Discussion:

Data, Markups, and Asset Prices

Alexandre Corhay, Kejia Hu, Jun E. Li, Jincheng Tong, and Chi-Yang Tsou

Nancy R. Xu

University of Cambridge & CEPR

EFA, Ghent

August 22, 2026

Motivation

- Data has become a central input into firms' production, pricing, and strategic decisions.
- The literature has examined some first-order effects: **productivity**, **quality**, or **scale** effects.

Motivation

- Data has become a central input into firms' production, pricing, and strategic decisions.
- The literature has examined some first-order effects: **productivity**, **quality**, or **scale** effects.
- **Central ongoing debate:** Moving beyond first-order effects, the more data always the better? Why do firms invest in data, and how does it affect market power and expected stock returns?

Motivation

- Data has become a central input into firms' production, pricing, and strategic decisions.
- The literature has examined some first-order effects: **productivity**, **quality**, or **scale** effects.
- **Central ongoing debate:** Moving beyond first-order effects, the more data always the better? Why do firms invest in data, and how does it affect market power and expected stock returns?
- **Existing view:** Data is an **intangible input** that improves efficiency and productivity.
  - Data-driven decision-making and productivity: Brynjolfsson and McElheran (2019).
  - Data as an informational and nonrival input: Jones and Tonetti (2020); Farboodi and Veldkamp (2021).

Motivation

- Data has become a central input into firms' production, pricing, and strategic decisions.
- The literature has examined some first-order effects: **productivity**, **quality**, or **scale** effects.
- **Central ongoing debate:** Moving beyond first-order effects, the more data always the better? Why do firms invest in data, and how does it affect market power and expected stock returns?
- **Existing view:** Data is an **intangible input** that improves efficiency and productivity.
  - Data-driven decision-making and productivity: Brynjolfsson and McElheran (2019).
  - Data as an informational and nonrival input: Jones and Tonetti (2020); Farboodi and Veldkamp (2021).
- **This paper:** Implications of data scientists on firm risk valuation.

What would an Asset Pricer predict?

What would an Asset Pricer predict?

I.e., me... :)

What would an Asset Pricer predict?

I.e., me... :)

1. **Clear prediction:** Data $\rightarrow$ better information about consumers $\rightarrow$ higher markups

What would an Asset Pricer predict?

I.e., me... :)

1. **Clear prediction:** Data $\rightarrow$ better information about consumers $\rightarrow$ higher markups
2. **Unclear prediction:** Risk implications

What would an Asset Pricer predict?

I.e., me... :)

1. **Clear prediction:** Data → better information about consumers → higher markups
2. **Unclear prediction:** Risk implications

Effect of having more data	**Implication for risk**
Better information allows higher markups and helps profitability in bad times	**Hedge** → lower risk
	Operating leverage → higher risk
Data scientists are a relatively persistent labor commitment. The firm pays their wages even when aggregate conditions deteriorate; higher cash flow risks.	

The paper addresses this tension.

What would an Asset Pricer predict?

I.e., me... :)

1. **Clear prediction:** Data → better information about consumers → higher markups
2. **Unclear prediction:** Risk implications

Effect of having more data	**Implication for risk**
Better information allows higher markups and helps profitability in bad times	**Hedge** → lower risk
	Operating leverage → higher risk
Data scientists are a relatively persistent labor commitment. The firm pays their wages even when aggregate conditions deteriorate; higher cash flow risks.	

What would an Asset Pricer predict?

I.e., me... :)

1. **Clear prediction:** Data $\rightarrow$ better information about consumers $\rightarrow$ higher markups
2. **Unclear prediction:** Risk implications

Effect of having more data	**Implication for risk**
Better information allows higher markups and helps profitability in bad times	**Hedge** $\rightarrow$ lower risk
	Operating leverage $\rightarrow$ higher risk
Data scientists are a relatively persistent labor commitment. The firm pays their wages even when aggregate conditions deteriorate; higher cash flow risks.	

The paper addresses this tension.

Model: From data to markups and risk

1. More data scientists improve demand forecasting

$$\underbrace{\omega_{it}}_{\text{data produced}} = \nu_0 N_{it}^{\nu_1} Y_{i,t-1}^{1-\nu_1}, \qquad \underbrace{\Omega_{it}}_{\text{forecast precision}} = \underbrace{\sigma_\theta^{-2}}_{\text{prior precision}} + \underbrace{\sigma_s^{-2}\omega_{it}}_{\text{precision from data}}.$$

N_{it}: N_{it}: share of data scientists in firm i's total workforce;; $Y_{i,t-1}$: lagged sales.

