A Convenient Omitted Variable Bias Formula for Treatment Effect Models*

Damian Clarke

October 6, 2018

Abstract

Generally, determining the size and magnitude of the omitted variable bias (OVB) in regression models is challenging when multiple included and omitted variables are present. Here, I describe a convenient OVB formula for treatment effect models with potentially many included and omitted variables. I show that in these circumstances it is simple to infer the direction, and potentially the magnitude, of the bias. In a simple setting, this OVB is based on mutually exclusive binary variables, however I provide an extension which loosens the need for mutual exclusivity of variables, deriving the bias in difference-in-differences style models with an arbitrary number of included and excluded "treatment" indicators.

JEL codes: C21, C22, C13.

Keywords: Omitted variable bias; Ordinary Least Squares Regression; Treatment Effects; Difference-in-Differences.

*This research was supported by FONDECYT (grant number 11160200) of the Government of Chile. I thank Serafima Chirkova and an anonymous referee for useful comments which have improved this work. Author's Affiliation: Department of Economics, Universidad de Santiago de Chile and IZA. Avenida Libertador Bernardo O'Higgins, Estación Central, Santiago, Chile. Contact email: damian.clarke@usach.cl. Contact telephone: +56227180750.

1 Introduction

The omitted variable bias (OVB) is a staple of econometrics courses, and applied research across all fields of economics, appearing as early as Theil (1957). In its most basic form, the omission of a single relevant explanatory variable in a linear model leads to an elegant bias formula providing a simple link between parameter estimates, true values, and underlying relationships between variables. This formula is often amenable to analysis using intuition from economic models. However, once outside a simple text-book case of a single included and excluded independent variable, the OVB can become increasingly complex, such that inferring even the direction of bias is impractical in models with multiple included and excluded variables (Clarke, 2005).[1]

In this paper I provide a convenient representation of the OVB in a model with arbitrarily many included and omitted variables, and an intuitive link to the underlying data generating process (or treatment assignation). The convenience of this representation comes at the cost of the *class* of models for which it serves. This representation is provided for models based on a series of mutually exclusive binary "treatment" variables. After documenting the bias for a case where potentially many treatment variables are included and excluded in a linear model, I then provide an extension to a more complicated setting: the difference-in-differences model. In this setting, while treatment effect indicators may be mutually exclusive among themselves, common fixed effects (e.g. for time) are shared. I document in this case that the simple OVB formula holds, and follows the same logic as in static models.

While this is a restrictive model, it is nonetheless frequently observed in empirical applications with a wide set of real-use cases. Models designed to estimate treatment effects with multiple treatment statuses—where a population is split into various treatment groups and a control group—are often encountered. A number of such cases are found in Kremer (2003); Banerjee et al. (2007), (these are referred to as "cross-cutting designs" in Duflo et al. (2007)) and also in fields beyond economics, including medicine (Baron et al., 2013). This is particularly relevant where concerns exist about imperfect observation of treatment status, for example where treatment externalities occur (Miguel and Kremer, 2004), or where environmental shocks diffuse over space (Almond et al., 2009). Finally, the OVB derived here extends to *any* model based on mutually exclusive (or largely mutually exclusive) binary independent variables, such as fixed-effect models. In the discussion section of this paper, I describe a wider set of circumstances in which models of this type are encountered in a range of fields of the economic literature, and also lay out a benefit of consistently deriving these biases using the

[1] For example, consider the presentation of Greene (2002, p. 180) in his widely studied text-book, who states, "if more than one variable is included, then the terms in the omitted variable formula involve multiple regression coefficients, which themselves have the signs of partial, not simple, correlations. …This requirement might not be obvious, and it would become even less so as more regressors were added to the equation."

OVB formula; namely, the ability to extend to an even wider range of models, such as those where treatment measures are continuous rather than binary.

2 The Traditional Omitted Variable Bias in Linear Models

To illustrate the traditional OVB model, consider a correctly specified model of the form:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{1}$$

Here the independent variables $X = [X_1 \quad X_2]$ are split into two matrices. The X_1 matrix consists of $1 + k_1$ variables (a constant plus k_1 other variables) and X_2 consists of k_2 variables. The error term ε in the full model is orthogonal to each column of X. Throughout this paper matrices are denoted in upper-case letters, while particular variables are denoted in lower-case.

