

Jesper B Sørensen

August 2012

Endogeneity is a fancy word for a simple problem. So fancy, in fact, that the Microsoft Word spell-checker does not recognize it.

Technically, in a statistical model you have an endogeneity problem when there is a correlation between your X variable and the error term in your model.

What does this mean? Well remember that the error term in your model is due to all of the stuff in your dependent variable that is not due to the variables you have in your model.

So in the broadest sense an endogeneity problem arises when there is something that is related to your Y variable that is also related to your X variable, and you do not have that something in your model. Call that something Z, although notice that I have not claimed that Z is a variable – it is just a “something.”

What this broad meaning of endogeneity suggests is that there are a wide variety of sources of endogeneity problems. In fact one of the problems with the use of the term is that people use it to cover many different things, yet the solutions to those different things can be quite different.

For example, endogeneity in this broad sense can be caused by omitted variables, or unobserved heterogeneity. In this case, the endogeneity complaint is a complaint that you left a variable (or two) out of your model.

This is obviously very familiar to everyone in this room. One reason why it is familiar is that we all know how to deal with it: measure the variable and put it in. And we all know how to fight with the reviewer about such things.

Unfortunately, there are other sources of endogeneity that are not so easily dealt with. And in fact, I think that in most cases where the charge of endogeneity is filed, people are not so much worried about omitted variables. Rather, what they are worried about is things like simultaneity – i.e., X causes Y but Y also causes X, -- and self-selection. The problem with such endogeneity problems is that no amount of control variables will address them.

For an example of simultaneity, consider a very nice paper by Simcoe and Waguespack on status signals. Sociologists are somewhat obsessed with the idea that rewards accrue to actors because of their status, and claim that status affects the performance of those actors – i.e., quality. This is the Matthew Effect – because people defer to high-status actors and wish to affiliate them, these actors reap higher rewards and get more recognition. Yet the problem is that quality also affects status – people get recognized because they do good work.

But if all there is to the status effect is the effect of accumulated quality, then there really is not a whole lot for sociologists to talk about, at least with respect to status. So trying to see whether the status signal has an independent effect is very important – but controlling for variables (even measures of quality) is not good enough.

Simcoe and Waguespack take advantage of a situation in which the identity of authors is sometimes obscured. The reason why the identity is obscured in some cases and not others is unrelated to the quality of the author's work. So if you can detect a difference in attention when the identity is obscured and when it is not, you can plausibly attribute that to an effect of status.

Self-selection is another source of endogeneity. Firm size and entrepreneurship...

I think we are starting to see in organizations and management research an increased concern with these kinds of endogeneity issues. If you think about the different kinds of endogeneity concerns, what you see is an increasing concern with more complex forms of endogeneity – everyone takes it for granted that we should be worried about omitted variables, a fair number of people think seriously about self-selection problems, and we are seeing the beginnings of an emphasis in management research on worrying about simultaneity problems.

I think this is a good thing. But we have to be careful not to get too obsessed.

One of the reasons it is a good thing is that I think a lot of quantitative researchers – both in the organizations and management fields, and in my own field of sociology – take too much comfort in the power of control variables.

The important thing to take away from the examples of simultaneity and self-selection is that it is impossible to include enough control variables in a regression model, but that does not mean that you should try to collect more.

In general I think a lot of researchers, in thinking about their research designs, misallocate effort. They dedicate too much time to collecting a wide range of control variables, even though it is often the case that many of those variables either don't affect the X or affect the Y (or both). Simcoe and Waguespack did not add value by expanding the list of control variables.

Instead, the efforts of researchers are often better focused on thinking about actual research design -- how they might anticipate and address concerns about endogeneity in the form of things like simultaneity and self-selection.

This involves thinking deeply about the X and the Y – the two constructs you are most interested in. Of course we already do that when we develop our theoretical claims. So really what I am saying is that we need to think deeply about the X – and in particular the processes that bring about the X.

What you will notice about simultaneity and self-selection is that they are both processes that bring about the level of X. And the extent to which we are concerned about them is the extent to which those processes are related to the Y variable.

Yet all too often, I think, researchers don't think very deeply about the social processes that generate their independent variables of interest. I think this is due to a false analogy between regression models on observational or field data and experimental data. (An observation made long ago by Stan Lieberson.)

In an experiment – of course the beauty of an experiment – different people are assigned to different levels of X by chance. This is the beauty of an experiment – the reason why it allows us to feel confident in our causal inferences.

In other words, when we look at our X variable, we all too often think that different people are assigned to different levels of the X variable by chance. Or we think that once we have controlled for enough variables, the assignment to different levels of X is by chance.

Of course if you asked people that explicitly, they would probably be appropriately skeptical and say that the assignment process is non-random. But the problem is that people don't think about this explicitly.

And I think a real benefit of the growing emphasis on endogeneity concerns is to force people to be explicit about this assumption. Right now, people often come to face this issue in the review process. But we can hope that over time people will internalize this concern in the same way that they have internalized the concerns with omitted variables.

What I really want to emphasize is this last point: that you want to internalize the concern with selectivity and self-selection *early* in the process: at the research design stage. A lot of the language and techniques around endogeneity is statistical – instrumental variables, sample selection models, etc. But these techniques are often times solutions to problems caused by the limitations of the research design.

I've emphasized some of the positives of the growing emphasis on endogeneity concerns. I do not have any doubt that this is a positive force in management research. And the creativity and effort that has been put into a lot of this work is impressive.

But there can always be too much of a good thing. It is important to think about the tradeoffs.

In thinking about the possible tradeoffs, I think it is most helpful to think about the difference between theory generation and theory testing. A lot of people think this is a distinction between abstract, conceptual work – thinking – on the one hand, and empirical analysis on the other. But I think this is wrong, since the abstract conceptual work does not happen without engagement with the empirical world, and since the empirical testing that we do does not happen without a conceptual overlay.

From a research design perspective, there is a tradeoff between theory generation and theory testing. In my experience, the more compelling the test that an author performs, the less creative is the theorizing. This is not to say that the theorizing is dull or uninteresting. But rather it is to say that the paper with the awesomely convincing test of the causal mechanism typically does not do much to generate a new insight – typically you are testing an idea that already exists. In short, as many others have noted, there is a tradeoff between fertility and rigor.

There are several reasons for this. One is that you only have so much creative energy, and things need to get published and get out into the debate. And so in any given project you have to allocate your creative energy, I think, either to generating new ideas and theories, or to designing the compelling test of the causal mechanism. Second, creativity demands imagination and speculation; a willingness to say “what if” and work out the details later. Yet a rigorous causal test is all about making the causal inference as immune as possible to the imaginative speculations of your critics. The better they can do in terms of dreaming up alternative interpretations, the worse you have done in providing a robust test. There are likely other explanations as well.

If you believe that this tradeoff is real – and you should – then it has a number of important implications. The first is that if there are real tradeoffs, the worst papers are those that pretend that there is no tradeoff – i.e., that pretend that it does not exist. For example, they are theoretically creative papers that vastly over-claim with respect to the strength of their evidence. Or they are papers with elegantly designed critical tests of existing theory that simultaneously try to convince you that they contain path-breaking ideas. Second, as a consequence, you should try to decide for yourself which side of the tradeoff you prefer. This does not have to be a lifelong decision, but for any given paper you should make a decision. Don't do a half-assed job on theory and a half-assed job on the empirics. Decide where your strengths lie, and where your project's strengths lie, and have the courage of your convictions. Finally, as a reviewer and critic of other people's work, keep the tradeoff in mind. If you review a paper that is clearly positioned as providing a more rigorous test of an important idea, it is not fair game to complain about the fact that there is no new theory. Similarly, if a paper goes a bit beyond what the

empirical evidence might support, but is circumspect about its limitations, don't pile on because the paper lacks true random assignment. Any one paper cannot do everything; our business is all about triangulating.

