Comments on "Unobservable Selection and Coefficient Stability: Theory and Evidence" and "Poorly Measured Confounders are More Useful on the Left Than on the Right"*

Giuseppe De Luca
University of Palermo, Palermo, Italy (giuseppe.deluca@unipa.it)

Jan R. Magnus
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands (jan@janmagnus.nl)

Franco Peracchi
Georgetown University, Washington, USA (fp211@georgetown.edu)

September 12, 2018

1 Introduction

The papers by Oster (2017) (henceforth Oster) and Pei, Pischke and Schwandt (2018) (henceforth PPS) contribute to the development of methods of inference about causal effects in the challenging and empirically relevant situation where the unknown data-generation process (DGP) is not included in the set of regression models considered by the investigator.

Building on Altonji, Elder and Table (2005), Oster analyzes the link between omitted variable bias in estimating a causal effect of interest and coefficient stability, defined as the change in the OLS estimates of the causal effect when imperfect controls are added to an initial model. PPS instead analyze the power properties of two alternative strategies for testing the consistency of the OLS estimator of the causal effect when the controls in the intermediate model are subject to measurement error. The two papers are in fact closely related, as they involve comparing the bias or the sampling variance of OLS estimators from misspecified models with different sets of regressors. The general misspecification framework recently proposed by De Luca, Magnus and Peracchi (2018) (henceforth DMP) is therefore particularly suited to analyze and understand the restrictions imposed by the two papers.

*We thank the Editor, an Associate Editor and two anonymous referees for their extremely valuable comments. We also thank Josh Angrist, Arik Levinson and Salvatore Modica for useful discussions. Giuseppe De Luca and Franco Peracchi acknowledge financial support from MIUR PRIN 2015FMRE5X.

Our comments are organized as follows. Section 2 presents the general misspecification framework developed in DMP. Section 3 discusses some results on inconsistencies and regression R^2 that are important for Oster's paper. Section 4 draws some implications of these results for empirical strategies. Section 5 discusses some results on testing strategies that are important for PPS's paper. Finally, Section 6 offers some conclusions. Proofs are collected in the Appendix.

2 A general misspecification framework

Oster and PPS focus on the case in which there is a single regressor of interest, so we consider the following simplified version of the DGP proposed in DMP:

$$y = \beta_1 x_1 + \beta_2' X_2 + \xi + \epsilon, \tag{1}$$

where x_1 is an observable scalar treatment, X_2 is a set of k_2 observable controls, β_1 and β_2 are unknown parameters, ξ is an unobservable specification error capturing, for example, the contributions of omitted variables or measurement errors, and ϵ is an unobservable regression error satisfying $\mathbb{E}(\epsilon|x_1, X_2, \xi) = 0$. We assume, without loss of generality, that all variables are centered to have mean zero. We also assume that (x_1, X_2, ξ) have positive definite second moment matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{21}' & \sigma_{1\xi} \\ \sigma_{21} & \Sigma_{22} & \sigma_{2\xi} \\ \sigma_{1\xi} & \sigma_{2\xi}' & \sigma_\xi^2 \end{bmatrix}.$$

When $k_2 = 1$, we write x_2 instead of X_2 and σ_2^2 instead of Σ_{22}. The parameter of interest is the scalar β_1 which, under our assumptions, is interpreted as the causal effect of x_1 on y. If ξ were observable, then β_1 could be estimated unbiasedly by an OLS regression of y on x_1, X_2 and ξ. The key statistical problem is that ξ is not observable.

Given a random sample from (1), we consider two alternative estimators of β_1: the restricted OLS estimator $\hat{\beta}_{1r}$ from the short regression of y on x_1, with probability limit denoted by β_{1r}, and the unrestricted OLS estimator $\hat{\beta}_{1u}$ from the intermediate regression of y on x_1 and X_2, with probability limit denoted by β_{1u}. From DMP, the inconsistencies of these two estimators are

$$b_{1r} = \beta_{1r} - \beta_1 = \tau_1 + \psi'(\beta_2 + \tau_2), \qquad b_{1u} = \beta_{1u} - \beta_1 = \tau_1, \tag{2}$$

