

THE UNIVERSITY OF CHICAGO

FIRM-SPECIFIC SHOCKS AND AGGREGATE VOLATILITY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

BY
CHEN YEH

CHICAGO, ILLINOIS

JUNE 2017

Copyright © 2017 by Chen Yeh

All Rights Reserved

Voor mijn ouders

Do not weep; do not wax indignant. Understand. - Baruch Spinoza

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	viii
ACKNOWLEDGMENTS	ix
ABSTRACT	x
1 INTRODUCTION	1
1.1 Contribution to the literature	4
1.2 Overview of the dissertation	7
2 U.S. CENSUS BUREAU DATA	8
2.1 Firm LBD	8
2.2 Firm-destination LBD	9
2.3 Plant ASM/CM	10
2.4 Product CM	10
3 VARIANCE DECOMPOSITION OF AGGREGATE SALES	12
3.1 Empirical framework	12
3.1.1 Theoretical motivation	12
3.1.2 Variance decomposition	15
3.1.3 Estimation	17
3.2 Data preview	18
3.3 Results	19
3.3.1 Decomposition of aggregate volatility	22
3.3.2 Decomposition of firm-specific component for aggregate volatility	23
3.4 Robustness exercises	27
4 THE REDUCED ROLE OF THE GRANULAR CHANNEL	29
4.1 Estimation results	29
4.2 Robustness and underlying economic mechanisms	32
4.3 Granularity	35
5 A MICROFOUNDATION FOR THE SIZE-VARIANCE RELATIONSHIP	40
5.1 Size-varying markups	40
5.1.1 Parametric method	40
5.1.2 Semi-parametric method	41
5.1.3 Estimation results	44
5.2 Identification: alternative mechanisms	47
5.3 Framework	49
5.3.1 Consumers	49
5.3.2 Firms	51

5.3.3	Balanced growth path equilibrium	54
5.3.4	Deviations from Gibrat’s law	56
5.4	Direct evidence on size-varying price elasticities	60
6	CONCLUSION	64
	REFERENCES	67
A	DATA FROM U.S. CENSUS BUREAU	73
A.1	Longitudinal Business Database (LBD)	73
A.2	Standard Statistical Establishment List (SSEL)	74
A.3	Longitudinal Firm Trade Transactions Database (LFTTD)	75
A.4	Annual Survey of Manufactures/Census of Manufactures (ASM/CM)	76
A.5	Homogeneous product-level data in CM	83
A.6	LBD-SSEL-LFTTD merging process	86
B	ROBUSTNESS AND ESTIMATION STRATEGIES	89
B.1	Variance decomposition	89
B.2	Size-variance relationship	94
B.3	Size-varying price elasticities	96
B.4	Markup estimation: de Loecker and Warzynski (2012)	97
C	ACCOUNTING FRAMEWORK FOR GRANULAR ECONOMY	102
C.1	Calibration	104
C.2	Alternative calibration strategies	106
D	PROOFS OF THEORETICAL FRAMEWORK	107
D.1	Properties of the Lambert W -function	107
D.2	Proof of lemma 1	107
D.3	Proof of proposition 1	108
D.4	Proof of proposition 2	108
D.5	Proof of proposition 3	109
D.6	Class of preferences with size-decreasing price elasticities	110

LIST OF FIGURES

3.1	Decomposing aggregate volatility into firm-specific and macro-sectoral components.	22
3.2	Disentangling granular and network components in the firm-specific term $\sigma_{F\tau}^2$.	25
4.1	Impact of the size-variance relationship on the granular channel.	37
5.1	Size-variance relationship in a random growth model with Kimball demand.	58
B.1	Including the financial crisis years 2008 and 2009.	89

LIST OF TABLES

3.1	Summary statistics of estimation sample.	19
3.2	Summary statistics of firm-destination level growth rates and its components.	21
3.3	Summary statistics of firm-specific components.	21
3.4	Relative contribution of firm-specific and macro-sectoral components.	23
3.5	Relative contribution of granular and network components.	26
4.1	Estimation results on the size-variance relationship as a power law.	30
4.2	Potential explanations for the size-variance relationship.	33
5.1	Estimation results of markup distributions by manufacturing industry group.	45
5.2	Estimation results of relationship between firm-level markups and size.	47
5.3	Estimation results of size-varying price elasticities.	63
A.1	Definition of product categories.	85
B.1	Definitions of constructed industries	92
C.1	Internally calibrated parameters.	105

ACKNOWLEDGMENTS

The list of acknowledgments for this dissertation consists of many people, but I am obliged to start by being extremely grateful to the members of my committee Lars Peter Hansen, Chad Syverson and Ufuk Akcigit. Without their encouragement, patience and invaluable support, this dissertation would not have seen the light of day.

To Aaron, Bong Geun and David who did not only help me get through the past few years but are also wonderful friends.

To my beautiful fiancée Claudia who I met in Chicago and I hope to be with forever.

To my sister Fan who has supported me ever since I was little and helped me get all the way here.

To my parents who have sacrificed many aspects of their life to ensure that I reap all the benefits of my education. I hope my graduation will be just one of the fruits of your incredible efforts. This dissertation is for you.

ABSTRACT

This dissertation examines the quantitative importance of firm-specific shocks on aggregate volatility. Firm-specific shocks to the largest firms can directly contribute to aggregate fluctuations whenever the firm size distribution is fat-tailed giving rise to the granular hypothesis. The first section of this dissertation identifies the variation in U.S. aggregate sales that can be explained by firm-level idiosyncratic shocks on the one hand and macro-sectoral forces on the other. By using a novel, comprehensive data set compiled from administrative sources that contains the universe of firms and trade transactions, I find that the granular hypothesis accounts at most for 16 percent of the variation in aggregate sales growth. This is about half of that found by previous studies that imposed Gibrat’s law where all firms are equally volatile regardless of their size.

The second section focuses on the underlying reason for the apparent discrepancy between the results found in this dissertation and the previous literature. I argue that the relatively small role of granularity in the U.S. economy can be entirely rationalized by accounting for the negative power law relationship between firm-level volatility and size, i.e. the size-variance relationship. Using the full distribution of growth rates among U.S. firms, I find robust evidence of such a relationship between firm-level volatility and size. The largest firms, whose shocks are most prominent for the granular channel, are the least volatile under the size-variance relationship, thus their influence on aggregates is mitigated. I show that by taking this relationship into account the effect of firm-specific shocks on observed macroeconomic volatility is substantially reduced.

The last part of this dissertation investigates several plausible mechanisms that could explain the negative size-variance relationship. After empirically ruling out some of them, I suggest a “market power” channel in which large firms face smaller price elasticities and therefore respond less to a given-sized productivity shock than small firms do. I provide direct evidence for this mechanism by estimating demand elasticities among U.S. manufactures. Furthermore, this channel is consistent with the observation that markups are increasing in

size. The results of a structural estimation procedure indicate that this is indeed the case for U.S. manufactures. Lastly, these robust empirical findings can be rationalized in an analytically tractable framework featuring random growth and a Kimball demand aggregator.

The research in this dissertation was conducted while the author was a Special Sworn Status researcher of the U.S. Census Bureau at the Chicago Census Research Data Center. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

CHAPTER 1

INTRODUCTION

Can firm-specific shocks to large firms drive the business cycle? Modern economies feature large firms whose size can be so overwhelming that shocks to these firms can lead to non-trivial aggregate movements (Gabaix, 2011). Such a channel of “granularity” can thus provide the micro-level underpinnings of aggregate fluctuations. In particular for the U.S. economy however, the empirical support for this channel is scarce.¹

In this dissertation, I assess the quantitative role of firm-level idiosyncratic shocks on aggregate volatility in the U.S. economy. I provide extensive evidence that the role of granularity is fairly limited in the U.S. economy despite the presence of a fat-tailed firm size distribution. The granular channel can explain at most around 16 percent of the variation in aggregate sales growth. This is considerably less than previous studies that have argued that the granular channel accounts for approximately one-third of aggregate fluctuations.

I argue that the gap between previous findings and mine can be fully explained by deviations from Gibrat’s law. I show that the observed relationship between firm-level volatility and size is substantially more negative than documented before. In other words, large firms are considerably less volatile than small firms. As a result, the impact of shocks to the largest firms on macroeconomic volatility is reduced. To rationalize my empirical findings, I construct a tractable framework of firm dynamics that is consistent with several size-varying firm-level moments and the Pareto right tail of the firm size distribution. The key driving force behind this model is that demand elasticities are decreasing in firm size which generate heterogeneous responses to a given-sized productivity shock across firms of different sizes.

I start my analysis by decomposing firms’ annual sales growth into different components which include a firm-specific one. I do so by merging the Census Bureau’s Longitudinal

1. The exceptions are Gabaix (2011) and Stella (2015) who both use data from Compustat. Gabaix (2011) finds that firm-level idiosyncratic shocks can explain about one-third of the variation in output growth. Stella (2015) shows that firm-specific shocks can explain 20 to 30 percent of aggregate sales dynamics, but there is little to no role for idiosyncratic shocks in explaining GDP growth fluctuations.

Business Database (LBD), Standard Statistical Establishment List (SSEL) and Longitudinal Firm Trade Transactions Database (LFTTD) into one novel, comprehensive data set. This database contains the universe of U.S. firms and trade transactions, and allows me to observe revenue at the firm-*destination*-year level. Using this specific disaggregated level of firm-level revenues allows me to relax the assumption that firms' sales are perfectly correlated across destinations. Firms that sell to multiple destinations are subject to heterogeneous demand shocks. To properly identify these shocks, it is necessary to observe variation in a firm's sales across destinations. This is particularly important for the largest firms in the economy, which are central under the granular hypothesis, because large firms typically serve multiple destination markets. This approach is similar to the identification strategy of di Giovanni, Levchenko and Mejean (2014) for extracting the firm-specific component from firm's annual sales growth. Using this large scale data set, I find that the granular component explains about 16 percent of aggregate sales volatility.

In the second part of the analysis, I focus on the underlying reason for the apparent discrepancy between the results found in this dissertation and the previous literature. I argue that the relatively small role of granularity in the U.S. economy can be entirely rationalized by accounting for the negative power law relationship between firm-level volatility and size, i.e. the size-variance relationship. The intuition behind this is simple: in a granular economy, aggregate volatility is primarily affected by the largest firms. Because firm-level volatility falls sufficiently fast in size, the volatility of the largest firms becomes small. Hence, their prominence in terms of aggregate fluctuations declines substantially.

I formalize this intuition by estimating the size-variance relationship with data containing the universe of U.S. firms. I show that it is explained by a power law and implement its estimates in a stylized model with heterogeneous firms subject to firm-level idiosyncratic shocks alone. I find that the aggregate implications of the size-variance relationship are substantial. The explanatory power of granular origins declines to roughly 15% whenever the size-variance relationship, as observed in the data, is taken into account.

My estimates on the size-variance relationship not only provide additional evidence on the violation of Gibrat’s law, but also shed light on the underlying economic mechanisms that can explain it. The size-variance relationship stays remarkably robust when controlling for output and product diversification (e.g. Klette and Kortum, 2004) or age which rules out any form of learning (e.g. Jovanovic (1982) and Arkolakis, Papageorgiou and Timoshenko (2015b)). Therefore to provide a explanation for the size-variance relationship, I resort to demand-side fundamentals and suggest a mechanism of “market power” in which a firm’s demand elasticity varies in its size. To show this theoretically, I develop an analytical framework with random growth (Luttmer, 2007) and a Kimball demand aggregator (Kimball, 1995).

Under the Kimball demand structure of Klenow and Willis (2016), the framework becomes tractable. Most importantly, the assumption of the Kimball demand aggregator implies that the price elasticity of demand is decreasing in a firm’s size. The model predicts a size-variance relationship, because large firms respond less to a given percentage change in productivity vis-à-vis small firms under this assumption. To support the main assumption of the framework, I find direct evidence of size-decreasing price elasticities in the Census of Manufactures (CM). Following Foster, Haltiwanger and Syverson (2008), I use a unique subset of establishments producing physically homogeneous products. The availability of physical quantities allows me to construct a measure of *physical* productivity which serves as a powerful instrument in dealing with the simultaneity bias confronted in demand estimation. Moreover, by focusing on those products that are considered to be the most physically homogeneous in the manufacturing sector, I avoid any biases due to variation in unobserved product quality.

This framework has the additional feature of predicting a positive relationship between markups and size. I verify this prediction of the framework by structurally estimating plant-level markups from the Annual Survey of Manufactures (ASM) using the methodology of Hall (1986) and de Loecker and Warzynski (2012). This approach has the benefit of estimating markups without imposing parametric assumptions on demand specifications and the nature

of competition. I find a robust positive correlation between plant-level markups with size and interpret this stylized fact as additional evidence in favor of my framework with respect to previous models that deal with deviations from Gibrat's law.

1.1 Contribution to the literature

This dissertation contributes to a number of strands of literature broadly covering the research on granularity and firm dynamics. First, this dissertation quantifies the contribution of firm-level idiosyncratic shocks to aggregate volatility. The current literature is characterized by two explanations for why firm-specific shocks can matter in the aggregate. The standard diversification argument (Lucas, 1977) breaks down whenever the firm size distribution is fat-tailed. Then, firm-level idiosyncratic shocks can provide a microfoundation for aggregate shocks. Thus, the magnitude of the granular channel has important implications for understanding the nature of business cycles and the cost-benefit analysis of stabilization policies.² This granular hypothesis is formalized by Gabaix (2011), but most of the empirical evidence for the U.S. economy on this channel comes from Standard and Poor's Compustat database that can lead to biased conclusions on firm-level and aggregate volatility as emphasized by Davis et al. (2007).³ I circumvent these sample selection issues by using the most extensive database available for the U.S. economy based on resources from the U.S. Census Bureau. As a result, I identify the impact of granularity on aggregate volatility directly from the data (as in Carvalho and Gabaix (2013) and di Giovanni, Levchenko and Mejean (2014)).

2. For an extensive overview of the literature on the costs of business cycles and the benefits of stabilization, see Barlevy (2005).

3. This empirical evidence includes the contributions by Gabaix (2011) and Stella (2015). The studies by di Giovanni, Levchenko and Mejean (2014) and Magerman et al. (2015) use data sets that cover more firms than just those that are publicly listed, but focus on the importance of firm-specific shocks on aggregate volatility in France and Belgium respectively. Instead, I focus on the importance of the granular channel in a large diversified country such as the U.S. as opposed to these countries. In contrast to other studies, I find a fairly limited role for granularity and emphasize the underlying reason for why this is the case in the U.S. economy.

On the other hand, firm-level idiosyncratic shocks can lead to aggregate volatility through their propagation in the network of input-output linkages (Acemoglu et al., 2012). While my main focus is the granular channel, I provide suggestive evidence on the importance of idiosyncratic forces that induce positive comovements between firms' residual growth rates similar to di Giovanni, Levchenko and Mejean (2014).

Second, I contribute to the literature on the size-variance relationship by providing estimates using data underlying the universe of U.S. firms. I find substantial deviations from Gibrat's law in the sense that there is a strong negative relationship between firm-level volatility and size in the form of a power law. This has been suggested before by Stanley et al. (1996), Sutton (2002) and Koren and Tenreyro (2013). However, all of their empirical results are derived from Compustat which only contains a subsample of the distribution of firm-level growth rates. Axtell (2001) mentions that this database is not only far from representative for the U.S. economy, but also has a different qualitative character in terms of firm size. Next to the reasons noted by Davis et al. (2007), this can substantially bias any views on the relationship between firm-level volatility and size which I confirm by comparing my estimates with those of previous findings.

I not only provide unbiased estimates on the size-variance relationship, but also highlight its aggregate implications. Specifically, when taking the size-variance relationship into account, the explanatory power of granularity is cut by more than half compared to the previous literature which assumes constant firm-level volatilities (as in Gabaix (2011) and Carvalho and Grassi (2015)).⁴ I also provide suggestive evidence on the underlying economic mechanism that can drive the size-variance relationship. Understanding this mechanism is important because it is crucial to understand what causes firms of different sizes to respond

4. Moreover, my findings have implications for the link between macroeconomic volatility and international trade. di Giovanni and Levchenko (2012) propose a mechanism in which trade liberalizations can raise a country's volatility: new trade avenues make large firms more important because exporting activities are more concentrated at these firms (e.g. Melitz (2003) and Bernard et al. (2009b)). Whenever a country is granular, the increased importance of large firms can thus increase its volatility. However, my results imply that the quantitative significance of this channel is relatively small for the U.S. economy as large firms are at the same time considerably less volatile.

heterogeneously during economic downturns. For example, Fort et al. (2013) argue that young and small businesses were especially hit hard (in terms of net employment growth) during the Great Recession.

Third, this dissertation also speaks to the literature in firm dynamics that deals with deviations from Gibrat's law. My analytical framework is most closely related to the random growth models of Luttmer (2007), Luttmer (2012) and Arkolakis (2016). One of the first-order features of Luttmer's (2007) framework is that it generates a stationary size distribution approximating Zipf's law. However, the instantaneous standard deviation of firm growth is independent of size. Under my setup with Kimball demand, the instantaneous variance of firm growth falls in size and approximately follows a power law.

Lastly, I provide extensive evidence on additional firm-level outcomes that vary with size besides firm-level volatility and growth.⁵ My results indicate that firm size plays an important role for price elasticities and markups. In fact, I show that there is a robust, positive correlation between markups and size implying that pass-through is incomplete. Furthermore, there is substantial variation of markups across firms within industries (at the 3-digit NAICS level). As pointed out by Atkin et al. (2015), this has important implications on a broad range of questions such as the welfare gains from trade liberalizations (e.g. de Loecker et al. (2016) and Edmond et al. (2015)), the effects of industrial policy interventions, and the transmission of international price shocks to local markets (e.g. Atkeson and Burstein (2008) and Nakamura and Zerom (2010)).

5. Violations on Gibrat's law have been documented as early as Hymer and Pashigian (1962). Surveys on (deviations from) Gibrat's law can be found in Caves (1998) and Sutton (1997). The literature has then developed a rich set of explanations that can rationalize several observations in the micro-level data that include product innovation (e.g. Klette and Kortum (2004), Lentz and Mortensen (2008), Akcigit and Kerr (2015)), industry-specific human capital accumulation and mean reversion (Rossi-Hansberg and Wright, 2007), investment in organizational capital (Luttmer, 2011), input diversification (Koren and Tenreyro, 2013) and market penetration costs (Arkolakis, 2016).

1.2 Overview of the dissertation

Chapter 2 contains a short description of the U.S. Census Bureau's data sets used in this paper. In chapter 3, I set up the empirical framework and present the results on the decomposition of firms' sales growth. In chapter 4, I describe the main aggregate implication of the size-variance relationship on granularity. Additionally, I provide extensive empirical evidence on the size-variance relationship as a power law and rule out a subset of alternative mechanisms put forward by other studies that explain this relationship. Chapter 5 estimates markups at the plant-year level in the manufacturing sector and how these markups display a robust, positive correlation with size. In addition, it presents an analytical framework that can rationalize the previously found deviations from Gibrat's law and the fat-tailed firm size distribution as observed in the U.S. economy. Concluding remarks can be found in chapter 6. A detailed description of the used data sets can be found in Appendix A. Appendices B and C contain numerous robustness checks and additional derivations of the framework can be found in Appendix D.

CHAPTER 2

U.S. CENSUS BUREAU DATA

The results of this paper are based on a variety of U.S. Census Bureau data sets. I use the Longitudinal Business Database (LBD), Standard Statistical Establishment List (SSEL), Longitudinal Firm Trade Transactions Database (LFTTD), Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM). In the following subsections, I briefly describe how I combine these data sets and what observables are available and used for my analysis. A more detailed description of these data sets can be found in Appendix A.1 to A.6.

2.1 Firm LBD

The Longitudinal Business Database (LBD) is a data set covering employment statistics at the establishment-level that covers nearly all sectors of the U.S. economy and comprises all geographic areas for the period of 1976 to 2011. Its underlying source is the Business Register (BR) that contains administrative records on U.S. businesses. The main purpose of the LBD is to obtain yearly employment (i.e. employee count) and revenue to construct growth rates at the firm-level. The LBD is unique in the sense that it covers the universe of firms in the U.S. economy and contains firm identifiers at the establishment-level. Thus, each establishment can be connected to some parent firm. Hence, it is straightforward to aggregate statistics up to the firm-level. While the LBD covers nearly all the industries in the U.S. economy, there are some exceptions. Roughly speaking, the LBD covers the non-farm private economy with some coverage of government-owned or operated entities.¹

An establishment belongs to a parent firm based on operational control: an establishment's statistics are included in the parent firm's activity whenever this firm majority owns the establishment. Furthermore, the LBD covers only those firms with at least one employee

1. More details on the coverage of the LBD and how this database is constructed, can be found in Jarmin and Miranda (2002).

on the payroll over their life-cycle. I use the Standard Statistical Establishment List (SSEL) to obtain revenue at the firm-level. Revenues are defined as the “total value of shipments, sales, receipts or revenue (in dollars)”. Appendix A.6 describes in full detail how revenue information from the SSEL can be incorporated into the LBD. Finally, I obtain a firm-level data set containing the variables of employment, payroll, age, year of entry and exit, multi-unit status, 6-digit level NAICS/SIC code, state and county codes and identifiers at the establishment- and firm-level. Due to restrictions in the SSEL, revenues are only available from 1994 onwards. The results of chapter 4 are mainly based on this data set.

2.2 Firm-destination LBD

The analysis in chapter 3 decomposes a firm’s revenue growth rate into a firm-level idiosyncratic and aggregate-sectoral component. However, this decomposition is only valid at the firm-destination level. Thus, I need to determine how a firm’s total revenue is decomposed across its destinations. To do this, I use export data from the Longitudinal Firm Trade Transactions Database (LFTTD) and I define a destination at the country level. The LBD and LFTTD share a common firm-level identifier that allows me to link the two databases. This procedure is explained in more detail in Appendix A.6. Therefore, I do not only observe a firm’s total revenue but I can also decompose it into domestic and foreign sales at the country level. While the LFTTD is available from 1992 onwards, I only use data from 1994 to 2011 as total revenues are only available in the LBD from 1994 onwards. Thus, I have the FIRM LBD database described above that is extended with export status and, most importantly, revenues disaggregated at the country level over the period 1994 to 2011. Chapter 3 exploits this constructed data base to its full potential.

2.3 Plant ASM/CM

A large benefit of the LBD is that it consists of the universe of U.S. firms over a significant amount of time. However, the number of firm-level observables is rather limited. Therefore, I restrict my attention to manufacturing in chapter 5 by using the Annual Survey of Manufactures (ASM) and Census of Manufactures (CM). The rich variety of available observables allows me to construct measures of labor (in hours as opposed to employee count only), capital, intermediate and energy inputs. Furthermore, these panels contain information on revenues (total value of shipments), gross output and value added. Thus, I can construct productivity measures such as value added or revenue per worker and total factor productivity. Combined with the publicly available information from the NBER-CES Manufacturing Database, I can construct proxies for real measures of output. A complete description of all variables and how these are constructed can be found in Appendix A.4. Finally, I obtain a plant-level data set that covers manufacturing (NAICS 31 - 33) over the period from 1976 to 2009.

2.4 Product CM

The PLANT ASM/CM data set does not contain physical quantities and instead relies on proxies for real measures of output. However, the Census Bureau does collect physical quantities for a subset of industries in Census years (ending with either “2” or “7”). More importantly, it collects this information at an extremely disaggregated level (7-digit SIC for 1977, 1982, 1987, 1992 and 1997 and 10-digit NAICS for 1997, 2002 and 2007). In addition to the PLANT ASM/CM data base, this results in information on physical quantities and prices which will prove to be extremely useful in section 5.4. I can construct measures of physical total factor productivity (TFPQ) which serves as a strong instrument in dealing with the simultaneity biases present in price-quantity regressions. I restrict myself to only a handful of product categories to focus on those establishments that produce a subset of

physically homogeneous products. By focusing on these products that are considered to be the most physically homogeneous in the manufacturing sector, I avoid any biases due to variation in unobserved product quality. The procedure that I follow to select these products is similar to Foster, Haltiwanger and Syverson (2008). Its details can be found in Appendix A.5.

CHAPTER 3

VARIANCE DECOMPOSITION OF AGGREGATE SALES

3.1 Empirical framework

In this section, I set up a framework that is the theoretical foundation of the variance decomposition. This model allows for a decomposition of individual growth rates into a firm-level idiosyncratic component and a component that contains aggregate and sectoral forces. As a result, I am able to quantify what fraction of aggregate sales volatility is due to granular forces. The structure of the model allows me to interpret these shocks in a meaningful way. However, this decomposition is only valid at the firm-*destination* level. Therefore, observing total sales at the firm level is not sufficient. Instead, I require a firm's sales to be broken down across different destinations. This additional source of variation allows me to relax the assumption that firms' sales are perfectly correlated across its destinations. Under this setup, the impact of sector level shocks is then allowed to be heterogeneous across firms.¹

3.1.1 Theoretical motivation

Aggregate sales in the U.S. are defined as $X_t = \sum_{f \in \mathcal{F}_t} x_{ft} = \sum_{(f,n) \in \mathcal{I}_t} x_{fnt}$ where x_{fnt} denotes the sales of some U.S. firm f to destination n in year t . Furthermore, \mathcal{F}_t and \mathcal{I}_t denote the set of active firms and firm-destination pairs in year t respectively. As noted before, the unit of observation is a firm-destination *pair*. By definition, aggregate sales growth is $\gamma_{At} = X_t/X_{t-1} - 1$ where I impose that the set of firms are identical in years t and $t - 1$.²

1. To illustrate the benefit of observing a firm's sales by destination, consider the following example. Suppose there are two types of firms within an industry. The first set of firms only serves domestic customers whereas the second set of firms also sells its product to Chinese customers. In the event of a positive income shock in China, this will only affect the foreign sales of those firms in the industry that export to China. To capture this industry-level demand shock that only affects exporting firms, it is necessary to observe sales at the firm-*destination* level.

2. By construction, this definition of growth only focuses on the intensive margin of growth. In section 3.3, I argue that most of the variation of aggregate sales growth is due to the intensive margin. Furthermore,

A firm f in the U.S. that is active in sector $j \in \{1, 2, \dots, J\}$ can sell its product to some country $n \in \{1, 2, \dots, N\}$ where $J, N > 1$. In each country n , the representative consumer is characterized by within-period Cobb-Douglas preferences over a set of J sectoral goods with income Y_{nt} . Preference parameters at this aggregation level can vary across countries and over time. Thus, the representative consumer's utility in country n at time t equals:

$$U_{nt} = \prod_{j=1}^J C_{jnt}^{\varphi_{jnt}}$$

where φ_{jnt} is a demand shock at the sectoral level that varies over time. The Cobb-Douglas structure directly implies that total expenditure on goods in sector j is equal to $Y_{jnt} = \varphi_{jnt} Y_{nt}$. In turn, each sectoral good j is a constant elasticity of substitution (CES) composite over varieties in Ω_{jnt} . Each firm f is associated with one sector and produces a unique variety in a monopolistically competitive fashion, thus I get:

$$C_{jnt} = \left(\sum_{f' \in \Omega_{jnt}} \omega_{f'nt}^{1/\theta} c_{f'nt}^{(\theta-1)/\theta} \right)^{\theta/(\theta-1)}$$

Let the U.S. be denoted by h ("home") and define $|\Omega_{jht}| = I_{jht}$. Firms are heterogeneous in their productivity that is characterized by the time-varying unit input requirement a_{fht} . Therefore, a firm f 's production technology at home is $y_{ft}(\iota) = \frac{1}{a_{fht}} \iota$ where ι denotes some composite input bundle that is priced at c_{jht} . Implicitly, I assume that the prices of input factors do not vary across firms *within* a sector. As a result, this setup in which firms face constant marginal costs of production and a downward-sloping CES demand curve leads to the well-known constant CES markup over marginal costs. Conditional on selling to

appendix B.1 discusses the robustness of my results with respect to the growth rate measure as suggested by Davis, Haltiwanger and Schuh (1996) which treats entry and exit symmetrically.

destination n , a firm f 's revenue at the home destination is then equal to:

$$x_{fnt} = \omega_{fnt} \frac{\varphi_{jnt} Y_{nt}}{P_{jnt}^{1-\theta}} \left[\frac{\theta}{\theta-1} \kappa_{jnh} c_{jht} a_{fht} \right]^{1-\theta}$$

where $P_{jnt} = \left(\sum_{f' \in \Omega_{jnt}} \omega_{f'nt} p_{f'nt}^{1-\theta} \right)^{1/(1-\theta)}$ denotes the ideal price aggregator at the sectoral level in country n at time t . Furthermore, $\kappa_{jnh} \geq 1$ denotes the iceberg shipping cost from h to n . By construction, it is the case that $\kappa_{jnh} = 1$ for all $n = h$. Implicitly, I assume that all the variation in a firm f 's marginal cost across destinations is captured by trade costs alone. These iceberg costs κ_{jnh} can vary across sectors within a country but are constant over time. This does not come with much loss of generality as any time variation in these iceberg costs will be absorbed by sectoral shocks otherwise. Whenever growth rates at the firm-destination level γ_{fnt} are approximated by log differences, I obtain:

$$\begin{aligned} \gamma_{fnt} &\simeq \ln \left(\frac{x_{fnt}}{x_{fnt-1}} \right) \\ &= \tilde{\delta}_{nt} + \tilde{\delta}_{jnt} + \varepsilon_{fnt} \\ &= \delta_{jnt} + \varepsilon_{fnt} \end{aligned}$$

where $\tilde{\delta}_{nt} = \Delta \ln Y_{nt}$ denotes the aggregate (or “macroeconomic”) shock to destination n and $\tilde{\delta}_{jnt} = \Delta \ln \varphi_{jnt} + (1 - \theta) (\Delta \ln c_{jdt} - \Delta \ln P_{jnt})$ captures sectoral-level demand and supply shocks that are specific to country n . On the other hand, $\varepsilon_{fnt} = \Delta \ln \omega_{fnt} + (1 - \theta) \Delta \ln a_{fdt}$ captures firm-specific demand and supply shocks. My main point of interest lies in the firm-specific component; thus, I combine the aggregate and sectoral level components into one term which I will refer to as the “macro-sectoral” component.

