

Causal Inferences in Capital Markets Research

Other titles in Foundations and Trends® in Accounting

Corporate Governance Research on Listed Firms in China: Institutions, Governance and Accountability

T. J. Wong

ISBN: 978-1-68083-098-9

Alphanomics: The Informational Underpinnings of Market Efficiency

Charles M. C. Lee and Eric C. So

ISBN: 978-1-60198-892-8

International Transfer Pricing

Richard Sansing

ISBN: 978-1-60198-894-2

Fair Value Measurement in Financial Reporting

Leslie Hodder, Patrick Hopkins, and Katherine Schipper

ISBN: 978-1-60198-886-7

The Role of Management Controls in Transforming Firm Boundaries and Sustaining Hybrid Organizational Forms

Shannon W. Anderson and Henri C. Dekker

ISBN: 978-1-60198-868-3

Corporate Governance, Board Oversight, and CEO Turnover

Volker Laux

ISBN: 978-1-60198-866-9

Causal Inferences in Capital Markets Research

Edited by

Iván Marinovic

Stanford Graduate School of Business

USA

imvial@stanford.edu

now

the essence of knowledge

Boston — Delft

Foundations and Trends® in Accounting

Published, sold and distributed by:

now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:

now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The content of the book was originally published in Foundations and Trends® in Accounting, vol. 10, nos. 2-4.

ISBN: 978-1-68083-160-3
© 2016 Now Publishers Inc

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Foundations and Trends[®] in Accounting

Volume 10, Issue 2-4, 2016

Editorial Board

Editor-in-Chief

Stefan J. Reichelstein

Stanford University

United States

Associate Editors

Ronald Dye

Northwestern University

David Larcker

Stanford University

Stephen Penman

Columbia University

Editorial Scope

Topics

Foundations and Trends[®] in Accounting publishes survey and tutorial articles in the following topics:

- Auditing
- Corporate Governance
- Cost Management
- Disclosure
- Event Studies/Market Efficiency Studies
- Executive Compensation
- Financial Reporting
- Management Control
- Performance Measurement
- Taxation

Information for Librarians

Foundations and Trends[®] in Accounting, 2016, Volume 10, 4 issues. ISSN paper version 1554-0642. ISSN online version 1554-0650. Also available as a combined paper and online subscription.

Contents

Causal Inferences in Capital Markets Research	1
<i>Iván Marinovic</i>	
Where’s the Rigor When You Need It?	4
<i>Nancy Cartwright</i>	
Mostly Useless Econometrics? Assessing the Causal Effect of Econometric Theory	23
<i>John Rust</i>	
Just How Sensitive are Instrumental Variable Estimates?	102
<i>Peter C. Reiss</i>	
Interpreting Point Predictions: Some Logical Issues	136
<i>Charles F. Manski</i>	
From Casual to Causal Inference in Accounting Research: The Need for Theoretical Foundations	160
<i>Jeremy Bertomeu, Anne Beyer, and Daniel J. Taylor</i>	

Comments and Observations Regarding the Relation Between Theory and Empirical Research in Contemporary Accounting Research	212
<i>Qi Chen and Katherine Schipper</i>	
Identification with Models and Exogenous Data Variation	259
<i>R. Jay Kahn and Toni M. Whited</i>	
Plausibility: A Fair & Balanced View of 30 Years of Progress in Ecologics	274
<i>Ivo Welch</i>	

Preface

In December 2014 the Stanford Graduate School of Business hosted the conference “Causality in the Social Sciences” in an attempt to promote a broad interdisciplinary debate about the notion of causality, and the role of causal inference in the social sciences.

The issue of causality is at the center of all scientific disciplines, but has become particularly contentious in accounting research. Despite usual disclaimers, there often is a gap between the causal language researchers use to describe empirical findings, and the extent to which causal claims are backed by evidence. At the risk of oversimplifying, the issue of causality divides the accounting research community in two polar views: 1) the view that causality is an unattainable ideal for the social sciences and must be given up as a standard, and 2) the view that, on one hand, causality should be the ultimate goal of all scientific endeavors and, on the other hand, theory and causal inference are inextricable. Reflecting and discussing about these views was the main motivation for organizing the Stanford’s causality conference.

The conference gathered some of the most distinguished scientists across five disciplines: philosophy, economics, accounting, finance, and marketing. For two days, the conference’s participants were able to reflect and discuss about the best methods for causal inference of social phenomena.

This volume summarizes the conclusions of the conference and is organized in three sections: I) Econometrics; II) Accounting, and III) Finance.

First, Nancy Cartwright addresses the problem of external validity and the reliability of scientific claims that generalize individual cases. Then, John Rust discusses the role of assumptions in empirical research and the possibility of assumption-free inference. Peter Reiss considers the question of how sensitive are instrumental variables to functional form transformations. Charles Manski studies the logical issues that affect the interpretation of point predictions, questioning prediction practices that use a single combined prediction such as the consensus forecast to summarize the beliefs of multiple forecasters.

Second, Bertomeu, Beyer and Taylor provide a critical overview of empirical accounting research, focusing on the benefits of theory-based estimation, while Chen and Schipper consider the question whether all research should be causal, and assess the existing gap between theory and empirical research in accounting.

Third, R. Jay Kahn and Toni Whited clarify and contrasts the notions of identification and causality, whereas Ivo Welch adopts a sociology of science approach to understand the consequences of the researchers' race for discovering novel and surprising results.

We hope this volume will allow researchers and Ph.D students in accounting (and the social sciences in general) to acquire a deeper understanding of the notion of causality and the nature, limits, and scope of empirical research in the social sciences.

Iván Marinovic
Stanford Graduate School of Business

Where's the Rigor When You Need It?

Nancy Cartwright

University of California, San Diego, USA and Durham University, UK;
ncartwright@ucsd.edu

ABSTRACT

When it comes to causal conclusions, rigor matters. To this end we impose high standards for how studies from which we draw causal conclusions are conducted. For instance, we are widely urged to prefer randomized controlled trials (RCTs) or instrumental variable (IV) models to observational studies relying just on correlations, and we have explicit criteria for what counts as a good RCT or a good IV model. But we tend to be shockingly sloppy when it comes to making explicit just what the causal conclusions we draw mean, why the methods we employ are good for establishing conclusions with just that meaning, and what can defensibly be taken to follow from these claims. With respect to what can be inferred from the limited causal conclusions our studies support, we are far too prone to over reach, to 'generalize' that what holds in a study or handful of studies holds widely. But, I shall argue, we do not get arrant for general claims by generalizing. Rather it takes a great tangle of scientific work to support a general claim, including a great deal of conceptual development, theory and the confirmation of a variety of different kinds of effects that the general claim implies.

1

Introduction

I was charged with the question: When it comes to causality, ‘What do social scientists know?’ –and, presumably, ‘How do they know it?’ In expanding on that, Iván (Marinovic) in his invitation added: “Often the issue of causality and identification is ignored or ‘resolved’ by adding explanatory variables (given the large amount of data available). If you have some specific thoughts about this type of research I would really appreciate that you discuss them.” There are issues here that need to be talked about. Many are technical, and you have here real experts to discuss them. But there are other issues that are not technical, that really matter, that do not get discussed, and we repeatedly get into trouble because we do not pay attention to them. To unearth some of these issues, there are two points I want to urge today:

1. Do not do inference by pun.
2. In general, well substantiated, reliable general claims in science \neq generalizations, i.e. claims warranted by generalizing from individual instances.

The first must be totally uncontroversial. The second certainly should be.

References

- Broadbent, A. 2014. “Risk relativism and physical law”. *Journal of Epidemiology and Community Health*. Published Online First: 14 August 2014.
- Fuller, J. 2015. *Myths and Fallacies of Simple Extrapolation in Medicine*. Ms.
- Imbens, G. 2014. “Instrumental variables: An econometrician’s perspective”. *Statistical Science*. 29(3): 323–358.
- Pearl, J. and E. Bareinboim. 2011a. “External validity and transportability: A formal approach”. In: *JSM Proceedings, Miami Beach FL*. July 30 – August 4. 157–171.
- Pearl, J. and E. Bareinboim. 2011b. “Transportability of causal and statistical relations: A formal approach”. In: *Proceedings of the 25th AAAI Conference on Artificial Intelligence*. Ed. by C. AAAI Press Menlo Park. August 7–11, San Francisco. 247–254.
- Petticrew, M. and I. Chalmers. 2011. “Use of research evidence in practice”. *The Lancet*. 378(9804): 1696.
- Post, N., H. de Beer, and G. H. Guyatt. 2013. “How to generalize efficacy results of randomized trials: Recommendations based on a systematic review of possible approaches”. *Journal of Evaluation in Clinical Practice*. 19(4): 638–643.

Mostly Useless Econometrics? Assessing the Causal Effect of Econometric Theory

John Rust

*Department of Economics, Georgetown University, USA;
jr1393@georgetown.edu*

ABSTRACT

Economics is highly invested in sophisticated mathematics and *empirical methodologies*. Yet the payoff to these investments in terms of uncontroverted *empirical knowledge* is much less clear. I argue that leading economics journals err by imposing an unrealistic burden of proof on empirical work: there is an obsession with establishing *causal relationships* that must be proven beyond the shadow of a doubt. It is far easier to publish theoretical econometrics, an increasingly arid subject that meets the burden of mathematical proof. But the overabundance of econometric theory has not paid off in terms of empirical knowledge, and may paradoxically hinder empirical work by obligating empirical researchers to employ the latest methods that are often difficult to understand and use and fail to address the problems that researchers actually confront. I argue that a change in the professional culture and incentives can help econometrics from losing its empirical relevance. Econometric theory needs to be more empirically motivated and problem-driven. Economics journals should lower the burden of proof for empirical work and raise the burden of proof for econometric theory. Specifically, there should be more room for descriptive empirical work in our journals. It should not be necessary to establish a causal mechanism or a non-parametrically identified structural model that provides an unambiguous explanation of empirical phenomena as a litmus test for publication. On the other hand, journals should increase the

burden on econometric theory by requiring more of them to show how the new methods they propose are likely to be *used* and be *useful* for generating new empirical knowledge.

