Are firm-level idiosyncratic shocks important for U.S. aggregate volatility?∗

Chen Yeh
University of Chicago

Abstract

This paper assesses the quantitative impact of firm-level idiosyncratic shocks on aggregate volatility in the U.S. economy and evaluates the underlying economic mechanisms of the negative relationship between firm-level volatility and size. I argue that the role of the granular channel in the U.S. economy is fairly limited despite its fat-tailed firm size distribution. Using a novel, comprehensive data set compiled from several sources of the U.S. Census Bureau, I find that the granular component accounts for only a relatively small fraction of the variation in aggregate sales growth which is in stark contrast with previous studies. This result can be rationalized by the size-variance relationship. Firm-level volatility declines at a substantially higher rate in firm size than previously found. Hence, the prominence of the largest firms in terms of aggregate fluctuations reduces drastically. I argue that the explanatory power of granular origins gets cut by about half whenever the size-variance relationship, as estimated in the micro-level data, is taken into account. To explain this deviation from Gibrat’s Law, I construct an analytically tractable framework featuring random growth and a Kimball aggregator. Under this setup, larger firms respond less to productivity shocks as the elasticity of demand is decreasing in its relative market share. Additionally, the model predicts a positive (negative) relationship between firm-level mark-ups (growth) and size. I confirm the predictions of the model by estimating size-varying price elasticities on unique product-level data from the Census of Manufacturers (CM) and structurally estimating mark-ups using plant-level information from the Annual Survey of Manufacturers (ASM).

∗Contact: chenyeh@uchicago.edu. The research in this paper was conducted while the author was a Special Sworn Status researcher of the U.S. Census Bureau at the Chicago Census Research Data Center. Research results and conclusions expressed are those of the author and do not necessarily reflect the views of the Census Bureau. This paper is preliminary and incomplete. The results based on confidential data are censored in the current version as the mandatory process for disclosure is still in progress but expected to finish soon. Do not cite without permission. All remaining errors are purely my own. I would like to thank my committee members Lars Peter Hansen, Chad Syverson and Ufuk Akcigit for invaluable input and support. I thank Cheryl Grim, Frank Limehouse, Javier Miranda and Kirk White for helping me with the U.S. Census Bureau data. For helpful comments and discussions, I thank David Argente, Chanont Banerthangansa, Vasco Carvalho, Bong Geun Choi, Allan Collard-Wexler, Xavier Gabaix, Veronica Guerrieri, Paymon Khorrami, Mary Li, Robert Lucas Jr., Claudia Macaluso, Sara Moreira, Aaron Pancost, Larry Schmidt, Andrea Stella and seminar participants at the University of Chicago.
1 Introduction

The use of exogenous aggregate shocks to explain aggregate fluctuations has a long tradition in macroeconomics. Despite its relative success, our understanding of the underpinnings of aggregate fluctuations is limited. This has lead a recent line of research to emphasize the importance of firm-level idiosyncratic shocks for aggregate volatility. Modern economies feature large firms and their size can be so overwhelming that shocks to these large firms can lead to non-trivial aggregate movements as they no longer cancel each other out (Gabaix (2011)). The empirical support for this “granular” channel however, in particular for the U.S. economy, is scarce.

In this paper, I assess the quantitative role of firm-level idiosyncratic shocks on aggregate volatility in the U.S. economy. Despite a fat-tailed firm size distribution, I provide extensive evidence that the role of granularity is fairly limited in the U.S. economy. The granular channel can explain at most XX percent of the variation in aggregate sales growth. This holds in stark contrast with previous studies that argue that the granular channel can account for approximately one third of aggregate fluctuations.

To get to this result, I use the methodology by di Giovanni, Levchenko and Mejean (2014) to perform a theoretically founded decomposition of firms’ annual sales growth into different components. This allows me to break down aggregate volatility into a macroeconomic, sectoral and firm-specific component. Unlike the results of di Giovanni, Levchenko and Mejean (2014), the firm-specific component does not nearly contribute as much to aggregate sales volatility as the macroeconomic and sectoral components. More importantly however, the decomposition of the firm-specific component indicates that input-output linkages are significantly more important than granular forces.

The underlying identifying assumption is that firms sell to multiple, imperfectly correlated destinations (i.e. countries). Thus, only revenues at the firm-destination level allow for a decomposition into a macroeconomic, sectoral and firm-specific component. As a result, I construct a database consisting of information on the universe of firms and international trade transactions. To achieve this, I merge the Census Bureau’s Longitudinal Business Database (LBD), Standard Statistical Establishment List (SSEL) and Longitudinal Firm Trade Transactions Database (LFTTD) into one novel, comprehensive data set.

In the second section, I focus on what the underlying reason is for the apparent discrepancy between the results found in this paper and the previous literature. I argue that the relatively small role of granularity in the U.S. economy can be rationalized by the negative relationship between firm-level volatility and size, i.e. the size-variance relationship, alone. The intuition behind this 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 substantially.

I estimate this power law with the LBD and implement its estimates in a stylized model with heterogeneous firms subject to firm-specific 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 XX% whenever the size-variance relationship, as observed in the data, is taken into account.

My estimates on the size-variance relationship do 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 (e.g. Jovanovic (1982) and Arkolakis, Papageorgiou and Timoshenko (2015)).

Therefore, I resort to demand-side fundamentals in which a firm’s demand elasticity varies in its size and develop an analytically tractable framework with random growth (Luttmer (2007)) and a Kimball demand aggregator (Kimball (1995)). Under the Kimball aggregator of Klenow and Willis (2016), the elasticity of effective demand is decreasing in a firm’s relative market share. To provide additional evidence in support of this framework, I use the Census of Manufacturers (CM). Following Foster, Haltiwanger and Syverson (2008), I use a unique subset of establishments producing physically homogeneous products. This has two main advantages. First, physical quantities are directly observed. This allows me to construct physical productivity (TFPQ) as opposed to measures of revenue-based productivity (TFPR) which serves as a powerful instrument in dealing with the simultaneity bias confronted in demand estimation. 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. Using this data set, I find that firms’ price elasticities are decreasing in firm size which is consistent with the Kimball specification.

I analytically show that a firm’s mean growth and volatility (i.e. standard deviation of growth) are declining in firm size. For the former, the negative relationship is most present for small firms whereas it flattens out as firms become larger which is consistent with the data. Most importantly however, the Kimball specification is flexible enough to generate the size-variance relationship as a power law. Intuitively, the response of a large firm vis-à-vis a small firm in terms of revenue to a given percentage change in firm-level productivity is dampened as its demand elasticity is smaller under the Kimball aggregator. Lastly, the model predicts that mark-ups are increasing in a firm’s size. While it is intuitive that a firm’s market power (as measured by its mark-up) increases with its size, this feature is not present in most models of firm dynamics with either perfect or monopolistic competition. I verify this prediction of the framework by estimating mark-ups structurally from the Annual Survey of Manufacturers (ASM) using the methodology of de Loecker and Warzynski (2012). I interpret this stylized fact as additional evidence in favor of my framework which is not featured in models that deal with deviations from Gibrat’s Law such as Rossi-Hansberg and Wright (2007), Koren and Tenreyro (2013) and Arkolakis (2016).

