From Economics to R

Using R in Economics and Econometrics

Cloudly Chen

School of Economics, Shandong University

2nd Chinese R Conference,
Renmin University of China, 6th, Dec, 2009

Outline

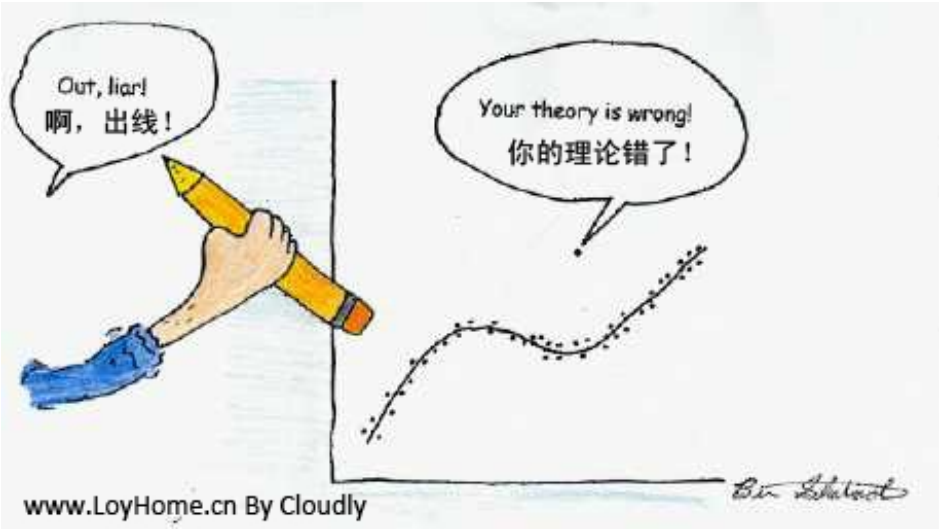
Outline

Economic Research

Why can't we leave statistics when doing economic research?

To what extend do economists need statistics?

- Is statistics the only tool for modern economic research?



Tool for Economics

- On the contrary, although economics has been deeply influenced by importing mathematical methods, including a large range of statistic applications, what an economist needs while doing a research is still originated from the real world – the observation of individual activities in the society and the concern about how the society functions.
- Personally speaking, I think nothing can replace the original economic thoughts, and although the help of statistics is needed, it contributes more to discovering the interesting relationships under the surface, than telling me how to think on a philosophical level.

The position of statistics

What's the position of statistics in economic research?

- Theoretical and empirical research
  In a complete economic paper, the use of two methods are necessary: a theoretical model to explain why something has happened, and an empirical result to persuade the readers that it is true.

statistics and econometics

- If you surf online, it is too common to see that some economic students complain about how hard econometrics is. However, econometrics is the most powerful (or ironically, easiest?) tool for them to write a paper, since it is more difficult to make an innovation in economic theories.

