

Simulation Based Estimation and Inference - Non-Normality and Effects of Small Samples: Lecture XXX

Charles B. Moss ¹

¹University of Florida

April 3, 2012

- 1 Readings
- 2 Estimation in Small (and Moderately) Sized Samples
- 3 Ordinary Least Squares and Non-Normal Errors
 - Linear Regression on Information Measures
 - Refresher on the Central Limit Theorem
 - Two Simulation Procedures
 - Jarque-Bera
 - Bootstrapping
 - Jackknifing
- 4 Sample Size and System Estimators
 - Simulation of Power of Test – Laitinen
 - Meisner – Symmetry

Readings

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition. (Chapter 15 pp. 603-621)
- * Laitinen, K. 1978. Why is Demand Homogeneity So Often Rejected. *Economics Letters* 1(3), 187-191.
- * Meisner, J.F. 1979. The Sad Fate of the Asymptotic Slutsky Symmetry Tests for Large Systems. *Economics Letters* 2(3), 231-233.
- * Mishra, A.K., G.T. Livanis, and C.B. Moss. 2011. Did the Federal Agricultural Improvement and Reform Act of 1996 Affect Farmland Values? *Entropy* 13(3), 668-682.
- * Moss, C.B. and H. Theil. 2003. Homogeneity Testing Revisited *Journal of Agricultural and Applied Economics* 35(Special Issue), 3-4.

Readings - Continued

- * Shonkwiler, J.S. and H. Theil. 1986. Some Evidence on the Power of Monte Carlo Tests in Systems of Equations. *Economics Letters* 20(1), 53-54.
- * Taylor, T.G., J.S. Shonkwiler, and H. Theil. 1986. Monte Carlo and Bootstrap Testing of Demand Homogeneity. *Economics Letters* 20(1), 55-57.

Estimation in Small (and Moderately) Sized Samples

- Most of our analysis so far has been based on two general assumptions:
 1. Normally distributed errors in small samples – Since most of our estimators are linear in the observed variables the estimator is then normally distributed.
 2. Asymptotic results based on large sample:
 - a. By applying the Central Limit Theorem to consistent estimators, we derive normality (Ordinary and Generalized Least Squares, and Generalized Method of Moments).
 - b. Apply the Cramer-Rao Lower Bound to Maximum Likelihood results.
- The first aspect of **Simulation Based Estimation and Inference** that I want to talk about is the **Inference** part.
- Most of this discussion follows from some failure in the small sample.

Ordinary Least Squares and Non-Normal Errors

- Consider the model estimated in Mishra, Livanis, and Moss (2011)

$$I_t = \alpha_0 + \alpha_1 t + \alpha_2 D_t + \alpha_3 IPR_t + \epsilon_t \quad (1)$$

where t is a time trend, D_t is a dummy variable which is equal to 0 if $t < 1996$ and equal to 1 if $t \geq 1996$, IPR_t is the index of prices received by farmers, and ϵ_t is the random variable.

Linear Regression on Information Measures

- A critical feature in this paper involved the distribution of ϵ_t and its impact on the test statistic in small sample.
 - Following Kullback

$$I_t = \int_{-\infty}^{\infty} f(x_t) \ln \left(\frac{f(x_t)}{g(x_t)} \right) dx_t \sim \chi_k^2 \quad (2)$$

where $f(x_t)$ and $g(x_t)$ are two different distributions with k parametric restrictions (typically $f(x_t)$ is the restricted model).

- Thus, it is unlikely that I_t is Normally distributed (i.e., subtracting a constant from a χ^2 variable cannot make it normal).
- Question: Is 59 observations sufficient to appeal to the Central Limit Theorem?

Refresher on the Central Limit Theorem

- On the test I am stickler for what the Central Limit Theorem acutally states.
- Given the regression

$$I_t = \alpha_0 + \alpha_1 t + \alpha_2 D_t + \alpha_3 IPR_t + \epsilon_t \quad (3)$$

what becomes normally distributed under the Central Limit Theorem – Not ϵ_t , it remains χ^2 .

- Answer to Question – I probably could have appealed to the Central Limit Theorem, but I did not.

Two Simulation Procedures

- Two common simulation procedures for constructing the confidence intervals (and test statistics) for potentially non-normal residuals in small samples are Bootstrapping and Jackknifing.
- As a starting place, assume that I have a regression model

$$y_t = \beta x_t + \epsilon_t \quad t = 1, \dots, T \quad (4)$$

where ϵ_t is Non-Normal and T is relatively small.

- The first thing I want to do is to run the regression and derive $\hat{\beta}$, $\hat{y}_t$, and $\hat{\epsilon}_t$.
- Using these estimated errors, you probably want to test for non-normality. A couple of popular tests are Kolmogorov-Smirnov test (based on the cumulative distribution function) and the Jarque-Bera tests

$$JB = \frac{T}{6} \left(S^2 + \frac{1}{4} (K - 3)^2 \right)$$
$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^3}{\left(\frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2 \right)^{3/2}}, \quad K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^4}{\left(\frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2 \right)^2} \quad (5)$$

Bootstrapping

- As a starting point, consider the fact that the estimated values of the dependent variables ($\hat{y}_t$) and residuals ($\hat{\epsilon}_t$) in Equation 4 are unbiased.
- The conjecture is that we can form an equivalently distributed sample by recombining the unbiased estimates of the dependent variables and residuals

$$\tilde{y}_t = \hat{y}_t + \hat{\epsilon}_s = \tilde{\beta}_k x_t \quad \exists: s \in \{1, 2, \dots, T\} \quad (6)$$

or where s is a randomly drawn sample between 1 and T .

