607
From Elemetary Econometrics to Advanced Time Series Econometrics in the Least Squares Family

Jon Faust

http://e105.org/e607

September 15, 2016

▶ **This lecture**

- This lecture starts with the simplest case of OLS and a linear model with fixed regressors, Gaussian, iid errors, etc.
- Under these assumptions, the OLS estimator is properly centered (in a certain sense), efficient (in a certain sense) and normally distributed, giving us access to a range of test statsitics distributed exactly according to t, F and χ^2 distributions under the null.
- Then we begin laying the groundwork for what I'll call advanced econometrics (including time series).
- In advanced econometrics, under wildly general assumptions and general nonlinear models we obtain estimators that are properly centered (in a certain sense), efficient (in a certain sense) and asymptotically normally distributed, giving us access to that same toolkit, but now they are justified as asymptotic approximations.

▶ **Well-actually**

- In the transition, we do take on one additional choice, which is the choice of whether to use methods we call efficient or methods we call robust.

  (This choice was present before, but the transition forces us to face it.)

▶ **Ok, back to that transition**

- The transition is the move to what McFadden calls the CAN framework (consistent asymptotically normal)

- Sketching this transition at its most general will hopefully help you see the simple superstructure of econometrics clearly.

▶ **And finally,**

- In my view (and the view of lots of folks) what distinguishes applied statistics in economics from statistics is a focus on causal inference.
- As you know that raises the statistical issue of identification.
- We'll only hint at identification a few times in this initial section.

  identification is a fairly cleanly separable topic, and we'll separate it.

▶ **Readings**

- I assigned a chunk of Jonathan Wright's notes.
- I like Hansen's web text, and all the material in the first 6 Chapters is fair game this lecture
- If you want the details on asymptotics, a great starting point is McFadden's chapter 4. I recommend you read it.

  go

  http://elsa.berkeley.edu/users/mcfadden/e240a%5Fsp01/ch4.pdf

- And Hal White's book is a classic on asymptotics, more generally

  Asymptotic theory for econometricians

▶ **The sketch again**

- In simplest OLS case with linear models, OLS delivers estimators that are
  - Properly centered (in the sense of being unbiased)
  - Efficient (in the sense of Gauss-Markov)
  - Gaussian, supporting a standard tool kit of inference

    (statistics distributed t, F, χ^2 under the null)

▶ **Elementary OLS case: DGP**

- Take the DGP:

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

or in matrix form:

$$Y = X\beta + \varepsilon$$

bn $Y, \varepsilon(T \times 1), X(T \times K), \beta(K \times 1)$

- Note x_t is a column vector made from the t^{th} row of X.
- The OLS estimator is

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

  and it is useful to define,

$$\hat{\varepsilon} = Y - X\hat{\beta}$$

- Substituting for Y using the DGP:

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

▶ $(\hat{\beta} - \beta)$ **is a func. of sample means**

$$\begin{aligned}(\hat{\beta} - \beta) &= (T^{-1}X'X)^{-1}T^{-1}X'\varepsilon \\ (\hat{\beta} - \beta) &= \bar{Q}^{-1}\bar{w}\end{aligned}$$

where

$$\begin{aligned}\bar{Q} &= T^{-1}\sum Q_t, \quad (K \times K) \\ \bar{w} &= T^{-1}\sum w_t, \quad (K \times 1) \\ Q_t &= x_t x_t' \\ w_t &= x_t \varepsilon_t\end{aligned}$$

▶ ...

- A1. The xs are nonstochastic and $X'X$ is full rank.
- A2. The ε are mean zero:

$$E\varepsilon_t = 0$$

- A3. The εs are homoskedastic and not serially correlated

$$\begin{aligned}E\varepsilon_t\varepsilon_s &= \sigma_\varepsilon^2 \quad \text{if } t = s \\ &= 0 \ \text{ otherwise}\end{aligned}$$

▶ **Question**

- Does our DGP combined with A1–A3 describe or imply a complete probability model?

  No. We have not fully described the stochastic properties of the εs.

- And hence, we will not be able to do much in the way of inference

And when I say inference, I mean have a basis for making point estimates and understanding their likely precision.

▶ **Inference in the simplest case**

- A4: $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$.