CONTRIBUTION TO THE LITERATURE. This paper contributes to a number of strands of literature. First, this paper 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. This granular channel is formalized by Gabaix (2011), but empirical evidence for the U.S. economy is mostly based on Standard and Poor’s CompuStat database which can lead to biased conclusions on firm-level and aggregate volatility as emphasized by Davis et al. (2007). I circumvent these sample selection issues by using the most extensive database imaginable for the U.S. economy based on resources from the U.S. Census Bureau. As a result, I identify the impact of granularity (as discussed by Carvalho and Gabaix (2013) and
Carvalho and Grassi (2015)) on aggregate volatility directly from the data. On the other hand, the results of section 3.3 provide suggestive evidence on the importance of buyer-supplier networks (Acemoglu et al. (2012)) and makes its results on buyer-supplier networks most closely related to di Giovanni, Levchenko and Mejean (2014). Moreover, the results of section 3 are also closely in spirit with Magerman et al. (2015). However, I focus on the importance of the granular channel in a large diversified country such as the U.S. as opposed to countries as France and Belgium. Instead, I find a limited role for the granular channel and emphasize the underlying reason for why this is the case in the U.S. economy.

The magnitude of the granular channel has important implications for understanding business cycles. In section 3, I show that the most dominant contributors to aggregate volatility consist of macroeconomic and sectoral components. In contrast to studies that leave a sizable role to firm-specific shocks, the results of this paper imply that there is considerable scope for stabilization policies that affect all firms in the economy (e.g. fiscal and monetary policies). Furthermore, my findings on the relatively small role of granularity indicate that small and medium-sized firms might need to receive a more prominent role in combatting recessions as emphasized by previous studies (e.g. Fort et al. (2013)). Moreover, my findings have implications on 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: trade openings make large firms more important as 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. as large firms are at the same time considerably less volatile.

Second, I contribute to the literature on the size-variance relationship. 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) and Sutton (2002). However, all of their empirical results are derived from CompuStat. 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 above, this can substantially bias our views on firm-level volatility.

I do not only provide unbiased estimates on the size-variance relationship, but also highlight its aggregate implications as the explanatory power of granularity is cut by more than half compared to the previous literature which assumes constant firm-level volatilities. More importantly however, I attempt to identify the underlying economic mechanism(s) that can drive this relationship. This is important as the impact of recessions can vary greatly across firms. Fort et al. (2013) argue that young and small businesses were especially hit hard during the Great Recession. As a result, it is crucial to understand what causes firms to respond heterogeneously during economic downturns.

This paper also speaks to the literature in firm dynamics. The analytical framework in section 5 is most closely related to the random growth models of Luttmer (2007) and Luttmer (2012). 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. In fact, the estimated standard deviation of firm growth is required to be high in order to approximate Zipf’s Law. Under the setup with Kimball demand, the instantaneous variance of firm growth falls in size and approximately
follows a power law. Lastly, I contribute to the rich literature that deals with deviations from Gibrat’s Law by providing extensive evidence on additional firm-level outcomes that vary with size besides firm-level volatility and growth.\(^1\) The results in section 5 also indicate that firm size plays an important role for price elasticities and mark-ups.

**Overview of this paper.** Section 2 contains a short description of the U.S. Census Bureau’s data sets used in this paper. In section 3, I set up the empirical framework and present the results of the variance decomposition. In section 4, I describe the aggregate implications of the size-variance relationship and in particular its consequences on granularity. Furthermore, I provide extensive empirical evidence on the size-variance relationship as a power law and rule out certain explanations put forward by other studies. Section 5 contains additional stylized facts and presents the analytical framework that can rationalize the previously found deviations from Gibrat’s Law. Concluding remarks can be found in section 6. A detailed description of the used data sets can be found in Appendix A. Appendix B contains proofs and additional derivations of the framework.

## 2 U.S. Census Bureau data

The results of this paper are based on a variety of U.S. Census Bureau data sets. I employ the Longitudinal Business Database (LBD), Standard Statistical Establishment List (SSEL), Longitudinal Firm Trade Transactions Database (LFTTD), Annual Survey of Manufacturers (ASM) and the Census of Manufacturers (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.

**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 includes all geographic areas over the period of 1976 - 2011. Its underlying source is the Business Register (BR) which contains administrative records on U.S. businesses. My main purpose with 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. As a result, each establishment can be connected to some parent firm and thus, it is straightforward to aggregate statistics 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 over government-owned or operated entities.\(^2\)

---

\(^1\)Violations on Gibrat’s Law were 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 which 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)).

\(^2\)More details on the coverage of the LBD and how this database is constructed, can be found in Jarmin and Miranda (2002).
An establishment belongs to a particular firm based on operational control: an establishment’s statistics are included in the parent firm’s activity whenever this parent firm majority owns the establishment. Furthermore, the LBD covers only those firms with at least one employee on payroll over their life-cycle. Lastly, 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 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 section 4 are mainly based on this data set.

**Firm-destination LBD.** The analysis in section 3 decomposes a firm’s revenue growth rate into a firm-level idiosyncratic and aggregate-sectoral component. However as will become clear then, this decomposition is only valid at the firm-destination level. As a result, I need to determine how a firm’s total revenue is decomposed across its destinations. To do this, I exploit export data from the Longitudinal Firm Trade Transactions Database (LFTTD). Thus, a destination is defined at the country level. Fortunately, the LBD and LFTTD share a common firm-level identifier which allows me to link the two databases. This procedure is explained in more detail in Appendix A.6. As a result, 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 - 2011 as total revenues are only available in the LBD from 1994 onwards. This leaves me with the Firm LBD database described above but extended with export status and, most importantly, revenues disaggregated at the country level over the period 1994 - 2011. Section 3 exploits this constructed data base at its full potential.

**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. As a result, I restrict my attention to manufacturing for section 5 by using the Annual Survey of Manufacturers (ASM) and Census of Manufacturers (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. As a result, I can construct productivity measures such as value added or revenue per worker and total factor productivity. Combined with publicly available information from the NBER-CES Manufacturing Database, I can construct proxies for real measures of output. A complete description of used 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 1976 - 2009.

**Product CM.** The Plant ASM/CM data set does not contain quantities and relies on proxies for real measures of output instead. 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, this results in information on physical quantities and prices which will prove to be extremely useful in section 5.1. Most importantly, I am able to 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. Lastly, I restrict myself to only a handful of product categories. In particular, I 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.

3 Variance decomposition of aggregate sales

3.1 Empirical framework

In the following, I set up a simple accounting framework that will serve as the theoretical foundation of the variance decomposition. While the model is simple, it is rich enough to allow for a decomposition of individual growth rates into a firm-level idiosyncratic component and a component that contains aggregate and sectoral forces. It is important to note however that this decomposition is only valid at the firm-destination level. As a result, observing total sales at the firm level is not sufficient. Instead, I require a firm’s sales to be broken down across destinations for identification. This will become clear in the section below. Moreover, firm-level idiosyncratic shocks can be interpreted as a combination of demand and supply shocks through the lens of the model.

**THEORETICAL MOTIVATION.** Aggregate sales in the U.S. are defined as

\[ X_t = \sum_{f \in F_t} x_{ft} = \sum_{(f,n) \in I_t} x_{fnt} \]

where \(x_{fnt}\) denotes the sales of some U.S. firm \(f\) to destination \(n\) in year \(t\). Furthermore, \(F_t\) and \(I_t\) denote the set of active firms and firm-destination pairs in year \(t\) respectively. As noted before, the unit of observation will be a firm-destination pair. By definition, aggregate sales growth is

\[ \gamma_{At} = \frac{X_t}{X_{t-1}} - 1 \]

where I impose that the set of firms are identical in years \(t\) and \(t - 1\).