where $\psi = \sigma_{21}/\sigma_1^2$ contains the population coefficients in the linear projection of X_2 on x_1 (or "balancing regression," using the terminology of PPS), $\tau_1 = \sigma^{11}\sigma_{1\xi} - \psi'\Sigma^{22}\sigma_{2\xi}$ and $\tau_2 = \Sigma^{22}(\sigma_{2\xi} - \sigma_{1\xi}\psi)$ are the population coefficients in the linear projection of ξ on x_1 and X_2, $\sigma^{11} = 1/\sigma_1^2 + \psi'\Sigma^{22}\psi$, and $\Sigma^{22} = (\Sigma_{22} - \sigma_{21}\sigma_{21}'/\sigma_1^2)^{-1}$. Notice that while the inconsistency of the unrestricted OLS estimator of β_1 from the intermediate regression is equal to τ_1, the inconsistency of the unrestricted OLS estimator of β_2 from the same regression is equal to τ_2. The expression for b_{1r} in (2) generalizes the classical omitted variables bias formula to settings where the intermediate regression is smaller than the unknown DGP. Since the DGP (1) encompasses a variety of misspecification problems, the expressions for b_{1r} and b_{1u} are completely general and can easily be extended to the case when x_1 contains more than one regressor. An immediate implication of (2) is that $b_{1r} - b_{1u} = \beta_{1r} - \beta_{1u} = \psi'(\beta_2 + \tau_2)$, which shows that the strategy of evaluating coefficient stability by augmenting the short regression with an additional set of regressors is only informative about the sign and magnitude of the difference of the inconsistencies, not about the sign and magnitude of the two inconsistencies separately. In fact, depending on the conditions discussed in DMP, the difference $b_{1r} - b_{1u}$ can be large or small, positive or negative. Thus, the two estimators may differ by little even when their inconsistencies are large. Furthermore, lack of coefficient stability may arise when the inconsistencies of the two estimators have opposite signs and $|b_{1u}| > |b_{1r}|$.

Sharper results may be obtained if stronger assumptions are imposed but, according to the Law of Decreasing Credibility (Manski 2003), the credibility of inference decreases with the strength of the assumptions maintained. Thus, in the next three sections, we focus on the additional assumptions employed by Oster and PPS to obtain their results.

3 Inconsistencies and regression R-squares

Oster writes her model as $Y = \beta X + \Psi'\omega^o + W_2 + \epsilon$, where X is an observable scalar treatment, ω^o is a set of observable controls, β and Ψ are unknown parameters, and W_2 and ϵ are unobservable random terms. This is the same as model (1) with $y = Y$, $\beta_1 x_1 = \beta X$, $\beta_2' X_2 = \Psi'\omega^o$ and $\xi = W_2$. It is useful to define the linear combination $\eta = \beta_2' X_2$ of the available controls, the vector μ of coefficients in the linear projection of x_1 on X_2, and the additional set of population second moments $\sigma_\eta^2 = \text{var}(\eta) = \beta_2'\Sigma_{22}\beta_2$, $\sigma_{1\eta} = \text{cov}(x_1, \eta) = \sigma_{21}'\beta_2$ and $\sigma_{\eta\xi} = \text{cov}(\eta, \xi) = \beta_2'\sigma_{2\xi}$. As in Oster, we describe the link between x_1, ξ and η through the "proportional selection relationship"

$$\frac{\sigma_{1\xi}}{\sigma_\xi^2} = \varphi \frac{\sigma_{1\eta}}{\sigma_\eta^2}, \tag{3}$$

for some value of the proportionality coefficient φ.

Oster's contribution is to provide various characterizations of the inconsistency $b_{1u} = \beta_{1u} - \beta_1$ of the unrestricted OLS estimator of β_1. Although these characterizations do not come for free, they have attracted considerable interest because of their simplicity and their possible use for sensitivity analysis or for deriving bias-corrected estimators of β_1. Her first main result (Proposition 1) is an explicit representation of b_{1u} based on the following set of assumptions:

Assumption A *The covariance $\sigma_{1\eta}$ between x_1 and η is nonzero.*

Assumption B *The controls in X_2 are uncorrelated with the specification error ξ.*

Assumption C *The controls in X_2 are mutually uncorrelated.*

Assumption D (Equal selection relationship) *The relationship (3) holds with $\varphi = 1$.*

Assumption E *The elements of $\beta_2 = (\beta_{21}, \ldots, \beta_{2k_2})'$ are proportional to the elements of $\mu = (\mu_1, \ldots, \mu_{k_2})'$.*

Assumption A is fundamental but never formally stated in Oster's paper. Assumption C helps simplify the proofs but, as shown below, is unnecessary. Assumption D is the same as Oster's Assumption 1. Assumption E corresponds to Oster's Assumption 2 but our formulation is slightly different to guarantee that the assumption also holds when there is only one control. This is a strong assumption and Oster points out that "with multiple controls it is very unlikely to hold except in pathological cases" (p. 6). As for Assumption B, Oster admits that it is controversial because "somewhat at odds with the intuition that the observables and the unobservables are related" (p. 6). In fact, when imposed jointly with Assumptions A and D, it implies that (i) b_{1r} and b_{1u} are proportional to each other, (ii) $\sigma_{1\eta}$, b_{1r} and b_{1u} have the same sign, and (iii) $\sigma_{1\eta}$ and $\beta_{1r} - \beta_{1u} = b_{1r} - b_{1u}$ have the same sign. The first two results follow from (2) after imposing the restrictions $\sigma_{2\xi} = 0$ and $\sigma_{1\xi} = \sigma_\xi^2 \sigma_{1\eta}/\sigma_\eta^2$, while the third follows from our Corollary 1 below.