The Importance of Being Levered

***Pierre Azoulay
Massachusetts Institute of Technology
Sloan School of Management***



**Massachusetts
Institute of
Technology**

***“In God I trust, everybody else bring
their data”***



[At Least] Two Empirical Cultures in Strategy & Innovation Research

- ***Grounded Theory/Ethnography/Field Studies/Case Studies***
 - *Where else are worthwhile hypotheses going to come from? (e.g. Fiona Murray's work on tissue engineering and the oncomouse)*

- ***Econometric Studies***
 - *Descriptive studies ("Just the facts, ma'am", as in Gans et al. 2002; Jones on Age and Great Invention ReSTAT 2009)*

 - *Theory-driven studies (e.g., Lerner 1997 on technology races; Klepper and coauthors on the product-life cycle; Lafontaine & Shaw on Franchising)*

 - ***Identification of causal relationships***
 - *What is the effect of VC status on the performance of biotech startups?*
 - *Are knowledge spillovers geographically localized?*
 - *Do long-term incentives really stimulate exploration?*

Which ingredients do papers (of the causal inference type) need?

[David Romer, quoted by Brad DeLong]

1. *A viewpoint*
2. *A lever*
3. *A result*

Types of Lever

➤ **Randomized experiments**

- *Nagin et al. 2002 on the “rational cheater” model*

➤ **Quasi-experiments**

- *DD and its variants (e.g., Furman & Stern 2010)*
- *Regression-discontinuity design (e.g., Keys et al. 2010)*
- *Clever survey design (Stern 2004; Hsu 2004)*
- *Clever archival data collection (Simcoe & Waguespack 2009; Williams 2010; Fernandez and various coauthors on social networks in hiring)*
- *Instrumental variables (Doyle 2010 on the “vacation from hell”)*

➤ **Poor Man’s experiment**

- *Non-parametric matching (Jaffe et al. 1993; Thompson & Fox-Kean 2004)*
- *Parametric matching (Azoulay et al. 2009)*

➤ **No experiment**

- *Insider econometrics (Lazear 2000; Ichniowski & Shaw, various papers)*

Leverage Signs

- ***The parent/sibling test***
- ***The graphical test***
 - *Source of variation*
 - *Main effect*
- ***The design-to-method ratio test***

Almond et al. (QJE 2009)

Chernobyl's Subclinical Legacy

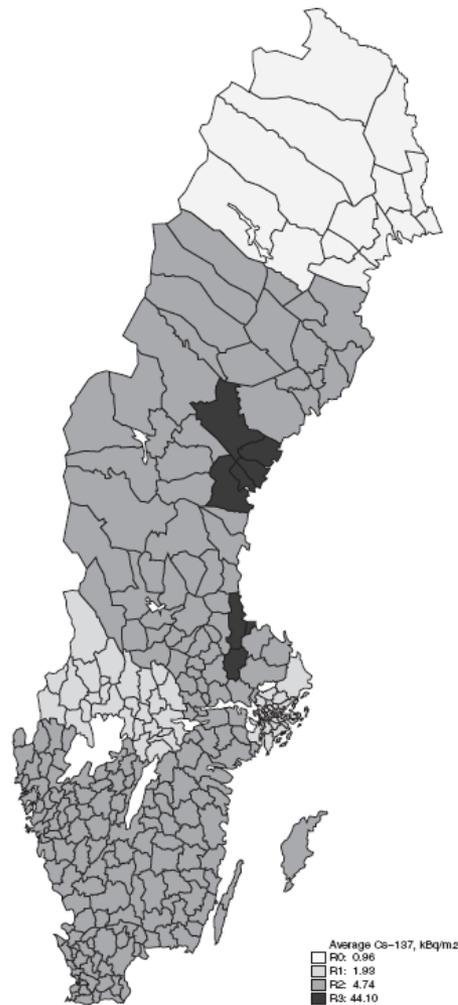


FIGURE III
Cesium-137 Ground Deposition in kBq/m² by Area (cf. Table I)

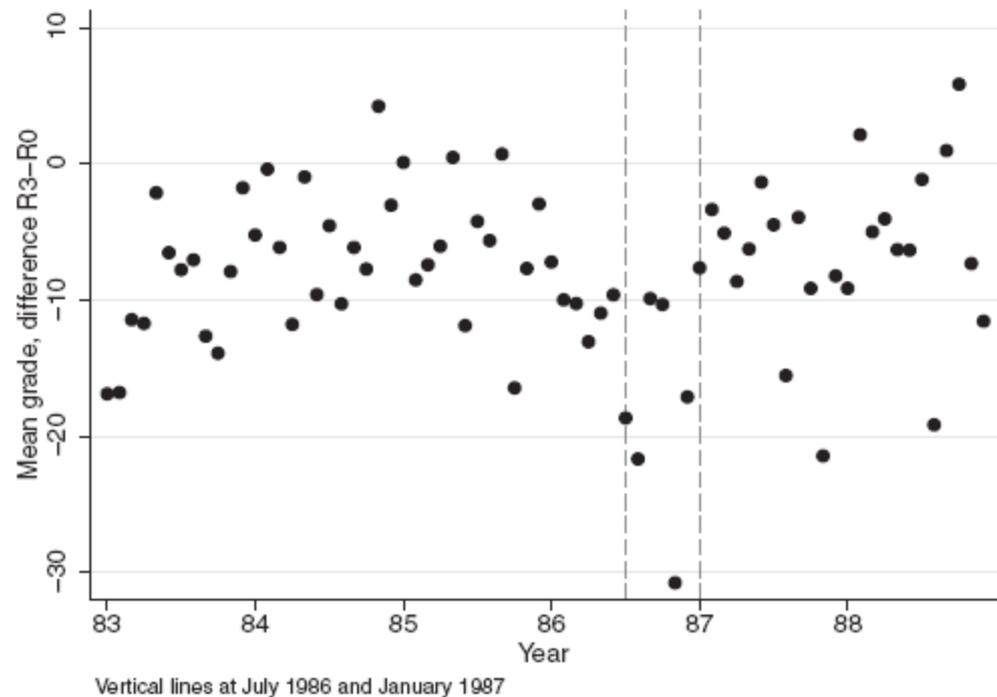
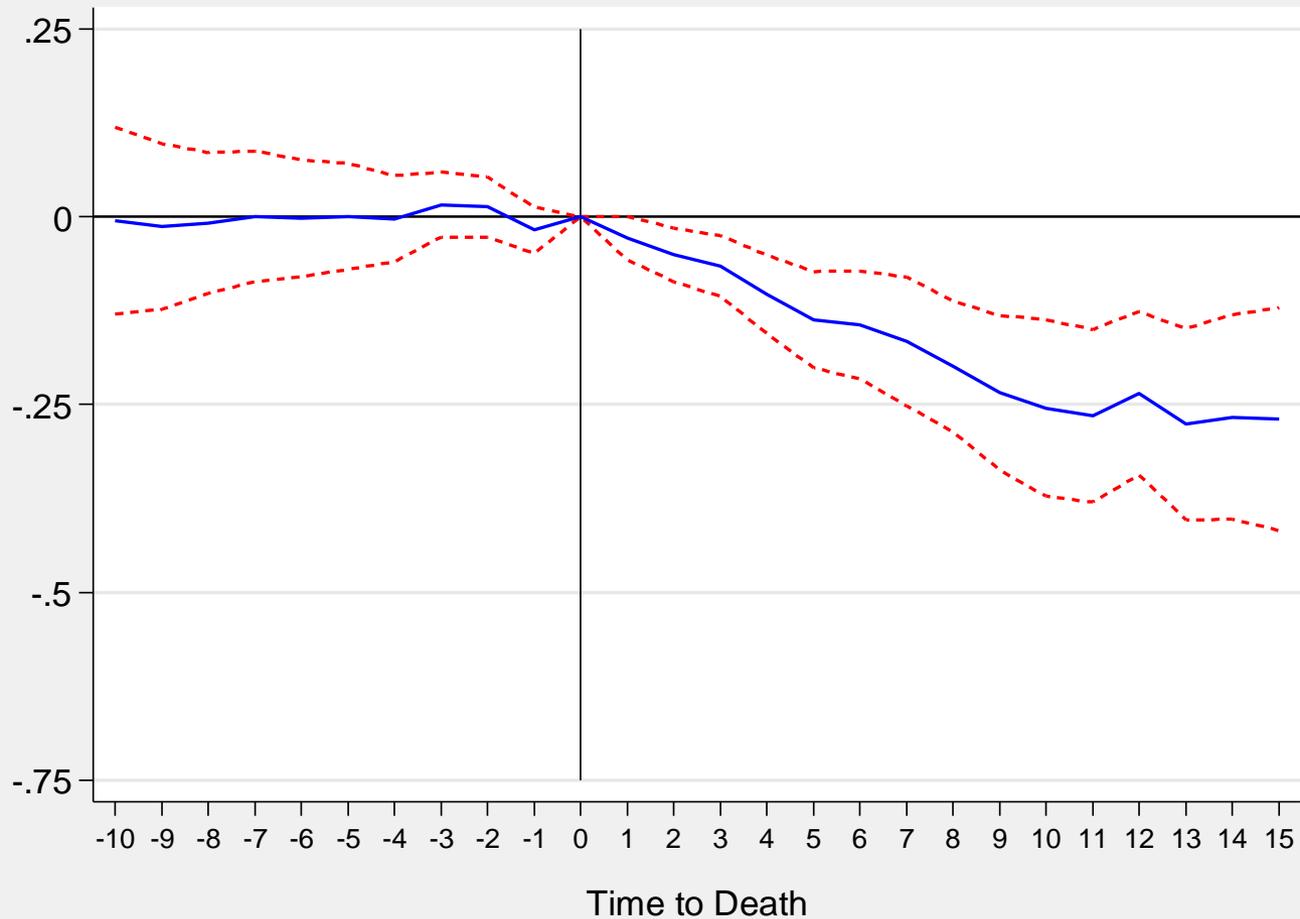