A firm \(f\) in the U.S. that is active in sector \(j \in \{1, 2, \ldots, J\}\) can sell its product to some country \(n \in \{1, 2, \ldots, N\}\) 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 are allowed to vary across countries and over time. As a result, the representative consumer’s utility in country \(n\) at time \(t\) equals:

\[ U_{nt} = \prod_{j=1}^{J} C_{jnt}^{\phi_j} \]

\[ \text{By construction, this definition of growth only focuses on the intensive margin of growth. In section 3.3, I argue that most of the variation of aggregate sales growth is due to the intensive margin. Furthermore, appendix B.1 discusses the robustness of my results with respect to the growth rate measure as suggested by Davis, Haltiwanger and Schuh (1996) which treats entry and exit symmetrically.} \]
where $\varphi_{jnt}$ is a demand shock at the sectoral level that varies over time. Obviously, the Cobb-Douglas structure implies that expenditure on goods in sector $j$ are 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 $\Omega_{jnt}$. 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 which is characterized by the time-varying unit input requirement $a_{fht}$. Therefore, a firm $f$’s production technology at home is $y_{f}(t) = \frac{1}{a_{fht} t}$ where $t$ 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 mark-up over marginal costs. Conditional on selling to destination $n$, a firm $f$’s revenue at the home destination is then equal to:

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

where $P_{jnt} = \left( \frac{\sum_{f' \in \Omega_{jnt}} \omega_{f'nt}^{1/\theta} c_{f'nt}^{\theta-1}}{\theta-1} \right)^{1/(\theta-1)}$ denotes the ideal price aggregator at the sectoral level in country $n$ and 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 am assuming that all the variation in a firm $f$’s marginal cost across destinations is captured by trade costs alone. These iceberg costs $\kappa_{jnh}$ 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. Whenever growth rates at the firm-destination level $\gamma_{fnt}$ are approximated by log differences, I obtain:

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

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 which are specific to country $n$. More importantly however, $\varepsilon_{fnt} = \Delta \ln \omega_{fnt} + (1 - \theta) \Delta \ln a_{fht}$ 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.

The 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 destina-
tions. As will become clear in the next subsection, demand shocks at the sectoral level can substantially differ across destinations; in particular between domestic and foreign sales. As a result, the proposed decomposition is only valid at the firm-destination level which 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. Thus, in addition to combining the Census Bureau’s LBD and SSEL (e.g. Moreira (2016)), I also use data on the universe of firm-level exports from the LFTTD.

**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 \( \gamma_{At} \) in year \( t \):

\[
\gamma_{At} = \sum_{f,n} \frac{x_{nf_{t-1}}}{X_{t-1}} \gamma_{nt} \\
= \sum_{f,n} \frac{x_{nf_{t-1}}}{X_{t-1}} \delta_{jnt} + \sum_{f,n} \frac{x_{nf_{t-1}}}{X_{t-1}} \varepsilon_{fnt} \\
= \sum_{j,n} \left( \sum_{f \in \Omega_{jnt}} \frac{x_{nf_{t-1}}}{X_{t-1}} \right) \delta_{jnt} + \sum_{f,n} \frac{x_{nf_{t-1}}}{X_{t-1}} \varepsilon_{fnt} \\
\equiv \sum_{j,n} w_{jn-1} \delta_{jnt} + \sum_{f,n} w_{fn-1} \varepsilon_{fnt}
\]

where \( w_{fnt} \) is the share of firm \( f \)’s sales to destination \( n \) in total sales and \( w_{jnt} = \sum_{f \in \Omega_{jnt}} w_{fnt} \) denotes the share of sales by those subset of firms active in sector \( j \) in total sales. In the following, I will 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 to be cross-sectionally and serially correlated. By construction, \( \gamma_{At} \) is stochastic as well. Performing a meaningful variance decomposition on \( \gamma_{At} \) is complicated by \( w_{fnt} \) and \( w_{jnt} \) however as these weights are time-varying. Analyzing the time-series properties of \( \gamma_{At} \) becomes difficult as it is hard to disentangle movements over time by the shocks from the 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 \( \gamma_{At} \) with the exception of the weights that are fixed at their \( t = \tau \) value. Given some value for \( \tau \), the weights \( w_{jnt} \) and \( w_{fnt} \) become predetermined variables which simplifies the analysis considerably. Note that \( \gamma_{At|\tau} \) explicitly depends on calendar time \( t \) as the weights from period \( \tau \) are combined with the realization of shocks in period \( t \). By construction, we have that \( \gamma_{At|\tau} = \gamma_{At} \).

---

4To see this, note that a firm’s total revenue growth does not allow for a decomposition that is loglinear. Under a log differences approximation, a firm’s total revenue growth would be equal to \( \gamma_{ft} = \ln \left( \sum_{n} x_{fnt} \right) - \ln \left( \sum_{n} x_{fnt-1} \right) \). 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 \( x_{fnt} \) would have to be independent of \( n \).
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:

$$
\sigma_{A\tau}^2 = \sigma_{M\tau}^2 + \sigma_{F\tau}^2 + COV_{\tau} \quad \text{with}
$$

$$
\sigma_{M\tau}^2 = V \left( \sum_{j,n} w_{jn\tau-1} \delta_{jnt} \right),
$$

$$
\sigma_{F\tau}^2 = V \left( \sum_{f,n} w_{f\tau-1} \varepsilon_{fnt} \right),
$$

$$
COV_{\tau} = cov \left( \sum_{j,n} w_{jn\tau-1} \delta_{jnt}, \sum_{f,n} w_{f\tau-1} \varepsilon_{fnt} \right).
$$

In the following, I will provide estimates for $\sigma_{A\tau}^2$ and its components for each $\tau \in \{1995, \ldots, 2011\}$. 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 aggregate sales growth in year $\tau$. Furthermore, I am particularly interested in their averages across $\tau$ and how the average of the firm-specific component relates to the average of the overall component, i.e. $\left( \frac{1}{T} \sum_{\tau=1995}^{2011} \sigma_{F\tau} \right) / \left( \frac{1}{T} \sum_{\tau=1995}^{2011} \sigma_{A\tau} \right)$. While $COV_{\tau}$ must be included in the decomposition for $\sigma_{A\tau}$, I will rarely report results for this component as this term is quantitatively negligible.

**Estimation.** The theoretical framework above implies that firm-destination level growth rates can be decomposed into a macro-sectoral and firm-specific component. To obtain these components, I adopt the methodology by Stockman (1988) in which firm-destination level growth rates are regressed on a constant and a set of sector-level dummies.\(^5\) Its fitted values form the macro-sectoral component while its residual will be interpreted as the firm-specific shock. More formally, I follow the regression specification:

$$
\gamma_{fnt} = c_n + \chi'X_{ft} + \varepsilon_{fnt}
$$

where $X_{ft}$ contains a set of sectoral dummies. Thus, shocks are constructed as:

$$
\hat{\delta}_{jnt} = \hat{c}_n + \hat{\chi}'X_{ft}
$$

$$
\hat{\varepsilon}_{fnt} = \gamma_{fnt} - \hat{\delta}_{jnt}
$$

Recall that the proposed decomposition is only valid at the firm-destination level. Hence, I perform the above regression for each destination-year pair separately which follows Koren and Tenreyro (2007). Once the shocks are constructed, it is straightforward to estimate synthetic growth rates and its components as

---

\(^5\)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 am assuming 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.
the weights \( w_{jn} \) and \( w_{fn} \) can be taken straight from the data. Lastly, the estimate for \( \sigma_{A\tau}^2 \) consists of the sample variance of \( \hat{\gamma}_{A\tau} \):

\[
\hat{\sigma}_{A\tau}^2 = \frac{1}{T-1} \sum_{\tau=1995}^{2011} \left( \hat{\gamma}_{A\tau} - \frac{1}{T} \sum_{\tau=1995}^{2011} \hat{\gamma}_{A\tau} \right)^2
\]

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

**AGGREGATE CONSISTENCY.** In this section, I will focus on some key patterns in the data. First, I show that the aggregate variables constructed from U.S. micro-level data exhibit similar patterns as those in official public sources. Recall that the underlying sources of the estimation sample consist of the LBD, SSEL and LFTTD. In section 2, I pointed out that the merge rates between the LBD and SSEL average to roughly XX% across years. Similarly, I am able to merge about XX% of the amount of firm-level transactions in the LFTTD export data.\(^6\) While these match rates are relatively high, potential concerns could arise due to the fact that aggregate variables are nevertheless poorly tracked.

