Problem set 5
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. Definitions

   (a) Consistent test
   Answer/comment
   We asked this before, but it is relevant to the next two parts, so what's wrong with a little repetition?

   (b) Noncentral χ^2 distribution.
   Answer/comment
   The noncentral χ^2 with k degrees of freedom can be defined as the sum of k independent squared normals that have unit variance but nonzero mean. Thus, the noncentral χ^2 differs from the conventional χ^2 in that the underlying normals variables have nonzero mean. The distribution has two parameters, the degress of freedom, k, and the *noncentrality* parameter, which is not uniformly defined in the literature, but is always something like,

   $$\sum_{j=1}^{k} \mu_j^2$$

   where μ_j is the mean of the j^{th} normal.

   (c) What does the noncentral χ^2 distribution have to do with the consistency of standard tests that are asymptotically χ^2 under the null hypothesis.

Answer/comment

Note: this answer is a sketch and presumes that you have seen this in an earlier course. If not, you should read up on this. An excellent source is McFadden's discussion (e.g. Sec. 7.6, p. 167, `http://eml.berkeley.edu/~mcfadden/e240a_sp01/ch7.pdf`.

Take a standard test statistic, $\phi(Y)$ that is distributed exactly $\chi^2_{(k)}$ under the null for every sample size. Exactness is only a simplification here, we can repeat the asymptotic version of the argument. This test will reject when $\phi > c$ for some critical value c. If the test is consistent, then it must be the case that however the null is false, the probability that $\phi > c$ goes to 1 with sample size. In other words, all of the mass of the distribution of ϕ must be moving rightward with sample size.

This χ^2-distributed test statistic generally arises through our ubiquitous quadratic form. In the simplest case:

$$W = T(\hat{\beta} - \beta_0)'V^{-1}(\hat{\beta} - \beta_0)$$

Where the null is $H_0 : \beta = \beta_0$, and under the null hypothesis,

$$\sqrt{T}(\hat{\beta} - \beta) \sim N(0, V)$$

(For simplicity, I am leaving aside that we usually have to use a consistent estimate of V.)

If truth is β_k so that

$$\sqrt{T}(\hat{\beta} - \beta_k) \sim N(0, V),$$

then our Wald statistic is distributed as a noncentral χ^2 with k degress of freedom. The noncentrality is $T\psi'\psi$ where $\psi = \beta_k - \beta_0$. Thus, we can compute the power using the noncentral χ^2 and we know that power will be determined by the noncentrality parameter. With the noncentrality parameter growing (moving to the right), we see consistency.

---

2. Suppose that the maintained model is an invertible, Gaussian MA(1):

$$y_t = \alpha + \varepsilon_t + \theta\varepsilon_{t-1}$$

$$(1 + .5L)\varepsilon_t \sim iidN(0, 1)$$

(a) You attempt to fit this using an AR model. What is the 'true' order of AR matching the MA(1).

Answer/comment

Infinite

(b) Run a Monte Carlo for $\theta = [-.9 - .5 - .1, 0 .1 .5 .9]$ and sample sizes $[20, 50, 200]$. Generate the samples using the DGP. On each sample:

- Estimate an AR selecting the lag length by both AIC and BIC, where the maximum lag is $T/5$. Save the selected lag length.
- For the chosen models, save the sum of the AR coefficients the estimated variance of ε.

Note: Once again, I've provided some code to start from. After this problem set, it'll mainly be up to you. There are some things to complete in the code, marked ???. There are probably other errors (this is a new problem and I'm letting you debug the code).

I've included my code to compute the ARs. It is VAR code, but an AR is just a 1 equation VAR, so that should be fine. The VAR routines (basicVAR, basicVARPickLag, prepVAR) should be fine (no errors). You may want to read them in any case.

(c) Summarize which criterion did best from the standpoint of

- Picking the correct lag length

  Answer/comment

  This is ill defined: both are arbitrarily bad. I ask this to remind you that picking the 'right' order is beside the point. We are always picking a finite order in a world in which the true order may be arbitrarily large. AIC has a smaller penalty for parameters, however, and picks a longer lag length.

- Bias, variance, and mean squared error in estimating the sum of the AR coefficients (hint: for each DGP, you'll have to figure out what the sum of AR coefficients should be.)

  Answer/comment

  The MA is $y_t = \alpha + \theta(L)\varepsilon_t$. The AR is $\theta(L)^{-1}y_t = \varepsilon_t$ so the sum of the AR coefs. is $\theta(1)^{-1}$. You should find that neither criterion really dominates in terms of estimating this sum. Both may lead to bad estimates, but neither is much better or worse.

- Bias, variance, and mean square error in estimating the variance of ε.

  Answer/comment

  In this case, the BIC sometimes leads to substantially smaller bias (e.g. 25% smaller). Of course, in this abstract setting, we can't say whether this is a meaningful difference.

