DEPARTMENT OF ECONOMICS

Working Paper

Bias of OLS Estimators due to Exclusion of Relevant Variables and Inclusion of Irrelevant Variables

Deepankar Basu

Working Paper 2018-19

UNIVERSITY OF MASSACHUSETTS AMHERST

Bias of OLS Estimators due to Exclusion of Relevant Variables and Inclusion of Irrelevant Variables

Deepankar Basu*

December 1, 2018

Abstract

In this paper I discuss three issues related to bias of OLS estimators in a general multivariate setting. First, I discuss the bias that arises from omitting relevant variables. I offer a geometric interpretation of such bias and derive sufficient conditions in terms of sign restrictions that allows us to determine the direction of bias. Second, I show that inclusion of some omitted variables will not necessarily reduce the magnitude of OVB as long as some others remain omitted. Third, I show that inclusion of irrelevant variables in a model with omitted variables can also have an impact on the bias of OLS estimators. I use the running example of a simple wage regression to illustrate my arguments.

JEL Codes: C20

Keywords: omitted variable; irrelevant variables; ordinary least squares; bias.

1 Introduction

This paper studies three issues related to the problem of bias of ordinary least squares (OLS) estimators that arise from errors of exclusion (of relevant variables) and inclusion (of irrelevant variables). The first issue relates to the possibility of determining the direction of omitted variable bias (OVB) in a general multivariate setting - a longstanding issue in econometrics; the second issue relates to the possibility of reducing bias of OLS estimators with the inclusion of *some* of the variables that were excluded in the first place; and the third issue relates to the possible interaction between the errors of exclusion and inclusion in determining the bias of OLS estimators. Since OLS estimation remains a work horse of applied econometrics research, the issues discussed in this paper will be of interest to a wide range of researchers in the social sciences, including in economics, sociology, political science.

*Department of Economics, University of Massachusetts, 310 Crotty Hall, Amherst, MA 01003, email: dbasu@econs.umass.edu. I would like to thank Michael Ash and Junshang Liang for comments and suggestions. This paper supersedes Basu (2018).

Let me introduce the first issue by considering a simple population regression function,

$$y = \beta_0 + \beta_1 x + \gamma_1 z + u, \tag{1}$$

where the expected value of u conditional on x and z, is zero, i.e. $\mathbb{E}(u|x,z) = 0$. Suppose a researcher does not have data on the variable z and ends up estimating the following model,

$$y = \alpha_0 + \alpha_1 x + v,$$

with OLS. Since the omitted variable, z, might be correlated with the included regressor, x, leaving it out of the estimated model induces correlation between the error term and the regressor. This makes the OLS estimator of the parameters biased. Moreover, conditional on x and z, the bias of OLS estimator of the slope parameter, β_1, is given by

$$\mathbb{E}\left(\hat{\alpha}_1\right) - \beta_1 = \left(\frac{\hat{\sigma}_z}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z}\gamma_1, \tag{2}$$

where $\hat{\rho}_{x,z}$ is the sample correlation coefficient between the omitted and included regressor, and $\hat{\sigma}_z$ and $\hat{\sigma}_x$ are sample standard deviations of z and x, respectively. It is often of interest to ascertain the direction of bias, even though the magnitude cannot be, in general, determined. Since $\hat{\sigma}_z$ and $\hat{\sigma}_x$ are always positive, it is possible, in this case, to determine the direction of bias only using information on the signs of γ_1 and $\hat{\rho}_{x,z}$: if both are of the same sign, the bias is positive; if both are of opposite signs, the bias is negative.

Let me introduce the second issue by considering a slight extension of the above framework and consider the following population regression function,

$$y = \beta_0 + \beta_1 x + \gamma_1 z_1 + \gamma_2 z_2 + u, \tag{3}$$

where $\mathbb{E}(u|x, z_1, z_2) = 0$. In the context of this model, let us define a estimated "short regression" model as

$$y = \alpha_0 + \alpha_1 x + v,$$

where both the variables z_1 and z_2 have been omitted, and an estimated "long regression" model as

$$y = \delta_0 + \delta_1 x + \delta_2 z_1 + w,$$

where only z_2 has been omitted.

Let $\hat{\alpha}_1$ and $\hat{\delta}_1$ denote the OLS estimators of the coefficient on x in the short and long regressions, respectively. Since x, z_1 and z_2 are likely to be correlated, both the estimated short and long regression models give biased OLS estimators of the parameters in the population regression function - because in both cases, the error term is likely to be correlated with the included regressors. Thus, both $\hat{\alpha}_1$ and $\hat{\delta}_1$ are biased estimators of β_1, the coefficient of x in the population regression function. Moreover, conditional on x, z_1, z_2, expressions for the bias are as follows:

$$\mathbb{E}\left(\hat{\alpha}_1\right) - \beta_1 = \left(\frac{\hat{\sigma}_{z_1}}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z_1}\gamma_1 + \left(\frac{\hat{\sigma}_{z_2}}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z_2}\gamma_2 \tag{4}$$

where $\hat{\rho}_{x,z_1}$ is the sample correlation coefficient between x and z_1, $\hat{\rho}_{x,z_2}$ is the sample correlation coefficient between x and z_2, $\hat{\sigma}_x, \hat{\sigma}_{z_1}$ and $\hat{\sigma}_{z_2}$ are sample standard deviations of x, z_1 and z_2, respectively; and

$$\mathbb{E}\left(\hat{\delta}_1\right) - \beta_1 = \left(\frac{\hat{\sigma}_{z_2}}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z_2}\gamma_2 + \left(\frac{\hat{\sigma}_{z_1}}{\hat{\sigma}_x}\right)\left(\frac{\hat{\sigma}_{z_2}}{\hat{\sigma}_{\hat{v}}}\right)\hat{\rho}_{x,z_1}\hat{\rho}_{\hat{v},z_2}\gamma_2 \tag{5}$$

where $\hat{\rho}_{x,z_2}$ and $\hat{\rho}_{\hat{v},z_2}$ are the sample correlation coefficients, respectively, between z_2 and x, and z_2 and $\hat{v}$, where the latter are the (negative of) the residuals from an auxiliary regression of z_1 on x, and $\hat{\sigma}_{\hat{v}}$ denotes the sample standard deviation of $\hat{v}$.

To introduce the third issue, let me return to the population regression function in (1),

$$y = \beta_0 + \beta_1 x + \gamma_1 z + u,$$

and consider the case where the estimated model is given by

$$y = \alpha_0 + \alpha_1 x + \alpha_2 w + v.$$

In this case, the estimated model is doubly misspecified: it excludes the relevant variable, z, and includes an irrelevant variable, w.[1] We assume that the error term in the population regression function satisfies the following condition:

$$\mathbb{E}(u|x, z, w) = 0.$$

If the doubly misspecified model is estimated with OLS, then all the parameters will be estimated with bias. In fact, letting $\hat{\alpha}_1$ denote the OLS estimator of α_1, a little algebra allows us to pin down the bias as,

$$\mathbb{E}\left(\hat{\alpha}_1\right) - \beta_1 = \left(\frac{\hat{\sigma}_z}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z}\gamma_1 + \hat{\rho}_{x,w}\left(\frac{\hat{\sigma}_w}{\hat{\sigma}_x}\right)\hat{\rho}_{z,w^x}\left(\frac{\hat{\sigma}_z}{\hat{\sigma}_{w^x}}\right)\gamma_1, \tag{6}$$

where w^x denotes the residuals that come from the regression of w on x, $\hat{\rho}_{z,w^x}$ denotes the sample correlation coefficient of z and w^x, and other symbols have their usual interpretation.

In moving from the simplest case with one omitted variable to even slightly more complex cases, we get a preview of three important results. First, a comparison of (2) and (4) shows that it is no longer possible to determine even the direction of OVB on the basis of the signs of parameters *only* when we have more than one omitted variable. This is because, as we see from (4), the bias of the OLS estimator,

$$\left(\frac{\hat{\sigma}_{z_1}}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z_1}\gamma_1 + \left(\frac{\hat{\sigma}_{z_2}}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z_2}\gamma_2,$$

is the sum of two terms, each of which can be of any sign. One can immediately generalize this to see that, in a multivariate case, it is no longer possible to unambiguously determine

[1] A variable is considered irrelevant if it does not appear in the population regression function, i.e. its coefficient in the population regression function (the true model) is zero.

the direction of OVB on the basis of signs of parameters only, as has been known for long (Forbes, 2000; Greene, 2012). An interesting question, and one that is investigated in this paper, is if we can determine the direction of the OVB in some special cases on the basis of signs of parameters only.

Second, a comparison of (4) and (5) shows that it is no longer possible to ensure bias reduction by the inclusion of an omitted variable when some others remain omitted. This is because, using (4) and (5), we get

$$\mathbb{E}\left(\hat{\delta}_1\right) - \beta_1 = \left[\mathbb{E}\left(\hat{\alpha}_1\right) - \beta_1\right] + \left(\frac{\hat{\sigma}_{z_1}}{\hat{\sigma}_x}\right)\hat{\rho}_{x,z_1}\gamma_1 - \left(\frac{\hat{\sigma}_{z_1}}{\hat{\sigma}_x}\right)\left(\frac{\hat{\sigma}_{z_2}}{\hat{\sigma}_{\hat{v}}}\right)\hat{\rho}_{x,z_1}\hat{\rho}_{\hat{v},z_2}\gamma_2,$$

so that the bias in the long regression, $\mathbb{E}\left(\hat{\delta}_1\right) - \beta_1$, can be larger or smaller than the bias in the short regression, $\mathbb{E}\left(\hat{\alpha}_1\right) - \beta_1$.

This result goes against the common perception that including omitted variables will always lead to a reduction in bias and arises from the fact that both the short and long regressions are mis-specified. In textbook treatments, bias is reduced because the long regression includes *all* the omitted variables. But if, as seems quite realistic, the long regression also suffers from the problem of omitted variables, then it is no longer possible to ensure bias reduction unambiguously by inclusion of omitted variables. One can easily generalize this to see that, in a multivariate setting, inclusion of omitted variables will not necessarily lead to a reduction in OVB if some variables remain omitted, a result that has been highlighted in recent work (Clarke, 2005; Luca et al., 2018).[2]

Third, a comparison of (2) and (6) shows that the bias in the doubly misspecified model, where a relevant variable is omitted and an irrelevant variable is included, is the sum of the omitted variable bias - which is the first term in (6) - and an additional term that comes due to the inclusion of the irrelevant variable - which is the second term in (6). Thus, in the doubly misspecified model, the overall bias of OLS estimators can be decomposed into two terms, the first being the direct effect of the omitted variable, and the second being the indirect effect of the omitted variable, the latter working its way through the irrelevant variable.