![Figure around here](image1.png)

(a) Aggregate sales growth - total sales

![Figure around here](image2.png)

(b) Aggregate sales growth - foreign sales

Figure 1: The left graph compares the aggregate sales growth rate as constructed from U.S. Census Bureau micro-level data with official statistics on GDP and gross output from the BEA. The graph on the right compares the aggregate sales growth on exports with official statistics on exports by IMF’s International Financial Statistics and Census’ Foreign Trade Division.

As a result, I compare the growth rate of aggregate sales as constructed from the micro-level data with official growth rate statistics on gross output and GDP from the BEA.\(^7\) Figure 1a shows that the micro-level data tracks aggregate activity quite well as aggregate sales growth follows both time-series from BEA sources. The correlation coefficient between the time-series constructed from the micro-data and gross output from

\(^6\)This percentage for the LFTTD can also be found in Bernard, Jensen and Schott (2009a) which alleviates concerns on potential biases due to non-matched observations.

\(^7\)These statistics are taken from the BEA's GDP-by-Industry data.
the BEA is 0.\text{xxxx}. To confirm the validity of the merging process between the LBD-SSEL and LFTTD, I construct the growth rate of aggregate export sales from the micro-level data and compare it with official statistics from the Census’ Foreign Trade Division and the International Monetary Fund’s (IMF) International Financial Statistics (IFS). As can be observed from figure 1b, the micro-level data also tracks aggregate export activity well. Thus, the small set of unmatched observations does not seem to have substantial aggregate implications.

**SUMMARY STATISTICS.** Table 1 displays some summary statistics for 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 section 4.

<table>
<thead>
<tr>
<th>Table I. Summary statistics of estimation sample.\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AVERAGE OF GROWTH</strong></td>
</tr>
<tr>
<td>Aggregate sales</td>
</tr>
<tr>
<td>Firm-destination level sales</td>
</tr>
<tr>
<td><strong>STANDARD DEVIATION OF SALES GROWTH RATE</strong></td>
</tr>
<tr>
<td>Whole sample (average)</td>
</tr>
<tr>
<td>0 - 20 size percentile</td>
</tr>
<tr>
<td>21 - 40 size percentile</td>
</tr>
<tr>
<td>41 - 60 size percentile</td>
</tr>
<tr>
<td>61 - 80 size percentile</td>
</tr>
<tr>
<td>81 - 100 size percentile</td>
</tr>
<tr>
<td><strong>AVERAGE OF HERFINDAHL INDEX</strong></td>
</tr>
<tr>
<td>Firm-destination $\sqrt{H(f, n)}$</td>
</tr>
<tr>
<td>Firm $\sqrt{H(f)}$</td>
</tr>
<tr>
<td>Industry (SIC2) $\sqrt{H(j)}$</td>
</tr>
</tbody>
</table>

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

First, the average growth of aggregate sales is smaller than the unweighted mean of the individual firm-destination level growth rates. This is not surprising as smaller firms, conditional on survival, tend to grow at higher rates than larger firms. Second, I report averages of firm-destination level sales volatility, as measured by its standard deviation, by size quintile. It can be observed that smaller firms are more volatile than their larger sized counterparts.\textsuperscript{8} Lastly, the table shows the time averages of Herfindahl indices defined at the firm-destination and firm level. These are in the same order of magnitude as reported in Gabaix (2011). 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 implied by low Herfindahl indices, their contribution to

\textsuperscript{8}As mentioned by Bernard et al. (2009b), exporting activities in the U.S. are primarily concentrated at the largest firms who might expose themselves to more volatile markets. This might explain the results for the 81 - 100 size percentile.
aggregate volatility declines as well. Thus, these summary statistics already suggest that the U.S. economy is less granular than previously conjectured. The following section will show this result in a more rigorous quantitative fashion.

### 3.3 Results

Table 2 shows the summary statistics of each growth rate component as implied by the decomposition of section 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 seems to be dominating most of the variation of 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. This implies that the role of aggregate or sector-level shocks could be important.

<table>
<thead>
<tr>
<th>Total Sales</th>
<th>Observations</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual $\gamma_{fnt}$</td>
<td>$\approx XX \cdot 10^6$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Firm-specific $\hat{\varepsilon}_{fnt}$</td>
<td>$\approx XX \cdot 10^6$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Sector-destination $\hat{\delta}_{jnt}$</td>
<td>$\approx XXXX$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domestic Sales</th>
<th>Observations</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual $\gamma_{fnt}$</td>
<td>$\approx XX \cdot 10^6$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Firm-specific $\hat{\varepsilon}_{fnt}$</td>
<td>$\approx XX \cdot 10^6$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Sector-destination $\hat{\delta}_{jnt}$</td>
<td>$\approx XXXX$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Foreign Sales</th>
<th>Observations</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual $\gamma_{fnt}$</td>
<td>$\approx XX \cdot 10^6$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Firm-specific $\hat{\varepsilon}_{fnt}$</td>
<td>$\approx XX \cdot 10^6$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Sector-destination $\hat{\delta}_{jnt}$</td>
<td>$\approx XXXX$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
</tbody>
</table>

The 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.

However, the table shows that these shocks are particularly volatile for export sales. This 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 2 indicate that this does matter as the average standard deviation of the macro-sectoral component is substantially different between domestic and foreign sales.

The empirical framework implies that firm-destination level specific shocks $\varepsilon_{fnt}$ consist of a demand and supply component. In particular, I constructed these shocks as:

$$\varepsilon_{fnt} = \Delta \ln \omega_{fnt} + (1 - \theta) \Delta \ln (a_{fdt})$$
Note that 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 \( \varepsilon_{fnt} \) 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 proxied by the difference between \( \varepsilon_{fnt} \) and the previously constructed destination-level average.

These measures are proxies at best as the proxy for supply shocks is also capturing demand shocks that are common across a firm’s destinations. Table 3 reports the summary statistics of the proxies to these supply and demand shocks. It is then clear that most of the variation for firm-specific shocks are coming from shocks that are destination-specific: the average standard deviation of \( \varepsilon_{fnt}^1 \) is more than \( XX \) percent larger than the average standard deviation of \( \varepsilon_{fnt}^2 \).

Table III. Summary statistics of firm-specific components.\(^c\)

<table>
<thead>
<tr>
<th></th>
<th>Observations</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm-destination</td>
<td>( \varepsilon_{fnt}^1 )</td>
<td>( \approx XX \cdot 10^6 )</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Firm-common</td>
<td>( \varepsilon_{fnt}^2 )</td>
<td>( \approx XXXXXX )</td>
<td>0.XXXX</td>
</tr>
</tbody>
</table>

\(^c\)The 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.

Thus, this is suggestive evidence of the importance of demand-side fundamentals which supports the view on the size-variance relationship laid down in sections 5.1 and 5.2.\(^9\)

**DECOMPOSITION OF AGGREGATE VOLATILITY.** Figure 2 displays the estimates \( \hat{\sigma}_{AR} \) (black), \( \hat{\sigma}_{MR} \) (red) and \( \hat{\sigma}_{Fr} \) (blue) for each \( \tau \in \{ 1995, \ldots, 2011 \} \). Table 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 \( \frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{Fr}}{\sigma_{Ar}} \) and \( \frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{MR}}{\sigma_{Ar}} \) respectively. Most importantly, 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 \( XX.XX \% \) over \( XX.XX \% \)). Furthermore, the firm-specific component is much smaller than previously found. Previous estimates average to approximately 80% whereas I only find a contribution of \( XX.XX \% \) which consists of a decrease by more than \( XX \% \). This is already suggestive of the fact that granularity plays a much smaller role in the U.S. economy than previously conjectured.

