

Kim, Dongwoo; Wilhelm, Daniel

**Working Paper**

## Powerful t-tests in the presence of nonclassical measurement error

cemmap working paper, No. CWP18/21

**Provided in Cooperation with:**

The Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Kim, Dongwoo; Wilhelm, Daniel (2021) : Powerful t-tests in the presence of nonclassical measurement error, cemmap working paper, No. CWP18/21, Centre for Microdata Methods and Practice (cemmap), London, <https://doi.org/10.47004/wp.cem.2021.1821>

This Version is available at:

<https://hdl.handle.net/10419/241954>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Powerful t-tests in the presence of nonclassical measurement error

---

Dongwoo Kim  
Daniel Wilhelm

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP18/21



Economic  
and Social  
Research Council

# Powerful t-Tests in the Presence of Nonclassical Measurement Error

Dongwoo Kim\*      Daniel Wilhelm†

March 28, 2021

## Abstract

This paper proposes a powerful alternative to the t-test in linear regressions when a regressor is mismeasured. We assume there is a second contaminated measurement of the regressor of interest. We allow the two measurement errors to be nonclassical in the sense that they may both be correlated with the true regressor, they may be correlated with each other, and we do not require any location normalizations on the measurement errors. We propose a new maximal t-statistic that is formed from the regression of the outcome onto a maximally weighted linear combination of the two measurements. Critical values of the test are easily computed via a multiplier bootstrap. In simulations, we show that this new test can be significantly more powerful than t-statistics based on OLS or IV estimates. Finally, we apply the proposed test to the studies of returns to education based on twins data from the US and the UK. With our maximal t-test, we are able to discover statistically significant returns to education when standard t-tests do not.

---

\*Corresponding author. Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada; E-Mail address: [dongwook@sfu.ca](mailto:dongwook@sfu.ca).

†Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom; E-Mail address: [d.wilhelm@ucl.ac.uk](mailto:d.wilhelm@ucl.ac.uk). This work was supported in part by the European Research Council (ERC-2014-CoG-646917-ROMIA) and the ESRC Centre for Microdata Methods and Practice at IFS (RES-589-28-0001).

# 1 Introduction

The presence of measurement error in regressors causes the variance of OLS and IV estimators to be larger than those in the case without measurement error. As a result, t-tests about the regression coefficients may suffer from poor power. This problem can be particularly severe when the measurement errors are nonclassical, i.e. when they depend on the true level of the regressor, and/or when they depend on measurement errors of other measurements. In most economic applications in which the presence of measurement error is a first-order concern, neither of these two types of dependence can be ruled out.

For example, earnings have been shown to contain measurement error so that high-earners under-reported and low-earners over-report true earnings (Bound, Brown, and Mathiowetz (2001)). Therefore, measurement error must be nonclassical. In twins data such as that in Ashenfelter and Krueger (1994) and Bonjour, Cherkas, Haskel, Hawkes, and Spector (2003), the twins’ reports on education have been argued to be mismeasured and nonclassical (e.g. Black, Berger, and Scott (2000), Hu and Sasaki (2017)). In this example, one would not want to rule out the possibility that the measurement errors of the twins are correlated because the twins are likely to possess correlated abilities to report correctly. In the literature on estimation of the skill production function (e.g. Cunha, Heckman, and Schennach (2010), Heckman, Pinto, and Savelyev (2013), Attanasio, Cattani, Fitzsimons, Meghir, and Rubio-Codina (2015), Attanasio, Meghir, and Nix (2017)), the measurements of skill production inputs are often elicited through the same data collection process, e.g. measurements of parental inputs may come from survey responses by the mother, which are responses recorded in short succession, on the same day, by the same interviewer. Therefore, mistakes in such responses are likely correlated. More generally, any study that uses repeated measurements of a latent variable is likely to incur correlated measurement errors. Black, Berger, and Scott (2000) provide additional economic examples for which there is evidence for the presence of nonclassical measurement error.

In this paper, we propose to combine two measurements of the unobserved, true regressor so as to construct a new (“maximal”) t-statistic that we show, in simulations, leads to a more powerful test compared to standard t-tests. Specifically, consider the linear regression

$$Y = \beta X^* + \varepsilon,$$

where  $Y$  is an outcome variable and  $X^*$  an unobserved regressor. Suppose we observe two measurements  $X$  and  $Z$  of  $X^*$  that are uncorrelated with the regression error  $\varepsilon$ . We want to test the null hypothesis of no effect,

$$H_0: \beta = 0 \quad \text{vs.} \quad H_1: \beta \neq 0$$

using data on  $Y$ ,  $X$ , and  $Z$ . If we knew a priori that the measurement errors in  $X$  and  $Z$  were classical and that the one in  $Z$  had a larger variance than that in  $X$ , then a t-test based on the OLS regression of  $Y$  on  $X$  would lead to a more powerful test than one based on the OLS regression of  $Y$  on  $Z$ . In that case, one might, for example, discard the noisier measurement  $Z$  and report only results from the regression of  $Y$  on  $X$ . The problem with this approach is that it crucially depends on the assumption of classical measurement error and on the knowledge of which measurement is less contaminated. Both of these are typically difficult, if not impossible, to justify.

The new “maximal” t-statistic proposed in this paper combines the two measurements in a data-driven fashion so as to make the t-statistic as large as possible. In the example above with classical and independent measurement errors, the maximal t-statistic will automatically put more weight on the precise measurement  $X$  relative to the noisy measurement  $Z$  without the user having to know a priori which one is more precise. However, for the new test to be valid we do not need to assume the measurement errors to be classical, i.e. they may depend on the true level of the regressor and the errors in the two measurements may depend on each other. In fact, we will see in the simulations that power gains relative to standard t-tests may be particularly large in the presence of nonclassical measurement error.

To develop the new maximal t-test we first introduce the class of IV estimands from a regression of  $Y$  onto an arbitrary linear combination of  $X$  and  $Z$ , using a possibly different linear combination as instrument. Standard OLS and IV estimands are special cases of this class, corresponding to particular weights in the linear combinations. The t-statistic derived from the estimator of the general estimand can then be optimized over the four weights (two in each of the two linear combinations of  $X$  and  $Z$ ), leading to our maximal t-statistic. By maximizing over the weights we test the null of no effect for all weights simultaneously, whereas standard OLS and IV t-tests test the null only for specific choices of those weights.

We show that, with homoskedastic regression errors, there exists a closed-form solu-

tion for the maximal weights in the linear combinations, leading also to a closed-form expression of the maximal t-statistic. With more general heteroskedasticity-robust variance estimators, such closed-form expressions may not exist. It is then immediate to see that the maximal t-statistic is in general strictly larger than standard OLS- or IV-based t-statistics. One might therefore wonder whether it is possible to construct a more powerful test based on this test statistic. We propose a multiplier bootstrap procedure for the calculation of critical values. Extensive Monte Carlo experiments confirm that the resulting maximal t-test is indeed very powerful. In all our simulation scenarios, the maximal t-test dominates all other OLS- and IV-based tests, sometimes by substantial margins, or dominates all but one, where this latter one performs similarly well.

We also show that it is not possible to improve [Black, Berger, and Scott \(2000\)](#)'s OLS and IV-based identification bounds on  $\beta$  by considering linear combinations of  $X$  and  $Z$ . The maximal t-test approach does therefore not lead to improved inference procedures in the more general hypothesis testing problem of

$$H_0: \beta = b \quad \text{vs.} \quad H_1: \beta \neq b$$

for some value  $b \in \mathbb{R}$ .

We apply our new maximal t-test to revisit studies of returns to schooling using twins data from the US ([Ashenfelter and Krueger \(1994\)](#)) and the UK ([Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#)). Some recent studies (e.g. [Rouse \(1999\)](#), [Amin \(2011\)](#)) use different datasets or remove outliers and find that the estimated returns in the original papers may be too large and that they may not be statistically significant. By employing our maximal t-test, which remains powerful even in the presence of non-classical ME and in small samples, we are able to recover statistically significant returns to education when standard t-tests as in [Amin \(2011\)](#) do not (e.g. even after removing outliers). We therefore view our results as providing robustness in support of the original results in [Ashenfelter and Krueger \(1994\)](#) and [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#).

There is a large literature on identification and estimation of parametric and nonparametric models with measurement error. See, for example, the surveys by [Aigner, Hsiao, Kapteyn, and Wansbeek \(1984\)](#), [Bound, Brown, and Mathiowetz \(2001\)](#), [Chen, Hong, and Nekipelov \(2011\)](#), and [Schennach \(2013\)](#). Valid inference in parametric models is usually rather standard, once an identification argument and a consistent estimator have been

provided. For example, the general approach by [Schennach \(2014\)](#) allows GMM-like moment conditions to depend on observables and unobservables, and shows how to convert them into equivalent moment conditions that depend only on observables. The testing of general hypotheses then proceeds by using standard GMM inference procedures. However, we are not aware of any work trying to *improve* measurement error-robust inference methods by exploiting information in multiple measurements, which is the subject of this paper.

## 2 Maximal Combination of Measurements

In this section, we explore the possibility of maximizing the t-statistic by optimally combining two measurements of the latent regressor. First, we consider testing the hypothesis of no effect and then the more general hypothesis for a pre-specified effect size. We show that, in the former problem, the t-statistic based on maximally weighted linear combinations is, in general, larger than those based on the OLS and IV counterparts, but also that this is not possible in the latter problem.

### 2.1 Testing the Hypothesis of No Effect

To explain the main idea of this section consider first the following simple regression model for a scalar outcome  $Y$  and a scalar regressor  $X^*$  satisfying

$$Y = \beta X^* + \varepsilon. \tag{1}$$

The outcome  $Y$  is observed, but  $X^*$  is not. Instead, we observe two measurements  $X$  and  $Z$  of  $X^*$  that may contain measurement errors (MEs)  $U := X - X^*$  and  $V := Z - X^*$ . Defining ME to be additive is without loss of generality because we will neither impose any localization restriction on the ME, nor will we assume that it is independent of the true regressor  $X^*$ . In fact, we will not even restrict the dependence among the two MEs  $U$  and  $V$ .

Suppose we are interested in testing

$$H_0: \beta = 0 \quad \text{vs.} \quad H_1: \beta \neq 0. \tag{2}$$

The OLS estimand of a regression of  $Y$  on  $X$  and the IV estimand from a regression of  $Y$  on  $X$  using  $Z$  as instrument are

$$\beta_{OLS} := \frac{E[YX]}{E[X^2]} = \beta \frac{E[(X^*)^2] + E[X^*U]}{E[(X^*)^2] + 2E[X^*U] + E[U^2]}$$

$$\beta_{IV} := \frac{E[YZ]}{E[XZ]} = \beta \frac{E[(X^*)^2] + E[X^*V]}{E[(X^*)^2] + E[X^*U] + E[X^*V] + E[UV]}$$

Under the null,  $\beta_{OLS} = \beta_{IV} = 0$ , so the standard t-tests based on the OLS or IV estimators are asymptotically of correct level regardless of the dependence structure of the measurement system (i.e. the values of  $E[X^*U]$ ,  $E[X^*V]$ ,  $E[UV]$ ). However, power of these two tests may be poor for two reasons: under any fixed and arbitrarily large alternative  $\beta \neq 0$ , at least one of the estimands  $\beta_{OLS}$  and  $\beta_{IV}$  is close to zero when (a)  $E[X^*U]$  or  $E[X^*V]$  is negative and large in absolute values or (b) one of  $E[U^2]$ ,  $E[X^*U]$ ,  $E[X^*V]$ ,  $E[UV]$  is large. Since the OLS and IV estimands depend on those quantities in different ways, the resulting t-tests may possess poor power under different data-generating processes.

We propose a new “maximal t-test” that is based on a class of estimands that contains OLS and IV as special cases. For any  $a, b \in \mathbb{R}$ , let  $W(a, b) := aX + bZ$  be an arbitrarily weighted linear combination of the two measurements  $X$  and  $Z$ . For  $\omega := (\omega_1, \omega_2, \omega_3, \omega_4)' \in \mathbb{R}^4$ , consider the IV estimand from the regression of  $Y$  onto the linear combination  $W(\omega_3, \omega_4)$  using a possibly different linear combination  $W(\omega_1, \omega_2)$  as instrument,

$$\beta(\omega) := \frac{E[W(\omega_1, \omega_2)Y]}{E[W(\omega_1, \omega_2)W(\omega_3, \omega_4)]}. \quad (3)$$

The OLS estimand of regressing  $Y$  onto  $X$  ( $Z$ ) corresponds to  $\beta(1, 0, 1, 0)$  ( $\beta(0, 1, 0, 1)$ ) and the IV estimand of regressing  $Y$  onto  $X$  ( $Z$ ) using  $Z$  ( $X$ ) as IV corresponds to  $\beta(1, 0, 0, 1)$  ( $\beta(0, 1, 1, 0)$ ).

The null implies that  $\beta(\omega) = 0$  for all  $\omega \in \mathbb{R}^4$  and, as long as  $E[(X^*)^2] + E[X^*V] \neq 0$  and  $E[(X^*)^2] + E[X^*U] \neq 0$ , one can easily show that (4) is equivalent to

$$H_0: \sup_{\omega \in \mathbb{R}} |\beta(\omega)| = 0 \quad \text{vs.} \quad H_1: \sup_{\omega \in \mathbb{R}} |\beta(\omega)| \neq 0. \quad (4)$$

If the regression errors are homoskedastic, we standardize the estimand by

$$\sigma^2(\omega) := \frac{\sigma_\varepsilon^2(\omega) E[W(\omega_1, \omega_2)^2]}{(E[W(\omega_1, \omega_2)W(\omega_3, \omega_4)])^2},$$



where  $\sigma_\varepsilon^2(\omega) := E[\varepsilon(\omega)^2]$  with  $\varepsilon(\omega) := Y - \beta(\omega)W(\omega_3, \omega_4)$ , is the residual variance. Given the reformulation of the null in (4), we consider the maximal t-ratio  $\sup_{\omega \in \mathbb{R}} |t(\omega)|$  with

$$t(\omega) := \frac{\beta(\omega)}{\sigma(\omega)} \quad (5)$$

because it imposes all restrictions under the null, unlike the OLS and IV estimands which only impose the null for one specific choice of  $\omega$ .