This is a novel result and goes against the common perception that inclusion of irrelevant variables has no impact on the bias of OLS estimators (Greene, 2012, pp. 58).[3] This, perhaps surprising, result comes from the fact that the model suffers from both problems at the same time - omitting a relevant variable and including an irrelevant variable. The presence of an omitted variable interacts with the correlation between the irrelevant and included regressor - which is also mediated by the correlation between the omitted variable and the part of the irrelevant variable that is not explained by the included regressor - to contribute an additional term to the overall bias. Of course, one can see the standard result, that inclusion of irrelevant variables have no effect on bias, as a special case of this more general framework. If the estimated model does not suffer from the problem of omitted variable problems, which

[2] I would like to thank Weikai Chen for pointing me to this literature.

[3] In the standard analysis, inclusion of irrelevant variables does have an implication on the second moment of the OLS estimator: it increases the variance of the OLS estimator (Fomby, 1981; Greene, 2012).

can be captured by positing that $\gamma_1 = 0$, then the inclusion of irrelevant variables will not have any impact on the bias.[4]

This paper pursues these three ideas related to bias of OLS estimators arising from errors of exclusion and inclusion in a multivariate context. The first contribution of the paper is to offer a simple geometric interpretation of the OVB in a general setting, with many included and omitted regressors. This helps us derive a set of sufficient conditions in terms of sign restrictions on partial effects that allow us to unambiguously determine the direction of OVB. I show that these conditions are natural multivariate generalizations of the simplest univariate case. I illustrate my argument with a canonical wage regression. The second contribution of the paper is to reiterate the negative result in Clarke (2005) and Luca et al. (2018) that inclusion of omitted variables will not always lead to a reduction in OVB.[5] In fact, I emphasize that we are not even able to derive sufficient conditions for bias reduction using sign restrictions. The third, and novel, contribution of the paper is to show that the inclusion of irrelevant variables can have an impact on the bias if the estimated model also suffers from the problem of omitted variables. While we cannot determine the direction of the impact in general, I derive sufficient conditions using sign restrictions that allow us to make assertions about the direction of the impact. I illustrate the arguments in the paper with a running example of a simple wage regression.

The rest of the paper is organized as follows: in section 2, I provide a general result on bias coming from omitting relevant variables and from including irrelevant variables in a multivariate context; in section 3, I offer a geometric interpretation and derive sufficient conditions using sign restrictions that allow us to unambiguously determine the direction of OVB; in section 4, I reiterate the fact that, in general, we cannot unambiguously ensure reduction in bias with inclusion of excluded variables if some variables remain omitted; in section 5, I discuss the doubly misspecified model and decompose the overall bias of OLS estimators into a component coming from omitting relevant variables and another coming from including irrelevant variables. I conclude in the final section by drawing some conclusions of the analysis in this paper. Proofs of propositions are collected together in Appendix A, and in Appendix B, I give some examples of the use of direction-of-bias arguments from the applied economics literature of the past several decades.

2 Omitted Variable Bias in a General Setting

To fix ideas, let the population regression function of interest be denoted as

$$\boldsymbol{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{Z_1}\boldsymbol{\gamma_1} + \boldsymbol{Z_2}\boldsymbol{\gamma_2} + \boldsymbol{u} \tag{7}$$

where $\boldsymbol{y}$ is a $N \times 1$ vector representing the dependent variable, $\mathbf{X}$, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ denote $(N \times J)$, $(N \times K)$ and $(N \times L)$ matrices, respectively, of regressors, $\boldsymbol{\beta}$, $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$ denote $(J \times 1)$,

[4] I would like to thank Michael Ash for suggesting this idea for investigation.

[5] In certain special cases, it is possible to derive explicit expressions for the OVB even in the multivariate case (Clarke, 2019).

$(K \times 1)$ and $(L \times 1)$ denote vectors of population regression coefficients, and $\boldsymbol{u}$ is the $N \times 1$ vector of errors which satisfy

$$\mathbb{E}\left[\boldsymbol{u}|\boldsymbol{X}, \boldsymbol{Z_1}, \boldsymbol{Z_2}\right] = \mathbf{0}. \tag{8}$$

If we collect $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ into the $N \times M$ matrix $\boldsymbol{Z} = [\boldsymbol{Z_1} \quad \boldsymbol{Z_2}]$, where $M = K + L$, and similarly collect together $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$ into the $M \times 1$ vector of coefficients $\boldsymbol{\gamma} = \left(\boldsymbol{\gamma_1'} \quad \boldsymbol{\gamma_2'}\right)'$, then we can also write the population regression function as

$$\boldsymbol{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{u}. \tag{9}$$

We would like to compare two scenarios. In the first scenario, the researcher is not able to include $\mathbf{Z_1}$ and $\mathbf{Z_2}$ in the estimated model. Let us call this the "short" regression model:

$$\boldsymbol{y} = \mathrm{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{v_S}. \tag{10}$$

In the second scenario, the researcher is able to include, $\mathbf{Z_1}$, in the estimated regression, but is not able to include the regressors, $\mathbf{Z_2}$. Let us call this the "long" regression model:

$$\boldsymbol{y} = \mathrm{X}\tilde{\tilde{\boldsymbol{\beta}}} + \mathbf{Z_1}\gamma_1 + \boldsymbol{v_L}. \tag{11}$$

Let us call the OLS estimator of $\tilde{\boldsymbol{\beta}}$ in (10) as $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$, and the OLS estimator of $\tilde{\tilde{\boldsymbol{\beta}}}$ in (11) as $\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}$, and note that both are likely to be biased estimators of $\boldsymbol{\beta}$ because of the possible correlation between $\boldsymbol{X}, \boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$.[6]

We would like to address two questions related to the bias of OLS estimators that arise due the error of excluding relevant variables from the estimated model. First, what is the omitted variable bias of an included regressor in the short regression when it is estimated with OLS, i.e. what is the OVB of $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$? Can we derive sufficient conditions using only signs of parameters to determine the direction of the OVB? Second, can we compare the omitted variable bias of an included regressor between the short and long regressions, when both are estimated with OLS, i.e. can we compare the OVB of $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}$? The following proposition gives a set of results to answer these two questions.

Proposition 1. *Conditional on the regressors $\boldsymbol{X}, \boldsymbol{Z}$, the omitted variable bias in $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$ is given by*

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}\right) - \boldsymbol{\beta} = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z}\gamma = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1}\gamma_1 + \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_2}\boldsymbol{\gamma_2} \tag{12}$$

and the omitted variable bias in $\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}$ is given by

$$\mathbb{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{L}} - \boldsymbol{\beta} = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_2}\boldsymbol{\gamma_2} + \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1}\left(\boldsymbol{Z_{1X}'Z_{1X}}\right)^{-1}\boldsymbol{Z_{1X}'Z_2}\boldsymbol{\gamma_2}. \tag{13}$$

where

$$\boldsymbol{Z_{1X}} = -\boldsymbol{M_X}\boldsymbol{Z_1} \equiv \left[\boldsymbol{X}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'} - \boldsymbol{I}\right]\boldsymbol{Z_1}.$$

is the $N \times K$ matrix of the (negative of) the residuals from auxiliary regressions of the columns of $\boldsymbol{Z_1}$ on $\boldsymbol{X}$, and $\boldsymbol{M_X}$ is the 'residual maker' matrix that is symmetric and idempotent.[7]

Proof. A proof is available in the appendix. □

[6] I assume the standard full-rank conditions - which rule out perfect collinearity among the regressors - that are necessary to ensure the existence of the OLS estimators.

[7] Since $\boldsymbol{M_X}$ is symmetric and idempotent, and is not the identity matrix, it is noninvertible.

3 OVB in the Short Regression

Let us begin with an investigation of the first question related to the problem of excluding relevant variables from an estimated model: what is the OVB of $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$? Inspecting the expression for the omitted variable bias in the short regression, we get the familiar formula for the OVB (Wooldridge, 2002; Angrist and Pischke, 2009; Greene, 2012), as

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}\right) - \boldsymbol{\beta} = \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'Z\gamma} = \hat{\boldsymbol{\delta}}\boldsymbol{\gamma}. \tag{14}$$

To interpret this formula for OVB, let $\hat{\boldsymbol{\delta}}^m$ denote the m-th column of the $J \times M$ matrix $\hat{\boldsymbol{\delta}}$. For $m = 1, 2, \ldots, M$, let $\boldsymbol{\delta}_m$ denote the coefficient vector in the auxiliary regression of the m-th omitted variable, $\boldsymbol{z}^m$, on the whole set of included regressors in the short regression, i.e.

$$\boldsymbol{z}^m = \boldsymbol{X}\boldsymbol{\delta}^m + \boldsymbol{v_m}, \tag{15}$$

so that the OLS estimator of $\boldsymbol{\delta}_m$ is given by

$$\hat{\boldsymbol{\delta}}^m = \left(\mathbf{X'X}\right)^{-1} \boldsymbol{X'z}^m$$

and stacking $\hat{\boldsymbol{\delta}}^m$ column-wise gives $\hat{\boldsymbol{\delta}}$. Hence,

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}\right) - \boldsymbol{\beta} = \hat{\boldsymbol{\delta}}^{\boldsymbol{1}}\gamma_1 + \hat{\boldsymbol{\delta}}^{\boldsymbol{2}}\gamma_2 + \cdots + \hat{\boldsymbol{\delta}}^{\boldsymbol{M-1}}\gamma_{M-1} + \hat{\boldsymbol{\delta}}^{\boldsymbol{M}}\gamma_M, \tag{16}$$

where γ_m is the m-th element of the M-vector $\boldsymbol{\gamma}$. This shows that the OVB of the j-th included regressor in the short regression, which is the j-th element of the vector in (16), is given by

$$OVB_j \equiv \mathbb{E}\left(\hat{\beta}_j\right) - \beta_j = \gamma_1\hat{\delta}_{1j} + \gamma_2\hat{\delta}_{2j} + \cdots + \gamma_M\hat{\delta}_{Mj} = \sum_{m=1}^{M} \gamma_m\hat{\delta}_{mj} \tag{17}$$

where $\hat{\delta}_{mj}$ is the j-th element of the coefficient vector $\hat{\boldsymbol{\delta}}_m$ in (15), with $j = 1, 2, \ldots, J$ and $m = 1, 2, \ldots, M$. Since this is a sum of M terms, each of which can be of any sign, we cannot determine the direction of OVB *in general*. But there are some configurations of partial effects that allow us to determine the direction of the OVB unambiguously just by knowing the sign of the partial effects. To see this, let $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}} = \begin{pmatrix} \hat{\delta}_{1j} & \hat{\delta}_{2j} & \cdots & \hat{\delta}_{Mj} \end{pmatrix}$ denote the j-th row of the $J \times M$ matrix, $\hat{\boldsymbol{\delta}}$. Since $\boldsymbol{\gamma}$ is a $M \times 1$ vector, the expression for the omitted variable bias in (17) is the inner product of the two vectors, $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$. Hence,

$$OVB_j = \hat{\boldsymbol{\delta}}^{\boldsymbol{j}}.\boldsymbol{\gamma} = \left\|\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}\right\|\left\|\boldsymbol{\gamma}\right\|\cos(\theta) \tag{18}$$

where $\left\|\boldsymbol{x}\right\|$ denotes the L_2-norm of the vector, $\boldsymbol{x}$, θ is the angle - measured in radians - between $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$, each considered as an element in $\mathbb{R}^M$, and $0 \leq \theta \leq \pi$.

