EIEF WORKING PAPER SERIES

Einaudi Institute for Economics and Finance

EIEF Working Paper 18/02
February 2018

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

by

Giuseppe De Luca

(University of Palermo)

Jan R. Magnus

(Vrije Universiteit Amsterdam)

Franco Peracchi

(Georgetown University and EIEF)

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)

February 11, 2018

1 Introduction

The papers by Oster (2017) (henceforth Oster) and Pei, Pischke and Schwandt (2017) (henceforth PPS) contribute to the development of inferential procedures for causal effects in the challenging and empirically relevant situation where the unknown data-generation process (DGP) is not included in the set of models considered by the investigator. Based on Altonji, Elder and Taber (2005), Oster analyzes the relationship between the change in the OLS estimates of a causal effect due to the inclusion of additional controls and the omitted variable bias in the associated long regression. She shows that, under certain conditions, this change depends on the improvements in the regression R-squares going from the short regression to the long regression and from the long regression to the unknown DGP.

In contrast, PPS analyze the power properties of two alternative strategies for testing the consistency of the OLS estimator of a causal effect when the control variables in the long regression are subject to various forms of measurement error. The two approaches are 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

*We thank Josh Angrist, Arik Levinson and Salvatore Modica for useful discussions. Giuseppe De Luca and Franco Peracchi acknowledge financial support from MIUR PRIN 2015FMRE5X.

Luca, Magnus and Peracchi (2018) (henceforth DMP) is therefore particularly suited to analyze and understand the restrictions needed by the two approaches.

2 A general misspecification framework

The models in Oster and PPS are both special cases of the misspecification framework in DMP. Since Oster and PPS focus on the case where there is a single regressor of interest, we consider a simplified version of the DGP in DMP, namely

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

where x_1 is an observable scalar treatment, X_2 is a set of k_2 observable control variables, β_1 and β_2 contain the unknown parameters, ξ is an unobservable misspecification term capturing, for example, the contributions of omitted variables (as in Oster) or measurement errors (as in PPS), and ϵ is an unobservable error term satisfying $\mathbb{E}(\epsilon|x_1, X_2, \xi) = 0$. Without loss of generality, all variables are centered to have mean zero. The parameter of interest is the scalar β_1, which is interpreted as the causal effect of x_1 on y. The population second moments of (x_1, X_2, ξ) are denoted by

$$\Sigma = \operatorname{var}\begin{pmatrix} x_1 \\ X_2 \\ \xi \end{pmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{21}' & \sigma_{1\xi} \\ \sigma_{21} & \Sigma_{22} & \sigma_{2\xi} \\ \sigma_{1\xi} & \sigma_{2\xi}' & \sigma_\xi^2 \end{bmatrix}.$$

When $k_2 = 1$, we write x_2 instead of X_2 and σ_2^2 instead of Σ_{22}.

Since ξ is unknown, we consider two alternative estimators of β_1: the restricted OLS estimator from the "short regression" of y on x_1 only, with probability limit denoted by β_{1r}, and the unrestricted OLS estimator from the "long 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 $\Sigma^{22} = (\Sigma_{22} - \sigma_{21}\sigma_{21}'/\sigma_1^2)^{-1}$. The expression for b_{1r} generalizes the classical omitted variables bias formula to settings where the long 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, coefficient instability may arise when the inconsistencies of the two estimators have opposite signs and $|b_{1u}| > |b_{1r}|$.

3 Inconsistencies and regression R-squares

To map our notation into Oster's notation, let $x_1 = X$, $\beta_1 = \beta$, $X_2 = \omega^o$, $\beta_2 = \Psi$, and $\xi = W_2$. Also define the linear combination $\eta = \beta_2' X_2 = W_1$ of the k_2 control variables in X_2. This gives the additional set of population second moments $\sigma_\eta^2 = \mathrm{var}(\eta) = \beta_2'\Sigma_{22}\beta_2$, $\sigma_{1\eta} = \mathrm{cov}(x_1, \eta) = \sigma_{21}'\beta_2$, and $\sigma_{\eta\xi} = \mathrm{cov}(\eta, \xi) = \beta_2'\sigma_{2\xi}$.