This framework is flexible enough to allow for demand shocks that differ across countries at the sectoral and firm level. Thus, I allow for a scenario in which firms sell to multiple, imperfectly correlated destinations. Thus, demand shocks at the sectoral level can substantially differ across destinations, in particular between domestic and foreign sales. Therefore,

the proposed decomposition is only valid at the firm-destination level and requires data on export sales as well. For this reason, observing a firm’s total sales is not sufficient to identify the macro-sectoral and firm-specific shocks.³ Thus, in addition to combining the Census Bureau’s LBD and SSEL, I also use data on the universe of firm-level exports from the LFTTD.⁴

3.1.2 Variance decomposition

My aim is to quantify the contribution of firm-specific shocks to aggregate volatility. First, I relate the firm-specific shocks $\{\varepsilon_{fnt}\}_{f,n}$ to the growth rate of aggregate sales γ_{At} in year t :

$$\begin{aligned} \gamma_{At} &= \sum_{f,n} \frac{x_{nft-1}}{X_{t-1}} \gamma_{nft} \\ &= \sum_{f,n} \frac{x_{nft-1}}{X_{t-1}} \delta_{jnt} + \sum_{f,n} \frac{x_{nft-1}}{X_{t-1}} \varepsilon_{fnt} \\ &= \sum_{j,n} \left(\sum_{f \in \Omega_{jnt}} \frac{x_{nft-1}}{X_{t-1}} \right) \delta_{jnt} + \sum_{f,n} \frac{x_{nft-1}}{X_{t-1}} \varepsilon_{fnt} \\ &\equiv \sum_{j,n} w_{jnt-1} \delta_{jnt} + \sum_{f,n} w_{fnt-1} \varepsilon_{fnt} \end{aligned}$$

where w_{fnt} is the share of firm f ’s sales to destination n in total sales and w_{jnt} denotes the share of sales by those subset of firms active in sector j in total sales. Thus, we have $w_{jnt} = \sum_{f \in \Omega_{jnt}} w_{fnt}$. In the following, I view the set of macro-sectoral and firm-specific components as stochastic processes. These processes $\{\delta_{jnt}\}_{j,n,t}$ and $\{\varepsilon_{jnt}\}_{j,n,t}$ are allowed

3. To see this, note that a firm’s total revenue growth does not allow for a decomposition that is log-linear. Under an approximation with log differences, a firm’s *total* revenue growth would be equal to $\gamma_{ft} = \ln(\sum_n x_{fnt}) - \ln(\sum_n x_{fnt-1})$. This term cannot be further decomposed in a meaningful way unless sales are perfectly correlated across destinations for every firm f . This is equivalent to saying that x_{fnt} must be independent across each destination n .

4. As far as I am aware of, the only studies that enhance the LBD with measures of (real) revenue are Moreira (2016) and Haltiwanger et al. (2016).

to be cross-sectionally and serially correlated.⁵ By construction, γ_{At} is then stochastic as well. Performing a meaningful variance decomposition on γ_{At} is complicated by w_{fnt} and w_{jnt} because these weights are time-varying. Analyzing the time-series properties of γ_{At} becomes difficult because it is hard to disentangle movements over time by the shocks from its weights. To overcome this problem, I follow Carvalho and Gabaix (2013) and di Giovanni, Levchenko and Mejean (2014) by considering a “synthetic growth” rate:

$$\gamma_{At|\tau} = \sum_{j,n} w_{jn\tau-1} \delta_{jnt} + \sum_{f,n} w_{fn\tau-1} \varepsilon_{fnt}$$

The synthetic growth rate $\gamma_{At|\tau}$ is identical to the actual growth rate γ_{At} with the exception of the weights that are fixed at their $t = \tau$ value. Given some value for τ , the weights w_{jnt} and w_{fnt} become predetermined variables that simplify the analysis considerably. Note that $\gamma_{At|\tau}$ explicitly depends on calendar time t as the weights from period τ are combined with the realization of shocks in period t . By construction, I get $\gamma_{At|t} = \gamma_{At}$.

Unlike for the actual growth rate, constructing the variance of the synthetic growth rate $\sigma_{A\tau}^2$ is relatively straightforward and allows for the following decomposition:

$$\begin{aligned} \sigma_{A\tau}^2 &= \sigma_{F\tau}^2 + \sigma_{M\tau}^2 + COV_{\tau} \quad \text{with} \\ \sigma_{F\tau}^2 &= V \left(\sum_{f,n} w_{fn\tau-1} \varepsilon_{fnt} \right), \\ \sigma_{M\tau}^2 &= V \left(\sum_{j,n} w_{jn\tau-1} \delta_{jnt} \right), \\ COV_{\tau} &= 2cov \left(\sum_{j,n} w_{jn\tau-1} \delta_{jnt}, \sum_{f,n} w_{fn\tau-1} \varepsilon_{fnt} \right). \end{aligned}$$

In the following, I provide estimates for $\sigma_{A\tau}^2$ and its components for each $\tau \in \{1995, \dots, 2011\}$.

5. See Appendix C of di Giovanni, Levchenko and Mejean (2014) for a more technical discussion on the assumptions behind these stochastic processes.

Formally, $\sigma_{A\tau}^2$ captures the variance of aggregate sales growth whenever there are no composition effects across firms and/or sectors. Thus, I will loosely interpret $\sigma_{A\tau}^2$ as the variance of aggregates sales growth in year τ . Furthermore, I examine their averages across τ and in particular the average relative contribution of the firm-specific component, that is $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{F\tau}}{\sigma_{A\tau}}$. While COV_{τ} must be included in the decomposition for $\sigma_{A\tau}$, I do not report the results for this component because this term is quantitatively negligible.

3.1.3 Estimation

The theoretical framework decomposes firm-destination level growth rates into macro-sectoral and firm-specific components. To obtain these components, I adopt the method in Stockman (1988) in which firm-destination level growth rates are regressed on a constant and a set of sector-level dummies for each *destination-year* subsample.⁶ Its fitted values form the macro-sectoral component while its residual is interpreted as the firm-specific shock. This procedure is akin to the following regression specification:

$$\gamma_{fnt} = \boldsymbol{\chi}'\mathbf{X}_{fnt} + \varepsilon_{fnt}$$

where \mathbf{X}_{fnt} contains a set of sectoral-destination-year fixed effects. Thus, the shocks are constructed as:

$$\begin{aligned}\widehat{\delta}_{jnt} &= \widehat{\boldsymbol{\chi}}'\mathbf{X}_{fnt} \\ \widehat{\varepsilon}_{fnt} &= \gamma_{fnt} - \widehat{\delta}_{jnt}\end{aligned}$$

6. This methodology is adopted by the majority of the granularity literature including Gabaix (2011) and Magerman et al. (2015). However, I deviate slightly from this approach by estimating this specification separately for each destination-year combination. Implicitly, I assume that firms' sales destinations are imperfectly correlated as in di Giovanni, Levchenko and Mejean (2014). While the difference seems benign methodologically, the outcomes could turn out to be very different. Under this setup, sectoral shocks are allowed to impact firms within an industry heterogeneously. In particular, exporting firms are subject to a wider variety of sector level shocks. This can especially be important for the largest firms in the economy as there is a strong correlation between firm size and export status.

Recall that the proposed decomposition is only valid at the firm-destination level. Hence, I demean individual growth rates at the sectoral level for each destination-year subsample separately which follows Koren and Tenreyro (2007). Once the shocks are constructed, it is straightforward to estimate synthetic growth rates and their components because the weights w_{jnt} and w_{fnt} come straight from the data. Lastly, the estimate for $\sigma_{A\tau}^2$ consists of the sample variance of $\hat{\gamma}_{At|\tau}$:

$$\hat{\sigma}_{A\tau}^2 = \frac{1}{T-1} \sum_{t=1995}^{2011} \left(\hat{\gamma}_{At|\tau} - \frac{1}{T} \sum_{t=1995}^{2011} \hat{\gamma}_{At|\tau} \right)^2$$

The estimates for $\sigma_{M\tau}^2$ and $\sigma_{F\tau}^2$ are generated in a similar fashion. di Giovanni, Levchenko and Mejean (2014) show that this estimator is consistent under a mild set of assumptions.

3.2 Data preview

Table 3.1 displays some summary statistics for the firm-destination level growth rates in my sample. While it is only suggestive, the table already shows some signs of deviations from Gibrat's law which will be crucial in my analysis in chapter 4.

First, the average growth of aggregate sales is smaller than the unweighted mean of the individual firm-destination level growth rates. The reasoning is because smaller firms, conditional on survival, tend to grow at higher rates than larger firms. Second, I report the averages of the volatility in firm-destination sales, as measured by its standard deviation, for each size quintile. The results show that smaller firms are more volatile than their larger sized counterparts.⁷

Lastly, the table shows the time averages of the Herfindahl indices defined at the firm-destination and firm levels. These are in the same order of magnitude as reported in Gabaix

7. As mentioned by Bernard et al. (2009b), exporting activities in the U.S. are primarily concentrated in the largest firms who might expose themselves to more volatile markets. This concentration might explain the results for the 81 - 100 size percentile.

(2011) who finds a value of 0.053. Intuitively, aggregate volatility is mostly determined by the largest firms in a granular economy. Whenever large firms are not as dominant in terms of size, as indicated by low Herfindahl indices, their contribution to aggregate volatility declines as well. Thus, these summary statistics indicate that the U.S. economy is less granular than previously thought. The following section shows this result in a more rigorous quantitative fashion.

Table 3.1: Summary statistics of estimation sample.^a

	Value
AVERAGE OF GROWTH	
Aggregate sales	0.046
Firm-destination level sales	0.051
STANDARD DEVIATION OF SALES GROWTH RATE	
Whole sample (average)	0.208
0 - 20 size percentile	0.240
21 - 40 size percentile	0.209
41 - 60 size percentile	0.200
61 - 80 size percentile	0.197
81 - 100 size percentile	0.201
AVERAGE OF HERFINDAHL INDEX	
Firm-destination $\sqrt{\mathcal{H}(f, n)}$	0.034
Firm $\sqrt{\mathcal{H}(f)}$	0.036
Industry (SIC2) $\sqrt{\mathcal{H}(j)}$	0.208

^a “Standard deviation of sales growth rate” reports the average standard deviation in the firm-destination sales growth rates within a percentile category. “Average of Herfindahl index” summarizes the average of a Herfindahl index (square root) across time.

3.3 Results

Table 3.2 shows the summary statistics of each growth rate component as generated by the decomposition of subsection 3.1. The average standard deviation of the firm-specific component is nearly equal to the standard deviation of the actual growth rate. Thus, the

firm-specific component is dominating most of the variation in sales growth at the firm-destination level. However, in contrast to the findings of di Giovanni, Levchenko and Mejean (2014), the estimated macro-sectoral shocks are quite volatile and display, on average, the same order of magnitude as the firm-specific component. These characteristics show that the role of aggregate or sector-level shocks could be important. However, the table shows that these shocks are particularly volatile for export sales. This volatility highlights the importance of my identification strategy in which I allow demand shocks at the sectoral level to vary across destinations. The summary statistics in table 3.2 indicate that this variation does matter because the average standard deviation in the macro-sectoral component is substantially different between domestic and foreign sales.

The empirical framework shows that firm-destination level specific shocks ε_{fnt} consist of a demand and supply component. In particular, I construct these shocks as:

$$\varepsilon_{fnt} = \Delta \ln \omega_{fnt} + (1 - \theta) \Delta \ln(a_{fdt})$$

By construction, supply shocks do not vary across destinations for a U.S. based firm. As a result, I proxy $\varepsilon_{fnt}^2 = (1 - \theta) \Delta \ln(a_{fdt})$ by calculating it as the average value of ε_{fnt} across its destinations n for each firm f that serves at least two destinations in some year t . The firm-destination level specific demand shocks $\varepsilon_{fnt}^1 = \Delta \ln \omega_{fnt}$ are then represented by the difference between ε_{fnt} and the previously constructed destination-level average.

As mentioned by di Giovanni et al. (2014), these measures are proxies at best because the constructed supply shocks also capture variation in demand factors that are common across a firm's destinations. Table 3.3 reports the summary statistics of the proxies for these supply and demand shocks. Most of the variation in the firm-specific shocks come from shocks that are destination-specific: the average standard deviation in ε_{fnt}^1 is more than 50 percent larger than the average standard deviation in ε_{fnt}^2 . Thus, this is suggestive evidence of the importance of demand-side fundamentals which supports the view on the size-variance

Table 3.2: Summary statistics of firm-destination level growth rates and its components.^b

	Observations	Mean	SD
TOTAL SALES			
Actual γ_{fnt}	$\simeq 46.2 \cdot 10^6$	0.052	0.239
Firm-specific $\widehat{\varepsilon}_{fnt}$	$\simeq 46.2 \cdot 10^6$	0.000	0.229
Sector-destination $\widehat{\delta}_{jnt}$	$\simeq 95000$	0.184	0.262
DOMESTIC SALES			
Actual γ_{fnt}	$\simeq 44.2 \cdot 10^6$	0.045	0.228
Firm-specific $\widehat{\varepsilon}_{fnt}$	$\simeq 44.2 \cdot 10^6$	0.000	0.219
Sector-destination $\widehat{\delta}_{jnt}$	$\simeq 1300$	0.043	0.052
FOREIGN SALES			
Actual γ_{fnt}	$\simeq 2.0 \cdot 10^6$	0.187	0.393
Firm-specific $\widehat{\varepsilon}_{fnt}$	$\simeq 2.0 \cdot 10^6$	0.000	0.381
Sector-destination $\widehat{\delta}_{jnt}$	$\simeq 93000$	0.186	0.264

^bThe column “Mean” denotes the average value of the growth rate component in the sample of firm-destination pairs and years. Similarly, “SD” denotes the average standard deviation of the growth rate component in this sample.

Table 3.3: Summary statistics of firm-specific components.^c

	Observations	Mean	SD
TOTAL SALES			
Firm $\widehat{\varepsilon}_{fnt}$	$\simeq 2.1 \cdot 10^6$	0.001	0.360
Firm-destination $\widehat{\varepsilon}_{fnt}^1$	$\simeq 2.1 \cdot 10^6$	0.000	0.321
Firm-common $\widehat{\varepsilon}_{fnt}^2$	$\simeq 433000$	0.006	0.207

^cThe column “Mean” denotes the average value of the growth rate component in the sample of firm-destination pairs and years. Similarly, “SD” denotes the average standard deviation of the growth rate component in this sample. Firm-destination-specific and firm-common components are only estimated on the subsample of firms that ship to at least two distinct countries.

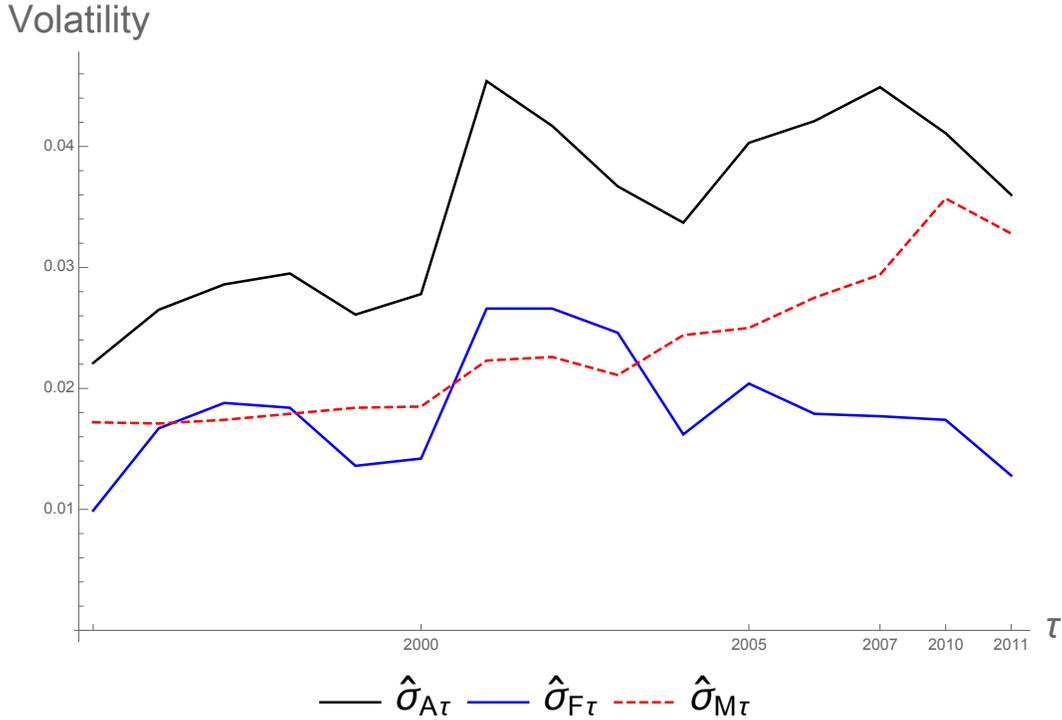
relationship laid down in section 5.3.⁸

8. Note that the estimation sample for ε_{fnt}^1 and ε_{fnt}^2 in table 3.3 is substantially smaller than in table 3.2. This is because I only focus on those firms that serve at least two different destinations when estimating ε_{fnt}^1 and ε_{fnt}^2 .

3.3.1 Decomposition of aggregate volatility

Figure 3.1 displays the estimates $\hat{\sigma}_{A\tau}$ (black), $\hat{\sigma}_{M\tau}$ (red) and $\hat{\sigma}_{F\tau}$ (blue) for each base year $\tau \in \{1995, \dots, 2011\}$. Table 3.4 summarizes one of the key results in this section. This table reports the average of the relative contribution of each component. Therefore, the contribution of the firm-specific and macro-sectoral component consists of $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{F\tau}}{\sigma_{A\tau}}$ and $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{M\tau}}{\sigma_{A\tau}}$ respectively. In contrast to di Giovanni, Levchenko and Mejean (2014), the macro-sectoral component is more important than the firm-specific component (as its relative standard deviation averages to 67 percent over 52.47 percent). Furthermore, the firm-specific component is much smaller than previously found in the literature. Those estimates average to approximately 80 percent whereas I only find a contribution of 52.47 percent which is a decrease of more than 30 percent. This is already suggestive of the fact that the granular hypothesis plays a much smaller role in the U.S. economy than previously conjectured.

Figure 3.1: Decomposing aggregate volatility into firm-specific and macro-sectoral components.



Even though the firm-specific component contributes relatively less than the macro-sectoral component, the standard deviation of the firm-specific component comoves over time with the standard deviation in aggregate sales. This is less clear for the standard deviation in the macro-sectoral component. However, its component does seem to increase slightly over time. By construction of these synthetic standard deviations, sales shares at the sector-destination level must then also increase over time. However, these do not change substantially over time because the increasing trend for $\widehat{\sigma}_{M\tau}$ over time is modest.

Table 3.4: Relative contribution of firm-specific and macro-sectoral components.^d

	SD	Relative SD
Aggregate $\widehat{\sigma}_{A\tau}$	0.0348	1.0000
Firm-specific $\widehat{\sigma}_{F\tau}$	0.0181	0.5247
Macro-sectoral $\widehat{\sigma}_{M\tau}$	0.0232	0.6700

^dThe column “SD” represents the average values of the aggregate, firm-specific and macro-sectoral components over the sample period, i.e. I report $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\sigma}_{A\tau}$, $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\sigma}_{F\tau}$ and $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\sigma}_{M\tau}$ respectively. The column “Relative SD” reports the average value of the relative standard deviations, i.e. $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\widehat{\sigma}_{F\tau}}{\widehat{\sigma}_{A\tau}}$ and $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\widehat{\sigma}_{M\tau}}{\widehat{\sigma}_{A\tau}}$.

3.3.2 *Decomposition of firm-specific component for aggregate volatility*

While the contribution of the firm-specific component is smaller relative to the literature, its magnitude is still non-negligible. Hence, it is worthwhile to pin down the economic forces behind this term. However, the firm-specific component does not necessarily only represent the granular mechanism as in Gabaix (2011). This is clarified by decomposing the

firm-specific term as:

$$\begin{aligned}
\sigma_{F\tau}^2 &= V \left(\sum_{f,n} w_{fn\tau-1} \varepsilon_{fnt} \right) \\
&= \sum_{g,m} \sum_{f,n} w_{gm\tau-1} w_{fn\tau-1} \text{cov}(\varepsilon_{gmt}, \varepsilon_{fnt}) \\
&= \sum_{f,n} w_{fn\tau-1}^2 V(\varepsilon_{fnt}) + \sum_{(g \neq f) \vee (m \neq n)} \sum_{f,n} w_{gm\tau-1} w_{fn\tau-1} \text{cov}(\varepsilon_{gmt}, \varepsilon_{fnt})
\end{aligned}$$

This decomposition follows Carvalho and Gabaix (2013) and di Giovanni, Levchenko and Mejean (2014) and it identifies two important components within the firm-specific term. Specifically, the contribution of the individual variances represents the direct effect of firm-level idiosyncratic shocks under an environment in which these are not propagated through any mechanism. This is the basic concept of granularity in Gabaix (2011). In the absence of, for example, firm-to-firm linkages, independently distributed firm-level idiosyncratic shocks can still have an impact on aggregate volatility whenever the firm size distribution is fat-tailed. The sheer magnitude of the largest firms in the economy can make their idiosyncratic shocks sufficiently important which contradicts the conventional logic behind the Law of Large Numbers.⁹ It is not surprising that this term captures granular forces because it is the variance of Gabaix’ (2011) granular residual whenever the firm-specific shocks ε_{fnt} are distributed independently from each other. Hence, I will denote its component by $\text{GRAN}_\tau = \sum_{f,n} w_{fn\tau-1}^2 V(\varepsilon_{fnt})$.¹⁰

9. Gabaix (2011) shows that aggregate volatility declines at the rate $1/N^{(\zeta-1)/\zeta}$ in the number of firms N where ζ denotes the Pareto right-tail of the firm-size distribution. The most extreme case involves Zipf’s law (i.e. $\zeta = 1$) under which aggregate volatility declines at the rate $1/\ln N$ instead. These results depend on a “fat-tailed” central limit theorem for infinite-variance random variables (see appendix A of Gabaix (2011)).

10. The impact of firm-level idiosyncratic shocks can be propagated through several amplification mechanisms, e.g. endogenous factor usage or input-output linkages. As a result, it could be argued that the firm-specific component is equivalent to the granular component. However, I define the granular component in a more narrow fashion that corresponds to a scenario without these amplification mechanisms. This is also done in di Giovanni, Levchenko and Mejean (2014), but more importantly my results can be compared with other granular studies that abstract from these propagation channels, e.g. the baseline economy in Gabaix (2011) and Carvalho and Grassi (2015).

Furthermore, there is the component that contains the comovements between firms' residual growth rates. Traditionally, comovements between firms were interpreted as the by-products of aggregate or industry-level shocks. However, there is a large and growing literature that shows that comovements between firms can be induced by independent firm-level idiosyncratic shocks through input-output linkages (e.g. Acemoglu et al. (2012) and Foerster, Sarte and Watson (2011)). Whenever there exists a sufficient degree of asymmetry in the network, in terms of how firms supply inputs to each other, then firm-level idiosyncratic shocks do not necessarily wash out in the aggregate.

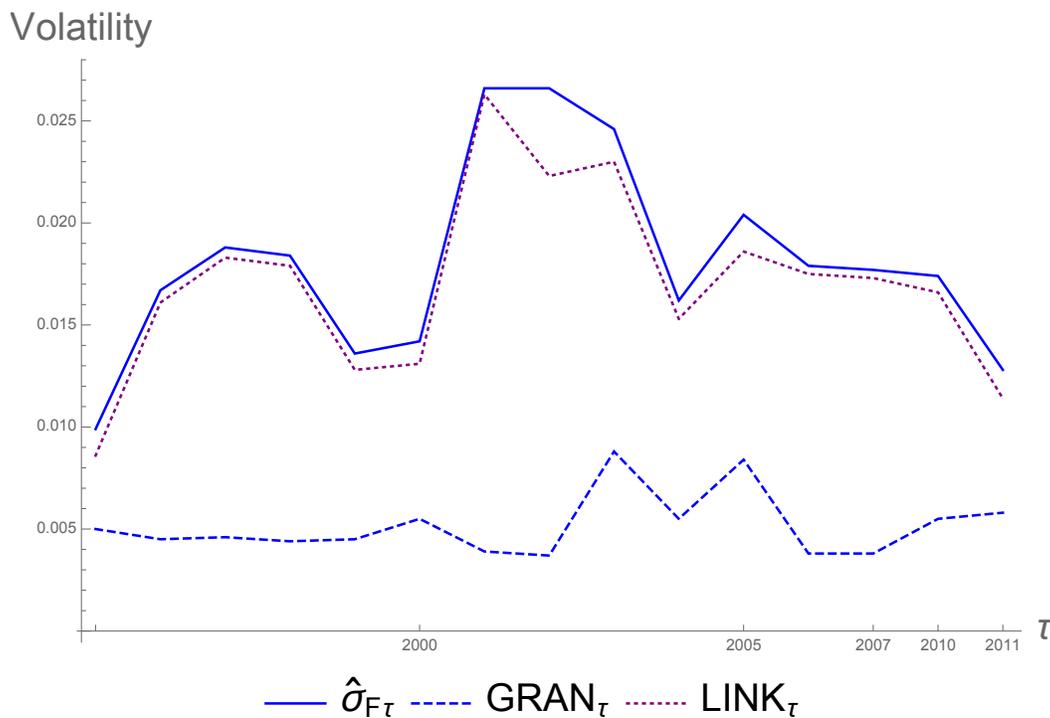


Figure 3.2: Disentangling granular and network components in the firm-specific term $\sigma_{F\tau}^2$.

Most contributions in the literature feature a setup in which shocks to upstream firms affect their downstream partners. The main idea though is that not only the direct partners of the upstream firms are affected but also the downstream firms of these partners and so forth. Higher order connections make firm-level idiosyncratic shocks cascade; therefore, propagating them downwards through the supply chain. Thus, firm-level idiosyncratic shocks are amplified and can affect aggregate volatility generating positive covariances in

the residual growth rates of connected firms. However, it is important to note that observing positive covariances between these residual growth rates are only necessary and not sufficient for the existence of these type of input-output linkages. In appendix B.1, I provide a method on how these covariances can possibly be interpreted as networks. However, my main focus is on the granular component, thus I denote this component as $\text{LINK}_\tau = \sum_{(g \neq f) \vee (m \neq n)} \sum_{f,n} w_{gm\tau-1} w_{fn\tau-1} \text{cov}(\varepsilon_{gmt}, \varepsilon_{fnt})$ in the mean time.

Table 3.5: Relative contribution of granular and network components.^e

	Volatility	Relative SD
$\widehat{\sigma}_{F\tau}$	0.0181	1.0000
$\widehat{\text{GRAN}}_\tau$ (- -)	0.0052	0.3045
$\widehat{\text{LINK}}_\tau$ (· ·)	0.0170	0.9374

^eThe column “SD” represents the average values of the firm-specific, granular and linkage components over the sample period, i.e. I report $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\sigma}_{F\tau}$, $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\text{GRAN}}_\tau$ and $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\text{LINK}}_\tau$ respectively. The column “Relative SD” reports the average of the relative standard deviations, i.e. $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\widehat{\text{GRAN}}_\tau}{\widehat{\sigma}_{F\tau}}$ and $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\widehat{\text{LINK}}_\tau}{\widehat{\sigma}_{F\tau}}$.

The results in figure 3.2 and table 3.5 depict a clear picture. The vast majority of the firm-specific component $\widehat{\sigma}_{F\tau}$ is dominated by the term that reflects comovements between firms as its average relative contribution $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\text{LINK}}_\tau / \widehat{\sigma}_{F\tau}$ averages at 93.74 percent. Furthermore, it comoves almost perfectly with the total firm-specific volatility. While the relative contribution of the granular component $\widehat{\text{GRAN}}_\tau$ is much smaller, it is still non-negligible.¹¹ Its relative contribution over the whole period averages at 30.45 percent. A quick back-of-the-envelope calculation then results in one of the main results of this paper. The relative contribution of the granular component to aggregate volatility equals $52.47\% \times 30.45\% = 15.97$ percent which is significantly less than 30 to 35 percent in Gabaix (2011) and di Giovanni

11. An interesting exception is 2009 in which the U.S. economy experienced a decline in gross output by 8.08 percent: the firm-specific component now primarily consists of the granular component. In appendix B.1, I show the results including the financial crisis in 2008 to 2009. The main conclusions of this section remain unaffected though.

and Levchenko (2012).¹²

3.4 Robustness exercises

The baseline empirical framework assumes that all firms respond identically to aggregate and sectoral shocks. This is consistent with a plethora of heterogeneous firms models.¹³ However, there is some evidence (e.g. Fort et al. (2013) and Moscarini and Postel-Vinay (2012)) that firms' cyclical dynamics are heterogeneous, i.e. firms' responses to aggregate and sectoral shocks are systematically different in the cross-section. This would imply that the firm-specific component is systematically underestimated because it should not only include firm-level idiosyncratic shocks but also heterogeneous responses to aggregate and sectoral shocks that differ by some firm-specific characteristic.

