Problem set 3
ANSWERS
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. Define the following terms
   (a) Nominal size of a test
   (b) Power of a test
   (c) Neyman-Pearson Lemma
   (d) Define a consistent test
2. χ^2.
   (a) Define the χ^2 in terms of normals.
   Answer/comment
   If $x_j \sim iidN(0,1)$, then
   $$\sum_{j=1}^{p} x_j^2 \sim \chi^2_{(p)}$$
   (b) If $z \sim N(0,\Sigma)$ is a $(k \times 1)$ vector and Σ is full rank, sketch the argument as to why $z'\Sigma^{-1}z$ is $\chi^2_{(k)}$.
   Answer/comment
   This uses the ideas in our disucssion of generalized distance and quadratic forms. Take a square L such that $L\Sigma L' = I$. Then
   $$w = Lz \sim iidN(0,1)$$
   and
   $$w'w = z'L'Lz = z'\Sigma^{-1}z$$
   is the sum of iid N(0,1) variables.

(c) Sketch the $\chi^2_{(k)}$ density for $k = 1$. How does this differ from the density for higher k?

Answer/comment

The key here is that it is positive with a global maximum at zero. For all other degress of freedom, the mode occurs at a strictly positive point, that point moving rightward as the degrees of freedom grow.

(d) State the mean and variance of the $\chi^2_{(k)}$ as a function of k.

Answer/comment

Using your deep knowledge of $N(0, 1)$ variables and sums thereof, we have that the expectation of a $\chi^2_{(k)}$ is k and the variance is $2k$. (Remember that for $z \sim \chi^2_{(k)}$, $\text{var}(z) = E(z - k)^2$.)

(e) If $x \sim \chi^2_{(k)}$, what happens to the mean and variance of x/k as $k \to \infty$.

Answer/comment

Notice that x/k is the sample mean of k iid variables that have mean 1 and variance 2. Thus, the mean is 1 for all k and the variance is $2/k$, and

$$\sqrt{k}(x/k - 1) \to_d N(0, 2)$$

---

3. F. Suppose $x \sim F_{(k,\ell)}$.

(a) Define the F distribution in terms of other familiar distributions.

Answer/comment

The F distribution can be characterized as the ratio of two independent χ^2 random variables, each divided by its degrees of freedom. This is generally denoted $F_{(d_1,d_2)}$ where d_1 and d_2 are the numerator and denominator degrees of freedom respectively.

(b) For fixed $k = 1$, what does increasing ℓ do to the shape of the density? In the limit for large ℓ?

Answer/comment

The numerator is unaffected, but the denominator, a χ^2 divided by its degrees of freedom is converging to a constant = 1. Thus, as ℓ rises, an $F_{1,\ell}$ random variable converges to a $\chi^2_{(1)}$ random variable.

Many of our stats (e.g., the standard regression F tests) are (asymptotically) distributed $F_{M,T}$, where M is fixed and equal to, say, the number of restrictions being tested and the T in the denominator is the sample size.

As T increases, the denominator becomes a constant and we are left with the numerator $\chi^2_{(M)}$ divided by its degrees of freedom. This gives rise to the common parlance 'the F-form of the statistic' or 'the χ^2 form of the statistic.'

We multiply the F form by the numerator degress of freedom to get the χ^2 form.

You will occasionally see this language relating, say, to Wald tests. Our standard derivation often gives an asymptotically χ^2 statistic, but we can divide by the degrees of freedom and treat the statistic instead as asymptotically F.

Think about which of these gives a more conservative view of rejecting the null hypothesis.

---

4. Take the AR(2) process,

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

where $A(L) = (1 - \rho_1 L - \rho_2 L^2)$. If the process is stationary, then we can also write the process as

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

where $B(L) = A(L)^{-1}$, or

$$B(L)A(L) = 1$$

(a) Assuming $B(L) = \sum_{j=0}^{\infty} b_j L^j$ and using the definition of $A(L)$ and $B(L)A(L) = 1$ (and by matching coefficients on powers of L) find expressions for $b_0, \ldots, b_3$ in terms of ρ_1 and ρ_2.