If the dependent variable y is regressed *only* on X_1 the expectation of the OLS estimator of the parameter β_1 is:

$$E[\widehat{\beta}_1^{ovb}|X] = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2. \tag{2}$$

This is the well-known OVB formula. The bias term of $(X_1'X_1)^{-1}X_1'X_2\beta_2$ is sometimes written as $\delta\beta_2$, where $\delta \equiv (X_1'X_1)^{-1}X_1'X_2$ is a $(1 + k_1) \times k_2$ projection matrix from the regression of each variable in X_2 on the full set of X_1 variables. This "textbook" OVB has a very simple interpretation in the case where $k_1 = 1$ and $k_2 = 1$: it is the product of the simple correlation between variables x_1 and x_2 (δ) and the direct effect of omitted x_2 on y in the population model (β_2). However, this formula can quickly become unwieldy when $k_1 > 1$ as each element of the $\widehat{\beta}_1^{ovb}$ vector differs from β_1 in expectation depending on the partial correlation between the relevant variable in X_1, conditional on the remainder of X_1. Economic models and intuition often have less to say about *partial* correlations between (potentially many) variables.

3 A Convenient Omitted Variable Bias Formula for Treatment Effects Models

Consider now model 1, where each element of X is a binary variable. This is what we refer to here as a treatment effect model which examines the impact of multiple "treatments" on an outcome of interest. Further, assume that each variable is mutually exclusive. Such a model is common when a pool of subjects are split into various treatment groups and a control group, each receiving at most one treatment. While this mutual exclusivity assumption may appear limiting, it is actually more flexible than it appears, as receipt of multiple treatments can be considered a treatment unto itself.[2] In this case, we can show that the OVB above has a very convenient and intuitive form, even in cases with an arbitrary number of included and excluded variables. In section 4 we document a range of real-world cases from the literature, and discuss how this model can extend to an even wider set of real-world econometric settings.

3.1 A Single Omitted Variable and Included Variable

In the simplest case, the true treatment model consists of a single omitted variable (an $N \times 1$ vector x_2) and single included explanatory variable (an $N \times 1$ vector x_1), plus an intercept term. Such a situation would apply, for example, in a Randomized Control Trial (RCT) with treatment externalities if a single treatment indicator is included, but no indicator is included to capture treatment externalities. In such a setting the OVB is easily interpretable as per equation 2, however I briefly document an alternative interpretation of the bias in this setting, before moving to the more complicated multivariate setting in sections 3.2-3.3.

Consider the bias term of equation 2. This can be further simplified if we resolve the matrix product $(X_1'X_1)^{-1}X_1'X_2\beta_2 = \delta\beta_2$. Given that each of x_1 and x_2 are binary, we denote N_{x_1} and N_{x_2} as the quantity of observations for which $x_1 = 1$ and $x_2 = 1$ respectively. This gives the following matrices for $X_1'X_1$ and $X_1'X_2$:

$$(X_1'X_1) = \begin{bmatrix} N & N_{x_1} \\ N_{x_1} & N_{x_1} \end{bmatrix} \qquad (X_1'X_2) = \begin{bmatrix} N_{x_2} \\ 0 \end{bmatrix}.$$

Inverting $(X_1'X_1)$ and post-multiplying by $(X_1'X_2)$ gives the matrix of partial correlations, δ, as:

$$\delta = \begin{bmatrix} \frac{N_{x_2}}{N-N_{x_1}} \\ -\frac{N_{x_2}}{N-N_{x_1}} \end{bmatrix},$$

[2] This is laid out explicitly in Duflo et al. (2007) who states "If a researcher is cross-cutting interventions A and B, each of which has a comparison group, she obtains four groups: no interventions (pure control); A only; B only; and A and B together (full intervention)" These four groups are mutually exclusive.

and from equation 2, we thus express the OVB formula for the included variable x_1 as[3]:

$$E[\widehat{\beta}_1^{1,ovb}|X] = \beta_1^1 - \beta_2 \left(\frac{N_{x_2}}{N - N_{x_1}} \right). \tag{3}$$

In equation 3 we add a superscript 1 to β_1 to denote that we are displaying the bias on the "treatment" indicator x_1 only, however given the matrices it is a trivial extension to calculate the bias on the constant term also. Throughout this paper we consistently report the bias on the treatment indicators of interest only, relegating additional algebra and bias on other terms to the Supplemental Appendix.

