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

Vladimir Asriyan[1] and Alexandre Kohlhas[2]

[1]CREI

[2]University of Oxford

Frontier Topics in Macro-Finance
Chicago, May 1-2, 2025

Motivation

Common perception: use of data by firms has exploded in recent decades

- Largest firms rely heavily on data
- Surveys suggest increasing adoption of data-driven decision making:
  - how much to produce
  - what to produce
  - how to price products

  Brynjolfsson and McElheran (2016, 2024), McKinsey (2022, 2023)

Aggregate implications of this rise of data use not well understood

This paper: empirical evidence + macroeconomic framework
from positive & normative perspectives

Motivation

Common perception: use of data by firms has exploded in recent decades

- Largest firms rely heavily on data
- Surveys suggest increasing adoption of data-driven decision making:
    - how much to produce
    - what to produce
    - how to price products

  Brynjolfsson and McElheran (2016, 2024), McKinsey (2022, 2023)

Aggregate implications of this rise of data use not well understood

This paper: empirical evidence + macroeconomic framework
from positive & normative perspectives

This Paper

Premise: data is information that helps firms predict economic fundamentals

1. Evidence: substantial increase in accuracy of firms' expectations
   - Larger firms are more accurate than smaller ones
   - Improvements in accuracy closely linked to firm-size dynamics
   - Accuracy impacts firm choices and outcomes

This Paper

Premise: data is information that helps firms predict economic fundamentals

2. Theory: a macroeconomic model of information production

- Information helps firms learn...

  — what to produce, how much to produce, and how to price products

- Dual role of information: it boosts...

  — efficiency of resource allocation within and across firms

  — rent-extraction by firms, which may be socially harmful

- Laissez faire is inefficient:

  — advances in data-processing technologies are ambiguous for welfare

  — limiting data use, especially by large firms, may be desirable (e.g., GDPR)

This Paper

Premise: data is information that helps firms predict economic fundamentals

3. Quantification of effects of advances in data-processing technologies:
   - Model validation and identification of key parameters using firm-level data
   - ΔTFP $\approx$ 5.3%–6.7% with channel decomposition (+1.7pp,+5.0pp,-1.5pp)
   - ΔWelf $\approx$ 0.1%–2.1% with 1.2pp eaten up by excessive information production

Motivating Evidence

Data and Empirics

Empirical data:

- Compustat Fundamentals Annual: firm-level outcomes
- I/B/E/S Guidance Data: managerial forecasts of firm-level outcomes
  - most comprehensive data set on managerial forecasts

Merged data: Compustat-I/B/E/S   12,423 obs. from 2,472 firms (2002-2022)

Supplementary data:

- The Duke CFO Survey: firm-level expectations of economy-wide outcomes

Summary statistics

Data and Empirics

Compustat-I/B/E/S details:

- SEC requirement to have calls with equity analyst
- Use one-year-ahead expectations ($Q1 - Q4$)...
- ... which are stated in conjunction with firms' annual report
- Use either mid-point of stated range or point estimates
- Revenue expectations provide the bulk of forecasts ($> 75\%$)

Revenue expectation errors:

$$\epsilon_{j,t+1|t} = rev_{j,t+1} - \mathbb{E}_{j,t}\left[rev_{j,t+1}\right]$$

- Use logs or scale revenues by the capital stock

Additional data comments

Fact 1: Accuracy of Expectations Have Improved

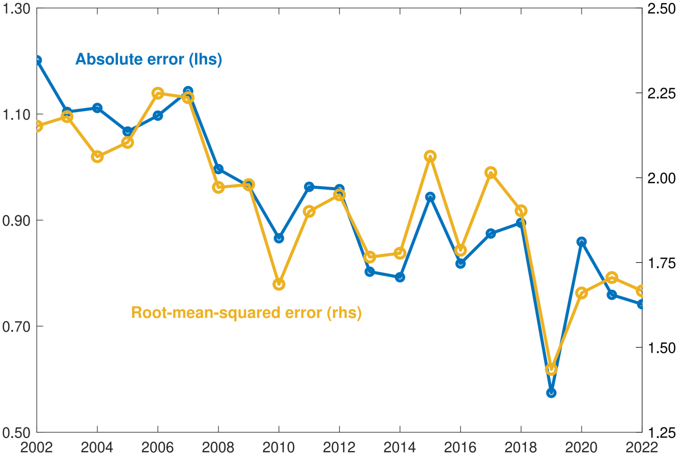


Figure: accuracy of one-year-ahead revenue expectations

Table
Extensions

Fact 2: Larger Firms Have More Accurate Expectations

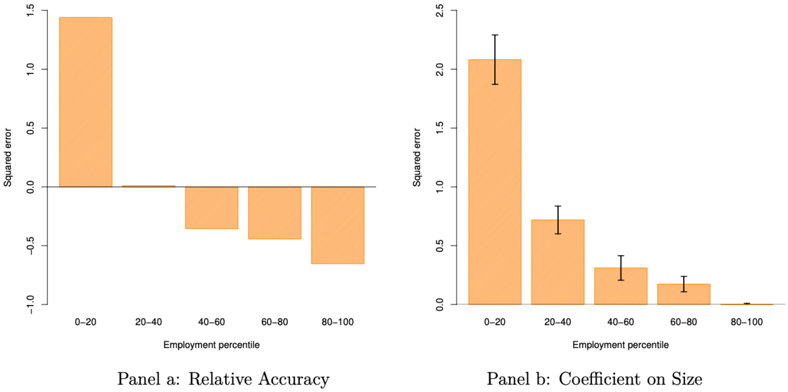


Figure: accuracy across the firm-size distribution

Panel b controls: size, age, rev. volatility, time and sector fixed effects

Table

Fact 3: Firms Have Become Larger

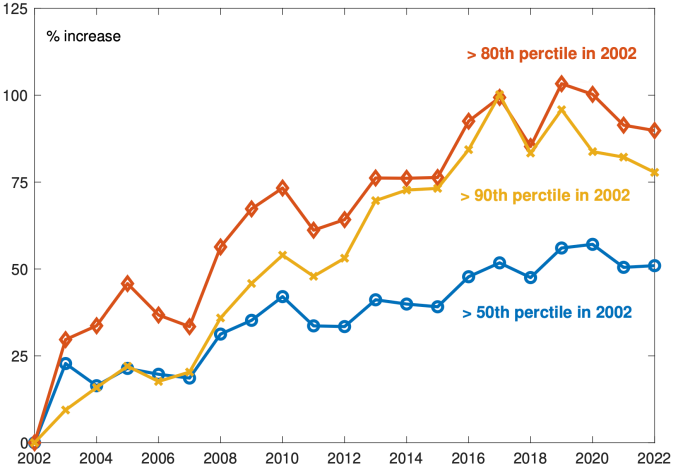


Figure: firm size by number of employees

Table

Quantitative Strength: Accuracy and Size

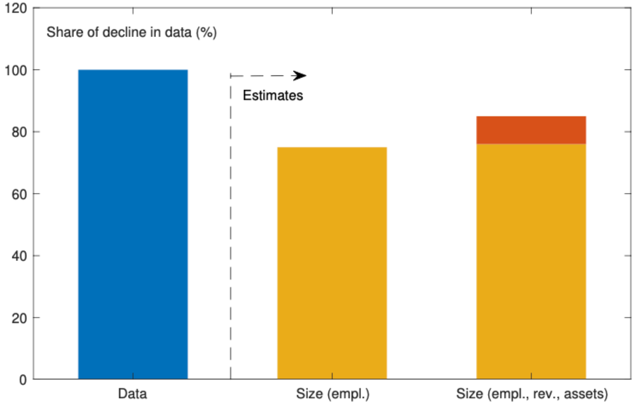


Figure: size-based simulation (data vs regression)

Estimates: take size-bin estimates and roll forward with time evolution of size bins

Size measure: both number of employees and assets—the former is more important

Taking Stock

Motivating facts:

1. Firms have become substantially more informed since early 2000's
2. Large firms are more informed than small ones
3. Improvements in information closely linked to firm-size dynamics

Robustness and inputs:

- Alternative regression specifications robust1
- Alternative measures of accuracy and size robust2
- Alternative measures of revenue: growth and levels robust3
- Alternative variables: firm profits and capex robust4
- Alternative variables: real GDP from the Duke-CFO Survey robust5
- Improved accuracy after large acquisition robust6
- Heterogeneity across sectors in accuracy and size evolution robust7
- Larger stock of acq. intangibles (e.g., software/patents) $\Longrightarrow$ more accurate input1

A Macroeconomic Model of Information Production

Household Preferences

- Unit mass of ex-ante identical households with preferences:

$$\mathcal{U}_i = C_i = \left[\int_0^1 (\delta_{ij} \cdot c_{ij})^{\frac{\theta-1}{\theta}} \cdot dj\right]^{\frac{\theta}{\theta-1}}$$

  where $\theta > 1$ and $\delta_{ij} \in \{\delta_H, \delta_L\}$ with $\delta_H > \delta_L > 0$

- Share $\gamma_j = \gamma(\omega_j | x_j)$ of δ_H-households depends on:
  - Demand state $\omega_j \in \{\text{red}, \text{blue}\}$ with $\mathbb{P}(\omega_j = \text{red}) = 1/2$
  - Variety type $x_j \in \{\text{red}, \text{blue}\}$ available in the market

- Variety j commands higher demand if customized to household tastes:

$$\gamma(\omega_j | x_j = \omega_j) = \bar{\gamma} > \underline{\gamma} = \gamma(\omega_j | x_j \neq \omega_j) \quad \forall j, \omega_j$$

- Each household is endowed with N units of labor

Household Preferences

- Unit mass of ex-ante identical households with preferences:

$$\mathcal{U}_i = C_i = \left[\int_0^1 (\delta_{ij} \cdot c_{ij})^{\frac{\theta-1}{\theta}} \cdot dj \right]^{\frac{\theta}{\theta-1}}$$

  where $\theta > 1$ and $\delta_{ij} \in \{\delta_H, \delta_L\}$ with $\delta_H > \delta_L > 0$

- Share $\gamma_j = \gamma(\omega_j | x_j)$ of δ_H-households depends on:
  - Demand state $\omega_j \in \{\text{red}, \text{blue}\}$ with $\mathbb{P}(\omega_j = \text{red}) = 1/2$
  - Variety type $x_j \in \{\text{red}, \text{blue}\}$ available in the market

- Variety j commands higher demand if customized to household tastes:

$$\gamma(\omega_j | x_j = \omega_j) = \bar{\gamma} > \underline{\gamma} = \gamma(\omega_j | x_j \neq \omega_j) \quad \forall j, \omega_j$$

- Each household is endowed with N units of labor

Household Preferences

- Unit mass of ex-ante identical households with preferences:

$$\mathcal{U}_i = C_i = \left[\int_0^1 (\delta_{ij} \cdot c_{ij})^{\frac{\theta-1}{\theta}} \cdot dj \right]^{\frac{\theta}{\theta-1}}$$

  where $\theta > 1$ and $\delta_{ij} \in \{\delta_H, \delta_L\}$ with $\delta_H > \delta_L > 0$

- Share $\gamma_j = \gamma(\omega_j|x_j)$ of δ_H-households depends on:
  - Demand state $\omega_j \in \{\text{red}, \text{blue}\}$ with $\mathbb{P}(\omega_j = \text{red}) = 1/2$
  - Variety type $x_j \in \{\text{red}, \text{blue}\}$ available in the market

- Variety j commands higher demand if customized to household tastes:

$$\gamma(\omega_j|x_j = \omega_j) = \bar{\gamma} > \underline{\gamma} = \gamma(\omega_j|x_j \neq \omega_j) \quad \forall j, \omega_j$$

- Each household is endowed with N units of labor

Technology for Goods Production

- Each variety is produced by a monopolistic firm, owned by households
- Firm j chooses type x_j and quantity y_j of its variety according to:

$$y_j = A_j \cdot n_j$$

  where n_j are units of labor employed and A_j is the firm's productivity

- Firms face productivity shocks:

$$a_j = \log(A_j) = \mu_j + \upsilon_j \quad \text{with} \quad \upsilon_j \sim \mathcal{N}\left(0, \tau_a^{-1}\right)$$

  where $\mu_j \sim \mathcal{N}\left(0, \tau_\mu^{-1}\right)$ is a source of ex-ante heterogeneity

Technology for Goods Production

- Each variety is produced by a monopolistic firm, owned by households
- Firm j chooses type x_j and quantity y_j of its variety according to:

$$y_j = A_j \cdot n_j$$

  where n_j are units of labor employed and A_j is the firm's productivity

- Firms face productivity shocks:

$$a_j = \log(A_j) = \mu_j + \upsilon_j \quad \text{with} \quad \upsilon_j \sim \mathcal{N}\left(0, \tau_a^{-1}\right)$$

  where $\mu_j \sim \mathcal{N}\left(0, \tau_\mu^{-1}\right)$ is a source of ex-ante heterogeneity

