The Private Value of Open-Source Innovation

Logan Emery[1] Chan Lim[2] Shiwei Ye[1]

[1]Rotterdam School of Management

[2]University at Buffalo, SUNY

19 May, 2025

Contents

Introduction

- Since Schumpeter, innovation has been recognized as a driver of economic growth.

Introduction

- Since Schumpeter, innovation has been recognized as a driver of economic growth.
- Costly to create, but non-rival in use.

Introduction

- Since Schumpeter, innovation has been recognized as a driver of economic growth.
- Costly to create, but non-rival in use.
- To encourage investment, conventional systems (e.g., patents) grant exclusive rights to monetize innovation.
  - This excludability is seen as the main source of private value.

Introduction

- Since Schumpeter, innovation has been recognized as a driver of economic growth.
- Costly to create, but non-rival in use.
- To encourage investment, conventional systems (e.g., patents) grant exclusive rights to monetize innovation.
  - This excludability is seen as the main source of private value.
- Open-source innovation lacks legal excludability, yet corporate investment is growing.
  - 90% of Fortune 100 companies use GitHub (GitHub Report, 2022).

Introduction

- Since Schumpeter, innovation has been recognized as a driver of economic growth.
- Costly to create, but non-rival in use.
- To encourage investment, conventional systems (e.g., patents) grant exclusive rights to monetize innovation.
  - This excludability is seen as the main source of private value.
- Open-source innovation lacks legal excludability, yet corporate investment is growing.
  - 90% of Fortune 100 companies use GitHub (GitHub Report, 2022).
- The decision to voluntarily freely release valuable intellectual property to the public appears puzzling (Lerner et al., 2006).

This paper

- We investigate open-source firms, the private value generated by their innovation, and it's association with firm growth.

This paper

- We investigate open-source firms, the private value generated by their innovation, and it's association with firm growth.
- Previous studies suggest private value is derived through various channels:
  - Broad adoption & network effects
  - Community benefits
  - Competition concerns

This paper

- We investigate open-source firms, the private value generated by their innovation, and it's association with firm growth.
- Previous studies suggest private value is derived through various channels:
  - Broad adoption & network effects
  - Community benefits
  - Competition concerns
- New data on open-source activities on GitHub creates new opportunities:
  - We compile a comprehensive project-level dataset of public firms' GitHub activity during 2015–2023.
  - Provide a new measure of the economic value of open-source innovations based on stock market reactions to project release (Kogan et al., 2017).

Results

- Firms active on GitHub account for a large portion of the U.S. public market and span a broad set of industries.

Results

- Firms active on GitHub account for a large portion of the U.S. public market and span a broad set of industries.
- Open-source firms are large, innovative, and face less competition. Most differences are absorbed by firm fixed effects.

Results

- Firms active on GitHub account for a large portion of the U.S. public market and span a broad set of industries.
- Open-source firms are large, innovative, and face less competition. Most differences are absorbed by firm fixed effects.
- The average private value of a project is $842,849, totaling nearly $25 billion, driven by the software industry and AI-related technologies.

Results

- Firms active on GitHub account for a large portion of the U.S. public market and span a broad set of industries.
- Open-source firms are large, innovative, and face less competition. Most differences are absorbed by firm fixed effects.
- The average private value of a project is $842,849, totaling nearly $25 billion, driven by the software industry and AI-related technologies.
- Competition and licensing excludability are important drivers of private value, but complementarity with commercialized products is not.

Results

- Firms active on GitHub account for a large portion of the U.S. public market and span a broad set of industries.
- Open-source firms are large, innovative, and face less competition. Most differences are absorbed by firm fixed effects.
- The average private value of a project is $842,849, totaling nearly $25 billion, driven by the software industry and AI-related technologies.
- Competition and licensing excludability are important drivers of private value, but complementarity with commercialized products is not.
- Releasing valuable projects is associated with substantial firm growth but also creative destruction.

Literature

- Valuing innovation
  - R&D (Sougiannis, 1994; Lev and Sougiannis, 1996) Patents (Kogan et al., 2017; Chen et al., 2019); Drugs (Aryal et al., 2022); Trademarks (Desai et al., 2023); Open source (Greenstein and Nagle, 2014; Murciano-Goroff et al., 2021; Robbins et al., 2021; Blind et al., 2021; Hoffmann et al., 2024)

    This paper estimates the *private* value of individual repositories using financial markets

Literature

- Valuing innovation
  - R&D (Sougiannis, 1994; Lev and Sougiannis, 1996) Patents (Kogan et al., 2017; Chen et al., 2019); Drugs (Aryal et al., 2022); Trademarks (Desai et al., 2023); Open source (Greenstein and Nagle, 2014; Murciano-Goroff et al., 2021; Robbins et al., 2021; Blind et al., 2021; Hoffmann et al., 2024)

    This paper estimates the *private* value of individual repositories using financial markets

- Sources of private value generated by open-source innovation
  - Lerner and Tirole (2002) discusses channels including complementarity, talent acquisition, reputation, cost reduction, and inherent excludability, while follow-up research empirically explores these incentives (e.g., Alexy et al. (2018), Nagle (2018), Lin and Maruping (2022))

    This paper finds significant value for inherent excludability, less value for complementarity.

Literature

- Valuing innovation
  - R&D (Sougiannis, 1994; Lev and Sougiannis, 1996) Patents (Kogan et al., 2017; Chen et al., 2019); Drugs (Aryal et al., 2022); Trademarks (Desai et al., 2023); Open source (Greenstein and Nagle, 2014; Murciano-Goroff et al., 2021; Robbins et al., 2021; Blind et al., 2021; Hoffmann et al., 2024)

    This paper estimates the *private* value of individual repositories using financial markets

- Sources of private value generated by open-source innovation
  - Lerner and Tirole (2002) discusses channels including complementarity, talent acquisition, reputation, cost reduction, and inherent excludability, while follow-up research empirically explores these incentives (e.g., Alexy et al. (2018), Nagle (2018), Lin and Maruping (2022))

    This paper finds significant value for inherent excludability, less value for complementarity.

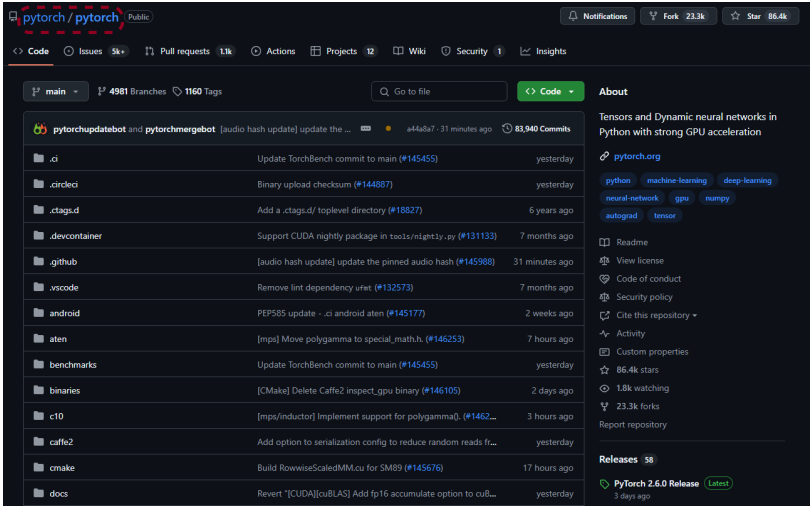
- Technological innovation and growth
  - Patents (Kogan et al. (2017); Acemoglu et al. (2018)); Trademarks (Desai et al. (2023)); Open source (Nagle, 2018)

