_{0}^{\infty} \{[1 - F_{\Delta_{12}X^*}(\Delta_{21}\epsilon_X | y_1^*, y_2^*)] + [1 - F_{\Delta_{12}X^*}(-\Delta_{21}\epsilon_X | y_1^*, y_2^*)]\} \, dG_{\Delta_{21}\epsilon_X} \\
&= 1 - 2\int_{0}^{\infty} \left\{ \frac{1}{2} F_{\Delta_{12}X^*}(\Delta_{21}\epsilon_X | y_1^*, y_2^*) + \frac{1}{2} F_{\Delta_{12}X^*}(-\Delta_{21}\epsilon_X | y_1^*, y_2^*) \right\} dG_{\Delta_{21}\epsilon_X},
\end{aligned}$$

where the second line follows by the symmetry of the distribution of $\Delta_{21}\epsilon_X$. Consider bounding the integrand for each $\Delta_{21}\epsilon_X \geq 0$. For $\Delta_{21}\epsilon_X \in [0, m]$, since $F_{\Delta_{12}X^*}(\cdot | y_1^*, y_2^*)$ is convex in the left-tail relative to the mode, $\frac{1}{2} F_{\Delta_{12}X^*}(\Delta_{21}\epsilon_X | y_1^*, y_2^*) + \frac{1}{2} F_{\Delta_{12}X^*}(-\Delta_{21}\epsilon_X | y_1^*, y_2^*) \geq F_{\Delta_{12}X^*}(0 | y_1^*, y_2^*)$ holds.

For $\Delta_{21}\epsilon_X = m + \delta > m$, by the symmetry and unimodality of the distribution of $\Delta_{12}X^* | (y_1^*, y_2^*)$, $F_{\Delta_{12}X^*}(m + \delta | y_1^*, y_2^*) + F_{\Delta_{12}X^*}(-m - \delta | y_1^*, y_2^*)$ is monotonically increasing in δ, as

$$\begin{aligned}
\frac{d}{d\delta} \left(F_{\Delta_{12}X^*}(m + \delta | y_1^*, y_2^*) + F_{\Delta_{12}X^*}(-m - \delta | y_1^*, y_2^*) \right) &= f_{\Delta_{12}X^*}(m + \delta | y_1^*, y_2^*) - f_{\Delta_{12}X^*}(-m - \delta | y_1^*, y_2^*) \\
&\geq 0.
\end{aligned}$$

Hence, $\frac{1}{2} F_{\Delta_{12}X^*}(m+\delta | y_1^*, y_2^*) + \frac{1}{2} F_{\Delta_{12}X^*}(-m-\delta | y_1^*, y_2^*) \geq \frac{1}{2} F_{\Delta_{12}X^*}(m | y_1^*, y_2^*) + \frac{1}{2} F_{\Delta_{12}X^*}(-m | y_1^*, y_2^*) \geq F_{\Delta_{12}X^*}(0 | y_1^*, y_2^*)$ holds by the convexity of $F_{\Delta_{12}X^*}(\cdot | y_1^*, y_2^*)$ in the left-tail relative to the mode.

Combining these inequalities, we obtain

$$\begin{aligned}
\Pr(X_1 > X_2 | y_1^*, y_2^*) &\leq 1 - 2\left[\int_{0}^{m} F_{\Delta_{12}X^*}(0 | y_1^*, y_2^*) dG_{\Delta_{21}\epsilon_X} + \int_{m}^{\infty} F_{\Delta_{12}X^*}(0 | y_1^*, y_2^*) dG_{\Delta_{21}\epsilon_X} \right] \\
&= 1 - F_{\Delta_{12}X^*}(0 | y_1^*, y_2^*) \\
&= \Pr(X_1^* > X_2^* | y_1^*, y_2^*).
\end{aligned}$$

This completes the proof. □

A2 Summary Statistics

Table A1: Summary statistics for empirical application

	Mean	Std. dev.	Skewness	p10	p90	N
X^*	0	1.000	0.644	-1.011	1.212	36135
X	0	1.148	-0.120	-1.135	1.311	36135
ε_X	0	0.509	-2.630	-0.391	0.411	36135
Y^*	0	1.000	0.909	-0.984	1.206	36135
Y	0	1.170	0.018	-1.169	1.328	36135
ε_Y	0	0.632	-2.432	-0.516	0.511	36135

Note: Variables demeaned and variances of X^* and Y^* standardized to one.

Table A2: Correlation matrix for empirical application

	X^*	X	ε_X	Y^*	Y	ε_Y
X^*	1.000					
X	0.897	1.000				
ε_X	0.058	0.494	1.000			
Y^*	0.295	0.253	-0.010	1.000		
Y	0.240	0.206	-0.008	0.841	1.000	
ε_Y	-0.022	-0.019	0.000	-0.025	0.519	1.000

Note: The table reports pairwise Pearson correlation coefficients.