DISCUSSION PAPER

November 2014 ■ RFF DP 14-41

The Bias from Misspecification of Control Variables as Linear

Leonard Goff

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

The Bias from Misspecification of Control Variables as Linear

Leonard Goff*

Resources for the Future
1616 P St. NW, Washington DC, USA

Abstract

We characterize the asymptotic bias that arises in ordinary least squares regression when control variables have nonlinear effects on an outcome variable, but are assumed to enter the regression equation linearly. We show that this bias can be larger than that from omitting such variables altogether, or may vanish entirely despite the nonlinearity. We find that under a natural assumption an upper bound to the magnitude of the bias may be estimated from the data, and consider examples of the bias through Monte Carlo simulations.

Key words: nonlinear, linear, control bias, misspecification, misspecified
JEL classifications: C1 C2 C13

1. Introduction

In applied econometrics, researchers often use regression techniques to estimate the effect of a single variable of interest s on some outcome y. Typically, a set of control variables $x_1 \dots x_k$ is included to isolate the ceteris paribus effect of s, unconfounded by variation in the x_j. Consider the standard multiple linear regression model

$$y_i = \alpha + \beta s_i + x_i \gamma + u_i, \tag{1}$$

where α and β are scalar parameters and γ is a conformable vector of parameters for $x_1 \dots x_k$. Although theory does not always suggest whether the effects of s and the x_j on y are actually linear, Equation 1 is often estimated in the absence of any particular expected alternative functional form. In the present work, we investigate the asymptotic bias of the ordinary least squares (OLS) estimator $\hat{\beta}_{lin,k}$ from Equation 1, when the true effects of the x_j are nonlinear. We refer to this phenomenon as *linear misspecification bias* from the control variables.

While bias due to functional form misspecification is a familiar topic in econometrics (see, e.g., Ramsey 1969), less attention has focused on the common situation in

*The author would like to thank Allen Blackman for his valuable comments, as well as Eric Adebayo, Alex Egorenkov, Rob Williams, Thomas Lemieux, and Juha Siikamaki for useful discussions on this research.

Email address: `goff@rff.org` (Leonard Goff)

which the researcher is only ultimately interested in the effect of a single variable s. This is typical in applied settings when one regression variable is of particular policy interest or is considered as a treatment variable. When there is nonlinearity in the variable s itself, Angrist and Krueger (1999) develop an interpretation of $\hat{\beta}_{lin,k}$ as a weighted average over its marginal effects, under a restriction on the joint distribution of s and the x_j.

In the present work, we focus on nonlinearity in the x_j, when the effect of s is linear as assumed by the researcher. Within this setting, consistent semi-parametric estimators exist for β (e.g. Robinson 1988 and Yatchew 1997), which in principle provide a means of avoiding linear misspecification bias due to control variables. However, given how common it is to estimate Equation 1 in applied settings, it is important to understand the nature of the bias that may result, and under what conditions it poses a significant problem. As a cautionary note, Achen (2005) constructs a hypothetical dataset where a small nonlinearity in the effect of a control variable causes Equation 1 to give a wildly incorrect estimate of β. We develop a broad characterization of linear misspecification bias from control variables, which provides an explanation for the large bias in Achen's example.

In the following section, we derive an expression for linear misspecification bias, and compare it with the asymptotic bias that occurs when one of the control variables is omitted from the regression altogether. We then consider a special case in which the bias vanishes, as well as the estimation of an upper bound to linear misspecification bias. Finally, we turn to Monte Carlo simulations to investigate the bias given specific data-generating processes.

2. The bias from linear misspecification of control variables

2.1. Characterizing the bias

Consider a true model of the form

$$y_i = \alpha + \beta s_i + g(x_{i1}, x_{i2}, \dots x_{il}) + u_i \tag{2}$$

with the strong exogeneity assumption satisfied: $E(u_i|s_i, x_{i1} \dots x_{il}) = 0$. For short, we refer to $g(x_{i1}, x_{i2}, \dots x_{il})$ as $g(x)$. The unknown function $g(x)$ may or may not be linear with respect to each x_j, and may or may not be additively separable among the x_j. This setting is not completely general, as we have assumed that the regression function is still additively separable and linear with respect to s. However, this assumption yields a well-defined single parameter β characterizing the effect of s on y, which we may consider the consistency of estimating.

