# Measuring Productivity: Lessons from Researcher-Designed Surveys and Productivity Benchmarking

David Atkin, MIT
Amit Khandelwal, Columbia GSB
Adam Osman, UIUC

# Motivation

Productivity differences can explain large differences in income per capita across and within countries

- Obvious Questions:
  - How large are productivity differences across firms?
  - What drives these differences?
  - What policies are most effective at raising productivity and reducing dispersion?
  - Etc.
- Active area of research (e.g. see Syverson 2011 review)

Central challenge for researchers analyzing productivity: it is *never observed*

# Measuring TFPQ and TFPR

Researchers typically want physical output conditional on physical inputs (TFPQ)

- But input and output quantities typically not available
- When physical quantities are available, substantial measurement error concerns
- Even then, large differences in specifications and quality make comparisons across firms difficult
    - Administrative product codes often coarse (e.g., 1000+ barcodes within single HS10 code)

# Measuring TFPQ and TFPR

Researchers typically want physical output conditional on physical inputs (TFPQ)

- But input and output quantities typically not available
- When physical quantities are available, substantial measurement error concerns
- Even then, large differences in specifications and quality make comparisons across firms difficult
    - Administrative product codes often coarse (e.g., 1000+ barcodes within single HS code)

Most firm-level datasets only provide expenditure/revenue data so lit. often uses TFPR

- Also captures differences in quality and markups
- If firm capabilities come from ability to produce both quantity and quality, TFPR may be reasonable

# Understanding and Mitigating Measurement Issues

Existing literature has pursued various approaches to understand/mitigate these measurement issues

- Typically draw on additional variation in same datasets alongside (often strong) identifying assumptions
  - E.g. Ackerberg, Caves and Frazer (2015), Kim, Petrin and Song (2016), de Loecker et al. (2016), Garcia-Marin and Voigtlander (2019), de Loecker and Goldberg (2014) survey
  - Hallak and Sividasan (2013) and Sutton and Trefler (2015) explore multidimensional firm productivity; Foster, Haltiwanger and Syverson (2018) separate demand shocks from TFPQ

# Understanding and Mitigating Measurement Issues

Our approach: focus on a specific industry—flat-weave rugs as in Atkin et al. (2017)—and design surveys to directly address these measurement issues

# Understanding and Mitigating Measurement Issues

Our approach: focus on a specific industry—flat-weave rugs as in Atkin et al. (2017)—and design surveys to directly address these measurement issues

- Detailed product specifications and external quality assessments allow us to calculate:
  - Quantity productivity (ability to produce quantity with a given set of inputs, TFPQ)
  - Quality productivity (ability to produce quality with a given set of inputs, TFPZ)
  - Capability (combine TFPQ and TFPZ using consumer's quality-quantity tradeoff, i.e. "TFPQ with quality-adjusted quantities")

# Understanding and Mitigating Measurement Issues

Our approach: focus on a specific industry—flat-weave rugs as in Atkin et al. (2017)—and design surveys to directly address these measurement issues

- Detailed product specifications and external quality assessments allow us to calculate:
  - Quantity productivity (ability to produce quantity with a given set of inputs, TFPQ)
  - Quality productivity (ability to produce quality with a given set of inputs, TFPZ)
  - Capability (combine TFPQ and TFPZ using consumer's quality-quantity tradeoff, i.e. "TFPQ with quality-adjusted quantities")

- To assess performance of various survey-based productivity measures, compare with productivity benchmarking exercises which we argue are closest to true productivity

# Preview of Results

1. Standard TFPQ performs poorly at measuring quantity productivity
   - Shows excessive dispersion across firms
   - Is inversely correlated with quality productivity!

