

## true economics

**True Economics is about original economic ideas and analysis concerning everyday events, news, policy views and their impact on the markets and you. Enjoy and engage!**

Friday, June 16, 2017

### 16/6/17: Replicating Scientific Research: Ugly Truth

Continuing with the theme on 'What I've been reading lately?', here is a smashing paper on 'accuracy' of empirical economic studies.

The paper, authored by Hou, Kewei and Xue, Chen and Zhang, Lu, and titled "Replicating Anomalies" (most recent version is from June 12, 2017, but it is also available in an earlier version via NBER) effectively blows a whistle on what is going on in empirical research in economics and finance. Per authors, the vast literature that detects financial markets anomalies (or deviations away from the efficient markets hypothesis / economic rationality) "is infested with widespread p-hacking".

What's p-hacking? Well, it's a shady practice whereby researchers manipulate (by selective inclusion or exclusion) sample criteria (which data points to exclude from estimation) and test procedures (including model specifications and selective reporting of favourable test results), until insignificant results become significant. In other words, under p-hacking, researchers attempt to superficially maximise model and explanatory variables significance, or, put differently, they attempt to achieve results that confirm their intuition or biases.

What's anomalies? Anomalies are departures in the markets (e.g. in share prices) from the predictions generated by the models consistent with rational expectations and the efficient markets hypothesis. In other words, anomalies occur when markets efficiency fails.

There are scores of anomalies detected in the academic literature, prompting many researchers to advocate abandonment (in all its forms, weak and strong) of the idea that markets are efficient.

Hou, Xue and Zhang take these anomalies to the test. The compile "a large data library with 447 anomalies". The authors then control for a key problem with data across many studies: microcaps. Microcaps - or small capitalization firms - are numerous in the markets (accounting for roughly 60% of all stocks), but represent only 3% of total market capitalization. This is true for key markets, such as NYSE, Amex and NASDAQ. Yet, as authors note, evidence shows that microcaps "not only have the highest equal-weighted returns, but also the largest cross-sectional standard deviations in returns and anomaly variables among microcaps, small stocks, and big stocks." In other words, these are higher risk, higher return class of securities. Yet, despite

### Disclaimer

This blog represents my personal views and is not reflective of the views or opinions held by any company, contractor, client or employer I work for currently or have worked for in the past. These views are not an endorsement to take any action in the markets or of any political position, figures or parties.

**My view on the future of economic development is outlined briefly in the TEDx talk I gave on [Human Capital and the Age of Change](#).**

### Twitter Updates



### Blog Archive

- ▼ **2017** (215)
  - ▶ **October** (30)
  - ▶ **September** (17)
  - ▶ **August** (17)
  - ▶ **July** (17)
  - ▼ **June** (22)
    - 30/6/17: Manning Financial: The End of the Trump B...**
    - 28/6/17: Tech Financing and NASDAQ: Divorce Procee...**
    - 28/6/17: Seattle's Minimum Wage Lessons for Califo...**
    - 27/6/17: Millennials' Support for Liberal Democrac...**
    - 22/6/17: Efficient Markets for H-bomb Fuel - 1954**
    - 22/6/17: Unwinding Monetary Excesses: FocusEconomi...**
    - 21/6/17: Azerbaijan Bank and Irish Saga of \$900 mi...**
    - 21/6/17: BusinessInsider: The Most Important Chart...**
    - 16/6/17: Replicating Scientific Research: Ugly Tru...**
    - 16/6/17: Trumpery & Knavery: New Paper on Washingt...**
    - 14/6/17: Unwinding the Mess: Fed's Road Map to Qun...**

this, "many studies overweight microcaps with equal-weighted returns, and often together with NYSE-Amex-NASDAQ breakpoints, in portfolio sorts." Worse, many (hundreds) of studies use 1970s regression technique that actually assigns more weight to microcaps. In simple terms, microcaps are the most common outlier and despite this they are given either same weight in analysis as non-outliers or their weight is actually elevated relative to normal assets, despite the fact that microcaps have little meaning in driving the actual markets (their weight in the total market is just about 3% in total).

So the study corrects for these problems and finds that, once microcaps are accounted for, the grand total of 286 anomalies (64% of all anomalies studied), and under more strict statistical significance test 380 (of 85% of all anomalies) "including 95 out of 102 liquidity variables (93%) are insignificant at the 5% level." In other words, the original studies claims that these anomalies were significant enough to warrant rejection of markets efficiency were not true when one recognizes one basic and simple problem with the data. Worse, per authors, "even for the 161 significant anomalies, their magnitudes are often much lower than originally reported. Among the 161, the q-factor model leaves 115 alphas insignificant (150 with  $t < 3$ )."

