

Discussion

The Macroeconomics of Data: Scale, Product Choice, and Pricing in the Information Age

(Asriyan-Kohlhas)

Joachim Hubmer

University of Pennsylvania

Frontier Topics in Macro-Finance, Chicago, May 2025

Table of Contents

The Paper

Broader Perspective on Production and Business Dynamism

Background and Motivation

- ▶ Firms increasingly invest in **intangible assets**
 - ▶ Intangible \approx tangible assets by 2019 (Crouzet-Eberly-Eisfeld-Papanikolaou 2022)
 - ▶ Effect on firm behavior, productivity, competition?

Background and Motivation

- ▶ Firms increasingly invest in **intangible assets**
 - ▶ Intangible \approx tangible assets by 2019 (Crouzet-Eberly-Eisfeld-Papanikolaou 2022)
 - ▶ Effect on firm behavior, productivity, competition?
- ▶ **Data-processing technologies** have become central production inputs
(Veldkamp-Chung 2024)

Background and Motivation

- ▶ Firms increasingly invest in **intangible assets**
 - ▶ Intangible \approx tangible assets by 2019 (Crouzet-Eberly-Eisfeld-Papanikolaou 2022)
 - ▶ Effect on firm behavior, productivity, competition?
- ▶ **Data-processing technologies** have become central production inputs
(Veldkamp-Chung 2024)
- ▶ Firms increasingly use data to optimize **scale**, **product choice**, and **pricing**
(Asriyan-Kohlhas 2025)

Background and Motivation

- ▶ Firms increasingly invest in **intangible assets**
 - ▶ Intangible \approx tangible assets by 2019 (Crouzet-Eberly-Eisfeld-Papanikolaou 2022)
 - ▶ Effect on firm behavior, productivity, competition?
- ▶ **Data-processing technologies** have become central production inputs
(Veldkamp-Chung 2024)
- ▶ Firms increasingly use data to optimize **scale**, **product choice**, and **pricing**
(Asriyan-Kohlhas 2025)
- ▶ Key tension: data **improves efficiency** but also facilitates **rent extraction**

Background and Motivation

- ▶ Firms increasingly invest in **intangible assets**
 - ▶ Intangible \approx tangible assets by 2019 (Crouzet-Eberly-Eisfeld-Papanikolaou 2022)
 - ▶ Effect on firm behavior, productivity, competition?
- ▶ **Data-processing technologies** have become central production inputs
(Veldkamp-Chung 2024)
- ▶ Firms increasingly use data to optimize **scale**, **product choice**, and **pricing**
(Asriyan-Kohlhas 2025)
- ▶ Key tension: data **improves efficiency** but also facilitates **rent extraction**

⇒ Goal: Assess **net effect** of productivity gains vs. rent extraction

Summary of Approach

- ▶ Measure **forecast accuracy** using I/B/E/S forecasts vs. realized revenue (y)
 - ▶ Forecast Error $_{i,t} = \left| \frac{y_{i,t} - \mathbb{E}_{i,t-1}[y_{i,t}]}{k_{i,t-1}} \right|^n$, $n \in \{1, 2\}$
 - ▶ Track forecast errors over time and by firm size