Oster's main contribution are expressions for the inconsistency b_{1u} of the unrestricted estimator of β_1. To derive these expressions, Oster imposes the following four assumptions:

Assumption 1 *The controls in X_2 are uncorrelated with the misspecification ξ, that is, $\sigma_{2\xi} = 0$.*

Assumption 2 *x_1, η and ξ are linked through the "proportional selection relationship"*

$$\varphi \frac{\sigma_{1\eta}}{\sigma_\eta^2} = \frac{\sigma_{1\xi}}{\sigma_\xi^2}.$$

Assumption 3 *The coefficients $\beta_2 = (\beta_{21}, \dots, \beta_{2k_2})'$ are linked to the coefficients $\mu = (\mu_1, \dots, \mu_{k_2})'$ in the population regression of x_1 on X_2 through the relationship $\beta_{2i}/\beta_{2j} = \mu_i/\mu_j$ for all i, j.*

Assumption 4 *$\sigma_{1\eta}$ has the same sign as $\mathrm{cov}(x_1, \eta^*)$, where $\eta^* = \beta_{2u}' X_2$ and β_{2u} is the vector of coefficients on X_2 in the population regression of y on x_1 and X_2.*

None of these assumptions is particularly intuitive, plausible, or easy to verify. Assumption 4 is particularly obscure as formulated, but it amounts to assuming that $\sigma_{1\eta}$ and $\beta_{1r} - \beta_{1u}$ have the same sign.

Oster presents two main results. Both require Assumption 1, but while the first (Proposition 1) also requires Assumption 2 with $\varphi = 1$ and Assumption 3, the second (Proposition 2) only requires Assumption 2 with φ unrestricted.

Oster's Proposition 1 gives the following representation of b_{1u}:

$$b_{1u} = (\beta_{1r} - \beta_{1u})\frac{R_{max} - R_u}{R_u - R_r}, \tag{3}$$

where R_{max} is the unknown population R-square from the DGP (1), and R_r and R_u are the population R-squares from the short and long regressions, respectively. Notice that $R_{max} - R_u$ and $R_u - R_r$ are both positive, so (3) implies that b_{1u} has the same sign as $\beta_{1r} - \beta_{1u}$. As stressed by Holly (1982), this is not generally true. Further, (3) also implies that $b_{1r}/b_{1u} > 1$, so adding X_2 to the short regression always reduces the bias in estimating β_1. As stressed by DMP, this is also not generally true.

Oster's Proposition 2 tells us that b_{1u} is a root of the cubic equation

$$a_3 z^3 + a_2 z^2 + a_1 z + a_0 = 0, \tag{4}$$

with real coefficients

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

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. This confirms that the inconsistency of the unrestricted estimator depends on the differences $\beta_{1r} - \beta_{1u}$, $R_{max} - R_u$, and $R_u - R_r$, but tells us little about the nature of this dependence. Further, when (4) admits three roots, it is unclear how to select one. Oster argues that the problem does not arise when $\varphi = 1$ and Assumptions 1, 2, and 4 hold, because the quadratic equation $a_2 z^2 + a_1 z + a_0 = 0$ has a unique root.

To clarify the relations between $\beta_{1r} - \beta_{1u}$, $R_{max} - R_u$, $R_u - R_r$, and b_{1u}, we offer the following result.

Theorem 1 *Under the DGP (1), if Assumptions 1 and 2 hold, then*

$$\begin{aligned}
\beta_{1r} - \beta_{1u} &= \frac{\sigma_{1\eta} - (\sigma_1^2 - \sigma_\nu^2)b_{1u}}{\sigma_1^2}, \\
\varphi\sigma_y^2(R_{max} - R_u) &= \left(\frac{\sigma_\eta^2}{\sigma_{1\eta}} - \varphi b_{1u}\right)\sigma_\nu^2 b_{1u}, \\
\sigma_y^2(R_u - R_r) &= \sigma_\eta^2 + \sigma_\nu^2 b_{1u}^2 - \frac{1}{\sigma_1^2}(\sigma_{1\eta} + \sigma_\nu^2 b_{1u})^2.
\end{aligned}$$

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 $c_0 = -\varphi\sigma_{21}^2\sigma_2^2\sigma_y^2(R_{max} - R_u)$, $c_1 = \sigma_1^2\sigma_2^2(\sigma_1^2\sigma_2^2 - \sigma_{21}^2)(\beta_{1r} - \beta_{1u})$, and $c_2 = (1-\varphi)\sigma_{21}^2(\sigma_1^2\sigma_2^2 - \sigma_{21}^2)$. If $k_2 > 1$, then b_{1u} is a root of the cubic equation (4).