    This paper provides evidence that open source innovation is associated with creative destruction

Contents

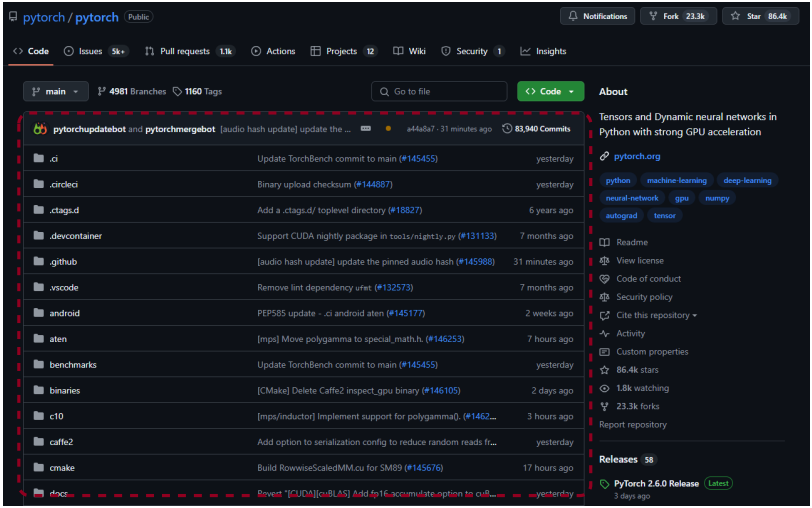
GitHub repositories

- Meta's PyTorch: ML framework for neural networks, under a permissive BSD license.



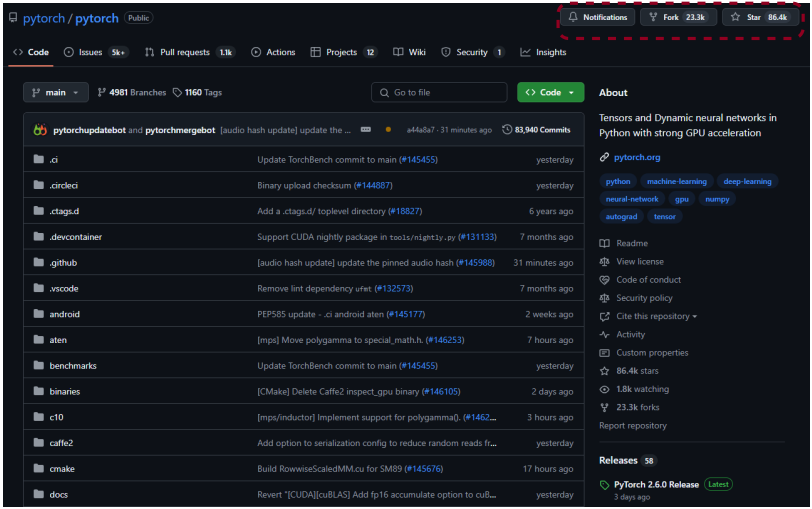
GitHub repositories

- Meta's PyTorch: ML framework for neural networks, under a permissive BSD license.



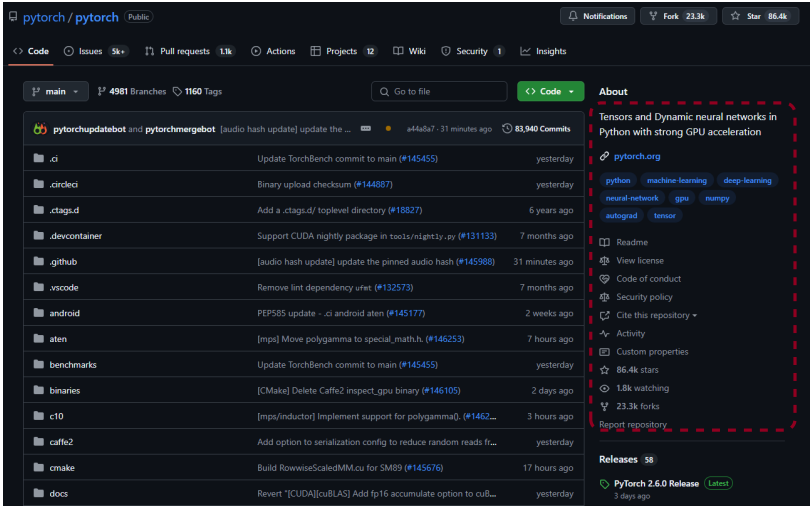
GitHub repositories

- Meta's PyTorch: ML framework for neural networks, under a permissive BSD license.



GitHub repositories

- Meta's PyTorch: ML framework for neural networks, under a permissive BSD license.



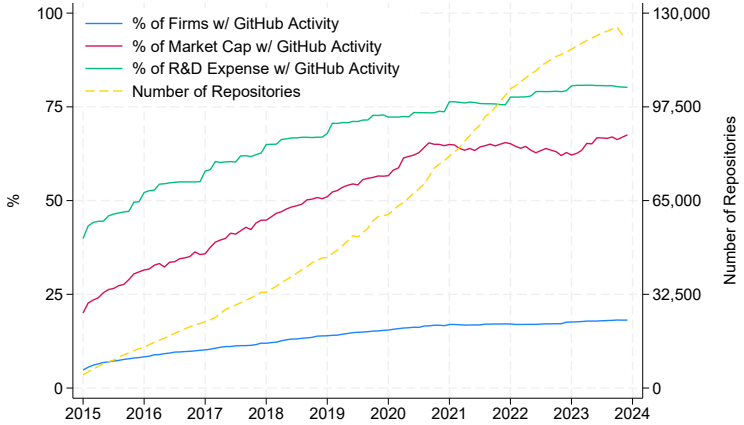
Contents

Data

- Match GitHub organization accounts (GHTorrent and GitHub API) to US public firms via website domains (Compustat and Orbis) and manual search.
  - 1,281 firms, 3,314 organization accounts, and 168,085 public repositories during 2015-2023.
- Timestamped public activity on GitHub from GHArchive
  - Identify the timestamps associated with the earliest activity, labeled as "Public Event."
- Static repository characteristics as of February, 2024, from the GitHub API, including licenses and number of stars.
- Large language models to classify or evaluate repositories based on topics, complementarity, and novelty (GPT-4o model).

