Inference and Instrumental Variables

References

Davidson, Russell and James G. MacKinnon (2004). *Econometric Theory and Methods.* Oxford University Press.
Hayashi, Fumio (2000). *Econometrics.* Princeton University Press.

1 Linear regression and exact inference

Consider the linear regression model

$$y_i = \beta_1 + \beta_2 x_i + u_i,\ \ i = 1, 2, ..., n,$$

in which $y|X \sim N(X\boldsymbol{\beta}, \sigma^2\mathbf{I})$*. Here* y *is an* $n \times 1$ *vector containing the observations on* y_i*,* X *is an* $n \times 2$ *matrix in which each row has the form* $(1, x_i)$*, the* 2×1 *vector* $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$*, and* $\mathbf{I}$ *is an* $n \times n$ *identity matrix.*

1.1

Derive the conditional distribution of the OLS estimator of the parameter vector $\boldsymbol{\beta}$*.*

Based on the assumptions given, we can rewrite the model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}). \tag{1}$$

We can write the OLS estimator for the parameter $\boldsymbol{\beta}$ as

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{OLS}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \\ &= \boldsymbol{\beta}_0 + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u},\end{aligned}$$

where $\boldsymbol{\beta}_0$ is the true value of the parameter vector. Since we assumed that $\mathbb{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta} \Leftrightarrow \mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ and $\text{Var}(\mathbf{y}|\mathbf{X}) = \sigma^2\mathbf{I} \implies \mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}] = \sigma^2\mathbf{I}$, we can evaluate the quadratic product of the bias term:

$$\begin{aligned}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}\mathbf{u}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ \mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top|\mathbf{X}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}, \\ \therefore \text{Var}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}.\end{aligned}$$

This of course assumes that the probability limit of $(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}$ is 0:

$$\operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{u} = \mathbf{S}_{\mathbf{X}^\top\mathbf{X}}^{-1}\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{u} = \mathbf{0},$$

where $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$ is a deterministic limiting matrix, with full rank k. Since we know that $\mathbb{E}[\mathbf{X}^\top\mathbf{u}|\mathbf{X}] = \mathbf{0}$, and by the law of iterated expectations $\mathbb{E}[\mathbf{X}^\top\mathbf{u}] = \mathbf{0}$, we are sure that the above expression tends to the null vector in the probability limit.

$$\therefore \hat{\boldsymbol{\beta}}|\mathbf{X} \sim N(\boldsymbol{\beta}_0, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}). \tag{2}$$

Why $\hat{\boldsymbol{\beta}}_2|\mathbf{X}$ is normally distributed is explained in the next section.

1.2

Briefly explain how this result can be used to obtain a test of the hypothesis that $\beta_2 = 0$, in the special case where σ^2 is known.

Using the FWL theorem, $\hat{\beta}_2$ can be attained by multiplying $\mathbf{y}$ by the elimination matrix, $\mathbf{M}_1$:

$$\mathbf{M}_1 = \mathbf{I} - \boldsymbol{\iota}(\boldsymbol{\iota}^\top \boldsymbol{\iota})^{-1}\boldsymbol{\iota}^\top,$$

where $\boldsymbol{\iota}$ is the unit vector, which is the first column of the matrix $\mathbf{X}$ in (1). This yields

$$\begin{aligned}\mathbf{M}_1\mathbf{y} &= \mathbf{M}_1\boldsymbol{\beta}_2\mathbf{x}_2 + \mathbf{r},\\ \hat{\beta}_2 &= (\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2)^{-1}\mathbf{x}_2^\top \mathbf{M}_1\mathbf{y}.\end{aligned}$$

In order to test the hypothesis that $\beta_2 = 0 = \beta_2^0$, we have to subtract β_2^0 from $\hat{\beta}_2$ and divide by the square root of the variance. By the FWL theorem we know

$$\text{Var}(\hat{\beta}_2|\mathbf{X}) = \sigma^2(\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2)^{-1},$$

and so our test statistic is thus

$$\begin{aligned}z_{\beta_2} &= \frac{1}{\sqrt{\sigma^2(\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2)^{-1}}}(\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2)^{-1}\mathbf{x}_2^\top \mathbf{M}_1\mathbf{y}\\ z_{\beta_2} &= \frac{\mathbf{x}_2^\top \mathbf{M}_1\mathbf{u}}{\sigma(\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2)^{\frac{1}{2}}} \sim N(0,1). \qquad (3)\end{aligned}$$