Answer/comment

The goal here is to simply get you some feel for the inverse of lag polynomials. Start with

$$(\sum_{j=0}^{\infty} b_j L^j)(\sum_{j=0}^{\infty} a_j L^j) = \sum c_j L^j = 1$$

so that $c_0 = 1$ and $c_j = 0$ for $j > 0$. Given our definition of $A(L)$ this simplifies to:

$$(\sum_{j=0}^{\infty} b_j L^j)(\sum_{j=0}^{2} -\rho_j L^j) = \sum_{j=0}^{\infty} c_j L^j = 1$$

with $\rho_0 = -1$.

We can multiply this out to form $B(L)A(L) = C(L)$ and then choose our b_js so that $c_0 = 1$ and $c_j = 0$ for $j > 0$. Since $c_0 = a_0 b_0$, we have, $b_0 = 1$.

The terms in L^1 will have coefficient:

$$c_1 = a_0 b_1 + a_1 b_0$$

or, using the definition of the as and $b_0 = 1$:

$$b_1 - \rho_1 = 0$$

or $b_1 = \rho_1$.

For L^2, we have

$$a_0 b_2 + a_1 b_1 + b_0 a_2 = 0$$

or

$$b_2 - \rho_1^2 - \rho_2 = 0$$

so

$$b_2 = \rho_1^2 + \rho_2$$

For L^3, you should see that we now need all the terms from the cross product that have subscripts adding to 3:

$$a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_1 = 0$$

but of course, $a_j = 0$ for $j > 3$, which will keep all the terms for $j > 2$ simple.

$$b_3 = \rho_1^3 + 2\rho_1\rho_2$$

Note that there are a couple other natural ways to get the MA representation. First, use the fact that the MA representation is the same as the impulse response function. Thus, set all εs to zero except $\varepsilon_0 = 1$. If I just substitute recursively moving forward, we have:

$$\begin{aligned}
y_{-2} &= 0 \\
y_{-1} &= 0 \\
y_0 &= \varepsilon_0 = 1 \\
y_1 &= \rho_1 y_0 = \rho_1 \\
y_2 &= \rho_1 y_1 + \rho_2 y_0 = \rho_1^2 + \rho_2 \\
y_3 &= \dots
\end{aligned}$$

The y_ks here are the impulse response function. Notice that this can be seen as the forecast for y_k seen from time zero when $\varepsilon_0 = 1$ is observed and assuming that all past ys and εs are zero.
By the way, this is the way many programs calculate the moving average representation.
Method 2. Alternatively, you could achieve the same result by inverting the lag polynomial by first factoring the 2nd order polynomial into two first order polynomials and inverting those separately by hand (since that is easy).

$$A(L) = (1 - mL)(1 - nL)$$

for some roots m, n (since this is a second-order polynomial, you can get a, b using the quadratic formula). Let's focus on the case of real roots. Expanding the RHS above, we see that

$$A(L) == 1 - (m + n)L + mnL^2$$

so $\rho_1 = m + n$ and $\rho_2 = -mn$.
Let's invert the factored polynomial using the fact that we know how to invert first order polynomials:

$$\begin{aligned}
A(L)^{-1} &= (1 - mL)^{-1}(1 - nL)^{-1} \\
&= (\sum_{i=0}^{\infty} m^i L^i)(\sum_{i=0}^{\infty} n^i L^i)
\end{aligned}$$

We can multiply this out and collect terms in powers, L^k.

$$\begin{aligned}
g_0 &= 1 \\
g_1 &= (m + n) \\
g_2 &= m^2 + n^2 + mn
\end{aligned}$$

and using the definitions of the ρs in terms of m and n.

BTW: this is a new problem and answer, so it'll need some proof-reading.

(b) Give an expression for the variance of the process in terms of the ρs and variance of ε.

