Problem set 2
607: Applied Macroeconometrics
Fall 2016
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. Review. Define...

   (a) Delta Method

   Answer/comment
   Given a statistic with an asymptotic distribution, the delta method is a way to establish the asymptotic distribution of functions of that statistic. Assume, θ a vector and $g(\theta)$ is a vector function. If we have

   $$\sqrt{T}(\hat{\theta} - \theta) \overset{a}{\sim} N(0, V)$$

   then

   $$\sqrt{T}(g(\hat{\theta}) - g(\theta)) \overset{a}{\sim} N(0, G(\theta)VG(\theta)')$$

   where G is a matrix first derivative of g. This holds where the derivatives are continuous. To sketch a proof, take a Taylor series approx:

   $$g(\hat{\theta}) \approx g(\theta) + G(\theta)(\hat{\theta} - \theta)$$

   Waving our hands and rearranging, we have

   $$\sqrt{T}(g(\hat{\theta}) - g(\theta)) \approx \sqrt{T}G(\theta)(\hat{\theta} - \theta)$$

   which we could imagine would give the result stated above.

   Like all asymptotic results, one must ask whether the approximation will be good in your case. Suppose we know that the the asymptotic normality of $\hat{\theta}$ describes the behavior of $\hat{\theta}$ well in the case at hand, then we can do some reasonable thinking about how accurate the Delta method approximation will be by investigating g in a region of what we believe to be the true θ to see how good the first order Taylor series approx. is. Basically, the approx. will be no better than the first order Taylor series approx. to g.

(b) Likelihood ratio test

Answer/comment

Here it is important to investigate two cases: when we have a point null and alternative hypothesis and when we have composite hypotheses instead. Of course, the Neyman-Pearson lemma shows that a likelihood ratio-based test will be most powerful under certain assumptions in the case of point hypotheses.

In the point hypotheses case, suppose $L(\theta|X)$ is the likelihood, null is $H_0 : \theta = \theta_0$, an alternative of interest is $H_K : \theta = \theta_K$. The LR test would reject H_0

$$\frac{L(\theta_0|X)}{L(\theta_K|X)} < \eta$$

for some fixed η.

Of course, any monotonic transform of the statistic gives an equivalent test, and we often consider minus twice the log of the likelihood ratio and reject for large values.

When the two hypotheses are composite instead, null is $H_0 : \theta \in \Theta_0$, an alternative of interest is $H_K : \theta \in \Theta_K = \Theta - \Theta_0$ (where $\theta \in \Theta$ is maintained), the LR test is generally computed as

$$\frac{\sup_{\theta \in \Theta_0} L(\theta|X)}{\sup_{\theta \in \Theta} L(\theta|X)}$$

and once again minus twice the log is the often the stat. reported.

Under certain assumptions, this statistic will be asymptotically $\chi^2_{(p)}$ under the null, when the null imposes p restrictions.

The Neyman-Pearson Lemma does not apply in the composite case: different tests will be most powerful against different θs in Θ_K. Various asymptotic optimality can be proven, however. For example, the LR test will often be *locally most powerful invariant*. See Engle's Chapter in the Handbook of Ecmet (Vol. 2, ch. 13). The modifiers 'locally' and 'invariant' limit the class of alternatives and of test statistics we are considering, respectively.

(c) Two stage least squares estimator (just give formula)

Answer/comment
Simplest case,

$$y_t = x_t'\beta + \varepsilon_t$$

where x_t may be correlated with ε_t. We also have z_t, which is correlated with x_t but not ε_t.

The 2SLS estimator of β is

$$\hat{\beta} = (\hat{X}'\hat{X})^{-1}(\hat{X}'y)$$

where $\hat{X}$ is the predicted value of X from a first stage regression of the Xs on the Zs:

$$\hat{X} = Z(Z'Z)^{-1}Z'X = N_Z X$$

This is an example of an instrumental variables estimator where the Z are instruments.

---

2. Suppose data are generated by,

$$\begin{aligned} w_t &= \rho w_{t-1} + \varepsilon_t \\ x_t &= \nu_t \\ y_t &= w_t + x_t \end{aligned}$$

where ν_t and ε_t are mutually orthogonal, not serially correlated, have mean zero and constant variance σ_ν^2 and σ_ε^2.

(a) What is the autocorrelation function for w?

Answer/comment
Since w_t is an $AR(1)$, we can follow the class notes to find that for w_t, $\gamma_i = \rho^i$.

(b) What is the autocorrelation function for y?

Answer/comment
Since $Ey_t = 0$ for all t, we have that the autocovariance at lag

k is $\sigma(k) = E[y_t y_{t-k}]$ and by definition, the autocorrelations are $\gamma(k) = \sigma(k)/\sigma(0)$.