Since we assume that the true value of the error term variance, σ^2, is known, and that errors are assumed to be normally distributed, we can use an exact test to test our hypothesis. Note that the numerator in (3) is dependent on $\mathbf{u}$ and not $\mathbf{y}$. This is because if the data was actually generated by the following DGP

$$\mathbf{y} = \beta_1\boldsymbol{\iota} + \beta_2\mathbf{x}_2 + \mathbf{u},$$

and $\beta_2 = 0$ under the null hypothesis, then

$$\mathbf{M}_1\boldsymbol{y} = \mathbf{M}_1(\beta_1\boldsymbol{\iota} + \mathbf{u}) = \mathbf{M}_1\mathbf{u},$$

and so

$$z_{\beta_2} = \frac{\mathbf{x}_2^\top \mathbf{M}_1\mathbf{y}}{\sigma(\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2)^{\frac{1}{2}}} = \frac{\mathbf{x}_2^\top \mathbf{M}_1\mathbf{u}}{\sigma(\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2)^{\frac{1}{2}}}.$$

It is rather easy to see that $z_{\beta_2} \sim N(0,1)$. Since we can condition on $\mathbf{X}$, the only thing left in z_{β_2} that is stochastic is $\mathbf{u}$. Since the numerator is just a linear combination of the components of $\mathbf{u}$, which is multivariate normal, the entire statistic must be normally distributed – analogous to why (2) follows a normal distribution. The variance of the numerator (3) is:

$$\begin{aligned}\mathbb{E}[\mathbf{x}_2^\top \mathbf{M}_1\mathbf{u}\mathbf{u}^\top \mathbf{M}_1\mathbf{x}_2|\mathbf{X}] &= \mathbf{x}_2^\top \mathbf{M}_1\mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{X}]\mathbf{M}_1\mathbf{x}_2\\ &= \sigma^2\mathbf{x}_2^\top \mathbf{M}_1\mathbf{x}_2.\end{aligned}$$

Since the denominator of z_{β_2} is just the square root of the variance of the numerator this implies that $z_{\beta_2} \sim N(0,1)$.

2 Asymptotic inference, omitted variables and instrumental variables

A researcher is interested in the effect of explanatory variable x_2 on an outcome y. She believes that both y and x_2 are influenced by a third variable x_3. She formulates a linear model

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

in which the subscript i indexes observation, and u_i denotes the error term. She is willing to assume that $\mathbb{E}[u_i] = \mathbb{E}[x_{2i}u_i] = \mathbb{E}[x_{3i}u_i] = 0$, and that $\mathbb{E}[u_i^2|x_{2i}, x_{3i}] = \sigma^2$ for all observations.

It's first worth defining some asymptotic properties that we will use for building confidence intervals and asymptotic tests. We can rewrite out the regression in matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbb{E}[\mathbf{u}] = 0,$$

and thus the OLS estimate for the parameter vector β is given as:

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}.$$

We could use the FWL theorem to attain an estimate of $\hat{\beta}_2$ using either the projection matrix $\mathbf{P}_2 = \mathbf{x}_2(\mathbf{x}_2^\top\mathbf{x}_2)^{-1}\mathbf{x}_2^\top$ or the elimination matrix $\mathbf{M}_{-2}$ (such that it eliminates all columns of $\mathbf{X}$ except $\mathbf{x}_2$). Then we can set up our hypothesis:

$$\begin{aligned} H_0 &: \ \beta_2 = 0 = \beta_2^0, \\ H_1 &: \ \beta_2 \neq 0, \end{aligned}$$

where the test statistic is

$$t_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2^0}{\text{SE}(\hat{\beta}_2)} \overset{a}{\sim} N(0,1). \tag{4}$$

Here, we would normally perform a t-test for the single restriction. But since we cannot assume that the error vector $\mathbf{u}$ is NID, and since we do not know the true value of the error vector variance σ^2, we cannot perform an exact test. We must instead rely on the assumption that our test statistic is asymptotically distributed as a standard normal distribution. The proof for why $t_{\beta_2} \overset{a}{\sim} N(0,1)$ is different to the proof we used in the previous section. Previously we relied on a finite sample/LLN argument; here, we will rely on the CLT to argue the asymptotic distribution of the test statistic. By the CLT we know that the random vector

$$\mathbf{v} = n^{-1/2}\mathbf{X}^\top\mathbf{u}$$

follows the normal distribution asymptotically with mean vector $\mathbf{0}$ and covariance matrix $\sigma_0^2\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$, where $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$ is the plim of $n^{-1}\mathbf{X}^\top\mathbf{X}$ as the sample size tends to infinity.