FIGURE V
Difference in Mean Grade Sums by Calendar Month of Birth: R3 (Eight Most Exposed Municipalities) Relative to R0 ("Norrbotten")

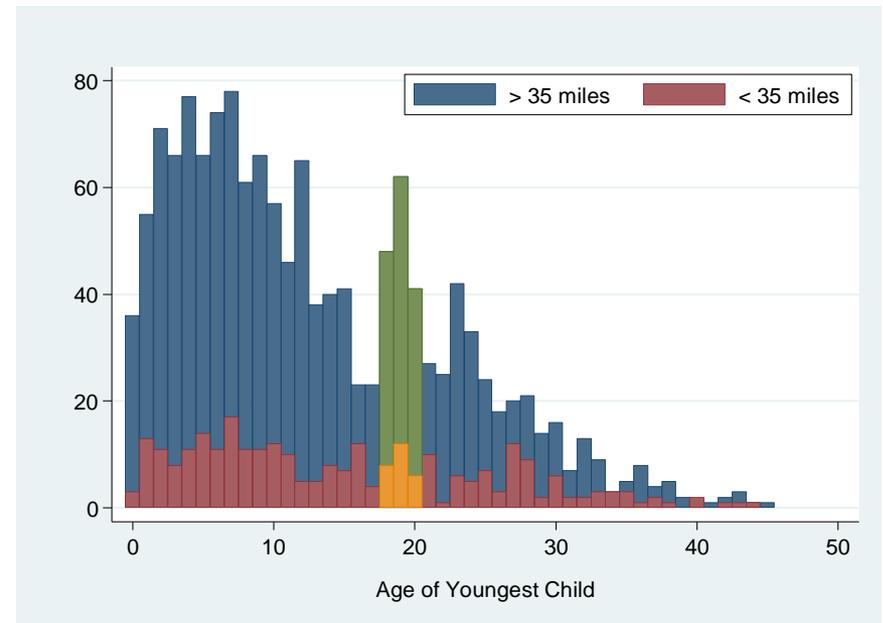
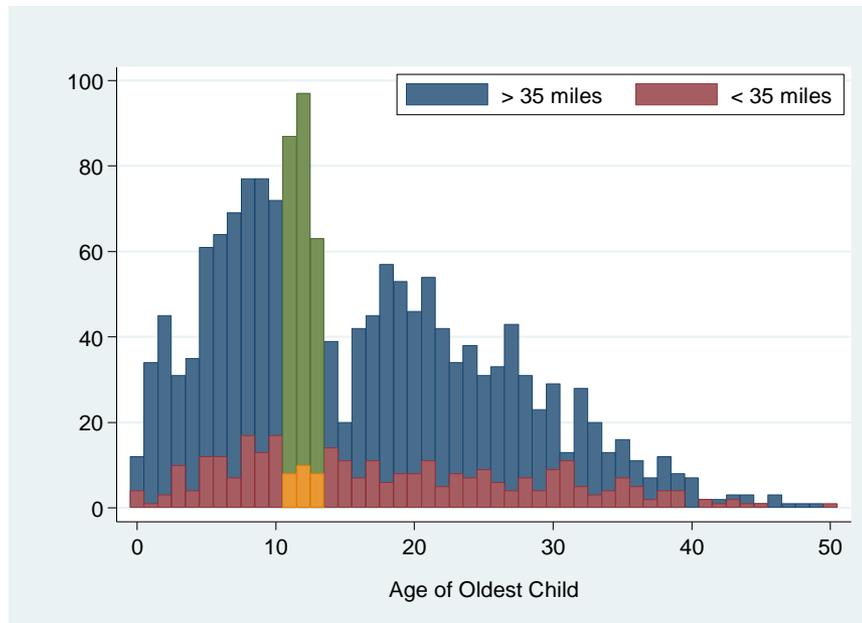
Azoulay et al. (QJE 2010)

The Importance of Being Alive



Azoulay et al.

Number of Job Switches for 3,500 “Superstars of Medicine”



Leverage Etiquette

- **Every method for causal inference in observational data relies on untestable assumptions; make them more plausible**
 - *RCTs: test that randomization was actually successful*
 - *Diff-in-Diffs: check the absence of pre-intervention trends*
 - *RD Design: check the absence of discontinuity*
 - *in the distribution of exogenous covariates around the threshold*
 - *in the distribution of the outcome variable in irrelevant subsamples*
 - *IV: does the effect disappear when the instrument shifts irrelevant margins?*
 - *Matching: test that propensity-score weighting balances “unused observables”*

The Pain of Leverlessness

- **“Not Even Wrong”**
 - *Fishing expeditions and the $n-1$ problem*

- **Econometrics as ceremony or obfuscation**
 - *Implausible instruments*
 - *Heckits without exclusion restrictions*

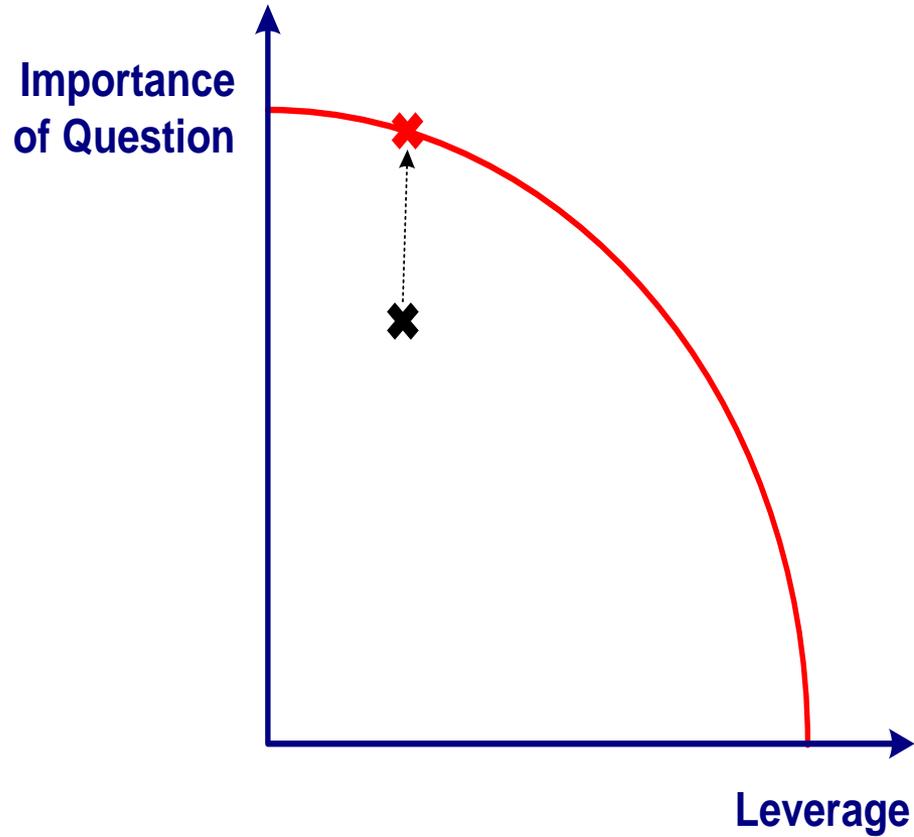
Why is leverage increasing?