Even though the firm-specific component contributes relatively less than the macro-sectoral component, the standard deviation of the firm-specific component does seem to comove over time with the standard deviation of aggregate sales. This is less clear for the standard deviation of the macro-sectoral component.

\(^9\)Note that the estimation sample for \( \varepsilon_{fnt}^1 \) and \( \varepsilon_{fnt}^2 \) in table 3 is substantially smaller than in table 2. This is because I only focus on those firms that serve at least two different destinations when estimating \( \varepsilon_{fnt}^1 \) and \( \varepsilon_{fnt}^2 \).
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.

Table IV. Relative contribution of firm-specific and macro-sectoral components.\(^d\)

<table>
<thead>
<tr>
<th>Component</th>
<th>SD</th>
<th>Relative SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate (\hat{\sigma}_A)</td>
<td>0.XXXX</td>
<td>1.0000</td>
</tr>
<tr>
<td>Firm-specific (\hat{\sigma}_F)</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>Macro-sectoral (\hat{\sigma}_M)</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
</tr>
</tbody>
</table>

\(^d\)The column “SD” represents the average values of the aggregate, firm-specific and macro-sectoral components over the sample period, i.e. I report \(\frac{1}{T} \sum_{\tau=1995}^{2011} \hat{\sigma}_A\), \(\frac{1}{T} \sum_{\tau=1995}^{2011} \hat{\sigma}_F\) and \(\frac{1}{T} \sum_{\tau=1995}^{2011} \hat{\sigma}_M\) respectively. The column “Relative SD” reports the average value of the relative standard deviations, i.e. \(\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\hat{\sigma}_F}{\hat{\sigma}_A}\) and \(\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\hat{\sigma}_M}{\hat{\sigma}_A}\).

**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. More importantly though, the firm-specific component does not necessarily only represent the granular mechanism as brought forth by
Gabaix (2011). This is clarified by decomposing the firm-specific term as:

\[
\sigma^2_{F_{\tau}} = V \left( \sum_{f,n} w_{fn_{\tau-1}} \varepsilon_{fnt} \right) = \sum_{g,m} \sum_{f,n} w_{gmn_{\tau-1}} w_{fn_{\tau-1}} \text{cov} (\varepsilon_{gmt}, \varepsilon_{fnt})
\]

\[
= \sum_{f,n} w_{fn_{\tau-1}}^2 V(\varepsilon_{fnt}) + \sum_{(g\neq f)(m\neq n)} \sum_{f,n} w_{gmn_{\tau-1}} w_{fn_{\tau-1}} \text{cov} (\varepsilon_{gmt}, \varepsilon_{fnt})
\]

This decomposition can be found in Carvalho and Gabaix (2013) and di Giovanni, Levchenko and Mejean (2014) and it identifies two important components within the firm-specific term. More specifically, the contribution of the individual variances represents the direct effect of firm-level idiosyncratic shocks under an environment in which firms are not connected, i.e. there are no firm-to-firm linkages. This is the key idea of Gabaix’ (2011) concept of granularity. In the absence of firm networks, 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 and steers us away of the conventional logic behind the Law of Large Numbers.

It is not surprising that this term captures granular forces as it is the variance of Gabaix’ (2011) granular residual whenever the firm-specific shocks \(\varepsilon_{fnt}\) are distributed independently from each other. Hence, I will denote its component by \(\text{GRAN}_{\tau} = \sum_{f,n} w_{fn_{\tau-1}}^2 V(\varepsilon_{fnt})\).

Lastly, there is the component containing movements between firms. Traditionally, movements between firms were interpreted as by-products of aggregate or industry-level shocks. However, there is a large and growing literature that movements 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.

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. As a result, 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 the following section, I will provide suggestive evidence of why these covariances can nevertheless be interpreted as networks. In the mean time, I denote this component as

\[\text{GRAN}_{\tau} = \sum_{f,n} w_{fn_{\tau-1}}^2 V(\varepsilon_{fnt}).\]

---

<sup>10</sup>Gabaix (2011) shows that aggregate volatility declines at the rate \(1/N(\zeta - 1)/\zeta\) in the number of firms \(N\) where \(\zeta\) 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)).
\[
\text{LINK}_\tau = \sum_{(g \neq f) \vee (m \neq n)} \sum_{f,n} w_{gmt} w_{fmt-1} \text{COV}(\varepsilon_{gmt}, \varepsilon_{fmt}).
\]

Figure 3: Disentangling granular and network components in the firm-specific term \(\sigma^2_{F\tau}\).

Table V. Relative contribution of granular and network components.

<table>
<thead>
<tr>
<th>Volatility</th>
<th>Relative SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\sigma}_{F\tau})</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>(\text{GRAN}_{\tau}) ((\cdot\cdot))</td>
<td>0.XXXX</td>
</tr>
<tr>
<td>(\text{LINK}_{\tau}) ((\cdot\cdot))</td>
<td>0.XXXX</td>
</tr>
</tbody>
</table>

The results in figure 3 and table 5 depict a clear picture. The vast majority of the firm-specific component \(\hat{\sigma}_{F\tau}\) is dominated by input-output linkages as its average relative contribution \(\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\text{LINK}_\tau}{\hat{\sigma}_{F\tau}}\) averages at XX.XX%. Furthermore, it seems to comove almost perfectly with total firm-specific volatility. While the relative contribution of the granular component \(\hat{\text{GRAN}}_{\tau}\) is much smaller, it is still non-negligible. Its relative contribution over the whole period averages at XX.XX%. 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 \(\frac{1}{100} \times (XX.XX \times XX.XX) = XX.XX\%\) which is significantly less than found by Gabaix (2011) and di Giovanni and Levchenko (2012).11

11Alternatively, I could have calculated the average of the relative granular contribution. This amounts to \(\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\text{GRAN}_{\tau}}{\hat{\sigma}_{A\tau}} = XX.XX\%\) which is almost identical to the back-of-the-envelope calculation.
IDENTIFYING INPUT-OUTPUT LINKAGES IN THE U.S. ECONOMY. It is clear from the previous decomposition that firm-specific volatility is dominated by the component containing comovements between firms. 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 argue though 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.

\[ \text{LINK}_{ijr} = \sum_{f \in I_{ir}, n} \sum_{g \in I_{jr}, m} w_{fnr-1} w_{gmr-1} \text{cov}(\varepsilon_{fnt}, \varepsilon_{gmt}) \]

where \( I_{ir} \) denotes the set of firms that are active in industry \( i \) at time \( r \). To create a measure of input-output linkages, an industry pair \((i, j)\)’s “mean network intensity” is defined as:

\[ \text{IO}_{ijr} = \frac{1}{2} [(1 - \lambda_{ir}) \rho_{ijr} + (1 - \lambda_{jr}) \rho_{jir}] \]

Figure 4: Identifying input-output linkages in the U.S. economy.

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:

\[ \text{LINK}_{ijr} = \sum_{f \in I_{ir}, n} \sum_{g \in I_{jr}, m} w_{fnr-1} w_{gmr-1} \text{cov}(\varepsilon_{fnt}, \varepsilon_{gmt}) \]
where $1 - \lambda_{i\tau}$ denotes the cost share of intermediate inputs in industry $i$ at time $\tau$. 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. 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. Appendix B.1 describes extensively how this is done.