**Assumption 1.** (i)  $E[\varepsilon X] = E[\varepsilon Z] = 0$ ,  $E[X^*X] \neq 0$  and  $E[ZX^*] \neq 0$

(ii)  $\text{Var}(X^*) > 0$ ,  $\text{Var}(X) > 0$ ,  $\text{Var}(Z) > 0$ ,  $\text{Var}(\varepsilon) > 0$

(iii)  $|\text{Corr}(X, Z)| \neq 1$  and  $|\text{Corr}(Y, X - Z)| \neq 1$

Notice that part (i) and (ii) of this assumption allows for endogeneity of the latent regressor ( $E[\varepsilon X^*] \neq 0$ ), requiring only that the measurements  $X$  and  $Z$  are exogenous. More importantly, the assumption does not restrict the covariances ( $E[X^*U]$ ,  $E[X^*V]$ ,  $E[UV]$ ). Part (iii) requires that the two measurements are not linearly dependent and  $Y$  is not perfectly correlated with the difference of the two MEs.

**Theorem 1.** Suppose model (1) and Assumption 1 hold. Then,  $\omega^* := \sup_{\omega \in \mathbb{R}^4} |t(\omega)|$  is of the form  $\omega^* = (a^*, 1 - a^*, a^*, 1 - a^*)$  with

$$a^* := \frac{E[XZ]E[ZY] - E[Z^2]E[XY]}{E[X(Z - X)]E[ZY] + E[Z(X - Z)]E[XY]}$$

with the possibility of  $|a^*| = \infty$  if the denominator is equal to zero.

Since the maximal elements  $(\omega_3^*, \omega_4^*)$  are equal to  $(\omega_1^*, \omega_2^*)$ , the estimand for the maximal weights becomes that of the OLS regression of  $Y$  onto  $W(a^*, 1 - a^*)$ :

$$\beta(\omega^*) = \frac{E[W(a^*, 1 - a^*)Y]}{E[W(a^*, 1 - a^*)^2]}.$$

Notice that, in general, the weight  $a^*$  may be outside  $[0, 1]$ , in fact it may be equal to  $\pm\infty$ . Interestingly the weight depends only on observable quantities based on the distribution of  $(Y, X, Z)$ , but leads to the largest possible t-ratio regardless of the unknown measurement covariance structure ( $E[X^*U]$ ,  $E[X^*V]$ ,  $E[UV]$ ) of the measurement system. In this sense,

the maximal t-ratio adapts to the unknown measurement covariance structure without the researcher having to place a priori restrictions on those covariances.

In the special case in which  $X$  contains no information about the latent regressor  $X^*$  (i.e.  $X$  is uncorrelated with  $X^*$ ),  $X$  is uncorrelated with the ME in  $Z$ , and  $E[X\varepsilon] = 0$ , then  $E[XZ] = E[XY] = 0$  and the maximal weight becomes  $a^* = 0$ . This weight corresponds to OLS regression of  $Y$  onto  $Z$ , ignoring the second measurement  $X$ . On the other hand, when  $Z$  contains no information about the latent regressor  $X^*$ ,  $Z$  is uncorrelated with the ME in  $X$ , and  $E[Z\varepsilon] = 0$ , then  $E[XZ] = E[ZY] = 0$  and the maximal weight becomes  $a^* = 1$ . This weight corresponds to OLS regression of  $Y$  onto  $X$ , ignoring the second measurement  $Z$ . Similarly, using the inverse function theorem one can show that there exists an  $a$  such that  $\beta(a, 1 - a, a, 1 - a)$  equals the IV estimand (for a regression of  $Y$  onto  $X$  ( $Z$ ), using  $Z$  ( $X$ ) as instrument), but the expression for that  $a$ -value is quite complicated.

Since the maximal weight  $\omega^*$  can attain any value on the extended real line, it is immediately clear that the maximal t-ratio  $|t(\omega^*)|$  is not smaller than the absolute value of the corresponding OLS and the IV t-ratios. Therefore, there exist data-generating processes for which the maximal t-ratio is strictly larger than any of them and one might expect that imposing all restrictions of the null by testing whether the maximal t-ratio is equal to zero might translate into favorable power properties of the test. Section 4 confirms this intuition.

Since  $\omega^*$  possesses a closed-form solution, the maximal t-ratio also does. For example, when  $|a^*| < \infty$ , then the maximal t-ratio can be written as

$$t(\omega^*) = \text{sign}(C) \sqrt{\frac{A}{B}}$$

with

$$\begin{aligned} A &:= E[(Z E[XY] - X E[ZY])^2] \\ B &:= A - E[Y^2] ((E[XZ])^2 - E[Z^2]E[X^2]) \\ C &:= E[X(Z - X)]E[ZY] + E[Z(X - Z)]E[XY] \end{aligned}$$

Under the null, the t-statistic is independent of the weight, so whether the weight is finite or infinite does not change its functional form. Under alternatives  $\beta \neq 0$ , the case  $|a^*| = \infty$  is unlikely and occurs only in extreme case in which the MEs are perfectly correlated with the true regressor  $X^*$  or the regression error  $\varepsilon$  (see Lemma 1 in the Appendix).

Now suppose there are additional, correctly measured, exogenous regressors  $R$ ,  $E(\varepsilon R) = 0$ , so that

$$Y = \beta X^* + \gamma' R + \varepsilon. \quad (6)$$

Denote by  $\tilde{X}$ ,  $\tilde{Z}$ ,  $\tilde{Y}$  the residuals of regressions of  $X$ ,  $Z$ ,  $Y$  onto  $R$  and let  $\tilde{W}(a, b) := a\tilde{X} + b\tilde{Z}$ . Define  $\tilde{\beta}(\omega)$  like  $\beta(\omega)$ , replacing  $W(a, b)$  by  $\tilde{W}(a, b)$ , and  $\tilde{\sigma}^2(\omega)$  like  $\sigma^2(\omega)$ , replacing  $Y, W(a, b), \beta(\omega)$  by  $\tilde{Y}, \tilde{W}(a, b), \tilde{\beta}(\omega)$ . The t-ratio is then defined as

$$\tilde{t}(\omega) := \frac{\tilde{\beta}(\omega)}{\tilde{\sigma}(\omega)}. \quad (7)$$

The OLS estimand of regressing  $Y$  onto  $(X, R)$  corresponds to  $\tilde{\beta}(1, 0, 1, 0)$ , the IV estimand regressing  $Y$  onto  $(X, R)$  using  $(Z, R)$  as IVs corresponds to  $\tilde{\beta}(1, 0, 0, 1)$ .

The maximal t-ratio and its maximizer can then be found using the result in Theorem 1:

**Corollary 1.** *Suppose (6) and Assumption 1 hold with  $(Y, X, Z)$  replaced by  $(\tilde{Y}, \tilde{X}, \tilde{Z})$ . Then,  $\tilde{\omega}^* := \sup_{\omega \in \mathbb{R}^4} |\tilde{t}(\omega)|$  is of the form  $\tilde{\omega}^* = (\tilde{a}^*, 1 - \tilde{a}^*, \tilde{a}^*, 1 - \tilde{a}^*)$  with*

$$\tilde{a}^* := \frac{E[\tilde{X}\tilde{Z}]E[\tilde{Z}\tilde{Y}] - E[\tilde{Z}^2]E[\tilde{X}\tilde{Y}]}{E[\tilde{X}(\tilde{Z} - \tilde{X})]E[\tilde{Z}\tilde{Y}] + E[\tilde{Z}(\tilde{X} - \tilde{Z})]E[\tilde{X}\tilde{Y}]}$$

with the possibility of  $|\tilde{a}^*| = \infty$  if the denominator is equal to zero.

The t-ratios defined in (5) and (7) use asymptotic variance expressions for homoskedastic regression errors. We now consider conditional heteroskedasticity in the regression model (6), i.e. when the conditional variance of the regression error given  $X^*$  is not independent of  $X^*$ . In this case, we are interested in the t-ratio

$$\bar{t}(\omega) := \frac{\tilde{\beta}(\omega)}{\bar{\sigma}(\omega)},$$

where the variance is

$$\bar{\sigma}^2(\omega) := \frac{E[\tilde{\varepsilon}(\omega)^2 \tilde{W}(\omega_3, \omega_4)^2]}{(E[\tilde{W}(\omega_1, \omega_2) \tilde{W}(\omega_3, \omega_4)])^2}$$

with  $\tilde{\varepsilon}(\omega) := \tilde{Y} - \tilde{\beta}(\omega)\tilde{W}(\omega_1, \omega_2)$ .

Without any further restrictions on the form of heteroskedasticity, there does not in general exist a closed-form expression for the maximal weights nor the maximal t-ratio. However, we can show that the four-dimensional optimization problem over  $\omega$  can be reduced to a two-dimensional one, which simplifies implementation significantly.

**Theorem 2.** Suppose model (6) and Assumption 1 hold with  $(Y, X, Z)$  replaced by  $(\tilde{Y}, \tilde{X}, \tilde{Z})$ . Then, for any  $\omega := (\omega_1, \omega_2, \omega_3, \omega_4) \in \mathbb{R}^4$ , there exist  $a_1, a_2 \in \mathbb{R}$  such that, for  $\omega' := (a_1, 1 - a_1, a_2, 1 - a_2)$ ,

$$\bar{t}(\omega) = \bar{t}(\omega').$$

**Remark 1.** In principle, one could construct a test statistic for  $H_0$  using more complicated, non-linear combinations of  $X$  and  $Z$  rather than the linear combinations as in the definition of  $W(a, b)$ . For example, one could consider  $W(a, b) := a'p(X) + b'q(Z)$ , where  $a, b \in \mathbb{R}^K$ , and  $p(\cdot)$  and  $q(\cdot)$  are  $K$ -dimensional vectors of basis functions (e.g. polynomials or splines). The results of this section can easily be extended to this case. However, allowing for interaction terms between the basis functions of  $X$  and  $Z$  would require more involved arguments and we leave these for future work. In this paper, we focus on linear combinations because we place more importance on practical usefulness of the method and because we want to avoid the maximal  $t$ -statistic to depend on higher-order moments of  $(Y, X, Z)$  which would make it more sensitive to outliers.  $\square$

## 2.2 Testing General Hypotheses

In this section, we consider testing the more general hypothesis

$$H_0 : \beta = b \quad \text{vs.} \quad H_1 : \beta \neq b \quad (8)$$

for some  $b \in \mathbb{R}$  in the simple regression model (1). Under the null of no effect, (2), the estimand  $\beta(a)$  considered in the previous section is zero for all weights  $a$ . This is not the case under the more general null (8) because  $\beta$  is not identified. However, one can construct bounds for  $\beta$  based on observable quantities. To describe these bounds consider the following notation. Because of Theorem 2 it suffices to consider linear combinations of measurements with weights  $a$  and  $1 - a$ , so we simplify the notation to  $W(a) := aX + (1 - a)Z$ . Let

$$\beta_{OLS}(a) := \frac{E[W(a)Y]}{E[W(a)^2]}$$

be the OLS estimand of a regression of  $Y$  on  $W(a)$ , and

$$\beta_{OLS-INV}(a) := \frac{E[Y^2]}{E[W(a)Y]}$$

that of the reverse regression. Similarly, the IV estimand from a regression of  $Y$  on  $W(a)$  using  $W(b)$  as an instrument is

$$\beta_{IV}(a, b) := \frac{E[W(b)Y]}{E[W(b)W(a)]}.$$

**Assumption 2.**  $E[\varepsilon X^*] = E[\varepsilon U] = E[\varepsilon V] = 0$

**Assumption 3.**  $E[X^*X] > 0$  and  $E[X^*Z] > 0$

**Assumption 4.**  $E[X^*X] \geq E[XZ] > 0$  and  $E[X^*Z] \geq E[XZ] > 0$

**Assumption 5.**  $E[UX] \geq 0$  and  $E[VZ] \geq 0$

**Assumption 6.**  $E[X^*U] \leq 0$  and  $E[X^*V] \leq 0$

For the discussion in this section, suppose  $\beta \geq 0$  (similar results hold for the case  $\beta \leq 0$ ). [Black, Berger, and Scott \(2000\)](#) show how to construct bounds on  $\beta$  using the above assumptions.<sup>1</sup> First, the OLS estimand and the inverse OLS estimand form lower and upper bounds on the regression coefficient:

$$\max_{a=0,1} \beta_{OLS}(a) \leq \beta \leq \min_{a=0,1} \beta_{OLS-INV}(a).$$

The IV estimand provides an additional bound that may or may not be tighter than the OLS bound,

$$\beta \leq \min_{a=0,1} \beta_{IV}(a, 1-a).$$

We now show that the bounds cannot be refined by considering linear combinations of measurements.

**Theorem 3.** *Suppose  $\beta \geq 0$ . Then:*

1. *Under Assumptions 2, 3, and 5,  $\beta_{OLS}(a) \leq \beta$  holds only for  $a \in \{0, 1\}$ .*
2. *Under Assumptions 2, 3, and 6,  $\beta \leq \beta_{OLS-INV}(a)$  holds only for  $a \in [0, 1]$ .*
3. *Under Assumptions 2 and 4,  $\beta \leq \beta_{IV}(a, 1-a)$  holds only for  $a \in \{0, 1\}$ .*

---

<sup>1</sup>Lemma 2 in the appendix provides a slightly modified statement and derivation for those bounds.