Summary of Approach

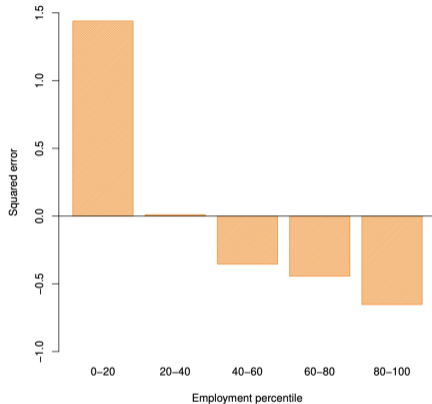
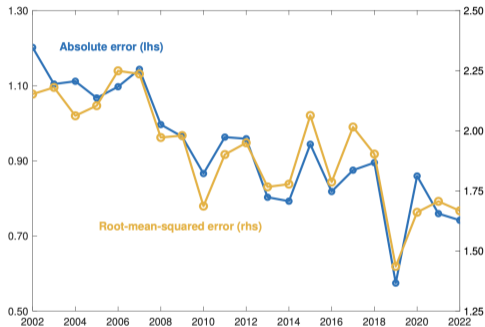
- ▶ Measure **forecast accuracy** using I/B/E/S forecasts vs. realized revenue (y)
 - ▶ Forecast Error $_{i,t} = \left| \frac{y_{i,t} - \mathbb{E}_{i,t-1}[y_{i,t}]}{k_{i,t-1}} \right|^n$, $n \in \{1, 2\}$
 - ▶ Track forecast errors over time and by firm size
- ▶ Develop GE model of firms choosing **scale**, **product**, and **pricing** under uncertainty
 - ▶ Firms uncertain about idiosyncratic TFP (ν): scale
 - ▶ Firms uncertain about demand state (ω): product choice
 - ▶ Model extension with price-discrimination to include rent extraction channel

Summary of Approach

- ▶ Measure **forecast accuracy** using I/B/E/S forecasts vs. realized revenue (y)
 - ▶ Forecast Error $_{i,t} = \left| \frac{y_{i,t} - \mathbb{E}_{i,t-1}[y_{i,t}]}{k_{i,t-1}} \right|^n$, $n \in \{1, 2\}$
 - ▶ Track forecast errors over time and by firm size
- ▶ Develop GE model of firms choosing **scale**, **product**, and **pricing** under uncertainty
 - ▶ Firms uncertain about idiosyncratic TFP (ν): scale
 - ▶ Firms uncertain about demand state (ω): product choice
 - ▶ Model extension with price-discrimination to include rent extraction channel
- ▶ Quantify effects of data advances on **TFP** and **welfare** using calibrated model

Finding 1: Empirical Motivation

Figure 1: Time Evolution of Revenue Accuracy



Panel (a): Relative Accuracy

► Forecast accuracy improves 2002–2022

► Larger firms have more accurate forecasts

Finding 2: Data Improves TFP in the Baseline Economy

- ▶ Firms face a **fixed cost of acquiring information**
- ▶ Large firms invest more in information optimally
 - ▶ Fixed cost χ calibrated to match 10% informed firms
⇒ model reproduces **firm size–accuracy gradient**

Finding 2: Data Improves TFP in the Baseline Economy

- ▶ Firms face a **fixed cost of acquiring information**
- ▶ Large firms invest more in information optimally
 - ▶ Fixed cost χ calibrated to match 10% informed firms
⇒ model reproduces **firm size–accuracy gradient**
- ▶ Over time: **Data-processing technologies improve**, reducing χ for everyone
 - ▶ Decline in fixed cost calibrated to match **time trend** in forecast accuracy

Finding 2: Data Improves TFP in the Baseline Economy

- ▶ Firms face a **fixed cost of acquiring information**
- ▶ Large firms invest more in information optimally
 - ▶ Fixed cost χ calibrated to match 10% informed firms
⇒ model reproduces **firm size–accuracy gradient**
- ▶ Over time: **Data-processing technologies improve**, reducing χ for everyone
 - ▶ Decline in fixed cost calibrated to match **time trend** in forecast accuracy
- ▶ Improved information leads to:
 - ▶ **Better scale choices** and **better product targeting**
 - ▶ “Aggregate TFP” $\equiv \frac{C}{L_{prod}}$ increases by 6.7%
 - ▶ “Welfare” $\equiv C$ increases by 2.1%

