Problem set 1
ANSWERS
607: Applied Macroeconometrics
Fall 2017
Jon Faust

The following is due at the beginning of next class. You can turn in any paper in my mailbox or in class; email me and requested computer work. You may work in groups; hand in a single submission for the group. The submission should list those who contributed.

1. Definitions

   (a) Define size and nominal size. For this definition assume that $Y \sim P_\theta$, $\theta \in \Theta$ and the null hypothesis is $H_0 : \theta \in \Theta_0 \subset \Theta$. The test rejects if the statistic $\phi(Y)$ satisfies $\phi(Y) > c$.

   Answer/comment

   Size: supremum of the probability you reject the null when it is true taken across all ways the null can be true:

   $$\sup_{\theta \in \Theta} P_\theta(E)$$

   . where E is the event 'reject the null' and P_θ is the probability measure associated with F_θ.

   Nominal size: The probability of E under some asymptotic distribution. That is, a version of 'size' in which we replace the sup over exact distributions with a single value under some asymptotic distribution.

   (b) Define *pivotal test statistic* and *asymptotically pivotal test statistic*.

   Answer/comment

   A statisic is pivotal if its distributuion does not depend on unknown parameters. Suppose E, the rejection event is defined by all outcomes in which $\phi > c$ for some test statistic ϕ and constant critical value c. If ϕ is pivotal under the null, then $P_\theta(E)$ does not vary across $\theta \in \Theta_0$, and the sup in the definition of size becomes superflous.

Similarly if a statistic is asymptotically pivotal, the asymptotic distribution does not vary with unknown parameters.

(c) Define *autocorrelation* and *autocovariance*

Answer/comment

These are the correlation or covariance of some variable through time (*auto* here meaning 'with itself'). Also called, e.g., serial correlation, which conveys more of the 'through time' aspect.'

These concepts are usually defined for a series with constant mean and variance through time.

Take a time series x_t with mean μ and variance σ^2, the autocovariance at lag k is:

$$\sigma(k) = E[(x_t - \mu)(x_{t-k} - \mu)]$$

and the autocorrelation at lag k is,

$$\gamma(k) = \sigma(k)/\sigma^2$$

(d) For each of *generalized least squares (GLS) estimator* and *instrumental variables (IV) estimator* give the textbook formula for the estimator of the slope coefficients and of the variance-covariance matrix of the slope coefficients. Briefly describe when these estimators are appropriate.

Answer/comment

In std. notation, $Y = X\beta + \varepsilon$, $E\varepsilon = 0$, $E\varepsilon\varepsilon' = \Sigma$, the GLS estimator of β is,

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$

and under standard (e.g., fixed Xs) assumptions have vcov matrix,

$$(X'\Sigma^{-1}X)^{-1}$$

Feasible GLS substitutes an estimator $\hat{\Sigma}$ for Σ.

As a rule of thumb, if we have made the assumptions such that OLS satisfies Gauss-Markov with $\Sigma = I\sigma^2$, then GLS will have that property with general Σ. Sketch of proof: premultiply the DGP equation by Φ s.t. $\Phi\Sigma\Phi' = I$. The new system is

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$$

, where, e.g., $\tilde{Y} = \Phi Y$. This system now satisfies the conditions of the Gauss-Markov thm.

Feasible GLS will also share the asymptotic properties of GLS if the estimate of Σ is good enough.

Feasible GLS is appropriate when the assumptions of the classical linear regression model are satisfied except that $E\varepsilon\varepsilon' = \Sigma \neq I\sigma^2$. That is, there may be cross correlation and heteroskedasticity of the εs.

Instrumental variables. Same model as above. Now presume that some of the x_t's may be correlated with ε_t. An instrumental variable is one correlated with the problematic x_ts, but not ε_t. Let's posit,

$$X = Z\gamma + u$$

and Z follows some well-behaved process unrelated to u and eps. Note it is standard (at least in this class) to call an instrument *relevant* if it is correlated with with the regressor in question and *valid* if it is not correlated with the error term. Thus, we'd like valid and relevant instruments.

If the number of Zs and Xs is the same (that is, the no. of columns is same), the IV estimator is,

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$$

Note, following the discussion above, we can think about this as transforming our equation back to one that is a simple case. In this case, the violation of the classical assumption of OLS is correlation of regressors and errors. How could we fix this? Suppose we pre-multiply the DGP by

$$\Phi = Z(Z'Z)^{-1}Z'$$

Now consider the implied tilde model as in the discussion of GLS: $\tilde{Y} = \Phi Y$, $\tilde{X} = \Phi X$ and $\tilde{\varepsilon} = \Phi\varepsilon$.