Consider now the estimation error of the vector of OLS estimates

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0 = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{u}, \tag{5}$$

$\hat{\boldsymbol{\beta}}$ is consistent under fairly weak conditions. If it is, then the estimator error tends to a limit of $\mathbf{0}$ as the sample size tends to infinity. Therefore, its limiting covariance matrix is a zero matrix. Thus, it would appear that asymptotic theory has nothing to say about limiting variances for consistent estimators. However, this is easily corrected by the usual device of introducing a few well–chosen powers of n whenever we use the CLT. If we rewrite (5) as

$$n^{1/2}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0 = (n^{-1}\mathbf{X}^\top\mathbf{X})^{-1}n^{-1/2}\mathbf{X}^\top\mathbf{u},$$

then the first factor on the RHS tends to $\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}$ as $n \to \infty$, and the second factor, which is just $\mathbf{v}$, tends to a random vector distributed as $N(\mathbf{0}, \sigma_0^2\mathbf{S}_{\mathbf{X}^\top\mathbf{X}})$. Because $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$ is deterministic, we find that, asymptotically

$$\mathrm{Var}(n^{1/2}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0)) = \sigma_0^2\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}} = \sigma_0^2\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}.$$

Moreover, since $n^{1/2}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0)$ is, asymptotically, just a deterministic linear combination of the components of the multivariate normal random vector $\mathbf{v}$, we may conclude that

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \overset{a}{\sim} N(\mathbf{0}, \sigma_0^2\mathbf{S}^{-1}_{\mathbf{X}^\top\mathbf{X}}).$$

This result tells us that the asymptotic covariance matrix of the vector $n^{1/2}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})$ is the limit of $\sigma_0^2(n^{-1}\mathbf{X}^\top\mathbf{X})^{-1}$ as $n \to \infty$. In practice, we divide this matrix by n and use $s^2(\mathbf{X}^\top\mathbf{X})$ to estimate $\mathrm{Var}(\hat{\boldsymbol{\beta}})$. It then follows that our test statistic to test the hypothesis that $\hat{\beta}_2 = 0 = \beta_2^0$ is

$$t_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2^0}{\mathrm{SE}(\hat{\beta}_2)} = \frac{\mathbf{x}_2^\top\mathbf{M}_1\mathbf{u}}{\sigma_0(\mathbf{x}_2^\top\mathbf{M}_1\mathbf{x}_2)^{\frac{1}{2}}} \overset{a}{\sim} N(0,1).$$

Based on this test statistic and corresponding critical value from the standard normal distribution (approximately 1.96 at a 5% level of significance), we would reject the null hypothesis if $|t_{\beta_2}| \geq 1.96$.

2.1

Using independent and identically distributed data on (y_i, x_{2i}, x_{3i}) for a large sample, explain how the researcher could obtain an approximate 95% confidence interval for the parameter β_2. In what sense is this confidence interval approximate? Any limit distribution results that are used in your argument should be stated formally but not proved.

To construct the 95% confidence interval, we start with the fact that

$$\Pr\left(z_{\frac{\alpha}{2}} \leq \frac{\hat{\beta}_2 - \beta_2^0}{\mathrm{SE}(\hat{\beta}_2)} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha,$$

where $\alpha = 0.05$. Then we can rearrange to get

$$\begin{aligned}
&\Pr\left(-z_{\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2) \leq \hat{\beta}_2 - \beta_2^0 \leq -z_{1-\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2)\right) \\
&= \Pr\left(-z_{\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2) \geq \beta_2^0 - \hat{\beta}_2 \geq -z_{1-\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2)\right) \\
&= \Pr\left(\hat{\beta}_2 - z_{\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2) \geq \beta_2^0 \geq \hat{\beta}_2 - z_{1-\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2)\right).
\end{aligned}$$

Our confidence interval is thus:

$$\left[\hat{\beta}_2 - z_{1-\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2), \hat{\beta}_2 - z_{\frac{\alpha}{2}}\mathrm{SE}(\hat{\beta}_2)\right].$$

Our confidence interval is an approximation in the sense that our standard errors are based on estimates of the true error vector variance, σ^2.

2.2

Suppose that x_{2i} is a choice variable that can be influenced by government policies, while x_{3i} is an innate characteristic of observation i that is not affected by government policies. Explain why it is important to control for x_3 here, if the objective is to estimate the effect of policy-induced changes in x_2 on the outcome y.