Model: From data to markups and risk

1. More data scientists improve demand forecasting

$$\underbrace{\omega_{it}}_{\text{data produced}} = \nu_0 N_{it}^{\nu_1} Y_{i,t-1}^{1-\nu_1}, \qquad \underbrace{\Omega_{it}}_{\text{forecast precision}} = \underbrace{\sigma_\theta^{-2}}_{\text{prior precision}} + \underbrace{\sigma_s^{-2}\omega_{it}}_{\text{precision from data}}.$$

N_{it}: N_{it}: share of data scientists in firm i's total workforce;; $Y_{i,t-1}$: lagged sales.

2. Better forecasts reduce effective demand elasticity

$$\underbrace{H_{it}}_{\text{inelastic demand}} = \bar{\Theta} - \Omega_{it}^{-1}, \qquad \underbrace{\varepsilon_{it}^{\text{eff}}}_{\text{effective demand elasticity}} = \eta\left(1 - \frac{H_{it}}{Y_{it}}\right).$$

$$N_{it} \uparrow \Rightarrow \Omega_{it} \uparrow \Rightarrow H_{it} \uparrow \Rightarrow \varepsilon_{it}^{\text{eff}} \downarrow \Rightarrow \underbrace{\mu_{it}}_{\textbf{markup}} \uparrow$$

Model: From data to markups and risk

1. More data scientists improve demand forecasting

$$\underbrace{\omega_{it}}_{\text{data produced}} = \nu_0 N_{it}^{\nu_1} Y_{i,t-1}^{1-\nu_1}, \qquad \underbrace{\Omega_{it}}_{\text{forecast precision}} = \underbrace{\sigma_\theta^{-2}}_{\text{prior precision}} + \underbrace{\sigma_s^{-2}\omega_{it}}_{\text{precision from data}}.$$

N_{it}: N_{it}: share of data scientists in firm i's total workforce;; $Y_{i,t-1}$: lagged sales.

2. Better forecasts reduce effective demand elasticity

$$\underbrace{H_{it}}_{\text{inelastic demand}} = \bar{\Theta} - \Omega_{it}^{-1}, \qquad \underbrace{\varepsilon_{it}^{\text{eff}}}_{\text{effective demand elasticity}} = \eta\left(1 - \frac{H_{it}}{Y_{it}}\right).$$

$$N_{it} \uparrow \Rightarrow \Omega_{it} \uparrow \Rightarrow H_{it} \uparrow \Rightarrow \varepsilon_{it}^{\text{eff}} \downarrow \Rightarrow \underbrace{\mu_{it}}_{\textbf{markup}} \uparrow$$

3. Persistent DS employment creates operating leverage

Assuming persistent and costly-to-adjust DS employment, countercyclical hiring amplifies cash-flow risk in downturns.

Model: From data to markups and risk

1. More data scientists improve demand forecasting

$$\underbrace{\omega_{it}}_{\text{data produced}} = \nu_0 N_{it}^{\nu_1} Y_{i,t-1}^{1-\nu_1}, \qquad \underbrace{\Omega_{it}}_{\text{forecast precision}} = \underbrace{\sigma_\theta^{-2}}_{\text{prior precision}} + \underbrace{\sigma_s^{-2}\omega_{it}}_{\text{precision from data}}.$$

N_{it}: N_{it}: share of data scientists in firm i's total workforce;; $Y_{i,t-1}$: lagged sales.

2. Better forecasts reduce effective demand elasticity

$$\underbrace{H_{it}}_{\text{inelastic demand}} = \bar{\Theta} - \Omega_{it}^{-1}, \qquad \underbrace{\varepsilon_{it}^{\text{eff}}}_{\text{effective demand elasticity}} = \eta\left(1 - \frac{H_{it}}{Y_{it}}\right).$$

$$N_{it} \uparrow \Rightarrow \Omega_{it} \uparrow \Rightarrow H_{it} \uparrow \Rightarrow \varepsilon_{it}^{\text{eff}} \downarrow \Rightarrow \underbrace{\mu_{it}}_{\textbf{markup}} \uparrow$$

3. Persistent DS employment creates operating leverage

Assuming persistent and costly-to-adjust DS employment, countercyclical hiring amplifies cash-flow risk in downturns.

Testable predictions

1. **Information quality:** More DS hiring improves management's sales-forecast accuracy.
2. **Market power:** A higher DS ratio is associated with higher markups.
3. **Risk and returns:** Countercyclical DS hiring creates operating leverage; if this dominates the markup hedge, high-DS firms earn higher expected returns.