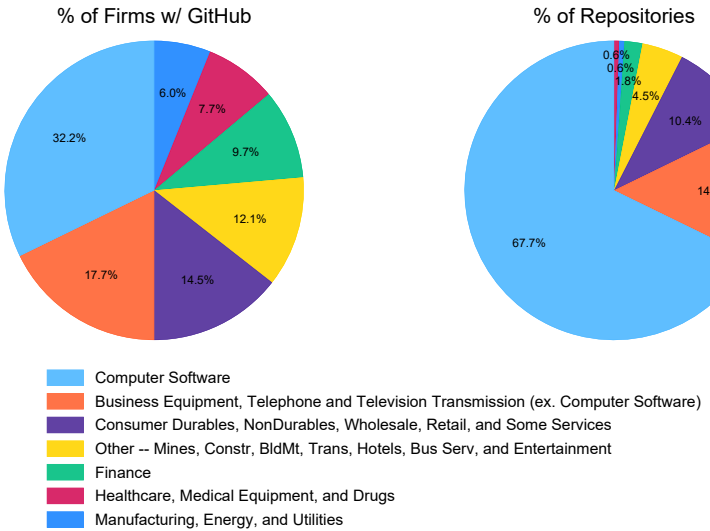
▸ Examples

Trend of firms with GitHub activity



- GitHub firms are an important part of the US economy
- GitHub participation rate: 4.8% (2015) to 18.1% (2023)
- 122,107 repositories, 67.5% of market cap, 80.2% of total R&D expenditure

Industry distribution of open-source engagement



Characteristics of firms with/without GitHub activity

	GitHub Firms		Non-GitHub Firms	
	Mean	Median	Mean	Median
Market Capitalization	32,220,913	3,743,408	1,418,477	71,523
Employees	32.5	4.7	7.9	1.1
Number of Patents	1,264	13	67	0
Investment	3.34%	2.06%	7.06%	4.30%
Market-to-Book	6.26	3.49	2.69	1.55
Return-on-Assets	-1.39%	2.26%	-1.35%	3.25%
Annual Returns	14.41%	6.50%	15.21%	5.74%
Sales Growth	14.93%	9.11%	17.27%	8.88%
Tangibility	14.87%	9.02%	27.35%	20.86%
R&D Exp / Total Assets	8.22%	4.79%	3.99%	0.00%
Market Power	3.13	2.21	2.29	1.64
Scope	11	10	8	7
Product Market Centrality	0.0043	0.0024	0.0086	0.0039
Product Market Similarity	4.16	1.74	11.64	2.00
Product Market Fluidity	5.23	4.91	7.67	6.92

Characteristics of firms with/without GitHub activity

	GitHub Firms		Non-GitHub Firms	
	Mean	Median	Mean	Median
Market Capitalization	32,220,913	3,743,408	1,418,477	71,523
Employees	32.5	4.7	7.9	1.1
Number of Patents	1,264	13	67	0
Investment	3.34%	2.06%	7.06%	4.30%
Market-to-Book	6.26	3.49	2.69	1.55
Return-on-Assets	-1.39%	2.26%	-1.35%	3.25%
Annual Returns	14.41%	6.50%	15.21%	5.74%
Sales Growth	14.93%	9.11%	17.27%	8.88%
Tangibility	14.87%	9.02%	27.35%	20.86%
R&D Exp / Total Assets	8.22%	4.79%	3.99%	0.00%
Market Power	3.13	2.21	2.29	1.64
Scope	11	10	8	7
Product Market Centrality	0.0043	0.0024	0.0086	0.0039
Product Market Similarity	4.16	1.74	11.64	2.00
Product Market Fluidity	5.23	4.91	7.67	6.92

- Larger, higher valuations, less tangible assets, larger R&D expenditure

Characteristics of firms with/without GitHub activity

	GitHub Firms		Non-GitHub Firms	
	Mean	Median	Mean	Median
Market Capitalization	32,220,913	3,743,408	1,418,477	71,523
Employees	32.5	4.7	7.9	1.1
Number of Patents	1,264	13	67	0
Investment	3.34%	2.06%	7.06%	4.30%
Market-to-Book	6.26	3.49	2.69	1.55
Return-on-Assets	-1.39%	2.26%	-1.35%	3.25%
Annual Returns	14.41%	6.50%	15.21%	5.74%
Sales Growth	14.93%	9.11%	17.27%	8.88%
Tangibility	14.87%	9.02%	27.35%	20.86%
R&D Exp / Total Assets	8.22%	4.79%	3.99%	0.00%
Market Power	3.13	2.21	2.29	1.64
Scope	11	10	8	7
Product Market Centrality	0.0043	0.0024	0.0086	0.0039
Product Market Similarity	4.16	1.74	11.64	2.00
Product Market Fluidity	5.23	4.91	7.67	6.92

- Larger, higher valuations, less tangible assets, larger R&D expenditure
- More market power, bigger scope, less competition

Characteristics of firms with/without GitHub activity

	GitHub Firms		Non-GitHub Firms	
	Mean	Median	Mean	Median
Market Capitalization	32,220,913	3,743,408	1,418,477	71,523
Employees	32.5	4.7	7.9	1.1
Number of Patents	1,264	13	67	0
Investment	3.34%	2.06%	7.06%	4.30%
Market-to-Book	6.26	3.49	2.69	1.55
Return-on-Assets	-1.39%	2.26%	-1.35%	3.25%
Annual Returns	14.41%	6.50%	15.21%	5.74%
Sales Growth	14.93%	9.11%	17.27%	8.88%
Tangibility	14.87%	9.02%	27.35%	20.86%
R&D Exp / Total Assets	8.22%	4.79%	3.99%	0.00%
Market Power	3.13	2.21	2.29	1.64
Scope	11	10	8	7
Product Market Centrality	0.0043	0.0024	0.0086	0.0039
Product Market Similarity	4.16	1.74	11.64	2.00
Product Market Fluidity	5.23	4.91	7.67	6.92

- Larger, higher valuations, less tangible assets, larger R&D expenditure
- More market power, bigger scope, less competition
- In regression, firm fixed effects absorb differences

▸ Full table

Contents

Repository announcements

- We estimate value using the stock market reaction to the repository announcement.

Repository announcements

- We estimate value using the stock market reaction to the repository announcement.
- Do investors *know* about repository announcements?
  - Firm's official websites (Microsoft, Amazon, Apple, Salesforce, Meta)
  - Media coverage (Wired, TechCrunch, The Verge)
  - Social media (Reddit, X/Twitter, HackerNews)

Repository announcements

- We estimate value using the stock market reaction to the repository announcement.
- Do investors *know* about repository announcements?
  - Firm's official websites (Microsoft, Amazon, Apple, Salesforce, Meta)
  - Media coverage (Wired, TechCrunch, The Verge)
  - Social media (Reddit, X/Twitter, HackerNews)
- Do investors *care* about repository announcements?
  - Alphabet Q1-2024 earnings call: "On top of our infrastructure, we offer more than 130 models, including our own models, open source models and third-party models. We made Gemini 1.5 Pro available to customers, as well as Imagen 2.0 at Cloud Next."
  - Meta Q2-2024 earnings call: "This quarter we released Llama 3.1, which includes the first frontier-level open source model, as well as new and industry-leading small and medium-sized models."

Repository announcements

- We estimate value using the stock market reaction to the repository announcement.
- Do investors *know* about repository announcements?
  - Firm's official websites (Microsoft, Amazon, Apple, Salesforce, Meta)
  - Media coverage (Wired, TechCrunch, The Verge)
  - Social media (Reddit, X/Twitter, HackerNews)
- Do investors *care* about repository announcements?
  - Alphabet Q1-2024 earnings call: "On top of our infrastructure, we offer more than 130 models, including our own models, open source models and third-party models. We made Gemini 1.5 Pro available to customers, as well as Imagen 2.0 at Cloud Next."
  - Meta Q2-2024 earnings call: "This quarter we released Llama 3.1, which includes the first frontier-level open source model, as well as new and industry-leading small and medium-sized models."
- Abnormal trading volume on repository announcements for large firms. ▸ Figure

Estimating repository value

- We follow the procedure developed by Kogan et al. (2017) to estimate patent value.

