Big Data and Bigger Firms: A Labor Market Channel

Discussion!

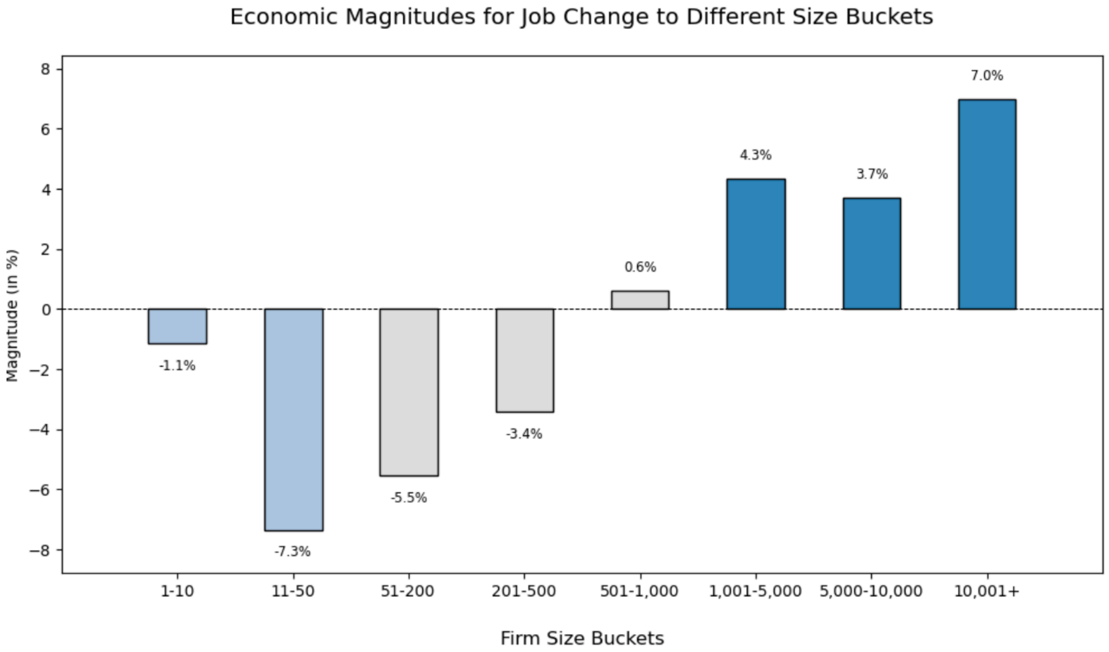
Paul Goldsmith-Pinkham

Yale SOM & NBER

2024-12-13

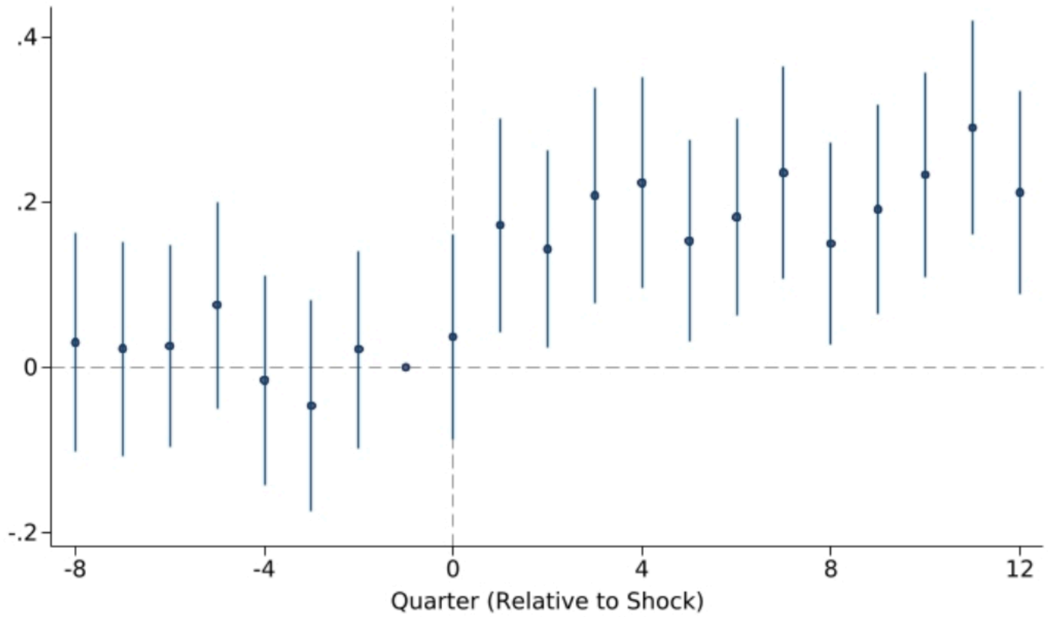
Key takeaways

- Following change in Github policy, workers with more private repo contributions moved more to big firms



Key takeaways

- Following change in Github policy, workers with more private repo contributions moved more to big firms
- Effect is sharp and immediate
- Also big!



Probability of moving to large firm (1000+ employees) (p.p.)

$\beta = 0.0017, \overline{y} = 0.03 \longrightarrow 5\%$

What is done in the paper

Econometrician:

1. Measure individuals' (noisy) productivity, private and public
2. Examine how these individuals move firms following revelation of public information
3. How does this vary across types of firm?
4. How does this vary within firm?

What is done in the paper

Econometrician:

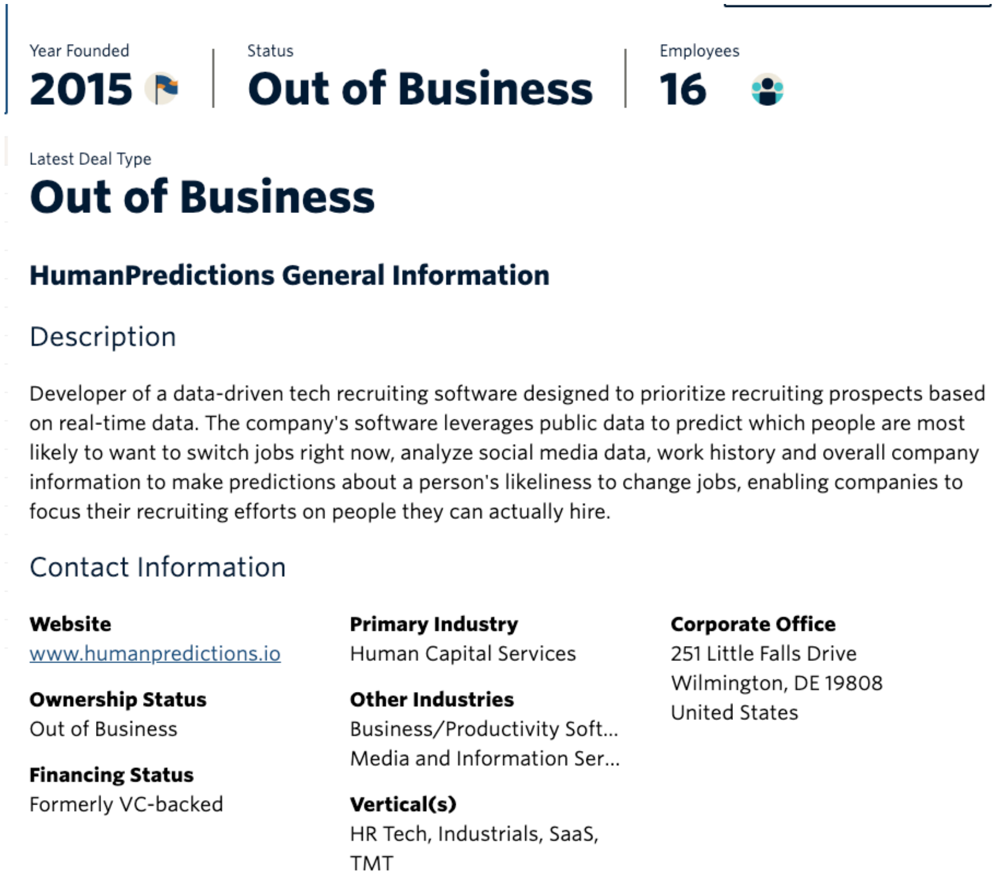
1. Measure individuals' (noisy) productivity, private and public
2. Examine how these individuals move firms following revelation of public information
3. How does this vary across types of firm?
4. How does this vary within firm?