1

Introduction

“As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired from ideas coming from ‘reality’, it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely l’art pour l’art. This need not be bad, if the field is surrounded by correlated subjects, which still have closer empirical connections, or if the discipline is under the influence of men with an exceptionally well-developed taste. But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant branches, and that the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical source, or after much ‘abstract’ inbreeding, a mathematical subject is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up. It would be easy to give examples, to trace specific evolutions into the baroque and the very high baroque, but this, again, would be too technical. In any event, whenever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the re-injection of more or less directly empirical ideas.”

von Neumann (1947)

This essay is a warning that von Neumann's danger signal is up for econometric theory, and his suggested remedy, "the re-injection of more or less directly empirical ideas", is overdue. In my opinion, econometric theory has run into seriously diminishing returns. It is increasingly abstract, technical, and difficult to understand: "baroque" is a good adjective for some of it. I am not anti-theory and realize that mathematical subjects are not easy. But too many econometric theory articles are poorly motivated, and the value of trying to implement the new estimators they propose is unclear.

Even though I teach econometrics to graduate students, I would have to admit that the subject can easily come across as a "disorganized mass of details and complexities" that does not really prepare them go out and discover new empirical knowledge. Instead I fear many of them regard it as an obstacle to be overcome: a vast catalog of *limit theorems* that archive the endless ways to use and reuse the Law of Large Numbers and Central Limit Theorem to prove even more arcane limit theorems that only tenuously pretend to have anything to do with the real challenges facing empirical researchers.

Though the title of this essay was motivated by *Mostly Harmless Econometrics* by Angrist and Pischke (2009), this is not a critique of their book. On the contrary, I would have titled their book "Mostly Useful Econometrics" because Angrist and Pischke are part of a group of (mostly labor-oriented) applied econometricians who *practice what they preach*. The enormous popularity of this book is due in part to the fact it is *not* written in the abstract style of econometric theory texts. Instead it provides compelling *empirical motivation* for a set of relatively easy to apply econometric methods that are actually used by applied researchers. As such, it is an essential part of the "tool kit" for doing good empirical work in economics.

On the other hand, our journals and many econometric textbooks devote too much space to estimation methods that very few of us actually use. Does anyone really use k -class estimators, or 3 stage least squares, or maximum score? Do we really need to know how to test for unit roots or cointegration in time series? Do we need to know how to non-parametrically estimate simultaneous equation systems with

non-separable errors or know about “identification at infinity” in order to do good empirical work?

Consider a typical econometric theory article in a typical issue of *Econometrica*. Rarely is there any specific empirical motivation as to why yet another estimation method is necessary, or a discussion of a *suggested* empirical application, much less an *actual* empirical application and a demonstration of how the new method changes a conclusion about a substantive empirical issue. Instead all too many are motivated by *previous econometric theory papers* and the main contribution is to generalize a previous model or proof (e.g. allow for non-additive errors instead of additive ones or re-prove a result using weaker assumptions).

Though I will avoid naming names, many econometric theorists are more like pure statisticians or mathematicians who do little empirical work themselves. Some of them have no apparent interest in economics: they call themselves “econometricians” mainly because salaries in economics are much higher than in statistics departments which have long been in decline and in some cases eliminated (such as at Princeton), or nearly eliminated (such as at Yale). Yet the celibate priesthood of econometric theorists yield great power over empirical researchers: it is difficult to publish empirical work in leading economics journals unless it is blessed by the high priests. So empirical papers published in *Econometrica* tend to be *illustrations of the latest methods* (justifying a demand for even more methods), rather than work that is focused on important economic problems or issues.

In comparison, the “hard sciences” such as physics or biology are less methodologically focused than economics, yet they appear to be much more productive of useful empirical knowledge. One only needs to look in the daily newspaper to see amazing new medical advances, huge advances in communications and computer technologies, and fundamental discoveries about the universe at both the smallest and largest scales. The hard sciences make progress because most of the research is *empirically motivated*: what causes cancer? why are atmospheric CO_2 levels rising? what does cosmic background radiation tell us about Big Bang asymmetries that kept matter from being completely annihilated by anti-matter?

Rather than endlessly debate *how to do science* people in the hard sciences just *do science*. They are much more focused on *data generation* particularly through the creation of sophisticated instruments and well designed experiments, and data gathering is often far more focused and theory-driven than in economics. For example, one of the last remaining fundamental particles predicted by the standard model of physics, the Higgs Boson (which in conjunction with the Higgs field gives particles their mass), was finally discovered in 2013 using the large hadron collider at CERN. Certainly a great deal of statistical sophistication is required to analyze the terabytes of data from billions of subatomic collisions to filter signal from noise, or to infer the presence of dark matter from doppler shifts in light from distant galaxies, or to infer small asymmetries 10^{-20} seconds after the big bang from cosmic background radiation. Even so, the average physicist is less statistically sophisticated and tooled-up than the average economist.

Economists, on the other hand, try to create an illusion of great scientific prowess by using advanced mathematics to study abstruse topics such as large square economies, unit roots, partial identification, moment inequalities and triangular models. Even topics that seem to have a useful goal, such as implementation of social choice rules, are analyzed at such a high level of abstraction that it is hard to see their practical value. Economists lionize ultra-mathematical theories, even if they have no clear real-world applicability or provide poor approximations to reality. This is certainly true of a lot of economic theory that assumes that people are rational expected utility maximizers, that firms maximize expected profits, that interactions between individuals and firms always occur in a state of Nash or competitive equilibrium, and that financial markets are complete and informationally efficient. Our math obsession has deluded us into thinking that these are good approximations to reality when there is lots of evidence that they aren't.¹

¹My critique differs from the “mathiness” critique of Romer (2015) who argues that some economists try to use mathematics to masquerade politically motivated viewpoints. I do not believe econometric theorists are “political” in this sense, and the mathematics they do is of very high quality. To the extent there is a masquerade, it is to foster the impression that all econometric theory must be useful because it ultimately enables economists to do better empirical work. Instead, my critique is

There is still time to change the orientation of theoretical econometrics to avoid the fate of pure economic theory, which has also suffered the consequences of being mostly useless. Economic theory had immense prestige in 1980s when I received my PhD, an era where empirical researchers (and even applied theorists) were viewed as distinctly second class citizens. However theoretical elitism turned out to be an unsustainable equilibrium, as the profession ignored von Neumann's danger signal and allowed economic theory to become increasingly abstract, baroque, and disconnected from reality. Hamermesh (2013) documents a significant decline in the amount of economic theory published by the top economics journals since the 1960s and speculates that "Economic theory may have become so abstruse that editors of the leading general journals, recognizing that very few of their readers could comprehend the theory, have cut back on publishing work of this type." (p. 169).

The problem of uselessness of economic theory became sufficiently severe to come to the attention of the popular press in a *New Yorker* article in 1996 titled "The Decline of Economics." It was motivated by the 1996 Nobel Prizes to James Mirrlees and William Vickrey. Vickrey had mentioned to a reporter at the *New York Times* that his famous paper on auctions (Vickrey, 1961) was "one of my digressions into abstract economics" and "At best, it's of minor significance in terms of human welfare." In response, Cassidy (1996) lamented that "Here is a world-renowned theorist confirming what many outsiders had long suspected — that a good deal of economic theory, even the kind that wins Nobel Prizes, simply doesn't matter much. That is a great pity, since economics is supposed to be a useful subject, and its intellectual founders stressed its practical importance." (p. 50).

Economic theory would be in a much better state today had profession followed the advice of Alvin Roth, who issued his own version of von Neumann's danger signal back in 1991: "if we do not take steps in the direction of adding a solid empirical base to game theory, but instead continue to rely on game theory primarily for conceptual insights

similar in many respects to the critique of McCloskey (2005) who concludes that the profession's obsession with "mathematical and statistical reasoning" "is a waste of time" and unless this changes "our understanding of the economic world will continue to stagnate."

(deep and satisfying as these may be), then it is likely that long before a hundred years game theory will have experienced sharply diminishing returns. In this respect, I think the next hundred years will likely bring about a change in the way theoretical and empirical work are related in economics generally, and that, if not, then the entire discipline of economics may also fail to realize its potential.” (Roth, 1991a, p. 108). One only needs to look at current job statistics on *Econ.JobMarket.org* to see that Roth underestimated how rapidly economic theory would collapse: both the number of positions and the number of candidates who list economic theory as their primary field has dwindled to less than one fourth the corresponding numbers for econometrics.

Theoretical econometrics has a superficial appearance of usefulness since it is a field that supposed to provide us with the methods and tools for doing empirical work, but it is infected by the same theoretical elitism and detachment from reality that lead to the decline of economic theory. The current professional culture still offers far greater rewards for doing econometric theory and publishing new estimation methods than for doing applied work, especially when we recognize that data gathering and analysis is a far more laborious and less glamorous task than proving theorems. Similar to economic theory, econometric theory has become sort of an elite sport that can be played at elite upper tier departments that are rich enough to afford it (e.g. Harvard, Princeton, Yale, MIT, Stanford, Berkeley, Chicago, etc.). Among the trophies in this sport is the ability to publish in the elite upper tier journals, such as *Econometrica* which is devoted to publishing the most arid types of economic and econometric theory.

Economists should be more concerned about our collective influence and impact. One metric is citations, and compared to other sciences, economists have few citations. For example the Thomson Reuters Essential Science Indicators database from 2000 to 2010 ranks economics 17th out of its list of 21 sciences in terms of citations. Only engineering, other social sciences, computer science and mathematics had lower average citations over this period than economics. The science with the most citations according to the ESI index was molecular biology, followed by immunology. Within economics, the least cited subfields are economic

theory and (micro) econometric theory according to Ellison (2013). He finds that “The fields in which papers are estimated to have the fewest citations — political economy, history, micro theory, cross-section econometrics, and industrial organization — are all fields where we estimated that researchers have relatively low citation indexes.” (p. 82).

The leading journals such as *Econometrica* have great influence on the type of work done in economics due to the hierarchical way the profession is self-organized. There is a hierarchy in the profession that is akin to a bee hive where a small number of theorists are the queen bees who have the influence to set the overall direction of research via their roles as editors at the top ranked journals. The sustainability of the bee hive depends on having cadres of worker bees who are willing to follow the directions they set in exchange for the chance to publish, get tenure, and try to make it slightly further up in the hierarchy. As von Neumann noted, this sort of academic hierarchy can be sustainable “if the field is surrounded by correlated subjects, which still have closer empirical connections, or if the discipline is under the influence of men with an exceptionally well-developed taste.” Unfortunately too few leading theorists have a real interest in the real world application of their theories: this is a task left for the worker bees. The minority of theorists who do attempt to pay homage to reality often suffer from the hubris of confusing knowledge of their theories with knowledge of the real world.