This OVB also has a simple interpretation when cast in terms of the underlying treatment effect models. If x_2 (a treatment indicator) is omitted from the model, this group will be confounded with the true controls. Thus, the treatment effect on x_1 will be biased by any non-zero impact of x_2 on y (β_2), multiplied by the degree to which these x_2 units dilute the true control group: $N_{x_2}/(N - N_{x_1})$.

3.2 An Arbitrary Quantity of Omitted and Excluded Variables

While this simple OVB formula is intuitive, it is more interesting to be able to generalise this representation to a case with multiple omitted and included variables, which are much less frequently amenable to a clear interpretation using the original OVB formula, and economic logic. For example, in a treatment effects model based on cross-cutting interventions, if each treatment is subject to externalities and these are not controlled for, we will be in the presence of a model with multiple included and multiple excluded treatment indicators. Let each of the multiple (k_1) variables in X_1 be called $x_1^k \forall k = 1, \ldots, k_1$, (a constant is also included in X_1, implying that this matrix has 1+k_1 columns) and similarly, the (multiple) k_2 variables in X_2 be called $x_2^k \forall k = 1, \ldots, k_2$.

Given the mutual exclusivity of included treatment indicators plus a shared constant term, the matrix $(X_1'X_1)$ consists of non-zero entries on the main diagonal, first column and first row, and zeros elsewhere. The general class of matrices of this form (arrowhead matrices) has a simple inverse formula (Najafi et al., 2014), and we derive the particular inverse $(X_1'X_1)^{-1}$ for an arbitrary number of variables in the Supplemental Appendix. Based on this inverse and $(X_1'X_2)$, the matrix of partial correlations

[3] Note that as this is a bivariate regression model, this can also be derived using simply the covariance and variance. We start with the simple bivariate version of the OVB, and as each variable is binary, the covariance and variance have the closed form solutions below, where $N_{x_1x_2}$ refers to the quantity of observations for which both $x_1 = 1$ and $x_2 = 1$ (none). As in 3, this gives:

$$E[\widehat{\beta}_1^{1,ovb}|X] = \beta_1^1 + \beta_2 \left(\frac{Cov(x_1, x_2)}{Var(x_1)} \right) = \beta_1^1 + \beta_2 \left(\frac{(N_{x_1x_2} \cdot N - N_{x_1}N_{x_2})/N^2}{N_{x_1}(N - N_{x_1})/N^2} \right) = \beta_1^1 - \beta_2 \left(\frac{N_{x_2}}{N - N_{x_1}} \right).$$

between omitted and included variables is solved as:

$$\delta = (X_1'X_1)^{-1}X_1'X_2 = \begin{bmatrix} \frac{N_{x_2^1}}{\lambda} & \frac{N_{x_2^2}}{\lambda} & \cdots & \frac{N_{x_2^{k_2}}}{\lambda} \\ -\frac{N_{x_2^1}}{\lambda} & -\frac{N_{x_2^2}}{\lambda} & \cdots & -\frac{N_{x_2^{k_2}}}{\lambda} \\ -\frac{N_{x_2^1}}{\lambda} & -\frac{N_{x_2^2}}{\lambda} & \cdots & -\frac{N_{x_2^{k_2}}}{\lambda} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{N_{x_2^1}}{\lambda} & -\frac{N_{x_2^2}}{\lambda} & \cdots & -\frac{N_{x_2^{k_2}}}{\lambda} \end{bmatrix},$$

where $\lambda = N - N_{x_1^1} - N_{x_1^2} - \cdots - N_{x_1^{k_1}}$ (full details are available in the Supplementary Appendix). Finally, substituting this matrix into the OVB formula 2 leads to the multivariate generalisation of 3 for each of the coefficients on the k_1 included treatment indicators:

$$E[\widehat{\beta}_1^{k,ovb}|X] = \beta_1^k - \beta_2^1\left(\frac{N_{x_2^1}}{\lambda}\right) - \beta_2^2\left(\frac{N_{x_2^2}}{\lambda}\right) - \cdots - \beta_2^{k_2}\left(\frac{N_{x_2^{k_2}}}{\lambda}\right), \tag{4}$$

where each β_2^k is the true impact of x_2^k on y.