We have

$$\sigma(0) = Ew_t^2 + Ex_t^2 + [\text{terms that are 0 in expectation}]$$

So using our knowledge of the variance of an AR(1) we have,

$$\sigma(0) = \sigma_\varepsilon^2/(1-\rho^2) + \sigma_\nu^2$$

Then we have

$$\sigma(1) = E[y_t y_{t-1}] = E[w_t w_{t-1}] + [\text{terms that are 0 in expectation}]$$

Or,

$$\sigma(1) = \rho\sigma_\varepsilon^2/(1-\rho^2)$$

and similarly

$$\sigma(k) = \rho\sigma(k-1) \quad k > 1$$

or more explicitly,

$$\sigma(k) = \rho^k \sigma_\varepsilon^2/(1-\rho^2) \;\; k \geq 1$$

The autocorrelations follow by direct division.

(c) Suppose $\rho = 1$. The process for $z_t = y_t - y_{t-1}$ is an ARMA(p, q) , where p is the order of the AR part and q is the order of the the MA part. What are p and q?

Answer/comment
Note that $(1-L)y_t = \varepsilon_t + (1-L)\nu_t$. The autocorrelations are zero after lag 1; thus, this is an MA(1): $p = 0$, $q = 1$.

Side note: that there is always an ambiguity about what a process is in that one can always multiply both sides by any common factor:

If,

$$A(L)y_t = B(L)\varepsilon_t$$

then it is also true that

$$C(L)A(L)y_t = C(L)B(L)\varepsilon_t$$

So when we ask what are p and q, we are implicitly including 'after cancellation of any common factors.'

---

3. Suppose that income follows the process

$$y_t = A(L)\varepsilon_t$$

$$A(L) = \sum_{i=0}^{\infty} a_i L^i, \quad a_0 = 1$$

(a) What is the variance of income?

Answer/comment

$$\text{var}(y_t) = \sigma_\varepsilon^2 (\sum_{i=0}^{\infty} a_i^2)$$

(b) What effect does a shock, an ε, have on the level of income in the long-run?

Answer/comment
Any shock ε_t has effect $a_i\varepsilon_t$ on y at time $t+i$, as $i \to \infty$ this effect goes to zero if the process is stationary. That is, if the variance is finite, then $\lim_{i\to\infty} a_i = 0$.

(c) Suppose we receive a shock $\varepsilon_t = 3000$. Give a simple expression for the change in the present value of income at t using the discount factor β to discount future income.

Answer/comment
We have that the shock adds income a_i at $t + i$ so, in total the shock adds

$$\sum_{j=0}^{\infty} \beta^j a_j \varepsilon_t$$

I probably was not clear enough in asking the question, but the simplest expression of this present value is probably,

$$A(\beta)\varepsilon_t$$

That is, replace the L in the lag polynomial with β. This question is here just to get you used to the fact that these lag polynomials are polynomials.

(d) Suppose instead that

$$(1-L)y_t = B(L)\varepsilon_t$$

where $B(1) \neq 0$. In this process, what is the long-run effect of a shock on the level of income?

Answer/comment
Note that $(1-L)y_t$ is simply the change in y_t. The change at $t+j$ is given by $b_j\varepsilon_t$, and the total change is just the sum of all the individual changes. This will be $B(1)\varepsilon_t = \sum b_j\varepsilon_t$. Once again B is a polynomial and we can stick a number in for L.

4. Variance-covariance matrices. In this problem, we have a vector, Y, $(T \times 1)$. $EY = 0$, $EYY' = \Omega$, $Y = (y_1, \ldots, y_T)'$.

(a) Show that the variance of the sample mean is the average value of all elements in Ω.

Answer/comment
The mean is $\bar{Y} = T^{-1}i'Y$ where i is a column of ones. Since the mean is zero, by definition, the variance will be

$$E\bar{Y}^2 = T^{-2}Ei'YY'i = T^{-2}i'\Omega i$$

which is the average element of Ω.

(b) Suppose that the ys form a Martingale Difference Sequence. What is the structure of Ω. Give an expression for the average element.

Answer/comment
For an MDS, $Ey_t|y_{t-} = 0$. Each off-diagonal element of Ω is of the form,

$$Ey_ty_s$$

where $t \neq s$. Thus, the off-diagonal elements are zero. The main diagonal will have positive numbers. The average element will be $T^{-1}\bar{\sigma}^2$ where $\bar{\sigma}^2$ is the average of the variances (on the main diagonal).

(c) Suppose in addition that the ys are covariance stationary. What is the structure of Ω. Give an expression for the average element.

Answer/comment
In this case, the variances down the main diagonal are all the same value, say, σ^2 and so the sum is $T^{-1}\sigma^2$.

(d) Suppose that we drop the MDS assumption, but maintain covariance stationarity. What is the structure of Ω. Give an expression for the average element.

Answer/comment
In the covariance-stationary case, Ω has the j^{th} autocovariance on the j^{th} diagonal numbering out from zero on the main diagonal. The average element will be:

$$T^{-1}(\sigma(0) + 2\sum_{j=1}^{T-1} \frac{T-j}{T}\sigma(j))$$

where j is the j^{th} autocovoriance.

