Essays on Firm Heterogeneity in Macroeconomics

by

Dimitrije Ruzic

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2018

Doctoral Committee:

Professor Andrei A. Levchenko, Co-Chair
Professor Matthew D. Shapiro, Co-Chair
Assistant Professor Joshua K. Hausman
Professor Stefan Nagel, University of Chicago

druzic@umich.edu

ORCID: 0000-0002-9879-9740

For my parents, Ivan Ružić and Milica Mitić-Ružić.

ACKNOWLEDGEMENTS

My dissertation committee's advice, enthusiasm, and support are core building blocks of this dissertation. Andrei Levchenko and Matthew Shapiro patiently and tirelessly guided me as I honed my research interests. Their emphasis on combining economic theory with careful data work has not only improved this dissertation, but has also given me a model to follow in pursuing research more broadly. I am also indebted to Joshua Hausman and Stefan Nagel for pushing me to think about and contextualize my research more broadly. Their feedback at all stages of the research process proved invaluable.

The collaborations that underlie this dissertation have been deeply rewarding, both professionally and personally. Chapter 1 began as a conversation with Sui-Jade Ho in the windowless basement that houses the Michigan Census Research Data Center. Her curiosity, diligence and kindness turned the RDC into an exciting workplace and sustained this rewarding collaboration even while we worked continents apart. My involvement in the project that resulted in Chapter 2 began over a lunch with Lin Ma. His sharp insights and boundless optimism made the research process fun and dynamic, especially on late-night Skype calls. Working with Rudi Bachmann and Gabe Ehrlich on Chapter 3 taught me immensely about research and provided me with wonderful mentors and friends.

My time in Ann Arbor has been replete with wonderful people who have in many different ways contributed to this dissertation. The Research Seminar in Quantitative Economics provided me a home in the economics department, allowed me to work alongside excellent colleagues, and generously supported my research. I learned an immense amount from my fellow graduate students and am especially indebted to Alberto Arredondo Chavez, Chris Boehm, Laurien Gilbert, Steve Hou, Yeliz Kacamak, Salma Khalid, Ben Lipsius, Nicolas Morales, Nitya Pandalai Nayar, Daniel Schaffa, Christina Synn, and Hope Thompson. I am particularly grateful for the non-economists in my Ann Arbor life, especially Drew Kozulniski, Tejas Navaratna, Steve Robinson, Rachel Sun, and Yining Zhang.

The staff at the University of Michigan Department of Economics have done a fantastic job of managing the Ph.D. program. I would especially like to thank Laura Flak and Julie Heinz for their help and enthusiasm during the rough patches of doctoral-student life.

Various iterations of the works included in this dissertation were presented at the University of Michigan, INSEAD, the Board of Governors of the Federal Reserve, the Federal Reserve Bank of Cleveland, HEC Montreal, Queens College, Williams College, and the U.S. Department of Treasury. I thank the audiences for insightful questions and comments.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

The three essays in this dissertation examine how differences across firms shape macroeconomic outcomes. These essays combine models of heterogeneous firms with detailed micro data to answer long-running questions about the U.S. economy: how well are resources allocated across firms? Does globalization drive income inequality? How large are the economic spillovers of firm reputation? Each essay highlights a different aspect of firm or industry heterogeneity that meaningfully changes the answer to the question at hand.

Chapter 1, from a work with Sui-Jade Ho, quantifies misallocation in the U.S. manufacturing sector using a structural model and restricted microdata from the U.S. Census. The estimated misallocation—the distance between aggregate productivity and a frontier where marginal products are equalized—declined 13 percent between 1982 and 2007. Key features of the model and estimation are markups and returns to scale that vary across industries and time. Strikingly, imposing a common markup and constant returns to scale, as is commonly done in the literature, implies that misallocation increased 29 percent over the same period. The essay rationalizes these discrepancies and shows that they arise from overlooking variation in markups and returns to scale.

Chapter 2, from a work with Lin Ma, provides causal evidence that access to global markets leads to a higher executive-to worker pay ratio within the firm. Specifically, the essay uses China's 2001 accession to the World Trade Organization as a trade shock to show that firms that exported to China prior to 2001 subsequently exported more, grew larger, and grew more unequal in terms of executive-to-worker pay. Counterfactual exercises in the accompanying structural model suggest that trade and FDI liberalizations can explain around 52 percent of the surge in top 0.1 percent income shares in the U.S. data between 1988 and 2008.

Chapter 3, joint with Ruediger Bachmann and Gabriel Ehrlich, uses the 2015 Volkswagen emissions scandal as a natural experiment to provide causal evidence that group reputation externalities matter for firms. The essay estimates that the scandal reduced the U.S. sales of non-Volkswagen German auto manufacturers by approximately 76,000 vehicles over the following year, leading to a loss of approximately $3.7 billion of revenue. These declines in sales accompanied declines in stock returns and in social-media sentiment to-

ward these firms. Volkswagen's malfeasance materially harmed the group reputation of "German car engineering" in the United States.

CHAPTER I

Returns to Scale, Productivity Measurement, and Trends in U.S. Manufacturing Misallocation

From a work with Sui-Jade Ho

Abstract

Aggregate productivity suffers when workers and machines are not matched with their most productive uses. This paper builds a model that features industry-specific markups, industry-specific returns to scale, and establishment-specific distortions, and uses it to measure the extent of this misallocation in the economy. Applying the model to restricted U.S. census microdata on the manufacturing sector suggests that misallocation declined by 13% between 1982 and 2007. The jointly estimated markup and returns to scale parameters vary substantially across industries. Furthermore, while the average markup has been relatively constant, the average returns to scale declined over this period. The finding of declining misallocation starkly contrasts the 29% increase implied by the widely used Hsieh and Klenow (2009) model, which assumes that all establishments charge the same markup and have constant returns to scale. Accounting for the variation in markups and returns to scale leads to the divergence of misallocation estimates in this paper from those implied by the Hsieh-Klenow model.

JEL Codes: D24, E23, E25, O47
Keywords:returns to scale, productivity, misallocation, manufacturing

1.1 Introduction

Aggregate productivity retreats from its frontier when workers and machines mismatch with their most productive uses. Formalized elegantly by Restuccia and Rogerson (2008), this notion of misallocation has the potential to explain why countries differ in their incomes, or why aggregate productivity changes over time. Yet, quantifying the extent of this misallocation is challenging, in part because we do not observe productivity directly. Most commonly, we must infer an establishment's total factor productivity from its revenue. This inference is a two-step process: first, we must deduce how establishments set prices, so we can map revenue to output; then, we must deduce how they produce, so we can map output to productivity.

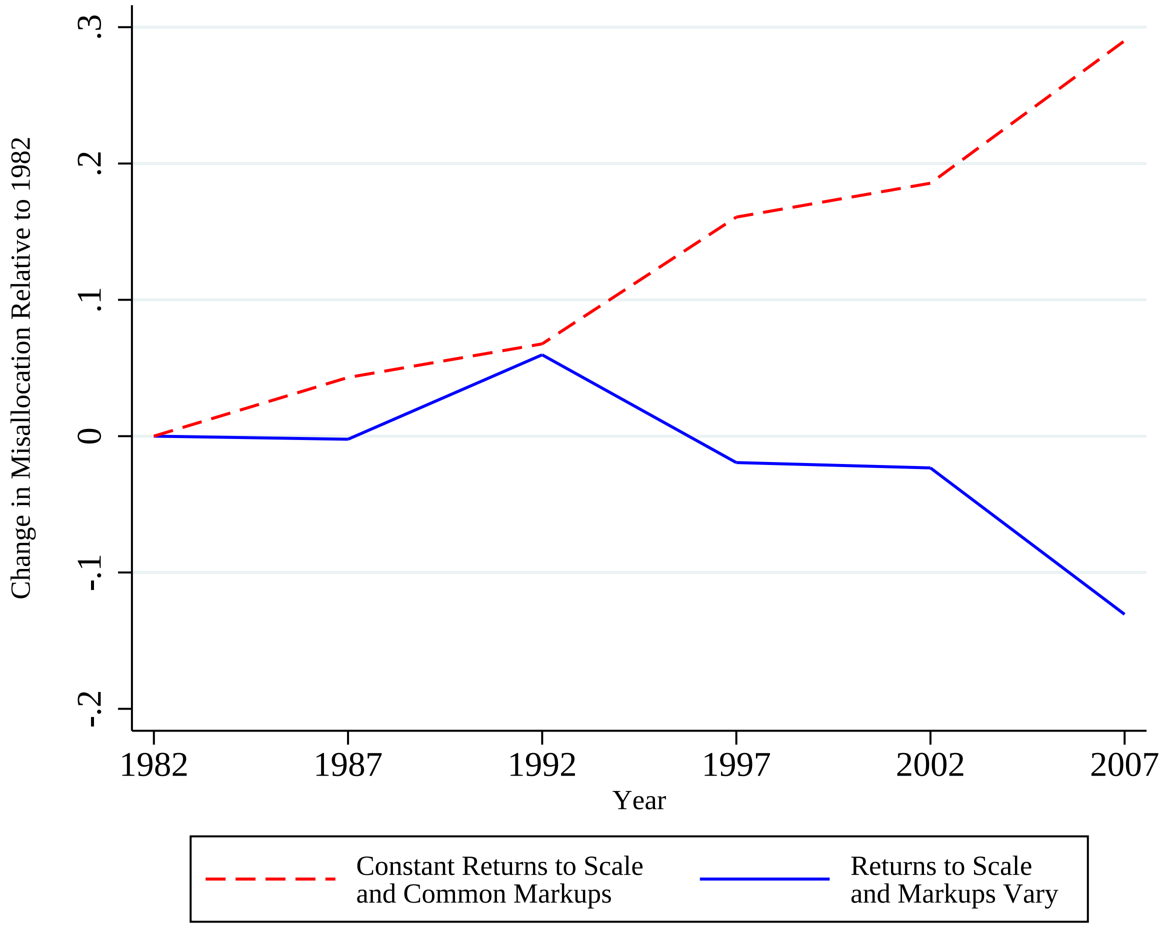
To infer productivity and measure misallocation, this paper builds a quantitative model in which markups of price over marginal cost and returns to scale differ across industries and time. We implement the model on restricted U.S. Census microdata covering the U.S. manufacturing sector from 1977 through 2007. In the process, we jointly estimate markups and returns to scale for individual industries within U.S. manufacturing. Our estimates show that industries differ meaningfully in both markups and returns to scale, with standard deviations across industries of about one-third the level of the respective parameters. Moreover, while the average markup remained relatively constant over this period, the average returns to scale fell, starting off as increasing and ending as nearly constant. We use these parameters to infer productivity, and find that misallocation in U.S. manufacturing declined 13% between 1982 and 2007.

Allowing for heterogeneous markups and returns to scale is crucial when estimating productivity and misallocation. The widely-used Hsieh and Klenow (2009) model is a special case of our framework in which all industries have a common markup and constant returns to scale. Figure 1 contrasts the downward trend in misallocation under our estimated parameters with the upward trend implied by the Hsieh-Klenow assumptions. Both measures of misallocation answer the question: how much more productive would the U.S. manufacturing sector be if it were as misallocated today as it was in 1982? If misallocation by this measure has increased, productivity today would be higher at 1982 levels of misallocation. Indeed, as the dashed red line shows, the assumptions of a common markup and constant returns to scale suggest a 29% increase in misallocation over the last 25 years. By contrast, the solid blue line traces out the declining trend in misallocation from our model.

We arrive at the declining trend in misallocation by estimating markups and returns to scale using a control-function approach rooted in Olley and Pakes (1996) and Levinsohn

Figure 1.1: Misallocation in U.S. Manufacturing

Change in U.S. Manufacturing TFP at 1982 Levels of Misallocation



Note: Misallocation is the distance between aggregate productivity and a frontier where marginal revenue products are equalized across establishments in each industry. Positive (negative) values indicate an increase (decrease) in misallocation relative to 1982.

and Petrin (2003). Our estimating procedure infers markups and returns to scale even in datasets, like the U.S. Census microdata, where we observe revenues, but not output or prices. For this procedure, we derive a model-based estimating equation that relates establishment revenue to its inputs and to industry size, as in De Loecker (2011). We map the reduced-form revenue elasticities to the markup and returns-to-scale parameters using model equations. In line with prior empirical work [e.g. Hall (1990), Basu and Fernald (1997), Basu, Fernald and Kimball (2006), Broda and Weinstein (2006)], we find that both markups and returns to scale indeed vary across industries. Moreover, the average markup for U.S. manufacturing has remained relatively constant over time, while returns to scale have declined, starting off as increasing in 1982 and ending as nearly constant by 2007.

We show that the decline in returns to scale is the key to rationalizing the different trends in misallocation between our model and the Hsieh-Klenow model. In short, ignoring the variation in markups and returns to scale leads to measures of productivity that conflate productivity and distortion. These conflated measures of productivity lead to incorrect inferences about the extent to which the most productive establishments bear the largest distortions, and hence lead to incorrect measures of misallocation. Our estimates suggest that the Hsieh-Klenow model understates misallocation on average. Over time, as the assumption of constant returns better fits the data for the U.S. manufacturing sector, the Hsieh-Klenow model understates misallocation less and less. This better fit drives the upward trend in misallocation under a common markup and constant returns.

Outside their relevance for measuring productivity and misallocation, the patterns we document for markups and returns to scale also fit with the recent literature on the decline of the labor share, and, more broadly, the changing division of value added. For instance, a large literature documents a thirty-year decline in labor's share of value added both for the United States and for other economies [e.g Elsby, Hobijn and Sahin (2013), Karabarbounis and Neiman (2014), Barkai (2016)]; we find this decline to be even larger for the U.S. manufacturing sector. Within that literature, using different approaches, both Karabarbounis and Neiman (2014) and Barkai (2016) suggest that the decline in labor's share of value added might not have been offset by an equivalent increase in the capital share. The resulting implication is that the share of profits in value added increased over time. Indeed, De Loecker and Eeckhout (2017) find evidence of rising profit rates both among U.S. publicly traded firms and in the national income accounts. For the U.S. manufacturing sector, we find that the rising profit share has been driven primarily by the decline in the returns to scale.

After presenting the main results, we generalize our model further, in the spirit of Atkeson and Burstein (2008), and introduce markups that vary across establishments in an industry; this added generality supplements the work of a growing literature that continues to refine the measurement of distortions (see Hopenhayn (2014) for a review). We show that, conditional on an industry-specific demand elasticity, the additional variation in markups changes the marginal revenue products of establishments and hence their measured distortions. Taking this additional feature to the data, we find that the levels of misallocation are lower, and that the divergent trends in misallocation between out model and the Hsieh-Klenow model persist. In changing the level of measured misallocation, the generalization to variable markups within an industry is similar to other work where richer depictions of establishment behavior reduce the level of measured misallocation [e.g. Bartelsman, Haltiwanger and Scarpetta (2013), Asker, Collard-Wexler and De Loecker (2014), Gopinath et al. (2017)], and those who emphasize measurement issues that make inferring misallocation challenging [e.g. White, Reiter and Petrin (2016), Bils, Klenow and Ruane (2017), Haltiwanger, Kulick and Syverson (2018)].

Within the recent literature on misallocation, our paper's closest counterparts are two works that emphasize the importance of measurement within the Hsieh-Klenow model: Bils, Klenow and Ruane (2017) and Haltiwanger, Kulick and Syverson (2018). The former explains the upward trend in U.S. manufacturing misallocation as an artefact of measurement error that increased over time. While we think measurement error is an important topic to address in the microdata, we show in Appendix A.4 that the Bils, Klenow and Ruane (2017) procedure risks conflating measurement error with model misspecification if returns to scale are not constant: ignoring a decline in returns to scale, like the one we document, could lead an econometrician to infer an increase in measurement error. The latter paper, Haltiwanger, Kulick and Syverson (2018), uses eleven manufacturing products to show that deviations from production and demand assumptions in the Hsieh-Klenow model lead to estimates of establishment-level distortions that behave differently than the distortions in the baseline model. We share their emphasis on deviations from standard Hsieh-Klenow assumption and view the works as complementary.

The remainder of the paper proceeds as follows. In section 2 we derive a measure of misallocation in a model that allows for variation in markups and returns to scale; we then develop a toolkit to understand the discrepancies in measured productivity and misallocation that arise from ignoring the variation in these parameters. We map the model to the data, detail the estimation procedure, and present the estimates of markups and returns to scale in section 3. Section 4 presents our measure of misallocation and uses the toolkit to explain why our measure deviates from the Hsieh-Klenow measure that assumes a common markup and constant returns. Section 5 highlights the robust difference between the trends in misallocation in the two models by showing that this difference persists for six meaningful changes in model structure and estimation. Section 6 concludes.

1.2 Model

We build a model that features industry-specific markups, industry-specific returns to scale, and establishment-specific distortions. We then show how ignoring the variation in markups and returns to scale leads to measures of productivity that conflate productivity and distortions, and leads to incorrect measures of misallocation.

1.2.1 Deriving a Measure of Misallocation

In this section, we derive a measure of misallocation for the aggregate economy, accounting for industry variation in markups and returns to scale. We measure misallocation as the distance between aggregate productivity and a frontier where inputs are reallocated so that marginal revenue products are equal across establishments in each industry. We proceed in three steps. First, we show the aggregation in the model, allowing us to map from the distortions that establishments face to aggregate misallocation. Second, we show how establishments optimally respond to the distortions they face; these expressions allow us to characterize establishment behavior when we reallocate resources and change the distortions that they face. Third, we derive a measure of misallocation by comparing aggregate productivity before and after resources are reallocated. Since much of this derivation is standard in the literature, here we highlight the structure of the model and the key inputs into the measure of misallocation. We refer interested readers to appendix A for more details.

Aggregation

A representative firm aggregates the output Y_i of I different industries using a Cobb-Douglas production technology, and sells the aggregate output Y in a perfectly-competitive market, as in (1.1):

$$\text{Aggregate} \qquad Y = \prod_{i=1}^{I} Y_i^{\theta_i} \text{ with } \sum_{i=1}^{I} \theta_i = 1 \qquad\qquad P = \prod_{i=1}^{I} (P_i/\theta_i)^{\theta_i} = 1. \tag{1.1}$$

Cost minimization by this aggregating firm implies that the elasticities θ_i from the production function correspond to the share of each industry's value added (P_iY_i) in aggregate value added (PY). This insight allows us to define the aggregate price index P, which we choose as the numeraire.

Within each industry, an aggregating firm combines the output Y_{ie} of N_i differentiated establishments using a constant-elasticity-of-substitution (CES) technology, as in (1.2):

$$\text{Industry} \qquad Y_i = \left(\sum_{e=1}^{N_i} Y_{ie}^{\frac{\sigma_i - 1}{\sigma_i}} \right)^{\frac{\sigma_i}{\sigma_i - 1}} \qquad\qquad P_i = \left[\sum_{e=1}^{N_i} \left(\frac{1}{P_{ie}} \right)^{\sigma_i - 1} \right]^{\frac{-1}{\sigma_i - 1}}. \tag{1.2}$$

The CES aggregator implies that each establishment in the industry faces a downward-sloping demand curve for its output. Cost minimization by the industry aggregating firm leads to the standard CES price index P_i. Note that that the elasticity σ_i can potentially vary across industries.

Each establishment in the industry produces value-added output Y_{ie} by combining its total factor productivity A_{ie}, capital K_{ie}, and labor L_{ie} using the Cobb-Douglas production function in equation (1.3):

$$\text{Establishment} \qquad Y_{ie} = A_{ie} K_{ie}^{\alpha_{K_i}} L_{ie}^{\alpha_{L_i}}, \quad \alpha_i = \alpha_{K_i} + \alpha_{L_i} \,. \tag{1.3}$$

The returns to scale in production are α_i, the sum of output elasticities α_{K_i} and α_{L_i}; when returns to scale differ from unity, we have non-constant returns to scale. Moreover, returns to scale can differ across industries.

Optimization

Each establishment maximizes profits π_{ie} by choosing how much capital and labor to hire:

$$\pi_{ie} = P_{ie} Y_{ie} - (1 + \tau_{L_{ie}}) w_i L_{ie} - (1 + \tau_{K_{ie}}) R_i K_{ie}. \tag{1.4}$$

The establishment takes as given the input prices R_i and w_i from perfectly competitive input markets; however, the effective cost of an input varies across establishments, with $\tau_{K_{ie}}$ and $\tau_{L_{ie}}$ capturing the input-specific distortions for capital and labor. Consider, for instance, regulations that mandate the benefits that establishments have to provide to workers. These regulations change the effective cost of hiring labor. If two establishments are subject to different regulations, then these establishments also differ in their $\tau_{L_{ie}}$.

Establishments that face large distortions have high marginal revenue products. The first-order conditions from profit maximization, shown in equation (1.5) for capital and equation (1.6) for labor,

$$MRPK_{ie} = \frac{\alpha_{K_i}}{\frac{\sigma_i}{\sigma_i - 1}} \frac{P_{ie} Y_{ie}}{K_{ie}} = (1 + \tau_{K_{ie}}) R_i \tag{1.5}$$

$$MRPL_{ie} = \frac{\alpha_{L_i}}{\frac{\sigma_i}{\sigma_i - 1}} \frac{P_{ie} Y_{ie}}{L_{ie}} = (1 + \tau_{L_{ie}}) w_i, \tag{1.6}$$

show that establishments trade off the marginal contribution to revenue of a given input ($MRPK_{ie}$ or $MRPL_{ie}$) against the effective cost of hiring it. For instance, an establishment facing a cost-increasing labor regulation has a large $\tau_{L_{ie}}$; this establishment will hire labor until the contribution to revenue of the last unit hired, $MRPL_{ie}$, exactly offsets the effective cost of hiring labor $(1 + \tau_{L_{ie}}) w_i$. In short, faced with larger distortion, the establishment requires larger marginal revenue products to justify hiring inputs. Moreover, in the absence of distortions, marginal revenue products are equalized across establishments in an industry. This notion will help us define a productivity frontier and subsequently misallocation.

Optimal responses to larger distortions lead establishments to charge higher prices. The establishment price in equation (1.7) is a markup over marginal cost:

$$P_{ie} = \underbrace{\frac{\sigma_i}{\sigma_i - 1}}_{Markup} \underbrace{\left[\left(\frac{R_i}{\alpha_{K_i}} \right)^{\alpha_{K_i}} \left(\frac{w_i}{\alpha_{L_i}} \right)^{\alpha_{L_i}} \right]^{\frac{1}{\alpha_i}} \left(Y_{ie} \right)^{\frac{1-\alpha_i}{\alpha_i}} \left[\frac{(1+\tau_{K_{ie}})^{\alpha_{K_i}} (1+\tau_{L_{ie}})^{\alpha_{L_i}}}{A_{ie}} \right]^{\frac{1}{\alpha_i}}}_{Marginal\ Cost}. \tag{1.7}$$

The model allows the markup $\sigma_i/(\sigma_i - 1)$ in equation (1.7) to be industry specific. Furthermore, the introduction of potentially non-constant returns to scale allows the marginal cost to change with the establishment's scale of production. Under the standard assumption of constant returns to scale ($\alpha_i = 1$), marginal cost is constant and independent of output Y_{ie}. However, if returns to scale deviate from unity ($\alpha_i \neq 1$), then marginal cost is increasing in output for decreasing returns to scale, and vice versa. Lastly, larger distortions increase the marginal cost of production and thus force the establishment to charge a higher price.[1]

An establishment responds to large distortions by choosing a smaller input bundle and shrinking in size. Since much of this paper is about the allocation of resources across establishments in an industry, the relevant measure of size captures the establishment's value added relative to the value added of the industry, s_{ie} in equation (1.8):

$$s_{ie} = \frac{P_{ie} Y_{ie}}{P_i Y_i} = \frac{\left[A_{ie} \left(\frac{\overline{1+\tau_{K,i}}}{1+\tau_{K_{ie}}} \right)^{\alpha_{K_i}} \left(\frac{\overline{1+\tau_{L,i}}}{1+\tau_{L_{ie}}} \right)^{\alpha_{L_i}} \right]^{\frac{1}{\frac{\sigma_i}{\sigma_i - 1} - \alpha_i}}}{\sum_{e=1}^{N_i} \left[A_{ie} \left(\frac{\overline{1+\tau_{K,i}}}{1+\tau_{K_{ie}}} \right)^{\alpha_{K_i}} \left(\frac{\overline{1+\tau_{L,i}}}{1+\tau_{L_{ie}}} \right)^{\alpha_{L_i}} \right]^{\frac{1}{\frac{\sigma_i}{\sigma_i - 1} - \alpha_i}}}. \tag{1.8}$$

For instance, if the labor distortion faced by the establishment $(1+\tau_{L_{ie}})$ increases relative to the average labor distortion in the industry $\overline{(1+\tau_{L_i})}$, the establishment declines in size.

We can also see from equation (1.8) that the size of the establishment after we reallocate resources depends solely on its productivity A_{ie}. From the earlier first-order conditions, we know that equalizing marginal products is akin to equalizing distortions. The reallocation of resources would then eliminate the relative distortions in equation (1.8), and the counterfactual size of the $s_{ie}|_{\tau=\overline{\tau}}$ would be strictly increasing in productivity A_{ie}.

[1] Formally, this statement is based on a rewriting of the price that eliminates the output term Y_{ie}; we show that $P_{ie} = \Omega_{P_i} \left[\frac{(1+\tau_{K_{ie}})^{\alpha_{K_i}} (1+\tau_{L_{ie}})^{\alpha_{L_i}}}{A_{ie}} \right]^{\frac{\frac{1}{\sigma_i - 1}}{\frac{\sigma_i}{\sigma_i - 1} - \alpha_i}}$ with Ω_{P_i} an industry-specific constant.

Misallocation

By combining the model aggregation with the establishment responses to distortions, we follow the literature and measure misallocation as the distance between aggregate productivity and its frontier. At this frontier, all establishments in the industry have the same marginal revenue products. The more that actual productivity lags from its frontier, the larger is the measure of misallocation.

$$\Phi_i = \frac{TFP_i\big|_{\tau=\bar{\tau}}}{TFP_i} = \frac{\left[\sum_{e=1}^{N_i}\left(A_{ie} \times \Omega_{TFP,\tau=\bar{\tau},ie}\right)^{\sigma_i-1}\right]^{\frac{1}{\sigma_i-1}}}{\left[\sum_{e=1}^{N_i}\left(A_{ie} \times \Omega_{TFP,ie}\right)^{\sigma_i-1}\right]^{\frac{1}{\sigma_i-1}}}. \tag{1.9}$$

Formally, industry misallocation Φ_i in equation (1.9) captures the distance between actual industry total factor productivity TFP_i and its frontier where distortions, and hence marginal revenue products, are equalized across establishments $TFP_i|_{\tau=\bar{\tau}}$. Since industry output is produced using a CES technology, as per equation (1.2), the industry total factor productivity TFP_i is also a CES aggregate of establishment productivity A_{ie}. The scaling factor $\Omega_{TFP,ie}$ captures the extent to which each establishment shapes industry productivity. When we reallocate resources to equalize marginal revenue products, each establishment's scaling factor changes from $\Omega_{TFP,ie}$ to $\Omega_{TFP,\tau=\bar{\tau},ie}$. We now provide some intuition about this change in scaling parameters and then define them in terms of model objects.

Since a highly distorted establishment becomes more integral to industry productivity when its distortions are removed, the extent of misallocation depends on which establishments bear the greatest distortions. If the most productive establishments also bear the largest distortions, we measure more misallocation than if less productive establishments bear the same distortions. In short, the correlation between productivity and distortion shapes the extent of misallocation, a notion first emphasized by Restuccia and Rogerson (2008).[2] In our model, this notion relies on the claim that the scaling factor $\Omega_{TFP,\tau=\bar{\tau},ie}$ increases more relative to the scaling factor $\Omega_{TFP,ie}$ when an establishment is highly distorted. We substantiate this claim below after relating the scaling factors to model objects.

The scaling factors are based on establishments' *revenue* productivity $TFPR_{ie}$, which

[2]Hopenhayn (2014) makes clear that a discussion of correlations in this setting requires the comparison of the same proportional distortion. In his summary and re-framing of the literature, correlations matter because the same proportional distortion $\tau_{L_{ie}}$ would displace more labor at a productive establishment than at an unproductive one.

summarizes the impact of distortions on the establishments. As in Foster, Haltiwanger and Syverson (2008), $TFPR_{ie}$ measures an establishment's ability to generate revenue per input bundle:

$$TFPR_{ie} = \frac{P_{ie} Y_{ie}}{K_{ie}^{\alpha_{K_i}} L_{ie}^{\alpha_{L_i}}} = P_{ie} A_{ie}. \tag{1.10}$$

Equation (1.10) highlights the implication that, when comparing two establishments with the same physical productivity A_{ie}, a higher revenue productivity $TFPR_{ie}$ reflects a higher price. As we showed earlier, a higher price reflects larger distortions.

As the model focuses on the allocation of resources across establishments, the scaling factors compare the average revenue productivity of the industry, $\overline{TFPR_i}$, with the revenue productivity of an establishment, $TFPR_{ie}$. Equation (1.11) shows that this relative revenue productivity depends on the size of the establishment and the relative distortions that it faces. In a comparison of two equally productive establishments, the more distorted establishment would have a smaller $\overline{TFPR_i}/TFPR_{ie}$ ratio. Equation (1.12) shows that the relative revenue productivity after equalizing marginal products is a function of the post-reallocation size of the establishment.

$$\Omega_{TFP,ie} = \frac{\overline{TFPR_i}}{TFPR_{ie}} = \left(\frac{P_{ie} Y_{ie}}{P_i Y_i}\right)^{\alpha_i - 1} \left(\frac{\overline{1+\tau_{K,i}}}{1+\tau_{K_{ie}}}\right)^{\alpha_{K_i}} \left(\frac{\overline{1+\tau_{L,i}}}{1+\tau_{L_{ie}}}\right)^{\alpha_{L_i}} \tag{1.11}$$

$$\Omega_{TFP,\tau=\bar{\tau},ie} = \left.\frac{\overline{TFPR_i}}{TFPR_{ie}}\right|_{\tau=\bar{\tau}} = \left(\left.\frac{P_{ie} Y_{ie}}{P_i Y_i}\right|_{\tau=\bar{\tau}}\right)^{\alpha_i - 1} = \left(\frac{\left[A_{ie}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i - 1} - \alpha_i}}}{\sum_{e=1}^{N_i} \left[A_{ie}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i - 1} - \alpha_i}}}\right)^{\alpha_i - 1}. \tag{1.12}$$

Before formally characterizing how the scaling factors in equations (1.11) and (1.12) differ from each other, we want to emphasize how they are shaped by variations in markups and returns to scale. First, deviations from constant returns to scale (i.e. $\alpha_i \neq 1$) imply that the size of the establishment affects its revenue productivity. By contrast, in the Hsieh-Klenow model, returns to scale are constant and the size term drops out of the scaling factors; for instance, the counterfactual TFPR ratio in equation (1.12) is then unity for all establishments, regardless of industry. Second, the difference between the markup $\sigma_i/(\sigma_i - 1)$ and the returns to scale α_i shapes the counterfactual size of the establishment in (1.12). In our model, two industries could be populated by equally productive establishments, and yet different wedges between markups and returns to scale would lead the industries to differ in their counterfactual size distributions. Under the Hsieh-Klenow assumptions, the counterfactual size distribution would be the same in both industries.

We examine the impact of these types of differences on misallocation in greater detail in section 1.2.2.

Returning now to the measure of misallocation, we show that, when rid of its distortions, a more distorted establishment becomes more integral to industry productivity. In equation (1.13) we isolate the establishment-specific components of the relative scaling factors:

$$\frac{\Omega_{TFP,\tau=\bar{\tau},ie}}{\Omega_{TFP,ie}} \propto \left[\left(\frac{1+\tau_{K_{ie}}}{1+\tau_{K,i}}\right)^{\alpha_{K_i}}\left(\frac{1+\tau_{L_{ie}}}{1+\tau_{L,i}}\right)^{\alpha_{L_i}}\right]^{\frac{\frac{\sigma_i}{\sigma_i-1}-1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}. \tag{1.13}$$

Since establishment productivity A_{ie} enters both scaling factors in the same manner, the only establishment-specific difference between the two comes from distortions. Note that the exponent on the distortions in (1.13) is positive, so that the derivative of $\Omega_{TFP,\tau=\bar{\tau},ie}/\Omega_{TFP,ie}$ with respect to the distortions is positive. In other words, the relative increase in scaling factor $\Omega_{TFP,\tau=\bar{\tau},ie}$ is greater for a more distorted establishment.

Having defined all elements of industry-level misallocation, we use the model structure to express the economy-wide misallocation Φ as the geometric average of the industry measures Φ_i, as per equation (1.14):

$$\Phi = \prod_{i\in I}\Phi_i^{\theta_i}. \tag{1.14}$$

Misallocation here captures the aggregate productivity loss from distortions faced by establishments within industries.

While this measure is standard within the literature, its construction implicitly relies on some additional assumptions. For instance, by focusing on equalizing distortions within industries, we leave average distortions unchanged across industries. This assumption overlooks the potential productivity improvement from reallocating resources across industries. Moreover, this measure of misallocation assumes no changes in entry and exit of establishments when we alter distortions. Another potential concern might be the absence of taste (i.e., demand) shocks from the benchmark model. For that particular case, however, we show in appendix A.3 that the measure of misallocation is unchanged for a simple extension where we allow establishment-specific taste parameters. In short, misallocation is a counterfactual that holds all non-distortion parameters—including tastes—fixed; the measure of misallocation above would correctly capture productivity losses even in that extended model.

1.2.2 Ignoring the Variation in Returns to Scale and Markups

In this section, we show that inappropriately imposing constant returns to scale and a common markup leads to incorrect measures of productivity and misallocation. Imposing constant returns to scale when returns to scale are decreasing, or understating the markup of price over marginal cost, leads us to measure more distorted establishments as more productive. This spurious positive correlation between productivity and distortion leads us to overstate misallocation. We use the expressions we derive in this section to help us explain in section 3.3 why and when the divergent trends in misallocation arise.

The discrepancies we highlight arise from inappropriate mappings from the observable establishment revenue to the unobservable establishment productivity. As we emphasized in the introduction, mapping from revenue to productivity is a two-step process: first we map revenue to output with the help of a pricing model, and then we map output to productivity with the help of a production function. We begin to formalize this notion by combining the demand for establishment output with the establishment production function, and derive the expression for establishment productivity A_{ie} in equation (1.15):

$$\ln A_{ie} = \frac{\sigma_i}{\sigma_i - 1} \ln\left(\frac{P_{ie}Y_{ie}}{P_iY_i}\right) - \alpha_i \ln\left[K_{ie}^{\frac{\alpha_{K_i}}{\alpha_i}} L_{ie}^{\frac{\alpha_{L_i}}{\alpha_i}}\right] + \ln Y_i. \tag{1.15}$$

This expression clarifies the first mapping by showing the markup $\sigma_i/(\sigma_i - 1)$ as the elasticity of productivity with respect to the revenue-based measure of size $P_{ie}Y_{ie}/(P_iY_i)$. Furthermore, returns to scale in production α_i highlight the second mapping, as α_i is the elasticity of productivity with respect to the input bundle under the assumption of constant returns to scale $K_{ie}^{\alpha_{K_i}/\alpha_i} L_{ie}^{\alpha_{L_i}/\alpha_i}$. We now explore the discrepancies in measures of productivity and misallocation from imposing constant returns to scale and a common markup.

Discrepancy from Imposing Constant Returns to Scale

To measure total factor productivity A_{ie}, we need to impose a production function on the data; as suggested by equation (1.15), if we mismeasure the returns to scale in production, we incorrectly measure productivity. We formalize this notion in equation (1.16) by comparing the productivity $\widehat{A_{ie}}$ measured under constant returns to scale to the productivity A_{ie} measured under returns to scale α_i:

$$\left(\frac{\widehat{A_{ie}}}{A_{ie}}\right)_{CRTS\ Discrepancy} = \left(K_{ie}^{\frac{\alpha_{K_i}}{\alpha_i}} L_{ie}^{\frac{\alpha_{L_i}}{\alpha_i}}\right)^{\alpha_i - 1}. \tag{1.16}$$

For example, if we impose constant returns to scale on an industry where returns to scale are decreasing, then the exponent on the input bundle in equation (1.16) is negative. As a result, if we compare two equally productive establishments in this decreasing returns to scale industry, then the more distorted establishment with the smaller input bundle would be perceived as more productive. The discrepancy works in the opposite direction when returns to scale are increasing: more distorted establishments with smaller input bundles appear less productive than they are.

These discrepancies in measured productivity lead us to discrepancies in measured misallocation. In equation (1.17) we compare the misallocation $\widehat{\Phi}_i$ derived under the imposition of constant returns to scale with the misallocation Φ_i derived under the returns to scale α_i:

$$\left(\frac{\widehat{\Phi}_i}{\Phi_i}\right)_{CRTS\ Discrepancy} = \frac{\left[\sum_{e=1}^{N_i}\left(A_{ie}\frac{\overline{TFPR_i}}{TFPR_{ie}}\bigg|_{\tau=\bar{\tau}} \Xi_{crts,ie}{}^{1-\alpha_i}\right)^{\sigma_i-1}\right]^{\frac{1}{\sigma_i-1}}}{\left[\sum_{e=1}^{N_i}\left(A_{ie}\frac{\overline{TFPR_i}}{TFPR_{ie}}\bigg|_{\tau=\bar{\tau}}\right)^{\sigma_i-1}\right]^{\frac{1}{\sigma_i-1}}}, \tag{1.17}$$

$$\textit{where } \Xi_{crts,ie} = \left\{\left[\frac{1+\tau_{K_{ie}}}{\overline{1+\tau_{K,i}}}\right]^{\frac{\alpha_{K_i}}{\alpha_i}}\left[\frac{1+\tau_{L_{ie}}}{\overline{1+\tau_{L,i}}}\right]^{\frac{\alpha_{L_i}}{\alpha_i}}\frac{s_{ie}|_{\tau=\bar{\tau}}}{s_{ie}}\right\}$$

When returns to scale are constant so that $\alpha_i = 1$, then the exponent on the establishment-specific scaling factor $\Xi_{crts,ie}$ is 0, and the ratio in (1.17) collapses to 1: the two measures of misallocation are identical. However, deviations from constant returns to scale lead to incorrect measures of misallocation.

The size of the discrepancy in misallocation depends on the extent to which returns to scale are not constant, and on the correlation between productivity and distortion. Note that the scaling factor $\Xi_{crts,ie}$ takes values above 1 for heavily distorted establishments; each of the three ratios defining the scaling factor exceeds 1 for a heavily distorted establishment. Under decreasing returns to scale, the positive exponent on $\Xi_{crts,ie}$ puts larger weights on the distorted establishments. If productivity and distortions are positively correlated, then the numerator in (1.17) exceeds the denominator, and we overstate misallocation. For the same positive correlation of productivity and distortion, an industry in which returns to scale are increasing would induce a negative exponent on $\Xi_{crts,ie}$ and lead us to understate misallocation if we inappropriately impose constant returns. After estimating returns to scale, we use these expressions to understand how imposing constant returns leads the Hsieh-Klenow measure of misallocation to deviate from our measure.

Discrepancy from Imposing a Common Markup

We also need the markup so as to map establishment revenue to establishment productivity; as hinted by equation (1.15), an incorrect markup leads to incorrect measures of output and productivity. We formalize this notion in equation (1.18) where we compare the productivity $\widehat{A_{ie}}$, measured under the markup generated by $\widehat{\sigma_i}$, with the productivity A_{ie}, measured under the true markup σ_i:

$$\left(\frac{\widehat{A_{ie}}}{A_{ie}}\right)_{Markup\ Discrepancy} = \left(\frac{P_{ie}Y_{ie}}{P_iY_i}\right)^{\frac{\widehat{\sigma}_i}{\widehat{\sigma}_i-1}-\frac{\sigma_i}{\sigma_i-1}}. \tag{1.18}$$

In short, imposing an incorrect markup leads to a measure of productivity that is a function of the establishment size $P_{ie}Y_{ie}/(P_iY_i)$. For instance, when the imposed markup overstates the true markup, then the exponent on establishment size is positive. Consequently, if we compare two equally productive establishments, then the more distorted establishment will be smaller in size, and would be incorrectly perceived as less productive. In this respect, overstating the markup induces similar discrepancies in measuring productivity as does understating the returns to scale in equation (1.16).

The imposition of an incorrect markup results in an incorrect measure of misallocation. To anticipate our subsequent decomposition, we formalize this notion under the assumption of constant returns to scale. In equation (1.19), we compare the misallocation $\widehat{\Phi}_i$ measured under the incorrect markup to the misallocation Φ_i measured under the true markup:

$$\left(\frac{\widehat{\Phi}_i}{\Phi_i}\right)_{Markup\ Discrepancy} = \left[\sum_{e=1}^{N_i} s_{ie}\big|_{\tau=\bar{\tau}}\ \Xi_{markup,ie}{}^{\frac{\widehat{\sigma}_i-\sigma_i}{\sigma_i-1}}\right]^{\frac{1}{\widehat{\sigma}_i-1}}, \tag{1.19}$$

$$\text{where } \Xi_{markup,ie} = \frac{s_{ie}\big|_{\tau=\bar{\tau}}}{s_{ie}}$$

If the markup is measured correctly, so that $\widehat{\sigma_i} = \sigma_i$, then the establishment-specific scaling factor $\Xi_{markup,ie}$ disappears; and, since the relative establishment sizes $s_{ie}|_{\tau=\bar{\tau}}$ sum to 1 by definition, there is no error in measuring misallocation. However, deviations from the correct markup lead to discrepancies in measured misallocation.

The magnitude of the discrepancy in measured misallocation depends on the direction in which we mismeasure the markup, and the correlation of productivity and distortion. We note that the scaling factor $\Xi_{markup,ie}$ takes values greater than 1 for distorted establishments since distorted establishments grow larger in size when the distortions are removed. Consider a setting in which productivity and distortion are positively correlated.

If we understate the markup, the scaling factor puts more weight on the large, productive establishments, and puts less weight on the small, unproductive establishments. This rescaling of establishment size makes the expression in equation (1.19) exceed 1, leading to a measure of misallocation that is too large. By contrast, overstating the markup makes the exponent on the scaling factor negative, reversing the impact of the scaling on the relative establishment sizes, and leading us to understate misallocation. Below we use these expressions to understand the forces that differentiate our measure of misallocation from the Hsieh-Klenow measure that imposes a common markup and constant returns to scale.

1.3 Mapping the Model to Data

In this section, we show how to map the available U.S. Census microdata to measure distortions and productivity in U.S. manufacturing. With data only on establishment revenue—not output or prices—we emphasize the need for an estimating equation that jointly estimates returns to scale and price markups. We show that the reduced-form elasticities from this estimating equation inform us about profit shares, and that the model can be used to translate these reduced-form elasticities into returns to scale and markups. We then provide estimates of returns to scale and markups that are consistent with the estimated profit shares.

1.3.1 Data

Our analysis relies on two core data sets from the U.S. Census Bureau: the Census of Manufactures (CMF) and the Annual Survey of Manufactures (ASM). The Census data sets provide us with the establishment-level variables from which we infer productivity and distortions. The CMF is conducted every five years (for years ending in 2 and 7) and contains information about all manufacturing establishments in the United States. The ASM is conducted in all non-Census years and covers establishments with at least 250 employees, as well as a randomly sampled panel of smaller establishments. On average, the ASM surveys 50,000–65,000 establishments selected from the approximately 350,000 establishments in the CMF. From these datasets, we obtain measures of value added, hours worked, materials expenditures, capital stock, and the relevant price deflators.[3] Our sample period spans 1977 through 2007. We exclude establishments whose information is imputed from administrative records, as well as those with missing information.

[3]The industry price deflators come from the NBER-CES database, and the capital stocks are constructed by the Census staff, following Foster, Grim and Haltiwanger (2016).

As industry classification in the U.S. changed during the sample period, we build off the concordance made by Fort and Klimek (2015) that assigns establishments a time-consistent NAICS (North American Industrial Classification System) 2002 code. For a small number of the 400+ 6-digit NAICS industries, we identify discontinuities in industry employment and establishment counts around the years where industry classification changed.[4] If the NAICS dictionaries suggest that the industries in question are cross-listed, we attempt to merge them into a single industry. When the merging eliminates discontinuities, we use the merged industries; otherwise, we exclude the industries from analysis. We also exclude industries that contain fewer than five establishments in any given year.

To construct more comprehensive industry measures of expenditures on labor, we supplement the Census data on salaries and wages with BLS measures of benefit payments. While the ASM and the CMF exhaustively cover many aspects of manufacturing establishments, the U.S. Census microdata on total labor compensation is much sparser; only direct payments to labor for services in production (i.e., salaries and wages) are widely documented. By contrast, for a smaller sample of establishments, the BLS-run National Compensation Survey collects data on wages, paid leave, insurance, retirement contributions, legally required benefits, and supplemental pay. From these data, the BLS constructed for us unpublished estimates of the hourly wage and the hourly total benefit cost. Using these data, we construct a BLS Adjustment with which we can adjust the Census industry labor payment to reflect payments to labor:

$$\text{BLS Adjustment}_{i,t} = \frac{\text{BLS hourly wage}_{i,t} + \text{BLS hourly benefits}_{i,t}}{\text{BLS hourly wage}_{i,t}}.$$

Given the survey size, to pass BLS disclosure review, our BLS Adjustment factors are constructed at the NAICS 3-digit level for five-year intervals spanning 1983–2007.[5]

1.3.2 Step 1: Measuring Distortions

To measure misallocation, we need to know the distortions faced by an establishment relative to the average distortions in the industry. We derive relative distortions by rearranging the first-order conditions from equations (1.5) and (1.6) and dividing by their

[4]We construct mid-point growth rates, and flag growth rates of establishment counts or hours worked that exceed 0.5 in absolute value.

[5]We apply the BLS Adjustment factors created with data from 1983–1987 to the Census data in both 1987 and 1982.

weighted averages over all establishments in the industry:

$$\frac{\overline{1+\tau_{K_{ie}}}}{\overline{1+\tau_{K,i}}} = \frac{\frac{P_{ie}Y_{ie}}{K_{ie}}}{\left[\sum_{e=1}^{N_i} \frac{P_{ie}Y_{ie}}{P_iY_i} \left(\frac{P_{ie}Y_{ie}}{K_{ie}}\right)^{-1}\right]^{-1}} = \frac{\frac{\text{Value Added}_{ie}}{\text{Capital Stock}_{ie}}}{\left[\sum_{e=1}^{N_i} \frac{\text{Value Added}_{ie}}{\text{Value Added}_{i}} \left(\frac{\text{Value Added}_{ie}}{\text{Capital Stock}_{ie}}\right)^{-1}\right]^{-1}} \tag{1.20}$$

$$\frac{\overline{1+\tau_{L_{ie}}}}{\overline{1+\tau_{L,i}}} = \frac{\frac{P_{ie}Y_{ie}}{L_{ie}}}{\left[\sum_{e=1}^{N_i} \frac{P_{ie}Y_{ie}}{P_iY_i} \left(\frac{P_{ie}Y_{ie}}{L_{ie}}\right)^{-1}\right]^{-1}} = \frac{\frac{\text{Value Added}_{ie}}{\text{Labor Hours}_{ie}}}{\left[\sum_{e=1}^{N_i} \frac{\text{Value Added}_{ie}}{\text{Value Added}_{i}} \left(\frac{\text{Value Added}_{ie}}{\text{Labor Hours}_{ie}}\right)\right]^{-1}}. \tag{1.21}$$

The resulting expressions, in equations (1.20) and (1.21), are independent of the returns to scale and markup parameters, which are common to all establishments in the industry, and map transparently to Census data.[6]

The model interprets high revenue productivity in inputs as an indicator for the presence of distortions. In a world without distortions, this model suggests that all establishments hire inputs so as to equalize their average capital $P_{ie}Y_{ie}/K_{ie}$ and average labor $P_{ie}Y_{ie}/L_{ie}$ revenue productivities. If an establishment has a high revenue productivity in a certain input, it would maximize profits by continuing to hire that input until this measure of revenue productivity declined and equaled that of the other establishments in the industry. If an establishment in the data has a high average revenue productivity in a given input, it must have been prevented from hiring more of the input; hence, the model assigns this establishment a high distortion.

These strong assumptions identify distortions and reflect the model's attempt to describe a steady-state economy. In a dynamic setting, we can think of frictions that might prevent an establishment from hiring the steady-state profit-maximizing quantity of an input. Asker, Collard-Wexler and De Loecker (2014), for instance, focus on adjustment costs in the hiring of capital as one reason that an establishment's choice might deviate from these steady-state predictions. Nonetheless, for the purpose of measuring misallocation across longer periods of time, we think these assumptions are a reasonable starting point. To match this view of the model's purpose, our estimates of model parameters and misallocation are based on five-year periods; we also document the robustness of the main results in section C.2 by extending this estimating window to ten years.

[6] Aggregation in the model leads to expressions that make it more natural to define the average $1/(1+\tau)$. Therefore we define $\overline{1+\tau}$ as the inverse of the average $1/(1+\tau)$. We present the details in appendix A.

1.3.3 Step 2: Measuring Productivity

With data on establishment revenue, not output and prices separately, we cannot directly estimate the returns to scale and markup we need to infer productivity. Instead, the revenue elasticities from our estimating equation inform us about the division of value added among labor, capital, and profits. Nonetheless, using model equations we can indirectly map these reduced-form revenue elasticities into returns to scale and markups, and then infer establishment productivity.

A common approach to measuring returns to scale in data sets with establishment revenue entails creating a proxy for output by dividing revenue $P_{ie}Y_{ie}$ with an industry price index P_i; this common practice leads to a downward bias in estimated returns to scale that was first pointed out by Marschak and Andrews (1944) and later made particularly salient by Klette and Griliches (1996). Intuitively, this bias arises because we expect the most productive establishments to hire the largest input bundles, to produce the most output, and—when output markets are imperfectly competitive—to charge the lowest prices. If the most productive establishments charge the lowest prices, then the proxy for output is likely to understate output most for these productive establishments. A cross-sectional estimator using this output proxy would understate the increase in output from having the large input bundles, and hence underestimate returns to scale.

The derivation of our estimating equation highlights this downward bias in returns-to-scale estimates. Specifically, we follow De Loecker (2011) and combine two model equations: the establishment's production function and the demand for its output. Rearranging this combined expression to solve for the ratio of revenue $P_{ie}Y_{ie}$ and the price index P_i, and taking logs, we derive the estimating equation (1.22):

$$\ln\left(\frac{P_{ie}Y_{ie}}{P_i}\right) = \beta_{K_i}\ln(K_{ie}) + \beta_{L_i}\ln(L_{ie}) + \beta_{Y_i}\ln(Y_i) + \beta_{A_i}\ln(A_{ie}), \tag{1.22}$$

$$\text{where } \beta_{K_i} = \frac{\alpha_{K_i}}{\frac{\sigma_i}{\sigma_i - 1}},\ \beta_{L_i} = \frac{\alpha_{L_i}}{\frac{\sigma_i}{\sigma_i - 1}},\ \beta_{Y_i} = \frac{1}{\sigma_i},\ \text{and } \beta_{A_i} = \frac{\sigma_i - 1}{\sigma_i}$$

$$\text{and } P_{ie}Y_{ie} = \text{Value Added}_{ie},\ K_{ie} = \text{Capital Stock}_{ie},\ L_{ie} = \text{Labor Hours},^{7}$$

$$P_i = \text{NBER-CES Industry Price Index}_i,\ P_iY_i = \sum_{e=1}^{N_i}\text{Value Added}_{ie},\ Y_i = \frac{P_iY_i}{P_i}.$$

The revenue elasticities $\beta_{i,L}$ and $\beta_{i,K}$ are quotients of the returns-to-scale parameters and the markup of price over marginal cost. Since we expect establishments to price at or above

[7] We compute total labor hours as the sum of the reported production-worker hours and the calculated non-production-worker hours following Kehrig (2011).

marginal cost, the gross markup exceeds 1. As a result, even when correctly estimated, the revenue elasticities understate the returns-to-scale parameters.