Answer/comment

We could solve for the whole MA representation (the bs) and then write,

$$\text{var}(y) = \sigma_\varepsilon^2 \sum b_j^2$$

but an easier approach is to write:

$$\begin{aligned} y_t^2 &= y_t(\rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t) \\ y_t^2 &= \rho_1 y_t y_{t-1} + \rho_2 y_t y_{t-2} + y_t \varepsilon_t \\ \sigma(0) &= \rho_1 \sigma(1) + \rho_2 \sigma(2) + \sigma_\varepsilon^2 \end{aligned}$$

using our standard definition of the autocovariance function.

Then, multiplying the definition of y_t by y_{t-1} and taking expectations:

$$\begin{aligned} E y_t y_t - 1 &= \rho_1 \sigma(0) + \rho_2 \sigma(1) \\ \sigma(1) &= \rho_1 \sigma(0) + \rho_2 \sigma(1) \\ \sigma(1) &= \frac{\rho_1}{1-\rho_2} \sigma(0) \end{aligned}$$

and similarly:

$$\begin{aligned} \sigma(2) &= \rho_1 \sigma(1) + \rho_2 \sigma(0) \\ \sigma(2) &= (\frac{\rho_1^2}{1-\rho_2} + \rho_2) \sigma(0) \end{aligned}$$

(Note: you can verify that $\sigma(j) = \rho_1 \sigma(j-1) + \rho_2 \sigma(j-2)$ for $j > 1$.)

Now we have $\sigma(1)$ and $\sigma(2)$ in terms of $\sigma(0)$. Substitute in the expression for $\sigma(0)$ and solve to get,

$$\sigma(0) = \sigma_\varepsilon^2 / \Big((1+\rho_2)(1-\rho_2) - \rho_1^2/(1-\rho_2))\Big)$$

All ARs can be solved in this way. If you are interested in AR math, you might go read about the Yule-walker equations, which we are flirting with here.

---

5. MA(1) processes.

(a) Take the MA(1) process:

$$y_t = (1 + \theta L)\varepsilon_t$$

with $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$.

Note: You may want to consult the lecture notes on covariance stationary math.

What is the variance of the process? And the first autocovariance?

$$\begin{aligned} \sigma(0) &= (1 + \theta^2)\sigma_\varepsilon^2 \\ \sigma(1) &= \theta\sigma_\varepsilon^2 \end{aligned}$$

(b) Now consider the process

$$y_t = (1 + \theta^{-1} L)\varepsilon_t$$

with $\varepsilon_t \sim iidN(0, \theta^2\sigma_\varepsilon^2)$.

What is the variance of the process? And the first autocovariance?

Answer/comment

Same as for the previous.

(c) Using a sample of, say, 100 observations could you distinguish these processes?

Answer/comment

No. The observable implications of the two are identical. The first order MA is locally identified, but not globally identified. That all θs in a small neighborhood of the true θ have somewhat different observable implications (local identification), but looking across the parameter space as a whole, for each θ, there is another θ that delivers the same observable implications.

Note that only one of these θs gives rise to an invertible polynomial. One has a root outside the unit circle, one inside.

This fact is closely associated with the pileup problem with maximum likelihood in the MA case. We saw this problem in lecture 2. I don't have a favorite reference here (should fix that), but the following is a nice paper: DeJong, D. N. and C. H. Whiteman, 1993, "Estimating Moving Average Parameters: Classical Pile-ups and Bayesian Posteriors," Journal of Business and Economic Statistics 64, 183-206. http://www.jstor.org/stable/1391955

This gives rise to a deep and often confusing (and sometimes confused) discussion of nonfundamental representations of MA processes. E.g., Lippi, Marco and Reichlin, Lucrezia, 1994. VAR analysis, nonfundamental representations, blaschke matrices, Journal of Econometrics, Elsevier, vol. 63(1), pages 307-325, July. http://www.sciencedirect.com/science/article/B6VC0-4582GY2-2D/2/f79503e0fc4de618ddd1eaf035deee71

(d) Suppose one of these two processes describes some real world process. Can it matter which is 'true'? Explain.

