# Tringular Systems and Instrumental Variables

Bryan S. Graham, UC - Berkeley & NBER

February 2, 2022

Consider the following statistical model

$$\begin{aligned} Y &= \alpha_0 + \beta_0 X + Z_1' \gamma_0 + U & (1) \\ X &= \eta_0 + Z' \pi_0 + V & (2) \end{aligned}$$

where $Z = (Z_1', Z_2')'$ and

$$\left( \begin{array}{c} U \\ V \end{array} \right) \Bigg| Z \sim \mathcal{N} \left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} \sigma_U^2 & \rho \sigma_U \sigma_V \\ \rho \sigma_U \sigma_V & \sigma_V^2 \end{array} \right) \right). \tag{3}$$

Acemoglu et al. (2001) work with a model similar to the one defined by (1), (2) and (3). The focus of their study is the causal, or structural, effect of institutions on long-run economic development. They work with a sample of former European colonies ($N = 64$), where $Y$, is logarithm of per capita GDP, $X$, a measure of institutional quality (specifically of security of property rights), $Z_1$, includes additional country-specific characteristics (e.g., latitude/distance from the equator), and $Z_2$, is the logarithm of the 19$^{\text{th}}$ century European settler mortality rate. For the purposes of exposition it is helpful to reify the unobserved disturbances appearing in equations (1) and (2). We will call $U$ unobserved productivity and $V$ a country's 'taste for lawfulness'. Acemoglu, Johnson and Robinson (2000) are interested in the parameter $\beta_0$, which is the semi-elasticity of GDP per capita with respect to institutional quality (holding observed country characteristics $Z_1$ and unobserved productivity $U$ constant). That is, the percentage effect on GDP per capita of a unit increase in institutional quality, $X$.

Equation (3) contains a lot of information. First, note that it specifies a distribution for $(U, V)$ conditional on $Z$. However, $Z$ does not enter this conditional distribution anywhere. Therefore (3) implies that $Z$ is independent of $(U, V)$. This feature of the model will be exploited below. Second, $U$ and $V$ are correlated if $\rho \neq 0$. Therefore countries with a high 'taste for lawfulness' (high $V$) may systematically differ in terms of unobserved productivity, $U$, from those countries with a low 'taste for lawfulness' (low $V$). This covariance implies that $X$ is a (right-hand-side) **endogenous variable**. A third important feature of the model is embedded in (1) and (2): while $Z_2$ enters the **first stage equation** describing institutional quality, it does not directly enter the **structural equation** describing GDP. The combination of independence of $Z$ and $U$, and the fact that $Z_2$ does not directly enter it means that $Z_2$ is **excludable** from equation (1). This implies that the entire effect of settler mortality, $Z_2$, on current per capita GDP levels, $Y$, is mediated through the latter's influence on institutions, $X$. Acemoglu et al. (2001) argue that colonies with high settler mortality rates were more 'extractive' in nature. This, in turn, led to a path of institutional development that was less focused on property rights. In contrast, colonies with low mortality rates became settler colonies or 'neo-Europes' that imported pro-property rights institutions.

## Estimation by least squares

Assume $\dim(Z_1) = 0$ and let $Z = Z_2$ for simplicity. Now consider the coefficient on $X$ in the linear predictor of $Y$ given $X$. We have, by our formula for the bivariate linear predictor coefficient,

$$
\begin{aligned}
b_{LP} &= \tfrac{\mathbb{C}(X,Y)}{\mathbb{V}(X)} & \\
&= \tfrac{\mathbb{C}(X,\beta_0 X+U)}{\mathbb{V}(X)} & \text{Equation (1)} \\
&= \beta_0 \tfrac{\mathbb{C}(X,X)}{\mathbb{V}(X)} + \tfrac{\mathbb{C}(X,U)}{\mathbb{V}(X)} & \text{Linearity of } \mathbb{C}(\cdot,\cdot) \\
&= \beta_0 + \tfrac{\mathbb{C}(Z'\pi_0+V,U)}{\mathbb{V}(Z'\pi_0+V)} & \mathbb{C}(X,X) = \mathbb{V}(X),\ \text{Equation (2)} \\
&= \beta_0 + \tfrac{\mathbb{C}(Z'\pi_0,U)+\mathbb{C}(V,U)}{\mathbb{V}(Z'\pi_0)+\mathbb{V}(V)+2\mathbb{C}(Z'\pi_0,V)} & \text{Expansion of covariance/variance} \\
&= \beta_0 + \tfrac{\mathbb{C}(V,U)}{\mathbb{V}(Z'\pi_0)+\mathbb{V}(V)} & \text{Independence of } Z \text{ and } (U,V) \\
&= \beta_0 + \tfrac{\mathbb{C}(V,U)}{\mathbb{V}(V)} \tfrac{1}{\frac{\mathbb{V}(Z'\pi_0)}{\mathbb{V}(V)}+1} & \text{Rearrangement} \\
&= \beta_0 + \rho \tfrac{\sigma_U}{\sigma_V} \tfrac{1}{\frac{\pi_0'\mathbb{V}(Z)\pi_0}{\mathbb{V}(V)}+1} & \mathbb{V}(Z'\pi_0) = \pi_0'\mathbb{V}(Z)\pi_0 \\
&= \beta_0 + \rho \tfrac{\sigma_U}{\sigma_V} \tfrac{1}{\mu^2+1} & \text{Rearrangement,}
\end{aligned}
$$