For example, in his 2003 Presidential Address to the American Economic Association, Robert E. Lucas (2003) proclaimed that “central problem of depression prevention has been solved, for all practical purposes, and has in fact been solved for many decades.” (p. 1). Oops! The 2008 financial crash and ensuing “Great Recession” revealed that economists actually knew far less about the world than they thought they knew. It suddenly became painfully clear that a large part of the profession was virtually clueless what was actually going on in the economy because few of the ugly real world complexities are captured in our oversimplified mathematical models.

Indeed, back in 2007 few economists knew what a “swap” was or how dangerously overlevered and interdependent most of Wall Street firms

were. This ignorance is not surprising: most real business cycle models assume away the financial sector as irrelevant. However leading experts in finance, such as Eugene Fama, are so wedded to idealized views such as the efficient markets hypothesis that they deny the possibility that a collapse of a credit/housing bubble could have led to the 2008 Wall Street crash: “I don’t know what a credit bubble means. I don’t even know what a bubble means. These words have become popular. I don’t think they have any meaning.” (quoted from interview in Cassidy, 2010). In short, too many of us can be accused of being “egg heads” who are not in touch with the real world.

Of course some economists are/were in touch with reality, and several issued prescient warnings of stock market and housing bubbles and an impending financial meltdown well before the 2008 crash. However most of these economists, such as Robert Shiller or Noriel Roubini, were ignored or treated as “flakes” or “doctors of doom” — the “Chicken Littles” who always predict that the sky is falling. I believe Shiller and Roubini would confirm that it is hard to publish or be heard if you dare speak out against the conventional wisdom and orthodoxy in economics. While there is a nascent literature on behavioral economics that does try to learn and understand what people and firms actually do (as opposed to rationalizing behavior to force it to conform to prediction of existing theories) it is hard to publish empirical work that provides evidence of behavior that is inconsistent with orthodox theory unless it is accompanied by an elaborate non-expected utility theory that “explains” this behavior. Leading journals such as *Econometrica* are much more likely to publish *behavioral theories* but much less likely to publish *behavioral evidence*.

This is part of the burden of proof on empirical work that tends to keep economists focused on theory, but less aware of reality. The profession ought to be more receptive to self-evaluation, given that our sense of smugness and self-confidence has been shaken by dramatic embarrassments such as our failure to predict the crash of 2008 and the Great Recession, or to even agree on the best policies to deal with it after the fact. There is a surprising level of ignorance, or at least a glaring lack of professional agreement, on a host of other important questions

as well. For example there is huge disagreement about how strongly taxes affect labor supply. Despite intensive study for over three decades, there is still disagreement about the efficacy of job training programs on various outcomes such as unemployment duration and subsequent earnings and about how best to deal with crumbling institutions such as Social Security and Medicare. The profession does not agree about whether financial markets are inherently unstable and whether they need to be regulated, and if so, how. I believe that economists need to be considerably more humble and admit that there is very little that we can confidently say that we know and agree upon. Given this, perhaps it is time to have a discussion about whether this state of affairs is inherent in the subject (i.e. economics is a more difficult topic to study than physics or biology) or whether the profession's bias for deductive versus inductive inference is partly to blame for our lack of agreement and results.

A full discussion is beyond the scope of this essay because it deals with the interaction between econometric theory, economic theory, and empirical work. I have discussed the tensions between *economic theory* and empirical work elsewhere (Rust, 2014) and want to stress that I am not anti-theory, though I feel that the profession refuses to let go of favorite theories despite considerable evidence that they are at odds with reality. My focus in this essay is on *econometric theory*. I also stress I am not against econometric theory and certainly do not claim that all econometric theory is useless. However economics is “data poor” relative to other sciences, but not because economic data are inherently harder to collect. Rather, due to the professional culture and a methodological bias at the top departments, the rewards for data gathering and data analysis are low, whereas the payoff to econometric theory is much higher. This equilibrium is good for the elite departments, but bad for the profession as a whole.

The rest of this essay is organized as follows. Section 2 documents the theory bias at *Econometrica* and their efforts to correct this bias to stay relevant. Though I argue that the causal impact of econometric theory is small, the amount of econometric theory devoted to causality is huge — particularly the burgeoning literature on *treatment effects*. I

discuss this literature in Section 3, but argue that it has not resulted in the sort of *credibility revolution* in applied economics claimed by Angrist and Pischke (2010). In Section 4 I discuss a personal example that illustrates how even an incomplete understanding of causality can be *extremely useful*. Since my example is from medicine, I also discuss an economic example. In Section 5 I review a vast literature on the causal effect of training programs and conclude that the useful knowledge from decades of study by the leading econometricians has been disappointingly meager. To avoid writing a completely hopeless and depressing essay, I devote Section 6 to discussing several success stories where economics has produced useful knowledge. Unfortunately, these are not examples where econometric theory played much of a role. Section 7 concludes with some ideas about how things can be turned around to make econometric theory more useful to economics.

Although this essay questions the uselessness of some ultra mathematical economic and econometric theory, I do not pretend to judge it from an superior vantage point. I readily admit that my own academic work has proved to be mostly useless. My views will no doubt cost me professionally, so why write this? While there are many prizes and ways economists self-congratulate themselves, there are fewer avenues for self-evaluation and few economists willing to take the professional risk to publicly voice their true concerns about the shortcomings of the discipline. I hope this essay will lead to a constructive discussion rather than be seen as an unhelpful rant on econometric theory and the state of the economics profession.

References

- Anderson, R., I. Ashlagi, D. Gamarnik, and A. E. Roth. 2015. “Finding long chains in kidney exchange using the traveling salesman problem”. *Proceedings of the National Academy of Sciences*. 112(3): 663–668.
- Angrist, J. and J.-S. Pischke. 2009. *Mostly Harmless Econometrics An Empiricist’s Companion*. Princeton, Princeton University Press.
- Angrist, J. D. and J.-S. Pischke. 2010. “The credibility revolution in empirical economics: How better research design is taking the con out of econometrics”. *Journal of Economic Perspectives*. 24(2): 3–30.
- Ashenfelter, O. 1987. “The case for evaluating training programs with randomized trials”. *Economics of Education Review*. 6(4): 333–338.
- Basu, K. 2014. “Randomisation, causality and the role of reasoned intuition”. *Oxford Development Studies*. 42-4.
- Bontemps, C., T. Magnac, and E. Maurin. 2012. “Set identified linear models”. *Econometrica*. 80(3): 1129–1155.
- Card, D. and S. DellaVigna. 2013. “Nine facts about top journals in economics”. *Journal of Economic Literature*. 51(1): 144–161.
- Card, D., J. Kluge, and A. Weber. 2010. “Active labour market policy evaluations: A meta-analysis”. *Economic Journal*. 120(548): F452–F457.

- Carrell, S. E., B. I. Sacerdote, and J. E. West. 2013. "From natural variation to optimal policy? The importance of endogenous Peer group formation". *Econometrica*. 81(3): 855–882.
- Cassidy, J. 1996. "The decline of economics". *New Yorker*. December(2).
- Cassidy, J. 2010. "Interview with Eugene Fama". *New Yorker*. January(13).
- Chan, T. Y. and B. H. Hamilton. 2006. "Learning, private information, and the economic evaluation of randomized experiments". *Journal of Political Economy*. 114(6): 997–1040.
- Crépon, B., E. Dufflo, M. Gurgand, R. Rathelot, and P. Zamora. 2013. "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment". *Quarterly Journal of Economics*. 128(2): 531–580.
- Cunha, F., J. J. Heckman, and S. M. Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation". *Econometrica*. 78(3): 883–931.
- Deaton, A. 2010. "Instruments, randomization, and learning about development". *Journal of Economic Literature*. 48: 424–455.
- Donohue, J. J. I. and S. D. Levitt. 2001. "The impact of legalized abortion on crime". *Quarterly Journal of Economics*. 116(2): 379–420.
- Ellison, G. 2013. "How does the market use citation data? The Hirsch index in economics". *American Economic Journal: Applied Economics*. 5: 63–90.
- Foote, C. L. and C. F. Goetz. 2006. "The impact of legalized abortion on crime: Comment". *Quarterly Journal of Economics*. 123(1): 407–423.
- Fourcade, M., E. Ollion, and Y. Algan. 2015. "The superiority of economists". *Journal of Economic Perspectives*. 29(1): 89–114.
- El-Gamal, M. A. and D. M. Grether. 1995. "Are people Bayesian? Uncovering behavioral strategies". *Journal of the American Statistical Association*. 90(432): 1137–1145.
- Gottlieb, D. J. 2014. "Sleep apnea and the risk of atrial fibrillation recurrence: Structural or functional effects?" *Journal of the American Heart Association*. 113(654): 1–3.

- Griffiths, T. L. and J. B. Tenenbaum. 2009. "Theory-based causal induction". *Psychological Review*. 116(4): 661–716.
- Hamermesh, D. S. 2013. "Six decades of top economics publishing: Who and how?" *Journal of Economic Literature*. 51(1): 162–172.
- Hayashi, F. and E. C. Prescott. 2002. "The 1990s in Japan: A lost decade". *Review of Economic Dynamics*. 5: 206–235.
- Heckman, J. J. 1979. "Sample selection bias as specification error". *Econometrica*. 47(1): 153–161.
- Heckman, J. J. 1992a. "Haavelmo and the birth of modern econometrics: A review of the history of econometric ideas by Mary Morgan". *Journal of Economic Literature*. 30(2): 876–886.
- Heckman, J. J. 1992b. "Randomization and social policy evaluation". In: *Evaluating Welfare and Training Programs*. Ed. by C. F. Manski and I. Garfinkel. Harvard University Press. 201–230.
- Heckman, J. J. 2005. "The scientific model of causality". *Sociological Methodology*. 35: 1–97.
- Heckman, J. J. and V. J. Hotz. 1989. "Choosing among alternative non-experimental methods for estimating the impact of social programs: The case of manpower training". *Journal of the American Statistical Association*. 84(408): 862–874.
- Heckman, J. J., V. J. Hotz, and M. Dabos. 1987. "Do we need experimental data to evaluate the impact of manpower training on earnings?" *Evaluation Review*. 11: 395–427.
- Heckman, J. J., R. J. LaLonde, and J. A. Smith. 1999. "The economics and econometrics of active labor market programs". In: *Handbook of Labor Economics, Volume 3*. Ed. by O. Ashenfelter and D. Card. New York: Elsevier. 1866–2097.
- Heckman, J. J. and R. Pinto. 2014. "Causal inference after Haavelmo". *Econometric Theory*. 31(1): 115–151.
- Heckman, J. J. and J. Smith. 1998. "Evaluating the welfare state". In: *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. Ed. by S. Strom. Cambridge University Press.