The researcher assumes that the effect of each variable x_j on y is linear and there are no interaction effects among them, and estimates Equation 1. The researcher may also omit some of the necessary control variables: i.e., $k \leq l$. The estimator for β from Equation 1, given Equation 2, is

$$\begin{aligned}
\hat{\beta}_{lin,k} &= (\mathbf{s}' M_{1x_1\dots x_k}\mathbf{s})^{-1}\mathbf{s}' M_{1x_1\dots x_k}\,(\alpha\mathbf{1} + \beta\mathbf{s} + g(\mathbf{x}) + \mathbf{u}) && (3)\\
&= \beta + \frac{\frac{1}{N}\mathbf{s}' M_{1x_1\dots x_k} g(\mathbf{x})}{\hat{var}(s)(1 - \hat{R}^2_{s\cdot 1,x_1\dots x_k})} + \frac{\frac{1}{N}\mathbf{s}' M_{1x_1\dots x_k}\mathbf{u}}{\hat{var}(s)(1 - \hat{R}^2_{s\cdot 1,x_1\dots x_k})}, && (4)
\end{aligned}$$

where boldface denotes vectors of N observations, $M_{1x_1\ldots x_k} = I - X(X'X)^{-1}X'$ with $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k, \mathbf{1})$, and $\mathbf{1} = (1, 1, \ldots 1)'$. Finally, $\hat{R}^2_{s\cdot 1,x_1\ldots x_k}$ denotes the R^2 statistic from a linear regression of s on the variables $x_1 \ldots x_k$ and a constant.

Equation 4 reveals that in a finite sample the estimator error $\hat{\beta}_{lin,k} - \beta$ has two components. The right-most term of Equation 4 is a familiar one from the vector of residuals $\boldsymbol{u}$, which under the strong exogeneity assumption converges in probability to zero as $N \to \infty$. The middle term of Equation 4, however, may not vanish in large samples. Rather, it is proportional to the *partial correlation* of s and $g(x)$, controlling for $x_1 \ldots x_k$ and a constant. The partial correlation between two variables X and Y after controlling for a set of variables Z, which we denote as $\hat{\rho}_{XY\cdot Z}$, is the correlation between the residuals of X regressed on Z and the residuals from Y regressed on Z (Cohen 2003). We use the hat notation $\hat{\rho}_{XY\cdot Z}$ for the partial correlation within a finite sample and $\rho_{XY\cdot Z}$ for its probability limit, and similarly for R^2.[1]

By the definition of partial correlation,

$$\hat{\rho}_{g(x),s\cdot 1x_1\ldots x_k} = \frac{\frac{1}{N}\mathbf{s}' M_{1x_1\ldots x_k} g(\mathbf{x})}{\sqrt{\hat{var}(g(x))(1 - \hat{R}^2_{g(x)\cdot 1,x_1\ldots x_k})}\sqrt{\hat{var}(s)(1 - \hat{R}^2_{s\cdot 1,x_1\ldots x_k})}},$$

where $\hat{var}(\cdot)$ is the sample variance. Using this and exogeneity, Equation 4 implies

$$\hat{\beta}_{lin,k} \xrightarrow{p} \beta + \frac{\sigma_{g(x)}}{\sigma_s}\sqrt{\frac{1 - R^2_{g(x)\cdot 1,x_1\ldots x_k}}{1 - R^2_{s\cdot 1,x_1\ldots x_k}}} \cdot \rho_{g(x),s\cdot 1x_1\ldots x_k}. \tag{5}$$

Letting $\mathbb{B}_{lin,k} = \text{plim}(\hat{\beta}_{lin,k}) - \beta$, we define *linear misspecification bias* from control variables as $\mathbb{B}_{lin,l}$, the second term in Equation 5 when none of the l control variables are omitted. Since $\rho_{g(x),s\cdot 1x_1\ldots x_l}$ is a Pearson correlation coefficient, its absolute value is bounded from above by one, and thus the asymptotic bias is bounded by

$$|\mathbb{B}_{lin,l}| \leq \frac{\sigma_{g(x)}}{\sigma_s}\sqrt{\frac{1 - R^2_{g(x)\cdot 1,x_1\ldots x_l}}{1 - R^2_{s\cdot 1,x_1\ldots x_l}}}. \tag{6}$$