Estimating repository value

- We follow the procedure developed by Kogan et al. (2017) to estimate patent value.
- Assume the **three-day cumulative market-adjusted return** following repository announcements is a function of both **investor reaction** and **idiosyncratic noise**:

$$R_i = v_i + \varepsilon_i.$$

Estimating repository value

- We follow the procedure developed by Kogan et al. (2017) to estimate patent value.
- Assume the **three-day cumulative market-adjusted return** following repository announcements is a function of both **investor reaction** and **idiosyncratic noise**:

$$R_i = v_i + \varepsilon_i.$$

- Assuming $v_i \sim \mathcal{N}^+(0, \sigma^2_{vft})$, the value of repository i, ξ_i, is

$$\xi_i = \frac{1}{N_i} E[v_i | R_i] M_i,$$

Estimating repository value

- We follow the procedure developed by Kogan et al. (2017) to estimate patent value.
- Assume the **three-day cumulative market-adjusted return** following repository announcements is a function of both **investor reaction** and **idiosyncratic noise**:

$$R_i = v_i + \varepsilon_i.$$

- Assuming $v_i \sim \mathcal{N}^+(0, \sigma^2_{vft})$, the value of repository i, ξ_i, is

$$\xi_i = \frac{1}{N_i} E[v_i | R_i] M_i,$$

- Results are robust to using $[t-2, t+2]$, $[t]$, or R_i.

▸ More details

Summary of repository values

- R: Mean = 0.12% (t=5.91), Median = 0.04%
- ξ: Mean = \$842,849, Median = \$562,022

Summary of repository values

- R: Mean = 0.12% (t=5.91), Median = 0.04%
- ξ: Mean = \$842,849, Median = \$562,022
- Industries with most value: Software, Consumer Products, Business Equipment
  - Least value: Healthcare

Summary of repository values

- R: Mean = 0.12% (t=5.91), Median = 0.04%
- ξ: Mean = \$842,849, Median = \$562,022
- Industries with most value: Software, Consumer Products, Business Equipment
  - Least value: Healthcare
- Amazon and Microsoft have most valuable portfolios, $>$ \$7B each.

Summary of repository values

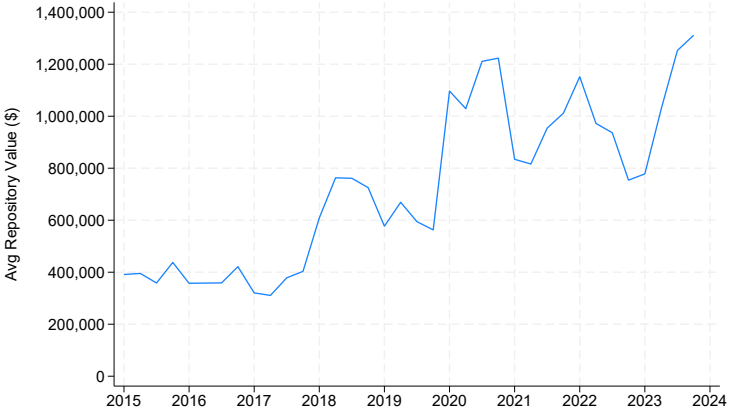
- R: Mean = 0.12% (t=5.91), Median = 0.04%
- ξ: Mean = \$842,849, Median = \$562,022
- Industries with most value: Software, Consumer Products, Business Equipment
  - Least value: Healthcare
- Amazon and Microsoft have most valuable portfolios, $>$ \$7B each.
- Python repositories produce most value.

Summary of repository values

- R: Mean = 0.12% (t=5.91), Median = 0.04%
- ξ: Mean = \$842,849, Median = \$562,022
- Industries with most value: Software, Consumer Products, Business Equipment
  - Least value: Healthcare
- Amazon and Microsoft have most valuable portfolios, > \$7B each.
- Python repositories produce most value.
- Using LLM for topic classification, ML & AI repositories produce most value.

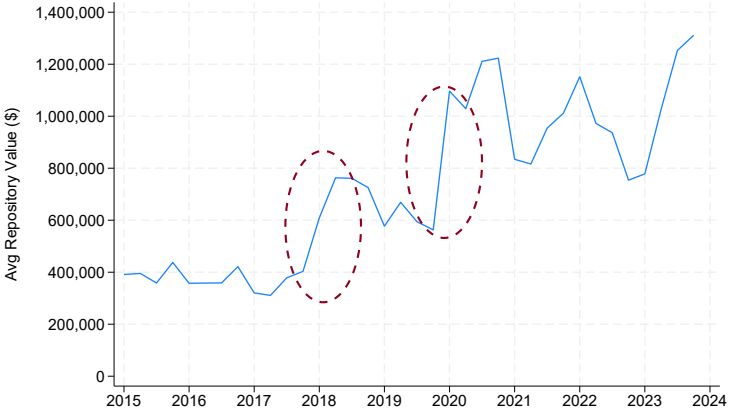
Tables

Average estimated repository value by quarter



- The average value grew from $400,000 (2015) to $1,300,000 (2023)

Average estimated repository value by quarter



- The average value grew from $400,000 (2015) to $1,300,000 (2023)
- Microsoft's acquisition of GitHub (2018) & COVID shutdown (2020)

Validating our measure

- Do the value estimates contain actual information?

Validating our measure

- Do the value estimates contain actual information?
- Two validation tests to ensure that we capture $\boldsymbol{\nu}$ (investor reaction), not ϵ (noise):

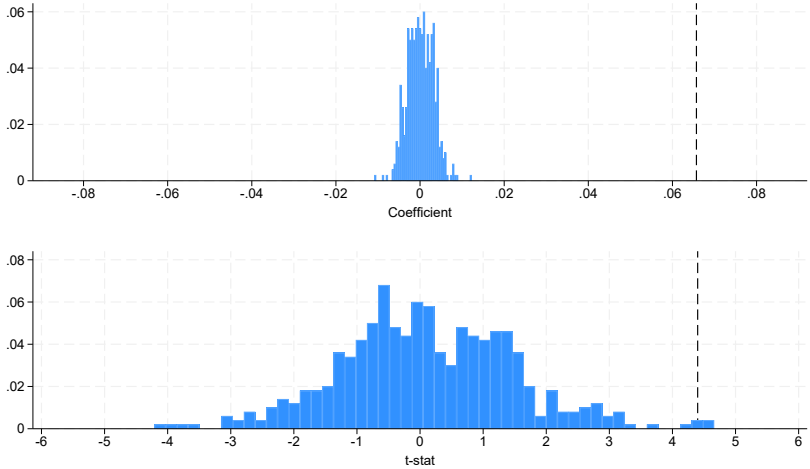
Validating our measure

- Do the value estimates contain actual information?
- Two validation tests to ensure that we capture $\boldsymbol{\nu}$ (investor reaction), not ϵ (noise):
  1. Regress repository value on its future popularity
     - Measure popularity using the number of stars (bookmarks)
     - More-valuable repositories are significantly more popular in the future ▸ Table