Economic agent:

- Worker: public and private information
- Own-Firm: Sees public + private information
- Other-Firm: Sees public information
  - Following policy change, see more private info

My takeaways

- Qualitative fact showing that workers with more private contributions move to larger firms seems true to me
- Fascinating data
- Interesting application of AKM approach



Year Founded
2015

Status
Out of Business

Employees
16

Latest Deal Type
Out of Business

HumanPredictions General Information

Description

Developer of a data-driven tech recruiting software designed to prioritize recruiting prospects based on real-time data. The company's software leverages public data to predict which people are most likely to want to switch jobs right now, analyze social media data, work history and overall company information to make predictions about a person's likeliness to change jobs, enabling companies to focus their recruiting efforts on people they can actually hire.

Contact Information

Website
www.humanpredictions.io

Ownership Status
Out of Business

Financing Status
Formerly VC-backed

Primary Industry
Human Capital Services

Other Industries
Business/Productivity Soft...
Media and Information Ser...

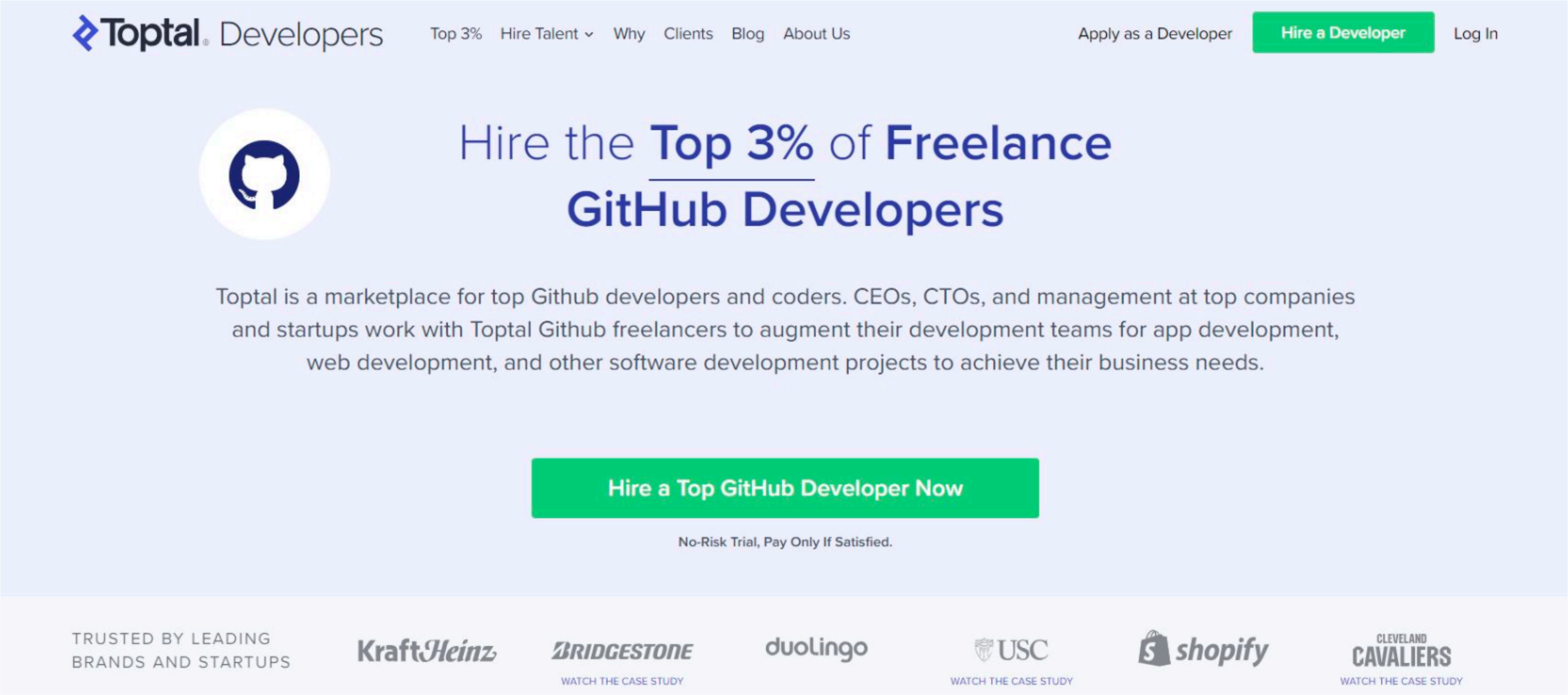
Vertical(s)
HR Tech, Industrials, SaaS, TMT

Corporate Office
251 Little Falls Drive
Wilmington, DE 19808
United States

My takeaways

- Is this a sufficiently good measure of productivity?
  - Correct interpretation may be qualitative, rather than quantitative
  - And that's fine! Not obvious at all this was true

Not obvious this story was true

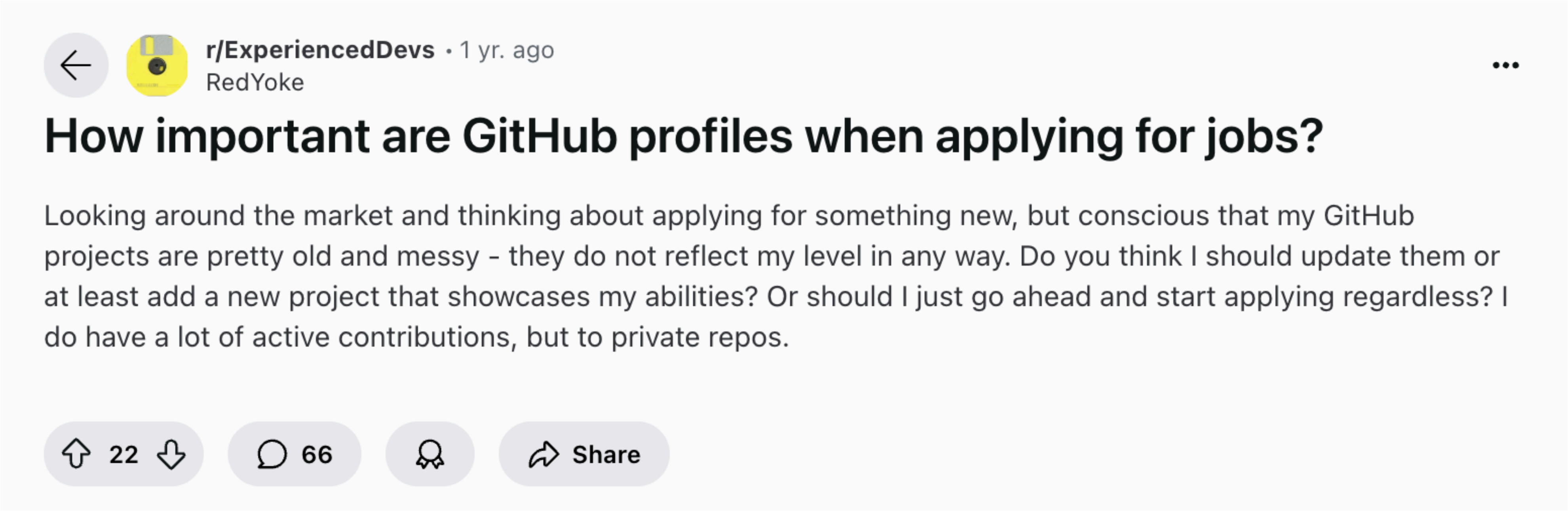


Hire the Top 3% of Freelance GitHub Developers

Toptal is a marketplace for top Github developers and coders. CEOs, CTOs, and management at top companies and startups work with Toptal Github freelancers to augment their development teams for app development, web development, and other software development projects to achieve their business needs.

No-Risk Trial, Pay Only If Satisfied.