We note two things about this OVB. Firstly, it is identical for each β_1^k term. Secondly, as in the univariate case, this has an intuitive explanation when cast in terms of treatment effects. The treatment effect on each included variable will be biased by any non-zero impact of each excluded treatment group (the β_2^k terms), multiplied by the degree that each of these omitted treatment indicators biases the formation of the control group ($N_{x_2^k}/(N - N_{x_1^1} - \cdots - N_{x_1^{k_1}})$). If each omitted treatment effect has the same sign as included treatment effects (this may be the case, for example, in spillovers of environmental shocks), estimated effects will be universally attenuated. Alternatively, if excluded treatments have opposite effects to included treatments, estimated effects will be consistently overstated in magnitude.

3.3 Extension to Alternative Treatment Models

Equation 4 provides an intuitive solution to the OVB in treatment effect models with multiple treatments, but only holds under the somewhat restrictive assumption that each treatment variable is mutually exclusive. However, the OVB derived above can be logically extended to other important settings. To see this, we consider a difference-in-differences model. Consider the correctly specified model:

$$y_{it} = \beta_0 + \beta_1 X_{1i} + \beta_2(X_1 \cdot \text{Post})_{it} + \beta_3\text{Post}_t + \gamma_1 X_{2i} + \gamma_2(X_2 \cdot \text{Post})_{it} + \varepsilon_{it} \tag{5}$$

with observations in two periods $t \in \{0,1\}$. The variable Post_t takes the value of 1 for all observations in period $t = 1$ and 0 when $t = 0$. Units for whom $X_{1i} = 1$ or $X_{2i} = 1$ receive treatment, which switches on only when interacted with the Post-treatment dummy. Units for whom both $X_{1i} = 0$ and $X_{2i} = 0$ are pure controls. Thus, as is standard in difference-in-difference models, treatment effects are estimated as the terms on the interaction: β_2 and γ_2. Baseline differences between treated and untreated individuals are captured by fixed effects β_1 and γ_1, and any generalised temporal impacts are captured by the time fixed effect β_3. This multi-group difference-in-differences model is similar to that in Imbens and Wooldridge (2009).

Suppose we estimate this model, omitting the vectors $(X_2 \cdot \text{Post})_{it}$ and corresponding fixed effect X_{2i}. We denote this vector $X_{2D} = [X_{2i} \quad (X_2 \cdot \text{Post})_{it}]$, and denote the vector $X_{1D} = [1 \quad X_{1i} \quad (X_1 \cdot \text{Post})_{it} \quad \text{Post}_t]$. We define k_1 as the quantity of treatment variables $(X_1 \cdot \text{Post})_{it}$ in X_{1D}, and k_2 the quantity of treatment variables in X_{2D}.[4] Here, the traditional OVB formula is given as:

$$E[\hat{\beta}^{ovb}|X] = \beta + (X_{1D}'X_{1D})^{-1}(X_{1D}'X_{2D})\gamma = \beta + \delta\gamma, \tag{6}$$

where δ once again refers to the projection matrix, here of X_{2D} on X_{1D}. To show that the same type of convenient OVB formula as in sections 3.1-3.2 exists for a difference-in-differences model, I first consider the case of a single included and omitted treatment variable (and corresponding fixed effects), before extending to the case with an arbitrary quantity of included and omitted variables.

Although the $(X_{1D}'X_{1D})$ matrix is now more complex than in the static treatment model, once again we can show that it has a reasonably simple closed form solution for the inverse, and, consequently, a simple matrix of partial correlations between included and excluded variables. Given the model's structure, $(X_{1D}'X_{1D})$ is a 2×2 block matrix, and can be inverted (see Supplemental Appendix and Lu and Shiou (2002)), and post-multiplied by $(X_{1D}'X_{2D})$ to give the partial correlation matrix as:

$$\delta = \begin{bmatrix} \frac{N_{x_{2t}}}{N_t - N_{x_{1t}}} & 0 \\ -\frac{N_{x_{2t}}}{N_t - N_{x_{1t}}} & 0 \\ 0 & -\frac{N_{x_{2t}}}{N_t - N_{x_{1t}}} \\ 0 & \frac{N_{x_{2t}}}{N_t - N_{x_{1t}}} \end{bmatrix}. \tag{7}$$