Although they do not directly estimate returns to scale, the revenue elasticities β_{K_i} and β_{L_i} are useful descriptors of differences across industries: they correspond to capital's and labor's share of value added and together imply an industry's profit share. Rearranging the first-order conditions from equations (1.5) and (1.6), and summing across establishments within an industry, we show in (1.23) that β_{K_i} and β_{L_i} are the distortion-inclusive expenditures on inputs as a share of value added:

$$\beta_{K_i} = \frac{\sum_{e=1}^{N_i}(1+\tau_{K_{ie}})R_i K_{ie}}{P_i Y_i} \text{ and } \beta_{L_i} = \frac{\sum_{e=1}^{N_i}(1+\tau_{L_{ie}})w_i L_{ie}}{P_i Y_i}. \tag{1.23}$$

In addition, we show in equation (1.24) that industry profits are the residual share of value added (i.e., the difference between one and the sum of the revenue elasticities):

$$\frac{\Pi_i}{P_i Y_i} = 1 - (\beta_{K_i} + \beta_{L_i}). \tag{1.24}$$

Since we expect establishments to earn weakly positive profits, the expression in (1.24) emphasizes that the sum of revenue elasticities is bounded from above by 1 in this model. This is yet another way to see the bias emphasized by Klette and Griliches (1996): if this model correctly characterizes the world, and if we lived in a world with returns to scale α_i in excess of 1, the standard estimating equation would still produce revenue elasticities that sum to less than 1.

The third revenue elasticity β_{Y_i}, the elasticity of establishment revenue with respect to industry output, is key to identifying the returns to scale and markup parameters from the revenue elasticities β_{K_i} and β_{L_i}. Specifically, the inverse of β_{Y_i} is the elasticity of substitution σ_i, from which we can construct the markups $\sigma_i/(\sigma_i - 1)$. With the estimated markup we can then back out the returns to scale parameters α_{K_i} and α_{L_i} as the products of the markup and the respective revenue elasticities. With the parameters for the markup and the returns to scale in hand, we can infer productivity.

We estimate $\widehat{\beta_{L_i}}$, the first of the three key elasticities, using the rearranged first-order condition for labor in expression (1.23). We map this expression to the data by multiplying the sum of salaries and wages reported in the U.S. Census microdata by the BLS Adjustment factors we detailed in section 1.3.1. In this way, our measure of industry labor expenditures attempts to capture not only the wage payments to labor, but also the benefits and indirect payments, from insurance to retirement contributions, that are not widely reported to the

Census. We then divide this measure of labor costs by the industry value added, as in equation (1.25):

$$\widehat{\beta_{L_i}} = \frac{\left[\sum_{e=1}^{N_i} \text{Salaries and Wages}_{ie}\right] \times \text{BLS Adjustment}_i}{\sum_{e=1}^{N_i} \text{Value Added}_{ie}}. \tag{1.25}$$

This estimate implicitly assumes that the labor distortions faced by establishments are priced into the wages and benefits that establishments pay workers. While we think this a reasonable assumption, we cannot rule out the possibility that some distortions are not priced. As a robustness check in section C.2, we also estimate this elasticity from the variation in labor usage across establishments. Even under these different assumptions required to thus estimate the elasticity, we find the path of U.S. manufacturing misallocation to look very different under the assumptions of our model and those of the Hsieh-Klenow model.

We estimate the remaining two elasticities β_{K_i} and β_{Y_i} using a two-step Generalized Methods of Moments (GMM) procedure based on the control-function approach in Levinsohn and Petrin (2003). This approach addresses the issue that productivity is unobserved in the estimating equation (1.26) by substituting out the unobserved productivity with a function of observable variables. Specifically, the control function is the choice of intermediate inputs, assumed to increase in establishment productivity: $ln(M_{ie}) = m(\ln K_{ie}, \ln Y_i, \ln A_{ie})$. If we can invert the expression characterizing this choice to express productivity as a function of the intermediate inputs, then we can substitute the unobservable A_{ie} in equation (1.22) with observables K_{ie}, M_{ie}, and Y_i as follows:

$$\underbrace{\ln\left(\frac{P_{ie}Y_{ie}}{P_i}\right) - \widehat{\beta_{L_i}}\ln(L_{ie})}_{pynet_{ie}} = \beta_{K_i}\ln(K_{ie}) + \beta_{Y_i}\ln(Y_i) + \beta_{A_i}m^{-1}(\ln K_{ie}, \ln Y_i, \ln M_{ie}) + u_{ie}, \tag{1.26}$$

where u_{ie} represents idiosyncratic shocks to production. For this substitution to be feasible and useful, we need to assume that the choice of intermediate inputs is invertible, and that productivity is the only unobservable component in the choice of intermediate inputs.[8] The first step of the procedure regresses the left-hand-side term of equation (1.26) $pynet_{ie}$ on a flexible polynomial of the observables to construct the predicted $\widehat{pynet_{ie}}$.

The second step of the procedure uses the assumption that log productivity $\ln A_{ie}$ evolves following a general first-order Markov process to construct moment conditions

[8]While common, these assumptions are strong and not directly testable. For instance, the second assumption eliminates the possibility that there are distortions in the intermediate input markets that are correlated with establishment productivity.

with which to estimate the elasticities β_{K_i} and β_{Y_i}. Specifically, we let $\varepsilon_{ie,t}$ correspond to the mean-zero innovations in productivity realized at time t. For a given guess $(\widehat{\beta_{K_i}}, \widehat{\beta_{Y_i}})$ of the elasticities, we construct an implied measure of log productivity by differencing $\widehat{pynet_{ie}}$ and $\widehat{\beta_{K_i}} \ln(K_{ie}) + \widehat{\beta_{Y_i}} \ln(Y_i)$. Regressing the implied productivity on a polynomial of its past value gives us the implied innovation to productivity $\varepsilon_{ie,t}(\widehat{\beta_{K_i}}, \widehat{\beta_{Y_i}})$, and the following moment conditions with which to estimate the two key elasticites:

$$\frac{1}{N}\frac{1}{T}\sum_{e\in N_i}\sum_{t\in T}\begin{pmatrix}\widehat{\varepsilon_{ie,t}}(\widehat{\beta_{K_i}}, \widehat{\beta_{Y_i}}) \ln K_{ie} \\ \widehat{\varepsilon_{ie,t}}(\widehat{\beta_{K_i}}, \widehat{\beta_{Y_i}}) \ln Y_i\end{pmatrix} = 0. \tag{1.27}$$

To estimate the elasticities in a model-consistent way, we constrain the parameter space to meet three criteria. First, to ensure that industry profits are weakly positive and less than 1 as a share of value added, we impose that $\widehat{\beta_{K_i}}$ and $\widehat{\beta_{L_i}}$ sum to a value between 0 and 1. Second, to estimate labor and capital shares of value that are strictly positive, we require that $\widehat{\beta_{K_i}}$ and $\widehat{\beta_{L_i}}$ are strictly positive. Third, to back out gross markups with values between 1 and 2, we impose that $\widehat{\beta_{Y_i}}$ be strictly positive and less than 0.5.[9] We implement these parameter restrictions by modifying the code made available by Jagadeesh Sivadasan for implementing the Ackerberg, Caves and Frazer (2015) variant of the Levinson-Petrin procedure, which refines the assumptions for identifying the model parameters and updates the estimator accordingly.

1.3.4 Division of Value Added in U.S. Manufacturing

By our estimates in panel A of table 1.1, labor's share of value added in U.S. manufacturing declined from 64% in 1982 to 39% in 2007; over the same period, the capital share increased from 20% to 25%. Together, these changes in the labor and capital shares imply that the profit share increased 20 percentage points, rising from 16% in 1982 to 36% in 2007. While, to our knowledge, this is the first paper to document these dynamics of industry profits for U.S. manufacturing, the findings are broadly consistent with other recent work. The decline of the labor share has been widely documented for the U.S. and for the global economy [e.g. Karabarbounis and Neiman (2014), Elsby, Hobijn and Sahin (2013), Barkai (2016)].[10] Moreover, using data on the U.S. non-financial corporate sector, Barkai (2016) finds that both the labor and capital shares declined, leading to an increase

[9] We think this is a reasonable parameter range as common choices for the elasticity σ_i range between 3 and 11, and imply markups between 1.1 and 1.5.

[10] The economy-wide decline has been of a smaller magnitude, around 6%. The much more pronounced decline of the labor share in the U.S. manufacturing sector that we document with the combined Census and BLS data is consistent with recent work by Autor et al. (2017) who also use U.S. Census data.

Table 1.1: U.S. Manufacturing – Division of Value Added

Panel A	Weighted Average across Industries		
	Capital Share β_{K_i}	Labor Share β_{L_i}	Profit Share $1-(\beta_{L_i}+\beta_{K_i})$
1982	0.20	0.64	0.16
1987	0.21	0.61	0.18
1992	0.27	0.55	0.18
1997	0.25	0.49	0.26
2002	0.31	0.46	0.23
2007	0.25	0.39	0.36

Panel B	Standard Deviation across Industries		
	Capital Share β_{K_i}	Labor Share β_{L_i}	Profit Share $1-(\beta_{L_i}+\beta_{K_i})$
1982	0.19	0.20	0.17
1987	0.20	0.19	0.20
1992	0.25	0.20	0.20
1997	0.23	0.18	0.24
2002	0.25	0.19	0.21
2007	0.26	0.19	0.30

Note: Reported values in panel A are weighted averages of industry-level coefficients, with the weights comprising industry value added. The underlying coefficients are estimated using five-year panels. Data for the estimation comes from the Annual Survey of Manufactures from the U.S. Census and the National Compensation Survey from the U.S. Bureau of Labor Statistics.

in profits over the last 30 years. Complementary exercises in Karabarbounis and Neiman (2014) also suggest that the capital share increased insufficiently to offset the decline in the labor share, implying that profits increased.

In addition to documenting the evolution of these shares across time, we document in panel B large variations in capital, labor, and profit shares across industries. At all points in time, the standard deviation of profit shares across industries is roughly as large as the average level of the profit shares. The standard deviations of capital and labor shares are of quantitatively similar magnitudes. These large standard deviations imply that the U.S. manufacturing sector is populated both by industries where profit margins are slim, as well as by industries in which establishments earn large profits as shares of value added.

1.3.5 Returns to Scale and Markups in U.S. Manufacturing

Accommodating these estimates of the profit shares requires deviating from the standard assumptions of constant returns to scale and a common markup. From Basu and Fernald (1997) we know that profits drive a wedge between markups and returns to scale under very general assumptions on the functional forms for production and demand. In our model, this relationship takes the following form:

$$1 - \frac{\Pi_i}{P_i Y_i} = \frac{\alpha_i}{\frac{\sigma_i}{\sigma_i - 1}}, \tag{1.28}$$

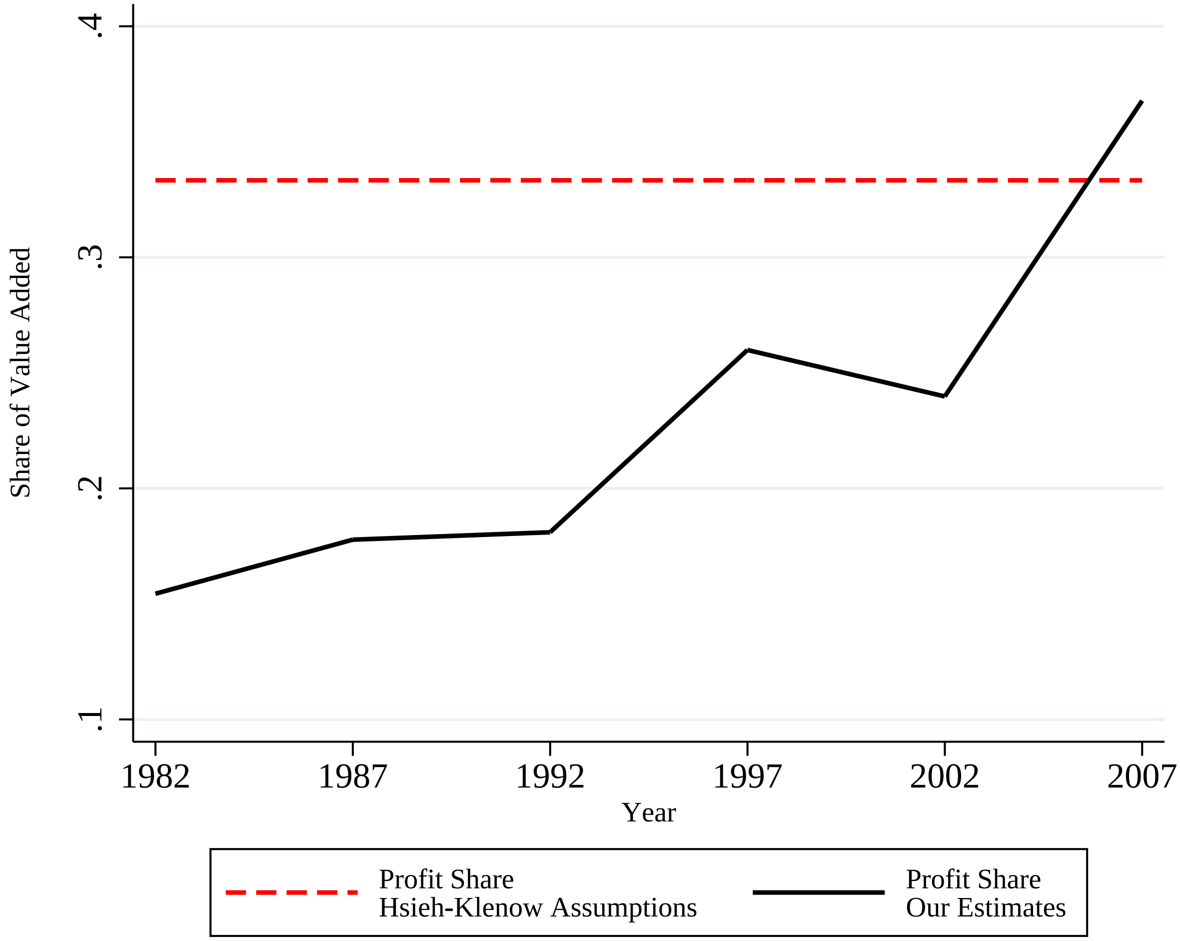
where the industry profit shares $\Pi_i/(P_i Y_i)$ act as a wedge between the returns to scale α_i and markup $\sigma_i/(\sigma_i - 1)$. By imposing constant returns to scale and a markup of 1.5 in every industry, the Hsieh-Klenow model implies that all establishments in all industries earn a third of their value added as profits. We emphasize this point in figure 1.2. The solid black line plots our estimated share of profits in value added. This rising measure of profits contrasts with the invariance of profit shares in the Hsieh-Klenow model, plotted as the dashed red line. The average profit share in 2007 of 0.36 roughly matches the Hsieh-Klenow assumptions. However, the variation across industries and the smaller profit shares throughout the 1980s and 1990s fit these assumptions less well.

To understand why markups and returns to scale can rationalize these variations in profit shares, we focus on the fact that an establishment earns profits when its price exceeds the average cost of production:

$$\frac{\pi_{ie}}{Y_{ie}} = \underbrace{\frac{\sigma_i}{\sigma_i - 1}\text{Marginal Cost}_{ie}}_{\text{Price}_{ie}} - \text{Average Cost}_{ie}. \tag{1.29}$$

The profits per unit sold, as per equation (1.29), can increase either if the markup increases or if the returns to scale decline. First, an establishment could increase its profit margin by charging a higher markup over marginal cost. Second, an establishment could increase its profit margin if average cost falls relative to marginal cost. A reduction in returns to scale drives such a shift in costs. For example, constant returns imply a constant marginal cost, while decreasing returns imply a marginal cost that increases with each unit produced. As a result, if returns to scale decline from constant to decreasing, the marginal cost for the last unit would exceed the average cost of all units produced, increasing the profit margin. Some combination of an increase in markups and a reduction in returns to scale drives the increase in profit shares in the data.

Figure 1.2: Profits as a Share of Value Added in U.S. Manufacturing



In panel A of table 1.2, we show that, while markups increased from 1.46 to 1.48, the decline in returns to scale from 1.23 to 0.96 is the primary driver of rising profit shares between 1892 and 2007. In short, the U.S. manufacturing sector exhibited meaningfully increasing returns to scale in the early 1980s. Since then, returns to scale have declined, driving up marginal cost relative to the average cost of production. By increasing the profit margin on each unit sold, this decline in returns to scale led to the rise in profit shares for U.S. manufacturing.[11]

[11] In a recent paper, De Loecker and Eeckhout (2017) explain the rise in profits as a consequence of rising markups. Our paper and De Loecker and Eeckhout (2017) share a common idea: changes in the shares of value added can be understood as either changes in markups or changes in returns to scale. In this paper, we jointly estimate the parameters underlying demand and production, and we allow the parameters to change over time. De Loecker and Eeckhout (2017) estimate time-invariant production parameters and infer markups as residuals that explain the observed value-added shares. However, if profits increase but returns-to-scale are kept time invariant, then increasing markups are the only way to rationalize increasing profits. This methodological difference, coupled with different data coverage in terms of sectors and aggregation, likely explains the different findings.

Table 1.2: U.S. Manufacturing – Returns to Scale and Markups

Panel A	Average Level across Industries	
	Returns to Scale α_i	Markups $\frac{\sigma_i}{\sigma_i - 1}$
1982	1.23	1.46
1987	1.20	1.44
1992	1.20	1.44
1997	1.12	1.51
2002	1.11	1.47
2007	0.96	1.48

Panel B	Standard Deviation across Industries	
	Returns to Scale α_i	Markups $\frac{\sigma_i}{\sigma_i - 1}$
1982	0.42	0.41
1987	0.46	0.40
1992	0.48	0.41
1997	0.49	0.43
2002	0.44	0.42
2007	0.58	0.42

Note: Reported values in Panel A are weighted averages of industry-level coefficients, with the weights comprising industry value added. Data for the estimation comes from the Annual Survey of Manufactures from the U.S. Census, and the National Compensation Survey from the U.S. Bureau of Labor Statistics.

Much like profit shares, both returns to scale and markups vary widely across industries. The standard deviations of both measures range between one third and one half the average values of their respective variables. For returns to scale, this variation suggests that, even as returns to scale have declined on average, the U.S. manufacturing sector is still comprised of both increasing and decreasing returns-to-scale industries. Similarly, while the average markup may be large, there are many industries with markups low enough to approximate perfect competition, as well as many industries where the degree of imperfect competition, and hence the markup, is large.

While table 1.1 showed how changing capital and labor shares drive the evolution of profits, here we show how the same evolution can be understood in terms of changing returns to scale and markups:

$$\Delta \frac{\Pi_i}{P_i Y_i} = \underbrace{-\Delta \alpha_{K_i} \frac{1}{\left(\frac{\sigma_i}{\sigma_i - 1}\big|_{2007}\right)} - \Delta \alpha_{L_i} \frac{1}{\left(\frac{\sigma_i}{\sigma_i - 1}\big|_{2007}\right)}}_{\text{Contribution of Returns to Scale}} + \underbrace{\left(\Delta \frac{\sigma_i}{\sigma_i - 1}\right) \frac{1 - \left(\frac{\Pi_i}{P_i Y_i}\big|_{1982}\right)}{\left(\frac{\sigma_i}{\sigma_i - 1}\big|_{2007}\right)}}_{\text{Contribution of Markup}}. \tag{1.30}$$

An increase in an industry's profit share between 2007 and 1982, $\Delta\Pi_i/(P_iY_i)$, is driven either by a decline in returns to scale $-\Delta\alpha_i$ or an increase in the markup $\Delta\sigma_i/(\sigma_i - 1)$, as per equation (1.30). Applying this decomposition to the manufacturing-sector data in tables 1.1 and 1.2, we show that of the 20-percentage-point increase in the manufacturing profit share, 18 percentage points come from the decline in returns to scale and 1–2 percentage points from the rise in the markup.[12] We can further decompose the 18 percentage points to emphasize separately the contributions of the capital and labor elasticities, α_{K_i} and α_{L_i}. The increase in the capital elasticity, reflected principally by the rising capital share, put downward pressure on the profit share of about –6 percentage points. Meanwhile, the sharp decline in the labor elasticity, reflected in the falling labor share, contributed 24 percentage points to the increase in the manufacturing profit share. We next turn to misallocation and emphasize the importance of incorporating this variation in markups and returns to scale.

1.4 Misallocation

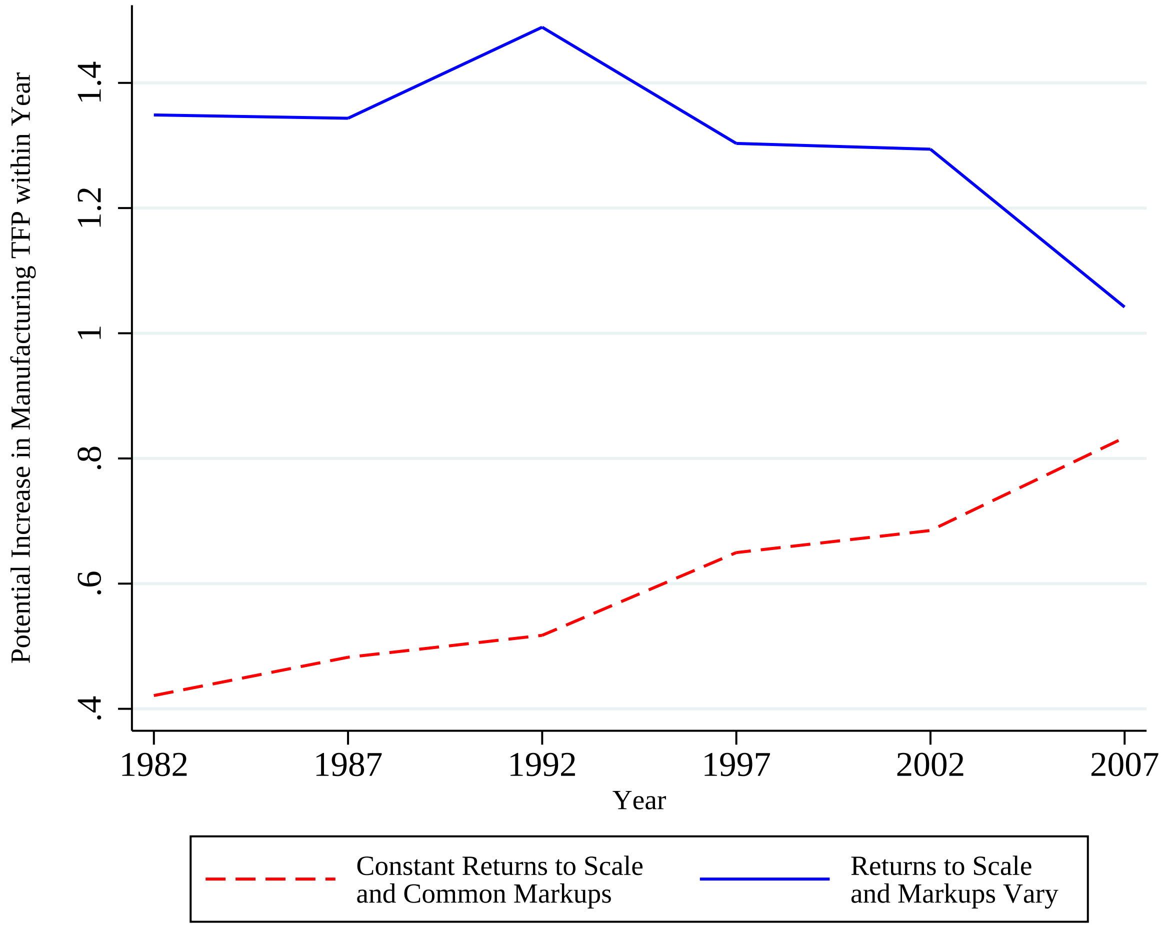
In this section, we present our measure of misallocation and contrast it with the Hsieh-Klenow measure that ignores variation in markups and returns to scale. We then decompose the discrepancy in measurement and show that the divergent trends in misallocation are driven by the decline in returns to scale over time.

1.4.1 Misallocation Has Not Been Increasing

Our estimates suggest that misallocation in U.S. manufacturing decreased over the last 30 years. Figure 1.3 quantifies misallocation as the potential increases in U.S. manufacturing TFP from equalizing the distortions establishments face within an industry, as per

[12] A Jensen's inequality term leads to the small discrepancy. The manufacturing profit share in table 1.1 is the weighted average of industry profit shares, which is equal to $1 - \sum_{i \in I} \theta_i \frac{\alpha_i}{\frac{\sigma_i}{\sigma_i - 1}}$. Meanwhile, the average returns to scale and markup reported in table 1.2 are also weighted averages and do not imply exactly the same manufacturing profit share $1 - \left(\sum_{i \in I} \theta_i \alpha_i\right) / \left(\sum_{i \in I} \theta_i \frac{\sigma_i}{\sigma_i - 1}\right)$.

Figure 1.3: Misallocation in U.S. Manufacturing
Change in U.S. Manufacturing TFP from Equalizing Within-Industry Distortions

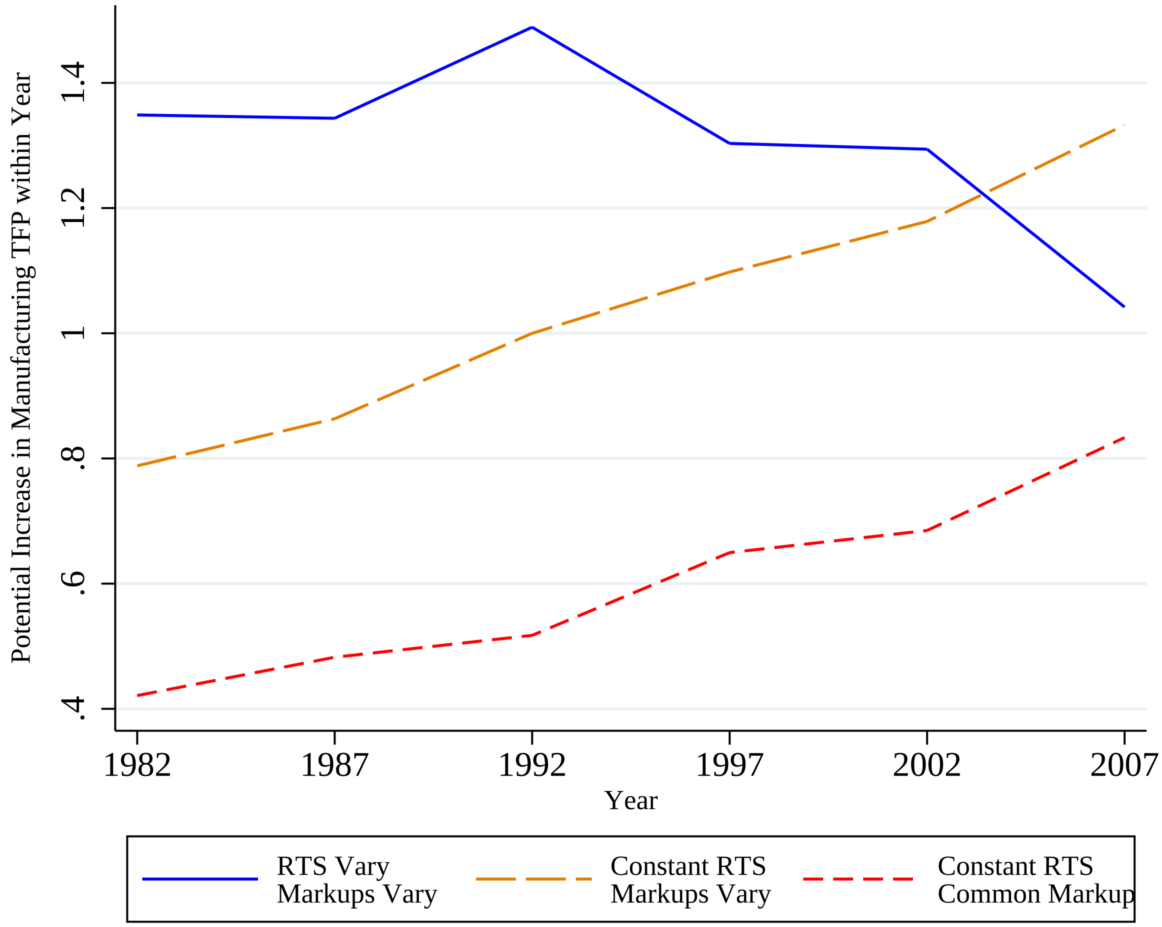


equation (1.9). The solid blue line depicts our model while the dashed red line depicts the Hsieh-Klenow model. By our estimates, the level of misallocation declined from 135% in 1982 to 104% in 2007. Meanwhile, misallocation increased under the Hsieh-Klenow assumptions, so that in 2007 the U.S. manufacturing sector could have been 83% more productive, nearly twice the potential increase of 42% in 1982. Figure 1.1 presented the same results expressed as changes relative to 1982.

We focus on trends in misallocation, rather than levels, because the model is static and consequently imposes the long-run steady state at each point in time. As we described in section 1.3.2, the model infers distortions by assuming that, in a world without misallocation, establishments hire inputs until their average revenue products are equalized across establishments. Short-run considerations can change that inference: for instance, adjustment costs or the time required to build productive capital could lead non-distorted establishments to differ in their average revenue products at a point in time. Despite these costs, we follow the literature and impose the steady-state assumption for two reasons.

Figure 1.4: Misallocation in U.S. Manufacturing

Change in U.S. Manufacturing TFP from Equalizing Within-Industry Distortions



First, by using a static model we can transparently document the role that industry-varying markups and returns to scale play in changing the measure of misallocation. Second, while these short-run considerations may lead us to misstate the level of misallocation, they likely have a smaller impact on trends across long periods of time.

To understand the source of the divergent trends in misallocation, we next decompose the discrepancy in measured misallocation into a component from imposing a common markup across industries and a component from imposing constant returns to scale. In figure 1.4 we preview the formal decomposition by plotting an intermediate measure of misallocation in which we include only one source of industry variation. In the long-dashed orange line we impose constant returns to scale, but maintain the estimated markups that vary across industries. The discrepancy in measured misallocation between our model and the Hsieh-Klenow model can now be split into two parts. The discrepancy from imposing the common markup is the distance from the intermediate model's long-dashed line and the Hsieh-Klenow model's short-dashed line. The discrepancy from imposing constant returns to scale is the distance between the our model's solid line and the new intermediate model's long-dashed one.

As we formally show over the next two sections, the divergent trends in misallocation are driven by the reduction in returns to scale between 1982 and 2007. As the U.S. manufacturing sector began to better approximate the assumed constant returns in the Hsieh-Klenow model, the discrepancy from imposing constant returns declined, leading to a perceived rise in misallocation. Figure 1.4 shows that most of the discrepancy in 1982 came from imposing constant returns to scale. By 2007, the discrepancy from imposing constant returns was less than half its initial value in absolute terms, while the discrepancy from imposing a common markup remained relatively unchanged. This reversal is reflected in the changing distances between the three lines. The shrinking distance between our solid blue line and the intermediate model's orange long-dashed line reflects the declining discrepancy in misallocation from imposing constant returns to scale. By contrast, the relatively stable distance between the Hsieh-Klenow model's and the intermediate model's lines suggests a more stable discrepancy over time from imposing a common markup.

1.4.2 Discrepancies in Establishment-Level Productivity

Incorrect measures of misallocation, both from imposing constant returns to scale and from imposing a common markup, are rooted in spurious correlations between productivity and the distortions that establishments face. As we did in figure 1.4, we document these spurious correlations in turn, focusing first on returns to scale and then on markups.

In panel A of table 1.3 we show that inappropriately imposing constant returns to scale leads to measures of productivity that conflate productivity and distortion. The regressions in panel A control for the productivity estimated when returns to scale vary, and compare the constant-returns productivity of establishments with different input bundles. This conditioning allows us to compare establishments that have the same productivity under our model, but that face different distortions, and hence have different input bundles. The key regression coefficients are conditional correlations of constant-returns productivity and input bundles, shown separately for industries with decreasing and increasing returns. As suggested by equation (1.16), these correlations should be opposite in sign.

Columns 2 and 3 support model predictions that imposing constant returns to scale on industries where returns to scale are not constant leads to predictable spurious correlations between productivity and distortions. Column 2 emphasizes that imposing constant returns in place of decreasing returns leads us to perceive more distorted establishments (i.e., those with smaller input bundles) as more productive. Specifically, a 1-standard-deviation

Table 1.3: Productivity Mismeasurement at the Establishment Level

Panel A: Imposing Constant Returns to Scale

Dependent Variable	Normalized Log Productivity (A_{ie}) (Constant Returns to Scale)		
	(1)	(2)	(3)
Normalized Log Input Bundle	0.1460	-0.3238	0.4465
	(0.0149)	(0.0147)	(0.0106)
Normalized Log Productivity (A_{ie}) (Variable Returns to Scale)	0.8241	1.0527	0.7528
	(0.0068)	(0.0067)	(0.0074)
Industry-Year Sample	All	Decreasing RTS	Increasing RTS
Industry×Year Fixed Effects	Yes	Yes	Yes
Observations	292000	126000	166000
R-squared	0.8130	0.9268	0.9338

Panel B: Imposing a Common Markup across Industries

Dependent Variable	Normalized Log Productivity (A_{ie}) (Common Markup)		
	(1)	(2)	(3)
Normalized Log Value Added	0.2514	-2.0286	0.5993
	(0.0243)	(0.0566)	(0.0137)
Normalized Log Productivity (A_{ie}) (Heterogeneous Markups)	0.4839	2.3961	0.4678
	(0.0155)	(0.0436)	(0.0104)
Industry-Year Sample	All	Understated Markup	Overstated Markup
Industry×Year Fixed Effects	Yes	Yes	Yes
Observations	292000	116000	176000
R-squared	0.5630	0.7099	0.9143

Note: Unit of observation is an establishment-year. The time period comprises 1982, 1987, 1992, 1997, 2002, and 2007. Standard errors are clustered at the industry-year level. To normalize the values within each industry, we demean the variable and divide by its standard deviation.

decrease in the log input bundle leads to a measure of productivity that is 0.32 standard deviations larger under the constant returns to scale model. Column 3 emphasizes the opposite pattern for increasing-returns industries. Following a 1-standard-deviation decrease in the log input bundle, productivity is 0.45 standard deviations smaller under constant returns to scale. In this case, imposing constant returns on industries where returns to scale are increasing leads us to perceive more distorted establishments as less productive.

In panel B of table 1.3, we show that understating the markup in an industry leads us to perceive more distorted establishments as more produtive, while overstating the markup leads us to perceive more distorted establishments as less productive. We document this pattern through the predictions from equation (1.18) by linking the mismeasurement of productivity to establishment size. In a parallel with panel A, we control for the productivity measured under the estimated markup, and then compare the common-markup productivity of establishments that differ in distortions, and hence in their sizes.

Columns 2 and 3 partition the sample by estimated markup size and back the model predictions. In particular, column 2 suggests that, indeed, understating the markup leads us to a spurious positive correlation between productivity and distortions: a 1-standard-deviation decrease in size (i.e. an increase in distortion) leads us to a 2.03-standard-deviation increase in common-markup productivity. Column 3 presents the opposite result for instances where we overstate the markup: a decrease in size leads to a 0.60-standard-deviation decrease in common-markup productivity.

1.4.3 Aggregate Decomposition

Having emphasized in the last section that incorrect markups and returns to scale lead to spurious correlations between productivity and distortion, we now show how those spurious correlations lead to discrepancies between our measure of misallocation and the Hsieh-Klenow measure. We emphasize that these discrepancies are positive when we overstate the correlation of productivity and distortion, and that the discrepancies are negative when we understate the correlation of productivity and distortion.

In the following schematic, we present the theoretical decomposition where the second row splits the aggregate discrepancy into one component from imposing constant returns to scale and another component from imposing the common markup. The aggregate discrepancy measures the difference in misallocation between the Hsieh-Klenow model (constant returns to scale [CRTS] and a common markup of 1.5 everywhere [$\sigma = 3$]) and our own (returns to scale [VRTS] and markups [$\widehat{\sigma}$] can both vary). We first capture the component from imposing constant returns by comparing the CRTS and VRTS measures of misallocation under the estimated markups $\widehat{\sigma}$. We then capture the component from the common markup by comparing the common markup ($\sigma = 3$) misallocation to the variable markup ($\widehat{\sigma}$) misallocation under CRTS. We can further decompose each component to understand the contribution of decreasing versus increasing returns to scale, as well as understating versus overstating the markup.

Aggregate Discrepancy

$$\ln \frac{\Phi_{CRTS,\sigma=3}}{\Phi_{VRTS,\widehat{\sigma}}}$$

Driven by Constant Returns to Scale

$$\ln \frac{\Phi_{CRTS,\widehat{\sigma}}}{\Phi_{VRTS,\widehat{\sigma}}}$$

Driven by the Common Markup

$$\ln \frac{\Phi_{CRTS,\sigma=3}}{\Phi_{CRTS,\widehat{\sigma}}}$$

Decreasing RTS

$$\ln \sum_{DRTS} \theta_i \left(\frac{\Phi_{i,CRTS,\widehat{\sigma}}}{\Phi_{i,VRTS,\widehat{\sigma}}} \right)$$

Increasing RTS

$$\ln \sum_{IRTS} \theta_i \left(\frac{\Phi_{i,CRTS,\widehat{\sigma}}}{\Phi_{i,VRTS,\widehat{\sigma}}} \right)$$

Understated

$$\ln \sum_{\widehat{\sigma}>3} \theta_i \left(\frac{\Phi_{i,CRTS,\sigma=3}}{\Phi_{i,CRTS,\widehat{\sigma}}} \right)$$

Overstated

$$\ln \sum_{\widehat{\sigma}\leq 3} \theta_i \left(\frac{\Phi_{i,CRTS,\sigma=3}}{\Phi_{i,CRTS,\widehat{\sigma}}} \right)$$

Table 1.4 decomposes the aggregate discrepancy in misallocation and shows that a decline in returns to scale explains why the discrepancy is smaller in 2007 than in 1982. The first two rows of panel A show that the 50% difference in misallocation between the Hsieh-Klenow model and our own in 1982 is split rather evenly between the imposition of constant returns to scale and the imposition of a common markup. By 2007, the aggregate discrepancy of 12% is split unevenly: the returns-to-scale component is half its previous value in absolute terms, while the markup component is essentially unchanged in size. These values quantify the visual decomposition from figure 1.4; the values in 1982 and 2007 capture the vertical distances among the three lines in the figure.

The third row of table 1.4 relates the discrepancy in misallocation to spurious correlations of productivity and distortion. In parentheses, the third row reports the difference in the correlation of productivity and distortion between the Hsieh-Klenow model and our own. For instance, the top of panel A indicates that imposing constant returns to scale on decreasing-returns industries in 1982 leads us to overstate the correlation of productivity and distortion by 0.13. By overstating this correlation, the constant-returns model also overstates misallocation, in this instance by 17%.[13] Across all deviations from the Hsieh-Klenow assumptions and across both years, inducing spurious positive correlation

[13] The 17% is scaled by the size of industries with decreasing returns to scale in U.S. manufacturing. By contrast, not as many industries have their markup understated relative to the Hsieh-Klenow markup of 1.5. Hence, even though understating the markup leads to a larger 0.28 overstatement of the productivity-distortion correlation, the overstated misallocation in those industries amounts to a smaller 3%.

Table 1.4: Decomposing the Differences in Misallocation

Panel A: 1982

Aggregate Discrepancy			
-0.4968			
Driven by Constant Returns to Scale		Driven by the Common Markup	
-0.2589		-0.2378	
Decreasing RTS	Increasing RTS	Understated	Overstated
0.1761	-0.4470	0.0317	-0.2599
(0.1315)	(-0.2805)	(0.2847)	(-0.2218)

Panel B: 2007

Aggregate Discrepancy			
-0.1212			
Driven by Constant Returns to Scale		Driven by the Common Markup	
0.1249		-0.2461	
Decreasing RTS	Increasing RTS	Understated	Overstated
0.4676	-0.3349	0.0455	-0.2853
(0.1850)	(-0.2641)	(0.1290)	(-0.2891)

of productivity and distortion leads us to overstate misallocation, and inducing spurious negative correlations leads us to understate misallocation.

1.5 Robustness

In this section, we argue that different trends in misallocation persist even when we incorporate additional modifications to the model and the data. The resulting trends in misallocation are both qualitatively and quantitatively similar to our baseline results. We then present ongoing work to estimate model parameters under different assumptions and with different estimators. Although different parameter estimates lead to different point estimates for the growth in misallocation, the stark qualitative differences between our model and the Hsieh-Klenow model remain.

In our first robustness exercise, we emphasize the need for time-varying model parameters for capturing the evolution of the profit shares. While our estimates match the rising profit shares through a decline in returns to scale, we consider an alternative parametrization: we impose constant returns to scale, and calculate hypothetical markups that account for all the industry and time variation in profit shares. In panel A of table 1.5, we show that matching industry profits through markups alone also does away with the increasing trend in misallocation from the Hsieh-Klenow model. By this alternative calculation, misallocation between 1982 and 2007 is virtually unchanged, increasing by 3%. By contrast, the baseline misallocation from the Hsieh-Klenow model increased 29% over the same period. We view the elimination of this upward trend in misallocation as evidence that accounting for industry profits is of first-order importance for measuring misallocation.

In our second robustness exercise, we allow establishments to charge different markups *within* an industry. Formally, we follow Atkeson and Burstein (2008) in assuming that establishments sell their output in oligopolistically competitive markets instead of monopolistically competitive ones. In this setting, an establishment is aware that its choice of how much to produce affects both its own price *and* also the price level of the whole industry. Larger establishments exert a larger impact on the industry price level and this influence is reflected in larger markups. This establishment-specific markup depends on the elasticity of substitution σ_i, which is common to all industries in the Hsieh-Klenow model and varies across industries in our model. We present full details of the model in appendix A.2. The key challenge in this extension is to solve for the establishment-specific markup in the counterfactual where we eliminate distortions. This problem is akin to a contraction mapping, and we solve it by iterating on an initial guess.

Panel A of table 1.5 shows that the additional generalization to markups that vary within the industry leaves trends in misallocation essentially unchanged. Relative to the baseline 29% increase and the 13% decline, allowing markups to vary across establishments leads to a 28% increase and an 11% decline, respectively, in the Hsieh-Klenow model and in our own. While the trends in misallocation remain unchanged, the levels of misallocation decline with heterogeneous markups within the industry. The decline is more notable in our model, with misallocation some 10% lower per year (e.g., from 104% to 96% in 2007), while the level in the Hsieh-Klenow model declines about 3% (e.g., 83% to 81% in 2007).

In a third robustness exercise, also reported in panel A, we argue that the different patterns of misallocation are robust to accounting for sample selection in the Annual Survey of Manufactures. The survey covers all large establishments and a random sample of smaller ones. Our baseline estimates of misallocation account for this sample selection by weighting establishments by their Census-provided sampling weights in calculating industry and aggregate misallocation. For this exercise, we construct the measure of misallocation using the full Census of Manufactures in 1982 and 2007, two of the years for which we have such data available. At a 27% increase and a 9% decline, the results of this extension replicate the baseline patterns.

We next consider alternative ways, and sets of assumptions, for estimating markups and returns to scale, and argue that introducing industry and time variation in these parameters continues to remove the sharp increase in misallocation from the Hsieh-Klenow

Table 1.5: U.S. Manufacturing Misallocation Relative to 1982, Robustness

Panel A:	Baseline Estimates	
	Hsieh-Klenow Model	Our Model
Baseline	0.29	-0.13
Model Change: impose constant returns to scale with implicit markups to match profit shares		0.03
Model Change: allow markups to vary across establishments in an industry	0.28	-0.11
Sample Change: use Census of Manufactures instead of Annual Survey of Manufactures	0.27	-0.09
Panel B:	**Alternate Estimates**	
	Hsieh-Klenow Model	Our Model
Estimation Change: estimate labor share of value added using Ackerberg et al (2015) instead of FOC	0.22	0.09
Estimation Change: define industries more broadly as NAICS 4-digit instead of NAICS 6-digit	0.26	-0.32
Estimation Change: use ten-year panels instead of five-year panels and compare 2007 to 1987	0.18	-0.02

model. First, instead of calculating the labor share of value added β_{L_i} directly as the share of labor expenditures, we estimate β_{L_i} in a control-function procedure alongside the two other elasticities. Second, we estimate markups and returns to scale for more broadly defined industries. Lastly, we lengthen the time frame of the estimation, using ten-year panels instead of five-year panels of data to estimate markups and returns to scale.

While our baseline estimates directly measure the labor share as the ratio of labor costs to value added, at the top of panel B we instead estimate the labor share as a revenue elasticity using the Ackerberg, Caves and Frazer (2015) correction to the Levinsohn and Petrin (2003) control-function procedure. To estimate this labor elasticity, we need additional assumptions that justify the use of intermediate inputs as proxies for productivity. One possibility is that some unobserved component of productivity is realized after an establishment chooses its labor and before it chooses its intermediate inputs. Hence, we now have to assume that establishments choose the labor they hire before they choose their intermediate inputs, and that unobserved productivity is realized before the intermediate-input choice. Our estimates of this labor elasticity suggest an 11% decline in labor's share of value added, compared to our direct calculation of a 25% decline. With a smaller decline of the labor share, we also find a smaller reduction in returns to scale over time. Ultimately, this more modest change in returns to scale over time leads to a smaller departure from the Hsieh-Klenow model's trend in misallocation; these alternate estimates imply a 9% increase in misallocation, a bit less than half the increase in the Hsieh-Klenow model.

We next estimate markups and returns to scale for more broadly-defined industries, and find that the divergent patterns of misallocation are amplified. Specifically, we use the NAICS-4 industry code instead of the more detailed NAICS-6. For instance, an industry now corresponds to "Dairy Product" instead of "Ice Cream and Frozen Dessert." The second entry in panel B shows that while misallocation in the Hsieh-Klenow model increases a bit over of 20%, misallocation in our model falls 32%, more than twice our baseline decline. This larger decline reflects an interaction of two forces. First, our measure of misallocation focuses on within-industry reallocation of resources. When we broaden the industry definition, we implicitly allow resources to be allocated across the NAICS-6 industries that comprise a NAICS-4. Second, returns to scale determine how large an establishment grows as a share of the industry when its distortions are removed. The larger are the returns to scale, the greater is the share of industry revenue generated by the most productive establishment. The interaction of larger industries and the reduction in returns to scale over time amplifies the decline in misallocation relative to our baseline results.

Lastly, we use ten-year instead of five-year panels to estimate the model parameters; this procedure attenuates the differences in parameter values across time and hence reduces the differences in misallocation trends between the two models. Under these parameter estimates, our model suggests that misallocation decreased 2% between 1987 and 2007 while the Hsieh-Klenow model implies an increase of 18%. We contextualize these estimates by reference to table 1.2, panel A, in which we document a continuous decline in returns to scale over the same period. By pooling the last decade of data together in this robustness exercise, our estimate of the decline in returns to scale is smaller than when we compare returns to scale only using the first five and the last five years of the sample. Nonetheless, even this smoothing of parameter estimates preserves the divergent trends in misallocation.

1.6 Conclusion

We argue in this paper that accounting for industry and time variation in markups and returns to scale leads to a measure of misallocation in U.S. manufacturing that is decreasing over time; this result stands in contrast to the increasing measure of misallocation under the widely-applied assumptions of a common markup and constant returns to scale, as in the Hsieh-Klenow model. To quantify these differences, we use five-year panels of restricted U.S. Census microdata to estimate markups and returns to scale across manufacturing industries. We find that industries differ meaningfully in these parameters at a given point in time, and that the average returns to scale in U.S. manufacturing declined between 1982 and 2007.

We decompose the differences in misallocation between the two models, and identify the decline in returns to scale as the primary driver of the divergent trends in misallocation. The Hsieh-Klenow measure on average understates our measure of misallocation. The assumption of constant returns to scale is a better fit for the data in 2007 than it is for 1982. Consequently, as the U.S. manufacturing sector began to reflect more closely the assumption of constant returns, the discrepancy in measuring misallocation declined. As this discrepancy declined, the Hsieh-Klenow measure of misallocation asymptoted toward our measure from below and hence drove the upward trend in misallocation.

We formalize the source of these differences in misallocation and show that, by ignoring the variation in markups and returns to scale, the Hsieh-Klenow model measures productivity in a way that conflates productivity and distortions. These spurious correlations lead us to incorrectly infer the extent to which the most productive establishments bear the most burdensome distortions, and hence to an incorrect measure of misallocation.

We think the patterns we identify in markups and returns to scale, and the discrepancies we highlight in measuring productivity, could be of broader interest. Outside the literature on misallocation, the measurement of establishment-level productivity is a key input in other attempts to trace the impacts of policies and shocks from affected establishments to aggregate outcomes.

CHAPTER II

Globalization and Top Income Shares

From a work with Lin Ma

Abstract

This paper documents empirically that access to global markets is associated with a higher executive-to worker pay ratio within the firm. It then uses China's 2001 accession to the World Trade Organization as a trade shock to show that firms that exported to China prior to 2001 subsequently exported more, grew larger, and grew more unequal in terms of executive-to-worker pay. To analytically and quantitatively evaluate the impacts of globalization on top income inequality, this paper builds a model with heterogeneous firms, occupational choice, and executive compensation. In the model, executive compensation grows with the size of the firm, while the wage paid to ordinary workers is determined in a country-wide labor market. As a result, the extra profits earned in the foreign markets benefit the executives more than the average workers. We calibrate the model to the U.S. economy and match the income distribution closely in the data. Counterfactual exercises suggest that trade and FDI liberalizations can explain around 52 percent of the surge in top 0.1 percent income shares in the data between 1988 and 2008.

JEL Codes: E25 F12 F62 J33

Keywords: trade, income inequality, occupational choice, CEO compensation

2.1 Introduction

The real income of the top 0.1 percent of the population increased by 85.8 percent between 1993 and 2011 in the United States; the real income of the bottom 99 percent increased by only 5.8 percent over the same period.[1] At the same time, the past several decades witnessed the fastest pace of globalization since the start of the First World War. The existing literature does not provide a clear link between globalization and the runaway top income shares. Researchers working on the distributional effects of trade usually focus on wage inequality and especially on the "skill premium," the wage difference between skilled and unskilled workers.[2] However, the income of the top 0.1 percent—which usually consists of executive compensation, business profits, and capital gains—cannot be easily explained using the "skill premium."[3]

Complementing the literature on the skill premium, this paper examines how globalization shapes the income gap between the very rich and the rest of the population. We first document a novel empirical pattern: income gaps between the top executives and the average workers are higher among exporting firms than among non-exporting firms in the United States. We then provide causal evidence on this relationship using China's accession to the World Trade Organization (WTO) as a trade shock: firms that benefit from lower trade barriers export more, grow larger, and grow more unequal. Motivated by the empirical findings, we develop a new model that incorporates occupational choice and executive compensation into a heterogeneous firms model of trade. The model reproduces the empirical patterns we documented at the firm level, and generates income and firm size distributions that closely resemble the aggregate U.S. data. We then perform a quantitative assessment of the impacts of globalization on the CEO-to-worker pay ratio within a firm and on the top income shares within the country as a whole. Counterfactual simulation suggests that globalization can explain up to 52 percent of the surge in top 0.1 percent income shares in the U.S. between 1988 and 2008.