Not obvious this story was true

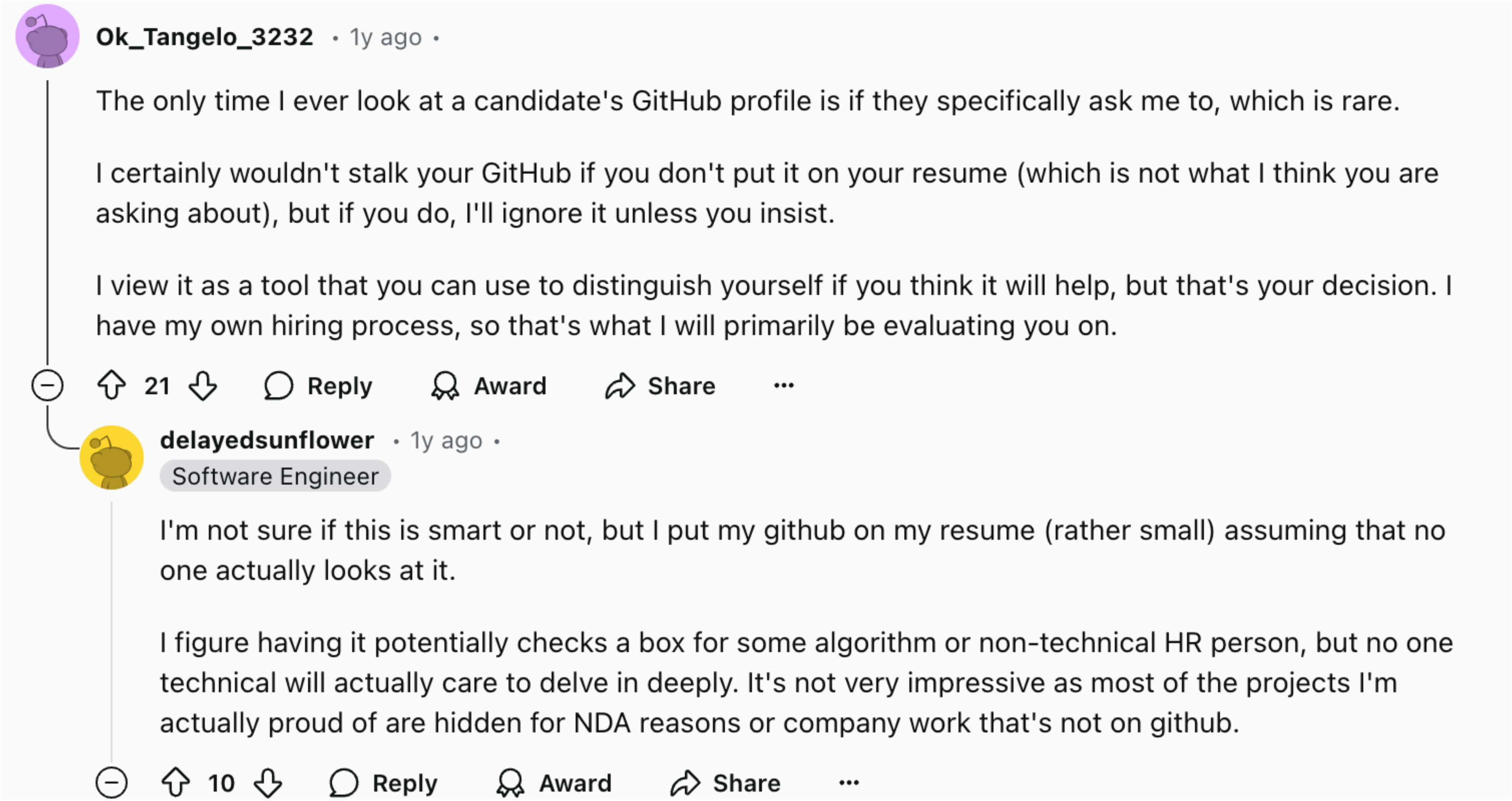


r/ExperiencedDevs • 1 yr. ago
RedYoke

How important are GitHub profiles when applying for jobs?

Looking around the market and thinking about applying for something new, but conscious that my GitHub projects are pretty old and messy - they do not reflect my level in any way. Do you think I should update them or at least add a new project that showcases my abilities? Or should I just go ahead and start applying regardless? I do have a lot of active contributions, but to private repos.

Not obvious this story was true



Ok_Tangelo_3232 • 1y ago •

The only time I ever look at a candidate's GitHub profile is if they specifically ask me to, which is rare.

I certainly wouldn't stalk your GitHub if you don't put it on your resume (which is not what I think you are asking about), but if you do, I'll ignore it unless you insist.

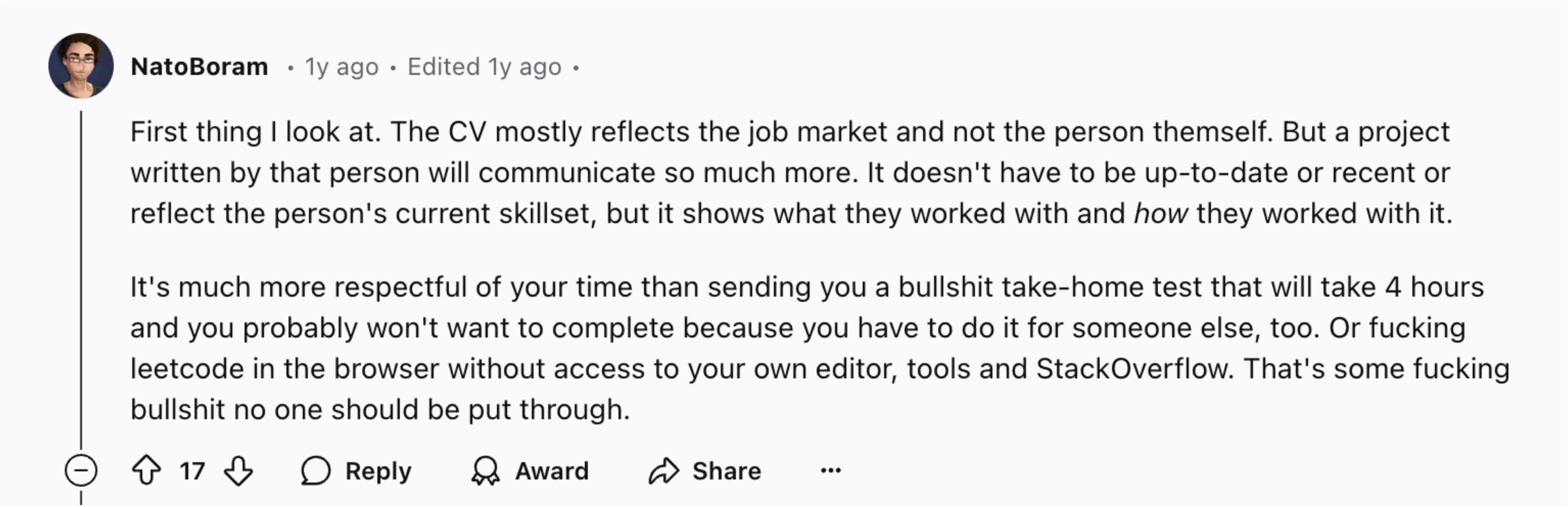
I view it as a tool that you can use to distinguish yourself if you think it will help, but that's your decision. I have my own hiring process, so that's what I will primarily be evaluating you on.

delayedsunflower • 1y ago •
Software Engineer

I'm not sure if this is smart or not, but I put my github on my resume (rather small) assuming that no one actually looks at it.

I figure having it potentially checks a box for some algorithm or non-technical HR person, but no one technical will actually care to delve in deeply. It's not very impressive as most of the projects I'm actually proud of are hidden for NDA reasons or company work that's not on github.