The position of statistics

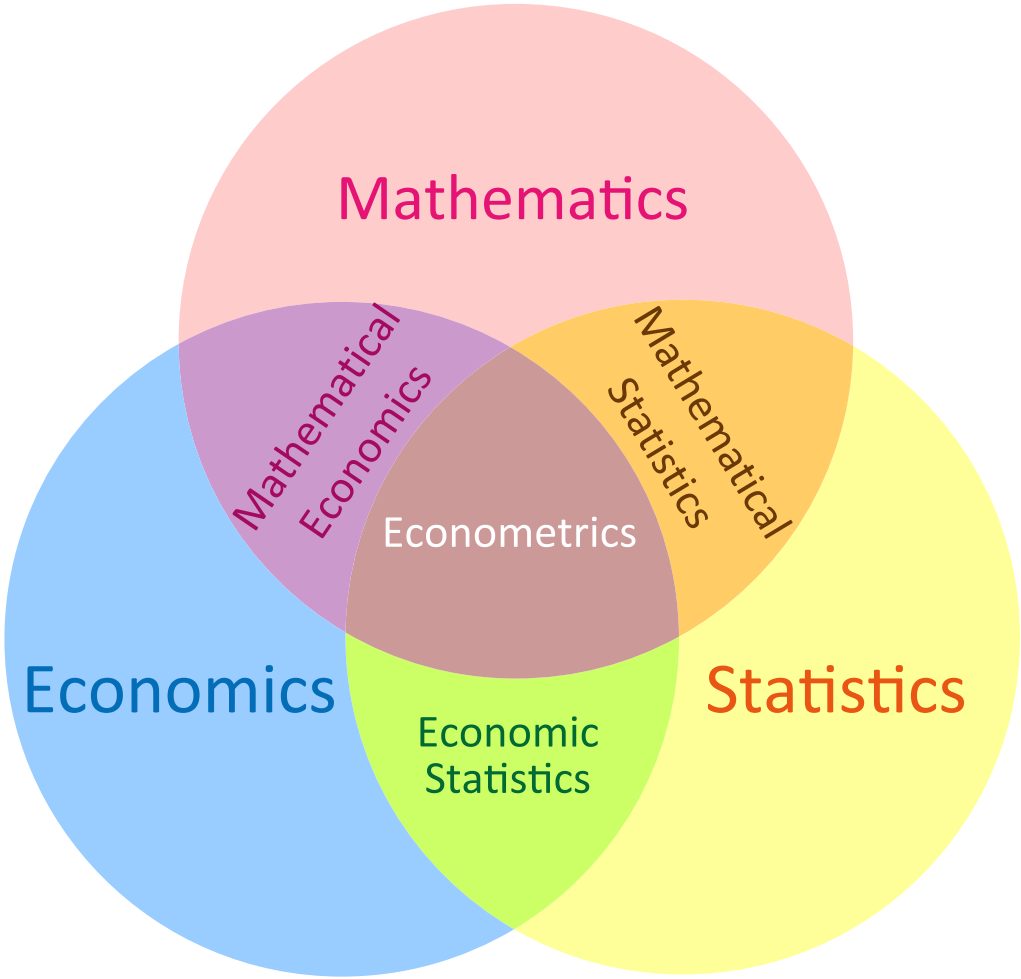
What's the position of statistics in economic research?

- Theoretical and empirical research
  In a complete economic paper, the use of two methods are necessary: a theoretical model to explain why something has happened, and an empirical result to persuade the readers that it is true.

statistics and econometics

- If you surf online, it is too common to see that some economic students complain about how hard econometrics is. However, econometrics is the most powerful (or ironically, easiest?) tool for them to write a paper, since it is more difficult to make an innovation in economic theories.

What is econometrics?



What is econometrics?

What is econometrics?

- Econometrics is concerned with the tasks of developing and applying quantitative or statistical methods to the study and elucidation of economic principles (From wikipedia).

What does an econometrician do?

One day, I talked with a professor in my school who was experienced with econometrics. I tried to persuade him how powerful R was, but he replied me simply:

- (Most) statisticians use R, but (most) economists use Stata.
- Its seems that statisticians don't really do regressions, and

Economists do regressions ONLY

What is econometrics?

What is econometrics?

- Econometrics is concerned with the tasks of developing and applying quantitative or statistical methods to the study and elucidation of economic principles (From wikipedia).

What does an econometrician do?

One day, I talked with a professor in my school who was experienced with econometrics. I tried to persuade him how powerful R was, but he replied me simply:

- (Most) statisticians use R, but (most) economists use Stata.
- Its seems that statisticians don't really do regressions, and

Economists do regressions ONLY

Outline

A simple list

What can I see around me?

1. Stata: No matter Macro- or Micro- econometric research, Stata can handle. Too many packages available makes it easy to do a regression or an estimation.
2. SPSS: Maybe a managemet student loves it better: a foolproof way (just click, click and click) of using a variety of statistic functions
3. Eviews: Too old to remember, but what can we do with the out-of-fashion textbooks? Anyway, it specializes more in time-series analysis.

A simple list

What can I see around me?

- SAS: I have no idea with it, since most economic students are afraid of learning programming, and I have even been told that "I do not want to see it anymore!". That is an economist: maximize his own utility or reach a particular goal with minimal cost.
- Matlab: Give me a reason to refuse it! –>R?