Answer/comment

Let's suppose t is measured in days, that our εs are known to have fat tails, and that our MA(1) describes the impact and first aftershock of an earthquake. Now take $\theta = .1$ so that $1/\theta = 10$. We have a modest earthquake today. If the true MA process has coefficient θ, then tomorrow we'll tend to have a small aftershock. If the true process is driven by $1/\theta$, we should all pack our bags and leave town. Sadly, we never can know for sure without further information.

The problem here is that in this process, big shocks tend to be both preceded by and followed by smaller shocks. You never know if the small shock led to the big shock or vice versa.

Yes, it does matter to you. After an unusually big earthquake today (which by luck you survived), you want to know if a really, really big one is coming tomorrow.

Of course, 'structural analysis,' (in this case, presumably in the form of plate tectonics, etc.) could provide additional information about whether aftershocks should in principle be smaller or larger than the original.

In general, the way to make progress on observational equivalence problems in causal systems is to think harder about the strucutres through which causality manifests itself.

---

6. The class presentation.

   Obtain monthly data for 3-month nominal interest rates in the US and UK and data for the dollar-pound exchange rate.

   (a) Plot the interest rates and the log change ($ln(x_t) - ln(x_{t-1})$ in the exchange rate.

   (b) When s_t is the log exchange rate and i_t and i_t^* are home and foreign 1-period interest rates, the uncovered interest rate parity relation is:

   $$E_t s_{t+1} - s_t = i_t - i_t^*$$

   Give an interpretation of this relation and explain why we might expect something like this relation to hold.

   Answer/comment

   Think of home and foreign currencies as dollars and pounds, respectively, and define e_t as the price of foreign currency (that is, e is stated as dollars per pound).

   Suppose I take a dollar today and invest it for one period at the rate i_t. I will get back $1 + i_t$ dollars tomorrow. If this is a nominally riskless investment, then I get $1 + i_t$ for certain.

   Instead, I could change the dollar into pounds, invest it at i_t^* and change back to dollars at $t + 1$.

   In expectation, this will give me

   $$E(1 + i_t^*)e_{t+1}/e_t = (1 + i_t^*)Ee_{t+1}/e_t$$

   dollars. The way I have written this, both i_t and i_t^* are known at t (nominally riskless) and so are outside expectations. Suppose that we knew e_{t+1} at time t as well. In this case, if I were to be indifferent between the two investments, it would have to be true that

   $$1 + i_t = (1 + i_t^*)e_{t+1}/e_t$$

   Rearranging gives,

   $$\frac{1 + i_t}{1 + i_t^*} = \frac{e_{t+1}}{e_t} \tag{1}$$

   Taking logs, and using the fact that $\ln(1 + x_t) \approx x_t$ for smallish x (note: this is a first order Taylor series approx.), we have

   $$s_{t+1} - s_t \approx i_t - i_t^*$$

where $\ln(e_t) = s_t$.

We don't know the future exchange rate in reality, to get the standard UIP relation in this case we need that

$$\ln(Ee_{t+1}) \approx E\ln(e_{t+1})$$

which essentially requires that the Jensen's inequality effect is small.

In all this approximating, what we are doing in economic terms is neglecting risk premia. To see this in a simple model, suppose a rational agent's choices satisfy the consumption Euler equation:

$$u'_t/p_t = \beta Eu'_{t+1}(1+i_t)/p_{t+1}$$

u'_t is marginal utility at t, p_t price of consumption, β discount factor, i_t the nominal interest rate between t and $t+1$. All prices and nominal rates in home currency.

If the agent's Euler equation also holds for foreign investments,

$$u'_t/p_t = \beta E(u'_{t+1}/p_{t+1})(1+i^*_t)\frac{e_{t+1}}{e_t}$$

The two Euler equations jointly imply

$$(1+i_t)Eu'_{t+1}/p_{t+1} = (1+i^*_t)E(u'_{t+1}/p_{t+1})\frac{e_{t+1}}{e_t}$$

where I have moved the interest rates outside the expectations operator. This is appropriate if they are known at t (that is nominally riskless). Re-arrange,