This is pretty damning for those of us who believe, based on empirical results published over the years, that markets are bounded-efficient, and it is outright savaging for those who claim that markets are perfectly inefficient. But, this tendency of researchers to silverplate statistics is hardly new.

Hou, Xue and Zhang provide a nice summary of research on p-hacking and non-replicability of statistical results across a range of fields. It is worth reading, because it dents significantly ones confidence in the quality of peer review and the quality of scientific research.

As the authors note, "in economics, Leamer (1983) exposes the fragility of empirical results to small specification changes, and proposes to "take the con out of econometrics" by reporting extensive sensitivity analysis to show how key results vary with perturbations in regression specification and in functional form." The latter call was never implemented in the research community.

"In an influential study, Dewald, Thursby, and Anderson (1986) attempt to replicate empirical results published at *Journal of Money, Credit, and Banking* [a top-tier journal], and find that inadvertent errors are so commonplace that the original results often cannot be reproduced."

"McCullough and Vinod (2003) report that nonlinear maximization routines from different software packages often produce very different estimates, and many articles published at *American Economic Review* [highest rated journal in economics] fail to test their solutions across different software packages."

"Chang and Li (2015) report a success rate of less than 50% from replicating 67 published papers from 13 economics journals, and Camerer et al. (2016) show a success rate of 61% from replicating 18 studies in experimental economics."

14/6/17: **The Fed: Bravely Going Somewhere Amidst R...**

13/6/17: **Unwinding the Mess: ECB vs Fed**

13/6/17: **Four Months of the Invisible Fiscal Disci...**

10/6/2017: **And the Ship [of Monetary Excesses] Sai...**

10/6/17: **Visualizing Cyber Security Attacks**

10/6/17: **Cart & Rails of the U.S. Monetary Policy**

7/6/17: **TrueEconomics Makes Top 100 List... again**

7/6/17: **European Policy Uncertainty: Still Above P...**

7/6/17: **Equity Markets Continue to Mis-price Polic...**

7/6/17: **Markets, Investors Exuberance and Fundamen...**

6/6/17: **Trump, Paris, Climate: The Problem is Bigg...**

▶ **May** (17)

▶ **April** (29)

▶ **March** (5)

▶ **February** (24)

▶ **January** (37)

▶ **2016** (293)

▶ **2015** (727)

▶ **2014** (724)

▶ **2013** (763)

▶ **2012** (584)

▶ **2011** (370)

▶ **2010** (382)

▶ **2009** (384)

▶ **2008** (10)

"It is not true that people stop pursuing dreams because they grow old, they grow old because they stop pursuing dreams." **Gabriel García Márquez**

**Nassim Nicholas Taleb** was asked whether public protests in Athens is a Black Swan Event. He replied: "No. The real Black Swan Event is that people are not rioting against the banks in London and New York."

"Getting worse more slowly is not the same as getting better", Prof. **Brad DeLong**

## A must-read

My thoughts on **Economics and Culture**: productive linkages and mutually-

"Collecting more than 50,000 tests published in *American Economic Review*, *Journal of Political Economy*, and *Quarterly Journal of Economics*, [three top rated journals in economics] Brodeur, L'e, Sangnier, and Zylberberg (2016) document a troubling two-humped pattern of test statistics. The pattern features a first hump with high p-values, a sizeable under-representation of p-values just above 5%, and a second hump with p-values slightly below 5%. The evidence indicates p-hacking that authors search for specifications that deliver just-significant results and ignore those that give just-insignificant results to make their work more publishable."

If you think this phenomena is encountered only in economics and finance, think again. Here are some findings from other 'hard science' disciplines where, you know, lab coats do not lie.

"...replication failures have been widely documented across scientific disciplines in the past decade. Fanelli (2010) reports that "positive" results increase down the hierarchy of sciences, with hard sciences such as space science and physics at the top and soft sciences such as psychology, economics, and business at the bottom. In oncology, Prinz, Schlange, and Asadullah (2011) report that scientists at Bayer fail to reproduce two thirds of 67 published studies. Begley and Ellis (2012) report that scientists at Amgen attempt to replicate 53 landmark studies in cancer research, but reproduce the original results in only six. Freedman, Cockburn, and Simcoe (2015) estimate the economic costs of irreproducible preclinical studies amount to about 28 billion dollars in the U.S. alone. In psychology, Open Science Collaboration (2015), which consists of about 270 researchers, conducts replications of 100 studies published in top three academic journals, and reports a success rate of only 36%."

Let's get down to real farce: everyone in sciences knows the above: "Baker (2016) reports that 80% of the respondents in a survey of 1,576 scientists conducted by Nature believe that there exists a reproducibility crisis in the published scientific literature. The surveyed scientists cover diverse fields such as chemistry, biology, physics and engineering, medicine, earth sciences, and others. More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than 50% have failed to reproduce their own experiments. Selective reporting, pressure to publish, and poor use of statistics are three leading causes."