The first and third statement of this theorem imply that it is not possible to refine the OLS and IV bounds using linear combinations of measurements. The proof shows that, for any  $a \notin \{0, 1\}$ , we can find a data-generating process satisfying the assumptions of the theorem for which the inequalities are violated. The second part of the theorem states that the inverse OLS bound holds for all  $a \in [0, 1]$ . However, because the inverse OLS estimand is inversely related to a linear function of the weight  $a$ , it is easy to see that the tightest bound,  $\min_{a \in [0, 1]} \beta_{OLS-INV}(a)$ , is achieved by either  $a = 0$  or  $a = 1$ . Therefore, the inverse OLS bound also cannot be refined by considering linear combinations of the two measurements.

Under the null hypothesis of no effect, (2), the estimand  $\beta(a)$  is equal to zero for all weights  $a$ , but under the alternative it varies with  $a$ . This is why it was possible to improve power of the t-test by choosing  $a$  carefully. In the case of testing the general null (8), however, the estimand of interest is the set of possible  $\beta$  values consistent with the data,

$$\Theta(a_1, a_2, a_3) := \{\beta \in \mathbb{R} : \beta_{OLS}(a_1) \leq \beta \leq \min\{\beta_{OLS-INV}(a_2), \beta_{IV}(a_3, 1 - a_3)\}\}.$$

These sets contain the identified set for  $\beta$  and thus are valid outer sets for any  $(a_1, a_2, a_3) \in \{0, 1\}^3$ , but they are not for any other value of  $(a_1, a_2, a_3)$ . Therefore, we cannot take linear combinations of the measurements to tighten the identified set or improve inference as in the approach of the previous section.

### 3 The Maximal t-Test

In this section, we propose a powerful t-test of the hypothesis (2), making use of the maximal combination of two measurements. We want to allow for heteroskedasticity-robust variance estimators in the construction of the t-statistic. As we have shown in Theorem 2, for the calculation of the maximal t-statistic, we do not have to consider general weights  $\omega$  in  $\mathbb{R}^4$ , but only weights of the form  $(a_1, 1 - a_1, a_2, 1 - a_2)$ . A direct implementation of this result could consider the test statistic

$$\max_{a=(a_1, a_2) \in \mathbb{R}^2} |\hat{t}(a_1, 1 - a_1, a_2, 1 - a_2)|,$$

where  $\hat{t}$  is an estimator of  $t$  multiplied by  $\sqrt{n}$ . Under standard conditions, it is easy to show that  $\hat{t}(\cdot)$  weakly converges to a Gaussian process, so that by the delta method the maximal

t-statistic converges to the supremum of that limiting process. Unfortunately, it is difficult to construct critical values from this process because its covariance function depends upon the unknown data-generating process. We therefore propose a simple multiplier bootstrap method to calculate critical values.

Suppose we observe an i.i.d. sample  $\mathcal{Y}_n := \{(Y_i, X_i, Z_i)\}_{i=1}^n$  from the distribution of  $(Y, X, Z)$  and ignore additional, correctly measured covariates. We describe a procedure that applies to numerical optimization of the t-ratio. Define a grid of weights  $\omega_1, \dots, \omega_p$  of the form  $\omega_j = (a_{j,1}, 1 - a_{j,1}, a_{j,2}, 1 - a_{j,2})'$  for some  $(a_{j,1}, a_{j,2})' \in \mathbb{R}^2$ . Let  $W_{1ij} := \omega_{j,1}X_i + \omega_{j,2}Z_i$ ,  $W_{2ij} := \omega_{j,3}X_i + \omega_{j,4}Z_i$ ,  $\sigma_j^2 := E[\varepsilon_i^2 W_{1ij}^2]$ , and  $\hat{\sigma}_j^2$  some estimator of  $\sigma_j^2$ . The notation  $\mathbb{E}_n[\cdot]$  denotes the average over the index  $1 \leq i \leq n$ , so for example  $\mathbb{E}_n[x_{ij}] = n^{-1} \sum_{i=1}^n x_{ij}$ .

Our maximal t-statistic is defined as  $|\hat{t}(\omega)|$  optimized over the grid  $\omega_1, \dots, \omega_p$ :

$$T := \max_{\omega \in \{\omega_1, \dots, \omega_p\}} |\hat{t}(\omega)| = \max_{1 \leq j \leq p} \left| \frac{\sqrt{n} \mathbb{E}_n[W_{1ij} Y_i]}{\hat{\sigma}_j} \right|.$$

We now describe the construction of critical values using a multiplier bootstrap. Let  $\{e_i^b\}_{i=1}^n$ ,  $b = 1, \dots, B$ , be an i.i.d. sequence of standard normally distributed random variables that are independent of  $\mathcal{Y}_n$  and the residual  $\hat{\varepsilon}_{ij} := Y_i - \hat{\beta}_j W_{2ij}$  where  $\hat{\beta}_j$  is the IV estimator from a regression of  $Y_i$  onto  $W_{2ij}$  using  $W_{1ij}$  as instrument. We define the bootstrap statistic

$$T^b := \max_{1 \leq j \leq p} \left| \frac{\sqrt{n} \mathbb{E}_n[e_i W_{1ij} \hat{\varepsilon}_{ij}]}{\hat{\sigma}_j} \right|, \quad b = 1, \dots, B,$$

and the bootstrap critical value

$$c_\alpha := \text{conditional } (1 - \alpha) - \text{quantile of } T^b \text{ given the data } \mathcal{Y}_n.$$

We reject  $H_0$  if and only if  $T > c_\alpha$ .

The validity of this bootstrap critical value can be shown using the high-dimensional central limit theorem in Corollary 3.1 of [Chernozhukov, Chetverikov, and Kato \(2013\)](#). It implies that the test has limiting rejection probability under the null equal to the nominal level  $\alpha$  and is consistent against any fixed alternative violating the the null. The result requires only mild conditions on the data-generating process and, in particular, allows the vector  $W_{1ij}$  to grow at a rate that is an exponential function of the sample size. This means that we can consider grids for the weights  $\omega_1, \dots, \omega_p$  of very large size  $p$ .

**Remark 2.** In simulations and empirical applications, we also consider a version of the maximal t-test in which we impose homoskedasticity ( $E[\varepsilon_i^2|W_{1ij} = w]$  does not vary with  $w$ ) by using the variance estimator  $\hat{\sigma}_j^2 = \hat{\sigma}_{\varepsilon,j}^2 \mathbb{E}_n[W_{1ij}^2]$  where  $\hat{\sigma}_{\varepsilon,j}^2$  is the sample variance of  $\hat{\varepsilon}_{ij}$ .  $\square$

## 4 Simulations

This section studies the finite sample performance of the maximal t-test described in Section 3. We consider the simple regression model in (1) with the two measurements  $X$  and  $Z$  generated from  $X = X^* + U$  and  $Z = X^* + V$  with

$$\begin{bmatrix} X^* \\ U \\ V \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{X^*U} & \sigma_{X^*V} \\ \sigma_{X^*U} & \sigma_U^2 & \sigma_{UV} \\ \sigma_{X^*V} & \sigma_{UV} & \sigma_V^2 \end{bmatrix} \right)$$

and  $\varepsilon \sim N(0, 1)$  is independent of  $(X^*, U, V)$ . The null hypothesis holds with  $\beta = 0$ . To generate alternatives we increase  $\beta$  on a grid up to 0.8. We define six different scenarios in which we vary the parameters  $\sigma_U^2$ ,  $\sigma_V^2$ ,  $\sigma_{UV}$ . Table 1 provides the definition of those scenarios. Within each of these we vary the covariances  $\sigma_{X^*U}$  and  $\sigma_{X^*V}$ . We generate 1,000 Monte Carlo samples of size  $n = 200$ .

We consider four tests. The first two are t-tests based on the OLS estimator from a regression of  $Y$  on  $X$  (“OLS w/ x”) and of  $Y$  on  $Z$  (“OLS w/ z”), respectively. The third (“IV x|z”) is the t-test based on the IV estimator from a regression of  $Y$  on  $X$  using  $Z$  as instrument. The fourth (“tmax”) is the maximal t-test with the multiplier bootstrap critical value described in Section 3. We use  $B = 1,000$  bootstrap samples and an equally spaced grid of weights in the interval  $[0, 1]$  with a distance of 0.2 between the grid points. All tests are implemented with the homoskedastic variance estimator described in Remark 2 and nominal level of 0.05.

Table 2 provides the null rejection frequencies of the tests and Figures 1–6 the power curves for each of the six scenarios.

All tests control size well as expected. The power, however, varies significantly among them. The findings can be summarized as follows. First, in almost all scenarios and parameter combinations, the IV-based t-test performs poorest. Second, in none of the scenarios and parameter combinations, the maximal t-test is dominated by another test.



Scenario	$\sigma_U^2, \sigma_V^2$	$\sigma_{UV}$	$\sigma_{X*U}$	$\sigma_{X*V}$
1	2	0	-0.3 or -0.7	-0.3 or -0.7
2	2	0.5	-0.3 or -0.7	-0.3 or -0.7
3	2	-0.5	-0.3 or -0.7	-0.3 or -0.7
4	1	0	-0.3 or -0.5	-0.3 or -0.5
5	1	0.3	-0.3 or -0.5	-0.3 or -0.5
6	1	-0.3	-0.3 or -0.5	-0.3 or -0.5

Table 1: Parameter values in each scenario.

In most of them, the maximal t-test strictly dominates all other tests (e.g. in scenarios 3 and 6). When the maximal t-test does not strictly dominate all other tests, then it does strictly dominate all but one of the tests, with the latter possessing power roughly equal to that of the maximal t-test. In most of the scenarios and parameter combinations, the power gains of the maximal t-test relative to the others is significant (e.g. scenarios 3 and 6). Third, a common feature among all simulation results is that when the measurement  $X$  is more precise than  $Z$ , then the OLS-based test using  $X$  dominates the OLS-based test using  $Z$  and the former is closer to the maximal t-test in terms of power (compare, for example, the upper-left and the lower-right panels in Figure 1). Our maximal t-test either dominates both or is equal to the more powerful of the two tests. In this sense, it automatically adapts to the unknown covariance structure in the measurement system and puts more weight on the more informative measurement.

## 5 Empirical Applications: Returns to Schooling

In this section, we revisit the empirical work by [Ashenfelter and Krueger \(1994\)](#) and [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#) who studied the returns to schooling using data sets from monozygotic twins.<sup>2</sup> While [Ashenfelter and Krueger \(1994\)](#) find that each additional year of schooling increases hourly wage by 12-16%, which is statistically significant at the 1% level, [Rouse \(1999\)](#) subsequently suggested that the returns may be much smaller (9.5%) and found mixed evidence on their statistical significance. Similarly,

<sup>2</sup>[Ashenfelter and Krueger \(1994\)](#) use twins data surveyed at the 16th Annual Twins Days Festival held in Twinsburg, Ohio, the United States. [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#) conduct a similar analysis using UK twins data collected from St. Thomas' U.K. Adult Twin Registry.

Bonjour, Cherkas, Haskel, Hawkes, and Spector (2003) estimate a return of 7.7%, which is statistically significant at the 5% level. Amin (2011) argues that this result is largely driven by a few outliers in the data set. In particular, one pair of twins have a very large difference in their hourly earnings (£94.18) with only a 2-year schooling gap, whereas the average difference of hourly wages with a 2-year difference in education is only £7.74. Amin (2011) shows that, after removing this outlier, the estimated return to schooling is only 5.1% and it is no longer significant at the 5% level.

The goal of this section is to apply our new t-test, which in simulations has been shown to me more powerful than standard OLS- and IV-based t-tests, and see whether we can find statistically significant returns to schooling in the US and UK datasets.

To this end, for a pair of twins, let  $Y_1$  and  $Y_2$  be log wages of the first and second twin. Denote by  $X_1$  and  $X_2$  their respective self-reported levels of education. In addition, the second twin is asked to report the education of the first twin,  $Z_1$ , and the first to report the education of the second,  $Z_2$ . We consider the regression models (1) and (6) with  $Y = Y_1 - Y_2$  the difference of log wages,  $X^*$  the difference of true education levels,  $X = X_1 - X_2$  the difference of self-reported education, and  $Z = Z_1 - Z_2$  the difference of cross-reported education.  $R$  contains additional covariates such as marital status and work experience.

## 5.1 US Twins Data

To study the returns to schooling in the US, we use Ashenfelter and Krueger (1994)'s dataset which is publicly available.

First, we test the null of no ME in self-reported education levels using the nonparametric test in Wilhelm (2019).<sup>3</sup> The test statistic is 0.00060 and the critical value 0.00156, leading to a p-value of 0.376. Hence we cannot reject the null hypothesis of no ME in self-reported education. There is not enough information in the data set to reject the null. This could simply be because the sample size (147) is too small so the nonparametric test does not have enough power. Therefore, in a second step we impose the linearity of the wage equation and perform the test of no ME again. To do this we regress  $Y$  onto both  $X$  and  $Z$  (and a constant). We find that the coefficient of  $Z$  is statistically significant at the

---

<sup>3</sup>We implement the test based on Delgado and Manteiga (2001)'s Cramér-von Mises statistic for the null hypothesis (9) in Wilhelm (2019) with  $\mu(y) = y$ , a gaussian kernel, and 1,000 bootstrap replications.

5% level (t-statistic of 2.21). Therefore, with the assumption of linearity, we reject the null of no ME in self-reported education (Hausman (1978) and Lee and Wilhelm (2020)). This finding is consistent with the possibility that the nonparametric test has low power because of the small sample size and thus fails to reject, but the test imposing linearity of the conditional expectation is more powerful and thus does reject. Since the education variable is discrete, we know that by construction ME (if there is any) cannot be classical because it has to depend on the true level of education.