Empirical and quantitative design

1. Data and sample

- Revelio Labs covers 150 occupations and is comprehensive since 2008; the final annual firm-level sample covers **2008–2021** and combines Revelio data with Compustat, CRSP, IBES Guidance, and Lightcast job postings.
- **29,462 firm-year observations** with non-missing DS ratios; mean = 6.90%, median = 4.44%, and P95 = 21.00%.

2. Test the model's three predictions

- **Information quality:** Does DS hiring improve management's forecast-revision accuracy?

$$\mathrm{RI}_{i,t+h} = a_i + \delta_t + b\,\mathrm{DS\ Hiring\ Rate}_{i,t} + c\,\mathrm{Controls}_{i,t} + \varepsilon_{i,t}, \qquad h \in \{0, 1\}.$$

- **Market power:** Is a higher DS ratio associated with higher De Loecker-style markups?

$$\mathrm{Markup}_{i,t} = a_i + \delta_t + b\,\mathrm{DS}_{i,t} + c\,\mathrm{Controls}_{i,t} + \varepsilon_{i,t}.$$

- **Expected returns:** Do high-DS firms earn higher subsequent returns in portfolio sorts and Fama–MacBeth regressions?

$$R_{i,t+1} - R_{f,t+1} = a_j + b\,\mathrm{DS\ Ratio}_{i,t} + c\,\mathrm{Controls}_{i,t} + \varepsilon_{i,t}.$$

3. Quantitative model

- Calibrate and simulate the model to see whether its data-learning and operating-leverage channels reproduce the observed markup and return patterns.

Main findings

$$DS\ Ratio_{it} = \frac{\text{Number of data scientists/data workers at firm } i}{\text{Total employees at firm } i}$$

Main findings

$$DS\ Ratio_{it} = \frac{\text{Number of data scientists/data workers at firm } i}{\text{Total employees at firm } i}$$

Prediction	**Evidence**	**Estimate**	***t*-stat.**
1. Information quality	DS hiring → forecast-revision improvement	0.215–0.352	1.79–3.27
2. Market power	One-SD higher DS ratio → markup	+0.155	3.70
3. Expected returns	DS-ratio high-minus-low excess return	4.13%	2.00
Risk-adjusted returns	Fama–French five-factor alpha	4.81%	3.03
Quantitative model	Model high-minus-low return spread	4.54%	—

- The empirical patterns line up with the model's three central predictions.
- The calibrated model closely matches the observed return spread.

A very interesting, timely paper!

The paper offers a novel view: more data can increase both market power and risk.

My comments:

1. Alternative interpretation for the risk story
2. A persistent labor commitment assumption

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

- The paper interprets DS employment as firms' capacity to learn about demand.

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

- The paper interprets DS employment as firms' capacity to learn about demand.
- But a higher DS ratio may also reflect stronger exposure to local digital-talent supply, wage pressure, and worker mobility (uncertainty).
- It may correlate with firms undergoing uncertain technological transformation.

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

- The paper interprets DS employment as firms' capacity to learn about demand.
- But a higher DS ratio may also reflect stronger exposure to local digital-talent supply, wage pressure, and worker mobility (uncertainty).
- It may correlate with firms undergoing uncertain technological transformation.

Suggestions:

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

- The paper interprets DS employment as firms' capacity to learn about demand.
- But a higher DS ratio may also reflect stronger exposure to local digital-talent supply, wage pressure, and worker mobility (uncertainty).
- It may correlate with firms undergoing uncertain technological transformation.

Suggestions:

- Decompose DS growth into a firm component (labor demand) and a local component (labor supply). Alternatively, add area-time fixed effects. [Control for the alternative story]

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

- The paper interprets DS employment as firms' capacity to learn about demand.
- But a higher DS ratio may also reflect stronger exposure to local digital-talent supply, wage pressure, and worker mobility (uncertainty).
- It may correlate with firms undergoing uncertain technological transformation.

Suggestions:

- Decompose DS growth into a firm component (labor demand) and a local component (labor supply). Alternatively, add area-time fixed effects. [Control for the alternative story]
- Test whether the DS risk premium is stronger where digital-talent supply is less elastic or technological uncertainty is higher. [For your story, you want the former]

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

- The paper interprets DS employment as firms' capacity to learn about demand.
- But a higher DS ratio may also reflect stronger exposure to local digital-talent supply, wage pressure, and worker mobility (uncertainty).
- It may correlate with firms undergoing uncertain technological transformation.