2. Controlling for product specifications–effectively making apples-to-apples comparisons–goes some way towards remedying deficiencies in (1)

3. Although standard TFPR does better than TFPQ at capturing a firm's broad capabilities, it performs worse than methods combining survey information with explicit quality measures

4. Quality-productivity dispersion similar to quantity-productivity dispersion, capability more dispersed

# Outline

1. Industry Background
2. Survey Design and Productivity Benchmarks
3. Measuring Productivity
4. Comparing Productivity Measures: Six Results
5. Conclusion

# Industry: Flat-Weave Rug Producers in Fowa, Egypt

Focus on firms producing a type of kilim rug called “duble”

Producers receive orders with particular set of specifications

- Codifiable attributes of the rug that are typically chosen by the buyer
- E.g. design, thread type, thread count, colors, duble subcategory
- Duble are already subset of 10 digit HS, we observe 435 spec combinations in surveys

Producers prepare inputs, install threads on double-treddle foot powered-looms, weave rug

# Industry: Flat-Weave Rug Producers in Fowa, Egypt

Focus on firms producing a type of kilim rug called “duble”

Producers receive orders with particular set of specifications

- Codifiable attributes of the rug that are typically chosen by the buyer
- E.g. design, thread type, thread count, colors, duble subcategory
- Duble are already subset of 10 digit HS, we observe 435 spec combinations in surveys

Producers prepare inputs, install threads on double-treddle foot powered-looms, weave rug

Rugs also differ in quality

- Depends on weaving technique and is difficult to codify or contract on
- E.g. how flat a rug lies determined by how skillfully thread is installed on loom, whether threads held correctly while weaving

# Survey Design and Data

Survey all 219 firms between 1-5 workers

- Same sample as Atkin, Khandelwal and Osman (2017)
- Exclude baseline/pre-treatment rounds to ensure export status not changing

Quarterly firm surveys (5 post-baseline rounds):

- Surveys at product-line level (capturing rugs produced in last month)
- Output and input prices/quantities
- Labor hours, capital equipment, materials used
- Rug specifications: thread type, thread count, design, colors, market segment, duble subcategory
- External quality assessor graded each rug at time of survey on 11 dimensions
  - Corners, waviness, weight, touch, packedness, warp thread tightness, firmness, design accuracy, warp thread packedness, inputs, and loom quality

Will construct multiple productivity measures using these survey data

# Productivity Benchmarking in a Laboratory

Set up a controlled laboratory:

- Firms paid a flat fee to produce an identical rug (domestic design, 140cm by 70cm, should weigh 1750g)
- We provide all firms with identical material inputs and loom in rented facility
- Recorded dimensions, weight, time to produce the rug
  - Provides direct measure of quantity productivity: $m^2$ per labor hour (**Lab TFPQ**)
  - "Benchmark" as lab ensures inputs, product specifications identical across firms
- Rug quality scored anonymously by quality assessor and prof of handicraft science
  - Will use these scores to benchmark quality productivity

# Outline

1. Industry Background

2. Survey Design and Productivity Benchmarks

**3. Measuring Productivity**

4. Comparing Productivity Measures: Six Results

5. Conclusion

# Measuring Productivity – Quantity Productivity

Cobb-Douglas production function

$$x = \phi_u l^{\alpha_l} k^{\alpha_k} e^{\epsilon}$$

- Output ($x$), Labor hours ($l$), Looms ($k$).
- For transparency, estimate via OLS (similar results using Olley Pakes control function)
- Recover **Unadjusted TFPQ** ($\phi_u$) — does not adjust for fact different firms produce different varieties

# Measuring Productivity – Quantity Productivity

Cobb-Douglas production function

$$x = \phi_u l^{\alpha_l} k^{\alpha_k} e^{\epsilon}$$

- Output (*x*), Labor hours (*l*), Looms (*k*).
- For transparency, estimate via OLS (similar results using Olley Pakes control function)
- Recover **Unadjusted TFPQ** ($\phi_u$) — does not adjust for fact different firms produce different varieties

Features of setting reduce measurement concerns:

1. Observe quantities of *x*, *l* and *k* rather than revenues and expenditures.
2. Given simple technology, plausibly no other inputs (e.g. no accounting, logistics, human resources).
3. Record inputs used for each specific rug produced so no error in allocating inputs to outputs
4. Leontief in materials.

# Measuring Productivity – Quantity Productivity

Cobb-Douglas production function

$$x = \phi_u l^{\alpha_l} k^{\alpha_k} e^{\epsilon}$$

- Output (*x*), Labor hours (*l*), Looms (*k*).
- For transparency, estimate via OLS (similar results using Olley Pakes control function)
- Recover **Unadjusted TFPQ** ($\phi_u$) — does not adjust for fact different firms produce different varieties

Adjusting for product specification: Consider $\phi_u = \phi(\lambda; \phi_a, \gamma) = \phi_a e^{\lambda\gamma}$

- Unit input requirements vary with vector of specifications ($\lambda$)
  - E.g. a high thread count rug requires more inputs
- Residual after conditioning on specification controls uncovers **Specification-Adjusted TFPQ** ($\phi_a$)

# Measuring Productivity – Quality Productivity

Similar specification rugs have very different quality, and like quantity, quality raise revenue

But how to measure **Quality Productivity (TFPZ)**?