We now apply our maximal t-test imposing homoskedasticity as described in Remark 2. We implement our maximal t-test with  $B = 5,000$  bootstrap replications and an equally-spaced grid of weights  $\omega_1, \dots, \omega_p$  such that  $\omega_j = ((j-1)/n, 1 - (j-1)/n, (j-1)/n, 1 - (j-1)/n)$  for  $j = 1, \dots, p = n+1$ . We have also used grids that are twice the size as well as half the size, but the test results were almost identical.

We consider several subsamples of the data, using from 30 to 147 (the full sample) observations. Table 3 shows the results of our maximal t-test (“maxt”) and compares it to three other tests: the t-test based on the OLS estimator from a regression of  $Y$  on  $X$  (“OLS w/ x”), the t-test based on the OLS estimator from a regression of  $Y$  on  $Z$  (“OLS w/ z”), and the t-test based on the IV estimator from a regression of  $Y$  on  $X$  using  $Z$  as instrument (“IV”). The latter three tests also employ variance estimators imposing homoskedasticity. The table also shows the maximizer of our t-statistic (“ $\hat{a}^*$ ”) defined as the first element of the grid point  $\omega_j$  at which the t-statistic is maximized, the 95%-critical value (“95% CV”), and the p-value of our test from the multiplier bootstrap.

Our maximal t-test rejects the null of no effect for all sample sizes, even when  $n = 30$ . Ashenfelter and Krueger (1994) consider the OLS estimator using  $X$  and the IV estimator, among others. These do reject for the full sample, but not for all smaller samples. The OLS-based test rejects for sample sizes of 90 or larger and the IV-based test for sizes of 60 or larger. The OLS estimator using  $Z$  was not considered in Ashenfelter and Krueger (1994), but it turns out to also reject for all sample sizes.

The findings in this empirical application are consistent with the simulation results. They could, for example, be interpreted similarly to the upper-left panel of Figure 2. The maximal t-test rejects the null at all sample sizes and so does OLS w/z. The other two tests fail to reject the null at smaller sample sizes but the IV test shows stronger power. It is plausible that twins may have similar “ability to report correctly” in the sense that  $U$  and  $V$  are positively correlated. As the OLS-based test using  $Z$  seems to

reject more often than the one using  $X$ , the findings are consistent with  $Z$  being a more precise measurement than  $X$ . The covariance structure of  $(X^*, U, V)$  is not identified and thus cannot be consistently estimated, so we cannot learn from data a priori which measurement is more precise, whether linear combinations of measurements help improve power, and therefore which of the many different possible OLS and IV tests to use. However, our maximal t-test adapts to the unknown covariance structure, thus does not require such knowledge, and is more powerful.

Finally, we test the null hypothesis without imposing homoskedasticity. Table 4 reports the results. The OLS and IV t-tests are implemented with Eicker-Huber-White heteroskedasticity robust standard errors. The key finding of this experiment is that “OLS w/z” and our test continue to reject for all sample sizes, while the “OLS w/x” and “IV” tests require much larger sample sizes (at least 110 and 120 respectively) to reject than under the homoskedasticity assumption. This finding is consistent with the presence of conditional heteroskedasticity in the error term.

## 5.2 UK Twins Data

To study the returns to schooling in the UK, we use [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#)’s data set.

First, we test the null of no ME in self-reported education levels using the same two tests as in the previous subsection. The nonparametric test in [Wilhelm \(2019\)](#) leads to a test statistic of 0.00115 and the critical value 0.00361, which is translated into a p-value of 0.4. The regression-based test imposing linearity of the conditional mean shows that the p-value of the coefficient estimate of  $Z$  is around 0.1 (t-statistic of 1.628). Therefore, we still suspect the presence of measurement errors while the tests do not strongly reject the null of no ME.

Analogously as with the US data, we first compare our maximal t-test to the three OLS- and IV-based t-tests imposing homoskedasticity (both [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#) and [Amin \(2011\)](#) assume homoskedasticity, too). The tests are implemented in the same fashion as for the US data. The results are reported in Table 5. We then allow for heteroskedasticity with the results reported in Table 6. The qualitative patterns are very similar to those for the US data. Under the homoskedasticity assumption, our maximal t-test rejects for smaller sample sizes (70 or larger) than the

other tests (110 or larger). Without assuming homoskedasticity, all tests perform similarly and start rejecting around sample sizes of 50 – 70.

Second, we test the null hypothesis of no returns to schooling after removing outliers. As [Amin \(2011\)](#) suggests, we first drop an observation that has the absolute difference in hourly wages  $> 90$ . Then, we sequentially drop further outliers with the absolute difference greater than 75, 65, and 60. In each step, we lose 1 observation. The results are displayed in [Table 7](#). After dispensing with the most extreme outlier, the OLS and IV estimates are no longer significant even at the 10% level. Our maximal t-test, on the contrary, rejects the null hypothesis of no effect at around the 6.5% level. The maximal t-test results remain significant at the 5.6% and 7.7% levels after we drop the second and third most extreme outliers respectively. And thereby our test supports the original conclusion in [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#) that the returns to schooling is significantly different from 0 even when a few extreme outliers are removed from the sample.

As a final robustness check, we also include additional covariates in the regression. The dataset contains various characteristics of the twins. As in [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#) and [Amin \(2011\)](#), we control for marital status, working part-time, region to live and current work experience; see [Table 8](#). The inclusion of these covariates further decreases our maximal t-test’s p-values so that it rejects the null at the 5% level even when all four most extreme observations are removed. The OLS and IV estimates, on the other hand, are only significant in 1 or 2 subsamples at the 10% level.

In conclusion, employing our maximal t-test, which remains powerful even in the presence of nonclassical ME in schooling, we have been able to recover statistically significant returns to schooling even after removing the most extreme observations. This finding is in contrast to those in [Amin \(2011\)](#) and provides further robustness in support of the original findings by [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#).

## A Finiteness of the Optimal Weights

**Assumption 7.** (i)  $E[XY] \neq 0$ ,  $E[ZY] \neq 0$ , and (ii)  $|corr(X^*, (Z - X))| \neq 1$ .

The possibility of  $|a^*| = \infty$  can be ruled out under a set of weak assumptions. The first part implies that  $\beta \neq 0$ . It also rules out the cases in which the MEs are perfectly negatively correlated with the true variable ( $E[(X^*)^2] = -E[X^*U] = -E[X^*V]$ ). The last part means that the true variable is not recovered from  $Z - X$ .

**Lemma 1.** Suppose (1) and Assumption 1 and 7 hold. Then, the maximal weight  $a^*$  is finite.

$$|a^*| < \infty$$

**Proof of Lemma 1.** The numerator of  $a^*$  in Theorem 1 is finite if the denominator is nonzero, then  $|a^*| < \infty$ .

$$C := E[X(Z - X)]E[ZY] - E[Z(Z - X)]E[XY]$$

By Assumption 1 (i),  $E[XY] = \beta E[XX^*]$  and  $E[ZY] = \beta E[ZX^*]$ . Then  $C$  trivially becomes zero when  $\beta = 0$  or  $E[(X^*)^2] + E[X^*U] = E[(X^*)^2] + E[X^*V] = 0$ . These are ruled out by Assumption 7. The other cases in which  $C$  is zero are following.

1.  $E[Z(Z - X)] = 0$  and  $E[X(Z - X)] = 0$
2.  $\frac{E[X(Z - X)]}{E[Z(Z - X)]} = \frac{E[XY]}{E[ZY]} = \frac{E[(X^*)^2] + E[X^*U]}{E[(X^*)^2] + E[X^*V]}$

The first case implies that  $X = Z$  and is ruled out by (iii) of Assumption 1. The second case indicates that the ratio of  $E[X(Z - X)]$  to  $E[Z(Z - X)]$  is equal to the ratio of  $E[(X^*)^2] + E[X^*U]$  to  $E[(X^*)^2] + E[X^*V]$ . Therefore, for a constant  $\gamma \neq 0$ ,

$$E[XZ - X^2] = \gamma(E[(X^*)^2] + E[X^*U]), \quad E[Z^2 - XZ] = \gamma(E[(X^*)^2] + E[X^*V]).$$

By simple algebra,

$$E[X^*V] - (1 + \gamma)E[X^*U] + E[UV] - E[U^2] = \gamma E[(X^*)^2]$$

$$(1 - \gamma)E[X^*V] - E[X^*U] + E[V^2] - E[UV] = \gamma E[(X^*)^2]$$

and therefore,

$$\gamma(E[X^*V] - E[X^*U]) = E[V^2] - 2E[UV] + E[U^2]$$

and this implies that

$$\gamma E[X^*(V - U)] = E[(V - U)^2] \implies X^* = \frac{1}{\gamma}(V - U) = \frac{1}{\gamma}(Z - X).$$

As Assumption 7 rules out perfect positive and negative correlation between  $X^*$  and  $Z - X$ , such a possibility is ruled out. Therefore, the denominator  $C$  is non-zero and hence  $a^*$  is finite. Q.E.D.

## B Proofs

### B.1 Proofs for Section 2

**Proof of Theorem 1.** First, we show that

$$\sup_{\omega \in \mathbb{R}^4} t(\omega)^2 = \sup_{a \in \mathbb{R}} t(a, 1 - a, a, 1 - a)^2. \quad (9)$$

To this end, notice that when  $(\omega_3, \omega_4) = (\omega_1, \omega_2)$

$$\beta(\omega_1, \omega_2, \omega_1, \omega_2) = \frac{E[W(\omega_1, \omega_2)Y]}{E[W(\omega_1, \omega_2)^2]}, \quad \sigma^2(\omega_1, \omega_2, \omega_1, \omega_2) = \frac{\sigma_\varepsilon^2(\omega_1, \omega_2, \omega_1, \omega_2)}{E[W(\omega_1, \omega_2)^2]}$$

and the t-ratio becomes

$$t(\omega_1, \omega_2, \omega_1, \omega_2) = \frac{E[W(\omega_1, \omega_2)Y]}{\sigma_\varepsilon(\omega_1, \omega_2, \omega_1, \omega_2) \sqrt{E[W(\omega_1, \omega_2)^2]}}.$$

Since  $\beta(0, 0, \omega_3, \omega_4) = 0$  for any  $\omega_3, \omega_4 \in \mathbb{R}^2$ ,  $\omega_1 = \omega_2 = 0$  cannot be a maximizer of the t-ratio. Therefore, in the remainder of the proof we assume without loss of generality that  $(\omega_1, \omega_2) \neq (0, 0)$ . By Assumption 1(ii), there is no linear combination of  $X$  and  $Z$  that is equal to zero so that, together with the positive variance Assumption 1(i), we have  $E[W(a, b)^2] \neq 0$  for all  $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ . Therefore, the residual variance  $E[(Y - \beta W(\omega_1, \omega_2))^2]$  is minimized at the OLS estimand  $\beta = E[W(\omega_1, \omega_2)Y]/E[W(\omega_1, \omega_2)^2]$ , which in turn implies  $t^2(\omega_1, \omega_2, \omega_3, \omega_4) \leq t^2(\omega_1, \omega_2, \omega_1, \omega_2)$  for all  $\omega \in \mathbb{R}^4$ ,  $(\omega_1, \omega_2) \neq (0, 0)$ , with equality when  $(\omega_1, \omega_2) = (\omega_3, \omega_4)$ . Therefore,

$$\sup_{\omega \in \mathbb{R}^4} t^2(\omega_1, \omega_2, \omega_3, \omega_4) = \sup_{(\omega_1, \omega_2) \in \mathbb{R}^2} t^2(\omega_1, \omega_2, \omega_1, \omega_2) \quad \forall \omega \in \mathbb{R}^4.$$

Next, notice that

$$\beta \left( \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2}, \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2} \right) = \beta(\omega_1, \omega_2, \omega_1, \omega_2)(\omega_1 + \omega_2)$$

and

$$\sigma^2 \left( \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2}, \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2} \right) = \sigma^2(\omega_1, \omega_2, \omega_1, \omega_2)(\omega_1 + \omega_2)^2.$$

Therefore, the weights in the t-ratio can be standardized to sum up to one,

$$t^2 \left( \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2}, \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2} \right) = t^2(\omega_1, \omega_2, \omega_1, \omega_2),$$

and (9) follows.

