

Hong, Joonkyo; Luparello, Davide

Working Paper

Robustness Report on "Within-Firm Productivity Dispersion: Estimates and Implications," by Scott Orr (2022)

I4R Discussion Paper Series, No. 181

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Hong, Joonkyo; Luparello, Davide (2024) : Robustness Report on "Within-Firm Productivity Dispersion: Estimates and Implications," by Scott Orr (2022), I4R Discussion Paper Series, No. 181, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/305229>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



No. 181

DISCUSSION PAPER SERIES

Robustness Report on “Within-Firm Productivity Dispersion: Estimates and Implications,” by Scott Orr (2022)

Joonkyo Hong

Davide Luparello

This paper received a response:

Orr, Scott. 2024. Response to Hong and Luparello (2024). *I4R Discussion Paper Series* No. 182. Institute for Replication.

October 2024

I4R DISCUSSION PAPER SERIES

I4R DP No. 181

Robustness Report on “Within-Firm Productivity Dispersion: Estimates and Implications,” by Scott Orr (2022)

Joonkyo Hong¹, Davide Luparello²

¹National Taiwan University, Taipeh/Taiwan

²Pennsylvania State University, University Park/USA

OCTOBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research

Hohenzollernstraße 1-3
45128 Essen/Germany

www.i4replication.org

ISSN: 2752-1931

Robustness Report on “Within-Firm Productivity Dispersion: Estimates and Implications,” by Scott Orr (2022)*

Joonkyo Hong

Davide Luparello

August 11, 2024

Abstract

In this study, we evaluate the reproducibility and replicability of Scott Orr’s (2022) innovative approach for identifying within-plant productivity differences across product lines. Orr’s methodology allows the estimation of plant-product level productivity, contingent upon a well-behaved pre-estimated demand system, which requires the use of carefully chosen instrumental variables (IVs) for output prices. Using Orr’s STATA replication package, we successfully replicate all primary estimates with the ASI Indian plant-level panel data from 2000 to 2007. Additionally, applying Orr’s replication codes to a sample from 2011 to 2020 reveals that the suggested IVs do not perform as expected.

KEYWORDS: Replication, Robustness, Productivity

JEL CODES: D22, D24

*Authors: Hong: National Taiwan University. E-mail: jkhong@ntu.edu.tw; Luparello: The Pennsylvania State University. E-mail: dluparello@psu.edu. We extend our sincere gratitude to Scott Orr, whose insightful feedback, thoughtful suggestions, and comments have significantly enhanced the quality of our work. We are equally indebted to Abel Brodeur for his ongoing support throughout this project. We also thank the Ministry of Statistics and Programme Implementation, Government of India, for providing the data used in this project.

1 Introduction

[Orr \(2022\)](#) introduces an innovative identification strategy for evaluating within-plant productivity across different product lines, utilizing data from the ASI Indian plant-product level panel from 2000 to 2007. This methodology involves two stages. First, the demand function is estimated using plant-product level data. Next, the researcher combines the first-order conditions from the cost minimization problem with these demand estimates to determine the allocation of input use across product lines within a plant. By recovering plant-product level input uses, the researcher can estimate the production function for each product, thereby deriving plant-product level productivity estimates. A key requirement for this approach is that the estimated demand function in the first stage must produce an invertible price elasticity matrix. In this report prepared for the Institute for Replication ([Brodeur et al. \(2024\)](#)), we first computationally reproduce the original results from the paper and then replicate them using new data.

Using the replication package provided on the JPE website, we first reproduced all the tables from [Orr \(2022\)](#) computationally. As the raw dataset was not included in the replication package, we obtained it by directly contacting the MOSPI. During this process, we identified an error in one of the data processing codes, which mistakenly dropped all plants in two key industries—74 and 75 of the AS-ICC 2-digit classification. After correcting this error, we obtained estimates that closely matched the published results. The numerical discrepancies between our reproduced estimates and the original results were negligible and did not affect the qualitative conclusions of the study.

Accurate estimation of the demand function is crucial to the methodology proposed by [Orr \(2022\)](#), necessitating carefully constructed instrumental variables (IVs). One such IV, which varies temporally within a 5-digit product code, exploits the average input price growth experienced by plants in other output markets using similar inputs. For this instrument to be valid, input price variations should stem from demand and supply shocks in other industries, orthogonal to machinery demand shocks or general machinery quality changes by product code. This validity hinges on the assumption that changes in average input prices are not driven by machinery demand—an assumption clearly violated in input markets where the ma-

chinery industry is the primary downstream consumer. To address this, [Orr \(2022\)](#) excludes observations with machinery cost shares exceeding 0.3. Our sensitivity analysis on this threshold choice reveals that the explanatory power of IVs is highly responsive to this cutoff. Notably, when the threshold is closer to 0 (e.g., 0.01), the resulting demand estimates imply an upward-sloping demand curve, preventing further progression in [Orr \(2022\)](#) methodological approach.

