607
Parsimonious approximation of dynamics and model selection

Jon Faust

http://e105.org/e607

October 7, 2015

▶ **Readings**

- Jonathan's notes and Hansen's text are both good on this
- I also like Herman Bierens's notes

  See the model selection section of course website

▶ **Background**

- Cov. stationarity is a pretty strong restriction
- But suppose we are willing to make this assumption
- We still face a difficult problem: The covariance stationary model is still infinitely parameterized (there is a process for each element in ℓ^2)
- Remember the lecture on why applied macro time series is hard
- Samples are short relative to the duration of relevant dependence
- And theory does not provide much specific guidance about dynamics
- Thus, we won't get a lot of help from data or theory in nailing down our infinite paramater model.

▶ **Aside:: In particular,**

- In particular, we might view our problem in the standard semi-parametric way

  Infinite-dimensional nuisance parameter.

- And then elaborate adaptive estimators could be explored.
- But in our short samples, elaborate often doesn't help

▶ **Note also**

- Many of the questions of interest do not involve getting the details of the dynamics exactly right
- Often we are interested in just getting an adequate parsimonious approximation to the dynamics as a background step in doing reliable inference on a question of interest.
- This is our goal, for example, when we are fitting the dependence in the ws in order to compute HAC variance-covariance matrices.

  That is, the ws from writing $(\hat{\beta} - \beta) = \bar{Q}^{-1}\bar{w}$.

▶ **Earlier stated fact**

- Let's start from an earlier stated fact
- The second order properties of any cov. stat. process can be approximated arbitrarilyvery well by some finite-order AR and some finite-order MA process
- Thus, in much of our work, we choose which family to use (AR(p),MA(q),ARMA(p, q)) and then model selection comes down to choosing p and/or q.
- Let's turn to this topic for some intution

▶ **Choosing AR or MA**

- What does autocorrelation fuction of AR(1) ($\rho > 0$) look like?

  ρ, ρ^2, ...

- And an MA(1)?

  Nonzero at lag 1, zero otherwise.

- What order AR does it take to exactly match an MA(1)?

  AR(∞)

- Similarly, takes $MA(\infty)$ to exactly match AR(1)

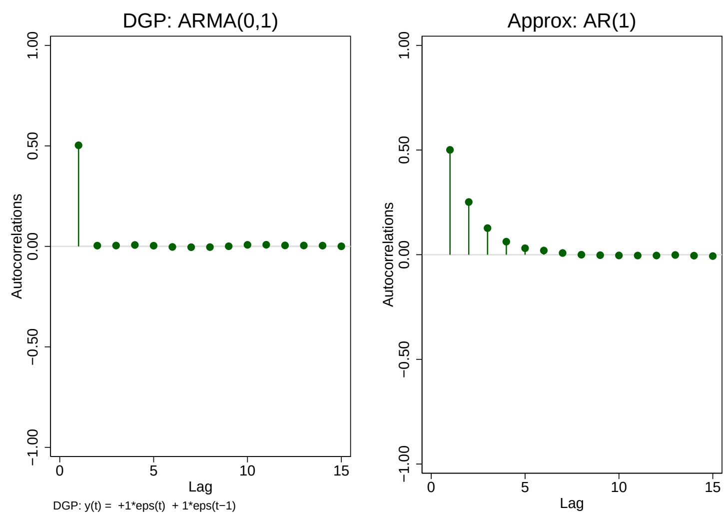
▶ **Thus,**

- We have several infinitely parameterized models of the cov. stat. AR, MA, etc.
- Whether or not an exact parsimonious models exists depends very heavily on happening to pick the right model
- It may seem like we also need to pick our approxmating model very astutely if we want a good *parsimonious* approx.