To establish a link between globalization and the income gap between the very rich and the rest, we create a new dataset that matches executive compensation to confidential U.S. Census microdata on payroll and international transactions. The resulting dataset focuses on publicly-listed firms and provides detailed information on executive compensa-

[1] Piketty and Saez (2003), with data updated to 2011.

[2] Among many others, see Goldberg and Pavcnik (2007), Helpman, Itskhoki and Redding (2010), and Burstein and Vogel (2012).

[3] For example, numerous studies have shown that education level, a widely used measure of skill, has no clear correlation with CEO compensation (Belliveau, O'Reilly and Wade, 1996; Geletkanycz, Boyd and Finkelstein, 2001) .

tion, employment, payroll, and export sales.[4] We focus on the top executives because they constitute a large fraction of the top earners. Around 40 percent of the top 0.1 percent income earners in the U.S. are professional executives, and these 0.04 percent earners are responsible for about 4 percent of the national income (Bakija et al., 2012). To our knowledge, this is the first dataset assembled that can be used to study the relationship between international trade and the CEO-to-worker pay ratio.

With this new dataset, we document that globalization disproportionately benefits the top executives relative to the workers within the same firm. Specifically, the gap in compensation between the CEO and the average worker in the firm—the CEO-to-worker pay ratio—is 50 percent larger for exporting firms compared to domestic firms. This globalization premium for the CEO-to-worker pay ratio holds for manufacturing and non-manufacturing firms, for multinationals as well as exporters, for private as well as public firms, for comparisons across firms, and for comparisons within firms as they transition to exporting. As in the simplest Melitz (2003) model of trade, the impact of globalization is intermediated through firm size. We show that accounting for the firm-size channel is sufficient to explain the differential in CEO-to-worker pay ratios between exporters and non-exporters. One can imagine other channels—special CEO skill for exporting or compensation for greater risk—that would increase the exporting CEO premium over and above what is implied by size. We do not find evidence of a meaningful premium in excess of that explained by size. Hence, our quantitative assessment focuses on the size channel.

We use China's 2001 accession to the WTO to provide direct evidence from a trade shock to firm exports, firm size, and CEO-to-worker pay ratios. Following its accession to the WTO, China gradually lowered import tariffs from an average of 15 percent in 2000 to 10 percent by 2007 (Lu and Yu, 2015). The tariff reductions potentially benefit firms with existing export links to China more than those without such existing relationships. Using a difference-in-differences methodology, we compare a treatment group of firms with China-specific trading relationships prior to 2001 to control groups drawn from the remaining firms. We find that, in the aftermath of China's WTO accession, the firms in the treatment group exported 57 percent more, grew 40 percent larger in terms of employment and payroll, and grew 13 percent more unequal as measured by the CEO-to-worker pay ratio. These results suggest that globalization might be responsible for the widening income gaps between the rich and the poor through within-firm inequality.

To evaluate the aggregate implications of the firm-level findings, we develop a framework that bridges the heterogeneous firm trade model based on Melitz (2003) with the

[4] Appendixes B.1 and B.2.1 construct the data and repeat the analysis for a limited set of privately-held firms.

literature on occupational choice and executive compensation. The model world consists of two countries. Each country is populated by a fixed measure of individuals who are endowed with different levels of human capital. An individual chooses between different occupations, as in Lucas (1978). She can either (1) create a new firm and become the founder and CEO of the firm or (2) work for an existing firm. If she chooses to create a new firm, her human capital determines the productivity of the firm, and her income depends positively on the size of the firm.[5] If she chooses to be a worker, her human capital determines the amount of efficiency labor she supplies to the market. The wage rate of efficiency labor is determined in a competitive countrywide labor market and equalized across firms within the same country. In equilibrium, only the individuals with human capital above a certain threshold choose to create firms, while the majority of the population chooses to work for an existing firm. Each firm produces a distinct variety and sells it in a monopolistically competitive market. Firms can choose to export to the foreign market after incurring fixed costs.

The model replicates the new empirical pattern documented in this paper; in equilibrium, within-firm inequality is higher among the firms that sell to the foreign market. The key mechanism is that the extra profits earned in the foreign market are not distributed evenly within the same firm. The compensation paid to the CEO of a firm is linked to the sales of the firm, while the wage rate of a typical worker is determined in a countrywide labor market. Any extra profits earned in the foreign market benefit the CEO directly, but benefit the workers only through general equilibrium effects. In the end, as the firm sells to the foreign market, its within-firm inequality will be higher. At the aggregate level, trade creates a gap in within-firm inequality between the exporting and domestic firms. Consistent with the empirical patterns described above, in the model, the size of the firm solely determines the level of within-firm inequality; once the size is controlled for, the exporting status of a firm has no impact on its CEO-to-worker pay ratio.

Before using the model to quantify the impact of globalization on top income shares, we show that the model can parsimoniously and precisely characterize the U.S. income and firm distributions at the same time. Empirically, the U.S. income distribution is well approximated by an exponential distribution for the majority at the left end and a Pareto distribution for the right tail.[6] At the same time, the U.S. firm size distribution can also be

[5]In the appendix, we provide an extension of the model to micro-found a market in which existing heterogeneous firms match with potential entrepreneurs. The micro-foundation delivers assortative matching between the CEOs and the firms, which in turn leads to the same compensation function that is exogenously assumed in the benchmark model.

[6]See Drăgulescu and Yakovenko (2001*a*), Drăgulescu and Yakovenko (2001*b*), Clementi and Gallegati (2005), and Yakovenko and Silva (2005) for details.

well described by a fat-tailed Pareto distribution (Axtell (2001)). These two distributions are captured simultaneously within the model by two assumptions: (1) human capital is distributed exponentially, and (2) firm productivity is an exponential function of the founder's human capital. The model then features a Pareto firm size distribution and a two-class-structured income distribution. The workers' wage depends on their human capital, which implies an exponentially distributed income outside of the very rich. The individuals at the right tail of the income distribution are the CEOs, whose income is linked to the size of the firm they manage. This implies that the right tail of the income distribution will follow the firm size distribution and thus be Pareto. Once the model is calibrated, it reproduces both the firm size and the income distribution observed in the data with reasonable precision.

Model counterfactuals suggest that trade liberalizations can explain 52 percent of the surge in the top 0.1 percent income shares in the United States between 1988 and 2008. To arrive at these values, we match one country in the model to the U.S. economy and the other to the rest of the world; we then calibrate their trade barriers and their relative TFP to match the data for each year. Targeting these moments alone, we compare the income distribution in the model to the income distribution in the data to quantify the potential explanatory power of our channel, linking globalization, firm size, and inequality. In other counterfactual exercises, we also study how income inequality responds to changes in trade barriers as we move from autarky to the observed level of trade openness. For 2008, this latter exercise more than triples the CEO-to-worker pay ratio at the largest firms in the United States. At the aggregate level, this opening to trade skews the income distribution rightward: the top 0.1 percent income share increases from 9.1 percent to 10.2 percent between autarky and trade.

By linking globalization and top income shares, this paper contributes to the literature on the distributional effects of globalization and the discussion on rising income inequality in the United States. The majority of the existing research in the international trade literature focuses on how globalization affects wage inequality, and particularly the wage and income gap between skilled and unskilled workers.[7] Top income inequality, such as the income gap between top managers and workers or the overall top income shares, is often overlooked in the trade literature. At the same time, researchers working on income inequality documented that the rising income inequality in the U.S. is mainly driven by the

[7]For example, see Feenstra and Hanson (1996), Manasse and Turrini (2001),Yeaple (2005), Helpman, Itskhoki and Redding (2010), and Egger and Kreickemeier (2012). Bernard and Jensen (1997) documented that exporting is associated with higher within-firm inequality in terms of the wage gap between skilled and unskilled workers. This paper focuses on another dimension of within-firm inequality: the wage gap between top managers and workers.

widening gaps between the top 1 percent and the bottom 99 percent, not by the income inequality within the bottom 99 percent themselves. Moreover, papers in this literature showed that a substantial part of the rise in U.S. top income inequality is due to the rise in labor income inequality, especially when business income is included in the category of labor income.[8] This current paper bridges the gap between the two literature by focusing on the impact of globalization on top income inequality. It is the first paper to show empirically that the access to the world markets increases CEO-to-worker pay ratio within the same firm, and thus trade can potentially affect top income shares. This paper also quantitatively shows that a large part of the surge in top income shares in the United States can potentially be attributed to globalization.

In broadening the focus to inequality between the very rich and the rest of the population, this paper complements an existing literature on inequality across executives. Monte (2011) and Meckl and Weigert (2011) developed models exploring the effects of trade on income inequality among the managers. By contrast, the model here is designed to generate a realistic income distribution that spans the entire population in general equilibrium, which has not been done before in the trade literature. This broader scope enables quantitative analysis of the aggregate impacts of globalization on income inequality, both within the right tail, and over the entire population.

By introducing Census data to the study of executive compensation, this paper also interfaces with the large literature on corporate governance and executive compensation (Roberts, 1956; Baker and Hall, 2004; Gabaix and Landier, 2008; Frydman and Saks, 2010). Compared to the existing literature, which mostly focuses on the level of executive compensation, the census data allow us both to measure the magnitude of executive compensation relative to the wages of ordinary workers within the same firm, and to do so on a large and comprehensive sample. In the process, we provide a new perspective to understand the implications of surging executive pay on inequality.[9] This paper is also the first to study executive-to-worker pay ratio among privately-held firms. A small strand of this literature, such as Sanders and Carpenter (1998), Oxelheim and Randøy (2005), Cuat and Guadalupe (2009), and Gerakos, Piotroski and Srinivasan (2009) documented that executive compensation in public firms increases as the firms start to participate in the global markets. This paper further documents that the positive link between executive compensation and globalization can also be observed at privately-held firms, though the magnitude is smaller.

[8]Among many others, see Piketty and Saez (2003) and Atkinson, Piketty and Saez (2011).

[9]It is possible to measure CEO-to-worker pay ratio without using the Census data as well. However, this usually leads to a biased and small sample of firms. This is discussed in detail in Section 2.2.

The rest of this paper is organized as follows. Section 2.2 presents the empirical results. Section 2.3 presents the model and Section 2.4 focuses on the analytical results. Section 2.5 provides details of the calibration and quantitative results. Section 2.6 concludes.

2.2 Empirical Evidence: Trade and Within-Firm Inequality

In this section, we first describe a new data set linking executive compensation to administrative firm-level data; we then document the robust relationships among within-firm inequality, export status, and size; finally, we provide direct evidence that trade shocks drive within-firm inequality using China's access to the WTO. These new empirical patterns motivate our modeling choices in section 2.3.

2.2.1 Data

Our empirical evidence focuses on public firms and is based on a linked data set that has three components: ExecuCompustat from Standard & Poor, the Longitudinal Business Database (LBD) from the Census Bureau, and the Longitudinal Firm Trade Transactions Database (LFTTD) from the U.S. Customs and the Census Bureau. Appendix B.1 details the construction of the data set used in the body of the paper, and introduces the data on privately held firms to which we return later.

This novel linked data set provides more comprehensive coverage of both payroll and export statistics relative to the data used in the existing literature. First, U.S. public firms are not required to disclose non-executive compensations. As a result, the majority of firms do not report total payroll expenditure in SEC filings, making it almost impossible to compute wages at the firm level and within-firm inequality. For example, as reported by Faleye, Reis and Venkateswaran (2013), around 87 percent of firms have to be dropped from ExecuCompustat due to this missing value problem in their study of the CEO-to-worker pay ratio. The under-reporting also leads to distortions of sectoral representation in the sample. For example, around 43 percent of the sample in ExecuCompustat are manufacturing firms, but they only constitute 16 percent of the sample in Faleye, Reis and Venkateswaran (2013). By contrast, the LBD provides universal coverage of employment and payroll and thus minimizes the loss of observations. Overall, around 50 percent of the ExecuCompustat observations can be matched with the linked LBD-LFTTD, which is on par with most studies that use the Compustat-SSEL bridge provided by the census. The sectoral representation in ExecuCompustat is also preserved in the data set used in this paper (See Table B.1 for details). For example, in the linked data set, manufacturing firms constitute 47 percent of the sample, a significant improvement over the sample used in the existing literature. Second, as firms are not required to report export sales separately, the missing value problem is prevalent in non-administrative data sets, forcing researchers to discard a large proportion of the data in studies that involve exporting behavior. By using the LFTTD, which provides universal coverage of U.S. international transactions, we minimize this reduction in sample size.

The final linked data set contains a sample of 17,233 firm-year observations between 1992 and 2007 with 2,561 unique firms. A total of 13,169 firm-year observations are classified as exporters and the remaining 4,054 as non-exporters. Overall the combined dataset contains around half of the US public firms over the period. Due to the nature of publicly-traded firms, large firms are over-represented in the dataset, as compared to the universe of U.S. firms. As a result of this, this dataset is also heavily skewed toward exporting firms: around 76 percent of the observations are exporting firms, and this is higher than the overall percentage of firms that export in the U.S.[10] Over-representation of large firms naturally leads to problems if one wish to make inferences for the overall economy. This problem is mitigated here, since it is reasonable to believe that the CEO-to-worker pay ratios are much smaller and less variable in small firms, and thus the results for the overall economy will be mainly driven by large firms.

The key variable of interest is the CEO-to-worker pay ratio. We construct this ratio as the total realized compensation (TDC2) divided by the average non-executive wage. To construct the average non-executive wage we subtract the salary and bonus of the CEO from the firm's total payroll for the year and then divide this difference by the total employment less the CEO. We rationalize this construction of the non-executive wage as follows: "Total payroll" as reported in the LBD comes from the Business Register, which is in turn based on IRS tax records. The salary and bonus of the CEO are reported as part of the total payroll for tax purposes, while the income earned from stock options is not.[11] Therefore, we need subtract only the salary and bonus of the CEO when computing the non-executive wage. The denominator is one less the total employment to account for the fact that the CEO is also counted as an employee in tax filings.

2.2.2 Export Status and Within-Firm Inequality

Over the course of our sample, an average CEO earns 89 times more than an average worker in the same firm; this CEO-to-worker pay ratio varies by exporting status: it is on

[10] For example, Bernard, Jensen and Schott (2009) reports that 18 percent of U.S. manufacturing firms exported in 2002.

[11] The "total payroll" and "employment" items in LBD are compiled from filings of IRS Form-941/943. See IRS Publications 15, 15-A, and 15-B for the details of tax deductions and exemptions.

average 92 for exporting firms and 81 for non-exporting firms. Appendix table B.2 reports these and other summary statistics.

We test these differences in within-firm inequality across exporters and non-exporters by estimating the following equation on the pooled panel data:

$$\log\left(\mathrm{CEO_{it}}/\mathrm{WAGE_{it}}\right) = \beta_0 + \beta_1 \mathrm{EXP}_{it} + \mathbf{b'}_2 \cdot \mathbf{g} + \mathbf{b'}_3 \cdot \mathbf{y} + \epsilon_{it}. \tag{2.1}$$

We define $\mathrm{CEO_{it}}/\mathrm{WAGE_{it}}$ as the CEO-to-worker pay ratio, EXP_{it} as the exporter status indicator for firm i at year t, $\mathbf{g}$ as a vector of group fixed effects (e.g., four-digit Standard Industrial Classification (SIC) codes, or firm identifiers), and $\mathbf{y}$ as a vector of year fixed effects. The standard errors are clustered at the year-sector level in the baseline specification. The coefficient of interest is β_1: if the CEO-to-worker pay ratio is significantly higher for exporters, we expect this parameter to be positive.

Table 2.1 shows that the CEO-to-worker pay ratio is higher among exporters than among non-exporters across a variety of specifications. Column 1 looks at manufacturing firms, includes sector fixed effects and finds that this measure of within-firm inequality is 73.3% higher for exporters than for non-exporters. Column 2 repeats the comparison across all firms and finds an exporter premium of 50.7% for within-firm inequality. We look

Table 2.1: Within-Firm Inequality and Export Status

Dependent Variable	Log CEO-to-Worker Pay Ratio			
	(1)	(2)	(3)	(4)
Exporter	0.733***	0.507***		0.093***
	0.108	(0.030)		(0.028)
Log Exports			0.119***	
			(0.005)	
Sample	Manufacturing	All	All	All
Year Fixed Effects	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Firm
Observations	8,000	17,000	13,000	17,000
R-squared	0.219	0.270	0.323	0.628

Note: The left-hand side variable for each of the regressions is the (log of) CEO-to-worker pay ratio. "Exporter" is the exporter indicator computed from LFTTD. Exports are dollar values of shipments from LFTTD. The unit of observation is firm-year and year varies between 1992 and 2007. See Table B.1 for sector distribution of the sample. Standard errors are clustered at the year-sector level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

at the intensive margin of exporting in column 3 by replacing the indicator EXP_{it} with the log of firm exports. In that setting we find that a 1-percent increase in firm exports is associated with an 0.12-percent increase in within-firm inequality. Lastly, column 4 replaces sector-level fixed effects with firm-level fixed effects to identify the key correlation from the time-series variation in export status within a firm; we find the exporter premium on within-firm inequality drops drastically to 9.3% in that case, suggesting firm characteristics can explain a large proportion of the CEO-to-worker pay ratio. All the estimates are statistically significant at conventional levels. In appendix table B.5, we show that the strong positive correlation between within-firm inequality and export status holds for changes in data (e.g., private firms, other executives, different components and different measures of CEO compensation) as well as changes in the specification of the estimating equation (e.g., different clustering choices, different fixed effects, inclusion of firm-specific time trends).

Table 2.2 suggests that differences in firm size drive most of the differences in the CEO-to-worker pay ratio. We proceed in two steps. First, we show in panel A that larger firms have greater within-firm inequality. Specifically, we replace the exporter indicator in equation 2.1 with measures of firm size: employment and payroll in the United States. Column 1, focusing on the manufacturing sector, suggests that a 1% increase in the firm's employment is associated with a CEO-to-Worker pay ratio that is 0.39% larger. In the sample with all firms, the same 1% increase in employment also coincides with a 0.39% increase in our measure of within-firm inequality, as per column 2. The remaining two columns affirm that this size-inequality relationship is positive and robust when we use the firm's payroll as a measure of size. Second, in panel B we include in equation 2.1 both the exporter indicator and a measure of firm size. In each specification, the coefficient on the size measure is positive, statistically significant, and dwarfs in magnitude the coefficient on exporter status. Furthermore, the coefficients on exporter status are close to zero and only sometimes significant. For instance, including employment as a measure of size reduces the coefficient on exporter status in column 2 by an order of magnitude: column 2 now suggests that exporters have a 5.0% larger CEO-to-Worker pay ratio than non-exporters, while same coefficient was 50.7% in table 2.1 when we did not control for firm size. These patterns repeat throughout the table, both when we change the sample of firms, as well as when we replace employment with payroll as the measure of firm size. In appendix table B.6, we show that the results also hold for firm sales and asset holdings, measures of firm size from COMPUSTAT.[12]

[12]We focus primarily on measures of employment and payroll from the LBD because these measures reflect a firm's size in the United States. For COMPUSTAT, it is not clear whether a given data point reflects a firm's global or U.S.-based sales and assets; consequently, we relegate the analysis of COMPUSTAT size measures to the appendix.

Table 2.2: Within-Firm Inequality, Export Status, and Firm Size

Panel A: Within-Firm Inequality and Firm Size in the United States

Dependent Variable	Log CEO-to-Worker Pay Ratio			
	(1)	(2)	(3)	(4)
Log Employment	0.388***	0.391***		
	(0.010)	(0.007)		
Log Payroll			0.378***	0.370***
			(0.010)	(0.008)
Sample	Manufacturing	All	Manufacturing	All
Year Fixed Effects	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Sector
Observations	8,000	17,000	8,000	17,000
R-squared	0.376	0.407	0.362	0.385

Panel B: Within-Firm Inequality and Export Status, Controlling for Firm Size

Dependent Variable	Log CEO-to-Worker Pay Ratio			
	(1)	(2)	(3)	(4)
Exporter	-0.060	0.050*	0.049	0.070***
	(0.095)	(0.026)	(0.101)	(0.027)
Log Employment	0.389***	0.388***		
	(0.010)	(0.007)		
Log Payroll			0.377***	0.364***
			(0.011)	(0.009)
Sample	Manufacturing	All	Manufacturing	All
Year Fixed Effects	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Sector
Observations	8,000	17000	8,000	17000
R-squared	0.376	0.407	0.362	0.385

Note: The left-hand side variable for each of the regressions is the (log of) CEO-to-worker pay ratio. "Exporter" is the exporter indicator computed from LFTTD. Exports are dollar values of shipments from LFTTD. Employment is the total annual employment reported in LBD. Payroll is the total annual payroll reported in LBD. The unit of observation is firm-year and the time period spans 1992 through 2007. See Table B.1 for sector distribution of the sample. Robust standard errors are clustered at the year-sector level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

These exercises convey a consistent message: the "exporter premium" for within-firm inequality is driven by the size of the exporters. Larger firms have higher within-firm inequality, and the reason we observe higher within-firm inequality among exporters is precisely because those firms are larger – a stylized fact confirmed by the empirical trade literature that motivated the new generation of heterogeneous firms trade models.[13] These results suggest that within-firm inequality can be naturally incorporated into a Melitz trade model, where exporting behavior and size are linked.

The insignificance of exporting status conditional on size does not imply that trade is irrelevant for within-firm inequality. Without trade, many of the large firms in the sample would not have been able to grow to the size that we observe in the data. In a counterfactual world where all the firms can only sell to the domestic market, many of the large firms would be smaller and, thus, their within-firm inequality would be lower. The insignificance of the exporter dummy implies only that whatever effect trade might have on within-firm inequality, the main channel goes through the size of the firm. In some cases, the coefficient on the exporter dummy is significantly positive after controlling for size, indicating that there are other factors that predict higher within-firm inequality among exporters. For example, exporting firms might need different managerial skills than domestic firms and thus are recruiting their CEOs in a different market. However, as the size of the coefficients suggests, no matter what these factors are, their explanatory power is small relative to firm size. Therefore, the model presented in Section 2.3 focuses solely on the size of the firm and leaves the other factors to future research.

2.2.3 Within-Firm Inequality: Evidence from a Trade Shock

Having shown that within-firm inequality is higher for exporters than for non-exporters, we now provide causal evidence on this channel using a trade shock: China's 2001 accession to the World Trade Organization (WTO). Around the time of its accession, China agreed to lower its import tariffs. Following a 6-percentage-point tariff reduction to 17% between 1996 and 1997, China's average tariff rate remained stable until the WTO accession, whereupon the average tariff declined to 10% by 2007 (Lu and Yu, 2015). Moreover, when China became a WTO member on December 11, 2001, it received permanent, reciprocal access to the *most favored nation* status, which reduced policy uncertainty and the threat of trade wars (Handley and Limão, 2017). These changes in trade policy made China a more attractive destination for U.S. exporters and contributed to a rise in exports to China. We next show that the reductions in China's import tariffs primarily benefit the U.S. firms that export to the Chinese market, increasing their export sales, firm size, and within-firm inequality.

[13] For example, see Bernard and Jensen (1999).

To estimate the causal effects of this trade shock, we compare the outcomes of firms that had existing trading relationships with China prior to its WTO accession to firms that did not have China-specific trading relationships. Our starting point is a standard difference-in-differences regression specification, e.g., Angrist and Pischke (2008):

$$\begin{aligned} \text{Outcome}_{\text{it}} = \delta_0 + \mathbf{d'}_1 \cdot \mathbf{f} + \mathbf{d'}_2 \cdot \mathbf{y} \\ + \delta_1 \text{Treatment}_i \times \text{Post China WTO Accession}_t + \epsilon_{it}, \end{aligned} \tag{2.2}$$

where Outcome_{it} is firm i's exports, CEO-to-Worker pay ratio, or firm size in year t; $\mathbf{f}$ and $\mathbf{y}$ are respectively the vectors of firm and year fixed effects; Treatment_i is an indicator taking value one for firms that exported to China prior to the WTO accession, and zero otherwise; and, "Post China WTO Accession_t" is an indicator taking value one for years 2002 and onward, and zero otherwise. The coefficient of interest, δ_1, captures the differential impact of China's accession to the WTO for firms that had a China-specific trading relationships prior to the event.

We rely on the "parallel trends" assumption to give this comparison a causal interpretation. Specifically, we posit that the observed outcomes of the untreated firms after the shock parallel the counterfactual outcomes of the treated firms had the shock never taken place. The causal effect is then the difference between the change in outcome for the firms with a China-specific trade relationship and the change in outcome for the firms without such a relationship. In appendix B.2.2, we provide more details on the timing of the shock, the plausibility of the parallel trends assumption, and robustness for the results that follow. Note also that—unlike the cross-sectional comparison in the previous section—the estimating procedure we use here relies on *within-firm* variation in the CEO-to-worker pay ratio. Since the variation within firms is meaningfully smaller than the variation across firms, this procedure is more demanding of the data.[14]

China's accession to the WTO resulted in increased exports and higher within-firm inequality for firms that exported to China prior to 2001, as per panel A of Table 2.3. We begin by focusing on the manufacturing sector and in columns 1 and 2 we define firms as "treated" if they exported to China in the three years between 1998 and 2000.[15] For firms with this pre-existing China-specific relationship, China's accession to the WTO led to a 56.8% increase in exports and a 12.5% increase in the CEO-to-worker pay ratio. In

[14]The cross-sectional variance of $\ln$(CEO-to-Worker pay ratio) is 1.205; the within-firm variance is roughly half the size at 0.659.

[15]We choose a three-year window based on the work of Monarch and Schmidt-Eisenlohr (2016), who document that import-export relationships in the United States that last at least three years account for 47% of the value of trade. We show in appendix B.2.2 that shifting the treatment window to include 2001, the year of the WTO accession, does not change our conclusions.

Table 2.3: China's Accession to the World Trade Organization and Within-Firm Inequality

Panel A: Exports and CEO-to-Worker Pay Ratio

Dependent Variable (log)	Exports	CEO-to-Worker Pay Ratio	Exports	CEO-to-Worker Pay Ratio
	(1)	(2)	(3)	(4)
Treatment × Post China WTO Accession	0.568***	0.125**	0.768***	0.076*
	(0.120)	(0.060)	(0.112)	(0.043)
Treatment	Exporter to China 1998-2000		Exporter to China 1998-2000	
Sample	Manufacturing		All	
Year Fixed Effects	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Observations	8,000	8,000	13,000	17,000
R-squared	0.908	0.714	0.932	0.755

Panel B: Employment and Payroll

Dependent Variable (log)	Employment	Payroll	Employment	Payroll
	(1)	(2)	(3)	(4)
Treatment × Post China WTO Accession	0.385***	0.406***	0.505***	0.500***
	(0.067)	(0.067)	(0.061)	(0.062)
Treatment	Exporter to China 1998-2000		Exporter to China 1998-2000	
Sample	Manufacturing		All	
Year Fixed Effects	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Observations	8,000	8,000	17,000	17,000
R-squared	0.926	0.917	0.937	0.922

Note: The left-hand side variable for each of the regressions is the (log of) CEO-to-worker pay ratio. "Exporter" is the exporter indicator computed from LFTTD. Exports are dollar values of shipments from LFTTD. Employment is the total annual employment reported in LBD. Payroll is the total annual payroll reported in LBD. The unit of observation is firm-year and the time period spans 1992 through 2007. See Table B.1 for sector distribution of the sample. Robust standard errors are clustered at the year-sector level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

columns 3 and 4 we expand the sample to cover firms in all sectors, and we estimate a 76.8% increase in exports and a 7.6% increase in within-firm inequality.

The pass-through of trade shocks to within-firm inequality takes place through firm size. Complementing the evidence in panel A that exports increased, panel B shows that China's accession to the WTO also led to increases in employment and payroll for firms with China-specific relationships. Columns 1 and 2 focus on the manufacturing sector and document that China's accession resulted in roughly 40% increases in both employment and payroll for those firms. Maintaining the same 1998-2000 treatment window and expanding the sample to all firms, as in columns 3 and 4, increases these estimates to roughly 50% increases in employment and payroll for the treated firms. For both sets of firms, the trade shocks increase employment and payroll by roughly the same magnitude, suggesting that the average firm wage is not responsive to the trade shocks. This in turn implies that the observed surge in CEO-to-worker pay ratio is driven by the fact that the trade shocks primarily increase executive compensation but not the average wage.

Taking the ratio of our estimates for trade-driven changes in inequality and firm size, we derive an implied elasticity: a 1% increase in firm size leads to a 0.2-0.3% increase in within-firm inequality and executive compensation. This range falls slightly below the elasticity of 0.4 that we estimated by regressing log size on log inequality directly in panel A of table 2.2. As we argued above, the workers' wages are unaffected by the trade shocks; hence, this elasticity also describes how CEO compensation changes with firm size, a relationship that is known as Roberts' Law (Roberts, 1956). In fact, Gabaix (2009) lists the standard range of Roberts' Law elasticities as 0.2-0.4, which encompasses our trade-driven estimates.

2.3 The Model

In this section, we build a model of heterogeneous firm in which international trade shapes within-firm inequality by changing firm size. The model setup is based on Helpman, Melitz and Yeaple (2004). We introduce occupational choice and executive compensation into the framework. The contribution of the model is two-fold. First, it offers a tractable framework to analyze the effects of trade on the CEO-to-worker pay ratio within each firm and overall income inequality. Second, the simple framework is also empirically relevant: it is able to generate income distribution and firm-size distribution with full support that closely resemble the data. Within this framework, we then carry out a quantitative analysis to evaluate the impacts of globalization on income inequality, both within the right-tail of the income distribution, and between the right-tail and the general population.

2.3.1 Model Setup

The model world consists of two countries indexed by i. Each country i is populated by individuals with measure n_i. People in each country are endowed with human capital x. As x uniquely identifies each individual, with a slight abuse of notations, we also use x as index for individuals within a country. The distribution of human capital in each country follows an exponential distribution with shape parameter λ. The cumulative distribution function (CDF) of human capital is as follows:

$$F(x) = 1 - e^{-\lambda x}.$$

We use the exponential distribution, together with other assumptions explained later in this section, to capture the structure and shape of the income distribution and the firm size distribution at the same time. We characterize the distributions in more detail when we present the analytical results in section 2.4.

An individual can choose between two careers. She can either work for an existing firm or she can create a new firm. If she chooses to be a worker, then her human capital directly translates into the amount of efficiency labor that will be inelastically supplied to the market. In this case, the individual's income will be $w_i x$, where w_i is the prevailing wage rate per efficiency unit of labor in country i. Individuals cannot move between countries and the wage rate w_i is determined in a country-wide competitive labor market.

The individual can also create a new firm to start producing a new variety of good. In doing so she becomes the founder and CEO of the firm. The productivity of the firm, denoted by $A_i(x)$, depends on the human capital of the founder and takes the following form:

$$A_i(x) = b_i e^x, \tag{2.3}$$

where b_i is the total factor productivity (TFP) in country i. With the assumption on the distribution of x, the above function implies that firm productivity, A_i, follows a Type-I Pareto distribution with location parameter b_i and shape parameter λ (see appendix for the proof). Subsequently this also implies that firm sales, employment, and profit distributions will also be Paretian.

The payoff to the founder and CEO of the firm is a function of the profit of the firm, denoted as $k(\pi) \leq \pi$, where π is the profit. For simplicity, we assume that the residual profit after the CEO compensation is distributed back to the entire population in country i evenly (i.e. all the people in the country own the firms through a mutual fund). This assumption does not affect the analysis of income inequality, and the main results of the paper do not change meaningfully if we relax this assumption.

Reflecting Roberts' Law (Roberts, 1956) from the corporate governance literature, we assume that $k(\pi)$ is exogenously determined, monotonically increasing, and regularly-varying in π. By definition, a function $k(\pi)$ is regularly-varying with tail index β if and only if for any $z > 0$, the following relationship holds:

$$\lim_{\pi\to\infty} \frac{k(z\pi)}{k(\pi)} = z^{\beta}.$$

Intuitively, regularly-varying functions are functions that behave like power functions at the limit.[16] In our context, the assumption of regular variation delivers the empirically robust Roberts' Law so CEO compensation is proportional to a power function of firm profits asymptotically. The assumption of regular variation also implies that the right-tail of the income distribution will exhibit Paretian behavior, though the vast majority of the distribution follows an exponential distribution — again, an empirically relevant result, which will be discussed in detail in the next section.

We show in appendix B.5 how these stylized, tractable features of the labor market emerge from richer microfoundations. In the extended model, CEOs and firms match in the market, and endogenously determine a compensation function for managerial talents, $k(x)$, much like in Gabaix and Landier (2008). In equilibrium, CEOs with higher talents will be matched with firms with higher productivity, and thus the compensation function will be monotonically increasing and regularly varying in both the managerial talent and the size of the firm. Positive assortative matching also implies that the matching pattern between CEOs and firms in equilibrium will be the same as if the best managers founded the best firms and remained as CEOs thereafter, the assumption that we relied on in the baseline model. As the endogenous labor market for CEOs delivers compensation functions and matching patterns identical to those exogenously assumed, the extension can be considered orthogonal to other parts of the model. For this very reason, we abstract away from a full-fledged labor market and compensation model for the CEOs in the baseline model, and refer the readers to appendix B.5 for more details.

The production side of the economy is modeled after Melitz (2003), with firms that are heterogeneous in their productivity $A_i(x)$ each producing a single variety of a good indexed by x. Each firm produces a quantity $q_i(x)$ of its variety using the following production function:

$$q_i(x) = A_i(x) \cdot [L_i(x) - f_{ii}],$$

[16] For more details, see Resnick (1987).

where $L_i(x)$ is the labor demand and f_{ii} is the fixed cost of production, paid in the units of labor of country i. The firms operate in a monopolistically-competitive market and earn positive profit in equilibrium.

Firms in country i can export to country j by paying a fixed cost f_{ji}, denominated in units of labor, to set up the distribution network. Trading incurs an iceberg cost of $\tau_{ji} > 1$: in order to supply country j with one unit of good from country i, the firm needs to ship τ_{ji} units.

Individuals in country i consume a CES aggregate of all the varieties available in country i. Their utility function is as follows:

$$U_i = \left(\int_{m \in \Theta_i} q_i(m)^{\frac{\epsilon-1}{\epsilon}} dm \right)^{\frac{\epsilon}{\epsilon-1}},$$

where ϵ is the elasticity of substitution, and Θ_i is the set of goods that are available in country i.

2.3.2 Solution and Equilibrium Conditions

The solution to the firm's problem is similar to Melitz (2003). Denote the total spending in country i as H_i and the ideal price index as P_i. The maximum profit a firm in country i can earn in its domestic market is:

$$\pi_{ii}(x) = \frac{H_i}{\epsilon} \left[\frac{\epsilon - 1}{\epsilon} \frac{P_i}{w_i} \right]^{\epsilon-1} A_i(x)^{\epsilon-1} - f_{ii} w_i.$$

The *additional* profit a firm in country i can earn from exporting to country j is:

$$\pi_{ji}^e(x) = \frac{H_j}{\epsilon} \left[\frac{\epsilon - 1}{\epsilon} \frac{P_j}{\tau_{ji} w_i} \right]^{\epsilon-1} A_i(x)^{\epsilon-1} - f_{ji} w_i, \tag{2.4}$$

The details of the solution to the firm's problem can be found in Appendix B.3.

Similar to Melitz (2003), under some loose parameter restrictions, firms sort into two groups. All the firms founded in country i serve the domestic market first. Moreover, the least productive firms only serve the domestic market. The more productive firms serve the domestic market and the foreign market through export. Denote the human capital of the founder of the least productive exporting firm in country i as x_{ji}^e, the cutoff must be the solution to the following equation respectively:

$$\pi_{ji}^e(x_{ji}^e) = 0. \tag{2.5}$$

The condition means that the marginal exporter earns zero profit from exporting.

The solution of the occupational choice problem is a single cutoff rule. There exists a human capital level x_i^* in country i such that all the individuals with human capital smaller than x_i^* choose to be workers and all the other individuals choose to create firms. The cutoff x_i^* is the solution to the following equation:

$$k(\pi(x_i^*)) = w_i x_i^*, \tag{2.6}$$

which requires that in equilibrium the founder of the marginal firm to be indifferent between creating a new firm or working for an existing firm. The sufficient and necessary condition for the existence of the solution is that $k(\pi_i(0)) < 0$, which means that the individual with the least amount of human capital must find creating a new firm unprofitable.

Figure 2.1 presents the solution in a simple setting where $k(\pi) = \pi$. The solid line is the income of a worker as a function of his/her human capital. The dashed line is the income of a CEO as a function of his/her human capital. Under the assumption that $k(\pi(0)) < 0$, the two curves cross once and only once at the cutoff human capital level x_i^*.

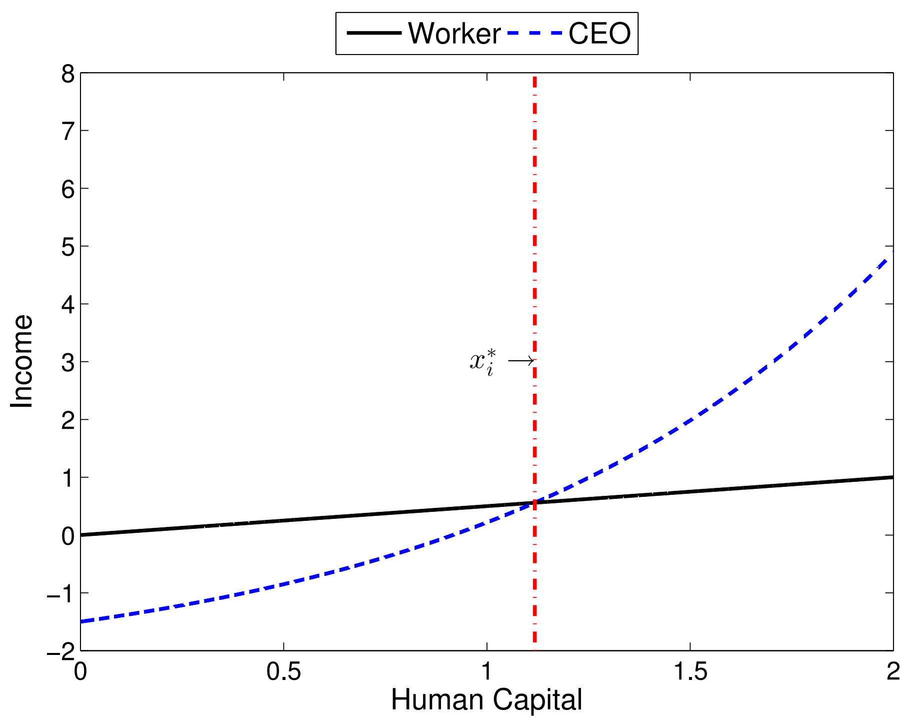


Figure 2.1: Solution of the Occupational Choice Problem

The graph plots the solution of the occupational choice problem. The black solid line is the income of a worker, and the blue dashed line is the income of a CEO. The vertical line indicates the cutoff human capital that is indifferent between being a worker or a CEO. This graph assumes that $k(\pi) = \pi$.

The equilibrium of the world economy is a vector of wages, $\{w_i\}$, a vector of the occupational choice cutoffs $\{x_i^*\}$, a vector of exporting cutoffs $\{x_{ji}^e\}$, a vector of ideal price levels $\{P_i\}$, and a vector of total expenditures $\{H_i\}$ such that for $i = 1, 2$ and $j = 1, 2$:

1. Every individual in country i maximizes their income by solving the occupational choice problem (equation (2.6) holds).

2. Every firm optimally chooses to be a non-exporter or exporter, (equations (2.5) holds).

3. Total income equals to total expenditure in each country:

$$H_i = n_i w_i \int_0^{x_i^*} x f_i(x) dx + n_i \int_{x_i^*}^{\infty} \pi_i(x) f_i(x) dx. \tag{2.7}$$

4. Aggregate price level and the individual prices satisfy the rational expectation condition:

$$P_i = \left(\int_{m \in \Theta_i} p(m)^{1-\epsilon} dm \right)^{\frac{1}{1-\epsilon}}. \tag{2.8}$$

5. Labor market clears in each country.

Equation (2.7) is the income-expenditure identity in country i. In equilibrium, the total expenditure in country i must equal the total income in country i, which is the sum of all the wage and profit income[17]. Equation (2.8) is the definition of the ideal price index, which is the cost of one unit of utility in country i. Appendix B.3 provides the details on these two equilibrium conditions, as well as the details on the labor market clearance condition.

2.4 Analytical Results

2.4.1 Firm Size Distribution and Income Distribution

As we detailed in the previous section, the distribution of firm productivity arises from the distribution of entrepreneurs' human capital; this productivity distribution in turn leads to plausible and tractable distributions of firm sales, employment and profits. Specifically,

[17]The CEO compensation function does not enter the total income function, because the difference between profit and CEO compensation at a given firm will be distributed back to the individuals in country i, which implies that we only need to consider total profit when accounting for total income in a given country.

the productivity distribution in country i follows a Type-I Pareto distribution with shape parameter λ and location parameter $b_i e^{x_i^*}$, where x_i^* is the marginal entrepreneur. Firm sales are linear functions of $A^{\epsilon-1}$. As a result, the distribution of sales follows a Type-I Pareto distribution with shape parameter $\lambda/(\epsilon-1)$. Moreover, as noted in di Giovanni, Levchenko and Rancire (2011), international trade systematically changes the size distribution of firms. In our framework, this influence of international trade on the distribution is reflected in the location parameters: the location parameters are small for domestic and large for exporting firms. Firm employment and profit are affine functions of $A^{\epsilon-1}$ due to the fixed costs of operating and exporting. They follow Type-II Pareto distributions with shape parameter $\lambda/(\epsilon-1)$. As in the distribution of sales, location parameters in the distributions of employment and profits vary by the market size accessible to a firm. Appendix B.3 provides details on the distributions of firms.

Individual income is ranked by occupations: the workers earn the lowest income, followed by the CEOs at domestic firms, and the CEOs at exporting firms. The income distribution follows a two-class structure. All the workers earn the same wage rate per efficiency labor unit; therefore, their income distribution is exponential with a shape parameter λ/w_i. The income of the CEOs depends on the CEO compensation function. By assumption, the compensation function $k(\pi)$ is monotonically increasing in π and regularly varying. Under these two assumptions, the income distribution of the CEOs adopts the following CDF:

$$U(y) = 1 - y^{-\frac{\lambda}{\beta(\epsilon-1)}} R(y), y > 0,$$

where y is the income, β is the tail index of $k(\pi)$, $\frac{\lambda}{\beta(\epsilon-1)}$ is the shape parameter of the distribution, and $R(y)$ is a slowly-varying function.[18] Distributions with this form of CDF are Pareto-Type distributions and exhibit fat-tail behavior at the right end similar to Type-I Pareto distributions. Appendix B.3 provides details on the derivation of the income distributions of different groups of individuals.

2.4.2 Partial Equilibrium

The main mechanism of the model is most clearly demonstrated in partial equilibrium with wages, prices and total expenditures fixed at their autarky levels: following an opening to trade, the most productive firms export, grow larger, and the compensation of the exporting CEOs far outpaces the domestic wages, leading to increased inequality.

In figure 2.2 we present these partial equilibrium results for a simplified model where the CEO compensation equals profits, $k(\pi) = \pi$. The black solid line and the blue dashed

[18] Slowly-varying functions are regularly-varying functions with tail index of 1. Intuitively, slowly-varying functions are functions that behave like linear functions at the limit.

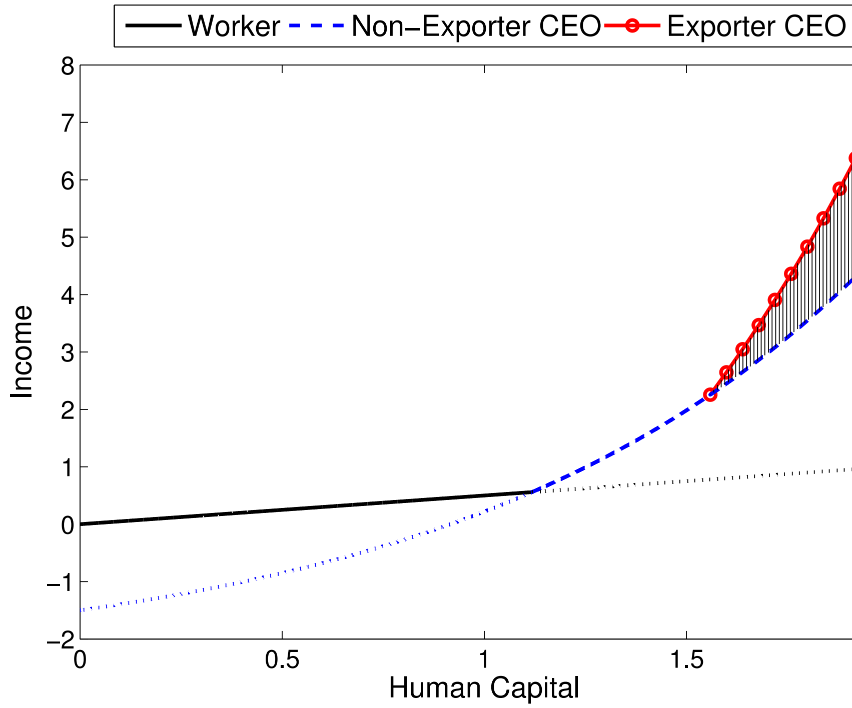


Figure 2.2: Trade and Top Income Shares in Partial Equilibrium

This graph plots the income of different individuals against their human capital for different occupations under autarky and under trade. The black solid line is the income of a worker. The blue dashed line is the income of CEOs at non-exporting firms. The red circled line is the income of CEOs at exporting firms. The shaded area is the extra profit earned from exporting. This partial equilibrium assumes that $k(\pi) = \pi$ and that wage, total expenditure, and prices are all fixed. It also abstracts away from FDI.

lines are the same as in figure 2.1: they are respectively the incomes of workers and CEOs in autarky for the home country. When the world opens up to trade, only the most productive firms export. In the graph, the right end of the CEO income function tilts up into the red circled line, which is the income of CEOs at the exporting firms. The shaded area between the red circled line and the blue dashed line is the extra profit (and extra compensation to the CEO) earned in the foreign country. In this simple case, all the benefits of globalization are claimed by the CEOs at the exporting firms, and none of the benefits trickle down to the workers in those firms. At the aggregate level, top income shares will be higher because the CEOs at the exporting firms are originally the richest people in autarky.

2.4.3 General Equilibrium

The main mechanism discussed in partial equilibrium above persists in general equilibrium as well. We first present a simple result characterizing the cross-sectional intra-firm inequality of the model in general equilibrium:

Proposition 1 *If the set of exporting firms in country i is non-empty, then the average CEO-to-worker pay ratio among domestic firms is strictly smaller than the average CEO-to-worker pay ratio among exporting firms.*

Proof 1 *See Appendix B.4.*

Proposition 1 replicates in general equilibrium the empirical findings from section 2.2. If an econometrician observes the model world and estimates equation (2.1) without any size control, she will find that the CEO-to-worker pay ratio is significantly higher among firms that sell to the foreign market than those who do not. In addition, in general equilibrium, the CEO-to-worker pay ratio is proportional to the size of the firm. Therefore, if the econometrician can also observe the size of the firm and controls for it when estimating equation (2.1), the observed between-group difference will disappear, just the same as we observed in the U.S. data.

In the next proposition, we show that the cutoff points of human capital among different groups of firms — x_i^* for firms that produce domestically and x_{ji}^e for exporters — are sufficient statistics for the profit-to-wage ratios, which in turn shape the within-firm inequality measured as CEO compensation relative to worker wages.

Proposition 2 *In general equilibrium, the domestic-profit-to-wage ratio, defined as*

$$\frac{\pi_{ii}(x)}{w_i} = \frac{H_i}{w_i \epsilon} \left(\frac{P_i}{w_i} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} A_i(x)^{\epsilon - 1} - f_{ii},$$

will be lower when x_i^* is higher; The exporting-profit-to-wage ratio, defined as

$$\frac{\pi_{ji}^e(x)}{w_i} = \frac{H_j}{w_i \epsilon} \left(\frac{P_j}{\tau_{ji} w_i} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} A_i(x)^{\epsilon - 1} - f_{ji},$$

will be lower when x_{ji}^e is higher.

Proof 2 *See Appendix B.4.*

Intuitively, as the trade costs τ_{ji} decrease, bilateral trade between i and j increases. The increased access to a foreign market makes exporting profitable for more firms, lowering the productivity cutoff for exporting x_{ji}^e. Proposition 2 then establishes that the exporting-profit-to-wage ratio will be higher among all the exporters as a result of the lower trade costs. Consequently, those whose income is linked to the profit of the firm—the top executives—will see their income increasing faster than the income of the workers. Trade liberalization also puts competitive pressures on the least productive domestic firms, leading them to exit and leading the domestic-production cutoff x_i^* to rise. Higher x_i^* in turn leads to lower CEO-to-worker pay ratio among the domestic firms. This proposition suggests that the top executives at the exporting firms stand to benefit from from trade liberalization, which might lead to higher top income shares at the aggregate level. We formally establish this result in the next proposition.

Proposition 3 *In a symmetric-country setup of the model, if $\tau_{ji} = \tau_{ij}$ drops, then there exists a percentile $p^* \in (0, 100)$ such that the top p^*-percent income share will be higher. Specifically:*

$$p^* = 100 \times (e^{\lambda x_{ji}^e}),$$

where x_{ji}^e is the exporting cut-off before the changes in τ_{ji} and τ_{ij}

Proof 3 *See Appendix B.4.*

Proposition 3 establishes that bilateral trade liberalization leads to higher income concentration at the right tail of the income distribution. Put differently, CEOs of exporting firms benefit more from trade than both the CEOs of domestic firms and the workers. As the CEOs of the exporting firms were already richer than the other groups of people before the trade liberalization, lower $\tau_{ij} = \tau_{ji}$ will lead to higher income shares at the top. Outside of a symmetric country setup, analytical results on top income shares are difficult to establish; for this reason, we next turn to the quantification of our model to study the relationship between globalization and top income shares.

2.5 Quantitative Analysis

In this section, we quantify the impact of trade liberalizations on top income inequality in general equilibrium. We first extend the benchmark model to incorporate multinational firms (MNEs), a choice motivated both by the relevance of MNEs for understanding international flows and by the fact that their CEOs are also well-paid relative to their workers, as per appendix table B.5. We then calibrate the model to resemble the U.S. economy in the 2000s, and show that the model provides a reasonably good approximation for the U.S. income distribution. We then study how different measures of income inequality respond to changes in trade barriers, and show that globalization might be responsible for a substantial part of the surge in top income shares in recent decades. In the end we show that the main results of the model are robust to changes in certain parameter values.

2.5.1 Multinational Firms

As multinational firms are important players in both international trade and capital flows, we first extend the model to allow for MNEs before we carry out the quantitative analysis. In addition to exporting, the firms in country i can also serve country j via horizontal foreign direct investment (FDI), as in Helpman, Melitz and Yeaple (2004). In order to serve country j from country i through FDI, the firm needs to pay the fixed overhead costs g_{ji} in units of labor in country i. The labor costs are interpreted as the overhead costs of starting operation, as well as the costs introduced by policy barriers. The *additional* profit a firm in country i can earn from FDI in country j is:

$$\pi^f_{ji}(x) = \frac{H_j}{\epsilon}\left[\frac{\epsilon-1}{\epsilon}\frac{P_j}{w_j}\right]^{\epsilon-1} A_i(x)^{\epsilon-1} - g_{ji}w_i. \tag{2.9}$$

Subject to some standard parameter restrictions, firms sort into three groups in the extended model. All the firms founded in country i serve the domestic market first. Moreover, the least productive firms only serve the domestic market. The more productive firms serve the domestic market and the foreign market through export. The most productive firms serve the domestic market and the foreign market through FDI. Denote the human capital of the CEO in the least productive MNE in country i as x^f_{ji}. The cutoff between exporters and MNEs must be the solution to the following equation:

$$\pi^e_{ji}(x^f_{ji}) = \pi^f_{ji}(x^f_{ji}), \tag{2.10}$$

which says that the marginal MNE finds it equally profitable to serve the foreign market by FDI and by exporting.

