

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Methodological Mistakes & Econometric Consequences

Dr. Asad Zaman

Nurturing Minds Seminar Series – PIDE

6th February 2014

Two Fundamental Errors in Methodology

GOAL OF TALK:

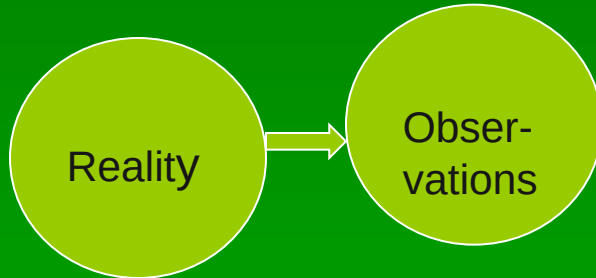
- What are these fundamental mistakes?
 - Focus on observables, instead of reality
- Why did they occur?
 - Influence of Logical Positivism
- What are the consequences in Econometrics?
 - Looking for patterns= good fit, Ignoring Cause/Structure

Rise & Fall of Logical Positivism

- The most spectacular story of 20th Century.
- Fight between Christianity & Science: Galileo
- Quest to prove Science superior.
- Positivism: All knowledge is based on observables (facts) + reason
- Knowledge about unobservables is impossible.

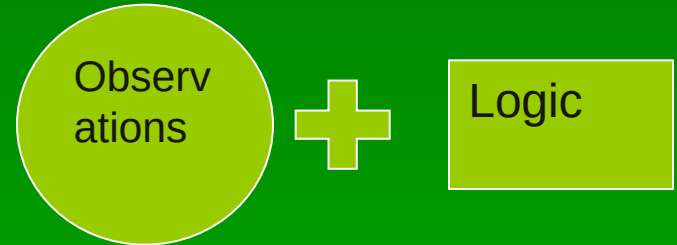
Picture of the World:

- PRE-Positivist:



- Observations are CLUES to Unobserved Reality. SKILL needed to put together COHERENT PICTURE

- POSITIVIST



- Facts/Observations or Data IS ALL THERE IS. We cannot have knowledge of Reality.

Wrong Picture of Science

- Earth revolves around Sun due to GRAVITY.
- Empiricist: Short-Hand for equation of observed motion of Earth.
- Realist: Equation holds because gravity exists.
- All Physicists are realists.
- Only Philosophers of Science are not.

Wrong Picture of Methodology

Positivist Myth;

- Scientists look at facts. Netural & Detached.
- Scientific Laws are patterns in the data.

Truth:

- Scientists look for REAL Explanations – structures of reality which can explain patterns in the data. Ontology, Causality

Consequences of this Wrong Description

- Behavioral Psychology.
- Probability – Frequentist & Bayesian
- Artificial Intelligence
- Causality
- Economics – any model which matches data.
- MASSIVE MISMATCH between what is done, and what is CLAIMED is being done.

Rhetoric: Assemble Facts into Picture

- Rhetoric, a Useful and Valued skill. How can we assemble the facts to create a **convincing** picture of **reality**.
- Post-Positivist View: Rhetoric is BAD. Since facts are SUFFICIENT to determine Truth, Rhetoric is just a way to distort facts for the sake of some (external, hidden) agenda.

Illustration: Samuelson's Mistake

- Revealed Preference Theory. Eliminates the unobservable preferences, and based on observable choices.
- HOWEVER, axioms for choice make sense ONLY if there exist (unobservable) preferences in reality.

Looking for Patterns in Data

Versus

Looking for Clues to Real Structures

DIFFERENT ACTIVITIES.

The Difference:

Looking for Patterns

- Forcing Patterns Onto Data
- No way to compare or differentiate between alternatives.

Looking for Clues

- Causality
- Structure

“Lost Causes” Kevin Hoover

VAR models

Finding Patterns in Data

Forecast Competitions: M1, M2, M3

Exponential Smoothing

- Single, Double, Holt-Winters, Damped

Time Series: ARMA, ARMAX, ARARMA

Neural Networks

State Space Methods, Kalman Filter.