Where is R?

- In fact, seldom can I hear from a person who says he/she has heard about R.

A simple list

What can I see around me?

- SAS: I have no idea with it, since most economic students are afraid of learning programming, and I have even been told that "I do not want to see it anymore!". That is an economist: maximize his own utility or reach a particular goal with minimal cost.
- Matlab: Give me a reason to refuse it! –>R?

Where is R?

- In fact, seldom can I hear from a person who says he/she has heard about R.

Outline

Core Methods

Example

Regression:

- Ordinary Least Squares (OLS)
- GLS/FGLS (Heteroskedasticity...)
- 2SLS (IV model)
- Quantile Regression

and figures (e.g. scatter graph)

Example

Parameter Estimation:

- Maximum Likelihood Estimate (MLE)
- Generalized Method of Moment (GMM)

Time-series Analysis:

- ARMA, ARCH...

Core Methods

Example

Regression:

- Ordinary Least Squares (OLS)
- GLS/FGLS (Heteroskedasticity...)
- 2SLS (IV model)
- Quantile Regression

and figures (e.g. scatter graph)

Example

Parameter Estimation:

- Maximum Likelihood Estimate (MLE)
- Generalized Method of Moment (GMM)

Time-series Analysis:

- ARMA, ARCH...

Leading?

American Economic Review

- The *American Economic Review* is the best economic journal around the world. I chose the issue of Jun, 2009.
  - Total articles: 18
  - Pure theoretical: 8
  - Main econometric methods: GMM+Panel Data, Poisson Regression, GMM+3SLS, OLS+Probit, Ordered Probit, Hedonic Regression, Dynamic OLS.

Leading?

- However, AER is a comprehensive journal rather than a econometric one. The methods may not be the most cutting-edge. Therefore, the *The Review of Economics and Statistics* shall be more convincing (I'm sorry that I cannot find *Econometrica* in my school's resource room).

The Review of Economics and Statistics

- Issue: Feb, 2009
- Main econometric methods: Bayesian and fat-tail, ZIP(zero-inflated possion), OLS+IV, Logit, Probit, Panel (fixed effect)

Chinese Journals

Then Let's have a look at domestic journals.

Economic Research Journal

- The *Economic Research Journal* is the best economic journal in China.
- Issue: Jul-Oct, 2009
- Total articles: about 50
- Empirical Research: about 35
- Main econometric methods:
  - MLE,LIML, GMM (2), Systematic GMM (2)
  - OLS (8), Dynamic OLS, 2SLS (4), B2SLS, WLS, 3SLS
  - Panel (fixed effect, 3), Dynamic Panel (2), Granger Causality Test, Var (2), VECM
  - Bootstrap, Stochastic Probit (2), PEA (2), MGARCH

Chinese Journals

Then let's pay attention to a more theoretical journal: *China Economic Quarterly*

China Economic Quarterly

- Issue: Apr, 2009
- Total article: 18
- Empirical research: 8
- Methods: GLS, 2SLS, Probit, GMM (3), HP, VAR, Panel (fixed effect)

World Economy, The Journal of Quantitative and Technical Economics

- Empirical/Total:5/7, 10/10
- SVAR (2), Dynamic Panel (2), FGLS, Ordered probit, DAG, 3SLS, MLE (2),SFA (2), Markov chain, TARCH

Chinese Journals

Then let's pay attention to a more theoretical journal: *China Economic Quarterly*

China Economic Quarterly

- Issue: Apr, 2009
- Total article: 18
- Empirical research: 8
- Methods: GLS, 2SLS, Probit, GMM (3), HP, VAR, Panel (fixed effect)