An implication of Theorem 1 is that Oster's Proposition 1 holds if and only if there is only one control variable in X_2.

Corollary 1 *When $\varphi = 1$ and Σ_{22} is nonsingular, the relationship (3) holds if and only if $k_2 = 1$.*

When $k_2 = 1$ but $\varphi \neq 1$, Theorem 1 implies the following result.

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

$$\varphi_1^* = 1 - \sqrt{1 + \frac{1}{\rho_{21}^2}\frac{R_u - R_r}{R_{max} - R_u}}, \qquad \varphi_2^* = 1 + \sqrt{1 + \frac{1}{\rho_{21}^2}\frac{R_u - R_r}{R_{max} - R_u}},$$

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.

When $k_2 > 1$, Theorem 1 implies Oster's Proposition 2 but does not require the controls in X_2 to be orthogonal to each other. In this more general case, the restriction $\varphi = 1$ yields $a_3 = 0$ and $a_2 a_0 = -\sigma_\nu^2\sigma_y^2(R_{max} - R_u)\sigma_1^4(\beta_{1r} - \beta_{1u})^2 < 0$, so (4) reduces to a quadratic equation with two real roots of opposite sign. When $\varphi = 1$, Assumption 4 allows one to select a unique root because it restricts b_{1u} to have the same sign as $\beta_{1r} - \beta_{1u}$. However, when $\varphi \neq 1$ or Assumption 4 does not hold, then one may select the wrong solution even when the values of φ and R_{max} are known.

Based on Proposition 2, Oster discusses three possible empirical strategies: (i) find the bounds on β_1 implied by given bounds on φ and R_{max}; (ii) find the value of φ that is consistent with given values of β_1 and R_{max}; (iii) find the value of R_{max} that is consistent with given values of β_1 and φ.

One problem with strategy (i) is that we would need a unique value of β_1 for any possible choice of φ and R_{max}. When there are multiple roots, we have a problem. To illustrate, suppose that $y = z_1 + z_2 - z_3 + z_4 + \epsilon$, where the z_j's are jointly normal with mean zero and second moment matrix

$$\Sigma = \text{var}\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{bmatrix} 1 & 0.35 & -0.30 & -0.40 \\ 0.35 & 1 & -0.25 & 0 \\ -0.30 & -0.25 & 1 & 0 \\ -0.40 & 0 & 0 & 1 \end{bmatrix},$$

and $\epsilon \sim N(0,1)$ independently of the z_j's. Setting $x_1 = z_1$, $X_2 = (z_2, z_3)$ and $\xi = z_4$, we find $\varphi = -1.537$ and $R_{max} = 0.833$. Table 1 shows the OLS estimates of, respectively, the true DGP,

Table 1: OLS estimates of the DGP, the long and the short regressions

	DGP		Long		Short	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE
z_1	0.995	(0.039)	0.509	(0.047)	1.231	(0.066)
z_2	0.994	(0.036)	1.140	(0.047)		
z_3	−0.975	(0.034)	−1.085	(0.045)		
z_4	0.963	(0.035)				
R^2	0.825		0.689		0.261	

the long regression of y on x_1 and X_2, and the short regression of y on x_1, from a pseudo-random sample of 1,000 observations. In this example, Assumption 4 fails because adding X_2 to the short regression increases the magnitude of the bias in estimating β_1 (in the terminology of DMP, X_2 is not a balanced addition). Following Oster, we set the lower bound for β_1 equal to the value 0.509 from the long regression. To determine the upper bound, we employ Oster's Stata routine with $\varphi = -1.537$ and $R_{max} = 0.833$. This gives three solutions for β_1, namely $\beta_1^{(1)} = 4.590$, $\beta_1^{(2)} = 1.087$ and $\beta_1^{(3)} = 1.582$. While the second solution is close to the true value $\beta_1 = 1$, the routine selects the first solution due to the failure of Assumption 4.

