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

\[ y = \beta_1 x_1 + \beta_2' X_2 + \xi + \epsilon, \]

(1)

where \( x_1 \) is an observable scalar treatment, \( X_2 \) is a set of \( k_2 \) observable control variables, \( \beta_1 \) and \( \beta_2 \) contain the unknown parameters, \( \xi \) is an unobservable misspecification term capturing, for example, the contributions of omitted variables (as in Oster) or measurement errors (as in PPS), and \( \epsilon \) 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 \( \beta_1 \), which is interpreted as the causal effect of \( x_1 \) on \( y \). The population second moments of \((x_1,X_2,\xi)\) are denoted by

\[ \Sigma = \text{var} \begin{pmatrix} x_1 \\ X_2 \\ \xi \end{pmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12}' & \sigma_{1\xi} \\ \sigma_{21} & \Sigma_{22} & \sigma_{2\xi}' \\ \sigma_{1\xi} & \sigma_{2\xi}' & \sigma_{\xi\xi} \end{bmatrix}. \]

When \( k_2 = 1 \), we write \( x_2 \) instead of \( X_2 \) and \( \sigma_2^2 \) instead of \( \Sigma_{22} \).

Since \( \xi \) is unknown, we consider two alternative estimators of \( \beta_1 \): the restricted OLS estimator from the “short regression” of \( y \) on \( x_1 \) only, with probability limit denoted by \( \beta_{1r} \), and the unrestricted OLS estimator from the “long regression” of \( y \) on \( x_1 \) and \( X_2 \), with probability limit denoted by \( \beta_{1u} \). From DMP, the inconsistencies of these two estimators are

\[ b_{1r} = \beta_{1r} - \beta_1 = \tau_1 + \psi'(\beta_2 + \tau_2), \quad b_{1u} = \beta_{1u} - \beta_1 = \tau_1, \]

(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 \( \xi \) 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 \( b_{1r} - b_{1u} = \beta_1 - \beta_{1u} = \psi'(\beta_2 + \tau_2), \) 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 \( \sigma^2_\eta = \text{var}(\eta) = \beta_2'\Sigma_{22}\beta_2, \sigma_{1\eta} = \text{cov}(x_1, \eta) = \sigma'_{21}\beta_2, \) and \( \sigma_{\eta\xi} = \text{cov}(\eta, \xi) = \beta_2'\sigma_{2\xi}. \)

Oster’s main contribution are expressions for the inconsistency \( b_{1u} \) of the unrestricted estimator of \( \beta_1. \) To derive these expressions, Oster imposes the following four assumptions:

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

**Assumption 2** \( x_1, \eta \) and \( \xi \) are linked through the “proportional selection relationship”

\[
\varphi \frac{\sigma_{1\eta}}{\sigma^2_{\eta}} = \frac{\sigma_{1\xi}}{\sigma^2_{\xi}}.
\]

**Assumption 3** The coefficients \( \beta_2 = (\beta_{21}, \ldots, \beta_{2k_2})' \) are linked to the coefficients \( \mu = (\mu_1, \ldots, \mu_{k_2})' \) 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 \( \text{cov}(x_1, \eta^*), \) where \( \eta^* = \beta_{2u}'X_2 \) and \( \beta_{2u} \) 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 \( \varphi \) unrestricted.
Oster’s Proposition 1 gives the following representation of $b_{1u}$:

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

where $R_{\text{max}}$ 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 $R_{\text{max}} - R_u$ 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 $\beta_1$. 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

$$a_3z^3 + a_2z^2 + a_1z + a_0 = 0, \quad (4)$$

with real coefficients

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

where $\sigma_y^2$ and $\sigma_\nu^2 = \sigma_1^2 - \sigma_2^1 \Sigma_{22}^{-1} \sigma_{21}$ 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 $\beta_{1r} - \beta_{1u}$, $R_{\text{max}} - R_u$, 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 $a_2z^2 + a_1z + a_0 = 0$ has a unique root.