Under Assumptions A–E, Oster's Proposition 1 shows that

$$b_{1u} = (\beta_{1r} - \beta_{1u}) \frac{R_{max}^2 - R_u^2}{R_u^2 - R_r^2}, \tag{4}$$

where R_{max}^2 is the unknown population R^2 from the DGP (1), while R_r^2 and R_u^2 are the population R^2 from, respectively, the short regression of y on x_1 and the intermediate regression of y on x_1 and X_2. An important implication of (4) is that $(\beta_{1u} - \beta_1)/(\beta_{1r} - \beta_{1u}) = (R_{max}^2 - R_u^2)/(R_u^2 - R_r^2)$, that is, "the ratio of the movement in coefficients is equal to the ratio of the movement in R-squared" (Oster, p. 7). Since $\hat{\beta}_{1r} - \hat{\beta}_{1u}$ is consistent for $\beta_{1r} - \beta_{1u}$, another implication of (4) for the special case when R_{max}^2 is known, is the following bias-corrected estimator of β_1

$$\tilde{\beta}_1 = \hat{\beta}_{1u} - (\hat{\beta}_{1r} - \hat{\beta}_{1u}) \frac{R_{max}^2 - R_u^2}{R_u^2 - R_r^2}.$$

This second result may help explain the appeal of Proposition 1 among practitioners despite the warning that "[g]iven the restrictiveness of the assumptions [...] it is not appropriate to suggest that researchers use this as an estimator directly" (Oster, p. 7).

To appreciate why Assumptions A–E are restrictive, notice that, under the plausible assumption that $R^2_{max} > R^2_u > R^2_r$, (4) implies that b_{1u} has the same sign as $\beta_{1r} - \beta_{1u}$. As stressed by Holly (1982) this is not generally true. Further, since $\beta_{1r} - \beta_{1u} = b_{1r} - b_{1u}$, we also have

$$\frac{b_{1r}}{b_{1u}} = 1 + \frac{R^2_u - R^2_r}{R^2_{max} - R^2_u} > 1.$$

In other words, Assumptions A–E together amount to assuming that adding the controls in X_2 decreases the bias in estimating β_1 or, in the terminology of DMP, that X_2 is a balanced addition. As stressed by DMP, this is also not generally true.

If Assumptions D and E are relaxed, then Oster's second main result (Proposition 2) shows that b_{1u} is a root of the cubic equation

$$a_3 z^3 + a_2 z^2 + a_1 z + a_0 = 0, \tag{5}$$

with real coefficients

$$\begin{aligned}
a_0 &= \varphi \sigma_1^2 \sigma_y^2 (R^2_{max} - R^2_u)(\beta_{1r} - \beta_{1u}), \\
a_1 &= \varphi(\sigma_1^2 - \sigma_\nu^2)\sigma_y^2 (R^2_{max} - R^2_u) - \sigma_\nu^2 \left(\sigma_y^2 (R^2_u - R^2_r) + \sigma_1^2 (\beta_{1r} - \beta_{1u})^2\right), \\
a_2 &= (\varphi - 2)\sigma_1^2 (\beta_{1r} - \beta_{1u})\sigma_\nu^2, \\
a_3 &= (\varphi - 1)(\sigma_1^2 - \sigma_\nu^2)\sigma_\nu^2,
\end{aligned}$$

where σ_y^2 and $\sigma_\nu^2 = \sigma_1^2 - \sigma_{21}' \Sigma_{22}^{-1} \sigma_{21}$ are the population variances of y and $\nu = x_1 - \mu' X_2$ respectively. If only Assumption E is relaxed, then $a_3 = 0$ and Proposition 2 implies that b_{1r} is a root of the quadratic equation

$$a_2 z^2 + a_1 z + a_0 = 0. \tag{6}$$

Notice that, while (5) admits either one or three real roots, (6) always admits two real roots of opposite sign.

The more general Proposition 2 confirms that the inconsistency of the unrestricted OLS estimator depends on the differences $\beta_{1r} - \beta_{1u}$, $R^2_{max} - R^2_u$ and $R^2_u - R^2_r$, but no longer provides an explicit representation. Further, when equations (5) or (6) admit multiple roots, it is unclear how to select one. To obtain a unique root of (6), Oster introduces an additional assumption (Assumption 3, p. 8):

Assumption F *The covariance $\sigma_{1\eta}$ between x_1 and η has the same sign as the covariance between x_1 and $\eta^* = \beta_{2u}' X_2$, where β_{2u} is the vector of coefficients on X_2 in the linear projection of y on x_1 and X_2.*

To gain intuition about this assumption, Oster offers the following interpretation: "[e]ffectively, this assumes that the bias from the unobservables is not so large that it biases the *direction* of the covariance between the observable index and the treatment" (p. 8). Notice however that $\eta^* = \beta_{2u}' X_2 = (\beta_2 + \tau_2)' X_2$. Hence, from (2),

$$\operatorname{cov}(x_1, \eta^*) = \operatorname{cov}(x_1, (\beta_2 + \tau_2)' X_2) = \sigma_{21}' (\beta_2 + \tau_2) = \sigma_1^2 \psi' (\beta_2 + \tau_2) = \sigma_1^2 (\beta_{1r} - \beta_{1u}).$$

This shows that Assumption F is equivalent to the assumption that $\sigma_{1\eta}$ and $\beta_{1r} - \beta_{1u}$ have the same sign which, as already mentioned, is an implication of Assumptions A, B and D. Thus Assumption F is in fact redundant.