Definition 1. *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors in $\mathbb{R}^M$ with θ denoting the angle between the two vectors.*

1. *We will say that* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *are similar in orientation if the angle between them is acute, i.e.,* $0 < \theta < \pi/2$.

2. *We will say that* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *are dissimilar in orientation if the angle between them is obtuse, i.e.* $\pi/2 < \theta < \pi$.

This definition is inspired by the notion of "cosine similarity" in the machine learning literature and can help us ascertain the direction of OVB.

Proposition 2. *The direction of omitted variable bias of the OLS estimator of the* j*-th included regressor in a misspecified model with many omitted variables is positive (negative) if the vectors* $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ *and* $\boldsymbol{\gamma}$ *are (dis)similar in orientation.*

Proof. The proof follows from an inspection of (18). □

3.1 Unambiguous Sign of OVB

The result in Proposition 2 shows that in general we will not be able to ascertain the sign of the OVB. Nonetheless, there are special configurations, as noted in Proposition 2, where we *will be* able to make unambiguous sign statements. The special configurations will depend on the signs of the vector $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$. What do these vectors signify? The m-th element of $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ gives the OLS estimator of the coefficient on the j-th included regressor in the auxiliary regression of the m-th omitted variable, $\boldsymbol{z}^m$, on the whole set of included regressors, $\boldsymbol{X}$. Hence, the vector $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ collects together the OLS estimators of the coefficients on the j-th included regressor in auxiliary regressions, successively, of the 1-st, 2-nd, $\cdots$, M-th omitted variable on the whole set of included regressors. On the other hand, the vector $\boldsymbol{\gamma}$ gives the partial effects of the omitted variables on the dependent variable in the population regression function. Hence, γ_m is the partial effect of the m-th omitted variable on the dependent variable in the population regression function in (11).

3.1.1 No Bias

We will be able to assert that there is no bias if the M-vectors $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$ are orthogonal or if one of them is a null vector. The two vectors are orthogonal when all omitted variables are orthogonal to all included regressors, and hence leaving out the omitted variables does not induce any correlation between the error term and the included regressors. That is why OLS is able to consistently estimate all the parameters. On the other hand, if either of the vectors is a null vector, it means that either the omitted variables are irrelevant or that the included regressors have no partial effect on the omitted variables (in the relevant auxiliary regression). That is why OLS is able to, once again, estimate the parameters consistently. Note that this is a multivariate generalization of the simplest case with one omitted variable: leaving out the omitted variable does not lead to bias if the omitted variable is orthogonal to all included regressors, or if its partial effect on the dependent variable is zero, or if the partial effects of included variables are zero.

3.1.2 Positive Bias

We will be able to unambiguously determine the sign of the OVB to be positive if both the M-vectors $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$ lie in the same orthant of $\mathbb{R}^M$. This is because, in this case, the two vectors will be similar in orientation according to Definition 1. If the two vectors lie in the same orthant, they will have the same sign for each of their corresponding elements, i.e.

$$sign(\hat{\delta}_{jm}) = sign(\gamma_m), \text{ for } m = 1, 2, \ldots, M.$$

In this case, we will be able to determine the sign of the OVB as positive irrespective of the magnitude of the elements of the two vectors because the inner product of the two vectors will be positive. How do we interpret this case? An unambiguously positive OVB will arise for the OLS estimate of k-th included regressor's coefficient in the mis-specified model in (10) if the partial effect of each omitted variable on the dependent variable has the *same sign* as the OLS estimator of the partial effect of the k-th included regressor on that omitted variable (in a auxiliary regression of the omitted variable on all the included regressors).

It is important to note that this is a multivariate generalization of the simplest case with one omitted variable. In that case, the OVB is positive if the sample correlation coefficient between the omitted and included variable is of the same sign as the partial effect of the omitted variable on the dependent variable in the population regression function in (2). In the multivariate case given in (11), we need to consider partial effects of omitted variables on the dependent variable in the population regression function, as in the univariate case. But, in place of the sample correlation coefficient between the omitted and the included regressor, we need to think in terms of OLS estimators of coefficients in the auxiliary regressions of all the omitted variables, in turn, on all the included regressors. And what is relevant is a comparison of the OLS estimator of the coefficient on the relevant included regressor in each of these auxiliary regressions with the partial effect of the corresponding omitted variable on the dependent variable. If these two are of the same sign, we will be able to assert that the OVB is positive.

3.1.3 Negative Bias

We will be able to unambiguously determine the sign of the OVB to be negative if the two M-vectors, $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$, lie in "opposite" orthants, by which I mean that the sign of each element in $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ is exactly opposite in sign of the corresponding element in $\boldsymbol{\gamma}$, i.e.

$$sign(\hat{\delta}_{jm}) = -sign(\gamma_m), \text{ for } m = 1, 2, \ldots, M.$$

This is because, in this case, the two vectors will be dissimilar in orientation, according to Definition 1. To see this, note that the inner product of the two vectors in this case will result in a negative scalar because each of the terms in the inner product is negative. Hence, the angle between the two vectors will be between $\pi/2$ and π. How should this case be interpreted? An unambiguously negative OVB will arise for the OLS estimate of j-th included regressor's coefficient in the misspecified model in (10) if the partial effect of each omitted variable on the dependent variable has the opposite sign of the partial effect of the

j-th included regressor on that omitted variable (in a auxiliary regression of the omitted variable on all the included regressors).

Note, again, that this is a multivariate generalization of the simplest case with one omitted variable. In that case, the OVB is negative if the sample correlation coefficient between the omitted and included variable is of the opposite sign of the partial effect of the omitted variable on the dependent variable in the population regression function in (2). In the multivariate case in (11), we need to consider partial effects of omitted variables on the dependent variable in the population regression function, as in the univariate case. But, in place of the sample correlation coefficient between the omitted and the included regressor, we need to think in terms of the auxiliary regressions of all the omitted variables, in turn, on all the included regressors. And what is relevant is a comparison of the OLS estimator of the coefficient on the relevant included regressor in each of these auxiliary regressions with the partial effect of the corresponding omitted variable on the dependent variable. If these two are of exactly the opposite sign, we will be able to assert that the OVB is negative.

3.2 Summary

Suppose the population regression function of y includes $x_1, x_2, \ldots, x_J$, and $z_1, z_2, \ldots, z_M$. Consider a scenario where the researcher is able to include only the J regressors, $x_1, x_2, \ldots, x_J$, in the estimated model. To determine the direction of omitted variable bias of the OLS estimator of any included regressor, x_j, in this general setting, a researcher can do the following thought experiment.

1. Form an M-vector, $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$, where the m-th element is the OLS estimator of the partial effect of x_j on z_m in an auxiliary regression of z_m on all the included regressors $x_1, x_2, \ldots, x_J$.
2. Form an M-vector, $\boldsymbol{\gamma}$, where the m-th element is the partial effect of z_m on the dependent variable in the population regression function (the correctly specified model).
3. The omitted variable bias of the OLS estimator of x_j is the inner product of the two vectors, $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$. Hence, we have the following:
   (a) If $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$ lie in the same orthant, the bias is postive.
   (b) If $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$ lie in opposite orthants, the bias is negative.
   (c) If $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$ are orthogonal, the bias is zero.
   (d) If $\hat{\boldsymbol{\delta}}^{\boldsymbol{j}}$ and $\boldsymbol{\gamma}$ are neither orthogonal, nor lie in the same orthant or in opposite orthants, then the direction of bias cannot be determined unambiguously on the basis of the signs of partial effects only.

3.3 Example: Returns to Education

Let me illustrate the argument outlined above using the canonical example of the returns to education. Let the population regression function of interest be a wage regression

$$\log wage = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 exper + \gamma_1 ability + \gamma_2 motiv + u$$

where $\log wage$ is regressed on *educ* (years of schooling), *age* (age in years), *exper* (years in the labour force), *ability* (intrinsic ability) and *motiv* (motivation). Suppose the last two variables are omitted due to lack of data, so that the following model is estimated with OLS

$$\log wage = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 exper + v.$$

In this case, the bias of the OLS estimator of the return to education will be

$$\mathbb{E}\hat{\beta}_1 - \beta_1 = \gamma_1 \hat{\delta}_{11} + \gamma_2 \hat{\delta}_{21}$$

where $\hat{\delta}_{11}$ is the coefficient on *educ* in the auxiliary regression of *ability* on $educ, age, exper$

$$ability = \hat{\delta}_{10} + \hat{\delta}_{11} educ + \hat{\delta}_{12} age + \hat{\delta}_{13} exper$$

and $\hat{\delta}_{21}$ is the coefficient on *educ* in the auxiliary regression of *motiv* on $educ, age, exper$

$$motiv = \hat{\delta}_{20} + \hat{\delta}_{21} educ + \hat{\delta}_{22} age + \hat{\delta}_{23} exper.$$

Since the bias of the OLS estimator of the return to education is given by $\gamma_1 \hat{\delta}_{11} + \gamma_2 \hat{\delta}_{21}$, which is the inner product of the two vectors (γ_1, γ_2) and $(\hat{\delta}_{11}, \hat{\delta}_{21})$, we can make the following claims:

- The bias is positive if the vectors (γ_1, γ_2) and $(\hat{\delta}_{11}, \hat{\delta}_{21})$ lie in the same quadrant, i.e. $sign(\gamma_1) = sign(\hat{\delta}_{11})$ and $sign(\gamma_2) = sign(\hat{\delta}_{21})$.
- The bias is negative if the vectors (γ_1, γ_2) and $(\hat{\delta}_{11}, \hat{\delta}_{21})$ lie in opposite quadrants, i.e. $sign(\gamma_1) = -sign(\hat{\delta}_{11})$ and $sign(\gamma_2) = -sign(\hat{\delta}_{21})$.
- The bias is zero if the vectors (γ_1, γ_2) and $(\hat{\delta}_{11}, \hat{\delta}_{21})$ are orthogonal or if one of them is a null vector.