  We will only apply this when we need it. A4 embeds A2 and A3 and adds much more.

▶ **The probability model**

- In our usual notation, the full probability model has $P_\theta, \theta \in \Theta$.
- Under A1, A2, A3, $\theta = \{\beta, \sigma_\varepsilon^2, \psi\}$, where ψ indexes all the stochastic processes for ε that are mean zero, homoskedastic, and not serially correlated.
- A1–A4 together form a complete prob. model, and ψ indexes all the ε satisfying A2–A4 with additional specification to fill out the family of distributions for ε form a complete prob. model.

▶ **Properties**

- Under A1–A3 assumptions, $\hat{\beta}$ will be properly centered in the sense of being unbiased.
- And will have exact variance,

  $$\text{var}(\hat{\beta}) = (X'X)^{-1}\sigma_\varepsilon^2$$

  which is consistently estimable by,

  $$(X'X)^{-1}s^2$$

  where $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/T$
- And, $\hat{\beta}$ is efficient in the sense of the Gauss-Markov theorem.
- Under A4 in addition,

  $$(\hat{\beta} - \beta) \sim N(0, (X'X)^{-1}\sigma_\varepsilon^2)$$

  which provides the basis for exact probability statements derived from the distributions of $\hat{\beta}$ and $(X'X)^{-1}s^2$.

  for example, exact t, χ^2 and F tests.

▶ **Let's race across the derivation of these results**

▶ $\hat{\beta}$ **unbiased**

- Under the assumptions $Ew = 0$ and $\bar{Q}$ is full rank and hence invertible.

  $$E(\hat{\beta} - \beta) = \bar{Q}^{-1}E\bar{w} = 0$$

  .

- The variance-covariance matrix in the unbiased case, is

$$\begin{aligned}\text{vcov}(\hat{\beta}) &\equiv E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' \\ &= E(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\end{aligned}$$

- With fixed x:

$$\text{vcov}(\hat{\beta}) = (X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1}$$

- With homosked. and not serially correlated εs, $E\varepsilon\varepsilon' = I\sigma_\varepsilon^2$ so,

$$\text{vcov}(\hat{\beta}) = \sigma_\varepsilon^2(X'X)^{-1}$$

▶ **Optimality**

- Under the assumptions, the Gauss-Markov theorem says that OLS is the best linear unbiased estimator.

▶ **Aside:: Best!?! Huh?**

- You should never hear words like 'best','optimal', 'efficient', 'large', 'small' without asking yourself 'in what sense?'
- Of course, in the Gauss-Markov case, we mean the minimum variance unbiased linear estimator.

▶ **Best, but how good?**

- We'd like to do inference about, e.g., pr$(R\beta > 0)$

  R here is a matrix stating linear restrictions among the βs.

- Under A4, we have fairly directly that

$$(\hat{\beta}-\beta) \sim N(0, (X'X)^{-1}\sigma_\varepsilon^2)$$

  and we can estimate this variance by

$$(X'X)^{-1}s^2$$

  as noted above

▶ **Aside:: Why normal?**

- Note that $(\hat{\beta}-\beta)$ is a deterministic linear combination of the iid normal εs.

- Such items are also normal

▶ **From elementary to advanced**

▶ **Beyond elementary ecmet**

- We are moving to the case in which εs are not iid and, indeed, may be heteroskedastic and dependent.

  We are allowing, $E\varepsilon\varepsilon' \equiv \Sigma \neq I\sigma_{\varepsilon}^2$. The values on the main diagonal of Σ may vary (heterskedasticity) and the off diagonal elements need not be zero (serial correlation). There may also be dependence reflected in higher moments.

- We also move to the case of stochastic xs that are not independent of εs.
- We'll also drop A4: we can seldom plausibly assert Gaussian errors.
- In making these moves, we lose our exact, finite sample results.

  and we turn to asymptotic approximations

▶ **The program**

- Re-state weaker versions of A1–A3 that guide us to asymptotic approximations to the distribution of $\hat{\beta}$.
- This will allow us to explore asymptotic versions of all the properties discussed above for OLS
- At a generic level, only trivial changes are needed.