Not obvious this story was true

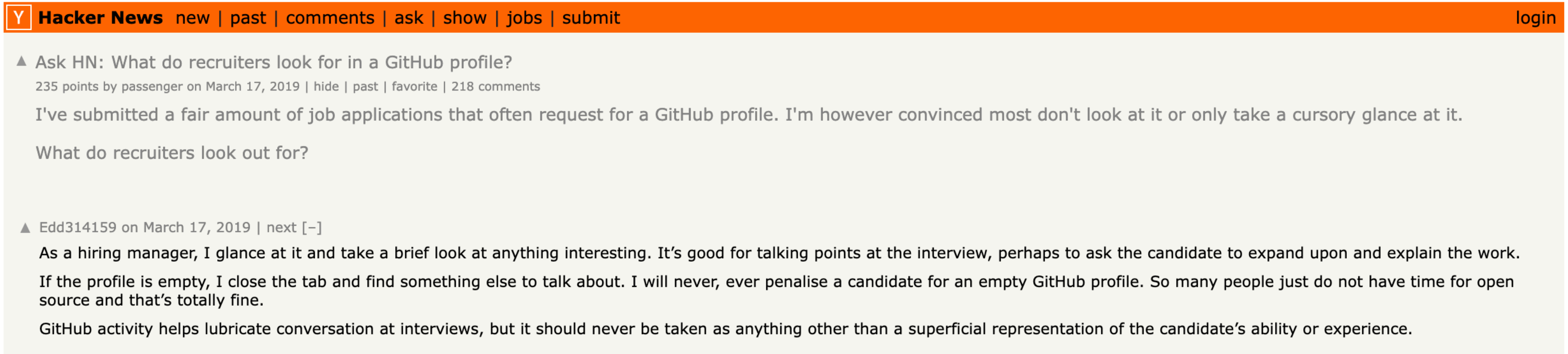


NatoBoram • 1y ago • Edited 1y ago •

First thing I look at. The CV mostly reflects the job market and not the person themself. But a project written by that person will communicate so much more. It doesn't have to be up-to-date or recent or reflect the person's current skillset, but it shows what they worked with and *how* they worked with it.

It's much more respectful of your time than sending you a bullshit take-home test that will take 4 hours and you probably won't want to complete because you have to do it for someone else, too. Or fucking leetcode in the browser without access to your own editor, tools and StackOverflow. That's some fucking bullshit no one should be put through.

Not obvious this story was true



Y Hacker News new | past | comments | ask | show | jobs | submit
login

▲ Ask HN: What do recruiters look for in a GitHub profile?

235 points by passenger on March 17, 2019 | hide | past | favorite | 218 comments

I've submitted a fair amount of job applications that often request for a GitHub profile. I'm however convinced most don't look at it or only take a cursory glance at it.

What do recruiters look out for?

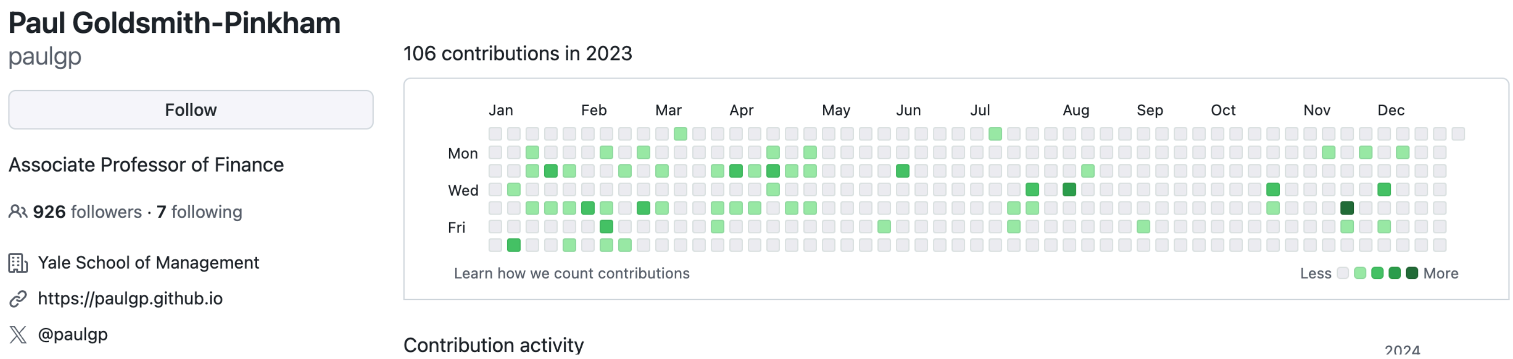
▲ Edd314159 on March 17, 2019 | next [–]

As a hiring manager, I glance at it and take a brief look at anything interesting. It’s good for talking points at the interview, perhaps to ask the candidate to expand upon and explain the work.

If the profile is empty, I close the tab and find something else to talk about. I will never, ever penalise a candidate for an empty GitHub profile. So many people just do not have time for open source and that’s totally fine.

GitHub activity helps lubricate conversation at interviews, but it should never be taken as anything other than a superficial representation of the candidate’s ability or experience.