While di Giovanni and Levchenko (2012) argue that the latter channel is quantitatively minor, I perform several robustness checks in which firms are allowed to respond heterogeneously to aggregate and sectoral shocks. To do this, I augment the decomposition of firm-destination level growth rates in subsection 3.1 as follows:

$$\gamma_{fnt} = \delta_{jnt} + \delta_{jnt} \times z_{ft} + \beta z_{ft} + \varepsilon_{fnt}$$

where z_{ft} denotes some firm-specific characteristic that is allowed to vary over time. This means that the augmented firm-specific component equals $\tilde{\varepsilon}_{fnt} \equiv \delta_{jnt} \times z_{ft} + \beta z_{ft} + \varepsilon_{fnt}$.

The firm-specific characteristic z_{ft} either comprises of firm size (as defined by a sales or employment quintile dummy), firm age (dummy for whether the firm is younger than five years) or trade openness. The latter is a dummy that equals one whenever a firm's ratio of export sales to total sales exceeds 10 percent. While the contribution of the firm-

12. Alternatively, I could have calculated the average of the relative granular contribution. This average amounts to $\frac{1}{T} \sum_{\tau=1995}^{2011} \widehat{\text{GRAN}}_{\tau} / \widehat{\sigma}_{A\tau} = 15.54$ percent which is almost identical to the back-of-the-envelope calculation.

13. This includes the majority of frameworks that build upon Dixit and Stiglitz (1977), Krugman (1980) and Melitz (2003).

specific component increases somewhat whenever aggregate and sectoral shocks are subject to heterogeneous responses in firm size, the relative contribution of the granular component decreases as well. Thus, my main conclusions on the contribution of the granular channel hold.

For age and trade openness, the results in subsection 3.3 are instead reinforced as the relative contribution of the firm-specific components are lower than in the benchmark case. These results are robust whenever I use a continuous measure of size or actual age instead of dummies. Overall, the contribution of the granular channel is reduced under these “heterogeneous response” specifications as well.

Lastly, I implement a decomposition of firm-destination level growth rates that takes heterogeneity in geographical location into account. Local factor markets (e.g. labor) could expose different regions in the U.S. to location-specific shocks. Specifically, this is a shock that affects all firms located within a specific geographical area. Thus, a firm-destination level growth rate gets decomposed as:

$$\gamma_{fnt} = \delta_{jnt} + \alpha_{\ell t} + \varepsilon_{fnt}$$

In this specification, regions (indexed by ℓ) are defined at the county level.¹⁴ The main results are also unaffected under this specification. In appendix B.1, I consider other robustness tests. First, I verify that my results on the diminished role of the granular channel are robust when considering firm-level entry and exit. Second, I deal with issues related to measurement error in year-to-year growth rates. Third, the use of year-to-year growth rates could underestimate the firm-specific component as firm-level idiosyncratic shocks could propagate in a relatively slow fashion through the buyer-supplier network. However, I show that my results are also robust when using three-year growth rates instead.

14. Counties are defined through their five-digit Federal Information Processing Standard (FIPS) code.

CHAPTER 4

THE REDUCED ROLE OF THE GRANULAR CHANNEL

The variance decomposition exercise depicts a clear picture on the relatively small role of granularity in the U.S. economy. However, this result seems at odds with the evidence in other studies where the granular channel explains approximately one third of observed aggregate volatility. In the following, I use a simple mechanism that can bridge this discrepancy, i.e. the size-variance relationship. First, I estimate the negative relationship between firm-level volatility and size by using data that covers the universe of U.S. firms. Second, I show that previous explanations of this relationship are not consistent with the data. Then, I show that the aggregate implications of the size-variance relationship are sizable. Most importantly, this relationship alone rationalizes the findings in chapter 3.

4.1 Estimation results

The negative relationship between a firm's volatility (i.e. standard deviation of a firm's sales or employment growth) and its size, as measured by the average volume of sales or employee count, was first documented by Meyer and Kuh (1957) and Hymer and Pashigian (1962). Additional or supporting evidence for this fact was later found by Hall (1987), Stanley et al. (1996), Sutton (2002) and Koren and Tenreyro (2013). However, all of these studies relied on a sample of publicly traded firms. This can be problematic as long-run patterns or characteristics of firm-level volatility can be vastly different between public and private firms. In particular, this is emphasized by Davis et al. (2007) who show that even though the volatility of publicly traded firms has been trending up over time, firm-level volatility overall has been declining. The impact of publicly traded firms is completely overwhelmed by the declining volatility amongst privately held firms.

To overcome this selection bias, I will estimate the size-variance relationship using the LBD which contains annual information on employment and payroll for the universe of the

U.S. economy. In particular, I focus on employment. Thus, firm-level volatility is defined as the standard deviation of a firm’s employment growth and size is measured as the average number of employees. Even though revenues are available from 1994 onwards, the use of employment comes with several advantages. First, the time dimension of employment is significantly longer as it is available from 1976 to 2011. Second, the comparison of employment across industries is easier than revenues as the components that constitute sales and receipts (underlying the revenue data) can vary substantially by industry. Furthermore, the content of the revenue data can also differ by the legal structure of the firm or its tax treatment status (see Moreira, 2016). Nevertheless, my results related to the diminished role of the granular channel are robust to using revenues as a measure of size instead.

While Hymer and Pashigian (1962) already conjectured a negative relationship between firm-level volatility and size, it was not until the study by Stanley et al. (1996) that this relationship was formulated as a power law. More precisely, the rate at which firm-level volatility falls in size is constant. Thus, it must be that firm-level volatility, conditional on size, follows the log-linear form $\sigma(g|S) \propto S^{-\alpha}$. The estimates for α under several regression specifications can be found in table 4.1.

Table 4.1: Estimation results on the size-variance relationship as a power law.^g

	EMP	EMP	EMP	EMP
SIZE	-0.249 (0.009)	-0.214 (0.015)	-0.249 (0.011)	-0.213 (0.015)
Fixed effects				
FIRM	N	Y	N	Y
INDUSTRY	N	N	Y	Y
ESTIMATION	$25.95 \cdot 10^6$	$25.95 \cdot 10^6$	$25.95 \cdot 10^6$	$25.95 \cdot 10^6$
SAMPLE				

^gFirm size is defined as employment count. Industry fixed effects are defined at the 3-digit NAICS level. Each specification uses the 5-year standard deviation of annual employment growth rates from 1977 to 2011. All regression specifications contain year fixed effects. Standard errors are clustered at the industry (3-digit NAICS) level. All estimates are statistically significant at the 1 percent level.

This table shows the estimates of α under regression specifications of the natural log of the volatility of employment growth on the natural log of average size. I follow Koren and Tenreyro (2013) and define volatility as the standard deviation of employment growth for non-overlapping five year periods from 1977 to 2011. This procedure also allows me to consider within firm-level variation only. The results depict a clear deviation from Gibrat law's as $\hat{\alpha}$ is obviously bounded away from zero. Under the most basic regression specification, the estimate for α is equal to -0.249 . More importantly though, *all* of the estimates are significantly more negative than found in previous studies using publicly listed firms in Standard and Poor's Compustat database. Previous lower bounds range from -0.16 (Stanley et al. (1996)) to -0.21 (Sutton (2002)) or -0.22 (Koren and Tenreyro (2013)).¹ Table 4.1 shows that even the highest estimates for α are more negative than the lower bounds set by the previous literature. This negative correlation stays remarkably strong even when I control for firm-level fixed effects restricting myself to within-firm variation only. This results in $\hat{\alpha} = -0.214$. An identical picture is displayed when I control for industry-level fixed effects ($\hat{\alpha} = -0.249$) or by including both firm- and industry-level fixed effects ($\hat{\alpha} = -0.213$).² Thus, firm-level volatility declines at a substantially higher rate in size than previously conjectured. I show in section 4.3 that this has substantial aggregate implications in granular economies.

1. Differing estimates on the size-variance relationship between the LBD and Compustat can be rooted in two causes. First, Compustat only covers publicly listed firms. As a result, the firm size distribution is significantly skewed towards the right tail which is highlighted by Axtell (2001). Second, information on employment and revenues are not necessarily restricted to the U.S. as the Compustat database collects information on companies' *international* activities.

2. Alternatively, I estimate the size-variance relationship using the methodology in Stanley et al. (1996) and Sutton (2002). The results are similar and the average estimate for $\hat{\alpha}$ is always contained within the interval $[-0.249, -0.213]$. Its estimate depends on the specification of the size bins, but these estimates do not differ much across different size bin specifications. Details on this estimation procedure can be found in Appendix B.2.

4.2 Robustness and underlying economic mechanisms

The previous set of regressions indicate a clear and robust negative correlation between firm-level volatility and size. Obviously, these reduced form specifications cannot say much about the underlying economic mechanism(s) that might be driving this relationship. Even though the amount of firm-level observables are somewhat limited in the LBD, it is rich enough to rule out a substantial amount of mechanisms that have been proposed in the literature.

In the following, I regress the natural log of the volatility of employment growth on average size (in natural logs) and some additional firm-level characteristic. Table 4.2 shows that the relationship between firm-level volatility and size stays robust even when controlling for a variety of firm-level characteristics.

A popular conjecture states that under a setting in which firms constitute of independent establishments (or business segments), firm-level volatility declines in size as large firms tend to operate with a larger number of establishments. Thus, the size-variance relationship could be driven by a form of output diversification. The first two rows of table 4.2 indicate that the previous estimates are robust to controlling for the number of establishments operating under the firm. Surprisingly, the results indicate that the number of establishments is positively (rather than negatively) related to firm-level volatility. A similar result with the number of business segments in Compustat can be found in Koren and Tenreyro (2013) though. Appendix B.2 contains an additional set of robustness checks on output diversification. The results do not change whenever I control with an indicator variable (1 if firm is multi-unit establishment and 0 otherwise) or estimate the size-variance relationship in the subsample of single-unit firms. Firms with only one establishment display the same elasticity with respect to size. Thus, I conclude that output diversification cannot be driving the size-variance relationship. Note that any form of financial diversification across a firm's establishments or the "partitions by integers" mechanism by Sutton (2002) and Wyart and Bouchaud (2003) are also indirectly ruled out by this procedure.

It has been suggested that firms do not diversify across establishments in terms of output

Table 4.2: Potential explanations for the size-variance relationship.^h

	EMP	EMP
Output diversification		
SIZE	-0.268 (0.012)	-0.221 (0.016)
ESTABLISHMENT COUNT	+0.162 (0.028)	+0.152 (0.015)
Product diversification		
SIZE	-0.266 (0.011)	-0.218 (0.015)
PRODUCT COUNT	+0.307 (0.039)	+0.204 (0.017)
Cohort fixed effects		
SIZE	-0.227 (0.009)	
ESTIMATION SAMPLE	$25.95 \cdot 10^6$	$25.95 \cdot 10^6$
Age blocks		
SIZE	-0.229 (0.009)	-0.214 (0.014)
ESTIMATION SAMPLE	$27.05 \cdot 10^6$	$27.05 \cdot 10^6$
Fixed effects		
FIRM	N	Y

^hFirm size is defined as employment count. Each row denoted by “SIZE” displays the estimate $\hat{\alpha}$ whenever the regression specification controls for output diversification (proxied by number of establishments within a firm), product diversification (proxied by number of unique 6-digit SIC establishments within a firm), age or cohort fixed effects. The exceptions are the rows “ESTABLISHMENTS” and “PRODUCTS (SIC6)” which display the regression coefficient on the number of establishments and products respectively. All regression specifications contain year and industry fixed effects. Standard errors are clustered at the industry (3-digit NAICS) level. All estimates are statistically significant at the 1 percent level.

but through the number of unique products they produce. This is the driving force behind the size-variance relationship in the family of models based on Klette and Kortum (2004).³

3. These are the product innovation frameworks build upon Klette and Kortum (2004), e.g. Lentz and

Unfortunately, I do not directly observe a firm's number of unique products. Instead, I observe the industry code of each of its establishments. The finest level of industry disaggregation at the establishment level is 6-digit SIC which proxies a product relatively well as these are typically defined at the 7-digit SIC level (see Foster et al., 2008). Thus, I proxy a firm's number of unique products by its amount of unique 6-digit SIC establishments. The second category of results in table 4.2 indicates that the size-variance relationship is basically unchanged.⁴ As a result, I also rule out product diversification as a potential mechanism behind the size-variance relationship.

Alternatively, it is possible that the main observable of interest is not firm size but rather firm age. In a world with some form of incomplete information, older firms might be less volatile as they have learned about either their own or competitors' type over time. In the presence of some unobservable fixed effects, young firms have incentives to learn about them (for example as in Jovanovic, 1982) and are willing to trade off higher volatilities in return. If this would be the case, then the size-variance relationship should be much weaker or even disappear when controlling for firm age. However, this does not seem to be case as the third category of results in table 4.2 indicates that the coefficient on size is barely affected when firm age is controlled for. I deal with any issues due to selection by controlling for cohort fixed effects or by estimating the size-variance relationship on a balanced panel. The results in table 4.2 and Appendix B.2 indicate that the results are identical. Lastly, I check whether the size-variance relationship displays a different regime for larger firms by testing for a structural break. The results can be found in Appendix B.2. I find no evidence for such a structural break.⁵

Mortensen (2008), Seker (2012) and Akcigit and Kerr (2015).

4. A similar set of robustness tests (as for the output diversification channel) can be found in Appendix B.2.

5. This is potentially important as the coefficient α could be severely underestimated with ordinary least squares (OLS) in the presence of a structural break for large firms. This can then impact the role of the granular channel for aggregate volatility whenever the size-variance relationship is taken into account.

4.3 Granularity

The results of the variance decomposition exercise in section 3.3 indicate that the granular component of aggregate sales volatility is approximately only 16 percent. This is substantially smaller than the quantitative findings in the literature.⁶ This result seems counterintuitive as the firm size distribution in the U.S. is substantially skewed. Many studies find that the U.S. firm size distribution displays a Pareto tail with a coefficient near one (i.e. Zipf's law), thus the role of granularity should be substantial according to Gabaix' (2011) theorem.⁷

I will argue on the other hand that the importance of the size-variance relationship for granularity has been severely understated. The intuition behind my argument is simple: in a granular economy, aggregate volatility is primarily affected by the largest firms. However, I observe that there is a declining power law relationship between firm-level volatility and size. Whenever the former falls in firm size at a sufficiently high rate, the volatility of the largest firms in the economy becomes small and, hence, their prominence in terms of aggregate fluctuations declines significantly. This seems plausible given the results of section 4.1.

Whenever the firm size distribution is fat-tailed, Gabaix (2011) argues that the $1/\sqrt{N}$ diversification rule is no longer applicable as the conditions for the central limit theorem are violated. Instead aggregate volatility decreases at the rate $1/N^{(\zeta-1)/\zeta}$ where ζ is the Pareto right tail of the firm size distribution. Obviously, the rate of convergence becomes slower and slower as $\zeta \rightarrow 1$. Even though Gabaix (2011) formalizes the case in which firm-level volatility falls in size at the rate α , this case is mostly ignored by the literature. It can be shown that the rate of convergence can be significantly higher for non-trivial values of α (i.e. $\alpha \neq 0$) as aggregate volatility behaves asymptotically as $1/N^{\zeta'}$ where

6. A large section of the literature finds that the granular mechanism can explain roughly one-third of aggregate fluctuations, e.g. Gabaix (2011), di Giovanni and Levchenko (2012), di Giovanni, Levchenko and Mejean (2014) and Carvalho and Grassi (2015). However, there does not seem to be a clear consensus as other studies conclude that the role of granularity is either larger (e.g. Magerman et al., 2015) or substantially smaller (Stella, 2015).

7. Axtell (2001) finds a Pareto tail of 1.058 using the U.S. Census Bureau employment cross-section for 1997. Luttmer's (2007) estimate for 2002 is extremely similar as it equals 1.06.

$\zeta' = \min\{(\zeta - 1 + \alpha)/\zeta, \frac{1}{2}\}$. However, this result is not sufficient to determine the quantitative decline of aggregate volatility in a granular economy as α increases.

In the following, I employ a stylized framework with heterogeneous firms subject to firm-specific idiosyncratic shocks alone. Most importantly, it features a fat-tailed firm size distribution and a declining power law relationship between firm-level volatility σ and its size S in reduced form:

$$\begin{aligned}\mathbb{P}(S > x) &\propto x^{-\zeta} \\ \sigma(g|S) &= AS^{-\alpha}\end{aligned}$$

Note that any framework with firm-specific idiosyncratic shocks alone that delivers these relationships in reduced form is sufficient for my quantitative exercise. Most importantly, my results do not hinge on a particular microfoundation for the size-variance relationship. Appendix C lays down the explicit details of a model that delivers these relationships in reduced form.⁸ In this framework, it can be shown that aggregate volatility, which is defined as the standard deviation of aggregate sales growth γ_A , can be calculated as:

$$\text{SD}(\gamma_A; \alpha) = A \cdot \sqrt{\sum_k (x_k^{-\alpha} s_k)^2}$$

where x_k and $s_k = x_k / \sum_{k'} x_{k'}$ denote a firm k 's sales and its sales share in the economy respectively. This formula nests the case in Gabaix (2011).

Under his baseline scenario by Gabaix (2011), the size-variance relationship is ignored (i.e. $\alpha = 0$) and the economy's baseline firm volatility is set to $A = \bar{\sigma}$ where $\bar{\sigma}$ usually equals the average volatility of the top 100 firms in Compustat.⁹ Even though the average volatility

8. This model is in line with the autarky case of di Giovanni and Levchenko (2012). I chose the autarky case in particular for simplicity. It is relatively straightforward to consider a framework with international trade and an additional non-tradeable sector. di Giovanni and Levchenko (2012) show however that any conclusions on aggregate volatility for large countries such as the U.S. are barely affected by adding international trade.

9. In this particular case, aggregate volatility is equal to the product of firm-level volatility $\bar{\sigma}$, which is set

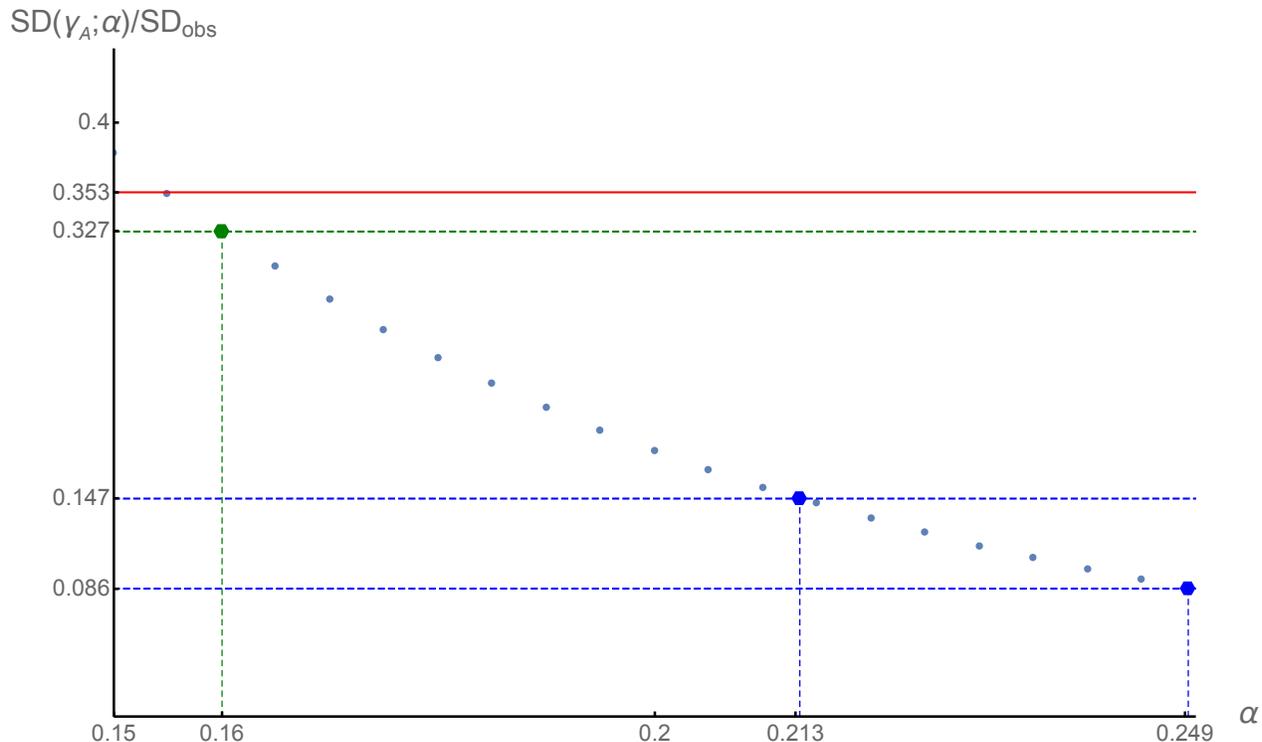


Figure 4.1: The volatility of the simulated granular economy $SD(\gamma_A; \alpha)$ relative to observed volatility in the U.S. is displayed for several values of $-\alpha$. Observed volatility SD_{obs} is defined as the standard deviation of real GDP growth in the U.S. over the period 1947 - 2011. The blue lines indicate the explanatory power of the granular channel whenever α is set equal to either the upper (-0.213) or lower bound (-0.249) of estimates in table 4.1. The green line displays the explanatory power of granularity under the previous estimate of α by Stanley et al. (1996). Lastly, the red line depicts the explanatory power of granularity whenever I abide by the calibration strategy of the previous literature.

of the top 100 firms in Compustat was previously considered to be a conservative choice, I show that for sufficiently high values of α , most importantly as those observed in section 4.1, this calibration choice can still substantially overestimate the contribution of the granular channel.

The results can be found in figure 4.1. The figure displays the ratio of the volatility of the simulated economy and observed volatility in the U.S. for a range of α . Furthermore,

equal to 0.12, and the square root of the economy's sales Herfindahl. My results indicate that the volatility of large firms is considerably lower than 12 percent. This could be due to two reasons. First, my results are derived from the LBD instead of Compustat. Second, Gabaix (2011) constructed firm-level volatility by measuring the average cross-sectional dispersion in firm-level growth rates among the top 100 firms. Instead, I measure firm-level volatility by considering the variation of a firm's growth rate over *time*.

I set $A = 1.3$ which is well above the upper bound reported in Davis et al. (2007). While this value is high, it guarantees that the volatility of a typical large firm implied by the estimates of section 4.1 and 4.2 is not underestimated.¹⁰ Observed volatility is calculated as the standard deviation of real GDP growth over the period 1947 - 2011 and equals 0.0238. This value is calculated using the Federal Reserve Economic Database from the St. Louis Federal Reserve.

I find that the aggregate implications of the size-variance relationship are substantial. The explanatory power of granular origins declines to roughly 15% whenever the size-variance relationship, as observed in the data, is taken into account. Whenever I use the most conservative estimate for α , i.e. $\hat{\alpha} = -0.213$, the granular channel only accounts for approximately 14.7% of observed aggregate volatility. On the other hand, the role of granularity is substantially smaller whenever I use the lower bound estimate for α . Table 4.1 indicates that this is $\hat{\alpha} = -0.249$. Under this scenario, the contribution of granularity to aggregate observed volatility reduces to about 8.6% instead.

This is consistent with the results of the variance decomposition exercise of section 3.3. Most importantly, my results indicate that this decline can purely be attributed to the incorporation of the size-variance relationship. A potential concern is that the estimated size-variance relationship overstates the implied volatility of the largest firms in the economy. Even though this is taken into account in the calibration strategy above by using an extremely conservative choice for A , I could also implement the values $(\hat{A}, \hat{\alpha})$ that are estimated from a subsample containing only the largest firms in the economy. In Appendix C.2, I show that the results are extremely similar. Under some specifications, the contribution of the granular channel is even lower than shown above.

Whenever I follow the calibration strategy of most granularity studies, the framework pre-

10. To illustrate this, the implied volatility of a firm with 200 000 employees which roughly corresponds to a top 20 firm in the U.S., approximately equals 10%. Note that if I set A equal to the value implied by the *actual* estimate of section 4.1, then the role of the granular channel becomes even smaller. As a result, the conclusions drawn from my quantitative exercise are extremely conservative.

dicts that the granular channel can approximately generate 35.3% of the observed volatility in the U.S. (red line in figure 4.1) as consistent with findings by Gabaix (2011), di Giovanni, Levchenko and Mejean (2014) and Carvalho and Grassi (2015). This implies that my results are purely driven by the size-variance relationship alone. Note that it is important to use the estimates on the universe of the U.S. economy (i.e. LBD) as previous estimates based on Compustat (e.g. Stanley et al. (1996)) find a value of $\hat{\alpha} = -0.16$ do not have a substantial impact on granularity. Whenever these previous estimates are implemented instead, the role of the granular channel is still significant as it can explain 32.7% of observed aggregate volatility (as captured by the green line in figure 4.1).

CHAPTER 5

A MICROFOUNDATION FOR THE SIZE-VARIANCE RELATIONSHIP

In chapter 4, I provided estimates on the size-variance relationship using the most extensive data set covering the U.S. economy. Moreover, I shed light on the potential economic mechanisms that could be driving this deviation from Gibrat’s law and found that several explanations brought forth by the literature can be rejected by the empirical results. However, there remain a handful of mechanisms that have been explored before that cannot be directly tested with the available data. To deal with this, I will argue that these remaining explanations are not consistent with markups being increasing in size. First, I structurally estimate markups without making assumptions on the structure of demand and type of market competition. Second, I establish that there is a robust positive correlation between markups and size. To accommodate these findings that are also consistent with the size-variance relationship, I propose a “market power” mechanism in which large firms face smaller price elasticities of demand. To avoid confounding competing mechanisms in the data that cannot be observed, I propose a tractable model of firm dynamics featuring size-varying demand elasticities that is consistent with the size-variance relationship as a power law and the positive correlation between markups and size. Lastly, I provide direct evidence from a subset of manufacturing industries that is consistent with the main assumption of the proposed framework.

5.1 Size-varying markups

5.1.1 Parametric method

I will estimate markups using two distinct methods. Under the first approach, I impose that a firm’s production technology is Cobb-Douglas. This is extremely convenient as the

firm's marginal cost function $c'(Q)$ is proportional to its total cost function over its level of total output $c(Q)/Q$. Let the returns to scale parameter be denoted by β , then it is straightforward to derive that a firm's markup must satisfy:

$$\beta \frac{\text{rev}(Q)}{c(Q)} = \mu(Q)$$

where $\text{rev}(Q)$ denotes a firm's revenue as a function of its total level of output Q . This identity is insightful as revenue and total costs can be observed directly from the data.¹ However, I can only observe total costs for plants in the manufacturing sector from 1976 to 2009. As a result, I will restrict my analysis on size-varying markups at the plant-year level to the manufacturing sector. Therefore, I am able to obtain estimates of markups up to a constant by only using information on revenue and total costs. Under constant returns to scale, this approximation is even exact as $\beta = 1$. Whenever the technology displays decreasing returns to scale, i.e. $\beta < 1$, the ratio of revenue and total costs overestimate markups as it then satisfies $\mu(Q)/\beta > \mu(Q)$. Most importantly however, the relationship between this measure of markups and size should be unaffected as the ratio of revenue and total costs is only off by a constant. In the following, I will denote this specific measure of markups by μ^{CRS} .

5.1.2 *Semi-parametric method*

The previous method strongly relies on the Cobb-Douglas production structure. To strengthen my results, I will also estimate markups structurally using the GMM-IV methodology by de Loecker and Warzynski (2012). I will only describe the estimation procedure briefly in this section, but a fully detailed overview can be found in Appendix B.4. The main advantage of this methodology is that markups at the plant-year level can be retrieved from a minimal set of assumptions. This involves plants engaging in cost minimization and the existence of

1. See Appendix A.4 for an extensive discussion of what is contained in a plant's measure of total cost.

at least one flexible input X . The key insight is then that markups at the plant-year level μ_{it} satisfy the following equality:

$$\theta_{it}^X = \mu_{it} \frac{P_{it}^X X_{it}}{P_{it} Q_{it}}$$

where $\theta_{it}^X = \frac{\partial \ln Q_{it}(\cdot)}{\partial \ln X_{it}}$ denotes a plant i 's output elasticity with respect to the flexible input X in year t . This equality is valid as long as input X is static because significant adjustment costs would drive a wedge between a plant's elasticity θ_{it}^X , its revenue share of input X and its markup. This equality is useful as the revenue share of several inputs can be directly constructed from the ASM/CM data. Thus, markups at the plant-year level can then be constructed by obtaining an estimate for θ_{it}^X .

To do this, I will make the following simplifying assumptions. First, I will assume that material inputs M_{it} are flexible.² Second, I impose $\theta_{it}^M = \beta_m^{j(i)}$. This means that output elasticities with respect to inputs are constant over time and do not vary across firms within a industry j . This assumption boils down to an industry-by-industry Cobb-Douglas production function estimation procedure.³ Thus, production technologies are allowed to differ across, but not within, industries. To allow for a sufficient degree of heterogeneity in production technologies, I will estimate technology parameters separately at the 3-digit NAICS level.

Furthermore, I allow for measurement error in gross output which results in observed log output satisfying $y_{it} = \ln Q_{it} + \epsilon_{it}$. Note that firms do observe their level of productivity ω_{it} but ϵ_{it} is neither observed by the firm nor the econometrician. In the ASM/CM data, material inputs and energy inputs are registered separately. Thus, the estimated production

2. Alternatively, I could have estimated a value added production function and identified markups from labor. This could potentially be problematic in the presence of labor adjustment costs as these drive a wedge between a plant's markup and the ratio of its output elasticity with respect to labor and revenue share of labor. Caballero et al. (1997) do find evidence in favor of non-convexities in labor adjustment cost technologies in a sample of U.S. manufacturing plants. Furthermore, the findings by Gandhi et al. (2013) highlight the empirical importance of the misspecification one is faced with when using value added.