We stress the fact that restricting $\sigma_{1\eta}$ and $\beta_{1r} - \beta_{1u} = b_{1r} - b_{1u}$ to have the same sign allows one to select a unique root of the quadratic equation (6). As with the explicit solution (4), the selected root depends on the implicit assumption that augmenting the short regression with X_2 always decreases the bias in estimating β_1. If this assumption is incorrect, one may select the wrong root even when the values of φ and R^2_{max} are known, as illustrated in Section 4.

The next theorem completely summarizes the relationships between b_{1u}, $\beta_{1r} - \beta_{1u}$, $R^2_{max} - R^2_u$ and $R^2_u - R^2_r$ implied by Assumptions A and B:

Theorem 1 *If Assumptions A and B hold, then*

$$\beta_{1r} - \beta_{1u} = \frac{\sigma_{1\eta} - (\sigma_1^2 - \sigma_\nu^2) b_{1u}}{\sigma_1^2},$$

$$\sigma_y^2 (R_u^2 - R_r^2) = \sigma_\eta^2 + \sigma_\nu^2 b_{1u}^2 - \frac{1}{\sigma_1^2} (\sigma_{1\eta} + \sigma_\nu^2 b_{1u})^2,$$

$$\varphi \sigma_y^2 (R^2_{max} - R^2_u) = \left(\frac{\sigma_\eta^2}{\sigma_{1\eta}} - \varphi b_{1u} \right) \sigma_\nu^2 b_{1u}.$$

If $k_2 > 1$, then b_{1u} is a root of the cubic equation (5). If $k_2 = 1$, then b_{1u} is a root of the quadratic equation $c_2 z^2 + c_1 z + c_0 = 0$ with real coefficients $c_0 = -\varphi \sigma_{21}^2 \sigma_2^2 \sigma_y^2 (R^2_{max} - R^2_u)$, $c_1 = \sigma_1^2 \sigma_2^2 (\sigma_1^2 \sigma_2^2 - \sigma_{21}^2)(\beta_{1r} - \beta_{1u})$, and $c_2 = (1 - \varphi) \sigma_{21}^2 (\sigma_1^2 \sigma_2^2 - \sigma_{21}^2)$.

Theorem 1 corrects an error in the equation system on p. 7 of Oster's paper and delivers the same conclusions of her Proposition 2 without assuming that the controls are mutually uncorrelated. The link between this theorem and Oster's Propositions 1 is made clear by the next corollary:

Corollary 1 *If Assumptions A, B and D hold, then*

$$(\beta_{1r} - \beta_{1u})\frac{R^2_{max} - R^2_u}{R^2_u - R^2_r} = \frac{z'\Omega z}{z'\Xi z}\, b_{1u},$$

where $z = \beta_2 - \Sigma_{22}^{-1}\sigma_{21} b_{1u}$, $\Omega = (\sigma_1^2 - \sigma_{21}'\Sigma_{22}^{-1}\sigma_{21})\Sigma_{22}\beta_2\sigma_{21}'$, *and* $\Xi = \beta_2'\sigma_{21}[\sigma_1^2\Sigma_{22} - \sigma_{21}\sigma_{21}']$. *The relationship* (4) *holds if and only Assumption E also holds.*

The ratios $(R^2_{max} - R^2_u)/(R^2_u - R^2_r)$ and $(z'\Omega z)/(z'\Xi z)$ are positive under general conditions, so this corollary also shows that $\beta_{1r} - \beta_{1u}$ and b_{1u} have the same sign. Notice that Assumption E is trivially satisfied when $k_2 = 1$ but, as already argued, is unlikely to hold when $k_2 > 1$. When $k_2 = 1$ but Assumption D does not hold, another corollary of Theorem 1 is the following:

Corollary 2 *When* $k_2 = 1$ *and* $\varphi \neq 1$, *define*

$$\varphi_1^* = 1 - \sqrt{1 + \frac{1}{\rho_{21}^2}\frac{R^2_u - R^2_r}{R^2_{max} - R^2_u}}, \qquad \varphi_2^* = 1 + \sqrt{1 + \frac{1}{\rho_{21}^2}\frac{R^2_u - R^2_r}{R^2_{max} - R^2_u}},$$

with $\rho_{21} = \sigma_{21}/(\sigma_1\sigma_2)$. *Then the quadratic equation* $c_2 z^2 + c_1 z + c_0 = 0$ *admits two distinct real roots if* $\varphi_1^* < \varphi < \varphi_2^*$, *one real root if* $\varphi = \varphi_1^*$ *or* $\varphi = \varphi_2^*$, *and no real root otherwise.*

4 Implications for empirical strategies

Based on Proposition 2, Oster discusses three possible strategies: (i) find the value of β_1 for given values of φ and R^2_{max}; (ii) find the value of φ for given values of β_1 and R^2_{max}; (iii) find the value of R^2_{max} for given values of β_1 and φ. These three strategies are easily implemented using Oster's Stata routine `psacal` but require Assumptions A and B to characterize the inconsistency of the unrestricted estimator of β_1 as a root of the cubic equation (5). Notice that the coefficients in this equation can all be estimated consistently provided φ and R^2_{max} are known or can be estimated consistently.