Technology for Goods Production

- Each variety is produced by a monopolistic firm, owned by households
- Firm j chooses type x_j and quantity y_j of its variety according to:

$$y_j = A_j \cdot n_j$$

where n_j are units of labor employed and A_j is the firm's productivity

- Firms face productivity shocks:

$$a_j = \log(A_j) = \mu_j + \upsilon_j \quad \text{with} \quad \upsilon_j \sim \mathcal{N}\left(0, \tau_a^{-1}\right)$$

where $\mu_j \sim \mathcal{N}\left(0, \tau_\mu^{-1}\right)$ is a source of ex-ante heterogeneity

Uncertainty and Information Production

- When choosing (x_j, n_j), firm j knows μ_j but not (υ_j, ω_j)
  — firm is uncertain about what and how much to produce

- To overcome uncertainty, firm j may obtain signals:

$$s_j^{\omega} \in \{\text{red}, \text{blue}\}, \quad \mathbb{P}(s_j^{\omega} = \omega_j | \omega_j) = \tau_j^{\omega} \in \left[\tfrac{1}{2}, 1\right]$$

  and

$$s_j^{\upsilon} = \upsilon_j + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}\left(0, (\tau_j^{\upsilon})^{-1}\right)$$

  where $\tau_j^{\upsilon} \in \{\underline{\tau}^{\upsilon}, \bar{\tau}^{\upsilon}\}$ and $\tau_j^{\omega} \in \{\underline{\tau}^{\omega}, \bar{\tau}^{\omega}\}$ are information precisions

- $(\underline{\tau}^{\upsilon}, \underline{\tau}^{\omega})$ are free, but firm j must allocate χ units of labor to increase them

Notation: $\quad \mathbf{s}_j \equiv [s_j^{\upsilon}, s_j^{\omega}] \qquad \boldsymbol{\tau}_j \equiv [\tau_j^{\upsilon}, \tau_j^{\omega}]$

Uncertainty and Information Production

- When choosing (x_j, n_j), firm j knows μ_j but not (υ_j, ω_j)
  — firm is uncertain about what and how much to produce

- To overcome uncertainty, firm j may obtain signals:

$$s_j^\omega \in \{\text{red}, \text{blue}\}, \quad \mathbb{P}(s_j^\omega = \omega_j | \omega_j) = \tau_j^\omega \in \left[\tfrac{1}{2}, 1\right]$$

  and

$$s_j^\upsilon = \upsilon_j + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}\left(0, (\tau_j^\upsilon)^{-1}\right)$$

  where $\tau_j^\upsilon \in \{\underline{\tau}^\upsilon, \bar{\tau}^\upsilon\}$ and $\tau_j^\omega \in \{\underline{\tau}^\omega, \bar{\tau}^\omega\}$ are information precisions

- $(\underline{\tau}^\upsilon, \underline{\tau}^\omega)$ are free, but firm j must allocate χ units of labor to increase them

Notation: $s_j \equiv [s_j^\upsilon, s_j^\omega]$ $\quad \tau_j \equiv [\tau_j^\upsilon, \tau_j^\omega]$

Uncertainty and Information Production

- When choosing (x_j, n_j), firm j knows μ_j but not (υ_j, ω_j)
  — firm is uncertain about what and how much to produce

- To overcome uncertainty, firm j may obtain signals:

$$s_j^{\omega} \in \{\text{red}, \text{blue}\}, \quad \mathbb{P}(s_j^{\omega} = \omega_j | \omega_j) = \tau_j^{\omega} \in \left[\tfrac{1}{2}, 1\right]$$

  and

$$s_j^{\upsilon} = \upsilon_j + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}\left(0, (\tau_j^{\upsilon})^{-1}\right)$$

  where $\tau_j^{\upsilon} \in \{\underline{\tau}^{\upsilon}, \bar{\tau}^{\upsilon}\}$ and $\tau_j^{\omega} \in \{\underline{\tau}^{\omega}, \bar{\tau}^{\omega}\}$ are information precisions

- $(\underline{\tau}^{\upsilon}, \underline{\tau}^{\omega})$ are free, but firm j must allocate χ units of labor to increase them

Notation: $s_j \equiv [s_j^{\upsilon}, s_j^{\omega}]$ $\quad \tau_j \equiv [\tau_j^{\upsilon}, \tau_j^{\omega}]$

Uncertainty and Information Production

- When choosing (x_j, n_j), firm j knows μ_j but not (υ_j, ω_j)
  — firm is uncertain about what and how much to produce
- To overcome uncertainty, firm j may obtain signals:

$$s_j^{\omega} \in \{\text{red}, \text{blue}\}, \quad \mathbb{P}(s_j^{\omega} = \omega_j | \omega_j) = \tau_j^{\omega} \in \left[\tfrac{1}{2}, 1\right]$$

  and

$$s_j^{\upsilon} = \upsilon_j + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}\left(0, (\tau_j^{\upsilon})^{-1}\right)$$

  where $\tau_j^{\upsilon} \in \{\underline{\tau}^{\upsilon}, \bar{\tau}^{\upsilon}\}$ and $\tau_j^{\omega} \in \{\underline{\tau}^{\omega}, \bar{\tau}^{\omega}\}$ are information precisions
- $(\underline{\tau}^{\upsilon}, \underline{\tau}^{\omega})$ are free, but firm j must allocate χ units of labor to increase them

Notation: $\quad \mathbf{s}_j \equiv \left[s_j^{\upsilon}, s_j^{\omega}\right] \quad \boldsymbol{\tau}_j \equiv \left[\tau_j^{\upsilon}, \tau_j^{\omega}\right]$

Timing and Markets

- Timing:

  Stage 0: nature draws $\left\{\mu_j, \upsilon_j, \omega_j, \{\delta_{ij}\}_i, \{\mathbf{s}_j\}_{\boldsymbol{\tau}_j}\right\}_j$

  Stage 1: firm j learns μ_j and makes information choice ι_j

  Stage 2: conditional on $(\mu_j, \mathbf{s}_j, \boldsymbol{\tau}_j)$, firms j decides (x_j, n_j)

  Stage 3: firm j learns its productivity υ_j and demand ω_j states;
  production takes place; markets "clear"; and households consume

- Markets:

  Labor market: competitive with wage w

  Variety-j market:

  - Baseline: market clears at a single price p_j
  - Rent-extracting economy: firms engage in discriminatory pricing
    — information helps firms learn how to price products
    — nests the baseline model as a special case

Timing and Markets

- Timing:

  Stage 0: nature draws $\left\{\mu_j, \upsilon_j, \omega_j, \{\delta_{ij}\}_i, \{\mathbf{s}_j\}_{\boldsymbol{\tau}_j}\right\}_j$

  Stage 1: firm j learns μ_j and makes information choice ι_j

  Stage 2: conditional on $(\mu_j, \mathbf{s}_j, \boldsymbol{\tau}_j)$, firms j decides (x_j, n_j)

  Stage 3: firm j learns its productivity υ_j and demand ω_j states;
  production takes place; markets "clear"; and households consume

- Markets:

  Labor market: competitive with wage w

  Variety-j market:

  - Baseline: market clears at a single price p_j
  - Rent-extracting economy: firms engage in discriminatory pricing
    — information helps firms learn how to price products
    — nests the baseline model as a special case

Timing and Markets

- Timing:

  `Stage 0:` nature draws $\left\{\mu_j, \upsilon_j, \omega_j, \{\delta_{ij}\}_i, \{\mathbf{s}_j\}_{\boldsymbol{\tau}_j}\right\}_j$

  `Stage 1:` firm j learns μ_j and makes information choice ι_j

  `Stage 2:` conditional on $(\mu_j, \mathbf{s}_j, \boldsymbol{\tau}_j)$, firms j decides (x_j, n_j)

  `Stage 3:` firm j learns its productivity υ_j and demand ω_j states;
  production takes place; markets "clear"; and households consume

- Markets:

  Labor market: competitive with wage w

  Variety-j market:

  - Baseline: market clears at a single price p_j
  - Rent-extracting economy: firms engage in discriminatory pricing
    — information helps firms learn how to price products
    — nests the baseline model as a special case

Information in the Cross Section

Result. *In equilibrium, firm j produces information if and only if* $\mu_j \geq \mu^\star$. *More informed firms, i.e.,* $\{j : \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\}$, *have on average higher and less dispersed tfp, less dispersed mrp, and they grow larger and more profitable.*

Intuition:

- Product-choice channel: tfp is higher and less volatile among informed firms
- Product-choice + Scale channels: mrp less dispersed among informed firms
- Both channels increase with μ_j and interact to boost firm size and profits
- Increasing returns to information production: $\boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}} \iff \mu_j \geq \mu^\star$

Caveat principalis:

- Measured tfp and misallocation may be misleading of social value creation
- Pricing channel: profit maximization $\neq$ social surplus maximization

Information in the Cross Section

Result. *In equilibrium, firm j produces information if and only if $\mu_j \geq \mu^\star$. More informed firms, i.e., $\{j : \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\}$, have on average higher and less dispersed tfp, less dispersed mrp, and they grow larger and more profitable.*

Intuition:

- Product-choice channel: tfp is higher and less volatile among informed firms
- Product-choice + Scale channels: mrp less dispersed among informed firms
- Both channels increase with μ_j and interact to boost firm size and profits
- Increasing returns to information production: $\boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}} \iff \mu_j \geq \mu^\star$

Caveat principalis:

- Measured tfp and misallocation may be misleading of social value creation
- Pricing channel: profit maximization $\neq$ social surplus maximization

Information and Rent Extraction

Optimal mechanism: at the optimum, firm j

- Extracts full surplus from δ_L-valuation consumers, but...
- Fails to capture full surplus from δ_H's
- Inefficiently reduces trade with δ_L's to maximize rent extraction from δ_H's

Information production:

- Improves product choices, and alters distribution of valuations (i.e., of γ_j)
- Thus, it also affects severity of distortions due to price discrimination

Information and Rent Extraction

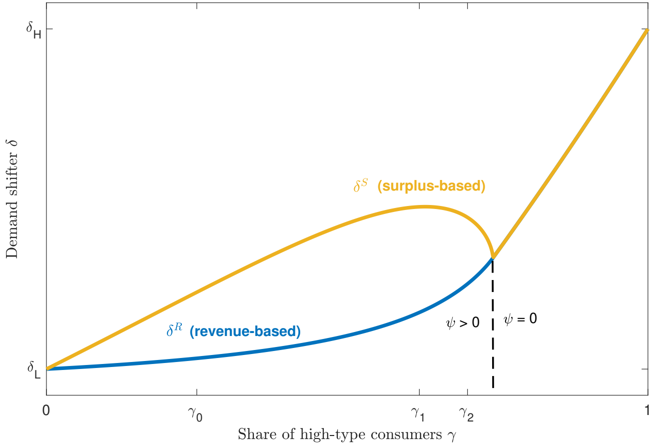
Optimal mechanism: at the optimum, firm j

- Extracts full surplus from δ_L-valuation consumers, but...
- Fails to capture full surplus from δ_H's
- Inefficiently reduces trade with δ_L's to maximize rent extraction from δ_H's

Information production:

- Improves product choices, and alters distribution of valuations (i.e., of γ_j)
- Thus, it also affects severity of distortions due to price discrimination

Information and Rent Extraction



Comment:

- Information production raises the likelihood of $\gamma_j = \bar{\gamma} > \underline{\gamma}$
- δ_j^R is the portion of social surplus captured that firm j captures $\neq \delta_j^S$

Information in General Equilibrium

Result. *In equilibrium, aggregate welfare satisfies:*

$$\mathcal{U}(\mu^\star) = \mathcal{A}(\mu^\star) \cdot [N - \mathcal{I}(\mu^\star)],$$

where $\mathcal{I}(\cdot)$ *is the aggregate information cost. Laissez faire implies:*

$$\mu^\star = \arg\max_{\mu} \ \mathcal{E}(\mu^\star)^{-1} \cdot \mathcal{U}(\mu^\star).$$

Comments:

- $\mathcal{A}(\mu^\star)$ is the aggregate TFP
- $\mathcal{E}(\mu^\star)$ is a macro-level externality:
  - Aggregate conflict between rent-extraction vs surplus creation (δ^S vs δ^R)
  - Central for understanding the benefits/costs of technological changes

  $\Longrightarrow$ Guides design of optimal corrective policy

Advancements in Data-Processing Technologies

We consider effects of a:

1) Fall in the information cost parameter χ

2) Rise in the information precisions $\bar{\boldsymbol{\tau}} = (\bar{\tau}_{\upsilon}, \bar{\tau}_{\omega})$

- Common narrative behind the rise of the data economy
  See, e.g., Brynjolfsson and McElheran (2016, 2024) and Baley and Veldkamp (2024)
- Substantial declines in computing costs and improvements in processing speeds
  See, e.g., Nordhaus (2008), Coyle and Hampton (2024), Gill et al. (2024)

Result. *An improvement in data-processing technologies, such as a fall in information cost parameter, χ, or a rise in information precisions, $\bar{\tau}$:*

- *Leads to an increase in the share of information producers*
- *But, has an ambiguous effect on aggregate TFP and welfare*

Advancements in Data-Processing Technologies

We consider effects of a:

1) Fall in the information cost parameter χ

2) Rise in the information precisions $\bar{\boldsymbol{\tau}} = (\bar{\tau}_\upsilon, \bar{\tau}_\omega)$

- Common narrative behind the rise of the data economy
  See, e.g., Brynjolfsson and McElheran (2016, 2024) and Baley and Veldkamp (2024)
- Substantial declines in computing costs and improvements in processing speeds
  See, e.g., Nordhaus (2008), Coyle and Hampton (2024), Gill et al. (2024)

Result. *An improvement in data-processing technologies, such as a fall in information cost parameter, χ, or a rise in information precisions, $\bar{\boldsymbol{\tau}}$:*

- *Leads to an increase in the share of information producers*
- *But, has an ambiguous effect on aggregate TFP and welfare*

Optimal Corrective Policy

Laissez-faire is inefficient:

- Restoring efficiency, however, requires implausible market interventions

Consider more constrained, data-regulation policies

1) Tax on information production: selects $\mu^{\star}$

2) "Garble" information: reduction in $\bar{\boldsymbol{\tau}} = (\bar{\tau}_{\upsilon}, \bar{\tau}_{\omega})$

— e.g., GDPR restricts access to and storage/sharing of consumer data by firms

Result.