(d) As a general rule, which criterion should you use?

Answer/comment

There is no general rule. Your question generally isn't 'what is the right lag length.' For any given question of interest, which criterion leads to the most reliable answer may differ. As Amemiya said in the lecture notes: choose wisely.

---

3. Read Campbell, J.Y. and N.G. Mankiw (1987): Are Output Fluctuations Transitory?, Quarterly Journal of Economics, 102, pp.857-880. You may also want to consider Gagnon, Joeseph, Short-Run Models and Long-Run Forecasts: A Note on the Permanence of Output Fluctuations The Quarterly Journal of Economics, Vol. 103, No. 2 (May, 1988), pp. 415-424. (Links to both articles on the course reading list.)

   (a) Are output fluctuations permanent or transitory?

   (b) Discuss virtues and limitations of Campbell and Mankiw's approach especially noting the Gagnon argument.

   (c) Make a case that macroeconomists should be permanently banned from using the word *permanent*.

   Answer/comment

   We discuss this paper in class, but let me add some note here on all three questions.

   The Campbell Mankiw paper is a beautiful exposition of how thoughtful, well-trained economists were approaching applied macroeconometrics at the time. In particular, it is a primer in all the basic time series issues we have been discussing: ARMA modelling, model selection, unit MA roots and the pileup problem, Monte Carlo as an approach to shedding light on issues not currently resolved in theory. It flirts with a topic we have also only flirted with so far, the way unit AR roots affect the distribution of statistics.

The paper is fundamentally misguided in two simple ways.

First, before pulling out cutting-edge statistical theory and techniques and producing lots of tables, you should always ask whether the answer to the question at hand could possibly be in the data. That is, can the available sample be informative regarding the question at hand?

In this case, the answer is clearly no. Put it this way, can a 30 or 50 year snapshot tell us anything reliable about the permanence of things? No. If it did, we'd probably conclude that incandescent light bulbs were a permanent feature of human society.

A deeper question is whether practical people should care about permanence. This is debatable, but the answers to most policy questions (personal policy and/or aggregate macro policy) don't much turn on this question. This perspective led to a famous paper, Unit Roots in GDP. Do we know and Do we care? We probably do care about dynamics over horizons like a quarter or a few years. For fiscal sustainability issues, we may even care about 100 years or so. But permanence is another matter. Think precisely about the question you are asking and then consider whether the available data have any hope of shedding light.

Second, this paper does not fully take on board the idea that the approach you take to specification search may strongly affect inference about your question of interest. Suppose your goal is to estimate the first order autocorrelation in quarterly data. If you have, say 50 years of data, then about any reasonable method will work. All of the models estimated by Campbell and Mankiw would have given about the same answer on this question so the answer is not at all sensitive to the model selection choice.

The long-run affect of shocks, however, is extremely sensitive to model selection. A good way to proceed in this regard is to think about worst case scenarios for the approach you are taking. If you find a case that you cannot rule out *a priori* but in which your method completely fails, then you probably should go find another approach or topic.

For Campbell and Mankiw, the worst case is intuitively pretty clear: shocks decay slowly beginning, say, after 5 years. Thus, a shock that has an effect of +1 on growth initially is followed much later by, say, 10 years of tiny negative effects on growth that totally offset the original impulse. Gagnon discusses how this would imply a long sequence of very small negative autocorrelations. In relevant sample sizes, these will be poorly estimated and insignificantly different from zero. There is no hope of answering the question Campbell and Mankiw ask. Put narrowly, the sum of autocorrelations 10 through 50 is very sensitive to model selection. Indeed, there is probably no reliable way to estimate these in relevant sample sizes.

As an aside, students often wonder about the process of getting projects. One way is to focus on fundamental issues and ask whether you can advance the discussion by bringing fundamental issues to bear: What is the worst case? Under what assumptions would standard approaches be optimal? These two questions led fairly directly to two of my first papers, one of which started as a thesis paper.

Lawrence J. Christiano, Martin Eichenbaum, Unit roots in real GNP: Do we know, and do we care?, Carnegie-Rochester Conference Series on Public Policy, Volume 32, 1990, Pages 7-61, ISSN 0167-2231, http://dx.doi.org/10.1016/0167-2231(90)90021-C. http://www.sciencedirect.com/science/article/pii/016722319090021C

Faust, Jon. When Are Variance Ratio Tests for Serial Dependence Optimal?; Econometrica, September 1992, v. 60, iss. 5, pp. 1215-26. http://www.jstor.org/stable/2999547

Faust, Jon. Conventional Confidence Intervals for Points on Spectrum Have Confidence Level Zero; Econometrica, May 1999, v. 67, iss. 3, http://www.jstor.org/stable/2951545

. . .