Strategy (i) may be used to derive a bias-corrected estimate of β_1 given knowledge of φ and R^2_{max} or, alternatively, to obtain bounds on β_1 given bounds on φ and R^2_{max}. In either case one needs a unique value of β_1 for any choice of φ and R^2_{max}. With multiple roots ($\varphi \neq 1$), Oster's Stata routine selects the root closest to the unrestricted OLS estimate $\hat{\beta}_{1u}$ and results in b_{1u} having the same sign as $\hat{\beta}_{1r} - \hat{\beta}_{1u}$. This creates two issues. First, when no root satisfies the sign condition on b_{1u}, it is not clear how the routine selects the solution for β_1. Second, when one or more root satisfies the sign condition, the root closest to $\hat{\beta}_{1u}$ is not necessarily the correct solution.

Table 1: OLS estimates of the coefficients in the DGP, the intermediate and the short regressions from a pseudo-random sample of 100,000 observations

	$\sigma_{1\xi} = \sigma_{14} = -0.40$			$\sigma_{1\xi} = \sigma_{14} = 0.80$		
Variable	DGP	Interm.	Short	DGP	Interm.	Short
z_1	1.001	0.519	1.247	1.003	1.963	2.445
z_2	1.007	1.149		1.007	0.725	
z_3	−0.994	−1.104		−0.994	−0.775	
z_4	0.998			0.995		
R^2	0.833	0.698	0.260	0.881	0.854	0.715
φ	−1.538			3.077		

To illustrate, we present an example where Assumptions A and B hold but Assumption D does not. Suppose that $y = z_1 + z_2 - z_3 + z_4 + \epsilon$, where the z_j's are jointly normal (Gaussian) with mean zero and second moment matrix

$$\Sigma = \operatorname{var}\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{bmatrix} 1 & 0.35 & -0.30 & \sigma_{14} \\ 0.35 & 1 & -0.25 & 0 \\ -0.30 & -0.25 & 1 & 0 \\ \sigma_{14} & 0 & 0 & 1 \end{bmatrix},$$

and $\epsilon \sim N(0,1)$ independently of the z_j's. We set $x_1 = z_1$, $X_2 = (z_2, z_3)$ and $\xi = z_4$, and consider two cases that differ by the value of $\sigma_{1\xi} = \sigma_{14}$, namely $\sigma_{14} = -0.40$ and $\sigma_{14} = 0.80$. Table 1 shows, for each case, the OLS estimates of the true DGP, the intermediate regression of y on x_1 and X_2, and the short regression of y on x_1 from a pseudo-random sample of 100,000 observations. Notice that adding X_2 to the short regression lowers the size of the bias in estimating β_1 in the second case ($\sigma_{1\xi} = 0.80$) but not in the first ($\sigma_{1\xi} = -0.40$). Of course, the investigator does not know this because ξ is unobservable.

In the first case, employing Oster's Stata routine with the true values of φ and R^2_{max} gives three possible bias-corrected estimates of β_1: $\hat{\beta}_1^{(1)} = 4.797$, $\hat{\beta}_1^{(2)} = 1.000$ and $\hat{\beta}_1^{(3)} = 1.747$. Although the second is equal to the true value, the routine selects the first because none of the $\hat{\beta}_1^{(j)}$ satisfies the sign condition on b_{1u}. The selected estimate is clearly severely upward biased. The interval [0.509, 4.798] for β_1 implied by the restrictions $-1.60 \leq \varphi \leq 0$ and $R^2_{max} = 0.85$ contains the true value of β_1 but is not sharp. Notice, however, that all the $\hat{\beta}_1^{(j)}$ fall in the region where $|b_{1r}/b_{1u}| < 1$ so, in this case, one knows for sure that $\hat{\beta}_{1r}$ has less bias than $\hat{\beta}_{1u}$.

In the second case, Oster's Stata routine gives $\hat{\beta}_1^{(1)} = 1.563$, $\hat{\beta}_1^{(2)} = 0.979$ and $\hat{\beta}_1^{(3)} = 4.825$. The first two values now satisfy the sign condition on b_{1u}. Although the second value is closer to the true value, the routine again selects the first because it is the closer to $\hat{\beta}_{1u} = 1.963$. The selected estimate is still upward biased but now the interval [1.963, 4.824] for β_1 implied by the restrictions $0 \leq \varphi \leq 3.10$ and $R^2_{max} = 0.90$ no longer contains the true value of β_1.