Which method works best on real time series?

Effort Wasted on Nonsense Question

- Patterns are clues to Causal Mechanism
- Different Causal mechanisms generate different types of series.
- Method which comes closest to capturing underlying causal mechanism will work best.
- Performance in these competitions will vary at random without rhyme or reason.

Finding the best fit

- With sufficiently large collections of models, we can fit ANY data set.
- High degree polynomial will pass through any finite collection of data.
- In looking for clues, MAIN issue is to find a SURPRISING fit.
- Hendry Methodology creates greater chance for surprise, and works marginally better.

Tools to prove anything

- Unit Root tests: Get any result you want.
- Causality tests: Get any result you want.
- Effect of X on Y: Get any result you want.
- Auxiliary assumptions, specifics details of modeling, testing, transforming, treating of outliers, error process, etc. provide complete control. Infinite numbers of curves fit any finite data set.

SURPRISE: Correlation

- Take stationary series X,Y. H is High, L is Low

X		H	H	H	L	H	H	L	H	H	L	H	H	L	L
Y		H	H	L	L	H	H	H	H	H	L	L	H	L	L

- Why does it happen that whenever X is Hi, Y is also Hi, and when X is Lo, Y is also Lo?
- This is surprising. There must be some REASON for it. Look for a reason.

No Surprise: Spurious Correlation

- Take two upward trending series:

X	L	L	L	L	L	L	L	H	H	H	H	H	H	H
Y	L	L	L	L	L	L	L	H	H	H	H	H	H	H

- Perfect match. No surprise.

Lesson: Fit does not matter, REASON for Fit Matters

- Reason: Use a Very Large Number of Models.
- Reason: Guided Specification Searches.

In these cases, good fit is no surprise. Tests of Residuals = Index of Conformity.

When is a good fit a surprise:

- Theory predicts a fit. We find a good match.
- PROCESS by which match is found matters.
- Need to calculate SURPRISE FACTOR

Bad Results from looking for good fits:

JUST LOOKING FOR A FIT can lead to any result.

No way to discriminate between large numbers of equally good fits.

Data can provide misleading clues.

Must refer to REALITY to assess.

Misleading Clues from Data:

- Confounding Factors May Exist
- Real World Info: Admission $\neq$ ► Gender

	Male Applicants	Female Applicants
Total	1000	1000
Admitted	300	700
Percentage	30%	70%

VAR models cannot yield info about real world.
Need to go look at real world mechanisms.

Visible Variable: Departments

- Discrimination is against females:

	Female Applicants		Male Applicants	
	Social Science	Engineering	Social Science	Engineering
Applied	900	100	100	900
Admitted	690	10	100	200
Percentage	77%	10%	100%	22%

- Observational studies: it is impossible to avoid such possibilities. Data only gives CLUES

Bad Effects of Wrong Methodology:

- Searching for Fits, without looking at mechanisms, structure, does not generate knowledge—i.e. VAR

Does USA GNP have Unit Root?

- Diff stationary; Nelson and Plosser (1982),
- Trend Stationary; Perron (1989),
- Trend Stationary; Zivot and Andrews (1992),
- Don't know; Rudebusch (1993),
- Trend stationary; Diebold and Senhadji (1996),
- Diff stationary; Murray and Nelson (2002),
Kilian and Ohanian (2002),
- Trend stationary; Papell and Prodan (2003)

Does Export Led Growth Hold?

