Math Review Notes—Causal Inference and Econometrics

Gregory Faletto

Contents

Last updated August 12, 2020

Chapter 1

Causal Inference and Econometrics

1.1 Generalized Method of Moments (Chapter 13 of Hansen [2020])

1.1.1 Overidentified Moment Equations (Section 13.4 of Hansen [2020])

Consider the instrumental variables model (see Section 1.2.2). The estimator $\hat{\beta}$ is the solution of the moment condition

$$\overline{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i(Y_i - X_i^\top \beta) = \frac{1}{n}\left(\boldsymbol{Z}^\top \boldsymbol{Y} - \boldsymbol{Z}^\top \boldsymbol{X}\boldsymbol{\beta}\right).$$

If this model is overidentified (that is, the number of instruments ℓ—and therefore moment conditions to satisfy—exceeds the number of variables p in $\boldsymbol{X}$—and therefore the number of parameters to estimate in $\boldsymbol{\beta}$), in general this estimator does not exist, so the method of moments estimator is not defined.

The idea of the generalized method of moments estimator is to make $\overline{g}_n(\beta)$ as close to zero as possible. Define the vector $\boldsymbol{\mu} := \boldsymbol{Z}^\top \boldsymbol{Y} \in \mathbb{R}^\ell$, the matrix $\boldsymbol{G} := \boldsymbol{Z}^\top \boldsymbol{X} \in \mathbb{R}^{\ell \times p}$, and the "error" $\boldsymbol{\eta} := \boldsymbol{\mu} - \boldsymbol{G}\boldsymbol{\beta}$. Then we can write the finite-sample analogue of the above equation as

$$\begin{aligned} & \boldsymbol{Z}^\top \boldsymbol{Y} = \boldsymbol{Z}^\top \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta} \\ \iff \quad & \boldsymbol{\mu} = \boldsymbol{G}\boldsymbol{\beta} + \boldsymbol{\eta}. \end{aligned}$$

Therefore the least squares estimator (if we take all moment conditions to be equally important) is $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{G}^\top \boldsymbol{G}\right)^{-1} \boldsymbol{G}^\top \boldsymbol{\mu}$. In general, we may want to weigh some moment conditions as more important than others (possibly because errors are non-homogeneous, in which case this increases efficiency). Then by analogy to weighted least squares (see Section **??**), for some positive definite weight matrix $\boldsymbol{W}$ we have the **generalized method of moments estimator**

$$\hat{\boldsymbol{\beta}} := \left(\boldsymbol{G}^\top \boldsymbol{W} \boldsymbol{G}\right)^{-1} \boldsymbol{G}^\top \boldsymbol{W} \boldsymbol{\mu} = \left(\boldsymbol{X}^\top \boldsymbol{Z} \boldsymbol{W} \boldsymbol{Z}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{Z} \boldsymbol{W} \boldsymbol{Z}^\top \boldsymbol{Y}. \tag{1.1}$$

This minimizes the weighted sum of squares $\boldsymbol{\eta}^\top \boldsymbol{W} \boldsymbol{\eta}$.

Definition 1.1.1 (**Generalized Method of Moments estimator; Definition 13.1 in Hansen [2020]**)**.** For a positive definite square weight matrix $\boldsymbol{W}$, define the GMM criterion function

$$J(\boldsymbol{\beta}) := n \overline{g}_n(\boldsymbol{\beta})^\top \boldsymbol{W} \overline{g}_n(\boldsymbol{\beta}). \tag{1.2}$$

Then the **generalized method of moments estimator** is

$$\hat{\boldsymbol{\beta}}_{\text{gmm}} := \arg\min_{\beta} \left\{ J_n(\boldsymbol{\beta}) \right\}.$$

Note that GMM includes the method of moments estimator as a special case. This implies that all results for GMM apply to any method of moments estimators. In this case $\boldsymbol{W}$ does not matter. In the overidentified case, the choice of $\boldsymbol{W}$ is important.