4 Comparison of OVB in Short and Long Regressions

Let us now turn to the second question about the bias of OLS estimators that arise from excluding relevant variables from an estimated model: can we compare the omitted variable bias of an included regressor between the short and long regressions, when both are estimated with OLS, i.e. can we compare the OVB of $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$, which is the the OLS estimator of $\tilde{\boldsymbol{\beta}}$ in (10), and $\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}$, which is the OLS estimator of $\tilde{\tilde{\boldsymbol{\beta}}}$ in (11)?

4.1 Expression for Difference in Bias

Proposition 1 allows us find an expression for the difference in bias because we have

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}\right) - \boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{\gamma_1} + \boldsymbol{C}\boldsymbol{\gamma_2} \tag{19}$$

and

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}\right) - \boldsymbol{\beta} = \boldsymbol{C}\boldsymbol{\gamma_2} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{\gamma_2} \tag{20}$$

where

$$\boldsymbol{A} = \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'Z_1} \tag{21}$$

is the $J \times K$ matrix of the column-wise stacked OLS estimators of the coefficient vectors for auxiliary regressions of the columns of $\boldsymbol{Z_1}$ on $\boldsymbol{X}$, and

$$\boldsymbol{B} = \left(\boldsymbol{Z'_{1X}Z_{1X}}\right)^{-1} \boldsymbol{Z'_{1X}Z_2} \tag{22}$$

is the $K \times L$ matrix of the column-wise stacked OLS estimators of the coefficient vectors for auxiliary regressions of the columns of $\boldsymbol{Z_2}$ on $\boldsymbol{Z_{1X}}$, where

$$\boldsymbol{Z_{1X}} = -\boldsymbol{M_X Z_1} \equiv \left[\boldsymbol{X}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'} - \boldsymbol{I}\right]\boldsymbol{Z_1} \tag{23}$$

is the $N \times K$ matrix of the column-wise stacked residuals from auxiliary regressions of the columns of $\boldsymbol{Z_1}$ on $\boldsymbol{X}$, and

$$\boldsymbol{C} = \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'Z_2} \tag{24}$$

is the $J \times L$ matrix of the column-wise stacked OLS estimators of the coefficient vectors for auxiliary regression of the columns of $\boldsymbol{Z_2}$ on $\boldsymbol{X}$.

Let $\boldsymbol{\tau_S} = \mathbb{E}\left(\boldsymbol{\hat{\beta}_S}\right) - \boldsymbol{\beta}$ be the bias of the OLS estimator in the short regression in (10), and $\boldsymbol{\tau_L} = \mathbb{E}\left(\boldsymbol{\hat{\beta}_L}\right) - \boldsymbol{\beta}$ be the bias in the long regression in (11). Using (19) and (20), we see that the difference in bias is given by

$$\boldsymbol{\tau_S} = \boldsymbol{\tau_L} + \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1}\left[\boldsymbol{\gamma_1} - \left(\boldsymbol{Z'_{1X}Z_{1X}}\right)^{-1}\boldsymbol{Z'_{1X}Z_2\gamma_2}\right]. \tag{25}$$

4.2 Sign of the Difference in Bias

From the expression in (25), we see that it is not possible, in general, to make any claims about bias reduction while comparing the short and long regressions.[8] This rather surprising result has been highlighted in Clarke (2005) and Luca et al. (2018). It arises from the fact that both the short and the long regressions remain mis-specified, i.e. both suffer from omitted variable problems. If the long regression included all the omitted variables - the regular textbook case - then the bias would be unambiguously reduced in comparison to the short regression because the bias in the long regression would become zero. We can make this point by looking at individual elements of $\boldsymbol{\beta}$ too.

Since the columns of $\boldsymbol{A\gamma_1}$ is a linear combination of the columns of $\boldsymbol{A}$, and the columns of $\boldsymbol{C\gamma_2}$ is a linear combination of the columns of $\boldsymbol{C}$, we have

$$\mathbb{E}\left(\boldsymbol{\hat{\beta}_S}\right) - \boldsymbol{\beta} = \left(\boldsymbol{a^1}\gamma_{11} + \cdots + \boldsymbol{a^K}\gamma_{1K}\right) + \left(\boldsymbol{c^1}\gamma_{21} + \cdots + \boldsymbol{c^L}\gamma_{2L}\right) \tag{26}$$

[8] If $\boldsymbol{X}$ and $\boldsymbol{Z_1}$ are orthogonal, then the bias from the short and long regressions are the same. This is intuitively clear: the part of the omitted variables set that has been included in the long regression is orthogonal to the regressors that were part of the short regression. Hence, even when they are omitted, that has no effect on the bias of the included regressors in the short regression.

where $\boldsymbol{a^k}$ and $\boldsymbol{c^l}$ are the k-th and l-th columns of $J \times K$ matrix $\boldsymbol{A}$ and the $J \times L$ matrix $\boldsymbol{C}$, and γ_{1k} and γ_{2l} are elements of the K-vector $\boldsymbol{\gamma_1}$ and the L-vector $\boldsymbol{\gamma_2}$, respectively. Let τ_{Sj} denote the OVB of the j-th included regressor, $\boldsymbol{X}$, in the short regression. Hence

$$\tau_{Sj} = \boldsymbol{a_j.\gamma_1} + \boldsymbol{c_j.\gamma_2} \tag{27}$$

where $\boldsymbol{a_j}$ denotes the j-th row of $\boldsymbol{A}$, $\boldsymbol{a_j.\gamma_1}$ denotes the inner product of the two K-vectors, $\boldsymbol{a_j}$ and $\boldsymbol{\gamma_1}$, $\boldsymbol{c_j}$ denotes the j-th row of $\boldsymbol{C}$, and $\boldsymbol{c_j.\gamma_2}$ denotes the inner product of the two L-vectors, $\boldsymbol{c_j}$ and $\boldsymbol{\gamma_2}$.

Using the same facts about matrix multiplication, we get

$$\mathbb{E}\left(\boldsymbol{\hat{\beta}_L}\right) - \boldsymbol{\beta} = \left(\boldsymbol{c^1}\gamma_{21} + \cdots + \boldsymbol{c^L}\gamma_{2L}\right) + \left(\boldsymbol{Ab^1}\gamma_{21} + \cdots + \boldsymbol{Ab^L}\gamma_{2L}\right), \tag{28}$$

where $\boldsymbol{b^k}$ denotes the k-th column of $\boldsymbol{B}$, with $k = 1, 2 \ldots, K$. Let τ_{Lj} denote the OVB of the j-th included regressor, $\boldsymbol{X}$, in the long regression. Hence

$$\tau_{Lj} = \boldsymbol{c_j.\gamma_2} + \left(\boldsymbol{a_j.b^1}\gamma_{21} + \cdots + \boldsymbol{a_j.b^L}\gamma_{2L}\right). \tag{29}$$

We are interested in finding conditions when absolute value of the bias is reduced as the researcher moves from the short to the long regression, i.e. $\left|\tau_{Lj}\right| < \left|\tau_{Sj}\right|$. Comparing the expressions in (27) and (29), we see that, in general, the bias from the long regression will not be smaller than the one in the short regression.

Is it possible to use arguments based only on the signs of partial effects, as we did in the previous section, to identify scenarios when we can make unambiguous claims about bias reduction? The answer is in the negative. To develop this argument, we will need the following L-vector

$$\boldsymbol{ab} = \begin{pmatrix} \boldsymbol{a_j.b^1} & \boldsymbol{a_j.b^2} \ldots & \boldsymbol{a_j.b^{L-1}} & \boldsymbol{a_j.b^L} \end{pmatrix}, \tag{30}$$

where each element of the above vector is an inner product of $\boldsymbol{a_j}$ and a column of $\boldsymbol{B}$.

An investigation of the conditions to ensure $\left|\tau_{Lj}\right| < \left|\tau_{Sj}\right|$ means that we need to consider the following four cases: (a) $\tau_{Lj} > 0$ and $\tau_{Sj} > 0$; (b) $\tau_{Lj} < 0$ and $\tau_{Sj} > 0$; (c) $\tau_{Lj} > 0$ and $\tau_{Sj} < 0$; and (d) $\tau_{Lj} < 0$ and $\tau_{Sj} < 0$.

4.3 Case 1

Consider the case when both biases are positive, i.e. $\tau_{Lj} > 0$ and $\tau_{Sj} > 0$. When $\tau_{Lj} > 0$ and $\tau_{Sj} > 0$, the absolute value of the bias will be reduced if $\tau_{Lj} - \tau_{Sj} < 0$.

What are sufficient sign restrictions that will ensure that $\tau_{Lj} > 0$ and $\tau_{Sj} > 0$? From the expressions in (27) and (29), we see that $\tau_{Lj} > 0$ and $\tau_{Sj} > 0$ if the following conditions hold: (a) the vectors $\boldsymbol{a_j}$ and $\boldsymbol{\gamma_1}$ lie in the same orthant of $\mathbb{R}^K$, i.e. $sign(a_{jk}) = sign(\gamma_{1k})$, for $k = 1, 2, \ldots, K$; (b) the vectors $\boldsymbol{c_j}$ and $\boldsymbol{\gamma_2}$ lie in the same orthant of $\mathbb{R}^L$, i.e. $sign(a_{jl}) = sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$; and (c) $\boldsymbol{ab}$ and $\boldsymbol{\gamma_2}$ lie in the same orthant of $\mathbb{R}^L$, i.e. $sign(\boldsymbol{a_j.b^l}) = -sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$.

What are sufficient sign restrictions that will ensure that $\tau_{Lj} - \tau_{Sj} < 0$? Using the expressions from (27) and (29), we see that the absolute value of the bias will be reduced if

$$\boldsymbol{a_j}.\boldsymbol{\gamma_1} - \left(\boldsymbol{a_j}.\boldsymbol{b^1}\gamma_{21} + \cdots + \boldsymbol{a_j}.\boldsymbol{b^L}\gamma_{2L}\right) > 0.$$

This inequality will be satisfied if the following two conditions hold: (d) $\boldsymbol{ab}$ and $\boldsymbol{\gamma_2}$ lie in opposite orthants of $\mathbb{R}^L$, i.e. $sign(\boldsymbol{a_j}.\boldsymbol{b^l}) = sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$, and (e) the vectors $\boldsymbol{a_j}$ and $\boldsymbol{\gamma_1}$ lie in the same orthant of $\mathbb{R}^K$, i.e. $sign(a_{jk}) = sign(\gamma_{1k})$, for $k = 1, 2, \ldots, K$.