$\mathbf{x}_3$ is assumed to have some influence on $\mathbf{x}_2$ and $\mathbf{y}$ (implying a potential multicollinearity problem), which could cause potential issues with the OLS estimated standard errors. Since $\mathbf{x}_3$ is unable to be controlled by policy, the researcher should attempt to remove any influence of $\mathbf{x}_3$ on the dependent variable and the other independent variables, so that regression parameter estimates are accurate for potential comparative statics of changes in $\mathbf{x}_2$.

The tradeoff here between dealing with potential multicollinearity (when $\mathbf{x}_3$ is included in the model) and potential omitted variable bias (endogeneity between $\mathbf{x}_2$ and the modified error term after controlling for $\mathbf{x}_3$). A best case scenario would be for the researcher to control for the effects of $\mathbf{x}_3$ and finding effective instruments to run the regression.

2.3

Now suppose that there is no data on x_{3i}, but the researcher has independent and identically distributed data on $(y_i, x_{2i}, z_{1i}, z_{2i})$ for a large sample. The researcher reformulates her model as

$$y_i = \tilde{\beta}_1 + \beta_2 x_{2i} + \tilde{u}_i$$

in which $\tilde{\beta}_1 = \beta_1 + \beta_3\mu_3$ and $\tilde{u}_i = u_i + \beta_3(x_{3i} - \mu_3)$, where μ_3 is the mean of the unobserved x_{3i} characteristic in this sample. She is willing to assume that $\mathbb{E}[\tilde{u}_i] = 0$ and that $\mathbb{E}[\tilde{u}_i^2|z_{1i}, z_{2i}] = \tilde{\sigma}^2$ for all observations, but that $\mathbb{E}[x_{2i}\tilde{u}_i] = 0$.

2.3.1

Explain how the researcher could investigate whether one or both of (z_{1i}, z_{2i}) are informative instruments for x_{2i} in this setting.

If we wish to test for informative instruments, then we can construct the following linear projection

$$\mathbf{x}_2 = \mathbf{Z}\boldsymbol{\pi} + \mathbf{v}, \tag{6}$$

where $\mathbf{Z} = \begin{bmatrix}\mathbf{z}_1 & \mathbf{z}_2\end{bmatrix}$, $\boldsymbol{\pi} = \begin{bmatrix}\pi_1 & \pi_2\end{bmatrix}^\top$, and $\mathbb{E}[\mathbf{v}|\mathbf{Z}] = \mathbf{0}$, and test if $\mathbb{E}[\mathbf{Z}^\top\mathbf{x}] \neq \mathbf{0} \leftrightarrow \Pi \neq \mathbf{0}$. This can be done by a Wald test after running a first stage OLS regression of (6). The estimated coefficient vector from our first stage regression is given by

$$\hat{\boldsymbol{\pi}} = (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{x}_2,$$

where, by construction,

$$\begin{aligned}\hat{\boldsymbol{\pi}} &= \boldsymbol{\pi}_0 + (\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{v} \\ \mathbb{E}[\hat{\boldsymbol{\pi}}] &= \boldsymbol{\pi}_0.\end{aligned}$$

Thus, our hypothesis is

$$\begin{aligned}&H_0 : \ \boldsymbol{\pi} = \mathbf{0} = \boldsymbol{\pi}_0, \\ &H_1 : \ \boldsymbol{\pi} \neq \mathbf{0},\end{aligned}$$

and our Wald test statistic is

$$W = (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0)^\top \left(\hat{\mathrm{Var}}(\hat{\boldsymbol{\pi}})\right)^{-1} (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) \overset{a}{\sim} \frac{\chi^2(p)}{p},$$

where $p = 2$, the rank of our restriction matrix or simply the number of instruments for which we wish to test that are significantly different from 0. Our decision rule, like before, is that we reject the null hypothesis if our test statistic lies in the rejection region (values greater than $\chi^2(p)$ multiplied by $1/p$). In addition, as a general rule of thumb, Staiger and Stock (1997) suggest that the value of the F-test statistic should be at least 10 here for the standard asymptotic approximation of the distribution of the 2SLS estimator to be considered reliable. For this linear model with IID data, a single endogenous explanatory variable and conditional homoskedasticity, Stock and Yogo (2005) provide alternative critical values for this F-test. These seek to control either the bias in the 2SLS estimator of $\tilde{\beta}$ relative to the bias of the OLS estimator of $\tilde{\beta}$ in the same model, or to control the rejection frequency of standard asymptotic t-tests of the null hypothesis that $\beta_2 = 0$ based on the 2SLS estimator.