3. The assumption of time-invariant production technologies can be relaxed. See footnote 15 in de Loecker and Warzynski (2012). Furthermore, I could allow for output elasticities that vary at the plant level by estimating a translog production function which can be shown to be a (local) second-order approximation to an arbitrary production function. This implies that output elasticities also vary *within* industries.

function specification (in log levels) becomes:

$$y_{it} = \omega_{it} + \beta_k k_{it} + \beta_\ell \ell_{it} + \beta_m m_{it} + \beta_e e_{it} + \epsilon_{it}$$

Obviously, this specification cannot be estimated with least squares as productivity ω_{it} is unobserved by the econometrician leading to the well-known “transmission bias” problem: ω_{it} is correlated with the set of inputs resulting in biased and inconsistent estimates of $\boldsymbol{\beta} = (\beta_k, \beta_\ell, \beta_m, \beta_e)$.⁴ To resolve this issue, I follow Levinsohn and Petrin (2003) and use a “proxy” method by exploiting material inputs which can, in turn, be used to construct a set of identifying moments. To do this, I make a semi-parametric assumption on plant-level productivity:

$$\omega_{it} = g_t(\omega_{it-1}) + \xi_{it}$$

where ξ_{it} denote the innovations to productivity. Identification of $\boldsymbol{\beta}$ is obtained by relying on the following four moment conditions:

$$\mathbb{E} \left[\begin{array}{c} \xi_{it}(\boldsymbol{\beta}) \\ \left(\begin{array}{c} k_{it} \\ \ell_{it-1} \\ m_{it-1} \\ e_{it-1} \end{array} \right) \end{array} \right] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Estimates of $\boldsymbol{\beta}$ are then obtained by using standard GMM techniques which involves the minimization of a quadratic loss function. The main assumptions underlying identification exploit the fact that capital is chosen a period ahead and should therefore be orthogonal to

4. It is not clear ex-ante whether OLS estimates will under- or overestimate the production function coefficients. More productive firms are able to produce the same level of output with, for example, less labor compared to less productive firms. Whenever production technologies are assumed to be identical across firms within a industry, this implies that β_ℓ will be underestimated under OLS. On the other hand, firms experiencing positive productivity shocks will most likely hire more labor. However, the estimate of β_ℓ under OLS will incorrectly contribute the increase in output to this change in labor. As a result, this means that $\widehat{\beta}_\ell^{OLS} > \beta_\ell$. More thorough discussions on the “transmission bias” problem in production function estimation can be found in Griliches and Mairesse (1996) and Eberhardt and Helmers (2016).

future innovations to productivity. A similar logic holds for the lagged inputs ℓ_{it-1} , m_{it-1} and e_{it-1} . While the plead for exogeneity is clear, the argument to satisfy rank conditions is not necessarily straightforward. For a lagged input to be a valid instrument for current input, some persistence in input prices is required. Recall from the discussion above that markups are obtained through the output elasticity of intermediate inputs. Using plant-level data in the U.S. manufacturing sector, Atalay (2014) shows that plant-level prices for material inputs are highly persistent. This seems to confirm the validity of the rank condition as material input prices indeed appear to be serially correlated over time.

Markups at the plant-year level are then simply constructed through:

$$\hat{\mu}_{it}^{GMM} = \hat{\beta}_m^{j(i)} \frac{P_{it}^M M_{it}}{P_{it} Q_{it} / \hat{\epsilon}_{it}}$$

where $j(i)$ denotes the industry to which firm i belongs to. Note that I explicitly correct observed gross output for measurement error with $\hat{\epsilon}_{it}$.⁵

5.1.3 Estimation results

The estimation results, obtained industry-by-industry (3 digit NAICS), are shown in table 5.1. The first two columns present the mean of the markup distribution obtained for each industry.⁶ Mean level of markups vary substantially across industries and the average value of these means is equal to 1.329. These values seem to be consistent with estimates using Slovenian (de Loecker and Warzynski, 2012), Chilean (Lamorgese et al., 2014), Chinese (Lu et al., 2014) and Indian (de Loecker et al., 2016) data. While there is a significant amount

5. This is potentially important as this correction will eliminate any variation in expenditure shares that comes from variation in output *not* related to variables impacting input demand (see de Loecker and Warzynski (2012)). This is particularly crucial whenever inputs are constructed by deflating input expenditures with industry-level deflators as mentioned by Klette and Griliches (1996).

6. Under decreasing returns to scale, the methodology using the ratio of revenues over total costs overestimates markups. The GMM-IV estimation procedure seems to be a decent job of correcting this bias as I find that, in general, $\text{med}(\hat{\mu}_{it}^{CRS}) \geq \text{med}(\hat{\mu}_{it}^{GMM})$. Due to disclosure restrictions however, I can only report mean rather than median values of the markup distributions.

of variation in markups between industries, the within-industry variation is also substantial.

Table 5.1: Estimation results of markup distributions by manufacturing industry group.^j

INDUSTRY GROUP	μ^{CRS}	μ^{GMM}	SD(μ^{GMM})	$\hat{\beta}_m^{OLS}$	$\hat{\beta}_m^{GMM}$
Food and Kindred Products	1.319	1.161	0.605	0.585	0.595
Textile Mill Products	1.229	1.236	0.620	0.527	0.565
Apparel and Leather	1.331	1.570	1.035	0.440	0.569
Lumber	1.186	1.169	0.519	0.558	0.612
Furniture and Fixtures	1.303	1.428	0.488	0.544	0.588
Paper and Allied Products	1.241	1.136	0.382	0.564	0.582
Printing and Publishing	1.386	1.296	0.652	0.363	0.376
Chemicals	1.403	1.420	0.770	0.560	0.597
Petroleum Refining	1.190	1.220	0.412	0.722	0.735
Plastics and Rubber	1.303	1.366	0.512	0.545	0.593
Non-metallic Minerals	1.333	1.433	0.755	0.504	0.534
Primary Metals	1.275	1.420	0.713	0.565	0.619
Fabricated Metal Products	1.280	0.999	0.548	0.433	0.353
Non-electrical Machinery	1.322	1.437	0.695	0.482	0.518
Electrical Machinery	1.358	1.404	0.555	0.539	0.573
Motor Vehicles and Other Transportation equipment	1.287	1.366	0.640	0.558	0.614
Computer and Electronic Products	1.406	1.249	0.635	0.410	0.407
Miscellaneous Manufacturing	1.387	1.613	0.812	0.472	0.547
AVERAGE	1.308	1.329	0.630	0.521	0.554
ESTIMATION SAMPLE	1.157 · 10 ⁶				

^jThe estimation sample covers the U.S. manufacturing industries (NAICS 31 - 33) over the period 1976 to 2009. The second and third columns depict the mean estimate of a plant's markup within an industry group. μ^{CRS} denotes this value under the assumption of constant returns to scale and a Cobb-Douglas production function. μ^{GMM} is obtained using the GMM-IV estimation procedure and reports the mean under the assumption of a Cobb-Douglas specification for gross output only. The column SD(μ^{GMM}) reports the standard deviation of the markup distribution within the industry group. The columns $\hat{\beta}_m^{OLS}$ and $\hat{\beta}_m^{GMM}$ report the estimated coefficients of the production function on intermediate inputs using OLS and the GMM-IV estimation procedure. All estimation procedures use deflated wage bill as labor input. Lastly, the table trims observations with markups above the 99th percentile.

The third column displays the standard deviation of the markup distribution within each industry: the mean of this standard deviation across industries amounts to 0.630. Lastly, columns 4 and 5 show the estimates of the output elasticity with respect to intermediate inputs. As can be seen from table 5.1, the OLS estimator tends to slightly underestimate

β_m highlighting the transmission bias in production function estimation.

While these estimates are interesting in their own right, I am primarily focused on whether markups vary with size. To assess this relationship, I perform the following regression in the pooled sample:

$$\ln \widehat{\mu}_{it} = \gamma_0 + \gamma_1 \ln S_{it} + \mathbf{\Gamma}'_{it} \boldsymbol{\sigma} + \nu_{it}$$

where a plant’s size is reflected by its total number of employees and $\mathbf{\Gamma}_{it}$ contains a full set of industry-year interactions to control for industry-specific aggregate trends in markups. Depending on the specification, $\mathbf{\Gamma}_{it}$ can also include plant-level fixed effects. I run this regression specification for both measures of markups (either $\widehat{\mu}_{it} = \widehat{\mu}_{it}^{CRS}$ or $\widehat{\mu}_{it} = \widehat{\mu}_{it}^{GMM}$). The results are summarized in table 5.2.

This table depicts a clear picture: plant-level markups increase in size regardless of the way markups are measured. Most importantly, this relationship even seems to hold whenever I restrict myself to variation in markups within plants. The markup elasticity with respect to size varies from 0.013 to 0.057 depending on whether I control for plant fixed effects. Note that this result is not obvious as the markup estimation procedure does not impose any restrictions on demand structures or the nature of competition. As a result, there is no reason to believe that the estimated markups are mechanically related to size.⁷

These estimates are furthermore sizable. Under the specification with industry-year and plant-level fixed effects for $\widehat{\mu}_{it} = \widehat{\mu}_{it}^{GMM}$, a typical “large” manufacturing plant with 100 employees charges a 6.25% higher markup than its smaller counterpart with only 10 employees which is relatively high considering markups are defined as profit per produced unit. Thus, I conclude the positive correlation between markups and size is robust and economically sizable.

7. This is different from the case in which, for example, the logit demand structure is imposed. Under this setup, a firm’s own price elasticity at price p with market share s is characterized by $dp(1 - s)$ where d denotes the disutility to a consumer from paying an additional unit for the firm’s product. Thus, price elasticities are decreasing in size which implies that markups are, by construction, increasing in size.

Table 5.2: Estimation results of relationship between firm-level markups and size.^k

	μ^{CRS}	μ^{CRS}	μ^{GMM}	μ^{GMM}
SIZE	0.013 (0.000)	0.057 (0.001)	0.018 (0.001)	0.022 (0.002)
CONSTANT	0.111 (0.009)	-0.005 [†] (0.012)	0.162 (0.008)	0.135 (0.013)
Fixed effects				
INDUSTRY-YEAR	Y	Y	Y	Y
PLANT	N	Y	N	Y

^kThe estimation sample covers the U.S. manufacturing industries (NAICS 31 - 33) over 1976 to 2009. The independent variable is always the natural log of markups. In columns 2 and 3, markups are constructed by using the ratio of revenues over total costs. Columns 3 and 4 employ markups as constructed by the GMM-IV methodology contained in the text. Standard errors are clustered at the industry (3-digit NAICS) level. All estimates are significant at the 1 percent level unless denoted by [†].

5.2 Identification: alternative mechanisms

The results of section 4 and the stylized facts mentioned above depict a clear picture: a rich set of firm-level outcomes are varying in firm size. This seems to be in stark contrast with a substantial fraction of the firm dynamics literature in which the strong version of Gibrat's law holds. The frameworks based on Klette and Kortum (2004) are one of the few exceptions as these models predict that both firm-level growth and volatility decline in firm size. However, the main driving mechanism behind the size-variance relationship in this set of product innovation models is product diversification. The results in section 4.2 clearly showed that the size-variance relationship stays robust when controlling for product diversification.⁸

As far as I am aware of, there are three alternative explanations that deal with deviations from Gibrat's law. Rossi-Hansberg and Wright (2007) focus on a mechanism in

8. This also rules out the mechanism described in Luttmer (2011) whenever "blueprints" are interpreted as either production lines or plants, see section 2.6.2 of Luttmer (2011).

mean reversion in industry-specific human capital accumulation. Under decreasing returns to scale, high levels of human capital lead to low rates of return and slower accumulation. Whenever establishment sizes respond monotonically in changes in factor prices, this creates mean reversion in establishment sizes as well. While establishment-level growth rates are declining in size, Rossi-Hansberg and Wright (2007) use a competitive model. Thus, there are no predictions on establishment-level markups. In their appendix A, it is shown that their results are robust to a setting with Dixit-Stiglitz monopolistic competition. However, this only creates variation in markups across industries. More importantly, markups within industries do not vary with size which contradicts the findings of section 5.1.

Koren and Tenreyro (2013) also conclude that output or product diversification cannot account for a firm's decline of volatility in size. As a result, they focus on *input* diversification: larger firms are less volatile as they can use a wider number of input factors to smooth out firm-level idiosyncratic shocks. However, it is hard to reconcile their framework with the stylized fact on size-varying markups. Firms face marginal costs that decrease in the number of input varieties. Hence, larger firms are characterized by lower marginal costs. Due to the use of CES aggregators however, a firm's markup remains constant in size. Lastly, Arkolakis (2016) adopts a setup with market penetration costs in which firms pay a cost that is convex in the share of consumers of the destination market. His elegant framework is analytically tractable and most importantly generates deviations from Gibrat's law. In particular, large firms are less volatile as they make relatively small adjustments in the extensive margin of reaching consumers. As a result, the effective demand elasticity is small for larger firms. While this generates a declining size-variance relationship qualitatively, it is hard to reconcile this with a power law as observed in the LBD data. Under Arkolakis' (2016) setup, a firm's instantaneous variance of sales growth approaches a constant as its size increases in the limit to infinity. This implies that the relationship between firm-level volatility and size must display a kink at a certain size. I show in Appendix B.2 however that there is no evidence for a structural break in the size-variance relationship. Lastly, prices in this framework are

characterized by a constant markup over marginal costs which is contradictory with the evidence of section 5.1.

Therefore, I suggest a new channel of “market power” in which large firms face smaller price elasticities of demand than small firms do. To show that this mechanism is capable of generating the size-variance relationship, I construct an analytically tractable framework in continuous time featuring random growth à la Luttmer (2007) and a Kimball aggregator (Kimball, 1995).

This aggregator results in a price elasticity that is declining in firm size. For a given percentage change in firm-level productivity, percentage changes in revenues are then lower for large firms compared to small firms which is equivalent to a declining size-variance relationship. Despite its parsimonious parametrization, which is taken from Klenow and Willis (2016), it is sufficiently flexible to generate a size-variance relationship that behaves as a power law. Lastly, the framework also predicts that markups are increasing in firm size as consistent with the results in section 5.1.⁹

5.3 Framework

5.3.1 Consumers

Time is continuous and indexed by t . There is a continuum of consumers $L_t = L \cdot \exp(\eta t)$ each of whom supply one unit of labor inelastically. Population growth η is non-negative. Dynastic preferences over per-capita consumption flows $\{C_t\}_{t \geq 0}$ are given by:

$$\left(\mathbb{E}_t \left[\int_{\tau=t}^{\infty} \exp(-\rho\tau) C_{\tau}^{1-\gamma} d\tau \right] \right)^{1/(1-\gamma)}$$

9. The Kimball aggregator is only one particular demand specification in which price elasticities are decreasing in firm size. However, it comes with a few advantages. As demonstrated in section 5.3, it allows for a *tractable* setup of firm dynamics in a random growth model. Furthermore, its parameterization is flexible enough to generate a size-variance relationship as a power law which is consistent with my findings of section 4. Nevertheless, I present a class of preferences in appendix D.6 that is qualitatively compatible with a negative (positive) relationship between firm-level volatility (markups) and size.

where ρ is the discount rate and the intertemporal elasticity of substitution $1/\gamma$ are positive. The composite bundle C_t consist of a continuum of differentiated goods and is aggregated in a costless fashion through the Kimball (1995) structure:

$$\frac{1}{|\Omega|} \int_{\omega \in \Omega} \psi \left(\frac{|\Omega| \cdot c_t(\omega)}{C_t} \right) d\omega = 1$$

where $\psi(1) = 1$, $\psi'(\cdot) > 0$ and $\psi''(\cdot) < 0$. Under these assumptions, this Kimball aggregator features a price elasticity that decreases in the relative quantity consumed of a monopolistically competitive good.¹⁰ In general, the composite good C_t is only implicitly defined through the Kimball structure. For ease of exposition, I will drop time subscripts unless denoted otherwise. In the balanced growth path constructed below, per-capita consumption and the wage rate grow at a constant rate κ . Whenever the composite good functions as the numéraire, the interest rate is constant at $r = \rho + \gamma\kappa$.

In the following, I will choose a specific functional form for the Kimball aggregator $\psi(x)$. Let x denote the *relative* quantity consumed of some individual variety, then I follow Klenow and Willis (2016) by choosing:

$$\psi(x) = 1 + (\theta - 1) \exp \left(\frac{1}{\varepsilon} \right) \varepsilon^{\theta/\varepsilon - 1} \left(\Gamma \left(\frac{\theta}{\varepsilon}, \frac{1}{\varepsilon} \right) - \Gamma \left(\frac{\theta}{\varepsilon}, \frac{1}{\varepsilon} x^{\varepsilon/\theta} \right) \right)$$

where $\Gamma(u, z) = \int_z^{+\infty} s^{u-1} \exp(-s) ds$ is the incomplete gamma function. This specific aggregator implies demand curves of the following form:

$$c(\omega) = \varphi \left(\frac{p(\omega)}{P} \right) \frac{C}{|\Omega|} \quad \text{with} \quad \varphi(x) = \left[1 + \varepsilon \ln \left(\frac{\theta - 1}{\theta} \frac{1}{x} \right) \right]^{\theta/\varepsilon}$$

This specification has several advantages. First, the framework allows for a tractable specification. Second, the aggregator is parsimoniously specified by only two parameters θ and ε

10. The standard model of monopolistic competition which features a constant elasticity of substitution (CES) aggregator is nested within this specification. This is done through $\psi(x) = x^{(\theta-1)/\theta}$ for some $\theta > 1$.

which control the rate at which price elasticities vary with relative quantities $x = c/C$. This result can be shown analytically as:

$$\begin{aligned}\sigma(x) &\equiv -\frac{d\ln\varphi(p)}{d\ln(p)} \\ &= \theta x^{-\varepsilon/\theta}\end{aligned}$$

Third, the specification has a natural interpretation. In the limit of $\varepsilon \rightarrow 0$, the Kimball aggregator converges to the standard CES aggregator. As a result, the parameter ε can be interpreted as the “deviation” from a CES specification with elasticity $\theta > 0$.

5.3.2 Firms

A firm is characterized by a unique technology that allows it to produce a particular commodity as a monopolist. Each firm with productivity z produces $z\ell$ units of output with ℓ units of labor. Then, its optimal pricing decision is characterized as:

$$p^*(z) = \arg \max_{p \geq 0} \left(p - \frac{w}{z} \right) \varphi \left(\frac{p}{P} \right)$$

where each firm takes the aggregate price level P as given. The necessary first order condition can be rearranged in the form of an endogenous markup function over marginal costs. This results in:

$$\left[\frac{\tilde{\sigma} \left(\frac{p}{P} \right)}{\tilde{\sigma} \left(\frac{p}{P} \right) - 1} \right] \frac{w}{z} = p \tag{A}$$

where $\tilde{\sigma}(x) = 1 / \left(1 - \frac{1}{\theta} \left[1 + \varepsilon \ln \left(\frac{\theta-1}{\theta} \frac{1}{x} \right) \right] \right)$.¹¹ Even though it appears that equation A does not have a closed form solution, it can still be expressed in a meaningful way by using

11. Note that the CES case of $\varepsilon \rightarrow 0$ would lead to the constant markup over marginal costs in which $p^* = \frac{\theta}{\theta-1} \frac{w}{z}$ as $\tilde{\sigma}(x) \rightarrow \theta$ for all x .

the Lambert W -function. In particular, the Lambert W -function (or product logarithm) is defined as the inverse of the mapping $x \mapsto x \cdot \exp(x)$ which has some properties that I can exploit later.¹² As a result, it can be shown that a firm's optimal price equals:

$$p^*(z) = \frac{\theta}{\varepsilon} \frac{1}{W\left[\Theta \frac{w}{z} \frac{1}{P}\right]} \frac{w}{z} \quad \text{where} \quad \Theta \equiv \frac{\theta}{\varepsilon} \frac{\theta}{\theta - 1} \exp\left(\frac{\theta - 1}{\varepsilon}\right)$$

Using the expression for a firm's optimal price, firm-level revenue can be simplified to:

$$\begin{aligned} \text{rev}(z) &\equiv p^*(z) \varphi\left(\frac{p^*(z)}{P}\right) \frac{C}{|\Omega|} \\ &= \frac{\theta}{\varepsilon} \frac{1}{W\left[\Theta \frac{w}{z} \frac{1}{P}\right]} \frac{w}{z} \left(\theta - \varepsilon W\left[\Theta \frac{w}{z} \frac{1}{P}\right]\right)^{\theta/\varepsilon} \frac{C}{|\Omega|} \end{aligned} \quad (\text{B})$$

In the following, firm-level revenue will be the measure for a firm's size.¹³

PRODUCTIVITY DYNAMICS AND SELECTION. I follow Luttmer (2007) and Arkolakis (2016) and assume that the productivity of a firm born at time t^b with age a evolves according to the following stochastic process:

$$z_{t^b,a} = \bar{z} \cdot A_t \cdot \exp\left(\theta_E t^b + \theta_I a + \sigma_F \mathcal{B}_{t^b,a}\right)$$

where $\{\mathcal{B}_{t^b,a}\}_{a \geq 0}$ is a standard Brownian motion. A new firm at time t enters the economy with productivity $z_{t,0} = \bar{z} \cdot \exp(\theta_E t)$. This implies that more recent entrants are of higher quality.¹⁴ Log productivity of incumbents trend upwards at rate θ_I . Furthermore,

12. In particular, this function $W(x)$ is single-valued for any $x \geq 0$. Furthermore, it satisfies $W(0) = 0$ and is strictly increasing and concave for any $x > 0$. Appendix D.1 contains a collection of identities involving the Lambert W -function that will prove to be extremely useful in proving propositions 1 to 3.

13. Alternatively, firm size can be defined as a firm's total labor input, but it is analytically more convenient to work with firm-level revenue. This is also done in Arkolakis (2016).

14. As mentioned in Arkolakis (2016), this incorporates a form of creative destruction. Alternatively, new firms can enter the economy with different levels of productivity that are drawn from a distribution with full support. This is relatively straightforward to incorporate and can be done in a similar fashion as Reed (2001) or the exogenous entry case in Luttmer (2007).

A_t denotes the aggregate technology shock. Note that the expected growth and volatility of productivities are *independent* of firm size. Nevertheless, the Kimball structure of demand will induce firm-level growth and volatility to vary with firm size as firms of different sizes respond differently to a given-sized shock.¹⁵

My framework does not feature a fixed cost of production or entry unlike the setup of Luttmer (2007). Instead, I assume that the economy innovates at some exogenous rate ν . Then, the measure of existing firms at each period t is given by $N_t = N \cdot \exp(\nu t)$. Despite the lack of a flow fixed cost of production, the economy does feature exit. The chosen Kimball specification induces demand curves with a *finite* choke price. The characterized optimal price can be shown to be decreasing in productivity z . As a result, there exists some cut-off productivity z^* such that firm-level profits are zero for all $z < \bar{z}^*$. Formally, this is cut-off is given by:

$$\bar{z} \equiv \inf \left\{ z \in \mathbb{R}_+ \mid \left(p^*(z) - \frac{w}{z} \right) \varphi \left(\frac{p^*(z)}{P} \right) = 0 \right\}$$

Under the specification of Klenow and Willis (2016), the choke price $\bar{\phi}$ and cut-off productivity z^* allow for a closed-form solution. These are given by:

$$\begin{aligned} \bar{\phi} &= \exp \left(\frac{1}{\varepsilon} \right) \frac{\theta - 1}{\theta} \\ z_t^* &= \frac{w_t}{P_t} \frac{\theta}{\bar{\phi} \left(\theta - 1 + \varepsilon \ln \left(\frac{\theta - 1}{\bar{\phi} \theta} \right) \right)} \propto \frac{w_t}{P_t} \end{aligned}$$

The latter observation will prove to be useful in constructing a balanced growth path equilibrium as cut-off productivities are known to grow at the rate of the real wage. Lastly, exogenous entry and a finite choke price lead to a substantially more simple problem for the

15. The size-variance relationship can also be rationalized by the fact that large firms simply face smaller sized shocks. However, recent findings by Berger and Vavra (2016) and Decker et al. (2016) favor explanations that are motivated by the responsiveness of firms rather than the size of their shocks.

firm than in Luttmer (2007) or Luttmer (2011) as a firm essentially faces a static problem.¹⁶ Nevertheless, the framework can characterize firm-level moments that vary in size and a stationary firm size distribution in closed form.

5.3.3 *Balanced growth path equilibrium*

I will focus on a balanced growth path equilibrium in which per-capita consumption and the wage rate grow at a constant rate $\kappa > 0$. Along the balanced growth path, the aggregate price level P_t is constant. Recall from the derivation above that the cut-off productivity is proportional to the real wage. This implies that firm-level revenue can be expressed as a function of detrended productivity $z_{t^b,a}/z_t^*$. Additionally, I assume that aggregate technology grows at a constant rate and satisfies $A_t = A \cdot \exp(g_A t)$ as in Arkolakis (2016). Then, it is straightforward to construct a balanced growth path in which $\kappa = g_A + \theta_E$. To do this, it is useful to determine the stationary distribution of $s_{t^b,a} \equiv \ln\left(\frac{z_{t^b,a}}{z_t^*}\right)$. Along the proposed balanced growth path, it must be that:

$$s_{t^b,a} = \bar{s} + (\theta_I - \theta_E)a + \sigma_F \mathcal{B}_{t^b,a}$$

where $\bar{s} \equiv \ln(\bar{z}) + \ln(w/P) + \ln\left(\frac{\theta}{\bar{\phi}(\theta - 1 + \varepsilon \ln(\frac{\theta-1}{\bar{\phi}\theta}))}\right)$. Then, let $\mu = \theta_I - \theta_E$ and it is straightforward to deduce that the natural log of detrended productivity follows a standard Brownian motion with drift:

$$ds_a = \mu da + \sigma_F \mathcal{B}_{t^b,a}$$

In a stationary setting, out-flows due to the stochastic evolution of firm-level productivi-

16. Arkolakis (2016) also dramatically simplifies the firm's problem by specifying exogenous entry. However, productivity cut-offs are determined without the use of fixed costs but through the specification of market penetration costs.

ties must be equal to the rate of exogenous, deterministic entry. This is reflected by the Kolmogorov forward equation in which $\frac{\partial}{\partial t} f(s, t) = 0$:

$$-\mu D_s \left[s \cdot f(s) \right] + \frac{\sigma_F^2}{2} D_{ss} \left[s^2 \cdot f(s) \right] = \nu f(s)$$

for each $s \in (0, \bar{s}) \cup (\bar{s}, +\infty)$. The solution concept is identical to Luttmer (2007). Let the roots ζ^* and ζ be defined as:

$$\zeta^* = \frac{\mu}{\sigma_F^2} + \sqrt{\left(\frac{\mu}{\sigma_F^2}\right)^2 + \frac{\nu}{\sigma_F^2/2}} \quad \text{and} \quad \zeta = -\frac{\mu}{\sigma_F^2} + \sqrt{\left(\frac{\mu}{\sigma_F^2}\right)^2 + \frac{\nu}{\sigma_F^2/2}}$$

Then, the stationary distribution for detrended productivity s is given by:

$$f(s) = \frac{\zeta^* \zeta \exp(-\zeta s)}{\exp(\zeta^* \bar{s})} \min \left\{ \frac{\exp([\zeta + \zeta^*]s) - 1}{\zeta + \zeta^*}, \frac{\exp([\zeta + \zeta^*]\bar{s}) - 1}{\zeta + \zeta^*} \right\}$$

which directly implies that the firm size distribution is characterized by a Pareto right tail equal to ζ . Given the numéraire of $P = 1$, the economy is closed in general equilibrium through budget balance and labor market clearing which pin down the values for C and w . Consistency along the balanced growth path requires $\nu = \eta$. Then, budget balance (i.e. total expenditure equals total revenue) and labor market clearing are given by:

$$C = N \cdot \left(\int_{s \in \mathbb{R}_+} r(s) f(s) ds \right) \tag{BB}$$

$$w \cdot L = \frac{N \cdot C}{|\Omega|} \left(\int_{s \in \mathbb{R}_+} \varphi(p^*(s)) \exp(-s) f(s) ds \right) \tag{LMC}$$

$$\text{where } \varphi(p) = \left[1 + \varepsilon \ln \left(\frac{\theta - 1}{\theta} \frac{1}{p} \right) \right]^{\theta/\varepsilon}$$

5.3.4 Deviations from Gibrat's law

The stationary firm size distribution is characterized by a Pareto right tail as observed in the U.S. data (see Axtell (2001)) since $f(s) \propto \exp(-\zeta s)$ for $s \geq \bar{s}$. In the following, I will show that the framework also predicts firm-level outcomes that are consistent with sections 4.1, 4.2 and 5.1. To do this, it will be convenient to establish the following lemma in which firm size, defined as firm-level revenue, and (log) productivity are related one-to-one:

LEMMA 1. A firm's size is increasing in its log level of detrended productivity.