- ***Better and more data***
- ***Pressures from referees and editors***
 - *Optimistic view: reflects fundamentals; design and data availability are complements*
 - *Cynical view: a passing fad*

Leverage skeptics

- ***Design-based research leads to boring papers***
 - *Speak for yourself, Kemo Sabe!*
- ***Design-based papers lack external validity***
 - *So do designless papers...*

The Leverage Frontier



Detecting Cause-and-Effect through Triangulation: Some Tips and Examples

Rosemarie Ziedonis

University of Oregon
AOM PDW August 2012

Step 1: Reframe the issue

“Identification PROBLEM”

“Endogeneity PROBLEM”

“Selection PROBLEM”



Be a Social Science Detective!



- I have a theory
- I see clues
- Could something other than my theory explain those clues?

How?



- ✓+ Find a natural experiment, valid instrument, etc...
- ✓- Over-assert
 - My theory is A, the clues confirm it!
 - Add a quiet "P.S." that the clues are consistent with Theories B and C as well
 - Hang your hat on flimsy evidence (2SLS w/o exclusion; IV with no logic)
- ✓ Refuse the job
- ✓ Triangulate!

Triangulation



“In the social sciences, triangulation is often used to indicate use of more than two methods in a study with a view to double (or triple) checking the results.”

Also known as “cross-examination”

--wikipedia

Two Examples

MANAGEMENT SCIENCE
Vol. 50, No. 6, June 2004, pp. 804–820
ISSN 0025-1909 | eISSN 1526-5501 | 04 | 5006 | 0804

informs.
doi 10.1287/mnsc.1040.0208
© 2004 INFORMS

Don't Fence Me In: Fragmented Markets for
Technology and the Patent Acquisition
Strategies of Firms

Rosemarie Ham Ziedonis
University of Michigan Business School, 701 Tappan, Ann Arbor, Michigan, rzied@umich.edu

Journal of Financial Economics 98 (2010) 478–499

Contents lists available at ScienceDirect

Journal of Financial Economics
journal homepage: www.elsevier.com/locate/jfec

Corporate venture capital and the returns to acquiring
portfolio companies^{a, c}

David Benson^a, Rosemarie H. Ziedonis^{b,*}

^a Marriott School of Management, Brigham Young University, Provo, UT 84602, USA
^b Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA

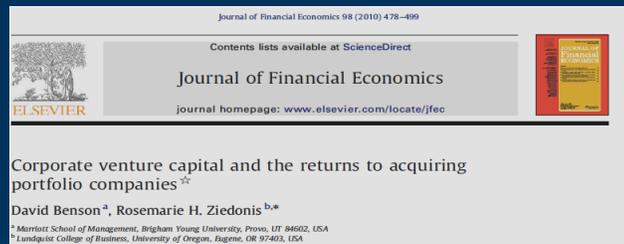
Example 1: “Don’t Fence Me In”

- The “Crime”: a surprising surge in patenting by IT companies (Y)
- Theory: a strategic response to market frictions and concerns of hold-up
 - X (external rights ↑) → W (friction ↑) → Y (Self-insure ↑↑)
- Clues: X, Y but not W
- Alt Theory: X and Y are caused by Z (technological opp), not W

Approach

- Dig deeper!
 - If my theory is correct, what else should I find?
 - Effects = amplified post “regime shift”
 - Effects for Type A firms > than for Type B firms
- Revisit the phenomenon
 - Did a technological shift **coincide** with the Regime Shift? Or **disproportionately** affect Type A v. B?
 - Read books & talk with people!
- Find an indirect proxy (trends in same sector in countries w/o “regime shift”)

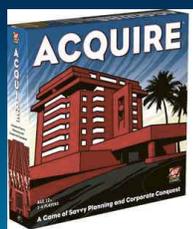
Example Two



An Anomaly Paper



- Expect Theory A
- Clues AT ODDS with Theory A
- Investigate Theories B, C, and D instead
 - Read books, go talk to people!
- Shift settings and re-test



Summary



- It's cool to be a social science detective!
- To be a good one, you must...
 - Do some digging
 - Be creative
 - Be honest
 - Sharpen your tools...and invest in new ones



Some References

Less Obvious

1. Deaton A. 2009. Instruments, Randomization, and Learning about Development. *Jl of Economic Literature*, 48(2): 424-455.
2. Diamond J, Robinson JA, ed. 2010 Natural Experiments of History. Harvard Univ Press.

More Obvious

3. Angrist JD, Pischke JS. 2008. Mostly Harmless Econometrics. Princeton Univ Press
4. Hamilton B, Nickerson J 2003. Correcting for Endogeneity in Strategic Management Research. *Strategic Organization*, 1, 51-78.
5. Shaver JM. 1998. Accounting for Endogeneity When Assessing Strategy Performance: Does Entry Mode Choice Affect FDI Survival? *Management Science*. 44(4): 571-585.
6. Bascle G. 2008. Controlling for Endogeneity with Instrumental Variables in Strategic Management Research. *Strategic Organization* 6(3): 285-327.
7. Reeb D, Sakakibara M, Mahmood IP. 2012. From the Editors: Endogeneity in International Business Research, *Jl International Business Studies* 43: 211-218.

My Examples

8. Ziedonis RH. 2004. Don't Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms. *Management Science* 50(6): 804-820.
9. Benson D, Ziedonis RH 2010. Corporate Venture Capital and the Returns to Acquiring Portfolio Companies. *Jl of Financial Economics*, 98: 478-499.

The AE says: “You need a natural experiment...”

- What is it?
 - Random assignment for construct of interest
- How do you find one?
 - Know your context
- Why does it matter?
 - Conclusive empirical results
 - Theoretical precision

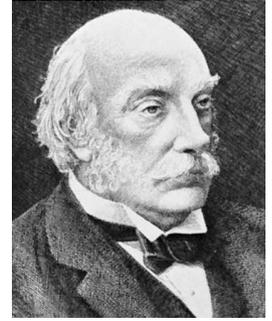
Status, Quality, and Attention: What's in a (Missing) Name?

Management Science 57(2), pp. 274–290

Tim Simcoe, Boston U. and NBER

David Waguespack, U. of Maryland

Status and Performance



- Theory: Social Signals matter for exchange when quality is uncertain
- Cumulative advantages for the well positioned via attention and resources
- Typical estimation approach: regress observed position plus controls on performance
- Identification Problems: omitted variables (quality), reverse causation
- Solution: disentangle status and signal
 - Merton 1968: “Rayleigh’s name was either omitted or accidentally detached [from a manuscript] and the Committee turned it down as the work of one of those curious persons called paradoxers. However, when the authorship was discovered, the paper was found to have merits after all.”



The Internet Engineering Task Force (IETF)



- The *de facto* Internet standards body
- Public data: 90K proposals, 3.5K publications, 7K authors, 800K listserv messages, 7.5K leadership appointments, 60K conference attendees over ~20 years.
- Authors submit manuscripts
- IETF posts and announces the manuscript
- Non-blind community review
- Revise, quit, or gets published
- “Et al” natural experiment – from 2000 to 2003 some prominent author names replaced with “et al” on email announcements

Email Announcement

* To: IETF-Announce ;
* Subject: I-D ACTION:draft-duffield-framework-papame-00.txt
* From: Internet-Drafts@ietf.org
* Date: Mon, 19 Nov 2001 08:33:38 -0500
* Reply-to: Internet-Drafts@ietf.org
* Sender: nsyracus@cnri.reston.va.us

A New Internet-Draft is available from the on-line Internet-Drafts directories.

Title: A Framework for Passive Packet Measurement

Author(s): R. Bush, N. Duffield, A. Greenberg, M. Grossglauser, J. Rexford

Filename: draft-duffield-framework-papame-00.txt

Pages:

Date: 16-Nov-01

A wide range of traffic engineering and troubleshooting tasks rely on reliable, timely, and detailed traffic measurements. We describe a passive packet measurement framework that is (a) general enough to serve as the basis for a wide range of operational tasks, and (b) relies on a small set of primitives that facilitate uniform deployment in router interfaces or dedicated measurement devices, even at very high speeds. This document describes the motivation for such a framework through several operational examples, defines the measurement primitives (filtering, sampling, and hashing), and illustrates their use.