# Measuring Productivity – Quality Productivity

Similar specification rugs have very different quality, and like quantity, quality raise revenue

But how to measure **Quality Productivity (TFPZ)**?

Consumer valuation of quantity and quality (where $q_j$ are our 11 quality dimensions):

$$U = ((\Pi_j q_j^{\theta_j} x)^{\frac{\sigma-1}{\sigma}} + (y)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$$

Implied demand:

$$\ln x = (\sigma - 1) \sum_j \theta_j \ln q_j - \sigma \ln p + c$$

- Impose σ=2.74 (Broda & Weinstein 2006)
- *θ*’s capture consumer’s valuation of 11 quality metrics vis a vis quantity

With quality aggregate, can formulate quality production function: $\Pi_j q_j^{\theta_j} = \zeta_u l^{\beta_l} k^{\beta_k} e^{\varepsilon}$

- Recover **Unadjusted TFPZ** ($\zeta_u$) via OLS
- Similarly, recover **Specification-Adjusted TFPZ** ($\zeta_a$) after imposing: $\zeta_u = \zeta(\lambda; \zeta_a, \delta) = \zeta_a e^{\lambda\delta}$

# Measuring Productivity – Capability

Aggregate quality and quantity production functions to form a firm's **Capability (TFPC)**

$$\Pi_j q_j^{\theta_j} x = \zeta_a \phi_a e^{\lambda(\gamma+\delta)} l^{\alpha_l+\beta_l} k^{\alpha_k+\beta_k} e^{\epsilon+\varepsilon}$$

**Unadjusted TFPC** ($\phi_u \zeta_u$)

**Specification-Adjusted TFPC** ($\phi_a \zeta_a$)

# Measuring Productivity - Lab Benchmarks

**Lab TFPQ** = $0.98m^2/l$

- Specifications fixed by lab design

**Lab TFPZ** = $\Pi_j q_{lab,j}^{\hat{\theta}_j}$

- For quality productivity, aggregate quality metrics using same $\theta$'s as survey measures

**Lab TFPC** = **Lab TFPQ ✕ Lab TFPZ**

- Capabilities is product of quality and quantity productivity

# Measuring Productivity: Summary

Quantity Productivity: **Unadjusted TFPQ** ($\phi_u$), **Specification-Adjusted TFPQ** ($\phi_a$)

Quality Productivity: **Unadjusted TFPZ** ($\zeta_u$), **Specification-Adjusted TFPZ** ($\zeta_a$)

Capability: **Unadjusted TFPC** ($\phi_u \zeta_u$), **Specification-Adjusted TFPC** ($\phi_a \zeta_a$)

Also: estimate **TFPR** using revenues and input expenditures instead of quantities $x$, $l$, $k$