- Herndon, T., M. Ash, and R. Pollin. 2014. "Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff". *Cambridge Journal of Economics*. 38(2): 257–279.
- Hoxby, C. and S. Turner. 2013. "Expanding college opportunities for high-achieving, low income students". *SIEPR Discussion Paper*. 12(014).
- Kapoor, S. and A. Magesan. 2014. "Paging inspector Sands: The costs of public information". *American Economic Journal: Economic Policy*. 6(1): 92–113.
- Kristensen, D. and B. Schjerning. 2015. "Implementation and estimation of discrete Markov decision models by Sieve approximations". *manuscript, University of Copenhagen*.
- Labor, U. D. of. 2014. *What Works in Job Training: A Synthesis of the Evidence*. U.S. Department of Labor.
- LaLonde, R. J. 1986. "Evaluating the econometric evaluations of training programs with experimental data". *American Economic Review*. 76(4): 604–620.
- Leamer, E. E. 1982. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley & Sons.
- Levitt, S. D. 1997. "Using electoral cycles in police hiring to estimate the effect of police on crime". *American Economic Review*. 87(3): 270–290.
- Ljungqvist, L. and T. J. Sargent. 2014. "An open letter to professors Heckman and Prescott". *manuscript, New York University*.
- MaCurdy, T. 1985. "Interpreting empirical models of labor supply in an intertemporal framework with uncertainty". In: *Longitudinal Analysis of Labor Market Data*. Ed. by J. Heckman and B. Singer. Cambridge University Press. 148–170.
- Manski, C. F. 2003. *Partial Identification of Probability Distributions*. Springer Verlag.
- Manski, C. F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Harvard University Press.
- Manski, C. F. 2014. "Identification of income-leisure preferences and evaluation of income tax policy". *Quantitative Economics*. 5: 145–174.

- McCloskey, D. 2005. "The trouble with mathematics and statistics in economics". *History of Economic Ideas*. XIII(3): 85–102.
- McCrary, J. 2002. "Do electoral cycles in police hiring really help us estimate the effect of police on crime? Comment". *American Economic Review*. 92(4): 1236–1243.
- Misra, S. and H. Nair. 2011. "A structural model of sales-force compensation dynamics: Estimation and field implementation". *Quantitative Marketing and Economics*. 9: 211–257.
- Neyman, J. 1923. "Sur les applications de la theorie des probabilités aux expériences agricoles: Essai des principes." *Masters Thesis*.
- Pakes, A. and P. McGuire. 2001. "Stochastic algorithms, symmetric Markov perfect equilibrium, and the 'Curse' of dimensionality". *Econometrica*. 69(5): 1261–1281.
- Phillips, P. C. B., S.-P. Shi, and J. Yu. 2015. "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500". *International Economic Review*: forthcoming.
- Popper, K. R. 2002. *The Logic of Scientific Discovery*. Psychology Press.
- Reinhart, C. M. and K. S. Rogoff. 2009. *This Time is Different Eight Centuries of Financial Folly*. Princeton University Press.
- Reinhart, C. M. and K. S. Rogoff. 2010. "Growth in a time of debt". *American Economic Review*. 100(2): 573–578.
- Robert E. Lucas, J. 2003. "Macroeconomic priorities". *American Economic Review*. 93(1): 1–14.
- Romer, P. M. 2015. "Mathiness in the theory of economic growth". *American Economic Review: Papers and Proceedings*. 105(5): 89–93.
- Roth, A. E. 1991. "A natural experiment in the organization of entry level labor markets: regional markets for new physicians and surgeons in the UK". *American Economic Review*. 81: 415–440.
- Roth, A. E. 1991a. "Game theory as part of empirical economics". *Economic Journal*. 101: 107–114.
- Roth, A. E. 2008. "What have we learned from market design?" *Economic Journal*. 118: 285–310.
- Rubin, D. 1980. "Discussion of randomization analysis of experimental data in the Fisher randomization test". *Journal of the American Statistical Association*. 75(591).

- Rubin, D. 2005. "Causal inference using potential outcomes". *Journal of the American Statistical Association*. 100(469): 322–331.
- Rust, J. 1987. "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher". *Econometrica*. 55(3): 999–1033.
- Rust, J. 1988. "Maximum likelihood estimation of discrete control processes". *SIAM Journal on Control and Optimization*. 26(5): 1006–1024.
- Rust, J. 2014. "The limits to inference *with* theory: A review of Wolpin (2013)". *Journal of Economic Literature*. 52(3): 820–850.
- Shiller, R. J. 2015. *Irrational Exuberance*. Princeton University Press.
- Stahl, A. E. and L. Feigenson. 2015. "Observing the unexpected enhances infants' learning and exploration". *Science*. 348(6230): 91–94.
- Su, C.-L. and K. L. Judd. 2012. "Constrained optimization approaches to estimation of structural models". *Econometrica*. 80(5): 2213–2230.
- Theil, H. 1971. *Principles of Econometrics*. New York: John Wiley & Sons.
- Todd, P. E. and K. I. Wolpin. 2006. "Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility". *American Economic Review*. 96(5): 1384–1417.
- Vickrey, W. 1961. "Counterspeculation, auctions, and competitive sealed tenders". *Journal of Finance*. 166(1): 8–37.
- von Neumann, J. 1947. "The mathematician". In: *Works of the Mind, Volume 1*. Ed. by R. Heywood. Chicago: University of Chicago Press. 180–196.
- Widerquist, K. 2005. "A retrospective on the negative income tax experiments: Looking back at the most innovative field studies in social policy". In: *The Ethics and Economics of the Basic Income Guarantee*. Ed. by W. Lewis and Pressman. Aldershot: Ashgate. 95–106.
- Wollan, M. 2015. "Pay it forward: Sophisticated software has enabled a 'market' for organ donation, without any money changing hands". *New York Times Magazine*. May(3).
- Wolpin, K. 2013. *The Limits of Interference without Theory*. Cambridge: MIT Press.

Just How Sensitive are Instrumental Variable Estimates?

Peter C. Reiss

Stanford Graduate School of Business, USA; preiss@stanford.edu

ABSTRACT

Researchers regularly use instrumental variables to resolve concerns about regressor endogeneity. The existing literature has correctly emphasized that the choice of instrumental variables matter for the resulting estimates. This paper shows that researchers should also be concerned that the functional form of the instrument matters as well for the resulting estimates. For example, simply changing an instrumental variable from the level to the logarithm can change estimates directly. This article documents the problem, suggests why the problem occurs and suggests different approaches to the problem.

1

Introduction

Many social scientists work with linear regression models in which some or all of the regressors are thought to be correlated with the regression error. Although in this case ordinary least squares (OLS) delivers the best linear predictor of the dependent variable given the right hand side regressors, this predictor differs from the best linear predictor one would obtain if there were no correlation between the regressors and the regression error. Researchers often are more interested in this second model because they see its coefficients as revealing “causal” effects. That is, they see the coefficients as suggesting how much the dependent variable would change if the corresponding right hand side variable changed by one unit and nothing else changed. Of course, the true causal model may not be linear. However, many researchers nevertheless believe that linear regressions still can reveal the signs and magnitudes of causal relations.

Instrumental variable (IV) methods are the primary means by which social scientists estimate regression models with endogenous regressors. IV methods require the researcher to identify auxiliary variables that minimally are uncorrelated with the regression error (e.g., exogenous) and yet correlated with the right hand side endogenous regressors (i.e.,

relevant).¹ Under standard assumptions, these instruments can be used to construct consistent estimates of the coefficients. Although the use of valid IVs produces consistent estimates, it is well known that the finite and large sample distributions of the IV estimator are impacted by the choice of instruments.² Further, it is known that when the instruments only slightly violate the exogeneity and relevance conditions, there can be dramatic, adverse consequences for the estimator's small and large sample distribution. Violations of concern include: having too few relevant instruments; using instruments that are correlated with the regression error; and, relying on instruments that are weakly correlated with the endogenous regressors.³

This paper documents another issue that has received little or no attention – that seemingly irrelevant changes in the functional forms of the same instruments can lead to *vastly* different IV estimates. This is not just a sampling issue, it is present in a *given sample*. This potential sensitivity should be concerning. Two (or more) researchers could be on solid ground arguing that their IV estimates are consistent, and yet their estimates might differ dramatically. Indeed, their estimates may differ in sign! Section 2 provides such an example. Ultimately this difference in the IV estimates prompts the difficult question of which estimate(s) to report. Alternatively, how might a researcher make others aware of any sensitivity?

These issues are illustrated and addressed in what follows. Section 2 shows that an instrument's functional form can matter. It relates the sensitivity of IV estimates to an instrument relevance condition. Section 3 discusses possible approaches to the problem based on existing relevance and weak instrument diagnostics. These approaches include reporting sensitivity analyses or measures of the local variation in the estimated coefficients. Section 4 discusses possible “efficient” instrument approaches to the problem. Section 5 illustrates the problem is general. Section 6 concludes.

¹While the term “instruments” refers to both the exogenous variables in the equation of interest and the excluded auxiliary variables, I will primarily use it to refer to the auxiliary variables.

²See for example Phillips (1983).

³See for example Bound *et al.* (1995) and Murray (2006).