We also applied [Orr \(2022\)](#)'s approach to a sample of Indian plants from 2011 to 2020. As in the first replication, we encountered issues with the proposed IVs for demand estimation. We were unable to obtain reasonable demand estimates, resulting in a singular price elasticity matrix. Despite trying various threshold values, the issue persisted. In some specifications, we managed to obtain a downward-sloping demand curve; however, the other key demand parameter (i.e., the nesting parameter) fell outside the theoretically acceptable range.

2 Central Idea: The Use of First Order Conditions

In the literature, a common practice for recovering unit-level productivity is to estimate it as production function residuals. Consequently, estimating plant-product level productivity involves estimating the following production function, as shown in equation (18) in [Orr \(2022\)](#):

$$Y_{it}^j = \exp(\omega_{it}^j)(L_{it}^j)^{\beta_L}(K_{it}^j)^{\beta_K}(M_{it}^j)^{\beta_M}, \quad (1)$$

where Y_{it}^j represents plant i 's total output of product j in period t ; L_{it}^j , K_{it}^j , and M_{it}^j denote the labor, capital, and material inputs used by plant i to produce product j , respectively; and ω_{it}^j stands for plant i 's total factor productivity for producing product j (TFPQ).

The production function approach necessitates input uses across within-plant product lines, $(L_{it}^j, K_{it}^j, M_{it}^j)$, while typically only plant-level input uses, (L_{it}, K_{it}, M_{it}) , are observed in datasets on producers. The central idea of [Orr \(2022\)](#)'s methodology is to employ the first-order conditions of cost minimization to reveal these unobserved input allocations across within-plant product lines. As detailed in equation

(7) of [Orr \(2022\)](#):

$$S_{it}^j = \frac{MC_{it}^j Y_{it}^j}{\sum_{j \in \mathbb{Y}_{it}} MC_{it}^j Y_{it}^j}, \quad (2)$$

where S_{it}^j represents the share of input uses allocated to product j ; Y_{it}^j is plant i 's total output of product j in period t ; MC_{it}^j denotes plant i 's short-run marginal cost of producing j in period t ; and \mathbb{Y}_{it} is the set of products produced by plant i in period t .

Given that Y_{it}^j is available in the data, equation (2) allows the researcher to estimate the production function by product lines, thereby recovering plant-product level productivity, provided MC_{it}^j is known. To obtain MC_{it}^j , the researcher first estimates the demand system using plant-product level output and prices. If the price elasticity matrix corresponding to the estimated demand system is invertible, MC_{it}^j can be recovered under a specific market conduct assumption, such as static Bertrand-Nash competition.

In his empirical application, [Orr \(2022\)](#) employs the nested logit demand system, as illustrated in equation (20) of [Orr \(2022\)](#):

$$rs_{it}^j - rs_t^0 = (1 - \sigma)rs_{it}^{j|g(j)} - \alpha p_{it}^j + \eta_{it}^j, \quad (3)$$

where $rs_{it}^j = \ln \frac{R_{it}^j}{I_t^{h(j)}}$, with R_{it}^j representing the revenue from product j of plant i at time t , and $I_t^{h(j)}$ denoting the total revenue of the 3-digit ASICC code $h(j)$. Additionally, $rs_{it}^{j|g(j)} = \ln \frac{R_{it}^j}{\Lambda_t^{g(j)}}$, where $\Lambda_t^{g(j)}$ is the total revenue of the 5-digit ASICC code $g(j)$. The variable p_{it}^j represents the logged output price for product j , and η_{it}^j denotes product appeal. Under this specification, the price elasticity matrix is invertible if and only if $\alpha > 0$ and $0 < \sigma < 1$.

3 Computational Reproducibility

Using [the journal-provided replication package](#), we successfully completed the computational reproducibility of [Orr \(2022\)](#). The package includes the cleaning codes, enabling us to generate the estimates from the raw data. However, it does not include the raw or analysis datasets. To address this, we obtained the raw data for

Indian plants from 2000 to 2020 by directly contacting the Ministry of Statistics and Programme Implementation (MOSPI). For computational reproduction, we applied the provided codes to the 2000 to 2007 sample, as analyzed in [Orr \(2022\)](#).

The replication process revealed a challenge within the provided package, specifically concerning the retention of plant samples in industries 74 and 75:

- In `asicc_code_cleaning.do`, line 130 fails to retain plants from industries 74 and 75 when utilizing the supplied ASICC09 codes CSV file.

This issue stems from a discrepancy in the official MOSPI ASICC09 codes file. For industries lacking distinct three-digit codes within a two-digit classification—as is the case for industries 74 and 75—the file omits the three-digit code in the sub-division column. Consequently, line 130 of `asicc_code_cleaning.do` unintentionally excludes rows with missing sub-division values, effectively omitting plants from these industries.

To address this, we created a revised version of the do file, which corrects this issue and works seamlessly with the MOSPI-provided ASICC09 codes. Using this revised do file, we successfully reproduced all the published tables and figures in the main text of [Orr \(2022\)](#), with only minimal discrepancies in production function estimates that do not alter the qualitative implications¹.