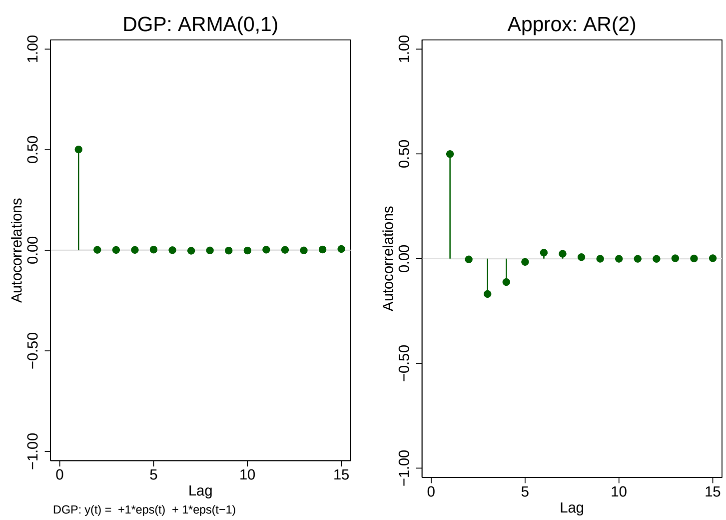
- This is true

  But perhaps not as true as one might think

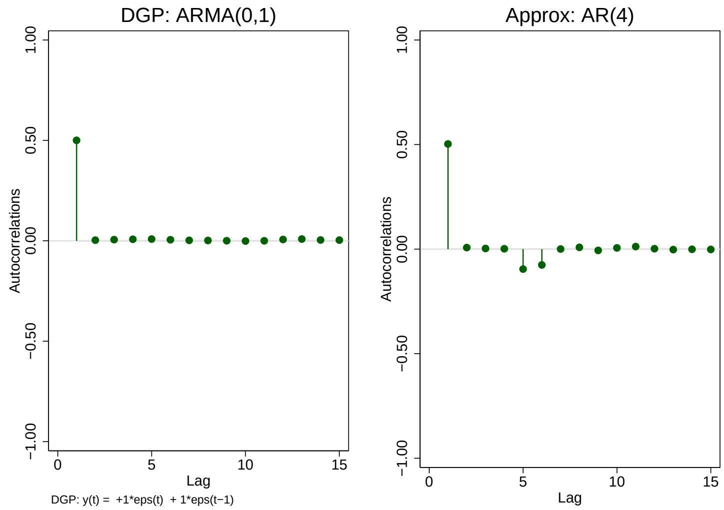
▶ **MA(1) approx. by AR(1)**



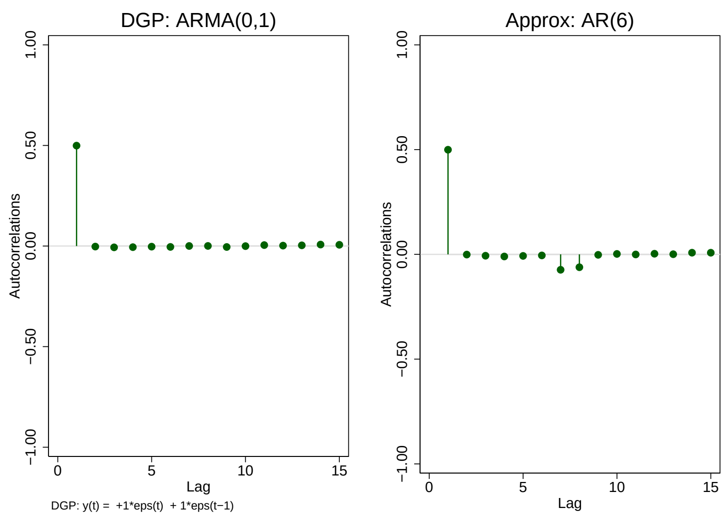
▶ **MA(1) approx. by AR(2)**



▶ **MA(1) approx. by AR(4)**



▶ **MA(1) approx. by AR(6)**



▶ **Aside:: Note**

- The STATA program that generated these pictures is saved with the lecture

  `viewARMA.do`

- The program allows you to state any ARMA and order for the AR approx. and will produce pictures like those just reported
- This is a 'quick and dirty' program, read the warning labels in the code

▶ **Lesson?**

- In much of applied macro, we tend to count one, two, three, infinity.