By a slight abuse of notation denote  $t(a) := t(a, 1 - a, a, 1 - a)$ .  $t(a)$  is a twice continuously differentiable function. To find the maximizer of  $t(a)^2$  we consider first the first-order condition with respect to  $a$ ,

$$\frac{\partial}{\partial a} t(a)^2 = 0. \quad (10)$$

Suppose  $C \neq 0$ , then this equation has two solutions,

$$a_1^* := \frac{E[YZ]}{E[YZ] - E[XZ]}$$

and

$$a_2^* := \frac{E[XZ]E[ZY] - E[Z^2]E[XY]}{C}.$$

The first solution yields  $t(a_1^*)^2 = 0$  and, therefore, cannot be the supremum. The second solution yields

$$t(a_2^*)^2 = n \frac{A}{B}.$$

By the Cauchy Schwarz inequality,  $(E[XZ])^2 \leq E[Z^2]E[X^2]$ , but equality is ruled out by Assumption 1(ii). Since  $A \geq 0$  and  $E[Y^2] > 0$ , we have  $B > 0$ . It remains to show that this is indeed the supremum. For  $\beta = 0$ , the t-ratio equals zero and is independent of the weight  $a$ . Therefore, we can restrict attention to the case  $\beta \neq 0$ . First, consider the second derivative at the postulated maximizer:

$$\left. \frac{\partial^2}{\partial a^2} t(a)^2 \right|_{a=a_2^*} = -\frac{2E[Y^2]C^4}{AB^2} < 0.$$



The inequality follows because  $A > 0$  whenever  $\beta \neq 0$ , otherwise  $\text{Var}(Z - \lambda X) = 0$  for  $\lambda = E(ZY)/E(XY)$  which is ruled out by Assumption 1(ii). Also,  $B > 0$ ,  $C \neq 0$ , and Assumption 1(i) guarantees that  $E(Y^2) > 0$ . Therefore,  $a_2^*$  is a local maximizer. Since the t-ratio is a continuously differentiable function,  $a_2^*$  is also a global maximizer of the squared t-ratio as long as the squared t-ratio's limit as  $a \rightarrow \pm\infty$  is smaller than (or equal to) its value at  $a_2^*$ . One can show that the two limits are equal and that they are smaller than (or equal to) the squared t-ratio at  $a_2^*$  if the following condition holds:

$$-\frac{E[Y^2]C^2}{BD} \geq 0 \quad (11)$$

where

$$D := (E[Y(X - Z)])^2 - E[Y^2]E[(X - Z)^2].$$

By the Cauchy-Schwarz inequality, we have  $D \leq 0$ , but equality is ruled out by Assumption 1(iii), so  $D < 0$ .  $D < 0$  and  $B > 0$  then imply that (11) holds and we thus have shown that  $a_2^*$  is indeed the global maximizer of the squared t-ratio when  $C \neq 0$ . We now turn to the case  $C = 0$ . In this case, the first-order condition (10) has only one solution,

$$a_3^* := \frac{E(ZY) [E(XZ)E(ZY) - E(XY)E(Z^2)]}{E(XY)(E(XY) - 2E(ZY))E(Z^2) + (2(XZ) - E(X^2))(EZY)^2},$$

which, when substituted back into the t-ratio yields  $t(a_3^*) = 0$  and therefore cannot be a maximizer. Therefore, the maximum of the t-ratio is

$$\lim_{a \rightarrow \infty} t(a)^2 = \lim_{a \rightarrow -\infty} t(a)^2 = -\frac{(E[(X - Z)Y])^2}{D} > 0$$

which is also equal to  $t(a_2^*)^2$  because  $a_2^* = \infty$  when  $C = 0$ . In conclusion,  $a_2^*$  is the global maximizer of the squared t-ratio for any value of  $C$ . The expression for  $t(a_2^*)$  follows simply by evaluating the t-ratio at the global maximizer and some straight-forward algebra. Q.E.D.

**Proof of Theorem 2.** The t-ratio can be written as

$$\bar{t}(\omega) = \frac{\tilde{\beta}(\omega)}{\bar{\sigma}(\omega)} = \frac{E[\tilde{W}(\omega_3, \omega_4)\tilde{Y}]}{\sqrt{E[\tilde{\varepsilon}(\omega)^2\tilde{W}(\omega_3, \omega_4)^2]}}.$$

Define  $\tilde{\omega} = \left( \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2}, \frac{\omega_3}{\omega_3 + \omega_4}, \frac{\omega_4}{\omega_3 + \omega_4} \right)$ . Then,

$$\begin{aligned}
\tilde{\varepsilon}(\tilde{\omega}) &= \tilde{Y} - \tilde{\beta}(\tilde{\omega})\tilde{W} \left( \frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2} \right) \\
&= \tilde{Y} - \frac{E \left[ \frac{\tilde{W}(\omega_3, \omega_4)}{\omega_3 + \omega_4} \tilde{Y} \right]}{E \left[ \frac{\tilde{W}(\omega_1, \omega_2)}{\omega_1 + \omega_2} \frac{\tilde{W}(\omega_3, \omega_4)}{\omega_3 + \omega_4} \right]} \frac{\tilde{W}(\omega_1, \omega_2)}{\omega_1 + \omega_2} \\
&= \tilde{Y} - \tilde{\beta}(\omega)\tilde{W}(\omega_1, \omega_2) \\
&= \tilde{\varepsilon}(\omega)
\end{aligned}$$

and therefore it is obvious that

$$\bar{t}(\tilde{\omega}) = \frac{\frac{E[\tilde{W}(\omega_3, \omega_4)\tilde{Y}]}{\omega_3 + \omega_4}}{\sqrt{E \left[ \tilde{\varepsilon}(\omega)^2 \left[ \frac{\tilde{W}(\omega_3, \omega_4)}{\omega_3 + \omega_4} \right]^2 \right]}} = \frac{E[\tilde{W}(\omega_3, \omega_4)\tilde{Y}]}{\sqrt{E[\tilde{\varepsilon}(\omega)^2 \tilde{W}(\omega_3, \omega_4)^2]}} = \bar{t}(\omega).$$

Define that  $a_1 := \frac{\omega_1}{\omega_1 + \omega_2}$  and  $a_2 := \frac{\omega_3}{\omega_3 + \omega_4}$ . Then,

$$\bar{t}(\omega) = \bar{t}(a_1, 1 - a_1, a_2, 1 - a_2).$$

Q.E.D.

The following lemma is a slightly modified version of results in [Black, Berger, and Scott \(2000\)](#):

**Lemma 2.** 1. Under Assumptions [2](#), [3](#), and [5](#), if  $\beta \geq 0$ ,

$$\max_{a=0,1} \beta_{OLS}(a) \leq \beta, \tag{12}$$

and, if  $\beta \leq 0$ , then

$$\beta \leq \min_{a=0,1} \beta_{OLS}(a). \tag{13}$$

2. Under Assumptions [2](#), [3](#), and [6](#), if  $\beta \geq 0$ ,

$$\beta \leq \min_{a=0,1} \beta_{OLS-INV}(a), \tag{14}$$

and, if  $\beta \leq 0$ , then

$$\max_{a=0,1} \beta_{OLS-INV}(a) \leq \beta. \tag{15}$$

3. Under Assumptions 2 and 4, if  $\beta \geq 0$ ,

$$\beta \leq \min_{a=0,1} \beta_{IV}(a, 1-a), \quad (16)$$

and, if  $\beta \leq 0$ , then

$$\max_{a=0,1} \beta_{IV}(a, 1-a) \leq \beta. \quad (17)$$

*Proof.* First, notice that by Assumption 2,

$$\beta_{OLS}(a) = \beta \cdot \frac{E[X^*W(a)]}{E[X^*W(a)] + E[U(a)W(a)]} \quad (18)$$

where  $U(a) := aU + (1-a)V$ . By Assumptions 3 and 5,  $E[X^*W(a)] > 0$  and  $E[U(a)W(a)] \geq 0$  for  $a = 0, 1$ , so

$$\frac{E[X^*W(a)]}{E[X^*W(a)] + E[U(a)W(a)]} \leq 1$$

and the lower bound in (14) and the upper bound in (15) follow. Now, notice that Assumption 2 also implies

$$\beta_{OLS-INV}(a) = \frac{\beta^2 E[(X^*)^2] + E[\varepsilon^2]}{\beta E[X^*W(a)]} \quad (19)$$

By Assumptions 3 and 6, if  $\beta \geq 0$ , then

$$\beta_{OLS-INV}(a) \geq \frac{\beta^2 E[(X^*)^2]}{\beta E[X^*W(a)]} \geq \beta,$$

yielding the upper bound in (14). The lower bound in (15) for  $\beta \leq 0$  follows from the same argument by reversing the above inequalities.

Now, we show that the IV bounds hold. Notice that by Assumption 2,

$$\beta_{IV}(a, 1-a) = \beta \cdot \frac{E[X^*W(1-a)]}{E[W(a)W(1-a)]}. \quad (20)$$

Assumption 4 then implies that, for  $a = 0, 1$ , the numerator and denominator are both positive and the latter is not larger than the former, so the IV bounds (16) and (17) follow. Q.E.D.

**Proof of Theorem 3.** To simplify the presentation we use the notation  $\sigma_{AB}$  for  $E[AB]$  and  $\sigma_A^2$  for  $E[A^2]$ , where  $A$  and  $B$  are random variables. Consider the first statement of

the theorem. First, notice that the OLS bound holds for  $a \in \{0, 1\}$  by Lemma 2. From the proof of Lemma 2, we know that  $\beta_{OLS}(a) \leq \beta$  holds if, and only if,

$$\frac{E[X^*W(a)]}{E[X^*W(a)] + E[U(a)W(a)]} \leq 1. \quad (21)$$

To show that this inequality holds only for  $a \in \{0, 1\}$  and not for any other value of  $a$ , we assume that  $a \notin \{0, 1\}$  and then construct a data-generating process characterized by

$$(\sigma_{\varepsilon X^*}, \sigma_{\varepsilon U}, \sigma_{\varepsilon V}, \sigma_{X^*}^2, \sigma_U^2, \sigma_V^2, \sigma_{UV}, \sigma_{X^*U}, \sigma_{X^*V})$$

which satisfies all assumptions of the theorem, but violates (21).

Suppose  $0 < a < 1$ . Let  $\epsilon > 0$  and  $\sigma_U^2 > 0$ . Set  $\sigma_{UV} = \sigma_{X^*U} = \sigma_{\varepsilon X^*} = \sigma_{\varepsilon U} = \sigma_{\varepsilon V} = 0$ ,  $\sigma_{X^*}^2 = \sigma_V^2 = a\sigma_U^2/(1-a) + \epsilon$ , and  $\sigma_{X^*V} = -\sigma_V^2 + a\epsilon/2$ . Then Assumption 2 holds. Since  $\sigma_{X^*V} + \sigma_{UV} \leq 0$ ,  $\sigma_{X^*U} + \sigma_{UV} \leq 0$ , and

$$E[XZ] = \sigma_{X^*}^2 + \sigma_{X^*U} + \sigma_{X^*V} + \sigma_{UV} = \frac{a\epsilon}{2} > 0,$$

Assumption 4 holds, which implies Assumption 3. Assumption 5 is satisfied by definition of the above quantities. We also have

$$E[X^*W(a)] = aE[X^*X] + (1-a)E[X^*Z] > 0.$$

By the definition of  $\sigma_V^2$ ,

$$\sigma_V^2 = \frac{a^2\sigma_U^2 + (1-a)^2\sigma_V^2}{1-a} + a\epsilon,$$

so that

$$(1-a)\sigma_{X^*V} < -(a^2\sigma_U^2 + (1-a)^2\sigma_V^2)$$

and thus

$$E[U(a)W(a)] = a\sigma_U^2 + (1-a)^2\sigma_V^2 + (1-a)\sigma_{X^*V} < 0.$$

Therefore, (21) is violated.

Suppose  $a < 0$ . Let  $\epsilon > 0$ ,  $\delta > 0$ , and  $\sigma_U^2 > 0$ . If  $a < -2$ , then pick  $\delta$  small enough so that  $\delta \leq -4\epsilon/(2+a)$ . Set  $\sigma_{X^*U} = \sigma_{\varepsilon X^*} = \sigma_{\varepsilon U} = \sigma_{\varepsilon V} = 0$ ,  $\sigma_{UV} = -a\sigma_U^2/(2(1-a)) + \epsilon$ ,  $\sigma_V^2 = 2\sigma_{UV} + a\sigma_U^2/(1-a) - \delta$ ,  $\sigma_{X^*V} = -\sigma_V^2 - \delta a/2$ , and  $\sigma_{X^*}^2 = \sigma_V^2(1-a)$ .

Then Assumption 2 holds. Assumption 3 holds because  $E[XX^*] = \sigma_{X^*}^2 + \sigma_{X^*U} = \sigma_{X^*}^2 > 0$  and

$$E[ZX^*] = \sigma_{X^*}^2 + \sigma_{X^*V} = \sigma_V^2(1-a) - \sigma_V^2 - \frac{\delta a}{2} > 0.$$

Assumption 5 is satisfied by definition of the above quantities. We also have

$$E[X^*W(a)] = \sigma_{X^*}^2 + a\sigma_{X^*U} + (1-a)\sigma_{X^*V} = \sigma_{X^*}^2 - (1-a)\sigma_V^2 - \frac{(1-a)a\delta}{2} = -\frac{(1-a)a\delta}{2} > 0.$$

By the definition of  $\sigma_V^2$ ,

$$\sigma_V^2 = \frac{a^2\sigma_U^2 + (1-a)^2\sigma_V^2 + 2a(1-a)\sigma_{UV}}{1-a},$$

so that

$$(1-a)\sigma_{X^*V} < -(a^2\sigma_U^2 + (1-a)^2\sigma_V^2 + 2a(1-a)\sigma_{UV})$$

and thus

$$E[U(a)W(a)] = a\sigma_U^2 + (1-a)^2\sigma_V^2 + 2a(1-a)\sigma_{UV} + (1-a)\sigma_{X^*V} < 0.$$

Therefore, (21) is violated.

In the case  $a > 1$ , we can construct a data-generating process violating (21) in a similar fashion as in the case  $a < 0$ .

Now, consider the second statement of the theorem. From the proof of Lemma 2, we know that the inverse OLS bound holds if, and only if,

$$\frac{\beta^2 E[(X^*)^2] + E[\varepsilon^2]}{\beta E[X^*W(a)]} \geq \beta. \quad (22)$$

This inequality holds for all  $a \in [0, 1]$  by the same reasoning as in the proof of Lemma 2 because  $E[X^*W(a)] > E[(X^*)^2]$  for all  $a \in [0, 1]$ .

To show that (22) does not hold for any  $a$  outside the interval  $[0, 1]$ , we construct a data-generating process that satisfies all assumptions of the theorem but violates (22). If  $a < 0$ , pick  $\sigma_{X^*}^2 > 0$ ,  $\sigma_{X^*U} \leq 0$ ,  $\sigma_{X^*V} \leq 0$ , such that  $E[X^*X] > (1 - 1/a)E[X^*Z]$ ,  $\sigma_{X^*}^2 + \sigma_{X^*U} > 0$ , and  $\sigma_{X^*}^2 + \sigma_{X^*V} > 0$ . If  $a > 1$ , then pick  $\sigma_{X^*}^2 > 0$ ,  $\sigma_{X^*U} \leq 0$ ,  $\sigma_{X^*V} \leq 0$  such that  $E[X^*X] < (1 - 1/a)E[X^*Z]$ ,  $\sigma_{X^*}^2 + \sigma_{X^*U} > 0$ , and  $\sigma_{X^*}^2 + \sigma_{X^*V} > 0$ . In both cases, Assumptions 3 and 6 hold and Assumption 2 can, as above, be satisfied by simply choosing the three covariances to be zero. However,

$$E[X^*W(a)] = aE[XX^*] + (1-a)E[ZX^*] < 0$$

and thus (22) is violated.