In this report, we present the reproductions of key results:

- Table [1](#): Demand Estimates
- Table [2](#): Production Function Estimates
- Figure [1](#): Negative Correlation between TFPQ and Product Appeal
- Figure [2](#): Revenue Efficiency (TFPR) Growth after Removal of Worst Product

Our reproduction exercise yielded results that align precisely with the demand estimates of the original study. However, we observed minor discrepancies in the production function coefficients compared to the published estimates. In the baseline specification, for instance, our reproduced output elasticities for labor, capital,

¹We are grateful to Scott Orr for the explanation about the feature of the MOSPI ASICC09 code file and for providing us with a modified version of the ASICC09 codes that is compatible with the original `asicc_code_cleaning.do` in the replication package.

and material are 0.325, 0.106, and 0.789, respectively. These values differ slightly from the original estimates of 0.331, 0.101, and 0.790 (Columns 1 and 2 of Table 2).

We attribute these minor variations to potential inconsistencies in STATA's Mata function for numerical matrix inversion across different versions and operating systems. Such inconsistencies could lead to subtle differences in marginal cost calculations, subsequently affecting input allocation rules. Consequently, these small deviations in intermediate calculations result in slightly different production function estimates.

Nonetheless, the differences in production function estimates are small and do not significantly alter the key qualitative message. As displayed in Figure 1, we observed a negative correlation between TFPQ (ω_{it}^j) and product appeal (η_{it}^j) of -0.283 with a standard error of 0.151, compared to -0.282 (0.151) in Orr (2022). Our reproduction exercise confirmed the key result that removing the worst-performing product leads to a larger growth in TFPR than removing the second-worst product (white bars in Figure 2).

4 Robust Replication

Accurate estimation of the demand function is essential for implementing the approach outlined by Orr (2022), which relies on carefully selected instrumental variables (IVs) for p_{it}^j and $rs_{it}^{j|g(j)}$ in Equation 3. Orr (2022) estimates Equation (3) by instrumenting the two endogenous variables, p_{it}^j and $rs_{it}^{j|g(j)}$, using variations in input prices faced by plant i , denoted as $Z_t^{g(j)}$ and Z_{it}^{-jg} . For construction details, refer to Equations (21) and (22) in Orr (2022).

The proposed instruments are constructed using the script `generate_input_ins.do`. This script excludes observations where machinery cost shares exceed 0.3, aiming to mitigate endogeneity issues stemming from instrumental variable changes driven by machinery demand. Our primary task in this replication report is to assess the robustness of these instruments with respect to this threshold. This assessment involves a crucial trade-off. More stringent (i.e., lower) thresholds are likely to reduce endogeneity concerns but at the cost of dropping more observations, potentially weakening instrument strength. Conversely, less strict (i.e., higher) thresholds could

bolster instrument strength but risk incorporating machinery demand and supply shocks, potentially invalidating the instrument.

The second part of our replication involves applying Orr (2022)'s method to data from 2011 to 2020. Unlike the first part, this decision was made prior to examining the code and programs. In 2011, the item codes were updated from ASICC09 to NPCMS-2011. To ensure comparability with Orr (2022), we converted the product codes to ASICC09 using the ASICC09-NPCMS2011 concordance provided by the Indian MOSPI. After converting the product codes, we followed the same procedures and specifications to replicate the main results of Orr (2022). Since the replication package does not include trade flows, nominal exchange rates, and price indices for the recent period, we sourced the necessary datasets from UN Comtrade, the Federal Research Data Center, and the Indian KLEMS Database, respectively.

Summary statistics for the recent sample are presented in Tables A1 and A2, and Figure A1 in the Appendix, mirroring Tables 1 and 2, and Figure 3 from Orr (2022). These statistics for the recent sample are quite similar to those from the original study. However, there are two notable differences. First, plants in the Machinery, Equipment, and Parts industry have shifted towards producing a single product. Specifically, the proportion of multiproduct plants during the period from 2000 to 2007 was 0.26 (see the last row of Table 2 in Orr (2022)), but this figure has nearly halved to 0.14 in the recent period. Second, the distribution of the number of products produced by multiproduct plants is more right-skewed in 2011-2020 than it was in 2000-2007.

4.1 Robustness of Instrument Construction Threshold

We observed that the IV construction threshold significantly impacts the demand estimates. To explore this, we conducted replications using several threshold values ranging from 0.01 to 0.5. Notably, with a stringent threshold (i.e., a threshold of 0.01), the instruments did not perform as expected: the estimated coefficient for p_{it}^j was 0.151 with a standard error of 0.420, indicating an upward-sloping demand curve (see Column (1) of Table 3). Due to this ill-behaved demand function, we could not obtain MC_{it}^j , thus failing to compute input allocations across the within-

plant product lines (Equation (2)). Consequently, we were unable to estimate the plant-product level productivity.

When using less stringent threshold values, we obtained a downward-sloping demand curve (see Columns (2)-(5) of Table 3). However, our replication faced issues with imprecise inference. Although the price coefficient estimates were closer to the original, ranging from -0.179 to -0.261, the standard errors were higher, varying from 0.125 to 0.433. Specifically, with threshold values greater than 0.3, we were unable to obtain statistically significant demand estimates.

