Multiple Regression: Omitted Variable Bias

Dr. Patrick Toche

Textbook:

    James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

    Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.

    Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Problems and Applications

A researcher plans to study the causal effect of police on crime, using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the county's police force.

a. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?

b. Use the previous answer to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?

Problems and Applications

A researcher plans to study the causal effect of police on crime, using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the county's police force.

a. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?

b. Use the previous answer to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?

Problems and Applications

a. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?

A county's crime rate depends on a host of factors other than the size of the police force. For instance, the average level of income, the distribution of income, the unemployment rate, and the level of education, among others. Furthermore, the size of the police force may be a poor proxy for police effectiveness: the level of education in the police force, the quality of police training (much below international standards in the United States), racial diversity (to reduce wrongful arrests), police force salaries (higher salaries would attract better people to join the police force), police resources (better cars, better computers, better labs, better sniffer dogs, etc.). Leaving these out essentially "forces" the size variable to explain the crime rate even in situations where other factors would be more important, thus producing biased estimates.

Problems and Applications

a. Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important omitted variables?

A county's crime rate depends on a host of factors other than the size of the police force. For instance, the average level of income, the distribution of income, the unemployment rate, and the level of education, among others. Furthermore, the size of the police force may be a poor proxy for police effectiveness: the level of education in the police force, the quality of police training (much below international standards in the United States), racial diversity (to reduce wrongful arrests), police force salaries (higher salaries would attract better people to join the police force), police resources (better cars, better computers, better labs, better sniffer dogs, etc.). Leaving these out essentially "forces" the size variable to explain the crime rate even in situations where other factors would be more important, thus producing biased estimates.

Problems and Applications

b. Use the previous answer to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?

The regression is likely to overestimate β_1, that is $\hat{\beta}_1 > \beta_1$. Consider the formula for the omitted variable bias:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \cdot \sigma_u/\sigma_X$$

Since we expect crime rates to vary across counties irrespective of policing, we can expect that counties with higher crime rates would tend to respond by having a larger police force. Since many of the potentially omitted variables would be positively associated with the crime rate, we can expect that the size of the police force would be positively correlated with the error term, $\rho_{Xu} > 0$s, implying a potential overestimate of $\hat{\beta}_1$. Since the true coefficient is likely to be negative $\beta_1 < 0$ (greater police force causes a smaller crime rate), it follows that an over-estimate of $\hat{\beta}_1$ would result in an under-estimate of the effect of police on crime. In practice, this intuition would need to be verified with the data by computing ρ_{Xu}. To sum up, we expect:

$$\hat{\beta}_1 > \beta_1 \implies |\hat{\beta}_1| < |\beta_1|$$

Problems and Applications

b. Use the previous answer to determine whether the regression will likely over- or underestimate the effect of police on the crime rate. That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?

The regression is likely to overestimate β_1, that is $\hat{\beta}_1 > \beta_1$. Consider the formula for the omitted variable bias:

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \cdot \sigma_u / \sigma_X$$

Since we expect crime rates to vary across counties irrespective of policing, we can expect that counties with higher crime rates would tend to respond by having a larger police force. Since many of the potentially omitted variables would be positively associated with the crime rate, we can expect that the size of the police force would be positively correlated with the error term, $\rho_{Xu} > 0$s, implying a potential overestimate of $\hat{\beta}_1$. Since the true coefficient is likely to be negative $\beta_1 < 0$ (greater police force causes a smaller crime rate), it follows that an over-estimate of $\hat{\beta}_1$ would result in an under-estimate of the effect of police on crime. In practice, this intuition would need to be verified with the data by computing $\rho_{X\hat{u}}$. To sum up, we expect:

$$\boldsymbol{\hat{\beta}_1 > \beta_1 \implies |\hat{\beta}_1| < |\beta_1|}$$