Suggestions:

- Decompose DS growth into a firm component (labor demand) and a local component (labor supply). Alternatively, add area-time fixed effects. [Control for the alternative story]
- Test whether the DS risk premium is stronger where digital-talent supply is less elastic or technological uncertainty is higher. [For your story, you want the former]
- Compare demand-analytics with software/infrastructure roles: only the former should strongly predict demand forecasts and markups under demand learning.
  - Sidenote: DS employees may perform different functions: (1) demand-facing analytics, such as customer analytics, pricing, and sales forecasting; and (2) general software and data-infrastructure tasks, such as engineering, database management, and IT systems.

Comment 1: Economic interpretations of DS

Alternative interpretation: DS may capture labor-market and technology risk.

- The paper interprets DS employment as firms' capacity to learn about demand.
- But a higher DS ratio may also reflect stronger exposure to local digital-talent supply, wage pressure, and worker mobility (uncertainty).
- It may correlate with firms undergoing uncertain technological transformation.

Suggestions:

- Decompose DS growth into a firm component (labor demand) and a local component (labor supply). Alternatively, add area-time fixed effects. [Control for the alternative story]
- Test whether the DS risk premium is stronger where digital-talent supply is less elastic or technological uncertainty is higher. [For your story, you want the former]
- Compare demand-analytics with software/infrastructure roles: only the former should strongly predict demand forecasts and markups under demand learning.
  - Sidenote: DS employees may perform different functions: (1) demand-facing analytics, such as customer analytics, pricing, and sales forecasting; and (2) general software and data-infrastructure tasks, such as engineering, database management, and IT systems.

Comment 2: A persistent commitment assumption

An important assumption: There is a persistent DS labor commitment.

Comment 2: A persistent commitment assumption

An important assumption: There is a persistent DS labor commitment.

- Main story: Data scientists are a relatively persistent labor commitment $\Rightarrow$ The firm pays their wages even when aggregate conditions deteriorate $\Rightarrow$ higher cash flow risks.

Comment 2: A persistent commitment assumption

An important assumption: There is a persistent DS labor commitment.

- Main story: Data scientists are a relatively persistent labor commitment $\Rightarrow$ The firm pays their wages even when aggregate conditions deteriorate $\Rightarrow$ higher cash flow risks.
- Missing evidence: Is DS really having a persistent commitment?

Comment 2: A persistent commitment assumption

An important assumption: There is a persistent DS labor commitment.

- Main story: Data scientists are a relatively persistent labor commitment $\Rightarrow$ The firm pays their wages even when aggregate conditions deteriorate $\Rightarrow$ higher cash flow risks.
- Missing evidence: Is DS really having a persistent commitment?

Suggestions:

- Construct a DS-specific wage bill using headcount and occupation/location wages; track it around negative TFP, sales, and aggregate shocks.

Comment 2: A persistent commitment assumption

An important assumption: There is a persistent DS labor commitment.

- Main story: Data scientists are a relatively persistent labor commitment $\Rightarrow$ The firm pays their wages even when aggregate conditions deteriorate $\Rightarrow$ higher cash flow risks.
- Missing evidence: Is DS really having a persistent commitment?

Suggestions:

- Construct a DS-specific wage bill using headcount and occupation/location wages; track it around negative TFP, sales, and aggregate shocks.
- Construct a DS-specific commitment measure; the return spread results should be stronger for firms with stronger commitment measure.

Comment 3: Minor comments

1. **Cross-reference:** Table 3 should refer to Equation (14), rather than Equation (13).
2. **Internal consistency:** In Section 4.3, the discussion after Table 10 states that low-productivity firms hire more DS and earn "lower expected returns." Since the model reports a positive 4.54% high-minus-low DS return spread, "lower" should likely be "higher."
3. **Notation and proofreading:** In Section 4.1, the DS adjustment-cost function is mistakenly written as $G^N(I_{it}, K_{it})$; its arguments should be $G^N(H_{it}, N_{it})$, consistent with DS hiring and employment. A final proofread could also correct spelling errors such as "amplifiying" and "orthogonoalized."

These are purely editorial points and do not affect the paper's main results.

Conclusion

Overall: The paper offers a novel framework in which data can increase both **market power** and **firm risk**.

1. Clarify the economic content of the DS ratio.
2. Validate the DS labor-commitment assumption.

Thank you!

nancy.xu@jbs.cam.ac.uk