The variability in demand estimates is significantly influenced by the choice of threshold value for instrumenting input prices. Table 4 demonstrates that setting the IV construction threshold at or below 0.1 results in Sanderson-Windmeijer F-statistics falling well short of the widely accepted threshold of 10, as proposed by Staiger and Stock (1997). Moreover, elevating the threshold to 0.5 diminishes instrument strength, presumably due to increased conflation of demand and supply shocks. More promising results emerge from threshold values of 0.2 and 0.4, which lie closer to 0.3. As Table 3 illustrates, these thresholds yield Sanderson-Windmeijer F-statistics approaching 10, correlating with more precise demand estimates.

4.2 Using the 2011-2020 Period

We now turn our attention to applying Orr (2022)'s methodology to the ASI sample from 2011 to 2020. Despite using a threshold of 0.3 for constructing IVs, we were unable to obtain reasonable estimates for the demand function. Table 5 replicates Table 3 from Orr (2022). The IV estimates indicate that the demand function is upward-sloping, with a coefficient of 0.074. Furthermore, the coefficient for $rs_{it}^{j|g(j)}$ is not statistically significant.

We examined the performance of the IVs by analyzing the first-stage regression. The third column shows that $Z_t^{g(j)}$ performs as expected for instrumenting p_{it}^j : the coefficient is 0.820 with a standard error of 0.484, indicating that the cost shifter leads to higher prices. However, the other cost-shifting instrument, Z_{it}^{-jg} , performs poorly. Additionally, neither of these IVs successfully instruments $rs_{it}^{j|g(j)}$. As shown in Table 5, the estimated coefficients are statistically insignificant: 0.082 with a

standard error of 0.205 and -0.283 with a standard error of 0.294, respectively. Consistent with these findings, the Sanderson-Windmeijer first-stage F-statistics for weak instruments are 1.36 and 0.90 for p_{it}^j and $rs_{it}^{j|g(j)}$, respectively, which are well below the rule of thumb threshold of 10.

We also experimented with constructing IVs using different threshold values (see Table 6). However, the resulting IV demand estimates were consistently unreasonable. Although a negative coefficient for p_{it}^j was obtained with a threshold of 0.2, the coefficient for $rs_{it}^{j|g(j)}$ exceeded 1, violating the invertibility condition of the price elasticity matrix. Furthermore, as shown in Table 7, the IVs exhibited minimal explanatory power for both p_{it}^j and $rs_{it}^{j|g(j)}$, as the Sanderson-Windmeijer F-statistics remained far below 10 across all specifications, except for p_{it}^j when the threshold was set to 0.01. However, even in this case, the first-stage coefficients were estimated to be unreasonably negative.

Given that the price elasticity matrices for the demand estimates were not invertible, we were unable to estimate plant-product level productivity using the 2011 to 2020 sample. The first-stage estimates suggest that the diminished strength of the input prices in the later sample periods is likely the primary factor contributing to these findings.