Aside from the intuitive case of $R^2_{g(x)\cdot 1,x_1\ldots x_l}$ near unity, Equation 6 indicates another limit in which linear misspecification bias will be small. If $g(x)$ is much less variable than s, such that $\sigma_{g(x)}/\sigma_s \leq \epsilon$ for some small value ϵ, and the control variables are simply *better* linear predictors of $g(x)$ than they are of s (in the sense of $R^2_{g(x)\cdot 1,x_1\ldots x_l} > R^2_{s\cdot 1,x_1\ldots x_l}$), then $|\mathbb{B}_{lin,l}| < \epsilon$.

In the case when a single control variable is included ($k = 1$), Equation 5 becomes

$$\hat{\beta}_{lin,1} \xrightarrow{p} \beta + \frac{\sigma_{g(x)}}{\sigma_s} \cdot \frac{\rho_{g(x),s} - \rho_{g(x),x}\rho_{s,x}}{1 - \rho^2_{s,x}}. \tag{7}$$

If $g(x)$ is in fact a linear function of the single regressor x, then $\rho_{g(x),s} = \pm\rho_{x,s}$ and $\rho_{g(x),x} = \pm 1$ (where in both cases the sign is that of the slope of $g(x)$), and linear misspecification bias vanishes as expected.

[1] A population partial correlation can be defined from ordinary population correlation coefficients and Equation 9.

2.2. Adding the kth control variable

Consider the bias that would arise when omitting the last control variable x_k entirely, estimating instead a regression with just $x_1 ... x_{k-1}$. Using the identity $1 - R^2_{X \cdot YZ} = \left(1 - R^2_{X \cdot Z}\right)\left(1 - \rho^2_{X,Y \cdot Z}\right)$ for any X, Y, Z (Cohen 2003), we have

$$\hat{\beta}_{lin,k-1} \quad \xrightarrow{p} \quad \beta + \frac{\sigma_{g(x)}}{\sigma_s} \sqrt{\frac{1 - R^2_{g(x) \cdot 1, x_1 ... x_k}}{1 - R^2_{s \cdot 1, x_1 ... x_k}}} \sqrt{\frac{1 - \rho^2_{s, x_k \cdot 1 x_1 ... x_{k-1}}}{1 - \rho^2_{g(x), x_k \cdot 1 x_1 ... x_{k-1}}}} \cdot \rho_{g(x), s \cdot 1 x_1 ... x_{k-1}}, \quad (8)$$

Applying the following recursive formula for partial correlation (Cohen 2003) to Equation 5,

$$\hat{\rho}_{g(x), s \cdot 1 x_1 ... x_k} = \frac{\hat{\rho}_{g(x), s \cdot 1 x_1 ... x_{k-1}} - \hat{\rho}_{g(x), x_k \cdot 1 x_1 ... x_{k-1}} \hat{\rho}_{s, x_k \cdot 1 x_1 ... x_{k-1}}}{\sqrt{1 - \hat{\rho}^2_{s, x_k \cdot 1 x_1 ... x_{k-1}}} \sqrt{1 - \hat{\rho}^2_{g(x), x_k \cdot 1 x_1 ... x_{k-1}}}}, \quad (9)$$

Equation 8 implies that

$$\frac{\mathbb{B}_{lin,k}}{\mathbb{B}_{lin,k-1}} \quad = \quad \frac{1 - \frac{\rho_{g(x), x_k \cdot 1 x_1 ... x_{k-1}} \rho_{s, x_k \cdot 1 x_1 ... x_{k-1}}}{\rho_{g(x), s \cdot 1 x_1 ... x_{k-1}}}}{1 - \rho^2_{s, x_k \cdot 1 x_1 ... x_{k-1}}}. \quad (10)$$