References

- Anderson, T. 1984. *Introduction to Multivariate Statistical Analysis*. 2nd. New York: Wiley.
- Bound, J., D. Jaeger, and R. Baker. 1995. “Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak”. *Journal of the American Statistical Association*. 90: 443–450.
- Cragg, J. and S. Donald. 1993. “Testing identifiability and specification in instrumental variables models”. *Econometric Theory*. 9: 222–240.
- Davidson, R. and J. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Epple, D. and B. McCallum. 2006. “Simultaneous equation econometrics: The missing example”. *Economic Inquiry*. 44(2): 374–384.
- Kelejian, H. 1971. “Two-stage least squares and econometric systems linear in the parameters but nonlinear in the endogenous variables”. *Journal of the American Statistical Association*. 66: 373–374.
- Kleibergen, F. and R. Paap. 2006. “Generalized reduced rank tests using the singular value decomposition”. *Journal of Econometrics*. 133: 97–126.
- Murray, M. 2006. “Avoiding invalid instruments and coping with weak instruments”. *Journal of Economic Perspectives*. 20(4): 111–132.
- Newey, W. 1990. “Efficient instrumental variables estimation of nonlinear models”. *Econometrica*. 58(4): 809–837.

- Olea, J. and C. Paueger. 2013. "A robust test for weak instruments". *Journal of Business and Economic Statistics*. 31: 358–368.
- Phillips, P. 1983. "Exact small sample theory in the simultaneous equations model". In: *Handbook of Econometrics*. Ed. by Z. Griliches and M. D. Intriligator. Vol. I, Amsterdam: North-Holland. 449–516.
- Robinson, P. 1988. "Root-N consistent semiparametric regression". *Econometrica*. 56: 931–954.
- Sanderson, E. and F. Windmeijer. 2013. "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables". *Working Paper* No. 13/315. University of Bristol.
- Staiger, D. and J. Stock. 1997. "Instrumental variables regression with weak instruments". *Econometrica*. 65: 557–586.
- Stock, J. H. and M. Yogo. 2005. "Testing for weak instruments in linear IV regression". In: *Identification and Inference for Econometric Models*. Ed. by D. W. K. Andrews and J. H. Stock. Essays in Honor of Thomas Rothenberg,. New York: Cambridge University Press. 80–108.
- Yeo, I. and R. Johnson. 2000. "A new family of power transformations to improve normality or symmetry". *Biometrika*. 87: 954–959.

Interpreting Point Predictions: Some Logical Issues

Charles F. Manski

*Department of Economics and Institute for Policy Research,
Northwestern University, USA; cfmanski@northwestern.edu*

ABSTRACT

Forecasters regularly make point predictions of future events. Recipients of the predictions may use them to inform their own assessments and decisions. This paper integrates and extends my past analyses of several simple but inadequately appreciated logical issues that affect interpretation of point predictions. I explain the algebraic basis for a pervasive empirical finding that the cross-sectional mean or median of a set of point predictions is more accurate than the individual predictions used to form the mean or median, a phenomenon sometimes called the “wisdom of crowds.” I call attention to difficulties in interpretation of point predictions expressed by forecasters who are uncertain about the future. I consider the connection between predictions and reality. In toto, the analysis questions prevalent prediction practices that use a single combined prediction to summarize the beliefs of multiple forecasters.

1

Introduction

Persons, firms, and governments regularly make point predictions of future events and estimates of past conditions. Recipients of predictions and estimates may use them to inform their own assessments and decisions. This paper integrates and extends my past analyses of several simple but inadequately appreciated logical issues that affect interpretation of predictions and estimates. I use the prospective term “prediction” (or forecast) for simplicity of terminology, but the paper applies to retrospective estimates as well.

I first explain the logical basis for a pervasive empirical finding on the performance of combined predictions of real quantities. Empirical researchers have long reported that the cross-sectional mean or median of a set of point predictions is more accurate than the individual predictions used to form the mean or median. This phenomenon is sometimes colloquially called the “wisdom of crowds.” It has only occasionally been recognized that these regularities have algebraic foundations. The one concerning mean predictions holds whenever a convex loss function (or concave welfare function) is used to measure prediction accuracy, by Jensen’s inequality. The one concerning median predictions holds whenever a unimodal loss or welfare function is used to measure accuracy.

I have called attention to the algebra underlying the wisdom of crowds in Manski (2010) and Manski (2011). Here, in Section 2, I paraphrase these earlier discussions and extend them to cover weighted averages of predictions, as advocated in Bayesian model averaging. I also caution that the algebra has limited scope of application. In particular, I show that combining predictions of treatment response need not outperform individual predictions when a planner makes binary treatment decisions.

The algebra in Section 2 implies nothing about the informativeness of the predictions that forecasters provide. Predictions may imperfectly convey the expectations that forecasters hold for future events and they may imperfectly anticipate future realities. To interpret predictions requires assumptions on the decision processes that forecasters use to generate their predictions. Sections 3 and 4 interpret predictions under various assumptions.

Section 3 concerns interpretation of point predictions of uncertain events. Economists commonly assume that persons hold probabilistic beliefs about uncertain events. A point prediction at most provides some measure of the location of a probability distribution—it cannot reveal anything about the shape of the distribution. Users of predictions typically do not know how forecasters choose points to summarize their beliefs. This generates an unavoidable problem in interpretation of point predictions, one that arises even if forecasters seek to honestly convey their beliefs. Other problems, which are avoidable, arise when researchers make logical errors in their interpretation of point predictions. A frequent error has been to use the dispersion of point predictions across forecasters to measure the uncertainty that forecasters perceive. Engelberg *et al.* (2009) and others have called attention to these matters. Section 3 explains.

Predicting binary outcomes is an important special case that is instructive to study in some depth. Manski (1990) observed that a point prediction of a binary outcome at most yields a bound on the probability that the forecaster holds for the outcome. Manski (2006) applied this simple finding to interpret the bids in prediction markets where traders bet on occurrence of a binary outcome. I summarize here.

The analyses of Sections 2 and 3 imply no conclusions about the connection between predictions and reality. Assessment of realism requires assumptions about the process generating predictions. Section 4 discusses two cases: unbiased and rationalizable predictions.

Predictions of a real quantity are unbiased if they are generated by a process such that the mean prediction equals the actual value of the quantity. This assumption, often made in research that combines forecasts, makes the wisdom of crowds a statistical rather than simply algebraic phenomenon. The assumption has strong consequences but is rarely credible.

Rationalizable predictions are ones that follows logically from some plausible model, without requiring that the model be accurate. Such predictions pose possible futures. In the absence of knowledge of the process generating predictions, there is no logical reason to average a set of rationalizable predictions as recommended in research on unbiased prediction. However, if predictions are formed in an adversarial environment, one might find it reasonable to conclude that the actual value of the quantity of interest lies in the interval between the smallest and largest prediction.

The concluding Section 5 observes that prevalent practices in prediction of future events summarize the beliefs of forecasters in two respects. First, individual forecasters commonly provide point predictions even though they may perceive considerable uncertainty. Second, recipients of predictions from multiple forecasters often combine them to form the mean or median prediction. I question both practices.

References

- Barron, O., M. Stanford, and Y. Yu. 2009. "Further evidence on the relation between analysts' forecast dispersion and stock returns". *Contemporary Accounting Research*. 26: 329–357.
- Bates, J. and C. Granger. 1969. "The combination of forecasts". *Operations Research Quarterly*. 20: 451–468.
- Brenner, M. and Y. Landskroner. 1983. "Inflation uncertainties and returns on bonds". *Economica*. 50: 463–468.
- Clemen, R. 1989. "Combining forecasts: A review and annotated bibliography". *International Journal of Forecasting*. 5: 559–583.
- Clements, M. 2009. "'Internal consistency of survey respondents' forecasts: Evidence based on the survey of professional forecasters". *The Methodology and Practice of Econometrics*, J. Castle and N. Shephard (editors).
- Cukierman, A. and P. Wachtel. 1979. "Differential inflationary expectations and the variability of the rate of inflation: Theory and evidence". *American Economic Review*. 69: 595–609.
- Diether, K., C. Malloy, and A. Scherbina. 2002. "Differences of opinion and the cross section of stock returns". *The Journal of Finance*. 57: 2113–2141.
- Elliott, G., I. Komunjer, and A. Timmermann. 2005. "Estimation and testing of forecast rationality under flexible loss". *Review of Economic Studies*. 72: 1107–1125.

- Elliott, G., I. Komunjer, and A. Timmermann. 2008. "Biases in macroeconomic forecasts: Irrationality or asymmetric loss?" *Journal of European Economic Association*. 6: 122–157.
- Engelberg, J., C. Manski, and J. Williams. 2009. "Comparing the point predictions and subjective probability distributions of professional forecasters". *Journal of Business and Economic Statistics*. 27: 30–41.
- Galton, F. 1907. "Vox populi". *Nature*. 75: 450–451.
- Giordani, P. and P. Söderlind. 2003. "Inflation forecast uncertainty". *European Economic Review*. 47: 1037–1059.
- Güntay, L. and D. Hackbarth. 2010. "Corporate bond credit spreads and forecast dispersion". *Journal of Banking & Finance*. 34: 2328–2345.
- Hahn, J. and D. Steigerwald. 1999. "Consumption adjustment under time varying income uncertainty". *Review of Economics and Statistics*. 81: 32–40.
- Hayford, M. 2000. "Inflation uncertainty, unemployment uncertainty, and economic activity". *Journal of Macroeconomics*. 22: 315–329.
- Hodges, J. and E. Lehmann. 1950. "Some problems in minimax point estimation". *Annals of Mathematical Statistics*. 21: 182–197.
- Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky. 1999. "Bayesian model averaging: A tutorial". *Statistical Science*. 14: 382–417.
- Johnson, T. 2004. "Forecast dispersion and the cross section of expected returns". *Journal of Finance*. 59: 1957–1978.
- Juster, T. 1966. "Consumer buying intentions and purchase probability: An experiment in survey design". *Journal of the American Statistical Association*. 61: 658–696.
- Larrick, R. and J. Soll. 2006. "Intuitions about combining opinions: Misappreciation of the averaging principle". *Management Science*. 52: 111–127.
- Levi, M. and J. Makin. 1979. "Fisher, Phillips, Friedman and the measured impact of inflation on interest". *Journal of Finance*. 34: 36–52.
- Levi, M. and J. Makin. 1980. "Inflation uncertainty and the Phillips curve: Some empirical evidence". *American Economic Review*. 70: 1022–1027.