1) *The planner imposes a positive (negative) tax if rent extraction is severe (mild)*

2) *If she can also garble information, then she does so only about demand and if rent extraction is socially destructive, in which case the information tax is not used*

⋆ More elaborate policies are also size dependent...

Optimal Corrective Policy

Laissez-faire is inefficient:

- Restoring efficiency, however, requires implausible market interventions

Consider more constrained, data-regulation policies

1) Tax on information production: selects $\mu^{\star}$

2) "Garble" information: reduction in $\bar{\boldsymbol{\tau}} = (\bar{\tau}_{\upsilon}, \bar{\tau}_{\omega})$

— e.g., GDPR restricts access to and storage/sharing of consumer data by firms

Result.

1) *The planner imposes a positive (negative) tax if rent extraction is severe (mild)*

2) *If she can also garble information, then she does so only about demand and if rent extraction is socially destructive, in which case the information tax is not used*

⋆ More elaborate policies are also size dependent...

Optimal Corrective Policy

Laissez-faire is inefficient:

- Restoring efficiency, however, requires implausible market interventions

Consider more constrained, data-regulation policies

1) Tax on information production: selects $\mu^\star$

2) "Garble" information: reduction in $\bar{\boldsymbol{\tau}} = (\bar{\tau}_\upsilon, \bar{\tau}_\omega)$

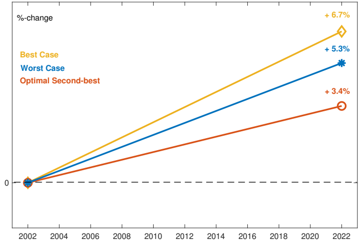
— e.g., GDPR restricts access to and storage/sharing of consumer data by firms

Result.

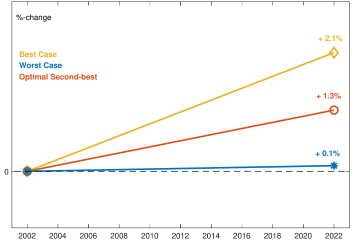
1) *The planner imposes a positive (negative) tax if rent extraction is <u>severe</u> (<u>mild</u>)*
2) *If she can also garble information, then she does so only about demand and if rent extraction is <u>socially destructive</u>, in which case the information tax is not used*

⋆ More elaborate policies are also size dependent...

Quantification



Panel a: TFP



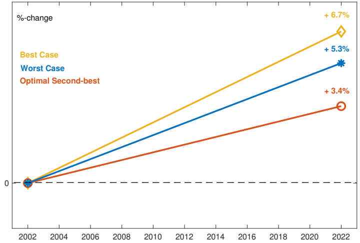
Panel b: Welfare

Strategy:

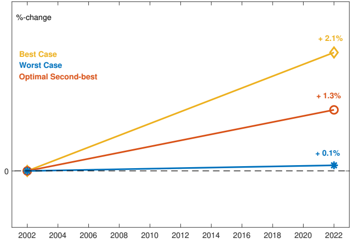
- Calibration using micro moments on expectations and productivity
- Data-processing improvements match increase in expectations accuracy
- Best- vs worst-case scenarios to due to indeterminacy inherent to firm-level data

Details

Quantification



Panel a: TFP

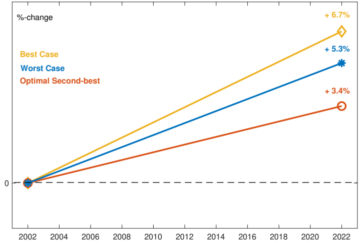


Panel b: Welfare

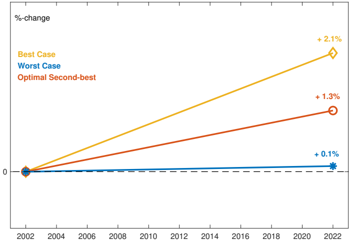
Results:

- ΔTFP $\approx$ 5.3%–6.7%, with channel decomposition (+1.7pp,+5.0pp,-1.5pp)
  - Scale channel contributes the lion-share of estimated TFP improvements
  - TFP increased by $\approx$ 15% over the sample period
- ΔWelf $\approx$ 0.1%–2.1%, with 1.2pp eaten up by excessive information production

Quantification



Panel a: TFP



Panel b: welfare

Data-regulation policy:

- Worst-case scenario in the severe rent-extraction region
- Optimal data-regulation lowers share of information producers by 5pp...

...and halves the rise in the share of employment by large firms

Quantitative Refinements

We consider several quantitative refinements of our baseline results:

Quantitative refinements:

1. Capital and variety accumulation: information in the long run? Ext1
2. Re-calibrated productivity parameters: information or productivity? Ext2
3. Retail sector estimates: where in the range are we likely to be? Ext3

Main takeaways remain robust across the three refinements

What Have We Learned?

1. Evidence: substantial increase in accuracy of firms' expectations since early 2000s
   - Large firms are better informed than small ones
   - Improvements in accuracy closely linked to firm-size dynamics

2. Theory: a macroeconomic model of information production
   - Information helps firms learn...

     — what to produce, how much to produce, and how to price products
   - Dual role of information:

     — tradeoff between efficiency vs rent extraction
   - Laissez faire inefficient:

     — advances in data-processing technologies are ambiguous for welfare

     — limiting data use by large firms may be desirable

3. Quantification: advances in data-processing technologies have led to...
   - ΔTFP $\approx$ 5.3%–6.7% with channel decomposition (+1.7pp,+5.0pp,-1.5pp)
   - ΔWelf $\approx$ 0.1%–2.1% with 1.2pp eaten up by excessive information production

Appendix

2. The Baseline Economy

Household and Firm Problems

- Third stage: household optimization implies that:

$$c_j \equiv \int_0^1 c_{ij} \cdot dj = \delta_j^{\theta-1} \cdot p_j^{-\theta} \cdot C$$

where the demand shifter of the as-if representative consumer is:

$$\delta_j \equiv \delta(\gamma_j) = \left[\gamma_j \cdot \delta_H^{\theta-1} + (1-\gamma_j) \cdot \delta_L^{\theta-1}\right]^{\frac{1}{\theta-1}}$$

- Second stage: conditional on (μ_j, s_j, τ_j), firm j chooses (x_j, n_j) to maximize:

$$\widehat{\pi}_j = \mathbb{E}\left[p_j \cdot y_j - w \cdot n_j | \mu_j, s_j, \tau_j\right],$$

subject to demand $y_j = \delta_j^{\theta-1} \cdot p_j^{-\theta} \cdot C$ and feasibility $y_j = A_j \cdot n_j$

- First stage: Conditional on μ_j, firm j chooses information production ι_j:

$$\iota_j = \begin{cases} 1 & \text{if } \mathbb{E}\left[\widehat{\pi}_j | \mu_j, \tau_j = \overline{\tau}\right] - w \cdot \chi \geq \mathbb{E}\left[\widehat{\pi}_j | \mu_j, \tau_j = \underline{\tau}\right] \\ 0 & \text{otherwise} \end{cases}$$

Household and Firm Problems

- Third stage: household optimization implies that:

$$c_j \equiv \int_0^1 c_{ij} \cdot dj = \delta_j^{\theta-1} \cdot p_j^{-\theta} \cdot C$$

where the demand shifter of the as-if representative consumer is:

$$\delta_j \equiv \delta(\gamma_j) = \left[\gamma_j \cdot \delta_H^{\theta-1} + (1-\gamma_j) \cdot \delta_L^{\theta-1}\right]^{\frac{1}{\theta-1}}$$

- Second stage: conditional on $(\mu_j, \mathbf{s}_j, \boldsymbol{\tau}_j)$, firm j chooses (x_j, n_j) to maximize:

$$\widehat{\pi}_j = \mathbb{E}\left[p_j \cdot y_j - w \cdot n_j | \mu_j, \mathbf{s}_j, \boldsymbol{\tau}_j\right],$$

subject to demand $y_j = \delta_j^{\theta-1} \cdot p_j^{-\theta} \cdot C$ and feasibility $y_j = A_j \cdot n_j$

- First stage: Conditional on μ_j, firm j chooses information production ι_j:

$$\iota_j = \begin{cases} 1 & \text{if } \mathbb{E}\left[\widehat{\pi}_j | \mu_j, \tau_j = \overline{\tau}\right] - w \cdot \chi \geq \mathbb{E}\left[\widehat{\pi}_j | \mu_j, \tau_j = \underline{\tau}\right] \\ 0 & \text{otherwise} \end{cases}$$

Household and Firm Problems

- Third stage: household optimization implies that:

$$c_j \equiv \int_0^1 c_{ij} \cdot dj = \delta_j^{\theta-1} \cdot p_j^{-\theta} \cdot C$$

  where the demand shifter of the as-if representative consumer is:

$$\delta_j \equiv \delta(\gamma_j) = \left[\gamma_j \cdot \delta_H^{\theta-1} + (1-\gamma_j) \cdot \delta_L^{\theta-1}\right]^{\frac{1}{\theta-1}}$$

- Second stage: conditional on $(\mu_j, \mathbf{s}_j, \boldsymbol{\tau}_j)$, firm j chooses (x_j, n_j) to maximize:

$$\widehat{\pi}_j = \mathbb{E}\left[p_j \cdot y_j - w \cdot n_j | \mu_j, \mathbf{s}_j, \boldsymbol{\tau}_j\right],$$

  subject to demand $y_j = \delta_j^{\theta-1} \cdot p_j^{-\theta} \cdot C$ and feasibility $y_j = A_j \cdot n_j$

- First stage: Conditional on μ_j, firm j chooses information production ι_j:

$$\iota_j = \begin{cases} 1 & \text{if } \mathbb{E}\left[\widehat{\pi}_j | \mu_j, \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\right] - w \cdot \chi \geq \mathbb{E}\left[\widehat{\pi}_j | \mu_j, \boldsymbol{\tau}_j = \underline{\boldsymbol{\tau}}\right] \\ 0 & \text{otherwise} \end{cases}$$

Optimal Input and Information Choices

Proposition

In any equilibrium:

(i) *Firm j with type μ_j, observing signals $\boldsymbol{s}_j$ with precisions $\boldsymbol{\tau}_j$, chooses:*

$$x_j = s_j^{\omega} \quad \text{and} \quad n_j = \mathbb{E}\left[(\delta_j \cdot A_j)^{\frac{\theta-1}{\theta}} \,|\mu_j, \boldsymbol{s}_j, \boldsymbol{\tau}_j \right]^{\theta} \cdot \Omega$$

(ii) *Firm j with type μ_j chooses:*

$$\iota_j = \begin{cases} 1 & \text{if} \quad \frac{1}{\theta-1} \cdot \left(\mathbb{E}\left[n_j | \mu_j, \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\right] - \mathbb{E}\left[n_j | \mu_j, \boldsymbol{\tau}_j = \underline{\boldsymbol{\tau}}\right] \right) \geq \chi \\ 0 & \text{otherwise} \end{cases}$$

where:

$$\mathbb{E}\left[n_j | \mu_j, \boldsymbol{\tau}_j\right] = \exp^{(\theta-1)\cdot\mu_j} \cdot g(\boldsymbol{\tau}_j)^{\theta-1} \cdot \Omega$$