Github policy change in the wild



Paul Goldsmith-Pinkham

paulgp

Associate Professor of Finance

926 followers · 7 following

Yale School of Management

https://paulgp.github.io

@paulgp

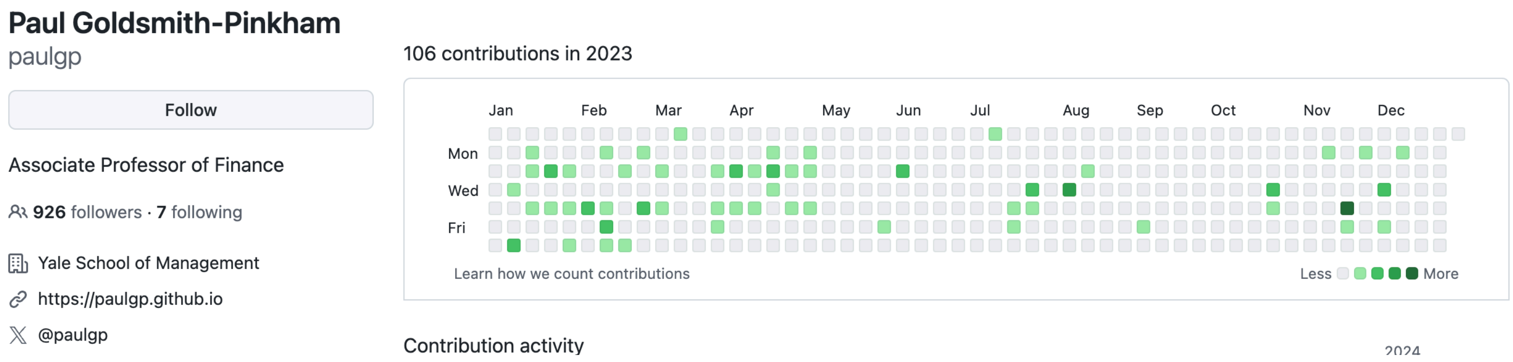
106 contributions in 2023

Learn how we count contributions

Contribution activity

2024

Github policy change in the wild



Paul Goldsmith-Pinkham

paulgp

Associate Professor of Finance

926 followers · 7 following

Yale School of Management

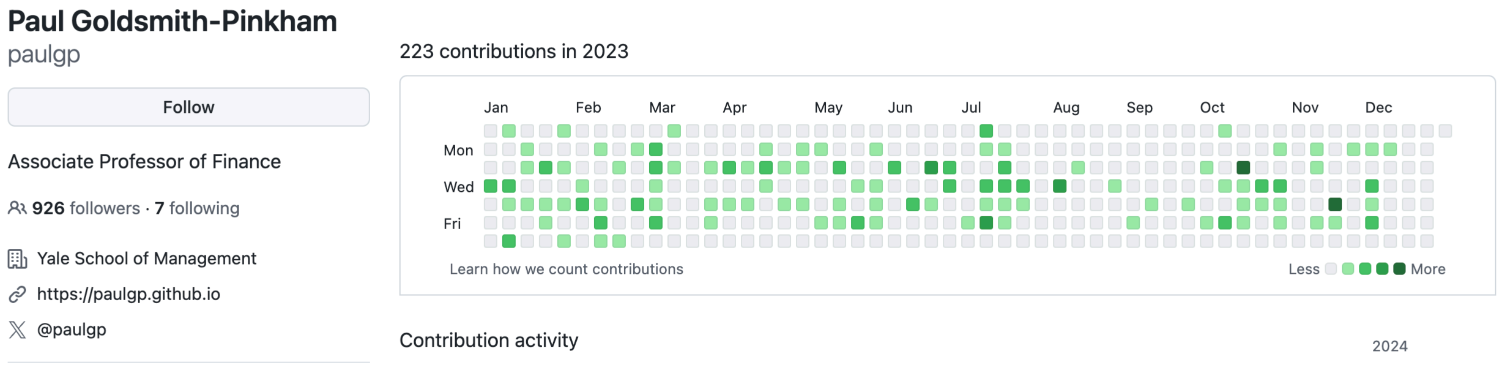
https://paulgp.github.io

@paulgp

106 contributions in 2023

Learn how we count contributions

Contribution activity



Paul Goldsmith-Pinkham

paulgp

Associate Professor of Finance

926 followers · 7 following

Yale School of Management

https://paulgp.github.io

@paulgp

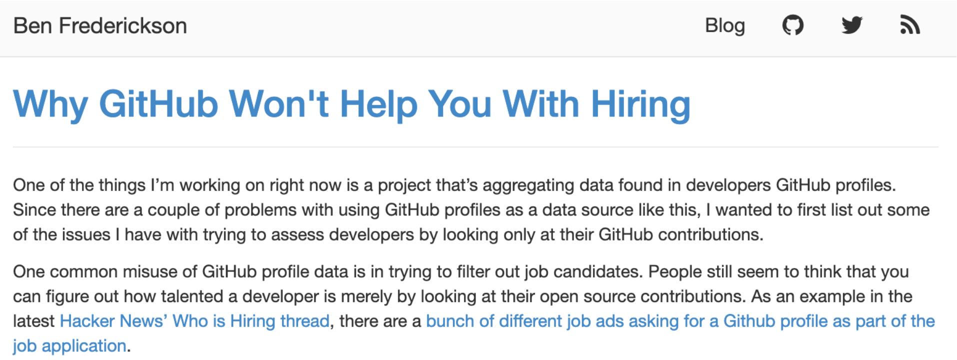
223 contributions in 2023

Learn how we count contributions

Contribution activity

Github policy change was not obvious to me

- My default setting was to have private contributions hidden
- I’m not alone in my ignorance



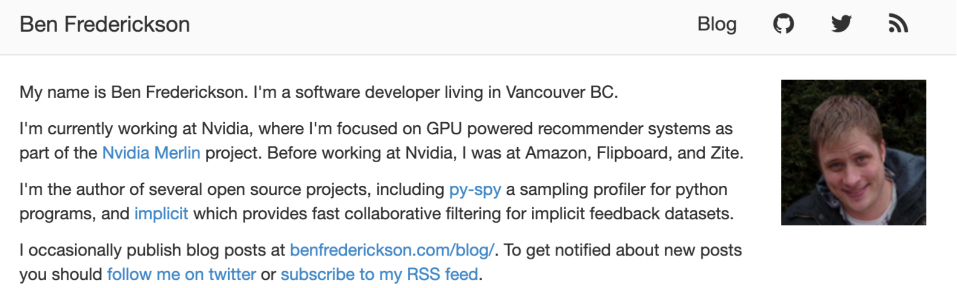
Ben Frederickson   Blog

Why GitHub Won't Help You With Hiring

One of the things I’m working on right now is a project that’s aggregating data found in developers GitHub profiles. Since there are a couple of problems with using GitHub profiles as a data source like this, I wanted to first list out some of the issues I have with trying to assess developers by looking only at their GitHub contributions.

One common misuse of GitHub profile data is in trying to filter out job candidates. People still seem to think that you can figure out how talented a developer is merely by looking at their open source contributions. As an example in the latest Hacker News’ Who is Hiring thread, there are a bunch of different job ads asking for a Github profile as part of the job application.

Published on 08 March 2018



Ben Frederickson   Blog

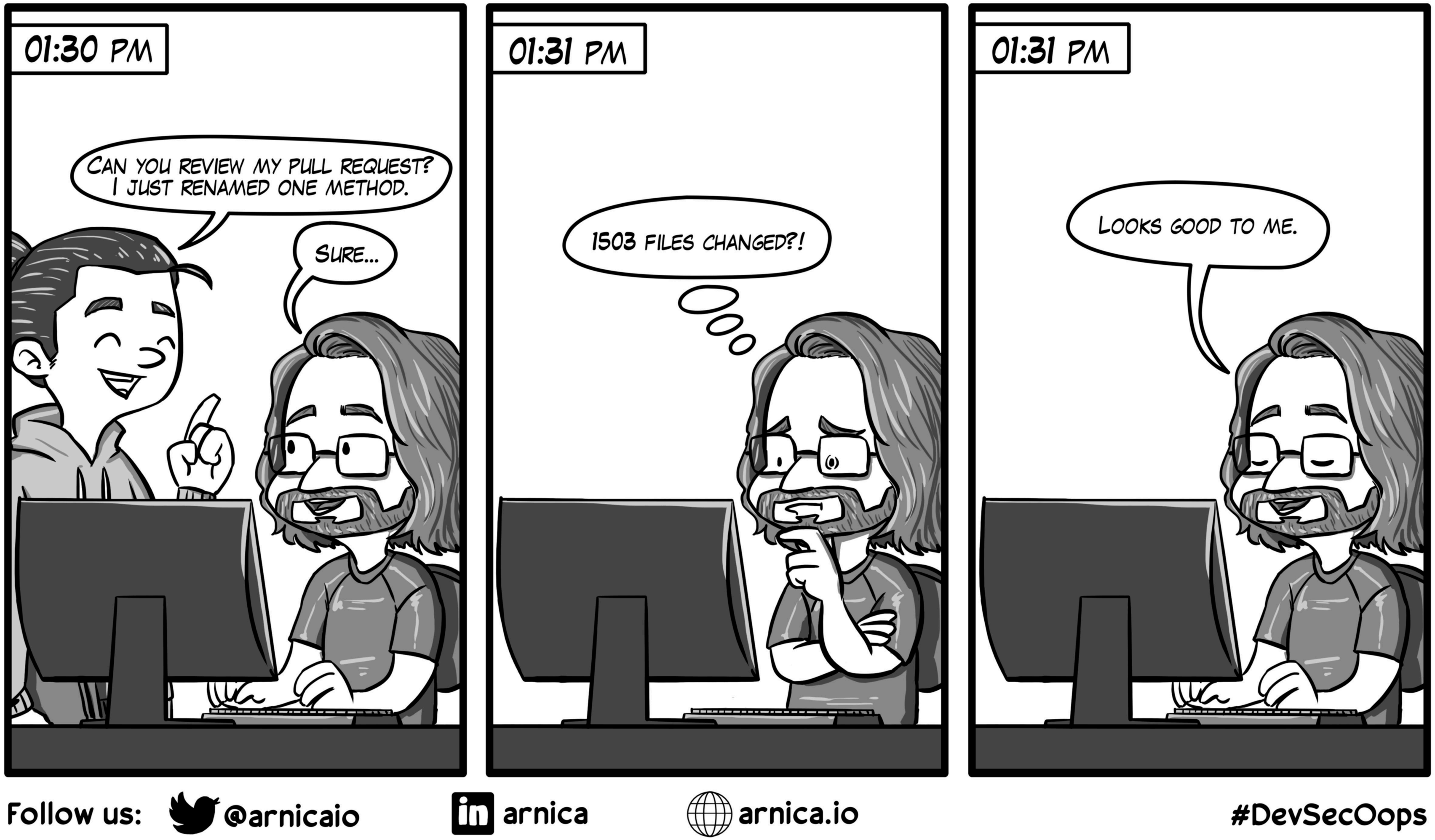
My name is Ben Frederickson. I'm a software developer living in Vancouver BC.

I'm currently working at Nvidia, where I'm focused on GPU powered recommender systems as part of the Nvidia Merlin project. Before working at Nvidia, I was at Amazon, Flipboard, and Zite.