We see from Equation 10 that adding the last variable x_k as a linear control has two effects on the asymptotic bias that exists before x_k is included in the regression: although the numerator of Equation 10 will be small if $\frac{\rho_{g(x), x_k \cdot 1 x_1 ... x_{k-1}} \rho_{s, x_k \cdot 1 x_1 ... x_{k-1}}}{\rho_{g(x), s \cdot 1 x_1 ... x_{k-1}}} \approx 1$, the denominator amplifies any remaining bias by a factor that increases without bound as $\rho^2_{s, x_k \cdot 1 x_1 ... x_{k-1}}$ approaches unity.[2] Whether controlling linearly for x_k (conditional on already controlling linearly for $x_1 \ldots x_{k-1}$) reduces the asymptotic bias in estimating β depends on the relative magnitude of these two effects. That is, $|\mathbb{B}_{lin,k}| < |\mathbb{B}_{lin,k-1}|$ iff

$$\rho^2_{s, x_k \cdot 1 x_1 ... x_{k-1}} < \frac{\rho_{g(x), x_k \cdot 1 x_1 ... x_{k-1}} \rho_{s, x_k \cdot 1 x_1 ... x_{k-1}}}{\rho_{g(x), s \cdot 1 x_1 ... x_{k-1}}} < 2 - \rho^2_{s, x_k \cdot 1 x_1 ... x_{k-1}}. \quad (11)$$

2.3. A special case of zero bias

A special case in which there is no linear misspecification bias for arbitrarily nonlinear $g(x)$ occurs when the conditional expectation of s happens to be linear in the x_j, i.e.,

$$E(s|x_1, x_2 \ldots x_k) = \pi_0 + \pi_1 x_1 + \pi_2 x_2 + \ldots \pi_k x_k \quad (12)$$

for some scalars $\pi_0 \ldots \pi_k$. This condition is assumed in Angrist and Krueger (1999) to derive an average derivative interpretation of $\hat{\beta}_{lin,k}$ when there is nonlinearity and/or heterogeneity in the effect of s. It is an important special case because it occurs whenever s and the x_j are jointly normally distributed (Lindgren et al. 2013), which is a testable condition. To see how Equation 12 implies $\hat{\beta}_{lin,k} \xrightarrow{p} \beta$ in our setting, write

[2] Achen (2005) demonstrates a large linear misspecification bias despite a very mildly nonlinear $g(x)$ through an example in which $\rho_{s,x} \approx 0.88$, and $\frac{\sigma_{g(x)}}{\sigma_s} \approx 7$.

$s_i = \pi_0 + \pi_1 x_1 + \pi_2 x_2 + \ldots \pi_k x_k + \epsilon_i$, where Equation 12 implies that $E(\epsilon_i | x_i) = 0$. Letting $\hat{\epsilon} = M_{1x_1 \ldots x_k}\mathbf{s}$, Equation 4 yields $\mathbb{B}_{lin,k} \propto \frac{1}{N}\mathbf{s}' M_{1x_1\ldots x_k} g(\mathbf{x}) = \frac{1}{N}\hat{\epsilon}' g(\mathbf{x}) \xrightarrow{p} E(\epsilon_i g(x_i)) = E(g(x_i) E(\epsilon_i | x_i)) = 0$.

We also note that in Equation 12, some of the x_j could be powers of some of the other x_j. This suggests that the common technique of adding quadratic or higher powers of control variables in OLS can help reduce linear misspecification bias in either of two ways: it may help provide a good approximation of the nonlinear function $g(x)$, or it may help approximate the nonlinear form of $E(s|x)$ to satisfy Equation 12. As a toy example, suppose $l = 1$ and $g(x) = e^x$, but that $E(s|x) = \pi_0 + \pi_1 x + \pi_2 x^2$. Then, the regression $y = \beta s + \gamma_0 + \gamma_1 x + \gamma_2 x^2 + u$ would consistently estimate β, despite the fact that $E(u|x) \neq 0$.