As for the other two empirical strategies suggested by Oster, 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 Assumption 1, this allows us to identify $\sigma_{1\xi}$ and $\sigma_{1\eta}$, and therefore also σ_η^2 from the third equation in Theorem 1. By restricting either R_{max} or φ, we can then identify σ_ξ^2. Thus, under Assumptions 1 and 2, the empirical strategies (ii) and (iii) amount to imposing arbitrary restrictions on all the unidentified model parameters. Finally, the results obtained are very sensitive to the choice of R_{max} for strategy (ii) and of φ for strategy (iii). To illustrate, consider again our example with $\beta_1 = 1$ and four different values of R_{max}, namely 0.7, 0.8, 0.9 and 1. In this case, the solutions for φ obtained from Oster's Stata routine range widely, being equal to −5.662, −1.765, −1.046, and −0.743, respectively.

4 Testing strategies

To map our notation into the notation in PPS, let $x_1 = s$, $\beta_1 = \beta^l$, $X_2 = \boldsymbol{x}^m$, $\beta_2 = 0$, and $\xi = \boldsymbol{\gamma}'\boldsymbol{x} = \boldsymbol{\gamma}'(\boldsymbol{\delta}s + \boldsymbol{u})$. Given a classical measurement error model, PPS specify their balancing regression as $\boldsymbol{x}^m = \boldsymbol{x} + \boldsymbol{m} = \boldsymbol{\delta}s + \boldsymbol{u} + \boldsymbol{m}$, where all components of $(\boldsymbol{u}, \boldsymbol{m})$ are uncorrelated with s so that $\psi = \boldsymbol{\delta}$. When $k_2 = 1$, they also consider a mean-reverting measurement error model of the form $x_2 = x^m = (1+\kappa)x + \mu = (1+\kappa)\delta s + (1+\kappa)u + \mu$, where $-1 < \kappa < 0$ and all components of

(s, u, μ) are uncorrelated with each other. In this case, the population coefficient of the balancing regression is $\psi = (1 + \kappa)\delta$.

The main contribution of PPS is to provide power comparisons between two alternative strategies for testing the consistency of the restricted OLS estimator: an F-test on the population coefficients $\psi = \sigma_{21}/\sigma_1^2$ in the balancing regression, and a Hausman-type test on the difference $\beta_{1r} - \beta_{1u} = b_{1r} - b_{1u} = \psi'(\beta_2 + \tau_2)$ between the coefficient of interest in the short and the long regressions. PPS refer to these tests as the balancing test (BT) and the coefficient comparison test (CCT), respectively. Their results show that, when the long regression is misspecified (i.e. $\boldsymbol{\gamma} \neq \mathbf{0}$), BT is generally more powerful than CCT because measurement errors are comparatively less harmful when mismeasured variables are employed as outcome variables in the balancing regression rather than as additional regressors in the long regression.

This useful insight reinforces our conclusion that adding control variables to the short regression does not necessarily improve the estimation of the causal effect of interest. While DMP and Oster are mainly concerned with the statistical properties of the restricted and unrestricted estimators of β_1, PPS focus on the implications of using the available control variables for testing purposes. However, as shown by the MSE comparisons in DMP, these two approaches are closely related to each other. Under certain conditions, MSE comparisons depend crucially on the noncentrality parameter in the distribution of the statistic (either the classical F-statistic or the Hausman-type statistic) used for testing the hypothesis $H_0 : \beta_2 = 0$ in the long regression.

Our general framework allows us to assess whether the poor power performance of CCT depends on the particular specification of the measurement error models considered by PPS. It follows immediately from (2) that BT and CCT provide tests of the null hypothesis of interest,

$$H_0 : b_{1r} = \tau_1 + \psi'(\beta_2 + \tau_2) = 0, \tag{5}$$

only if suitable restrictions are placed on $b_{1u} = \tau_1$. If these restrictions are not valid, then there exist regions of the parameter space where both BT and CCT have large size distortion and low power. BT is concerned with the null hypothesis $H_0 : \psi = 0$. Writing $\tau_1 = \sigma_{1\xi}/\sigma_1^2 - \psi'\tau_2$, we see that this is equivalent to (5) if and only if there exists a k_2-vector $\omega \neq -\beta_2$ such that $\sigma_{1\xi} = \sigma_{21}'\omega$, so that

$$b_{1u} = \tau_1 = \psi'(\omega - \tau_2), \qquad b_{1r} = \psi'(\beta_2 + \omega). \tag{6}$$