Through proposition 4, we show that the cutoff x^f_{ji} for serving the foreign market through FDI is a sufficient statistics for the ratio of FDI profits to the wage ratio:

Proposition 4 *In general equilibrium, the FDI-profit-to-wage ratio, defined as:*

$$\frac{\pi^f_{ji}(x)}{w_i} = \frac{H_j}{w_i\epsilon}\left(\frac{P_j}{w_j}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} A_i(x)^{\epsilon-1} - g_{ji},$$

will be lower when x^f_{ji} is higher.

Proof 4 *See Appendix B.4.*

As with the cutoffs for domestics production and exporting, we can use this relationship to understand the impact of FDI liberalization on inequality by tracking the changes in x^f_{ji}.

2.5.2 Calibration

We interpret the two countries in the model world as the United States and the rest-of-the-world (ROW). We treat 109 economies combined as the ROW. These countries, together with the U.S., are responsible for around 74 percent of the world population and 82 percent of the world GDP in 2008. The selection of countries is due to data availability, and the countries included in ROW are reported in Table B.8.[19]

In our measure of population, n_i, we want to account for differences in worker productivity across countries, difference that arise from the variation in both human capital and the physical capital associated with each worker. As a result, we follow the methods outlined in Caselli (2005) and first compute the "quality-adjusted workforce" using the Penn World Table 7.0 and the educational attainment data from Barro and Lee (2010). We then augment this measure of total workforce with the estimated capital stock and arrive at the final measure of the size of "population." For details on the population measures, see Appendix B.6.

With this measure of population, we then calibrate the country TFP b_i to match the relative size of the United States and the ROW. We normalize U.S. TFP to 1 and report the calibrated b_i for ROW in Table B.9. Furthermore, we set the elasticity of substitution to 4 so that the average markup charged by firms is 33 percent. This level of mark-up is in the middle of plausible estimates, and we provide robustness checks with ϵ between 2 and 6 in a later section.[20] The shape parameter of the human capital distribution, λ, is set to 3.18. This implies that the Pareto shape parameter of the firm employment distribution is $\lambda/(\epsilon - 1) = 1.06$, the estimation provided by Axtell (2001).

We calibrate the fixed costs of operation and export using the Doing Business database from the World Bank following the methods outlined in di Giovanni and Levchenko (2012*a*, 2013). Specifically, we use the days of starting a business in the U.S. as the raw measure of the fixed costs of operation in the home country. The fixed costs of operation in the ROW are the average across the rest 109 countries weighted by GDP. We use the *Trading across Borders* module of the Doing Business Indicators database to measure the fixed costs of international trade. Define ϕ_{ij} as the sum of days required to export a 20-foot dry-cargo container from country i and to import the same kind of container into country j. The

[19] A country is included in the sample if and only if its data from 1988 to 2008 are available both in Penn World Table 7.0 and Barro and Lee (2010).

[20] For example, Domowitz, Hubbard and Petersen (1988) estimated the average markup for U.S. manufacturing firms to be 0.37. Rotemberg and Woodford (1991) used steady-state markups between 0.2 and 0.6, while Feenstra and Weinstein (2010) estimated the average markup to be 0.3 in 2005 in the U.S. The elasticity of substitution used here is slightly lower than the estimates based on gravity equations, which are usually between 5 and 10, as reported by Anderson and van Wincoop (2004). Robustness checks show that the main results of the paper hold true with higher levels of ϵ.

fixed cost of exporting from the U.S. to the ROW is computed as the weighted average of $\phi_{\mathrm{i,US}}, i = 1, 2 \cdots, 109$:

$$f_{21} = \frac{\sum_{i=1}^{109} E_{\mathrm{i,US}} \cdot \phi_{i,US}}{\sum_{i=1}^{109} E_{\mathrm{i,US}}},$$

where the weight, $E_{\mathrm{i,US}}$, is the export from the U.S. to country i. Similarly, the fixed cost of export from the ROW to the U.S. is:

$$f_{12} = \frac{\sum_{i=1}^{109} E_{\mathrm{US,i}} \cdot \phi_{US,i}}{\sum_{i=1}^{109} E_{\mathrm{US,i}}},$$

where the weight, $E_{\mathrm{US,i}}$, is the export from country i to the U.S. The f_{ij} matrix at this stage is measured in the unit of time. At the end, we normalize the entire f_{ij} matrix so that around 0.83 percent of individuals in the U.S. choose to create firms. This statistics matches the ratio of chief executives to working population in 2000 Public Use Microdata Series (PUMS) 5 percent sample obtained from IPUMS.[21]

To capture differences in ownership structures of firms, we use the following functional form of $k(\pi)$ as CEO compensation:

$$k(\pi) = \begin{cases} \pi & \text{if} \quad \pi \leq \alpha \\ \alpha^{1-\beta}\pi^{\beta} & \text{if} \quad \pi > \alpha \end{cases}, \tag{2.11}$$

This function is monotonically increasing in π and regularly varying; therefore, all the analytical results in Section 2.4 carry over. Intuitively, the function captures the idea that firms with profit less than or equal to α are "sole proprietorship" firms: the founder and CEO owns the firm and claims all the profit. Firms with profit larger than α are "corporations," and the founder can only claim a proportion of the profit. The power function form for larger firms implies that the right tail of the income distribution follows a Pareto distribution with tail index $\frac{\lambda}{(\epsilon-1)\beta}$.

As noted in Section 2.3, equation (2.11) is based on the empirical findings in the literature that CEO compensation is proportional to the power function of the firm size, $k \sim \pi^{\beta}$, otherwise known as the "Roberts law" (Roberts, 1956). This function also arises naturally as an equilibrium compensation function from a matching model where the managers with higher ability are matched with larger and more productive firms in equilibrium. Specifically, this function is a special case of the duo-scaling equation in Gabaix and Landier

[21]This statistics measures chief executives, not self-employed to working population ratio, which around 10.9 percent as reported in Hipple (2010). See Ruggles et al. (2010) for details.

(2008), where α is the size of the reference firm. Within the context of this paper, the reference firm is the smallest corporation in each country. The calibration strategy described below ensures that the smallest firm in the model is always smaller than α in the benchmark model. This further implies that both types of firms exist in equilibrium.

We calibrate the ownership threshold α to match the ratio of sales of all the corporations to the sales of all the firms; this ratio is 62 percent in the U.S. in 2007.[22] We calibrate β to match the right tail index of the U.S. income distribution. Drăgulescu and Yakovenko (2001*b*) documents that the Pareto index of the U.S. income distribution is around 1.7. This implies that in this model, conditional on the tail index of the firm size distribution, β is $\frac{1.06}{1.7} \approx 0.747$.

We impose an upper bound, s, on the human capital distribution to eliminate unrealistically large corporations. We calibrate s to match the highest CEO-to-worker pay ratios in the data. We first compute the ratio between the highest CEO compensation in Execu-Compustat and the average U.S. wage from national income and product accounts (NIPA) in each year between 1992 and 2007[23]. We then set $s = 3.249$ so that the same ratio in the model is matched to the median of the data sequence, which is around 2,903.

We assume that both the iceberg trade costs and the fixed costs of starting foreign subsidiaries are symmetric: $\tau_{12} = \tau_{21}$ and $g_{12} = g_{21}$. We then jointly calibrate the two cost parameters, $\{\tau_{21}, g_{21}\}$, to match the exports-to-GDP ratio and the multinational-firm-sales-to-GDP ratio in the U.S. in year 2008. The first moment condition can be directly estimated using GDP data from NIPA. The second moment condition come from the Bureau of Economic Analysis's *Direct Investment and Multinational Corporations* data set.[24] These two parameters have to be jointly calibrated because iceberg trade costs affect not only the volume of trade but also the multinational sales through the extensive margin. Similarly, the fixed costs of FDI affect the volume of trade as well through the extensive margin. At the end we have $\tau_{21} = 1.720$ and $g_{21} = 1020$. All the above parameters are reported in Table 2.4.

[22]The sales of U.S. firms by legal form come from the *Statistics of U.S. Businesses, 2007* from the Census Bureau. The definition of "corporation" in this paper follows the legal form of "corporation" used by the Census. The other legal forms in the Census definition are classified as "proprietorship", which includes "S-corporations", "tax-exempt corporations", "partnership", "sole proprietorship", "other" and "tax-exempt other". The receipts of "government" are subtracted from the total firm sales.

[23]The wage data comes from NIPA Table 6.6A-D. The census does not allow disclosure of extreme values (maximum and minimum) that involve confidential data. Therefore we use the ratio between CEO compensation and the average U.S. wage instead of the CEO-to-worker pay ratio at the firm level in the empirical part.

[24]We use "All non-bank foreign affiliates" sales data up to 2008 as the estimate for the sales of multinational firms.

Parameter	Benchmark	Target/Source
λ	3.81	Axtell (2001)
ϵ	4.0	Average mark-up
α	23.2	Corporate sales as a percentage of all firms sales
β	0.747	Tail index of income dist., Drăgulescu and Yakovenko (2001*b*)
f_{11}	6.0	World Bank Doing Business Index
f_{12}	19.6	World Bank Doing Business Index
f_{21}	24.6	World Bank Doing Business Index
f_{22}	38.9	World Bank Doing Business Index
f-Scale	0.756	Percentage. of chief execu. in work force.
n_{ROW}	6.0	Caselli (2005), Barro and Lee (2010)
b_{ROW}	0.58	Relative country size
s	3.249	Highest-CEO-to-average-wage ratio among public firms
τ	1.720	Export-GDP ratio in 2008
g	1020	Multinational-sales-GDP ratio in 2008

Table 2.4: Calibration Targets and Results

Note: λ is the shape parameter of the exponential distribution. ϵ is the elasticity of substitution in the utility functions. α is the size of the smallest public firm. β is the tail index of the compensation function. f_{ij} is the fixed cost of exporting from country j to country i. f-Scale is the normalizing factor of the entire f_{ij} matrix. We divide the f_{ij} matrix by this number. n_i is the measure of capital-adjusted endowment of human capital in country i. b_i is the TFP in country i. s is the upper bound of human capital distribution. See Section 2.5.2 and the appendix for the details of calibration. See Table B.9 for the calibrated values of τ, g and TFP by year used in the counter-factual.

2.5.3 Model Fit

Even though it is calibrated to the *tail* index of U.S. income distribution, the model generates a good fit for the *overall* U.S. income distribution in general equilibrium. Figure 2.3 compares the model-generated income shares with the data in 2008.[25] The model provides a good approximation of the U.S. income distribution for the right tail. For example, the top 0.01 percent income share is 3.4 percent in the data and 4.2 percent in the model in 2008. The top 5 percent income share is 33.8 percent in the data and 29.6 percent in the model. Outside of the top income decile, the model also captures the overall shape of the income distribution reasonably well. The top 25 and 50 percent income shares in the model and the data only differ within a few percentage points. Overall, the difference between the model and the data for the top income shares reported in Figure 2.3 is around 7.8 percent when measured in Euclidean 2-norm.

[25]The data for income shares above the top 10 percent come from the updated Table A.1 in Piketty and Saez (2003). The income share outside of the top 10 percent comes from the Tax Foundation report (Greenberg, 2017), which is in turn based on IRS tax return data.

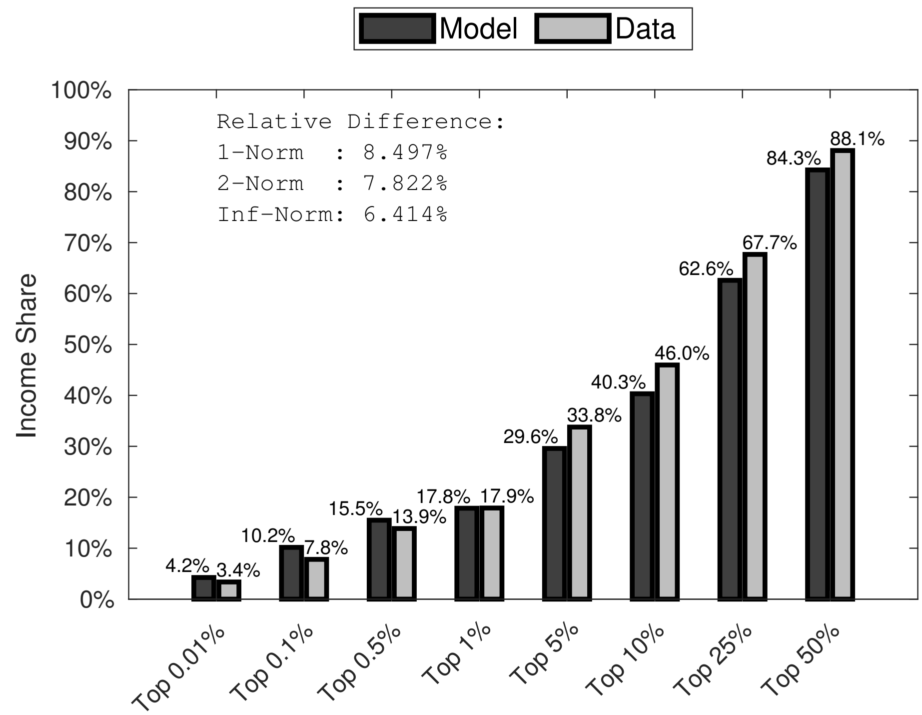


Figure 2.3: Top Income Shares: Model vs. Data (2008)

Note: This graph compares the top income shares between the model and the data in 2008. The top income shares in the model are described by the dark grey bars and those in the data described by light gray bars. The parameters behind the model simulation can be found in Section 2.5.2. The source of data is the updated Table A.1 from Piketty and Saez (2003). The average difference between the model and the data across the six top income shares is measured in Euclidean 2-norm. The differences are reported in percentage terms.

The model also compares favorably to the other moments of the data not targeted in the calibration; table 2.5 presents these comparisons for the mean-median ratio of the U.S. income distribution, the workers' share of income, and the CEO-to-worker pay ratio. The mean-to-median ratio of the U.S. economy in the model is 2.79, and the counterpart in the data is 1.61(Rodriguez et al., 2002). The second row in table 2.5 compares the workers' share of income in the model and the data. In the model the corresponding statistics is computed as the total wage payment to workers (CEO not included) divided by total output. In the data the statistics is computed as wage compensation divided by the gross domestic income of the private sector.[26] Again, the model closely resembles the data: the workers' share of income is 0.76 in the model, and 0.71 in the data. The last row compares the CEO-to-worker pay ratio. The model statistics is computed for the sample of "public"

[26]These data come from NIPA table 1.10. The gross domestic income of the private sector is defined as compensation of employees plus net operating surplus of private enterprises.

firms whose profit is higher than α. The counter-part in the data is based on the dataset described in Section 2.2. In the model the CEO-to-worker pay ratio is 95 while in the data it is 89.

Moments	Model	Data
Mean-to-median ratio, income	2.70	1.61
Workers' share of income	0.760	0.711
CEO-worker pay ratio	95	89

Table 2.5: Model Fit, Additional Measures

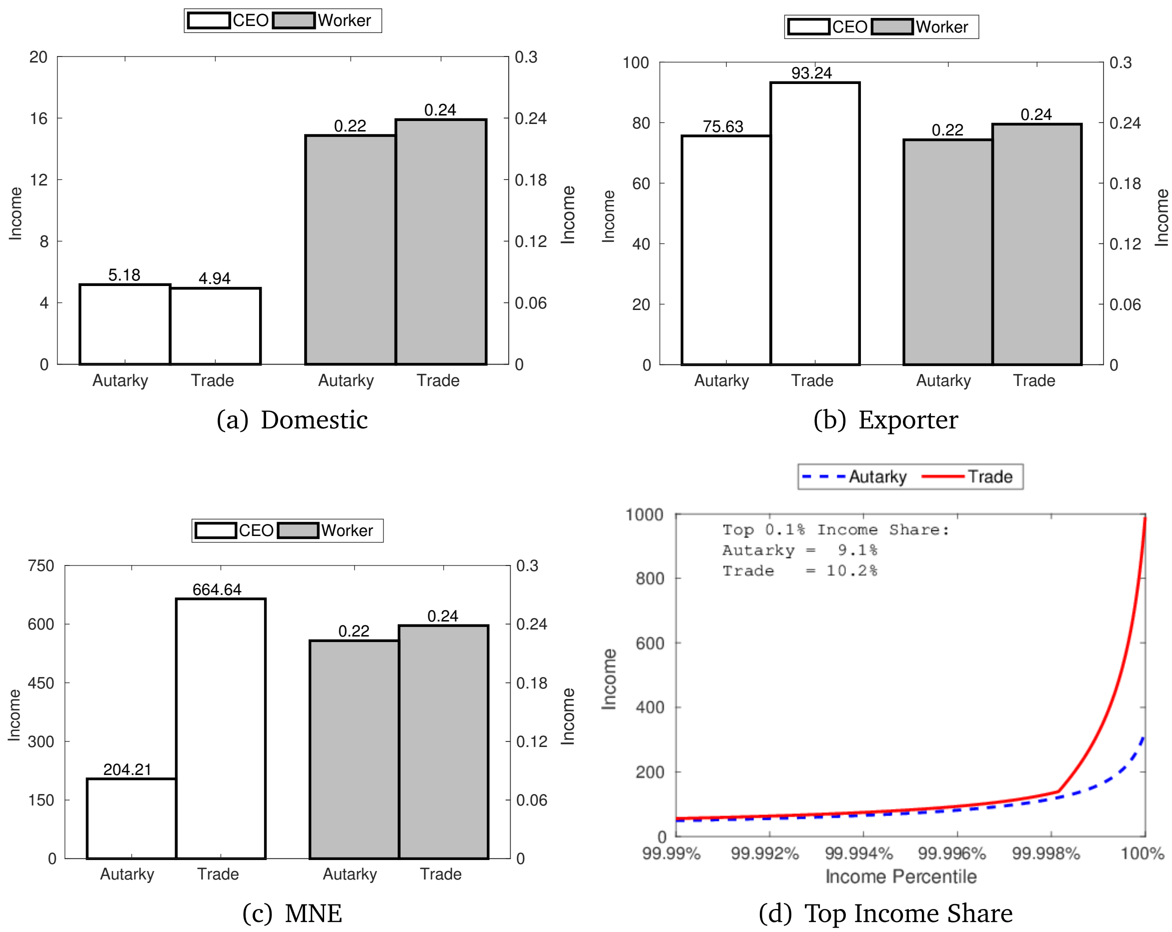
Note: The Mean-to-median ratio and the percentile location of the mean come from Rodriguez et al. (2002). The workers' share of income is computed from NIPA. Table 1.10. The CEO-worker pay ratio is computed in Section 2.2.

2.5.4 Openness and Income Inequality

While the benchmark calibration used the iceberg trade costs τ and the fixed cost of starting foreign subsidiaries g to match the moments of trade volume and multinational sales in the data, we now examine how different measures of income inequality vary with τ and g. The first set of results compare the autarky equilibrium with the benchmark model, and the second set of results report the sensitivity of income inequality to continuous changes in the openness of trade.

Autarky and Trade We first show that opening to trade widens within-firm inequality. To do so, we compare the income of different individuals between autarky and the benchmark model. In "autarky," we set τ and g matrices high enough such that no trade and foreign investment takes place, while keeping all the other parameters the same as in the benchmark model. The first three panels in Figure 2.4 compare the income of the CEO and a worker with average human capital across three different firms in autarky and in trade. The firm in panel (a) is a domestic firm in trade equilibrium, the firm in panel (b) an exporter, and the firm in panel (c) a multinational firm.[27] The income of the average worker increases by around 9.1 percent from 0.22 to 0.24 in all three firms between autarky and trade. However, different CEOs see different income paths. The CEO at the domestic firm sees his/her income decrease by around 4.6 percent, the CEO at the exporting firm sees his/her income increase by around 23.3 percent, while the CEO at the multinational firm sees his/her income surge by as much as 225 percent. As a result, trade widens within-firm

[27] To keep the results comparable between this section and the robustness check sections, we report the income of the CEO from the largest domestic, exporting, and multinational firm respectively in each graph.



(a) Domestic (b) Exporter

(c) MNE (d) Top Income Share

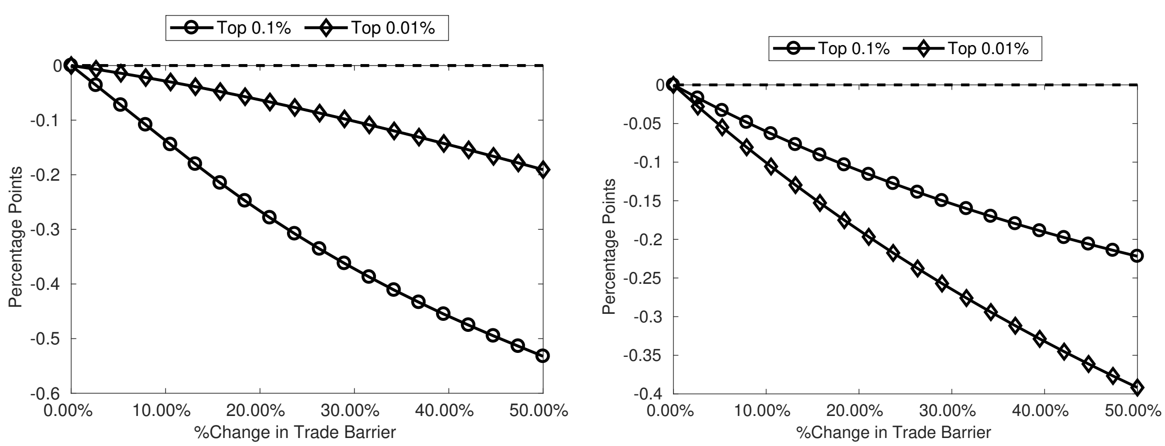
Figure 2.4: Income Inequality between Autarky and Trade, Benchmark Model

Note: The first three panels of the figure compares the income of the CEO and a worker with average endowment of human capital at three different firms in the economy. The last panel plots the income of top 0.1 percent in autarky v.s. in trade. "Autarky" means both τ and g are set to a large number so trade and FDI fall to 0. "Trade" means the benchmark model when both τ and g are calibrated so the exports-to-GDP ratio and multinational-sales-to-GDP ratio match the U.S. data in 2008.

inequality for the large firms that sell to ROW: the CEO-to-worker pay ratio increases from 340 to 388 in the exporting firm, and from 928 to 2,769 in the multinational firm.

Higher within-firm inequality translates into higher top income shares. The last panel in figure 2.4 compares the income of the top 0.1 percent of the population between autarky and the benchmark model. The income distribution is already skewed to the right in autarky, with the top 0.1 percent of the population claiming around 9.1 percent of total income. In trade equilibrium, the distribution is even more skewed to the right, with the top 0.1 percent income share increasing to 10.2 percent. This is a 1.1 percentage point change in absolute income shares, or a 12 percent increase in relative terms. In comparison, the top 0.1 percent income share increased by 2.6 percentage points between 1988 and 2008 in the U.S. data. Overall, the model seems to be able to explain a significant proportion of the change in top income share using the change in the volume of trade and FDI sales.

Top Income Shares, τ, and g In the next set of simulations, we study how different income shares respond to gradual changes in τ and g. We first gradually increase τ from the benchmark value, $\tau = 1.72$, by 50 percent to $\tau = 2.08$, while keeping all the other parameters at the benchmark value. As τ increases, the exports-to-GDP ratio drops from 0.129 (2008 value) to 0.042, which is roughly the level in early 1970s. Panel (a) of figure 2.5 presents how top 0.1 and top 0.01 income shares in the U.S. respond to changes in τ. Higher trade barriers hurt the top earners more than the rest of the distribution and lead to lower top income shares. For example, the top 0.1 percent income share drops by 0.53 percentage point, and the top 0.01 income share drops by 0.19 percentage point. Similarly, Panel (b) in the same figure presents the changes in income shares responding to the changes in g. Again, higher fixed costs to set-up foreign subsidiaries hurt the top income earners more: top 0.1 percent income share decreases by 0.22 percentage point, while the top 0.01 percent income share decreases by 0.39 percentage point, when g is 50 percent higher than the benchmark model.



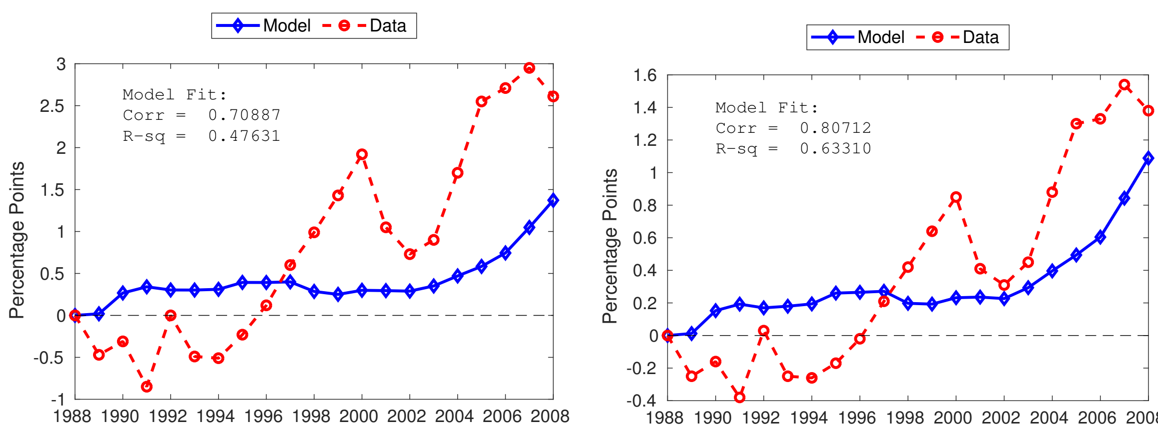
(a) τ (b) g

Figure 2.5: Income Shares and Barriers to Trade, Model Simulations

Note: This figure plots how income shares respond to changes in trade barriers τ and g. The vertical axis is the change in income shares as compared to the benchmark model. The horizontal axis is the percentage changes in τ and g as compared to the benchmark model.

Top Income Shares between 1988 and 2008 To quantify the impact of globalization on top income shares, we calibrate the model to match data on trade flows, multinational sales, and GDP and then compare the predicted income shares to those in the data. Specifically, we calibrate τ and g to match the export-to-GDP ratio and the multinational-sales-to-GDP ratio, and b_i to match the GDP ratio between ROW and the U.S. in each year between 1988 and 2008. All the other parameters are fixed at the benchmark value. The values of τ, g and b_i are reported in Table B.9. Conditional on the calibrated τ, g, and b_i in each year, we solve the general equilibrium of the model, compute the measures of income inequality, and compare them to the data.

The model captures changes in top income shares between 1988 and 2008: the correlation between the annual changes in top 0.1 percent income shares in the model and the data is 0.71, and the adjusted R-squared of regressing the data series on the model series is 0.48. Panel (a) in figure 2.6 compares the data and model series over the 20-year period. The red dashed line is the change in the income share data between the year on the x-axis and 1988, expressed in percentage points, with the data coming from Table A.1 in Piketty and Saez (2003) updated through 2008. For example, the last point on this curve indicates that compared to 1988, the top 0.1 percent income share in 2008 is 2.61



(a) Top 0.1 Percent (b) Top 0.01 Percent

Note: This graph shows the change in top 0.1 percent and top 0.01 percent income shares in percentage points between 1988 and 2008. In the model simulation τ and g are calibrated to match the imports-to-GDP ratio and multinational-sales-to-GDP ratio in each year. For other model parameters behind this simulation, see Section 2.5.2. The source of the data is Table A.1 in the updated tables of Piketty and Saez (2003). Two measures of model fit are computed: the Pearson correlation between the two curves and the adjusted R-squared of estimating a linear relationship with data sequence on the left-hand-side and model sequence on the right (with constant term).

Figure 2.6: Top Income Share and Globalization over the Years

percentage points higher. The blue solid line is the same measure in the model. Each point on the blue solid line is based on the top income share computed with the calibrated τ, g, and b_i in that year. The last point on the graph with the parameters calibrated to the moments in 2008 is the benchmark model.

Our model explains roughly half of the changes in top 0.1 percent income shares in the data. For example, between 2008 and 1988 the top 0.1 percent income share increased by 2.61 percentage points in the data and 1.37 percentage point in the model, indicating that 1.37/2.61$\approx$ 52 percent of the change in top income shares can be explained using the changes in trade volumes and relative productivity. This result suggests that a large proportion of the observed change in aggregate income inequality can be explained through the channel of within-firm inequality: better access to foreign markets benefits the top executives more than it benefits the average workers, widening the income gap between the rich and the poor.

The explanatory power of the model varies from period to period. During the first period, from the beginning of the sample to around 1994, the top income shares fluctuate greatly from year to year in the data. This variation is largely driven by the short and long term effects of the 1986 Tax Reform Act.[28] This tax reform drastically changed the marginal tax rates and tax brackets for the top income earners, thus changing the tax reporting incentives. The short-term consequences of the 1986 TRA are reflected in the sharp increase in top income shares measured in the tax return data between 1986 and 1988 (not shown in the graph). The long-term consequences of the tax reform are less clear, but they can still be observed in the volatility of the data curve in Figure 2.6 before 1994. By contrast, the model economy exhibits a steady increase in income shares driven by the slow increases in trade and multinational sales. The discrepancy between the model and the data is expected because the model does not consider various effects of income tax. In the second phase, starting from 1994 until the 2001-2002 stock market crash, we start to observe a rapid increase in the top income share in the data, but only a modest increase in the model. The surge in top income shares in the data can probably be attributed to the rapid economic growth and the stock market boom. In the model world where no equity market exists, top income shares only respond to the changes in the volumes of trade and multinational sales, which grow slowly during this period. For example, the trade-to-GDP ratio in the U.S. only increased by around 0.15 percent point each year between 1994 and 2002. The low explanatory power of the model is again, expected, because the model is not designed to capture capital gains in the stock market. In the last phase from 2002 onwards, the explanatory power of the model is high. This is a period during which the

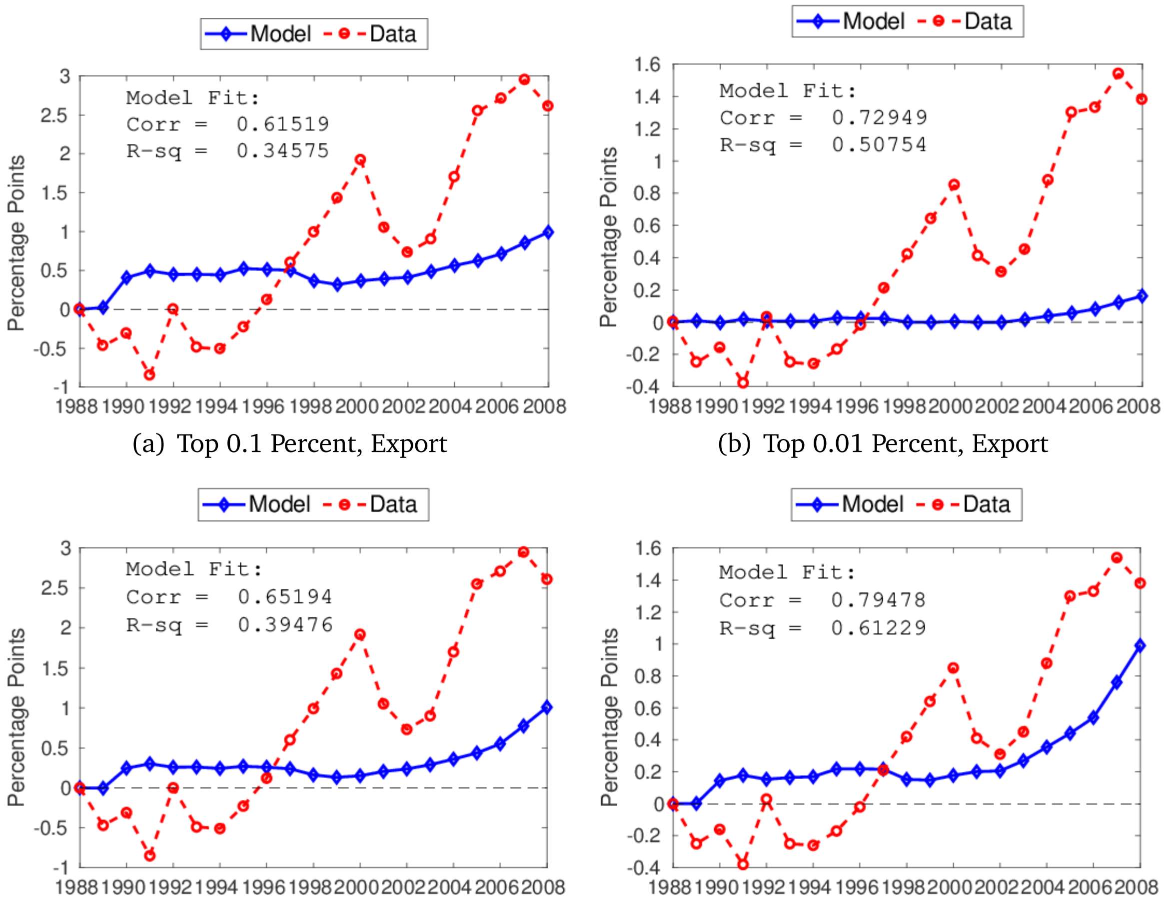
[28] See Slemrod (1996) and Poterba and Feenberg (2000) for details.

trade-to-GDP ratio increases at the fastest pace (1.32 percentage points per year) after World War II. As a result, the trade-induced inequality increases rapidly in the model, matching the concurrent surge in top income shares in the data to a large extent.

Repeating this analysis for the top 0.01 percent of the income distributions returns similar results as shown in panel (b) of figure 2.6. Between 1988 and 2008, the income share of the top 0.01 percent increased by 1.38 percentage points in the data, while it increased by 1.08 percentage points in the model; this suggests that the model can explain a much higher ($1.08/1.38 \approx$ 78) percent of the change in the data in the top 0.01 percent as compared to the top 0.1 percent. With the model designed to explain the income of top executives in large corporations—executives who happen to occupy the pinnacle of the income pyramid—the explanatory power of the model declines as we move down the income ladder. For example, other occupations such as working professionals are among the top 0.1 or 1 percent of the income ladder in the real world, and they are missing in the model. The lower we move down the income ladder, the more frequent are these cases, and the lower the explanatory power of the model will be. It is important to understand how globalization affects different occupations differently, however, this is beyond the scope of this paper. The model does not attempt to provide a comprehensive theory to explain the surge of top income shares in developed countries. Instead, it highlights a particular channel through which globalization can affect the top income earners differently from the way it affects the general population.

The two channels of globalization, exporting and multinational sales, exert roughly equal influence on the top 0.1 percent income share. Figure 2.7 reports two counterfactual simulations in which we only allow one channel of globalization to move while fixing the other parameters. With the cost of starting foreign subsidiaries g_{ij} fixed, the movements in trade costs τ_{ij} can generate a 0.99 percentage point surge in the top 0.1 percent income share between 1988 and 2008. The movements in g_{ij} generate a similar percentage (1.05 percentage points) in the top 0.1 percent income share. At the top 0.01 percent, the reductions in g are more effective than those in τ (0.97 v.s. 0.16 percentage points), probably due to the higher concentrations of CEOs from the multinational firms within the top 0.01 percent.

Lastly, we show that changes in relative TFP across countries cannot explain the dynamics of top income shares alone without the expansion of the volume of trade and multinational sales. We run two counterfactual simulations to highlight this point. In the first simulation, we fix both τ and g to their benchmark values, and allow only the TFP vector b_i to vary from year to year. Conditional on year-specific b_i, we solve the model and compute the top income shares for each year, and compare them to the data. Without the



(a) Top 0.1 Percent, Export

(b) Top 0.01 Percent, Export

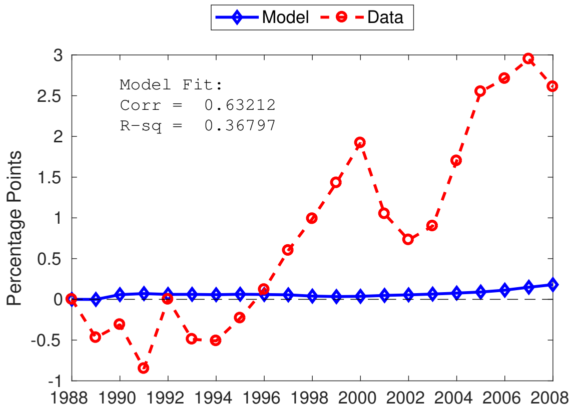
(c) Top 0.1 Percent, FDI

(d) Top 0.01 Percent, FDI

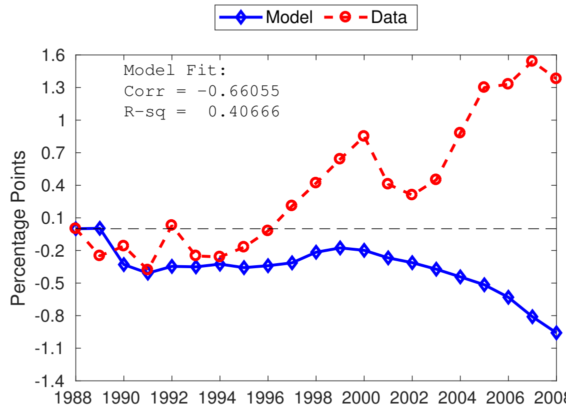
Note: This graph shows the change in top 0.01 percent and top 0.1 percent income shares in percentage points between 1988 and 2008 while we only allow for export (τ) or FDI (g) liberalization separately. For more details, see notes to Figure 2.6.

Figure 2.7: Exporting and FDI Liberalization

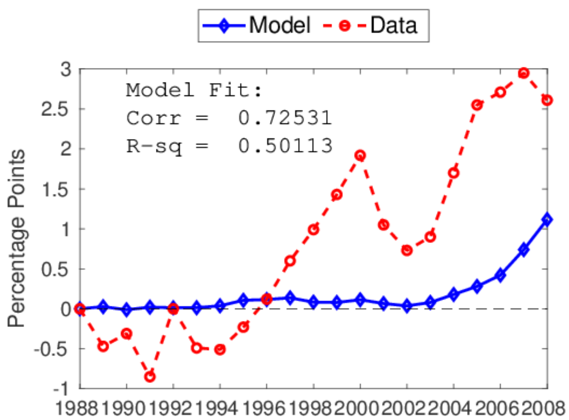
expansion of the volume of trade and multinational sales, top income shares in the model do not follow the data as shown in the two two panels of Figure 2.8. The top 0.1 percent income share barely moves over time, and the top 0.01 income share actually decreases when τ and g are fixed at the benchmark value. In the second counter-factual, we do exactly the opposite exercise: we fix TFP at the 1988 level and allow τ and g to move. The results reported in the bottom two panels of figure 2.8 are basically identical to the baseline simulations in figure 2.6. This confirms the message from the top two panels: it is the the evolution of trade barriers, not the relative productivity that drives the pattern of top income shares.



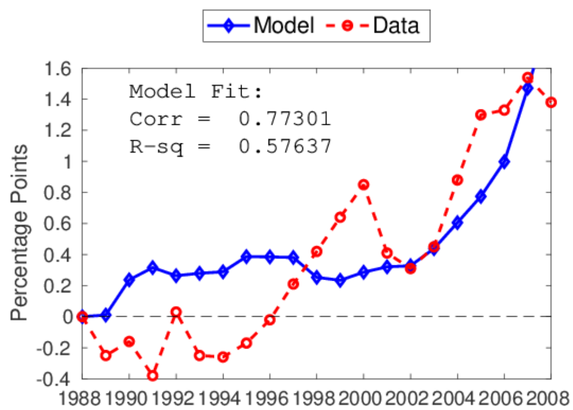
(a) Top 0.1 %, Only TFP Changes



(b) Top 0.01 %, Only TFP Changes



(c) Top 0.1 Percent, No TFP Changes



(d) Top 0.01 Percent, No TFP Changes

Figure 2.8: The Effects of TFP Changes

Note: This graph shows the change in top 0.1 percent and top 0.01 percent income shares in percentage points between 1988 and 2008. The change in the model is shown on the left axis, and the change in the data is shown on the right axis. In the model simulation, τ and g matrices are fixed at 1988 level, while TFP varies from year to year. For other model parameters behind this simulation, see Section 2.5.2. The source of the data is Table A.1 in the updated tables of Piketty and Saez (2003). Two measures of model fit is computed: the Pearson correlation between the two curves and the adjusted R-squared of estimating a linear relationship with data sequence on the left-hand-side and model sequence on the right (with constant term).

2.5.5 Robustness Checks

In this section, we show that the earlier analysis is robust to different values of the elasticity ϵ and the implied markup. In the benchmark model we calibrate $\epsilon = 4$ to capture the average markup. In this section we set ϵ to 2 and 6 and repeat the earlier analysis. In each of the robustness checks we re-calibrate every parameter to match the same moments as in the benchmark model; the re-calibrated parameters are reported in Table 2.6.

Parameter	Benchmark	High ϵ	Low ϵ	Target/Source
λ	3.81	5.3	1.06	Axtell (2001)
ϵ	4.0	6.0	2.0	Average mark-up
α	23.2	29.5	461.1	Corporate sales as a percentage of all firms sales
β	0.747	0.747	0.747	Tail index of income dist., Drăgulescu and Yakovenko (2001*b*)
f_{11}	6.0	6.0	6.0	World Bank Doing Business Index
f_{12}	19.6	19.6	19.6	World Bank Doing Business Index
f_{21}	24.6	24.6	24.6	World Bank Doing Business Index
f_{22}	38.9	38.9	38.9	World Bank Doing Business Index
f-Scale	0.756	0.0476	10.56	Percentage. of chief execu. in work force.
n_{ROW}	6.0	6.0	6.0	Caselli (2005), Barro and Lee (2010)
b_{ROW}	0.58	0.58	0.47	Relative country size
s	3.249	2.202	8.72	Highest-CEO-to-average-wage ratio among public firms
τ	1.720	1.566	3.323	Export-GDP ratio in 2008
g	1020	3548	702	Multinational-sales-GDP ratio in 2008

Table 2.6: Calibration Targets and Results

Note: λ is the shape parameter of the exponential distribution. ϵ is the elasticity of substitution in the utility functions. α is the size of the smallest public firm. β is the tail index of the compensation function. f_{ij} is the fixed cost of exporting from country j to country i. f-Scale is the normalizing factor of the entire f_{ij} matrix. We divide the f_{ij} matrix by this number. n_i is the measure of capital-adjusted endowment of human capital in country i. b_i is the TFP in country i. s is the upper bound of human capital distribution. See Section 2.5.2 and the appendix for the details of calibration. See Table B.9 for the calibrated values of τ, g and TFP by year used in the counter-factual.

In general, when the elasticity of substitution is higher, the markup and profit margins of the firms decrease, the income distribution is less concentrated in the hands of the executives, and top income shares are less responsive to changes in trade barriers. Figure B.2 in the appendix reports the results when $\epsilon = 6$. The main results of the benchmark model carry through in this case with a smaller magnitude. In this case, the real income of workers increase by around 4.7 percent between autarky and trade, while the CEO at a multinational firm sees his/her income increasing by approximately 181 percent. The impact of trade can also be observed at the aggregate level: top 0.01 percent income share increases from 3.5 to 3.7 percentage points. Figure B.3 in the appendix reports the results when $\epsilon = 2$. Again, the main results of the benchmark model are preserved and even strengthened in this case due to the same reason outlined above. Between autarky and trade, the real income of the workers increases by 39 percent, while the income of the CEO at the multinational firm increases by 211 percent. At the aggregate level, the top 0.01 percent income share increases from 5.0 to 8.0 percentage points between autarky and trade.

2.6 Conclusion

This paper studies the relationship between globalization and income inequality with a special focus on the gap between the rich and the poor. Empirically, this paper presents a new fact that within-firm inequality is higher among the firms that have access to global markets. On average, the CEO-to-worker pay ratio is about 50 percent higher among the exporting firms than among domestic firms. The differences in within-firm inequality are mainly driven by differences in firm size. Using China's 2001 accession into the WTO as a trade shock, we also show that the U.S. firms with prior linkage to the Chinese market experienced higher exports and within-firm inequality during the years after China's accession using a difference-in-difference method.

This paper presents a new framework to study the distributional effect of trade. It merges the heterogeneous firms trade model with a model of occupational choice and executive compensation. The key mechanism to generate higher within-firm inequality among exporters and MNEs is through the size effect. On the one hand, CEO compensation is positively linked to the performance of the firm, and only the large and productive firms find it profitable to sell to the global markets. On the other hand, the wage rate is determined in a countrywide labor market and is not linked to each specific firm. These two forces imply that within-firm inequality is higher among the firms that have access to the global markets. We analytically show that trade liberalization leads to higher top income shares in general equilibrium. Using counterfactual analysis, we argue that the changes in trade barriers are able to quantitatively explain a large fraction of the surge in top income shares in the U.S. data.

CHAPTER III

Firms and Collective Reputation: the Volkswagen Emissions Scandal as a Case Study

From a work with Rüdiger Bachmann and Gabriel Ehrlich

Abstract

This paper uses the 2015 Volkswagen emissions scandal as a natural experiment to provide causal evidence that group reputation externalities matter for firms. Our estimates show statistically and economically significant declines in the U.S. sales and stock returns of, as well as public sentiment towards, BMW, Mercedes-Benz, and Smart as a result of the Volkswagen scandal. In particular, the scandal reduced the sales of these non-Volkswagen German manufacturers by approximately 76,000 vehicles over the following year, leading to a loss of approximately $3.7 billion of revenue. Volkswagen's malfeasance materially harmed the group reputation of "German car engineering" in the United States.

JEL Codes: D12, D90, F23, L14, L62.
Keywords: automobiles, collective reputation, country reputation, difference-in-differences, event study, Google trends, firm reputation, natural experiment, reputation externalities, Twitter sentiment, Volkswagen emissions scandal.

3.1 Introduction

A group's collective reputation can influence the outcomes for its members independently of their individual characteristics or behavior. In a seminal paper, Tirole (1996) develops a theoretical framework for modeling collective reputation showing how stereotypes can arise through history dependence so that an original sin by elder group members can have long-lasting effects on a group. Tadelis (1999), writing in the large theoretical literature on firm reputation, highlights the importance of names as intangible assets conveying reputation. In this paper, we combine these two perspectives and investigate empirically a case in which firms have a collective reputation by virtue of their association with a particular country.

Despite the theoretical interest in the issue, there is limited empirical evidence that group reputation and reputational externalities matter economically at the level of the firm.[1] This evidence is limited by many obstacles. First is the scarcity of large and prominent shocks that directly implicate only a subset of group members so that the existence of reputational spillovers on the other group members can be demonstrated. Second is the difficulty of identifying natural and reputationally salient groups of firms. Third is the rarity of direct measures of group reputation, which requires researchers to make indirect inferences about the effects of reputation. To overcome these issues, we use the 2015 Volkswagen (VW) emissions scandal as a natural experiment, which directly implicated VW but not the other German automakers. We study the scandal's effects on their vehicle sales, stock returns, and social media sentiment. In doing so, we provide empirical support for the theoretical literature and the existence of group reputation externalities in an important setting.

On September 18, 2015, the U.S. Environmental Protection Agency (EPA) served a Notice of Violation to the VW Group alleging that approximately 500,000 VW and Audi diesel-engine cars sold between 2009 and 2015 in the United States contained a defeat device that allowed these cars to comply with emissions regulations in the test box, while

[1]There is a growing empirical literature on reputation management on online platforms (e.g., Mayzlin, Dover and Chevalier (2014), Fan, Ju and Xiao (2016) and Li, Tadelis and Zhou (2016)), but it does not generally focus on collective reputation or reputational externalities. There is an agricultural economics literature with an emphasis on group reputation effects of regional appellations such as Bordeaux Wines (Castriota and Delmastra (2014) and Landon and Smith (1998)), but it typically does not include causal or quasi-experimental analysis. As Castriota and Delmastra (2014) state: "Despite the ubiquity of the concept, the economic literature concerning collective reputation is still in its infancy. [. . .] to our knowledge, there is no work that has tested the determinants of the process of collective reputation building." Noskoy and Tadelis (2015) provide a field experiment with an online platform that also includes an emphasis on causal analysis of reputational spillovers.

having higher on-road emissions.[2] This date marks the public eruption of one of the major industrial scandals in recent history, with a prolonged legal fallout in the United States.

Several features of the scandal make it an appealing natural experiment: (1) For the general public the scandal was a clear surprise in September 2015, and it immediately generated extensive media coverage. There was also no concurrent notion in the public that the non-Volkswagen German car manufacturers—BMW, Mercedes-Benz, and Smart—were manipulating vehicle emissions.[3] (2) The German auto manufacturers featured the notion of "German car engineering" prominently in their U.S. advertising, creating a natural reputational group. Additionally, we show that pre-scandal trends in reputation and business outcomes appear to have been similar among automakers from all countries, suggesting that non-German automakers can serve as a control group for understanding the effects of a German-specific shock. (3) Individual automotive makes are salient to consumers, enabling us to use novel company-specific data on U.S. social-media sentiment and internet searches to directly establish the existence of reputational externalities.

Adding to its appeal as a natural experiment, the scandal occurred within an important setting: (4) The car manufacturing industry is large and important in Germany. In 2014, the year prior to the scandal, cars amounted to 18 percent of Germany's total exports according to the German Federal Statistical Office (Destatis (2015)), and were thus Germany's largest export category. Also as of 2014, Germany captured by far the largest share of world car exports in UN trade statistics (United Nations (2017)), with 22.7 percent in dollar and 18.5 percent in unit terms, followed by Japan with, respectively, 12.5 percent and 10.7 percent. (5) German vehicles are a large share of the U.S. market: in 2014 German car manufacturers comprised 8.1 percent of all U.S. light vehicle sales, making Germany the second-largest source for foreign-branded vehicles. (6) The scandal's reputational consequences were amplified by the damage the excess emissions caused to the public. Oldenkamp, van Zelm and Huijbregts (2016) estimate that the excess emissions caused by VW diesel cars cost 45,000 disability adjusted life years, with a value of life lost of approximately \$39 billion.[4] We add to this a calculation of the economic damage for the other German car manufacturers. Finally, (7) the scandal also sparked a widespread

[2]The Volkswagen Group consists of Volkswagen proper plus Audi and Porsche.

[3]On September 22, 2015, CNN published (Petroff (2015)): "But before you start worrying about the complete collapse of the German auto industry, it's worth repeating that – at least for now – the scandal is limited to Volkswagen. Other German automakers such as Daimler, which owns Mercedes-Benz, and BMW have said they're not affected." Only in the summer of 2017, after our study period, was it suggested that Mercedes-Benz had also manipulated emissions (Zeit Online (2017)), although Mercedes-Benz never admitted wrongdoing in the United States.

[4]Barrett et al. (2015) estimate 59 premature deaths and a social cost of \$450 million; Holland et al. (2016) estimate similar numbers.

public discussion regarding the mechanism at the center of our paper: country-related reputational spillovers. Our paper provides numbers to this debate.

We find that the VW scandal caused a loss of 76,000 vehicle sales worth roughly $3.7 billion of revenue for the non-VW German car manufacturers along with a decline in their stock returns relative to expected market outcomes. We reach these conclusions using a difference-in-differences approach (e.g., Angrist and Krueger (1999)) that compares how key outcomes changed over time for the treated group, non-VW German car manufacturers, versus a control group of non-German car manufacturers. The differential responses to the scandal provide causal evidence on the scandal's economic consequences.

