607
Generalized Distance, Quadratic Forms, and the GLS principle

Jon Faust

http://e105.org/e607

October 7, 2015

- **Readings!?!**
  - There are no readings for this brief note.
  - But you might want to look at Mahalanobis (1936) as mentioned at the end.

    Proceedings of the National Institute of Sciences of India 2 (1): 49-55. go

    http://www.new.dli.ernet.in/rawdataupload/upload/insa/INSA%5F1/20006193%5F49.pdf

- **Quadratic forms, distance, units, etc.**
  - Suppose ϕ is a $(K \times 1)$ real-valued vector
  - ϕ defines a point in $\mathcal{R}^K$
  - And the Euclidean distance between the origin and ϕ is $D^{1/2}$ where

$$D = (\phi'\phi)$$

- **Coordinate systems**
  - Take any full rank matrix L and define

$$\tilde{\phi} = L\phi$$

  - We can think of $\tilde{\phi}$ as picking out the same point as before in a new coordinate system.
  - This system has the same origin as the original, but the axes that are moved around in some way.
  - And we may have proportionally stretched or shrunken the units on the axes.

E.g., for $L = \kappa I$ we have stretched/shrunken the units on every axis proportionally.

- In our new units, the distance to the origin is $\tilde{D}^{1/2}$ where

$$\tilde{D} = \tilde{\phi}'\tilde{\phi} = \phi' L' L \phi$$

  Or defining $W = L'L$,

$$\tilde{\phi}'\tilde{\phi} = \phi' W \phi$$

- $W = L'L$ is obviously positive definite (since L is full rank).
- Thus, we can think of $\phi' W \phi$ where W varies across all the positive definite matrices, as giving the distance to the origin in all coordinate systems of this type.

▶ **Statistics**

- If ϕ is a drawn from some joint distribution with mean zero elements, $D^{1/2}$ is a measure of how far this draw is from the mean.
- With symetric, single peaked, iid distributions, ϕs farther from the mean are less likely to be observed.
- Thus, D forms a natural measure of how unlikely ϕ is to be observed.

  bigger D implies less likely

- Of course, we can always ask 'unlikely relative to what?'
- And this is essentially a matter of units
- If the elements of ϕ are mean zero with variance-covariance matrix $I\sigma^2$ we have one natural metric: the variance (or standard deviation).
- $K^{-1} E \phi' \phi = K^{-1} E[D] = \sigma^2$
- Thus, it is often useful to talk in standard deviation units. How many standard deviations away from the origin does ϕ fall.
- This is given by $(D/E[D])^{1/2}$.

▶ **Heterogeneity**

- If the ϕs are heterogeously distributed (but still mean zero), then it is not necessarily the case that larger D means less likely.

▶ **Aside:: Should be obvious, but...**

- Take the case with $K = 2$ and $E[\phi_i^2] = \sigma_i^2$, $i = 1, 2$ and σ_1^2 much larger than σ_2^2.
- In this case $\phi = (2, 1)'$ (for which $D = 5$) may be more likely to be observed than $\phi = (1, 1.5)'$, for which $D = 3.25$.

- Thus, with heterogeous variances, relative Ds are not a good indicator of relative likelihoods of being observed.

▶ **But generalized distance, however,...**

- But we could shift units to some give a metric for which D has the nice properties it does when the ϕs are mutually uncorrelated and homoskedastic.
- As you probably know, if $E\phi\phi' = \Sigma$, then the distance metric we want is $\phi' W \phi$ where $W = \Sigma^{-1}$.
- In particular, pick L such that,
$$W \equiv L'L = \Sigma^{-1}$$
- Now measure distance in terms of $\tilde{\phi} = L\phi$
- We have that

$$\begin{aligned} E\tilde{\phi}\tilde{\phi}' &= EL\phi\phi' L \\ &= L\Sigma L \\ &= L(L'L)^{-1}L \\ &= I \end{aligned}$$

- That is, the $\tilde{\phi}$s are orthonormal

  (mutually orthogonal with unit variance)

- In these new units, a larger value of $\tilde{D}$ unambiguously means that such $\tilde{\phi}$s are less likely to be observed when the underlying distributions are single-peaked, symmetric distributions, and differ only up to scale.

▶ **The GLS principle**

▶ **The GLS principle**

- GLS principle: When you see a stochastic expression such as $\phi'\phi$, you should consider a more general expression, $\phi' W \phi$, and, in particular, to consider $W = \Sigma^{-1}$ where Σ is the variance-covariance matrix of ϕ.
- Perhaps better called the GLS rule of thumb, but that doesn't have the gravitas of a principle.

▶ **Examples: OLS and GLS**

- Consider OLS in the case with fixed regressors.

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

- OLS is defined by minimizing $\hat{\varepsilon}'\hat{\varepsilon}$.

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

- Of course, OLS is not efficient when $E\varepsilon\varepsilon' = \Sigma \neq I\sigma_\varepsilon^2$

  But GLS, is efficient

- GLS chooses $\hat{\beta}$ to minimize

$$\hat{\varepsilon}'\Sigma^{-1}\hat{\varepsilon}$$

- This is the first application of the GLS principle that we usually see in econometrics and it is from this result I take the name.

▶ **Aside:: GLS and changed units**

- GLS can be be viewed as OLS on the system

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$$

  where in each case, the tilde version of the term involves premultiplying the original by L where $L'L = \Sigma^{-1}$.

▶ **Wald test statistics**

▶ **Wald test statistics**

- As you may recall, the conventional Wald test statistic is precisely a generalized distance measure
- We estimate an unrestricted model and then take a generalized distance between the estimate, $\hat{\beta}$ and the value hypothesized under the null, β_0.
- That is, the Wald statistic is:

$$(\hat{\beta} - \beta_0)'V^{-1}(\hat{\beta} - \beta_0)$$

  where V is (a consistent estimator for) the variance-covariance matrix of $\hat{\beta} - \beta$.

▶ **Efficient GMM**

- We'll see below that efficient GMM essentially arises from picking the GMM weight matrix to satisfy the GLS principle.

▶ **Generalized distance and χ^2 distributions**

▶ **Generalized distance and χ^2 distributions**

- If $\phi \sim N(0, \Sigma)$, then $\tilde{\phi} = L\phi$ where $L'L = \Sigma^{-1}$ satistifes

$$\tilde{\phi} \sim N(0, I)$$

  is

- Thus, $\tilde{\phi}'\tilde{\phi} = \phi'\Sigma^{-1}\phi \sim \chi^2_{(p)}$
- This is where χ^2 distributions arise in testing.

► **Asymptotics**

- The results just given all follow as asymptotic results when ϕ is asymptotically $N(0, \Sigma)$.

► **Much econometrics**

- As we are seeing, much of our advanced econometrics involves quadratic approximations and asymptotically normal random variables.
- Thus the general perspective reviewed here, and the GLS principle, in particular, will get a lot of play.

► **To nuture you scholarly nature**

- Prasanta Mahalanobis wrote a nice article entitled, 'On the generalised distance in statistics' in 1936.

  Proceedings of the National Institute of Sciences of India 2 (1): 49-55. go

  http://www.new.dli.ernet.in/rawdataupload/upload/insa/INSA%5F1/20006193%5F49.pdf

- What I've discussed as a generalized distance measure is sometimes called the Mahalandobis distance.
- There now, don't you feel you roots in statistics growing down and reaching out?