Proof. See Appendix D.2. □

This property is present in a large body of work in firm dynamics. Nevertheless, this framework is able to *jointly* predict the set of stylized facts presented in sections 4.1, 4.2 and 5.1 along the balanced growth path constructed above. Equation B indicates that firm-level revenue is the product of a deterministic term, i.e. the composite good C , and a stochastic term. Firm-level revenue can simply be expressed as $\text{rev}(s_a) = g(a) \times h(s_a)$ where $g(a) = C_a$ and $h(s_a) = \frac{1}{|\Omega|} \frac{\theta}{\varepsilon} \frac{\exp(-s)}{W[\Theta \cdot \exp(-s)]} (\theta - \varepsilon W[\Theta \cdot \exp(-s)])^{\theta/\varepsilon}$. Applying Itô's lemma to this expression, it is immediate that:

$$\frac{d\text{rev}(s_a)}{\text{rev}(s_a)} = \kappa + \left[\mu \frac{h'(s_a)}{h(s_a)} + \frac{\sigma_F^2}{2} \frac{h''(s_a)}{h(s_a)} \right] da + \sigma_F \frac{h'(s_a)}{h(s_a)} d\mathcal{B}_a$$

Firm-level volatility and growth for a firm with initial size s_0 is defined as the instantaneous variance and expected growth rate conditional on $s = s_0$ respectively. As a result, it is possible to derive that firm-level volatility is declining in firm size. This is formalized in proposition 1.

PROPOSITION 1. Firm-level volatility $\text{vol}(s_0) = \sqrt{V \left(\frac{d\text{rev}(s)}{\text{rev}(s)} \Big|_{s=s_0} \right) / da}$ declines in firm size $\text{rev}(s_0)$ whenever $\theta > \varepsilon > 0$. Furthermore, firm-level volatility is equal to $\sigma_F(\theta - 1)$ for all $s_0 > 0$ whenever $\varepsilon \rightarrow 0$.

Proof. See Appendix D.3. □

The intuition for proposition 1 is simple. Recall that the elasticity of demand is decreasing in the relative quantity consumed of a firm's variety. It is straightforward to derive that relative quantities are positively correlated with productivity. Thus, a firm's elasticity of demand decreases in its size as well. For a given percentage change in firm-level productivity, percentage changes in revenues are then smaller for larger firms. As a result, large firms are less volatile than small firms.

Proposition 1 is only a qualitative statement, so it is unclear whether the generated size-variance relationship is also a declining *power* law. However, figure 5.1 shows that the Kimball specification is flexible enough to generate this. Even though it is difficult to establish this analytically, it is not completely surprising that the Kimball aggregator is able to generate the size-variance relationship as a power law. Recall from before that the price elasticity of a differentiated good is decreasing in its relative market share at the rate ε/θ . Both firm-level revenue and this relative market share are one-to-one with firm-level (log) productivity. This implies that a declining size-variance relationship can be generated by letting ε/θ be fairly small. To see this, consider the extreme cases. Whenever $\varepsilon \rightarrow 0$, price elasticities are constant and there is no differential response to firm-specific productivity shocks between large and small firms. Hence, Gibrat law holds and the size-variance relationship is flat.

On the other hand, a relatively high value for ε/θ implies that price elasticities fall steeply in firm size which would imply that large firms respond so strongly to a shock such that it is almost completely mitigated. In other words, this means that the levels of firm-level volatility are too low for large firms compared to what is observed in the data. This illustrates the strength of the Kimball aggregator as the degree to which firms of different sizes respond to firm-level productivity shocks can be controlled by the ratio ε/θ only.

A declining size-variance relationship will also generate a negative relationship between

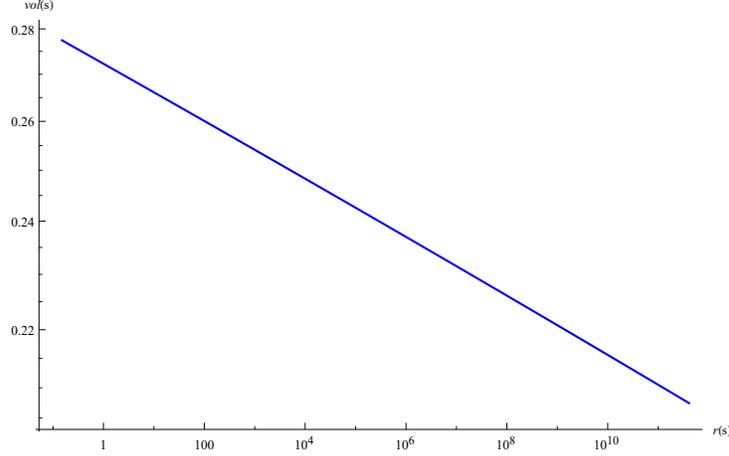


Figure 5.1: Size-variance relationship under the constructed random growth model with Kimball demand whenever $\sigma_F = 0.025$, $\theta = 12$ and $\varepsilon = 0.1$. The vertical (firm-level volatility) and horizontal (firm size) axes are in log scale.

firm-level growth and size. However, the framework also predicts that Gibrat’s law (in terms of growth) approximately holds among large firms. This is formalized in proposition 2.

PROPOSITION 2. Firm-level growth is constant among large firms as expected growth $m(s_0) = \mathbb{E} \left(\frac{d\text{rev}(s)}{\text{rev}(s)} \Big|_{s=s_0} \right) / da$ converges to κ as $s_0 \rightarrow +\infty$. For $\varepsilon \rightarrow 0$, firm-level growth is constant as $m(s_0) = \kappa + \mu(\theta - 1) + \frac{\sigma_F^2}{2}(\theta - 1)^2$ for all $s_0 > 0$.

Proof. See Appendix D.4. □

Proposition 2 highlights another qualitative feature on the relationship between firm-level growth and size that the framework is able to generate. As mentioned by Klette and Kortum (2004), smaller firms grow faster than larger firms, but Gibrat’s law cannot be rejected amongst the largest firms.¹⁷ Lastly, this random growth framework with a Kimball aggregator is able to generate a positive relationship between firm-level markups and size. This is depicted in proposition 3.

17. Many studies find that smaller firms grow faster than larger firms. Several contributions in the literature have documented this particular fact, see the surveys by Caves (1998) and Sutton (1997). This violation of Gibrat’s law is then often cited to emphasize the importance of small firms for economic growth (e.g. Birch (2010) and Neumark, Wall and Zhang (2011)). More recently, Akcigit and Kerr (2015) find evidence that firm-level growth rates are negatively correlated with firm size in their subsample of innovating firms in the LBD. In the following, I will take this stylized fact as stated in Klette and Kortum (2004) as given.

PROPOSITION 3. A firm’s markup (i.e. optimal price over marginal cost) declines in firm size $\text{rev}(s_0)$. Whenever $\varepsilon \rightarrow 0$, markups are constant at $\theta/(\theta - 1)$ for all $s_0 > 0$ and thus do not vary in firm size.

Proof. See Appendix D.5. □

This result is intuitive as larger firms possess, by definition, a larger share of the market and hence have more market power. This is exactly reflected in a firm’s markup. In this specific framework, larger firms set lower prices but their marginal costs are also lower. Proposition 3 then implies that the cost effect dominates the lower price set by a larger firm.¹⁸

While there are models that can account for variable markups, none of these frameworks, as far as I am aware of, can generate the stylized facts from sections 4.1, 4.2 and 5.1 *simultaneously* in one stylized setup.¹⁹ This is potentially important as there are numerous aggregate implications from deviating from Gibrat’s law. First, the framework’s prediction on the declining relationship between firm-level growth and size are consistent with the importance of small firms for economic growth (e.g. Birch (2010) and Neumark et al. (2011)).

Second, as demonstrated in section 4.3, the size-variance relationship has drastic implications on the quantitative importance of the granular channel as the role of granularity is cut by more than half compared to the previous literature. Furthermore, the size-variance relationship can have implications that extend to international trade. di Giovanni and Levchenko (2012) propose a novel mechanism in which trade liberalizations raise aggregate volatility. Trade openings typically only induce the more productive (i.e. larger) firms to enter the

18. Through the lens of the results in Appendix D.6, this result is not surprising. Appendix D.6 characterizes a class of preferences in which markups are increasing in firm-level productivity. A sufficient condition is that the superelasticity of demand $-d\ln \varepsilon_D(p)/d\ln p$ is negative. This is the case with Kimball preferences as the superelasticity is conveniently characterized as $-\varepsilon x^{-\varepsilon/\theta} < 0$ for all relative quantities $x > 0$.

19. Most models in firm dynamics and international trade have abstracted from variable markups, but there is an increasing stream in the international trade literature that emphasizes size-varying markups. This requires the necessity of variable markups which can be found in Bernard et al. (2003), Atkeson and Burstein (2008), Melitz and Ottaviano (2008) and Edmond et al. (2015).

export market and the least productive (i.e. smallest) firms are forced to exit.²⁰ Hence, a trade liberalization gives a more prominent role to the largest firms. Whenever the firm size distribution is close to Zipf’s law, granular forces then increase aggregate volatility as macroeconomic fluctuations are mainly driven by the largest firms in a granular economy. di Giovanni and Levchenko (2012) argue that their results are robust to the size-variance relationship, but the used estimates are based on the values provided by Stanley et al. (1996) and Sutton (2002). I showed in sections 4.1 and 4.2 however that the declining relationship between firm-level volatility and size is substantially more negative in the LBD. As a result, this mechanism under which international trade affects macroeconomic fluctuations can be significantly more limited than conjectured before.

Another application to international trade lies in the fact that markups are size-varying. Pass-through of changes in tariffs induced by trade liberalizations are complete under constant markups. As a result, most of the benchmark models in trade completely abstract from the markup channel as a potential source of gains from trade.²¹ Finally, the existence of size-varying markups can improve our understanding on firms’ responses to industrial policy interventions (see Atkin et al., 2015) or firms’ pass-through policies of international price shocks (e.g. Nakamura and Zerom, 2010).

5.4 Direct evidence on size-varying price elasticities

The framework of section 5.3 provides a microfoundation for the observed deviations from Gibrat’s law and is consistent with a Pareto right tail of the firm size distribution. In this setup, the Kimball aggregator induces demand curves in which price elasticities are decreasing in a firm’s size. However, this is a direct consequence of the imposed assumption of the Kimball aggregator. To strengthen the credibility of this assumption, I will show

20. This mechanism is demonstrated in Melitz (2003) and its implications have empirically been confirmed by Bernard et al. (2009a).

21. The reader is referred to de Loecker et al. (2016) for a more thorough discussion on the markup channel as a potential source of gains from trade.

direct evidence that supports the main assumption of my framework.

Most contributions in the literature are unable to verify size-varying price elasticities directly in the data as it is usually complicated by at least one of the following factors: (1) firm- or plant-level data usually do not contain any information on prices and/or physical quantities and (2) when these are available, it is challenging to find a suitable instrument to overcome the simultaneity bias present in price-quantity regressions.

To overcome all these issues, I use a unique subset of establishments producing physically homogeneous products in the Census of Manufactures (CM) as in Foster, Haltiwanger and Syverson (2008).²² This has two main advantages. First, physical quantities are directly observed. Combined with revenue data, this allows me to construct average prices.²³ A huge advantage of observing physical quantities combined with extensive data on input factors is that I can construct measures of *physical* productivity (TFPQ) as opposed to measures of revenue-based productivity (TFPR). Physical productivity then lends itself as a strong instrument in dealing with the endogeneity problem present in price-quantity regressions. Typically, physical levels of productivity (reflecting physical production costs) are strongly negatively correlated with prices satisfying rank conditions. Furthermore, the exclusion restriction imposes that TFPQ is not correlated with any short-run plant-specific demand shocks.²⁴ A condition that is most likely to hold. As a result, a plant's TFPQ level serves as an ideal instrument to overcome typical simultaneity biases.

Second, by focusing on those products that are considered to be the most physically homogeneous in the manufacturing sector, I avoid any biases due to variation in unobserved product quality. Therefore, this strategy allows me to obtain reliable estimates of price elasticities and how they vary in size. To test this hypothesis, I turn to the following

22. The reader is referred to Appendix A.5 for details on this data set.

23. Revenues and quantities are observed at the yearly level, so constructed unit prices are yearly averages. These can be interpreted as quantity-weighted averages of all prices charged by a plant in a given year as mentioned by Foster, Haltiwanger and Syverson (2008).

24. Note that it is not obvious that the exclusion restriction holds for revenue-based measures of productivity as revenue typically does reflect these types of demand shocks through unit prices.

empirical specification which will be estimated separately for each product:

$$\ln q_{it} = \sigma_0 + \sigma_1 \ln p_{it} + \sigma_2 \ln p_{it} \times \mathbf{d}_{it} + \sigma_3 \mathbf{d}_{it} + \sum_t \alpha_t \text{YEAR}_t + \beta \ln(\text{INCOME}_{m(i)t}) + v_{it}$$

where \mathbf{d}_{it} denotes a dummy that is equal to 1 whenever a plant i 's size (as measured by employee count) in year t is larger than the 75th percentile. I also include a set of year fixed effects to control for economy-wide shifts in the demand curve. In the following, I will mainly focus on one product: ready-mixed concrete. As mentioned by Syverson (2004), ready-mixed concrete is a physically homogeneous output good. Nevertheless, this good is monopolistically competitive in a spatial sense as ready-mixed concrete is subject to high transportation costs. To control for demand shifts at the local level, I also include the average income of the local market m in which plant i is situated denoted by INCOME_{mt} . Local markets are defined according to the BEA's Economic Areas.

A profit-maximizing firm always operates on the elastic part of their demand curve, thus I expect $\sigma_1 < -1$ to hold. Whenever larger firms face lower price elasticities, then it should be the case that $\sigma_2 > 0$ but $\sigma_1 + \sigma_2 < -1$. To deal with the simultaneity bias, I perform a two-stage least squares (2SLS) exercise in which the set of endogenous variables $(\ln p_{it}, \ln p_{it} \times \mathbf{d}_{it})$ is instrumented with $(\ln TFPQ_{it}, \ln TFPQ_{it} \times \mathbf{d}_{it})$. The results of this procedure are displayed in table 5.3.

The first two columns are nearly identical to Foster, Haltiwanger and Syverson (2008) and serve as a validity test only. For concrete, they find estimates of -0.82 (OLS) and -5.93 (IV) which are identical to the values I find. More interestingly, I find that larger firms indeed face lower price elasticities as $\hat{\sigma}_2 > 0$. I also verify whether large firms face lower price elasticities by restricting the sample to either small and middle sized firms or large firms only and imposing the restriction $\sigma_2 = \sigma_3 = 0$. This specification is also consistent with the previous estimates as the estimate $\hat{\sigma}_1$ is significantly more negative for small and middle sized firms than for large firms. Lastly, I repeat the exercise for four other product

Table 5.3: Estimation results of size-varying price elasticities.ⁱ

	OLS	IV	IV - SMALL	IV - LARGE	IV - \mathbf{d}_{it}
CONCRETE					
p_{it}	-0.820 (0.088)	-5.931 (0.359)	-8.047 (0.488)	-4.013 (0.377)	-8.030 (0.489)
$p_{it} \times \mathbf{d}_{it}$					+3.998 (0.625)
Fixed effects					
YEAR	Y	Y	Y	Y	Y
SAMPLE	$\simeq 12000$	$\simeq 12000$	$\simeq 9000$	$\simeq 3000$	$\simeq 12000$

ⁱThe first two columns estimate the restricted regression specification in which $\sigma_2 = \sigma_3 = 0$. This is done with OLS and IV (2SLS) respectively which replicates the results of Foster, Haltiwanger and Syverson (2008). The third and fourth column display the results of the IV regression in which $\sigma_2 = \sigma_3 = 0$ but the sample is restricted to small/medium and large firms respectively. In the last column, the results of the unrestricted IV regression are displayed. The column “IV - \mathbf{d}_{it} ” displays the estimated coefficients when the interaction dummy equals 1 whenever a plant’s size exceeds the 75th percentile. Standard errors are clustered at the plant level. All estimates are statistically significant at the 1 percent level.

categories. The results for boxes, bread, processed ice and gasoline display the same picture qualitatively. The results can be found in Appendix B.3.²⁵

25. I concentrate on ready-mixed concrete as the bulk of the total sample consists of plants producing this particular good. While I also consider other goods, I do not consider all homogeneous goods covered in Foster, Haltiwanger and Syverson (2008). This is because I need a sufficient amount of cross-sectional variation in plant size within a product category to allow for size-varying price elasticities. Details can be found in Appendix A.5.

CHAPTER 6

CONCLUSION

The granular hypothesis postulates that a significant portion of aggregate fluctuations can be attributed to firm-level idiosyncratic shocks. This channel is in particular supposed to be powerful in the presence of a firm size distribution with a tail parameter that is close to one which is consistent with U.S. Census data (e.g. Axtell (2001) and Luttmer (2007)). However, I find that the role of firm-level idiosyncratic shocks to aggregate volatility in the U.S. economy is fairly limited despite its fat-tailed firm size distribution. The contribution of the granular channel as proposed by Gabaix (2011) can at most explain about 16 percent of observed aggregate volatility which is about half what previous studies find.

To come to this conclusion, I construct a novel, comprehensive data set comprised of several sources from the U.S. Census Bureau which allows me to perform a theoretically founded variance decomposition of aggregate sales volatility into a firm-specific and macro-sectoral component. In stark contrast to previous findings, I observe that the component capturing macroeconomic and sectoral forces is more important for aggregate volatility than its firm-specific counterpart. A further decomposition of the firm-specific component indicates that the role of granularity (relative to aggregate volatility) is capped around 16 percent. More interestingly though, the majority of the firm-specific component seems to be generated by forces that generate positive comovements between firms' residual growth rates. While I cannot perfectly identify the exact driving forces behind these patterns, preliminary correlation exercises do seem to be suggestive of input-output linkages which gives additional empirical support for the potential importance of buyer-supplier networks as described in Acemoglu, Carvalho, Ozdaglar and Tahbaz-Selehi (2012).

My results on the role of granular channel vis-à-vis the previous literature can be entirely rationalized by the size-variance relationship. Firm-level volatility falls at a substantially higher rate in size than found before and, hence the prominence of the largest firms in terms of macroeconomic volatility is strongly diminished. The negative log-linear relationship be-

tween firm-level volatility and its size is remarkably robust to a variety of specifications and the empirical results indicate that traditional explanations of the literature cannot explain the size-variance relationship which include narratives on output, input or product diversification, firm learning, industry-specific human capital accumulation and market penetration costs.

Instead, I propose a theory of firm dynamics that incorporates a demand specification with price elasticities that are declining in firm size for which I find direct evidence using product-level data for a select number of industries in the Census of Manufactures. Most importantly, this tractable framework does not only predict “standard” deviations from Gibrat’s law (i.e. size-varying firm-level growth and volatility), but it is also consistent with firm-level markups being increasing in firm size. This finding is directly supported by the empirical findings from section 5.1 using data spanning the whole U.S. manufacturing sector.

The aggregate implications of deviations from Gibrat’s law have usually been confined to first moments: smaller firms grow at a substantially faster rate than large firms and indicates the importance of small firms for economic growth. However, I show that deviations in second moments (through the size-variance relationship) have drastic consequences for granular economies. In particular, the role of the granular channel is cut by more than half whenever the size-variance relationship, as observed in the data, is taken into account. Additionally as mentioned by Axtell (2001), this result places “important limits on models of firm dynamics”.

Finally, I find that markups are increasing in size which could have important implications on several distinct strands of the literature. Size-varying markups give rise to the markup channel as a potential source of gains from trade. Moreover, my results are strongly indicative of size-varying pass-through which could have strong implications for our understanding on firms’ responses to industrial policy interventions or firms’ pass-through policies of exchange rate shocks. As a result, it is important for future macroeconomic models of firm dynamics to consider deviations from the strong version of Gibrat’s law as opposed to the majority of

current and past work.

REFERENCES

- Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tahbaz-Selehi**, “The Network Origins of Aggregate Fluctuations,” *Econometrica*, 2012, 80 (5), 1977–2016.
- Akerberg, D., K. Caves, and G. Frazer**, “Identification Properties of Recent Production Function Estimators,” *Econometrica*, 2015, 83 (6), 2411 – 2451.
- Akcigit, U. and W. R. Kerr**, “Growth through Heterogeneous Innovations,” *Working Paper*, 2015.
- Anderson, J.E. and E. van Wincoop**, “Trade Costs,” *Journal of Economic Literature*, 2004, 42 (3), 691–751.
- Arkolakis, C.**, “A Unified Theory of Firm Selection and Growth,” *Quarterly Journal of Economics*, 2016, *forthcoming*.
- , **A. Costinot, D. Donaldson, and A. Rodriguez-Clare**, “The Elusive Pro-Competitive Effects of Trade,” *Working Paper*, 2015.
- , **T. Papageorgiou, and O. A. Timoshenko**, “Firm Learning and Growth,” *Working Paper*, 2015.
- Atalay, E.**, “Materials Prices and Productivity,” *Journal of the European Economic Association*, 2014, 12 (3), 575 – 611.
- , “How Important Are Sectoral Shocks?,” *Working Paper*, 2015.
- Atkeson, A. and A. Burstein**, “Pricing-to-Market, Trade Costs and International Relative Prices,” *American Economic Review*, 2008, 98 (5), 1998 – 2031.
- Atkin, D., A. Chaudhry, S. Chaudhry, A. Khandelwal, and E. Verhoogen**, “Markup and Cost Dispersion across Firms: Direct Evidence from Producer Surveys in Pakistan,” *American Economic Review: Papers and Proceedings*, 2015, 105 (5), 537 – 544.
- Axtell, R.L.**, “Zipf Distribution of U.S. Firm Sizes,” *Science*, 2001, 293 (5536), 1818–1820.
- Bailey, M., C. Hulten, and D. Campbell**, “Productivity Dynamics in Manufacturing Plants,” *Brookings Papers on Economic Activity: Microeconomics*, 1992, pp. 187–267.
- Barlevy, G.**, “The Cost of Business Cycles and the Benefits of Stabilization,” *Federal Reserve Bank of Chicago Economic Perspectives*, 2005, pp. 32–49.
- Basu, B.**, “Intermediate Goods and Business Cycles: Implications for Productivity and Welfare,” *American Economic Review*, 1995, 85 (3), 512–531.
- Berger, D. and J. Vavra**, “Shocks vs Responsiveness: What Drives Time-Varying Dispersion?,” *Working Paper*, 2016.
- Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum**, “Plants and Productivity in International Trade,” *American Economic Review*, 2003, 93 (4), 1268 – 1290.

- Bernard, A.B., J.B. Jensen, and P.K. Schott**, “Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. that Trade Goods,” *NBER Working Paper*, 2009.
- , – , **S.J. Redding, and P.K. Schott**, “The Margins of US Trade,” *American Economic Review*, 2009, *99* (2), 487–493.
- Birch, D.L.**, *Job Creation in America: How Our Smallest Companies Put the Most People to Work*, New York: Free Press, 2010.
- Caballero, R. J., E. M. R. A. Engel, and J. Haltiwanger**, “Aggregate Employment Dynamics: Building from Microeconomic Evidence,” *American Economic Review*, 1997, *87* (1), 115 – 137.
- Carvalho, V. and B. Grassi**, “Large Firm Dynamics and the Business Cycle,” *Working Paper*, 2015.
- **and X. Gabaix**, “The Great Diversification and its Undoing,” *American Economic Review*, 2013, *103* (5), 1697–1727.
- Caves, R.E.**, “Industrial Organization and New Findings on the Turnover and Mobility of Firms,” *Journal of Economic Literature*, 1998, *36* (4), 1947 – 1982.
- Clementi, G. L. and D. Palazzo**, “Entry, Exit, Firm Dynamics and Aggregate Fluctuations,” *American Economic Journal: Macroeconomics*, 2016, *8* (3), 1 – 41.
- Davis, S.J., J. Haltiwanger, R. Jarmin, and J. Miranda**, “Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms,” *NBER Macroeconomics Annual 2006*, 2007, *2*.
- , **J.C. Haltiwanger, and S. Schuh**, “Job Creation and Destruction,” *MIT Press Books*, 1996, *1*.
- de Loecker, J. and F. Warzynski**, “Markups and Firm-Level Export Status,” *American Economic Review*, 2012, *102* (6), 2437 – 2471.
- , **P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik**, “Prices, Markups and Trade Reform,” *Econometrica*, 2016, *84* (2), 445 – 510.
- Decker, R., J. Haltiwanger, R. S. Jarmin, and J. Miranda**, “Changing business dynamism and productivity: Shocks vs. responsiveness,” *Working Paper*, 2016.
- di Giovanni, J., A. Levchenko, and I. Mejean**, “Firms, Destinations and Aggregate Fluctuations,” *Econometrica*, 2014, *82* (4), 1303–1340.
- **and –**, “Country Size, International Trade, and Aggregate Fluctuations in Granular Economies,” *Journal of Political Economy*, 2012, *120* (6), 1083–1132.
- Dixit, A. K. and J. E Stiglitz**, “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review*, 1977, *67* (3), 297 – 308.

- Eberhardt, M. and C. Helmers**, “Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators,” *Working Paper*, 2016.
- Edmond, C., V. Midrigan, and D. Y. Xu**, “Competition, Markups, and the Gains from International Trade,” *American Economic Review*, 2015, *105* (10), 3183 – 3221.
- Feenstra, R. C.**, “Restoring the Product Variety and Pro-Competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity,” *NBER Working Paper*, 2014.
- Foerster, A.T., P.-D.G. Sarte, and M.W. Watson**, “Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production,” *Journal of Political Economy*, 2011, *119* (1), 1–38.
- Fort, T. and S. D. Klimek**, “The Effects of Industry Classification Changes on US Employment Composition,” *Working Paper*, 2016.
- , **J. Haltiwanger, R. Jarmin, and J. Miranda**, “How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size,” *IMF Economic Review*, 2013, *61* (3), 520 – 559.
- Foster, L., J. Haltiwanger, and C. Syverson**, “Selection on Productivity or Profitability?,” *American Economic Review*, 2008, *98* (1), 394 – 425.
- Gabaix, X.**, “The Granular Origins of Aggregate Fluctuations,” *Econometrica*, 2011, *79* (3), 733–772.
- Gandhi, A., S. Navarro, and D. Rivers**, “On the Identification of Production Functions: How Heterogeneous is Productivity?,” *Working Paper*, 2013.
- Gaubert, C. and O. Itskhoki**, “Granular Comparative Advantage,” *Working Paper*, 2015.
- Griliches, Z. and J. Mairesse**, *Production Functions: The Search for Identification*, New York: Cambridge University Press, 1996.
- Hall, B.H.**, “The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector,” *Journal of Industrial Economics*, 1987, *35* (4), 583 – 606.
- Hall, R. E.**, “Market Structure and Macroeconomic Fluctuations,” *Brookings Papers on Economic Activity*, 1986, *2*, 285 – 322.
- Haltiwanger, J., R. S. Jarmin, R. Kulick, and J. Miranda**, “High Growth Young Firms: Contribution to Job, Output and Productivity Growth,” *Working Paper*, 2016.
- Hymer, S. and P. Pashigian**, “Firm Size and Rate of Growth,” *Journal of Political Economy*, 1962, *70*, 556 – 569.
- Jarmin, R. and J. Miranda**, “The Longitudinal Business Database,” *Census Bureau Working Paper*, 2002.

- Jovanovic, B.**, “Selection and the Evolution of Industry,” *Econometrica*, 1982, 50 (3), 649 – 670.
- Kehrig, M.**, “The Cyclical Nature of the Productivity Distribution,” *Working Paper*, 2015.
- Kimball, M.**, “The Quantitative Analytics of the Basic Neomonetarist Model,” *Journal of Money, Credit and Banking*, 1995, 27 (4), 1241 – 1277.
- Klenow, P. and J. Willis**, “Real Rigidities and Nominal Price Changes,” *Economica*, 2016, 83 (331), 443 – 472.
- Klette, T.J. and S.S. Kortum**, “Innovating Firms and Aggregate Innovation,” *Journal of Political Economy*, 2004, 112 (5), 986–1018.
- **and Z. Griliches**, “The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous,” *Journal of Applied Econometrics*, 1996, 11 (4), 343 – 361.
- Koren, M. and S. Tenreyro**, “Sectoral and National Aggregate Disturbances to Industrial Output in Seven European Countries,” *Quarterly Journal of Economics*, 2007, 122 (1), 243 – 287.
- **and –**, “Technological Diversification,” *American Economic Review*, 2013, 103 (1), 378–414.
- Krugman, P.**, “Scale Economies, Product Differentiation and the Pattern of Trade,” *American Economic Review*, 1980, 70 (5), 950 – 959.
- Lamorgese, A.R., A. Linarello, and F. Warzynski**, “Free Trade Agreements and Firm-Product Markups in Chilean Manufacturing,” *Working Paper*, 2014.
- Lentz, R. and D. T. Mortensen**, “An Empirical Model of Growth through Product Innovation,” *Econometrica*, 2008, 76 (6), 1317 – 1373.
- Levinsohn, J.A. and A. Petrin**, “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, 2003, 70 (2), 317 – 340.
- Lu, Y., Z. Tao, and L. Yu**, “The Markup Effect of Agglomeration,” *Working Paper*, 2014.
- Lucas, R. E.**, “Understanding Business Cycles,” in “Carnegie-Rochester Conference Series on Public Policy,” Vol. 5 Elsevier 1977, pp. 7–29.
- Luttmer, E.G.J.**, “Selection, Growth, and the Size Distribution of Firms,” *Quarterly Journal of Economics*, 2007, 122 (3), 1103–1144.
- , “On the Mechanics of Firm Growth,” *Review of Economic Studies*, 2011, 78.
- , “Technology Diffusion and Growth,” *Journal of Economic Theory*, 2012, 147, 602 – 622.
- Magerman, G., K. de Bruyne, E. Dhyne, and J. van Hove**, “Heterogeneous Firms and the Micro Origins of Aggregate Fluctuations,” *KU Leuven Working Paper*, 2015.