I'm the author of several open source projects, including py-spy a sampling profiler for python programs, and implicit which provides fast collaborative filtering for implicit feedback datasets.

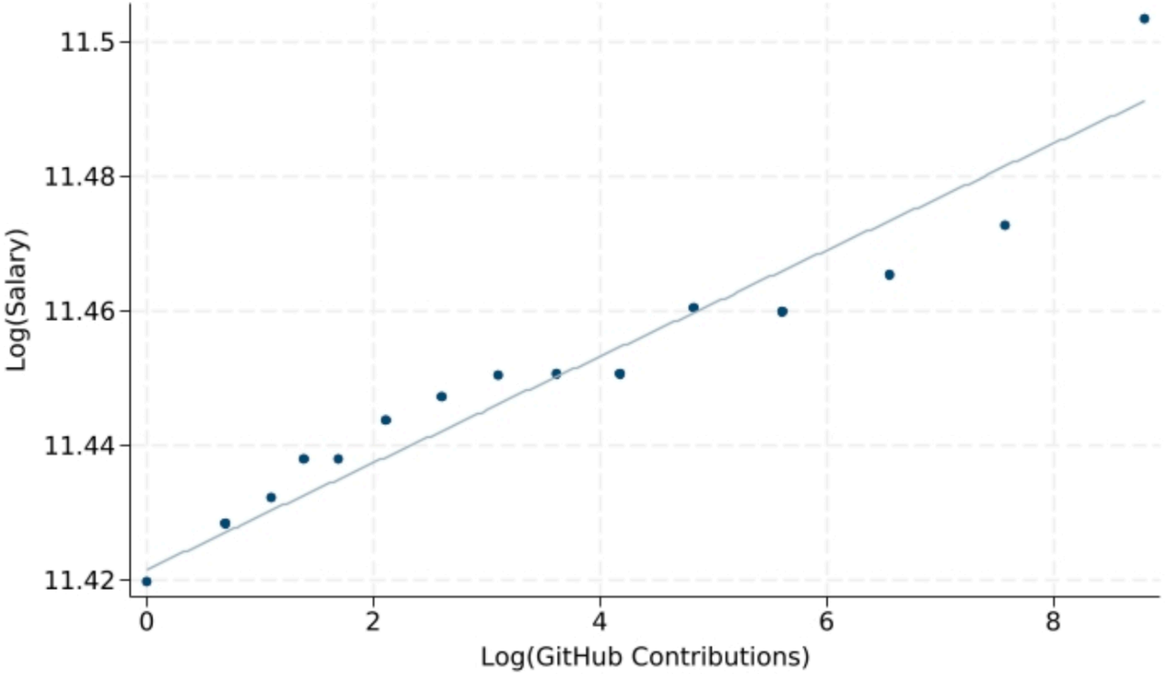
I occasionally publish blog posts at benfrederickson.com/blog/. To get notified about new posts you should follow me on twitter or subscribe to my RSS feed.

Does more output mean you're more productive?



But that doesn’t mean it’s not a signal!

- Verification of signal: elasticity of salary with total Github contributions
- $\varepsilon = 0.01$
- Would love to know this with private vs. public
- Where is income from?



Empirical approach

Abowd Kramarz Margolis (1999) [AKM] approach, but using productivity instead of wages:

$$P_{\{i,f,t\}} = \alpha_i + \alpha_f + \alpha_t + X_i \beta$$

where P is log of private contributions – the private signal of productivity. Then, standard dynamics DiD:

$$Y = \pi_{\{i\}} + X_i \gamma + \sum \gamma_s \theta_s \times D + \varepsilon_{\{i,t\}}$$

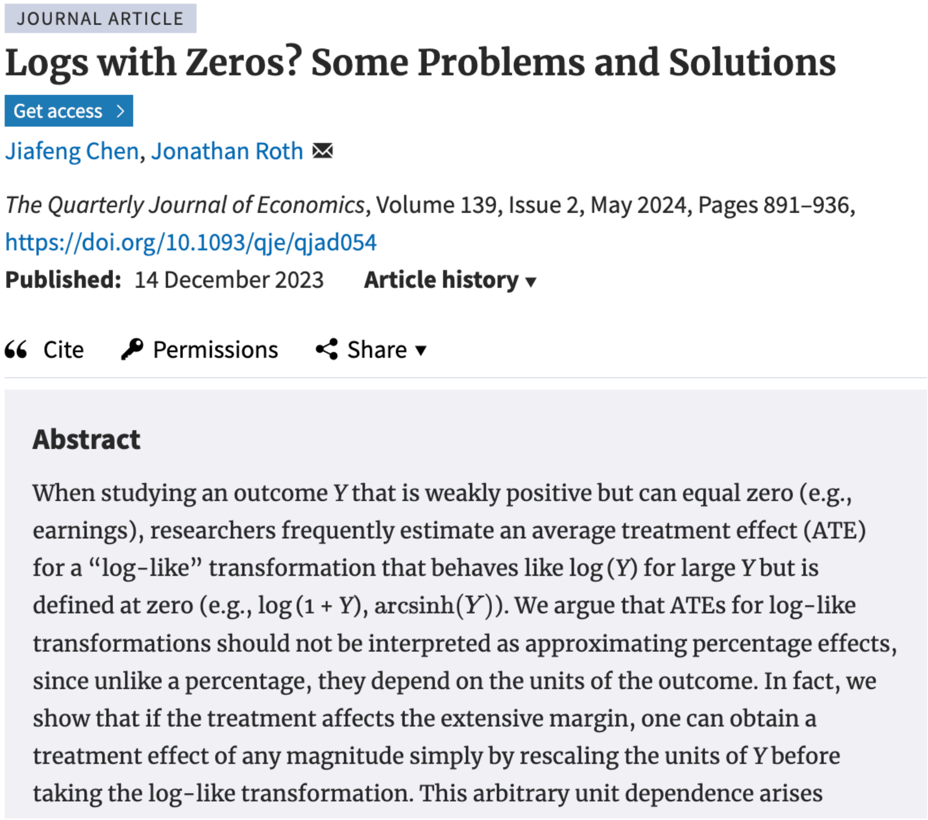
Unpacking the empirical approach

First, let's understand the AKM approach.

1. The main focus of outcome is log(private contributions + 1)
2. We are focusing in on one clean productivity measure (but highly noisy) α_i
   - Important question on what this captures

A brief journey into log(1+Y) regressions

- Log(1+Private Contributions) is used because contribtuions are:
  - Zero-inflated
  - Right-skewed
- Historically common approach but some issues on interpretation
  - Authors aware of this and discuss
  - But I think insufficiently appreciative!



JOURNAL ARTICLE

Logs with Zeros? Some Problems and Solutions

Jiafeng Chen, Jonathan Roth ✉

The Quarterly Journal of Economics, Volume 139, Issue 2, May 2024, Pages 891–936, https://doi.org/10.1093/qje/qjad054