- Makin, J. 1982. "Anticipated money, inflation uncertainty, and real economic activity". *Review of Economics and Statistics*. 64: 126–134.
- Mankiw, G. R. R. and J. Wolfers. 2003. "Disagreement about inflation expectations". *NBER Macroeconomics Annual 18*. Chicago: University of Chicago Press: 209–248.
- Manski, C. 1990. "The use of intentions data to predict behavior: A best case analysis". *Journal of the American Statistical Association*. 85: 934–940.
- Manski, C. 2004. "Measuring expectations". *Econometrica*. 72: 1329–1376.
- Manski, C. 2006. "Interpreting the predictions of prediction markets". *Economic Letters*. 91: 425–429.
- Manski, C. 2010. "When consensus choice dominates individualism: Jensen's inequality and collective decisions under uncertainty". *Quantitative Economics*. 1: 187–202.
- Manski, C. 2011. "Interpreting and combining heterogeneous survey forecasts". *Oxford Handbook on Economic Forecasting*: 457–472. Ed. by M. Clements and D. Hendry.
- Manski, C. 2013. *Public Policy in an Uncertain World*. Cambridge, MA: Harvard University Press.
- McNees, S. 1992. "The uses and abuses of 'Consensus' forecasts". *Journal of Forecasting*. 11: 703–710.
- Surowiecki, J. 2004. *The Wisdom of Crowds*. New York: Random House.
- Wikipedia. 2014. "Wisdom of the crowd". accessed August 19, 2014. URL: en.wikipedia.org/wiki/Wisdom_of_the_crowd#cite_note-WoC-Combi-1.
- Zarnowitz, V. and L. Lambros. 1987. "Consensus and uncertainty in economic prediction". *Journal of Political Economy*. 95: 591–621.

From Casual to Causal Inference in Accounting Research: The Need for Theoretical Foundations

Jeremy Bertomeu¹, Anne Beyer² and Daniel J. Taylor³

¹*Baruch College; City University of New York, USA;*
jeremy.bertomeu@baruch.cuny.edu

²*Graduate School of Business; Stanford University, USA;*
abeyer@stanford.edu

³*The Wharton School; University of Pennsylvania, USA;*
dtayl@wharton.upenn.edu

ABSTRACT

On December 5th and 6th 2014, the Stanford Graduate School of Business hosted the *Causality in the Social Sciences Conference*. The conference brought together several distinguished speakers from philosophy, economics, finance, accounting and marketing with the bold mission of debating scientific methods that support causal inferences. We highlight key themes from the conference as relevant for accounting researchers. First, we emphasize the role of formal economic theory in informing empirical research that seeks to draw causal inferences, and offer a skeptical perspective on attempts to draw causal inferences in the absence of well-defined constructs and assumptions. Next, we highlight some of the conceptual limitations of quasi-natural experimental methods that were discussed at the conference, and discuss the role of structural estimation. Finally, we illustrate many of the points from the conference by estimating a novel, theoretically grounded measure of disclosure costs.

1

Introduction

On December 5th and 6th 2014, the Stanford Graduate School of Business hosted the *Causality in the Social Sciences Conference*. The conference brought together distinguished speakers from philosophy, economics, finance, accounting and marketing with the bold mission of debating scientific methods that support causal inferences in the social sciences. The conference was structured around a keynote by philosopher Nancy Cartwright and presentations by five notable economists: Joshua Angrist, Guido Imbens, Charles Manski, Peter Reiss and John Rust. Three panel discussions offered complementary views about causality in the areas of finance, accounting and marketing. Each panel was composed of journal editors and leading researchers, and focused on the state-of-the-art and opportunities for future research. The conference approached several issues of interest to accounting researchers, such as the role of theoretical foundations, quasi-natural experimental methods, and structural estimation.

The conference organizers asked us to report on and synthesize the proceedings from three unique perspectives: audience members, accounting researchers, and a unique blend of individuals with empirical and theoretical research interests. We were asked to illustrate some of

the themes of the conference using examples drawn from the accounting literature and to include a brief application. While initially challenging, we found it intellectually rewarding to step outside the silos of “empirical researcher” and “theoretical researcher” and discuss issues of common interest. Perhaps this will be the conference’s legacy: the conference showed methodological and cross-disciplinary academic fertilization at its best, and has the potential to disseminate new perspectives about accounting topics.

Accounting researchers may look at the vast and divergent views being espoused in the proceedings of the conference published in this issue of *Foundation and Trends*, and wonder about the takeaways as it relates to accounting. Our synthesis aims to highlight what we think were a few of the key takeaways, and expand on each using several examples and settings familiar to accounting researchers. In this respect, we caveat that our synthesis necessarily reflects our own views as accounting researchers. We view this report as contributing to the debate *in accounting* surrounding causal inferences, methodology, and the role of formal theory in informing empirical work—certainly not as the final word on that debate. Our synthesis is intended to complement a growing body of literature exploring issues related to causality within accounting. Given the shared topic—causality in accounting research—many of the issues raised at the conference and included in our synthesis also appear in prior work.¹

Two key themes emerged at the conference, and these themes pervade our discussion. First, regardless of method, causal inferences rely on *untestable* core assumptions (see Nancy Cartwright’s paper in this issue). Identification of causal channels does not come from statistical techniques but from assumptions. For example, if the assumptions behind instrumental variable regression (IV) are satisfied (e.g., the instruments satisfy the exclusion restriction), then IV can be used to estimate a causal effect. While institutional knowledge might tell us

¹See, for example, the Special Issue on Causality in the October 2014 issue of *Accounting, Organizations, and Society*, Gow *et al.* (2016), and Qi Chen and Katherine Schipper’s paper in this issue.

whether certain assumptions are more plausible than others, certain *untestable* assumptions will always be necessary.²

Second, there was considerable skepticism about statistical techniques commonly referred to as “quasi-natural experimental methods”, and whether strong, causal inferences typically associated with the use of such methods are reasonable. We found this particularly surprising given the increasing emphasis on these methods within the accounting literature. Concerns about these methods focused on generalizability and the ability of these methods to identify underlying causal mechanisms in the absence of formal theory.³

At this point, let us simply note two practical implications of these themes for originating research that pushes the frontiers of the accounting literature. First, making assumptions should not be taken as a scientific compromise, or something to hide; instead, assumptions should be presented and opened for discussion. The message is simple: assumptions should be explicit, transparent, and deliver new insights. Without a clear understanding of the assumptions of empirical tests and measures, causal inferences will remain elusive.

Second, regardless of empirical method, a link between formal economic theory and empirical work is essential in helping researchers identify causal effects of interest. For example, when researchers measure systematic risk, they employ estimates of β from the Capital Asset Pricing Model—an empirical measure of systematic risk derived from formal theory; when researchers measure information asymmetry, they employ estimates of Kyle’s λ or the probability of informed trade (PIN)—empirical measures of information asymmetry between market participants derived from formal theory. It strikes us that none of these measures is particularly intuitive in the absence of the corresponding formal economic theory and assumptions. Formal theory provides the

² For example, the exclusion restriction of instrumentals variable, the parallel trends assumption of difference-in-differences, and the continuity assumption of regression discontinuity are all inherently untestable (e.g., Larcker and Rusticus, 2010; Roberts and Whited, 2013).

³ These issues are not new to the economics literature (e.g., Heckman, 2000, Heckman, 2005; Heckman and Vytlačil, 2007).

assumptions that guide how we interpret the relations in the data—for example, how we interpret the covariance between a firm’s returns and the market return. In this regard, formal theory makes transparent the assumptions that underlie each of the above measures.

Relative to the other fields represented at the conference, accounting research tends not to be very clear about what assumptions it is making. For example, researchers in accounting generally rely on intuition and less formal verbal descriptions to motivate empirical measures of theoretical constructs rather than mathematical descriptions derived from formal theory. The advantage of this approach is clear: it does not constrain empirical work to topics studied by theorists. The disadvantage of this approach is also clear: great confusion—about the validity of empirical proxies and what theoretical constructs they are intended to capture.

As a result, the accounting literature simultaneously features a great variety of empirical proxies and fundamental disagreements about what these proxies capture. Few of the numerous proxies of accounting quality, conservatism, or proprietary costs are derived from formal theory. Consequently, the theoretical construct of interest is only defined at an intuitive level, and the assumptions that underlie various proxies are not transparent. For example, what assumptions are necessary for the asymmetric timeliness coefficient of Basu (1997) to measure conservatism? While the proxy has great intuitive appeal, it is only more than 15 years later—and only *after* its widespread use—that we are starting to get a sense for the assumptions that underlie this measure (e.g., Ball *et al.*, 2013). Understanding these assumptions, and their validity, is critical to reliably estimating causal effects. More generally, it seems difficult to credibly identify the causal effect of an intervention (e.g., regulation) on a theoretical construct, if that construct is only defined at an intuitive level, or if we do not understand the assumptions that underlie our empirical measures of the construct.

This is now being given the opportunity to change, as more researchers within accounting are interested in probing the foundations of common measures of accounting quality, conservatism, and proprietary costs—among other examples—each with the potential to affect a dra-

matic rethink in the measures and relations that have been exhausted in prior research.

We structure the remainder of the paper as follows. In Section 2, we discuss the role of formal economic theory in informing empirical research that seeks to make causal statements, and offer a skeptical perspective on attempts to draw causal inferences in the absence of well-defined constructs and assumptions. In Section 3, we highlight some of the conceptual limitations of quasi-natural experimental methods that were discussed at the conference; limitations that appear to be underappreciated within the accounting literature. In Section 4, we discuss the role of structural estimation as an emerging method within accounting research, and provide a simple application that illustrates many of the themes discussed at the conference. Specifically, we estimate a novel, theoretically grounded measure of voluntary disclosure costs based on Verrecchia (1983).

