

## Omitted Variables Problem

**True Model:**

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_r X_{rt} + \beta_{r+1} X_{r+1t} + \cdots + \beta_K X_{Kt} + \varepsilon_t$$

**Misspecified Model:**

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_r X_{rt} + u_t$$

where  $u_t = \beta_{r+1} X_{r+1t} + \cdots + \beta_K X_{Kt} + \varepsilon_t$

**Consider the regressions of the excluded variables on all of the included variables (in deviation form):**

$$x_{r+1t} = \gamma_{r+1,1,2,3,\dots,r} x_{1t} + \gamma_{r+1,2,1,3,\dots,r} x_{2t} + \cdots + \gamma_{r+1,r,1,2,\dots,r-1} x_{rt} + v_{r+1t}$$

.  
.  
.

$$x_{Kt} = \gamma_{K,1,2,3,\dots,r} x_{1t} + \gamma_{K,2,1,3,\dots,r} x_{2t} + \cdots + \gamma_{K,r,1,2,\dots,r-1} x_{rt} + v_{Kt}$$

where for example,  $\gamma_{r+1,1,2,3,\dots,r}$  is the regression coefficient on  $x_{1t}$  from the regression of  $x_{r+1t}$  on  $x_{1t}, x_{2t}, \dots, x_{rt}$ .

**It can be shown that**

$$E(\tilde{\beta}_1) = \beta_1 + \beta_{r+1} \gamma_{r+1,1,2,3,\dots,r} + \beta_{r+2} \gamma_{r+2,1,2,3,\dots,r} + \cdots + \beta_K \gamma_{K,1,2,3,\dots,r}$$

.  
.  
.

$$E(\tilde{\beta}_r) = \beta_r + \beta_{r+1} \gamma_{r+1,r,1,2,\dots,r-1} + \beta_{r+2} \gamma_{r+2,r,1,2,\dots,r-1} + \cdots + \beta_K \gamma_{K,r,1,2,3,\dots,r-1}$$

**Bias Expressions:**

$$E(\tilde{\beta}_1) - \beta_1 = \underbrace{\beta_{r+1}\gamma_{r+1,1,2,3,\dots,r} + \beta_{r+2}\gamma_{r+2,1,2,3,\dots,r} + \dots + \beta_K\gamma_{K,1,2,3,\dots,r}}_{\text{Bias}}$$

·  
·  
·

$$E(\tilde{\beta}_r) - \beta_r = \underbrace{\beta_{r+1}\gamma_{r+1,r,1,2,\dots,r-1} + \beta_{r+2}\gamma_{r+2,r,1,2,\dots,r-1} + \dots + \beta_K\gamma_{K,r,1,2,3,\dots,r-1}}_{\text{Bias}}$$