The OLS estimator on the tilde model is,

$$(X'\Phi'\Phi X)^{-1}(X'\Phi'\Phi Y)$$

which through the miracle of cancellation is the IV estimator. The new regressor $\tilde{X}$ is,

$$\tilde{X} = Z[(Z'Z)^{-1}Z'X]$$

You should recognize $\tilde{X}$ as the projected value of X from an OLS regression of X on Z. You should be able to produce the argument why this tilde regressor will (in the limit) be uncorrelated with the $\tilde{\varepsilon}$.

Under standard assumptions the asymptotic vcov matrix of IV will be

$$Q_{ZX}^{-1}Q_{ZZ}Q_{XZ}^{-1}\sigma_{\varepsilon}^{2}$$

where

$$Q_{AB} = \text{plim} T^{-1}A'B$$

IV is appropriate when some x variable is correlated with ε, but you have a valid and relevant instrument z. Valid means that z_t is not correlated with ε_t; relevant means that z_t is correlated with x_t.

---

2. Warm up for time series (without much mentioning time series). Suppose that $x_t, t = 1, \ldots, T$ satisfies,

$$\begin{aligned} Ex_t &= 0 \\ Ex_tx_s &= \sigma(t,s) \end{aligned}$$

where $\sigma(.)$ is finite.

Define $X = (x_1, \ldots, x_T)'$, the column vector containing the xs.

Define the variance-covariance matrix of X:

$$\Sigma = EXX'$$

The sample mean of the xs can be written,

$$\bar{x} = T^{-1}i'X$$

where i is a vector of 1s. Since the mean of X is zero, the variance of the sample mean is

$$E(\bar{x}^2) = T^{-2}i'(Exx')i = T^{-2}i'\Sigma i$$

which you'll recognize (at least upon brief reflection) is the average element of Σ.

(a) Give a formula for the variance of the sample mean in terms of the $\sigma()$s.

Answer/comment

$$E\bar{x}^2 = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \Sigma_{st}$$

(b) We are interested in conditions under which sample means converge in probability to population means (WLLNs). What additional assumptions beyond those given already do we need to show that $\bar{x} \rightarrow_p 0$ using Chebyshev's inequality?

(Hint 1: You may find it more convenient to use Tchebysheff's inequality. Hint 2: Hint 1 is a transliteration joke.)

Answer/comment

Using Chebyshev's inequality, we can prove $\bar{x} \rightarrow_p 0$ if the variance of $\bar{x}$ goes to 0. For the variance to go to zero, we need that the average element of Σ $(T \times T)$ goes to zero with T.

This has a nice intuition. WLLNs doe not require that the xs be indepdendent. But they always require some sense of 'most' items are 'mostly' unassociated with 'most' others. The version we have just derived is that the 'typical' covariance of any two xs goes to zero; most x need to be nearly uncorrelated with most others.

(c) Show that the sample mean is the coefficient from the simple OLS regression of a variable on a constant:

$$X = i'\beta + \varepsilon$$

Framed this way, discuss the merits of using OLS vs. GLS to estimate the mean of X in data that are not mutually uncorrelated.

Answer/comment

Our usual $X'X$ matrix for OLS is $i'i$ in this case or T, our usual $X'Y$ is $i'X$ or the sum of the xs. Thus, $\hat{\beta} = \bar{x}$.

As noted above, we typically think OLS is best when $E\varepsilon\varepsilon' = I\sigma^2$ and that GLS is more appropriate when $E\varepsilon\varepsilon' = \Sigma$ as in the problem we have here.

Thus, this general reasoning says we might prefer GLS to OLS on efficiency grounds. As we'll explore further in the next problem set, when x has constant second moments (that is, the variance and autocovariances are constant through time), we have a special case where OLS is asymptotically efficient despite the fact that Σ has nonzero off diagonals.

(d) Suppose that $\Sigma_{t,s}$ depends only on $|t-s|$: $\Sigma_{t,s} \equiv \sigma(k)$ where $k = |t-s|$. That is, the covariance between any two elements x_t and x_s depends only on how far apart the two observations are in the data series, $|t-s|$. Describe the structure of Σ. How does this assumption simplify your formula for the variance of the sample mean?

Answer/comment

This question takes up that case of constant second moments. In this case Σ is a Toeplitz matrix: a matrix where on any diagonal (running upper left to lower right) there is only a single number. The k^{th} autocovariance is on the k^{th} diagonal counting out from zero at the main diagonal. The matrix is also symmetric of course. Using these facts, you should see that the variance becomes

$$E\bar{x}^2 = T^{-1}\left((\sigma(0) + 2\sum_{k=1}^{T-1} \frac{(T-k)}{T}\sigma(k)\right)$$