Compare to Lab Benchmarks: **Lab TFPQ**, **Lab TFPZ**, **Lab TFPC**

# Outline

1. Industry Background

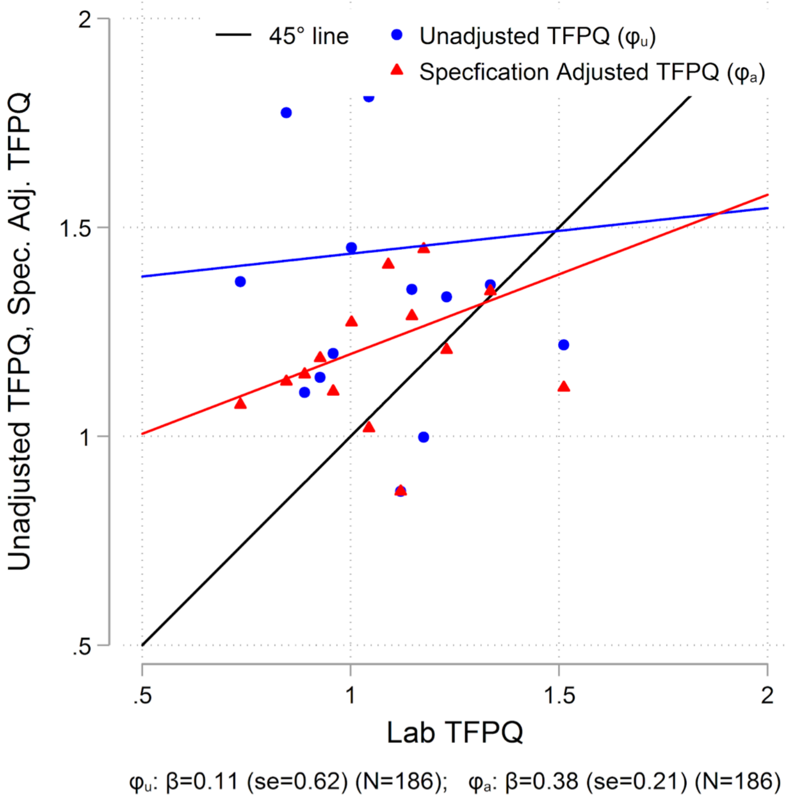
2. Survey Design and Productivity Benchmarks

3. Measuring Productivity

**4. Comparing Productivity Measures: Six Results**

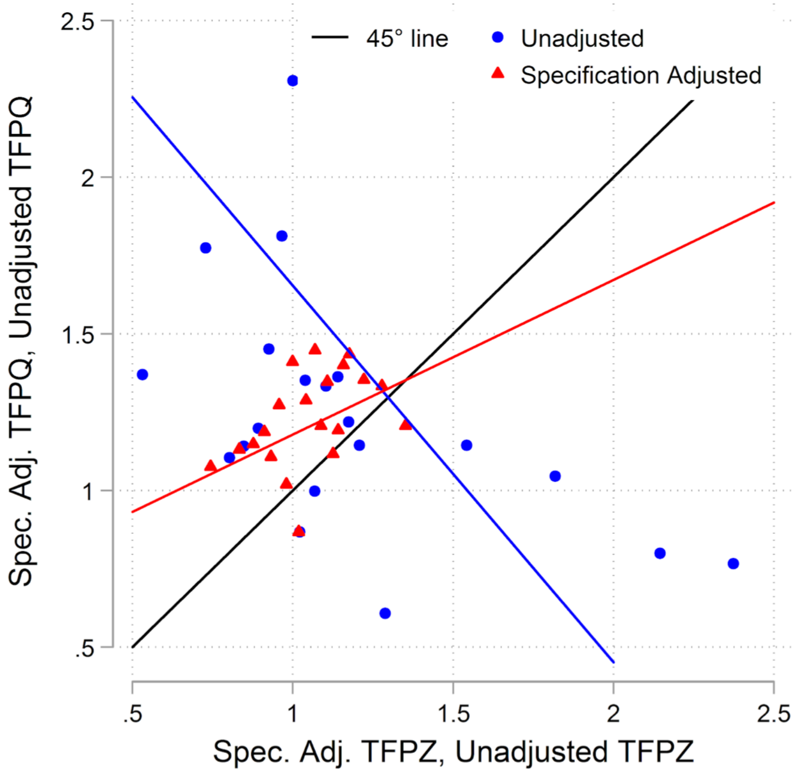
5. Conclusion

# Result 1: Importance of Adjusting for Product Spec.




- Shows value of finer product-category controls for measuring quantity productivity