Validating our measure

- Do the value estimates contain actual information?
- Two validation tests to ensure that we capture $\boldsymbol{v}$ (investor reaction), not ϵ (noise):
  1. Regress repository value on its future popularity
     - Measure popularity using the number of stars (bookmarks)
     - More-valuable repositories are significantly more popular in the future ▸ Table
  2. Placebo test using randomly generated repository announcement dates
     - 500 iterations, repeating the previous test each time
     - Compare the distribution of coefficients and t-stats to the true values

Placebo test



- Randomly assigned release date in the same year (500 iterations)

Contents

Determinants: Repository characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)
ln(Stars + 1)	0.090***	0.088***	0.054**	0.099***	0.097**
	(0.017)	(0.019)	(0.021)	(0.022)	(0.033)
Copyleft License	0.089***				0.105**
	(0.019)				(0.038)
Other License	0.037				0.019
	(0.027)				(0.025)
Complementarity		-0.302***			-0.221***
		(0.071)			(0.048)
Novelty			0.499***		0.379***
			(0.093)		(0.092)
ln(Repo Size + 1)				-0.026**	-0.014
				(0.008)	(0.014)
Observations	28,690	28,690	28,690	28,690	28,690
Adj. R^2	0.819	0.821	0.820	0.819	0.825
Controls	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value. Column (8) adds N Repos, Template, and N Issues Opened.

Determinants: Repository characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)
ln(Stars + 1)	0.090***	0.088***	0.054**	0.099***	0.097**
	(0.017)	(0.019)	(0.021)	(0.022)	(0.033)
Copyleft License	0.089***				0.105**
	(0.019)				(0.038)
Other License	0.037				0.019
	(0.027)				(0.025)
Complementarity		-0.302***			-0.221***
		(0.071)			(0.048)
Novelty			0.499***		0.379***
			(0.093)		(0.092)
ln(Repo Size + 1)				-0.026**	-0.014
				(0.008)	(0.014)
Observations	28,690	28,690	28,690	28,690	28,690
Adj. R^2	0.819	0.821	0.820	0.819	0.825
Controls	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value. Column (8) adds N Repos, Template, and N Issues Opened.

Determinants: Repository characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)
ln(Stars + 1)	0.090***	0.088***	0.054**	0.099***	0.097**
	(0.017)	(0.019)	(0.021)	(0.022)	(0.033)
Copyleft License	0.089***				0.105**
	(0.019)				(0.038)
Other License	0.037				0.019
	(0.027)				(0.025)
Complementarity		-0.302***			-0.221***
		(0.071)			(0.048)
Novelty			0.499***		0.379***
			(0.093)		(0.092)
ln(Repo Size + 1)				-0.026**	-0.014
				(0.008)	(0.014)
Observations	28,690	28,690	28,690	28,690	28,690
Adj. R^2	0.819	0.821	0.820	0.819	0.825
Controls	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value. Column (8) adds N Repos, Template, and N Issues Opened.

Determinants: Repository characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)
ln(Stars + 1)	0.090***	0.088***	0.054**	0.099***	0.097**
	(0.017)	(0.019)	(0.021)	(0.022)	(0.033)
Copyleft License	0.089***				0.105**
	(0.019)				(0.038)
Other License	0.037				0.019
	(0.027)				(0.025)
Complementarity		-0.302***			-0.221***
		(0.071)			(0.048)
Novelty			0.499***		0.379***
			(0.093)		(0.092)
ln(Repo Size + 1)				-0.026**	-0.014
				(0.008)	(0.014)
Observations	28,690	28,690	28,690	28,690	28,690
Adj. R^2	0.819	0.821	0.820	0.819	0.825
Controls	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value. Column (8) adds N Repos, Template, and N Issues Opened.

Determinants: Product market characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)	(6)
ln(Stars + 1)	0.068***	0.078***	0.065***	0.070***	0.069***	0.064***
	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)
Market Power	0.102***					0.067***
	(0.018)					(0.019)
Product Market Centrality		0.017				0.135***
		(0.011)				(0.022)
Scope			-0.149***			-0.100***
			(0.018)			(0.019)
Product Market Similarity				-0.088		-0.010
				(0.047)		(0.032)
Product Market Fluidity					-0.158***	-0.131**
					(0.022)	(0.044)
Observations	23,735	23,735	23,735	23,735	23,735	23,735
Adj. R^2	0.824	0.821	0.825	0.822	0.824	0.827
Controls	✓	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value

Determinants: Product market characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)	(6)
ln(Stars + 1)	0.068***	0.078***	0.065***	0.070***	0.069***	0.064***
	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)
Market Power	0.102***					0.067***
	(0.018)					(0.019)
Product Market Centrality		0.017				0.135***
		(0.011)				(0.022)
Scope			-0.149***			-0.100***
			(0.018)			(0.019)
Product Market Similarity				-0.088		-0.010
				(0.047)		(0.032)
Product Market Fluidity					-0.158***	-0.131**
					(0.022)	(0.044)
Observations	23,735	23,735	23,735	23,735	23,735	23,735
Adj. R^2	0.824	0.821	0.825	0.822	0.824	0.827
Controls	✓	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value

Determinants: Product market characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)	(6)
ln(Stars + 1)	0.068***	0.078***	0.065***	0.070***	0.069***	0.064***
	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)
Market Power	0.102***					0.067***
	(0.018)					(0.019)
Product Market Centrality		0.017				0.135***
		(0.011)				(0.022)
Scope			-0.149***			-0.100***
			(0.018)			(0.019)
Product Market Similarity				-0.088		-0.010
				(0.047)		(0.032)
Product Market Fluidity					-0.158***	-0.131**
					(0.022)	(0.044)
Observations	23,735	23,735	23,735	23,735	23,735	23,735
Adj. R^2	0.824	0.821	0.825	0.822	0.824	0.827
Controls	✓	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value

Determinants: Product market characteristics

Dep. var. = ln(ξ)	(1)	(2)	(3)	(4)	(5)	(6)
ln(Stars + 1)	0.068***	0.078***	0.065***	0.070***	0.069***	0.064***
	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)
Market Power	0.102***					0.067***
	(0.018)					(0.019)
Product Market Centrality		0.017				0.135***
		(0.011)				(0.022)
Scope			-0.149***			-0.100***
			(0.018)			(0.019)
Product Market Similarity				-0.088		-0.010
				(0.047)		(0.032)
Product Market Fluidity					-0.158***	-0.131**
					(0.022)	(0.044)
Observations	23,735	23,735	23,735	23,735	23,735	23,735
Adj. R^2	0.824	0.821	0.825	0.822	0.824	0.827
Controls	✓	✓	✓	✓	✓	✓
Industry x Year FE	✓	✓	✓	✓	✓	✓

Controls: repository popularity, market capitalization, volatility, employees, and total patent value

Repository values and firm output

$$\ln Y_{f,t+k} - \ln Y_{f,t} = \beta_k \textit{Repo Output}_{f,t} + \gamma_k \textit{Repo Output}_{I \setminus f,t} + \psi_k X_{f,t} + \epsilon_{f,t+k}$$