Our interpretation of these results is that the scandal harmed the collective reputation of German automakers in the United States. To support this interpretation, we proceed in several steps: First, we provide suggestive evidence that German car manufacturers constitute a reputationally salient group under the umbrella of "German engineering." Second, we document that the scandal reduced the sales of each non-VW German automaker individually. Third, we document a deterioration in positive public sentiment toward the non-VW German automakers in the social media data.[5] Fourth, we show that the results are not primarily driven by diesel cars, despite the scandal's origins in the diesel market. Fifth, we use internet search data to argue that consumers did not engage in increased information-seeking regarding the non-VW German automakers, which is inconsistent with suspicions of malfeasance similar to VW's.

Our results thus substantiate the opening claim in Tirole (1996) that: "Collective reputations play an important role in economics and the social sciences. Countries, ethnic, racial or religious groups are known to be hard-working, honest, corrupt, hospitable or belligerent." Further, we show that the actions of one member of a group can materially damage the group's reputation, producing reputation externalities from the standpoint of individual firms. We are not aware of any systematic investigation into the reputational spillover effects of major corporate scandals and their economic consequences, and thus hope that this paper fills this lacuna in the literature.

Our study is also related to four additional strands of literature: First, a recent literature in international macroeconomics, for instance di Giovanni and Levchenko (2012*b*), di Giovanni, Levchenko and Mejean (2014) and di Giovanni, Levchenko and Mejean (2015)), emphasizes granularity and the importance of large international firms; our results suggest that misbehavior at such firms can damage the collective reputation of particular national powerhouse industries. Second, the international economics literature has examined the

[5]The economic and reputational spillovers from the scandal on the non-VW German car manufacturers are separate from the deterioration in economic outcomes and reputational measures for VW itself.

extent to which taste shocks for domestic versus foreign goods can explain the comovement of international business cycles (Stockman and Tesar (1995)). Our results suggest that the misbehavior of large multinational firms might generate such taste shocks through reputational spillovers. Third, our results provide a case study for the recent macroeconomic literature on customer capital; our evidence shows how customer capital can decline through reputational spillovers and quantifies the economic consequences of such a loss (Drozd and Nosal (2012) and Gourio and Rudanko (2014)). Fourth, in what Berry, Levinsohn and Pakes (1995) call the "characteristics approach" to demand estimation, researchers must specify what properties of products enter into consumers' preferences. Our results suggest that country of origin might be an important such attribute, consistent with an existing marketing literature.[6]

Section 3.2 provides a more detailed timeline of the VW emissions scandal and describes the scandal's effect on VW. It also introduces our main data sources. Section 3.3 quantifies the economic fallout from the scandal for the other German car manufacturers, focusing on their vehicle sales and also examining their stock returns. Section 3.4 interprets this economic fallout as evidence for the existence of reputational spillovers to the non-VW German car manufacturers. A final section 3.5 concludes.

3.2 The VW Emissions Scandal as a Natural Experiment

In this section, we describe the timeline of the VW emissions scandal in detail, and argue that it represents a natural experiment with which to study the economic effects of reputation. We show that the scandal was largely unanticipated, both in the media and by stock market participants. We also document that the scandal was widely covered in the media and that Volkswagens sales growth, stock price and reputation declined substantially in the scandal's aftermath.

3.2.1 Timeline of the Scandal

In May 2014, West Virginia University's Center for Alternative Fuels Engines and Emissions found discrepancies between high on-road emissions by VW diesel cars and earlier test results. The U.S. Environmental Protection Agency's (EPA) replication of the tests led VW to order a voluntary recall of diesel cars in December 2014, citing a need to recalibrate

[6]This marketing literature has devoted substantial attention to "Country of Origin" (COO) effects on corporate reputation, as summarized in Newburry (2012), who states: "While COO has been studied extensively with respect to firm products, the impact of country on a firms overall reputation and the dimensions forming a countrys reputation are less well understood." In addition, Newburry (2012) calls for more causal analyses of how firms can shape a country's reputation and vice versa.

engines. In May 2015, the California Air Resources Board (CARB) conducted new tests, and again the on-road emissions failed to match the test-box results for VW diesel cars. In July 2015, the agencies informed VW about these tests and threatened not to certify the 2016 diesel cars. On September 3, 2015, VW admitted to the EPA and CARB that they had used a defeat device in their software which regulated emissions and produced fake test results in the test box (see Breitinger (2016) for a more complete timeline). The scandal entered its public phase on September 18, 2015, when the EPA served a Notice of Violation to the Volkswagen Group.

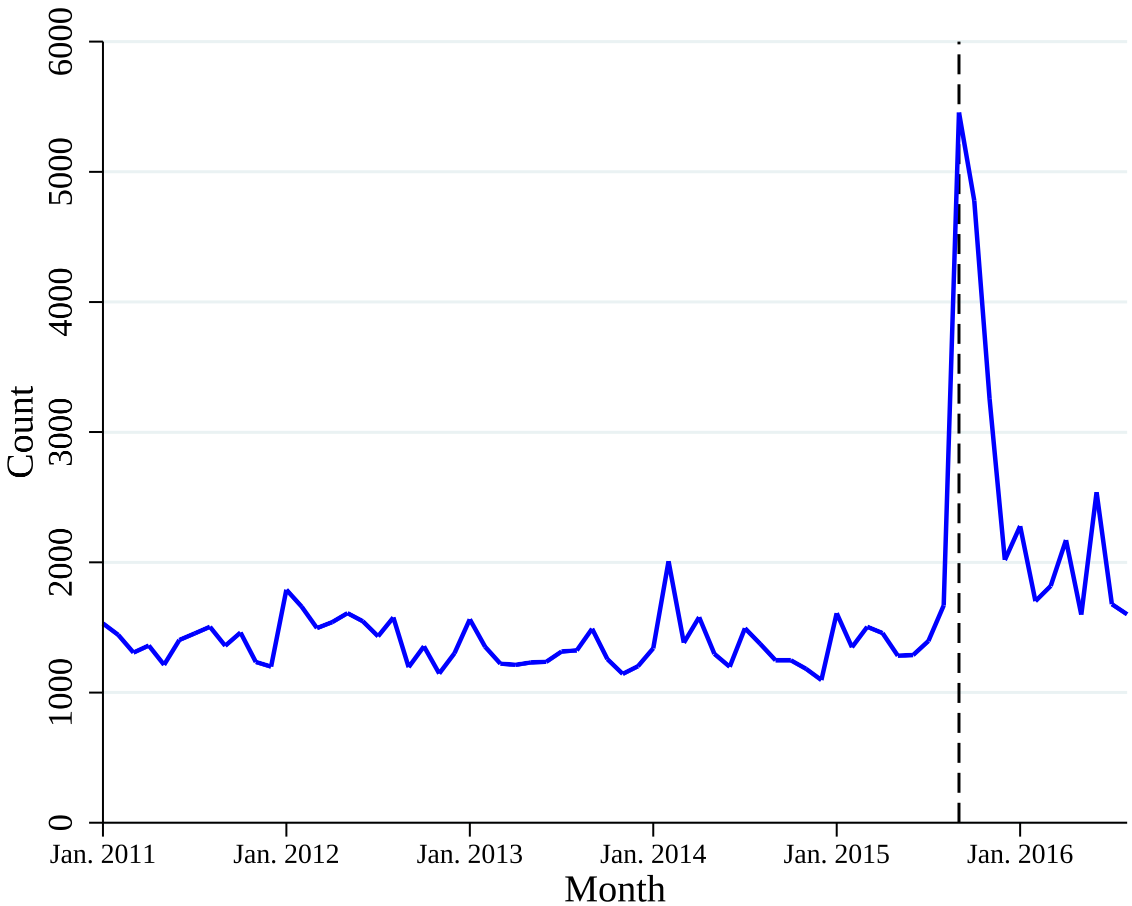
Volkswagen's culpability quickly became a matter of public knowledge: on September 20, two days after the start of the scandal, Volkswagen admitted publicly to the deception and issued an apology. VW Chief Executive Officer Martin Winterkorn resigned three days later. On September 28, German authorities opened a fraud investigation of the former CEO, and in October they authorized a police raid on the VW headquarters. The U.S. Congress called the VW U.S. CEO Michael Horn to testify on October 8, 2015, and he formally resigned his post in early March 2016. In anticipation of the fines and settlements associated with the scandal, VW set aside more than $18 billion in fiscal year 2015. The scandal's legal resolution in the United States began in April 2016. On July 26, 2016, VW and a U.S. court agreed on a civil settlement amounting to $15 billion.

Major news outlets across many countries covered the scandal and its aftermath. On September 19, the morning after the scandal, the front page of the New York Times read: "U.S. Orders Major VW Recall Over Emissions Test Trickery." The Wall Street Journal used a more accusatory tone: "Volkswagen Faked EPA Exhaust Test, U.S. Alleges." Spiegel Online and Zeit Online, the online platforms of two major German newspapers, frequently reported about the scandal. The scandal quickly spilled over into popular culture. On September 22, 2016, VW was awarded the satirical Ig Noble Prize in chemistry (Improbable Research (2016)), and on October 13, 2015, Paramount Pictures and Leonardo DiCaprio's production company announced that they had secured the rights to shoot a film about the scandal (Breitinger (2016)).

3.2.2 The Scandal's Effect on Volkswagen and Its Reputation

We quantify the immediate and longer-term media prominence of the scandal using data from the Newsbank news aggregator on print media mentions of "Volkswagen" in the United States. This database covers roughly 5,000 U.S. newspapers, newswires, journals, and magazines. Figure 3.1 shows that mentions of "Volkswagen" spiked after the scandal, more than tripling from the preceding months to 5,500 in September 2015. This suggests that the scandal came as a complete surprise to the general public. Media interest in Volkswagen remained elevated for most of the following year.

Figure 3.1: Monthly Print Media Mentions of "Volkswagen" in the United States



Note: Dashed line shows the date of the Volkswagen emissions scandal, dated September 2015. Data come from the Newsbank news aggregator, which covers roughly 5,000 U.S. newspapers, newswires, journals, and magazines. Time period covered is January 2011 to August 2016.

Economic consequences quickly followed suit. Table 3.1 displays U.S. light vehicle sales statistics over the period of January 2011 to August 2016, split into a pre-scandal period through August 2015 and a post-scandal period beginning September 2015. Volkswagen's light vehicle sales declined by an average of 3,000 units per month in the 12 months following the scandal relative to the pre-scandal average. Volkswagen's average 12-month natural log sales growth, $\ln \text{Sales}_t - \ln \text{Sales}_{t-12}$, was -0.12 in the year following the scandal, compared to 0.07 prior to the scandal.[7] By contrast, the sales of all automakers increased by an average of 7,000 units per month and make relative to the pre-scandal average, and their post-scandal log sales growth averaged 0.05.

[7]The VW Group shows increases in its average number of vehicle sales by month-make in the post-scandal period relative to the pre-scandal period; this is explained by an increasing trend in vehicles sold over the entire time period. More informative and relevant for our paper is the switch from positive average growth rates for VW Group before the scandal (0.12) to negative average growth rates after the scandal (-0.01).

Table 3.1: U.S. Light Vehicle Sales – Descriptive Statistics

	Average Vehicle Sales by Make-Month	Average 12-month Log Sales Growth by Make-Month	Number of Make-Months
All Makes			
Whole Sample	34,622	0.10	2,798
Pre-Scandal	33,567	0.11	2,378
Post-Scandal	40,595	0.05	420
Volkswagen			
Pre-Scandal	29,751	0.07	68
Post-Scandal	26,481	-0.12	12
VW Group			
Pre-Scandal	14,982	0.12	204
Post-Scandal	15,891	-0.01	36

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Volkswagen Group is defined as Volkswagen, Audi, and Porsche. Pre-scandal period is January 2011 to August 2015; post-scandal period is September 2015 to August 2016. Sales are measured in units sold. Data come from Ward's Automotive.

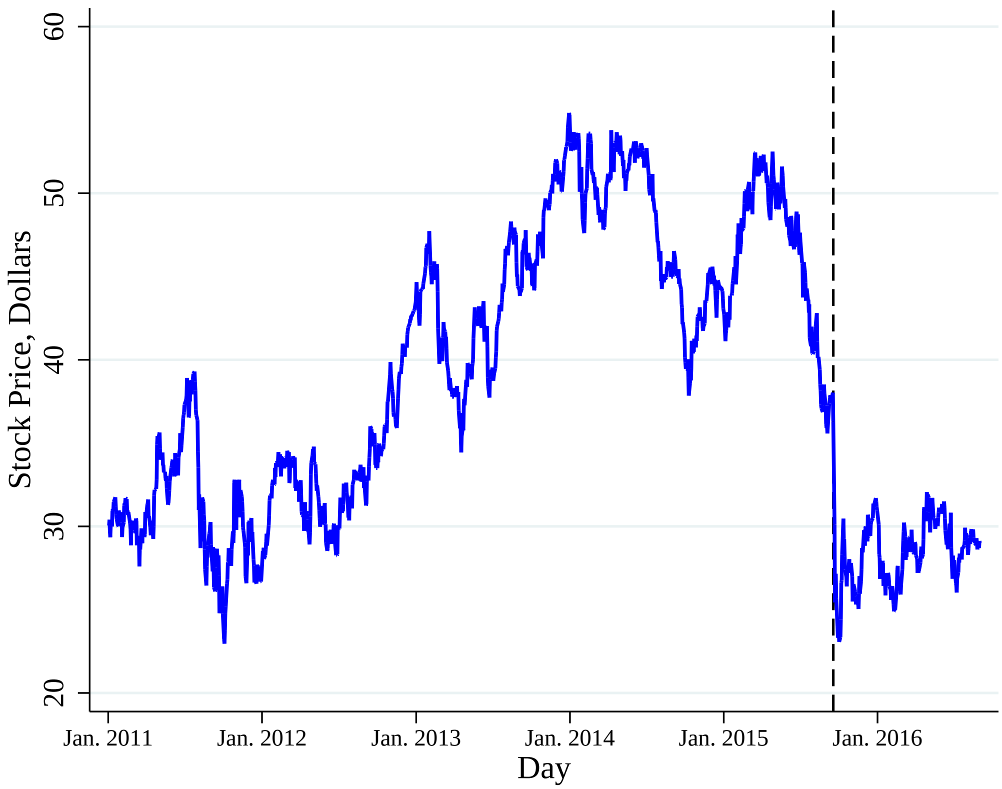
To construct these statistics, we obtain light vehicle sales data at the unit level from WardsAuto, one of the premier automotive industry publications. WardsAuto receives these unit-level sales data from all car manufacturers in the United States. It is thus in principle a complete count of light vehicle. The car manufacturers themselves use Ward's data for their own analyses. In addition, the official U.S. car sales statistics in the national accounting data are based on the same data we use. An individual observation underlying the statistics in table 3.1 contains identifiers for the vehicle make (e.g., Honda or Volkswagen), the vehicle model (e.g., Civic or Jetta), and the vehicle powertype (e.g., gas or diesel). The data set contains 37 makes, listed in appendix table C.1, and 357 distinct models. We identify six makes as of German origin: Audi, BMW, Mercedes-Benz, Porsche, Smart, and Volkswagen.[8]

Along with the adverse attention in the media and its reduced sales, VW's stock price declined precipitously following the EPA's announcement; the visually evident discontinuity on September 18 in figure 3.2 suggests that the scandal came as a surprise to market participants. Volkswagen's end-of-day stock price fell by 33 percent in the two trading

[8] Mini, the present-day incarnation of a line manufactured by the British Motor Corporation and its successors between 1959 and 2000, is currently owned by BMW. Given its historical association with Britain, we classify Mini as not of German origin. We consider alternative classifications in appendix C.2.1.

days following the scandal.[9] The stock price subsequently recovered some of its losses over the rest of the year, but at the end of August 2016 it remained 24 percent lower than its pre-scandal closing price.

Figure 3.2: End-of-Day Stock Price for Volkswagen Group



Note: Dashed line shows the date of the Volkswagen emissions scandal, dated September 2015. End-of-day price shown for Volkswagen ADR listed on U.S. stock exchanges. Data come from the Bloomberg database. Time period covered is January 2011 to August 2016.

Finally, we use novel sentiment measures from Networked Insights to measure the scandal's effect on Volkswagen's reputation. Networked Insights is a data analytics company, founded in 2006, that provides a platform for real-time semantic analyses of social media posts; its primary clients are consumer-facing companies that use the platform to manage their brands. We focus on sentiment data from Twitter, an online social media networking service where some 300 million active monthly users share short messages. The sentiment measures in our data set are calculated from a 10 percent random sample from Twitter.

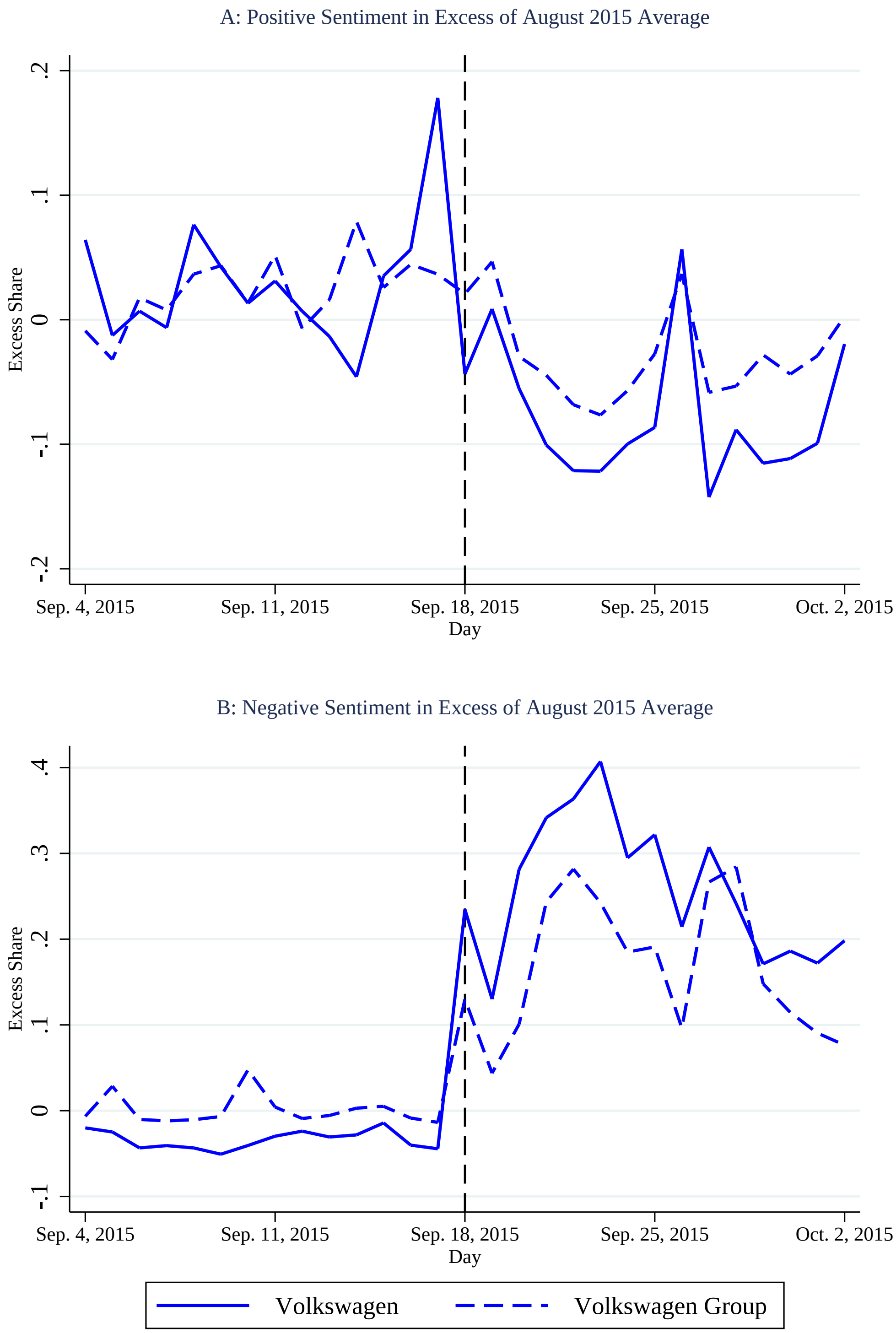
[9] To focus on the effects within the United States and to avoid currency effects from the euro-based VW listing on the Frankfurt Stock Exchange, we use the price of the VW American Depository Receipt (ADR) traded on U.S. markets. ADRs are issued by a U.S. depository bank and entitle the owner to shares in an international security; they are priced and pay dividends in U.S. dollars, and are traded through broker-dealers.

Networked Insights categorizes tweets as displaying positive, neutral, or negative sentiment toward the mentioned company. Posts are excluded from the analysis if they are not written in English or if the user accounts are associated with locations outside the United States. Networked Insights also constructs brand identifiers. An identifier for Volkswagen, for instance, is meant to collate mentions of "Volkswagen," "VW," "#Volkswagen," and the like. Given the size of the underlying data set, Networked Insights only retains the past 13 months of data. We requested the data in September 2016, so our time series begins on August 10, 2015. We first create average daily sentiment shares (positive/negative/neutral) for August 2015 for each vehicle make in our data. We then construct sentiment shares in excess of this August baseline for each day.

Figure 3.3 displays these sentiment metrics for VW and the VW group two weeks before and after the scandal: positive sentiment toward VW declined following the scandal, while negative sentiment spiked. Panel A shows a decrease in positive sentiment toward VW, from an average of 3 percentage points higher than its August baseline in the two weeks prior to the scandal to an average of 8 percentage points below in the two weeks following the scandal. Panel B displays an even sharper increase in negative sentiment toward VW: from an average of 3 percentage points below to an average of 26 percentage points above.[10] The results for the entire Volkswagen group (which includes Audi and Porsche) are similar. Together, these two panels suggest that Volkswagen's reputation suffered in the aftermath of the September 18 EPA announcement.

[10]The pre-scandal and post-scandal means are statistically different at the 1 percent significance level for both positive and negative sentiment.

Figure 3.3: Daily Twitter Sentiment Towards Volkswagen



Note: Dashed vertical lines show the date of the Volkswagen emissions scandal, dated September 18, 2015. Sentiment shares are normalized by subtracting the average sentiment share from August 10 to August 31, 2015. We show a window of ± 14 days around September 18, 2015. Volkswagen Group is defined as Volkswagen, Audi, and Porsche. Data come from Networked Insights.

3.3 The Scandal's Economic Effects on BMW, Mercedes-Benz, and Smart

In this section, we show that the Volkswagen emissions scandal had *economically* important spillovers on the other German auto manufacturers (BMW, Mercedes-Benz, and Smart). First, we present our main result, that the scandal substantially reduced the U.S. sales growth of the other German auto manufacturers relative to their non-German counterparts. Second, we show that those German auto manufacturers' U.S. stock returns fell immediately after the scandal relative to the returns of non-German auto manufacturers. In the ensuing section 3.4, we interpret these economic spillovers as reflective of *reputational* spillovers to the non-VW German car manufacturers.

We estimate the causal effects of the scandal on non-VW German auto manufacturers using a standard difference-in-differences regression specification:

$$y_{it} = \alpha + \beta_i + \gamma_t + \delta T_{it} + \varepsilon_{it}, \tag{3.1}$$

where y_{it} is an outcome of interest for an individual vehicle make or company i at time t; β_i an individual-specific fixed effect; γ_t a time fixed effect; and T_{it} is an indicator taking value one for the German manufacturers on and after the scandal date, and zero otherwise. We exclude the Volkswagen Group from the sample to focus the analysis on the economic consequences of reputation for German automakers not directly implicated by the scandal, that is, reputational externalities. The coefficient of interest, δ, captures the differential impact of being a non-VW German auto manufacturer after the scandal. To interpret the estimated coefficient as a causal effect, we invoke and support the so-called "parallel trends" assumption.

3.3.1 Sales Growth

To study the scandal's spillovers on vehicle sales, we define the outcome variable in equation (3.1) as the 12-month growth rate of unit sales, expressed in log points: $\ln \text{Sales}_{it} - \ln \text{Sales}_{it-12}$. The unit of observation is the vehicle make-month (e.g., Honda in January 2016), and the sample period is January 2011 to August 2016. In this setting, β_i is a make fixed effect, capturing potentially unobserved heterogeneity at the make level. γ_t is a fixed effect for each month in the sample, capturing seasonality in car sales and the potential impacts of time-varying gasoline and diesel prices. We estimate this equation as a weighted regression, with the square root of sales volumes as weights, to dampen the impact of highly volatile sales growth rates of small sales levels.

Table 3.2: U.S. Light Vehicle Sales Growth
German vs. Non-German Manufacturers, Excl. VW Group

Dependent Variable	12-month Log Sales Growth		
	Baseline	No saturated make effects	No saturated fixed effects
	(1)	(2)	(3)
German Manuf. × Post-Scandal	-0.104	-0.110	-0.106
	(0.035)	(0.035)	(0.033)
German Manuf.		0.020	0.017
		(0.011)	(0.010)
Post-Scandal			-0.038
			(0.011)
Time Fixed Effects	Yes	Yes	No
Make Fixed Effects	Yes	No	No
R^2	0.292	0.161	0.015
N	2150	2150	2150

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Standard errors clustered at vehicle make level in parentheses. Volkswagen Group (Volkswagen, Audi, and Porsche) excluded from all regressions. Volkswagen emissions scandal dated September 18, 2015. Sales are measured in units sold. All regressions include a constant and are weighted by the square root of sales volumes. Column (1) is based on a regression with make and time fixed effects. Column (2) is based on a regression with time fixed effects, and an indicator for non-VW German makes. Column (3) is based on a regression with an indicator for non-VW German makes and an indicator for the post-scandal period (September 2015 through August 2016). Data come from Ward's Automotive.

We estimate that the scandal reduced the sales growth rates of the non-VW German automakers by 10.4 percentage points, as shown in our baseline specification in table 3.2, column (1). Columns (2) and (3) present specifications with a coarser treatment of the time and make fixed effects. Specifically, in column (2) we replace the saturated make fixed effects with an indicator variable for non-VW German makes. Column (3) presents an even simpler difference-in-differences specification in which we additionally replace the saturated time fixed effects with an indicator for the post-scandal period. The estimated treatment effect δ is negative and stable across these specifications.[11]

Our estimates suggest that the scandal resulted in 76,070 fewer unit sales for the non-Volkswagen German auto manufacturers between September 2015 and August 2016. We

[11] Appendices C.2.2 and C.2.3 show that the results are also robust to variations in the control group and to several alternative econometric specifications.

calculate this decline by multiplying the manufacturers' total sales of 731,444 units during the twelve month period prior to the scandal by negative 10.4 percent, our estimate of the scandal's effect on their sales growth rate (column (1) of table 3.2). Actual sales for these manufacturers declined by 29,484 units in the twelve months after the scandal, suggesting that sales would have risen by about 45,000 units in the absence of the scandal.

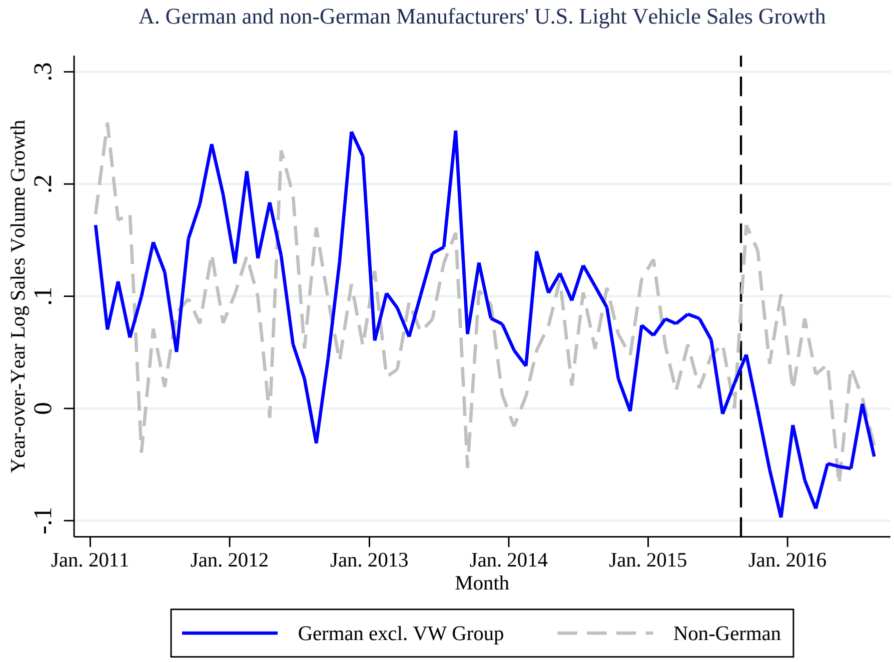
We quantify the joint revenue loss to BMW, Mercedes-Benz, and Smart as $3.70 billion dollars by multiplying the unit sales decline by the manufacturers' suggested retail prices (MSRPs) for calendar year 2015, also obtained from Ward's. Because actual transaction prices may have included post-scandal discounts not reflected in MSRPs, we view our estimated decline in revenue as a likely lower bound of the scandal's true revenue effects on the non-VW German car manufacturers.[12]

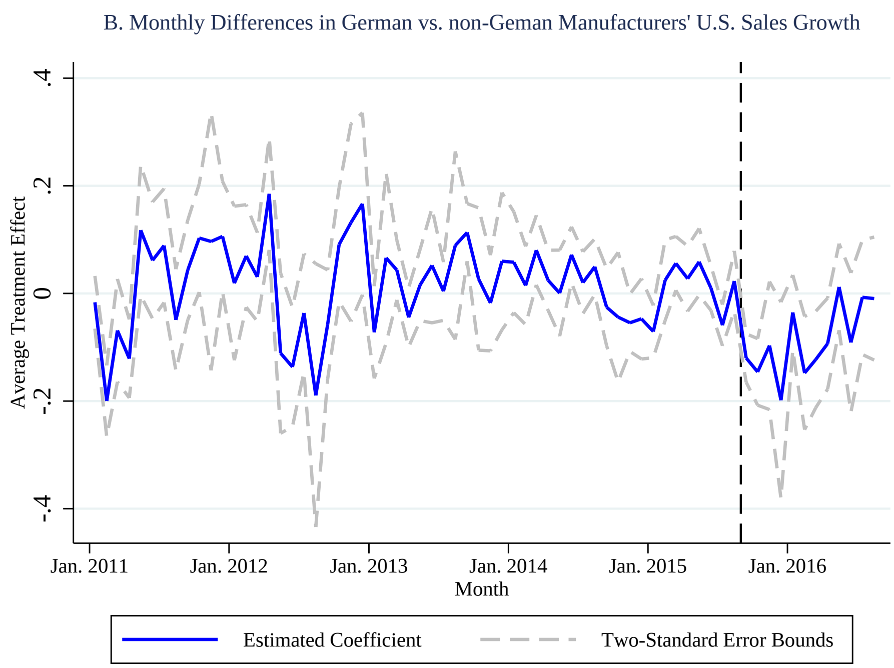
To allay potential concerns about the key identifying assumption underlying our estimation procedure, we provide suggestive evidence in figure 3.4 that the parallel trends assumption holds. Panel A plots the sales growth rates of the two groups, and panel B plots the monthly differences in sales growth between non-VW German and non-German manufacturers. Both panels show that the growth rates of the two groups' sales were essentially indistinguishable prior to the scandal, consistent with the parallel trends assumption.

Two additional patterns in figure 3.4 are worth noting. First, the solid blue line in panel A shows that the sales of the non-VW German automakers declined in the aftermath of the scandal. One might have expected the sales of non-VW German makes to increase as VW sales plummeted. The outright decline observed instead would be difficult to explain without reputational spillovers from VW to the other German car manufacturers. Second, panel B traces out the month-by-month sales differential between non-VW German and non-German automakers after the scandal, providing a more dynamic picture of the effects of the scandal: the sales declines were concentrated in the immediate aftermath in the scandal, as evidenced by an inverted hump-shaped response that faded out to zero by August 2016.

[12] More precisely, we calculate the revenue loss in the following manner: we first calculate model-level revenue by multiplying the quantity sold of a model in the year prior to the scandal by its median MSRP across its available trim levels for 2015. We use the median MSRP across trims for each model because the Ward's data does not break out sales volumes by trim. Second, we aggregate model-level revenues to compute the total revenue for the non-VW German auto manufacturers. Third, we divide total revenue by the total quantity sold to construct an average unit sales price for the non-VW German auto manufacturers. Finally, we multiply the estimated unit decline in sales by this average unit sales price. The estimated decline in revenue that we report here hardly changes when estimated using a difference-in-differences regression with MSRP-based revenues directly as the outcome variable. Transaction prices are not readily available to researchers (see Berry, Levinsohn and Pakes (1995), who use MSRPs as well). By contrast, papers in the literature that use transaction prices typically employ specifically designed and highly confidential surveys; see, for instance, Berry, Levinsohn and Pakes (2004).

Figure 3.4: Differences in U.S. Light Vehicle Sales Growth





Note: Dashed vertical lines show the date of the Volkswagen emissions scandal, September 2015. Panel B displays estimated regression coefficients of the month-by-month differences of German vs. non-German manufacturers' 12-month log U.S. light vehicle sales growth. These coefficients and the corresponding confidence bands are estimated by a pooled (and sales-weighted) regression of the form: $\ln(\text{Sales})_{it} - \ln(\text{Sales})_{it-12} = \sum_{s=1}^{T} \gamma_s \text{ Month}_s + \sum_{s=1}^{T} \beta_s \text{non-VW German}_i \times \text{Month}_s + \varepsilon_{it}$, where $\{\text{Month}_s\}_{s=1}^{T}$ is a complete set of month dummies, ranging from January 2011 to August 2016, and "non-VW German" is a dummy variable that takes a value of one if the make i is BMW, Mercedes-Benz, or Smart. Data come from Ward's Automotive.

3.3.2 Stock Returns

We next show that the scandal's effect on sales was mirrored in financial outcomes, as measured by stock returns. To this end, we combine two complementary data sources. We construct daily U.S. stock returns from the Center for Research in Securities Prices (CRSP) database, which covers primary listings on NYSE, NYSE MKT, NASDAQ, and NYSE Arca.[13] We supplement this data with American Depository Receipts (ADRs) for publicly-listed auto manufacturers from other countries.[14] By incorporating ADRs into our dataset, rather than using the home-country listings, we can compare daily returns coming from the same trading days and better capture the reaction of U.S. investors across a common set of securities. ADRs allow us to calculate the daily returns for foreign car manufacturers even on days when the underlying stocks are not traded in their home markets.[15]

We start by assuming that an individual stock's expected return depends only on the stock's covariance with the market return, its beta. The difference between the expected return on stock i at time t and its actual return is referred to as the "abnormal return," AR_{it}. Formally, we measure each stock's abnormal return by estimating the following regression:

$$R_{it} = \alpha_i + \beta_i R_{market,t} + \varepsilon_{it} \tag{3.2}$$

where i indexes individual stocks and t represents market trading days. The regression sample covers the trading year, approximately 250 trading days, ending thirty days before the scandal. In equation (3.2), R_{it} is the daily stock return for stock i from day $t-1$ to day t, and $R_{market,t}$ is the return on the CRSP value-weighted market portfolio (which does not include the ADRs). The abnormal return is the difference between the stock's actual return and its return predicted from the regression: $AR_{it} = R_{it} - \hat{R}_{it}$. The cumulative abnormal return, CAR_{it}, is then defined as $\sum_{s=0}^{t} AR_{is}$. The starting point in our definition of cumulative abnormal returns, $s = 0$, is September 16, 2015, two days before the scandal.

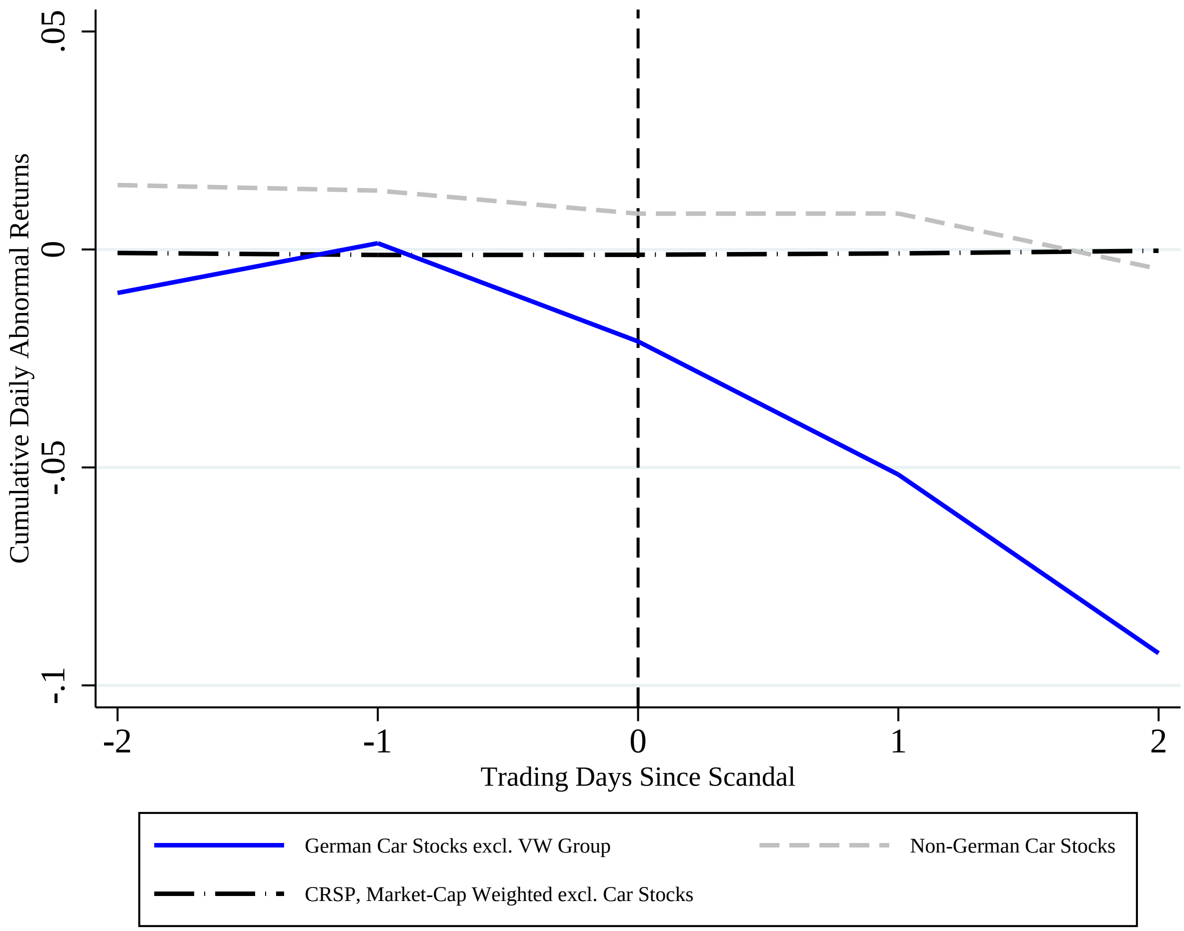
Figure 3.5 shows that the cumulative abnormal return for German auto manufacturers excluding Volkswagen was negative 10 percent within two trading days of the scandal. This decline contrasts sharply with the near-zero abnormal returns for the non-automotive

[13] NYSE is the New York Stock exchange and the premier market place. NYSE MKT is the marketplace within the NYSE for small market capitalization companies. NASDAQ is the second largest marketplace for stocks in the world after NYSE, with a certain specialization in high-tech companies. Finally, NYSE Arca is another specialized electronic-trading marketplace for U.S. stocks.

[14] Appendix table C.2 lists the holding companies of car makes used in our analysis (ADRs for the holding companies of some car makes in the Ward's sales data were insufficiently frequently traded to be used). Ford, General Motors, and Tesla are listed on U.S. stock exchanges; all remaining prices come from ADRs.

[15] For instance, the Tokyo Stock Exchange was closed for holidays September 21-23, 2015; U.S. exchanges were open on those days. If we constructed daily returns of Japanese securities, e.g., Mazda or Nissan, from the Japanese exchange, we would have no observations on those U.S. trading days.

Figure 3.5: Cumulative Abnormal Stock Returns, Market Model



Note: Dashed line shows the date of the VW emissions scandal, September 18, 2015. Automotive stock data come from the Bloomberg database; CRSP index comes from the Center for Research in Security Prices.

stocks in the CRSP database during the five days around the EPA announcement. Similarly, the non-German auto stocks exhibited only slight abnormal return movements on and around the event date. As the three groups had similar cumulative abnormal returns prior to the scandal date, the divergence of the non-VW German auto manufacturers' returns following the scandal reflects the scandal's causal impact on this group. We formalize this notion through two empirical exercises.

First, we use both the AR_{it} and the CAR_{it} as outcome variables in our difference-in-differences regression. Owing to the high-frequency nature of the data, we use data for September 16 and 17, 2015, as the pre-scandal period, and data for September 18, 21, and 22, 2015, as the post-scandal period.

Table 3.3 shows the results and quantifies the causal impact of the scandal: relative to non-German car stocks, the non-VW German automakers experienced roughly 2 percent

Table 3.3: Abnormal Stock Returns – German vs. Non-German Car Firms, Excl. VW Group

Dependent Variable	Abnormal Returns		Cumulative Abnormal Returns	
	(1)	(2)	(3)	(4)
German × Post-Scandal	-0.019	-0.019	-0.064	-0.061
	(0.004)	(0.005)	(0.013)	(0.015)
Weighting	None	Sales Volume	None	Sales Volume
Time Fixed Effects	Yes	Yes	Yes	Yes
Company Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.687	0.600	0.882	0.799
N	60	60	24	24

Note: Unit of observation in columns (1) and (2) is the daily abnormal stock return, and in columns (3) and (4) it is the cumulative abnormal returns for the periods before and after the event date. Abnormal returns are calculated using a market model (equation (3.2)). Automotive stock prices come from Bloomberg and include U.S.-listed stocks (Ford, General Motors, and Tesla) and ADRs (all other care make holding companies). The weighted regressions are sales-weighted, using the Ward's Automotive sales data, because for ADRs we do not have meaningful market capitalization data. Robust standard errors in parentheses. The pre-scandal period comprises September 16 and 17, 2015, and the post-scandal period comprises September 18, 21, and 22, 2015 (September 18 was a Friday in 2015). All regressions include a constant, company and time fixed effects.

lower daily abnormal returns. The latter two columns of the table compare cumulative abnormal returns at the end of the pre- and post- scandal periods: the non-VW German car stocks experienced roughly 6 percent lower cumulative abnormal return. Both results are statistically significant. This evidence suggests that the Volkswagen emissions scandal materially harmed the financial valuations of automakers linked to the scandal through their collective reputation as German automakers.

Second, we corroborate this analysis through the event-study methodology more commonly used in the finance literature (e.g., MacKinlay (1997)). The market model assumes that the return distributions are the same during the estimation period prior to the scandal and during the event window surrounding it. As a result, the (cumulative) abnormal returns for stocks on the three days of September 18, 21, and 22, 2015 should remain approximately zero if the scandal had no effect on stock returns. We can test this hypothesis by computing the following test statistic for each stock i:

$$\frac{CAR_{iT}}{\Big(T * Var[AR_i]\Big)^{\frac{1}{2}}}, \tag{3.3}$$

where CAR_{iT} is the cumulative abnormal return of stock i between September 16 and September 22, 2015; $T = 5$ (trading) days in the event window; and $Var[AR_i]$ is an estimate of the variance of the abnormal return of stock i. Following MacKinlay (1997), we use the abnormal return variance over the estimation window for equation 3.2. Assuming that stock returns are normally distributed, this statistic is distributed approximately standard normal. BMW's cumulative abnormal return of negative 7.2 percent and Daimler's negative 11.4 percent have respective test statistics of 2.42 and 4.17. We note that both Mercedes-Benz and Smart are subsidiaries of Daimler.

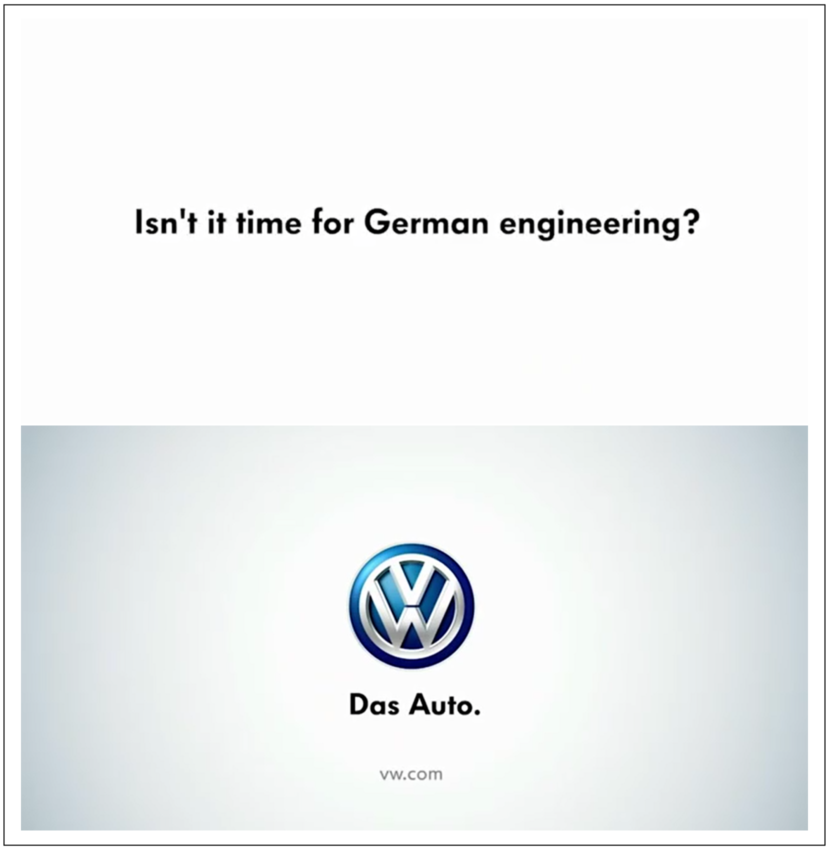
The visual and statistical evidence in this section suggests that the scandal reduced the stock returns of the non-VW German auto manufacturers. Furthermore, the evidence here corroborates the notion that market participants did not anticipate the EPA's notice of violation.

3.4 Mechanism: the Scandal and Reputational Spillovers

In this section, we interpret the economic spillovers documented in the previous section as evidence of the scandal's harm to the group reputation of "German Engineering." We proceed in five steps. First, we provide suggestive evidence that German automakers constitute a reputationally salient group under the umbrella construct of "German Engineering" in both marketing and the media. Second, we emphasize the scandal's collective harm by showing that each non-VW German automaker suffered individually as a result of the scandal. Third, we provide direct evidence of a decline in public sentiment toward the non-VW German automakers. After documenting this evidence for the collective reputation mechanism, we consider two alternative mechanisms for the scandal's spillover effects. We argue that neither can fully account for the scandal's spillovers without the collective reputation mechanism we document. To that end, we show in the fourth subsection that—despite the scandal's origins in the diesel market—the non-VW German automakers experienced adverse spillovers in both their diesel and their non-diesel vehicle sales. Lastly, we use differences in internet searches across German automakers to argue that consumers are unlikely to have ascribed malfeasance similar to VW's directly to the other German automakers.

3.4.1 "German Engineering" as a Group Identity

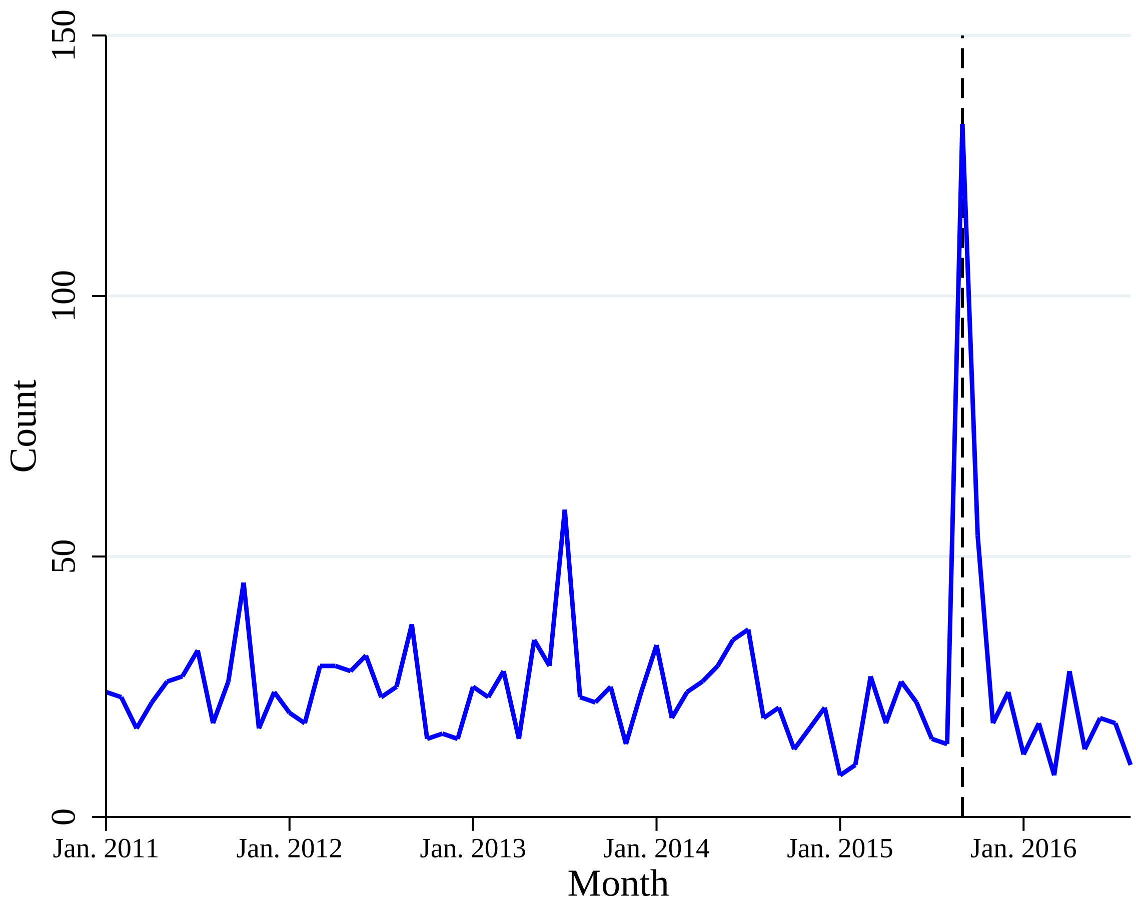
We first note that German manufacturing companies have historically leveraged the broader reputation of "German engineering" in their marketing. For instance, a Volkswagen commercial from 2014 states, "... Everyone knows that the best cars in the world come from Germany." The ad fades out to the question: "Isn't it time for German engineering?"



Following the scandal, media attention to "German engineering" spiked, with 130 print articles mentioning the term in September 2015, a five-fold increase over the preceding months. We illustrate this increase with data from the Newsbank aggregator in figure 3.6. A recurring theme in this coverage was the notion that the scandal might tarnish the broader reputation of German manufacturing firms. As part of this coverage of the scandal, Reuters published an article on September 22, 2015, "VW scandal threatens 'Made in Germany' image" (Chambers (2015)). A day later, Reuters doubled down with an article entitled "Volkswagen could pose bigger threat to German economy than Greek crisis" (Nienaber (2015)), which included the claim: "The broader concern for the German government is that other car makers such as Mercedes-Benz and BMW could suffer fallout from the Volkswagen disaster."[16]

[16]See also Bruckner (2015), Werz (2016), and Remsky (2017).