5 Conclusion

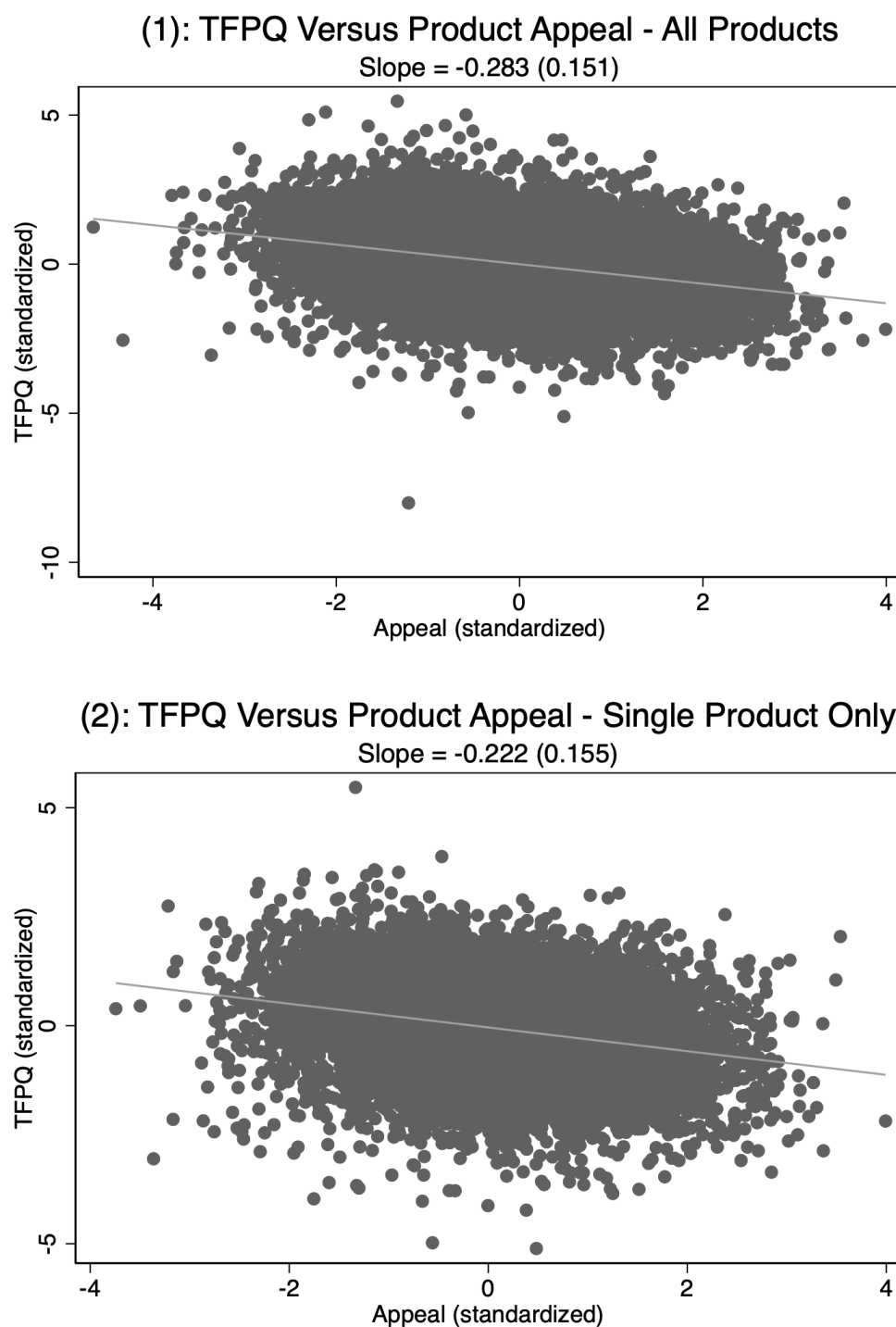
Overall, our findings indicate that the innovative method proposed by Orr (2022) hinges on the careful selection of IVs for accurately estimating the demand function. Without a robustly estimated demand function, proceeding to estimate the plant-level production function is not feasible. In this replication project, we found that the IVs used in the original study were sensitive to the choice of a threshold and did not perform as expected for the more recent sample periods. Future research could focus on identifying more reliable IVs to implement Orr (2022)'s novel methodology, making it applicable to alternative datasets.

References

- Brodeur, A., Mikola, D. and Cook, N.: 2024, Mass reproducibility and replicability: A new hope, *I4R Discussion Paper* **107**.
- Orr, S.: 2022, Within-firm productivity dispersion: Estimates and implications, *Journal of Political Economy* **130**(11), 2771–2828.
- Staiger, D. and Stock, J. H.: 1997, Instrumental variables regression with weak instruments, *Econometrica* **65**(3), 557–586.

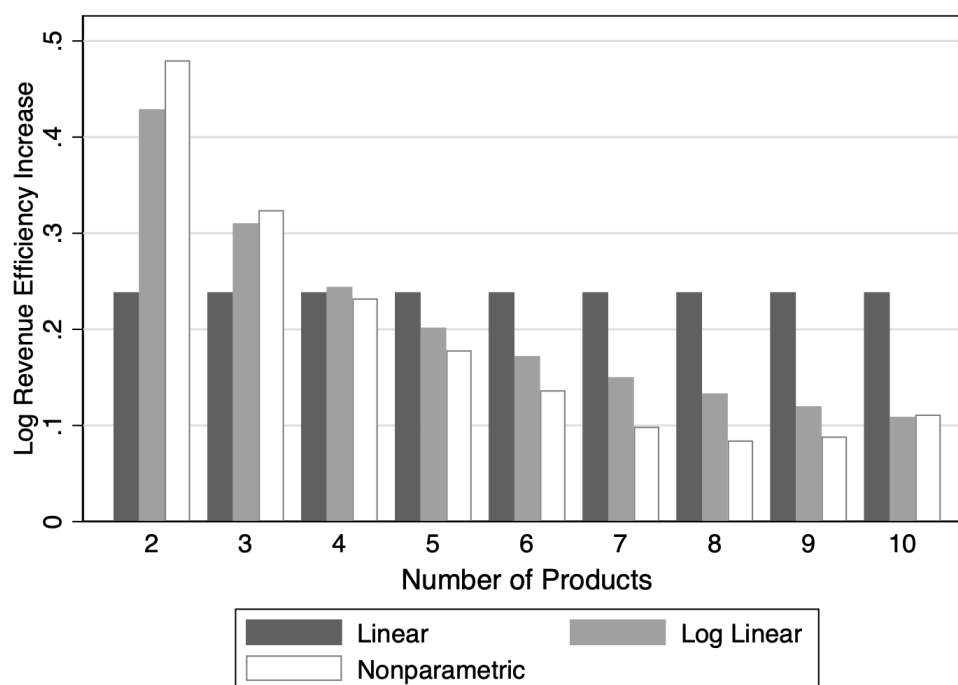
6 Figures

Figure 1: Computational Reproduction of Figure 1 in [Orr \(2022\)](#): TFPQ vs. Product Appeal



Notes: Calculations are based on the 2000–2007 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASIC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. TFPQ and demand appeal are standardized to have a mean of zero and a standard deviation of one. Plant-level block bootstrapped standard errors are presented in parentheses.