A URL for this Internet-Draft is:

<http://www.ietf.org/internet-drafts/draft-duffield-framework-papame-00.txt>

Draft

INTERNET-DRAFT
draft-duffield-framework-papame-00

Nov 14, 2001

Randy Bush
Nicholas G. Duffield
Albert Greenberg
Matthias Grossglauser
Jennifer Rexford
AT&T Labs - Research

A Framework for Passive Packet Measurement

Abstract

A wide range of traffic engineering and troubleshooting tasks rely on reliable, timely, and detailed traffic measurements. We describe a passive packet measurement framework that is (a) general enough to serve as the basis for a wide range of operational tasks, and (b) relies on a small set of primitives that facilitate uniform deployment in router interfaces or dedicated measurement devices, even at very high speeds. This document describes the motivation for such a framework through several operational examples, defines the measurement primitives (filtering, sampling, and hashing), and illustrates their use.

Email Announcement

* To: IETF-Announce ;
* Subject: I-D ACTION:draft-zelig-pw-mib-00.txt
* From: Internet-Drafts@ietf.org
* Date: Fri, 13 Jul 2001 06:54:42 -0400
* Reply-to: Internet-Drafts@ietf.org
* Sender: nsyracus@cnri.reston.va.us

A New Internet-Draft is available from the on-line Internet-Drafts directories.

Title: Pseudo Wire (PW) Management Information Base Using SMIPv2

Author(s): D. Zelig et al.

Filename: draft-zelig-pw-mib-00.txt

Pages: 58

Date: 12-Jul-01

This memo defines an experimental portion of the Management Information Base (MIB) for use with network management protocols in the Internet community. In particular, it describes managed objects for modeling of Pseudo Wire (PW) services on a general Packet Switched Net (PSN). In addition, the current revision of the draft describes MIB module for PW operation over Multi-Protocol Label Switching (MPLS) [MPLSArch] Label Switch Router (LSR). Future revisions will include other types of PSN, for example L2TP, GRE, etc.

A URL for this Internet-Draft is:

<http://www.ietf.org/internet-drafts/draft-zelig-pw-mib-00.txt>

Draft

Internet Draft
Expires: January 2002

David Zelig
Sharon Mantin
Corrigent Systems

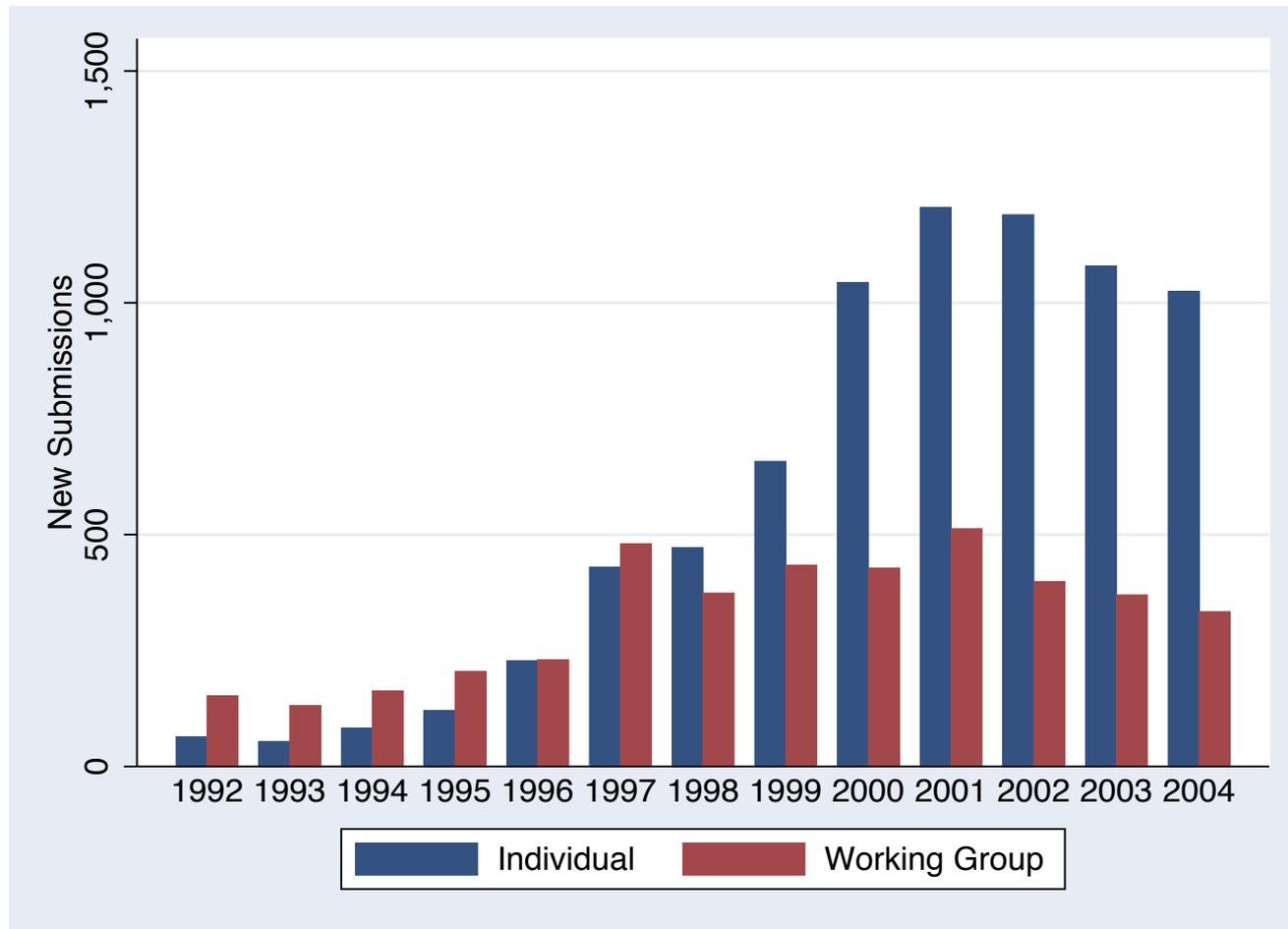
Thomas D. Nadeau
Cisco Systems, Inc.

Dave Danenberg
Litchfield Communications, Inc.
A. Malis
Vivace Networks, Inc.
July 2001

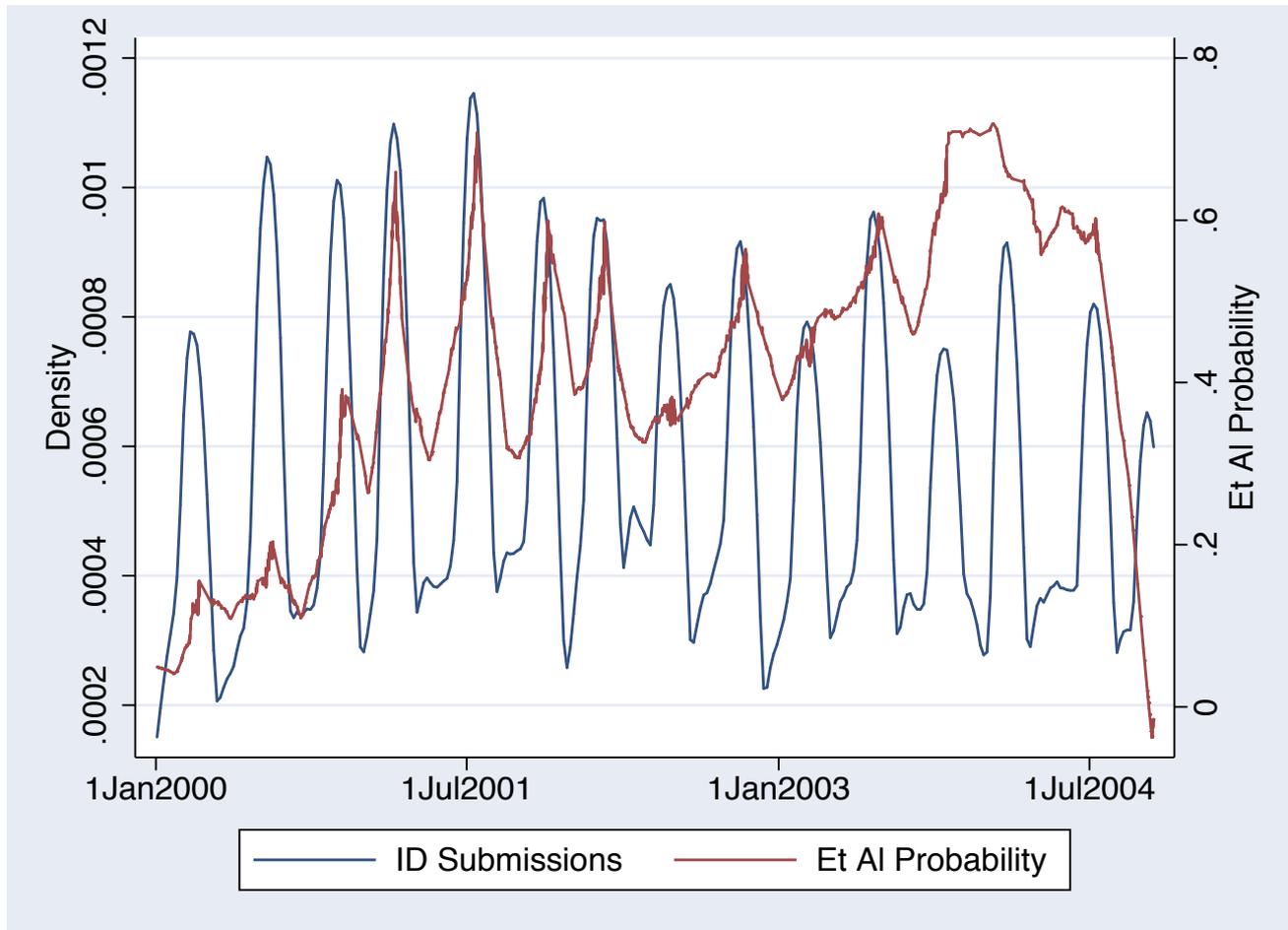
Pseudo Wire (PW) Management Information Base Using SMIPv2