Lastly, I plot $\text{LINK}_{ij} \equiv \frac{1}{T} \sum_{\tau=1995}^{2011} \text{LINK}_{ij\tau}$ and $\text{IO}_{ij} \equiv \frac{1}{T} \sum_{\tau=1995}^{2011} \text{IO}_{ij\tau}$ against each other. The result can be found in figure 4. [under construction]

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. However, there is some evidence (e.g. Fort et al. (2013)) 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 as this component should not only include firm-level idiosyncratic shocks but also heterogeneous responses to aggregate and sectoral shocks that differ by some firm-specific characteristic.

Table VI. Relative contributions under various robustness specifications.$^f$

<table>
<thead>
<tr>
<th></th>
<th>Firm-specific</th>
<th>Macro-sectoral</th>
<th>GRAN</th>
<th>LINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_{ft}$</td>
<td>0.XXXX</td>
<td>0.XXXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
</tr>
<tr>
<td>Size</td>
<td>0.XXXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
</tr>
<tr>
<td>Age</td>
<td>0.XXXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
</tr>
<tr>
<td>Trade openness</td>
<td>0.XXXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
</tr>
<tr>
<td>Local shocks</td>
<td>0.XXXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
<td>0.XXX</td>
</tr>
</tbody>
</table>

$^f$The columns “Firm-specific” and “Macro-sectoral” represent the average values of the firm-specific and macro-sectoral components over the sample period, i.e. I report $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{F\tau}}{\sigma_{A\tau}}$ and $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{M\tau}}{\sigma_{A\tau}}$ respectively. The columns “GRAN” and “LINK” report the average values of the standard deviations relative to the firm-specific component, i.e. $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{GRAN\tau}}{\sigma_{F\tau}}$ and $\frac{1}{T} \sum_{\tau=1995}^{2011} \frac{\sigma_{LINK\tau}}{\sigma_{F\tau}}$.

While di Giovanni and Levchenko (2012) argue that the latter channel is quantitatively minor, I will nevertheless 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 section 3.1 as follows:

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

$^{12}$See Appendix E of di Giovanni, Levchenko and Mejean (2014) for a framework that can motivate this measure.

$^{13}$This includes the majority of frameworks that build upon Dixit and Stiglitz (1977), Krugman (1980) and Melitz (2003).
where \( z_{ft} \) denotes some firm-specific characteristic that is allowed to vary over time. This implies that the augmented firm-specific component equals \( \tilde{\varepsilon}_{fnt} \equiv \delta_{jnt} \times z_{ft} + \beta z_{ft} + \varepsilon_{fnt} \).

The firm-characteristic \( z_{ft} \) either comprises of firm size (as defined by a sales quintile dummy), firm age (dummy for whether firm is younger than 5 years) or trade openness. The latter is a dummy that equals 1 whenever a firm’s ratio of export sales to total sales exceeds 10 percent. The results of the augmented specification are displayed in table 6 which shows that the main results are unaffected. While the contribution of the firm-specific component increases most whenever aggregate and sectoral shocks are subject to heterogeneous responses in firm size, the relative contribution of the granular component decreases as well.

For age and trade openness, the results of section 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 significantly 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. To be exact, this is a region-specific shock that affects all firms located within a specific geographical area. Thus, a firm-destination level growth rate gets decomposed as:

\[
\gamma_{fnt} = \delta_{jnt} + \alpha_{lt} + \varepsilon_{fnt}
\]

In this specification, regions are defined at the county level. The last row of table 6 shows that the main results are also unaffected under this specification.

4 Aggregate implications of the size-variance relationship

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 previous evidence as other studies found that the granular channel can explain approximately one third of observed volatility. In the following, I will suggest a simple mechanism that can bridge this apparent discrepancy, i.e. the size-variance relationship. First, I estimate the negative relationship between firm-level volatility and size. Second, I shed some light on the potential economic mechanism(s) that can drive this relationship. Lastly, I show that the aggregate implications of the size-variance relationship are sizable. Most importantly, this relationship alone is able to rationalize the findings of section 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 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)). To assuage any concerns on the robustness of my results, I also report the results using revenue data. These can be found in Appendix B.2.

While Hymer and Pashigian (1962) already established 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 $\alpha$ under several regression specifications can be found in table 7.

<table>
<thead>
<tr>
<th>SIZE $\hat{\alpha}$</th>
<th>EMP</th>
<th>EMP</th>
<th>EMP</th>
<th>EMP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.xxxx</td>
<td>-0.xxxx</td>
<td>-0.xxxx</td>
<td>-0.xxxx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>FIRM</th>
<th>INDUSTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ESTIMATION SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

9Firm 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.

This table shows the estimates of $\alpha$ 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. 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 $\alpha$ is equal to -0.xxxx. 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.

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 $\hat{\alpha} = -0.\text{xxxx}$. An identical picture is displayed when I control for industry-level fixed effects ($\hat{\alpha} = -0.\text{xxxx}$) or by including both firm- and industry-level fixed effects ($\hat{\alpha} = -0.\text{xxxx}$). 14 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.

### 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 8 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 in 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 8 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) is also indirectly ruled out by this procedure.

---

14 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 $\hat{\alpha}$ amounts to -0.\text{xxxx}. Details on this estimation procedure can be found in Appendix B.2.
Table VIII. Potential explanations for the size-variance relationship.\textsuperscript{h}

<table>
<thead>
<tr>
<th></th>
<th>EMP</th>
<th>EMP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OUTPUT DIVERSIFICATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
</tr>
<tr>
<td>Establishments</td>
<td>0.XXX</td>
<td>0.XXX</td>
</tr>
<tr>
<td><strong>PRODUCT DIVERSIFICATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
</tr>
<tr>
<td>Products (SIC6)</td>
<td>0.XXX</td>
<td>0.XXX</td>
</tr>
<tr>
<td><strong>FIRM AGE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.XXX</td>
<td>-0.XXX</td>
</tr>
<tr>
<td><strong>COHORT EFFECTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.XXX</td>
<td>-0.XXX</td>
</tr>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Estimation Sample</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{h}Firm size is defined as employment count. Standard errors are clustered at the industry (3-digit NAICS) level. Each row 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.

It has been suggested that firms do not diversify across establishments in terms of output 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).\textsuperscript{15} 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. 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 7 indicates that the size-variance relationship is basically unchanged.\textsuperscript{16} 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 when controlling for firm age. However, this does not seem to be case as the third category of results in table 8 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. Once more, it appears that

\textsuperscript{15}These are the product innovation frameworks build upon Klette and Kortum (2004), e.g. Lentz and Mortensen (2008), Seker (2012) and Akcigit and Kerr (2015).

\textsuperscript{16}A similar set of robustness tests (as for the output diversification channel) can be found in Appendix B.2.
the relationship between firm-level volatility and size is robust. Lastly, I check whether the size-variance relationship displays a different regime for larger firms by means of a Chow (1960) test. The results can be found in Appendix B.2. I find no evidence for such a structural break.  

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 18 percent. This is substantially smaller than the quantitative findings in the literature. 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.

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 $\zeta$ is the Pareto right tail of the firm size distribution. Obviously, the rate of convergence becomes slower and slower as $\zeta \to 1$. Even though Gabaix (2011) formalizes the case in which firm-level volatility falls in size at the rate $\alpha$, 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 $\alpha$ (i.e. $\alpha \neq 0$) as aggregate volatility behaves asymptotically as $1/N^{\zeta'}$ where $\zeta' = \min\{(\zeta - 1 + \alpha)/\zeta, 1/2\}$. However, this result is not sufficient to determine the quantitative decline of aggregate volatility in a granular economy as $\alpha$ 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 $\sigma$ and its size $S$ in reduced form:

$$\mathbb{P}(S > x) \propto x^{-\zeta}$$

$$\sigma(S) = AS^{-\alpha}$$

---

17 This is potentially important as the coefficient $\alpha$ 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. I thank Vasco Carvalho for pointing this out.