⋆ $\Omega \equiv C \cdot \left(\frac{\theta}{\theta-1} \cdot w \right)^{-\theta}$ is the market size faced by all firms

⋆ $g(\boldsymbol{\tau_j})$ is the information shifter, which increases in $\tau_{\upsilon,j}$ and $\tau_{\omega,j}$ info-shifter

Information in the Cross Section

Corollary

In equilibrium, firm j produces information if and only if:

$$\mu_j \geq \bar{\mu} \equiv \frac{1}{\theta - 1} \cdot \log \left[\frac{(\theta - 1) \cdot \chi}{\left(g(\bar{\boldsymbol{\tau}})^{\theta-1} - g(\underline{\boldsymbol{\tau}})^{\theta-1} \right) \cdot \Omega} \right]$$

More informed firms, i.e., $\{j : \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\}$, *have on average higher and less dispersed tfp, less dispersed mrp, and they grow larger and more profitable as measured by:*

$$\mathbb{E}[X_j | \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}] > \mathbb{E}[X_j | \boldsymbol{\tau}_j = \underline{\boldsymbol{\tau}}]$$

for $X_j \in \{n_j, n_j + \chi \cdot \iota_j, p_j \cdot y_j, p_j \cdot y_j - w \cdot n_j, \pi_j\}$

Remarks: given (x_j, n_j, ι_j), surplus created by firm j isomorphic to $(\delta_j \cdot A_j)^{\frac{\theta-1}{\theta}}$:

$$\text{tfp}_j \equiv \frac{\theta - 1}{\theta} \cdot \log(\delta_j \cdot A_j)$$

Product-choice channel $\Longrightarrow$ tfp is higher and less volatile among more informed firms

See, also, Farboodi and Veldkamp (2024)

Information in the Cross Section

Corollary

In equilibrium, firm j produces information if and only if:

$$\mu_j \geq \bar{\mu} \equiv \frac{1}{\theta - 1} \cdot \log \left[\frac{(\theta - 1) \cdot \chi}{\left(g(\bar{\boldsymbol{\tau}})^{\theta - 1} - g(\underline{\boldsymbol{\tau}})^{\theta - 1} \right) \cdot \Omega} \right]$$

More informed firms, i.e., $\{j : \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\}$*, have on average higher and less dispersed tfp, less dispersed mrp, and they grow larger and more profitable as measured by:*

$$\mathbb{E}[X_j | \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}] > \mathbb{E}[X_j | \boldsymbol{\tau}_j = \underline{\boldsymbol{\tau}}]$$

for $X_j \in \{n_j, n_j + \chi \cdot \iota_j, p_j \cdot y_j, p_j \cdot y_j - w \cdot n_j, \pi_j\}$

Remarks: given (x_j, n_j, ι_j), surplus created by firm j isomorphic to $(\delta_j \cdot A_j)^{\frac{\theta - 1}{\theta}}$:

$$\text{tfp}_j \equiv \frac{\theta - 1}{\theta} \cdot \log(\delta_j \cdot A_j)$$

Product-choice channel $\Longrightarrow$ tfp is higher and less volatile among more informed firms

See, also, Farboodi and Veldkamp (2024)

Information in the Cross Section

Corollary

In equilibrium, firm j produces information if and only if:

$$\mu_j \geq \bar{\mu} \equiv \frac{1}{\theta - 1} \cdot \log \left[\frac{(\theta - 1) \cdot \chi}{\left(g(\bar{\boldsymbol{\tau}})^{\theta-1} - g(\underline{\boldsymbol{\tau}})^{\theta-1} \right) \cdot \Omega} \right]$$

More informed firms, i.e., $\{j : \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\}$, have on average higher and less dispersed tfp, less dispersed mrp, and they grow larger and more profitable as measured by:

$$\mathbb{E}[X_j | \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}] > \mathbb{E}[X_j | \boldsymbol{\tau}_j = \underline{\boldsymbol{\tau}}]$$

for $X_j \in \{n_j, n_j + \chi \cdot \iota_j, p_j \cdot y_j, p_j \cdot y_j - w \cdot n_j, \pi_j\}$

Remarks: absent information frictions, we would have $mrp_j = mrp_j'$ for all j, j':

$$\text{mrp}_j \equiv \log(p_j \cdot y_j) - \log(n_j)$$

Product-choice and Scale channels $\Longrightarrow$ $\mathbb{VAR}[\text{mrp}_j]$ lower among more informed firms

See, also, David et al. (2016) and David and Venkateswaran (2019)

Information in the Cross Section

Corollary

In equilibrium, firm j produces information if and only if:

$$\mu_j \geq \bar{\mu} \equiv \frac{1}{\theta - 1} \cdot \log \left[\frac{(\theta - 1) \cdot \chi}{\left(g(\bar{\boldsymbol{\tau}})^{\theta - 1} - g(\underline{\boldsymbol{\tau}})^{\theta - 1} \right) \cdot \Omega} \right]$$

More informed firms, i.e., $\{j : \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}\}$*, have on average higher and less dispersed tfp, less dispersed mrp, and they grow larger and more profitable as measured by:*

$$\mathbb{E}[X_j | \boldsymbol{\tau}_j = \bar{\boldsymbol{\tau}}] > \mathbb{E}[X_j | \boldsymbol{\tau}_j = \underline{\boldsymbol{\tau}}]$$

for $X_j \in \{n_j, n_j + \chi \cdot \iota_j, p_j \cdot y_j, p_j \cdot y_j - w \cdot n_j, \pi_j\}$

Remarks: given the properties of the information shifter $g(\cdot)$:

$$\frac{\partial \mathbb{E}[n_j | \mu_j, \boldsymbol{\tau}_j]}{\partial \tau_{\upsilon,j}}, \frac{\partial \mathbb{E}[n_j | \mu_j, \boldsymbol{\tau}_j]}{\partial \tau_{\omega,j}}, \frac{\partial^2 \mathbb{E}[n_j | \mu_j, \boldsymbol{\tau}_j]}{\partial \tau_{\upsilon,j} \partial \tau_{\omega,j}} > 0$$

Product-choice and Scale channels reinforce each other's effects on firm size/profits

Define aggregate total factor productivity (TFP) as:

$$\mathcal{A} \equiv C \cdot \mathcal{N}^{-1}$$

where $\mathcal{N} = \int_0^1 n_j \cdot dj$ denotes aggregate employment in goods production

Lemma

Given the marginal-type firm, $\bar{\mu}$, aggregate TFP is given by:

$$\mathcal{A}(\bar{\mu}; g) \equiv \exp^{\frac{\theta-1}{2}\cdot\frac{1}{\tau_\mu}} \cdot \left[g(\underline{\boldsymbol{\tau}})^{\theta-1} \cdot (1-\xi(\bar{\mu})) + g(\bar{\boldsymbol{\tau}})^{\theta-1} \cdot \xi(\bar{\mu})\right]^{\frac{1}{\theta-1}}$$

where $\xi(\bar{\mu}) \in (0,1)$ decreases with $\bar{\mu}$

Information in General Equilibrium

- Optimality implies that, for a given market size, the marginal-type firm is:

$$\bar{\mu}(\Omega) = \frac{1}{\theta - 1} \cdot \log \left[\frac{(\theta - 1) \cdot \chi}{(g(\bar{\tau})^{\theta-1} - g(\underline{\tau})^{\theta-1}) \cdot \Omega} \right]$$

- Market clearing implies that, for a given marginal-type firm, the market size is:

$$\Omega(\bar{\mu}) = \frac{\mathcal{A}(\bar{\mu}, g) \cdot \left[N - \chi \cdot \Phi\left(-\bar{\mu} \cdot \sqrt{\tau_\mu}\right) \right]}{\mathcal{A}(\bar{\mu}, g)^{\theta}}$$

Equilibrium Determination in a Figure

Proposition

An equilibrium exists, is unique, and in it the marginal-type firm solves:

$$\bar{\mu}(\Omega(\mu^\star)) = \mu^\star$$

The aggregate TFP and welfare are in turn given by:

$$\mathcal{A}^\star = \mathcal{A}(\mu^\star, g) \text{ and } C^\star = \mathcal{A}^\star \cdot \left[N - \chi \cdot \Phi\left(-\mu^\star \cdot \sqrt{\tau_\mu}\right) \right]$$

Information in General Equilibrium

- Optimality implies that, for a given market size, the marginal-type firm is:

$$\bar{\mu}(\Omega) = \frac{1}{\theta - 1} \cdot \log \left[\frac{(\theta - 1) \cdot \chi}{(g(\bar{\tau})^{\theta-1} - g(\underline{\tau})^{\theta-1}) \cdot \Omega} \right]$$

- Market clearing implies that, for a given marginal-type firm, the market size is:

$$\Omega(\bar{\mu}) = \frac{\mathcal{A}(\bar{\mu}, g) \cdot \left[N - \chi \cdot \Phi \left(-\bar{\mu} \cdot \sqrt{\tau_\mu} \right) \right]}{\mathcal{A}(\bar{\mu}, g)^\theta}$$

Equilibrium Determination in a Figure

Proposition

An equilibrium exists, is unique, and in it the marginal-type firm solves:

$$\bar{\mu}(\Omega(\mu^\star)) = \mu^\star$$

The aggregate TFP and welfare are in turn given by:

$$\mathcal{A}^\star = \mathcal{A}(\mu^\star, g) \text{ and } C^\star = \mathcal{A}^\star \cdot \left[N - \chi \cdot \Phi \left(-\mu^\star \cdot \sqrt{\tau_\mu} \right) \right]$$

Advancements in Data-Processing Technologies

We consider effects of a:

1) Fall in the information cost parameter χ

2) Rise in the information precisions $\bar{\boldsymbol{\tau}} = (\bar{\tau}_\upsilon, \bar{\tau}_\omega)$

- Common narrative behind the rise of the data economy
  See, e.g., Brynjolfsson and McElheran (2016, 2024) and Baley and Veldkamp (2024)
- Substantial declines in computing costs and improvements in processing speeds
  See, e.g., Nordhaus (2008), Coyle and Hampton (2024), Gill et al. (2024)

Comparative Statics in a Figure

Proposition

An improvement in data-processing technologies, such as a fall in χ, a rise in $\bar{\tau}_\upsilon$, or a rise in $\bar{\tau}_\omega$ leads to:

- *An increase in the share of information producers, $\Phi\left(-\mu^* \cdot \sqrt{\tau_\mu}\right)$*
- *An increase in the aggregate TFP, A^*, and welfare, C^**

Advancements in Data-Processing Technologies

We consider effects of a:

1) Fall in the information cost parameter χ

2) Rise in the information precisions $\bar{\boldsymbol{\tau}} = (\bar{\tau}_\upsilon, \bar{\tau}_\omega)$

- Common narrative behind the rise of the data economy
  See, e.g., Brynjolfsson and McElheran (2016, 2024) and Baley and Veldkamp (2024)
- Substantial declines in computing costs and improvements in processing speeds
  See, e.g., Nordhaus (2008), Coyle and Hampton (2024), Gill et al. (2024)

Comparative Statics in a Figure

Proposition

An improvement in data-processing technologies, such as a fall in χ, a rise in $\bar{\tau}_\upsilon$, or a rise in $\bar{\tau}_\omega$ leads to:

- *An increase in the share of information producers, $\Phi\left(-\mu^\star \cdot \sqrt{\tau_\mu}\right)$*
- *An increase in the aggregate TFP, $\mathcal{A}^\star$, and welfare, $C^\star$*

Taking Stock

Information production boosts:

- Allocative efficiency at firm level
- TFP and welfare at aggregate level

A bird's eye view:

$$\mathcal{U}(I) = \mathcal{A}(I) \cdot [N - I]$$

where I are the information costs. Laissez-faire allocations select:

$$I^{\star} = \arg\max_{I} \; \mathcal{U}(I)$$

Unsurprisingly, advances in data-processing technologies are amazing

Too benign?

- Concerns that consumer data is being used for discriminatory purposes (e.g., European Commission Report, 2020)
- Ongoing debates on data-regulation policies (e.g., GDPR in the EU)

Taking Stock

Information production boosts:

- Allocative efficiency at firm level
- TFP and welfare at aggregate level

A bird's eye view:

$$\mathcal{U}(I) = \mathcal{A}(I) \cdot [N - I]$$

where I are the information costs. Laissez-faire allocations select:

$$I^{\star} = \arg\max_{I} \, \mathcal{U}(I)$$

Unsurprisingly, advances in data-processing technologies are amazing

Too benign?