Finding 2: Data Improves TFP in the Baseline Economy

- ▶ Firms face a **fixed cost of acquiring information**
- ▶ Large firms invest more in information optimally
 - ▶ Fixed cost χ calibrated to match 10% informed firms
⇒ model reproduces **firm size–accuracy gradient**
- ▶ Over time: **Data-processing technologies improve**, reducing χ for everyone
 - ▶ Decline in fixed cost calibrated to match **time trend** in forecast accuracy
- ▶ Improved information leads to:
 - ▶ **Better scale choices** and **better product targeting**
 - ▶ “Aggregate TFP” $\equiv \frac{C}{L_{prod}}$ increases by 6.7%
 - ▶ “Welfare” $\equiv C$ increases by 2.1%
 - ▶ Comment: fixed $\bar{L} = L_{prod} + L_{info} \Rightarrow$ difference is data processing cost (L_{info})
⇒ NIPA measured $TFP \approx \frac{C}{\bar{L}} + 2.1\%$ (info-processing is intermediate input)
⇒ Suggestion: De-emphasize framing of TFP vs welfare dis-connect

Finding 2: Data Improves TFP in the Baseline Economy

- ▶ Firms face a **fixed cost of acquiring information**
- ▶ Large firms invest more in information optimally
 - ▶ Fixed cost χ calibrated to match 10% informed firms
⇒ model reproduces **firm size–accuracy gradient**
- ▶ Over time: **Data-processing technologies improve**, reducing χ for everyone
 - ▶ Decline in fixed cost calibrated to match **time trend** in forecast accuracy
- ▶ Improved information leads to:
 - ▶ **Better scale choices** and **better product targeting**
 - ▶ “Aggregate TFP” $\equiv \frac{C}{L_{prod}}$ increases by 6.7%
 - ▶ “Welfare” $\equiv C$ increases by 2.1%
 - ▶ Comment: fixed $\bar{L} = L_{prod} + L_{info} \Rightarrow$ difference is data processing cost (L_{info})
⇒ NIPA measured $TFP \approx \frac{C}{\bar{L}} + 2.1\%$ (info-processing is intermediate input)
⇒ Suggestion: De-emphasize framing of TFP vs welfare dis-connect

Finding 3: Data Enables Rent Extraction

- ▶ Rent-extraction model: firms also use info to **price discriminate** (via quantities)
 - ▶ Extra choice: information affects **how prices are set**, not what is produced

Finding 3: Data Enables Rent Extraction

- ▶ Rent-extraction model: firms also use info to **price discriminate** (via quantities)
 - ▶ Extra choice: information affects **how prices are set**, not what is produced
- ▶ Key insight: **Information may be overproduced** in equilibrium
 - ▶ Information boosts firm profits but **does not raise total surplus as much**
 - ▶ Result: **welfare loss from excessive information use** (not about redistribution)

Finding 3: Data Enables Rent Extraction

- ▶ Rent-extraction model: firms also use info to **price discriminate** (via quantities)
 - ▶ Extra choice: information affects **how prices are set**, not what is produced
- ▶ Key insight: **Information may be overproduced** in equilibrium
 - ▶ Information boosts firm profits but **does not raise total surplus as much**
 - ▶ Result: **welfare loss from excessive information use** (not about redistribution)
- ▶ Baseline and rent-extraction models are **observationally equivalent**
 - ▶ Same firm-level choices and revenue—difference lies in use of information

Finding 3: Data Enables Rent Extraction

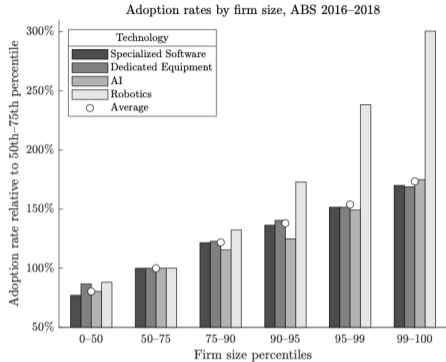
- ▶ Rent-extraction model: firms also use info to **price discriminate** (via quantities)
 - ▶ Extra choice: information affects **how prices are set**, not what is produced
- ▶ Key insight: **Information may be overproduced** in equilibrium
 - ▶ Information boosts firm profits but **does not raise total surplus as much**
 - ▶ Result: **welfare loss from excessive information use** (not about redistribution)
- ▶ Baseline and rent-extraction models are **observationally equivalent**
 - ▶ Same firm-level choices and revenue—difference lies in use of information
- ▶ Quantitative result:
 - ▶ **TFP gains** 5.3–6.7%
 - ▶ But **welfare gains only 0.1–2.1%**: sensitive to rent extraction