2.3.2

Assuming that both z_{1i} and z_{2i} are informative, explain how the researcher could investigate whether these variables are valid instruments for x_{2i} in this setting.

To test the validity of the instruments, the researcher can run the Sargan test based on the test statistic

$$J = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{2SLS}})^\top \mathbf{P}_{\mathbf{Z}} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{2SLS}})}{n^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{2SLS}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{2SLS}})} \overset{a}{\sim} \chi^2(p-d),$$

where $\mathbf{P}_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top$, d is the number of endogenous variables (here $d = 1$), and the 2SLS estimator is attained by running the second stage regression

$$\hat{\boldsymbol{\beta}}_{\text{2SLS}} = (\mathbf{X}^\top \mathbf{P}_{\mathbf{Z}} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P}_{\mathbf{Z}} \mathbf{y}.$$

The Sargan test can be used test the null hypothesis that $\mathbb{E}[\mathbf{Z}^\top \mathbf{u}] = \mathbf{0}$, against the alternative hypothesis that the instruments are invalid.

There is a caveat to the Sargan test, however. Suppose that both instruments of $\mathbf{Z}$ are informative, but only one of them is valid. The Sargan test is unable to distinguish between which instrument is valid and which is invalid, and would thus fail.

3 The asymptotic t-test with predetermined regressors

We begin by applying a CLT to the k-vector

$$\mathbf{v} = \frac{1}{\sqrt{n}}\mathbf{X}^\top \mathbf{u} = n^{-1/2}\sum_{t=1}^{n} u_t \mathbf{X}_t.$$

We assume that $\mathbb{E}_t(u_t|\mathbf{X}_t) = 0$ which implies that $\mathbb{E}_t(u_t\mathbf{X}_t^\top) = 0$, as required for the CLT, which then tells us that

$$\mathbf{v} \overset{a}{\sim} N\left(\mathbf{0}, \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathrm{Var}(u_t\mathbf{X}_t^\top)\right) = N\left(\mathbf{0}, \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_t(u_t^2\mathbf{X}_t^\top\mathbf{X}_t)\right);$$

recall that for the multivariate version of the CLT, suppose we have a sequence of uncorrelated random m-vectors $\mathbf{x}_t$, for some fixed m, with $\mathbb{E}_t(\mathbf{x}_t) = 0$, then

$$\operatorname*{plim}_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{x}_t = \mathbf{x}_0 \sim N\left(\mathbf{0}, \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathrm{Var}(\mathbf{x}_t)\right),$$

where $\mathbf{x}_0$ is multivariate normal, and each $\mathrm{Var}(\mathbf{x}_t)$ is an $m\times m$ matrix. Notice that because $\mathbf{X}_t$ is a $1\times k$ row vector, the covariance matrix here is $k\times k$. If we assume that $\mathbb{E}_t(u_t|\mathbf{X}_t) = 0$ and $\mathbb{E}_t(u_t^2|\mathbf{X}_t) = \sigma^2$, then

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_t(u_t^2\mathbf{X}_t^\top\mathbf{X}_t) &= \lim_{n\to\infty}\sigma_0^2\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_t(\mathbf{X}_t^\top\mathbf{X}_t)\\
&= \sigma_0^2\operatorname*{plim}_{n\to\infty}\sum_{t=1}^{n}\mathbf{X}_t^\top\mathbf{X}_t\\
&= \sigma_0^2\operatorname*{plim}_{n\to\infty}\mathbf{X}^\top\mathbf{X}\\
&= \sigma_0^2\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}.
\end{aligned}$$

Now, we can rewrite (4) as

$$t_{\beta_2} = \left(\frac{\mathbf{y}^\top\mathbf{M}_\mathbf{X}\mathbf{y}}{n-k}\right)^{-1/2}\frac{n^{-1/2}\mathbf{x}_2\mathbf{M}_{-2}^\top\mathbf{y}}{(n^{-1}\mathbf{x}_2\mathbf{M}_{-2}^\top\mathbf{x}_2)^{1/2}},$$

and under a correctly specified DGP, $s^2 = \mathbf{y}^\top\mathbf{M}_\mathbf{X}\mathbf{y}/(n-k)$ tends to σ_0^2 as $n\to\infty$. i.e. the OLS error variance estimator is consistent. If the data were generated with $\beta_2 = 0$, we would have that $\mathbf{M}_{-2}\mathbf{y} = \mathbf{M}_{-2}\mathbf{u}$, and so t_{β_2} would be asymptotically equivalent to

$$\frac{n^{-1/2}\mathbf{x}_2\mathbf{M}_{-2}^\top\mathbf{u}}{\sigma_0^2(n^{-1}\mathbf{x}_2\mathbf{M}_{-2}^\top\mathbf{x}_2)^{1/2}}. \tag{7}$$