- Concerns that consumer data is being used for discriminatory purposes (e.g., European Commission Report, 2020)
- Ongoing debates on data-regulation policies (e.g., GDPR in the EU)

Taking Stock

Information production boosts:

- Allocative efficiency at firm level
- TFP and welfare at aggregate level

A bird's eye view:

$$\mathcal{U}(I) = \mathcal{A}(I) \cdot [N - I]$$

where I are the information costs. Laissez-faire allocations select:

$$I^\star = \arg\max_I \mathcal{U}(I)$$

Unsurprisingly, advances in data-processing technologies are amazing

Too benign?

- Concerns that consumer data is being used for discriminatory purposes (e.g., European Commission Report, 2020)
- Ongoing debates on data-regulation policies (e.g., GDPR in the EU)

3. The Rent-Extracting Economy

Information and Price Discrimination

Stage 3: design optimal mechanism

- Consists of a menu $\mathcal{M}_j = \{(t_j^H, q_j^H), (t_j^L, q_j^L)\}$
- t_j^l is the payment from type-$l \in \{H, L\}$ consumer to firm j
- q_j^l is the allocation of units of the variety to type-l consumer

Firm j maximizes profits subject to (PC), (IC), and (FC), given $(x_j, n_j, \upsilon_j, \omega_j)$

Standard solution: at the optimum, firm j:

- Extracts full surplus from δ_L-valuation consumers, but...
- Fails to capture full surplus from δ_H's
- Inefficiently reduces trade with δ_L's to maximize rent extraction from δ_H's

Information and Price Discrimination

Stage 3: design optimal mechanism

- Consists of a menu $\mathcal{M}_j = \{(t_j^H, q_j^H), (t_j^L, q_j^L)\}$
- t_j^l is the payment from type-$l \in \{H, L\}$ consumer to firm j
- q_j^l is the allocation of units of the variety to type-l consumer

Firm j maximizes profits subject to (PC), (IC), and (FC), given $(x_j, n_j, \upsilon_j, \omega_j)$

Standard solution: at the optimum, firm j:

- Extracts full surplus from δ_L-valuation consumers, but...
- Fails to capture full surplus from δ_H's
- Inefficiently reduces trade with δ_L's to maximize rent extraction from δ_H's

Rent Extraction vs Social Surplus Maximization

Surplus-based demand shifter: social surplus created by firm j:

$$\delta_j^S \equiv \delta^S(\gamma_j) = \left[\gamma_j \cdot \left(\delta_H \cdot \alpha_j^H\right)^{\frac{\theta-1}{\theta}} + (1-\gamma_j) \cdot \left(\delta_L \cdot \alpha_j^L\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}$$

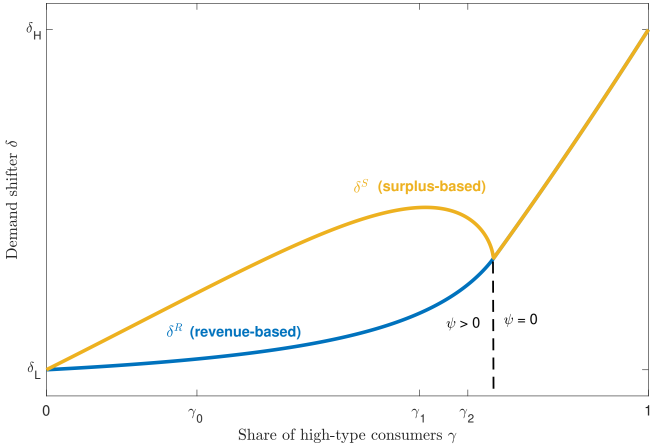
Revenue-based demand shifter: part that is captured by the firm:

$$\delta_j^R \equiv \delta^R(\gamma_j) = \left[\gamma_j \cdot \left(\delta_H \cdot \alpha_j^H\right)^{\frac{\theta-1}{\theta}} + (1-\gamma_j) \cdot \left(\psi_j \cdot \delta_L \cdot \alpha_j^L\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}$$

Comments:

- α_j^l is the share of variety j allocated to type-l consumer
- $\psi_j \leq 1$ reflects micro-level distortion: $\text{MU}_L/\text{MU}_H = \psi_j^{-\frac{\theta-1}{\theta}}$ Details

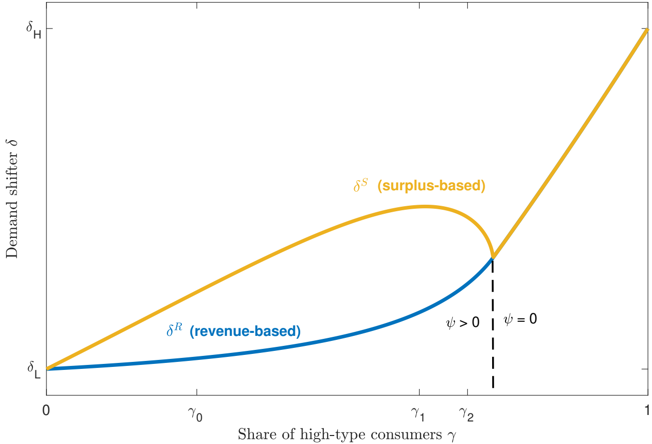
Rent Extraction vs Social Surplus Maximization



Note: information production raises the likelihood of $\gamma_j = \bar{\gamma} > \underline{\gamma}$

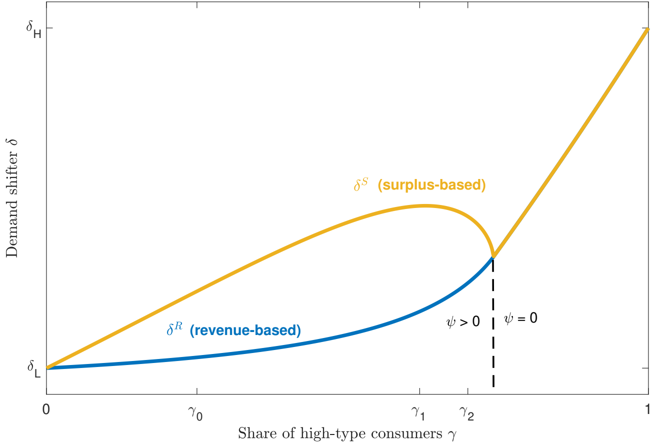
δ_j^R increases in γ_j, whereas δ_j^S/δ_j^R first increases and then falls in γ_j

Rent Extraction vs Social Surplus Maximization



Def'n: rent extraction is severe if $\delta^S(\bar{\gamma})/\delta^R(\bar{\gamma}) < \delta^S(\underline{\gamma})/\delta^R(\underline{\gamma})$; otherwise, it is mild.

Rent Extraction vs Social Surplus Maximization



Def'n: rent extraction is socially destructive if $\delta^S(\bar{\gamma}) < \delta^S(\underline{\gamma})$

Equilibrium of the Rent-Extracting Economy

1) Firm behavior same as in baseline economy:

Proposition

An equilibrium of the rent-extracting economy exists, is unique, and in it:

(i) *Firm j's choices of x_j, n_j and ι_j are the same as in the baseline economy, except that the demand shifter δ_j is replaced by δ_j^R*

(ii) *Marginal-type, $\mu^\star$, is determined as in the baseline economy, except in the expression for the information shifter $g(\cdot)$ the demand shifter δ_j is replaced with δ_j^R*

⋆ Same implications for firm-level measures of tfp, mrp, firm size and profits...

...but these now contain misleading information about social-value creation

⋆ Baseline $\cong$ rent-extracting economy with $0 = \underline{\gamma} < \bar{\gamma} = 1$ and adjusted (δ_H, δ_L)

Equilibrium of Rent-Extracting Economy

2) Aggregate behavior very different:

Proposition

The aggregate TFP and welfare of the rent-extracting economy are given by:

$$\mathcal{A}^\star = \mathcal{E}\left(\mu^\star\right) \cdot \mathcal{A}(\mu^\star, g^R) \quad \text{and} \quad C^\star = \mathcal{A}^\star \cdot \left[N - \Phi\left(-\mu^\star \cdot \sqrt{\tau_\mu}\right) \cdot \chi\right],$$

where $g^R(\cdot)$ is the same as the information shifter $g(\cdot)$, except replace δ_j with δ_j^R

- ⋆ Externality: firms make choices that maximize $\mathcal{E}^{-1} \cdot C$ instead of C
- ⋆ Arises as firms are unable to capture full surplus created by their choices
- ⋆ Can be strong so that advancements in information technologies reduce welfare
  - E.g., rent extraction socially destructive + scale channel unimportant [Illustration]

Laissez-faire equilibrium is inefficient

- Except when information is symmetric, i.e., $0 = \underline{\gamma} < \bar{\gamma} = 1$ ($\cong$ Baseline)
- However, restoring efficiency requires implausible market interventions

Consider more constrained interventions: data-regulation policies

1) Tax on information production: selects μ^*

2) "Garble" information: reduction in $\bar{\tau} = (\bar{\tau}_v, \bar{\tau}_\omega)$

— e.g., GDPR restricts access to and storage/sharing of consumer data by firms

Proposition

1) The planner imposes a positive (negative) tax if rent extraction is severe (mild)

2) If she can also garble information, then she does so only about demand and if rent extraction is socially destructive, in which case the information tax is not used

⋆ More elaborate policies are also size dependent...

Optimal Corrective Regulation

Laissez-faire equilibrium is inefficient

- Except when information is symmetric, i.e., $0 = \underline{\gamma} < \bar{\gamma} = 1$ ($\cong$ Baseline)
- However, restoring efficiency requires implausible market interventions

Consider more constrained interventions: data-regulation policies

1) Tax on information production: selects $\mu^\star$

2) "Garble" information: reduction in $\bar{\boldsymbol{\tau}} = (\bar{\tau}_\upsilon, \bar{\tau}_\omega)$

— e.g., GDPR restricts access to and storage/sharing of consumer data by firms

Proposition

1) The planner imposes a positive (negative) tax if rent extraction is severe (mild)

2) If she can also garble information, then she does so only about demand and if rent extraction is socially destructive, in which case the information tax is not used

⋆ More elaborate policies are also size dependent...

Laissez-faire equilibrium is inefficient

- Except when information is symmetric, i.e., $0 = \underline{\gamma} < \bar{\gamma} = 1$ ($\cong$ Baseline)
- However, restoring efficiency requires implausible market interventions

Consider more constrained interventions: data-regulation policies

1) Tax on information production: selects $\mu^\star$

2) "Garble" information: reduction in $\bar{\boldsymbol{\tau}} = (\bar{\tau}_\upsilon, \bar{\tau}_\omega)$

— e.g., GDPR restricts access to and storage/sharing of consumer data by firms

Proposition

1) The planner imposes a positive (negative) tax if rent extraction is severe (mild)

2) If she can also garble information, then she does so only about demand and if rent extraction is socially destructive, in which case the information tax is not used

⋆ More elaborate policies are also size dependent...

Optimal Corrective Regulation

Laissez-faire equilibrium is inefficient

- Except when information is symmetric, i.e., $0 = \underline{\gamma} < \bar{\gamma} = 1$ ($\cong$ Baseline)
- However, restoring efficiency requires implausible market interventions

Consider more constrained interventions: data-regulation policies

1) Tax on information production: selects $\mu^\star$

2) "Garble" information: reduction in $\bar{\boldsymbol{\tau}} = (\bar{\tau}_\upsilon, \bar{\tau}_\omega)$

— e.g., GDPR restricts access to and storage/sharing of consumer data by firms

Proposition

1) The planner imposes a positive (negative) tax if rent extraction is severe (mild)

2) If she can also garble information, then she does so only about demand and if rent extraction is socially destructive, in which case the information tax is not used

⋆ More elaborate policies are also size dependent...