2.4. Bounding the bias from the data

In this section, we propose a feasible statistic that generally estimates an upper bound on the magnitude of $\mathbb{B}_{lin,k}$:

$$\begin{aligned} \Omega_N = \frac{\frac{1}{N}\mathbf{y}' M_{1s}\mathbf{y} - \frac{1}{N}\mathbf{y}' M_{1sx_1\ldots x_k}\mathbf{y}}{\sigma_s^2 \left(1 - \hat{R}^2_{s\cdot 1, x_1 \ldots x_k}\right)} &\xrightarrow{p} \mathbb{B}^2_{lin,k} + \frac{\sigma^2_{g(x)}\left(R^2_{g(x)\cdot 1, x_1\ldots x_k} - R^2_{g(x)\cdot 1,s}\right)}{\sigma_s^2\left(1 - R^2_{s\cdot 1, x_1\ldots x_k}\right)} \\ &= \mathbb{B}^2_{lin,k}\left(1 + \frac{R^2_{g(x)\cdot 1, x_1\ldots x_k} - R^2_{g(x)\cdot 1,s}}{\left(1 - R^2_{g(x)\cdot 1, x_1 \ldots x_k}\right)\rho^2_{g(x), s\cdot 1, x_1\ldots x_k}}\right). \end{aligned} \tag{13}$$

Provided that the variables $x_1 \ldots x_k$ together provide a better linear predictor of the function $g(x)$ than the variable s does, $R^2_{g(x)\cdot 1, x_1 \ldots x_k}$ will be greater than $R^2_{g(x)\cdot 1, s}$ and the second term of Equation 13 thus positive, implying that Ω_N converges to an upper bound for the squared asymptotic bias $\mathbb{B}^2_{lin,k}$. This condition is not directly testable from the data, but can be expected to hold except in very particular circumstances. Equation 13 may, however, provide a gross overestimate of $\mathbb{B}^2_{lin,k}$, if $\left(1 - R^2_{g(x)\cdot 1, x_1\ldots x_k}\right) \cdot \rho^2_{g(x), s\cdot 1, x}$ is very small, for instance when $g(x)$ is very nearly linear.

3. Monte Carlo simulations

In this section, we use Monte Carlo simulations to analyze the behavior of the estimator $\hat{\beta}_{lin,k}$ for various $g(x)$ and distributions of the stochastic variables. For simplicity and brevity, we focus on the case of a single control variable x.

3.1. Omitting the control variable

First, we demonstrate by example the possibility that linear misspecification bias due to assuming linearity of x can in fact be greater in magnitude than the bias of omitting x from the regression altogether, as suggested by Equation 11. We denote the estimators as $\hat{\beta}_{lin}$ and $\hat{\beta}_{omit}$. We consider the following data generating process (DGP),

which models the correlation between s and x as coming from being jointly influenced by a third variable z:

$$\begin{aligned} s_i &= z_i/2 + \epsilon_{si} \\ x_i &= log(z_i^2) + \epsilon_{xi} \\ y_i &= \alpha + \beta s_i + g(x_i) + u_i, \end{aligned} \tag{14}$$

where $z_i, \epsilon_{si}, \epsilon_{xi}, u_i \sim \mathcal{N}(0, 1)$, and we take $g(x) = x^2, \alpha = 0, \beta = 1$. This DGP results in a correlation between s and x of $\rho_{s,x} \approx .28$, and $\sigma_s \approx 1.1, \sigma_x \approx 2.3$. We generate $1,000$ samples of $N = 10,000$ observations.

As shown in Figure 1, we obtain distributions of $\hat{\beta}_{lin}$ and $\hat{\beta}_{omit}$ that are biased in opposite directions around $\beta = 1$, with means of 2.2 and -0.075, respectively. The magnitude of the bias from misspecifying x as linear is in this case about 13% larger than that of omitting x from the regression altogether.