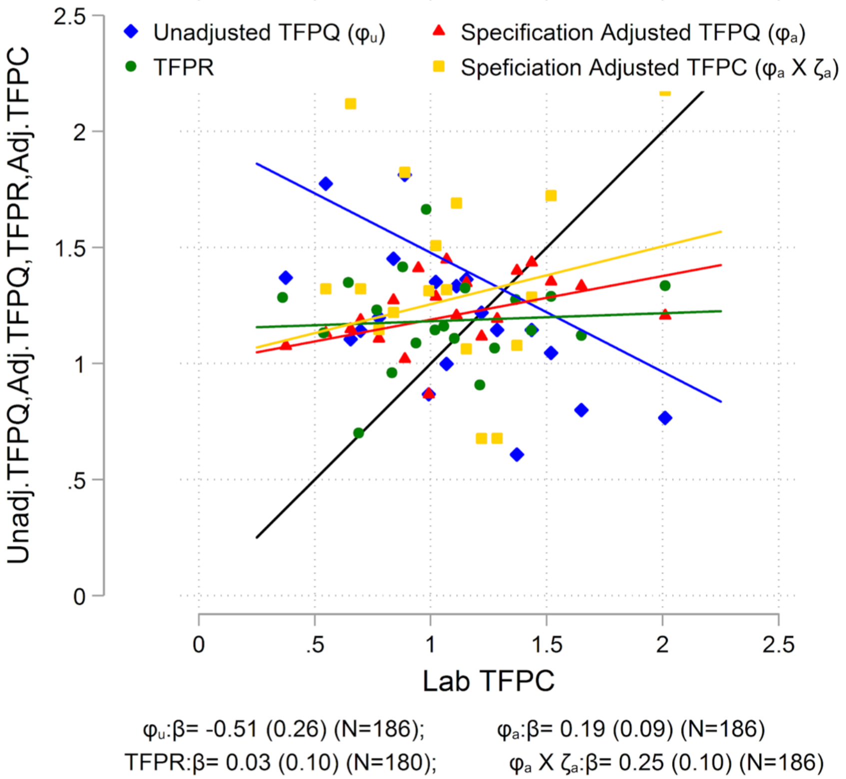
# Result 2: Quantity versus Quality Productivity




$\varphi_u$: β=-1.20 (0.19) (N=209); $\varphi_a$: β=0.49 (0.22) (N=209)

- TFPQ and TFPQ positively correlated
    - But only after spec adjusting, as more capable firms make varieties with more demanding specs

# Result 3: TFPR as a Proxy for Capabilities

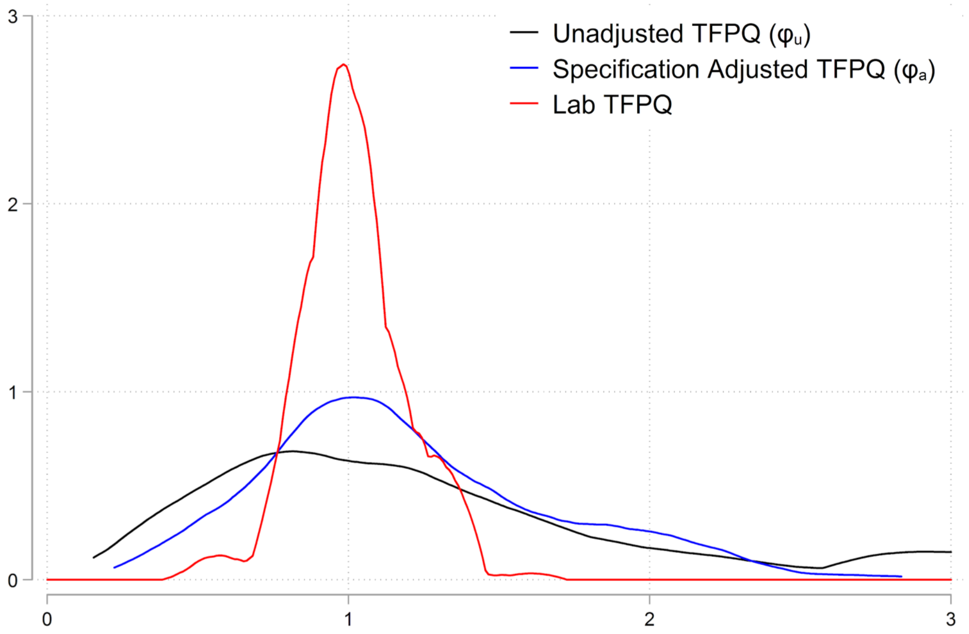



- Unadjusted TFPQ is misleading capability measure, TFPR more suitable proxy
- But specification adjusting, and/or combining quality and quantity measures superior