1.2 Instrumental Variables (Section 4.8 of Cameron and Trivedi [2005])

1.2.1 Inconsistency of OLS and Examples of Endogeneity (Section 4.8.1 of Cameron and Trivedi [2005], Section 12.3 in Hansen [2020])

- **Measurement error in the regressor.** Suppose $\mathbb{E}[Y \mid Z] = Z^\top \beta$, but Z is not observed; instead, $X = Z + u$ is observed, where u is measurement error with $\mathbb{E}(u) = 0$ and u is independent of e and Z. We have

  $$Y = Z^\top \beta + e = (X - u)^\top \beta + e = X^\top \beta + \nu$$

  where $\nu = e - u^\top \beta$. Therefore

  $$Y = X^\top \beta + \nu,$$

  but

  $$\mathbb{E}[X\nu] = \mathbb{E}[(Z + u)(e - u^\top \beta)] = -\mathbb{E}[uu^\top]\beta \neq 0.$$

  Therefore least squares estimation is inconsistent, and X is endogenous. The projection coefficient (the quantity least squares is consistent for) is (in the case $p = 1$)

  $$\beta^* = \beta + \frac{\mathbb{E}[X\nu]}{\mathbb{E}[X^2]} = \beta \left(1 - \frac{\mathbb{E}[u^2]}{\mathbb{E}[X^2]} \right).$$

  Since $\mathbb{E}[u^2]/\mathbb{E}[X^2] < 1$, the projection coefficient shrinks the structural parameter β towards zero. This is called **measurement error bias** or **attentuation bias**.

- **Simultaneous equations bias.** Suppose that quantity Q and price P are determined jointly by demand

$$Q = -\beta_1 P e_1$$

and supply

$$Q = \beta_2 P + e_2,$$

with (for simplicity) $e = (e_1, e_2)$ satisfying $\mathbb{E}[e] = 0$ and $\mathbb{E}[ee'] = I_2$. In matrix notation, we have

$$\begin{aligned}
& \begin{pmatrix} 1 & \beta_1 \\ 1 & -\beta_2 \end{pmatrix} \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \\
\iff \qquad & \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} 1 & \beta_1 \\ 1 & -\beta_2 \end{pmatrix}^{-1} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \\
& \qquad = \frac{1}{\beta_1 + \beta_2} \begin{pmatrix} \beta_2 & \beta_1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \\
& \qquad = \begin{pmatrix} (\beta_2 e_1 + \beta_1 e_2)/(\beta_1 + \beta_2) \\ (e_1 - e_2)/(\beta_1 + \beta_2) \end{pmatrix}.
\end{aligned}$$

The projection of Q on P yields $Q = \beta^* P + e^*$ with $\mathbb{E}[Pe^*] = 0$ and the coefficient defined by projection as

$$\beta^* = \mathbb{E}[P^2]^{-1}\mathbb{E}[PQ] = \frac{\beta_2 - \beta_1}{2}.$$

The projection coefficient β^* equals neither the demand slope β_1 nor the supply slope β_2, but equals an average of the two. (The fact that it is a simple average is an artifact of the covariance structure.) Hence the OLS estimate satisfies $\hat{\beta} \xrightarrow{p} \beta^*$, and the limit does not equal β_1 or β_2. The fact that the limit is neither the supply nor demand slope is called **simultaneous equations bias**. This occurs generally when Y and X are jointly determined, as in market equilibrium. Generally, when both the dependent variable and a regressor are simultaneously determined, the variables should be treated as endogenous.

- **Choice variables as regressors.** Suppose we are interested in outcome y, log-earnings, and we have predictor x, years of schooling. We are interested in the causal effect on y of an **exogenous** change in x—a change in amount of schooling that is not the choice of the individual; for example, an increase in the minimum age at which students leave school. The OLS regression model specifies

$$y = \beta x + u$$

where u is an error term. Regression of y on x yields OLS estimate $\hat{\beta}$ of β. If we assume that x is uncorrelated with u, OLS yields a consistent estimator for the true causal effect. However, u (which contains the effects of all variables besides schooling on earnings) could be correlated with x. For example, unobserved *ability* may be correlated with both earnings and increased levels of schooling. In that case, OLS will be consistent for

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \beta + \frac{\mathrm{d}u}{\mathrm{d}x} > \beta.$$

That is, the positive correlation between x and u means that the linear projection coefficient β^* is upwardly biased relative to the structural coefficient β. The OLS estimator is therefore biased and inconsistent for β, over-estimating the causal effect of education on wages.

This type of endogeneity occurs generally when Y and X are both choices made by an economic agent, even if they are made at different points in time. Generally, when both the dependent variable and a regressor are choice variables made by the same agent, the variables should be treated as endogenous.