where

$$\mu^2 = \frac{\pi_0'\mathbb{V}(Z)\pi_0}{\mathbb{V}(V)},$$

is closely related to something called the **concentration parameter** in the econometrics literature. This parameter measures the ratio of the explained variance in $X$ to the unexplained variance in $X$.

 

The above derivation shows that the linear predictor coefficient on $X$ differs from the parameter of interest. In particular we have a bias of

$$b_{LP} - \beta_0 = \rho \frac{\sigma_U}{\sigma_V} \frac{1}{\mu^2 + 1}. \tag{4}$$

This bias depends on a variety of features of the underlying data generating process. First, if $\rho = 0$ there is no bias. In that case the unobserved factors which determine institutional quality do not covary with the unobserved factors which drive long run GDP per capita. However, in practice we might expect the two to covary. Assume that $\rho > 0$ so that countries with a high 'taste for lawfulness' also tend to be more productive. In that case those countries with high realizations of $V$ will ten to have both high realizations of $X$ (see equation (2)) *and* high realizations of $U$. This means that $X$ and $U$ will positively covary in the population. If those countries with high levels of institutional quality tend also to be productive for other reasons, then the covariance between $X$ and $Y$ will reflect a combination of the structural effect of interest and the fact that high levels of institutional quality are associated with other things 'good' for GDP. The latter effects induce the above bias.

A second feature of the bias expression (4) is that it is decreasing in $\mu^2$. Note that the explained variance in $X$ is driven by $Z$. Since $Z$ is independent of $U$, the more of the variation in $X$ that is driven by $Z$, the less bias.

# Estimation by control function (CF) methods

Our linear predictor coefficient is biased for $\beta_0$ because countries with a high 'taste for lawfulness' tend to be more productive for reasons other than their good institutions. This suggests that perhaps we should directly 'control for' a country's 'taste for lawfulness' in our analysis. By equations (2) and (3), holding $V$ fixed all variation in $X$ is driven by $Z$, which varies independently of $U$. Consider the linear predictor regression function

$$\mathbb{E}^* [Y|X, V] = \alpha_0 + X'\beta_0 + \mathbb{E}^* [U|X, V].$$

 

Manipulating the third term to the right of the equality we have

$$\begin{aligned}
\mathbb{E}^*[U|X,V] &= \mathbb{E}^*[\mathbb{E}^*[U|X,Z,V]|X,V] && \text{Law of iterated linear predictors} \\
&= \mathbb{E}^*[\mathbb{E}^*[U|Z,V]|X,V] && X \text{ is colinear with } Z \text{ and } V \\
&= \mathbb{E}^*[\mathbb{E}^*[U|V]|X,V] && (U,V) \text{ is independent of } Z \\
&= \mathbb{E}^*\left[\rho\tfrac{\sigma_U}{\sigma_V}V\middle| X,V\right] && \begin{array}{c}\text{Definition of linear predictor} \\ \text{(note } U \text{ and } V \text{ are mean zero)}\end{array} \\
&= \rho\tfrac{\sigma_U}{\sigma_V}\mathbb{E}^*[V|X,V] && \text{Bring out the constant} \\
&= \rho\tfrac{\sigma_U}{\sigma_V}V && \text{Property of LP} \\
&= \phi_0 V && \text{Definition}
\end{aligned}$$

for $\phi_0 = \rho\frac{\sigma_U}{\sigma_V}$. Putting things together we have

$$\mathbb{E}^*[Y|X,V] = \alpha_0 + \beta_0 X + \phi_0 V.$$

So our augmented linear predictor does indeed identify the structural effect of interest. The intuition for this result is that conditional on $V$, a country's taste for lawfulness, variation in institutional quality across countries is driven by $Z$ and hence is idiosyncratic or exogenous (since $Z$ is independent of $U$).