(e) The ys are covariance stationary. Show that the sample mean can be computed as $\hat{\beta}$ from an OLS regression on a constant. Is OLS an efficient estimator for the population mean?

Answer/comment
A regression on a column of ones gives,

$$(i'i)^{-1}i'Y = T^{-1}i'Y = \bar{Y}$$

Writing this as a regression, we have

$$Y = i\mu + U$$

and the true μ is zero, so that $EUU' = EYY' = \Omega$.

It should cross your mind that, you should think that GLS would be more efficient:

$$(i'\Omega^{-1}i)^{-1}(i'\Omega^{-1}Y)$$

Folks seldom use GLS (that is, FGLS) to estimate sample means, a fact that might surprise you. As recently as the late 1960s, the best minds in the field were working out expressions for the relative efficiency of OLS and GLS in cases like the AR(1), see Chipman, et al. (1968).

One nice thing about these papers is that they reveal much about the structure of covariance stationary processes in cases that can be easily (by modern standards) understood.

For example, for the AR(1) case, the structure of the implied Ω implies that the GLS estimate of the mean is

$$\bar{y}_{GLS} = \frac{y_1 + y_T + (1-\rho)\sum_{t-1}^{T-1} y_t}{2\rho + T(1-\rho)}$$

This, of course, is almost the sample mean. Ignoring the first and last observation, this is,

$$\frac{\sum y_t}{T + \frac{2\rho}{1-\rho}}$$

You should be able to see that the sample mean is asymptotically efficient for any fixed $|\rho| < 1$.

More generally, the sample mean is asymptotically efficient when Y is covariance stationary. Grenander (1954) is a lovely article on this. Why does this asymptotic efficiency arise?

The answer starts with an exact result that OLS and GLS correspond when the K regressors are fixed and are each linear combinations of K of the eigenvectors of Ω. This is not a time series result, and can be stated many ways, but in reviewing this area, Puntanen and Stayen (1989) say that the great time series econometrician T.W. Anderson first established this and that it led to the famous papers by Durbin and Watson deriving, among other things, the Durbin-Watson statistic.

As it applies to the sample mean, this result about equivalence of GLS and OLS requires that our regressor, a column of 1s, is (proportional to) an eigenvector of Ω, which in turn means $\Omega i = \kappa i$ for some κ–thus, all row sums the same. (Note, Ω and Ω^{-1} have the same eigenvectors.)

For finite order, stationary $AR(p)$ processes, it is easy to show that all but the first and last P rows have the same row sum.

Thus, asymptotically almost all the rows have the same sum and OLS is efficient. (Note from above that OLS treats all but the first and last observation right in the AR(1) case.) To see that the same thing holds in the MA case is a bit trickier, but you might take it on faith it that every MA can be approximated arbitrarily well by some finite AR process.

There is a bit more history for you. If you want to understand covariance-stationarity more deeply, go through these quite readable papers.

References:

John S. Chipman, Koteswara Rao Kadiyala, Albert Madansky and John W. Pratt, Journal of the American Statistical Association, Vol. 63, No. 324 (Dec., 1968), pp. 1237- 1246.
http://www.jstor.org/stable/2285880

Grenander, Ulf. On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance. Ann. Math. Statist. 25 (1954), no. 2, 252–272.
http://projecteuclid.org/euclid.aoms/1177728784

Simo Puntanen and George P. H. Styan. The Equality of the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator, The American Statistician, Vol. 43, No. 3 (Aug., 1989), pp. 153-161
http://www.jstor.org/stable/2685062

5. Many macroeconomic data series are heavily revised after they are first released. Series such as GDP are never ‘final’ and continue to be revised indefinitely.

   (Note: For this problem you might want consult, News and Noise in G-7 GDP Announcements, Faust, Rogers, Wright Journal of Money, Credit and Banking, Vol. 37, No. 3 (Jun., 2005), pp. 403-419
   http://www.jstor.org/stable/3839161

   NOTE: Some discussion/an answer for this question is coming. . . .

   (a) About how many months following the end of a quarter is GDP growth for that quarter first announced in the United States? In the Euro area? In China?

(b) The Philadelphia Fed's website has a link for downloading the U.S. real GDP (or GNP) data series as it stood in each month since 1965.

Site:

http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files/ROUTPUT/

Data:

http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files/ROUTPUT/ROUTPUTQvQd.xls

Note: In the file, the date at the head of the column indicates the date of the data vintage (the date when the data in the column were current). As you look from left to right on any row you see the history of revision for of the data for the quarter whose date is in col. 1.

Plot the history of the revisions to the growth rate for 1957Q3, 1973Q4, and 1997Q2.

(Note: vertical axis is growth rate for the stated quarter, horizontal axis is the date of the data vintage.)

(c) What was the standard deviation of annualized quarterly GDP growth from 1954:1-1959:1 as measured in the 1965Q4 vintage? And in the 1989Q4 vintage? And in the most recent vintage in the file?