This memo defines an experimental portion of the Management Information Base (MIB) for use with network management protocols in the Internet community. In particular, it describes managed objects for modeling of Pseudo Wire (PW) services on a general Packet Switched Net (PSN). In addition, the current revision of the draft describes MIB module for PW operation over Multi-Protocol Label Switching (MPLS) [MPLSArch] Label Switch Router (LSR). Future revisions will include other types of PSN, for example L2TP, GRE, etc.

Internet Draft Submissions



Q: When is Darlene *really* busy?



A: Just before an IETF Meeting

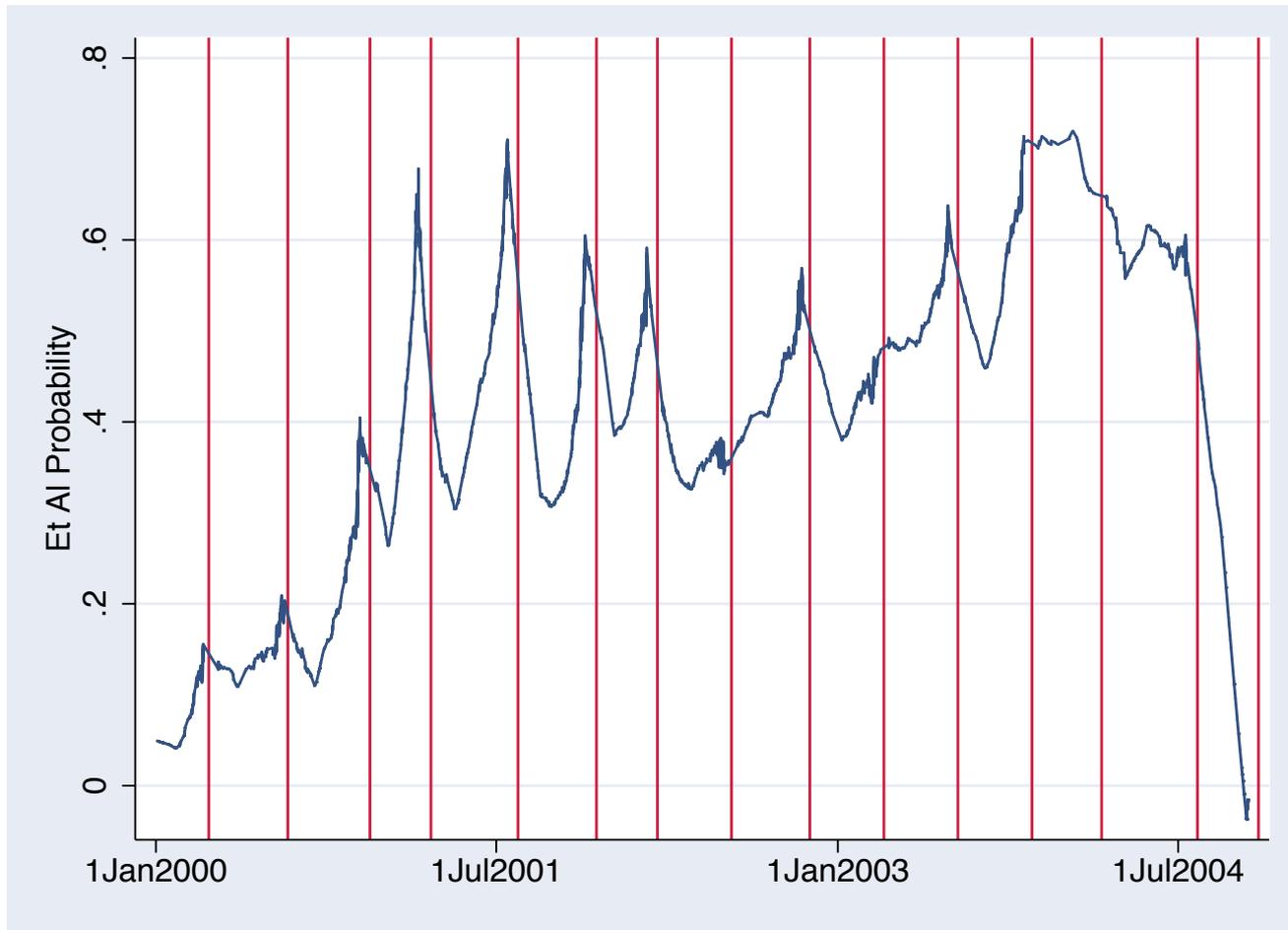


Table 3: Identity as a Signal[†]

Linear Probability Models of ID Publication
Unit of Observation = Internet Draft

Dependent Variable = Published as RFC

Sample	All Individual IDs		All WG IDs	
WG Chair Author	0.084** (0.01)	0.040* (0.02)	0.103** (0.03)	0.045 (0.04)
Unlisted WG Chair	-0.066** (0.02)	-0.056* (0.02)	-0.031 (0.07)	-0.011 (0.07)
Et Al Dummy	-0.026** (0.01)	0.006 (0.01)	0.048 (0.04)	0.071 (0.05)
Published RFCs		0.026** (0.01)		0.025+ (0.01)
Log Pages		0.004 (0.00)		0.027* (0.01)
Intl Author		-0.014+ (0.01)		-0.035 (0.03)
Multi-Sponsor		0.004 (0.01)		0.090* (0.05)
Days-to-Meeting		0.038** (0.01)		0.048* (0.02)
Constant	0.058** (0.00)	-0.076* (0.03)	0.385** (0.02)	0.240* (0.10)
Author Count Effects	N	Y	N	Y
IETF Meeting Effects	N	Y	N	Y
Observations	4013	4013	1405	1405
R-squared	0.021	0.045	0.012	0.035
Mean of DV	0.072	0.072	0.441	0.441

Publication Decision on 5418 Manuscripts submitted 2000-2003. ~80 new or revised per week.

High Status position = WG Chair

High Status Signal

Sample split into two groups:
1) High uncertainty: publication rate = 07%
2) Low uncertainty: publication rate = 44%

Name Signal matters when uncertainty high

Signal point estimate not sensitive to control variables

Large status position effect, but no signaling effect when uncertainty is low

Status position point estimates very sensitive to controls

Robust standard errors in parentheses; +10% significance; *5% significance; **1% significance. [†]See Table A3 for additional models that control for rank of Unlisted Chair in list of authors.