4. Model Validation and Quantification

Qualitative Validation

We use our firm-level data to validate key cross-sectional predictions:

1. More informed firms have higher and less dispersed tfp
2. More informed firms have less dispersed mrp
3. More informed firms grow larger and more profitable

model validation data

Parameterization

1. Externally calibrated parameters:
   - Elasticity of substitution $\theta = 3$ (Hsieh and Klenow, 2009)
   - Normalization: $N = 1$, $\delta_H = \exp^{+\widehat{\delta}}$, $\delta_L = \exp^{-\widehat{\delta}}$
2. Internally calibrated parameters:

$$\left(\tau_\mu, \tau_a, \underline{\boldsymbol{\tau}}, \overline{\boldsymbol{\tau}}, \widehat{\delta}, \underline{\gamma}, \bar{\gamma}\right)$$

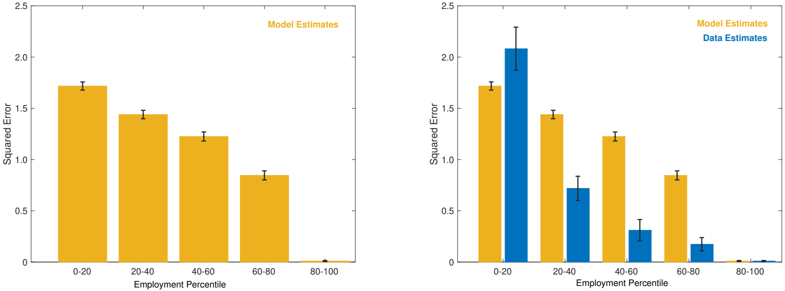
   to match productivity and forecast-error moments between 2002-2007:
   - Dispersion of fixed-effects of firm-level productivity
   - Mean of log-productivity of (un)informed firms
   - Unconditional variance of log-productivity
   - Conditional variance of log-productivity of (un)informed firms
   - MSE of (un)informed firms log-revenue forecasts

Informed vs uninformed:

- Assume 10 pct. of firms are informed in 2002-2007 $\Longrightarrow \chi$
  Brynjolfsson and McElheran, 2016
- Decrease χ and raise $\bar{\boldsymbol{\tau}}$ to hit (i) 41pct increase in forecast accuracy; and (ii) difference in forecast accuracy and volatility of informed vs uninformed Back

Quantitative Validation

We reproduce the close quantitative link between information and firm size:



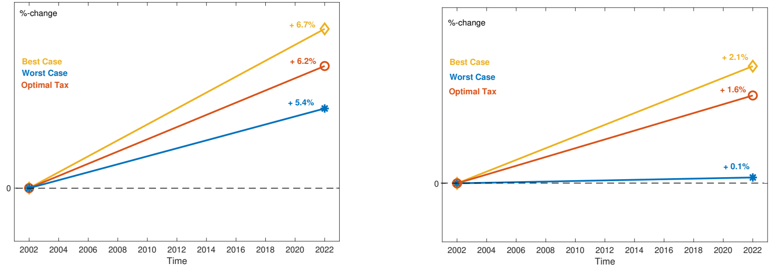
Panel a: model estimates   Panel b: data estimates

Figure: accuracy across the firm-size distribution

Additional cross-sectional moments:

- Share of information producers rises by 64pp (62pp) in model (data) Table
- Share of employment of large firms rises by 7.3pp (6pp) in model (data)

Model Quantification



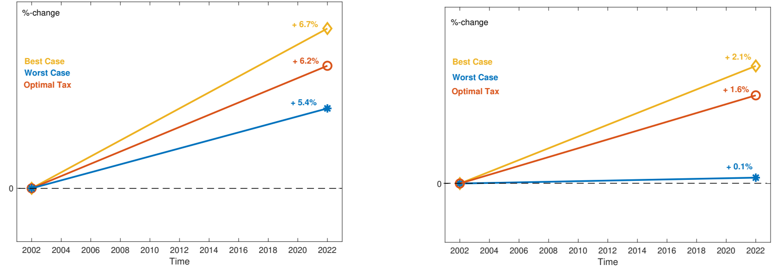
Panel a: TFP Panel b: Welfare

Figure: effects of advances in data-processing technologies

Decomposition:

- ΔTFP $\approx$ 5.4%–6.7%, with channel decomposition (+1.7pp,+5.0pp,-1.5pp)
  - Scale channel contributes the lion-share of estimated TFP improvements
  - TFP increased by $\approx$ 15% over the sample period
- ΔW $\approx$ 0.1%–2.1%, with 1.5pp eaten up by excessive information production

Model Quantification



Panel a: TFP   Panel b: welfare

Figure: effects of improvements in data-processing technologies

Data-regulation policy:

- Worst-case scenario in the severe rent-extraction region
- Optimal data-regulation lowers share of information producers by 5pp
- Halves the rise in the share of employment by large firms

Quantitative Refinements

We consider several quantitative refinements of our baseline results:

Quantitative refinements:

1. Capital and variety accumulation: information in the long run? Ext1
2. Re-calibrated productivity parameters: information or productivity? Ext2
3. Retail sector estimates: where in the range are we likely to be? Ext3

Main takeaways remain robust across the three refinements

What Have We Learned?

1. Evidence: substantial increase in accuracy of firms' expectations since early 2000's
   - Large firms are better informed than small ones
   - Improvements in accuracy closely linked to firm-size dynamics

2. Theory: a macroeconomic model of information production
   - Information helps firms learn...
     — what to produce, how much to produce, and how to price products
   - Dual role of information:
     — tradeoff between efficiency vs rent extraction
   - Laissez faire inefficient:
     — advances in data-processing technologies are ambiguous for welfare
     — limiting data use by large firms may be desirable

3. Quantification: advances in data-processing technologies have led to...
   - ΔTFP $\approx$ 5.4%–6.7% with channel decomposition (+1.7pp,+5.0pp,-1.5pp)
   - ΔW $\approx$ 0.1%–2.1% with 1.5pp eaten up by excessive information production

Thank you for your time and attention!

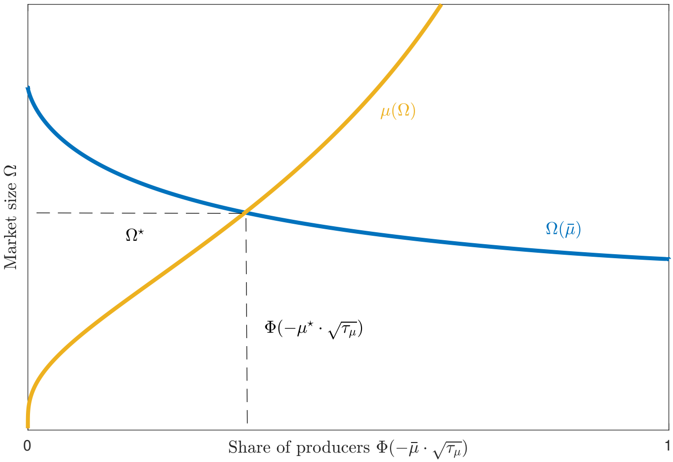
Information Shifter

For a given demand shifter $\delta(\cdot)$, the information-shifter is given by:

$$g(\boldsymbol{\tau}) \equiv \left[\tau_\omega \cdot \delta(\bar{\gamma})^{\frac{\theta-1}{\theta}} + (1-\tau_\omega) \cdot \delta(\underline{\gamma})^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}} \cdot \exp^{\frac{1}{2} \cdot \frac{\theta-1}{\theta} \cdot \frac{1}{\tau_a} \cdot \frac{\tau_a + \theta \cdot \tau_\upsilon}{\tau_a + \tau_\upsilon}}$$

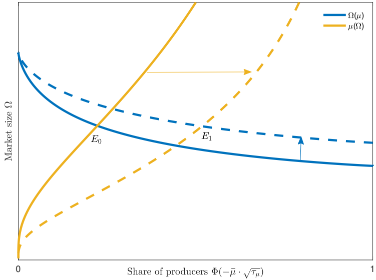
Back

Equilibrium in a Figure

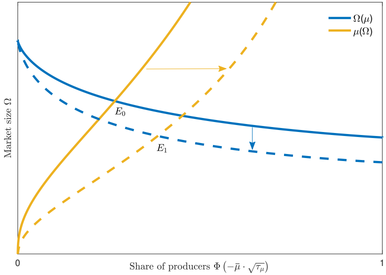


Back

Comparative Statics in a Figure



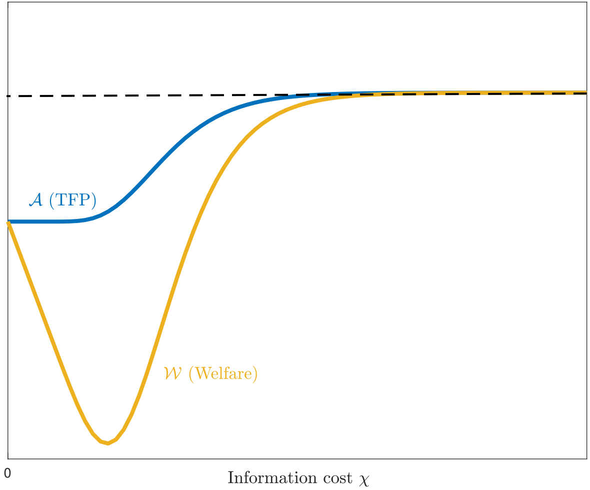
Panel (a): a fall in χ



Panel (b): a rise in $\bar{\boldsymbol{\tau}}$

Back

Socially Destructive Advancements in Data-Processing Technologies



Back

Allocations of Optimal Trading Mechanisms

Allocations: Share of variety j allocated to type-I consumer given by:

$$\alpha_j^H \equiv \alpha^H(\gamma_j) = \frac{\delta_H^{\theta-1}}{\gamma_j \cdot \delta_H^{\theta-1} + (1-\gamma_j)\cdot(\psi(\gamma_j)\cdot\delta_L)^{\theta-1}},$$

$$\alpha_j^L \equiv \alpha^L(\gamma_j) = \frac{(\psi(\gamma_j)\cdot\delta_L)^{\theta-1}}{\gamma_j \cdot \delta_H^{\theta-1} + (1-\gamma_j)\cdot(\psi(\gamma_j)\cdot\delta_L)^{\theta-1}},$$

where $\psi(\gamma_j) \leq 1$ is the distortion due to asymmetric information:

$$\psi_j \equiv \psi(\gamma_j) = \begin{cases} \left[\frac{1-\gamma_j\cdot\left(\frac{\delta_H}{\delta_L}\right)^{\frac{\theta-1}{\theta}}}{1-\gamma_j}\right]^{\frac{\theta}{\theta-1}} & \text{if } \delta_L^{\frac{\theta-1}{\theta}} \geq \gamma_j \cdot \delta_H^{\frac{\theta-1}{\theta}} \\ 0 & \text{otherwise} \end{cases}$$

Back

Summary statistics

Variable Name	Obs.	Mean	Std.	Median
Revenue	12,423	3,843	11,721	771.98
Profits	12,423	291.83	1,422	30.00
Capital	12,332	1,068	4,747	121.73
Investment	12,416	188.10	873	27.23
Wages	807	1,895	4,046	358.80
Debt	12,378	1,277	9,016	43.03
Assets	12,423	5,841	22,558	995.51
Employment	12,341	13.17	32.93	2.93
Revenue/capital	12,328	13.33	32.29	7.01
Forecast	12,328	14.81	51.83	7.04
Forecast Log	12,328	1.95	1.13	1.95
Forecast Error	12,084	-0.13	1.91	0.01
Forecast Error Log	12,080	-0.01	0.13	0.00

Notes: The table reports descriptive statistics for the sample of 2,472 firms from 2002-2022 in the combined Compustat-I/B/E/S database. The units of the first seven rows are USD millions. The employment row is in '000-employees. The first eight rows capture, respectively, firm revenue, GAAP net-profits, book value of the capital stock, total value of capital expenditures, end-of-period total liabilities and assets, overall expenditures to labor and related expenses, and the total number of employees. The next three rows measure revenue scaled by a firm's tangible capital and the (log) of the year-ahead forecast. The final two rows are for the year-ahead forecast error defined as realized future (log)-revenue minus (log of) the forecast. In the final two rows, observations have been removed that are in top and bottom one percent of the error distribution.