We see that conditions (c) and (d) contradict each other. Hence, it is not possible to derive sufficient conditions for bias reduction when both biases are positive using only signs of partial effects.

4.4 Case 2

The second case to consider is when $\tau_{Lj} < 0$ and $\tau_{Sj} > 0$. From the expression in (29), we see that $\tau_{Lj} < 0$ if the following conditions hold: (a) the vectors $\boldsymbol{c_j}$ and $\boldsymbol{\gamma_2}$ lie in opposite orthants of $\mathbb{R}^L$, i.e. $sign(a_{jl}) = -sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$; and (c) $\boldsymbol{ab}$ and $\boldsymbol{\gamma_2}$ lie in the same orthants of $\mathbb{R}^L$, i.e. $sign(\boldsymbol{a_j}.\boldsymbol{b^l}) = sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$. From the expressions in (27), we see that $\tau_{Sj} > 0$ if the following conditions hold: (c) the vectors $\boldsymbol{a_j}$ and $\boldsymbol{\gamma_1}$ lie in the same orthant of $\mathbb{R}^K$, i.e. $sign(a_{jk}) = sign(\gamma_{1k})$, for $k = 1, 2, \ldots, K$; (d) the vectors $\boldsymbol{c_j}$ and $\boldsymbol{\gamma_2}$ lie in the same orthant of $\mathbb{R}^L$, i.e. $sign(a_{jl}) = sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$. Since the conditions (a) and (c) contradict each other, we cannot generate sufficient conditions for the biases to be of opposite signs only on the basis of the signs of partial effects.

4.5 Case 3

The third case to consider is when $\tau_{Lj} > 0$ and $\tau_{Sj} < 0$. Just like in case 2, it is not possible to derive sufficient conditions in terms of sign restrictions to ensure that $\tau_{Lj} > 0$ and $\tau_{Sj} < 0$.

4.6 Case 4

The final case to consider is when both biases are negative, i.e. $\tau_{Lj} < 0$ and $\tau_{Sj} < 0$. In this case, the absolute value of the bias will be reduced if $\tau_{Sj} - \tau_{Lj} < 0$.

What are sufficient sign restrictions that will ensure that $\tau_{Lj} < 0$ and $\tau_{Sj} < 0$? From the expressions in (27) and (29), we see that $\tau_{Lj} < 0$ and $\tau_{Sj} < 0$ if the following conditions hold: (a) the vectors $\boldsymbol{a_j}$ and $\boldsymbol{\gamma_1}$ lie in opposite orthants of $\mathbb{R}^K$, i.e. $sign(a_{jk}) = -sign(\gamma_{1k})$, for $k = 1, 2, \ldots, K$; (b) the vectors $\boldsymbol{c_j}$ and $\boldsymbol{\gamma_2}$ lie in opposite orthants of $\mathbb{R}^L$, i.e. $sign(a_{jl}) = -sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$; and (c) $\boldsymbol{ab}$ and $\boldsymbol{\gamma_2}$ lie in opposite orthants of $\mathbb{R}^L$, i.e. $sign(\boldsymbol{a_j}.\boldsymbol{b^l}) = sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$.

What are sufficient sign restrictions that will ensure that $\tau_{Sj} - \tau_{Lj} < 0$? Using the expressions from (27) and (29), we that the condition reduces to

$$\boldsymbol{a_j}.\boldsymbol{\gamma_1} - \left(\boldsymbol{a_j}.\boldsymbol{b^1}\gamma_{21} + \cdots + \boldsymbol{a_j}.\boldsymbol{b^L}\gamma_{2L}\right) < 0.$$

This will be ensured if the following two conditions hold: (d) $\boldsymbol{ab}$ and $\boldsymbol{\gamma_2}$ lie in the same orthant of $\mathbb{R}^L$, i.e. $sign(\boldsymbol{a_j}.\boldsymbol{b^l}) = -sign(\gamma_{2l})$, for $l = 1, 2, \ldots, L$, and (e) the vectors $\boldsymbol{a_j}$ and $\boldsymbol{\gamma_1}$ lie in opposite orthants of $\mathbb{R}^K$, i.e. $sign(a_{jk}) = -sign(\gamma_{1k})$, for $k = 1, 2, \ldots, K$.

We see, again, that conditions (c) and (d) contradict each other. Hence, it is not possible to derive sufficient conditions for bias reduction when both biases are positive only using signs of partial effects.

4.7 Example: Returns to Education

Let me illustrate the argument outlined in this section by returning to the example of the returns to education. Let the population regression function of interest be the wage regression,

$$\log wage = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 exper + \gamma_1 ability + \gamma_2 motiv + u,$$

where $\log wage$ is regressed on $educ$ (years of schooling), age (age in years), $exper$ (years in the labour force), $ability$ (intrinsic ability) and $motiv$ (motivation). Suppose the short (estimated) regression model excludes both $ability$ and $motiv$, so that it is given by

$$\log wage = \alpha_0 + \alpha_1 educ + \alpha_2 age + \alpha_3 exper + v_S$$

and the long (estimated) regression model includes $ability$, so that it is given by

$$\log wage = \delta_0 + \delta_1 educ + \delta_2 age + \delta_3 exper + \gamma_1 ability + v_L,$$

and both these models are estimated by OLS.

In this case, we can express the bias of the OLS estimators from both regressions in terms of sample variances and sample correlation coefficients. Thus, the bias of the returns to schooling in the short regression model will be

$$\mathbb{E}\left(\hat{\alpha}_1\right) - \beta_1 = \left(\frac{\hat{\sigma}_{ability}}{\hat{\sigma}_{edu}}\right) \hat{\rho}_{educ,ability} \gamma_1 + \left(\frac{\hat{\sigma}_{motiv}}{\hat{\sigma}_{educ}}\right) \hat{\rho}_{educ,motiv} \gamma_2,$$

and the corresponding bias in the long regression will be

$$\begin{aligned} \mathbb{E}\left(\hat{\delta}_1\right) - \beta_1 = & \left(\frac{\hat{\sigma}_{motiv}}{\hat{\sigma}_{educ}}\right) \hat{\rho}_{educ,motiv} \times \gamma_2 \\ & + \left(\frac{\hat{\sigma}_{ability}}{\hat{\sigma}_{educ}}\right) \left(\frac{\hat{\sigma}_{motiv}}{\hat{\sigma}_{\hat{v}}}\right) \hat{\rho}_{educ,ability} \times \hat{\rho}_{\hat{v},motiv} \times \gamma_2 \end{aligned}$$

where $\hat{v}$ represent the (negative of) the residuals from an auxiliary regression of $ability$ on $educ$. Hence, the difference in the bias is given by

$$\begin{aligned} \mathbb{E}\left(\hat{\delta}_1\right) - \beta_1 = & \left[\mathbb{E}\left(\hat{\alpha}_1\right) - \beta_1\right] + \left(\frac{\hat{\sigma}_{ability}}{\hat{\sigma}_{edu}}\right) \hat{\rho}_{educ,ability} \gamma_1 \\ & - \left(\frac{\hat{\sigma}_{ability}}{\hat{\sigma}_{educ}}\right) \left(\frac{\hat{\sigma}_{motiv}}{\hat{\sigma}_{\hat{v}}}\right) \hat{\rho}_{educ,ability} \times \hat{\rho}_{\hat{v},motiv} \times \gamma_2 \end{aligned}$$

so that, in general, we cannot determine whether including *ability* in the long regression model reduces the bias of the OLS estimator of the returns to schooling - as long as *motiv* remains omitted. The difficulty of making any claims about bias reduction comes from two possible correlations: (a) between education and ability, and (b) between motivation and those factors that determine ability over and above education.

5 A Doubly Misspecified Model

Two specification errors are common and ubiquitous in applied econometric research - the problems of excluding relevant variables and of including irrelevant variables. In textbook treatments of these problems, it is common to discuss the two separately (Wooldridge, 2002; Greene, 2012). This separate treatment is motivated by the well known result that, whereas exclusion of relevant variables leads to biased OLS estimators, inclusion of irrelevant variables has no such bias implications. In this section, I demonstrate that this conventional understanding is not true in a general setting where an estimated model excludes relevant and included irrelevant variables.

5.1 Expression for Bias

To fix ideas, let the population regression function be given by

$$\boldsymbol{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{u}, \tag{31}$$

and the estimated model be given by

$$\boldsymbol{y} = \mathrm{X}\boldsymbol{\alpha} + \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{v} \tag{32}$$

where we have

$$\mathbb{E}\left(\boldsymbol{u}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{W}\right) = \boldsymbol{0}, \tag{33}$$

with $\boldsymbol{X}$ a $N\times J$ matrix, $\boldsymbol{Z}$ a $N\times K$ matrix, $\boldsymbol{W}$ a $N\times L$ matrix, and $\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha},\boldsymbol{\delta}$, corresponding and conformable vectors of coefficients. Note that the estimated model in (32) is doubly misspecified: it has excluded the relevant variables, $\boldsymbol{Z}$, and it has included the irrelevant variables, $\boldsymbol{W}$.

Proposition 3. *Let $\hat{\boldsymbol{\alpha}}$ denote the OLS estimator of $\boldsymbol{\alpha}$ in the doubly misspecified model (32). Then $\hat{\boldsymbol{\alpha}}$ is a biased estimator of $\boldsymbol{\beta}$ in (31) and, conditional on $\boldsymbol{X},\boldsymbol{Z},\boldsymbol{W}$, the bias is given by*

$$\mathbb{E}\left(\hat{\boldsymbol{\alpha}}\right) - \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{\gamma} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}\left(\boldsymbol{W}_{\boldsymbol{X}}'\boldsymbol{W}_{\boldsymbol{X}}\right)^{-1}\boldsymbol{W}_{\boldsymbol{X}}'\boldsymbol{Z}\boldsymbol{\gamma}. \tag{34}$$

where

$$\boldsymbol{W}_{\boldsymbol{X}} = -\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{W} \equiv \left[\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-\mathbf{1}}\boldsymbol{X}' - \boldsymbol{I}\right]\boldsymbol{W}$$

is the (negative of) the matrix of residuals that comes from the auxiliary regressions of the columns of $\boldsymbol{W}$ on the full set of included regressors $\boldsymbol{X}$, and $\boldsymbol{M}_{\boldsymbol{X}}$ is the 'residual maker' matrix that is symmetric and idempotent.

Proof. A proof is given in the appendix. □

The expression for bias in (34), shows that the bias of the OLS estimator of the doubly misspecified model - exclusion of relevant variables and inclusion of irrelevant variables - is the sum of two terms. The first term represents the *direct effect* on bias caused by omitting the relevant variables, $\boldsymbol{Z}$, and the second term represents the *indirect effect* on bias caused by omitting the relevant variable, which works through its interaction with the irrelevant variables, $\boldsymbol{W}$.