References

- Abdulkadiroglu, A., J. Angrist, P. Hull, and P. Parhak. 2016. “Charters without lotteries: Testing takeovers in New Orleans and Boston”. *American Economic Review*, forthcoming.
- Armstrong, C. 2013. “Discussion of CEO compensation and corporate risk-taking: Evidence from a natural experiment”. *Journal of Accounting and Economics*. 56: 102–111.
- Armstrong, C., D. Taylor, and R. Verrecchia. 2016. “Asymmetric reporting”. *Journal of Financial Reporting*. 1: 15–32.
- Arya, A., J. Glover, and S. Sunder. 2003. “Are unmanaged earnings always better for shareholders?” *Accounting Horizons*. 17: 111–116.
- Balakrishnan, K., M. Billings, B. Kelly, and A. Ljungqvist. 2014. “Shaping liquidity: On the causal effects of voluntary disclosure”. *Journal of Finance*. 69: 2237–2278.
- Ball, R., S. P. Kothari, and V. Nikolaev. 2013. “Econometrics of the Basu asymmetric timeliness coefficient and accounting conservatism”. *Journal of Accounting Research*. 51: 1071–1097.
- Basu, S. 1997. “The conservatism principle and the asymmetric timeliness of earnings”. *Journal of Accounting and Economics*. 24: 3–37.
- Beaver, B., M. Barth, and W. Landsman. 2001. “The relevance of the value relevance literature for financial accounting standard setting: Another view”. *Journal of Accounting and Economics*. 31: 77–104.

- Bertomeu, J., M. Darrrough, and W. Xue. 2016. "Optimal conservatism with earnings manipulation". *Contemporary Accounting Research*, forthcoming.
- Bertomeu, J., P. Ma, and I. Marinovic. 2015. "How often do managers withhold information?" Working Paper.
- Beyer, A., I. Guttman, and I. Marinovic. 2015. "Earnings management and earnings quality: Theory and evidence". Working paper.
- Chen, T., J. Harford, and C. Lin. 2015. "Do analysts matter for governance? Evidence from natural experiments". *Journal of Financial Economics*. 115: 383–410.
- Cheyne, E. and M. Liu-Watts. 2015. "A back-of-the-envelope estimator of disclosure costs". Working paper.
- Derrien, F. and A. Kecskes. 2013. "The real effects of financial shocks: Evidence from exogenous changes in analyst coverage". *Journal of Finance*. 68: 1407–1440.
- Easley, D., N. Kiefer, and M. O'Hara. 1997. "One day in the life of a very common stock". *Review of Financial Studies*. 10: 805–835.
- Easley, D. and M. O'Hara. 2004. "Information and the cost of capital". *Journal of Finance*. 59: 1553–1583.
- Feltham, G. and J. Ohlson. 1995. "Valuation and clean surplus accounting for operating and financial activities". *Contemporary Accounting Research*. 11: 689–731.
- Gebhardt, W., C. Lee, and B. Swaminathan. 2001. "Towards an implied cost of capital". *Journal of Accounting Research*. 39: 135–176.
- Gerakos, J. and C. Syverson. 2015. "Competition in the audit market: Policy implications". *Journal of Accounting Research*. 53: 725–775.
- Gigler, F., C. Kanodia, H. Saprà, and R. Venugopalan. 2009. "Accounting conservatism and the efficient of debt contracts". *Journal of Accounting Research*. 47: 767–797.
- Gormley, T. and D. Matsa. 2014. "Common errors: How to (and not to) control for unobserved heterogeneity". *Review of Financial Studies*. 27: 617–661.
- Gow, I., D. Larcker, and P. Reiss. 2016. "Causal inferences in accounting research". *Journal of Accounting Research*. 54: 477–523.

- Guay, W., D. Samuels, and D. Taylor. 2016. "Guiding through the fog: Financial statement complexity and voluntary disclosure". *Journal of Accounting and Economics*, forthcoming.
- Heckman, J. 2000. "Causal parameters and policy analysis in economics: A twentieth century retrospective". *Quarterly Journal of Economics*. 115: 45–97.
- Heckman, J. 2005. "The scientific model of causality". *Sociological Methodology*. 35: 1–97.
- Heckman, J. and E. Vytlacil. 2007. "Econometric evaluation of social programs, part I: Causal models, structural models, and econometric policy evaluation". In: *Handbook of Econometrics*. Chap. 70. 4779–4874.
- Holthausen, R. and R. Watts. 2001. "The relevance of the value relevance literature for financial accounting standard setting". *Journal of Accounting and Economics*. 31: 3–75.
- Iliev, P. 2010. "The effect of SOX section 404: Costs, earnings quality, and stock prices". *Journal of Finance*. 65: 1163–1196.
- Irani, R. and D. Oesch. 2013. "Monitoring and corporate disclosure: Evidence from a natural experiment". *Journal of Financial Economics*. 109: 398–418.
- Kelly, B. and A. Ljungqvist. 2012. "Testing asymmetric-information asset pricing models". *Review of Financial Studies*. 25: 1366–1413.
- LaFond, R. and R. Watts. 2008. "The information role of conservatism". *The Accounting Review*. 83: 447–478.
- Lambert, R., D. Larcker, and R. Verrecchia. 1991. "Portfolio considerations in valuing executive compensation". *Journal of Accounting Research*. 29: 129–149.
- Lambert, R., C. Leuz, and R. Verrecchia. 2007. "Accounting information, disclosure, and the cost of capital". *Journal of Accounting Research*. 45: 385–420.
- Lambert, R., C. Leuz, and R. Verrecchia. 2012. "Information asymmetry, information precision, and the cost of capital". *Review of Finance*. 16: 1–29.

- Larcker, D. and T. Rusticus. 2010. "On the use of instrumental variables in accounting research". *Journal of Accounting and Economics*. 49: 186–205.
- Leuz, C. and P. Wysocki. 2016. "The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research". *Journal of Accounting Research*. 54: 525–622.
- Lucas, R. 1976. "Econometric policy analysis: A critique". *Carnegie-Rochester Conference Series on Public Policy*. 1: 19–46.
- McNichols, M. and B. Trueman. 1994. "Public disclosure, private information collection, and short-term trading". *Journal of Accounting and Economics*. 17: 69–94.
- Michels, J. 2016. "Disclosure versus recognition: Inference from subsequent events". *Journal of Accounting Research*, forthcoming.
- Nikolaev, V. 2014. "Identifying accounting quality". Working paper.
- Ohlson, J. 1995. "Earnings, book values, and dividends in equity valuation". *Contemporary Accounting Research*. 11: 661–687.
- Payne, J. and W. Thomas. 2003. "The implications of using stock-split adjusted I/B/E/S data in empirical research". *Accounting Review*. 78: 1049–1067.
- Roberts, M. and T. Whited. 2013. "Endogeneity in empirical corporate finance". *Handbook of Economics and Finance*. 2: 493–572.
- Roll, R. 1977. "A critique of the asset pricing theory". *s tests? Journal of Financial Economics*. 4: 129–176.
- Romer, P. 2015. "Mathiness in the theory of economic growth". *American Economic Review: Papers and Proceedings*. 105: 89–93.
- Rust, J. 2014. "The limits of inferences with theory: A review of Wolpin (2013)". *Journal of Economic Literature*. 52: 820–850.
- Verrecchia, R. 1983. "Discretionary disclosure". *Journal of Accounting and Economics*. 5: 179–194.
- Wolpin, K. 2013. *The Limits of Inference without Theory*. Cambridge, MA: MIT Press.
- Zakolyukina, A. 2014. "Measuring intentional gaap violations: A structural approach". Working paper.

Comments and Observations Regarding the Relation Between Theory and Empirical Research in Contemporary Accounting Research

Qi Chen and Katherine Schipper

*Duke University: The Fuqua School of Business, USA;
qc2@duke.edu, schipper@duke.edu*

ABSTRACT

We offer some thoughts on the relation between theoretical and empirical accounting research in the context of causal inference, in response to two questions posed by Professor Ivan Marinovic, organizer of the 2014 Stanford University Graduate School of Business Causality conference. The two questions are: should causal inference be the objective of accounting research; and what is, and what should be, the relation between theory and empirical research in accounting? With regard to the latter, we point to two sources of difficulty: (1) confusion and disagreement about interpretation, advantages and disadvantages of various empirical identification strategies; and (2) a lack of progress on the part of empirical researchers in testing the implications of existing accounting theories and thereby providing discipline to those theories. We argue that published empirical accounting research relies too much on insufficiently precise verbal models or generic models that provide few or no new accounting-specific insights and tends to ignore recent advances made by theoretical researchers. As a result analytical models in accounting research are not sufficiently challenged by empirical research and analytical

Qi Chen and Katherine Schipper (2016), “Comments and Observations Regarding the Relation Between Theory and Empirical Research in Contemporary Accounting Research”, *Foundations and Trends® in Accounting*: Vol. 10, No. 2-4, pp 314–360. DOI: 10.1561/1400000046.

researchers have made slow progress in establishing a meaningful distinction between accounting information and other types of information provided by firms and their managers. Our concern is that accounting research is in danger of losing the healthy disciplining balance between theory and empirical research that is essential to any scientific field. Without this balance, the profession becomes a discipline of beliefs, rather than a discipline of scientific discovery.

1

Introduction

We present our views on two questions posed by Professor Ivan Marinovic, organizer of the 2014 Causality Conference at Stanford University's Graduate School of Business. The first question pertains to the role of causality in accounting research, specifically whether casual inference should be the objective of accounting research. The second question pertains to the relation between theory and empirical research in accounting, including whether theory should discipline, and has properly disciplined, empirical research; whether the gap between theory and empirical research is undesirably large; and whether the development of theory in accounting has been affected by empirical accounting evidence. We illustrate our views by reference to research that considers the relation between disclosures and the costs of capital, accounting measurement, and the necessity for and consequences of financial reporting standards.¹

¹ Although some discussion at the Stanford Causality conference focused on process issues, including specifically the role of theory in (1) the context of the publication process in accounting research, for example the roles of referees and editors, and in (2) the context of PhD education, this discussion abstracts from those issues. Also, although we illustrate some of our views by reference to specific

With regard to the first question, we do not view the role of causality to be different in accounting research than in other business-related disciplines such as finance or economics; what differs is the context in which researchers analyze causality. We take the position that accounting research seeks to understand the causes, uses and consequences of accounting information; some research focuses specifically on firm-specific accounting choices (accounting policies, judgments and estimates) and financial reporting standard setting decisions, with the goal of producing insights to enhance the role of accounting in improving the efficiency of resource allocation in the economy. Achieving these objectives requires both an understanding of the theoretical links among behaviors, accounting information and outcomes, and empirical evidence on the existence and magnitudes of the links. That said, and as explained later in this discussion, we believe there is a substantial role for descriptive evidence, referring to the provision of facts or evidence presented as facts, in accounting research, as long as the provision of descriptive evidence is motivated by a desire to test and challenge theories.