- Does open source innovation by a firm contribute to future growth for that firm?
  - Regress Y growth for year $t+1, t+2, t+3$ on repository value and patent value at t
  - Y: sales, profits, labor, proprietary innovation output (patents)
- Does open source innovation by a firm's competitors affect the future growth of that firm?
  - Open-source innovation by competitors is freely accessible
  - Open-source innovation by competitors may be difficult to implement (inherent excludability)

Repository values and firm output (firm)

	Firm (Horizon)		
	1	2	3
Panel A: Sales			
Repo Output	0.015***	0.033***	0.051**
	(0.002)	(0.006)	(0.018)
Panel B: Profits			
Repo Output	0.019***	0.036***	0.055***
	(0.003)	(0.007)	(0.013)
Panel C: Labor			
Repo Output	0.015***	0.028***	0.041**
	(0.002)	(0.007)	(0.012)
Panel D: Value of Patents			
Repo Output	0.049***	0.102***	0.150***
	(0.009)	(0.018)	(0.028)
Panel E: Number of Patents			
Repo Output	0.043***	0.072***	0.113***
	(0.010)	(0.013)	(0.019)
Controls	✓	✓	✓
Industry FE	✓	✓	✓
Year FE	✓	✓	✓

Controls: one lag of dependent variable, firm capital, the number of employees, idiosyncratic volatility, patent outputs of both the firm/competitors

- Firms engaging in more-valuable open-source innovation have higher growth rates
- Patent outputs on growth: 5.4% (sales), 5.9% (Profits), 2.6% (Labor)

Repository values and firm output (competitors)

	Competitors (Horizon)		
	1	2	3
Panel A: Sales			
Repo Output	−0.002	−0.006	−0.027
	(0.012)	(0.018)	(0.014)
Panel B: Profits			
Repo Output	−0.012	−0.029**	−0.040***
	(0.008)	(0.011)	(0.009)
Panel C: Labor			
Repo Output	−0.005	−0.011	−0.014**
	(0.003)	(0.007)	(0.005)
Panel D: Value of Patents			
Repo Output	−0.014	−0.036	−0.048*
	(0.015)	(0.020)	(0.020)
Panel E: Number of Patents			
Repo Output	−0.014**	−0.037***	−0.049**
	(0.006)	(0.006)	(0.014)
Controls	✓	✓	✓
Industry FE	✓	✓	✓
Year FE	✓	✓	✓

Controls: one lag of dependent variable, firm capital, the number of employees, idiosyncratic volatility, patent outputs of both the firm/competitors

- Competitors' open-source innovation is associated with creative destruction
- Note: Our results do not consider the public value of open-source innovation

Contents

Conclusion

- We compile a comprehensive dataset of open-source activities on GitHub, and estimate their **private** value based on stock market reaction.
- Private value generated by open-source innovation totals $25B, with the computer software industry and AI-related innovation contributing the most.
- Key Determinants of Open-Source Project Value:
  - Inherent excludability and innovation novelty are valued more than permissiveness and complementarity to commercial products.
  - Firms generate less value from open-source projects in highly competitive markets.
- The private value of open-source innovation predicts long-term firm growth and is associated with creative destruction on average.

References I

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N. and Kerr, W. (2018), 'Innovation, Reallocation, and Growth', *American Economic Review* **108**(11), 3450–3491.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20130470*

Alexy, O., West, J., Klapper, H. and Reitzig, M. (2018), 'Surrendering control to gain advantage: Reconciling openness and the resource-based view of the firm', *Strategic Management Journal* **39**(6), 1704–1727.

Aryal, G., Ciliberto, F., Farmer, L. E. and Khmelnitskaya, E. (2022), 'Valuing Pharmaceutical Drug Innovations'.
URL: *https://papers.ssrn.com/abstract=4300678*

Blind, K., Böhm, M., Grzegorzewska, P., Katz, A., Muto, S., Pätsch, S. and Schubert, T. (2021), The impact of Openb Source Software and Hardware on technological independence, competitiveness and innovation in the EU economy. European Commission, Ed.

Chen, M. A., Wu, Q. and Yang, B. (2019), 'How Valuable Is FinTech Innovation?', *Review of Financial Studies* **32**(5), 2062–2106.
URL: *https://doi.org/10.1093/rfs/hhy130*

Desai, P., Gavrilova, E., Silva, R. and Soares, M. (2023), The Value of Trademarks. Working Paper, Nova School of Business and Economics.

Greenstein, S. and Nagle, F. (2014), 'Digital dark matter and the economic contribution of Apache', *Research Policy* **43**(4), 623–631.

Hoffmann, M., Nagle, F. and Zhou, Y. (2024), The value of open source software. Working paper, Harvard University, University of Toronto.

Kogan, L., Papanikolaou, D., Seru, A. and Stoffman, N. (2017), 'Technological Innovation, Resource Allocation, and Growth', *Quarterly Journal of Economics* **132**(2), 665–712.

Lerner, J., Pathak, P. A. and Tirole, J. (2006), 'The Dynamics of Open-Source Contributors', *American Economic Review* **96**(2), 114–118.
URL: *https://pubs.aeaweb.org/doi/10.1257/000282806777211874*

Lerner, J. and Tirole, J. (2002), 'Some Simple Economics of Open Source', *Journal of Industrial Economics* **50**(2), 197–234.

Lev, B. and Sougiannis, T. (1996), 'The capitalization, amortization, and value-relevance of r&d', *Journal of Accounting and Economics* **21**(1), 107–138.

Lin, Y.-K. and Maruping, L. M. (2022), 'Open Source Collaboration in Digital Entrepreneurship', *Organization Science* **33**(1), 212–230.

Murciano-Goroff, R., Zhuo, R. and Greenstein, S. (2021), 'Hidden software and veiled value creation: Illustrations from server software usage', *Research Policy* **50**(9), 104333.

Nagle, F. (2018), 'Learning by Contributing: Gaining Competitive Advantage Through Contribution to Crowdsourced Public Goods', *Organization Science* **29**(4), 569–587.

Robbins, C., Korkmaz, G., Guci, L., Calderón, J. B. S. and Kramer, B. (2021), A First Look at Open-Source Software Investment in the United States and in Other Countries, 2009-2019. Paper prepared for the IARIW-ESCoE Conference.

Sougiannis, T. (1994), 'The accounting based valuation of corporate r&d', *Accounting Review* **69**(1), 44–68.

An example of a GitHub repository: Meta's PyTorch

- Open-source ML framework under a permissive BSD license, widely used for deep learning applications.
- How PyTorch Benefits Meta:
  - **Industry Standard**: Adopted by AI leaders (e.g., Tesla, OpenAI, Microsoft), reinforcing Meta's influence in AI.
  - **AI Talent Pipeline**: Researchers and engineers trained on PyTorch reduce onboarding and training costs for Meta.
  - **Ecosystem Control**: Meta directs PyTorch's development, ensuring compatibility with its AI infrastructure (e.g., MTIA chips).
  - **Community Contributions**: External companies enhance PyTorch at no cost to Meta, driving innovation.
- Strategic Trade-off: Competitors benefit, but Meta gains AI standardization, research leadership, and talent acquisition.