**Excludability** of $Z$ from (1) was central to getting the above result. A second important, but less transparent, assumption for getting a usable result is that $Z$ actually induces variation in $X$ (i.e., that $\pi_0 \neq 0$). If $\pi_0 = 0$ then our control function regression is

$$\begin{aligned}
\mathbb{E}^*[Y|X,V] &= \alpha_0 + \beta_0 X + \phi_0 V \\
&= \alpha_0 + \beta_0 X + \phi_0 (X - Z'\pi_0) \\
&= \alpha_0 + (\beta_0 + \phi_0) X \\
&= \mathbb{E}^*[Y|X],
\end{aligned}$$

since when $\pi_0 = 0$ we have $\mu^2 = 0$ and therefore $b_{LP} = \beta_0 + \rho\frac{\sigma_U}{\sigma_V}\frac{1}{\mu^2+1} = \beta_0 + \rho\frac{\sigma_U}{\sigma_V} = \beta_0 + \phi_0$. So a second requirement is a **first stage relationship** between $Z$ and $X$.

In practice $V$ is unobserved, but it may be estimated. This observation leads to the following feasible procedure for estimating $\beta_0$.

1. Compute the least squares fit of $X$ onto $Z$. Compute $\widehat{V}_i = X_i - \hat{\eta} - Z_i'\widehat{\pi}$ for $i = 1, \ldots, N$.

2. Compute the least squares fit of $Y$ on 1, $X$ and $\widehat{V}$. Use the coefficient on $X$ as the estimate of $\beta_0$, say $\widehat{\beta}_{CF}$.

 

This procedure is simple to implement, but it turns out that the sampling error in $\widehat{V}$ affects that in $\widehat{\beta}_{CF}$. Therefore our normal least squares variance estimator will be inconsistent. While this problem can be corrected, we will instead pursue a different approach to estimation.

## Estimation by instrumental variable (IV) methods

The control function approach to identifying $\beta_0$ is useful conceptually, but in the current set-up a simple **instrumental variables (IV)** estimator is available. Let $\psi(W,\theta) = Z(Y - R'\theta)$ with $\theta = (\alpha, \beta', \gamma')'$, $R = (1, X', Z_1')'$, $W = (X', Y, Z')'$ and observe that

$$\mathbb{E}[\psi(W,\theta_0)] = \mathbb{E}[Z(Y - R'\theta_0)] = \mathbb{E}[ZU] = 0, \tag{5}$$

where the last follows from independence of $Z$ and $U$ (at this stage it is convenient to assume that $Z$ includes a constant as well as $Z_1$ and $Z_2$).

Solving (5) for $\theta_0$ yields (assuming that $K = \dim(R) = \dim(Z) = J$)

$$\theta_0 = \mathbb{E}[ZR']^{-1} \times \mathbb{E}[ZY].$$

For $\theta_0$ to be well-defined we require that $\mathbb{E}[ZR']$ be invertible. How is this related to the requirement that $\pi_0 \neq 0$?

An analog estimator is therefore

$$\widehat{\theta}_{IV} = \left[\frac{1}{N}\sum_{i=1}^{N} Z_i R_i'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} Z_i Y_i\right]. \tag{6}$$

Substituting in for $Y_i$ and rearranging yields

$$\sqrt{N}\left(\widehat{\theta}_{IV} - \theta_0\right) = \left[\frac{1}{N}\sum_{i=1}^{N} Z_i R_i'\right]^{-1} \times \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i U_i\right].$$

By the law of large numbers and the central limit theory we respectively have

$$\frac{1}{N}\sum_{i=1}^{N} Z_i R_i' \xrightarrow{p} \Gamma_0$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i U_i \xrightarrow{D} \mathcal{N}(0,\Omega_0)$$

 

where $\Gamma_0 = \mathbb{E}[ZR']$ and

$$\Omega_0 = \mathbb{E}[ZUU'Z'] = \mathbb{E}\left[\psi(W,\theta_0)\,\psi(W,\theta_0)'\right] = \mathbb{V}(\psi(W,\theta_0)).$$

Slutsky's Theorem then gives

$$\sqrt{N}\left(\widehat{\theta}_{IV} - \theta_0\right) \xrightarrow{D} \mathcal{N}(0,\Lambda_0), \qquad \Lambda_0 = \left(\Gamma_0'\Omega_0^{-1}\Gamma_0\right)^{-1}.$$

For conducting inference we replace $\Lambda_0$ with the estimate

$$\widehat{\Lambda} = \left(\widehat{\Gamma}'\widehat{\Omega}^{-1}\widehat{\Gamma}\right)^{-1},$$

where

$$\widehat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N} Z_i R_i', \qquad \widehat{\Omega} = \frac{1}{N}\sum_{i=1}^{N} Z_i \widehat{U}_i \widehat{U}_i' Z_i'$$

with $\widehat{U}_i = Y_i - R_i'\widehat{\theta}_{IV}$.