- Melitz, M. J. and G. I. P. Ottaviano**, “Market Size, Trade and Productivity,” *Review of Economic Studies*, 2008, 75 (1), 295 – 316.
- Melitz, M.J.**, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 2003, 71 (6), 1695–1725.
- Meyer, J. R. and E. Kuh**, *The Investment Decision*, Cambridge, Massachusetts: Harvard University Press, 1957.
- Moreira, S.**, “Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles,” *Working Paper*, 2016.
- Moscarini, G. and F. Postel-Vinay**, “The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment,” *American Economic Review*, 2012, 102 (6), 2509 – 2539.
- Nakamura, E. and D. Zerom**, “Accounting for Incomplete Pass-Through,” *Review of Economic Studies*, 2010, 77 (3), 1192 – 1230.
- Neumark, D., B. Wall, and J. Zhang**, “Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series,” *Review of Economics and Statistics*, 2011, 93 (1), 16 – 29.
- Reed, W. J.**, “The Pareto, Zipf and Other Power Laws,” *Economic Letters*, 2001, 74 (1), 15 – 19.
- Rossi-Hansberg, E. and M.L.J. Wright**, “Establishment Size Dynamics in the Aggregate Economy,” *American Economic Review*, 2007, 97 (5), 1639 – 1666.
- Seker, M.**, “A Structural Model of Firm and Industry Evolution: Evidence from Chile,” *Journal of Economic Dynamics and Control*, 2012, 36.
- Stanley, M. H. R., N. H. Amaral, S. V. Buldyrev, S. Havlin, H. Leschorn, P. Maass, M. A. Salinger, and H. E. Stanley**, “Scaling Behavior in the Growth of Companies,” *Science*, 1996, 379 (29), 804 – 806.
- Stella, A.**, “Firm Dynamics and the Origins of Aggregate Fluctuations,” *Journal of Economic Dynamics and Control*, 2015, 55, 71 – 88.
- Stockman, A.C.**, “Sectoral and National Aggregate Disturbances to Industrial Output in Seven European Countries,” *Journal of Monetary Economics*, 1988, 21 (2-3), 387 – 409.
- Sutton, J.**, “Gibrat’s Legacy,” *Journal of Economic Literature*, 1997, 1.
- , “The variance of firm growth rates: the ‘scaling’ puzzle,” *Physica A*, 2002, 312 (3), 577 – 590.
- Syverson, C.**, “Market Structure and Productivity: A Concrete Example,” *Journal of Political Economy*, 2004, 112 (6), 1181 – 1222.

White, T. K., “Recovering the Item-level Edit and Imputation Flags in the 1977 - 1997 Censuses of Manufactures,” *CES Working Paper 14-37*, 2014.

Wyart, M. and J.-P. Bouchaud, “Statistical Models for Company Growth,” *Physica A*, 2003, *326* (1-2), 241 – 255.

APPENDIX A

DATA FROM U.S. CENSUS BUREAU

A.1 Longitudinal Business Database (LBD)

The Longitudinal Business Database (LBD) is an establishment-level data set covering employment statistics for nearly all sectors of the economy and includes all geographic areas over the period of 1976 - 2011. Its underlying source is the Standard Statistical Establishment List (SSEL) or Business Register (BR) of U.S. businesses with paid employees. The LBD is unique in the sense that it covers the universe of the U.S. economy and contains firm identifiers at the establishment-level. As a result, each establishment can be connected to some parent firm and thus, it is straightforward to aggregate statistics to the firm-level. An establishment belongs to a particular firm based on operational control. I follow Davis et al. (2007) in the sense that an establishment's statistics are included in the parent firm's activity whenever this parent firm majority owns the establishment. While the LBD covers nearly all the industries in the U.S. economy, there are some exceptions. Roughly speaking, the LBD covers the non-farm private economy with some coverage over government-owned or operated entities.¹

The main variables in the LBD are employment (`emp`), payroll (`pay`), year of entry (`firstyear`) and exit (`lastyear`), 6-digit level SIC code (`sic`), establishment age (`age`), type of operation (`toc`), legal form of organization (`lfo`), 5-digit Federal Information Processing Standard (FIPS) county codes (`state` and `county`), 6-digit zipcode (`zip`), establishment identifier (`lbdnum`) and a firm-level identifier (`firmed`). Industry codes are not consistent over time as these are constructed from two SIC regimes (i.e. SIC72 and SIC87) and the covered period contains the transition from SIC

1. This is described in detail by Jarmin and Miranda (2002). Out of scope industries include Agriculture, Forestry and Fishing (SIC Division A), railroads (SIC 40), U.S. Postal Service (SIC 43), Certificated Passenger Air Carriers (part of SIC 4512), Elementary and Secondary Schools (SIC 821), Colleges and Universities (SIC 822), Labor Organizations (SIC 863), Political Organizations (SIC 865), Religious Organizations (SIC 866) and Public Administration (SIC Division J). In general, government-owned and/or operated entities (SIC Division J) are out of scope. The only exceptions are Wholesale Distributors of Beer, Wine and Distilled Alcoholic Beverages (SIC 518), Liquor Stores (SIC 5912), Central Reserve Depository Institutions (SIC 601), Federal and Federally-sponsored Non-depository Institutions (SIC 611) and Hospitals (SIC 806). Despite all of these exceptions, the LBD tracks aggregate employment statistics (relatively to those provided by the County Business Patterns (CBP)) extremely well, see figures 5 and 6 in Jarmin and Miranda (2002).

to NAICS codes and the NAICS 2002 and 2007 regimes. As a result, I follow Fort and Klimek (2016) who developed a methodology to classify all establishments to a consistent 6-digit NAICS industry classification system from 2002 (`fk_naics02`). To determine the industry and location at the *firm* level, I follow Moreira (2016) by defining the main industry and location of the firm as that industry group and location with the largest share of employment within the firm. In case of a tie, the industry and county code mode across all establishments within the firm will function as the industry and location of the firm. Whenever this mode is not unique, that industry and county code with the smallest value is arbitrarily chosen instead.

The LBD covers only those firms with at least one employee *on payroll* over their life-cycle.² Employment is defined as employees subject to U.S. payroll taxes which is deduced from records by the Internal Revenue Service (IRS) available in the BR. Also, the LBD is measured yearly: the employment statistics reflect those numbers as measured on March 12 of each calendar year.

A.2 Standard Statistical Establishment List (SSEL)

The Standard Statistical Establishment List (SSEL) or Business Register (BR) is an establishment-level data set that covers all domestic businesses (with the exception of private households and governments) and organizational units of multi-establishment businesses. Unlike the LBD, it also provides coverage of non-employer businesses. The BR is a continuously updated database that consists of information from the Census Bureau and the Internal Revenue Service (IRS). It contains name and address, industry, employment and payroll data based on payroll tax records and serves as the underlying source of the LBD. More importantly, it contains receipts data based on income tax records from the IRS. These variables on receipts are the main sources to construct “revenues” which is a broad category capturing value of shipments, sales, receipts and/or revenue (in dollars). Data on receipts are not always available for all establishments. To circumvent this problem, the IRS asks every establishment to report their receipts data from the previous (two) year(s). As a result, missing information on receipts in year t can then be deduced from year $t + 1$ and/or $t + 2$.

2. Under the definition of the Census Bureau, paid employment consists of both full- and part-time employees, including salaried officers and executives of corporations. Employees on sick leave, holidays and vacations are included whereas proprietors and partners of unincorporated businesses are not.

Unfortunately, this procedure is only possible until 2001. After 2001, the Census Bureau constructs its own measure of receipts which can then be used to construct revenues.

A.3 Longitudinal Firm Trade Transactions Database (LFTTD)

The Longitudinal Firm Trade Transactions Database (LFTTD) is an international transaction-level data set that covers the universe of U.S. firms that make them. Thus, it captures all the exporting activity by U.S. firms and import transactions to U.S. importers. I am only utilizing export activities in my quantitative exercises, so I will restrict my attention to exports only in this section. The underlying sources of the LFTTD are comprised of the Foreign Trade Exports (FT-EXP) data and the LBD. The latter is described in more detail under section A.1 of the appendix.

The FT-EXP contains transactions on the exports of the universe of U.S. firms that engage in merchandise export to some foreign destination. This data containing the universe of trade transactions data is collected by U.S. Customs and Border Protection (CBP) during the export process. Every U.S. firm that engages in some exporting activity valued at more than 2500 dollars is obliged to fill out the Shipper’s Export Declaration (SED) form which contains the value (`value`) and quantity (`qty1`, `qty2`) of the exported good(s), transaction date (`date`), transport mode (`mot`), country of origin (`sourceid`) or destination (`country`, `isoctry`), 10-digit Harmonized System (HS) product codes (`hs`) and the name (`name1`, `name2`) and tax payer ID (`ein`) of the exporting establishment in question.³ Furthermore, information is collected on whether the transaction takes place at “arm’s length” or between two “related” parties. Export partners are defined to be “related” whenever either party directly (or indirectly) owns at least 10 percent of the other party.

The LFTTD is constructed by linking the FT-EXP with the LBD. This is possible through the use of the SSEL (see section A.2 of the appendix) as the tax payer ID (EIN) available in the FT-EXP is sufficient to link it to the SSEL. In turn, the SSEL and LBD are linked through a common record number (`recnum`). As a result, it is possible to construct a database at the firm-

3. The U.S. Census Bureau imputes a total value for transactions valued lower than 2500 dollars but the reliability of this imputed number is unknown. Hence, I exclude these imputed records from my analysis following Bernard et al. (2009a).

export transaction level. There is a caveat to this procedure as exports to Canada are not collected through the SED form. U.S. institutions use data on Canadian imports from U.S. firms to infer U.S. exports to Canada. This is done to reduce the reporting burden for U.S. and Canadian firms in accordance with a 1987 Memorandum of Understanding signed by the U.S. Census Bureau, U.S. CBP, Canadian customs and Statistics Canada. U.S. exports to Canada are then linked to U.S. firms through the business name field available in the SSEL. For more details on this “name matching” procedure, see the appendix of Bernard et al. (2009a).

Aggregation up to the firm-*destination* level is trivial as each transaction contains the destination country. While exporting numbers are available from 1992 - 2011, I only use the LFTTD from 1994 to 2011 as reliable measures of firm-level revenues are only available from 1994 onwards.

A.4 Annual Survey of Manufactures/Census of Manufactures (ASM/CM)

The Annual Survey of Manufactures (ASM) and Census of Manufactures (CM) are establishment-level data sets that cover the U.S. manufacturing sector (SIC2 20 - 39; NAICS2 31 - 33). Establishments are defined as a distinct physical location of a manufacturing firm in which the main activity is production. As a result, purely administrative establishments are excluded. An establishment is identified by its Permanent Plant Number (*ppn*) which is time-invariant and does not change in case of ownership change or temporary shutdown. The ASM/CM contain a rich set of variables, however I will only use a limited subset of them. In the following, I will discuss what adjustments I make to the raw variables of need. For consistency, all values will be deflated to a common 1997 dollars basis.

Real measure of goods produced. The ASM/CM neither provides physical quantities nor prices for a large set of plants. As a result, I am forced to construct a proxy for real production with industrial price deflators from the NBER-CES Manufacturing database. Fortunately, the ASM/CM does contain the total value of shipments (*tv*s) and detailed information on inventories. As a result, it is possible to construct a proxy for real *production* by adjusting the total value of shipments with the nominal values of finished goods and work-in-progress inventories and convert

them with the appropriate deflators.

I have data on the nominal values of beginning- and end-of-year finished goods (**fib** and **fie**) and work-in-progress (**wib** and **wie**) inventories. Then, real measures for each of these inventories can be obtained by using the appropriate deflators:

$$\text{invf}_{p,t} = \frac{\text{fie}_{p,t} - \text{fib}_{p,t}}{\text{pifi}_{i(p),t}} + \frac{\text{wie}_{p,t} - \text{wib}_{p,t}}{\text{piwi}_{i(p),t}}$$

A real measure of goods produced is then obtained by deflating the total value of shipments and adjusting them for inventories:

$$\begin{aligned} \mathbf{q}_{p,t} &= \frac{\text{tvs}_{p,t}}{\text{piship}_{i(p),t}} + \text{invf}_{p,t} \\ &= \frac{\text{tvs}_{p,t}}{\text{piship}_{i(p),t}} + \frac{\text{fie}_{p,t} - \text{fib}_{p,t}}{\text{pifi}_{i(p),t}} + \frac{\text{wie}_{p,t} - \text{wib}_{p,t}}{\text{piwi}_{i(p),t}} \end{aligned}$$

As mentioned by Kehrig (2015), deflators for inventories are unfortunately not publicly available; not even at the industry level. As a result, I proxy inventory price deflators with the 4-digit SIC industry-level shipment price deflator **piship** from the NBER-CES Manufacturing database. This results in:

$$\mathbf{q}_{p,t} = \frac{\text{tvs}_{p,t}}{\text{piship}_{i(p),t}} + \frac{\text{fie}_{p,t} - \text{fib}_{p,t}}{\text{piship}_{i(p),t}} + \frac{\text{wie}_{p,t} - \text{wib}_{p,t}}{\text{piship}_{i(p),t}}$$

A potentially important factor is the treatment of resales (**cr**). Kehrig (2015) classifies resales as finished goods rather than material inputs as, by definition, resales are goods that are bought and immediately resold by the plant. As a result, resales do not enter the production process at any point in time. Thus, an alternative measure for real output is defined as follows:

$$\mathbf{q}_{p,t}^k = \frac{\text{tvs}_{p,t} - \text{cr}_{p,t}}{\text{piship}_{i(p),t}} + \frac{\text{fie}_{p,t} - \text{fib}_{p,t}}{\text{piship}_{i(p),t}} + \frac{\text{wie}_{p,t} - \text{wib}_{p,t}}{\text{piship}_{i(p),t}}$$

Real value added. Value added (**va**) is defined as the contribution of labor and capital to gross

output. *Nominal* values are directly available from the ASM/CM. For a consistent comparison across years, I deflate nominal value added by using industry-level chain-type price indices for value added (**piva**) which can be found in the BEA’s GDP-by-Industry Value Added file. These are available for the manufacturing sector at roughly the 3-digit NAICS level.⁴ Then, real value added becomes:

$$\mathbf{rva}_{p,t} = \frac{\mathbf{va}_{p,t}}{\mathbf{piva}_{i(p),t}}$$

Labor. Labor is measured as the total number of worker hours. However, I only observe the number of hours by *production* workers (**ph**). As a result, I follow the adjustment method by Bailey, Hulten and Campbell (1992). Hours by production workers are multiplied by the ratio of total payroll (**sw**) over payroll to production workers (**ww**) whenever these are both strictly positive. Otherwise, total hours are just equal to production hours. Therefore, labor is constructed as:

$$\mathbf{l}_{p,t} = \begin{cases} \mathbf{ph}_{p,t} \times \frac{\mathbf{sw}_{p,t}}{\mathbf{ww}_{p,t}} & \text{if } \mathbf{sw}_{p,t}, \mathbf{ww}_{p,t} > 0, \\ \mathbf{ph}_{p,t} & \text{otherwise.} \end{cases}$$

An alternative measure for labor is a plant’s deflated wage bill. I deflate total payroll (**sw**) with the shipment deflator **pishop** from the NBER-CES Manufacturing database. Thus, I obtain:

$$\mathbf{l}_{p,t} = \frac{\mathbf{sw}_{p,t}}{\mathbf{pishop}_{i(p),t}}$$

Capital. The construction of a measure of capital with the manufacturing data is complicated by several factors. First, the value of capital stocks are only reported for a limited amount of years. Second, depreciation and return rates are not available at the plant level. Lastly, plants only report

4. The BEA constructs industry groups within the manufacturing sector at the 3-digit level. However, there are some exceptions as it groups industries with NAICS codes 311 and 312 into “Food and beverage and tobacco products”, 313 and 314 into “Textile mills and textile product mills” and 315 and 316 into “Apparel and leather and allied products”. As a result, I need to deflate value added for these industry groups with the appropriate price index for the industry *group* as defined by the BEA.

the book value of their assets rather than market values. As a result, I will deal with these issues as follows.

First, capital investment *expenditures* are reported annually. These can be denominated in 1997 dollars by using deflators from the NBER-CES Manufacturing database. These deflators are available at the 4-digit SIC level. Consequently, I can construct capital flows with the perpetual inventory method once capital stocks are initialized. This requires some assumptions on the initialization of capital though, but this will become clear below. Second, the perpetual inventory method requires depreciation and rental rates which are not available at the plant level. Thus, I will use data from the BLS Capital Tables which is reported at the 3-digit NAICS level. As a result, I will approximate depreciation and rental rates by using 3-digit NAICS industry-level analogues. Lastly, I transform book values into constant 1997 dollars market values by using data from the BEA's Fixed Asset Tables which is also reported at the 3-digit NAICS level. All industry-level deflators from the NBER-CES Manufacturing, BEA and BLS databases are normalized to a common 1997 basis (1997 = 1.0).

The ASM/CM contains a complete breakdown of beginning- and end-of-year total assets into structures (buildings) and equipment (machinery) for the period of 1976 - 1985, 1987 and 1992. These reported values are book values. Thus, I transform these to market values denominated in constant 1997 dollars by multiplying them with the ratio of current (*cke*_q and *ckst*) and historical (*hke*_q and *hkst*) cost of capital stocks found in the BEA's Fixed Asset Tables. Current cost of capital stocks are already deflated to 1997 dollars using the investment price deflators (*piinve* and *piinvs*) from the NBER-CES Manufacturing database. Historical costs are denominated in period *t* dollars. As a result, the ratio of current and historical costs are in $\frac{1997\$}{\text{year } t\$}$. This implies that capital stocks for equipment and structures are:

$$\begin{aligned} \text{ksteq}_{p,t} &= \text{mae}_{p,t} \left(\frac{\text{cke}_{q_{i(p),t}}}{\text{hke}_{q_{i(p),t}}} \right) \\ \text{kstst}_{p,t} &= \text{bae}_{p,t} \left(\frac{\text{ckst}_{i(p),t}}{\text{hkst}_{i(p),t}} \right) \end{aligned}$$

The breakdown of end-of-year total assets (*tae*) into structures and equipment is not available after 1987 (with the exception of 1992). However, capital expenditures on machines (*cme*) and buildings

(cbe) are available annually. As a result, I impute the capital stocks of a plant that entered the sample in or before 1987 with a *forward* perpetual inventory method:

$$\begin{aligned} \text{ksteq}_{p,t} &= (1 - \text{dprkeq}_{i(p),t})\text{ksteq}_{p,t-1} + \frac{\text{cme}_{p,t}}{\text{piinve}_{i(p),t}} \\ \text{kstst}_{p,t} &= (1 - \text{dprkst}_{i(p),t})\text{kstst}_{p,t-1} + \frac{\text{cbe}_{p,t}}{\text{piinvs}_{i(p),t}} \end{aligned}$$

where $\text{dprkeq}_{i(p),t}$ and $\text{dprkst}_{i(p),t}$ are 3-digit NAICS industry-level depreciation rates taken from the BLS. The nature of the data makes initialization of capital stocks for plants entering after 1987 non-trivial. Even though I do not observe end-of-year assets separately for structures and equipment after 1987, I do observe a plant's end-of-year *total* assets in Census years (i.e. 1992, 1997, 2002 and 2007). The breakdown between equipment and structures is then determined by using the fraction of equipment capital at the 4-digit SIC industry level observed in the NBER-CES Manufacturing database. Initialization occurs in the most recent Census year in which the plant is still active. This implies:

$$\begin{aligned} \text{ksteq}_{p,t} &= \left(\frac{\text{equip}_{i(p),t}}{\text{equip}_{i(p),t} + \text{plant}_{i(p),t}} \right) \text{tae}_{p,t} \left(\frac{\text{cke}_{i(p),t}}{\text{hke}_{i(p),t}} \right) \\ \text{kstst}_{p,t} &= \left(\frac{\text{plant}_{i(p),t}}{\text{equip}_{i(p),t} + \text{plant}_{i(p),t}} \right) \text{tae}_{p,t} \left(\frac{\text{ckst}_{i(p),t}}{\text{hkst}_{i(p),t}} \right) \end{aligned}$$

where, with some abuse of notation, t denotes the most recent Census year after plant p 's entry year in which it is still active.⁵ Then, capital stocks are once more imputed using forward and backward perpetual inventory methods.

Lastly, I need to construct capital rental rates for equipment and structures. To do this, I turn to the BLS Capital Tables. The rental rate for each type of capital is obtained by dividing corporate capital income by the stock of productive capital. The latter is the product of capital stock and the ratio of capital inputs to productive stock. The product of capital stock is deflated to 1997 dollars using piinve or piinvs . Thus, rental rates are in units of $\frac{\text{year } t\$}{1997\$}$.

5. If the most recent Census year is 1992, then I initialize $\text{ksteq}_{p,t} = \text{mae}_{p,t} \left(\frac{\text{cke}_{i(p),t}}{\text{hke}_{i(p),t}} \right)$ and $\text{kstst}_{p,t} = \text{bae}_{p,t} \left(\frac{\text{ckst}_{i(p),t}}{\text{hkst}_{i(p),t}} \right)$ as the breakdown of total assets is still available in 1992.

Intermediate inputs (materials). The value of materials is simply the sum of expenditures on materials and parts, resales and contract work.⁶

Thus, I get:

$$\mathbf{vm}_{p,t} = \mathbf{cp}_{p,t} + \mathbf{cr}_{p,t} + \mathbf{cw}_{p,t}$$

A measure for real material input is then obtained by deflating the previous value with the 4-digit SIC level deflator \mathbf{pimat} from the NBER-CES Manufacturing database. This results in:

$$m_{p,t} = \frac{\mathbf{vm}_{p,t}}{\mathbf{pimat}_{i(p),t}}$$

Energy. Total expenditure on energy is the sum of expenditure on fuels and electricity. To obtain a real measure of energy input, I deflate it by the 4-digit SIC level deflator \mathbf{pien} from the NBER-CES Manufacturing database. In total, this implies:

$$\begin{aligned} \mathbf{ve}_{p,t} &= \mathbf{cf}_{p,t} + \mathbf{ee}_{p,t} \\ e_{p,t} &= \frac{\mathbf{ve}_{p,t}}{\mathbf{pien}_{i(p),t}} \end{aligned}$$

Industry-level cost shares. To construct measures of TFP (or Solow residuals), I need the cost shares of all inputs. These will be calculated at the industry level to minimize measurement error at the plant level. To achieve this, I use data from the NBER-CES Manufacturing database and BLS Capital Tables:

6. Even though this is the conventional definition of material inputs, Kehrig (2015) argues that resales should not be considered as material inputs as they never enter the production process. As a result, Kehrig (2015) emphasizes a definition for the value of intermediate inputs that also includes material inventories. For robustness, I also consider his definition which is $\mathbf{vm}_{p,t}^k$ s.t. $\mathbf{mie}_{p,t} = \mathbf{mib}_{p,t} + \mathbf{cp}_{p,t} + \mathbf{cw}_{p,t} - \mathbf{vm}_{p,t}^k$. Obviously, I define the real input measure of materials accordingly as $\mathbf{m}_{p,t}^k = \mathbf{vm}_{p,t}^k / \mathbf{pimat}_{i(p),t}$

$$\begin{aligned}
al_{i,t} &= \frac{\text{pay}_{i,t}}{\text{tc}_{i,t}} \\
ake_{i,t} &= \frac{\text{eqrkl}_{i,t} \times \text{equip}_{i,t}}{\text{tc}_{i,t}} \\
aks_{i,t} &= \frac{\text{strkl}_{i,t} \times \text{plant}_{i,t}}{\text{tc}_{i,t}} \\
am_{i,t} &= \frac{\text{matcost}_{i,t} - \text{energy}_{i,t}}{\text{tc}_{i,t}} \\
ae_{i,t} &= \frac{\text{energy}_{i,t}}{\text{tc}_{i,t}}
\end{aligned}$$

where $\text{tc}_{i,t} = \text{pay}_{i,t} + \text{strkl}_{i,t} \times \text{plant}_{i,t} + \text{eqrkl}_{i,t} \times \text{equip}_{i,t} + \text{matcost}_{i,t}$. Note that the NBER-CES Manufacturing database includes energy expenditures in the expenditures for material inputs. However, this is not problematic as expenditures for energy are reported separately. Lastly, I resort to division-based cost shares as even cost shares at the industry level are considered to be noisy. Thus, I obtain:

$$\text{dax}_{i,t} = (\text{ax}_{i,t} + \text{ax}_{i,t-1})/2$$

for $x \in \{l, ke, ks, m, e\}$.

TFP. The constructed measure of TFP is a Solow residual based on deflated revenues and also referred to as TFPT by Foster, Haltiwanger and Syverson (2008). I construct the *natural log* of this measure as:

$$\begin{aligned}
\text{lftpt}_{p,t} &\equiv \ln(q_{p,t}) - \text{dal}_{i(p),t} \ln(l_{p,t}) - \text{dake}_{i(p),t} \ln(\text{ksteq}_{p,t}) \\
&\quad - \text{daks}_{i(p),t} \ln(\text{kstst}_{p,t}) - \text{dam}_{i(p),t} \ln(m_{p,t}) - \text{dae}_{i(p),t} \ln(e_{p,t})
\end{aligned}$$

Then, TFP is simply defined as $\text{tfpt}_{p,t} = \exp(\text{lftpt}_{p,t})$.

A.5 Homogeneous product-level data in CM

The Census Bureau collects product-level information for a subset of manufacturing plants at the 7-digit SIC level. The CM is conducted every five years in years ending with “2” and “7”. For these years only, data on physical output is available. I consider the Census years 1977, 1982, 1987, 1992 and 1997 as a baseline as the Census Bureau ceased denoting products at the 7-digit SIC level after 1997. Afterwards, products are only denoted at the 10-digit NAICS level. This makes it hard for me to compare products across time and as a result, I do not use any information from Census years 2002 or 2007 in my baseline results. As an extension, I create a bridge file that connects 7-digit SIC codes with 10-digit NAICS codes. The results are shown in table A.1.

The Census Bureau has a set of special product codes that are important to highlight. First, I drop all products from administrative records sources (“**ar** = 1”). These products are typically produced by very small plants (employment count < 5) and only contain information at the 4-digit SIC level. As a result, I am not able to assign these products to a single 7-digit SIC code. Second, the sum of the total value of shipments reported for each individual product (**pv**) does not always add up to the total value of shipments (**tvS**) reported at the plant level. As a result, the Census Bureau creates “artificial” products to align these numbers. These are denoted as “balancing” codes which I will drop. Third, I exclude any product that is associated with contract work, miscellaneous receipts and/or resales. Fourth, any product with a negative value of shipment is dropped as well. Finally, I drop any remaining observations that are imputed by the Census Bureau. To do this, I employ the item-level edit and imputation flags made available to me by the Census Bureau (see White, 2014).

To avoid any issues regarding unobserved heterogeneity in product quality, I will focus on a particular subset of products that are considered to be the most physically homogeneous in the manufacturing sector. To achieve this, I follow the procedure of Foster, Haltiwanger and Syverson (2008) almost to the letter. With some abuse of terminology, I will define “product industries”. Each product industry can consist of a single or multiple 7-digit SIC level products. Thus, each of the defined products will be the sum of quantities (**pqs**) over one or several 7-digit SIC level products. In the end, I consider five product industries which are defined in table A.1.

The construction of these specific product industries comes with some caveats. Physical quantities were not collected for ready-mix concrete in 1997. The unit of measurement for boxes changed over the sampling period such that data from 1992 and 1997 cannot be compared to other years. Lastly, there are recording errors in physical quantity for processed ice in 1992. As a result, I do not use data on these product industry-year pairs.

Regression specifications are run at the product level. However, the data on inputs (i.e. labor, capital, intermediate inputs and energy) is reported at the plant level. To deal with this, I impose a product specialization criterion: a plant must obtain at least 50 percent of its revenue from sales of the defined product of interest. This share will be denoted as the “primary product specialization ratio” (**ppsr**). As Foster, Haltiwanger and Syverson (2008) mention, this restriction reduces measurement problems in calculating TFPQ. Furthermore, I apportion the share of inputs used to make the product of interest. This is simply done by dividing physical quantities by the product’s share of plant sales. Since I am only focusing on primary products, physical quantities are defined as $q = \text{pqs}/\text{ppsr}$.