As for strategies (ii) and (iii), note that fixing the value of β_1 for given values of β_{1r} and β_{1u} is equivalent to fixing the values of b_{1r} and b_{1u}. Under Assumptions A and B, this allows one to identify $\sigma_{1\xi}$ and $\sigma_{1\eta}$, and therefore also σ^2_η from the third equation in Theorem 1. By restricting either R^2_{max} or φ, one can then identify σ^2_ξ. Thus, under Assumptions A and B, strategies (ii) and (iii) amount to imposing arbitrary restrictions on all the unidentified model parameters. The results obtained are also sensitive to the choice of R^2_{max} for strategy (ii) and φ for strategy (iii).

5 Testing strategies

PPS consider the model $y = \beta^l s + \boldsymbol{\gamma}'\boldsymbol{x} + e$, where s is a scalar treatment, β^l is the causal effect of interest, $\boldsymbol{x} = \boldsymbol{\delta} s + \boldsymbol{u}$ is a vector of unobservable controls, and $\boldsymbol{u}$ and e are random errors. This is a special case of model (1) where $\beta_1 x_1 = \beta^l s$, $X_2 = \boldsymbol{x}^m = \boldsymbol{x} + \boldsymbol{m}$ is a vector of observable error-ridden controls, $\beta_2 = 0$, $\xi = \boldsymbol{\gamma}'\boldsymbol{x}$ is the specification error, and $\epsilon = e$. The assumption of classical measurement error gives a balancing regression of the form $\boldsymbol{x}^m = \boldsymbol{\delta} s + \boldsymbol{u} + \boldsymbol{m}$, with $\boldsymbol{u}$ and $\boldsymbol{m}$ uncorrelated with s, and therefore $\psi = \boldsymbol{\delta}$. When $k_2 = 1$, PPS also consider a mean-reverting measurement error model resulting in a balancing regression of the form $x^m = (1+\kappa)\delta s + (1+\kappa)u + \mu$, where $-1 < \kappa < 0$ and s, u and μ are uncorrelated with each other. In this case, the coefficient of the balancing regression is $\psi = (1 + \kappa)\delta$.

The main contribution of PPS is to compare the power properties of two alternative strategies for testing whether the restricted OLS estimator is consistent: a classical F-test of significance of the population coefficient $\psi = \sigma_{21}/\sigma^2_1$ in the balancing regression and a Hausman-type test based on the difference $\beta_{1r} - \beta_{1u} = b_{1r} - b_{1u} = \psi'(\beta_2 + \tau_2)$ between the restricted and the unrestricted OLS estimators of β_1. PPS refer to these tests as the balancing test (BT) and the coefficient comparison test (CCT), respectively. Their results show that, if the intermediate regression is misspecified (i.e., $\boldsymbol{\gamma} \neq \boldsymbol{0}$), then BT is generally more powerful than CCT because measurement error is comparatively less harmful when mismeasured variables are employed as outcome variables in the balancing regression rather than as additional controls in the intermediate regression.

This insight reinforces our key point that adding controls to the short regression does not necessarily improve the estimates of the causal effect of interest. While DMP and Oster are mainly concerned with the statistical properties of the restricted and unrestricted OLS estimators of β_1, PPS focus on the implications of using the available controls for testing purposes. However, these two approaches are closely related through the relationship between mean squared error (MSE) comparisons and testing strategies. It is well-known that if we delete a single control from a correctly specified model, then $\mathrm{MSE}(\hat{\beta}_{1r}) \leq \mathrm{MSE}(\hat{\beta}_{1u})$ if and only if the t-statistic on the coefficient of the deleted control is smaller than one in absolute value. Similar results extend to the case of multiple controls, where MSE comparisons depend crucially on the noncentrality parameter in the distribution of either the classical F-statistic or the Hausman-type statistic used for testing the hypothesis $H_0\colon \beta_2 = 0$ in the intermediate regression (Toro-Vizcarrondo and Wallace 1968, Holly 1982). Additional results on MSE comparisons for the case when the intermediate regression is subject to specification errors can be found in Appendix A of DMP.

We now use the general framework in DMP to provide more insight into the restrictions required for the validity of BT and CCT. It follows immediately from (2) that BT and CCT provide tests of the null hypothesis of interest, namely

$$H_0\colon b_{1r} = \tau_1 + \psi'(\beta_2 + \tau_2) = 0, \tag{7}$$

only if suitable restrictions are placed on τ_1. BT is concerned with the null hypothesis $H_0\colon \psi = 0$. Writing $\tau_1 = \sigma_{1\xi}/\sigma_1^2 - \psi'\tau_2$, we see that this is equivalent to (7) if and only if there exists a k_2-vector $\omega \neq -\beta_2$ such that $\sigma_{1\xi} = \sigma_{21}'\omega$, so that

$$b_{1u} = \tau_1 = \psi'(\omega - \tau_2), \qquad b_{1r} = \psi'(\beta_2 + \omega). \tag{8}$$