Table of Contents

The Paper

Broader Perspective on Production and Business Dynamism

More Generally True: Larger Firms Use More Advanced Capital Technologies

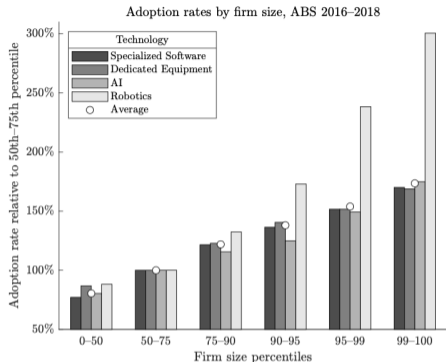


► Automation techn adoption skewed

e.g. AI, robotics, software

Hubmer-Restrepo 2024, using ABS data

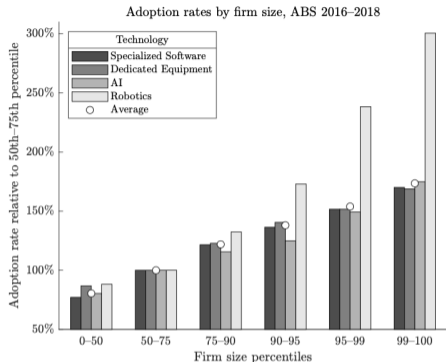
More Generally True: Larger Firms Use More Advanced Capital Technologies



- Consistent with AK data and story: larger firms generally automate more

- Automation techn adoption skewed
e.g. AI, robotics, software
Hubmer-Restrepo 2024, using ABS data

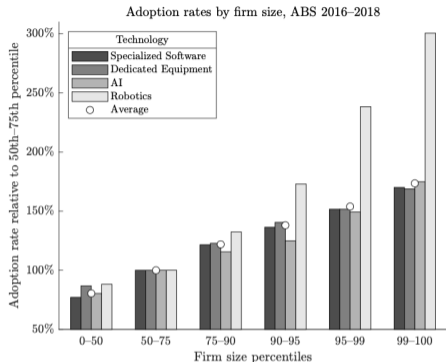
More Generally True: Larger Firms Use More Advanced Capital Technologies



- ▶ Consistent with AK data and story: larger firms generally automate more
- ▶ Same mechanism: fixed cost —alternatives?

- ▶ Automation techn adoption skewed
e.g. AI, robotics, software
Hubmer-Restrepo 2024, using ABS data

More Generally True: Larger Firms Use More Advanced Capital Technologies

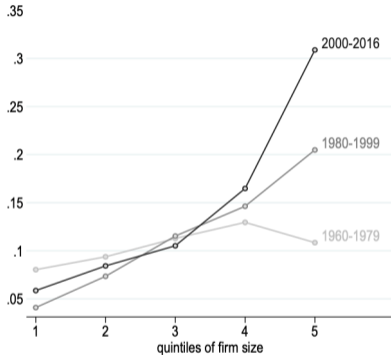


- ▶ Automation techn adoption skewed
e.g. AI, robotics, software
Hubmer-Restrepo 2024, using ABS data