Finally, consider the third statement of the theorem. The IV bound holds if, and only if,

$$\frac{E[X^*W(1-a)]}{E[W(a)W(1-a)]} \geq 1. \quad (23)$$

Suppose  $0 < a < 1$ . Set  $\sigma_{UV} = \sigma_{X^*V} = \sigma_{X^*U} = \sigma_{\varepsilon X^*} = \sigma_{\varepsilon U} = \sigma_{\varepsilon V} = 0$  and let  $\sigma_U^2, \sigma_V^2, \sigma_{X^*}^2$  any positive values. Then Assumptions 2 and 4 hold. However,

$$E[X^*W(1-a)] = \sigma_{X^*}^2 + E[X^*U(1-a)] = \sigma_{X^*}^2 > 0$$

and

$$E[U(a)W(1-a)] = a(1-a)(\sigma_U^2 + \sigma_V^2) > 0,$$

which means the bound (23) is violated.

Suppose  $a < 0$ . Set  $\sigma_{UV} = \sigma_{X^*V} = \sigma_{\varepsilon X^*} = \sigma_{\varepsilon U} = \sigma_{\varepsilon V} = 0$ . Pick  $\sigma_U^2, \sigma_V^2 > 0$  and  $\sigma_{X^*U}$  small enough so it satisfies  $\sigma_{X^*U} < -(1-a)(\sigma_U^2 + \sigma_V^2)$ . Choose  $\sigma_{X^*}^2 > -(1-a)\sigma_{X^*U}$ . Then Assumptions 2 and 4 are satisfied because

$$E[XZ] > a\sigma_{X^*U} > 0$$

and  $\sigma_{X^*U} + \sigma_{UV} \leq 0$  and  $\sigma_{X^*V} + \sigma_{UV} \leq 0$ . However,

$$E[X^*W(1-a)] = \sigma_{X^*}^2 + (1-a)\sigma_{X^*U} + a\sigma_{X^*V} > 0$$

and

$$E[U(a)W(1-a)] = a(1-a)(\sigma_U^2 + \sigma_V^2) + a\sigma_{X^*U} > 0,$$

which means, again, the bound (23) is violated. In the case  $a > 1$ , such a violation of (23) can be created in a similar fashion. Q.E.D.

$\sigma_{X*U}$	$\sigma_{X*V}$	Test	S1	S2	S3	S4	S5	S6
strong	strong	OLS w/ x	0.054	0.050	0.055	0.040	0.046	0.048
		OLS w/ z	0.039	0.038	0.052	0.050	0.059	0.039
		IV x z	0.018	0.000	0.051	0.000	0.035	0.023
		tmax	0.051	0.050	0.069	0.042	0.056	0.058
		$a^*$	0.507	0.520	0.482	0.491	0.480	0.483
strong	weak	OLS w/ x	0.049	0.051	0.043	0.056	0.049	0.058
		OLS w/ z	0.050	0.062	0.058	0.052	0.033	0.046
		IV x z	0.001	0.027	0.024	0.010	0.028	0.001
		tmax	0.059	0.056	0.056	0.056	0.047	0.057
		$a^*$	0.529	0.509	0.512	0.508	0.507	0.530
weak	strong	OLS w/ x	0.037	0.046	0.042	0.058	0.049	0.042
		OLS w/ z	0.051	0.049	0.047	0.055	0.062	0.052
		IV x z	0.001	0.025	0.016	0.006	0.052	0.001
		tmax	0.048	0.057	0.055	0.061	0.064	0.057
		$a^*$	0.470	0.476	0.477	0.509	0.465	0.461
weak	weak	OLS w/ x	0.051	0.059	0.050	0.048	0.049	0.040
		OLS w/ z	0.058	0.044	0.052	0.058	0.049	0.049
		IV x z	0.015	0.034	0.000	0.035	0.046	0.002
		tmax	0.067	0.056	0.059	0.063	0.051	0.054
		$a^*$	0.471	0.530	0.511	0.475	0.515	0.502

Table 2: Rejection probabilities in the six different scenarios. The values of  $\sigma_{X*U}$  and  $\sigma_{X*V}$  vary between “strong” (-0.7 for scenarios 1–3; -0.5 for scenarios 4–6) and “weak” (-0.3 for all scenarios).  $a^*$  denotes the maximal weight averaged across the MC samples.

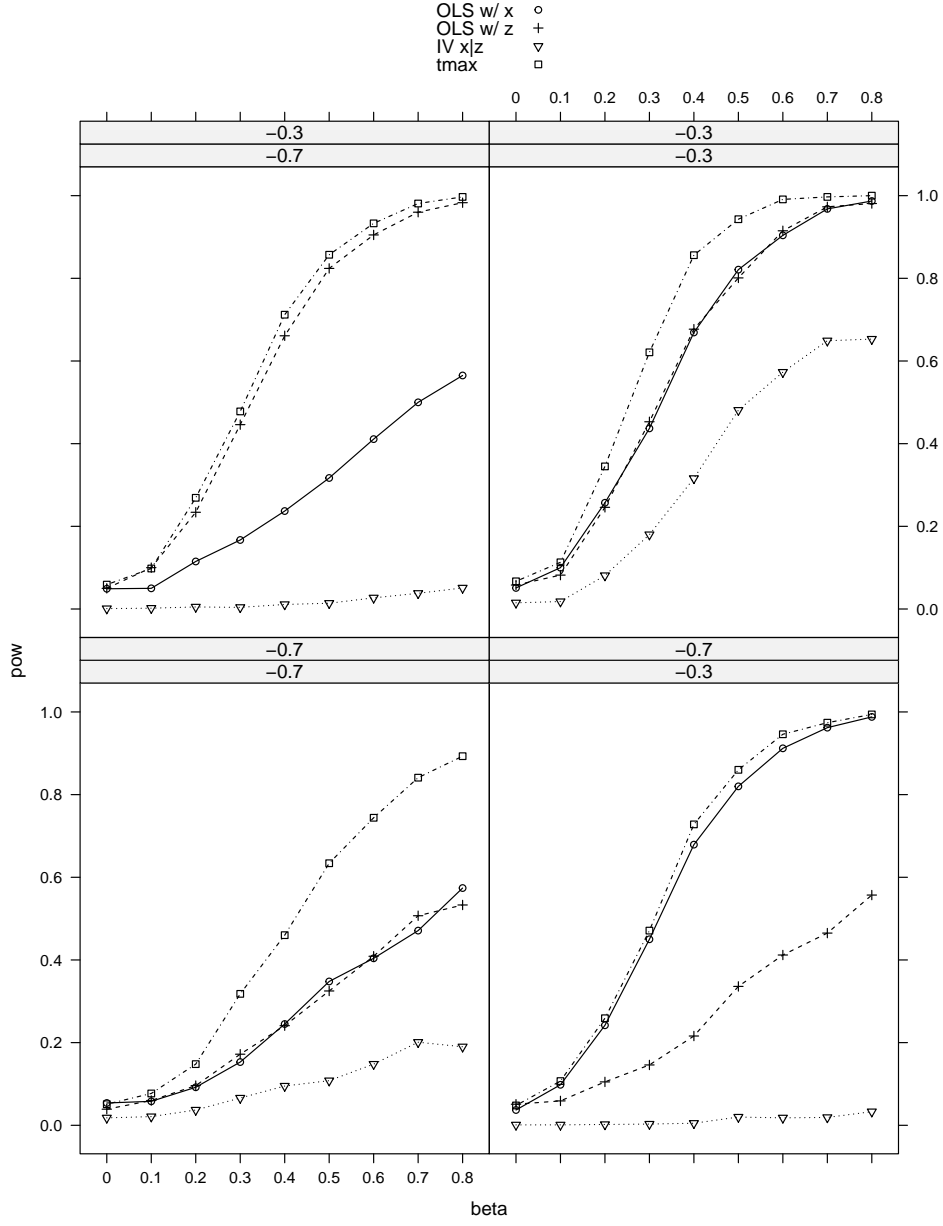


Figure 1: Rejection probabilities in scenario 1, for  $\sigma_{X*V} \in \{-0.3, -0.7\}$  varying from top to bottom and  $\sigma_{X*U} \in \{-0.3, -0.7\}$  varying from left to right.



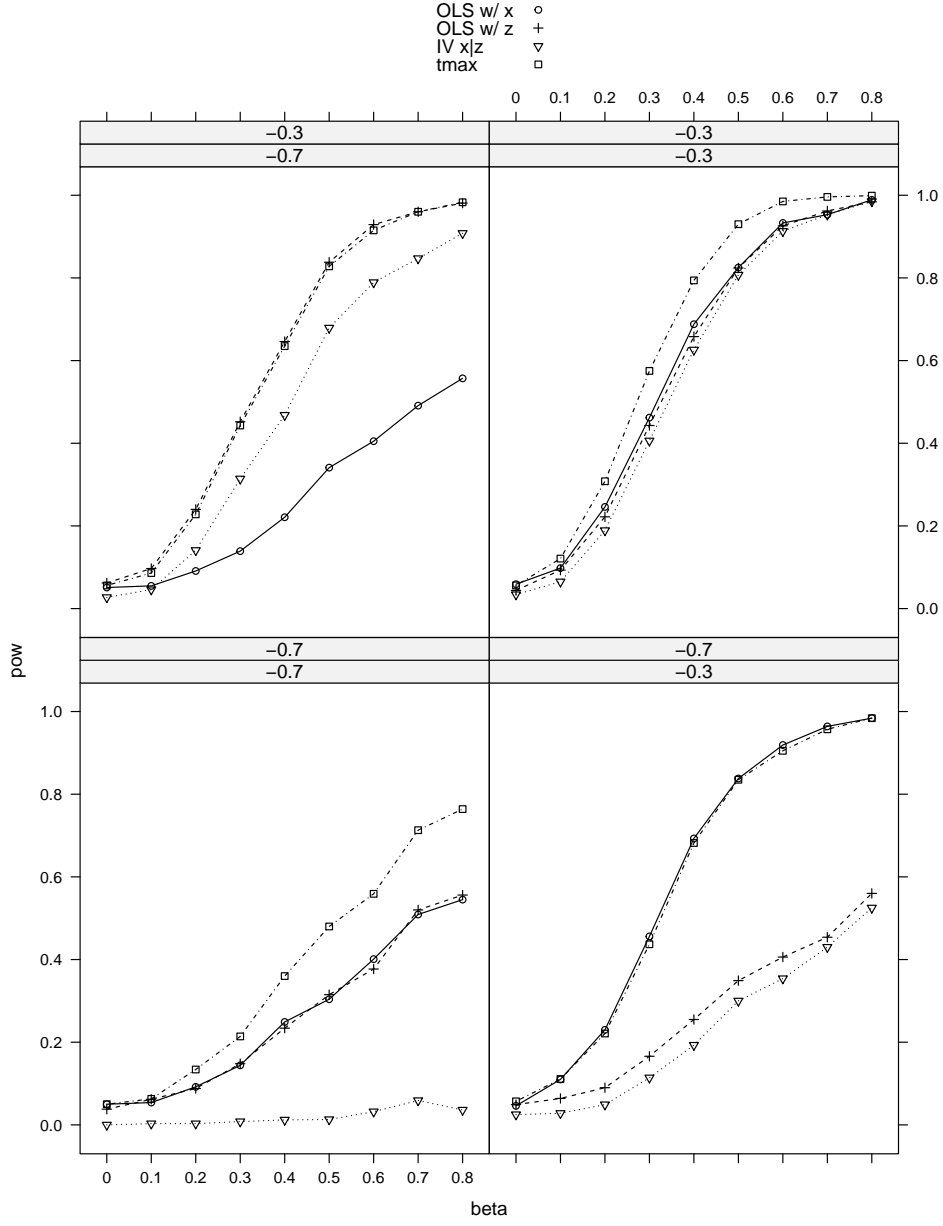


Figure 2: Rejection probabilities in scenario 2, for  $\sigma_{X*V} \in \{-0.3, -0.7\}$  varying from top to bottom and  $\sigma_{X*U} \in \{-0.3, -0.7\}$  varying from left to right.