Published: 14 December 2023 **Article history** ▾

Abstract

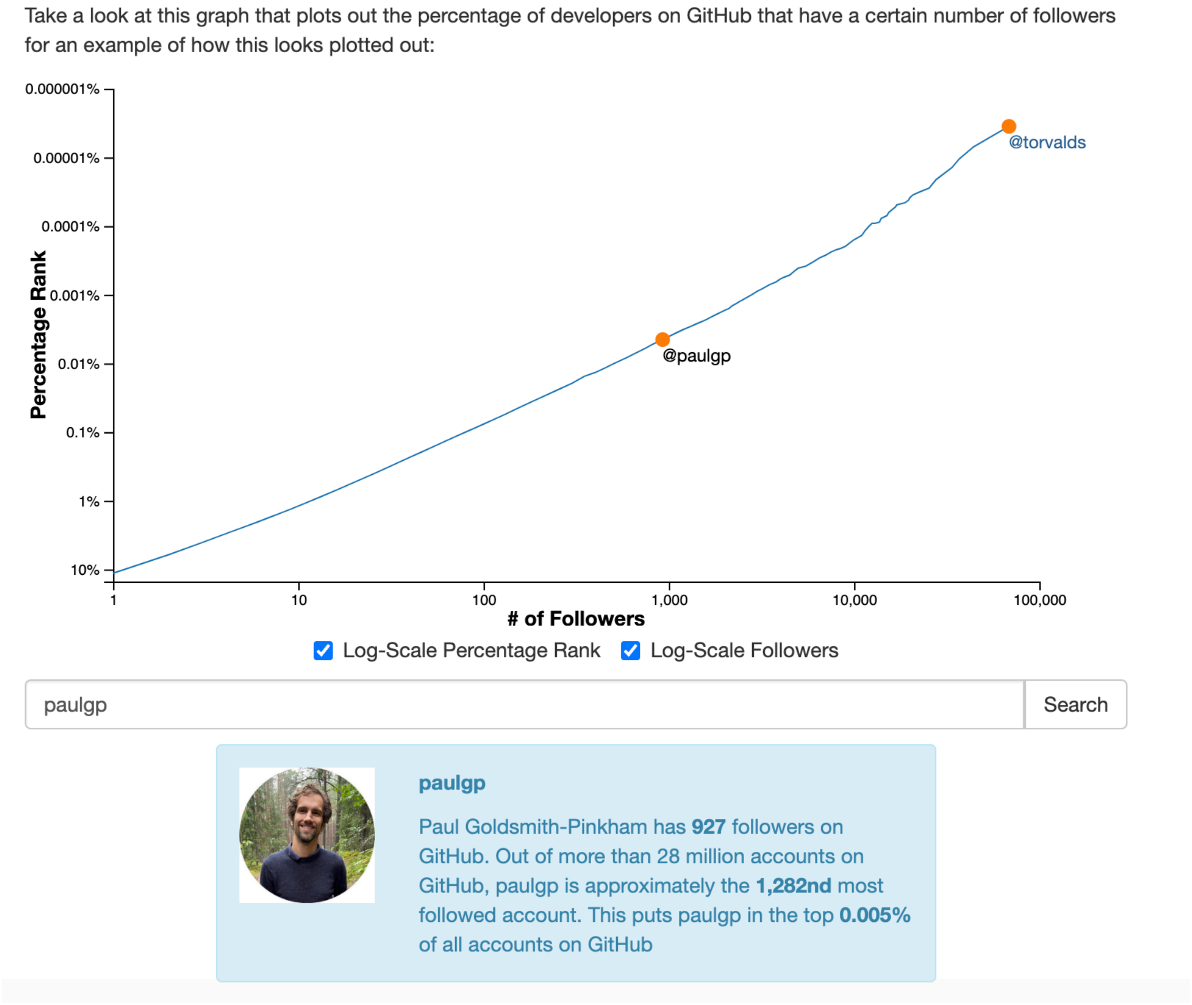
When studying an outcome Y that is weakly positive but can equal zero (e.g., earnings), researchers frequently estimate an average treatment effect (ATE) for a "log-like" transformation that behaves like $\log(Y)$ for large Y but is defined at zero (e.g., $\log(1 + Y)$, $\text{arcsinh}(Y)$). We argue that ATEs for log-like transformations should not be interpreted as approximating percentage effects, since unlike a percentage, they depend on the units of the outcome. In fact, we show that if the treatment affects the extensive margin, one can obtain a treatment effect of any magnitude simply by rescaling the units of Y before taking the log-like transformation. This arbitrary unit dependence arises

A brief journey into log(1+Y) regressions

Variable	Mean	Std. Dev.	P10	Median	P90
Total Log Contributions	2.60	2.16	0	2.08	5.87
Public Log Contributions	2.35	1.91	0	1.95	5.17
Private Log Contributions	0.66	1.86	0	0	3.22

- Implication: over 50% of observations have zero private contributions
- Key insight from Chen and Roth (2024):
  - put much more weight on the **extensive** margin
- Extensive margin changes will occur because of product shift, or firm shift
  - Measurement approach is not time-varying (snapshot)

Like most of the internet, there’s a lot of emptiness



Take a look at this graph that plots out the percentage of developers on GitHub that have a certain number of followers for an example of how this looks plotted out:

☑ Log-Scale Percentage Rank ☑ Log-Scale Followers

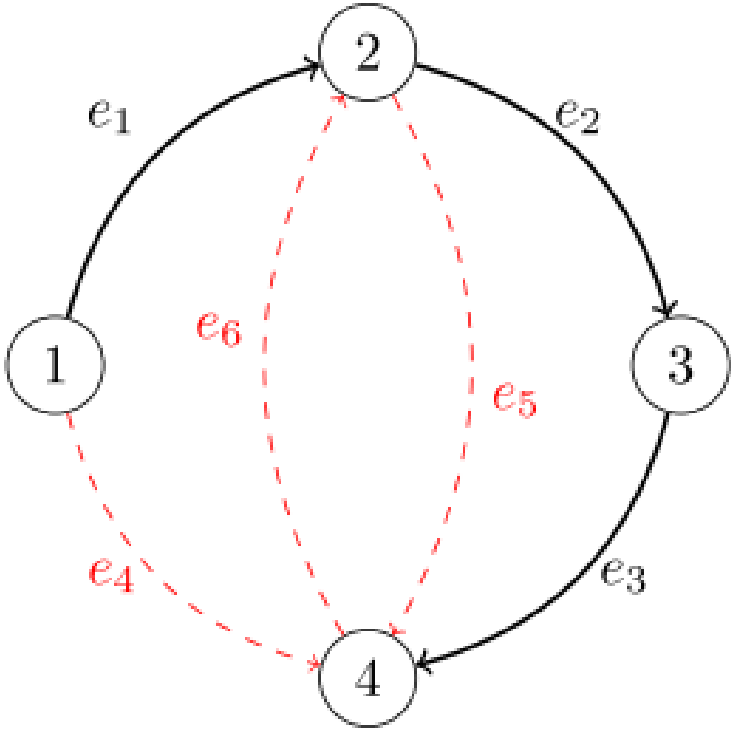
paulgp | Search

paulgp

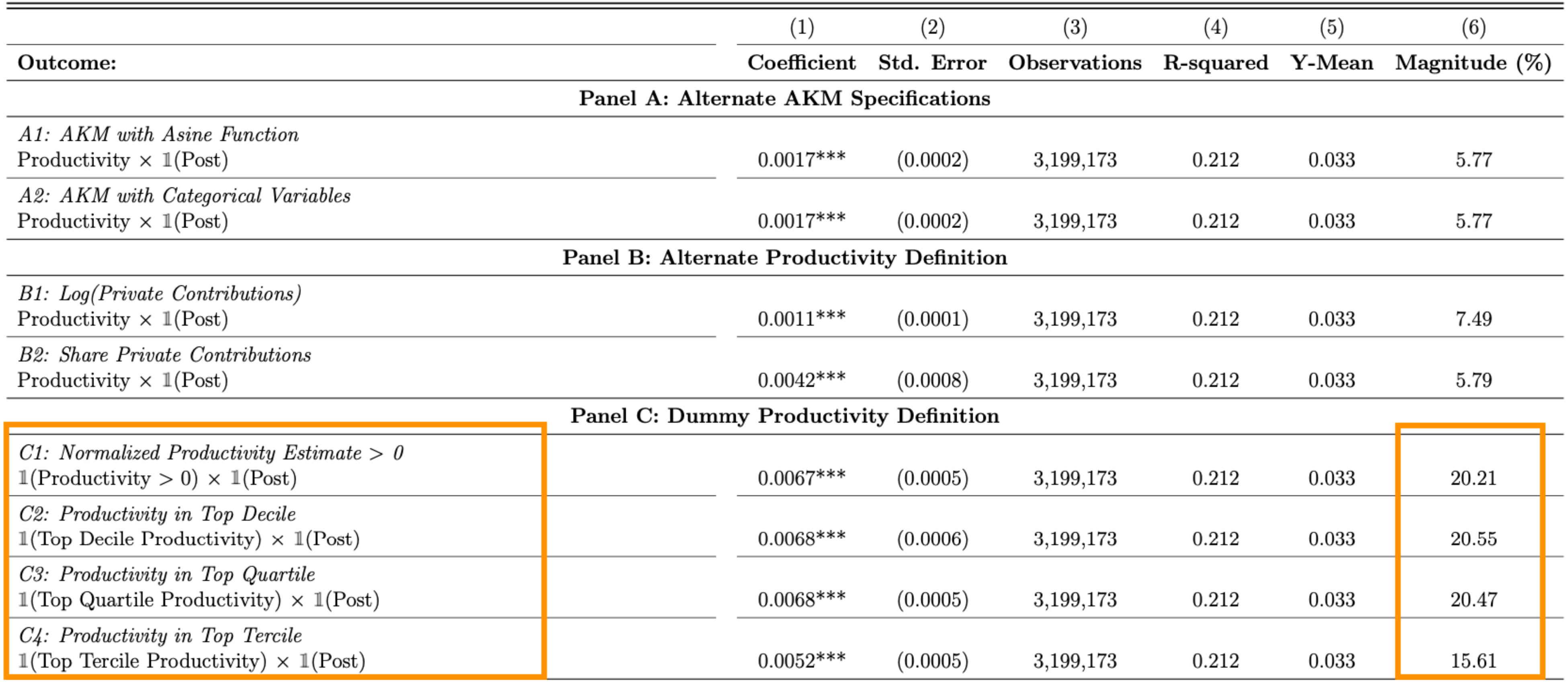
Paul Goldsmith-Pinkham has **927** followers on GitHub. Out of more than 28 million accounts on GitHub, paulgp is approximately the **1,282nd** most followed account. This puts paulgp in the top **0.005%** of all accounts on GitHub