Back

Auxiliary data comments

Comments about data validity:

1. Expectations are close to unbiased: dynamic considerations? data
2. Similar relationships hold in the Duke-CFO survey (GDP exp.) data
3. Expectation changes are associated with input changes data
4. Accuracy of expectations correlates with firm-level outcomes data

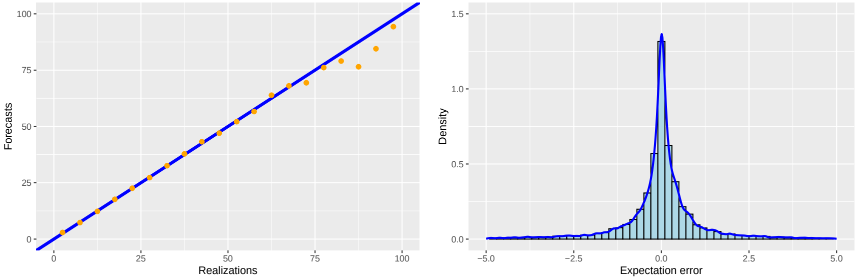
Additional reason:

- The arguments that we provide only require...
- ... that changes in forecasts in part reflect changes in information
- We do not require the complete absence of strategic/behavioral drivers

See also Tanaka et al (2021), Chen et al. (2023), and Chen et al. (2024) back

Data validation: unbiased

Figure: forecasts, realizations, and errors



Panel a: forecasts and realizations   Panel b: distribution of errors

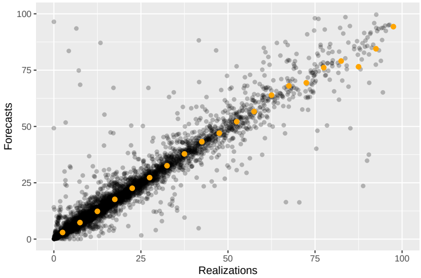
Comments:

- Errors are close to unbiased and the error distribution is symmetric

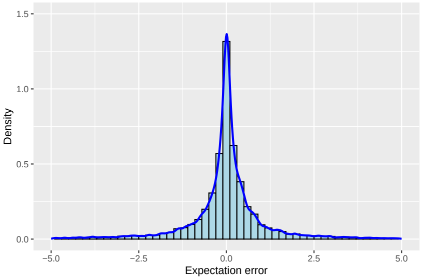
back

Data validation: unbiased

Figure: forecasts, realizations, and errors



Panel a: forecasts and realizations



Panel b: distribution of errors

Comments:

- Errors are close to unbiased and the error distribution is symmetric back

Data validation: expectations matter for input choices

Table: employment, investment, and forecasts

	Realized increase (%)	
	Employment	Investment
Revenue forecast (%)	0.069***	0.140**
	(0.007)	(0.085)
	(0.070)	(0.085)
Firm age (quintile)	-2.072	-5.794***
	(1.627)	(4.198)
Observations	10,260	10,277
Firm FE	✓	✓
Time FE	✓	✓
F statistic	52.856***	137.772***

Note: estimate of outcome (t) on revenue forecast (t), controlling for age and time and firm fixed effects

Data validation: expectations matter for input choices (cont'd)

Table: employment, investment, and forecasts errors

	Realized increase (%)		
	Employment	Investment	Forecast
log. revenue error	0.858***	1.001***	2.336**
	(0.074)	((0.211)	(1.013)
Firm age (quintile)	-0.395*	-0.259	-1.056
	(0.267)	(0.580)	(7,342)
Observations	10,034	10,088	8,722
Firm FE	✓	✓	✓
Time FE	✓	✓	✓
F statistic	69.462***	11.738***	14.294***

Note: estimate of outcome (t) on realized error (t),
controlling for age and time and firm fixed effects

back

Accuracy has improved over time 1/2

	Panel a: revenue errors and time			
	Absolute error		Squared error	
	(1)	(2)	(3)	(4)
Time	-0.024^{***}	-0.015^{***}	-0.028^{***}	-0.024^{***}
	(0.003)	(0.002)	(0.007)	(0.005)
Constant	1.251^{***}	1.135^{***}	1.303^{***}	1.219^{***}
	(0.037)	(0.026)	(0.085)	(0.065)
Observations	12,566	12,562	12,566	12,562
Covid dummy	×	✓	×	✓
Residual std. error	1.835	1.282	4.302	3.148
F statistic	66.99^{***}	22.21^{***}	17.68^{***}	22.57^{***}

	Panel b: size and time			
	50th perc.	70th perc.	80th perc.	90th perc.
	(1)	(2)	(3)	(4)
Time	0.011^{***}	0.012^{***}	0.009^{***}	0.005^{***}
	(0.002)	(0.001)	(0.001)	(0.001)
Constant	0.567^{***}	0.374^{***}	0.215^{***}	0.111^{***}
	(0.015)	(0.016)	(0.007)	(0.005)
Observations	21	21	21	21
Residual std. error	0.029	0.032	0.021	0.014
F statistic	119.56^{***}	116.05^{***}	153.64^{***}	99.41^{***}

Back

Accuracy has improved over time 2/2

	Panel a: sector fixed effects			
	Absolute error		Squared error	
	(1)	(2)	(3)	(4)
Time	-0.024^{***}	-0.016^{***}	-0.027^{***}	-0.026^{***}
	(0.003)	(0.002)	(0.007)	(0.005)
Constant	0.586^{***}	1.818^{**}	0.405^{***}	3.472
	(0.117)	(0.860)	(0.106)	(2.738)
Observations	12,567	12,563	12,567	12,563
Covid dummy	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓
Residual std. error	1.814	1.269	4.282	3.141
F statistic	17.32^{***}	13.71^{***}	7.15^{***}	5.66^{***}
	Panel b: sector×time fixed effects			
	Absolute error		Squared error	
	(1)	(2)	(3)	(4)
Time	-0.023^{***}	-0.011^{***}	-0.028^{***}	-0.020^{***}
	(0.003)	(0.002)	(0.007)	(0.005)
Constant	1.246^{***}	1.114^{***}	1.301^{***}	1.209^{***}
	(0.033)	(0.026)	(0.082)	(0.067)
Observations	12,567	12,563	12,567	12,563
Sector×time FE	✓	✓	✓	✓
Residual std. error	1.652	1.257	4.184	3.176
F statistic	79.09^{***}	29.37^{***}	18.52^{***}	15.54^{***}

Notes: Panel least-squares estimates from the merged I/B/E/S-Compustat sample. Panel a: estimate of the coefficient of the absolute value (squared value) of individual one-year ahead revenue errors on time, after having partialled out for sector (NAICS-2) fixed effects and a COVID dummy. Revenue errors are scaled by a firm's tangible capital stock and normalized by the overall average absolute (squared) error in the sample. The top and bottom 1 percent of errors have been removed. Panel b instead partials out for sector×time fixed effects beforehand.Columns (1) and (3) are in levels, whereas Columns (2) and (4) pertain to the logs of variables. Robust standard errors in parentheses. Sample: 2002–2022.

Back

Firms have become larger

	Panel a: revenue errors and time			
	Absolute error		Squared error	
	(1)	(2)	(3)	(4)
Time	-0.024^{***}	-0.015^{***}	-0.028^{***}	-0.024^{***}
	(0.003)	(0.002)	(0.007)	(0.005)
Constant	1.251^{***}	1.135^{***}	1.303^{***}	1.219^{***}
	(0.037)	(0.026)	(0.085)	(0.065)
Observations	12,566	12,562	12,566	12,562
Covid dummy	×	✓	×	✓
Residual std. error	1.835	1.282	4.302	3.148
F statistic	66.99^{***}	22.21^{***}	17.68^{***}	22.57^{***}
	Panel b: size and time			
	50th perc.	70th perc.	80th perc.	90th perc.
	(1)	(2)	(3)	(4)
Time	0.011^{***}	0.012^{***}	0.009^{***}	0.005^{***}
	(0.002)	(0.001)	(0.001)	(0.001)
Constant	0.567^{***}	0.374^{***}	0.215^{***}	0.111^{***}
	(0.015)	(0.016)	(0.007)	(0.005)
Observations	21	21	21	21
Residual std. error	0.029	0.032	0.021	0.014
F statistic	119.56^{***}	116.05^{***}	153.64^{***}	99.41^{***}

Back

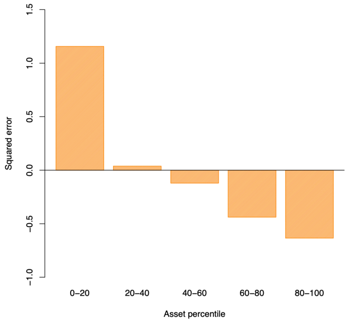
Larger firms have better forecasts

	Squared (log.) revenue errors					
	(1)	(2)	(3)	(4)	(5)	(6)
Firm size	−0.468***	−0.454***		−0.423**	−0.430*	−0.293***
	(0.055)	(0.052)		(0.135)	(0.226)	(0.067)
Firm size (1)			2.082***			
			(0.210)			
Firm size (2)			0.719***			
			(0.118)			
Firm size (3)			0.311***			
			(0.104)			
Firm size (4)			0.174***			
			(0.065)			
Time		0.007			0.117***	
		(0.007)			(0.028)	
Firm age		−0.063**	−0.073**	0.143**	0.117	0.140**
		(0.032)	(0.059)	(0.063)	(0.093)	(0.054)
Log rev. volatility				0.013*		
				(0.007)		
Log TFP volatility						1.089**
						(0.362)
Observations	12,488	12,488	12,488	6,817	2,569	5,238
Sector FE	✓	✓	✓	×	×	×
Firm FE	×	×	×	✓	✓	✓
Time FE	×	×	✓	✓	×	✓
Panel GMM	×	×	×	×	✓	×
F statistic	3.911***	3.922***	4.321***	9.240***	NA.	2.837***

Back

Alternative measures I/II: size and assets

	Panel b: asset size and time*			
	50th perc.	70th perc.	80th perc.	90th perc.
	(1)	(2)	(3)	(4)
Time	0.019***	0.025***	0.027***	0.018***
	(0.001)	(0.002)	(0.002)	(0.003)
Constant	0.557***	0.387***	0.232***	0.066***
	(0.014)	(0.020)	(0.018)	(0.023)
Observations	21	21	21	21
Residual std. error	0.026	0.039	0.048	0.064
F Statistic	445.104***	310.202***	239.202***	62.681***



Back

Alternative measures II/II: accuracy measures

	Absolute error		*Squared error*		*Squared error log*
	(1)	(2)	(3)	(4)	(5)
Firm size	-0.353^{***}	-0.293^{***}	-0.560^{***}	-0.450^{***}	
	(0.034)	(0.040)	(0.072)	(0.084)	
Firm assets					-0.272^{***}
					(0.056)
Time	-0.006		-0.003		
	(0.006)		(0.012)		
Firm age	0.004	0.023	0.046	0.058	0.022
	(0.016)	(0.024)	(0.035)	(0.052)	(0.031)
Rev. volatility		0.010^{***}		0.010^{**}	0.004
		(0.002		(0.002)	(0.011)
Observations	12,489	6,819	12,489	6,819	6,834
Time FE	×	✓	×	✓	✓
Sector FE	✓	✓	✓	✓	✓
F Statistic	10.460^{***}	7.704^{***}	5.494^{***}	5.043^{***}	2.083^{***}

Back

Other revenue measures: pct. increase

	Absolute error		*Squared error*		
	(1)	(2)	(3)	(4)	(5)
Firm size	−0.210***	−0.177***	−0.372***	−0.308***	
	(0.025)	(0.031)	(0.056)	(0.064)	
Time	0.0005		0.005		−0.014**
	(0.004)		(0.008)		(0.007)
Firm age	−0.050***	0.009	−0.095**	0.026	
	(0.016)	(0.025)	(0.039)	(0.051)	
Rev. volatility (pct.)		0.004***		0.005***	
		(0.001)		(0.002)	
Observations	10,083	5,700	10,083	5,700	10,138
Time FE	×	✓	×	✓	×
Sector FE	✓	✓	✓	✓	×
F Statistic	5.589***	4.648***	3.491***	3.163***	6.747***

Back

Other variables: firm profits and capex

	Panel a: errors and time			
	Profits		Capex	
	Abs. error	Sqr. error	Abs. error	Sqr. error
Time	-0.039^{***}	-0.068^{***}	-0.012^{***}	-0.001
	(0.010)	(0.024)	(0.003)	(0.011)
Constant	1.398^{***}	1.691^{***}	1.126^{***}	0.986^{***}
	(0.126)	(0.303)	(0.035)	(0.120)
Observations	2,487	2,487	1,839	1,839
Residual std. error	2.482	6.187	1.871	6.982
F statistic	15.27^{***}	7.385^{***}	12.91^{***}	0.011

	Panel b: errors and size			
	Profits		Capex	
	Abs. error	Sqr. error	Abs. error	Sqr. error
Firm size	-0.361^{***}	-0.578^{***}	-0.074^{***}	-0.130^{***}
	(0.082)	(0.195)	(0.018)	(0.047)
Firm age	-0.029	0.047	-0.070^{***}	-0.117^{*}
	(0.080)	(0.147)	(0.016)	(0.061)
Constant	0.456	0.680	-0.113^{***}	-0.144
	(0.314)	(0.486)	(0.042)	(0.165)
Observations	2,487	2,487	1,839	1,839
Sector FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Residual std. error	2.347	6.209	1.740	6.744
F statistic	2.694^{***}	1.056	7.906^{***}	2.849^{***}

Back

Other variables: real GDP growth

	Squared error		Absolute error	
	(1)	(2)	(3)	(4)
Firm size	−0.077**	−0.060*	−0.050***	−0.045***
	(0.031)	(0.032)	(0.016)	(0.016)
GDP familiarity		0.033		0.018
		(0.061)		(0.030)
Constant	1.639***	1.676***	1.648***	1.681***
	(0.561)	(0.650)	(0.286)	(0.331)
Observations	1,584	1,464	1,584	1,464
Sector FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
R^2	0.177	0.201	0.241	0.271
Residual std. error	1.578	1.544	0.812	0.793
F Statistic	19.774***	20.165***	29.175***	29.814***