Yeah, you get the idea: you need years of research, testing, re-testing and, more often than not, you get the results are not significant or weakly significant. Which means that after years of research you end up with unpublishable paper (no journal would welcome a paper without significant results, even though absence of evidence is as important in science as evidence of presence), no tenure, no job, no pension, no prospect of a career. So what do you do then? Ah, well... p-hack the shit out of data until the editor is happy and the referees are satisfied.

Which, for you, the reader, should mean the following: when we say that 'scientific research established fact A' based on reputable journals publishing high quality peer reviewed papers on the subject, know that around half of the findings claimed in these papers, on

reinforcing challenges

## My thoughts on European Federalism

### Economic Cycles: From Nanosecond Trading to Generational Oscillations

My presentation at TEDx Dublin: Human Capital and the Age of Change

Thomas Sargent on Everything You Need to Know About Economics in 297 Words

Ronald Coase: Saving Economics from Economists

Taleb on Euro crisis ([here](#))

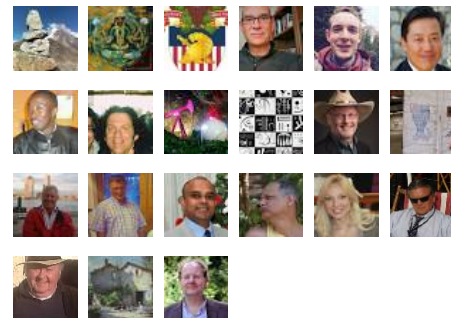
How influential are influential scientists? Brilliant research into the value of brilliance in research ([here](#))

When economics inspires artists - a must see ([here](#))

A fascinating article on mathematics of artistic beauty ([here](#))

## Followers

Followers (389) [Next](#)



[Follow](#)

## My Favorite Links

<http://financhill.com/>

Focus Economics

The Global Macro Monitor

OilPrice: Energy Economics

Karl Whelan's blog

BondWatch - Irish Banks Bonds

ReszatOnline - Risk & Financial Dynamics blog

Yanis Varoufakis

Irish Tax News

Cool mathematics blog

Solutions to Cool maths blog

A Credit Trader (US)

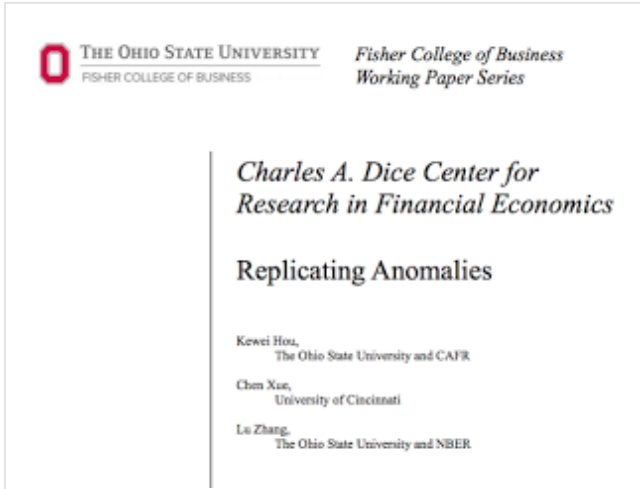
Crimson-Observer

Ferdinand von Prondzynski's blog

Gary Becker

average, most likely cannot be replicated or verified. And then remember, it takes one or two scientists to turn the world around from believing (based on scientific consensus at the time) that the Earth is flat and is the centre of the Universe, to believing in the world as we know it to be today.

Full link to the paper: Charles A. Dice Center Working Paper No. 2017-10; Fisher College of Business Working Paper No. 2017-03-010. Available at SSRN: <https://ssrn.com/abstract=2961979>.



Posted by **Constantin Gurdgiev** at **10:21 PM**

Reactions: boring (0) interesting (1) cool (1)



Labels: **anomalies**, **applied investment**, **efficient markets hypothesis**, **Finance**, **Investment Theory**, **markets efficiency**, **p-hacking**, **peer review**, **publishing**, **science**, **scientific publishing**, **Scientific research**

**George Jochnowitz**  
**Polyconomics**  
**Richard Delevan**  
**Roman Frydman: the 'Imperfect Knowledge Man'**  
**The Dangerous Economics Blog**  
**Tim Duy Fed Watch**  
**Turbulence Ahead**  
**Urbanophile**  
**Wise Up Journal**

## About Me



**Constantin Gurdgiev**

**G+** Follow 541

**View my complete profile**

## Academic Publishing

**My academic output:**

My SSRN webpage: **here**