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

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

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

$$\varphi \sigma_y^2 (R_{\text{max}} - R_u) = \left( \frac{\sigma_y^2}{\sigma_1 \eta} - \varphi b_{1u} \right) \sigma_\nu^2 b_{1u},$$

$$\sigma_y^2 (R_u - R_r) = \sigma_y^2 + \sigma_\nu^2 b_{1u}^2 - \frac{1}{\sigma_1^2} (\sigma_1 \eta + \sigma_\nu^2 b_{1u})^2.$$
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 $c_0 = -\varphi \sigma_{21}^2 \sigma_y^2 (R_{\text{max}} - R_u)$, $c_1 = \sigma_{11}^2 \sigma_y^2 (\sigma_{22}^2 - \sigma_{21}^2) (\beta_1 - \beta_{1u})$, and $c_2 = (1 - \varphi) \sigma_{21}^2 (\sigma_{22}^2 - \sigma_{21}^2)$. 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 $\Sigma_{22}$ 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

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

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 $a_2 a_0 = -\sigma_{2}^2 \sigma_y^2 (R_{\text{max}} - R_u) \sigma_{11}^2 (\beta_{1r} - \beta_{1u})^2 < 0$, 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 $\varphi$ and $R_{\text{max}}$ are known.

Based on Proposition 2, Oster discusses three possible empirical strategies: (i) find the bounds on $\beta_1$ implied by given bounds on $\varphi$ and $R_{\text{max}}$; (ii) find the value of $\varphi$ that is consistent with given values of $\beta_1$ and $R_{\text{max}}$; (iii) find the value of $R_{\text{max}}$ that is consistent with given values of $\beta_1$ and $\varphi$.

One problem with strategy (i) is that we would need a unique value of $\beta_1$ for any possible choice of $\varphi$ and $R_{\text{max}}$. 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 $R_{\text{max}} = 0.833$. Table 1 shows the OLS estimates of, respectively, the true DGP,
Table 1: OLS estimates of the DGP, the long and the short regressions

<table>
<thead>
<tr>
<th>Variable</th>
<th>DGP</th>
<th>Long</th>
<th>Short</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>SE</td>
<td>Coeff.</td>
</tr>
<tr>
<td>$z_1$</td>
<td>0.995 (0.039)</td>
<td>0.509 (0.047)</td>
<td>1.231 (0.066)</td>
</tr>
<tr>
<td>$z_2$</td>
<td>0.994 (0.036)</td>
<td>1.140 (0.047)</td>
<td></td>
</tr>
<tr>
<td>$z_3$</td>
<td>-0.975 (0.034)</td>
<td>-1.085 (0.045)</td>
<td></td>
</tr>
<tr>
<td>$z_4$</td>
<td>0.963 (0.035)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.825</td>
<td>0.689</td>
<td>0.261</td>
</tr>
</tbody>
</table>

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 $\beta_1$ (in the terminology of DMP, $X_2$ is not a balanced addition). Following Oster, we set the lower bound for $\beta_1$ equal to the value 0.509 from the long regression. To determine the upper bound, we employ Oster’s Stata routine with $\phi = -1.537$ and $R_{max} = 0.833$. This gives three solutions for $\beta_1$, 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 $\beta_1$ for given values of $\beta_{1r}$ and $\beta_{1u}$ 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 $\sigma_{\eta}^2$ from the third equation in Theorem 1. By restricting either $R_{max}$ or $\phi$, we can then identify $\sigma_{\xi}^2$. 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 $\phi$ 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 $\phi$ 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 $x_1 = s$, $\beta_1 = \beta^l$, $X_2 = x^m$, $\beta_2 = 0$, and $\xi = \gamma'x = \gamma'(\delta s + u)$. Given a classical measurement error model, PPS specify their balancing regression as $x^m = x + m = \delta s + u + m$, where all components of $(u, m)$ are uncorrelated with $s$ so that $\psi = \delta$. 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
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. $\gamma \neq 0$), 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 $\beta_1$, 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,

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

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

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

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

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
\[ \psi = (1 + \kappa) \delta, \quad \tau_1 = \delta \gamma \left( \frac{\theta}{(1 + k)^2 (1 - \theta) + \theta} \right), \quad \tau_2 = \frac{\gamma}{1 + k} \left[ 1 - \left( \frac{\theta}{(1 + k)^2 (1 - \theta) + \theta} \right) \right], \]
with $\theta = \sigma^2_{\mu} / (\sigma^2_{\mu} + \sigma^2_u)$. 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