▶ **Note: the CAN framework**

- A short-hand here is that we will make enough assumptions to remain within the CAN (Consistent Asympotically Normal) family of estimators.
- But as I'll emphasize, we also need to have a good estimate of the variance-covariance matrix of the asymptotic distribution.
- Thus, I view CAN as short for CAN-WCEAVCM

  With Consistently Estimable Asymptotic Variance-Covariance Matrix

- As we'll see, attaining CAN is trivial; that WCEAVCM part is a bit heavier going.

▶ **A first step**

- Let's take a first small step in this direction by allowing stochastic xs and by dropping $A4$.

▶ **We replace A1 with**

- A1'. The xs are stochastic and satisfy

$$Ex_t x_t' = Q_t^e \quad (fullrank)$$

and these Q_t^es (nonstochastic) converge in the sense that:

$$T^{-1} \sum Q_t^e \to Q \quad (fullrank)$$

and a WLLN applies to Q_t (stochastic) so that

$$\bar{Q} \to_p Q$$

- A2' (tentative). The ε are conditionally mean zero:

$$E[\varepsilon_t | x_t] = 0$$

> This implies that the x_t is not correlated with ε_t, a result that before followed simply from the xs being fixed.

▶ **Note**

- We'll need a bit more in the way of assumptions, but it easiest to see what we need as we derive results
- So we'll be sharpening A2' and call it tentative for now.

▶ **Properties under A1' and A2'**

- Remember:

$$(\hat{\beta} - \beta) = \bar{Q}^{-1} \bar{w}$$

- Because ε is conditionally mean zero under $A2'$, we have that $Ew_t = 0$

> But this no longer delivers that $\hat{\beta}$ is unbiased because with stochastic xs, $\bar{Q}$ and ε need not be independent implying that
>
> $$E\bar{Q}^{-1}\bar{w} \neq E\bar{Q}^{-1}E\bar{w}$$

▶ **Aside:: Some intuition regarding bias in time series.**

- ε_t may be unrelated to x_t, but ε_t may generally feed back on future xs.
- For example, consider the case in which one of the xs is lagged y. It is clear that ε_t is then part of y_{t+k} for $k > 0$.
- Important: For the reason just given, most of our estimators in time series are biased.
- Note: We sometimes can adjust a bit for the bias, but often do not.

for example, bootstrap methods may allow us to approximate, and then adjust for, the bias.

- **Consistency**
  - We give up unbiasedness, and fall back on an asymptotic notion of properly centered: consistency.
  - Since A1' implies $(\hat{\beta} - \beta) = \bar{Q}^{-1}\bar{w}$, and we have assumed that $Ew_t = 0$, all we need is that the ws are sufficiently well-behaved for a WLLN to have

$$\bar{w} \rightarrow_p 0$$

    Let's just assume this.
  - A2' $E[\varepsilon_t|x_t] = 0$, and $w_t = x_t\varepsilon_t$ obeys a WLLN.
  - Now since we already assumed that $\bar{Q} \rightarrow_p Q$, full rank, we have that $\hat{\beta}$ is consistent.
- **Aside:: Consistency and explosive $\bar{Q}$**
  - Note that so long as $\bar{w}$ remains bounded, $\hat{\beta}$ is also consistent if $\bar{Q}$ diverges to ∞

    This case can sometimes be ruled out as pathological and unlikely
  - In time series the case where $\bar{Q}$ diverges may often be relevant

    It arises in the famous unit root case; you saw this in the explosive variance in problem set 1.
  - Thus, we'll have to keep this case in mind.
- **Asymptotic distribution of $\hat{\beta}$**
  - For inference, we need some handle on the distribution of $\hat{\beta}$
  - We can't say anything about the mean of $\hat{\beta}$, and for essentially the same reason, we can't say much about the exact variance of $\hat{\beta}$;
  - But we'll rely on the variance of an asymptotic distribution.
  - We have,

$$\sqrt{T}(\hat{\beta} - \beta) = \bar{Q} \times \sqrt{T}\bar{w}$$

  - And we have already assumed $\bar{Q} \rightarrow_p Q$, that $Ew_t = 0$, and that a WLLN applies to $\bar{w}$.
  - Strengthen A2' to say that the ws are sufficiently well behaved for a CLT to apply to $\sqrt{T}\bar{w}$.
  - A2' (final). $E[\varepsilon_t|x_t] = 0$; $w_t = x_t\varepsilon_t$ are sufficiently well behaved that a WLLN and CLT apply:

$$\begin{aligned} \bar{w} &\rightarrow_p 0 \\ \sqrt{T}\bar{w} &\rightarrow_d N(0, \Omega) \end{aligned}$$

    for finite, full rank Ω.