5.2 Two Special Cases

Two special cases immediately fall out of the general result in (34). The first special case is one where the estimated model in (32) does not include any irrelevant variables, even though it might have excluded relevant variables. We can capture this with the restriction $\boldsymbol{M} = \boldsymbol{0}$ in (32). Hence, the expression for bias, in this special case, can be derived by plugging in $\boldsymbol{M} = \boldsymbol{0}$ in (34):

$$\mathbb{E}\left(\hat{\boldsymbol{\alpha}}\right) - \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{\gamma}.$$

This is the standard expression for omitted variable bias, as we have seen in section 2.

The second special case is one where the estimated model does not exclude any of the relevant variables, even though it might include irrelevant variables. We can capture this special case with the restriction $\boldsymbol{\gamma} = \boldsymbol{0}$ because this implies that that estimated model in (32) has not excluded any of the variables of the population regression function in (31). Hence, the expression for the bias in this case can be derived by plugging in $\boldsymbol{\gamma} = \boldsymbol{0}$ in (34):

$$\mathbb{E}\left(\hat{\boldsymbol{\alpha}}\right) - \boldsymbol{\beta} = \boldsymbol{0}.$$

Here we get the familiar textbook result: inclusion of irrelevant variables does not give biased OLS estimators (Greene, 2012, pp. 58).

5.3 Interpretation of Bias

Let us now turn to interpreting the expression for bias appearing in (34). The first component is the familiar OVB. It is the direct effect of the omitted variables, $\boldsymbol{Z}$, on the bias of the OLS estimators of the coefficients on the included variables, $\boldsymbol{X}$. As long as the two sets of variables are (partially) correlated, the first component in (34) will be non-zero. Intuitively, OLS will attribute some of the effect of the components of $\boldsymbol{Z}$ on the components of $\boldsymbol{Z}$. The second component is the indirect effect of the omitted variables, $\boldsymbol{Z}$, on the bias of the OLS estimators. This indirect effect works through the channel of the irrelevant variables, $\boldsymbol{W}$ - that is why the inclusion of irrelevant variables has an impact on bias.

To see this in more concrete terms, note that the second term in (32) can be written as

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W} \left(\boldsymbol{W}_{\boldsymbol{X}}'\boldsymbol{W}_{\boldsymbol{X}}\right)^{-1} \boldsymbol{W}_{\boldsymbol{X}}'\boldsymbol{Z}\boldsymbol{\gamma} = \boldsymbol{D}\boldsymbol{E}\boldsymbol{\gamma}$$

where each column of the $J \times L$ matrix

$$\boldsymbol{D} = \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'W} \tag{35}$$

is the OLS estimator of the coefficient vector from an auxiliary regression of the corresponding column of $\boldsymbol{W}$ on the whole set of included regressors $\boldsymbol{X}$, and each column of the $L \times K$ matrix

$$\boldsymbol{E} = \left(\boldsymbol{W'_X W_X}\right)^{-1} \boldsymbol{W'_X Z} \tag{36}$$

is the OLS estimator of the coefficient vector from an auxiliary regression of the corresponding column of $\boldsymbol{Z}$ on $\boldsymbol{W_X}$, where the latter are residuals that come from previous auxiliary regressions of the columns of $\boldsymbol{W}$ on the full set of included regressors $\boldsymbol{X}$.

To get an intuitive grasp of the terms, $\boldsymbol{D}$ and $\boldsymbol{E}$, that comprise the indirect effect, consider the auxiliary regression of the irrelevant variables, $\boldsymbol{W}$, on the included regressors, $\boldsymbol{X}$. This regression decomposes $\boldsymbol{W}$ into two orthogonal components,

$$\boldsymbol{W} = \boldsymbol{P_X W} + \boldsymbol{M_X W},$$

where $\boldsymbol{P_X = X \left(X'X\right)^{-1} X'}$ and $\boldsymbol{M_X = I - M_X}$, so that the first component, $\boldsymbol{P_X W}$, is the projection of $\boldsymbol{W}$ into the column space of $\boldsymbol{X}$, and the second component, $\boldsymbol{M_X W}$, is orthogonal to the first component. Thus, the first component is that part of $\boldsymbol{W}$ which can be explained by $\boldsymbol{X}$, and the second component is that part of $\boldsymbol{W}$ which cannot be explained by $\boldsymbol{X}$, i.e. it represents factors other than $\boldsymbol{X}$ that determine $\boldsymbol{W}$.

From (34), we see that the indirect effect - the second term on the RHS of (34) - is the product of three effects:

1. the projection of $\boldsymbol{W}$ into the column space of $\boldsymbol{X}$,

   $$\boldsymbol{P_X W} = \boldsymbol{XD},$$

   which will be non-zero as long as $\boldsymbol{X}$ has some explanatory power for $\boldsymbol{W}$;

2. the partial correlations between $\boldsymbol{M_X W}$ and the omitted variables, $\boldsymbol{Z}$,

   $$\left(\boldsymbol{W'_X W_X}\right)^{-1} \boldsymbol{W'_X Z} = \boldsymbol{E}$$

   which will be non-zero as long as factors other than $\boldsymbol{X}$ that determine $\boldsymbol{W}$ are also correlated with the omitted variables, $\boldsymbol{Z}$;

3. the partial effect of the omitted variables, $\boldsymbol{Z}$, on the dependent variable in the population regression function, $\boldsymbol{\gamma}$.

If the omitted variables are relevant, then $\boldsymbol{\gamma} \neq \boldsymbol{0}$. In this case, the indirect effect of these omitted variables on the dependent variable show up in terms of non-zero partial correlations with $\boldsymbol{M_X W}$ (factors other than $\boldsymbol{X}$ that determine $\boldsymbol{W}$, and the omitted variables, $\boldsymbol{Z}$), which is then relayed to the dependent variable through its product with the projection of $\boldsymbol{W}$ into the column space of $\boldsymbol{X}$. If these latter effects are non-zero, this indirect effect of the omitted variables will be ascribed to the included regressors, $\boldsymbol{X}$, and will thereby impact the overall bias of OLS estimators.

As long as the omitted variables are actually relevant, so that $\boldsymbol{\gamma}$ is nonzero, there are only two situations in which inclusion of irrelevant variables will not have any impact on bias.[9] The first case will arise when the partial correlations between the included regressors, $\boldsymbol{X}$, and the irrelevant variables, $\boldsymbol{W}$, are zero. In this case, the columns of $\boldsymbol{W}$ will be orthogonal to the column space of $\boldsymbol{X}$, so that $\boldsymbol{X'W} = \mathbf{0}$. Thus, the second term in (34) will be zero. The second case will arise when the partial correlation between the omitted variables, $\boldsymbol{Z}$, and $\boldsymbol{W_X}$ (the part of the irrelevant variables that is not explained by the included regressors) are zero. In this case the columns of $\boldsymbol{Z}$ will be orthogonal to the column space of $\boldsymbol{W_X}$, so that $\boldsymbol{W_X'Z} = \mathbf{0}$. Thus, the second term in (34) will be zero.

5.4 Example: Returns to Education

Let me illustrate the argument outlined above using a simple version of the canonical wage regression. Let the population regression function of interest be a wage regression

$$\log wage = \beta_0 + \beta_1 educ + \gamma_1 ability + u$$

where $\log wage$ is explained by *educ* (years of schooling) and *ability* (intrinsic ability). Suppose the estimated model leaves out *ability* and includes *coding* (proficiency in writing computer codes)

$$\log wage = \beta_0 + \beta_1 educ + \gamma_2 coding + v.$$

In this case, the estimated in doubly misspecified: it leaves out a relevant variable, *ability*, and includes an irrelevant variable, *coding*.

Let us think of the following auxiliary regression:

$$coding = a_0 + a_1 educ + \varepsilon_1.$$

It is plausible to argue that the OLS estimator of the coefficient on *educ* in the auxiliary regression, $\hat{a}_1$, will be non-zero. This is because those who have more years of schooling, will, on average be more proficient in coding - just because they might have been exposed to computer programming. Now let us think of some factors which determine the proficiency in coding that have been left out of the above auxiliary regression. The aptitude for logical reasoning might be one such omitted factor. It is plausible to argue that the aptitude for logical reasoning - which leads to better coding skills - is correlated with *ability*, the variable that has been omitted from the estimated model. If this were to be the case then the overall

[9]Note that $\boldsymbol{W_X}$, the part of $\boldsymbol{W}$ that is orthogonal to the column space of $\boldsymbol{X}$, cannot be $\mathbf{0}$ as long as we rule out perfect collinearity between $\boldsymbol{X}$ and $\boldsymbol{W}$ - which we do, to ensure the existence of the OLS estimator. This is because the columns of $\boldsymbol{W_X}$ are orthogonal to the column space of $\boldsymbol{X}$. If $\boldsymbol{W}$ is not perfectly linearly related to $\boldsymbol{X}$, then it's columns do not lie in the column space of $\boldsymbol{X}$. Hence, $\boldsymbol{W_X} \neq \mathbf{0}$.

bias of the OLS estimator for the returns to schooling would be impacted by the inclusion of the irrelevant variable, *coding*, in the estimated model. Moreover,

$$\mathbb{E}\left(\hat{\beta}_1\right) - \beta_1 = \left(\frac{\hat{\sigma}_{abil}}{\hat{\sigma}_{educ}}\right)\hat{\rho}_{educ,abil} \times \gamma_1 + \hat{\rho}_{educ,coding}\left(\frac{\hat{\sigma}_{coding}}{\hat{\sigma}_{educ}}\right)\hat{\rho}_{abil,coding^{educ}}\left(\frac{\hat{\sigma}_{abil}}{\hat{\sigma}_{coding^{educ}}}\right) \times \gamma_1,$$

where $\hat{\sigma}_{educ}, \hat{\sigma}_{abil}, \hat{\sigma}_{coding}$ are the sample standard deviations of years of schooling, ability and coding proficiency, $coding^{educ}$ denotes the residuals that come from the regression of *coding* on *educ*, $\hat{\rho}_{abil,coding^{educ}}$ denotes the sample correlation coefficient of *abil* and $coding^{educ}$. The first term on the RHS of the above expression for bias of the OLS estimator of the coefficient on *educ* is the direct effect of omitting *ability* on the bias; the second term is the indirect effect of omitting *ability*. This latter effect works through the correlation of *ability* with those factors that determine *coding*, over and above *educ*, which is then relayed to $\log(wages)$ through its product with the correlation between *educ* and *coding*.

Hence, we can make the following claims.