Figure 2: Computational Reproduction of Figure 2 in [Orr \(2022\)](#): TFPR Growth Following Removal of Least Efficient Product



Notes: Calculations are based on the 2000–2007 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). The y-axis represents the growth of log TFPR after the removal of a plant's least efficient product.

7 Tables

Table 1: Computational Reproduction of Column 2 in Table 3 of [Orr \(2022\)](#): Nested Logit Demand System

	IV Original Study	IV Reproduction
p_{it}^j	-0.220 (0.116)	-0.220 (0.116)
$rs_{it}^{j g(j)}$	0.621 (0.306)	0.621 (0.306)
Observations	64,917	64,917

Notes: Calculations are based on the 2000–2007 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. The dependent variable is the logarithm of the plant's product j revenue share relative to the total revenue generated by the 3-digit ASICC sector. Robust standard errors are presented in parentheses and are clustered by plant and product.

Table 2: Computational Reproduction of Columns 2, 3, and 4 in Table 4 of [Orr \(2022\)](#): Cobb-Douglas Production Function Estimates

	GMM					
	Original Study	Reproduction	Original Study	Reproduction	Original Study	Reproduction
β_L	0.331 (0.192)	0.325 (0.192)	0.321 (0.191)	0.315 (0.191)	0.626 (0.261)	0.617 (0.261)
β_K	0.101 (0.082)	0.106 (0.082)	0.097 (0.081)	0.102 (0.081)	0.236 (0.099)	0.239 (0.099)
β_M	0.790 (0.191)	0.789 (0.191)	0.806 (0.186)	0.806 (0.186)	0.217 (0.352)	0.223 (0.352)
ρ^{74}	0.757 (0.222)	0.757 (0.222)	0.747 (0.197)	0.747 (0.197)	0.842 (0.199)	0.841 (0.186)
ρ^{75}	0.657 (0.082)	0.658 (0.082)	0.661 (0.078)	0.661 (0.078)	0.670 (0.068)	0.670 (0.068)
ρ^{76}	0.651 (0.098)	0.652 (0.098)	0.653 (0.104)	0.653 (0.104)	0.623 (0.079)	0.623 (0.080)
ρ^{77}	0.420 (0.062)	0.420 (0.062)	0.422 (0.060)	0.422 (0.060)	0.541 (0.087)	0.540 (0.088)
ρ^{78}	0.194 (0.319)	0.195 (0.318)	0.181 (0.265)	0.182 (0.264)	0.569 (0.651)	0.568 (0.658)
RTS	1.222 (0.084)	1.220 (0.084)	1.224 (0.080)	1.223 (0.080)	1.078 (0.113)	1.078 (0.112)
Instruments						
$(Z_t^{g(j)}, Z_{t-1}^{g(j)})$	✓	✓			✓	✓
m_{it-1}	✓	✓	✓	✓		
Observations	3,620	3,620	3,620	3,620	3,620	3,620

Notes: Calculations are based on the 2000–2007 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. The dependent variable is the log quantity of product j produced by plant i . Plant-level block bootstrapped standard errors are presented in parentheses.

Table 3: Nested Demand Estimates: Impact of Varying IV Construction Thresholds (Original Study = 0.3)

	IV				
	Replication (1)	Replication (2)	Replication (3)	Replication (4)	Replication (5)
p_{it}^j	0.151 (0.420) [0.599]	-0.245 (0.180) [0.192]	-0.261 (0.125) [0.048]	-0.179 (0.137) [0.192]	-0.249 (0.433) [0.596]
$rs_{it}^{j g(j)}$	0.816 (0.619) [0.229]	0.530 (0.505) [0.301]	0.634 (0.342) [0.086]	0.539 (0.331) [0.131]	0.088 (0.370) [0.780]
Threshold	0.01	0.1	0.2	0.4	0.5
Observations	36,447	57,096	63,325	66,283	67,094

Notes: Calculations are based on the 2000–2007 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. The dependent variable is the log of the plant's product j revenue share relative to the total revenue generated by the 3-digit ASICC sector. The original study uses a threshold of 0.3 to construct instruments. Robust standard errors, clustered by plant and product, are presented in parentheses. P-values are presented in brackets.