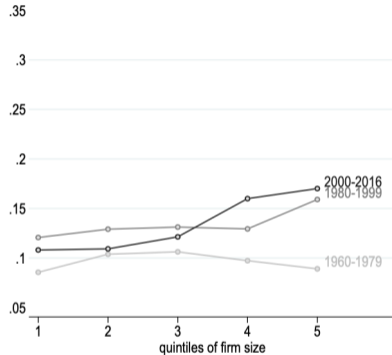
- ▶ Consistent with AK data and story:
larger firms generally automate more
- ▶ Same mechanism: fixed cost
—alternatives?
- ▶ Tiny suggestion: normalize forecast
errors by revenue, not capital, since
 y/k falls with firm size

Capital Intensity (Techn Adoption?) Gradient Becomes Steeper Over Time

Output-to-capital elasticity:
manufacturing firms in Compustat

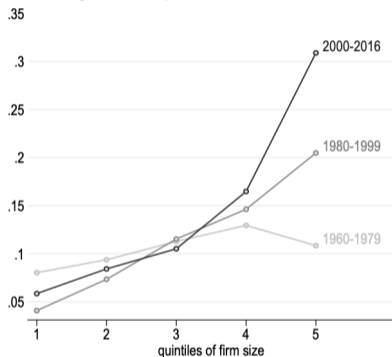


Output-to-capital elasticity:
non-manufacturing firms in Compustat

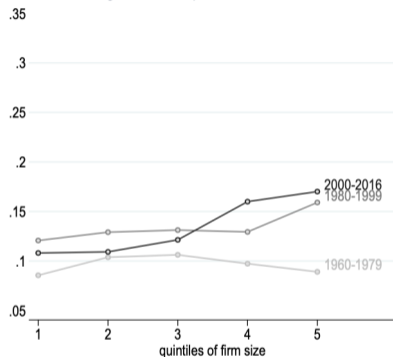


Capital Intensity (Techn Adoption?) Gradient Becomes Steeper Over Time

Output-to-capital elasticity:
manufacturing firms in Compustat



Output-to-capital elasticity:
non-manufacturing firms in Compustat



- Suggestion: Dis-entangle **time** vs. **firm size** in forecasting more directly:

$$\text{ForecastError}_{i,t+1} = \alpha + \beta_1 \cdot t + \beta_2 \cdot \text{size}_{i,t} + \beta_3 \cdot t \cdot \text{size}_{i,t} + \varepsilon_{i,t+1}$$

Can do (indirect) inference on β_n to inform model

Relation to Broader Business Dynamism Discussion

- ▶ Decker-Haltiwanger-Jarmin-Miranda 2020:

$$\underbrace{\text{Reallocation Rate}}_{\downarrow \text{ over time}} = \underbrace{\text{Dispersion of Shocks}}_{\text{flat or } \uparrow} \times \underbrace{\text{Firm Responsiveness}}_{\downarrow \text{ over time}}$$

Shock dispersion has not fallen, but **responsiveness** ↓

Relation to Broader Business Dynamism Discussion

- ▶ Decker-Haltiwanger-Jarmin-Miranda 2020:

$$\underbrace{\text{Reallocation Rate}}_{\downarrow \text{ over time}} = \underbrace{\text{Dispersion of Shocks}}_{\text{flat or } \uparrow} \times \underbrace{\text{Firm Responsiveness}}_{\downarrow \text{ over time}}$$

Shock dispersion has not fallen, but **responsiveness** ↓

- ▶ AK: Shocks not smaller, **forecast errors** ↓ ⇒ **information precision** ↑

Relation to Broader Business Dynamism Discussion

- ▶ Decker-Haltiwanger-Jarmin-Miranda 2020:

$$\underbrace{\text{Reallocation Rate}}_{\downarrow \text{ over time}} = \underbrace{\text{Dispersion of Shocks}}_{\text{flat or } \uparrow} \times \underbrace{\text{Firm Responsiveness}}_{\downarrow \text{ over time}}$$