To obtain full consistency with the results in Foster, Haltiwanger and Syverson (2008), all prices are denominated in 1987 dollars. All input variables are constructed using industry-level deflators denominated in 1987 dollars. As a result, the only variables to adjust to a common 1987 basis are product-level prices and revenues. First, nominal prices are obtained by dividing revenues by adjusted physical quantities ($p = \text{pv}/q$). Since the latter two are reported on a yearly basis, constructed unit prices are annual averages but can also be interpreted as a quantity-weighted average of all transactions’ prices charged by a plant during some year. Second, I need to construct deflators that have a common 1987 basis. To do this, I adopt a Cobb-Douglas aggregate price index:

$$\bar{P}_{jt} = \prod_{i=1}^{N_{jt}} p_{ijt}^{s_{ijt}}$$

where p_{ijt} denotes the unit price charged by plant i active in product industry j at time t . Similarly, s_{ijt} denotes the plant’s revenue share and N_{jt} counts the number of active plants in

Table A.1: Definition of product categories.¹

PRODUCT	YEAR					
	1977 (SIC7)	1982 (SIC7)	1987 (SIC7)	1992 (SIC7)	1997 (SIC7/NAICS10)	2002 (NAICS10)
CONCRETE	3273000 3273011	3273000	3273000	3273000	3273000100	3273000100
BOXES	2653012 2653013 2653015 2653016 2653018 2653021 2653022 2653029 2653051 2653067 2653068	2653012 2653013 2653015 2653016 2653018 2653021 2653022 2653029 2653051 2653067 2653068	2653012 2653013 2653014 2653015 2653016 2653018 2653021 2653022 2653022 2653030 2653051 2653067 2653068	2653012 2653013 2653014 2653015 2653016 2653018 2653018 2653021 2653022 2653030 2653051 2653067 2653068 2653098	3222110111 3222110114 3222110221 3222110341 3222110345 3222110431 3222110433 3222110435 3222110437 3222110551 3222110661 3222110666 3222110691	3222110111 3222110114 3222110221 3222110341 3222110345 3222110431 3222110437 3222110551 3222110661 3222110666 3222110691
BREAD	2051111	2051111	2051111	2051121 2051122	3118121111 3118121121	3118121111 3118121121
ICE	2097011 2097051	2097011 2097051	2097011 2097051	2097011 2097051	3121130111 3121130121	3121130111 3121130121
GASOLINE	2911131	2911131	2911131	2911131	3241101121	3241101121

¹Product categories are the sum of either 7-digit SIC products (1977 - 1997) or 10-digit NAICS products (1997 - 2002). The Census Bureau assigned both 7-digit SIC and 10-digit NAICS codes to products in 1997. The units of measurement for each product category are thousands of cubic yards (“Concrete”), short tons (“Boxes” and “Ice”) and thousands of pounds (“Bread”). The categories are almost identically defined as in Foster, Haltiwanger and Syverson (2008). The exception is for “Ice”. I combine the products “Manufactured can or block ice” and “Manufactured cubed, crushed or other processed ice”. This is also done by the Census Bureau from 2007 onwards, see “Numerical List for Manufactured and Mineral Products 2007”.

product industry j at time t . Then, real prices adjusted to a common 1987 basis are defined as $p_{ijt}^d = p_{ijt} / (\bar{P}_{jt} / \bar{P}_{j1987})$. Deflated revenues are defined in a similar fashion and are used in the computation of revenue-based TFP (TFPR). Alternatively, I use the shipments deflator from the NBER-CES Manufacturing database that is defined at the 4-digit SIC industry level. The results remain unaltered qualitatively whenever I use this alternative deflator.

A.6 LBD-SSEL-LFTTD merging process

Revenue is defined as the “total value of shipments, sales, receipts and/or revenue” which is denominated in dollars. This is different from the common measure of “total value of shipments” (`tvsl`) in the ASM/CM which encompasses less. The necessary variables to construct revenue are contained in the SSEL files. Revenue can consist of multiple components depending on the type of industry and legal structure which can cause some difficulties. Fortunately, the Census Bureau provides instructions on how to construct revenue by industry.⁷ While this is extremely helpful, there are nevertheless some considerable challenges.

The Business Register’s measures of revenue are based on administrative data from annual business tax returns. These are reported at the tax ID (or Employer Identification Number (EIN)) level. This can potentially cause problems as the notion of “establishment”, “tax ID (EIN)” and “firm” are all not identical. A firm can consist of multiple tax reporting identities whereas multiple establishments are allowed to file taxes under one EIN. Unfortunately, EIN is the most disaggregated level at which revenues are available. As a result, it is not possible to obtain revenues at the establishment level. Of course, single-unit firms are the exception as, by definition, a firm can only consist of one establishment and hence one tax-reporting entity. Thus, revenues at the firm level are easily obtained whenever a firm’s identity (`firmid`) and its tax-reporting entities (`ein`) are linked. While this is possible (as the SSEL files are the underlying source of the LBD), the structure of the Census Bureau’s files requires caution to do this.

Furthermore, there are complications with missing values for the revenue data which are more severe than for employment. Fortunately, the Census Bureau also asks each firm to report their revenues from the previous year which can then be used to fill in missing values.

The variable that links the LBD with the SSEL files is the SSEL’s record number. This variable is not uniquely defined in the LBD though. In fact, it is a mix of the record numbers of two databases. The Census Bureau splits each year’s SSEL file into two files: the single unit (`SSELSu`) and the multi-unit (`SSELmu`) file; each of which contain their own record number. As a result, the

7. Nevertheless, the constructed concept of “revenue” can be hard to compare across industries. As a result, the Census Bureau advises to only use variation *within* industries. This issue is less problematic for the current study as I am mostly focusing on the *growth* of revenue rather than its level.

record number in the LBD is unique up to the level of multi-unit status.

SSELSu contains information of all the business tax returns and is thus comprised of the universe of all EIN's in the U.S. economy. SSELmu contains all the establishments that belong to multi-unit firms.⁸ However, this file does *not* contain any information on revenues. It is necessary to use this file though as the link between the LBD and SSEL for multi-unit firms is based on the record number of the *multi*-unit SSEL file.

The revenue data in the SSEL files are available from 1992 onwards, but most of the values for the years 1992 and 1993 are imputed and not deemed reliable enough. Hence, I construct the measure for firm-level revenue for the period 1994 to 2011. The information on employment and payroll in LBD is very reliable and does not suffer from many missing observations. However, the resulting merging process for revenues does lead to some missing observations. On average, I am able to merge the revenue data for the majority of the firms though. Most importantly however, I do not seem to be missing a extraordinarily high amount of observations for middle-sized or large firms. The missing variable issue does seem somewhat more acute for the smallest firms in the LBD, but this seems less problematic for my purposes as I am focusing mostly on the largest firms in the economy. Moreover, these sampling concerns do not seem to be leading to significant biases as mentioned by Moreira (2016).

The constructed file contains *total* revenues at the firm level. However, the main identifying assumption of the exercise in chapter 3 requires a decomposition of a firm's revenues across destinations (i.e. countries). As a result, I turn to the Census Bureau's trade database. The LFTTD is a database at the transaction level. For every transaction however, it includes a firm identifier (`alpha` which is identical to the LBD's `firmid`) and an identifier that labels the country to which the firm exported (`country`). Given that the LFTTD contains `firmid`, constructing a version of the LBD with revenues at the firm-*destination* level is relatively straightforward.

Unfortunately, the Census Bureau does not provide `firmid` for every transaction in the trade database. This is most probably due to Canadian exports as transactions to Canada are not

8. To be exact, SSELSu contains all single-unit establishments and submasters which are defined as the aggregation of establishments belonging to multi-unit firms to the EIN level. In turn, SSELmu contains the establishment-level data for those EIN entities with multiple establishments within the submasters. This is also mentioned in Moreira (2016).

collected through standard methods. Nevertheless, I am able to match, on average, approximately 75% of the *value* of all transactions. These matching rates are identical to Bernard, Jensen and Schott (2009a) who also work with the LFTTD which should assuage any concerns on the reliability of the merging process.

APPENDIX B

ROBUSTNESS AND ESTIMATION STRATEGIES

B.1 Variance decomposition

FINANCIAL CRISIS OF 2008 - 2009. In chapter 3, I showed that the average granular contribution to aggregate volatility was 15.54%. To calculate this average, I excluded the weights τ from 2008 and 2009 which are the years of the financial crisis. The main reason to exclude these years is that the industry- and firm-level sales composition (and hence their weights) were not representative of the U.S. economy in those years. In particular, U.S. aggregate sales growth plummeted to -8.08% in 2009 according to public figures from the BEA. Note though that all my calculations in section 3.3 do include the shocks from these years as well. In the following, I will show that the results on the relatively small contribution of the granular channel remain qualitatively the same whenever I do include the financial crisis.

Figure B.1: Including the financial crisis years 2008 and 2009.

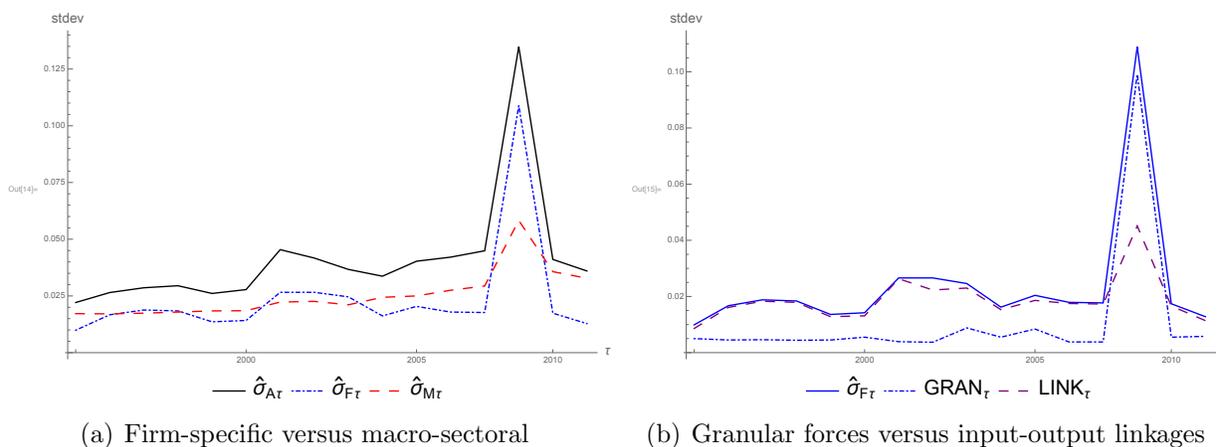


Figure B.1.a shows that the firm-specific component of aggregate volatility increases dramatically during the financial crisis. However, the relative contribution of the firm-specific and macro-sectoral components stay almost identical and amount to 54.24% (compared to 52.47%) and 65.51% (compared to 67%) respectively. A similar picture is depicted for the granular contribution relative to the firm-specific component in figure B.1.b. The linkage component \widehat{LINK}_τ dominates the firm-

specific component for the whole sample period with the exception of 2009 in which the granular component increases substantially. Nevertheless, the average relative contribution of the granular component only increases to 34.23% (compared to 30.45%). Thus, the total contribution of the granular channel amounts to $54.24\% \times 30.45\% = 18.56\%$ which is almost identical to what I find in section 3.3.

IDENTIFYING INPUT-OUTPUT LINKAGES IN THE U.S. ECONOMY. It is clear from section 3.3 that firm-specific volatility is dominated by the component containing comovements between firms' residual growth rates. However, the presence of positive covariances between the residual growth rates of firms does not necessarily imply the existence of input-output linkages in the U.S. economy. An identical pattern on these covariances could be observed due to, for example, local labor markets. In the following, I will suggest an exercise that can provide suggestive evidence of the fact that at least a fraction of these comovements between firms must be due to buyer-supplier networks. To do this, I follow the strategy suggested by di Giovanni, Levchenko and Mejean (2014). The intuition behind this procedure is simple. The component capturing comovements between firms is constructed at the industry pair level and is then correlated with a statistic that summarizes input-output linkages between two industries. Whenever this correlation is sufficiently high, it is concluded that a significant portion of the comovements between firms is due to input-output linkages.

Unfortunately, there is no extensive information available on firm-level interconnections in the U.S. economy which leads me to use measures that are defined at the industry level instead. I construct the linkage measure that captures comovements between firms at the industry-pair level through:

$$\text{LINK}_{ij\tau} = \sum_{f \in I_{i\tau}, n} \sum_{g \in I_{j\tau}, m} w_{fn\tau-1} w_{gm\tau-1} \text{COV}(\varepsilon_{fnt}, \varepsilon_{gmt})$$

where $I_{i\tau}$ denotes the set of firms that are active in industry i at time τ . To create a measure of

input-output linkages, an industry pair (i, j) 's “mean network intensity” is defined as:

$$\mathbf{IO}_{ij\tau} = \frac{1}{2}[(1 - \lambda_{i\tau})\rho_{ij\tau} + (1 - \lambda_{j\tau})\rho_{ji\tau}]$$

where $1 - \lambda_{i\tau}$ denotes the cost share of intermediate inputs in industry i at time τ . Similarly, $\rho_{ij\tau}$ is the share of inputs from sector j in sector i 's spending on intermediate inputs. This measure is attractive for two reasons. First, this measure can be motivated by a simple model of input-output linkages at the firm level.¹ Second, the measure can be constructed with data at the *sector* level. As a result, this measure can be implemented with data from the BEA's annual Input-Output Accounts.

To do this, I use public data from the BEA. First, I construct $\rho_{ij\tau}$ from the BEA's annual Input-Output Accounts data. In particular, I turn to the “Use Tables” for 1995 - 1996 and 1997 - 2011. Unfortunately, sectors are not defined in a consistent way over time. Most probably, this is due to the introduction of NAICS codes in 1997. As a result, I categorize sectors in a way such that they are time-consistent. To do this, I classify industries as in Atalay (2015). A full overview of these industries can be found in table B.1. In the end, I will be working with 30 industries as the LBD does not contain any economic activity by the government. Thus, the entries $\rho_{ij\tau}$ are constructed by only using information from industries 1 to 30 but are normalized such that the equality $\sum_{j=1}^{30} \rho_{ij\tau} = 1$ is satisfied for every industry i and period τ .

Furthermore, I need to calculate the share of intermediate inputs relative to gross output for each of the constructed industries. For this purpose, I use the BEA's annual GDP-by-Industry data. Constructing the required ratio is immediate as intermediate inputs (code **II**) and gross output (code **GO**) can be taken directly from the tables. Aggregation to the required industry level as defined above is straightforward as the BEA uses `IOcode` in both data sets.

Finally, I plot $\mathbf{LINK}_{ij} \equiv \frac{1}{T} \sum_{\tau=1995}^{2011} \mathbf{LINK}_{ij\tau}$ and $\mathbf{IO}_{ij} \equiv \frac{1}{T} \sum_{\tau=1995}^{2011} \mathbf{IO}_{ij\tau}$ against each other. A strong positive correlation between the empirically measured term that captures comovements between firms $\widehat{\mathbf{LINK}}_{ij}$ and the intensity of input-output linkages \mathbf{IO}_{ij} at the industry-pair level would be indicative of the presence of input-output linkages in the U.S. economy.

1. See Appendix E of di Giovanni, Levchenko and Mejean (2014) for a framework that can motivate this measure.

Table B.1: Definitions of constructed industries.¹

#	Industry name	NAICS02-03	IOcode
1	Agriculture and Fishery	111 - 113	111CA, 113FF
2	Mining	212	212
3	Oil and Gas Extraction	211, 213	211, 213
4	Construction	23	23
5	Food and Kindred Products	311, 312	311FT
6	Textile Mill Products	313, 314	313TT
7	Apparel and Leather	315, 316	315AL
8	Lumber	321	321
9	Furniture and Fixtures	337	337
10	Paper and Allied Products	322	322
11	Printing and Publishing	323, 511	323, 511
12	Chemicals	325	325
13	Petroleum Refining	324	324
14	Plastics and Rubber	326	326
15	Non-metallic Minerals	327	327
16	Primary Metals	331	331
17	Fabricated Metal Products	332	332
18	Non-electrical Machinery	333	333
19	Electrical Machinery	335	335
20	Motor Vehicles	336	3361MV
21	Other Transportation Equipment	336	3364OT
22	Computer and Electronic Products	334	334
23	Miscellaneous Manufacturing	339	339
24	Warehousing	493	493
25	Communications and Media	512, 515 - 519	512 - 514
26	Utilities	22	22
27	Wholesale and Retail	42, 441, 445, 452 - 454, 521, 522	42, 441, 445, 452, 4A0
28	Finance, Insurance and Real Estate	521 - 525, 531 - 533	521CI, 523, 524, 525, HS, ORE, 532RL
29	Other Services	541, 55, 561, 562, 611, 621 - 624, 711 - 713, 721, 722, 811 - 814	5411, 5415, 5412OP, 55, 561, 562, 61, 621 - 624, 711AS, 713, 721, 722, 81
30	Transportation	481 - 488, 492	481 - 486, 487OS
31	Government	-	GFGD, GFGN, GFE, GSLG, GSLE

¹This table is based on the BEA's annual Input-Output Accounts data of 1997 - 2014 which contains 71 industries. The data for 1963 - 1996 (of which I only use 1995 and 1996) is less disaggregated but nevertheless contains 65 industries. Converting these industries to the groups as defined above is identical to table B.1.2 with the exception of: Wholesale and Retail (IOcodes 42 and **44RT**), Finance, Insurance and Real Estate (IOcodes 521CI, 523, 524, 525, **531**, 532RL), Other Services (5411, 5415, 5412OP, 55, 561, 562, 61, 621, **622HO**, 624, 711AS, 713, 721, 722 and 81) and Government (IOcodes GFG, GFE, GSLG and GLE).

ADDITIONAL ROBUSTNESS SPECIFICATIONS. The results of section 3.3 are based on firm-destination level growth rates that are constructed through natural log differences. This is useful as it allows for

a decomposition of individual level growth rates into a firm-specific and macro-sectoral component. By construction however, this growth rate measure is not properly defined for entry and exit and thus ignores the extensive margin completely. As a result, I will resort to the growth rate measure suggested by Davis, Haltiwanger and Schuh (1996) that treats entry and exit symmetrically:

$$\tilde{\gamma}_{fnt} = \frac{x_{fnt} - x_{fnt-1}}{\frac{1}{2}(x_{fnt} + x_{fnt-1})} \in [-2, +2]$$

This implies that entrants and exiters have growth rates equal to -2 and $+2$ respectively. Technically speaking, the suggested decomposition of individual level growth rates in section 3.1 does not hold for this specific measure. Thus, the presented results should be interpreted with some care. To achieve consistent aggregation, the appropriate weights are:

$$\tilde{w}_{fnt} = \frac{x_{fnt} + x_{fnt-1}}{\sum_{f',n'} x_{f'n't} + x_{f'n't-1}}$$

which implies that aggregate sales growth must satisfy $\tilde{\gamma}_{At} = \frac{X_t - X_{t-1}}{0.5(X_t + X_{t-1})} = \sum_{f,n} \tilde{w}_{fnt} \tilde{\gamma}_{fnt}$. I perform two exercises with the growth rate by Davis, Haltiwanger and Schuh (1996). First, I consider the full sample with entrants and exiters. As a result, the estimation sample increases considerably. The relative contribution of the firm-specific component decreases substantially compared to the benchmark specification of section 3.3. Even though this is somewhat surprising at first sight, this could be explained by the fact that firm-level entry and exit amplify and propagate the effects of aggregate shocks as mentioned by Clementi and Palazzo (2016). This indirect effect can obviously not be identified in the baseline sample, but can prove to be important quantitatively as firm-level entry and exit rates are persistently high in the U.S. economy.² The average contribution of the granular component relative to the firm-specific component is nearly identical to the benchmark specification. As a result, the total average contribution of granularity is smaller than the baseline specification which only reinforces my conclusions.

Second, I use the growth rate by Davis, Haltiwanger and Schuh (1996) but trim outliers as in the baseline specification. This removes the effects of the extensive margin, but considers the sole

2. Statistics from the Business Dynamics Statistics (BDS) indicate that average firm-level entry and exit rates from 1976 to 2011 amount to 11.12% to 8.81% respectively.

effect of using a different growth rate. The results are extremely similar to the baseline results. This is not surprising as the growth rate of Davis, Haltiwanger and Schuh (1996) can be interpreted as a second-order approximation to natural log differences.

Lastly, I consider three-year growth rates instead of annual growth rates to deal with measurement error in year-to-year growth rates. The results are nearly identical compared to the baseline specification of section 3.3. Furthermore, three-year growth rates also capture any slow (i.e. longer than a year) propagation effects of firm-level idiosyncratic shocks through the production network. The results indicate that the firm-specific component is barely affected.

B.2 Size-variance relationship

ROBUSTNESS TESTS FOR KOREN-TENREYRO PROCEDURE. The results of table 4.2 in section 4.2 indicate that the size-variance relationship is robust when I control for the (natural log of the) number of establishments or products within the firm. This implies that the size-variance relationship cannot be driven entirely by either output or product diversification. In the following, I reinforce these findings by using different regression specifications. First, I regress (the natural log of) firm-level volatility on (the natural log of) size and a *dummy* $\mathbf{d}_{i\tau}$ that equals 1 whenever firm i consists of more than one establishment or product at year block τ instead of the *number* of establishments or products.

The results are consistent with the main text as the elasticity of firm-level volatility with size is basically unaffected. Furthermore, being a multi-unit or multi-product firm is positively (rather than negatively) related with firm-level volatility. This result is akin to Koren and Tenreyro (2013) who find a similar result with the number of business segments in Compustat for U.S. firms. Alternatively, I estimate the size-variance relationship on the subsample of firms that only have one establishment or product. It appears that these single-unit firms do not display different size elasticities with respect to firm-level volatility. Note that all of these observations hold regardless of whether I control for firm-level fixed effects or not.

Lastly, I verify that my results on the size-variance relationship are not completely driven by selection. Table 4.2 in section 4.2 already included a specification with cohort fixed effects to deal

with this, but I also estimate the same regression specifications with a balanced panel to ensure the robustness of my results. Obviously, this implies that my sample shrinks but I am nevertheless left with a substantial sample. The results indicate that the estimates from table 4.2 are robust which alleviates any concerns due to selection.

TESTING FOR STRUCTURAL BREAKS IN THE SIZE-VARIANCE RELATIONSHIP. In section 4.1 and 4.2, I assume that the size-variance relationship holds uniformly across the whole size distribution. However, it could be conjectured that firm-level volatility declines at a substantially slower rate in size for large firms. Formally, it is possible that firm-level volatility takes the form:

$$\sigma(g|S) \propto \begin{cases} S^{-\alpha}, & \text{if } S < B, \\ S^{-\alpha+\alpha_b}, & \text{otherwise.} \end{cases}$$

where $\alpha_b \in [-\infty, \alpha)$ and $B > 1$. This is potentially important as the imposition of a single regime for the size-variance relationship could substantially underestimate the volatility of larger firms.³ In turn, this can have consequences on the results of section 4.3. In that section, I argue that the relatively small role of the granular channel can be rationalized by the size-variance relationship alone as my estimates in section 4.1 and 4.2 clearly indicate that larger firms are substantially less volatile than their smaller sized counterparts.

Under a discontinuous regime for the size-variance relationship as conjectured above, this intuition does not necessarily have to hold. I will show though that this is not the case by means of testing for a structural break. More precisely, I test the null hypothesis of the coefficient on $\ln\text{SIZE} \times \mathbb{1}(S_{it} > B)$ being zero for several known breaking points in size B . The indicator $\mathbb{1}(S_{it} > B)$ denotes whether the firm at time block t is large or not. The results indicate that for large values of B (i.e. large firm size) the null hypothesis can never be rejected. I only focus on relatively large firms as aggregate volatility is primarily determined by the largest firms in a granular economy. Thus, I conclude that the size-variance relationship does not show any forms of a structural break for the upper right tail of the firm size distribution.

3. Obviously, this can also result in significantly overestimating the volatility of the largest firms. However, this would only reinforce the results of section 4.3. As a result, I do not deal with this scenario.

STANLEY-SUTTON PROCEDURE FOR SIZE-VARIANCE RELATIONSHIP. My results on the size-variance relationship are primarily based on the methodology by Koren and Tenreyro (2013). However, another popular methodology is used by Stanley et al. (1996) and Sutton (2002). These authors view the size-variance relationship as a purely cross-sectional phenomenon and therefore provide year-by-year estimates for α .

The procedure is carried out as follows. First, take the employment cross-sections of two consecutive years (e.g. year x and year $x + 1$). Second, calculate year-to-year growth rates at the firm level and order them by the size of the firm in the initial year (i.e. year x). Third, categorize firm size in $\mathcal{B} > 1$ bins, take the standard deviation of firm-level growth rates vol_b and calculate average size size_b within each bin b . This results in the set of pairs $(\text{vol}_b, \text{size}_b)_{b=1}^{\mathcal{B}}$. Finally, regress the natural log of vol_b on the natural log of size_b with OLS.

I implement this procedure by using the bin categories for employment as adopted by the Census Bureau and create the bins 1 – 4, 5 – 9, 10 – 19, 20 – 49, 50 – 99, 100 – 249, 250 – 499, 500 – 999, 1000 – 2499, 2500 – 4999, 5000 – 9999 and ≥ 10000 . These categories are, for example, used in the Business Dynamics Statistics (BDS).

Using the LBD, I repeat this procedure for every cross-section from 1976 – 1977, 1977 – 1978, ..., 2010 – 2011 resulting in 35 values for $\hat{\alpha}$. Its average estimate over time is extremely similar to the results shown in the main text. Lastly, I repeat the procedure above but allow for a more fine level of disaggregation at the right tail of the size distribution. Thus, the bin ≥ 10000 is further divided up into 10000 – 14999, 15000 – 24999 and ≥ 25000 . Once again, the results are similar to the benchmark specification of table 4.1 in section 4.1. Thus, the results for the size-variance relationship do not depend on its estimation methodology.

B.3 Size-varying price elasticities

I show in section 5.4 that large plants, defined as plants with more employees than the 75th percentile in a given year’s size distribution, are faced with a less elastic demand. However, the results are only based on plants producing ready-mixed concrete (7-digit SIC 3273000). Even though the

bulk of my estimation sample consists of plants producing concrete, I will show that most of the results in the main text also hold for a variety of other homogeneous products in the manufacturing sector. I follow Foster, Haltiwanger and Syverson (2008) and also estimate the main specification of section 5.1 for the product categories boxes, bread, processed ice and gasoline. The previous study considers a broader range of products, but I only focus on those products that have at least 50 observations for each Census year for at least three years to leave a sufficient amount of variation in size across plants and time.⁴

Smaller firms face a more elastic demand curve (i.e. more negative estimate for $\hat{\sigma}_1$). The results for the baseline specification in section 5.1 are identical for the category boxes. The estimates for bread, processed ice and gasoline are economically significant but not always statistically significant. Imprecise estimates might be obtained as the total number of large plants for these products is fairly low across time.

Lastly, I show that the results on size-varying elasticities do not depend on the definition of a “large” plant. The main text defines a plant to be large whenever its size exceeds the 75th percentile of the size distribution in a given year. The results are qualitatively the same whenever different cut-off percentiles are used (i.e. 50th and 90th percentile). As expected, larger plants face smaller price elasticities. A similar conclusion holds for the other product categories and are available on the author’s request.

B.4 Markup estimation: de Loecker and Warzynski (2012)

In this section, I will provide more details on the markup estimation procedure. To this extent, I will use the methodology by de Loecker and Warzynski (2012). Their procedure encompasses a relatively broad environment while resting on only a mild set of assumptions. Let the production function be given by $Q_{it} = Q_{it}(\mathbf{X}_{it}, K_{it}, \omega_{it})$ where $\mathbf{X}_{it} = (X_{it}^1, \dots, X_{it}^V)$ is a vector of inputs free of adjustment costs, K_{it} denotes capital and ω_{it} is a plant’s technology level. Following de Loecker and Warzynski (2012), I will impose the following assumptions.

4. See table A.1 of Appendix A.5 to see the exact product composition of these product groups.

B1. The production function $Q_{it}(\cdot)$ is continuous and twice differentiable with respect to its arguments.

B2. Productivity ω_{it} is one-dimensional and Hicks-neutral.

B3. Technology parameters are constant across time and common within a industry group.

Imposing assumption **B1** and using standard cost minimization arguments, it is straightforward to obtain:

$$\theta_{it}^X = \mu_{it} \frac{P_{it}^X X_{it}}{P_{it} Q_{it}}$$

where $\theta_{it}^X = \frac{\partial Q_{it}(\cdot)}{\partial X_{it}} \frac{X_{it}}{Q_{it}}$ is the output elasticity with respect to input X and μ_{it} is the plant-specific time-varying markup. Let $\alpha_{it}^X \equiv \frac{P_{it}^X X_{it}}{P_{it} Q_{it}}$ be the ratio of input X 's expenditure to total sales. Note that input expenditures and total sales are directly observed from the ASM/CM data. As a result, markups μ_{it} are easily obtained through the identity:

$$\mu_{it} = \frac{\theta_{it}^X}{\alpha_{it}^X}$$

Thus, the main goal is to estimate the elasticity θ_{it}^X . To accompany this, I impose assumptions **B2** and **B3** for the estimation of production technologies through proxy methods. Assumption **B3** gives me that $\theta_{it}^X = \beta_X^{j(i)}$ where $j(i)$ denotes the industry of some plant i . These assumptions jointly immediately imply that production can be written as $Q_{it} = F(\mathbf{X}_{it}, K_{it}; \boldsymbol{\beta}) \exp(\omega_{it})$. I also allow for measurement error and assume that *observed* logged output satisfies $y_{it} = \ln(Q_{it}) + \epsilon_{it}$. The error term ϵ_{it} is *not* observed by plants when they have to make their optimal input decisions. As a result, I can write:

$$y_{it} = f(\mathbf{x}_{it}, k_{it}; \boldsymbol{\beta}) + \omega_{it} + \epsilon_{it}$$

Obviously, plant-level productivities ω_{it} are not observed by the econometrician. As a result, I turn to the methodology by Levinsohn and Petrin (2003) and use material demand m_{it} to proxy for

productivity:

$$m_{it} = m_t(k_{it}, \omega_{it}, \mathbf{p}_{it})$$

where capital k_{it} and productivity ω_{it} are state variables at the time of input choice(s). Furthermore, the vector \mathbf{p}_{it} denotes any additional variables that can affect a plant's optimal demand for material inputs. Under a mild set of assumptions, as described in appendix A of Levinsohn and Petrin (2003), the material input demand function is invertible. Thus, I obtain $\omega_{it} = h_t(m_{it}, k_{it}, \mathbf{p}_{it})$. As a result, production y_{it} can be written in terms of observables only:

$$\begin{aligned} y_{it} &= f(\mathbf{x}_{it}, k_{it}; \boldsymbol{\beta}) + h_t(m_{it}, k_{it}, \mathbf{p}_{it}) \\ &= \phi_t(\mathbf{x}_{it}, k_{it}, m_{it}, \mathbf{p}_{it}) + \epsilon_{it} \end{aligned}$$

I will focus on gross output Cobb-Douglas production functions in my empirical applications. The case for translog production is conceptually almost identical. Estimating the production technology parameters is done in a three stage fashion which is in a similar spirit as Akerberg, Caves and Frazer (2015).