Table 4: First Stage Estimates: Impact of Varying IV Construction Thresholds (Original Study = 0.3)

	Replication (1)		Replication (2)		Replication (3)		Replication (4)		Replication (5)	
	p_{it}^j	$rs_{it}^{j g(j)}$	p_{it}^j	$rs_{it}^{j g(j)}$	p_{it}^j	$rs_{it}^{j g(j)}$	p_{it}^j	$rs_{it}^{j g(j)}$	p_{it}^j	$rs_{it}^{j g(j)}$
$Z_t^{g(j)}$	-0.353 (1.061) [0.748]	1.067 (0.519) [0.043]	0.258 (0.102) [0.012]	0.146 (0.042) [0.020]	0.279 (0.106) [0.011]	0.164 (0.047) [0.013]	0.221 (0.080) [0.017]	0.110 (0.041) [0.032]	0.033 (0.099) [0.740]	0.096 (0.033) [0.007]
Z_{it}^{-jg}	2.450 (0.848) [0.005]	0.640 (0.886) [0.494]	0.302 (0.244) [0.248]	-0.077 (0.111) [0.483]	0.347 (0.221) [0.136]	-0.168 (0.118) [0.304]	0.307 (0.187) [0.087]	-0.146 (0.106) [0.355]	0.169 (0.132) [0.259]	-0.082 (0.084) [0.341]
Sanderson-Windmeijer F-stat	5.67	3.75	2.49	5.30	9.95	10.27	10.23	8.28	1.61	9.00
Threshold	0.01		0.1		0.2		0.4		0.5	
Observations	36,447		57,096		63,325		66,283		67,094	

Notes: Calculations are based on the 2000–2007 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. The dependent variables are the logged price of the plant's product j and the log revenue share of the plant's product j within the 5-digit ASICC industry. The original study uses a threshold of 0.3 to construct instruments. Robust standard errors, clustered by plant and product, are presented in parentheses. P-values are presented in brackets. The Sanderson-Windmeijer first-stage F-statistic for weak instruments is evaluated using the rule of thumb proposed by [Staiger and Stock \(1997\)](#), which suggests a cutoff of 10.

Table 5: Replication of Table 3 in [Orr \(2022\)](#): Nested Logit Demand 2011-2020

	OLS	IV	p_{it}^j	$rs_{it}^{j g(j)}$
p_{it}^j	0.008 (0.002) [0.000]	0.074 (0.229) [0.756]		
$rs_{it}^{j g(j)}$	0.943 (0.005) [0.000]	0.642 (0.447) [0.882]		
First Stage				
$Z_t^{g(j)}$			0.820 (0.484) [0.088]	0.082 (0.205) [0.663]
Z_{it}^{-jg}			-0.050 (0.358) [0.903]	-0.283 (0.294) [0.235]
Sanderson-Windmeijer F-stat			1.36	0.90
Observations	85,357	85,357	85,357	85,357

Notes: Calculations are based on the 2011-2020 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. The dependent variable is the log of the plant's product j revenue share relative to the total revenue generated by the 3-digit ASICC sector. ASICC codes are recovered using the concordance between ASICC09 and NPCMS-2011 provided by MOSPI. Robust standard errors, clustered by plant and product, are presented in parentheses. P-values are presented in brackets. The Sanderson-Windmeijer first-stage F-statistic for weak instruments is evaluated using the rule of thumb proposed by [Staiger and Stock \(1997\)](#), which suggests a cutoff of 10.

Table 6: Nested Logit Demand Estimates 2011-2020: Changing IV Construction Threshold

	IV				
	Replication (1)	Replication (2)	Replication (3)	Replication (4)	Replication (5)
p_{it}^j	0.265 (0.483) [0.583]	5.439 (67.983) [0.935]	-0.029 (0.276) [0.916]	0.112 (0.212) [0.367]	0.193 (0.612) [0.857]
$rs_{it}^{j g(j)}$	1.057 (1.124) [0.347]	-4.248 (60.223) [0.945]	1.431 (0.834) [0.088]	0.049 (3.467) [0.990]	0.264 (2.350) [0.908]
Threshold	0.01	0.1	0.2	0.4	0.5
Observations	62,927	81,464	84,496	85,658	85,967

Notes: Calculations are based on the 2011–2020 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. The dependent variable is the logarithm of the plant's product j revenue share relative to the total revenue generated by the 3-digit ASICC sector. The original study uses a threshold of 0.3 to construct instruments. Robust standard errors, clustered by plant and product, are presented in parentheses. P-values are presented in brackets.

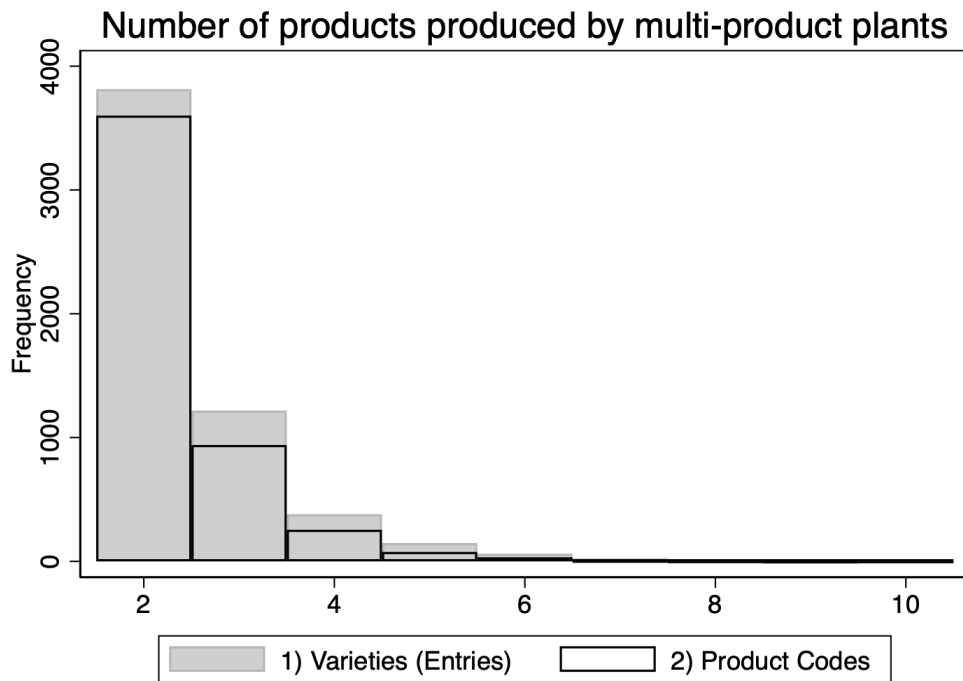
Table 7: First Stage Estimates 2011-2020: Changing IV Construction Thresholds