Conclusions

- Juliet (but not Shakespeare) was wrong...
 - “A rose by any other name would smell as sweet”
- Matthew Effects at the IETF
 - Individual vs. Working Group submissions
 - Volume + heterogeneity => screening heuristics
 - Attention and increasing returns
 - Name-based signals effect intermediate outcomes

Natural Experiment: Attribution in IETF publishing

- What is it?
 - Random assignment of author name visibility by Darlene
- How do you find one?
 - Know your context: IETF processes
- Why does it matter?
 - Conclusive empirical results: status matters for performance
 - Theoretical precision: signal and position are functionally distinct, but reinforcing

Caveat: can't be certain you will find a NE and can't prove randomness

Great news!: thinking experimentally will improve methods and theory... even if no NE found

Do I have an endogeneity problem? And does it matter?

Dealing with Endogeneity

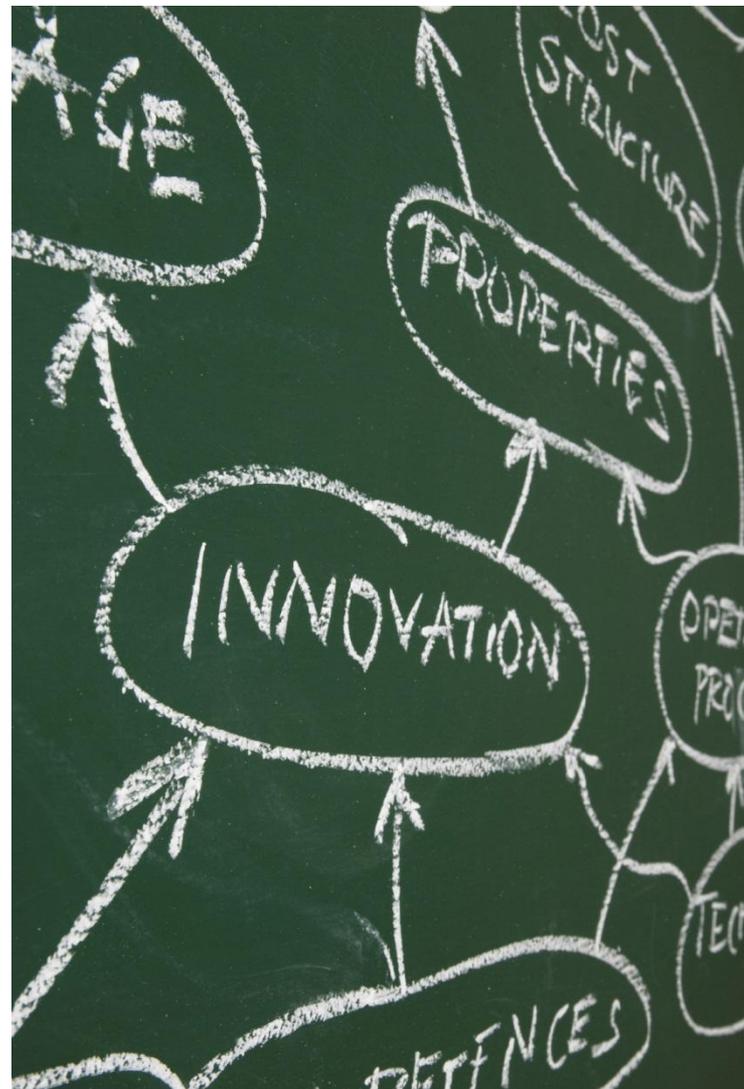
Alberto Galasso
University of Toronto

Academy of Management Annual Meeting, Boston, Aug 3 2012

My Research

- determinants of innovative activity;
- the management of innovation;
- the functioning of markets for technology.

*I use patent litigation data as
window on the market for
technologies*



My Perspective as Author/Referee

Instrumental Variables

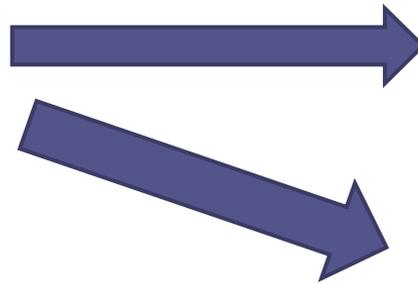
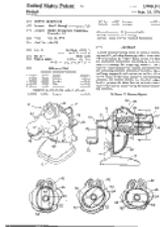
The Role of Theory



How does the market for innovation affect patent litigation?

The **market for innovation** –the licensing and sale of patents- is an important source of R&D incentives, especially for small firms and innovative entrepreneurs (Arora, Fosfuri and Gambardella, 2001; Gans, Hsu and Stern, 2002)

Growing concern in academic and policy debates that patent transactions can deter innovation if they take place for the purpose of extracting rents through **patent litigation**, and not associated with technology transfer (U.S. FTC 2011 report and U.S. Supreme Court)



Can we conclude that the market reallocates patent to entity with higher propensity to litigate?

OR

Increase in technology value made patent more likely to be traded and litigated??

RAL
X
TIONS

02

TIONS
K
1986
na
1,520

RE
WAL

FEDERAL
TAX
REGULATIONS

2002

REGULATIONS
UNDER
I.R.C. 1986
Sections
1.501 to 1.969

U.S. CODE
CONGRESSIONAL

FEDERAL
TAX
REGULATIONS

2002

REGULATIONS
UNDER
I.R.C. 1986
Sections
1.970 to 1. End

REGULATIONS
UNDER
I.R.C. 1986
Sections
30.0 to 601.901
31 C.F.R.
Parts 10, 19
INDEX

U.S. CODE
CONGRESSIONAL

FEDERAL
TAX
REGULATIONS

2002

REGULATIONS
UNDER
I.R.C. 1986
Sections
30.0 to 601.901
31 C.F.R.
Parts 10, 19
INDEX

U.S. CODE
CONGRESSIONAL

INTERNAL
REVENUE
CODE

1987

COMPLETE
TEXT
with
INDEX

U.S. CODE

INTERNAL
REVENUE
CODE

1988

COMPLETE
TEXT
with
INDEX

U.S. CODE

Galasso, Schankerman and Serrano (2012) “Trading and Enforcing Patent Rights”

Identification Strategy

According to section 1235 of the Internal Revenue Code, the **transfer of a patent** by an individual is treated as the sale of a capital asset and is subject to **capital gain taxes**. On the other hand, **patent litigation damages** (and licensing royalties) are **taxed as ordinary income**.

This means that the decision to trade a patent will be affected by the capital gains tax rate, but the decision to litigate will not!

Findings

- First, capital gains taxes strongly affect market transactions in patent rights granted to individual inventors.
- Second, the reallocation of these patent rights reduces litigation risk for individually-owned patents, on average, indicating that enforcement gains are more important than product market gains for such patents.
- Third, the marginal treatment effect of trade on litigation is heterogeneous. Patents with larger potential gains from trading are those with the highest estimated probability of changing ownership, suggesting that the market reallocates patent rights efficiently.

Do IP rights on existing technologies hinder subsequent innovation?

Cacophony of theories (Kitch, Green and Scotchmer, Heller and Eiseberg)

Some empirical evidence (Williams, 2012; Murray and Stern, 2007) but in most technology areas the relationship between innovation and IP remains unexplored.

Endogeneity Problem



Can we conclude that IP reduces cumulative innovation?

OR

Positive shock in the value of the underlying technology?



Galasso and Schankerman (2012) work in progress.