  That is, we treat rather smallish numbers as if they are close to infinite

- Often this is a bad practice, but AR and MA approximations, it is not so bad
- It takes an AR(∞) to exactly match an MA(1), but an $AR(4)$ does a pretty darn good job.

▶ **Aside:: Box-Jenkins**

- Box and Jenkins laid out an elaborate recipe for approximating dynamics using the ARMA family of models
- Google `Box Jenkins` for several summaries
- You made lots of subjective decisions in choosing the AR and MA orders

▶ **Aside::**

- This approach isn't emphasized much anymore

  Although current methods build heavily on insights gained by Box and Jenkins

- We have moved on for various reasons, good and bad

  But I leave this to the history of thought.

▶ **Model selection**

- Model selection in fitting dynamics often comes down to choice among AR, MA, and ARMA models
- And then choosing the order.
- Given the model, the choice of the order is an example of the much studied problem you may have seen in elementary economics: choice of regressors or model selection more generally.

▶ **Basic approach**

- We know that as we increase the number of regressors, the fit (as measured, say, in terms of the sum of squared errors) increases.
- And when $K = T$—that is, number of regressors equals number of observations—the fit is perfect

  ...and the estimates have become worthless.

- Thus, most recipes for data-based model selection involve maximizing some notion of fit while including some penalty for the number of parameters used to attain the fit.

▶ **Recipe**

- Choose a measure of fit

  For example, the sum of squared residuals or the likelihood. In Gaussian OLS case, these are equivlent.

- Choose a penalty for increasing model flexibility or degrees of freedom consumed
- Pick the model that maximizes penalized fit.

▶ **Example: Akaike Information Criterion**

- The Akaike information criterion (AIC) in regression case:

$$AIC = \ln(SSR/T) + 2p/T$$

  SSR, sum of squared residuals, p is number of parameters.

- Fit: $\ln(SSR/T)$, penalty: $2p/T$.

▶ **Others:**

- Bayesian information criterion:

$$BIC = \ln(SSR/T) + p\ln(T)/T$$

- Hannan-Quinn criteiron:

$$HQ = \ln(SSR/T) + 2p\ln(\ln(T))/T$$

▶ **Aside:: In practice**

- Choose a set of models, say, the AR family with order between $\underline{\text{P}}$ and $\bar{P}$.
- Estimate each model; calculate the loss under the chose criterion
- Pick model with lowest loss.

▶ **Optimality**

- We can prove various optimality properties for these criteria
- For example, we can show that BIC is consistent in that it if the truth is $AR(P^*)$, then BIC will choose P^* with probability approaching 1 as the sample size rises.
- In contrast, AIC asymptotically tends to choose a model that has order greater than P^*
- As always, once we have one item that is consistent, we have actually defined an arbitrarily large set of consistent criteria

▶ **Aside:: Consistency**

- As the Wright notes show, any criterion of the form

$$\ln(SSR/T) + pg(T)$$

will be consistent so long as

$$g(T) \to 0 \;\; \& \;\; Tg(T) \to \infty$$

as T rises.

▶ **Aside:: A warning about optimality**

- Once theorists have been able to prove something that might be called an *optimality* property, those properties get emphasized.

    Thus, BIC is often talked of as if it is better than AIC because it is consistent.

- But remember, our goal is not to pick the 'right' model
- Our goal is parsimonious approximation in order to move on to some question of interest about which we'd want to get the 'right' answer.

▶ **Aside::**

- For example, suppose you have 60 observations

    15 years of quarterly data.

- And God shares with you that the true process is AR(27).
- You know truth, so consistency in model selection is not your problem.
- But you have 60 observations: your problem is parsimonious approximation of an unknown 27 parameter model.

▶ **Aside::**

- In this area 'truth' is over rated.

▶ **An important perspective**

- You need to have an objective—say, a question of interest—when you do statistical work.
- When considering any optimality property, you can ask yourself 'does this have any good properties from the standpoint of my objective?'