18 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)).

19 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.
Note that any framework with firm-specific idiosyncratic shocks alone that delivers these relationships in reduced form is sufficient for my quantitative exercise. Appendix C lays down the explicit details of such a model. In this framework, it can be shown that aggregate volatility, which is defined as the standard deviation of aggregate sales growth $\gamma_A$, can be calculated as:

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

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).

Figure 5: The volatility of the simulated granular economy $\text{SD}(\gamma_A; \alpha)$ relative to observed volatility in the U.S. is displayed for several values of $-\alpha$. Observed volatility $\text{SD}_{\text{obs}}$ 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 $\alpha$ is set equal to either the upper or lower bound of estimates in table 7. The green line displays the explanatory power of granularity under the previous estimate of $\alpha$ by Stanley et al. (1996).

Under his baseline scenario, 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. Even though the average volatility of the top 100 firms in Compustat was previously considered to

---

20This 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.

21In this particular case, aggregate volatility is equal to the product of $\bar{\sigma}$, which is set equal to 0.12, and the square root of the economy’s sales Herfindahl.
be a conservative choice, I show that for sufficiently high values of $\alpha$, 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 5. The figure displays the ratio of the volatility of the simulated economy and observed volatility in the U.S. for a range of $\alpha$. Furthermore, 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 in line with what is empirically observed.\textsuperscript{22} 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 XX\% whenever the size-variance relationship, as observed in the data, is taken into account. Whenever I use the most conservative estimate for $\alpha$, i.e. $\hat{\alpha} = -0.xxxx$, the granular channel only accounts for approximately XX.X\% of observed aggregate volatility. On the other hand, the role of granularity is substantially smaller whenever I use the lower bound estimate for $\alpha$. Table 7 indicates that this is $\hat{\alpha} = -0.XXXX$. Under this scenario, the contribution of granularity to aggregate observed volatility reduces to about X.X\% 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. Whenever I follow the calibration strategy of most granularity studies, the framework predicts that the granular channel can approximately generate 35.3\% of the observed volatility in the U.S. as consistent with findings by Gabaix (2011), di Giovanni, Levchenko and Mejean (2014) and Carvalho and Grassi (2015). 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. Thus, I conclude that the relatively small role of the granular channel can be solely rationalized with the size-variance relationship.

\textsuperscript{22}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., equals about X.X\%. 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. I thank Xavier Gabaix for mentioning this issue.
5 A microfoundation for the size-variance relationship

5.1 Stylized facts

PRICE ELASTICITY. The results from section 4.1 demonstrate a clear deviation from Gibrat’s Law. More importantly, there seems to be little evidence for the conjectures that are commonly thought to drive the size-variance relationship. Most of these explanations are driven by supply-side fundamentals. However, it is well possible that larger firms respond in a less volatile manner to productivity shocks as their effective demand elasticity is simply smaller. This requires a price elasticity of demand that is varying in a firm’s size. Most contributions in the literature are unable to verify this directly in the data though 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 Manufacturers (CM) as in Foster, Haltiwanger and Syverson (2008).23 This has two main advantages. First, physical quantities are directly observed. Combined with revenue data, this allows me to construct (average) prices. 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). 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 a strong 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 firm size. To test this hypothesis, I turn to the following empirical specification which will be estimated separately for each product:

\[ \ln q_{it} = \sigma_0 + \sigma_1 \ln p_{it} + \sigma_2 \ln p_{it} \times d_{it} + \sigma_3 d_{it} + \sum_t \alpha_t \text{YEAR}_t + \nu_{it} \]

where \( d_{it} \) denotes a dummy that is equal to 1 whenever a plant \( i \)’s size (as measured by employee count) in year \( t \) crosses a certain threshold. This threshold will either be determined by some percentile in the size distribution or by some number of employees as will become clear below. I also include a set of year fixed effects to control for economy-wide shifts in the demand curve. This specification is then estimated separately for each product group. These products include concrete, boxes, bread and ice.

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

23The reader is referred to Appendix A.5 for details on this data set.
24Note 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.
\(\sigma_1 + \sigma_2 < -1\). To deal with the simultaneity bias, I perform a two-stage least squares (2SLS) exercise in which the set of endogenous variables \((\ln p_{it}, \ln p_{it} \times d_{it})\) is instrumented with \((\ln TFPQ_{it}, \ln TFPQ_{it} \times d_{it})\). The results of this procedure are displayed in table 9.

<table>
<thead>
<tr>
<th>CONCRETE</th>
<th>OLS</th>
<th>IV</th>
<th>IV - SMALL</th>
<th>IV - LARGE</th>
<th>IV - ( d_{it}^p )</th>
<th>IV - ( d_{it}^{TE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{it} )</td>
<td>-XXXX</td>
<td>-XXXX</td>
<td>-XXXX</td>
<td>-XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>( p_{it} \times d_{it} )</td>
<td>-XXXX</td>
<td>-XXXX</td>
<td>-XXXX</td>
<td>-XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table IX. Estimation results of size-varying price elasticities.\(^i\)

\(^i\)The 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 two columns, the results of the unrestricted IV regression are displayed. The column “IV - \( d_{it}^p \)” displays the estimated coefficients when the interaction dummy equals 1 whenever a plant’s size exceeds the 75th percentile. Similarly, the column “IV - \( d_{it}^{TE} \)” displays the estimated coefficients whenever the dummy threshold is set to 50 employees. Standard errors are clustered at the plant level.

The first two columns are almost identical to Foster, Haltiwanger and Syverson (2008) and serve as a validity test only. For concrete, they find estimates of \(-X.XX\) (OLS) and \(-X.XX\) (IV) which are extremely similar 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 the three other product categories. The results for boxes, bread and ice display the same picture qualitatively. The results can be found in table B.3 in Appendix B.3.

**MARK-UPS.** The first set of results in this section showed that a firm’s price elasticity is decreasing in firm size. As a result, it can be conjectured that firm-level mark-ups are increasing in firm size. To see this result more formally, consider the profit maximization problem of any firm with some price-setting power and an arbitrary differentiable demand \(\varphi(P)\) and cost function \(c(Q)\) where \(Q = \varphi(P)\).\(^{25}\) Furthermore, I assume that firms take factor prices \(v \geq 0\) as given. Thus, its profit function is defined as:

\[
\Pi(v) = \max_{P \geq 0} \{P \cdot \varphi(P) - c(\varphi(P); v)\}
\]

\(^{25}\)Strictly speaking, I will be assuming a set of regularity conditions on \(\varphi(\cdot)\) and \(c(\cdot)\) such that first-order conditions are sufficient to solve a firm’s profit maximization problem.
Its first-order condition with respect to price $P$ is easily rearranged to $P \left( \frac{\varphi(P)}{\varphi'(P)} + 1 \right) = c'(\varphi(P))$. Define the price elasticity function as $\tilde{\sigma}(P) = \varphi'(P) \cdot P/\varphi(P)$ which immediately implies that the previous first-order condition is equal to:

$$
\mu(P) = \frac{P}{c'(\varphi(P))} = \frac{\tilde{\sigma}(P)}{\tilde{\sigma}(P) - 1}
$$

It is straightforward to see that a firm’s mark-up $\mu$ is decreasing in the price elasticity $\tilde{\sigma}$. Given the previous results on a subset of homogeneous products in the manufacturing sector in which I found that $\tilde{\sigma}$ and firm size are negatively related, this implies that mark-ups are increasing in firm size. This result is intuitive as larger firms, by definition, have larger market shares. As a result, their market power as reflected by mark-ups should increase as well.