Figure 3.6: Monthly Print Media Mentions of “German Engineering” in the United States



Note: Dashed line shows the date of the Volkswagen emissions scandal, dated September 2015. Data come from the Newsbank news aggregator, which covers roughly 5,000 U.S. newspapers, newswires, journals, and magazines. Time period covered is January 2011 to August 2016.

3.4.2 Economic Spillovers to Individual German Manufacturers

As evidence of a change in *collective* reputation, we next document that the scandal reduced sales growth for each of the three non-VW German car manufacturers individually. To arrive at this result, in table 3.4 we estimate parallel difference-in-differences specifications in which the treatment group includes only one non-VW German auto manufacturer at a time; each regression excludes the others. Column (1) reproduces our baseline result that the scandal reduced the non-VW automakers’ sales growth overall by 10.4 percentage points. Looking across the individual automakers, columns (2) through (4) show that the scandal reduced BMW’s sales growth rate by 15.1 percentage points, Mercedes-Benz’s by 6 percentage points, and Smart’s by 30.8 percentage points. These estimates translate into sales losses of 53,000 units for BMW, 23,000 units for Mercedes-Benz, and 2,400

Table 3.4: U.S. Light Vehicle Sales Growth
German vs. Non-German Manufacturers, Excl. VW Group

Dependent Variable	12-month Log Sales Growth			
	Baseline	BMW	Mercedes-Benz	Smart
	(1)	(2)	(3)	(4)
German × Post-Scandal	-0.104	-0.151	-0.060	-0.308
	(0.035)	(0.012)	(0.011)	(0.012)
Time Fixed Effects	Yes	Yes	Yes	Yes
Make Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.292	0.296	0.294	0.295
N	2150	2014	2014	2014

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Standard errors clustered at vehicle make level in parentheses. Volkswagen Group (Volkswagen, Audi, and Porsche) excluded from all regressions. Volkswagen emissions scandal dated September 18, 2015. Sales are measured in units sold. All regressions include a constant, make and time fixed effects, and are weighted by the square root of sales volumes. Data come from Ward's Automotive.

for Smart. We view these pervasive declines as evidence of a change in non-VW German automakers' collective reputation following the scandal.

3.4.3 Direct Reputational Spillovers of the Scandal

Next, we use Twitter sentiment data to show that perceptions of the non-VW German automakers suffered in the aftermath of the VW emissions scandal. To show this result formally and to estimate the causal reputational effects of the scandal, we compare the outcomes of non-VW German and non-German auto manufacturers before and after the scandal, adapting our difference-in-differences regression specification. The unit of observation is now a make-day, and the estimation sample is a window of ± 14 days around September 18, 2015. The outcome variables are, respectively, positive and negative Twitter sentiment towards a particular make, relative to an August 2015 sentiment baseline. Both regressions are weighted by tweet volume.[17]

Column (1) of table 3.5 documents a statistically significant decline of 3.5 percentage points in positive sentiment toward non-VW German manufacturers as a result of the scan-

[17] Of the 37 auto makes with light vehicle sales in the Ward's U.S. data (our data source for car sales), Alfa Romeo, Saab, Suzuki and Tesla did not have identifiers in the Networked Insights database; see table C.3 in the appendix for details.

dal. To put that number in perspective, the share of tweets expressing positive sentiment toward those companies averaged 12.3 percent in August 2015. By contrast, the share of tweets expressing negative sentiment toward the non-VW German automakers did not change meaningfully following the scandal, as seen in column (2).

Table 3.5: Twitter Sentiment – German vs. Non-German Manufacturers, Excl. VW Group

Dependent Variable	Positive Sentiment	Negative Sentiment
	(1)	(2)
German $\times$ Post-Scandal	-0.035	0.002
	(0.006)	(0.006)
Time Fixed Effects	Yes	Yes
Make Fixed Effects	Yes	Yes
R^2	0.348	0.268
N	840	840

Note: Unit of observation is vehicle make-day. Sentiment shares are normalized by subtracting the average sentiment share from August 10 to August 31, 2015. The estimation period comprises 14 days before and after scandal date of September 18, 2015. Volkswagen Group is defined as Volkswagen, Audi, and Porsche. All regressions include a constant, make and time fixed effects, and are weighted by tweet volume. Data come from Networked Insights.

We interpret these changes in social-media sentiment as reflecting a decline in the collective reputation of German automakers, but as unlikely to reflect suspicions that the other manufacturers were guilty of the same malfeasance as VW. The sentiment evidence for VW shown in figure 3.3 suggests that evidence of malfeasance manifests mainly as an increase in negative sentiment toward the wrongdoer. The absence of an increase in negative sentiment toward the other German manufacturers is inconsistent with a suspicion of similar wrongdoing. The decrease in positive sentiment, however, suggests that the scandal did adversely affect their reputations. Whereas the original instigator of a reputational event suffers especially in terms of negative sentiment, the reputational spillover of that event appears to manifest mostly in a decline in positive sentiment.

3.4.4 The German Effect Is Not a Diesel Effect

A natural concern given the scandal's origins in the diesel market is that the sales declines we document in section 3.3.1 were driven by the reputation of or new information about diesel vehicles, rather than the collective reputation of German automakers. Indeed, table 3.6 shows that diesel sales declined by 85 percent for the average month and make after the scandal, compared to a 6 percent increase for non-diesel vehicles. This decline was driven primarily by the Volkswagen Group, which was legally prohibited from selling diesel vehicles in the U.S. in the aftermath of the scandal. Nonetheless, even excluding the Volkswagen group, diesel sales declined by 11 percent for the average make-month after the scandal, compared to a 5 percent increase for non-diesels.

Table 3.6: U.S. Light Vehicle Sales – Diesel vs. Non-Diesel

	Average Vehicle Sales by Make-Month	Average 12-month Log Sales Growth by Make-Month	Number of Make-Months
Non-Diesel			
Pre-Scandal	32,589	0.11	2,378
Post-Scandal	39,548	0.06	420
Diesel			
Pre-Scandal	3,836	0.11	606
Post-Scandal	3,577	-0.85	123
Non-Diesel, excl. VW Group			
Pre-Scandal	34,475	0.11	2,174
Post-Scandal	41,782	0.05	384
Diesel, excl. VW Group			
Pre-Scandal	4,197	0.13	433
Post-Scandal	4,169	-0.11	104

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Volkswagen Group is defined as Volkswagen, Audi, and Porsche. Pre-scandal period is January 2011 to August 2015; post-scandal period is September 2015 to August 2016. Sales are measured in units sold. Data come from Ward's Automotive.

To address the concern that the scandal's spillovers were driven by the diesel market, we document that the reputational consequences of the scandal are evident in both diesel and non-diesel sales. Our difference-in-differences estimates in table 3.7 show that the scandal reduced German automaker' sales growth of non-diesel vehicles by 9.6 percentage points and that of the diesel vehicles by 23.3 percentage points. We view the scandal's effect on non-VW automakers' non-diesel sales as evidence that a change in the collective reputation of diesel vehicles cannot be the sole driver of our results.

Table 3.7: U.S. Light Vehicle Sales Growth
German vs. Non-German Manufacturers, Excl. VW Group, By Power Type

Dependent Variable	12-month Log Sales Growth		
Power Type	Baseline	non-Diesel	Diesel
	(1)	(2)	(3)
German × Post-Scandal	-0.104	-0.096	-0.233
	(0.035)	(0.038)	(0.126)
Time Fixed Effects	Yes	Yes	Yes
Make Fixed Effects	Yes	Yes	Yes
R^2	0.292	0.289	0.284
N	2150	2150	428

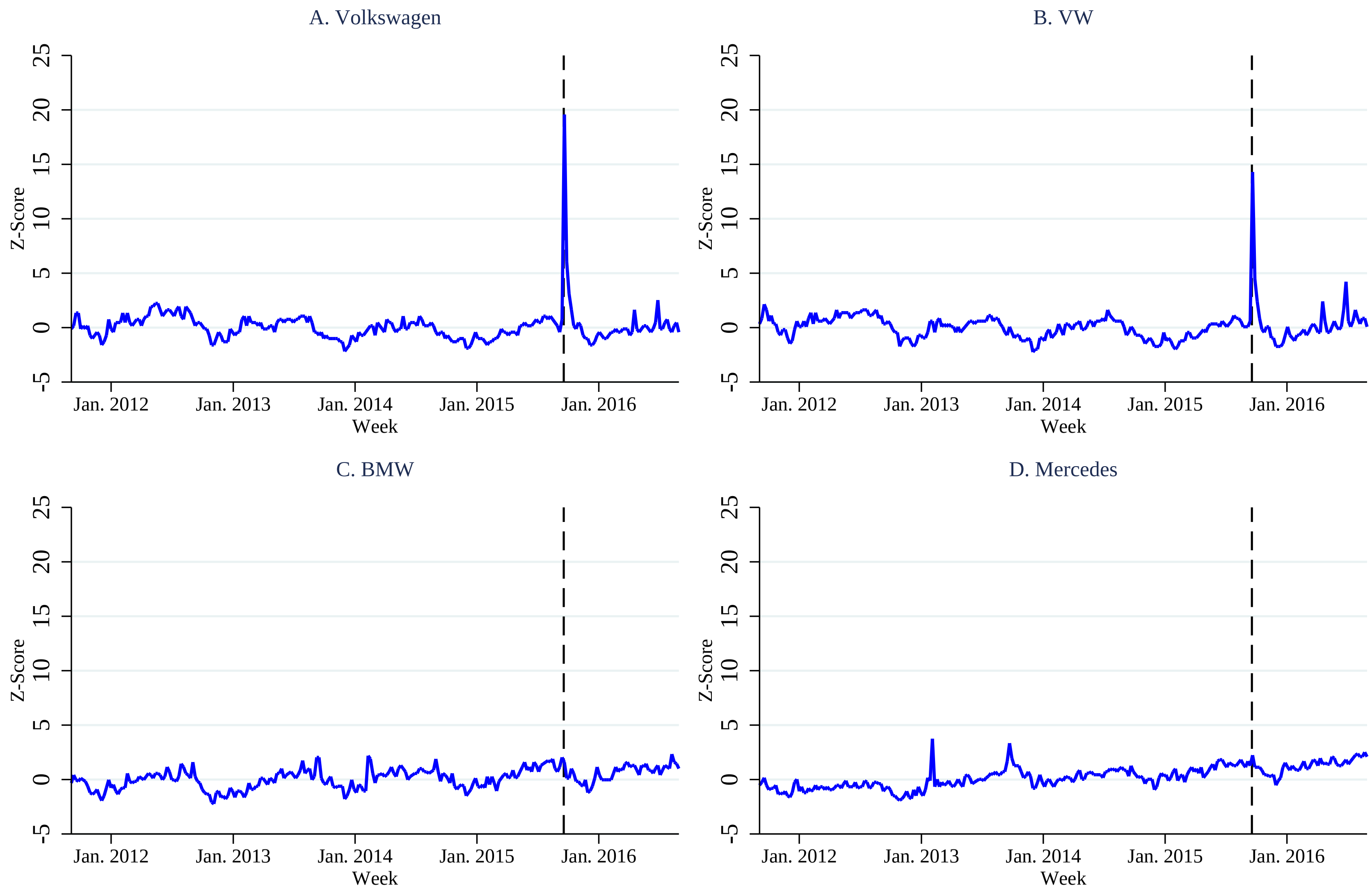
Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Standard errors clustered at vehicle make level in parentheses. Volkswagen Group (Volkswagen, Audi, and Porsche) excluded from all regressions. Volkswagen emissions scandal dated September 18, 2015. Sales are measured in units sold. All regressions include a constant, make and time fixed effects, and are weighted by the square root of sales volumes. Data come from Ward's Automotive.

3.4.5 Information Seeking about German Automakers

Finally, we use internet search data to argue that consumers did not engage in increased information-seeking regarding the non-VW German automakers, which is inconsistent with suspicions of malfeasance similar to VW's. This evidence suggests that the scandal's effects on non-VW German automakers are unlikely to be driven by informational spillovers. Figure 3.7 shows that the general public's information-seeking regarding Volkswagen spiked following the scandal, but did not change noticeably regarding BMW and Mercedes. Each of the four panels plots a time series of a single Google search term ("Volkswagen", "VW", "BMW", and "Mercedes"). The underlying data on Google trends is weekly, and it is scaled so that 100 corresponds to the largest number of searches per week in the search period.[18] We normalize the series and express weekly values as z-scores, deviations from the mean that are scaled by the standard deviation. A z-score equal to 1 indicates a 1-

[18]For weekly data, Google trends only allows users to download a few pre-defined search periods. We chose a five-year window from August 2011 to August 2016.

Figure 3.7: Google Trends of Searches for German Car Manufacturing Firms



Note: Dashed lines show the date of the Volkswagen emissions scandal, dated September 18, 2015. The underlying data on Google trends is weekly and it is scaled so that 100 corresponds to the largest number of searches per week in the search period (August 2011 to August 2016). We construct z-scores as the deviation from the mean of the series, normalized by its standard deviation. The z-scores on the vertical axis are constructed using the mean and standard deviation for each search term over the period prior to September 2015.

standard-deviation increase over the mean. Both the means and the standard deviations are constructed using the period prior to September 2015. Searches for “Volkswagen” and “VW”, in panels A and B, show dramatic increases in information-seeking in the aftermath of the scandal; the week of September 18, the date of the EPA announcement, coincides with z-scores of 22 and 15. By contrast, searches for the two main non-VW German makes, “BMW” in panel C and “Mercedes” in panel D, seem indistinguishable from their regular fluctuations. Together, these panels suggest that rather than precipitating new information-seeking about the individual manufacturers, the scandal changed the collective reputation of German automakers.

3.5 Conclusion

This paper uses the 2015 Volkswagen emissions scandal as a natural experiment to study the economic consequences of collective reputation. By documenting the changing outcomes of non-VW German car manufacturers relative to non-German car manufacturers, we capture the “German engineering” reputation of BMW, Mercedes-Benz, and Smart. We find that non-VW German car manufacturers suffered in three dimensions from the VW scandal: a loss of 76,000 vehicle sales worth roughly \$3.7 billion of revenue; a sharp decline in their stock returns relative to expected market outcomes; and a deterioration in positive public sentiment toward them. We interpret these outcomes as demonstrating the existence a country-specific collective reputation for German car manufacturing. We thus provide empirical support for the theoretical literature on collective reputation and the existence of group reputation externalities.

Our results also contextualize the economic harm of one of the largest U.S. industrial and public health scandals in recent U.S. history. As a complement to the literature on the public-health costs of the excess emissions, we trace the economic spillovers to companies not directly tainted by the scandal. The economically substantial spillovers we document suggest the need to understand what, if any, policy steps are required to address the centrality of national companies to the reputation of the country as a whole. Our results could thus provide an argument for policy instruments that would incentivize large companies to internalize their potential reputation spillovers.

APPENDIX A

Chapter I Supporting Material

A.1 Model Summary

Aggregation

We assume that the manufacturing sector is characterized by a representative establishment selling its output Y in a perfectly competitive market. This firm aggregates the output Y_i of I different industries using a Cobb-Douglas production technology with elasticities θ_i:

$$Y = \prod_{i=1}^{I} Y_i^{\theta_i}, \text{ with } \sum_{i=1}^{I} \theta_i = 1. \tag{A.1}$$

Cost minimization by this aggregating firm implies that θ_i is also each industry's share of aggregate expenditure

$$P_i Y_i = \theta_i P Y, \tag{A.2}$$

where P_i is the price of an industry composite good, and P is the price of the final good

$$P = \prod_{i=1}^{I} \left(\frac{P_i}{\theta_i} \right)^{\theta_i}. \tag{A.3}$$

An industry aggregating firm produces Y_i from the output of N_i differentiated establishments via a constant-elasticity-of-substitution (CES) technology with elasticity σ_i

$$Y_i = \left(\sum_{e=1}^{N_i} Y_{ie}^{\frac{\sigma_i - 1}{\sigma_i}} \right)^{\frac{\sigma_i}{\sigma_i - 1}}. \tag{A.4}$$

Cost minimization by the industry aggregating firm implies a standard CES price index P_i:

$$P_i = \left[\sum_{e=1}^{N_i} \left(\frac{1}{P_{ie}} \right)^{\sigma_i - 1} \right]^{\frac{-1}{\sigma_i - 1}}. \tag{A.5}$$

Establishment Optimization

Each establishment in the industry produces value-added output Y_{ie} by combining its TFP A_{ie}, capital K_{ie} and labor L_{ie} in a Cobb-Douglas production function

$$Y_{ie} = A_{ie} K_{ie}^{\alpha_{K_i}} L_{ie}^{\alpha_{L_i}}, \tag{A.6}$$

where the industry level returns to scale α_i are the sum of the output elasticities α_{K_i} and α_{L_i}. The establishment maximizes profits by taking as given the prices R and w from perfectly competitive input markets. However, the effective cost of an input varies across establishments, with the $\tau_{K_{ie}}$ and $\tau_{L_{ie}}$ capturing these input-specific distortions for capital and labor, respectively

$$\pi_{ie} = P_{ie} Y_{ie} - (1 + \tau_{L_{ie}}) w L_{ie} - (1 + \tau_{K_{ie}}) R K_{ie}. \tag{A.7}$$

By internalizing the demand for its variety, the establishment charges a price that is a constant markup over its marginal cost. Note that the marginal cost under variable RTS depends on the scale of production:

$$\begin{aligned} P_{ie} &= \Omega_{P_i} \left[\frac{(1+\tau_{K_{ie}})^{\alpha_{K_i}} (1+\tau_{L_{ie}})^{\alpha_{L_i}}}{A_{ie}} \right]^{\frac{1}{\alpha_i + \sigma_i(1-\alpha_i)}} \\ \text{where } \Omega_{P_i} &= \left(P_i^{\sigma} Y_i \right)^{\frac{1-\alpha_i}{\alpha_i + \sigma_i(1-\alpha_i)}} \left[\left(\frac{\sigma_i}{\sigma_i - 1} \right)^{\alpha_i} \left(\frac{R}{\alpha_{K_i}} \right)^{\alpha_{K_i}} \left(\frac{w}{\alpha_{L_i}} \right)^{\alpha_{L_i}} \right]^{\frac{1}{\alpha_i + \sigma_i(1-\alpha_i)}} \\ P_{ie} &= \frac{\sigma_i}{\sigma_i - 1} \left[\left(\frac{R}{\alpha_{K_i}} \right)^{\alpha_{K_i}} \left(\frac{w}{\alpha_{L_i}} \right)^{\alpha_{L_i}} \right]^{\frac{1}{\alpha_i}} \left(Y_{ie} \right)^{\frac{1-\alpha_i}{\alpha_i}} \left[\frac{(1+\tau_{K_{ie}})^{\alpha_{K_i}} (1+\tau_{L_{ie}})^{\alpha_{L_i}}}{A_{ie}} \right]^{\frac{1}{\alpha_i}}. \end{aligned} \tag{A.8}$$

Within the confines of this model, there is a natural restriction on the returns to scale parameter. As in Basu and Fernald (1997), standard cost-minimization requires that the RTS parameter α_i is (weakly) less than the markup $\sigma_i/(\sigma_i - 1)$. The returns to scale and the markup shape the price elasticities of supply and demand, respectively. The price elasticity of supply is increasing in the RTS parameter α_i: when RTS are sufficiently large, the supply curve becomes downward sloping. The restriction that α_i is smaller than the markup guarantees that a downward-sloping supply curve is not steeper than a downward-sloping demand curve. This restriction ensures that the willingness-to-pay reflected in the demand curve exceeds the cost of production embodied by the supply curve when establishments are deciding whether to produce. A rearrangement of this inequality guarantees that the often-recurring term $[\alpha_i + \sigma_i(1-\alpha_i)]$ is positive.

An establishment facing larger distortions uses less capital and labor.

$$K_{ie} \propto \left[\frac{A_{ie}^{\sigma_i-1}}{(1+\tau_{K_{ie}})^{[\alpha_i+\sigma_i(1-\alpha_i)]+\alpha_{K_i}(\sigma_i-1)}(1+\tau_{L_{ie}})^{\alpha_{L_i}(\sigma-1)}}\right]^{\frac{1}{\alpha_i+\sigma(1-\alpha_i)}} \tag{A.9}$$

$$L_{ie} \propto \left[\frac{A_{ie}^{\sigma_i-1}}{(1+\tau_{K_{ie}})^{\alpha_{K_i}(\sigma_i-1)}(1+\tau_{L_{ie}})^{[\alpha_i+\sigma(1-\alpha_i)]+\alpha_{L_i}(\sigma_i-1)}}\right]^{\frac{1}{\alpha_i+\sigma(1-\alpha_i)}}. \tag{A.10}$$

Moreover, measured in terms of either physical output or the establishment's revenue share in the industry, a more distorted establishment is also smaller in size.

$$\frac{P_{ie}Y_{ie}}{P_iY_i} = \frac{\left[A_{ie}\left(\frac{\frac{1}{1+\tau_{K_{ie}}}}{\frac{1}{1+\tau_{K,i}}}\right)^{\alpha_{K_i}}\left(\frac{\frac{1}{1+\tau_{L_{ie}}}}{\frac{1}{1+\tau_{L,i}}}\right)^{\alpha_{L_i}}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}{\sum_{e=1}^{N_i}\left[A_{ie}\left(\frac{\frac{1}{1+\tau_{K_{ie}}}}{\frac{1}{1+\tau_{K,i}}}\right)^{\alpha_{K_i}}\left(\frac{\frac{1}{1+\tau_{L_{ie}}}}{\frac{1}{1+\tau_{L,i}}}\right)^{\alpha_{L_i}}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}. \tag{A.11}$$

Marginal Revenue Products and Market Clearing

Distortions affect establishment choices by changing the marginal revenue gained from an additional unit of an input (e.g. $MRPK_{ie}$ for capital K_{ie}). In equilibrium, the marginal revenue product of an additional hired input equals the effective cost to the establishment of hiring the input. If an establishment faces barriers that make acquiring capital more expensive, then $(1+\tau_{K_{ie}})$ is high, and the establishment will only hire an additional unit of capital if its $MRPK_{ie}$ exceeds the cost $(1+\tau_{K_{ie}})R$. The same reasoning holds for all variable inputs in production.

$$MRPK_{ie} \triangleq MPK_{ie} \times P_{ie} \times \frac{\sigma_i - 1}{\sigma_i} = \alpha_{K_i}\frac{Y_{ie}}{K_{ie}}P_{ie}\frac{\sigma_i-1}{\sigma_i} = (1+\tau_{K_{ie}})R \tag{A.12}$$

$$MRPL_{ie} \triangleq MPL_{ie} \times P_{ie} \times \frac{\sigma_i - 1}{\sigma_i} = \alpha_{L_i}\frac{Y_{ie}}{L_{ie}}P_{ie}\frac{\sigma_i-1}{\sigma_i} = (1+\tau_{L_{ie}})w. \tag{A.13}$$

To understand the impact of establishment-level distortions for the productivity of the industry as a whole, we need to aggregate the establishment choices. Combining input-market-clearing conditions with establishment input choices, we can show that each industry uses capital and labor in proportion to the industry's share of the national economy θ_i, to the industry's input elasticity α_{X_i} for a given factor X, and in inverse proportion to that factor's average marginal revenue products across the industry's establishments $\overline{MRPX_i}$.

$$K_i = K \frac{\alpha_{K_i}\theta_i \frac{1}{\overline{MRPK_i}}}{\sum_{i'=1}^{I} \alpha_{K_{i'}}\theta_{i'} \frac{1}{\overline{MRPK_{i'}}}} \tag{A.14}$$

$$L_i = L \frac{\alpha_{L_i}\theta_i \frac{1}{\overline{MRPL_i}}}{\sum_{i'=1}^{I} \alpha_{L_{i'}}\theta_{i'} \frac{1}{\overline{MRPL_{i'}}}}. \tag{A.15}$$

The average marginal revenue products are weighted by establishment size. In the absence of distortions, or if all establishments faced the same distortion, $MRPX_{ie}$ would be equal across establishments and hence equal to the industry $\overline{MRPX_i}$. We revisit this point below when we define a counterfactual allocation of resources in which all establishments are equally distorted.

$$\frac{1}{\overline{MRPK_i}} = \sum_{f=1}^{N_i} \frac{1}{MRPK_{ie}} \frac{P_{ie}Y_{ie}}{P_iY_i} = \frac{1}{R} \sum_{f=1}^{N_i} \frac{1}{(1+\tau_{K_{ie}})} \frac{P_{ie}Y_{ie}}{P_iY_i} \tag{A.16}$$

$$\frac{1}{\overline{MRPL_i}} = \sum_{f=1}^{N_i} \frac{1}{MRPL_{ie}} \frac{P_{ie}Y_{ie}}{P_iY_i} = \frac{1}{w} \sum_{f=1}^{N_i} \frac{1}{(1+\tau_{L_{ie}})} \frac{P_{ie}Y_{ie}}{P_iY_i}. \tag{A.17}$$

Much like the above definitions of average $MRPX$ in the industry, we simplify the notation for the average distortion in an industry by defining

$$\overline{(1+\tau_{X_i})} = \left[\sum_{e=1}^{N_i} \frac{P_{ie}Y_{ie}}{P_iY_i} \frac{1}{1+\tau_{X_{ie}}} \right]^{-1} \text{ for } X \in \{K, L\}.$$

Toward a Measure of Industry Productivity

Industry output can now be expressed as

$$Y_i = A_i K_i^{\alpha_{K_i}} L_i^{\alpha_{L_i}}, \tag{A.18}$$

where A_i is the total factor productivity TFP_i of the industry. In thinking about how distortions affect industry productivity, we introduce notation based on Foster, Haltiwanger and Syverson (2008) and Hsieh and Klenow (2009) that distinguishes the productivity for producing a quantity of physical goods, A_{ie}, from the productivity for generating revenue, $TFPR_{ie}$.

$$TFPR_{ie} \triangleq P_{ie}A_{ie} = \frac{P_{ie}Y_{ie}}{K_{ie}^{\alpha_{K_i}} L_{ie}^{\alpha_{L_i}}}. \tag{A.19}$$

This distinction is helpful since two establishments with the same physical productivity A_{ie} can have different revenue productivities $TFPR_{ie}$ if they face different distortions. In other words, $TFPR$ can help summarize the impact of distortions on an establishment:

$$TFPR_{ie} = \left(\frac{\sigma_i}{\sigma_i - 1}\right)^{\alpha_i} \left(P_{ie}Y_{ie}\right)^{1-\alpha_i} \left[\frac{MRPK_{ie}}{\alpha_{K_i}}\right]^{\alpha_{K_i}} \left[\frac{MRPL_{ie}}{\alpha_{L_i}}\right]^{\alpha_{L_i}} \tag{A.20}$$

$$TFPR_{ie} \propto \left[(1+\tau_{K_{ie}})^{\alpha_{K_i}}(1+\tau_{L_{ie}})^{\alpha_{L_i}} A_{ie}^{(\sigma_i-1)(1-\beta_i)}\right]^{\frac{1}{\alpha_i+\sigma(1-\alpha_i)}}. \tag{A.21}$$

Revenue productivity increases in the level of distortions, as the establishment's input bundle has to compensate for a large effective cost of hiring the inputs.

We can define an industry revenue productivity following the establishment definition:

$$\overline{TFPR}_i \triangleq P_iA_i = \left(\frac{\sigma}{\sigma - 1}\right)^{\beta_i} \left(P_iY_i\right)^{1-\beta_i} \left[\frac{\overline{MRPK}_i}{\alpha_{K_i}}\right]^{\alpha_{K_i}} \left[\frac{\overline{MRPL}_i}{\alpha_{L_i}}\right]^{\alpha_{L_i}}. \tag{A.22}$$

This formulation of industry revenue productivity allows us to write industry TFP_i as the CES aggregate of establishment physical productivity A_{ie}, weighted by the difference between industry and establishment revenue productivity $\overline{TFPR}_i/TFPR_{ie}$.

$$TFP_i = P_iA_i\frac{1}{P_i} = \overline{TFPR}_i\frac{1}{P_i} = \left[\sum_{e=1}^{M_i}\left(A_{ie}\frac{\overline{TFPR}_i}{TFPR_{ie}}\right)^{\sigma-1}\right]^{\frac{1}{\sigma-1}}. \tag{A.23}$$

The weight captures the establishment's size as well as the deviations of its marginal revenue products from their respective industry averages:

$$\frac{\overline{TFPR}_i}{TFPR_{ie}} = \left(\frac{P_iY_i}{P_{ie}Y_{ie}}\right)^{1-\alpha_i} \left[\frac{\overline{MRPK}_i}{MRPK_{ie}}\right]^{\alpha_{K_i}} \left[\frac{\overline{MRPL}_i}{MRPL_{ie}}\right]^{\alpha_{L_i}} \tag{A.24}$$

$$\frac{\overline{TFPR}_i}{TFPR_{ie}} = \left(s_{ie}\right)^{\alpha_i-1} \left(\frac{\overline{1+\tau_{K,i}}}{1+\tau_{K_{ie}}}\right)^{\alpha_{K_i}} \left(\frac{\overline{1+\tau_{L,i}}}{1+\tau_{L_{ie}}}\right)^{\alpha_{L_i}}. \tag{A.25}$$

Misallocation

More distorted establishments have smaller weights in industry productivity. Consequently, the correlation of productivity and distortion is important for measuring gains from equalizing the distortions faced by different establishments within the industry. If more productive establishments are also more distorted, then equalizing distortions would give larger weights to the more productive establishments in the counterfactual. This tilting of weights toward more productive establishments would translate to large TFP gains from reallocating inputs.

More formally, if all establishments within an industry face the same distortions, so that $\tau = \overline{\tau}$, then the establishment weights for calculating industry TFP_i simplify in the following manner:

$$\left.\frac{\overline{TFPR_i}}{TFPR_{ie}}\right|_{\tau=\overline{\tau}} = \left(s_{ie}\big|_{\tau=\overline{\tau}}\right)^{\alpha_i-1} = \left(\frac{\left[A_{ie}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}{\sum_{e=1}^{N_i}\left[A_{ie}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}\right)^{\alpha_i-1}. \tag{A.26}$$

Note that under constant returns to scale ($\alpha_i = 1$) $TFPR_{ie}$ is identical across all establishments. This equality is at the center of Hsieh and Klenow (2009) intuition: "A key result we exploit is that *revenue* productivity... should be equated across firms in the absence of distortions. To the extent revenue productivity differs across firms, we can use it to recover a measure of firm-level distortions" (1404). Note, however, that if returns to scale in an industry are not constant, then revenue productivity can vary across undistorted establishments. As a result, there is not a direct mapping between the variance of TFPR and the misallocation within industry. To calculate the gains from eliminating distortions, the econometrician has to calculate the counterfactual weight for each establishment.

For every industry i, we then define misallocation as Φ_i, the net gain to industry TFP from equalizing distortions across establishments within the industry:

$$\Phi_i = \frac{TFP_i\big|_{\tau=\bar{\tau}}}{TFP_i} = \frac{\left[\sum_{e=1}^{N_i}\left(A_{ie}\left.\frac{\overline{TFPR_i}}{TFPR_{ie}}\right|_{\tau=\bar{\tau}}\right)^{\sigma_i-1}\right]^{\frac{1}{\sigma_i-1}}}{\left[\sum_{e=1}^{N_i}\left(A_{ie}\frac{\overline{TFPR_i}}{TFPR_{ie}}\right)^{\sigma_i-1}\right]^{\frac{1}{\sigma_i-1}}}. \tag{A.27}$$

The misallocation for all of US manufacturing in a given year is then

$$\Phi = \prod_{i \in I} \Phi_i^{\theta_i}, \tag{A.28}$$

where θ_i is industry i's revenue share in the manufacturing sector.

A.2 Heterogeneous Markups within Industry

In this appendix, we generalize our model to allow markups to vary across establishments in an industry. We introduce these heterogeneous markups by replacing monopolistic competition in output markets with oligopolistic competition, in the style of Atkeson and Burstein (2008). In short, we allow establishments to internalize their impact on the industry demand, leading them to change their price-setting behavior, with larger establishments now charging higher markups.

Previously, establishments internalized their own downward-sloping demand curves:

$$P_{ie} = P_i Y_i Y_i^{\frac{1-\sigma_i}{\sigma_i}} Y_{ie}^{\frac{-1}{\sigma_i}}. \tag{B.1}$$

Now they also internalize the demand for the industry aggregate, so we can write an individual establishment's demand curve as

$$P_{ie} = \theta_i PYY_i^{\frac{1-\sigma_i}{\sigma_i}} Y_{ie}^{\frac{-1}{\sigma_i}}. \tag{B.2}$$

Profit maximization on the part of these oligopolistic establishments leads to an updated expression for the equilibrium price, which is still a markup over marginal cost:

$$P_{ie} = \frac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1}\left[\left(\frac{R}{\alpha_{K_i}}\right)^{\alpha_{K_i}}\left(\frac{w}{\alpha_{L_i}}\right)^{\alpha_{L_i}}\right]^{\frac{1}{\alpha_i}}\left(Y_{ie}\right)^{\frac{1-\alpha_i}{\alpha_i}}\left[\frac{(1+\tau_{K_{ie}})^{\alpha_{K_i}}(1+\tau_{L_{ie}})^{\alpha_{L_i}}}{A_{ie}}\right]^{\frac{1}{\alpha_i}}. \tag{B.3}$$

The establishment-specific markup $\varepsilon(s_{ie})/(\varepsilon(s_{ie})-1)$ is now based on the elasticity $\varepsilon(s_{ie})$, whose inverse is defined as the weighted average of inverses of the industry CES elasticity of substitution σ_i and of the aggregate economy's Cobb-Douglas elasticity 1.

$$\frac{1}{\varepsilon(s_{ie})} = \frac{1}{\sigma_i}(1-s_{ie}) + s_{ie} \tag{B.4}$$

$$\frac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1} = \frac{\sigma_i}{\sigma_i-1}\frac{1}{1-s_{ie}}. \tag{B.5}$$

Larger establishments charge higher markups:

$$\frac{\partial\frac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1}}{\partial s_{ie}} = \left[\frac{1}{\varepsilon(s_{ie})-1} - \frac{\varepsilon(s_{ie})}{(\varepsilon(s_{ie})-1)^2}\right]\frac{\partial\varepsilon(s_{ie})}{\partial s_{ie}} = \frac{\sigma_i-1}{\sigma_i}\left[\frac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1}\right]^2 > 0. \tag{B.6}$$

Working through the model, we show that the establishment size now depends on the establishment markup:

$$s_{ie} = \frac{P_{ie}Y_{ie}}{P_iY_i} = \frac{\left[\left(\dfrac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1}\right)^{-\alpha_i} \dfrac{A_{ie}}{(1+\tau_{K_{ie}})^{\alpha_{K_i}}(1+\tau_{L_{ie}})^{\alpha_{L_i}}}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}{\sum_{i=1}^{N_i}\left[\left(\dfrac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1}\right)^{-\alpha_i} \dfrac{A_{ie}}{(1+\tau_{K_{ie}})^{\alpha_{K_i}}(1+\tau_{L_{ie}})^{\alpha_{L_i}}}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}. \tag{B.7}$$

To calculate misallocation in this generalized model, we derive the scaling factors with and without distortions. The scaling factors defined by the relative revenue productivity now depend on $\widetilde{MRPX}$, the average marginal revenue products that are scaled by the establishment-specific markups:

$$\frac{\overline{TFPR}_i}{TFPR_{ie}} = \left(\frac{P_iY_i}{P_{ie}Y_{ie}}\right)^{1-\beta} \left[\frac{\widetilde{MRPK}_i}{MRPK_{ie}\dfrac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1}}\right]^{\alpha_{K_i}} \left[\frac{\widetilde{MRPL}_i}{MRPL_{ie}\dfrac{\varepsilon(s_{ie})}{\varepsilon(s_{ie})-1}}\right]^{\alpha_{L_i}} \tag{B.8}$$

$$\frac{\overline{TFPR}_i}{TFPR_{ie}} = \left(\frac{P_iY_i}{P_{ie}Y_{ie}}\right)^{1-\beta} \left[\frac{\dfrac{K_{ie}}{P_{ie}Y_{ie}}}{\sum_{e=1}^{N_i}\dfrac{P_{ie}Y_{ie}}{P_iY_i}\dfrac{K_{ie}}{P_{ie}Y_{ie}}}\right]^{\alpha_{K_i}} \left[\frac{\dfrac{L_{ie}}{P_{ie}Y_{ie}}}{\sum_{e=1}^{N_i}\dfrac{P_{ie}Y_{ie}}{P_iY_i}\dfrac{L_{ie}}{P_{ie}Y_{ie}}}\right]^{\alpha_{L_i}}, \tag{B.9}$$

where the last expression above is now entirely in terms of data, making it straightforward to implement. In the absence of distortions, we can write the scaling factor as a function solely of the relative size in the absence of distortions $s_{ie}|_{\tau=\bar{\tau}}$:

$$\left.\frac{\overline{TFPR}_i}{TFPR_{ie}}\right|_{\tau=\bar{\tau}} = \left(s_{ie}|_{\tau=\bar{\tau}}\right)^{\alpha_i-1} \left[\frac{\left(1-s_{ie}|_{\tau=\bar{\tau}}\right)}{\sum_{e=1}^{N_i} s_{ie}|_{\tau=\bar{\tau}}\left(1-s_{ie}|_{\tau=\bar{\tau}}\right)}\right]^{\alpha_i} \tag{B.10}$$

$$\text{where } s_{ie}|_{\tau=\bar{\tau}} = \frac{\left[(1-s_{ie}|_{\tau=\bar{\tau}})^{\alpha_i}A_{ie}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}{\sum_{i=1}^{N_i}\left[(1-s_{ie}|_{\tau=\bar{\tau}})^{\alpha_i}A_{ie}\right]^{\frac{1}{\frac{\sigma_i}{\sigma_i-1}-\alpha_i}}}. \tag{B.11}$$

A.3 Demand Shocks and Misallocation

Our measure of misallocation is based on a counterfactual in which we change distortions but keep fundamentals (e.g., tastes/demand, productivity, etc.) unchanged. We show that our residual-based measure of establishment productivity $\widehat{A_{ie}}$ would conflate productivity A_{ie} and demand in an augmented model where we allow for establishment-specific taste parameters ψ_{ie}. We then show that we would correctly calculate misallocation even when we cannot separately measure productivity and tastes in the residual $\widehat{A_{ie}}$. In short, the measure of misallocation requires us to capture this combined object of productivity and demand; it does not require us to separate the two.

If we allowed for establishment-specific taste parameters, our residual $\widehat{A_{ie}}$ would be a product of the establishment productivity and the taste parameter. We show this by modifying the industry CES aggregator from equation (1.2) to include establishment-specific taste shifters ψ_{ie}:

$$Y_i = \left(\sum_{e=1}^{N_i} (\psi_{ie} Y_{ie})^{\frac{\sigma_i - 1}{\sigma_i}} \right)^{\frac{\sigma_i}{\sigma_i - 1}} . \tag{B.12}$$

In this augmented model, the demand for an establishment's revenue depends on the consumer's tastes for the variety in question:

$$P_{ie} Y_{ie} = P_i Y_i^{\frac{1}{\sigma_i}} (\psi_{ie} Y_{ie})^{\frac{\sigma_i - 1}{\sigma_i}} \tag{B.13}$$

Following the standard process for backing out the residual $\widehat{A_{ie}}$, we now back out a term that conflates productivity A_{ie} and the taste shifter ψ_{ie}:

$$\widehat{A_{ie}} = \frac{(P_{ie} Y_{ie})^{\frac{\sigma_i}{\sigma_i - 1}}}{K_{ie}^{\alpha_{K_i}} L_{ie}^{\alpha_{L_i}}} = \kappa_i \psi_{ie} A_{ie} \tag{B.14}$$

Since productivity and taste parameters always enter multiplicatively in the expression for misallocation, we would calculate misallocation correctly even though we could not separately measure productivity and demand shocks. We note first that the relative revenue productivity is unchanged from its expression in the baseline model:

$$\frac{\overline{TFPR}_i}{TFPR_{ie}} = \left(\frac{P_i Y_i}{P_{ie} Y_{ie}} \right)^{1-\beta} \left[\frac{\overline{\left(\frac{K_{ie}}{P_{ie} Y_{ie}} \right)}}{\left(\frac{K_{ie}}{P_{ie} Y_{ie}} \right)} \right]^{-\alpha_{K_i}} \left[\frac{\overline{\left(\frac{L_{ie}}{P_{ie} Y_{ie}} \right)}}{\left(\frac{L_{ie}}{P_{ie} Y_{ie}} \right)} \right]^{-\alpha_{L_i}} . \tag{B.15}$$

When we reallocate inputs to equalize distortions across establishments, the relative revenue productivity now depends on the product $\psi_{it} A_{ie}$:

$$\left.\frac{\overline{TFPR}_i}{TFPR_{ie}}\right|_{\tau=\bar{\tau}} = \left[\frac{\sum_{e=1}^{N_i} (\psi_{ie} A_{ie})^{\frac{1}{\frac{\sigma_i}{\sigma_i - 1} - \beta}}}{(\psi_{ie} A_{ie})^{\frac{1}{\frac{\sigma_i}{\sigma_i - 1} - \beta}}} \right]^{1-\beta} \tag{B.16}$$

Putting these pieces together, we show that we can calculate misallocation using the residual $\widehat{A_{ie}}$ even when we cannot separately measure productivity and demand shocks:

$$\Phi_i = \frac{TFP_i|_{\tau=\bar{\tau}}}{TFP_i} = \frac{\left[\sum_{e=1}^{N_i} \left((\psi_{ie} A_{ie}) \left.\frac{\overline{TFPR}_i}{TFPR_{ie}}\right|_{\tau=\bar{\tau}} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}}{\left[\sum_{e=1}^{N_i} \left((\psi_{ie} A_{ie}) \frac{\overline{TFPR}_i}{TFPR_{ie}} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}}. \tag{B.17}$$

A.4 Measurement Error in Bils, Klenow and Ruane (2017)

Bils, Klenow and Ruane (2017), henceforth BKR, highlight the possibility that measurement error could be misinterpreted as misallocation in microdata. They propose a correction for *additive* measurement error in establishment revenue R and input bundles I. Their estimates suggest that measurement error has increased in U.S. Census microdata, and that accounting for this change in measurement error eliminates the upward trend in misallocation from a gross-output Hsieh-Klenow model.

In this appendix, we show that deviations from constant returns to scale look like measurement error in the BKR procedure, and that a decline in returns to scale over time looks like an increase in measurement error. Informally, our argument emphasizes two points. First, a procedure that does not explicitly account for *multiplicative* measurement error will pick up this multiplicative measurement error as *additive* measurement error. Second, overlooking deviations from constant returns to scale leads to multiplicative measurement error in the input bundle. For instance, if the true returns to scale in an industry were α_i, then the input bundle under constant returns to scale $I_{crts,ie}$ would relate to the true input bundle I_{ie}, as shown below. As a result, the BKR procedure could interpret deviations from constant returns to scale as measurement error.

$$I_{crts,ie} = \underbrace{I_{ie}^{\frac{1-\alpha_i}{\alpha_i}}}_{multiplicative\ measurement\ error} I_{ie}$$

$$I_{crts,ie} = I_{ie} + \underbrace{[I_{ie}^{\frac{1-\alpha_i}{\alpha_i}} - 1]I_{ie}}_{additive\ measurement\ error} .$$

Formally, we focus on the key parameter λ in BKR estimating equation [2], reproduced below. BKR show that $\lambda = 1$ if there is no misallocation. Larger deviations from unity indicate a greater extent of measurement error. The key estimating equation relates the time-series change in revenue, R, to the change in the input bundle I, the revenue productivity $TFPR$, and the product of I and $TFPR$. Both I and $TFPR$ depend on the assumed returns to scale. The measure of change Δ defined as the "growth rate of a plant variable relative to the mean of its sector."

$$\Delta\widehat{R} = \Psi \cdot \Delta\widehat{I} + \Phi \cdot f(\ln TFPR) + \Psi(1-\lambda) \cdot \Delta\widehat{I} \cdot g(\ln TFPR).$$

We derive the below relationship between λ_{crts}, estimated under assumed constant returns to scale, and true λ, where $g(\cdot)$ is some polynomial. In short, the BKR procedure correctly captures measurement error under one of two conditions: either $\lambda = 1$, so there is no measurement error, or $\alpha_i = 1$, so that the assumed constant returns to scale hold in the data. As a result, if there is any measurement error in the data, the BKR estimates can conflate measurement error with model misspecification.

$$\lambda_{crts} = \lambda + (1-\lambda)[1-\gamma] \text{ where } \gamma = \frac{g\left(\ln R_{ie} - \alpha_i \ln I_{crts,ie}\right)}{g\left(\ln R_{ie} - \ln I_{crts,ie}\right)}.$$

We show that the mismeasurement of λ varies predictably with returns to scale α_i. Since $\widehat{\lambda}_{crts}$ in BKR takes values between 0.095 and 0.358 for U.S. data, we focus entirely on the case of $\lambda < 1$. In short, if λ_{crts} is closer to 1 than is λ, then we understate measurement error when we impose constant returns to scale. Indeed, this is the case when returns to scale are increasing: when $\alpha_i > 1$, then $1 > \lambda_{crts} > \lambda$, and we understate misallocation. By contrast, when returns to scale are decreasing, then we overstate misallocation, since $\alpha_i < 1$ leads to $1 > \lambda > \lambda_{crts}$.

Consider a setting in which measurement error does not change over time, but returns to scale decline from increasing to constant; imposing constant returns to scale in this setting would lead us to infer an increase in measurement error, even though no such increase has taken place. As detailed in the previous paragraph, overlooking increasing returns to scale leads us to understate measurement error. As returns to scale decline over time, our estimate of measurement error asymptotes to its true value from below. In short, imposing constant returns to scale here would lead us to understate measurement error early in the period and to see this measurement error grow toward its true value over time. However, by assumption, true measurement error has not changed; we only see it grow as the bias from imposing constant returns declines over time.

With the caveat that our estimates of returns to scale are for a value-added world, while BKR work in a gross-output world, we present the BKR estimates of λ_{crts} and our estimates of returns to scale α_i. By the arguments above, it is possible that a decline in returns to scale could explain the increase in measurement error over time that BKR find. If, as a result, there has not been a substantial change in measurement error over time, then measurement error is less capable of explaining the upward trend in misallocation.

Table A.1: U.S. Manufacturing – Division of Value Added

	1978-1982	1983-1987	1988-1992	1993-1997	1998-2002	2003-2007
λ_{crts}	0.358	0.336	0.326	0.326	0.192	0.095
$\alpha_{average}$	1.23	1.20	1.20	1.12	1.11	0.96

Supporting Algebra

- Key measurement objects:

$$\ln TFPR_{crts,ie} = \ln TFPR_{ie} + \frac{\alpha_i - 1}{\alpha_i} \ln I_{ie} = \ln TFPR_{ie} + (\alpha_i - 1) \ln I_{crts,ie}$$

$$\Delta \widehat{I_{crts,ie}} = \ln \left[\frac{\frac{I_{crts,ie,t}}{\overline{I_{crts,i,t}}}}{\frac{I_{crts,ie,t-1}}{\overline{I_{crts,i,t-1}}}} \right] = \ln \left[\frac{\frac{I_{ie,t} I_{ie,t}^{\frac{1-\alpha_i}{\alpha_i}}}{\overline{I_{crts,i,t}}}}{\frac{I_{ie,t-1} I_{ie,t-1}^{\frac{1-\alpha_i}{\alpha_i}}}{\overline{I_{crts,i,t-1}}}} \right] = \ln \left[\frac{\frac{I_{ie,t}}{\overline{I_{i,t}}}}{\frac{I_{ie,t-1}}{\overline{I_{i,t-1}}}} \frac{\frac{I_{ie,t}^{\frac{1-\alpha_i}{\alpha_i}}}{\left(\overline{I_{i,t}}\right)^{\frac{1-\alpha_i}{\alpha_i}}}}{\frac{I_{ie,t-1}^{\frac{1-\alpha_i}{\alpha_i}}}{\left(\overline{I_{i,t-1}}\right)^{\frac{1-\alpha_i}{\alpha_i}}}} \frac{\frac{\left(\overline{I_{i,t}}\right)^{\frac{1}{\alpha_i}}}{\overline{I_{crts,i,t}}}}{\frac{\left(\overline{I_{i,t-1}}\right)^{\frac{1}{\alpha_i}}}{\overline{I_{crts,i,t-1}}}} \right]$$

$$\Delta \widehat{I_{crts,ie}} = \Delta \widehat{I_{ie}} + \frac{1-\alpha_i}{\alpha_i} \Delta \widehat{I_{ie}} + \Omega_{\Delta I}, \text{ where } \Omega_{\Delta I} = \ln \left[\frac{\frac{\left(\overline{I_{i,t}}\right)^{\frac{1}{\alpha_i}}}{\overline{I_{crts,i,t}}}}{\frac{\left(\overline{I_{i,t-1}}\right)^{\frac{1}{\alpha_i}}}{\overline{I_{crts,i,t-1}}}} \right]$$

Rearranging, we have:

$$\ln TFPR_{ie} = \ln TFPR_{crts,ie} - \frac{\alpha_i - 1}{\alpha_i} \ln I_{ie} = \ln TFPR_{crts,ie} - (\alpha_i - 1) \ln I_{crts,ie}$$
$$\Delta \widehat{I_{ie}} = \alpha_i \Delta \widehat{I_{crts,ie}} - \alpha_i \Omega_{\Delta I}$$

- I will now plug these rearranged objects into the main estimating equation to derive the parameters Ψ_{crts}, Φ_{crts}, and $\Psi_{crts}(1-\lambda_{crts})$:

$$\begin{aligned}
\Delta \widehat{R} &= \Psi \cdot \Delta \widehat{I} + \Phi \cdot f(\ln TFPR) + \Psi(1-\lambda) \cdot \Delta \widehat{I} \cdot g(\ln TFPR) \\
&= \Psi \cdot \left[\alpha_i \Delta \widehat{I_{crts,ie}} - \alpha_i \Omega_{\Delta I} \right] + \Phi \cdot f \left(\ln TFPR_{crts,ie} - (\alpha_i - 1) \ln I_{crts,ie} \right) \\
&+ \Psi(1-\lambda) \cdot \left[\alpha_i \Delta \widehat{I_{crts,ie}} - \alpha_i \Omega_{\Delta I} \right] \cdot g \left(\ln TFPR_{crts,ie} - (\alpha_i - 1) \ln I_{crts,ie} \right) \\
&= \Psi \cdot \left[\alpha_i - \frac{\alpha_i \Omega_{\Delta I}}{\Delta \widehat{I_{crts,ie}}} \right] \Delta \widehat{I_{crts,ie}} + \Phi \cdot \frac{f \left(\ln TFPR_{crts,ie} - (\alpha_i - 1) \ln I_{crts,ie} \right)}{f \left(\ln TFPR_{crts,ie} \right)} f \left(\ln TFPR_{crts,ie} \right) \\
&+ \Psi(1-\lambda) \cdot \left[\alpha_i - \frac{\alpha_i \Omega_{\Delta I}}{\Delta \widehat{I_{crts,ie}}} \right] \cdot \frac{g \left(\ln TFPR_{crts,ie} - (\alpha_i - 1) \ln I_{crts,ie} \right)}{g \left(\ln TFPR_{crts,ie} \right)} \Delta \widehat{I_{crts,ie}} g \left(\ln TFPR_{crts,ie} \right)
\end{aligned}$$
$$\Delta \widehat{R} = \Psi_{crts} \cdot \Delta \widehat{I_{crts}} + \Phi_{crts} \cdot f(\ln TFPR_{crts}) + \Psi_{crts}(1-\lambda_{crts}) \cdot \Delta \widehat{I_{crts}} \cdot g(\ln TFPR_{crts})$$

- How do estimated Ψ_{crts}, Φ_{crts}, and $\Psi_{crts}(1-\lambda_{crts})$ relate to parameters Ψ, Φ, and $\Psi(1-\lambda)$?
  - $\Psi_{crts} = \Psi \cdot \left[\alpha_i - \frac{\alpha_i \Omega_{\Delta I}}{\Delta \widehat{I_{crts,ie}}}\right]$
  - $\Phi_{crts} = \Phi \cdot \frac{f\left(\ln TFPR_{crts,ie} - (\alpha_i - 1)\ln I_{crts,ie}\right)}{f\left(\ln TFPR_{crts,ie}\right)}$
  - $\Psi_{crts}(1-\lambda_{crts}) = \Psi(1-\lambda) \cdot \left[\alpha_i - \frac{\alpha_i \Omega_{\Delta I}}{\Delta \widehat{I_{crts,ie}}}\right] \cdot \frac{g\left(\ln TFPR_{crts,ie} - (\alpha_i - 1)\ln I_{crts,ie}\right)}{g\left(\ln TFPR_{crts,ie}\right)}$
- Without model misspecification, we can back out true λ as

$$\lambda = 1 - \frac{\Psi(1-\lambda)}{\Psi}$$

- With model misspecification, λ_{crts} provides a biased estimate of λ

$$\frac{\Psi_{crts}(1-\lambda_{crts})}{\Psi_{crts}} = \frac{\Psi(1-\lambda)}{\Psi} \frac{\left[\alpha_i - \frac{\alpha_i \Omega_{\Delta I}}{\Delta \widehat{I_{crts,ie}}}\right] \cdot \frac{g\left(\ln TFPR_{crts,ie} - (\alpha_i - 1)\ln I_{crts,ie}\right)}{g\left(\ln TFPR_{crts,ie}\right)}}{\left[\alpha_i - \frac{\alpha_i \Omega_{\Delta I}}{\Delta \widehat{I_{crts,ie}}}\right]}$$

$$1 - \lambda_{crts} = (1-\lambda)\frac{g\left(\ln TFPR_{crts,ie} - (\alpha_i - 1)\ln I_{crts,ie}\right)}{g\left(\ln TFPR_{crts,ie}\right)}$$

$$\lambda_{crts} = 1 + (\lambda - 1)\frac{g\left(\ln TFPR_{crts,ie} - (\alpha_i - 1)\ln I_{crts,ie}\right)}{g\left(\ln TFPR_{crts,ie}\right)}$$

$$\lambda_{crts} = 1 + (\lambda - 1)\frac{g\left(\ln R_{ie} - \ln I_{crts,ie} - (\alpha_i - 1)\ln I_{crts,ie}\right)}{g\left(\ln R_{ie} - \ln I_{crts,ie}\right)}$$

$$\lambda_{crts} = 1 + (\lambda - 1)\frac{g\left(\ln R_{ie} - \alpha_i \ln I_{crts,ie}\right)}{g\left(\ln R_{ie} - \ln I_{crts,ie}\right)}$$

$$\lambda_{crts} = \lambda + (1-\lambda)[1-\gamma] \text{ where } \gamma = \frac{g\left(\ln R_{ie} - \alpha_i \ln I_{crts,ie}\right)}{g\left(\ln R_{ie} - \ln I_{crts,ie}\right)}$$

APPENDIX B

Chapter II Supporting Material

B.1 Data Descriptions

B.1.1 Publicly-Traded Firms

The empirical evidence on public firms is based on a linked data set that has three components. In this appendix we describe the details of the dataset.