# Result 4: Unadjusted TFPQ Overstates Dispersion more than Spec-Adjusted TFPQ

<u>90-10 Ratios:</u>

Unadjusted TFPQ: 4.7
Spec-Adjusted TFPQ: 3.0
Lab TFPQ: 1.6

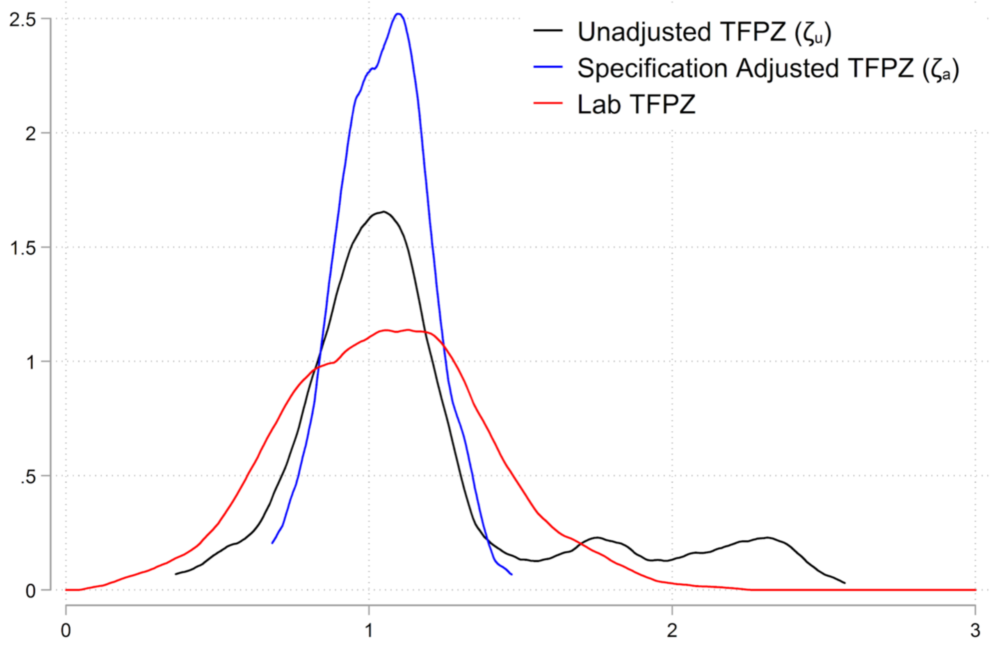



Dispersion in standard datasets may partially reflect product differentiation, rather than differences in underlying productivity

# Result 5: TFPZ Dispersion is Large

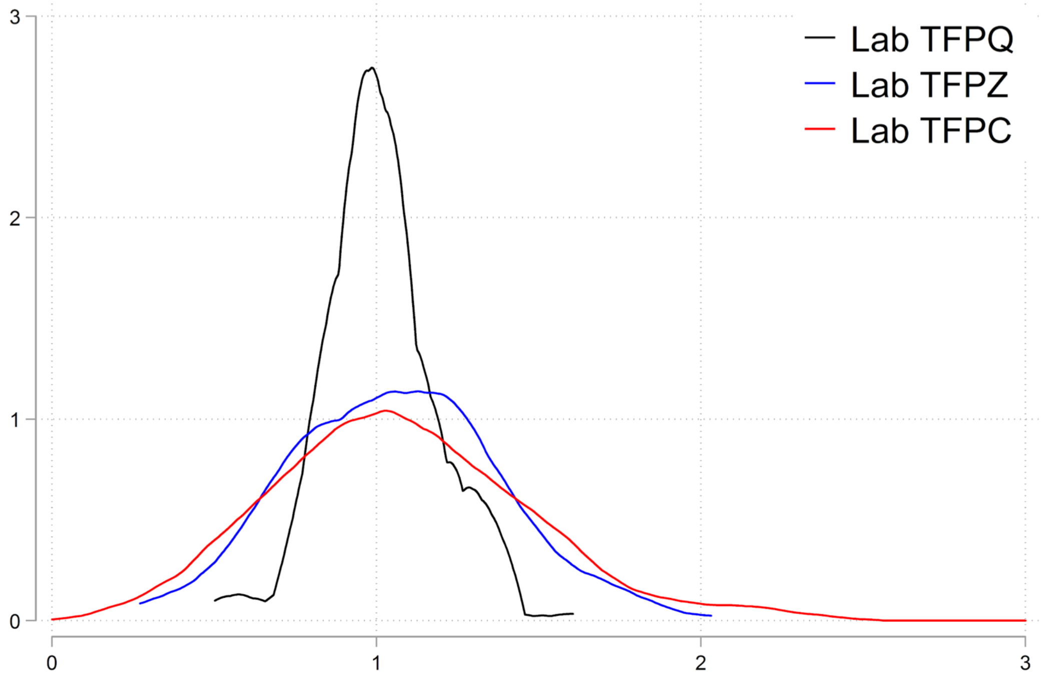
<u>90-10 Ratios:</u>

Unadjusted TFPZ: 2.6
Spec-Adjusted TFPZ: 1.5
Lab TFPZ: 2.9




Even within very narrowly defined product category, large variation in quality across firms