Now, consider the numerator of the above expression. It can be written as

$$n^{-1/2}\mathbf{x}_2^\top\mathbf{u} - n^{-1/2}\mathbf{x}_2^\top\mathbf{P}_{-2}\mathbf{u}. \tag{8}$$

The first term of this expression is just the last, or kth, component of $\mathbf{v}$, which we can denote by v_2. By writing out the projection matrix $\mathbf{P}_{-2}$, explicitly, and dividing various expressions by n in a way that cancels out, the second term can be rewritten as

$$\underbrace{n^{-1}\mathbf{x}_2^\top\mathbf{X}_{-2}}_{\xrightarrow{p}\mathbf{S}_{21}}\underbrace{(n^{-1}\mathbf{X}_{-2}^\top\mathbf{X}_{-2})^{-1}}_{\xrightarrow{p}\mathbf{S}_{11}^{-1}}n^{-1/2}\mathbf{X}_{-2}^\top\mathbf{u}.$$

$\mathbf{S}_{21}$ is a submatrix of $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$ and $\mathbf{S}_{11}^{-1}$ is the inverse of a submatrix of $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$. Thus, only the last factor remains random when $n \to \infty$. it is just the subvector of $\mathbf{v}$ consisting of the first $k-1$ components, which we denote as $\mathbf{v}_1$. Asymptotically, in partitioned matrix notation (8) becomes

$$v_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{v}_1 = \begin{bmatrix}\mathbf{S}_{21}\mathbf{S}_{11}^{-1} & 1\end{bmatrix}\begin{bmatrix}\mathbf{v}_1 \\ v_2\end{bmatrix}.$$

Since $\mathbf{v}$ is asymptotically multivariate normal, this scalar expression is also asymptotically normal, with mean zero and variance

$$\sigma_0^2\begin{bmatrix}\mathbf{S}_{21}\mathbf{S}_{11}^{-1} & 1\end{bmatrix}\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}\begin{bmatrix}\mathbf{S}_{21}\mathbf{S}_{11}^{-1} \\ 1\end{bmatrix},$$

where, since $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$ is symmetric, $\mathbf{S}_{12}$ is just the transpose of $\mathbf{S}_{21}$. If we now express $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$ as a partitioned matrix, the variance of (8) becomes

$$\sigma_0^2\begin{bmatrix}\mathbf{S}_{21}\mathbf{S}_{11}^{-1} & 1\end{bmatrix}\begin{bmatrix}\mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{S}_{21}\mathbf{S}_{11}^{-1} \\ 1\end{bmatrix} = \sigma_0^2(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}) \tag{9}$$

The denominator of (7) is easier to deal with. The square of the second factor is

$$\begin{aligned} n^{-1}\mathbf{x}_2\mathbf{M}_{-2}^\top\mathbf{x}_2 &= n^{-1}\mathbf{x}_2^\top\mathbf{x}_2 - n^{-1}\mathbf{x}_2^\top\mathbf{P}_{-2}\mathbf{x}_2 \\ &= n^{-1}\mathbf{x}_2^\top\mathbf{x}_2 - n^{-1}\mathbf{x}_2^\top\mathbf{X}_{-2}(n^{-1}\mathbf{X}_{-2}^\top\mathbf{X}_{-2})^{-1}n^{-1/2}\mathbf{X}_{-2}^\top\mathbf{x}_2. \end{aligned}$$

In the limit, all the pieces of this expression become submatrices of $\mathbf{S}_{\mathbf{X}^\top\mathbf{X}}$, and so

$$n^{-1}\mathbf{x}_2\mathbf{M}_{-2}\mathbf{x}_2 \to \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}.$$

When multiplying this by σ_0^2, we get (9), and so putting all the pieces together

$$\frac{\sigma_0^2(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})}{\sigma_0^2(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})} = 1,$$

proving that $t_{\beta_2} \overset{a}{\sim} N(0,1)$.[1]

[1] This derivation assumed that we merely have predetermined regressors. The derivation is much simpler if we assume exogenous regressors.