CCT is instead concerned with the null hypothesis $H_0 : \psi'(\beta_2 + \tau_2) = 0$, which is equivalent to (5) if and only if there exist a scalar $a \neq -1$ such that

$$b_{1u} = \tau_1 = a\psi'(\beta_2 + \tau_2), \qquad b_{1r} = (1 + a)\psi'(\beta_2 + \tau_2). \tag{7}$$

The restrictions (6) and (7) may constrain the sign and magnitude of b_{1u} and b_{1r}. For example, when $k_2 = 1$, 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, (6) and (7) 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 (6) and (7) 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, when there are multiple controls subject to measurement error (i.e. $k_2 > 1$), the condition $b_{1r}/b_{1u} > 1$ doesn't need to 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.

Finally, 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.

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*, forthcoming.

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.

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. (2017), "Poorly Measured Confounders Are More Useful on the Left Than on the Right," NBER Working Paper No. 23232.

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. The first three expressions follow from the fact that

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

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

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

When $k_2 = 1$, (A1) and (A3) do not allow separate identification of β_2 and b_{1u}. The result for $k_2 = 1$ follows by solving the system of equations (A1) and (A2) in the unknowns β_2 and b_{1u}, while the result for $k_2 > 1$ is the same as Oster's Proposition 2.

Proof of Corollary 1. When $\varphi = 1$, it follows from (A1)–(A3) that

$$(\beta_{1r}-\beta_{1u})\frac{R_{max}-R_u}{R_u-R_r} = \frac{z'\Omega z}{z'\Xi z}\,b_{1u},$$

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 ratio $z'\Omega z/z'\Xi z$ equals one for every z if and only if $z'[\Omega-\Xi]z = 0$ for every z, and this occurs if and only if $\Omega+\Omega' = \Xi+\Xi' = 2\,\Xi$. Hence, $z'\Omega z/z'\Xi z$ equals one for every z if and only if

$$(\sigma_1^2-\sigma_{21}'\Sigma_{22}^{-1}\sigma_{21})\left[\Sigma_{22}\beta_2\sigma_{21}'+\sigma_{21}\beta_2'\Sigma_{22}\right] = 2\beta_2'\sigma_{21}\left[\sigma_1^2\Sigma_{22}-\sigma_{21}\sigma_{21}'\right]. \tag{A4}$$

If (A4) holds, then postmultiplying by β_2 gives

$$(\sigma_1^2-\sigma_{21}'\Sigma_{22}^{-1}\sigma_{21})\left[\Sigma_{22}\beta_2(\sigma_{21}'\beta_2)+\sigma_{21}(\beta_2'\Sigma_{22}\beta_2)\right] = 2\beta_2'\sigma_{21}\left[\sigma_1^2\Sigma_{22}\beta_2-\sigma_{21}(\sigma_{21}'\beta_2)\right].$$

Hence, upon rearranging terms,

$$(\sigma_1^2+\sigma_{21}'\Sigma_{22}^{-1}\sigma_{21})(\beta_2'\sigma_{21})\Sigma_{22}\beta_2 = \left[(\sigma_1^2-\sigma_{21}'\Sigma_{22}^{-1}\sigma_{21})(\beta_2'\Sigma_{22}\beta_2)+2(\beta_2'\sigma_{21})^2\right]\sigma_{21}.$$

This shows that $\Sigma_{22}\beta_2$ is a multiple of σ_{21}, say $\Sigma_{22}\beta_2 = \alpha\sigma_{21}$. Inserting $\beta_2 = \alpha\Sigma_{22}^{-1}\sigma_{21}$ into (A4) gives $(\Sigma_{21}'\Sigma_{22}^{-1}\sigma_{21})\Sigma_{22} = \sigma_{21}\sigma_{21}'$, which implies that Σ_{22} has rank one. Since Σ_{22} is nonsingular, this is only possible if $k_2 = 1$.

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

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

and using the fact that $\sigma_y^2(R_u-R_r)/(\beta_{1r}-\beta_{1u})^2 = \sigma_1^2(\sigma_1^2\sigma_2^2-\sigma_{21}^2)/\sigma_{21}^2$, from (A1) and (A3).