- If the correlation between years of schooling and coding proficiency, in the sample, is zero, then the inclusion of the irrelevant variable, *coding*, in the estimated model will not have any impact on the bias of the OLS estimator of the returns to schooling.
- If those factors that determine the proficiency of coding, over and above years of schooling, are not correlated, in the sample, with *ability*, then the inclusion of the irrelevant variable, *coding*, in the estimated model will not have any impact on the bias of the OLS estimator of the returns to schooling.
- If neither of the above two conditions hold, which seems likely, then the inclusion of the irrelevant variable, *coding*, in the estimated model will have an impact on the bias of the OLS estimator of the returns to schooling. In general, it is not possible to determine the sign of this effect.

5.5 Direction of the Effect on Bias

While we cannot determine the direction of impact of irrelevant variables in the general case, there are some special cases where we might be able to use sign restrictions to pin it down using the methodology used in previous sections. To see this, let $\boldsymbol{d_j}$ denote the j-th row of the matrix $\boldsymbol{D}$ defined in (35), and let $\boldsymbol{e^k}$ denote the k-th column of the matrix $\boldsymbol{E}$ defined in (36). Then, the j-th element of the second term in (34) will be given by

$$\left(\boldsymbol{d_j.e^1}\gamma_1 + \boldsymbol{d_j.e^2}\gamma_2 + \cdots + \boldsymbol{d_j.e^{K-1}}\gamma_{K-1} + \boldsymbol{d_j.e^K}\gamma_K\right).$$

To proceed, let us define the following K vector

$$\boldsymbol{de_j} = \left(\boldsymbol{d_j.e^1} \quad \boldsymbol{d_j.e^2} \ldots \quad \boldsymbol{d_j.e^{K-1}} \quad \boldsymbol{d_j.e^K}\right), \tag{37}$$

where the k-th element of the above vector is the inner product of $\boldsymbol{d_j}$ and $\boldsymbol{e^k}$, with $k = 1, 2, \ldots, K$. Then, we have the following results:

- if the K vectors $\boldsymbol{de_j}$ and $\boldsymbol{\gamma}$ lie in the same orthant then the inclusion of irrelevant variables in the model with omitted variables will contribute a positive magnitude to the overall bias; this is because each of the terms in the j-th element of the second term in (34),

$$\left(\boldsymbol{d_j.e^1}\gamma_1 + \boldsymbol{d_j.e^2}\gamma_2 + \cdots + \boldsymbol{d_j.e^{K-1}}\gamma_{K-1} + \boldsymbol{d_j.e^K}\gamma_K\right),$$

  will be positive; if the OVB was positive, to begin with, this will increase the bias of the OLS estimator of the j-th included regressor;

- if the K vectors $\boldsymbol{de_j}$ and $\boldsymbol{\gamma}$ lie in opposite orthants, i.e. if each element of $\boldsymbol{de_j}$ is opposite in sign from the corresponding element of $\boldsymbol{\gamma}$, then the inclusion of irrelevant variables in the model with omitted variables will contribute a negative magnitude to the overall bias; this is because each of the terms in the j-th element of the second term in (34),

$$\left(\boldsymbol{d_j.e^1}\gamma_1 + \boldsymbol{d_j.e^2}\gamma_2 + \cdots + \boldsymbol{d_j.e^{K-1}}\gamma_{K-1} + \boldsymbol{d_j.e^K}\gamma_K\right),$$

  will be negative; if the OVB was negative, to begin with, this will increase the bias of the OLS estimator of the j-th included regressor;

- if the K vectors $\boldsymbol{de_j}$ and $\boldsymbol{\gamma}$ are orthogonal or one of them is a null vector, then the inclusion of irrelevant variables in the model with omitted variables will have no impact on the bias of the OLS estimator of the j-th included regressor.

6 Conclusion

In this paper, I have studied three issues related to the bias of OLS estimators arising from errors of exclusion (of relevant variables) and inclusion (of irrelevant variables): (1) omitted variable bias; (2) possible reduction of omitted variable bias with the inclusion of *some* of the omitted variables; (3) possible impact of the inclusion of irrelevant variables on the bias of OLS estimators in a model with omitted variables.

The first result of this paper is a derivation of some sufficient conditions, in terms of sign restrictions on parameters, to determine the direction of OVB. These are natural multivariate generalizations of the univariate case of one omitted variable and can partly address the longstanding problem of the difficulty of ascertaining the sign of OVB of OLS estimators in a multivariate context (Forbes, 2000; Greene, 2012). The second result of this paper is a reiteration of the results in Clarke (2005) and Luca et al. (2018) that inclusion of some omitted variables will not necessarily reduce the magnitude of OVB when some other relevant variables remain omitted. Moreover, we cannot derive sufficient conditions for bias reduction using sign restrictions only. The third result of this paper is to show that inclusion of irrelevant variables will have an impact, in general, on the bias of OLS estimators in a model with omitted variables. To the best of my knowledge, this is a novel result and is at odds with the common perception that inclusion of irrelevant variables does not have any impact on the bias - though they might have an impact on the variance - of OLS estimators (Fomby, 1981; Greene, 2012). At least three implications of this analysis are worth noting.

In discussing the problem of OVB, and of strategies to deal with it, researchers have frequently relied on arguments about the *direction* of the bias (Blackburn and Neumark, 1995; Tootell, 1996; Card, 2001; Hertz, 2003; Ahenfelter and Greenstone, 2004; Banerjee and Iyer, 2005; Autor et al., 2013). The first implication of the analysis in this paper is that researchers should use direction-of-bias arguments with caution because in most realistic cases, it is not possible to determine the sign of the OVB of OLS estimators. Unless the researcher carries out a detailed analysis of signs of parameters along the lines developed in section 3.1, it will not be possible to make any assertions about the direction of the OVB.

One common strategy to deal with the bias caused by omitted variables is to use instrumental variables estimators. In such contexts, it is standard in the literature to make comparisons of the direction and magnitude of bias of OLS and IV estimators (Angrist and Krueger, 2001, pp. 79). The second implication of the analysis in this paper points to the difficulty of making such comparisons - other than in very simple cases. If, in general, neither the magnitude nor the sign of OLS bias can be determined, then it is not clear how one would compare it with the possibly large bias of the IV estimator caused, for instance, by the use of weak instruments.

The third implication of the analysis in this paper is that the commonly used strategy to deal with bias of OLS estimators with the inclusion of more variables in regression models should be used with caution. Researchers should be aware of two important possibilities related to the inclusion of variables. First, if the estimated model continues to exclude some relevant variables, then inclusion of irrelevant variables will have an impact on the bias of OLS estimators - quite apart from the impact on its efficiency (Fomby, 1981; Greene, 2012). Second, if the estimated model continues to exclude relevant variables, then the inclusion of relevant variables cannot, in general, guarantee reduction in bias of OLS estimators.

References

Orley Ahenfelter and Michael Greenstone. Estimating the value of statistical life: The importance of omitted variables and publication bias. *AEA Papers and Proceedings*, 94 (2):454–460, 2004.

Joshua D. Angrist and Alan B. Krueger. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4):69–85, 2001.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, New York, 2009.

David H. Autor, David Dorn, and Gordon H. Hanson. The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–2168, 2013.

Abhijit Banerjee and Lakshmi Iyer. History, institutions, and economic performance: The legacy of colonial land tenure systems in India. *American Economic Review*, 95(4):1190–1213, 2005.

Deepankar Basu. When can we determine the direction of omitted variable bias of ols estimators? Working Paper 2018-16, Department of Economics, UMass Amherst. Retrieved from `https://www.umass.edu/economics/publications/2018-16.pdf`, 2018.

McKinley L. Blackburn and David Neumark. Are ols estimates of the return to schooling biased downward? another look. *The Review of Economics and Statistics*, 77(2):217–230, 1995.

David Card. Estimating the return to schooling; progress on some persistent econometric problems. *Econometrica*, 69(5):1121–1160, 2001.

Damian Clarke. A convenient omitted variable bias formula for treatment effect models. *Economics Letters*, 174:84–88, 2019.

Kevin A. Clarke. The phantom menace: Omitted variable bias in econometric research. *Conflict Management and Peace Science*, 22:341–352, 2005.

Thomas B. Fomby. Loss of efficiency in regression analysis due to irrelevant variables: A generalization. *Economics Letters*, 7:319–322, 1981.

Kristin J. Forbes. A reassessment of the relationship between inequality and growth. *American Economic Review*, 90(4):170–192, 2000.

William H. Greene. *Econometric Analysis*. Prentice Hall, seventh edition, 2012.

Tom Hertz. Upward bias in the estimated returns to education: Evidence from south africa. *American Economic Review*, 93(4):1121–1160, 2003.

Giuseppe De Luca, Jan R. Magnus, and Franco Peracchi. Balanced variable addition in linear models. *Journal of Economic Surveys*, 32(4):1183–1200, 2018.

Geoffrey M. B. Tootell. Redlining in boston: Do mortgage lenders discriminate against neighbourhoods. *Quarterly Journal of Economics*, 111(4):1049–1079, 1996.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2002.

Appendix A

Proof of Proposition 1.

Proof. For the omitted variable bias in $\hat{\boldsymbol{\beta}}_{\boldsymbol{S}}$, conditional on the regressors, $\boldsymbol{X}, \boldsymbol{Z}$, note, from the model in (10), that

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{S}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y},$$

so that

$$\mathbb{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{S}} = \mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\right] = \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{\gamma},$$

where we have plugged in the expression for $\boldsymbol{y}$ from the true model in (11), and the last step follows from the orthogonality of the error term given in (8).

Using the algebra of partitioned matrices, that

$$\boldsymbol{X}'\boldsymbol{Z} = \boldsymbol{X}'\begin{bmatrix}\boldsymbol{z}^1 & \boldsymbol{z}^2 & \cdots & \boldsymbol{z}^M\end{bmatrix} = \begin{bmatrix}\boldsymbol{X}'\boldsymbol{z}^1 & \boldsymbol{X}'\boldsymbol{z}^2 & \cdots & \boldsymbol{X}'\boldsymbol{z}^M\end{bmatrix}$$

where $\boldsymbol{z}^m$ refers to the $N\times1$ vector representing the m-th column of $\boldsymbol{Z}$, with $m = 1, 2, \ldots, M$. Hence

$$\begin{aligned}
\mathbb{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{S}} - \boldsymbol{\beta} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{\gamma} \\
&= \begin{bmatrix}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{z}^1 & \cdots & \left(\mathbf{X}'\mathbf{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{z}^M\end{bmatrix}\boldsymbol{\gamma} \\
&= \begin{bmatrix}\hat{\boldsymbol{\delta}}^1 & \cdots & \hat{\boldsymbol{\delta}}^M\end{bmatrix}\boldsymbol{\gamma} \\
&= \hat{\boldsymbol{\delta}}\boldsymbol{\gamma}
\end{aligned}$$

where, for $m = 1, 2, \ldots, M$, $\hat{\boldsymbol{\delta}}^m$ is OLS estimator of the coefficient vector in the linear projection of the m-th omitted variable on the whole set of included regressors, i.e.

$$\boldsymbol{z}^m = \boldsymbol{X}\boldsymbol{\delta}^m + v_m, \tag{38}$$

with $\mathbb{E}\left(\boldsymbol{X}'v_m\right) = \mathbf{0}$, so that

$$\hat{\boldsymbol{\delta}}^m = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{z}^m.$$

Columnwise stacking of $\hat{\boldsymbol{\delta}}_m$, then gives the $J \times M$ matrix $\hat{\boldsymbol{\delta}}$.