Mapping this back in to AKM

- AKM approach: firm effects are identified by works moving firms
  - Intuition: consider a network of firms and workers
- Identified effects in pre-period are estimated off of changes
- AKM typically done with log(wages)
  - I think $\log(1 + Y)$ will exacerbate the extensive margin effect

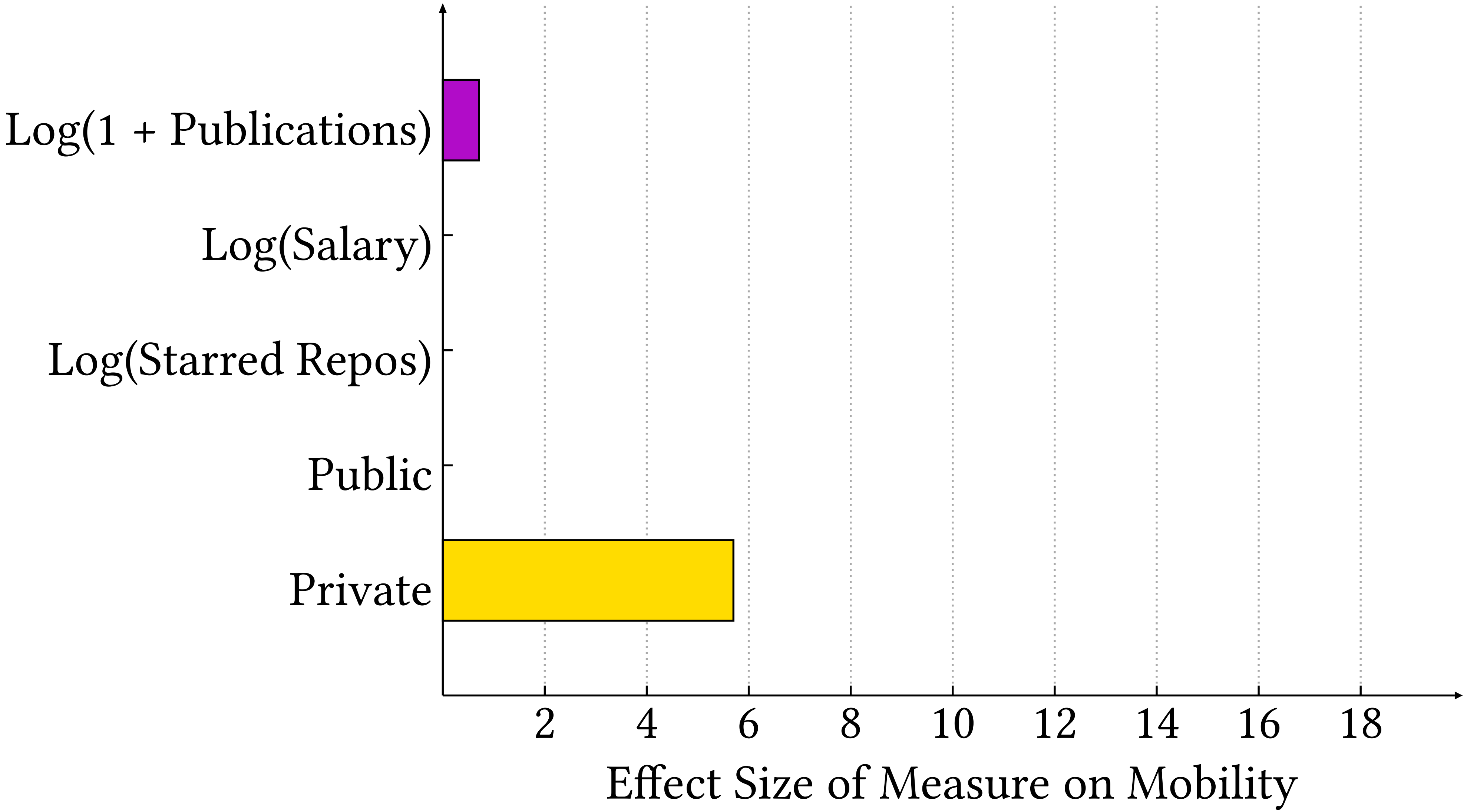


Authors address point in Appendix Table A.3



	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	**Coefficient**	**Std. Error**	**Observations**	**R-squared**	**Y-Mean**	**Magnitude (%)**
Panel A: Alternate AKM Specifications						
A1: AKM with Asine Function Productivity × $\mathbb{1}$(Post)	0.0017***	(0.0002)	3,199,173	0.212	0.033	5.77
A2: AKM with Categorical Variables Productivity × $\mathbb{1}$(Post)	0.0017***	(0.0002)	3,199,173	0.212	0.033	5.77
Panel B: Alternate Productivity Definition						
B1: Log(Private Contributions) Productivity × $\mathbb{1}$(Post)	0.0011***	(0.0001)	3,199,173	0.212	0.033	7.49
B2: Share Private Contributions Productivity × $\mathbb{1}$(Post)	0.0042***	(0.0008)	3,199,173	0.212	0.033	5.79
Panel C: Dummy Productivity Definition						
C1: Normalized Productivity Estimate > 0 $\mathbb{1}$(Productivity > 0) × $\mathbb{1}$(Post)	0.0067***	(0.0005)	3,199,173	0.212	0.033	20.21
C2: Productivity in Top Decile $\mathbb{1}$(Top Decile Productivity) × $\mathbb{1}$(Post)	0.0068***	(0.0006)	3,199,173	0.212	0.033	20.55
C3: Productivity in Top Quartile $\mathbb{1}$(Top Quartile Productivity) × $\mathbb{1}$(Post)	0.0068***	(0.0005)	3,199,173	0.212	0.033	20.47
C4: Productivity in Top Tercile $\mathbb{1}$(Top Tercile Productivity) × $\mathbb{1}$(Post)	0.0052***	(0.0005)	3,199,173	0.212	0.033	15.61

Placebo test suggests it’s really private contributions



So how should we interpret this?

- I interpret results as extensive measures of workers who have (visible) private repo contributions
- Contrast these workers with workers who have similar movements through firm network but no (visible) private contributions
- These workers far more likely to move to big firms
- I'm convinced that this policy definitely shifted workers
  - Important qualitative test of how information can be used by workers vs. firms

Big takeaways in economics

- Labor and corporate finance take on "owning your data"
- Can be very valuable for workers to demonstrate their value
- What would be knock-on impacts to smaller firms?
  - Acemoglu and Pischke (1999)
- Labor risk is serious for firms!