Shock dispersion has not fallen, but **responsiveness** ↓

- ▶ AK: Shocks not smaller, **forecast errors** ↓ ⇒ **information precision** ↑
- ▶ Puzzle: If firms know more, shouldn't **responsiveness** ↑?
 - ▶ Simple logic: if firms maximize expected profits

$$\max_{\ell_t} \mathbb{E}_{t-1} \left[\underbrace{z_t \cdot \ell_t^\eta}_{\text{rev prod fun}} - w \cdot \ell_t \right] \quad \text{with } \eta < 1,$$

and z is more accurately known, then **input** ℓ should **respond more** to changes in z

Relation to Broader Business Dynamism Discussion

- ▶ Decker-Haltiwanger-Jarmin-Miranda 2020:

$$\underbrace{\text{Reallocation Rate}}_{\downarrow \text{ over time}} = \underbrace{\text{Dispersion of Shocks}}_{\text{flat or } \uparrow} \times \underbrace{\text{Firm Responsiveness}}_{\downarrow \text{ over time}}$$

Shock dispersion has not fallen, but **responsiveness** ↓

- ▶ AK: Shocks not smaller, **forecast errors** ↓ ⇒ **information precision** ↑
- ▶ Puzzle: If firms know more, shouldn't **responsiveness** ↑?
 - ▶ Simple logic: if firms maximize expected profits

$$\max_{\ell_t} \mathbb{E}_{t-1} \left[\underbrace{z_t \cdot \ell_t^\eta}_{\text{rev prod fun}} - w \cdot \ell_t \right] \quad \text{with } \eta < 1,$$

and z is more accurately known, then **input ℓ should respond more to changes in z**

- ▶ Possible reconciliation: measured responsiveness understates true responsiveness
 - ▶ DeLeon-Macaluso-Yeh 2024: regular labor sticky; but staffed labor responds more

Responsiveness May Operate Through Intermediate Inputs

- ▶ Mertens-Schoefer 2025: Firms grow by scaling up **intermediate input use**, not labor

Responsiveness May Operate Through Intermediate Inputs

- ▶ Mertens-Schoefer 2025: Firms grow by scaling up **intermediate input use**, not labor
- ▶ Framework:

$$y = z \cdot f(k, \ell, m)$$

Finding: positive demand shock \Rightarrow little increase in k, ℓ , strong increase in m

Responsiveness May Operate Through Intermediate Inputs

- ▶ Mertens-Schoefer 2025: Firms grow by scaling up **intermediate input use**, not labor

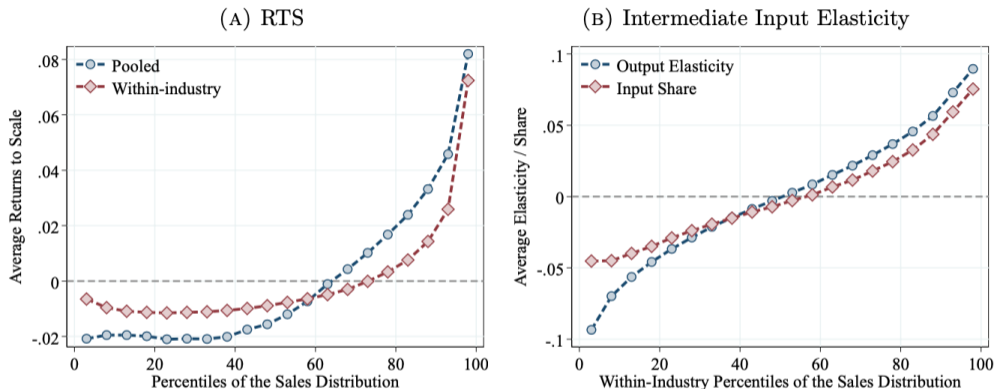
- ▶ Framework:

$$y = z \cdot f(k, \ell, m)$$

Finding: positive demand shock \Rightarrow little increase in k, ℓ , strong increase in m