Back Data comments

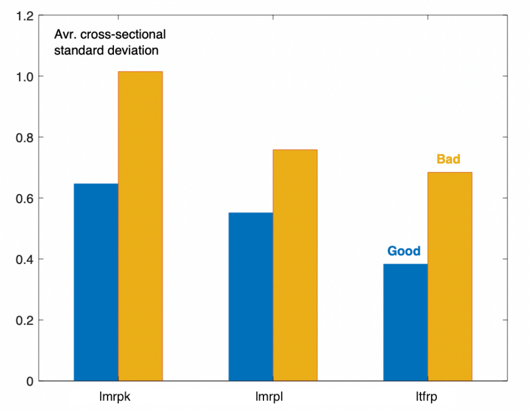
Large acquisitions (5 pct. of firm assets)

	Squared error	Absolute error
	(1)	(2)
Large acquisitions	-0.139^{*}	-0.060^{***}
	(0.007)	(0.033)
Large acquisitions(-1)	-0.120^{*}	-0.083^{***}
	(0.066)	(0.031)
Large acquisitions(-2)	-0.106^{*}	-0.079^{***}
	(0.063)	(0.031)
Large acquisitions(-3)	0.035	-0.043
	(0.067)	(0.031)
Firm age	0.037	0.004
	(0.034)	(0.017)
Revenue volatility	0.002^{***}	0.001^{***}
	(0.000)	(0.000)
Observations	5,108	5,108
Sector FE	✓	✓
Time FE	✓	✓
Residual std. error	2.325	1.050
F statistic	1.535^{***}	2.879^{***}

Note: time since large acquisition Back

Validation: better forecasters allocate inputs more efficiently

Figure: accuracy and dispersion in marginal revenue products



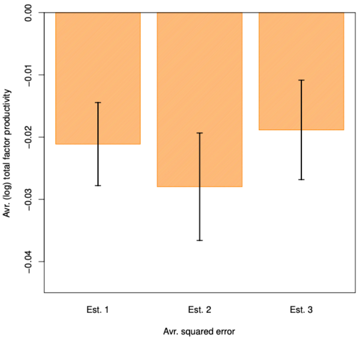
Misallocation: lower on all measures for better forecasters

Back

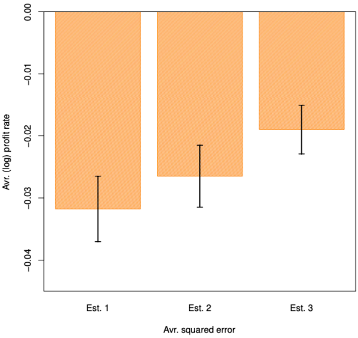
Production function: $\log y_t = \log A_t + \alpha \cdot \log k_t + (1-\alpha) \cdot \log l_t, \quad \alpha = 1/3$

Validation: better forecasters are more productive and more profitable

Figure: productivity, profits, and forecast accuracy



Panel a: productivity



Panel b: profitability

Est. 2 controls: size, age, and sector fixed effects Est. 3 controls: + trim 1%

Production function: $\log y_t = \log A_t + \alpha \cdot \log k_t + (1-\alpha) \cdot \log l_t$ (CES elast. $\theta = 5$)

Back

Validation: better forecasters grow faster

Table: Accuracy and firm growth

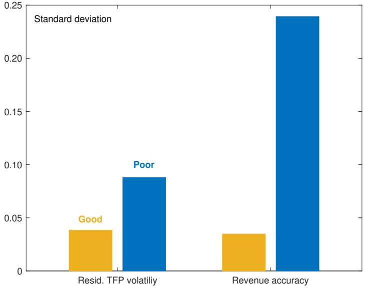
	Employment		*Revenue*	
	(1)	(2)	(3)	(4)
Good firms	0.100***	0.110***	0.087**	0.081**
	(0.037)	(0.037)	(0.037)	(0.039)
Initial employment	0.067***	0.066***		
	(0.003)	(0.003)		
Initial revenue			1.175***	1.152***
			(0.001)	(0.001)
Observations	10,186	10,186	10,234	10,234
Sector FE	✓	×	✓	×
Time FE	✓	✓	✓	✓
Industry FE	×	✓	×	✓
Added controls	age	age	age	age
F statistic	1,681***	272.7***	1,573***	271.2***

Notes: Panel estimates from the merged I/B/E/S-Compustat sample. Columns (1) and (2) report estimates of firm subsequent employment (2007-2022) on whether a firm was a good or bad forecaster in a initial period (2002-2007), firm initial age, as well firm initial employment (2002). We further control for time and sector (NAICS-2) or industry (NAICS-4) fixed effects. Columns (3) and (4) report estimates which instead focus on firm revenue. Robust (clustered) standard errors in parentheses. Sample: 2002-2022.

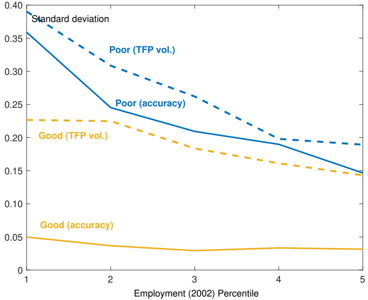
Back

Validation: better forecasters have less volatile TFP

Figure: Forecast accuracy and productivity volatility



Panel a: accuracy relationship



Panel b: size quintiles

Panel a controls: size and sector fixed effects.

Panel b controls: sector fixed effects

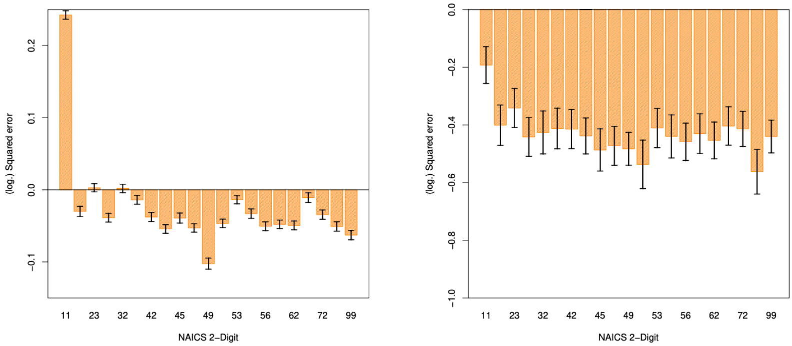
Good accuracy: below 20th percentile of the MSE of one-year revenue forecasts

Poor accuracy: above 80th percentile of the MSE of one-year revenue forecasts

Back

Across sector analysis

Figure A.1: sector-based heterogeneity



Sector heterogeneity: more along size-dimension than time-dimension Back

Firms have become larger

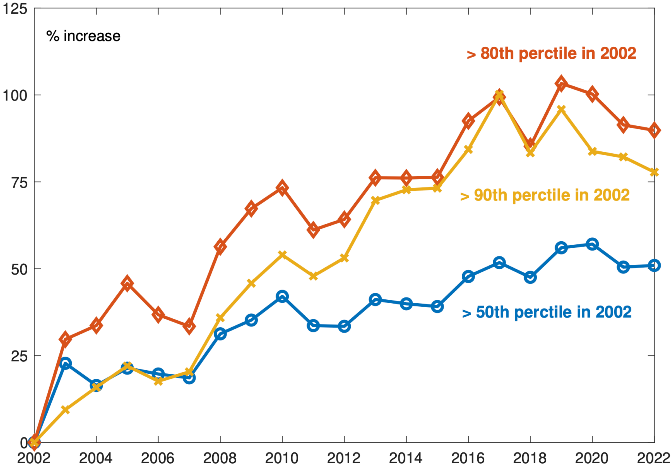


Figure: firm size by number of employees

Table Relative

Accuracy and intangible capital

	Sqr. error	Abs. error
Firm acq. stock of intangibles	-0.070*	-0.038***
	(0.040)	(0.015)
Firm size	-0.405***	-0.211***
	(0.041)	(0.019)
Firm age	-0.051	-0.034***
	(0.034)	(0.013)
Observations	11,371	11,371
Sector FE	✓	✓
Time FE	✓	✓
F statistic	3.721***	6.499***

Intangible capital: measured as in Chiavari and Goraya (2023). Back

Firms have become larger

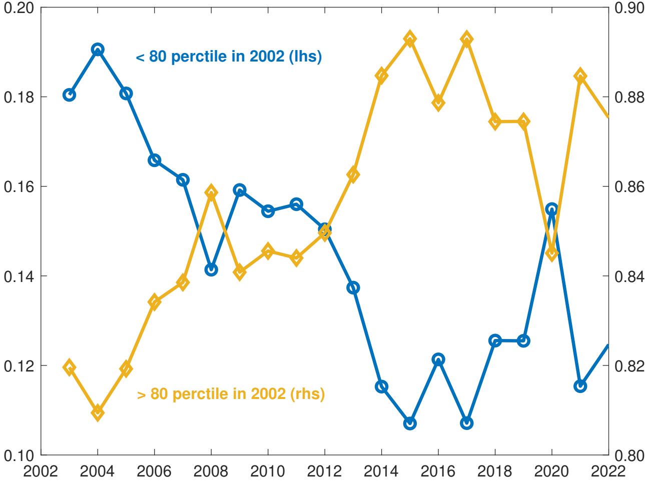


Figure: Relative firm size by number of employees

Back

Validation exercise

Table: accuracy and firm size

	log. squared errors	
	Data	Model
Size (labor)	-0.468^{***}	-0.408^{***}
	(0.052)	(0.001)
Time (years)	0.007	-0.072^{***}
	(0.007)	(0.001)

Back

Related literature

- **Evidence on effects of data and information**: Brynjolfsson and McElheran (2008, 2016), Philippon et al. (2016), Goldfarb and Tucker (2019), Tanaka et al. (2020), Coibion et al. (2023) ...
- **Macro implications of data and information**: David et al. (2016), Begenau et al. (2018), Farboodi et al. (2019), Li (2019), Baley and Veldkamp (2024), Eeckhout and Veldkamp (2024) ...
- **Intangibles and rising firm concentration**: Autor et al. (2017), Guttierez and Philippon (2017), DeLoecker et al. (2018), Chiavari and Goraya (2021), Hsieh and Rossi-Hansberg (2023) ...

Refinement 1: information in the long run

We enrich our baseline framework with capital and firm entry/exit

- Dynamics and capital: household preferences,

$$\mathcal{U} = \sum_t \beta^t \cdot \log\left(C_t\right),$$

where C_t is consumption at time t. We focus on the case with fixed labor.

△ Each variety is supplied by a monopolistically competitive firm:

$$y_{it} = A_{it} \cdot k_{it}^{\alpha} \cdot l_{it}^{1-\alpha}.$$

Capital k_{it} depreciates at rate $\gamma \in (0,1)$ and can be rented at rate r.

- Firm entry/exit: firms can activate a variety by paying a fixed cost f; varieties become obsolete at rate $\eta \in (0,1)$

Steady state: closely resembles baseline model with elastic input

Refinement 1: information in the long run

Figure: Capital and variety accumulation

Panel a: TFP   Panel b: welfare

Comments:

- Re-calibrate model to target the same moments
- The cost χ declines with a similar magnitude to before
- Extended quantification shows amplified benefits relative to before   Back
- Reason: elastic inputs + firm/varieties entry   ... but still $> 80\%$ of potential

Refinement 2: recalibrate productivity parameters

Maybe productivity process changed?

- Confound estimates due to two-sided relationship between size and information

Recalibration procedure:

- Recalibrate both productivity (i.e., τ_μ, τ_a) and information parameters (i.e., $\chi, \bar{\tau}$)
- Use last three years of sample as target moments

Parameters	TFP (%)	Welfare (%)
All parameters	(11.1, 13.1)	(5.5, 7.1)
Productivity only	(4.2, 4.3)	(4.5, 4.6)
Information only	(5.5, 6.9)	(0.8, 2.2)

Similar ballpark estimates: TFP improvements $\approx$ 5–7%; welfare $<$ 2.5%

Back

Refinement 3: non-linear pricing in the retail sector

Where within our range?

- Answer depends on the severity of price distortions
- Measuring these requires detailed product-level transaction data

Bornstein and Peter (2024):

- Estimate relationship between log of t_j^l/q_j^l and log of q_j^l
- Kilts-Nielsen Retail Scanner Data from October 2017

	Bornstein-Peter		*Model*	
	(1)	(2)	Best case	Worst case
Log quantity	-0.60	-0.39	0.0	-0.34
	(0.0001)	(0.0001)	(0.02)	(0.02)
Sample	All	Expensive	All	All
Fixed effects	✓	✓	✓	✓
Observations	88.3M	4.5M	826 x2	826 x2

Estimates: suggests economy may be close to the worst-case scenario

Back