IDENTIFICATION STRATEGY

*Judges are assigned to patent cases through **a computer program that randomly generates three-judge panels**, subject to the judges' availability and the requirement that each judge deals with a representative cross-section of the fields of law within the jurisdiction of the court (Fed. Cir. R. 47.2).*

We exploit the random allocation of patents to CAFC judges in validity cases and construct an index capturing the propensity of the three judge panel to vote in favour of patent invalidity as IV

(Very Preliminary) Findings

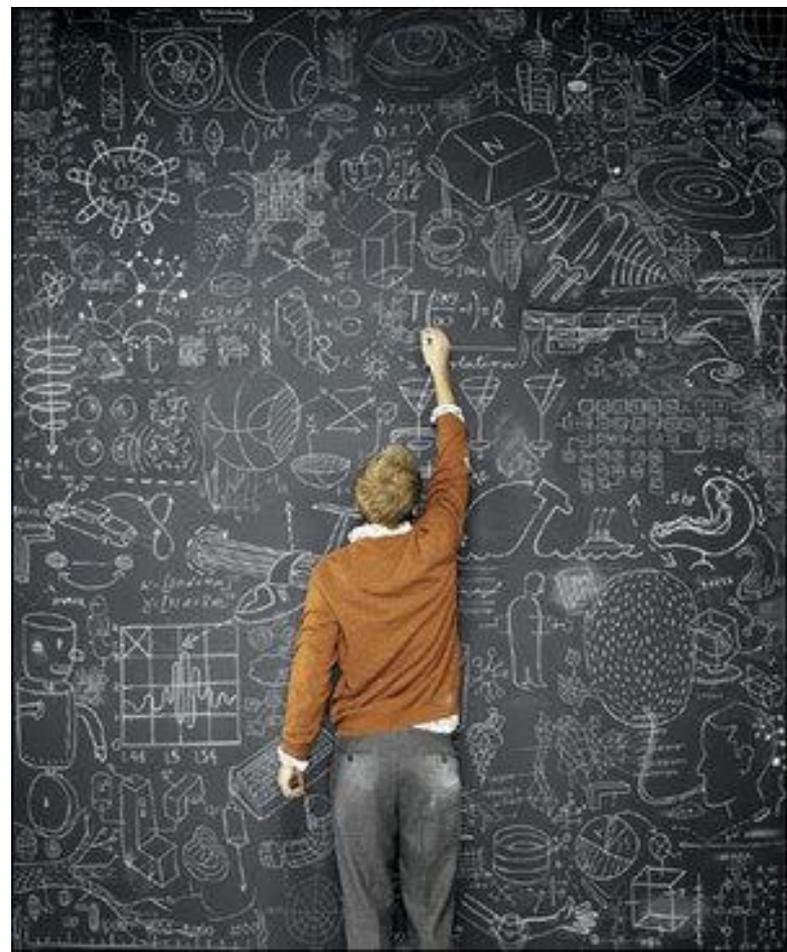
We find that patent invalidation is on average associated with roughly 50 % increase in citations received).

The marginal treatment effect of patent invalidation is highly heterogeneous:

- Across tech areas (effect very large for medical instruments but not statistically significant from zero in electronics);
- Within tech areas (invalidating a patent of a large firm has greater impact than small firm especially if lots of small firms operate in tech area).

Some lessons I've learned

1. Good IVs hide behind institutional details;
2. Good instrumental variables have to be complemented with good theory;
3. Bad instrumental variables have to be substituted with good theory;



Bandiera, Guiso, Prat and Sadun (2012) “*Matching firms, managers and incentives*”

How do firms and managers generate surplus by matching appropriately?

Exogenous variation very difficult to get...

Theory: managers differ in talent and risk aversion, firms differ in their cost structure and private benefits from control.

Their model offers detailed predictions on:

- contract offered by firms (high power incentives less likely when control benefits are high);
- type of contract accepted by managers (high talent choose high incentives);
- effort of managers, satisfaction and wages (higher with steep contract);
- profits of firms (higher with high power incentives).

Data strongly support their theory.

Alternative theories may be consistent with a subset of the correlations they report but not entire set.

Isn't this more convincing than bad IV? Or more instructive than good IV but unclear mechanism??

***IT JUST
MAKES
SENSE***

Thank you!

Detecting Cause-and-Effect through Triangulation: Some Tips and Examples

Rosemarie Ziedonis

University of Oregon
AOM PDW August 2012

Step 1: Reframe the issue

“Identification PROBLEM”

“Endogeneity PROBLEM”

“Selection PROBLEM”



Be a Social Science Detective!



- I have a theory
- I see clues
- Could something other than my theory explain those clues?

How?



- ✓+ Find a natural experiment, valid instrument, etc...
- ✓- Over-assert
 - My theory is A, the clues confirm it!
 - Add a quiet "P.S." that the clues are consistent with Theories B and C as well
 - Hang your hat on flimsy evidence (2SLS w/o exclusion; IV with no logic)
- ✓ Refuse the job
- ✓ Triangulate!

Triangulation



“In the social sciences, triangulation is often used to indicate use of more than two methods in a study with a view to double (or triple) checking the results.”

Also known as “cross-examination”

--wikipedia

Two Examples

MANAGEMENT SCIENCE
Vol. 50, No. 6, June 2004, pp. 804–820
ISSN 0025-1909 | eISSN 1526-5501 | 04 | 5006 | 0804

informs.
doi 10.1287/mnsc.1040.0208
© 2004 INFORMS

Don't Fence Me In: Fragmented Markets for
Technology and the Patent Acquisition
Strategies of Firms

Rosemarie Ham Ziedonis
University of Michigan Business School, 701 Tappan, Ann Arbor, Michigan, rzied@umich.edu

Journal of Financial Economics 98 (2010) 478–499

Contents lists available at ScienceDirect

Journal of Financial Economics
journal homepage: www.elsevier.com/locate/jfec

Corporate venture capital and the returns to acquiring
portfolio companies^{a, c}

David Benson^a, Rosemarie H. Ziedonis^{b,*}

^a Marriott School of Management, Brigham Young University, Provo, UT 84602, USA
^b Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA

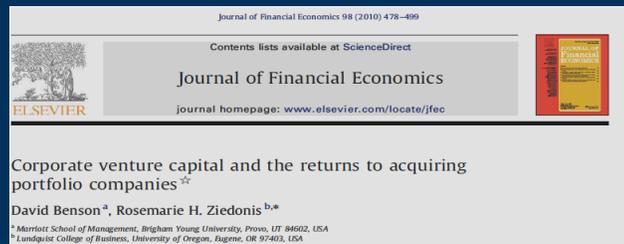
Example 1: “Don’t Fence Me In”

- The “Crime”: a surprising surge in patenting by IT companies (Y)
- Theory: a strategic response to market frictions and concerns of hold-up
 - X (external rights ↑) → W (friction ↑) → Y (Self-insure ↑↑)
- Clues: X, Y but not W
- Alt Theory: X and Y are caused by Z (technological opp), not W

Approach

- Dig deeper!
 - If my theory is correct, what else should I find?
 - Effects = amplified post “regime shift”
 - Effects for Type A firms > than for Type B firms
- Revisit the phenomenon
 - Did a technological shift **coincide** with the Regime Shift? Or **disproportionately** affect Type A v. B?
 - Read books & talk with people!
- Find an indirect proxy (trends in same sector in countries w/o “regime shift”)

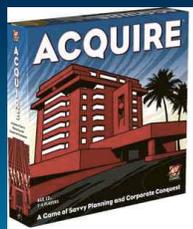
Example Two



An Anomaly Paper



- Expect Theory A
- Clues AT ODDS with Theory A
- Investigate Theories B, C, and D instead
 - Read books, go talk to people!
- Shift settings and re-test



Summary



- It's cool to be a social science detective!
- To be a good one, you must...
 - Do some digging
 - Be creative
 - Be honest
 - Sharpen your tools...and invest in new ones



Some References

Less Obvious

1. Deaton A. 2009. Instruments, Randomization, and Learning about Development. *Jl of Economic Literature*, 48(2): 424-455.
2. Diamond J, Robinson JA, ed. 2010 Natural Experiments of History. Harvard Univ Press.

More Obvious

3. Angrist JD, Pischke JS. 2008. Mostly Harmless Econometrics. Princeton Univ Press
4. Hamilton B, Nickerson J 2003. Correcting for Endogeneity in Strategic Management Research. *Strategic Organization*, 1, 51-78.
5. Shaver JM. 1998. Accounting for Endogeneity When Assessing Strategy Performance: Does Entry Mode Choice Affect FDI Survival? *Management Science*. 44(4): 571-585.
6. Bascle G. 2008. Controlling for Endogeneity with Instrumental Variables in Strategic Management Research. *Strategic Organization* 6(3): 285-327.
7. Reeb D, Sakakibara M, Mahmood IP. 2012. From the Editors: Endogeneity in International Business Research, *Jl International Business Studies* 43: 211-218.

My Examples

8. Ziedonis RH. 2004. Don't Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms. *Management Science* 50(6): 804-820.
9. Benson D, Ziedonis RH 2010. Corporate Venture Capital and the Returns to Acquiring Portfolio Companies. *Jl of Financial Economics*, 98: 478-499.