For the omitted variable bias in $\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}$, conditional on the regressors, $\boldsymbol{X}, \boldsymbol{Z}$, consider the model in (11) and note that

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}_{\boldsymbol{L}} \\ \hat{\gamma}_{\mathbf{1}}\end{pmatrix} = \left(\mathbf{U}'\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{y}$$

where $\boldsymbol{U} = [\boldsymbol{X} \quad \boldsymbol{Z}]$, so that

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}_{\boldsymbol{L}} \\ \hat{\gamma}_{\mathbf{1}}\end{pmatrix} = \begin{bmatrix}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z}_{\mathbf{1}} \\ \boldsymbol{Z}_{\mathbf{1}}'\boldsymbol{X} & \boldsymbol{Z}_{\mathbf{1}}'\boldsymbol{Z}_{\mathbf{1}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}_{\mathbf{1}}'\boldsymbol{y}\end{bmatrix}.$$

Substituting for $\boldsymbol{y}$ from (11) and using the fact that $\mathbb{E}\left(\boldsymbol{u}|\boldsymbol{X},\boldsymbol{Z}\right)=\boldsymbol{0}$, we have

$$\begin{pmatrix}\mathbb{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}-\boldsymbol{\beta}\\ \mathbb{E}\hat{\boldsymbol{\gamma}}_{\boldsymbol{1}}-\boldsymbol{\gamma}_{\boldsymbol{1}}\end{pmatrix}=\begin{bmatrix}\boldsymbol{X'X} & \boldsymbol{X'Z_1}\\ \boldsymbol{Z_1'X} & \boldsymbol{Z_1'Z_1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{X'Z_2}\\ \boldsymbol{Z_1'Z_2}\end{bmatrix}\boldsymbol{\gamma_2}. \tag{39}$$

Hence, using the expression for the inverse of partitioned matrices in Greene (2012, pp. 993–994), we have

$$\begin{pmatrix}\mathbb{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}-\boldsymbol{\beta}\\ \mathbb{E}\hat{\boldsymbol{\gamma}}_{\boldsymbol{1}}-\boldsymbol{\gamma}_{\boldsymbol{1}}\end{pmatrix}=\begin{bmatrix}\boldsymbol{W^{11}} & \boldsymbol{W^{12}}\\ \boldsymbol{W^{21}} & \boldsymbol{W^{22}}\end{bmatrix}\begin{bmatrix}\boldsymbol{X'Z_2}\\ \boldsymbol{Z_1'Z_2}\end{bmatrix}\boldsymbol{\gamma_2} \tag{40}$$

where

$$\boldsymbol{W^{11}}=\left(\boldsymbol{X'X}\right)^{-1}+\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1FZ_1'X}\left(\boldsymbol{X'X}\right)^{-1},$$

$$\boldsymbol{W^{12}}=\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1F},$$

$$\boldsymbol{W^{22}}=\boldsymbol{F},$$

and

$$\boldsymbol{F}=\left(\boldsymbol{Z_1'Z_1}-\boldsymbol{Z_1'X}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1}\right)^{-1}.$$

Hence,

$$\mathbb{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}-\boldsymbol{\beta}=\boldsymbol{W^{11}X'Z_2\gamma_2}+\boldsymbol{W^{12}Z_1'Z_2\gamma_2}.$$

Note that

$$\boldsymbol{F}=\left(\boldsymbol{Z_1'Z_1}-\boldsymbol{Z_1'X}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1}\right)^{-1}=\left[\boldsymbol{Z_1'M_XZ_1}\right]^{-1},$$

where $\boldsymbol{M_X}=\boldsymbol{I}-\boldsymbol{X}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'}$ is the “residual maker” matrix that is symmetric and idempotent (Greene, 2012, pp. 31). Hence $\boldsymbol{F}=\left(\boldsymbol{Z_{1X}'Z_{1X}}\right)^{-1}$, where $\boldsymbol{Z_{1X}}=-\boldsymbol{M_XZ_1}$. Collecting terms, we get

$$\mathbb{E}\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}-\boldsymbol{\beta}=\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_2\gamma_2}-\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'Z_1}\left(\boldsymbol{Z_{1X}'Z_{1X}}\right)^{-1}\boldsymbol{Z_{1X}'Z_2\gamma_2},$$

which is the expression in (13).

□

Proof of Proposition 3.

Proof. The estimated model in (32) is

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{W\delta}+\boldsymbol{v},$$

which can be written as

$$\boldsymbol{y}=\boldsymbol{U\alpha}+\boldsymbol{v},$$

where $\boldsymbol{U} = (\boldsymbol{X} \quad \boldsymbol{W})$ and $\boldsymbol{\alpha} = (\boldsymbol{\beta}' \quad \boldsymbol{\delta}')'$. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ denote the OLS estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, respectively. Then, we have, conditional on $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}$,

$$\begin{pmatrix} \mathbb{E}\hat{\boldsymbol{\beta}} \\ \mathbb{E}\hat{\boldsymbol{\delta}} \end{pmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W} \\ \boldsymbol{W}'\boldsymbol{X} & \boldsymbol{W}'\boldsymbol{W} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} \\ \boldsymbol{W}'\boldsymbol{Z} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W} \\ \boldsymbol{W}'\boldsymbol{X} & \boldsymbol{W}'\boldsymbol{W} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{W}'\boldsymbol{Z} \end{bmatrix} \gamma, \tag{41}$$

where we have used the fact that $\mathbb{E}\left(\boldsymbol{u}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W}\right) = \boldsymbol{0}$. Using the results for the inverse of partitioned matrices, as in the previous proof, we get

$$\mathbb{E}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\boldsymbol{W_{11}}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W_{12}}\boldsymbol{W}'\boldsymbol{X}\boldsymbol{\beta}\right) + \left(\boldsymbol{W_{11}}\boldsymbol{X}'\boldsymbol{Z}\gamma + \boldsymbol{W_{12}}\boldsymbol{W}'\boldsymbol{Z}\gamma\right)$$

where

$$\boldsymbol{W_{11}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{F}\boldsymbol{W}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},$$

$$\boldsymbol{W_{12}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{F}, \quad \boldsymbol{W_{22}} = \boldsymbol{F},$$

and

$$\boldsymbol{F} = \left[\boldsymbol{W}'\boldsymbol{W} - \boldsymbol{W}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}\right]^{-1}.$$

Using the expressions for $\boldsymbol{W_{11}}$ and $\boldsymbol{W_{12}}$, and simplifying gives

$$\mathbb{E}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Z}\gamma + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{F}\boldsymbol{W}'\left[\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' - \boldsymbol{I}\right]\boldsymbol{Z}\gamma.$$

Using $\boldsymbol{M_X} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' - \boldsymbol{I}$ and $\boldsymbol{W_X} = -\boldsymbol{M_X}\boldsymbol{W}$, we get

$$\mathbb{E}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Z}\gamma + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}\left(\boldsymbol{W}_{\boldsymbol{X}}'\boldsymbol{W_X}\right)^{-1}\boldsymbol{W}_{\boldsymbol{X}}'\boldsymbol{Z}\gamma,$$

which is the expression in (34). □

Appendix B

In discussing the problem of omitted variable bias, and of strategies to deal with it, researchers have frequently relied on arguments about the *direction* of the bias. Here are some examples of the use of direction-of-bias arguments in papers published over the last few decades.[10]

- "One of the longest-running debates in empirical labor economics regards bias in OLS estimates of the economic return to schooling. The overriding concern pertains to individual-specific productivity components not reflected in the usual human-capital measures, as these ability components may be positively correlated with both wages and schooling. If the return to schooling is estimated with no account taken of the role of ability, the estimate is *generally expected to be biased upward.* (Blackburn and Neumark, 1995, pp. 217, emphasis added).

[10] This list of examples is purely for the purposes of illustration and does not pretend to completeness.

- "Equation (7) generalizes the conventional analysis of ability bias in the relationship between schooling and earnings. Suppose that there is no heterogeneity in the marginal benefits of schooling (i.e., $b_i = \bar{b}$) and that log earnings are linear in schooling (i.e. $k_1 = 0$). Then (7) implies that

  $$\operatorname{plim} b_{ols} - \bar{b} = \lambda_0$$

  which is the standard expression for the asymptotic bias in the estimated return to schooling that arises by applying the omitted variables formula to an earnings model with a constant schooling coefficient $\bar{b}$. According to the model presented here, this bias arises through the correlation between the ability component a_i and the marginal cost of schooling r_i. If marginal costs are lower for people who would tend to earn more at any level of schooling, then $\sigma_{ra} < 0$, implying that $\lambda_0 > 0$." (Card, 2001, pp. 1134).

- "Ordinary least-squares (OLS) estimates of the proportionate increase in wages due to an extra year of education in the United States (the Mincerian rate of return) are believed to be reasonably consistent. It appears that *upward bias* due to omitted variables is roughly offset by attenuation bias due to errors in the measurement of schooling. Orley Ashenfelter and Cecilia Rouse (1998) find a net *upward bias* on the order of just 10 percent of the magnitude of the OLS estimate. David Card's (2001) survey of instrumental variables-based estimates reaches a similar conclusion, as do Ashenfeiter et al. (1999)." (Hertz, 2003, pp. 1354, emphasis added).

- "Our IV results, together with the results on neighboring districts and the historical data, lead us to conclude that our OLS results are *not biased upward* due to omitted district characteristics." (Banerjee and Iyer, 2005, pp. 1206, emphasis added).

- "There are several possible threats to our strategy. One is that product demand shocks may be correlated across high-income countries. In this event, both our OLS and IV estimates may be contaminated by correlation between import growth and unobserved components of product demand, making the impact of trade exposure on labor-market outcomes *appear smaller than it truly is*." (Autor et al., 2013, pp. 2129, emphasis added).