A more formal treatment of the linear regression model with K regressors leads to the same conclusion. Under standard assumptions, a necessary condition for consistency of OLS is that $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u} \xrightarrow{p} \boldsymbol{0}$; we can see this because

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y} \\
&= \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \frac{1}{n}\boldsymbol{X}^\top \left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\right) \\
&= \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} + \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u} \\
&= \boldsymbol{\beta} + \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u};
\end{aligned}
$$

we see this converges to $\boldsymbol{\beta}$ in probability if $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{u} \xrightarrow{p} \boldsymbol{0}$ (see also Section 4.7.1 of Cameron and Trivedi [2005]).

1.2.2 Instrumental Variable

The inconsistency of OLS is due to the endogeneity of x, meaning that changes in x are associated not only with changes in y bu also changes in the error u. What is needed is a method to generate only exogenous variation in x. An obvious way is through a randomized experiment, but for many economic applications such experiments are too expensive, infeasible, or unethical. One alternative approach is using an instrument.

An **instrument** z is a variable that is correlated with x but not with u or directly with y (that is, z is associated with y only through its effect on x).

Definition 1.2.1 (Instrumental variable; Definition 12.1 in Hansen [2020]). The random vector $Z \in \mathbb{R}^\ell$ is an **instrumental variable** if the following are true:

$$
\begin{aligned}
\mathbb{E}[Z^\top e] &= 0, \\
\mathbb{E}[ZZ^\top] &= 0, \qquad \text{and} \\
\operatorname{rank}(\mathbb{E}[ZX^\top]) &= p.
\end{aligned}
$$

The first component of this definition is that the instruments are uncorrelated with the regression error. Second, we must exclude linearly dependent instruments. The third condition is often called the **relevance condition** and is essential for the identification of the model. A necessary condition for the relevance condition is $\ell \geq p$.

1.2.3 Instrumental Variables Estimator

For regression with scalar regressor x and scalar instrument z, the **instrumental variables (IV) estimator** is defined as

$$\hat{\beta}_{IV} := \left(\boldsymbol{z}^\top \boldsymbol{x}\right)^{-1} \boldsymbol{z}^\top \boldsymbol{y}.$$

This estimator is consistent for the slope coefficient β in the linear model if z is correlated with x and uncorrelated with u.

We will derive this estimator. Note that under our assumptions,

$$\mathbb{E}\left[\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta} \mid \boldsymbol{z}\right] = \boldsymbol{0}.$$

Using this, we have

$$\boldsymbol{0} = \mathbb{E}\left[\boldsymbol{z}^\top \boldsymbol{0}\right] = \mathbb{E}\left[\boldsymbol{z}^\top \mathbb{E}\left[\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta} \mid \boldsymbol{z}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{z}^\top \left(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}\right) \mid \boldsymbol{z}\right]\right] = \mathbb{E}\left[\boldsymbol{z}^\top \left(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}\right)\right].$$

If the number of instruments equals the number of regressors ($\dim(\boldsymbol{z}) = p$), the method of moments estimator is then the solution to the corresponding sample moment condition

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i (y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}) = \boldsymbol{0} \\ \iff \quad & \boldsymbol{z}^\top \left(\boldsymbol{y} - \boldsymbol{x}\hat{\boldsymbol{\beta}}\right) = \boldsymbol{0} \\ \iff \quad & \boldsymbol{z}^\top \boldsymbol{y} = \boldsymbol{z}^\top \boldsymbol{x}\hat{\boldsymbol{\beta}} \\ \iff \quad & \hat{\boldsymbol{\beta}} = \left(\boldsymbol{z}^\top \boldsymbol{x}\right)^{-1} \boldsymbol{z}^\top \boldsymbol{y}, \end{aligned}$$

as shown in (1.4).

1.2.4 Two-Stage Least Squares (Section 8.3.4 of Greene [2003])

Suppose there may be more instruments than endogenous variables. Then $Z^\top X$ is not invertible (it is rank p but has ℓ rows), and a new analysis is required. Since Z is uncorrelated with e, we can express an approximation $\hat{X}$ of X in the column space of Z by projection:

$$\hat{X} = Z(Z^\top Z)^{-1} Z^\top X.$$

Then we can regress y against $\hat{X}$ to get a consistent estimator for the endogenous (structural) coefficient:

$$
\begin{aligned}
\beta_{\mathrm{IV}} &= \left(\hat{X}^\top \hat{X}\right)^{-1} \hat{X}^\top y \\
&= \left(\left[Z(Z^\top Z)^{-1} Z^\top X \right]^\top Z(Z^\top Z)^{-1} Z^\top X \right)^{-1} \left[Z(Z^\top Z)^{-1} Z^\top X \right]^\top y \\
&= \left(X^\top Z(Z^\top Z)^{-1} Z^\top Z(Z^\top Z)^{-1} Z^\top X \right)^{-1} X^\top Z(Z^\top Z)^{-1} Z^\top y \\
&= \left(X^\top Z(Z^\top Z)^{-1} Z^\top X \right)^{-1} X^\top Z(Z^\top Z)^{-1} Z^\top y. \qquad (1.3)
\end{aligned}
$$