- Drawing repeatedly we generate a sample of $\tilde{\beta}_k$ for $k = 1, \dots, K$.

Bootstrapping – Continued

- One concept is that we can use these estimated $\tilde{\beta}_k$ to construct new (more robust) estimates of the variance

$$\Omega_{\beta} = \sum_{k=1}^K \left(\tilde{\beta}_k - \hat{\beta} \right)' \left(\tilde{\beta}_k - \hat{\beta} \right) / K \quad (7)$$

- We will use this approach for system estimators where normality is really not the driving issue.

Bootstrapping – Continued

- A better concept is to use the results as a direct test of the parameter.
 - For example, assume that we want to test

$$H_0 : \beta_1 \geq 0, H_1 : \beta_1 < 0 \quad (8)$$

(say downward sloping demand curves). We could simply count the number of draws where the outcome satisfied the null hypothesis.

- Say that $K = 100$, but $\beta_1 \geq 0$ only twice out of 100 $\Rightarrow$ reject the null hypothesis.

Jackknifing

- Using a slightly different approach, we randomly draw subsamples of observations from the full sample.
- In Mishra, Livanis, and Moss (2011), we ran the regression over and over again randomly dropping 20 sample points (roughly $1/3$ of the sample).

Jackknifing – Continued

Table: Effect of FAIR on Land Value Distribution

Variable	Northeast	Lake States	Corn Belt
Constant	72.31 (107.10)	3.82 (60.95)	129.64 (167.88)
Trend	1.98** (2.04)	1.09** (1.34)	2.24** (2.75)
FAIR	23.51 (32.92)	22.39** (26.92)	55.78** (60.39)
Price Recieved	0.01 (0.28)	0.06 (0.24)	0.71** (0.86)

Sample Size and System Estimators

- I know that you are getting tired of this example, consider the Rotterdam demand system

$$\begin{aligned}\bar{f}_{1t}DX_{1t} &= \theta_1 DQ_t + \pi_{11} Dp_{1t} + \pi_{12} Dp_{2t} + \pi_{13} Dp_{3t} + \pi_{14} Dp_{4t} + \epsilon_{1t} \\ \bar{f}_{2t}DX_{2t} &= \theta_2 DQ_t + \pi_{21} Dp_{1t} + \pi_{22} Dp_{2t} + \pi_{23} Dp_{3t} + \pi_{24} Dp_{4t} + \epsilon_{2t} \\ \bar{f}_{3t}DX_{3t} &= \theta_3 DQ_t + \pi_{31} Dp_{1t} + \pi_{32} Dp_{2t} + \pi_{33} Dp_{3t} + \pi_{34} Dp_{4t} + \epsilon_{3t} \\ \bar{f}_{4t}DX_{4t} &= \theta_4 DQ_t + \pi_{41} Dp_{1t} + \pi_{42} Dp_{2t} + \pi_{43} Dp_{3t} + \pi_{44} Dp_{4t} + \epsilon_{4t}\end{aligned}\tag{9}$$

- Economic theory suggested two sets of hypothesis
 - Homogeneity

$$\sum_{j=1}^4 \theta_j = 1, \quad \sum_{j=1}^4 \pi_{ij} = 0 \forall i\tag{10}$$

where the first condition is met by singularity.

- Symmetry $\pi_{ij} = \pi_{ji} \forall i, j$.

Sample Size and System Estimators – Continued

- The problem is that the empirical estimators where rejecting both hypothesis on a regular basis – Leading to the question: Is our theory wrong?
- Theory is not incorrect. The difficulty is something you have seen.
 - The problem is with the estimator of the residual variance matrix ($\hat{\Omega}$) from 9

$$\hat{\Omega} = \sum_{t=1}^T \left(\bar{D}x_t - \theta DQ_t - \pi \begin{bmatrix} Dp_{1t} \\ Dp_{2t} \\ Dp_{3t} \\ Dp_{4t} \end{bmatrix} \right)' \left(\bar{D}x_t - \theta DQ_t - \pi \begin{bmatrix} Dp_{1t} \\ Dp_{2t} \\ Dp_{3t} \\ Dp_{4t} \end{bmatrix} \right) / T \quad (11)$$

Simulation of Power of Test – Laitinen

- Kenneth Laitinen proposed examining the power of the Wald test using simulation.
 - He simulated the Dutch data set imposing both symmetry and homogeneity and then tested for the hypothesis of homogeneity.

Table: Number of Rejections out of 100

Number of Goods	Rejections at		
	5%	1%	0.1%
$n = 5$	14	6	2
$n = 8$	30	16	7
$n = 11$	53	35	26
$n = 14$	87	81	79

Laitinen – Continued

- Laitinen proved that the problem was that the variance matrix was ill-scaled in small samples. The distribution is actually Hotelling's T^2 which is a multiple of the F statistic with $n - 1$ degrees of freedom.
- Specifically, taking the original Wald result to be T

$$\frac{(n-1)(T-n-1)}{T-2n+1} \sim F(n-1, T-2n+1) \quad (12)$$

- Moss and Theil (2003) demonstrate that empirical rejection rate is consistent with this Hotelling's T^2 distribution.

Meisner – Symmetry

- Next, James Meisner applied the same approach to the over-rejection of symmetry

Table: Number of Rejections out of 100

Number of Goods	Rejections of Symmetry	
	5%	1%
$n = 5$	9	3
$n = 8$	26	8
$n = 11$	50	37
$n = 14$	96	91