Examples of complementarity and novelty

- Complementarity
  - WhatsApp StringPacks - 0
    - Stores translation strings in a more efficient binary format for Android applications.
  - WhatsApp Stickers - 0.8
    - API for creating third-party sticker packs for WhatsApp.
- Novelty
  - LinkedInLearning - 0.1
    - Exercises associated with courses on the LinkedIn Learning platform.
  - AlphaFold - 0.8
    - Google AI that predicts a protein's 3D structure, developed by 2024 Chemistry Nobel Prize winners.

Back

GitHub activity across industries

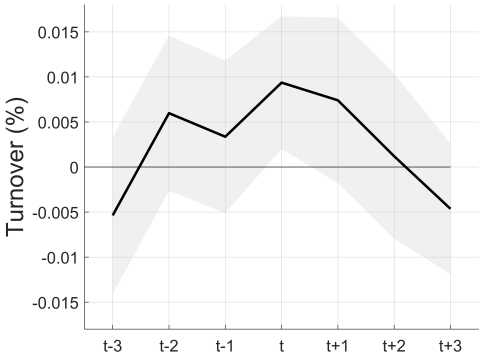
		Number of Repositories								
	% GitHub	Mean	Std	p25	p50	p75	p90	p95	p99	Total	N Firms
Total	18.1%	30.9	430.1	0	0	0	17	62	425	122,971	3,982
Computer Software	62.6%	226.5	1298.7	0	13	82	340	657	4551	82,896	366
Consumer Products	19.0%	23.6	393.4	0	0	0	13	36	197	12,757	541
Manufacturing, Energy, and Utilities	7.5%	1.2	6.4	0	0	0	0	6	38	6961	575
Business Equipment (ex. Computer Software)	34.9%	49.1	229.6	0	0	8	87	156	1253	17,726	361
Other	20.6%	13.3	65.8	0	0	0	14	46	336	5,558	418
Finance	10.8%	3.4	17.3	0	0	0	2	18	93	2,173	636
Healthcare	6.4%	0.8	5.7	0	0	0	0	2	24	720	861

Determinants of open-source activity

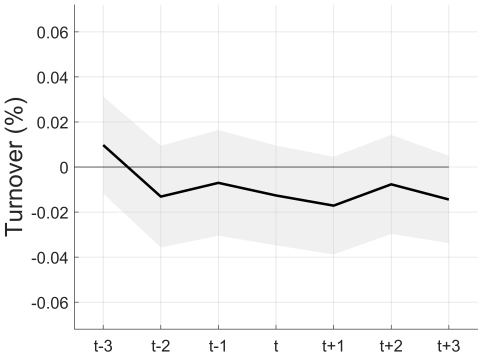
	(1) GitHub	(2) GitHub	(3) Commits	(4) Commits
ln(Mkt Cap)	0.087***	0.016*	1.127***	0.026
	(0.018)	(0.009)	(0.272)	(0.223)
ln(Employees)	0.003	0.024	-0.092	0.532*
	(0.011)	(0.019)	(0.248)	(0.273)
ln(N Patents + 1)	0.043***	0.017	0.613***	0.035
	(0.012)	(0.042)	(0.138)	(0.249)
Market-to-Book	0.014*	0.002	-0.025	0.081*
	(0.008)	(0.003)	(0.050)	(0.047)
Return-on-Assets	-0.008*	0.002	-0.178***	0.017
	(0.004)	(0.002)	(0.060)	(0.049)
Investment	0.002	-0.000	0.014	0.101
	(0.004)	(0.002)	(0.128)	(0.101)
Return (t-12 to t-1)	-0.007***	-0.003***	-0.072**	-0.041**
	(0.002)	(0.001)	(0.031)	(0.017)
Sales Growth	-0.001	-0.001	0.065	-0.060*
	(0.004)	(0.002)	(0.042)	(0.036)
Tangibility	-0.003	-0.010	0.130	-0.015
	(0.010)	(0.007)	(0.174)	(0.178)
R&D Exp/Total Assets	0.029**	-0.003	0.280*	0.056
	(0.012)	(0.005)	(0.149)	(0.091)
R&D Exp Missing	-0.041**	-0.014	0.416*	-0.777***
	(0.018)	(0.016)	(0.234)	(0.218)
Market Power	0.009	0.001	0.215***	-0.117***
	(0.007)	(0.004)	(0.044)	(0.030)
Scope	-0.005	0.003	0.066	-0.050
	(0.010)	(0.005)	(0.122)	(0.085)
Product Market Centrality	-0.037**	-0.001	-0.472**	0.168*
	(0.016)	(0.015)	(0.213)	(0.092)
Product Market Similarity	0.002	-0.005	0.149	-0.078**
	(0.017)	(0.011)	(0.239)	(0.034)
Product Market Fluidity	0.009	-0.006*	0.120**	-0.079*
	(0.007)	(0.003)	(0.052)	(0.046)
Observations	208,528	208,513	26,422	26,413
Adj. R^2	0.331	0.866	0.327	0.781
Industry x Time FE	✓	✓	✓	✓
Firm FE		✓		✓
Sample	All firms	All firms	GitHub = 1	GitHub = 1

Back

Share turnover around repository announcements



(A) Large Firms



(B) Small Firms

- Measurable trading reaction to repository announcements for large firms.

Back

Estimating repository value (continued)

- $E[v_i|R_i]$ is the expected return attributable to the repository announcement conditional on observing the three-day market-adjusted return.
- Assuming $v_i \sim \mathcal{N}^+(0, \sigma^2_{vft})$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2_{\varepsilon ft})$, the conditional expected return is

$$E[v_i|R_i] = \delta R_i + \sqrt{\delta}\sigma_{\varepsilon ft} \frac{\phi\left(-\sqrt{\delta}\frac{R_j}{\sigma_{\varepsilon ft}}\right)}{1 - \Phi\left(-\sqrt{\delta}\frac{R_j}{\sigma_{\varepsilon ft}}\right)},$$

- δ is the signal-to-noise ratio.

Estimating repository value (continued)

- We adopt the same simplifying assumption as Kogan et al. (2017) that δ is the same for all firms and all time periods.
  - This assumption still allows σ^2_{vft} and $\sigma^2_{\varepsilon ft}$ to vary across firms and time, but only in constant proportion.
- To estimate δ, we compare the variance of returns in the announcement window to that of returns over other three-day periods for the same firm within the same year:

$$\ln(R^2_{fd}) = \gamma I_{fd} + \lambda_{dow} + \eta_{fy} + u_{fd},$$

where R_{fd} is the three-day cumulative market-adjusted return for firm f on day d, I_{fd} is an indicator variable that equals one if there is a repository announcement by firm f on day d.

Estimation of noise-to-signal ratio (continued)

- The estimated $\hat{\delta}$ can be calculated from the resulting estimate $\hat{\lambda}$ as $\hat{\delta} = 1 - e^{-\hat{\gamma}}$.
- Finally, we estimate $\sigma^2_{\varepsilon ft}$ for each firm within each year as

$$\sigma^2_{\varepsilon ft} = \frac{3\sigma^2_{ft}}{1 + 3d_{ft}(e^{-\hat{\gamma}} - 1)},$$

where d_{ft} is the fraction of days in the given year that are announcement days for firm f and σ^2_{ft} is the variance of daily market-adjusted returns calculated within each firm for each year.