The ExecuCompustat provides data on executive compensation. It reports the total realized and estimated compensation of the CEO, CFO, and three other highly paid executives of U.S. public firms in the S&P Composite 1500 Index from 1992 onward.[1] The executive compensation consists of salary, bonus, stock options, long term incentive plans (LTIPs), restricted stock awards, and all others. "Realized" compensation (variable name: TDC2) measures the value of stock option awards at the time of execution, while "estimated" compensation (variable name: TDC1) measures the value of stock options at the time of granting using the Black-Scholes formula.[2]

The confidential Census Bureau databases provide the other key variables needed to measure within-firm inequality and exporting status. The LBD is compiled from the Census Bureau's Business Register, which covers the universe of U.S. firms at the establishment level. We aggregate it up to the firm level and extract annual employment and payroll variables, which are used to compute the average non-executive wage for each firm in a given year. The LBD is linked to the last component of the data set, the LFTTD, using the

[1]The Securities and Exchange Commission (SEC) requires public firms to disclose the total compensation of at least five said executives starting from 1992. Any firm that was once included in the S&P 1500 Index is included in the sample, even if the firm is later dropped from the index. The S&P 1500 Index is the union of three commonly used indices: S&P 500 (LargeCap), S&P MidCap 400 Index, and S&P SmallCap 600 Index. This index covers approximately 90 percent of the total U.S. public firm capitalization.

[2]In 2006, the SEC changed the disclosure rule on executive compensation, which makes the raw data before and after 2006 not directly comparable. The ExecuCompustat data set takes this into account when constructing TDC1 and TDC2 so these two variables can be used for the entire sample.

methods described in McCallum (2013). The LFTTD records the universe of individual international trade transactions made by U.S. firms based on the data collected by U.S. Customs from 1992 onward. It links each export transaction to the U.S. exporting firm and thus provides the base to identify exporting firms in each year. The final linkage between ExecuCompustat and the linked LBD-LFTTD is done through the Compustat-SSEL Bridge provided by the Census Bureau. Table B.1 and B.2 provide summary statistics of the combined data set.

	Matched Data		ExecuCompustat	
Sector	Percent	N.Obs.	Percent	N.Obs.
Mineral & Construction	4.39%	751	5.44%	1876
Manufacturing	46.15%	7892	42.51%	14649
Transportation, Communications and Utilities	10.79%	1845	11.24%	3873
Wholesale and Retail Trade	12.36%	2113	11.49%	3960
Finance, Insurance and Real Estate	13.91%	2379	15.28%	5265
Services	12.40%	2121	14.03%	4835
Other	0.71%	122	0.69%	239
Total	100.00%	17223	100.00%	34697

Table B.1: Sector Composition: Public Firm Sample

Note: This table reports the sectoral composition of the firm-year observations in the linked ExecuCompustat-LBD-LFTTD data set and compares the distribution with the original ExecuCompustat data set. The sector definition is based on a one-digit SIC code.

Mean	Exporters	Non-Exporters	Overall
CEO Compensation, Estimated	4487.7	3254.3	4197.1
CEO Compensation, Realized	4662.4	3340.4	4350.8
CEO-to-worker Pay Ratio, Estimated	91.9	80.8	89.3
CEO-to-worker Pay Ratio, Realized	91.8	79.6	88.9
N. Observations	13169	4054	17223

Table B.2: Summary Statistics: Public Firm Sample

Note: This table reports the mean of key variables of the linked ExecuCompustat-LBD-LFTTD data set. The unit of observation is firm-year. Executive compensations are measured in thousands of U.S. dollars. For the difference between estimated and realized compensation, see Section 2.2.

B.1.2 Privately-Held Firms

The evidence on privately-held firms in the US is based on the linked CIQ-LBD-LFTTD dataset. In this appendix we describe the details of the datasets.

To construct the dataset, we start with executives working in private U.S. firms between 2003 and 2007 from the CIQ data. This yields a data set that contains around 33,000 individuals working in 3,849 privately held firms and 11,706 firm-year level observations. We then link this data set to the Standard Statistical Establishment List (SSEL) in the Census Bureau. Unlike the ExecuCompustat, where the bridge files exist and firms can be matched using standardized identifiers, the CIQ data have not been linked to the census data sets before. Therefore, we carry out a fuzzy match based on name, street address, and zip code. We require that the weighted similarity has to be at least 95 percent for two entries to be considered a match and then hand-screen all the matched records to eliminate obvious errors. The matched CIQ records are then linked with LBD-LFTTD constructed by McCallum (2013).

Table B.3 summarizes the results of the fuzzy merge and compares the distribution of firms across sectors in the linked data set and the original CIQ data. The linked data set contains 6,002 firm-year observations and 2202 unique firms. A total of 3,366 firm-year observations and 1,207 unique firms are exporting firms, while the remaining 2,636 observations with 9,95 unique firms are non-exporters. Overall, 51 percent of the CIQ records are matched with the census data. The sectoral distribution of the CIQ is preserved in the linked data set. For example, manufacturing firms constitute 33.8 percent in the linked data and 34.4 percent in the original data; financial firms are responsible for 22.0 percent in the linked data and 18.9 percent in CIQ.

Instead of the CEO-to-worker pay ratio, we construct the ratio between the highest-paid executive and the non-executive wage as the benchmark measure of intra-firm inequality. The CIQ data does not report standardized job titles, and therefore, constructing the CEO title from the raw data would introduce unnecessary noise. Nevertheless, most of the highest-paid executives are indeed CEOs: in ExecuCompustat, more than 98 percent of the highest-paid executives are the CEOs. There is no strong reason to believe that this ratio will be significantly different in the CIQ sample.

The summary statistics of the top-1-to-worker pay ratio are reported in Table B.4. Overall, within-firm inequality is lower among private firms than among public firms. The top-1-to-worker pay ratio is 37.6 in the private firm sample compared with 89 in the public firm sample. Again, the top-1-to-worker pay ratio varies with exporting status. The ratio is 41.3 among exporters and only 32.8 among non-exporters.

Sector	Matched Data Percent	Matched Data N.Obs.	Capital IQ Percent	Capital IQ N.obs.
Mineral & Construction	3.32%	199	4.13%	483
Manufacturing	33.86%	2032	34.44%	4032
Transportation, Communications and Utilities	10.71%	643	10.23%	1197
Wholesale and Retail Trade	9.30%	558	9.18%	1075
Finance, Insurance and Real Estate	21.98%	1319	18.85%	2206
Services	19.99%	1200	21.80%	2552
Other	0.85%	51	1.38%	161
Total	100.00%	6002	100.00%	11706

Table B.3: Sector Composition: Private Firm Sample

Note: This table reports the sectoral composition of the firm-year observations in the linked CIQ-LBD-LFTTD data set and compares the distribution with the original Capital-IQ data set. The sector definition is based on one-digit SIC code.

Mean	Exporters	Non-Exporters	Overall
Top 1 Compensation, Estimated	2626.9	1731.2	2233.5
Top 1 Compensation, Realized	2157	1522.1	1878.2
Top-1-to-worker Pay Ratio, Estimated	49.8	36.7	44
Top-1-to-worker Pay Ratio, Realized	41.3	32.8	37.6
N. Observations	3366	2636	6002

Table B.4: Summary Statistics: Private Firm Sample

Note: This table reports the mean of key variables of the linked CIQ-LBD-LFTTD data set. The unit of observation is firm-year. Executive compensations are measured in thousands of U.S. dollars. For the difference between estimated and realized compensation, see Section 2.2.

B.1.3 Multinational Firms

The evidence on multinational firms is based on the same dataset as our baseline estimation. The multinational firm indicators are constructed from the geographic segment data in Compustat. We classify a firm-year observation as multinational if a U.S. firm reports the existence of a non-domestic geographic segment, such as a foreign division. The multinational indicators from segment data are then linked with the ExecuCompustat-LBD. The resulting data set contains 12,943 firm-year observations and 1,606 unique firms. Out of these firm-year observations, 5,885 records are classified as non-MNE and the rest 7,058 as MNE. On average, the CEO-to-worker pay ratio is 87.4 among the non-MNE group and 100.0 among the MNE group.

B.2 Additional Tables and Figures

B.2.1 Additional Evidence on Export Status

To complement section 2.2.2, we present additional evidence of the robust correlation between participation in international markets and within-firm inequality. In panel A of Table B.5, we show that exporting status is positively correlated with within-firm inequality even when CEO compensation is measured by its subcomponents: CEO salary, bonus, and stock & options. Column (6) shows that the key correlation is also positive when we measure inequality relative to the compensation of the top 5 executives. In panel B we first modify estimating equation (2.1) to include different forms of fixed effects, firm-specific time trends, and we employ different forms of clustering; the core positive relationship between exporting status within-firm inequality persists. In column (5) we show that multinational firms have 23.6% higher within-firm inequality than non-multinationals. Panel C uses the data on compensation within privately-held firms to document a positive relationship between exporting status and different measures of executive compensation among private firms. Table B.6 replicates table 2.2 while replacing employment and payroll with sales and assets as measures of firm size.

Table B.5: Robustness: Within-Firm Inequality and Export Status

Panel A: Public Firms	Log CEO-to-Worker Pay Ratio					
	(1)	(2)	(3)	(4)	(5)	(6)
Exporter	0.507***	0.486***	0.213***	0.508***	0.855***	0.477***
	0.030	(0.029)	(0.023)	(0.035)	(0.071)	(0.027)
CEO Compensation	Baseline (Actual)	Estimated	Salary	Bonus	Stock & Options	Top 5 Executives
Sample	All	All	All	All	All	All
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Sector	Sector	Sector
Observations	17,000	17,000	17,000	13,000	17,000	16,000
R-squared	0.270	0.266	0.370	0.340	0.182	0.341

Panel B: Public Firms	Log CEO-to-Worker Pay Ratio				
	(1)	(2)	(3)	(4)	(5)
Exporter	0.507***	0.537***	0.507***	0.087***	
	0.030	(0.032)	(0.047)	(0.028)	
MNE					0.236***
					(0.028)
Robustness	Baseline	Year×Sector Fixed Effects	Cluster at Firm Level	Firm-Specific Time Trend	Multinational Enterprise
Sample	All	All	All	All	All
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Sector	Sector
Observations	17,000	17,000	17,000	17,000	13,000
R-squared	0.270	0.413	0.270	0.962	0.277

Panel C: Private Firms	Log Highest-Paid-Executive-to-Worker Pay Ratio					
	(1)	(2)	(3)	(4)	(5)	(6)
Exporter	0.411***	0.442***	0.168***	0.313***	0.544***	0.390***
	0.055	(0.052)	(0.036)	(0.086)	(0.103)	(0.054)
Executive Compensation	Actual	Estimated	Salary	Bonus	Stock & Options	Top 5 Executives
Sample	All	All	All	All	All	All
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Sector	Sector	Sector
Observations	6,000	6,000	5,000	4,000	5,000	5,000
R-squared	0.402	0.363	0.497	0.388	0.393	0.429

Table B.6: Robustness: Within-Firm Inequality, Export Status, and Firm Size

Panel A: Within-Firm Inequality and Firm Size in the United States

Dependent Variable	Log CEO-to-Worker Pay Ratio			
	(1)	(2)	(3)	(4)
Log Sales	0.423***	0.437***		
	(0.010)	(0.007)		
Log Assets			0.427***	0.425***
			(0.009)	(0.007)
Sample	Manufacturing	All	Manufacturing	All
Year Fixed Effects	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Sector
Observations	8,000	17,000	8,000	17,000
R-squared	0.417	0.439	0.407	0.428

Panel B: Within-Firm Inequality and Export Status, Controlling for Firm Size

Dependent Variable	Log CEO-to-Worker Pay Ratio			
	(1)	(2)	(3)	(4)
Exporter	0.047	0.024	0.295***	0.062**
	(0.091)	(0.025)	(0.079)	(0.025)
Log Sales	0.422***	0.436***		
	(0.010)	(0.007)		
Log Assets			0.424***	0.420***
			(0.011)	(0.007)
Sample	Manufacturing	All	Manufacturing	All
Year Fixed Effects	Yes	Yes	Yes	Yes
Group Fixed Effects	Sector	Sector	Sector	Sector
Observations	8,000	17000	8,000	17000
R-squared	0.417	0.439	0.407	0.428

Note: The left-hand side variable for each of the regressions is the (log of) CEO-to-worker pay ratio. "Exporter" is the exporter indicator computed from LFTTD. Exports are dollar values of shipments from LFTTD. Sales are the total annual sales reported in COMPUSTAT. Assets are the total assets reported in COMPUSTAT. The unit of observation is firm-year and the time period spans 1992 through 2007. See Table B.1 for sector distribution of the sample. Robust standard errors are clustered at the year-sector level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.2.2 Additional Evidence on the Trade Shock

To highlight the source and timing of our identification, we compare across different periods the within-firm inequality for firms with and without China-specific trade relationships. To that effect, we plot in Figure B.1 the coefficients γ_p measuring the period-p difference in within-firm inequality from a pooled regression of the following form:

$$\text{Outcome}_{\text{it}} = \delta_0 + \mathbf{d'}_1 \cdot \mathbf{f} + \mathbf{d'}_2 \cdot \mathbf{y} + \sum_{p=1}^{P} \gamma_p \text{Treatment}_i \times \text{Period}_p + \epsilon_{it}, \tag{B.1}$$

where Period_p partitions our sample into consecutive three-year blocks of time and Treatment_t is an indicator for firms that exported to China between 1999 and 2001.[3] Prior to China's WTO accession, the relative within-firm inequality γ_p is not statistically distinguishable from zero. Nonetheless, over the pre-WTO period, the point estimates of γ_p increase from 3% at the start to 13% just before the WTO accession. The impact of the WTO accession is driven by the 2005-2007 period when the treated firms have 25% higher inequality—a differential effect that can also be statistically distinguished from zero.

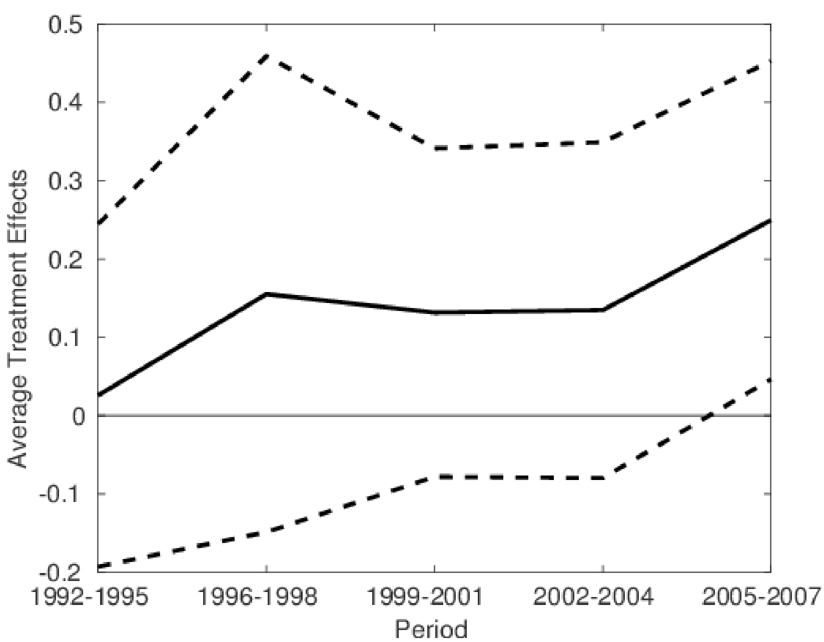


Figure B.1: Average Treatment Effect for CEO-to-Worker Pay Ratios across Time

Note: This figure plots estimated regression coefficients γ_p of the period-by-period difference in within-firm inequality for firms that exported to China between 1999 and 2001 relative to the control group. The dashed lines are 95 percent confidence intervals.

[3]More detailed definitions of time periods would not have permitted us to disclose all coefficients from regression (B.1).

To allay potential concerns about the timing of the trade shock, we repeat the earlier analysis using a window that includes 2001, the year of the WTO accession, into the definition of the treatment. The results are qualitatively and quantitatively similar to those from the baseline table 2.3. Specifically, in table B.7 we redefine the treated firms to be those that exported to China in the period 1999-2001 instead of the period 1998-2000. For manufacturing firms with this pre-existing China-specific relationship, China's accession to the WTO led to a 77.9% increase in exports and an 11.2% increase in the CEO-to-worker pay ratio. When we include firms in all sectors, we estimate a 43.9% increase in exports and a 10.3% increase in within-firm inequality. Taking the ratio of our estimates for trade-driven changes in inequality and firm size, we–as based on table 2.3–derive an elasticity implying that a 1% increase in firm size leads to a 0.2-0.3%.

Table B.7: China's Accession to the World Trade Organization and Within-Firm Inequality Robustness: Including Accession Year 2001 in Treatment Definition

Panel A: Exports and CEO-to-Worker Pay Ratio

Dependent Variable (log)	Exports	CEO-to-Worker Pay Ratio	Exports	CEO-to-Worker Pay Ratio
	(1)	(2)	(3)	(4)
Treatment × Post China WTO Accession	0.779***	0.112*	0.439***	0.103*
	(0.122)	(0.063)	(0.111)	(0.053)
Treatment	Exporter to China 1999-2001		Exporter to China 1999-2001	
Sample	Manufacturing		All	
Year Fixed Effects	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Observations	8,000	8,000	13,000	17,000
R-squared	0.909	0.714	0.930	0.755

Panel B: Employment and Payroll

Dependent Variable (log)	Employment	Payroll	Employment	Payroll
	(1)	(2)	(3)	(4)
Treatment × Post China WTO Accession	0.594***	0.601***	0.321***	0.326***
	(0.070)	(0.072)	(0.052)	(0.054)
Treatment	Exporter to China 1999-2001		Exporter to China 1999-2001	
Sample	Manufacturing		All	
Year Fixed Effects	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Observations	8,000	8,000	17,000	17,000
R-squared	0.927	0.919	0.936	0.921

Note: The left-hand side variable for each of the regressions is the (log of) CEO-to-worker pay ratio. "Exporter" is the exporter indicator computed from LFTTD. Exports are dollar values of shipments from LFTTD. Employment is the total annual employment reported in LBD. Payroll is the total annual payroll reported in LBD. The unit of observation is firm-year and the time period spans 1992 through 2007. See Table B.1 for sector distribution of the sample. Robust standard errors are clustered at the year-sector level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.2.3 Additional Information on the Calibration

Afghanistan	Cote d'Ivoire	Iraq	Nepal	Spain
Albania	Denmark	Ireland	Netherlands	Sri Lanka
Algeria	Dominican	Israel	New Zealand	Sudan
Argentina	Ecuador	Italy	Nicaragua	Sweden
Australia	Egypt	Jamaica	Niger	Switzerland
Austria	El Salvador	Japan	Norway	Syria
Bangladesh	Fiji	Jordan	Pakistan	Tanzania
Belgium	Finland	Kenya	Panama	Thailand
Benin	France	Korea	Papua New Guinea	Togo
Bolivia	Germany	Laos	Paraguay	Tonga
Botswana	Ghana	Lesotho	Peru	Trinidad &Tobago
Brazil	Greece	Malawi	Philippines	Tunisia
Bulgaria	Guatemala	Malaysia	Poland	Turkey
Burundi	Guyana	Maldives	Portugal	Uganda
Cameroon	Haiti	Mali	Romania	United Arab Emirates
Canada	Honduras	Mauritania	Rwanda	United Kingdom
Central African	Hong Kong	Mauritius	Saudi Arabia	United States
Chile	Hungary	Mexico	Senegal	Uruguay
China	Iceland	Mongolia	Sierra Leone	Venezuela
Colombia	India	Morocco	Singapore	Vietnam
Congo	Indonesia	Mozambique	Slovak	Zambia
Costa Rica	Iran	Namibia	South Africa	Zimbabwe

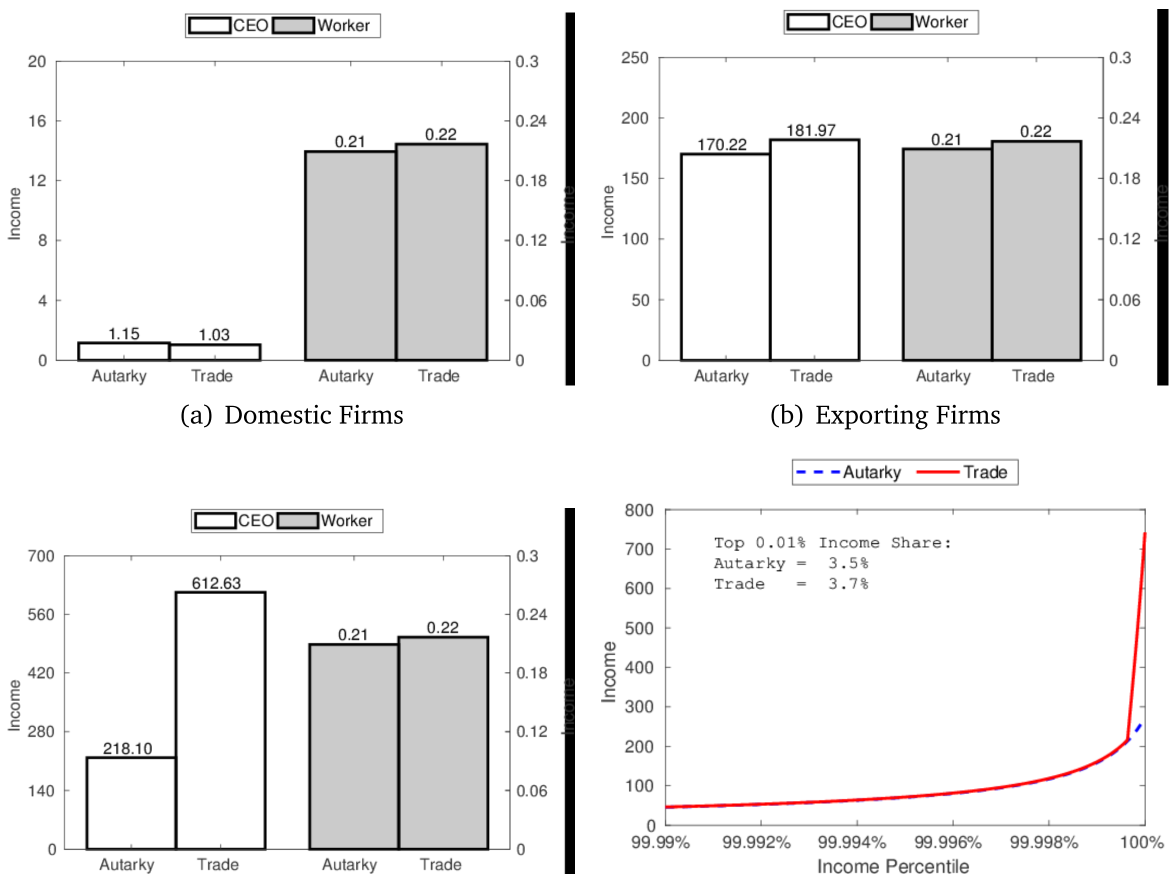
Table B.8: Countries Included in Calibration

Note: This table reports the list of countries (110 in total) included in the calibration. All the countries except the U.S. are included in ROW. The GDP and population data are based on Penn World Table 7.0 in the year 2008. GDP is in the unit of constant 2005 international dollar and calculated as the product of *RGDPL* and *POP*.

	τ	g	TFP, ROW
1988	1.930	6389.846	0.406
1989	1.906	6363.507	0.405
1990	1.911	4204.818	0.463
1991	1.896	3831.308	0.476
1992	1.891	4110.471	0.466
1993	1.894	4019.989	0.467
1994	1.873	4024.921	0.463
1995	1.834	3662.035	0.468
1996	1.826	3686.147	0.465
1997	1.808	3742.491	0.461
1998	1.830	4336.815	0.444
1999	1.832	4447.511	0.438
2000	1.812	4197.577	0.442
2001	1.854	3915.460	0.453
2002	1.884	3808.301	0.460
2003	1.878	3361.962	0.470
2004	1.843	2880.914	0.482
2005	1.821	2481.549	0.494
2006	1.797	2063.574	0.514
2007	1.760	1440.825	0.546
2008	1.720	1020.479	0.577

Table B.9: τ, g, and TFP

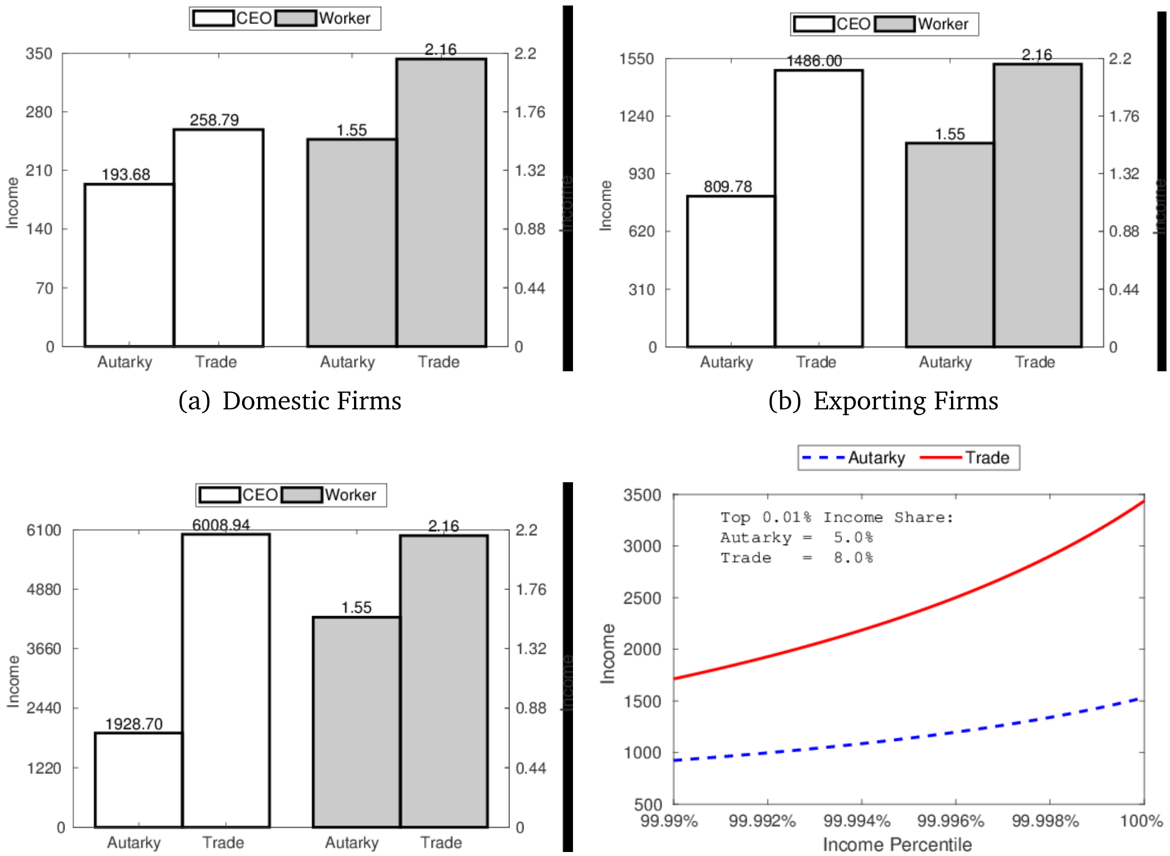
Note: This table reports the calibrated trade cost τ, g, and the estimated TFP. The τ and g matrices are assumed to be symmetric. The calibrated τ and g assume that the TFP for both countries is fixed at the 1988 level. The TFP reported is calculated to match the GDP ratio between the U.S. and ROW in each year. The TFP in the U.S. is always normalized to 1.



(a) Domestic Firms (b) Exporting Firms (c) MNE (d) Top Income Share

Figure B.2: Income Inequality Between Autarky and Trade, $\epsilon = 6$

Note: This figure plots how income inequality changes between autarky and trade for the case when $\epsilon = 6$. For more details, see the notes to Figure 2.4.



(a) Domestic Firms

(b) Exporting Firms

(c) MNE

(d) Top Income Share

Figure B.3: Income Inequality Between Autarky and Trade, $\epsilon = 2$

Note: This figure plots how income inequality changes between autarky and trade for the case when $\epsilon = 2$. For more details, see the notes to Figure 2.4.

B.3 Details of the Model

B.3.1 The Firm's Problem

Denote the total expenditure in country i as H_i, the ideal price level as P_i. If a firm in country j wants to sell to the market i, denote the price of the good as $p_{ij}(x)$ and the marginal cost (iceberg cost included) of selling to market i as $M_{ij}(x)$. The firm solves the following problem:

$$\max_{q_{ij}(x)} \quad p_{ij}(x)q_{ij}(x) - M_{ij}(x)q_{ij}(x),$$
$$\text{s.t.} \quad p_{ij}(x) = H_i^{\frac{1}{\epsilon}} P_i^{\frac{\epsilon-1}{\epsilon}} q_{ij}(x)^{-\frac{1}{\epsilon}},$$

where the constraint of the maximization problem is the inverse of the derived demand function from solving the consumer's problem in market i.

The solution of the above maximization problem is

$$q_{ij}(x) = H_i P_i^{\epsilon-1} \left(\frac{\epsilon}{\epsilon - 1} M_{ij}(x) \right)^{-\epsilon}, \tag{B.2}$$
$$p_{ij}(x) = \frac{\epsilon}{\epsilon - 1} M_{ij}(x). \tag{B.3}$$

Equation (B.3) is the result of plugging equation (B.2) into the inverse derived demand function.

The marginal cost of supplying to market i depends on the productivity of the firm, as well as the method through which the firm chooses to serve market i. If market i is served by a domestic firm or by an exporter in country j, then:

$$M_{ij}(x) = \frac{\tau_{ij} w_j}{A_j(x)}.$$

In the special case of $i = j$, market i is served by the domestic firm in country i:

$$M_{ii}(x) = \frac{w_i}{A_i(x)}.$$

If market i is served by an MNE founded in country j, then

$$M_{ij}(x) = \frac{w_i}{A_j(x)}.$$

The sales to market i, $\sigma_{ij}(x)$ is therefore

$$\sigma_{ij}(x) = p_{ij}(x)q_{ij}(x) = H_i P_i^{\epsilon-1}\left(\frac{\epsilon}{\epsilon-1}M_{ij}(x)\right)^{1-\epsilon}.$$

To supply $q_{ij}(x)$ to market i, the labor used in production is

$$L_{ij}(x) = H_i P_i^{\epsilon-1}\left(\frac{\epsilon}{\epsilon-1}M_{ij}(x)\right)^{\epsilon}\frac{\tau_{ij}}{A_j(x)},$$

with the understanding that when $i = j$, $\tau_{ij} = 1$.

The profit earned in market i before the fixed cost is

$$[p_{ij}(x) - M_{ij}(x)]q_{ij}(x) = \frac{H_i}{\epsilon}P_i^{\epsilon-1}\left(\frac{\epsilon}{\epsilon-1}M_{ij}(x)\right)^{1-\epsilon}$$

To ensure that firms sort into non-exporters, exporters, and multinational firms by productivity, we impose the following assumption similar to the one used in Helpman, Melitz and Yeaple (2004):

$$\frac{g_{ji}}{f_{ji}} \geq \left(\frac{\tau_{ji}w_i}{w_j}\right)^{\epsilon-1}$$

This equation implies that only the most productive firms will engage in FDI, while the other productive firms choose export over FDI.

A similar restriction needs to be imposed to ensure the separation of the domestic firms: we need to make sure that in equilibrium, not all the firms choose to sell to the foreign market. In a Melitz model, this condition can be written down explicitly. Unfortunately, it is not possible to do so for this paper. The reason is that x_i^* does not admit a closed-form solution. Nevertheless, characterization of the restriction is still possible. Generally, we need the market size of the home country to be above a certain level relative to the foreign country, or the variable trade cost to be above a certain level, so the firms in the home country will not find exporting to the foreign country too easy. In all the results presented in this paper, the separation of firms into domestic and exporting/multinational firms is checked and ensured.

B.3.2 The Equilibrium Conditions

The first three equilibrium conditions on cutoff human capital levels are self-evident. Here we explain the other two equilibrium conditions in detail. In this section, we derive the equilibrium conditions under truncation.

Income-Expenditure Identity The third equilibrium condition, equation (2.7), requires that the total expenditure and total income in country i must be the same:

$$H_i = n_i w_i \int_0^{x_i^*} x f_i(x) dx + n_i \int_{x_i^*}^{\infty} \pi_i(x) f_i(x) dx. \tag{B.4}$$

Total expenditure is denoted as H_i. Total income consists of two parts: the total labor income and the total profits. The CEO compensation function, $k(\pi)$, does not enter the accounting equation. The difference between the profit and the CEO compensation at each firm is distributed to all the individuals in the same country, and therefore $k(\pi)$ does not matter for total income.

The total labor income is easy to compute. It is the wage rate $w(i)$ times the total labor supply:

$$\begin{aligned}
w_i \cdot \left(n_i \int_0^{x_i^*} x f_i(x) dx \right) &= w_i n_i \frac{\lambda}{1 - e^{s_i \lambda}} \int_0^{x_i^*} x e^{-\lambda x} dx, && \text{(B.5)} \\
&= \frac{w_i n_i}{(1 - e^{-\lambda s_i} \lambda)} \left[e^{-\lambda x_i^*} (-\lambda x_i^* - 1) + 1 \right], && \text{(B.6)} \\
&= w_i n_i \frac{F(x_i^*)}{\lambda} - \frac{n_i x_i^* e^{-\lambda x_i^*}}{1 - e^{-\lambda s_i}}, && \text{(B.7)} \\
&= w_i \cdot \underbrace{\left\{ \frac{n_i}{\lambda} [F(x_i^*) - x_i^* f(x_i^*)] \right\}}_{\text{Labor Supply}}, && \text{(B.8)}
\end{aligned}$$

where $f(.)$ is the PDF of the truncated exponential distribution. The part in the curly brackets is the total labor supply in country i.

The total profit in country i is composed of three parts: the profit earned in the home country i, the profit earned in the other country j through export, and the profit earned in country j through FDI. This three-part separation is not the same as separating the profits into firms in the three corresponding groups. The difference is that, the profits earned in the home country i includes the profits from all the firms, as the exporters and MNEs also sell to the home market.

The total profit earned in the home market i is

$$n_i \int_{x_i^*}^{s} \frac{H_i}{\epsilon} P_i^{\epsilon - 1} \left(\frac{\epsilon}{\epsilon - 1} w_i \right)^{1-\epsilon} (b_i e^x)^{\epsilon - 1} f_i(x) dx - n_i f_{ii} w_i [1 - F(x_i^*)].$$

The total profit earned in the foreign market though exporting is

$$n_i \int_{x_{ji}^e}^{x_{ji}^f} \frac{H_j}{\epsilon} P_j^{\epsilon-1} \left(\frac{\epsilon}{\epsilon-1} \tau_{ji} w_i \right)^{1-\epsilon} (b_i e^x)^{\epsilon-1} f_i(x) dx - n_i f_{ji} w_i [F(x_{ji}^f) - F(x_{ji}^e)],$$

and the total profit earned in the foreign market through FDI is

$$n_i \int_{x_{ji}^f}^{s} \frac{H_j}{\epsilon} P_j^{\epsilon-1} \left(\frac{\epsilon}{\epsilon-1} w_j \right)^{1-\epsilon} (b_i e^x)^{\epsilon-1} f_i(x) dx - n_i g_{ji} w_i [1 - F(x_{ji}^f)].$$

The total profit in country i is the summation over these three parts. The income-expenditure identity here does not imply trade balance, as it usually does in a Melitz model. What it does imply is trade and financial balance. Trade in equilibrium is almost surely unbalanced, and the gap will be offset by the differences in capital flow: the differences between the profits the domestic MNEs collected from abroad and the foreign MNEs collected from the home market.

Ideal Price Level Equation (2.8) is the definition of the ideal price level in country i:

$$P_i = \left(\int_{m \in \Theta_i} p(m)^{1-\epsilon} dm \right)^{\frac{1}{1-\epsilon}}. \tag{B.9}$$

What needs further explanation is the set of goods available in country i: Θ_i. This set is the union of three mutually exclusive subsets: (1) the goods provided by all the firms created in country i, (2) the goods provided by all the exporting firms in country j, and (3) the goods provided by all the MNEs in country j. The price for every single variety in each of the subsets is a constant mark-up over the marginal cost in that subset. The marginal cost for goods in different subsets can be found in Appendix B.3.1. The ideal price level is a CES integration of all the individual prices over the set Θ_i.

After decomposing the set Θ_i into the three subsets mentioned above, the ideal price level can be expressed based on the firm productivity distribution directly:

$$P_i^{1-\epsilon} = \left\{ \sum_{j=1}^{2} \left[n_j \left(\frac{\epsilon}{\epsilon-1} \tau_{ij} w_j \right)^{1-\epsilon} \int_{x_{ij}^e}^{x_{ij}^f} b_i e^x f(x) dx + n_j \left(\frac{\epsilon}{\epsilon-1} w_i \right)^{1-\epsilon} \int_{x_{ij}^f}^{s} b_i e^x f(x) dx \right] \right\}.$$

Note that when $i = j$, $x_{ij}^e = x_i^*$. The first part in the square bracket includes all the goods provided by domestic firms, domestic exporters, and foreign exporters. The second part in the square bracket includes all the goods provided by the domestic and foreign MNEs.

Labor Market Clearing Condition The labor market clearing condition in country i requires that total supply of efficiency labor equals to total demand. Total supply equals the integral of x from 0 to x_i^* over the density function $f(x)$. Total labor demand is more complicated. It has four parts:

1. The labor used in the production of all the goods supplied to the home market i and exported to the foreign market j by the firms founded in country i:

$$L_i^{(1)} = n_i \sum_{j=1}^{2} \int_{x_{ji}}^{x_{ji}^f} \frac{H_j}{P_j^{1-\epsilon}} \left(\frac{\epsilon}{\epsilon - 1} \frac{\tau_{ji} w_i}{A_i(x)} \right)^{-\epsilon} \frac{\tau_{ji}}{A_i(x)} f(x) dx.$$

2. The labor used in fixed costs of operation and export incurred for the production in part 1:

$$L_i^{(2)} = n_i \sum_{j=1}^{n} f_{ji} \int_{x_{ji}}^{x_{ji}^f} f(x) dx.$$

3. The labor used in fixed costs for the goods supplied to country j through FDI by the firms created in country i:

$$L_i^{(3)} = n_i \sum_{j=1}^{2} g_{ji} \int_{x_{ji}^f}^{\infty} f(x) dx.$$

4. The labor used in the production of the goods supplied to country i by the foreign subsidiaries in country i from the firms founded in country j:

$$L_i^{(4)} = \sum_{j=1}^{2} n_j \int_{x_{ij}^f}^{\infty} \frac{H_i}{P_i^{1-\epsilon}} \left(\frac{\epsilon}{\epsilon - 1} \frac{w_i}{A_i(x)} \right)^{-\epsilon} \frac{1}{A_i(x)} f(x) dx.$$

B.3.3 Firm Size Distributions

In this appendix, we derive the CDF of firm productivity, sales, profit, and employment distributions for different groups of firms.

B.3.3.1 Productivity Distribution

The human capital, x, in country i is distributed exponentially with the following CDF:

$$F(x) = 1 - e^{-\lambda x},$$

and the firm founded by the individual with human capital x has the following productivity:

$$A_i(x) = b_i e^x.$$

Note that the marginal individual between entrepreneur and worker has human capital x^*, and thus will create a firm with productivity:

$$A_i(x^*) = b_i e^{x^*}.$$

For simplicity of notation we denote the lowest firm productivity as A_i^*. This implies that the human capital distribution underlying all the entrepreneurs is a shifted exponential distribution with the following CDF:

$$F^*(x) = 1 - e^{-\lambda(x - x^*)}.$$

The CDF of the firm productivity distribution conditional on the lower bound A_i^*, denoted as $F_A(y)$, can be derived as follows:

$$\begin{aligned}
F_A(y) &= \Pr(A_i(x) \le y) = \Pr(b_i e^x \le y) = \Pr(e^x \le \frac{y}{b_i}), \\
&= \Pr(x \le \log(y/b_i)) = F^*(\log(y/b_i)), \\
&= 1 - e^{-\lambda[\log(y/b_i) - x^*]}, \\
&= 1 - b_i^{\lambda} e^{\lambda x^*} y^{-\lambda}, \\
&= 1 - (A^*)^{\lambda} y^{-\lambda},
\end{aligned}$$

which is the CDF of a Type-I Pareto distribution with location parameter $A^* = b_i x^*$ and shape parameter λ. This CDF is shared by all the firms in country i whether they are non-exporting firms, exporting firms, or multinational firms.

Truncation If the exponential distribution is truncated from above at s, then the CDF of the human capital distribution for all entrepreneurs will be:

$$F(x) = \frac{1 - e^{-\lambda(x - x^*)}}{1 - e^{-\lambda s}}, x \in [x^*, s].$$

Given the same functional form of firm productivity, the CDF of the productivity distribution can be derived using similar methods outlined above. The distribution can be verified to be a truncated Pareto distribution,

$$F_A(y) = \frac{1-(A^*)^\lambda y^{-\lambda}}{1-b_i^\lambda u_i^{-\lambda}}, y \in [b_i, u_i],$$

where u_i is the country-specific upper bound of firm productivity:

$$u_i = b_i e^s.$$

In the rest of the this appendix, we use the original distribution without truncation.

B.3.3.2 Sales Distribution

The sales from country j to country i is derived in Appendix B.3.1 and repeated here:

$$p_{ij}(x)q_{ij}(x) = H_i P_i^{\epsilon-1}\left(\frac{\epsilon}{\epsilon-1}M_{ij}(x)\right)^{1-\epsilon}, \tag{B.10}$$

where $M_{ij}(x)$ is the marginal cost of production conditional on the mode of access (export or multinational production). Based on the market-specific sales, we derive the firm sales. We denote sales for a firm with CEO human capital x in country i as $\sigma_i(x)$ and rewrite it as a linear function of $A_i(x)^{\epsilon-1}$:

$$\sigma_i(x) = \Sigma_i(x)A_i(x)^{\epsilon-1}.$$

$\Sigma_i(x)$ summarizes the market size accessible to the firm. It is a step function depending on x:

$$\Sigma_i(x) = \begin{cases} H_i\left(\frac{P_i}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} & , x \in [x_i^*, x_{ji}^e), \\ H_i\left(\frac{P_i}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} + H_j\left(\frac{P_j}{\tau_{ji}w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} & , x \in [x_{ji}^e, x_{ji}^f), \\ H_i\left(\frac{P_i}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} + H_j\left(\frac{P_j}{w_j}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} & , x \in [x_{ji}^f, \infty). \end{cases}$$

The first line is the market accessible to the non-exporters, the second line the exporters, and the last line the multinational producers. The general formula for the CDF of the sales distribution is

$$
\begin{aligned}
F_\sigma(y) &= \Pr(\sigma < y), \\
&= \Pr(\Sigma_i(x) A_i(x)^{\epsilon-1} < y) = \Pr\left(A_i(x) < \left(\frac{y}{\Sigma_i(x)} \right)^{\frac{1}{\epsilon-1}} \right), \\
&= F_A\left(\left(\frac{y}{\Sigma_i(x)} \right)^{\frac{1}{\epsilon-1}} \right) = 1 - (A_i^*)^\lambda \left(\frac{y}{\Sigma_i(x)} \right)^{\frac{-\lambda}{\epsilon-1}}, \\
&= 1 - \left(\frac{\Sigma_i(x)}{(A_i^*)^{1-\epsilon}} \right)^\theta y^{-\theta},
\end{aligned}
$$

where

$$
\theta = \frac{\lambda}{\epsilon - 1}.
$$

The above equation defines Type-I Pareto distribution with shape parameter $\frac{\lambda}{\epsilon-1}$ and location parameter $\Sigma_i(x)(A_i^*)^{\epsilon-1}$. The location parameter differs by $\Sigma_i(x)$. The non-exporting firms have the smallest $\Sigma_i(x)$ and therefore the lowest location parameter. The exporting firms have higher $\Sigma_i(x)$ and the multinational firms have the highest $\Sigma_i(x)$. Note that within the same group (non-exporters, exporters, and multinationals), $\Sigma_i(x)$ is the same for all the firms.

B.3.3.3 Profit Distribution

The profit earned in each market is provided in Appendix B.3.1. Based on the market-specific profit, the firm profit can be written as an affine function of $A_i(x)^{\epsilon-1}$:

$$
\pi_i(x) = \Pi_i(x) A_i(x)^{\epsilon-1} - C_i(x).
$$

Similar to the sales distribution, $\Pi_i(x)$ takes three values depending on x:

$$
\Pi_i(x) = \begin{cases}
\frac{H_i}{\epsilon} \left(\frac{P_i}{w_i} \frac{\epsilon-1}{\epsilon} \right)^{\epsilon-1} & , x \in [x_i^*, x_{ji}^e), \\
\frac{H_i}{\epsilon} \left(\frac{P_i}{w_i} \frac{\epsilon-1}{\epsilon} \right)^{\epsilon-1} + \frac{H_j}{\epsilon} \left(\frac{P_j}{\tau_{ji} w_i} \frac{\epsilon-1}{\epsilon} \right)^{\epsilon-1} & , x \in [x_{ji}^e, x_{ji}^f), \\
\frac{H_i}{\epsilon} \left(\frac{P_i}{w_i} \frac{\epsilon-1}{\epsilon} \right)^{\epsilon-1} + \frac{H_j}{\epsilon} \left(\frac{P_j}{w_j} \frac{\epsilon-1}{\epsilon} \right)^{\epsilon-1} & , x \in [x_{ji}^f, \infty).
\end{cases}
$$

The first line is the market size accessible to a domestic firm. The second line is the market size for exporting firms, and the third line is the market size for multinational firms. Similarly, the fixed cost term $C_i(x)$ depends on the type of the firm

$$C_i(x) = \begin{cases} w_i f_{ii} & , x \in [x_i^*, x_{ji}^s), \\ w_i(f_{ii} + f_{ji}) & , x \in [x_{ji}^e, x_{ji}^f), \\ w_i(f_{ii} + g_{ji}) & , x \in [x_{ji}^f, \infty). \end{cases}$$

The distribution function of π takes the following general formula

$$\begin{aligned} F_\pi(y) = \Pr(\pi \leq y) &= \Pr(\Pi_i(x) \cdot A_i(x)^{\epsilon-1} - C_i(x) \leq y), \\ &= \Pr\left(A_i(x) \leq \left(\frac{y + C_i(x)}{\Pi_i(x)})\right)^{\frac{1}{\epsilon-1}}\right), \\ &= 1 - b_i^\lambda \left(\frac{y + C_i(x)}{\Pi_i(x)})\right)^{\frac{-\lambda}{\epsilon-1}} = 1 - \left(\frac{y + C_i(x)}{\Pi_i(x) b_i^{\epsilon-1}}\right)^{-\frac{\lambda}{\epsilon-1}}, \\ &= 1 - \left(1 + \frac{y + \mu_i(x)}{\chi_i(x)}\right)^{-\theta}, \end{aligned}$$

where

$$\begin{aligned} \mu_i(x) &= \chi_i(x) - C_i(x), \\ \chi_i(x) &= \Pi_i(x) \cdot (A_i^*)^{\epsilon-1}, \\ \theta &= \frac{\lambda}{\epsilon - 1}. \end{aligned}$$

This equation is the CDF of a Type-II Pareto distribution as defined in Arnold (1985). The shape index of the firm profit distribution is $\theta = \frac{\lambda}{\epsilon-1}$. The two location parameters $\mu_i(x)$ and $\chi_i(x)$ depend on the market that the firm can access to.