▶ **A Caveat about asymptotic optimality of selection criteria**

- One can prove that various natural goals of model selection are inherently in conflict, even asymptotically

Revealing title (Yang, 2005, Biometrika): Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation [get it]

http://dx.doi.org/10.1093/biomet/92.4.937

- Even asymptotically, certain appealing properties are in conflict.

▶ **Some good advice from a theorist**

- Amemiya defined his own model selection criterion
- Cite: Selection of Regressors,International Economic Review, Vol. 21, No. 2 (Jun., 1980), pp. 331-354

  [get it]

  http://www.jstor.org/stable/2526185

▶ **Amemiya' parting advice**

- p.352:

  It is not my intention to recommend any single criterion as a definitely superior criterion or as a panacea of the problem of selection. On the contrary, the general picture that has emerged from this paper is that all of the criteria considered are based on a somewhat arbitrary assumption which cannot be fully justified, and that by slightly varying the loss function and the decision strategy one can indefinitely go on inventing new criteria. [I smell a career here]

- cont.

  This is what one would expect, for there is no simple solution to a complex problem. Thus, as I have already said in the Introduction, the selection of regressors should be primarily based on one's knowledge of the underlying economic theory and one's intuition which cannot be easily quantified.

▶ **Pragmatic justification for model selection criteria**

- Let us suppose that you are attempting to answer an economic question and along the way you come to a practical need to choose some parsimonious way to characterize the dynamics

  As in Newey-West or VARHAC

- One good reason to use AIC or BIC or HQ is that this is systematic, reportable, and replicable.

  That is, folks will know what you did and how to replicate it.

- If your problem involves myriad such choices (as is often the case in macro) 'automating' many of them is often a good pragmatic way to get to a paper that folks find useful and comprehensible.

- **But...**
  - Whenever you are doing stuff for pragmatic reasons, you should make sure that the answers to the core questions in your work are not artifacts of these pragmatic choices.
  - That is, show that your answers are relatively robust to sensible variation in the pragmatic choices.
- **Other model selection selection approaches**
- **Testing-based approaches**
  - Testing approach 1:
    - Start with an AR for some given order presumed to be too large.
    - Test if $\hat{\rho}$ on last lag is statistically significantly diff. from zero by conventional standards.
    - If it is not, drop this variable and go back to first step
    - Otherwise, you are done
- **Aside:: Ok, one more topic to add to our list**
  - This is the first we've talked directly of testing in these notes.
  - We need to do a thorough, advanced econometrics, view of testing.
  - That lecture is coming
- **Testing as model selection**
  - Could alternatively start with 1 lag and raise P so long as last lag is signficant
  - This alternative is usually a bad idea

    That is testing up from small to large often has bad properties relative to testing down from large to small.
  - A standard mantra is that general to specific dominates specific to general.

    Google `Hendry generl to specific` for readings on this.
  - Intuition for these approaches: if the variable is not stat. sig. then you must not need it for a good approx. to the dynamics.
  - Alternative: run test on serial correlation of the $\hat{\varepsilon}$s.
  - Start with an AR(P) for large P;
  - Compute the first k sample autocorrelations of the $\hat{\varepsilon}$s (for some k you choose)
  - Test hypoth. that these are collectively zero.

Choose initial P large enought that you don't reject this hypoth

- Progressively lower the lag order until you do reject the null that the autocorrelations are zero.
- Then raise P by 1 (back to the last value where you didn't reject the null)
- Intuition: The shocks in the linear model are not serially correlated.

We should include lags until the $\hat{\varepsilon}$s also appear to have this property.

▶ **Discussion**

- All these approaches present similar ways to choose a reasonable parsimonious approximation to the dynamics present in the data.
- The 'selection criterion' approach dominates empirical work at this time
- My advice: use the selection approach unless you have good reason to use something less standard
- And then check whether your results are an artifact of this choice
- None of these approaches at this time has been shown to dominate in terms of reliable results
- And the selection criterion approach has merits in terms of being widely used, easy to report, and transparent.
- And these goals dominate absent any fundamental reason favoring other approaches.