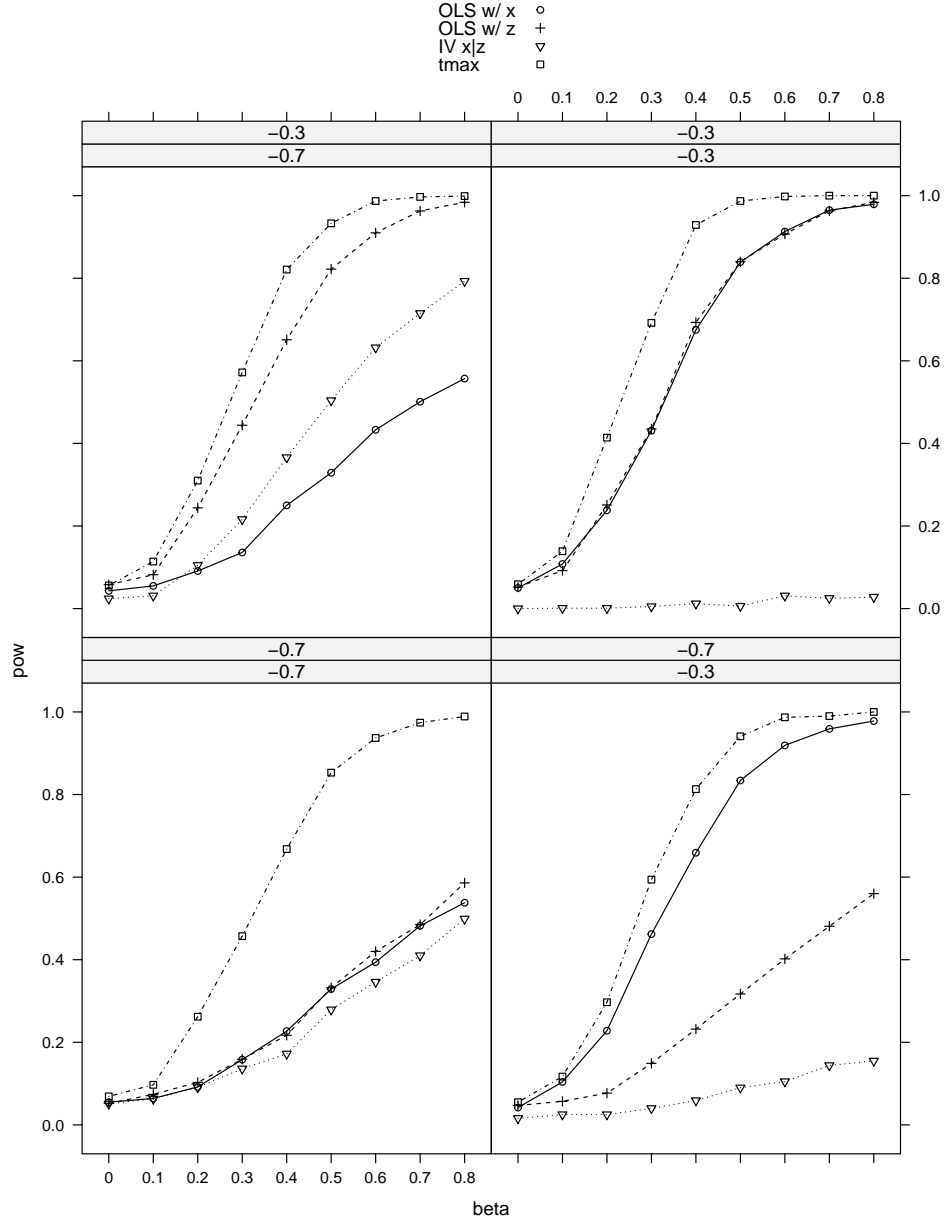


Figure 3: Rejection probabilities in scenario 3, for  $\sigma_{X*V} \in \{-0.3, -0.7\}$  varying from top to bottom and  $\sigma_{X*U} \in \{-0.3, -0.7\}$  varying from left to right.

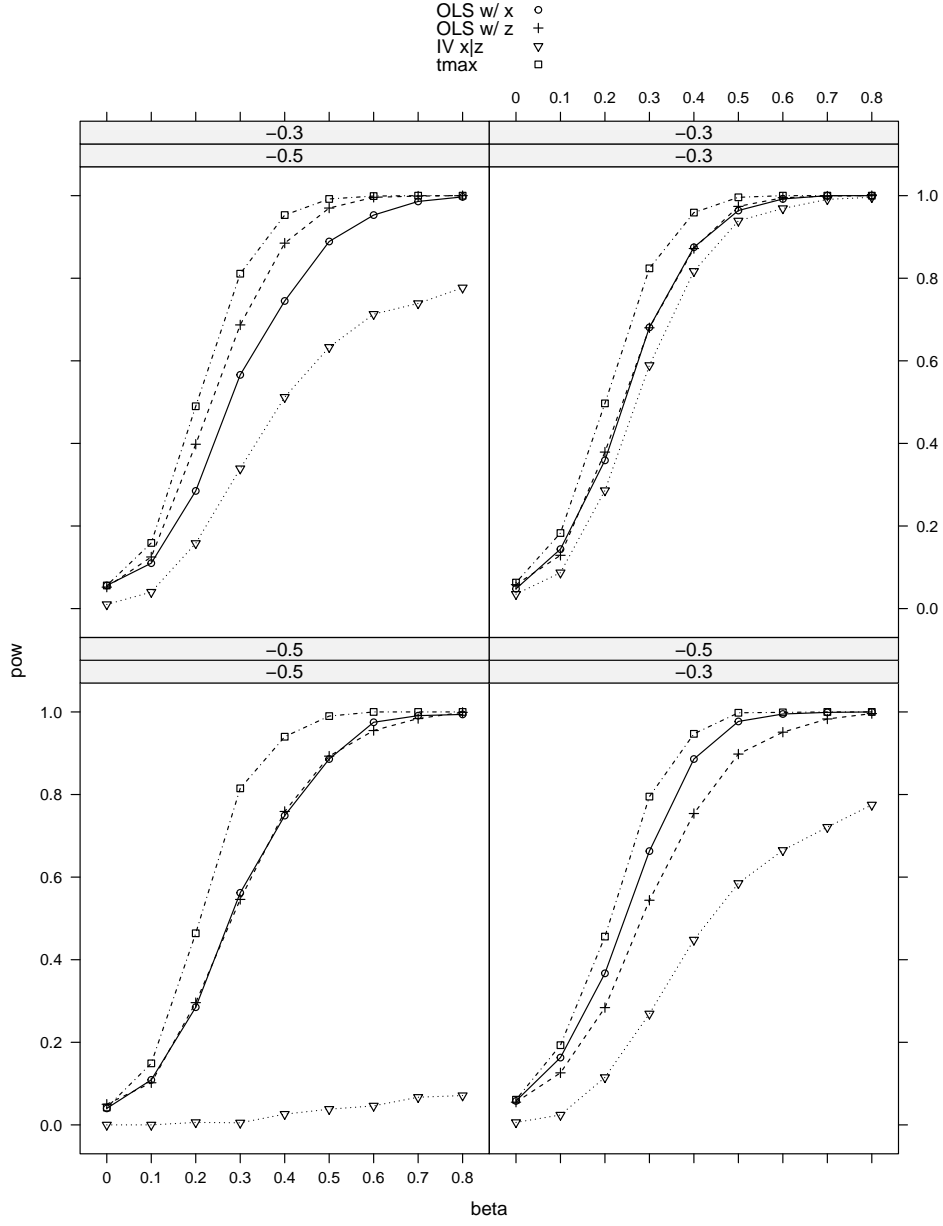


Figure 4: Rejection probabilities in scenario 4, for  $\sigma_{X*V} \in \{-0.3, -0.5\}$  varying from top to bottom and  $\sigma_{X*U} \in \{-0.3, -0.5\}$  varying from left to right.

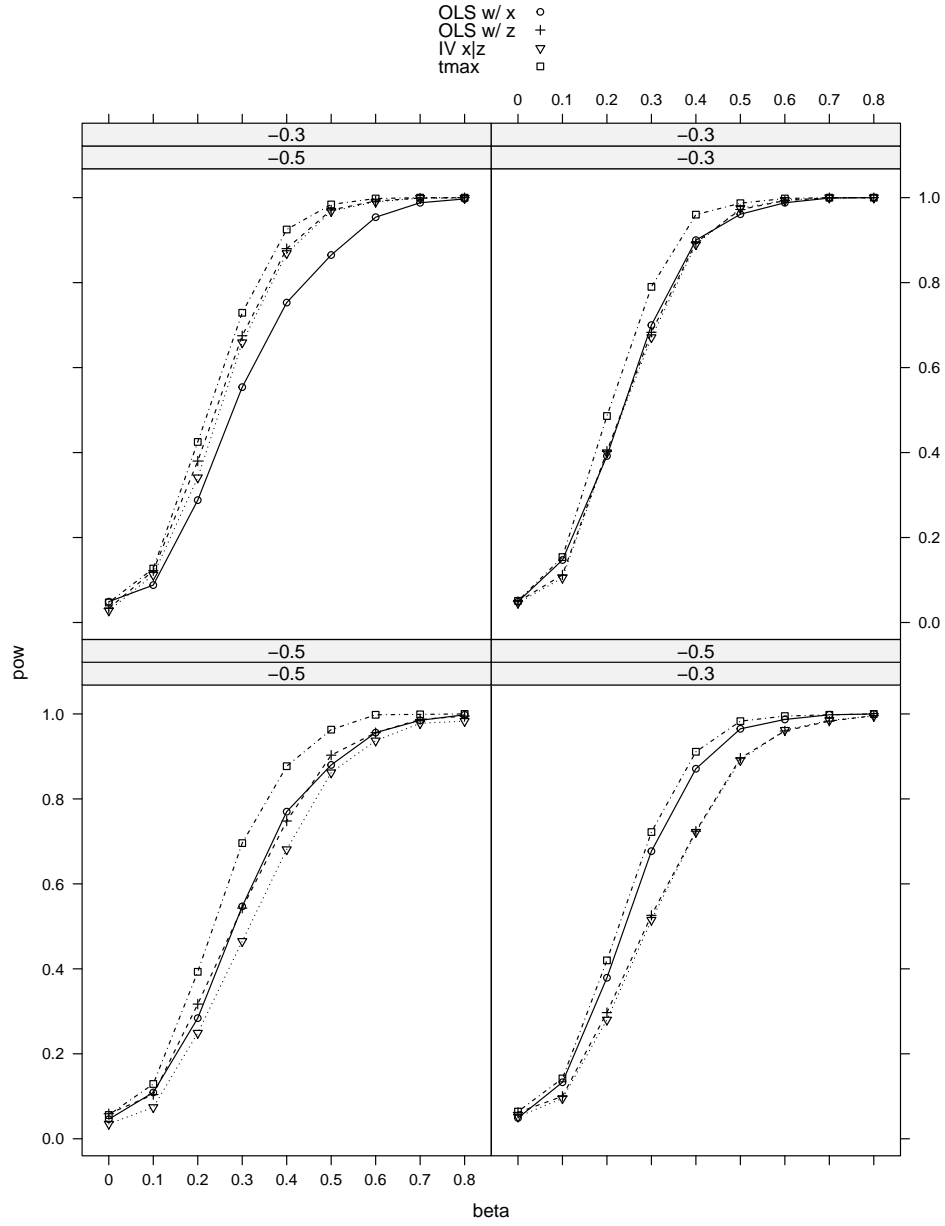


Figure 5: Rejection probabilities in scenario 5, for  $\sigma_{X*V} \in \{-0.3, -0.5\}$  varying from top to bottom and  $\sigma_{X*U} \in \{-0.3, -0.5\}$  varying from left to right.

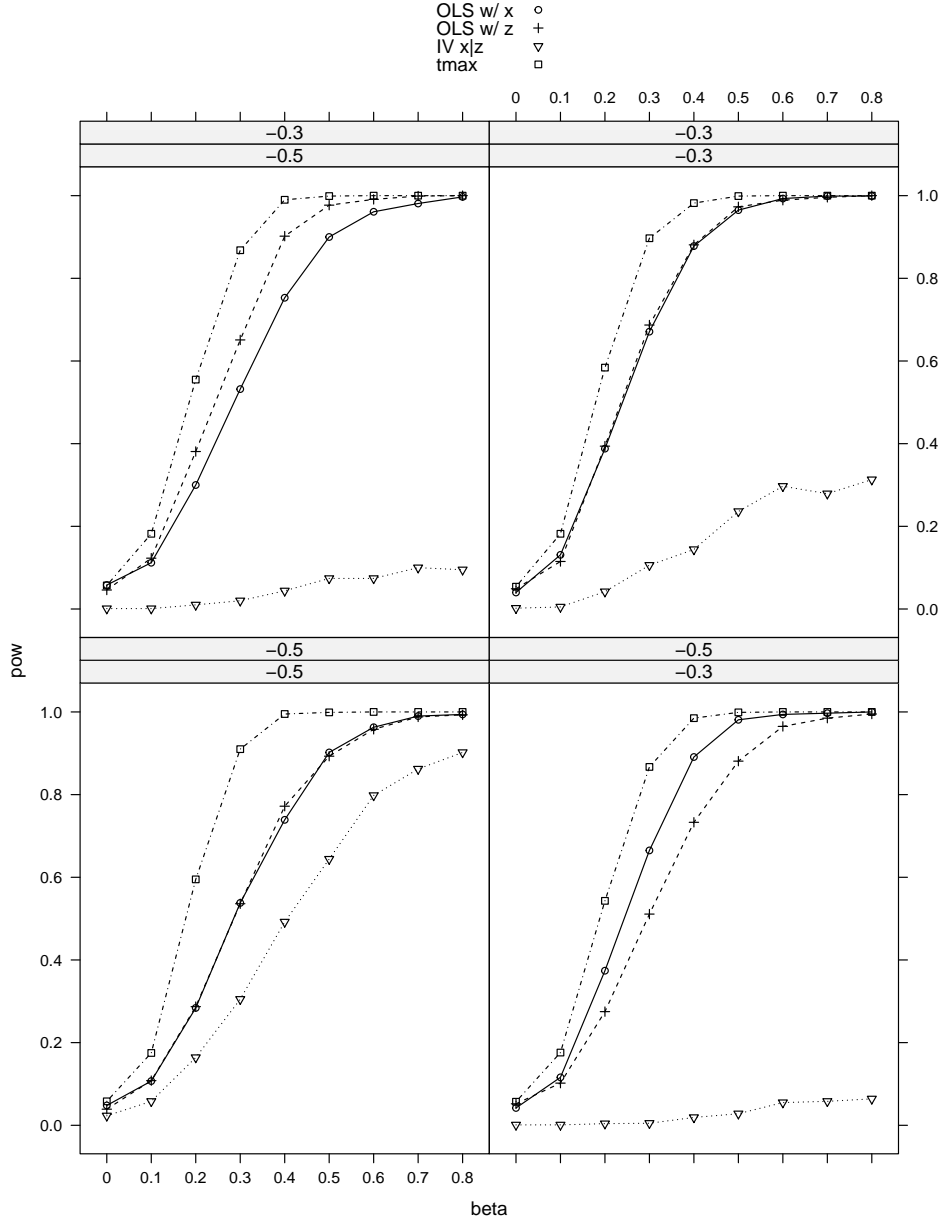


Figure 6: Rejection probabilities in scenario 6, for  $\sigma_{X*V} \in \{-0.3, -0.5\}$  varying from top to bottom and  $\sigma_{X*U} \in \{-0.3, -0.5\}$  varying from left to right.