Appendix

Proof of Theorem 1. The first three expressions follow from the fact that

\[
\sigma^2_1(\beta_{1r} - \beta_{1u}) = [\beta_2 - \Sigma_{22}^{-1}\sigma_{21} b_{1u}]' \sigma_{21}, \tag{A1}
\]

\[
\varphi \sigma^2_y(R_{\max} - R_u) = \frac{\sigma^2 - \sigma_{12}' \Sigma_{22}^{-1} \sigma_{21}}{\beta_2' \sigma_{21}} [\beta_2 - \varphi \Sigma_{22}^{-1} \sigma_{21} b_{1u}]' \Sigma_{22} \beta_2 b_{1u}, \quad (A2)
\]

\[
\sigma^2_y(R_u - R_r) = [\beta_2 - \Sigma_{22}^{-1} \sigma_{21} b_{1u}]' \Sigma_{22} \left[ \frac{1}{\sigma^2_1} \sigma_{21} \sigma_{21}' \right] \left[ \beta_2 - \Sigma_{22}^{-1} \sigma_{21} b_{1u} \right]. \tag{A3}
\]

When \( k_2 = 1 \), (A1) and (A3) do not allow separate identification of \( \beta_2 \) and \( b_{1u} \). The result for \( k_2 = 1 \) follows by solving the system of equations (A1) and (A2) in the unknowns \( \beta_2 \) 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

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

where \( z = \beta_2 - \Sigma_{22}^{-1} \sigma_{21} b_{1u}, \ \Omega = (\sigma^2_2 - \sigma_{12}' \Sigma_{22}^{-1} \sigma_{21}) \Sigma_{22} \beta_2 \sigma_{21}' \), and \( \Xi = \beta_2' \sigma_{21} [\sigma^2_1 \Sigma_{22} - \sigma_{21} \sigma_{21}'] \). The ratio \( z' \Omega / z' \Xi \) 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, \( z' \Omega / z' \Xi \) equals one for every \( z \) if and only if

\[
(\sigma^2_1 - \sigma_{12}' \Sigma_{22}^{-1} \sigma_{21}) \left[ \Sigma_{22} \beta_2 \sigma_{21}' + \sigma_{21} \beta_2' \Sigma_{22} \right] = 2 \beta_2' \sigma_{21} \left[ \sigma^2_1 \Sigma_{22} - \sigma_{21} \sigma_{21}' \right]. \tag{A4}
\]

If (A4) holds, then postmultiplying by \( \beta_2 \) gives

\[
(\sigma^2_1 - \sigma_{12}' \Sigma_{22}^{-1} \sigma_{21}) \left[ \Sigma_{22} \beta_2 \sigma_{21}' + \sigma_{21} \beta_2' \Sigma_{22} \right] = 2 \beta_2' \sigma_{21} \left[ \sigma^2_1 \Sigma_{22} \beta_2 - \sigma_{21} (\beta_2' \beta_2) \right].
\]

Hence, upon rearranging terms,

\[
(\sigma^2_1 + \sigma_{12}' \Sigma_{22}^{-1} \sigma_{21}) (\beta_2' \sigma_{21}) \Sigma_{22} \beta_2 = \left[ (\sigma^2_1 - \sigma_{12}' \Sigma_{22}^{-1} \sigma_{21}) (\beta_2' \Sigma_{22} \beta_2) + 2 (\beta_2' \sigma_{21})^2 \right] \sigma_{21}.
\]

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

Proof of Corollary 2. The result follows by solving \( \varphi \) from the equation

\[
0 = (\sigma^2_1 \sigma^2_2 - \sigma^2_{12}) \frac{\sigma^2_1}{\sigma^2_{21}} (\beta_{1r} - \beta_{1u})^2 + 4 \varphi (1 - \varphi) \frac{\sigma^2_y (R_{\max} - R_u)}{\sigma^2_y},
\]

and using the fact that \( \sigma^2_y (R_u - R_r) / (\beta_{1r} - \beta_{1u})^2 = \sigma^2_1 (\sigma^2_1 \sigma^2_2 - \sigma^2_{12}) \sigma^2_{21} \), from (A1) and (A3).