CCT is instead concerned with the null hypothesis $H_0\colon \psi'(\beta_2 + \tau_2) = 0$, which is equivalent to (7) if and only if there exist a scalar $a \neq -1$ such that

$$b_{1u} = \tau_1 = a\psi'(\beta_2 + \tau_2), \qquad b_{1r} = (1 + a)\psi'(\beta_2 + \tau_2). \tag{9}$$

For example, when $k_2 = 1$ and measurement error is classical, we have $\psi = \delta$, $\tau_1 = \delta\gamma\theta$ and $\tau_2 = (1 - \theta)\gamma$, with $\theta = \sigma_m^2/(\sigma_m^2 + \sigma_u^2)$. In this case, (8) and (9) hold when $\omega = \gamma \neq 0$ and $a = \theta/(1-\theta) > 0$, but this model is known to be restrictive because it implies that $b_{1r}/b_{1u} = 1/\theta > 1$. Similar considerations apply to the mean-reverting measurement error model, where

$$\psi = (1 + \kappa)\delta, \qquad \tau_1 = \delta\gamma\frac{\theta}{(1 + k)^2(1 - \theta) + \theta}, \qquad \tau_2 = \frac{\gamma}{1 + k}\left[1 - \frac{\theta}{(1 + k)^2(1 - \theta) + \theta}\right],$$

with $\theta = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_u^2)$. Here, the restrictions (8) and (9) hold when $\omega = \gamma/(1 + k) \neq 0$ and $a = \theta/[(1+k)^2(1 - \theta)] > 0$, but this implies that $b_{1r}/b_{1u} = 1 + (1 + k)^2(1 - \theta)/\theta > 1$. Like PPS, we stress that this result is special and does not extend to more realistic settings in which s and m are correlated (Frost 1979), or s is also measured with error (Barnow 1976). Also notice that, if there are multiple controls subject to measurement error (i.e., $k_2 > 1$), then the condition $b_{1r}/b_{1u} > 1$ need not hold (Garber and Klepper 1980). Although theoretical power comparisons for the case of multiple controls are still lacking, the Monte Carlo simulations in PPS provide convincing evidence in favor of the BT strategy.

As mentioned by PPS, pretesting may have nontrivial effects on the statistical properties of these tests. Strategies for addressing this issue, such as post-model-selection inference (see, e.g., Berk et al. 2013 and Leeb, Pötscher and Ewald 2015) and model-averaging estimation under a misspecified model space (see, e.g., Zhang et al. 2016 and Ando and Li 2017), deserve careful attention.

6 Conclusions

Oster's Proposition 1 delivers a very sharp result but requires strong assumptions and knowledge of the key parameter R^2_{max}. Her Proposition 2, as reformulated in our Theorem 1, weakens some of these assumptions but requires knowledge of both R^2_{max} and the additional parameter φ. Despite the strong assumptions, her characterization of the bias of the unrestricted OLS estimator as a root of a cubic equation is useful but, when this equation has three roots, it is unclear which to select. Our paper does not solve this problem, nor it offers other ways of correcting for bias, but we hope it helps clarify the nature of Oster's assumptions and properly evaluate the results of her Stata routine. Finally, the two testing strategies in PPS also require restrictions, but in this case the assumptions are fewer and more transparent, which makes it easier for practitioners to check whether they are indeed satisfied.

References

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151–184.

Ando, T., and Li, K.-C. (2017), "A Weight-Relaxed Model Averaging Approach for High-Dimensional Generalized Linear Models," *Annals of Statistics*, 45, 2654–2679.

Barnow, B. S. (1976), "The Use of Proxy Variables When One or Two Independent Variables Are Measured With Error," *American Statistician*, 30, 119–121.

Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013), "Valid Post-Selection Inference," *Annals of Statistics*, 41, 802–837.

De Luca, G., Magnus, J. R., and Peracchi, F. (2018), "Balanced Variable Addition in Linear Models," *Journal of Economic Surveys*, 32, 1183–1200.

Frost, P. A. (1979), "Proxy Variables and Specification Bias," *The Review of Economics and Statistics*, 61, 323–325.

Garber, S., and Klepper, S. (1980), "Extending the Classical Normal Errors-in-Variables Model," *Econometrica*, 48, 1541–1546.

Holly, A. (1982), "A Remark on Hausman's Specification Test," *Econometrica*, 50, 749–760.

Leeb, H., Pötscher, B. M., and Ewald, K. (2015). "On Various Confidence Intervals Post-Model-Selection," *Statistical Science*, 30, 216–227.

Manski, C. F. (2003), *Partial Identification of Probability Distributions,* New York: Springer.

Oster, E. (2017), "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2016.1227711.

Pei, Z., Pischke, J.-S., and Schwandt, H. (2018), "Poorly Measured Confounders Are More Useful on the Left Than on the Right," DOI: 10.1080/07350015.2018.1462710.

