607
From OLS to a large class of estimators: efficiency

Jon Faust

http://e105.org/e607

October 7, 2015

- **Readings**
  - You might review GMM and ML as presented, e.g., in Hansen's text and the Wright notes.
  - Also read Hansen's discussion of semi-parametric efficiency
  - And for those interested in the deeper theory of semi-parmetric efficiency, Newey has a beautiful and pretty readable paper.

    Newey, W. K. (1990), Semiparametric efficiency bounds. J. Appl. Econ., 5: 99-135. go

    http://onlinelibrary.wiley.com/doi/10.1002/jae.3950050202/epdf

- **Review**
  - We have reviewed simple assumptions in which OLS is best in the Gauss-Markov sense.
  - And in a slight generalization of those assumptions, GLS is best.
  - You should know, but we have not yet discussed:
    - the Cramer-Rao bound, which is attained asymptotically by the MLE
    - The choice of W to reach efficient GMM
- **This lecture provides a brief review/sketch of how these are related.**
- **Highest level perspective.**
  - We have shown the for OLS, MLE, GMM, and implicitly other extremum estimators,

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

  - The form of this asymptotic variance-covariance matrix is often called a sandwich form.

- When we have the 'efficient' version of a given estimator, this asymptotic variance-covariance expression collapses to something without the sandwich form.
- In particular, for the efficient case, we have some result such as $\Omega = Q^{-1}$ under which the asymptotic variance-covariance matrix is the relevant $-Q^{-1}$.

▶ **OLS with iid errors**

- With fixed Xs we have exact variance-covariance matrix of:

$$E(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}$$

- When OLS is efficient, $E\varepsilon\varepsilon' = I\sigma^2$ and,

$$\begin{aligned} E(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} &= (X'X)^{-1}X'X(X'X)^{-1}\sigma^2 \\ &= (X'X)^{-1}\sigma^2 \end{aligned}$$

▶ **Efficiency**

- It is easy to show that the GLS variance-covariance matrix is the smallest (in the matrix sense) among all estimators that minimize,

$$\hat{\varepsilon}'W\hat{\varepsilon}$$

  for some positive definite W.

▶ **Maximum likelihood**

- We sketched a derivation of the fact that

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta^*) \to_d N(0, \mathbf{Q^{-1}\Omega Q^{-1}})$$

  where Q is the limiting Hessian and,

$$\mathbf{\Omega} = \lim \mathbf{E T\bar{q}\bar{q}'}$$

- But the MLE satisfies the GLS property automatically, and

$$\mathbf{\Omega} = -\mathbf{Q}$$

  both of them equal to the information matrix.
- So, the sandwich form reduces to $\mathbf{Q^{-1}}$.

▶ **GMM**

- In GMM for general W we derived the estimator is CAN with asymptotic variance-covariance equal to

$$\mathbf{Q^{-1}\Omega Q^{-1}}$$

where

$$\begin{aligned}\bar{D}'W\bar{D} &\to_p \mathbf{Q} \quad (\mathbf{fullrank}) \\ \mathbf{\Omega} &= \lim E\bar{D}'WH_TW\bar{D}\end{aligned}$$

- In GMM, we minimize $\bar{g}'W\bar{g}$ for some positive definite W.
- The GLS principle says that we should consider $W = [\text{vcov}(\bar{g})]^{-1} \equiv H^{-1}$
- If we do this, the sandwich form asymptotic variance-covariance matrix collapses to Q^{-1}
- As usual, it is pretty straightforward to show that this minimizes the asymptotic variance-covariance matrix across the possible choices of W.

▶ **Efficient GMM vs. MLE**

- The Cramer-Rao bound theorem shows that the lower bound on the size of the asymptotic variance-covariance matrix is given by I^{-1}
- Since the MLE has the asymptotic variance-covariance matrix, ML is asymptotically efficient in the sense of attaining the Cramer-Rao bound.