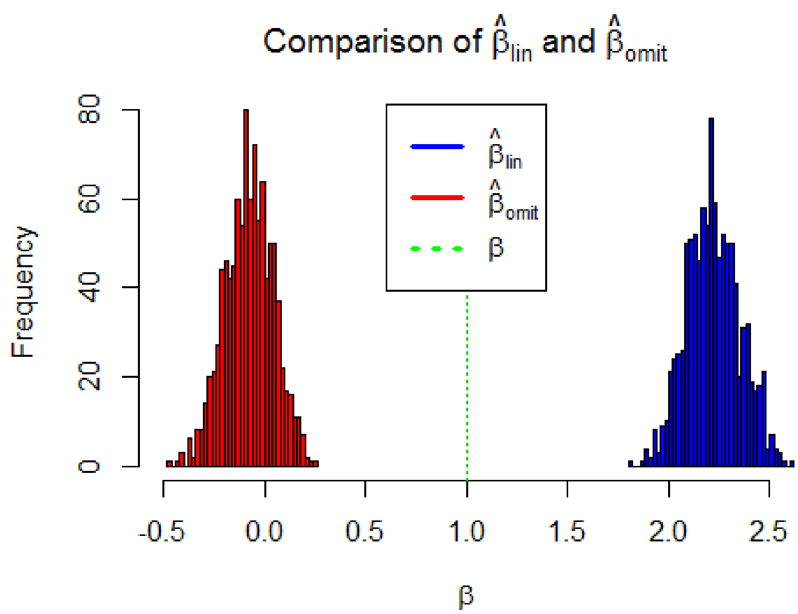

Figure 1: Sampling distributions of $\hat{\beta}_{lin}$ and $\hat{\beta}_{omit}$ for the DGP of Equation 14.

3.2. Effect of the functional form g and the dispersion of x

Lastly, we investigate the effect of various functional forms on the bias of $\hat{\beta}_{lin,l}$. As DGP, we use the bivariate logistic Gumbel distribution $\mathcal{G}(r)$:

$$\begin{pmatrix} s_i \\ x_i \end{pmatrix} = \begin{pmatrix} \tilde{s}_i + 3 \\ \gamma(\tilde{x}_i + 3) \end{pmatrix}, \quad \text{where} \quad \begin{pmatrix} \tilde{s}_i \\ \tilde{x}_i \end{pmatrix} \sim \mathcal{G}(0.7)\,. \tag{15}$$

The correlation between s and x is $\rho_{s,x} \approx .5$, the Gumbel dependence parameter $r = 0.7$, and γ is an overall scale parameter for x. As before, $y_i = \alpha + \beta s_i + g(x_i) + u_i$ with $u_i \sim \mathcal{N}(0, 1)$, $\alpha = 0, \beta = 1$. We again take 1,000 iterations with sample size of 10,000, and compute the mean bias of $\hat{\beta}_{lin}$.

On the basis of a Taylor approximation argument,[3] one may expect linear misspecification bias to be small in situations where σ_x is small, since a linear Taylor series approximation of $g(x)$ may be locally quite good. Thus, to investigate the robustness of each of six functional forms $g(x)$ to increasing σ_x, we sweep through a series of increasing values of σ_x by changing γ, beginning with $\gamma = 1$ and doubling γ iteratively. Figure 2 reveals that even slightly nonlinear functions such as $x^{0.99}$ can have diverging linear misspecification bias as σ_x increases. The linear misspecification bias with $g(x) = \log(x)$ is stable with increasing γ, since rescaling x simply adds a constant to $g(x)$.

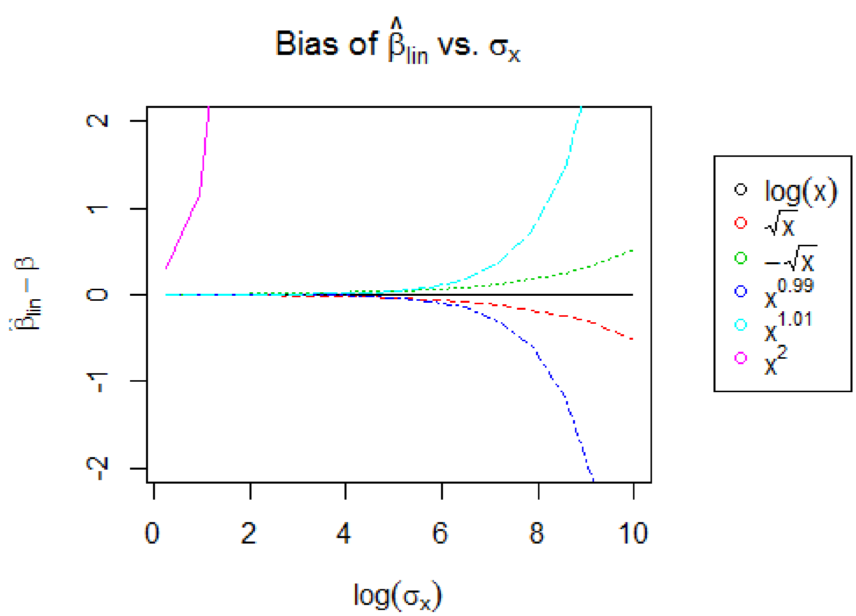

Figure 2: Mean bias vs. $log(\sigma_x)$ for six functional forms $g(x)$, for the DGP of Equation 15.