n	OLS w/ x	OLS w/ z	IV	$\hat{a}^*$	tmax	95% CV	p-value
30	1.39	<b>3.03</b>	1.36	0.23	<b>3.16</b>	2.74	0.03
40	1.61	<b>2.99</b>	1.90	0.22	<b>3.11</b>	2.75	0.02
50	1.61	<b>3.20</b>	1.94	0.22	<b>3.31</b>	2.75	0.01
60	1.82	<b>3.91</b>	<b>2.44</b>	0.17	<b>3.97</b>	2.73	0.00
70	1.60	<b>3.22</b>	<b>2.64</b>	0.07	<b>3.23</b>	2.54	0.01
80	1.75	<b>3.18</b>	<b>2.72</b>	0.11	<b>3.20</b>	2.43	0.01
90	<b>1.99</b>	<b>3.17</b>	<b>2.76</b>	0.23	<b>3.25</b>	2.52	0.01
100	<b>2.21</b>	<b>3.34</b>	<b>2.98</b>	0.23	<b>3.42</b>	2.69	0.01
110	<b>2.49</b>	<b>3.40</b>	<b>3.09</b>	0.31	<b>3.55</b>	2.66	0.01
120	<b>3.09</b>	<b>3.61</b>	<b>3.38</b>	0.40	<b>3.92</b>	2.55	0.00
130	<b>3.25</b>	<b>3.54</b>	<b>3.36</b>	0.45	<b>3.94</b>	2.59	0.00
147	<b>3.69</b>	<b>3.86</b>	<b>3.72</b>	0.47	<b>4.33</b>	2.51	0.00

Table 3: Tests using homoskedastic standard errors applied to first  $n$  observations of the data set from [Ashenfelter and Krueger \(1994\)](#); rejections at the 5% level are highlighted in bold face.

n	OLS w/ x	OLS w/ z	IV	$\hat{a}^*$	tmax	95% CV	p-value
30	1.02	<b>3.31</b>	0.59	0.10	<b>3.42</b>	2.34	0.00
40	1.20	<b>3.44</b>	0.86	0.08	<b>3.50</b>	2.30	0.00
50	1.19	<b>3.50</b>	0.85	0.06	<b>3.51</b>	2.33	0.00
60	1.38	<b>4.26</b>	1.17	0.02	<b>4.27</b>	2.35	0.00
70	1.40	<b>3.18</b>	1.45	0.00	<b>3.18</b>	2.23	0.00
80	1.58	<b>3.47</b>	1.60	0.00	<b>3.47</b>	2.26	0.00
90	1.70	<b>3.30</b>	1.63	0.00	<b>3.30</b>	2.31	0.00
100	1.82	<b>3.30</b>	1.80	0.00	<b>3.30</b>	2.25	0.00
110	<b>1.97</b>	<b>3.31</b>	1.85	0.00	<b>3.31</b>	2.27	0.00
120	<b>2.56</b>	<b>3.62</b>	<b>2.26</b>	0.10	<b>3.64</b>	2.28	0.00
130	<b>2.67</b>	<b>3.53</b>	<b>2.28</b>	0.20	<b>3.59</b>	2.31	0.00
147	<b>3.16</b>	<b>3.80</b>	<b>2.63</b>	0.36	<b>3.95</b>	2.30	0.00

Table 4: Tests using heteroskedasticity robust standard errors applied to first  $n$  observations of the data set from [Ashenfelter and Krueger \(1994\)](#); rejections at the 5% level are highlighted in bold face.

n	OLS w/ x	OLS w/ z	IV	$\hat{a}^*$	tmax	95% CV	p-value
30	0.35	0.22	0.22	0.77	0.36	1.41	0.77
40	0.26	0.24	0.24	0.55	0.30	1.24	0.79
50	0.86	0.77	0.77	0.58	0.94	1.26	0.16
60	1.00	0.94	0.94	0.57	1.08	1.26	0.10
70	1.17	1.32	1.31	0.36	<b>1.38</b>	1.17	0.02
80	1.04	1.35	1.33	0.21	<b>1.37</b>	1.16	0.02
90	1.22	1.24	1.24	0.47	<b>1.37</b>	1.21	0.02
100	1.27	1.23	1.23	0.55	<b>1.36</b>	1.18	0.02
110	<b>1.96</b>	<b>2.61</b>	<b>2.55</b>	0.04	<b>2.62</b>	2.20	0.02
120	<b>1.99</b>	<b>2.76</b>	<b>2.69</b>	0.00	<b>2.76</b>	2.25	0.02
130	<b>2.01</b>	<b>3.31</b>	<b>3.16</b>	0.00	<b>3.31</b>	2.28	0.00
140	<b>1.96</b>	<b>3.19</b>	<b>3.06</b>	0.00	<b>3.19</b>	2.22	0.01
150	<b>2.16</b>	<b>3.05</b>	<b>2.97</b>	0.00	<b>3.05</b>	2.20	0.01
160	<b>2.24</b>	<b>2.95</b>	<b>2.89</b>	0.09	<b>2.96</b>	2.27	0.01
170	<b>2.04</b>	<b>3.19</b>	<b>3.07</b>	0.02	<b>3.20</b>	2.18	0.00
180	1.64	<b>2.79</b>	<b>2.70</b>	0.00	<b>2.79</b>	2.25	0.02
190	<b>2.02</b>	<b>2.92</b>	<b>2.85</b>	0.06	<b>2.92</b>	2.14	0.01
200	<b>1.97</b>	<b>2.73</b>	<b>2.68</b>	0.09	<b>2.73</b>	2.28	0.01
214	1.74	<b>2.37</b>	<b>2.34</b>	0.10	<b>2.38</b>	2.22	0.04

Table 5: Tests using homoskedastic standard errors applied to first  $n$  observations of the data set from [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#); rejections at the 5% level are highlighted in bold face.



n	OLS w/ x	OLS w/ z	IV	$\hat{a}^*$	tmax	95% CV	p-value
30	0.49	0.46	0.48	0.43	0.56	2.25	0.81
40	0.43	0.59	0.61	0.20	0.61	2.23	0.77
50	1.34	1.77	<b>2.00</b>	0.28	1.93	2.27	0.11
60	1.58	<b>2.16</b>	<b>2.39</b>	0.27	<b>2.31</b>	2.23	0.04
70	<b>1.98</b>	<b>3.13</b>	<b>3.16</b>	0.23	<b>3.30</b>	2.28	0.00
80	1.78	<b>3.21</b>	<b>3.28</b>	0.18	<b>3.31</b>	2.29	0.00
90	<b>2.03</b>	<b>2.83</b>	<b>2.96</b>	0.26	<b>3.09</b>	2.28	0.01
100	<b>2.22</b>	<b>2.73</b>	<b>2.82</b>	0.29	<b>2.94</b>	2.29	0.01
110	<b>2.80</b>	<b>2.30</b>	<b>2.16</b>	0.94	<b>2.81</b>	2.24	0.01
120	<b>2.80</b>	<b>2.38</b>	<b>2.24</b>	0.87	<b>2.83</b>	2.18	0.01
130	<b>3.03</b>	<b>2.86</b>	<b>2.57</b>	0.75	<b>3.17</b>	2.23	0.00
140	<b>2.92</b>	<b>2.80</b>	<b>2.52</b>	0.72	<b>3.08</b>	2.22	0.00
150	<b>2.90</b>	<b>2.67</b>	<b>2.43</b>	0.73	<b>3.04</b>	2.24	0.01
160	<b>3.01</b>	<b>2.56</b>	<b>2.35</b>	0.82	<b>3.08</b>	2.26	0.00
170	<b>2.76</b>	<b>2.90</b>	<b>2.48</b>	0.60	<b>3.07</b>	2.27	0.01
180	<b>2.11</b>	<b>2.46</b>	<b>2.18</b>	0.34	<b>2.50</b>	2.19	0.02
190	<b>2.59</b>	<b>2.65</b>	<b>2.36</b>	0.60	<b>2.85</b>	2.22	0.01
200	<b>2.49</b>	<b>2.42</b>	<b>2.22</b>	0.67	<b>2.67</b>	2.22	0.02
214	<b>2.20</b>	<b>2.12</b>	<b>1.99</b>	0.67	<b>2.35</b>	2.22	0.04

Table 6: Tests using heteroskedasticity robust standard errors applied to first  $n$  observations of the data set from [Bonjour, Cherkas, Haskel, Hawkes, and Spector \(2003\)](#); rejections at the 5% level are highlighted in bold face.

	OLS	IV	$\hat{a}^*$	tmax
Full sample	1.744	2.340	0.10	2.376
( $n = 214$ )	(0.083)	(0.020)		(0.034)
abs. wage dif. < 90	1.495	1.616	0.39	1.705
( $n = 213$ )	(0.136)	(0.108)		(0.065)
abs. wage dif. < 75	1.528	1.651	0.39	1.742
( $n = 212$ )	(0.128)	(0.100)		(0.056)
abs. wage dif. < 65	1.572	1.228	0.84	1.584
( $n = 211$ )	(0.118)	(0.221)		(0.077)
abs. wage dif. < 60	1.496	1.313	0.68	1.545
( $n = 210$ )	(0.136)	(0.191)		(0.114)

Table 7: Robustness of the results when outliers are removed; the numbers in parentheses are p-values.

	OLS	IV	$\hat{a}^*$	tmax
Full sample	1.628	2.312	0.02	2.350
( $n = 214$ )	(0.105)	(0.022)		(0.037)
abs. wage dif. < 90	1.495	1.630	0.38	1.714
( $n = 213$ )	(0.137)	(0.105)		(0.065)
abs. wage dif. < 75	1.622	1.733	0.41	1.837
( $n = 212$ )	(0.107)	(0.085)		(0.050)
abs. wage dif. < 65	1.715	1.288	0.90	1.719
( $n = 211$ )	(0.088)	(0.199)		(0.047)
abs. wage dif. < 60	1.833	1.510	0.77	1.861
( $n = 210$ )	(0.068)	(0.133)		(0.047)

Table 8: Robustness of the results with additional covariates (marriage, working part-time, living in London or the South East, and work experience) when outliers are removed; the numbers in parentheses are p-values.

## References

- AIGNER, D. J., C. HSIAO, A. KAPTEYN, AND T. WANSBEEK (1984): “Latent Variable Models in Econometrics,” in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, pp. 1321–1393. Elsevier Science B.V.
- AMIN, V. (2011): “Returns to education: evidence from UK twins: comment,” *American Economic Review*, 101(4), 1629–35.
- ASHENFELTER, O., AND A. KRUEGER (1994): “Estimates of the Economic Return to Schooling from a New Sample of Twins,” *The American Economic Review*, 84(5), 1157–1173.
- ATTANASIO, O., S. CATTAN, E. FITZSIMONS, C. MEGHIR, AND M. RUBIO-CODINA (2015): “Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia,” Working Paper 20965, NBER.
- ATTANASIO, O., C. MEGHIR, AND E. NIX (2017): “Human Capital Development and Parental Investment in India,” Discussion paper.
- BLACK, D. A., M. C. BERGER, AND F. A. SCOTT (2000): “Bounding Parameter Estimates with Nonclassical Measurement Error,” *Journal of the American Statistical Association*, 95(451), 739–748.
- BONJOUR, D., L. F. CHERKAS, J. E. HASKEL, D. D. HAWKES, AND T. D. SPECTOR (2003): “Returns to education: Evidence from UK twins,” *American Economic Review*, 93(5), 1799–1812.
- BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): “Measurement Error in Survey Data,” in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. V, pp. 3705–3843. Elsevier Science B.V.
- CHEN, X., H. HONG, AND D. NEKIPELOV (2011): “Nonlinear Models of Measurement Errors,” *Journal of Economic Literature*, 49(4), 901–937.
- CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2013): “Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors,” *The Annals of Statistics*, 41(6), 2786–2819.

- CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): “Estimating the Technology of Cognitive and Noncognitive Skill Formation,” *Econometrica*, 78(3), 883–931.
- DELGADO, M. A., AND W. G. MANTEIGA (2001): “Significance Testing in Nonparametric Regression Based on the Bootstrap,” *The Annals of Statistics*, 29(5), 1469–1507.
- HAUSMAN, J. A. (1978): “Specification Tests in Econometrics,” *Econometrica*, 46(6), 1251–1271.
- HECKMAN, J., R. PINTO, AND P. SAVELYEV (2013): “Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes,” *American Economic Review*, 103(6), 2052–2086.
- HU, Y., AND Y. SASAKI (2017): “Identification of Paired Nonseparable Measurement Error Models,” *Econometric Theory*, 33(4), 955–979.
- LEE, Y. J., AND D. WILHELM (2020): “Testing for the presence of measurement error in Stata,” *The Stata Journal*, 20(2), 382–404.
- ROUSE, C. E. (1999): “Further estimates of the economic return to schooling from a new sample of twins,” *Economics of Education Review*, 18(2), 149–157.
- SCHENNACH, S. M. (2013): “Measurement Error in Nonlinear Models — A Review,” in *Advances in Economics and Econometrics*, vol. 3. Cambridge University Press.
- (2014): “Entropic Latent Variable Integration via Simulation,” *Econometrica*, 82(1), 345–385.
- WILHELM, D. (2019): “Testing for the Presence of Measurement Error,” Working Paper CWP48/19, CeMMAP.