As before, N refers to the number of observations, $N_{x_{1t}}$ to those for which $(X_1 \cdot \text{Post})_{it} = 1$ (ie those for whom $X_{1i} = 1$ in period $t = 1$), and $N_{x_{2t}}$ to the quantity for which $(X_2 \cdot \text{Post})_{it} = 1$. Additionally N_t is equal to the number of observations for which $\text{Post}_t = 1$. Putting this together following equation

[4] A subscript D is added to these matrices to indicate that they are based on difference-in-difference models, and so are different to the X_1 and X_2 matrices described in equation 1 given that they also include baseline fixed effects and, in the case of X_{1D}, a time dummy. We use k_1 and k_2 consistently to refer to the number of "treatment effects" of interest.

6 gives the OVB formula for the time-varying element of a difference-in-differences model as:

$$E[\widehat{\beta}_2^{ovb}|X] = \beta_2 - \gamma_2 \left(\frac{N_{x_{2t}}}{N_t - N_{x_{1t}}} \right). \tag{8}$$

Here we observe that the OVB follows virtually the same logic as in equation 3, however now conditional on treatment occurring in the second period. The omission of a relevant treatment indicator in difference-in-differences models biases included treatment effects by any effect that this treatment indicator has on the outcome of interest, multiplied by the proportion of the (naive) "control group" that were actually treated.

Finally, note that this OVB formula can be resolved in this way even in the extreme case of multiple included and multiple omitted treatment indicators. An example of a model such as this is the application of Miguel and Kremer (2004), where two treatment groups each potentially generate their own externalities. The structure of the time-specific treatment indicators in equation 5 implies that once again $(X_{1D}'X_{1D})$ is a 2×2 block matrix, and each block is an arrowhead matrix of dimension $(1+k_1) \times (1+k_1)$. Thus, each block of the $(X_{1D}'X_{1D})$ matrix is invertible, as is the underlying matrix (full details in Supplemental Appendix).

In the difference-in-difference model with multiple included and excluded treatment indicators, the matrix of partial correlations between excluded and included variables is:

$$\delta = \begin{bmatrix} \frac{N_{x_{2t}^1}}{\theta_t} & \cdots & \frac{N_{x_{2t}^{k_2}}}{\theta_t} & 0 & \cdots & 0 \\ -\frac{N_{x_{2t}^1}}{\theta_t} & \cdots & -\frac{N_{x_{2t}^{k_2}}}{\theta_t} & 0 & \cdots & 0 \\ -\frac{N_{x_{2t}^1}}{\theta_t} & \cdots & -\frac{N_{x_{2t}^{k_2}}}{\theta_t} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\frac{N_{x_{2t}^1}}{\theta_t} & \cdots & -\frac{N_{x_{2t}^{k_2}}}{\theta_t} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -\frac{N_{x_{2t}^1}}{\theta_t} & \cdots & -\frac{N_{x_{2t}^{k_2}}}{\theta_t} \\ 0 & \cdots & 0 & -\frac{N_{x_{2t}^1}}{\theta_t} & \cdots & -\frac{N_{x_{2t}^{k_2}}}{\theta_t} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -\frac{N_{x_{2t}^1}}{\theta_t} & \cdots & -\frac{N_{x_{2t}^{k_2}}}{\theta_t} \\ 0 & \cdots & 0 & \frac{N_{x_{2t}^1}}{\theta_t} & \cdots & \frac{N_{x_{2t}^{k_2}}}{\theta_t} \end{bmatrix}. \tag{9}$$

where $\theta_t = N_t - N_{x_{1t}^1} - N_{x_{1t}^2} - \cdots - N_{x_{1t}^{k_1}}$. Note that the matrix in equation 7 is just a special case of the above, where $k_1 = 1$. Now, to determine the OVB *on each* included variable we return to equation 6. From this, and the preceding matrices, we have that the OVB formula on each treatment effect in time-varying $(X_1 \cdot \text{Post})_{it}$ is:

$$E[\widehat{\beta}_2^{k,ovb}|X] \;=\; \beta_2^k - \gamma_2^1 \frac{N_{x_{2t}^1}}{\theta_t} - \gamma_2^2 \frac{N_{x_{2t}^2}}{\theta_t} - \gamma_2^{k_2} \frac{N_{x_{2t}^{k_2}}}{\theta_t} \quad \forall k \in 1, \ldots, k_1. \tag{10}$$