$$\frac{1+i_t}{1+i^*_t} = \frac{E(u'_{t+1}/p_{t+1})(e_{t+1/e_t)}}{Eu'_{t+1}/p_{t+1}}$$

$$\frac{1+i_t}{1+i^*_t} = \frac{E[e_{t+1}/e_t]E[u'_{t+1}/p_{t+1}] + \mathrm{cov}(u'_{t+1}/p_{t+1}, e_{t+1/e_t)}}{Eu'_{t+1}/p_{t+1}}$$

where the second line exploits $\mathrm{cov}(a,b) = Eab - EaEb$. Or cancelling,

$$\frac{1+i_t}{1+i^*_t} = Ee_{t+1}/e_t + \frac{\mathrm{cov}(u'_{t+1}/p_{t+1}, e_{t+1/e_t)}}{Eu'_{t+1}/p_{t+1}}$$

The covariance term is interpreted as a risk premium. If we ignore this, we have equation (1) above. Thus, this is the sense in which that equation ignores a risk premium. The final step of approximation to the standard statement of UIP is as before.

(c) To test the UIP relation, researchers often run a regression

$$s_{t+1} - s_t = \alpha + \beta(i_t - i_t^*) + \varepsilon_t$$

and test $\alpha = 0$ and $\beta = 1$. Under what assumptions about expectations is this a sensible test?

Answer/comment

Under rational expectations, ignoring the approx. error, and the risk premium,

$$Es_{t+1} - s_t = (i_t - i_t^*) + \varepsilon_t$$

where ε_t is uncorrelated with things known at t (and thereby, with $i_t - i_t^*$).

(d) What modifications/transformations to your data are needed to properly implement the test just described? Think about the following: Your interest rate data may be in annualized percent change. You have three-month rates measured at a monthly frequency. Exchange rate may be stated in dollars per pound or the other way. Would you rather use monthly data that is on an average-of-daily basis or end-of-month basis? Which do you have?

Answer/comment

We have three month interest rates, so the change in the exchange rate must be over three months. As for the units, both the change the exchange rate and the interest differential should be stated either as percent (e.g., 5 for 5 percent) or decimals (0.05 for 5 percent). Otherwise our interpretation of the magnitude of β is screwed up. A version of the equation should hold every day of the month; thus, the equation should hold in data measured as, say, the value for the last day of the month.

How about monthly average data? Well, the equation makes a valid regressions for each day in the month, and so you might think it would hold in monthly average data. The trouble is that time-averaging may create simultaneity bias.

Suppose there are two days in the month and a version of the equation holds for each day with the error orthogonal to the regressor (the interest differential). But suppose that on the first day of the month, there is a big ε affecting the exchange rate. On the second day, the central bank responds and raises interest rates. Because of the feedback between εs yesterday and interest rates today, the two day average of the interest differential and the daily ε will be correlated.

Write down the equations and verify this for yourself.

The idea of this exercise is that you have to be careful about minor data issues: units, which way is the exchange rate stated, monthly average vs. end of month, etc.

Care on these points is as important as care about statistical issues.

It turns out that this regression has many pathologies and we have just touched on some here. For example, under the model sketched above, the interest rate differential and change in the exchange rate should have similar variance. If the exchange rate variance is higher, it is due to ε: news arriving that makes the exchange rate different from expected. In practice, exchange rates are much more variable than the differential—so variable that our models have trouble explaining it. Not only is the exchange rate too variable, it's correlation with the differential is often wrong: we get a -1 when we expect $+1$ in the UIP regression. These are major puzzles in international finance and typical of what we find when we take simple theories to financial market data.

(e) Report results of the regression above for various sample periods in your data (e.g., whole sample period, split in the middle, etc.

Answer/comment

See the class presentation.

(f) For the presenter: using Fama's article show how the above results have implications for the sign and variance of the risk premium.

cite: Fama 1984 E. Fama, Spot and forward exchange rates. Journal of Monetary Economics, 14 (1984), pp. 319–338.

http://dx.doi.org/10.1016/0304-3932(84)90046-1

note: Google `UIP puzzle` for many papers on this topic.