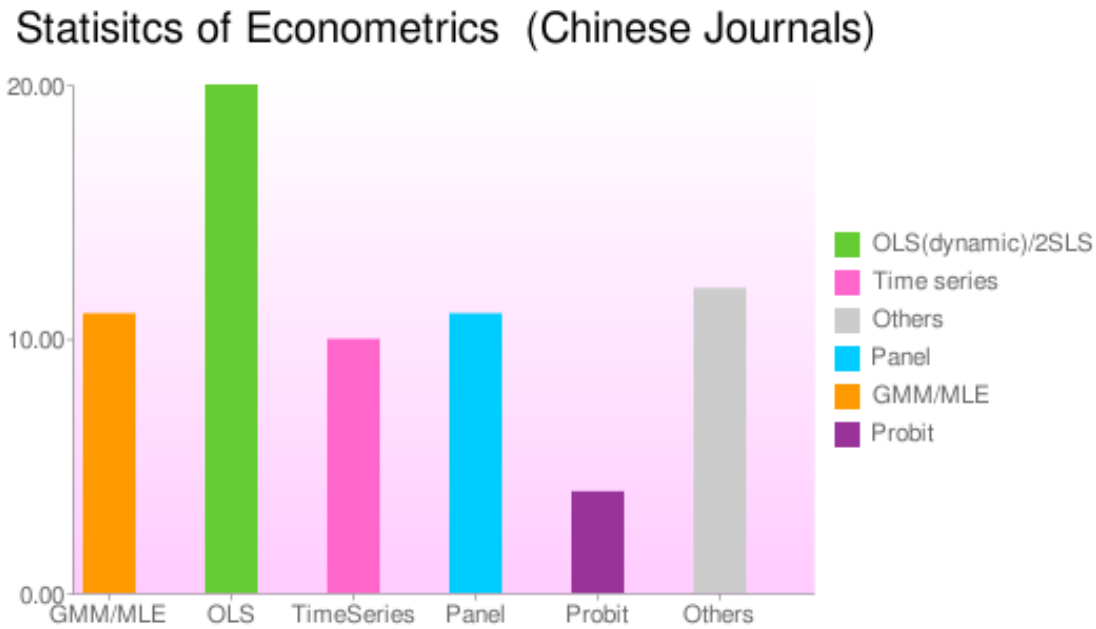
World Economy, The Journal of Quantitative and Technical Economics

- Empirical/Total:5/7, 10/10
- SVAR (2), Dynamic Panel (2), FGLS, Ordered probit, DAG, 3SLS, MLE (2),SFA (2), Markov chain, TARCH

Chinese Journals

Statistic journals

- *Application of Statistics and Management, Chinese Journal of Applied Probability and Statisties, Statistical Research*
- Related to economics/ Total: 1/10, 9/12, 10/16
- Nonparametric Estimation, SFA, Spatial econometrics (2), Bayes.



Outline

OLS and Regression

OLS in R

- OLS:
  - lm(crime ~ poverty + single, data=crime2)
- Robust regression (package MASS):
  - rlm(crime ~poverty + single, data=crime2)

Equation System Estimation in R

- Equation System Estimation (systemfit package)
  - OLS, Weighted Least Squares (WLS)
  - Seemingly Unrelated Regression (SUR)
  - Two-Stage Least Squares (2SLS), Weighted Two-Stage Least Squares (W2SLS), Three-Stage Least Squares (3SLS).

OLS and Regression

OLS in R

- OLS:
  - lm(crime ~ poverty + single, data=crime2)
- Robust regression (package MASS):
  - rlm(crime ~poverty + single, data=crime2)

Equation System Estimation in R

- Equation System Estimation (systemfit package)
  - OLS, Weighted Least Squares (WLS)
  - Seemingly Unrelated Regression (SUR)
  - Two-Stage Least Squares (2SLS), Weighted Two-Stage Least Squares (W2SLS), Three-Stage Least Squares (3SLS).

MLE & GMM

MLE in R

- mle(minuslogl, start = formals(minuslogl), method = "BFGS", fixed = list(), ...)

GMM in R (package GMM)

- GMM: gmm(g,x)

Quantile Regression (package quantreg)

- rq(stack.loss ~ stack.x,.5) #median (l1) regression fit for the stackloss data.
- rq(stack.loss ~ stack.x,.25) #the 1st quartile

MLE & GMM

MLE in R

- mle(minuslogl, start = formals(minuslogl), method = "BFGS", fixed = list(), ...)