[3]See White (1980) for a discussion and warning about interpreting OLS as a Taylor approximation to the unknown function $g(x)$.

4. Conclusion

In this paper we have investigated formally the intuitive fact that when control variables in OLS regression are misspecified as linear, the estimator for a single linear variable of interest is generally inconsistent. Since the true function $g(x)$ characterizing the control variables is not directly observable, accounting for it may not be a straightforward task. Semi-parametric estimators for β can offer protection against potential linear misspecification bias,[4] if the number of control variables is not so high as to render them infeasible. Alternatively, "binning" the controls into a set of dummy variables, or parametric nonlinear approximations to $g(x)$ (e.g. including quadratic terms), might help capture its nonlinear form in some cases.

Another approach would be to proceed with the linear specification Equation 1 with some assurance that linear misspecification bias does not cause a significant problem. For example, given the results of Section 2.3, s and the x_j could be tested for joint normality (Cox and Small 1978). Or, Equation 1 might be checked for general specification error with a test such as RESET (Ramsey 1969), before and after adding higher powers of the control variables. However, such tests may not give any assurance that linear misspecification bias is not present, even when it does in fact vanish or is very small. In this case, our results in Section 2.4 may be helpful if the practitioner is confident that $R^2_{g(x)\cdot 1,x_1\ldots x_k} > R^2_{g(x)\cdot 1,s}$ and Ω_N evaluates to a tolerably small possible bias.

References

Achen CH. Let's put garbage-can regressions and garbage-can probits where they belong. Conflict Management and Peace Science 2005;22(4):327–39. URL: http://cmp.sagepub.com/content/22/4/327.abstract. doi:10.1080/07388940500339167.

Angrist JD, Krueger AB. Empirical strategies in labor economics. In: Ashenfelter O, Card D, editors. Handbook of Labor Economics. Elsevier; volume 3; 1999. p. 1277–366. URL: http://ideas.repec.org/h/eee/labchp/3-23.html.

Cohen J. Applied Multiple Regression/correlation Analysis for the Behavioral Sciences. Number v. 1. Routledge, 2003. URL: http://books.google.com/books?id=fuq94a8C0ioC.

Cox DR, Small NJH. Testing multivariate normality. Biometrika 1978;65(2):263–72. URL: http://www.jstor.org/stable/2335204.

Lindgren G, Rootzén H, Sandsten M. Stationary Stochastic Processes for Scientists and Engineers. Chapman and Hall/CRC Taylor & Francis Group, 2013.

[4]Robinson (1988) exploits the fact that together Equation 2 and $E(u_i|s_i, x_{i1}\ldots x_{il}) = 0$ imply that $y_i - E(y_i|x_{i1}\ldots x_{il}) = \beta(s_i - E(s_i|x_{i1}\ldots x_{il})) + u_i$, and relies on flexible estimation of $E(s_i|x_{i1}\ldots x_{il})$ and $E(y_i|x_{i1}\ldots x_{il})$. A method by Yatchew (1997) relies on first-differencing the data after sorting by x, and relying on $g(x)$ having a finite first-derivative.

Ramsey JB. Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society Series B (Methodological) 1969;31(2):350–71. URL: http://www.jstor.org/stable/2984219.

Robinson P. Root-n-consistent semiparametric regression. Econometrica 1988;56(4):931–54.

White H. Using least squares to approximate unknown regression functions. International Economic Review 1980;21(1):149–70. URL: http://EconPapers.repec.org/RePEc:ier:iecrev:v:21:y:1980:i:1:p:149-70.

Yatchew A. An elementary estimator of the partial linear model. Economics Letters 1997;57(2):135–43. URL: http://www.sciencedirect.com/science/article/pii/S0165176597002188. doi:http://dx.doi.org/10.1016/S0165-1765(97)00218-8.