## Estimation of overidentified models

In many settings the number of excludable instruments exceeds the number of right-hand-side endogenous variables. This will, in turn, imply that $K = \dim(R) < \dim(Z) = J$. When there are more instruments than regressors we cannot set the sample average of $\psi(W_i,\widehat{\theta}_{IV})$ exactly equal to zero by choice of $\widehat{\theta}_{IV}$. We only have $\dim(\theta) = K$ free parameters to vary, but a total of $\dim(\psi(W,\theta)) = J$ moment conditions to satisfy.

Although we cannot choose $\widehat{\theta}_{IV}$ such that $\frac{1}{N}\sum_{i=1}^{N}\psi(W_i,\widehat{\theta}_{IV})$ equals a vector of zeros, we can choose $\widehat{\theta}_{IV}$ to set a *linear combination* of the $J$ sample moments equal to zero. Let $\widehat{C}$ be a $K \times J$ **weight matrix**. The notation emphasizes the fact that $\widehat{C}$ may be estimated using the data, we assume however that $\widehat{C} \xrightarrow{p} C_0$. We choose $\widehat{\theta}_{IV}$ such that

$$\begin{aligned} \widehat{C}\left[\frac{1}{N}\sum\nolimits_{i=1}^{N}\psi(W_i,\widehat{\theta}_{IV})\right] &= \widehat{C}\left[\frac{1}{N}\sum\nolimits_{i=1}^{N} Z_i Y_i - \frac{1}{N}\sum\nolimits_{i=1}^{N} Z_i R_i'\widehat{\theta}_{IV}\right] \\ &= 0, \end{aligned}$$

solving for $\widehat{\theta}_{IV}$ we have

$$\widehat{\theta}_{IV} = \left[\widehat{C}\left(\frac{1}{N}\sum\nolimits_{i=1}^{N} Z_i R_i'\right)\right]^{-1}\left[\widehat{C}\left(\frac{1}{N}\sum\nolimits_{i=1}^{N} Z_i Y_i\right)\right],$$

where we require that $\widehat{C}\left(\frac{1}{N}\sum_{i=1}^{N}Z_iR_i'\right)$ is non-singular.

To derive the large sample distribution of this estimator we proceed as before. Substituting in for $Y_i$ and rearranging yields

$$\sqrt{N}\left(\widehat{\theta}_{IV}-\theta_0\right)=\left[\widehat{C}\left(\frac{1}{N}\sum\nolimits_{i=1}^{N}Z_iR_i'\right)\right]^{-1}\widehat{C}\left[\left(\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}Z_iU_i\right)\right].$$

The law of large numbers, the central limit theorem and Slutsky's Theorem imply that

$$\left[\widehat{C}\left(\frac{1}{N}\sum\nolimits_{i=1}^{N}Z_iR_i'\right)\right]^{-1}\widehat{C}\xrightarrow{p}\left(C_0\Gamma_0\right)^{-1}C_0$$
$$\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}Z_iU_i\xrightarrow{D}\mathcal{N}\left(0,\Omega_0\right).$$

Combining these results using Slutsky's Theorem gives

$$\sqrt{N}\left(\widehat{\theta}_{IV}-\theta_0\right)\xrightarrow{D}\mathcal{N}(0,\Lambda_0),\qquad\Lambda_0=\Delta_0\Omega_0\Delta_0',$$

where $\Delta_0=\left(C_0\Gamma_0\right)^{-1}C_0$ and

$$\underset{J\times K}{\Gamma_0}=E\left[ZR'\right],\qquad\underset{J\times J}{\Omega_0}=\mathbb{E}\left[\psi\left(W,\theta_0\right)\psi\left(W,\theta_0\right)'\right]=E\left[ZUU'Z'\right].$$

To conduct inference we replace $\Lambda_0$ with an estimate.