This bias term is once again intuitive in the treatment effects framework. Here each omitted binary treatment indicator is incorrectly included in the control group, and hence biases the estimated treatment effects of included variables. This bias consists of the true treatment effect of non-included treatment variables (γ_2^k) scaled by the degree to which this treatment group contaminates the naive control group ($N_{x_{2t}^k}/(N_t - N_{x_{1t}^1} - N_{x_{1t}^2} - \cdots - N_{x_{1t}^{k_1}})$). As discussed previously, if a researcher has prior information related to omitted treatment units, this may allow an even finer consideration of this bias: producing an attenuation if the γ^k coefficients are of the same sign as included treatment effects β, while overstating the magnitude if excluded and included treatment effects are of opposite sign.

4 Discussion and Conclusion

The Omitted Variable Bias is frequently encountered in economics. While it is the base of a range of useful derivations, when multiple omitted variables are considered in regressions it is often presented as an *ex-post* test of model stability (e.g. Gelbach (2016)), rather than as providing a simple *ex-ante* formula for determining parameter bounds.

This paper provides a formula for the OVB which can be used *ex-ante* to infer biases in treatment models. It documents a simple and intuitive link, where the bias owes to the contamination of the control group with treated units. I show that this holds in simple static treatment models, as well as in more complicated difference-in-differences models. This OVB formula provides an uncomplicated way to present treatment bounds in models with imperfectly observed treatment variables, as are often encountered in natural experiments, or experiments with imperfect compliance. This derivation can serve as both a useful starting point, and as a benchmark of potential biases in applied settings, in quite a wide range of circumstances. To document the flexibility of this result, below I provide a number of illustrations in varying contexts, along with applied examples in which these considerations may be useful. In each case laid out below, the bias derivations given in this paper will be of use if a simple treatment model is estimated, while in reality the more complete diffusion model is the true model.

Geographic Diffusion of Treatment through Environmental or Social Channels One of the most common settings in which potentially omitted treatment indicators occur is in the case of treatment diffusion over space, where the precise diffusion function is unknown, and depends on environmental or social channels. These are common in difference-in-difference settings, bus also occur in experimental cases. A very small subset of such examples include (a) the diffusion of environmental pollutants based on climatic and geographic conditions (see for example Almond et al. (2009)'s analysis of the Chernobyl catastrophe on human health), (b) individuals from surrounding neighbourhoods travelling to treated communities to access public policies (see for example Bentancor and Clarke (2017) who analyse this phenomenon in the availability of the emergency contraceptive pill in Chile), and (c) information diffusing over space through formal or informal institutions (for example Armand et al. (2017)'s analysis of "the reach of the radio"). In each of these cases, the proportion of the population living within certain distances of treated areas is typically easily observable, and so a full bias calculation following equation 10 of this paper simply requires an economic model or an assumed range for the impact of any potential treatment spillovers on the outcome of interest.

Social Interaction Models However, these phenomena are certainly not only limited to simple spatial spillovers. Alternative settings in which one may commonly be concerned about such treatment effect biases, particularly in the case of randomized experiments, is when treatment diffuses through social networks. In this case, the simple model discussed in this paper can be extended to a full social interaction model, which can, by definition or construction, meet mutual exclusivity criterion. Consider the highly flexible social interaction model of Manski (2013). The bias definitions provided in this paper will be relevant any time that his "individualized treatment response" (ITR) assumption fails. As Manski (2013) lays out, spillovers diffuse through a network following a response function. Restrictions on response functions limit the way that spillovers can occur, limiting the number of omitted indicators (our k_2) which may be relevant in our equation 4. If a simple treatment effect model is estimated (ie the ITR assumption is upheld), equation 4 will quantify the bias in treatment estimates in case of any social network spillovers, where each spillover parameter in the vector β_2 can be partially inferred based on assumptions or observations related to the interaction between units, for examine reinforcing or opposing interactions. Manski (2013) suggests a number of clear examples, including the burning of clean fuels and vaccine provision. An alternative example comes from de-worming, and within and between-school interactions described in Miguel and Kremer (2004).