Toro-Vizcarrondo, C., and Wallace, T. D. (1968), "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," *Journal of the American Statistical Association*, 63, 558–572.

Zhang, X., Yu, D., Zou, G., and Liang, C. (2016), "Optimal Model Averaging Estimation for Generalized Linear Models and Generalized Linear Mixed-Effects Models," *Journal of the American Statistical Association*, 111, 1775–1790.

Appendix

Proof of Theorem 1. Assumption A and B imply the equation system

$$\sigma_1^2(\beta_{1r} - \beta_{1u}) = \left[\beta_2 - \Sigma_{22}^{-1}\sigma_{21}b_{1u}\right]' \sigma_{21}, \tag{A1}$$

$$\sigma_y^2(R_u^2 - R_r^2) = \left[\beta_2 - \Sigma_{22}^{-1}\sigma_{21}b_{1u}\right]' \left[\Sigma_{22} - \frac{1}{\sigma_1^2}\sigma_{21}\sigma_{21}'\right] \left[\beta_2 - \Sigma_{22}^{-1}\sigma_{21}b_{1u}\right], \tag{A2}$$

$$\varphi\sigma_y^2(R_{max}^2 - R_u^2)\beta_2'\sigma_{21} = (\sigma_1^2 - \sigma_{21}'\Sigma_{22}^{-1}\sigma_{21}) \left[\beta_2 - \varphi\Sigma_{22}^{-1}\sigma_{21}b_{1u}\right]' \Sigma_{22}\beta_2 b_{1u}, \tag{A3}$$

which in turns implies the equation system in the statement of the theorem. When $k_2 > 1$, we then obtain a cubic equation in b_{1u} following the argument in the proof of Oster's Proposition 2. When $k_2 = 1$, equation (A2) is redundant and the result follows by solving for β_2 and b_{1u} the pair of equations (A1) and (A3).

Proof of Corollary 1. The first result follows from (A1)–(A2) by setting $\varphi = 1$. To proof the second result, define $p = \Sigma_{22}^{1/2}\beta_2$, $q = \Sigma_{22}^{-1/2}\sigma_{21}$ and $\tau = b_{1u}$, so that $\Sigma_{22}^{1/2}z = p - \tau q$, $\Sigma_{22}^{-1/2}\Omega\Sigma_{22}^{-1/2} = (\sigma_1^2 - q'q)pq'$, and $\Sigma_{22}^{-1/2}\Xi\Sigma_{22}^{-1/2} = (p'q)(\sigma_1^2 I - qq')$. It then follows that

$$\begin{aligned} z'\Omega z &= (\sigma_1^2 - q'q)(p - \tau q)'pq'(p - \tau q) \\ &= (\sigma_1^2 - q'q)\left[\tau^2(p'q)(q'q) - \tau((p'p)(q'q) + (p'q)^2) + (p'p)(p'q)\right] \end{aligned}$$

and

$$\begin{aligned} z'\Xi z &= (p'q)(p - \tau q)'(\sigma_1^2 I - qq')(p - \tau q) \\ &= (\sigma_1^2 - q'q)(\tau^2(p'q)(q'q) - 2\tau(p'q)^2 + (p'p)(p'q)) + (p'q)((p'p)(q'q) - (p'q)^2). \end{aligned}$$

Hence, $z'\Omega z = z'\Xi z$ if and only if

$$\begin{aligned} 0 &= (\sigma_1^2 - q'q)\left[\tau^2(p'q)(q'q) - \tau((p'p)(q'q) + (p'q)^2) + (p'p)(p'q)\right] \\ &\quad - (\sigma_1^2 - q'q)\left[\tau^2(p'q)(q'q) - 2\tau(p'q)^2 + (p'p)(p'q)\right] - \left[(p'q)((p'p)(q'q) - (p'q)^2)\right] \\ &= -\left[(p'p)(q'q) - (p'q)^2\right]\left[(\sigma_1^2 - q'q)\tau + (p'q)\right]. \end{aligned}$$

Under Assumption A, the second term is always different from zero. Thus, $z'\Omega z = z'\Xi z$ if and only if p and q are proportional to each other, that is β_2 is proportional to $\mu = \Sigma_{22}^{-1}\sigma_{21}$.

Proof of Corollary 2. The result follows by solving for φ the equation

$$0 = (\sigma_1^2\sigma_2^2 - \sigma_{21}^2)\frac{\sigma_1^4}{\sigma_{21}^4}(\beta_{1r} - \beta_{1u})^2 + 4\varphi(1 - \varphi)\frac{\sigma_y^2(R_{max}^2 - R_u^2)}{\sigma_2^2},$$

and then using the fact that, from (A1) and (A2), $\sigma_y^2(R_u^2 - R_r^2)/(\beta_{1r} - \beta_{1u})^2 = \sigma_1^2(\sigma_1^2\sigma_2^2 - \sigma_{21}^2)/\sigma_{21}^2$.