## Two-Stage Least Squares

Let $R$ be a $K\times 1$ vector and $Z$ a $J\times 1$ vector as above. Let $\mathbb{E}^*\left[R_k|Z\right]=Z'\pi_k$ be the best linear predictor of the $k^{th}$ element of $R$ given $Z$. Consider the $K\times J$ matrix

$$\Pi=\begin{pmatrix}\pi_1'\\ \vdots\\ \pi_K'\end{pmatrix},$$

then

$$\mathbb{E}^*\left[R|\,Z\right]=\Pi X=\begin{pmatrix}\pi_1'Z\\ \vdots\\ \pi_K'Z\end{pmatrix},\qquad V=R-\Pi X$$

is the **multivariate linear predictor** of $R$ given $Z$. The multivariate LP simply stacks up the $k=1,\ldots,K$ univariate LPs. It is a useful exercise to show that $\Pi = \mathbb{E}\left[RZ'\right]\mathbb{E}\left[ZZ'\right]^{-1}$. Now consider the instrumental variables estimator based on the estimated weight matrix

$$\begin{aligned}\widehat{C} &= \left[\frac{1}{N}\sum\nolimits_{i=1}^{N} R_i Z_i'\right]\left[\frac{1}{N}\sum\nolimits_{i=1}^{N} Z_i Z_i'\right]^{-1} \\ &= \widehat{\Pi}.\end{aligned}$$

Note that $\widehat{\Pi}$ is the $K \times J$ matrix of least squares coefficients associated with the least squares regression of *each* of the $K$ elements in $R$ onto the $J$ instruments, $Z'$. That is $\widehat{\Pi}$ estimates $\Pi$ as defined above.

We can show that this particular instrumental variables estimator is equivalent to what is called the **two-stage least squares (TSLS)** estimator. We have

$$\begin{aligned}
\widehat{\theta}_{IV} &= \left[\widehat{C}\left(\tfrac{1}{N}\sum_{i=1}^{N} Z_i R_i'\right)\right]^{-1}\left[\widehat{C}\left(\tfrac{1}{N}\sum_{i=1}^{N} Z_i Y_i\right)\right] && \text{Definition of IV} \\
&= \left[\widehat{\Pi}\left(\tfrac{1}{N}\sum_{i=1}^{N} Z_i R_i'\right)\right]^{-1}\left[\widehat{\Pi}\left(\tfrac{1}{N}\sum_{i=1}^{N} Z_i Y_i\right)\right] && \text{Definition of } \widehat{C} \\
&= \left[\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i \left(\Pi Z_i + V_i\right)'\right]^{-1}\left[\left(\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i Y_i\right)\right] && \text{Substitute } \Pi Z_i + V_i \text{ for } R_i \\
&= \left[\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i \left(\widehat{\Pi} Z_i - \left(\widehat{\Pi} - \Pi\right) Z_i + V_i\right)'\right]^{-1}\left[\left(\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i Y_i\right)\right] && \text{Add and subtract } \widehat{\Pi} Z_i \\
&= \left[\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i \left(\widehat{\Pi} Z_i + \widehat{V}_i\right)'\right]^{-1}\left[\left(\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i Y_i\right)\right] && \begin{aligned}\widehat{V}_i &= R_i - \widehat{\Pi} Z_i \\ &= \Pi Z_i + V_i - \widehat{\Pi} Z_i\end{aligned} \\
&= \left[\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i Z_i' \widehat{\Pi}'\right]^{-1}\left[\left(\tfrac{1}{N}\sum_{i=1}^{N} \widehat{\Pi} Z_i Y_i\right)\right] && \tfrac{1}{N}\sum_{i=1}^{N} Z_i \widehat{V}_i' = 0 \\
&= \left[\left(\tfrac{1}{N}\sum_{i=1}^{N} \widehat{R}_i \widehat{R}_i'\right)\right]^{-1}\left[\left(\tfrac{1}{N}\sum_{i=1}^{N} \widehat{R}_i Y_i\right)\right] = \widehat{\theta}_{TSLS} && \text{Substitute } \widehat{R}_i \text{ for } \widehat{\Pi} Z_i
\end{aligned}$$

where $\widehat{R}_i = \widehat{\Pi} Z_i$ are the first stage fitted values. So our IV estimator is numerically equivalent to the following two-step procedure:

1. Perform $K$ least squares fits of each element of $R$ onto $Z$ and form the vector of fitted values, $\widehat{R}_i = \widehat{\Pi} Z_i$.
2. Compute the least square fit of $Y_i$ onto $\widehat{R}_i$.

# References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369 – 1401.

 