Back

Summary of repository values (continued)

	Mean	Std	p1	p5	p10	p25	p50	p75	p90	p95	p99	N
R	0.12%	3.49%	-9.17%	-4.59%	-3.06%	-1.31%	0.04%	1.48%	3.47%	5.03%	10.13%	29,543
E[r|R]	0.27%	0.31%	0.01%	0.03%	0.04%	0.06%	0.14%	0.39%	0.68%	0.89%	1.36%	29,543
ξ	842,849	1,044,529	1,637	9,514	23,006	140,577	562,022	1,146,063	1,953,618	2,742,288	5,161,436	29,543
Stars	212.2	2,226.6	0	0	0	2	10	44	207	565	3,752	29,543
Complementarity	0.41	0.29	0	0	0	0.1	0.5	0.6	0.8	0.8	0.8	29,508
Novelty	0.26	0.14	0	0.1	0.1	0.2	0.3	0.3	0.5	0.5	0.6	29,529
Repo Size	31,322.1	282,597.8	5	14	27	113	808	6,374	39,712	100,330	526,113	29,535
N Issues Opened	56.0	1,098.4	0	0	0	0	1	8	43	123	814	29,543

Back

Repository value by industry

	Total ξ	Mean ξ	Median ξ	N Repos	% Permissive
Computer Software	13,472,579,594	781,835	463,577	17,232	68.3%
Consumer Products	7,911,507,157	1,090,190	892,361	7,257	88.7%
Business Equipment (ex. Computer Software)	3,091,412,879	961,559	232,003	3,215	49.4%
Other	181,271,280	345,938	216,893	524	65.3%
Finance	89,538,721	255,825	140,748	350	77.7%
Healthcare	51,579,235	531,745	269,625	97	41.2%

Back

Firms with the most valuable GitHub portfolios

	Portfolio ξ	Mean ξ	Median ξ	N Repos	% Permissive
Amazon.com Inc	7,814,250,435	1,145,281	924,267	6,823	91.5%
Microsoft Corp	7,759,843,486	1,129,855	862,161	6,868	72.7%
Meta Platforms Inc	2,280,475,890	1,949,125	1,738,289	1,170	45.0%
Alphabet Inc	1,763,699,392	1,031,403	808,095	1,710	86.8%
NVIDIA Corporation	1,391,421,878	2,394,874	1,868,125	581	49.6%
Apple Inc	1,059,601,598	4,489,837	3,702,204	236	55.1%
Salesforce Inc	495,777,701	477,168	361,598	1,039	74.4%
Adobe Inc	225,007,179	646,572	578,176	348	77.6%
International Business Machines Corp	211,208,262	344,549	334,237	613	51.7%
Oracle Corp	206,261,704	661,095	612,226	312	71.2%
...	...	...	...	...	...
Total	24,900,292,747	842,849	562,022	29,543	70.7%

Back

Most valuable programming languages

	Total ξ	Mean ξ	Median ξ	N Repos	% Permissive
Python	6,853,266,877	1,154,526	826,076	5,936	75.1%
TypeScript	1,924,509,635	834,928	623,765	2,305	80.2%
JavaScript	1,738,402,866	552,927	272,963	3,144	72.2%
Jupyter Notebook	1,565,510,149	1,302,421	941,121	1,202	79.9%
C#	1,256,454,230	794,721	565,036	1,581	71.8%
Java	1,216,853,117	640,112	414,790	1,901	74.8%
C++	974,203,192	987,035	669,700	987	66.6%
Shell	782,874,993	794,797	555,364	985	73.4%
Go	754,409,476	517,428	228,037	1,458	80.0%
HTML	605,445,891	663,866	363,772	912	59.4%

Back

Repository value by topic (continued)

	Mean ξ	Median ξ	Total ξ	Mean Topic Score	N Repos	% Permissive
Core AI and ML	880,828	623,525	2,850,360,153	0.60	3,236	70.3%
AI Applications	824,267	584,499	2,853,612,365	0.58	3,462	70.4%
Digital Media	618,571	359,345	1,214,874,229	0.60	1,964	64.2%
Education and Learning	548,422	369,979	1,188,979,071	0.51	2,168	52.1%
Advanced Data Analysis	488,397	330,059	1,933,075,386	0.44	3,958	74.3%
Cloud Infrastructure and DevOps	482,957	360,650	5,319,776,510	0.56	11,015	82.9%
Security	425,193	261,813	1,436,727,161	0.53	3,379	77.2%
Configuration and Templates	370,301	258,140	856,135,980	0.45	2,312	81.8%
Development Tools	361,109	203,698	1,926,875,938	0.48	5,336	75.1%
OS and Platforms	356,741	180,554	682,089,698	0.47	1,912	58.3%
General Data Handling	325,652	211,294	2,492,538,843	0.38	7,654	76.2%
Software Engineering	322,105	206,260	5,649,729,492	0.40	17,540	73.7%
Back-End Web Development	308,481	171,495	587,655,412	0.46	1,905	66.6%
Front-End Web Development	303,391	141,752	700,226,773	0.48	2,308	64.5%
Documentation	277,836	156,394	1,188,581,911	0.39	4,278	67.9%
Community and Governance	275,448	126,018	91,173,314	0.33	331	65.9%

Back

Repository value and future popularity

	(1)	(2)	(3)	(4)	(5)
ln(Stars + 1)	0.123***	0.107***	0.089***	0.074***	0.066***
	(0.023)	(0.022)	(0.015)	(0.021)	(0.015)
ln(Mkt Cap)	1.815***	1.803***	1.734***	1.827***	1.667***
	(0.115)	(0.117)	(0.064)	(0.070)	(0.132)
ln(Volatility)	0.406***	0.596***	0.620***	0.463***	
	(0.056)	(0.049)	(0.035)	(0.022)	
ln(Employees)	-0.289*	0.102	0.196***	-0.157***	
	(0.151)	(0.146)	(0.053)	(0.046)	
ln(Total Patent Value + 1)	0.113	0.050	0.078**	0.211***	
	(0.075)	(0.048)	(0.029)	(0.028)	
Observations	28,388	28,388	28,388	28,388	28,388
Adj. R^2	0.782	0.800	0.813	0.850	0.858
Year FE	✓	✓			
Industry FE		✓			
Industry x Year FE			✓	✓	
Firm FE				✓	
Firm x Year FE					✓

- Valuable repositories are more popular in the future

Back

Robustness tests

	(1) Ex. Amazon, Microsoft, Alphabet	(2) Dep. = One-day ξ	(3) Dep. = R
ln(Stars + 1)	0.104***	0.088***	0.057*
	(0.012)	(0.015)	(0.027)
ln(Mkt Cap)	1.934***	1.781***	-0.200**
	(0.011)	(0.065)	(0.076)
ln(Volatility)	0.446***	0.608***	0.029
	(0.006)	(0.036)	(0.036)
ln(Employees)	0.051**	0.185**	0.258***
	(0.017)	(0.058)	(0.061)
ln(Total Patent Value + 1)	-0.009	0.067**	-0.003
	(0.015)	(0.028)	(0.050)
Observations	13,330	28,732	28,732
Adj. R^2	0.876	0.820	0.070
Industry x Year FE	✓	✓	✓