General Equilibrium Estimates and other Economic Externalities Frequently, randomized or natural experiments are assessed to estimate local impacts of treated populations, abstracting away from any general equilibrium impacts. However, estimating naive program impacts may considerably understate *or* overstate the true total impacts if the program or natural experiment has wider general equilibrium effects on related markets or areas (see for example Allcott and Keniston (2018); Crépon et al. (2013) for two such documented cases). The formal OVB formulas laid out in equations 4 and 10 of this paper can provide a useful foothold to consider true treatment impacts, if some maximum general equilibrium effect is expected, as, for example, discussed in Allcott and Keniston (2018).

The above three settings are certainly not the exclusive domain where this bias calculation may act as a useful way to benchmark naive treatment effects, but they are provided to show that the class of models for which the derivation holds is not as narrow as it may appear. It is important to note that one could arrive at this bias calculation in a number of ways. In this paper I present the derivation in a consistent manner, housing it in the logic of the well-known omitted variable bias formula. One considerable benefit of such a derivation is that it opens a number of routes to extend this to more complicated models in other settings. For example, the structure lends itself very well to a case where rather than binary treatment measures, continuous treatment measures are applied (see for example Duflo (2001) for such a model). Extending the models in the paper to continuous treatment measures will result in a similarly structured bias in treatment effects, where otherwise a clear and economically relevant closed-form solution to parameter biases would be difficult to determine.

References

H. Allcott and D. Keniston. Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America. *The Review of Economic Studies*, 85(2):695–731, 2018.

D. Almond, L. Edlund, and M. Palme. Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden. *The Quarterly Journal of Economics*, 124(4): 1729–1772, 2009.

A. Armand, P. Atwell, and J. Gomes. The reach of radio: Defection messaging and armed group behavior. HiCN Working Papers 249, Households in Conflict Network, 2017.

A. V. Banerjee, S. Cole, E. Duflo, and L. Linden. Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics*, 122(3):1235–1264, 2007.

G. Baron, E. Perrodeau, I. Boutron, and P. Ravaud. Reporting of analyses from randomized controlled trials with multiple arms: a systematic review. *BMC Medicine*, 11(84):1–7, 2013.

A. Bentancor and D. Clarke. Assessing Plan B: The Effect of the Morning After Pill on Women and Children. *Economic Journal*, 127(607):2525–2552, December 2017.

K. A. Clarke. The Phantom Menace: Omitted Variable Bias in Econometric Research. *Conflict Management and Peace Science*, 22(4):341–352, May 2005.

B. Crépon, E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora. Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment. *The Quarterly Journal of Economics*, 128(2):531–580, 2013.

E. Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4):795–813, September 2001.

E. Duflo, R. Glennerster, and M. Kremer. Chapter 61 Using Randomization in Development Economics Research: A Toolkit. In T. P. Schultz and J. A. Strauss, editors, *Handbook of Development Economics*, volume 4, pages 3895 – 3962. Elsevier, 2007.

J. B. Gelbach. When Do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics*, 34(2):509–543, 2016.

W. H. Greene. *Econometric Analysis*. Prentice Hill, New Jersey, 5 edition, 2002.

G. W. Imbens and J. M. Wooldridge. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86, 2009.

M. Kremer. Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons. *American Economic Review*, 93(2):102–106, May 2003.

T.-T. Lu and S.-H. Shiou. Inverses of 2×2 Block Matrices. *Computers & Mathematics with Applications*, 43(1–2):119–129, 2002.

C. F. Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, February 2013.

E. Miguel and M. Kremer. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1):159–217, 01 2004.

H. S. Najafi, S. A. Edalatpanah, and G. A. Gravvanis. An efficient method for computing the inverse of arrowhead matrices. *Applied Mathematics Letters*, 33(1):1–5, 2014.

H. Theil. Specification Errors and the Estimation of Economic Relationships. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 25(1/3):41–51, 1957.