# Result 6: TFPC More Dispersed than TFPQ & TFPZ




90-10 Ratios:

Lab TFPQ: 1.6
Lab TFPZ: 2.9
Lab TFPC: 3.2

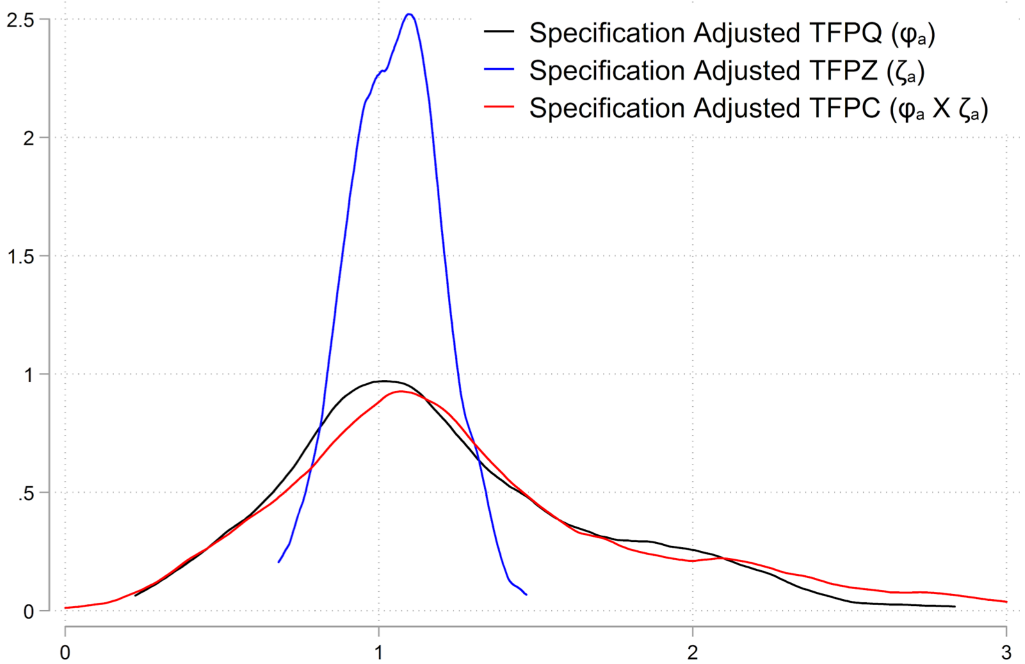
Broad capabilities of firms are more dispersed than a single dimension of productivity

- Implication of fact quantity and quality productivity positively correlated

# Result 6: TFPC More Dispersed than TFPQ & TFPZ

<u>90-10 Ratios:</u>

Spec-Adjusted TFPQ: 3.0
Spec-Adjusted TFPZ: 1.5
Spec-Adjusted TFPC: 3.6




Broad capabilities of firms are more dispersed than a single dimension of productivity

- Implication of fact quantity and quality productivity positively correlated

# Conclusions

Combination of tailored surveys and productivity benchmarks provides insight into productivity measurement

1. Standard TFPQ performs poorly at measuring quantity productivity
   - Weakly corr. with Lab TFPQ, inversely corr. with quality productivity
   - Using product specifications to make an apples-to-apples comparisons reduces measured productivity dispersion by a third

2. Substantial variation in quality productivity and even greater dispersion in capabilities
   - Implication of positive correlation between quality and quantity productivity

3. If researchers are interested in broader capabilities of firms:
   - TFPR—for all its imperfections—may be better proxy than (unadjusted) TFPQ
   - But tailored surveys and/or benchmarking may be best way to understand productivity differences across firms