- Now you should be able to show that

$$\sqrt{T}(\hat{\beta}-\beta) = \bar{Q}^{-1}\sqrt{T}\bar{w} \to_d N(0, Q^{-1}\Omega Q^{-1})$$

▶ **Thus,**

- By assuming A1', A2' we have that $\hat{\beta}$ is CAN
- But, we have to go further to get to WCEAVCM
- Since we already assumed $\bar{Q} \to_p Q$, we just need a consistent estimator for Ω to form a consistent estimator of $Q^{-1}\Omega Q^{-1}$.
- We'll spend a good deal of time on estimators for Ω.
- At this point, let's just make a high-level assumption.
- A3' The w_ts are sufficiently well behaved that we have a consistent estimator for $\hat{\Omega}$ for Ω.

▶ **Wrapping up**

▶ **The elementary case**

- A1. The xs are nonstochastic and $X'X$ is full rank.
- A2. The ε are mean zero:

$$E\varepsilon_t = 0$$

- A3. The εs are homoskedastic and not serially correlated

$$\begin{aligned} E\varepsilon_t\varepsilon_s &= \sigma_\varepsilon^2 \quad \text{if } t = s \\ &= 0 \ \ \text{otherwise} \end{aligned}$$

- A4. $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$
- Under A1–A3 the OLS $\hat{\beta}$ is unbiased, satisfies Gauss-Markov, and has variance-covariance matrix consistently estimable by $(X'X)^{-1}s^2$.
- Under A4 in addition, $\hat{\beta}-\beta \sim N(0, (X'X)^{-1}\sigma_\varepsilon^2)$ which we can use as a basis for probability statements about $\hat{\beta}$ and forming conventional test statistics.

▶ **The advanced case,**

- Under A1', A2', and A3', $\hat{\beta}$ is consistent and

$$\sqrt{T}(\hat{\beta}-\beta) \to_d N(0, Q^{-1}\Omega Q^{-1})$$

- Asymptotically valid probability statements can be based on a a normal distribution with variance-covariance matrix consistently estimated by $\bar{Q}^{-1}\hat{\Omega}\bar{Q}^{-1}$

Where we've waved our hands a bit on the consistent $\hat{\Omega}$.

- Mainly, we've made MUCH weaker assumptions, and lost finite-sample properties but replaced them with some asymptotic version that may be adequate for our purposes
- But we have not yet mentioned efficiency in the asymptotic case
- We will mainly set aside efficiency for now, but…

▶ **Aside:: Efficiency and the GLS principle**

- In the simple case, OLS satisfied Gauss-Markov
- As you should know, if we allow serial correlation and hetersked. of the εs, OLS is not efficient even in the case with fixed Xs
- And the usual OLS esimtator of the variance-covariance matrix won't be consistent
- But we can often find some other estimator of the variance covariance matrix of OLS that is consistent.

▶ **Aside::**

- Further, feasible GLS (FGLS) may be efficient
- Thus, you face a choice:
  - Stick with OLS

    In which case, you give up a claim to efficiency, and must find an appropriate estimator for the variance-covariance matrix.
  - Go to some feasible version of GLS

    It's more complicated and may bring other hassles, but allows you to claim efficiency.
- In elementary econometrics, we often did not face up to this choice very explicitly.
- Advanced econometrics, and the nature of the applications that drive us to advanced econometrics essentially forces us to face this choice more explicity.
- We'll mainly set this aside for a bit

▶ **Left on our agenda:**

- Get clearer on what we mean by 'sufficiently well-behaved' for a CLT and/or WLLN to apply to w_t or Q_t.

  with particular emphasis on well-behaved time series dependence
- Get clearer on what we mean by 'sufficiently well-behaved for a consistent estimator of Ω to exist,' and then find some actual consistent estimator we can use.

  A consistent estimator will be enough to justify asymptotic inference
- Finally, get clear on the efficiency issues

▶ **And then we'll have covered the basics of much of advanced time series econometrics**