B.3.3.4 Employment Distribution

Employment distribution is similar to the profit distribution. Market-specific employment is provided in Appendix B.3.1 and here we aggregate it up to firm-level employment. For each firm the employment, $L_i(x)$, can be written as an affine function of $A_i(x)^{\epsilon-1}$:

$$L_i(x) = \Lambda_i(x) A_i(x)^{\epsilon-1} + T_i(x).$$

$\Lambda_i(x)$, again, summarizes the market size accessible to a firm x and is a step function that takes three values:

$$\Lambda_i(x) = \begin{cases} \frac{H_i}{P_i^{1-\epsilon}}\left(\frac{1}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon} & , x \in [x_i^*, x_{ji}^e), \\ \frac{H_i}{P_i^{1-\epsilon}}\left(\frac{1}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon} + \frac{H_j}{P_j^{1-\epsilon}}\left(\frac{1}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\tau_{ji}^{1-\epsilon} & , x \in [x_{ji}^e, x_{ji}^f), \\ \frac{H_i}{P_i^{1-\epsilon}}\left(\frac{1}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon} + \frac{H_j}{P_j^{1-\epsilon}}\left(\frac{1}{w_j}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon} & , x \in [x_{ji}^f, \infty). \end{cases}$$

$T_i(x)$ is the labor used as fixed cost of operation, export, and multinational production:

$$T_i(x) = \begin{cases} f_{ii} & , x \in [x_i^*, x_{ji}^s), \\ f_{ii} + f_{ji} & , x \in [x_{ji}^e, x_{ji}^f), \\ f_{ii} + g_{ji} & , x \in [x_{ji}^f, \infty). \end{cases}$$

Because both the employment and the profit are affine transformations of $A_i(x)^{\epsilon-1}$, the steps to derive the general formula of CDF are exactly the same. In the end, employment distributions are also Type-II Pareto distributions with shape parameter θ. The two location parameters depend on the market size accessible to the firm as well.

B.3.4 Income Distribution

The equilibrium income distribution in the model follows a two-class structure: the worker's income distribution follows an exponential distribution, and the CEO's income follows various Pareto-Type distributions. In this appendix, we present the details of the income distributions of the model.

Workers Workers in country i receive w_i for each unit of efficiency labor supplied to the market. The income for a worker with human capital x is $w_i x$, which follows an exponential distribution, same as x. The shape parameter of the income distribution is $\frac{\lambda}{w_i}$. The CDF of the distribution is

$$\begin{aligned} V(y) = \Pr(w_i x \leq y) &= \Pr(x \leq \frac{y}{w_i}), \\ &= 1 - e^{-\frac{\lambda}{w_i}y}. \end{aligned}$$

CEOs If $k(\pi)$ is monotonic and regularly varying with tail index β, then the CEO income follows a Pareto-Type distribution with shape parameter θ/β. Given a compensation function $k(\pi)$, the CDF of the CEO income is

$$U(y) = \Pr(k(\pi) \leq y) = \Pr(\pi \leq k^{-1}(y)) = F_\pi(k^{-1}(y)),$$

where $k^{-1}(y)$ is the inverse of $k(\pi)$ and $F_\pi(\cdot)$ is the CDF of firm profit distribution derived in Appendix B.3.3. The inverse function exists because $k(\pi)$ is monotonic. Because $k(\pi)$ is a regularly varying function with tail index β, the inverse function $k^{-1}(\cdot)$ is also a regularly varying function with tail index $1/\beta$ (Proposition 0.8.5, Resnick (1987)).

The survival function of π is a regularly varying function, with tail index $-\theta$ as well. To see this:

$$\lim_{\pi\to\infty} \frac{1 - F_\pi(\eta\pi)}{1 - F_\pi(\pi)} = \frac{\left(1 + \frac{\eta\pi+\mu}{\chi}\right)^{-\theta}}{\left(1 + \frac{\pi+\mu}{\chi}\right)^{-\theta}} = \eta^{-\theta}.$$

The composition of two regularly varying functions is a regularly varying function, and the tail index of the composition function is the product of the two indices (Proposition 0.8.4, Resnick (1987)). Therefore $1 - U(y)$, as the composition of $k^{-1}(y)$ and $1 - F_\pi(\pi)$, is a regularly varying function with tail index $-\frac{\theta}{\beta}$. This defines $y = k(\pi)$ as a Pareto-Type distribution with shape parameter $\frac{\theta}{\beta}$ (Definition 7.25, Gulisashvili (2012)). Moreover, the CDF of $k(\pi)$ can be re-written as:

$$U(y) = 1 - y^{-\theta/\beta} R(y),$$

where $R(y)$ is a slowly varying function:

$$\lim_{y\to\infty} \frac{R(\eta y)}{R(y)} = 1.$$

Example The CEO compensation function is

$$k(\pi) = \alpha^{1-\beta}\pi^\beta = \alpha^{1-\beta}\left(\Pi \cdot A^{\epsilon-1} - C\right)^\beta.$$

The CDF of $k(\pi)$ is

$$\begin{aligned}U(y) &= \Pr(k \leq y) = \Pr(\alpha^{1-\beta}\left(\Pi \cdot A^{\epsilon-1} - C\right)^{\beta} \leq y),\\ &= \Pr\left(A^{\epsilon-1} \leq \frac{y^{\frac{1}{\beta}}\alpha^{\frac{\beta-1}{\beta}} + C}{\Pi}\right),\\ &= 1 - b^{\lambda}\left(\frac{y^{\frac{1}{\beta}}\alpha^{\frac{\beta-1}{\beta}} + C}{\Pi}\right)^{-\frac{\lambda}{\epsilon-1}}.\end{aligned}$$

Using the general result proved above, it is trivial to show that $k(\pi)$ follows a Pareto-Type distribution. Here we follow a different route and prove directly that the survival function $1 - U(y)$ is a regularly varying function. To see this:

$$\begin{aligned}\lim_{y\to\infty}\frac{1-U(\eta y)}{1-U(y)} &= \lim_{y\to\infty}\left(\frac{\eta^{\frac{1}{\beta}}y^{\frac{1}{\beta}}\alpha^{\frac{\beta-1}{\beta}} + C}{y^{\frac{1}{\beta}}\alpha^{\frac{\beta-1}{\beta}} + C}\right)^{-\frac{\lambda}{\epsilon-1}},\\ &= \lim_{y\to\infty}\left(\frac{\eta^{\frac{1}{\beta}} + \frac{C}{y^{\frac{1}{\beta}}\alpha^{\frac{\beta-1}{\beta}}}}{1 + \frac{C}{y^{\frac{1}{\beta}}\alpha^{\frac{\beta-1}{\beta}}}}\right)^{-\frac{\lambda}{\epsilon-1}}.\end{aligned}$$

As $y \to \infty$, $y^{\frac{1}{\beta}} \to \infty$, therefore

$$\lim_{y\to\infty}\frac{1-U(\eta y)}{1-U(y)} = \eta^{-\frac{\lambda}{\beta(\epsilon-1)}},$$

which defines $1 - U(y)$ as a regularly varying function with index $-\frac{\lambda}{\beta(\epsilon-1)}$. This further implies that the income distribution function of CEOs in corporations can be expressed as

$$U(y) = 1 - y^{-\frac{\lambda}{\beta(\epsilon-1)}}R(y).$$

The income distribution of the CEOs at sole proprietorship firms is the same as the profit distribution and therefore is Type-II Pareto.

See Feller (1966), Resnick (1987), and Gulisashvili (2012) for more details on regularly varying functions and Pareto-Type distributions.

B.4 Proofs

For completeness, we provide another proposition to establish the ranking in the extended model with multinational firms. We then provide the proof for this proposition along side with proposition 1:

Proposition 5 *If the sets of exporting firms and multinational firms in country i are non-empty, then the average CEO-to-worker pay ratio among domestic firms is strictly smaller than the average CEO-to-worker pay ratio among exporting firms, which in turn is strictly smaller than the average CEO-to-worker pay ratio among multinational firms.*

B.4.1 Proof of Proposition 1 and 5

The least productive CEOs manage the domestic firms, which implies that, on average, they receive the lowest compensation among all the CEOs. The more productive CEOs manage the exporting firms, and the most productive CEOs manage the multinational firms. Since wage is equalized across the firms, the ranking of the CEO-to-worker pay ratio is the same as the ranking of the CEO income. ∎

B.4.2 Proof of Proposition 2

Profit-to-wage ratios in this model only depends on the cutoff human capitals in general equilibrium. This property can be exploited to gain some insight into the basic mechanism of the model without quantification.

Domestic Profit The profit-to-wage ratio in the domestic market is the profit earned from the domestic market divided by domestic wage. This part of profit is earned by the domestic firms, the exporters, and the MNEs created in the home country.

The profit-to-wage ratio is

$$\frac{\pi_{ii}(x)}{w_i} = \frac{H_i}{w_i\epsilon}\left(\frac{P_i}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} A_i(x)^{\epsilon-1} - f_{ii}.$$

From the cutoff condition of the marginal firm, we know:

$$\frac{H_i}{w_i\epsilon}\left(\frac{P_i}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} b_i^{\epsilon-1}e^{(\epsilon-1)x_i^*} - f_{ii} = x_i^*,$$

and therefore

$$\frac{H_i}{w_i\epsilon}\left(\frac{P_i}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} = \frac{x_i^* + f_{ii}}{b_i e^{(\epsilon-1)x_i^*}}. \tag{B.11}$$

Plug this into the first equation, we have

$$\frac{\pi_{ii}(x)}{w_i} = (x_i^* + f_{ii})e^{(\epsilon-1)(x-x_i^*)} - f_{ii}.$$

The partial derivative of this ratio with respect to x is positive, so in general, the profit-to-wage ratio is higher when the firm is more productive and larger. All the general equilibrium movements affect this ratio through the only endogenous variable in this equation: the cutoff value x_i^*. The cutoff human capital is a measure of the competitiveness of the home market in general equilibrium: it will be higher when the market is more competitive due to highly productive foreign firms entering. The partial derivative of this ratio with respect to x_i^* is

$$\frac{\partial}{\partial x_i^*}\left(\frac{\pi_{ii}(x)}{w_i}\right) = e^{(\epsilon-1)(x-x_i^*)}[1-(\epsilon-1)(x_i^*+f_{ii})]. \tag{B.12}$$

The sign of this derivative is the same as $[1-(\epsilon-1)(x_i^*+f_{ii})]$. We claim that this sign is always negative under the assumption that the least productive individual in country i must not find creating a new firm profitable. This restriction is imposed to guarantee the existence and uniqueness of the occupational choice cutoff in the paper. This assumption means:

$$\frac{H_i}{\epsilon}P_i^{\epsilon-1}w_i^{1-\epsilon}\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1}A_i(0)^{\epsilon-1} - f_{ii}w_i < 0,$$
$$f_{ii} > \frac{H_i}{\epsilon w_i}^{\epsilon-1}\left(\frac{\epsilon-1}{\epsilon}\frac{P_i}{w_i}\right)^{\epsilon-1}A_i(0)^{\epsilon-1}.$$

Plug equation (B.11) into the above inequality, we have

$$f_{ii} > \frac{x_i^*+f_{ii}}{A_i(x_i^*)^{\epsilon-1}}A_i(0)^{\epsilon-1}$$
$$f_{ii} > \frac{x_i^*}{e^{(\epsilon-1)x_i^*}-1}.$$

Now we need to prove

$$x_i^* + f_{ii} > \frac{1}{\epsilon-1}. \tag{B.13}$$

To do this, we define

$$m(x_i^*) = x_i^* + \frac{x_i^*}{e^{(\epsilon-1)x_i^*} - 1} - \frac{1}{\epsilon - 1}.$$

It is easy to show that $m(x_i^*)$ is monotonically increasing,

$$\frac{\partial m(x_i^*)}{\partial x_i^*} = 1 + \frac{e^{(\epsilon-1)x_i^*}(1 + (\epsilon-1)x_i^*) - 1}{(e^{(\epsilon-1)x_i^*} - 1)^2} > 0,$$

because

$$((\epsilon - 1)x_i^* > 0) \wedge (e^{(\epsilon-1)x_i^*} > 1).$$

Therefore, the minimum of $m(x^*)$ is obtained at $x_i^* = 0$, which is precisely 0. To see this, we need to apply L'Hôpital's rule to the second term at $x_i^* = 0$:

$$\begin{aligned}\lim_{x_i^* \to 0} m(x^*) &= x_i^* + \frac{1}{e^{(\epsilon-1)x_i^*}(\epsilon - 1)} - \frac{1}{\epsilon - 1}, \\ &= \frac{1}{\epsilon - 1} - \frac{1}{\epsilon - 1} = 0.\end{aligned}$$

This implies that for all possible values of $x_i^* \in [0, \infty)$, equation (B.13) is true and therefore the profit-to-wage ratio decreases with x_i^*.

Exporting Profits The profits earned from exporting to the foreign country, divided by local wage, is

$$\frac{\pi_{ji}^e(x)}{w_i} = \frac{H_j}{w_i \epsilon} \left(\frac{P_j}{\tau_{ji} w_i} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} A_i(x)^{\epsilon-1} - f_{ji}.$$

Similar to the domestic profit, the cutoff human capital of the marginal exporter is a sufficient statistics for the size of the foreign market and the marginal cost of accessing to that market. To see this, we start with the cutoff condition:

$$\begin{aligned}\frac{H_j}{\epsilon} \left(\frac{P_j}{\tau_{ji} w_i} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} A_i(x_{ji}^e)^{\epsilon-1} - f_{ji} w_i = 0, \\ \frac{H_j}{w_i \epsilon} \left(\frac{P_j}{\tau_{ji} w_i} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} = \frac{f_{ji}}{b_i e^{(\epsilon-1)(x - x_{ji}^e)}}.\end{aligned}$$

Plugging the above equation into the original profit-to-wage ratio, we have:

$$\frac{\pi^e_{ji}(x)}{w_i} = f_{ji}[e^{(\epsilon-1)(x-x^e_{ji})} - 1].$$

This ratio depends positively on x and negatively on x^e_{ji}. x^e_{ji} is a measure of the access to the foreign market: it will be lower (easier to access) when τ_{ji} is lower, or the foreign market is larger (H_j or P_j higher). When τ_{ji} is lower, the profit-to-wage ratio from the exporting market will be higher. ∎

B.4.3 Proof of Proposition 3

If $\tau_{ji} = \tau_{ij}$ drops, then x^e_{ji} will be lower and x^*_i will be higher. Proposition 2 implies that $\frac{\pi_{ii}(x)}{w_i}$ will be lower, which futher implies that the income ratio between CEOs at the domestic firms and the workers will be smaller. Proposition 2 also implies that $\frac{\pi^e_{ji}(x)}{w_i}$ will be higher as a result. Note that the ratio between the profit of the exporting firms and wage rate is the sum of the domestic and the export ratios:

$$\frac{\pi_{ii}(x) + \pi^e_{ji}(x)}{w_i} = \frac{H_i}{w_i \epsilon}\left(\frac{P_i}{w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} A_i(x)^{\epsilon-1} - f_{ii} + \frac{H_j}{w_i \epsilon}\left(\frac{P_j}{\tau_{ji} w_i}\frac{\epsilon-1}{\epsilon}\right)^{\epsilon-1} A_i(x)^{\epsilon-1} - f_{ji}.$$

The symmetry assumption implies that H, P and w will be equalized across countries, and so will their partial derivative with respect to τ_{ji} and τ_{ij}:

$$\frac{\partial\left(H_i P_i^{\epsilon-1} w_i^{-\epsilon}\right)}{\partial \tau_{ji}} = \frac{\partial\left(H_j P_j^{\epsilon-1} w_i^{-\epsilon}\right)}{\partial \tau_{ij}}$$

It is also straight forward to show that the partial derivative must be positive in general equilibrium, otherwise, there will be negative aggregate gains from trade.

The above observations imply that the sign of

$$\frac{\partial\left(\frac{\pi_{ii}(x)+\pi^e_{ji}(x)}{w_i}\right)}{\partial \tau_{ji}}$$

is always positive independent of $A_i(x)$, because $\tau_{ji}^{1-\epsilon}$ will be higher if τ_{ji} is lower. The intuition is simple: the income ratio between CEOs at the exporting firm and the workers will be higher following bilateral trade liberalizations.

The income ratio between the CEOs at the exporting firms and the domestic firms will also be higher because $\frac{\pi_{ii}(x)}{w_i}$ is lower and $\frac{\pi_{ii}(x)+\pi^e_{ji}(x)}{w_i}$ is higher as shown above.

Proposition 1 implies that the any CEO at the exporting firms earn higher income that the CEOs at the domestic firms and the workers. Now consider any individual with $x > x^e_{ji}$. After the changes in τ_{ji}, the income gap between him and all the individuals below x^e_{ji} will be wider. This directly implies that the income share for all the individuals with $x > x^e_{ji}$ will be higher if τ_{ji} is lower, and thus p^* can be computed as the survial function of the human capital distribution:

$$p^* = 100 \times (1 - F(x^e_{ji})) = 100 \times (e^{\lambda x^e_{ji}}) \quad \blacksquare$$

B.4.4 Proof of Proposition 4

FDI Profits The profits earned from FDI to the foreign country, divided by local wage, is:

$$\frac{\pi^f_{ji}(x)}{w_i} = \frac{H_j}{w_i \epsilon} \left(\frac{P_j}{w_j} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} A_i(x)^{\epsilon - 1} - g_{ji}.$$

From the FDI cutoff condition, we know

$$\begin{aligned} \frac{H_j}{w_i \epsilon} \left(\frac{P_j}{w_j} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} A_i(x^f_{ji})^{\epsilon - 1} &= \frac{H_j}{w_i \epsilon} \left(\frac{P_j}{\tau_{ji} w_i} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} A_i(x^f_{ji})^{\epsilon - 1} + (g_{ji} - f_{ji}), \\ \frac{H_j}{w_i \epsilon} \left(\frac{P_j}{w_j} \frac{\epsilon - 1}{\epsilon} \right)^{\epsilon - 1} &= \left[f_{ji} \frac{A_i(x^f_{ji})^{\epsilon - 1}}{A_i(x^e_{ji})^{\epsilon - 1}} + g_{ji} - f_{ji} \right] \frac{1}{A_i(x^f_{ji})^{\epsilon - 1}}. \end{aligned}$$

Therefore

$$\frac{\pi^f_{ji}(x)}{w_i} = f_{ji} e^{(\epsilon - 1)(x - x^e_{ji})} + (g_{ji} - f_{ji}) e^{(\epsilon - 1)(x - x^f_{ji})} - g_{ji}.$$

This profit-to-wage ratio decreases with x^f_{ji}:

$$\frac{\partial \frac{\pi^f_{ji}(x)}{w_i}}{\partial x^f_{ji}} = e^{(\epsilon - 1)(x - x^f_{ji})} (g_{ji} - f_{ji})(1 - \epsilon) < 0. \quad \blacksquare$$

B.5 A Model for the CEO Market

In this section we extend the benchmark model to allow for a labor market for CEOs, and an endogenously-determined CEO compensation function. The model here closely follows the work of Gabaix and Landier (2008). The key message of the extended model is, as long as the CEO contributes to the productivity of the firm, the equilibrium compensation function will satisfy the key assumptions that were used to exogenously define the compensation functions in the benchmark model.

Instead of allowing the individuals to create firms, we start by assuming that there exists a continuum of potential firms with different innate productivity, denoted and indexed by $\phi \in \Phi$, where Φ is a subset of real numbers. A firm needs to hire a CEO in order to operate. A potential CEO comes from the pool of candidates who are differentiated by their human capital x. The distribution of x follows the same exponential distribution as in the benchmark model. The final productivity of the firm depends on both the innate productivity of the firm, and the ability of the CEO. Following the notation of the benchmark model, the final productivity of the firm is:

$$A(\phi, x) = \phi \cdot b \cdot e^x,$$

where b denotes the TFP of the country. CEO receives compensation k from the firm. The compensation as a function of talent, $k(x)$, will be determined in equilibrium. Following the notation of the benchmark model, the profit of the firm in this extension can be written as:

$$\pi(\phi, x) = \tilde{H} A(\phi, x)^{\epsilon-1} - fw - k(x),$$

where $\tilde{H}$ describes the size of the markets to which the firm has access:

$$\tilde{H} = \frac{H}{\epsilon} P^{\epsilon-1} w^{-\epsilon} \left(\frac{\epsilon}{\epsilon - 1} \right)^{1-\epsilon}.$$

When the firm determines which CEO to hire, it takes the market price of talent, $k(x)$ as given. The first order condition of the firm is:

$$\tilde{H}(\epsilon - 1) A(\phi, x)^{\epsilon} A(\phi, x)' = k'(x)$$

which is essentially balancing the benefit of hiring a slightly better CEO with the extra cost of doing so. The solution to the differential equation of $k'(x)$ is:

$$k(x) = \tilde{H} b(\epsilon - 1) \int_{\underline{x}}^{\infty} \phi(x) e^{x(\epsilon-1)} dx + C, \tag{B.14}$$

where $\phi(x) : \mathcal{R} \rightarrow \Phi$ is the equilibrium mapping between CEO with talent x and the firm with productivity ϕ. C is the integration constant, which can be pinned down by the outside option of the least talented CEO, $\underline{x}$:

$$C = \underline{x}w.$$

It is impossible to exactly solve equation (B.14) without specifying the functional form of $\phi(x)$. However, without a closed-form solution we can still establish a couple of properties of $k(x)$. Gabaix and Landier (2008) characterized $k(x)$ by re-mapping x and ϕ into sequential indices, and utilizing an approximate spacing function of x. Specifically, they showed that equation (B.14) can approximately obtain a closed form solution if x follows an exponential distribution, up to a slowly varying function. Their key insights are two-folds. First, efficient market implies that in equilibrium there must be assortative matching between firms and CEOs, and thus $\phi(x)$ must be monotonically increasing in x. This implies that $k(x)$ must be monotonically increasing in x as well. Further more, when x follows an exponential distribution, the spacing function of x is regularly varying. This implies that in equilibrium, $k(x)$ must be regularly varying as well.

The arguments above establish that in equilibrium, the endogenously-determined $k(x)$ must be 1) monotonically increasing in x, and 2) regularly varying in x. These two results are precisely the assumptions that we made in the benchmark model, where $k(x)$ is exogenously imposed on the market. Moreover, it shows that even if we separate CEOs and founders, and model the market between CEO talents and firms, the end result in terms of the compensation scheme and matching pattern, will not change.

B.6 Calibration

The measure of population are computed following the method in Caselli (2005). The computation is based on Penn World Table 7.0, and all undefined variable names in italics are the standard variable names in PWT. We first compute real GDP in year t, Y_t, as

$$Y_t = pop_t \cdot rgdpl_t.$$

The number of workers, L_t, is backed out by

$$L_t = Y_t / rgdpwok_t.$$

This raw measure of the stock of work-force is first adjusted by human capital. Using years of school attainment for both males and females 25 years old and above from Barro and Lee (2010), we construct human capital h_t as

$$h_t = e^{\phi(c_t)},$$

where c_t is the years of schooling and $\phi(c_t)$ is piece-wise linear:

$$\phi(c_t) = \begin{cases} 0.134 * c & \text{if} \quad c_t \leq 4 \\ 0.134 * 4 + 0.101 * (c_t - 4) & \text{if} \quad 4 < c_t \leq 8 \\ 0.134 * 4 + 0.101 * 4 + 0.068 * (c_t - 8) & \text{if} \quad 8 < c_t \end{cases}.$$

Because the year of schooling data are only available at five-year intervals, linear interpolation is used to fill in the gap years. c_t is a slow-moving variable; therefore, linear interpolation can provide reasonably smooth estimations.

To construct the stock of physical capital in each year, we first compute investment in each year as

$$I_t = Y_t * \mathrm{ki_t}/100,$$

and then back out the initial capital stock using perpetual inventory method. We assume that capital and output grow at the same rate, and the depreciation rate is 6 percent per year. The initial capital stock when $t = 0$ is

$$K_0 = I_0/(g_k + 0.06),$$

where g_k is the average growth rate of GDP in the first 10 years of data. Given the initial capital stock, the sequence of capital stock in year t is computed as

$$K_t = (1 - 0.06)K_{t-1} + I_t.$$

With a computed sequence of physical capital, the final measure of population year t, n_t, is computed as

$$n_t = K_t^a (h_t L_t)^{1-a},$$

where $a = 1/3$. The number of n used in the benchmark calibration is the average between 1988 and 2008.

APPENDIX C

Chapter III Supporting Material

C.1 Additional Tables

Table C.1: Makes in Sales Volume Data

Acura	Honda	Nissan
Alfa Romeo	Hyundai	Porsche
Audi	Infiniti	Ram
BMW	Jaguar	Saab
Buick	Jeep	Scion
Cadillac	Kia	Smart
Chevrolet	Land Rover	Subaru
Chrysler	Lexus	Suzuki
Dodge	Lincoln	Tesla
Fiat	Mazda	Toyota
Ford	Mercedes-Benz	Volkswagen
GMC	Mini	Volvo
	Mitsubishi	

Note: Data come from Ward's Automotive. We exclude Mercury from the dataset because it was discontinued in January 2011, the first month of our analysis.

Table C.2: Automakers in Stock Price Data

Automaker (Holding Company)	Ticker
BMW	BAMXY:US
Daimler (Mercedes-Benz)	DDAIY:US
Ford	F:US
Fiat-Chrysler	FCAU:US
Fuji-Subaru	FUJHY:US
General Motors	GM:US
Honda	HMC:US
Mazda	MZDAY:US
Nissan	NSANY:US
Toyota	TM:US
Tesla	TSLA:US
Tata Motors (Jaguar and Land Rover)	TTM:US
Volkswagen Group	VLKAY:US

Note: Stock prices for automakers come from Bloomberg. Ford, General Motors, and Tesla are listed on U.S. stock exchanges; all remaining prices come from American Depository Receipts. Sample restricted to companies that sell light vehicles in the United States.

Table C.3: Makes in Twitter Data

Acura	Honda	Mini
Audi	Hyundai	Mitsubishi
BMW	Infiniti	Nissan
Buick	Jaguar	Porsche
Cadillac	Jeep	Ram
Chevrolet	Kia	Scion
Chrysler	Land Rover	Smart
Dodge	Lexus	Subaru
Fiat	Lincoln	Toyota
Ford	Mazda	Volkswagen
GMC	Mercedes-Benz	Volvo

Note: Data come from Networked Insights.

C.2 Robustness

In this appendix, we first show that our treatment of Mini as a non-German make in the baseline specification does not drive our results. Next, we document that the results are very similar to the baseline, when we use different control groups, and in a number of alternative econometric specifications.

C.2.1 Classification of Mini

Table C.4: U.S. Light Vehicle Sales Growth, Robustness – Mini
German vs. Non-German Manufacturers, Excl. VW Group

Dependent Variable	12-month Log Sales Growth		
Treatment of Mini	Baseline (non-German)	Exclude	Include as German
	(1)	(2)	(3)
German × Post-Scandal	-0.104	-0.105	-0.111
	(0.035)	(0.035)	(0.034)
Time Fixed Effects	Yes	Yes	Yes
Make Fixed Effects	Yes	Yes	Yes
R^2	0.292	0.294	0.293
N	2150	2082	2150

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Standard errors clustered at vehicle make level in parentheses. Volkswagen Group (Volkswagen, Audi, and Porsche) excluded from all regressions. Volkswagen emissions scandal dated September 18, 2015. Sales are measured in units sold. All regressions include a constant, make and time fixed effects, and are weighted by the square root of sales volumes. Data come from Ward's Automotive.

In our country classification of car makes, we include Mini—a company with historical roots in Britain that is now owned by BMW—as a non-German make. Our classification is supported by BMW board member Peter Schwarzenbauer, who told Reuters in a 2017 interview that the "brand being perceived as British, that's important... Most people don't know where the cars are produced" (Pitas (2017)). This focus on the country of brand association rather than the country of production or ownership drives our baseline classification choice. Nonetheless, we show here that this choice does not impact our results. Column (2) of table C.4 excludes Mini from the analysis altogether; the resulting estimate of a 10.5 percentage point decline in non-VW German car sales growth hardly changes from the baseline. Classifying Mini as a German make through its ownership by BMW in column (3) of table C.4 leads to an estimated decline of 11.1 percentage points, which is also similar to the baseline result in column (1).

C.2.2 Alternative Control Groups

Our baseline estimates of the scandal's spillovers use all non-German makes as the control group. Here, we vary the control group along two dimensions to investigate the stability of our baseline results. First, we partition the control group along one specific country-of-origin dimension, and then along a "more or less expensive"-dimension.

Column (2) of table C.5 shows that using only foreign automakers as the control group (i.e., excluding U.S. makes altogether) leads to a growth decline estimate of 12.0 percentage points for non-VW German car makers. Using only U.S. makes as the control group (i.e., excluding non-U.S. non-German makes altogether), we find a 8.8 percentage point decline in column (3). These estimated effects are both similar to the baseline result of a 10.4 percentage point decline in column (1).

Table C.5: U.S. Light Vehicle Sales Growth, Robustness – Control Group, Country German vs. Non-German Manufacturers, Excl. VW Group

Dependent Variable	12-month Log Sales Growth		
Control Group Makes	Baseline (non-German)	non-German non-U.S.	U.S.
	(1)	(2)	(3)
German Manuf. × Post-Scandal	-0.104	-0.120	-0.088
	(0.035)	(0.036)	(0.040)
Time Fixed Effects	Yes	Yes	Yes
Make Fixed Effects	Yes	Yes	Yes
R^2	0.292	0.313	0.360
N	2150	1402	952

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Standard errors clustered at vehicle make level in parentheses. Volkswagen Group (Volkswagen, Audi, and Porsche) excluded from all regressions. Volkswagen emissions scandal dated September 18, 2015. Sales are measured in units sold. All regressions include a constant, make and time fixed effects, and are weighted by the square root of sales volumes. Data come from Ward's Automotive.

We next split the control group into more and less expensive non-German makes in table C.6. To construct the two control groups, we first calculate a sales-weighted average of the MSRPs across models for each make in 2015. We then partition the makes by whether their average MSRP is above or below the median non-German average MSRP, approximately $33,000. The treatment group continues to be defined as the three non-VW German automakers. Smart's average MSRP is below $33,000, while BMW's and Mercedes-Benz's are above. For comparison, VW's average MSRP is also below the median. When the control group consists of the more expensive makes, as in column (2), we estimate an 8.2 percentage point decline in the sales growth of non-VW German makes. When the control group consists of the less expensive makes, as in column (3), we estimate an 11.3 percentage point decline. These estimated declines narrowly span our baseline estimate of a 10.4 percentage-point decline that we report in column (1).

Table C.6: U.S. Light Vehicle Sales Growth, Robustness – Control Group, Price
German vs. Non-German Manufacturers, Excl. VW Group

Dependent Variable	12-month Log Sales Growth		
Control Group Makes	Baseline (non-German)	More Expensive non-German	Less Expensive non-German
	(1)	(2)	(3)
German Manuf. × Post-Scandal	-0.104	-0.082	-0.113
	(0.035)	(0.037)	(0.036)
Time Fixed Effects	Yes	Yes	Yes
Make Fixed Effects	Yes	Yes	Yes
R^2	0.292	0.318	0.311
N	2150	1098	1210

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Standard errors clustered at vehicle make level in parentheses. Volkswagen Group (Volkswagen, Audi, and Porsche) excluded from all regressions. Volkswagen emissions scandal dated September 18, 2015. Sales are measured in units sold. All regressions include a constant, make and time fixed effects, and are weighted by the square root of sales volumes. Data come from Ward's Automotive.

C.2.3 Alternative Econometric Specifications

We show here that our difference-in-differences estimates are not sensitive to several alternative econometric specifications. Throughout the paper, we have weighted observations by the square root of the make's monthly sales volume. Column (2) of table C.7 shows that our choice to weight the observations leads to a conservative estimate of the spillovers: the unweighted estimate of the sales growth decline is 15 percentage points. Moreover, instead of natural log-differences, in column (3) we consider mid-point growth rates, where the change in sales volume between period t and period $t - 12$ is divided by the average level of sales of the two periods. The estimated effect of the scandal on the German auto manufacturers under this alternative measure is a 10.6 percentage point decline in the sales growth rate, which is similar to the baseline result in column (1). Finally, column (4) shows that including make-specific linear time trends in addition to the make and time fixed effects of the baseline specification again yields a similar result.

Table C.7: U.S. Light Vehicle Sales Growth, Robustness – Econometrics
German vs. Non-German Manufacturers, Excl. VW Group

Dependent Variable	12-month Log Sales Growth			
Specification	Baseline	Unweighted	Mid Point	Make-specific Trends
	(1)	(2)	(3)	(4)
German Manuf. × Post-Scandal	-0.104	-0.150	-0.106	-0.112
	(0.035)	(0.057)	(0.035)	(0.044)
Time Fixed Effects	Yes	Yes	Yes	Yes
Make Fixed Effects	Yes	Yes	Yes	Yes
Make-Specific Linear Time Trends	No	No	No	Yes
R^2	0.292	0.130	0.310	0.397
N	2150	2150	2150	2150

Note: Unit of observation is vehicle make-month. Time period covered is January 2011 to August 2016. Standard errors clustered at vehicle make level in parentheses. Volkswagen Group (Volkswagen, Audi, and Porsche) excluded from all regressions. Volkswagen emissions scandal dated September 18, 2015. Sales are measured in units sold. All regressions include a constant and are weighted by the square root of sales volumes. Regressions in columns (1) through (3) include make and time fixed effects. The regression in column (4) includes, in addition, make-specific linear time trends. Data come from Ward's Automotive.

BIBLIOGRAPHY

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451.

Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature*, 42(3): 691–751.

Angrist, Joshua, and Alan Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economcis*. Vol. 3, , ed. Orley Ashenfelter and David Card, Chapter 23, 1277–1366. Elsevier.

Angrist, Joshua D., and Jrn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Arnold, Barry C. 1985. *Pareto Distributions.* International Co-operative Publishing House.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis)Allocation." *Journal of Political Economy*, 122(5): 1013–1063.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review*, 98(5): 1998–2031.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature*, 49(1): pp. 3–71.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. "The Fall of the Labor Share and the Rise of Superstar Firms." *National Bureau of Economic Research Working Paper Series*, No. 23396.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science*, 293(5536): 1818–1820.

Baker, George P., and Brian J. Hall. 2004. "CEO Incentives and Firm Size." *Journal of Labor Economics*, 22(4): 767–798.

Bakija, Jon, Adam Cole, Bradley T Heim, et al. 2012. "Jobs and income growth of top earners and the causes of changing income inequality: Evidence from US tax return data." *Unpublished manuscript, Williams College*.

Barkai, Simcha. 2016. "Declining labor and capital shares." *Working Paper, University of Chicago*.

Barrett, Steven, Raymond Speth, Sebastian Eastham, Irene Dedoussi, Akshay Ashok, Robert Malina, and David Keith. 2015. "Impact of the Volkswagen emissions control defeat device on US public health." *Environmental Research Letters*, 10(11): 114005.

Barro, Robert J., and Jong-Wha Lee. 2010. "A New Data Set of Educational Attainment in the World, 1950-2010." National Bureau of Economic Research, Inc NBER Working Papers 15902.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review*, 103(1): 305–34.

Basu, Susanto, and John G. Fernald. 1997. "Returns to Scale in U.S. Production: Estimates and Implications." *Journal of Political Economy*, 105(2): 249–283.

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. "Are Technology Improvements Contractionary?" *The American Economic Review*, 96(5): 1418–1448.

Belliveau, Maura A., Charles A. O'Reilly, and James B. Wade. 1996. "Social Capital at the Top: Effects of Social Similarity and Status on CEO Compensation." *The Academy of Management Journal*, 39(6): pp. 1568–1593.

Bernard, Andrew B., and J. Bradford Jensen. 1997. "Exporters, skill upgrading, and the wage gap." *Journal of International Economics*, 42(1-2): 3–31.

Bernard, Andrew B., and J. Bradford Jensen. 1999. "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*, 47(1): 1–25.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2009. "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods." In *Producer Dynamics: New Evidence from Micro Data*. *NBER Chapters*, 513–552. National Bureau of Economic Research, Inc.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy*, 112(1): 68–105.

Bils, Mark, Peter J. Klenow, and Cian Ruane. 2017. "Misallocation or Mismeasurement?" *Working Paper*.

Breitinger, Matthias. 2016. "Der Diesel Skandal." http://www.zeit.de/wirtschaft/diesel-skandal-volkswagen-abgase.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains From Variety." *The Quarterly Journal of Economics*, 121(2): 541–585.

Bruckner, Johanna. 2015. “Was die Skandale für das deutsche Image bedeuten.” *http://www.sueddeutsche.de/kultur/pegida-dfb-und-vw-was-die-skandale-fuer-das-deutsche-image-bedeuten-1.2732828*.

Burstein, Ariel, and Jonathan Vogel. 2012. “International Trade, Technology, and the Skill Premium.” Society for Economic Dynamics 2012 Meeting Papers 664.

Caselli, Francesco. 2005. “Accounting for Cross-Country Income Differences.” In *Handbook of Economic Growth*. Vol. 1 of *Handbook of Economic Growth*, , ed. Philippe Aghion and Steven Durlauf, Chapter 9, 679–741. Elsevier.

Castriota, Stefano, and Marco Delmastra. 2014. “The Economics of Collective Reputation: Evidence from the Wine Industry.” *American Journal of Agricultural Economics*, 97(2): 469–489.

Chambers, Madeline. 2015. “VW scandal threatens ‘Made in Germany’ image.” *http://www.reuters.com/article/usa-volkswagen-germany-image-idUSL5N11S38X20150922*.

Clementi, Fabio, and Mauro Gallegati. 2005. “Pareto’s Law of Income Distribution: Evidence for Grermany, the United Kingdom, and the United States.” EconWPA Microeconomics 0505006.

Cuat, Vicente, and Maria Guadalupe. 2009. “Globalization and the Provision of Incentives inside the Firm: The Effect of Foreign Competition.” *Journal of Labor Economics*, 27(2): 179–212.

De Loecker, Jan. 2011. “Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity.” *Econometrica*, 79(5): 1407–1451.

De Loecker, Jan, and Jan Eeckhout. 2017. “The Rise of Market Power and the Macroeconomic Implications.” *National Bureau of Economic Research Working Paper Series*, No. 23687. http://www.nber.org/papers/w23687.pdf.

Destatis. 2015. “Automobilindustrie liefert wichtigstes Exportgut.” *https://www.destatis.de/DE/ZahlenFakten/ImFokus/IndustrieVerarbeitendesGewerbe/AutomobilindustrieWirtschaftDeutschland.html*.

di Giovanni, Julian, and Andrei A. Levchenko. 2012*a*. “Country Size, International Trade, and Aggregate Fluctuations in Granular Economies.” *Journal of Political Economy*, 120(6): 1083 – 1132.

di Giovanni, Julian, and Andrei A. Levchenko. 2013. “Firm entry, trade, and welfare in Zipf’s world.” *Journal of International Economics*, 89(2): 283–296.

di Giovanni, Julian, and Andrei Levchenko. 2012*b*. “Country Size, International Trade, and Aggregate Fluctuations in Granular Economies.” *Journal of Political Economy*, 120(6): 1083–1132.

di Giovanni, Julian, Andrei A. Levchenko, and Romain Rancire. 2011. "Power laws in firm size and openness to trade: Measurement and implications." *Journal of International Economics*, 85(1): 42–52.

di Giovanni, Julian, Andrei Levchenko, and Isabelle Mejean. 2014. "Firms, Destinations, and Aggregate Fluctuations." *Econometrica*, 84(2): 1303–1340.

di Giovanni, Julian, Andrei Levchenko, and Isabelle Mejean. 2015. "The Micro Origins of International Business Cycle Comovement." *NBER WP 21885*.

Domowitz, Ian, R Glenn Hubbard, and Bruce C Petersen. 1988. "Market Structure and Cyclical Fluctuations in U.S. Manufacturing." *The Review of Economics and Statistics*, 70(1): 55–66.

Drăgulescu, Adrian, and Victor M Yakovenko. 2001*a*. "Evidence for the Exponential Distribution of Income in the USA." *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(4): 585–589.

Drăgulescu, Adrian, and Victor M Yakovenko. 2001*b*. "Exponential and Power-Law Probability Distributions of Wealth and Income in the United Kingdom and the United States." *Physica A: Statistical Mechanics and its Applications*, 299(1): 213–221.

Drozd, Lukasz, and Jaromir Nosal. 2012. "Understanding International Prices: Customers as Capital." *American Economic Review*, 102(1): 364–395.

Egger, Hartmut, and Udo Kreickemeier. 2012. "Fairness, Trade, and Inequality." *Journal of International Economics*, 86(2): 184–196.

Elsby, Michael W. L., Bart Hobijn, and Aysegul Sahin. 2013. "The decline of the U.S. labor share." *Brookings Papers on Economic Activity*, 1–42.

Faleye, Olubunmi, Ebru Reis, and Anand Venkateswaran. 2013. "The Determinants and Effects of CEO-Employee Pay Ratios." *Journal of Banking & Finance*, 37(8): 3258–3272.

Fan, Ying, Jiandong Ju, and Mo Xiao. 2016. "Reputation premium and reputation management: Evidence from the largest e-commerce platform in China." *International Journal of Industrial Organization*, 46: 63–76.

Feenstra, Robert C., and David E. Weinstein. 2010. "Globalization, Markups, and the U.S. Price Level." National Bureau of Economic Research, Inc NBER Working Papers 15749.

Feenstra, Robert C, and Gordon H Hanson. 1996. "Globalization, Outsourcing, and Wage Inequality." *American Economic Review*, 86(2): 240–45.

Feller, William. 1966. *An Introduction to Probability Theory and Its Applications.* Vol. 2, John Wiley & Sons, Inc.

Fort, Teresa C., and Shawn D. Klimek. 2015. "The Effect of Industry Classification Changes on U.S. Employment Composition." *Working Paper*.

Foster, Lucia, Cheryl Grim, and John Haltiwanger. 2016. "Reallocation in the Great Recession: Cleansing or Not?" *Journal of Labor Economics*, 34(S1): S293–S331.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98(1): 394–425.

Frydman, Carola, and Raven E. Saks. 2010. "Executive Compensation: A New View from a Long-Term Perspective, 1936–2005." *Review of Financial Studies*, 23(5): 2099–2138.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics*, 1(1): 255–294.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *The Quarterly Journal of Economics*, 123(1): 49–100.

Geletkanycz, Marta A, Brian K Boyd, and Sydney Finkelstein. 2001. "The Strategic Value of CEO External Directorate Networks: Implications for CEO Compensation." *Strategic Management Journal*, 22(9): 889–898.

Gerakos, Joseph J, Joseph D Piotroski, and Suraj Srinivasan. 2009. "Globalization and Executive Compensation: An Analysis of Pay Practices in UK Companies." Working paper. University of Chicago.

Goldberg, Pinelopi Koujianou, and Nina Pavcnik. 2007. "Distributional Effects of Globalization in Developing Countries." *Journal of Economic Literature*, 45(1): 39–82.

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. "Capital Allocation and Productivity in South Europe." *The Quarterly Journal of Economics*, 132(4): 1915–1967.

Gourio, Francois, and Leena Rudanko. 2014. "Customer Capital." *Review of Economic Studies*, 81(3): 1102–1136.

Greenberg, Scott. 2017. "Summary of the Latest Federal Income Tax Data, 2016 Update." Tax Foundation.

Gulisashvili, Archil. 2012. *Analytically Tractable Stochastic Stock Price Models.* Springer.

Hall, Robert E. 1990. "Invariance Properties of Solow's Productivity Residual." *Growth/ Productivity/ Unemployment: Essays to Celebrate Bob Solow's Birthday*, , ed. Peter Diamond. Cambridge, Mass.:MIT Press.

Haltiwanger, John, Robert Kulick, and Chad Syverson. 2018. "Misallocation Measures: The Distortion That Ate the Residual." *National Bureau of Economic Research Working Paper Series*, No. 24199.

Handley, Kyle, and Nuno Limão. 2017. "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States." *American Economic Review*, 107(9): 2731–83.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export Versus FDI with Heterogeneous Firms." *American Economic Review*, 94(1): 300–316.

Helpman, E., O. Itskhoki, and S. Redding. 2010. "Inequality and Unemployment in a Global Economy." *Econometrica*, 78(4): 1239–1283.

Heston, Alan, Robert Summers, and Bettina Aten. 2002. "Penn World Table." Center for International Comparisons at the University of Pennsylvania.

Hipple, Steven F. 2010. "Self-Employment in the United States." *Monthly Labor Review*, 133(9): 17–32.

Holland, Stephen, Erin Mansur, Nicholas Muller, and Andrew Yates. 2016. "Damages and Expected Deaths Due to Excess NO_X Emissions from 2009 to 2015 Volkswagen Diesel Vehicles." *Environmental Science & Technology*, 50: 1111–1117.

Hopenhayn, Hugo A. 2014. "Firms, Misallocation, and Aggregate Productivity: A Review." *Annual Review of Economics*, 6(1): 735–770.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124(4): 1403–1448.

Improbable Research. 2016. http://www.improbable.com/ig/winners/.

Internal Revenue Service. 2013*a*. "Publication 15-A, Employer's Supplemental Tax Guide." Circular E.

Internal Revenue Service. 2013*b*. "Publication 15-B, Employer's Tax Guide to Fringe Benefits." Circular E.

Internal Revenue Service. 2013*c*. "Publication 15, Employer's Tax Guide." Circular E.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *The Quarterly Journal of Economics*, 129(1): 61–103.

Kehrig, Matthias. 2011. "The Cyclicality of Productivity Dispersion." *US Census Bureau Center for Economic Studies Working Paper*, No. CES-WP-11-15: 67.

Klette, Tor Jakob, and Zvi Griliches. 1996. "The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and Endogenous." *Journal of Applied Econometrics*, 11(4): 343–361.

Landon, Stuart, and Constance Smith. 1998. "Quality Expectations, Reputation, and Price." *Southern Economic Journal*, 64(3): 628–647.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies*, 70(2): 317–341.

Li, Lingfan, Steven Tadelis, and Xiaolan Zhou. 2016. "Buying Reputation as a Signal of Quality: Evidence from an Online Marketplace." *Mimeo, University of California at Berkeley,*.

Lucas, Jr, Robert E. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics*, 9(2): 508–523.

Lu, Yi, and Linhui Yu. 2015. "Trade Liberalization and Markup Dispersion: Evidence from China's WTO Accession." *American Economic Journal: Applied Economics*, 7(4): 221–53.

MacKinlay, A. Craig. 1997. "Event Studies in Economics and Finance." *Journal of Economic Literature*, 35(1).

Manasse, Paolo, and Alessandro Turrini. 2001. "Trade, Wages, and 'Superstars'." *Journal of International Economics*, 54(1): 97–117.

Marschak, Jacob, and William H. Andrews. 1944. "Random Simultaneous Equations and the Theory of Production." *Econometrica*, 12(3/4): 143–205.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation." *American Economic Review*, 104(8): 2421–2455.

McCallum, Andrew H. 2013. "The Structure of Export Entry Costs." University of Michigan Econonomics Department mimeo.

Meckl, Jrgen, and Benjamin Weigert. 2011. "Occupational Choice, Aggregate Productivity, and Trade." Institute for the Study of Labor (IZA) IZA Discussion Papers 5497.

Melitz, M.J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Monarch, Ryan, and Tim Schmidt-Eisenlohr. 2016. "Learning and the Value of Relationships in International Trade." , (16-11).

Monte, Ferdinando. 2011. "Skill Bias, Trade, and Wage Dispersion." *Journal of International Economics*, 83(2): 202–218.

Newburry, William. 2012. "Waving the Flag: The Influence of Country of Origin on Corporate Reputation." In *The Oxford Handbook of Corporate Reputation*. , ed. Timothy Pollock and Michael Barnett. Oxford Handbooks Online, Oxford University Press.

Nienaber, Michael. 2015. "Volkswagen could pose bigger threat to German economy than Greek crisis." http://www.reuters.com/article/us-usa-volkswagen-germany-economy-idUSKCN0RN27S20150923.

Noskoy, Chris, and Steven Tadelis. 2015. "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment." *Mimeo, University of California at Berkeley,*.

Oldenkamp, Rik, Rosalie van Zelm, and Mark Huijbregts. 2016. "Valuing the human health damage caused by the fraud of Volkswagen." *Environmental Pollution*, 212: 121–127.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–1297.

Oxelheim, Lars, and Trond Randøy. 2005. "The Effects of Internationalization on CEO Compensation." *Journal of International Business Studies*, 36(4): 470–483.

Petroff, Alanna. 2015. "What Volkswagen means to the German economy." http://money.cnn.com/2015/09/22/news/economy/volkswagen-germany-cars-economy/.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998." *The Quarterly Journal of Economics*, 118(1): 1–39.

Pitas, Costas. 2017. "Mini boss says UK production not essential to brand." https://www.reuters.com/article/us-autoshow-geneva-mini-britain/mini-boss-says-uk-production-not-essential-to-brand-idUSKBN16E1F0.

Poterba, James M., and Daniel R. Feenberg. 2000. "The Income and Tax Share of Very High-Income Households, 1960-1995." *American Economic Review*, 90(2): 264–270.

Remsky, Sarah. 2017. "Eine Belastung fr das Label 'Made in Germany'." http://www.zeit.de/wirtschaft/unternehmen/2017-07/abgasskandal-deutsche-wirtschaft-krise-volkswagen-daimler-porsche-automobilindustrie

Resnick, Sidney I. 1987. *Extreme Values, Regular Variation, and Point Processes.* Springer.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy distortions and aggregate productivity with heterogeneous establishments." *Review of Economic Dynamics*, 11(4): 707–720.

Roberts, David R. 1956. "A General Theory of Executive Compensation Based on Statistically Tested Propositions." *The Quarterly Journal of Economics*, 70(2): 270–294.

Rodriguez, Santiago Budria, Javier Diaz-Gimenez, Vincenzo Quadrini, and Jose-Victor Rior-Rull. 2002. "Updated facts on the U.S. distributions of earnings, income, and wealth." *Quarterly Review*, 2002(Sum): 2–35.

Rotemberg, Julio J., and Michael Woodford. 1991. "Markups and the Business Cycle." In *NBER Macroeconomics Annual 1991, Volume 6. NBER Chapters*, 63–140. National Bureau of Economic Research, Inc.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]." Minneapolis, MN: Minnesota Population Center [producer and distributor].

Sanders, WM Gerard, and Mason A Carpenter. 1998. "Internationalization and Firm Governance: The Roles of CEO Compensation, Top Team Composition, and Board Structure." *Academy of Management Journal*, 41(2): 158–178.

Slemrod, Joel. 1996. "High-Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response." In *Empirical Foundations of Household Taxation*. *NBER Chapters*, 169–192. National Bureau of Economic Research, Inc.

Stockman, Alan C., and Linda L. Tesar. 1995. "Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements." *American Economic Review*, 85(1): 168–185.

Tadelis, Steven. 1999. "What's in a Name? Reputation as a Tradeable Asset." *American Economic Review*, 89(3): 548–563.

Tirole, Jean. 1996. "A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality)." *Review of Economic Studies*, 63(1): 1–22.

United Nations. 2017. "UN comtrade data base." *https://comtrade.un.org/db/*.

Werz, Michael. 2016. "Smarter Lobbyieren." *http://www.zeit.de/wirtschaft/unternehmen/2016-01/automobilindustrie-usa-volkswagen-audi-porsche-verfahren*.

White, T. Kirk, Jerome P. Reiter, and Amil Petrin. 2016. "Imputation in U.S. Manufacturing Data and Its Implications for Productivity Dispersion." *National Bureau of Economic Research Working Paper Series*, No. 22569. http://www.nber.org/papers/w22569.pdf.

Yakovenko, Victor M, and A Christian Silva. 2005. "Two-class Structure of Income Distribution in the USA: Exponential Bulk and Power-Law Tail." In *Econophysics of Wealth Distributions*. 15–23. Springer.

Yeaple, Stephen Ross. 2005. "A Simple Model of Firm Heterogeneity, International Trade, and Wages." *Journal of International Economics*, 65(1): 1–20.

Zeit Online. 2017. "Daimler soll mehr als eine Million Fahrzeuge manipuliert haben." *http://www.zeit.de/wirtschaft/unternehmen/2017-07/abgasskandal-daimler-ag-schadstoffe?*