Year of study publication	ELG	GLE	BD	NC
1985 -1990	Jung & Marshall (1985) Ram (1987)	Results from Time Series Causality Studies for Indonesia		
1991-95	Hutchison and Singh (1992)	Ahmad & Harnhirun (1992), Hutchison and Singh (1992)	Bahmani-Oskooee et al. (1991) Dodaro (1993)	Hutchison & Singh (1992), Ahmad and Harnhirun (1995), Arnade & Vasavada (1995)
1996-2000	Piazolo (1996), Xu (1996), Islam (1998)	Pomponio (1996), Ahmad et al. (1997)	Ekanayake (1999)	Riezman et al. (1996),
2001-05	Amir (2004)			Lihan & Yogi (2003).
2006-10	Liwan and Lau (2007)	Pramadhani et. (2007). Bahmani-Claire (2009).		Nushiwat (2008)

We Can (and do) get any result we want

- PPP
- Debt Sustainability
- Efficient Markets Hypothesis
- RBC
- Co-Integration: Implies Existence of Equilibrium Long Run Relationship.

Selecting Indicators & using Confounding

- Africa has low growth compared to ROW
- Pick ANY criteria in which Africa is behind.
- That criteria is significant for growth.
- Example: Open Economy Indicators. Volume of Exports + Imports.

How to Use Data Correctly

- Data + Auxiliary Assumptions → Conclusions.
- Minimize Assumptions, Simplest Models.
- Learning about CAUSAL Effects is crucial.
- Data can NEVER provide Causal evidence.
- Can provide strong clues, but must be supplemented by variety of additional evidence.

Elementary Descriptive Methods are good for providing CLUES.

- Box Plots
- ANOVA
- New Graphical Techniques.

AVOID OVERFITTING. Use SIMPLE Models. MEASURE Surprise – requires keeping track of PROCESS by which model was selected.

Causality is Key Factor Ignored in Standard Econometrics

- $C = a + b Y + \text{error}$
- $Y = e + f C + \text{error}$

First Equation is useful and informative.

Second Equation is not.

WHY?

How do we know which direction to run the regression?

*$\mathcal{Y}$ is a **STRONG** determinant of C
 C is **NOT** a strong determinant of $\mathcal{Y}$.*

- Data CANNOT differentiate between the two forms.
- Correlation is SYMMETRIC.
- Causation is UNI-Directional.
- Causality is REAL Structure, is UNOBSERVABLE. Use Variety of Clues, Shoe Leather.

Causality is EVERYTHING

Successful regression depicts causal law

- $\text{Ln}(\text{GDP}) = a + b \text{Ln}(\text{News})$
- $\text{Ln}(\text{GDP}) = a + b \text{Edu}(t-10)$ [Barro]
- First equation is statistically stronger, but meaningless.
- Second is important IF IT represents a causal relation.
- Finding out requires a RANGE of evidences – Qualitative, Historical.

Does Money Cause Inflation?

- Friedman: YES
- Hendry & Ericsson: NO
- Others: Sometimes
- Recent Evidence: Tremendous Printing of Money by USA has not led to corresponding inflation.
- Value of Money determined by TRUST & other non-economic (political) factors.

Can we find out by running regressions?

Conditions for Granger Causality to Work:

- Timing & effects must be STABLE
- Data Sampling Frequency must correspond to causal chains.
- Functional Form, Lags must be correctly chosen. All relevant variables must be included.

Does Printing Money Cause Inflation

- Trace Causal Chains:
- Who prints money?
- What happens to it after it is printed?
- Whose hands does it go to?
- What do they do with it?

Many Types of Clues Are Available

- If Y determines C then $C|Y$ is stable.
- If C determines Y then $Y|C$ is stable.
- One has to look for Breaks in structure.
- There are lots of other tricks of this type.
EXOGENEITY is very complex.

Screening: $X \Rightarrow Z \Rightarrow Y$

- Y on (X,Z) will give misleading results.
- Need to check partial correlation between X and Y holding Z fixed.
- Example: Migration & Crop Yield.

Keys: Causality & Structure

- Econometrics does not have any methods to detect causality. Granger Causality is not an effective tool.
- Econometrics does not have any methods to determine structure. Which variables affect which others. The famous inclusion-exclusion restriction.