With regard to the second question, we argue that while few accounting academics would disagree (at least publicly) that theories are important first steps to establish causal links, accounting researchers do not agree on what constitutes a rigorous theory, or at least a usable one, and to what extent existing theories can be used to guide empirical accounting research. Accounting empiricists criticize accounting theorists for making implausible and empirically meaningless assumptions about institutional features unique to accounting information, or ignoring them altogether, and for wholesale inattention to making their analyses as transparent as possible to those who would like to test them. Accounting theorists, in turn, criticize empiricists for excessive reliance on imprecise *ad hoc* verbal models and for overly-narrow viewpoints on what are the important institutional features of accounting information.

We believe these disagreements reflect a gap between empirical and analytical research in accounting, and one that is increasing, in that empirical accounting research displays a tendency to have little direct

examples of accounting research, this discussion is not intended to provide a survey or review of any area of accounting research.

connection to rigorous economic theory. In turn, accounting theorists are not sufficiently challenged by empirical facts. Without a concerted effort by the profession to bridge the gap, accounting research is in danger of losing the healthy disciplining balance between theory and empirical research that is essential to any scientific field. Without this balance, the profession becomes a discipline of beliefs, rather than a discipline of scientific discovery.

With regard to both questions posed by Professor Marinovic, we note that some disagreements about causality and causal inference in accounting research appear to focus on which empirical identification strategy is superior, for example, structural estimation or natural experiment, with some confusion about what exactly is considered a structural estimation.² Again, such disagreements are not unique to accounting; similar debate has taken place in economics and other disciplines (see, e.g., Heckman, 2000, Heckman and Vytlačil, 2007). Our view in this discussion is that at least some of the disagreement among accounting researchers is misplaced. Different definitions of structural estimations exist, and whatever the definition, both structural estimation and natural experiments each have their pros and cons, and neither approach can claim absolute superiority over a reduced form regression. The first implication of this view is that the type of research question addressed should dictate the choice of research design and method, which in turn affects what kinds of inferences the research can support. The second implication is that an assessment of whether a specific research paper makes a substantial incremental contribution should depend on whether the paper presents significant new insights or significant new methodologies or significant new evidence, arrived at as rigorously as possible given the constraints imposed by data existence and access, and should be largely decoupled from the specific methods applied. Debates over empirical identification strategies divert attention from the very real

² We do not seek to contribute to debates and discussions about what qualifies as a structural model. Our discussion takes the viewpoint expressed in Hood and Koopmans (1953). For purposes of this discussion, a structural model shows a decision maker's behavior under constraints and subject to economic forces, and is distinct from an empirical summary of that behavior.

problems that arise from a lack of rigorous theories that help establish causal links in the first place.

As previously noted, we use ideas from three streams of accounting research to illustrate our views: (1) the relation between voluntary and/or mandatory disclosure and costs of capital; (2) the possible effects of differences in accounting measurement; and (3) the necessity and consequences of financial reporting standards. We chose these research areas for two reasons. First, the research addresses issues that are important for the accounting profession. Second, the research streams differ in the extent to which theories are well developed and empirical work has been tied to those theories. For example, theorists have made substantial progress in establishing causal links between disclosure (construed broadly as information provision) and costs of capital, and empirical work has by and large been organized around those theories. In contrast, and despite their significance for accountancy as a discipline and as a profession, the second and third research streams remain at a preliminary stage theoretically and empirical work tends not to be organized around the existing theories.

Although we consider the three areas separately for expositional ease, we view them as related. For example, financial reporting standards specify required or permitted accounting measurement attributes for assets and liabilities, and that specification presumably has consequences for the nature of information provided. As another example, evidence on the consequences of disclosure (voluntary or mandatory), including for example its effect on costs of capital, should in turn inform standard setters.

The rest of this discussion proceeds as follows. Section 2 provides background to the discussions of accounting research which follow. Section 3 provides information about the number of theory papers published in accounting and citation-based evidence on the extent to which accounting theory papers appear to be influential in subsequent research. Section 4 presents our views on the relation between theory and empirical work in three areas of accounting research. Section 5 concludes.

References

- Angrist, J. D. and V. Lavy. 1999. "Using Maimonides' rule to estimate the effect of class size on scholastic achievement". *Quarterly Journal of Economics*. 114: 533–575.
- Angrist, J. D. and J.-S. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Ashenfelter, O. 1978. "Estimating the effect of training programs on earnings". *Review of Economics and Statistics*. 60: 47–57.
- Ashenfelter, O. and D. Card. 1985. "Using the longitudinal structure of earnings to estimate the effect of training programs". *Review of Economics and Statistics*. 67: 648–660.
- Bennedsen, M., K. M. Nielsen, F. Perez-Gonzalez, and D. Wolfenzon. 2007. "Inside the family firm: The role of families in succession decisions and performance". *Quarterly Journal of Economics*. 122: 647–691.
- Card, D. and A. B. Krueger. 1994. "Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania". *American Economic Review*. 84: 772–793.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR". *Quarterly Journal of Economics*. 126: 1593–1660.

- Davis, M. A., J. D. M. Fisher, and T. M. Whited. 2014. "Macroeconomic implications of agglomeration". *Econometrica*. 82: 731–764.
- Dube, A., T. W. Lester, and M. Reich. 2010. "Minimum wage effects across state borders: Estimates using contiguous counties". *Review of Economics and Statistics*. 92: 945–964.
- Heckman, J. J., H. Ichimura, and P. E. Todd. 1997. "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". *Review of Economic Studies*. 64: 605–654.
- Heckman, J. and R. Pinto. 2015. "Causal analysis after Haavelmo". *Econometric Theory*. 31: 115–151.
- Koopmans, T. C. 1949. "Identification problems in economic model construction". *Econometrica*. 17: 125–144.
- Krueger, A. B. 1999. "Experimental estimates of education production functions". *Quarterly Journal of Economics*. 114: 497–532.
- Krueger, A. B. and D. M. Whitmore. 2001. "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR". *Economic Journal*. 111: 1–28.
- Neumark, D. and W. Wascher. 1992. "Employment effects of minimum and subminimum wages: Panel data on state minimum wage laws". *Industrial and Labor Relations Review*. 46: 55–81.
- Sorkin, I. 2015. "Are there long-run effects of the minimum wage?" *Review of Economic Dynamics*. 18: 306–333.
- Staiger, D. and J. H. Stock. 1997. "Instrumental variables regression with weak instruments". *Econometrica*. 65: 557–586.

Plausibility: A Fair & Balanced View of 30 Years of Progress in Ecologics

Ivo Welch

Anderson School of Management, University of California, Los Angeles, USA; ivo.welch@anderson.ucla.edu

ABSTRACT

Theoretical model identifiability and universality are opposites. Only the strongest and most basic forces and models are universal enough to permit reasonable extrapolation above and beyond their specific historical contexts. Unfortunately, papers that admit to these limits are rarely considered interesting. Instead, researchers search for and find ever-more unlikely explanations and ever-more unlikely evidence in a competitive quest to be surprising, clever and published. The review processes have also incentivized lack of care, private and social failure to correct errors, and even outright misconduct. The discipline is veering towards a theater of the absurd. To help correct the problem, I suggest dedicating a third of every journal to independent replications and critique of prior papers.

1

Introduction

My paper belabors the obvious: First, plausible extrapolatable forces and models are limited to only the strongest of effects. Models either make sharp predictions that facilitate good empirical identification and tests, or they are suitable for extrapolation to other contexts, but not both. Second, published research claims are now so routinely violating reasonable plausibility limits that violations have become the routine. Third, ecologics provides superior competitive incentives to find implausible surprising results, small incentives to find plausible unsurprising results, and negative incentives to replicate, falsify, or verify existing results without prejudice.

To put the scientific state of affairs in economics, finance, and accounting into context, and to explain what is wrong with it in a simple metaphor, my paper begins with an extension of a fable about the field of ecologics invented by Leamer (1983). Leamer's metaphor lends itself naturally to a skeptical discussion of prominent research over the last 30 years. This history helps to frame simple axioms about plausible research. My paper ends with observations about the process that led us to this point and that will lead us beyond where we are today. As such, it contains warnings about the future of the profession and suggestions for radical change.

References

- Bennett, C. M., A. A. Baird, M. B. Miller, and G. L. Wolford. 2010. “Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: An argument for proper multiple comparisons correction”. *Journal of Serendipitous and Unexpected Results*. 1: 1–5.
- Henrich, J., S. J. Heine, and A. Norenzayan. 2010. “The weirdest people in the world?” *Behavioral and Brain Sciences*. 33(2-3): 61–83.
- Ingersoll, Jr., J. E. 2016. “Cumulative prospect theory, aggregation, and pricing”. *Critical Finance Review*. 5:2: forthcoming.
- Leamer, E. E. 1983. “Let’s take the con out of econometrics”. *The American Economic Review*. 73(1): 31–43.
- Lucas, R. 1976. “Econometric policy evaluation: A critique”. In: *The Phillips Curve and Labor Markets*. Ed. by K. Brunner and A. Meltzer.
- McAfee, R. P. 1983. “American economic growth and the voyage of Columbus”. *The American Economic Review*. 73(4): 735–740.
- McCloskey, D. N. 1986. *The Rhetoric of Economics*. University of Wisconsin Press. ISBN: 0-299-15814-4.
- Pfleiderer, P. 2014. “Chameleons: the misuse of theoretical models in finance and economics”. Technical report. Stanford University.

- Spitzer, M. and E. Talley. 2014. "On experimentation and real options in financial regulation". *The Journal of Legal Studies*. 43(S2): S121–S149.
- Welch, I. 2012. "A critique of recent quantitative and deep-structure modeling in capital structure research and beyond". *Critical Finance Review*. 2(1): 131–172.
- Welch, I. 2014. "Referee recommendations". *Review of Financial Studies*. 27(9): 2773–2804.