Step 1. Non-parametric estimation of ϕ_{it} and ϵ_{it} .

First, I estimate φ_{it} and ϵ_{it} non-parametrically by approximating them with a polynomial in $(\ell_{it}, k_{it}, m_{it}, e_{it})$. More precisely, I run the regression:

$$\begin{aligned} y_{it} = & \sum_{j=1}^M \ell_{it}^j + m_{it}^j + k_{it}^j + e_{it}^j + \left[\sum_{j'=1}^N \ell_{it}^j m_{it}^{j'} + \ell_{it}^j k_{it}^{j'} + \ell_{it}^j e_{it}^{j'} + k_{it}^j m_{it}^{j'} + k_{it}^j e_{it}^{j'} + m_{it}^j e_{it}^{j'} \right] \\ & + \ell_{it}^j m_{it}^j k_{it}^j e_{it}^j + \epsilon_{it} \end{aligned}$$

for some integers $M, N \geq 1$. Let its fitted values and residuals be denoted by $\hat{\phi}_{it}$ and $\hat{\epsilon}_{it}$ respectively.

Step 2. Construction of innovations ξ_{it} to productivity ω_{it} .

Second, I assume that productivity ω_{it} is only a function of its lagged value. As a result, I get $\omega_{it} = g_t(\omega_{it-1}) + \xi_{it}$. Under a Cobb-Douglas specification, productivity is approximated by:

$$\omega_{it}(\boldsymbol{\beta}) = \widehat{\phi}_{it} - \beta_\ell \ell_{it} - \beta_k k_{it} - \beta_m m_{it} - \beta_e e_{it}$$

where $\boldsymbol{\beta} \equiv (\beta_\ell, \beta_k, \beta_m, \beta_e)'$.⁵ Then, I approximate $g_t(\cdot)$ with a third order polynomial in its argument:

$$\omega_{it}(\boldsymbol{\beta}) = \rho_0 + \rho_1 \omega_{it-1}(\boldsymbol{\beta}) + \rho_2 \omega_{it-1}^2(\boldsymbol{\beta}) + \rho_3 \omega_{it-1}^3(\boldsymbol{\beta}) + \xi_{it}$$

Thus, the innovations to productivity can be constructed as a function of $\boldsymbol{\beta}$ through:

$$\xi_{it}(\boldsymbol{\beta}) = \omega_{it}(\boldsymbol{\beta}) - \Omega_{it-1}(\boldsymbol{\beta})' \widehat{\rho}(\boldsymbol{\beta})$$

where $\widehat{\rho}(\boldsymbol{\beta}) = (\widehat{\rho}_0, \widehat{\rho}_1, \widehat{\rho}_2, \widehat{\rho}_3)'$ is the OLS estimator associated with the set of regressors $\Omega_{it-1}(\boldsymbol{\beta})$ on $\omega_{it}(\boldsymbol{\beta})$. Thus, we have $\Omega_{it-1}(\boldsymbol{\beta}) = (1, \omega_{it-1}(\boldsymbol{\beta}), \omega_{it-1}^2(\boldsymbol{\beta}), \omega_{it-1}^3(\boldsymbol{\beta}))'$

Step 3. GMM-IV estimation of $\boldsymbol{\beta}$.

I assume that capital k_{it} is decided one period ahead, thus it is orthogonal to the innovation $\xi_{it}(\boldsymbol{\beta})$. Furthermore, lagged material inputs m_{it-1} are used to instrument for m_{it} as current period material inputs are decided after the realization of the innovation $\xi_{it}(\boldsymbol{\beta})$. As a result, I expect $\mathbb{E}(m_{it} \xi_{it}) \neq 0$. However, I require that m_{it} and m_{it-1} to be correlated with each other to satisfy the rank conditions. This is the case whenever input prices for materials inputs are correlated over time. Atalay (2015) finds empirical evidence for this using U.S. manufacturing data. A similar argument is made for labor inputs ℓ_{it} and energy inputs e_{it} . Define the instrument vector as $\mathbf{z}_{it} = (\ell_{it-1}, k_{it}, m_{it-1}, e_{it-1})'$,

5. The case for translog production is extremely similar. In fact, the first stage is identical to the case for Cobb-Douglas production technologies. The only difference is that the dimension of the vector $\boldsymbol{\beta}$ is substantially increased as $\boldsymbol{\beta} = (\beta_\ell, \beta_k, \beta_m, \beta_e, \beta_{\ell\ell}, \beta_{kk}, \beta_{mm}, \beta_{ee}, \beta_{\ell m}, \beta_{km}, \beta_{em}, \beta_{k\ell}, \beta_{ke}, \beta_{\ell e}, \beta_{\ell km}, \beta_{\ell ke}, \beta_{kme}, \beta_{\ell me})'$ whenever I adopt a polynomial approximation of degree $M = 3$.

then the parameters β are identified through the following $K = 4$ moment conditions:

$$\mathbb{E}(\xi_{it}(\beta)\mathbf{z}_{it}) = \mathbf{0}_{4 \times 1}$$

To obtain β , I rely on the minimization of a quadratic loss function which is standard in GMM estimation.⁶ Thus, I get:

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}_+^4} \sum_{k=1}^K \left(\sum_{i=1}^N \sum_{t=1}^T \xi_{it}(\beta) z_{it}^k \right)^2$$

Recall that markups are constructed using the output elasticity with respect to intermediate inputs:

$$\hat{\mu}_{it} = \hat{\beta}_m \left(\frac{\mathbf{vm}_{it}}{\mathbf{go}_{it}/\hat{\epsilon}_{it}} \right)^{-1}$$

where \mathbf{vm}_{it} and $\mathbf{go}_{it} = \mathbf{tvs}_{it} + (\mathbf{fie}_{it} - \mathbf{fib}_{it}) + (\mathbf{wib}_{it} - \mathbf{wib}_{it})$ denote a plant i 's total expenditure on intermediate inputs and value of gross output at year t . Note that gross output is defined as total sales *plus* the value of finished goods and work-in-progress inventories. My markup results are robust to whenever I do not include inventories.

6. By construction, the number of parameters in β is equal to the amount of identifying moments. This case of “just identification” renders the specification of a weighting matrix useless.

APPENDIX C

ACCOUNTING FRAMEWORK FOR GRANULAR ECONOMY

I show in section 4.3 that the aggregate implications of the size-variance relationship are substantial. To show this quantitatively, any framework with firm-specific shocks alone that delivers a firm size distribution with a Pareto right tail and the size-variance relationship as a power law is sufficient. In the following section, I show the explicit microfoundations of such a framework.

REPRESENTATIVE CONSUMER. The representative consumer's problem is characterized by CES preferences over a discrete number $N \geq 1$ of goods (indexed by k) with elasticity θ . Labor endowment L is supplied inelastically and compensated at the wage rate $w > 0$. Total expenditure X comprises of labor income and received profits $\tilde{\pi}$ from operating firms as they are fully owned by the representative consumer. As a result, expenditure on good k equals:

$$x(k) = X \cdot \left(\frac{p(k)}{P} \right)^{1-\theta}$$

where $P = \left(\sum_{k=1}^N p(k)^{1-\theta} \right)^{1/(1-\theta)}$ denotes the ideal price index and reflects the price of an additional unit of utility.

FIRMS. A firm k is defined by permanent productivity $a(k)$ that is determined upon entry and does not change after. However, a firm will also be subject to idiosyncratic shocks $z(k)$. Then, it can transform $a(k)z(k)$ input bundles into one unit of output. Aggregate volatility will be driven solely by the idiosyncratic shocks $z(k)$ alone in the constructed granular economy. The input bundle ι is a Cobb-Douglas composite of labor ℓ and some intermediate input bundle M , i.e.

$$\iota(k) = \left(\frac{\ell(k)}{\beta} \right)^{\beta} \left(\frac{M(k)}{1-\beta} \right)^{1-\beta}$$

where $M(k) = \left(\sum_{k'=1}^N m(k, k')^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}}$ and $m(k, k')$ denotes firm k 's demand for the good produced by firm k' . Thus, the production side of the economy features a roundabout structure as in

Basu (1995). By letting $\beta \rightarrow 1$, these linkages are absent. By standard arguments, the cost of one input bundle ι is $c \equiv w^\beta P^{1-\beta}$. A firm's production function is linear in this input bundle, thus its optimal price is characterized by the standard constant CES markup over marginal costs which in turn results in the following equilibrium supplied quantity:

$$\left(\frac{p^*(k)}{P}\right)^{-\theta} \frac{X}{P} = \frac{X}{P^{1-\theta}} \left[\frac{\theta}{\theta-1} ca(k)z(k)\right]^{-\theta}$$

ENTRANTS. Each potential entrant can discover the permanent component of its inverse productivity a after paying the entry cost f_e which is denominated in input bundles. Once paid, these costs are sunk and a firm, given its realization of a , has to decide whether to incur the fixed cost of production f_P which is also denoted in input bundles. I explicitly assume that a firm has to incur the cost f_P before the realization of the idiosyncratic shock z .

Due to these timing assumptions, a firm k will only decide to enter if the permanent component of its inverse productivity $a(k)$ is low enough.¹ Thus, there exists a cut-off value \bar{a} such that firms only incur the fixed cost of production f_P whenever $a \leq \bar{a}$. This selection procedure is similar to Melitz (2003) who relies on a *Law of Large Numbers* to validate his argument as his setup features a continuum of firms. However, I explicitly assumed that there is a *finite* amount of potential entrants and hence the number of active firms must be finite as well. Recall this is necessary for any granular framework as a continuum of firms would otherwise automatically imply an aggregate volatility of zero. Therefore, I need to impose two assumptions to validate this selection procedure.

ASSUMPTION 1. The *marginal* firm, which is small enough, ignores the impact of its own realization of z on aggregate expenditure X and price P .

ASSUMPTION 2. The *marginal* firm treats X and P as non-stochastic.

1. To see this, note that ex-interim expected profits, that is whenever the permanent component of productivity $a(k)$ is already realized, are equal to:

$$\bar{\pi}(a(k)) = \mathbb{E}_z \left[\frac{X}{\theta P^{1-\theta}} \left(\frac{\theta}{\theta-1}\right)^{1-\theta} (ca(k)z(k))^{1-\theta} - cf_P \middle| a(k) \right]$$

As $\theta > 1$, there exists a cut-off \bar{a} such that $\bar{\pi}(a) \geq 0$ if and only if $a \leq \bar{a}$.

These assumptions deliver substantial computational and analytical simplifications: solving the equilibrium is reduced to a static problem as will become clear.² The number of potential entrants \bar{N} will be pinned down by the free entry condition in which the *ex-ante* profits of a firm must be reduced to zero. To close the model in general equilibrium, I need to solve for the aggregate price P . To avoid a complex computational procedure, I will *approximate* the equilibrium price level P by ignoring firm-specific idiosyncratic shocks by fixing them at their expected values. As a result, P satisfies:

$$P = \left(\sum_{k=1}^N \mathbb{E}_{a,z} \left[(p^*(k))^{1-\theta} \right] \right)^{\frac{1}{1-\theta}} = \left(\Pr(a \leq \bar{a}) \bar{N} \left(\frac{\theta}{\theta-1} c \right)^{1-\theta} \mathbb{E}_a \left[a^{1-\theta} | a \leq \bar{a} \right] \right)^{\frac{1}{1-\theta}}$$

Then, I can solve for a *monopolistically competitive* equilibrium (w, P) by clearing the labor market $\sum_{k=1}^N \ell^*(k) = L$, good markets $y(k) = c(k) + \sum_{k'=1}^N m(k', k)$ for every $k \in \{1, 2, \dots, N\}$ and imposing free entry which brings net aggregate profits to zero such that $\beta X = wL$. The numéraire of the economy is the wage rate w , thus I set $w = 1$.

C.1 Calibration

Aggregate volatility can be driven by idiosyncratic shocks alone whenever the firm size distribution is fat-tailed. However, I argue in section 4.3 that it is equally important to take the size-variance relationship into account. To do this in the easiest way, I normalize $\mathbb{E}_z (z^{1-\theta}) \equiv \mathbb{E}_z (\tilde{z}) = 1$ without loss of generality and impose:

ASSUMPTION 3. The permanent component of productivity satisfies $1/a \sim \text{PAR}(b, \varphi)$ and the

2. While assumption 1 is standard and “not controversial” according to di Giovanni and Levchenko (2012), it is not completely trivial in a granularity context. If the behavior of the largest firms are the most important for aggregate volatility, then it is reasonable to assume that these firms are aware of their impact on aggregate prices. As a result, the standard CES markup over marginal cost would *not* be the firm’s optimal pricing function. However, di Giovanni and Levchenko (2012) suggest that deviations of fully flexible markups from constant CES markups are very small; even for the largest firms in the economy. This insight is shared by Gaubert and Itskhoki (2015) who note that competition through prices leads to substantially less markup variation than through quantities. The second assumption implies that the marginal firm ignores the fact that its entry into the market has an impact on aggregate volatility and prices. As a result, I am abstracting from the impact that the marginal firm has on the value of entry through its own entry decision. Furthermore, it only applies to the *marginal* firm. Assuming it is small enough, my conclusions on granularity should be affected minimally as only the large firms should dominate aggregate outcomes. In any numerical exercise, I will explicitly verify that the market share of the marginal firm is negligible.

conditional standard deviation of a firm’s idiosyncratic shock is decreasing in its size at the rate α , i.e. we have $\text{SD}_z [\tilde{z}|x] = Ax^{-\alpha}$.

Then, it is straightforward to derive that the constructed granular economy has a Pareto right-tail with coefficient $\zeta = \frac{\varphi}{\theta-1}$ and aggregate volatility is given by:

$$\text{SD}_z \left[\frac{\Delta X}{\mathbb{E}_z(X)} \right] = A \cdot \sqrt{\sum_k (\mathbb{E}_z(x(k))^{-\alpha} h(k))^2}$$

where firm k ’s expected share of total sales equals $h(k) \equiv \mathbb{E}_z(x(k))/\mathbb{E}_z(X)$. To implement the model quantitatively, I set 4 parameters as according to the literature. The elasticity of substitution satisfies $\theta = 6$ which is in the middle of the range $[3, 10]$ as mentioned by Anderson and van Wincoop (2004). Furthermore, I set $\beta = 0.56027$ which is the average of the ratio of gross output over GDP in the period 1950 - 2011 according to the BEA’s GDP-by-Industry tables. Labor endowment equals $L = 155 \cdot 10^6$ which is equal to the civilian labor force in 2012 according to the BLS. Lastly, I set $b = 0.1$ as in di Giovanni and Levchenko (2012) to ensure that the market share of the marginal firm is negligible. This ensures that assumption A.2 is satisfied. The parameters α and A are estimated from the LBD and taken from section 4.1. The remaining parameters are calibrated internally according to standard moments.

Table C.1: Internally calibrated parameters.⁶

PARAMETER	VALUE	TARGET	MODEL	SOURCE
φ	5.3	1.06	1.06	Pareto tail (Luttmer, 2007)
f_e	1.9189	0.1091	0.1091	Entry rate (BDS, 1977 - 2012)
f_P	1.0546	$5.0309 \cdot 10^6$	$5.0309 \cdot 10^6$	Firm count (BDS, 2012)

⁶The parameter φ governs the Pareto right-tail of the size distribution, thus it is set to match the empirical Pareto right-tail in Luttmer (2007). The fixed cost of entry f_e and production f_P jointly determine the fraction of potential entrants that enters the economy and thus also the absolute number of active firms. The empirical counterpart for the former is the firm entry rate which I set equal to the average entry rate over the period 1977 - 2012 in the Business Dynamics Statistics (BDS) whose underlying source is the LBD. Lastly, the absolute number of active firms is set equal to approximately 5 million which is the number of U.S. firms in 2012 according to the BDS.

C.2 Alternative calibration strategies

Section 4.3 displayed the aggregate implications of the size-variance relationship in a granular economy. More precisely, the explanatory power of the granular channel is capped at 14.7% whenever the size-variance relationship is taken into account as estimated from the data. However, it is argued that only the volatility of the largest firms matter in granular economies. Thus for the sake of robustness, I implement the values from the size-variance relationship that are estimated from the subsample of large firms only.

Even though the estimates vary somewhat across the regression specifications, the quantitative role of the granular channel is always about half or less relative to the results of the previous literature. More precisely, the contribution of the granular channel is bounded between 8.9% and 18.1% which is extremely similar to the baseline results of section 4.3. This alleviates any concerns related to the results on the size-variance relationship in table 4.1 underestimating the volatility of the largest firms in the U.S. economy.

APPENDIX D

PROOFS OF THEORETICAL FRAMEWORK

D.1 Properties of the Lambert W -function

The Lambert W -function is defined as that mapping W satisfying $z = W(z \cdot \exp(z))$. This function possesses the following properties which I exploit at several stages of the proofs for lemma 1 and propositions 1 to 3.

Property 1. $W(x) > 0$ and $W(0) = 0$.

Property 2. $\frac{dW(x)}{dx} = \frac{W(x)}{x(1+W(x))} > 0$ and $\frac{d^2W(x)}{dx^2} < 0$ for $x > 0$.

Property 3. $\ln(W(x)) = \ln(x) - W(x)$ for $x > 0$.

In the following, I will take these properties of the Lambert W -function as given.

D.2 Proof of lemma 1

Recall that firm-level revenue can be decomposed as:

$$\text{rev}(s_a) = g(a) \times h(s_a)$$

Thus, it is sufficient to show that $h'(s_0) > 0$. Straightforward differentiation and using property **3** of the Lambert W -function implies that:

$$\text{sgn} \{h'(s_0)\} = \text{sgn} \left\{ \frac{W(\Theta \cdot \exp(-s_0))}{1 + W(\Theta \cdot \exp(-s_0))} \right\}$$

which is always positive by property **1** of the Lambert W -function whenever $\Theta > 0$. □

D.3 Proof of proposition 1

I start by showing that the following expression is strictly decreasing in s_0 :

$$\sigma_F \frac{h'(s_0)}{h(s_0)}$$

Straightforward differentiation of the above expression and using property **3** of the Lambert W -function implies that:

$$\text{sgn} \left\{ \frac{d}{ds_0} \frac{h'(s_0)}{h(s_0)} \right\} = \text{sgn} \left\{ -2\theta + (\varepsilon - \theta)W(\Theta \cdot \exp(-s_0)) \right\}$$

By property **1** of the Lambert W -function, $W(\Theta \cdot \exp(-s_0)) > 0$ whenever $\Theta > 0$. Then, it is obvious that $\theta > \varepsilon$ is a sufficient condition for $\frac{d}{ds_0} \frac{h'(s_0)}{h(s_0)} < 0$ to hold. For $\varepsilon \rightarrow 0$, the Kimball demand specification converges to the CES case which gives me $h(s_0) \propto \exp((\theta - 1)s_0)$. This immediately implies that $\frac{h'(s_0)}{h(s_0)}$ is constant and equal to $\theta - 1$. The statement in the proposition then follows directly from lemma 1. \square

D.4 Proof of proposition 2

Along the balanced growth path, per capita consumption C_t with $t = t^b + a$ grows at the rate κ .

By lemma 1, it is then sufficient to show that:

$$\frac{h'(s_0)}{h(s_0)} \rightarrow 0 \text{ and } \frac{h''(s_0)}{h(s_0)} \rightarrow 0$$

as $s_0 \rightarrow +\infty$. Note that $h'(s_0)/h(s_0)$ can be simplified to:

$$\frac{h'(s_0)}{h(s_0)} = \varepsilon \frac{W(\Theta \cdot \exp(-s_0))^2}{[1 + W(\Theta \cdot \exp(-s_0))] \cdot [\theta - \varepsilon W(\Theta \cdot \exp(-s_0))]}$$

By property **1** of the Lambert W -function, it is true that $W(0) = 0$. This immediately implies that

$\lim_{s_0 \rightarrow +\infty} \frac{h'(s_0)}{h(s_0)} = 0$. The case for $\lim_{s_0 \rightarrow +\infty} \frac{h''(s_0)}{h(s_0)} = 0$ is identical. Lastly, the case of $\varepsilon \rightarrow 0$ is equivalent to CES preferences with substitution parameter $\theta > 0$. As a result, $h(s_0) \propto \exp((\theta - 1)s_0)$. This

implies that $\frac{h'(s_0)}{h(s_0)} = \theta - 1$ and $\frac{h''(s_0)}{h(s_0)} = (\theta - 1)^2$. By Itô's lemma and $\mathbb{E}(\mathcal{B}_a) = 0$, it is true that $m(s_0) = \kappa + \mu(\theta - 1) + \frac{\sigma_F^2}{2}(\theta - 1)^2$ along the constructed balanced growth path as $\varepsilon \rightarrow 0$. \square

D.5 Proof of proposition 3

The firm's optimal price as a function of productivity z is simply characterized as:

$$p^*(z) = \frac{\theta}{\varepsilon} \frac{1}{W\left[\Theta \frac{w}{z} \frac{1}{P}\right]} \frac{w}{z}$$

The production technology is linear in labor only, hence a firm's marginal cost of production must be equal to $MC(z) \equiv w/z$. Obviously, this implies a markup, defined as equilibrium price over marginal cost, of:

$$p^*(z)/MC(z) = \frac{\theta}{\varepsilon} \frac{1}{W\left[\Theta \frac{w}{z} \frac{1}{P}\right]}$$

Recall that I defined detrended log productivity as $s_{t^b,a} = \ln\left(\frac{z_{t^b,a}}{z_t^*}\right)$ and $z_t^* \propto w_t/P_t$. Hence, there exists a constant $\mathbf{z} > 0$ such that markups can be expressed as a function of detrended log productivity as:

$$p^*(s)/MC(s) \Big|_{s=s_0} = \frac{\theta}{\varepsilon} \frac{1}{W\left[\Theta \cdot \mathbf{z} \cdot \exp(-s_0)\right]}$$

Firm-level markups are then increasing in detrended log productivity s_0 by property **2** of the Lambert W -function. The first statement of the proposition then follows from lemma 1. Whenever $\varepsilon \rightarrow 0$, the Kimball demand specification converges to the CES case with substitution parameter $\theta > 0$ (see footnote 34). This gives rise to the well-known CES pricing rule in which markups are constant and a function of the parameter θ only through $p^*(s)/MC(s) \Big|_{s=s_0} = \frac{\theta}{\theta-1}$ for all $s_0 > 0$ which is the second and last statement of the proposition. \square

D.6 Class of preferences with size-decreasing price elasticities

Section 5.3 presents a random growth framework in the spirit of Luttmer (2007) and Arkolakis (2016) with a Kimball aggregator that is analytically tractable and consistent with the observed deviations from Gibrat's law. The key characteristic of the Kimball aggregator is that it allows for a price elasticity that is decreasing in firm size which is the crucial component for generating the size-variance relationship and the positive correlation between markups and size.

Even though it is unclear whether any demand specification with price elasticities that is decreasing in firm size can generate the size-variance relationship as a power law, it does seem to be sufficient to qualitatively generate the stylized facts from sections 4.1 and 5.1. As a result, I will characterize a class of demand functions that generates downward-sloping price elasticities in firm size. To do this, I use the demand system by Arkolakis, Costinot, Donaldson and Rodriguez-Clare (2015a).

There exists a continuum of differentiated goods Ω . A consumer with income M takes prices $\mathbf{p} = \{p_\omega\}_{\omega \in \Omega}$ as given and has a Marshallian demand curve for each differentiated good ω :

$$q_\omega(\mathbf{p}; M) = Q(\mathbf{p}, M) D\left(\frac{p_\omega}{P(\mathbf{p}, M)}\right)$$

By assumption, the aggregate demand shifters $Q = Q(\mathbf{p}, M)$ and $P = P(\mathbf{p}, M)$ are jointly determined in the aggregate as the solution of the following system:

$$\int_{\omega \in \Omega} \left(H\left(\frac{p_\omega}{P}\right)\right)^\beta \left(p_\omega Q D\left(\frac{p_\omega}{P}\right)\right)^{1-\beta} d\omega = M^{1-\beta} \quad (\text{D.1})$$

$$Q^{1-\beta} \left(\int_{\omega \in \Omega} p_\omega Q D\left(\frac{p_\omega}{P}\right) d\omega\right)^\beta = M^\beta \quad (\text{D.2})$$

where $\beta \in \{0, 1\}$, $H'(\cdot) > 0$ and $H''(\cdot) < 0$. In the following, I will restrict my attention to homothetic preferences that are not iso-elastic. Arkolakis, Costinot, Donaldson and Rodriguez-Clare (2015a) show that this is equivalent to $\beta = 1$. Note that this case nests Kimball preferences and Quadratic of Mean Order r (QMOR) expenditure functions (e.g. Feenstra (2014)). Under this restriction, equation D.1 determines the aggregate price level $P(\mathbf{p})$ whereas equation D.2 is the consumer's budget constraint. As a result of homothetic preferences, it can be shown that P is

independent of income M and the demand shifter $Q(\mathbf{p}, M)$ is proportional to it. Thus, it must be the case that:

$$P = P(\mathbf{p})$$

$$Q = MQ(\mathbf{p})$$

The restriction $\beta = 1$ only has consequences for the aggregate demand shifters P and Q . The only restriction that I will make on $D(\cdot)$ is that it features a *finite* choke price which is formalized in the following assumption.

ASSUMPTION D. There exists a finite $a > 0$ such that $D(x) = 0$ for all $x \geq a$.

As mentioned by Arkolakis, Costinot, Donaldson and Rodriguez-Clare (2015a), the normalization of $a = 1$ is without loss of generality. A firm i producing the differentiated good ω is characterized by a constant returns to scale production technology in some input bundle ι with productivity z_i . The input bundle ι is supplied competitively at the rate w . For ease of exposition, I will drop the firm index i and denote $c = w/z$. Furthermore, firms take the aggregate demand shifters $Q = MQ(\mathbf{p})$ and $P = P(\mathbf{p})$ as given. Hence, a firm's optimal pricing decision is characterized by:

$$p^*(c) = \arg \max_{p \geq 0} \left\{ (p - c) D\left(\frac{p}{P}\right) \right\}$$

By standard arguments, the first order condition can be rearranged and expressed as a function of its own price elasticity:

$$\begin{aligned} \frac{p - c}{p} &= - \frac{\partial \ln D\left(\frac{p}{P}\right)}{\partial \ln p} \\ &\equiv \epsilon_D\left(\frac{p}{P}\right) \end{aligned}$$

This is equivalent to the markup identity: $p/c = \epsilon_D\left(\frac{p}{P}\right) / (\epsilon_D\left(\frac{p}{P}\right) - 1)$. Define the transformed

variables $m = p/c$ and $v = P/c$, then the equilibrium markup is characterized as the solution of:

$$m = \frac{\epsilon_D(m/v)}{\epsilon_D(m/v) - 1}$$

In the following, I will provide necessary and sufficient conditions on the curvature of the demand function $D(\cdot)$ that induce more productive firms to charge higher markups. To do this, define the mapping $f(m, v) = m - \frac{\epsilon_D(m/v)}{\epsilon_D(m/v) - 1}$. Optimality implies that $f(m^*(v), v) = 0$ must hold. Straight-forward differentiation with respect to m and v delivers:

$$\begin{aligned} \frac{\partial f(m, v)}{\partial m} &= 1 + \frac{1}{v} \frac{\epsilon'_D(m/v)}{(\epsilon_D(m/v) - 1)^2} \\ \frac{\partial f(m, v)}{\partial v} &= -\frac{m}{v^2} \frac{\epsilon'_D(m/v)}{(\epsilon_D(m/v) - 1)^2} \end{aligned}$$

According to the *Implicit Function Theorem*, more productive firms (i.e. higher values of v) charge higher markups whenever:

$$\begin{aligned} \frac{dm^*(v)}{dv} &= -\frac{\frac{\partial f(m, v)}{\partial v}}{\frac{\partial f(m, v)}{\partial m}} \\ &= \frac{m}{v} \frac{\epsilon'_D(m/v)}{(\epsilon_D(m/v) - 1)^2 v + \epsilon'_D(m/v)} \\ &> 0 \end{aligned}$$

Thus, a necessary and sufficient condition for more productive firms charging higher markups is

$\frac{\epsilon'_D}{(\epsilon_D - 1)^2 v + \epsilon'_D} > 0$. Denote a demand function's *superelasticity* as:

$$\epsilon_S(x) \equiv -\frac{\epsilon'_D(x) \cdot x}{\epsilon_D(x)}$$

Then, the above condition can be rearranged as $\frac{\epsilon_S(m/v)}{\frac{(\epsilon_D(m/v) - 1)^2 m}{\epsilon_D(m/v)} + \epsilon_S(m/v)}$. Recall however that $m = \frac{\epsilon_D(m/v)}{\epsilon_D(m/v) - 1}$ must hold at the optimum. Thus, a necessary and sufficient condition can be expressed in terms of elasticities $\epsilon_D = \epsilon_D(m/v)$ and superelasticities $\epsilon_S = \epsilon_S(m/v)$ only:

$$\frac{\epsilon_S}{\epsilon_S - \epsilon_D + 1} > 0$$

The above inequality directly implies that $\epsilon_S < 0$ is a sufficient condition for more productive firms to charge higher markups as any firm always prices at the elastic portion of its demand curve (i.e. $\epsilon_D > 1$). Whenever $\epsilon_S > 0$ holds, it is easy to observe that the sufficient condition then reduces to $\epsilon_S > \epsilon_D - 1$.