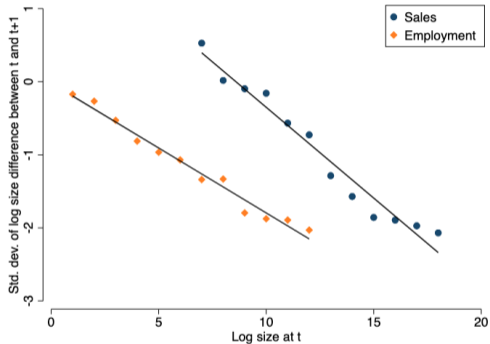
- ▶ Suggestion: Study responsiveness in **intermediate input use**
 - ▶ Is intermediate input volatility increasing with information precision?
 - ▶ Does it vary with firm size and over time?
 - ▶ Put measure of Δ (flexible) intermediate input on LHS of suggested regression
- ▶ Could reconcile high forecast accuracy with flat labor responsiveness...
 - ▶ ... if **adjustment occurs on m , not ℓ**

Larger Firms Use More Intermediates—and Have Higher Returns to Scale



- ▶ Hubmer-Chan-Ozkan-Salgado-Hong 2024: Larger firms have higher RTS in cross-section
 - ▶ Driven by greater use of intermediate inputs m
 - ▶ Also suggestive of responsiveness via m
 - ▶ In addition: large firms have higher RTS $\eta < 1 \Rightarrow \frac{d \ln y}{d \ln z} = \frac{1}{1-\eta}$ higher (w/o friction)

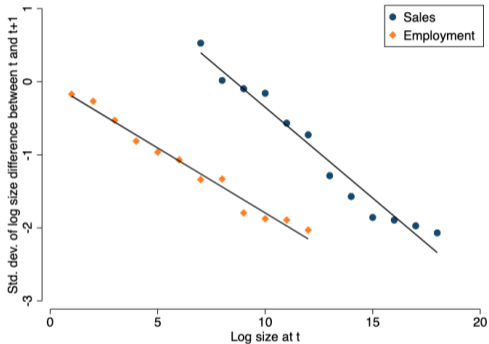
Another Reflection: Larger Firms Are Less Volatile Overall



(a) Data

- ▶ Boehm-Oberfield-South-Waseem 2024: Larger firms are systematically **less volatile**—violates Gibrat's law
 - ▶ Diversification (more customers)

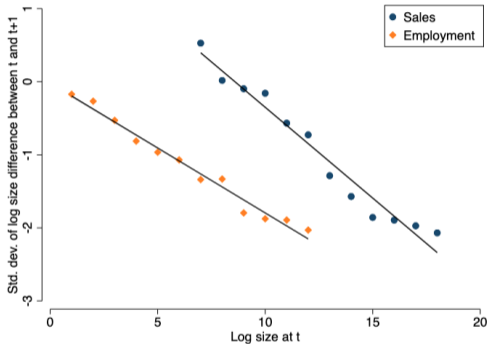
Another Reflection: Larger Firms Are Less Volatile Overall



(a) Data

- ▶ Boehm-Oberfield-South-Waseem 2024: Larger firms are systematically **less volatile**—violates Gibrat's law
 - ▶ Diversification (more customers)
- ▶ Could partly explain why **larger firms have lower forecast errors** in AK

Another Reflection: Larger Firms Are Less Volatile Overall



(a) Data

- ▶ Boehm-Oberfield-South-Waseem 2024: Larger firms are systematically **less volatile**—violates Gibrat's law
 - ▶ Diversification (more customers)
- ▶ Could partly explain why **larger firms have lower forecast errors** in AK
- ▶ Again, suggested regression separates:
 - ▶ **Time effect** (better info)
 - ▶ **Size effect** (less vola or better info)
 - ▶ **Interaction effect** (better info!)

- ▶ Very insightful and timely paper
- ▶ Opens new ground **connecting data processing improvements to macro outcomes**
- ▶ Some empirical tensions (puzzle!) with broader business dynamism literature
⇒ fruitful avenues for future research