GMM in R (package GMM)

- GMM: gmm(g,x)

Quantile Regression (package quantreg)

- rq(stack.loss ~ stack.x,.5) #median (l1) regression fit for the stackloss data.
- rq(stack.loss ~ stack.x,.25) #the 1st quartile

MLE & GMM

MLE in R

- mle(minuslogl, start = formals(minuslogl), method = "BFGS", fixed = list(), ...)

GMM in R (package GMM)

- GMM: gmm(g,x)

Quantile Regression (package quantreg)

- rq(stack.loss ~ stack.x,.5) #median (l1) regression fit for the stackloss data.
- rq(stack.loss ~ stack.x,.25) #the 1st quartile

Time Series Analysis

ARMA & ARIMA

- stats-package
  - arma(s, order=c(20,0)
  - arima(lh, order = c(1,0,0))

ARCH

- arch(x, lags.single = 16, lags.multi = 5, multivariate.only = TRUE)

Time Series Analysis

ARMA & ARIMA

- stats-package
  - arma(s, order=c(20,0)
  - arima(lh, order = c(1,0,0))

ARCH

- arch(x, lags.single = 16, lags.multi = 5, multivariate.only = TRUE)

Panel Data

plm Package

- Panel data econometrics is obviously one of the main fields in the profession, but most of the models used are difficult to estimate with R.
- plm is a package for R which intends to make the estimation of linear panel models straightforward. plm provides functions to estimate a wide variety of models and to make (robust) inference.

From: Giovanni Millo & Yves Croissant, 2008. "Panel Data Econometrics in R: The plm Package," *Journal of Statistical Software*, American Statistical Association, vol. 27(02), 07.

Outline

Data

- foreign (package): Read Data Stored by
- Minitab, S,
- SAS
- SPSS
- Stata
- Systat, dBase, ...

Outline

LaTeX

- Sweave allows R code to be included in a latex file.
- This is a good marriage
  - latex provides typeset text
  - R is statistically and graphic oriented

Outline

Micro and Macro Econometrics

- Maybe many people think that econometric methods are most essential in macroeconomics, where requires numerous analysis of time series data. The simplest example is to analysis (or predict) the correlations between GPD and something else, especially nowadays in China. Although the volume of available data on macro-level cannot be compared to the social survey's, the improvements of methods make the results more and more reliable, particularly those based on successful macroeconomic theories.

Micro and Macro Econometrics

- However, in recent year, more and more microeconomic researchers are concerning micro-level data. Is it only because "The bigger the sample is, the more convincing the result seems to be"? That is only a part of the reason. In fact, the rapid development of microeconomics requires more advanced micro-econometric methods to help the designer test the validation of their theoretical models.
- For example, in the field of labor economics, we often use Quantile Regression to illustrator the differences among different groups (based on percentile).

example about quantile regression

- Zhang Juwei and Xue Xinxin, 2008, "State and Non-state Sector Wage Differentials and Human Capital Contribution", *Economic Research Journal*, 2008(4).

Cooperation and Combination

- I've recently finished a paper about microeconomics, and the next step I need to do is dealing with thousands of transaction data. It is a big panel, and I would like to try some new methods to convince myself.
- I really think that economics needs the support of statistics, not only on the empirical analysis level, but also on the pattern of thoughts.
- At least, most economists lack the ability to find/collect useful data. Therefore, the knowledge of data mining shall be extremely helpful. Overall, while the economic student are studying statistic knowledges actively, I really hope that some statistic students (or if possible, professors are more welcome) can pay attention to the economic area and make a contribution.

Mutual Beneficial

- In economics we often talk about "social division of labor" and "comparative advantage". That is, if every individual does what they are most talent at, the final social production will be the highest. Here if we ignore transaction cost for a moment, and assume that we needn't to pay anything (money or time) for find a partner, then the probable cooperation between statisticians and economists will be mutual beneficial.
- In fact, the opportunity of communication like this conference is the most ideal case in an economic view, since it reduces the transaction cost in a large extent as well as the level of asymmetric information. (So I really appreciate the chance to participate here. Thank you all!)

Conclusion

- That is more easy than it seems to be.

Thank You!