  Attaining the Cramer-Rao bound would more precisely be stated as, the MLE converges in distribution to a random variable with $variance - covariance$ matrix given by I^{-1}.

▶ **Efficient GMM**

- IF (that, as they say, is a big if) the moment conditions for GMM are equivalent to the score, then efficient GMM also attains the Cramer-Rao bound.

  This is hand-wavy since we have not defined equivalent, but let's let that go for the moment for equivalence to ML is not really the important concept.

- We often use GMM after having derived some moment conditions in cases where we have not completely specified the model.
- For example, we may have specified a consumption Euler equation, but the rest of the model is vague.
- Or we have specified the full model, but we only want to take seriously the moment conditions.

  The other bits of the model are viewed as speculative or as being artifacts of assumptions of convenience.

- We'd like a notion of efficiency that applies in this case

▶ **The model and semi-parametric eficiency**

- I emphasized on the first day that we should always start with a complete probability model.
- We have done this for ML and I sketched this for our OLS model when OLS was first introduced.
- I have not done so explicitly for GMM.

▶ **The model**

- The model is implicitly defined by,
$$Eg(Y;\theta) = 0$$
where g is known and $\theta \in \Theta$.
- A completion of this is $Y \sim P_A$, $A = \{\theta, \gamma\}$ and $A \in \mathcal{A}$.
- And γ is the, presumably infinite-dimensional, parameter indexing all distributions for Y that satisfy $Eg(Y;\theta) = 0$.
- Thus, the truth has $\theta = \theta^*$ and has some γ^* describing all aspects of the distribution other than the moment condition.
- We call the estimation problem semi-parametric when part of the model is tightly parameterized and the remainder is vaguely specified—typically, specified with infinite-dimensional vagueness.

▶ **Semi-parametric efficiency**

- There is a notion of semi-parametric efficiency that has some intuitive link to the Cramer-Rao bound.
- Think of all possible parametric sub-models that are 'contained in' our semi-parametric model
- Each of those will have a Cramer-Rao bound.
- And the efficiency bound for our full model cannot be smaller than the biggest those for the sub-models.

  You cannot attain a better answer asymptotically by making less assumptions.

- For certain semi-parametric models, this bound can be explicitly derived
- Since the parametric part of our model in the GMM case is defined by simple moment conditions, the case is particularly amenable to analysis.
- And 'efficient GMM' attains the semi-parametric efficiency bound

  That is, it is efficient among estimators that exploit nothing more than those moment conditions.

▶ **'Nothing more'**

- When I say ‘nothing more’ I mean nothing more than the moment conditions and all those regularity conditions we needed for the estimator to be CAN and well-behaved.

  In saying well-behaved I am ruling out some anomalous cases analogous such as estimators that beat the Cramer-Rao bound in the parametric case.

▶ **Aside:: Cov. stat. and semi-parametrics**

- I have said regularly that we don’t aspire to estimating the true model in macro time series.

  Our best guess as to the family that truth comes from is infinite-dimensional and its hard to estimate infinite-dimensional models well using the short samples we have in macro

- Thus, we attempt to approximate the dynamics well having made some palatable assumptions such as covariance stationarity.
- Of course, the covariance stationary model is infinitely parameterized, say by elements of ℓ^2.
- Thus, we might say that most macro time-series work that is parametric at all is (implicitly at least) semi-parametric.

▶ **MLE and QMLE**

- Of course, applied work sometimes uses maximum likelihood techniques, usually using the Gaussian likelihood.
- Even in these cases, we typically view the chosen likelihood as an approximation to some richer true likelihood.
- When we viewing our chosen likelihood as an approximation and are being careful, we say we are using ‘quasi-maximum likelihood’ methods.

  When the approximating likelihood is Gaussian we refer to ‘the Gaussian QMLE’.

▶ **Aside:: Quasi**

- Whenever you see the word quasi in econometrics, you should probably at least try on substituting ‘not the’ for ‘quasi’

  That is, QMLE is ‘not the MLE’

- This just helps remind you of the possible approximation error.