	Replication (1)		Replication (2)		Replication (3)		Replication (4)		Replication (5)	
	p_{it}^j	$r s_{it}^{j g(j)}$	p_{it}^j	$r s_{it}^{j g(j)}$	p_{it}^j	$r s_{it}^{j g(j)}$	p_{it}^j	$r s_{it}^{j g(j)}$	p_{it}^j	$r s_{it}^{j g(j)}$
$Z_t^{g(j)}$	-6.473 (2.792) [0.019]	0.056 (1.808) [0.949]	0.196 (0.751) [0.794]	0.295 (0.316) [0.321]	1.271 (0.876) [0.134]	-0.025 (0.292) [0.928]	0.289 (0.470) [0.529]	-0.015 (0.142) [0.898]	0.794 (0.443) [0.087]	0.037 (0.127) [0.746]
Z_{it}^{-jg}	-2.419 (2.833) [0.416]	-1.167 (2.287) [0.609]	1.190 (1.277) [0.359]	1.308 (1.087) [0.239]	0.806 (0.753) [0.270]	0.448 (0.567) [0.518]	-0.011 (0.278) [0.963]	-0.073 (0.279) [0.903]	0.141 (0.209) [0.505]	-0.059 (0.237) [0.680]
Sanderson-Windmeijer F-stat	13.35	0.33	0.01	0.01	1.32	0.64	0.36	0.07	0.23	0.08
Threshold	0.01		0.1		0.2		0.4		0.5	
Observations	62,927		81,464		84,496		85,658		85,967	

Notes: Calculations are based on the 2011–2020 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. The dependent variable is the logarithm of the plant's product j revenue share relative to the total revenue generated by the 3-digit ASICC sector. The original study uses a threshold of 0.3 to construct instruments. Robust standard errors, clustered by plant and product, are presented in parentheses. P-values are presented in brackets. The Sanderson-Windmeijer first-stage F-statistic for weak instruments is evaluated using the rule of thumb proposed by [Staiger and Stock \(1997\)](#), which suggests a cutoff of 10.

8 Appendix Figures

Figure A1: Product Counts: ASI Entries versus Number of Product Codes



Notes: Calculations are based on the 2011-2020 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant.

9 Appendix Tables

Table A1: Plant-Product-Year Summary Statistics for Machinery, Equipment, and Parts (86,543 Observations)

Variable	Mean	Std. Dev.	Min	Max	Median
Log Revenue: r_{it}^j	17.86	2.5	0	26.45	17.99
Log Quantity Sold: q_{it}^j	9.38	4.06	-4.61	22.51	9.41
Log Prices: p_{it}^j	8.48	3.32	-4.12	22.35	8.28
Log Quantity Produced: y_{it}^j	9.34	3.97	-4.61	21.95	9.36
Multiproduct	.58	.49	0	1	1
Single Industry	.57	.49	0	1	1
Vertical Integration	.15	.36	0	1	0

Notes: Calculations are based on the 2011-2020 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant and a 5-digit ASICC variety. The sample is restricted to producers within the machinery, equipment, and parts industry. All variables are constructed following the methodology of [Orr \(2022\)](#).

Table A2: Plant-Year Summary Statistics for Machinery, Equipment, and Parts (41,322 Observations)

Variable	Mean	Std. Dev.	Min	Max	Median
Log Labor: l_{it}	9.43	1.66	2.3	15.64	9.37
Log Capital Stock: k_{it}	16.06	2.43	-.7	24.63	16.07
Log Materials: m_{it}	7.7	3.31	-3.22	21.49	7.33
No. of varieties: J_{it}	1.21	.65	1	10	1
Multiproduct	.14	.34	0	1	0

Notes: Calculations are based on the 2011-2020 Indian ASI dataset, provided by the Ministry of Statistics and Programme Implementation (MOSPI). Each observation represents a plant. The sample is restricted to producers within the machinery, equipment, and parts industry. All variables are constructed following the methodology of [Orr \(2022\)](#).