Similarly, when p endogenous regressors are in X and p (an equal number) of instruments are available, we have

$$
\hat{\beta}_{IV} := \left(\boldsymbol{Z}^\top \boldsymbol{X} \right)^{-1} \boldsymbol{Z}^\top \boldsymbol{y}. \qquad (1.4)
$$

1.2.5 GMM Estimator (Section 13.6 of Hansen [2020])

As discussed in Section 1.1.1, the moment equations for instrumental variables are

$$
\boldsymbol{Z}^\top \boldsymbol{Y} - \boldsymbol{Z}^\top \boldsymbol{X} \boldsymbol{\beta} = 0,
$$

so the GMM criterion (1.2) can be written as

$$
J(\beta) = n \left(\boldsymbol{Z}^\top \boldsymbol{Y} - \boldsymbol{Z}^\top \boldsymbol{X} \boldsymbol{\beta} \right)^\top \boldsymbol{W} \left(\boldsymbol{Z}^\top \boldsymbol{Y} - \boldsymbol{Z}^\top \boldsymbol{X} \boldsymbol{\beta} \right).
$$

The GMM estimator minimizes $J(\beta)$. The first order conditions are

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \beta} J(\hat{\beta}) \\
&= 2 \frac{\partial}{\partial \beta} \overline{g}_n(\hat{\beta})^\top \boldsymbol{W} \overline{g}_n(\hat{\beta}) \\
&= -2 \left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{Z} \right) \boldsymbol{W} \left(\frac{1}{n} \boldsymbol{Z}^\top (\boldsymbol{Y} - \boldsymbol{X} \hat{\boldsymbol{\beta}}) \right).
\end{aligned}
$$

The solution is the GMM estimator for the overidentified IV model,

$$
\hat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \left(\boldsymbol{X}^\top \boldsymbol{Z} \boldsymbol{W} \boldsymbol{Z}^\top \boldsymbol{X} \right)^{-1} \boldsymbol{X}^\top \boldsymbol{Z} \boldsymbol{W} \boldsymbol{Z}^\top \boldsymbol{Y},
$$

the same estimator as in (1.1). The dependence on the estimator $\boldsymbol{W}$ is only up to scale; that is, if $\boldsymbol{W}$ is replaced by $c\boldsymbol{W}$ for some $c > 0$, $\hat{\boldsymbol{\beta}}_{\mathrm{gmm}}$ does not change. When $\boldsymbol{W}$ is fixed by the user, we call $\hat{\boldsymbol{\beta}}_{\mathrm{gmm}}$ a **one-step GMM** estimator. Note that by comparison to (1.3), we see that if $\boldsymbol{W} = \left(\boldsymbol{Z}^\top \boldsymbol{Z} \right)^{-1}$ then we have the two stage least squares estimator. Also note that if $\ell = p$ then $\boldsymbol{X}^\top \boldsymbol{Z}$ is invertible (as is $\boldsymbol{W}$ since it is positive definite by assumption) and we have

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{gmm}} &= \left(\boldsymbol{Z}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{W}^{-1} \left(\boldsymbol{X}^\top \boldsymbol{Z}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{Z} \boldsymbol{W} \boldsymbol{Z}^\top \boldsymbol{Y} \\
&= \left(\boldsymbol{Z}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{W}^{-1} \boldsymbol{W} \boldsymbol{Z}^\top \boldsymbol{Y} \\
&= \left(\boldsymbol{Z}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{Y},
\end{aligned}
$$

which matches the estimator in (1.4).

Bibliography

A. C. Cameron and P. K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005. doi: 10.1017/CBO9780511811241.

W. H. Greene. *Econometric Analysis*. Pearson Education, fifth edition, 2003. ISBN 0-13-066189-9. URL http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm.

B. E. Hansen. *Econometrics*. August 2020.