However, due to data restrictions, it is unclear whether price elasticities are decreasing in firm size for a broader range of industries. I will show though that, despite not knowing price elasticities, mark-ups are increasing in firm size for a extremely broad range of manufacturing industries. To do this, I will estimate mark-ups using two methods.

Under the first approach, a firm’s production technology is assumed to be Cobb-Douglas. More precisely, a firm’s gross output technology is given by:

$$
Q = \exp(\omega)K^{\beta_k}L^{\beta_l}M^{\beta_m}
$$

where $\omega$ denotes a Hicks-neutral productivity shock and $M$ denotes intermediate inputs. Note that I am only imposing that $\beta_k + \beta_l + \beta_m > 0$, thus I am not necessarily restricted to specifications with constant returns to scale. By assumption of the Cobb-Douglas specification, a firm’s total cost function is equal to:

$$
c(Q) = Q^{1/\beta} \left[ \beta \left( \frac{r}{\beta_k} \right)^{\beta_k/\beta} \left( \frac{w}{\beta_l} \right)^{\beta_l/\beta} \left( \frac{p_m}{\beta_m} \right)^{\beta_m/\beta} \left( \frac{1}{\exp(\omega)} \right)^{1/\beta} \right]
$$

where, with some abuse of notation, $\beta \equiv \beta_k + \beta_l + \beta_m$. A firm’s marginal cost function is $c'(Q)$ and let its total revenue be equal to $\text{rev}(Q)$, then I can derive the following identity:

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

This identify is insightful as a firm’s revenue $\text{rev}(Q)$ is directly observed from the ASM/CM whereas it is relatively straightforward to construct its cost function.\(^{26}\) Therefore, I am able to obtain mark-up estimates 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 mark-ups as it then satisfies $\mu(Q)/\beta > \mu(Q)$. Most importantly however, the relationship between this measure of mark-ups and firm size should be unaffected.

\(^{26}\)See Appendix A.4 for an extensive discussion of what is contained in a plant’s measure of total cost.
as the ratio of revenue and total costs is only off by a constant. In the following, I will denote this specific measure of mark-ups by $\mu^{CRS}$.

To strengthen my results, I will also estimate mark-ups 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.5. Standard cost minimization techniques imply that that the output elasticity with respect to material inputs $\theta_{it}^M$ can be rewritten as:

$$\theta_{it}^M = \mu_{it} \frac{P_{it}^M M_{it}}{P_{it} Q_{it}}$$

where $\mu_{it}$ denotes a firm’s mark-up at time $t$. This equality is valid as long as material inputs are a static input, i.e. these inputs are not prone to significant adjustment costs. Total expenditures on material inputs $P_{it}^M M_{it}$ and the value of gross output $P_{it} Q_{it}$ are easily constructed in the ASM/CM. Thus, mark-ups at the firm-year level can then be constructed by obtaining an estimate for $\theta_{it}^M$. In the following, I lay out the empirical framework to do this. I impose $\theta_{it}^M = \theta_{j(i)}^M$. This means that output elasticities with respect to inputs are constant over time and do not vary within a industry.\(^{27}\)

I will use a gross output Cobb-Douglas specification as a baseline, but the procedure is easily extended to the translog case. The estimation procedure is carried out industry-by-industry. Thus, production technologies are allowed to differ across, but not within, industries. 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 $\omega_{it}$ but $\epsilon_{it}$ 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 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 $\omega_{it}$ is unobserved by the econometrician leading to the well-known “transmission bias” problem: $\omega_{it}$ is correlated with the set of inputs resulting in biased and inconsistent estimates of $\beta = (\beta_k, \beta_\ell, \beta_m, \beta_e)$.\(^{28}\) To resolve this issue, I follow Levinsohn and Petrin (2003) and use a “proxy” method by exploiting material inputs. It is assumed that capital is decided upon a period ahead. Therefore, a firm’s state at time $t$ consists of capital $k_{it}$ and productivity $\omega_{it}$. As a result, its policy function for material inputs can be described as:

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

\(^{27}\)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 translog production functions.

\(^{28}\)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 $\beta_\ell$ will be underestimated under OLS. On the other hand, firms experiencing positive productivity shocks will most likely hire more labor. However, the estimate of $\beta_\ell$ under OLS will incorrectly contribute the increase in output to this change in labor. As a result, this means that $\hat{\beta}_{\ell}^{OLS} > \beta_\ell$. 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).
Levinsohn and Petrin (2003) show that under a relatively mild set of conditions the mapping \( m_t(k_{it}, \cdot) \) is invertible. Therefore, I obtain \( \omega_{it} = h_t(m_{it}, k_{it}) \). This implies that unobserved productivity can be expressed as a function of observable inputs only. As a result, I run the following non-parametric regression in the first stage:

\[
y_{it} = \phi_{it}(k_{it}, \ell_{it}, m_{it}, e_{it}) + \epsilon_{it}
\]

Let its fitted values and residuals (i.e. estimates of expected output and measurement error respectively) be denoted by \( \hat{\phi}_{it} \) and \( \hat{\epsilon}_{it} \). In the second stage, I retrieve the production function coefficients \( \beta \). To do this, I assume that the law of motion for productivity is given by:

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

where \( \xi_{it} \) denote the innovations to productivity. For any value of \( \beta \), it must be that \( \omega_{it}(\beta) = \hat{\phi}_{it} - \beta_kk_{it} - \beta_\ell\ell_{it} - \beta_mm_{it} - \beta_e e_{it} \). Thus, it is straightforward to obtain a measure of \( \xi_{it}(\beta) \) by non-parametrically regressing \( \omega_{it}(\beta) \) on its lag. Identification of \( \beta \) is obtained by relying on the following four moment conditions:

\[
E \left[ \begin{bmatrix} k_{it} \\ \ell_{it-1} \\ m_{it-1} \\ e_{it-1} \end{bmatrix} \right] = \left[ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right]
\]

Estimates of \( \beta \) are then obtained by using standard GMM techniques which involves the minimization of a quadratic loss function. Standard errors are obtained through a block bootstrapping procedure. The main assumptions underlying identification exploit the fact that capital is chosen a period ahead and should therefore be orthogonal to future innovations to productivity. A similar logic holds for the lagged inputs \( \ell_{it-1}, m_{it-1} \) 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 mark-ups 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 appear to be serially correlated over time. Mark-ups at the firm-year level are then simply constructed through:

\[
\hat{\mu}_{it} = \beta_j(i) \frac{P_{it} M_{it}}{P_{it} Q_{it} \hat{\epsilon}_{it}}
\]

where \( j(i) \) denotes the industry to which firm \( i \) belongs to.

Note that I explicitly correct observed gross output for measurement error with \( \hat{\epsilon}_{it} \). The estimation results, \(^{29}\) are then obtained by using standard GMM techniques which involves the minimization of a quadratic loss function. Standard errors are obtained through a block bootstrapping procedure. The main assumptions underlying identification exploit the fact that capital is chosen a period ahead and should therefore be orthogonal to future innovations to productivity. A similar logic holds for the lagged inputs \( \ell_{it-1}, m_{it-1} \) 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 mark-ups 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 appear to be serially correlated over time. Mark-ups at the firm-year level are then simply constructed through:

\[
\hat{\mu}_{it} = \beta_j(i) \frac{P_{it} M_{it}}{P_{it} Q_{it} \hat{\epsilon}_{it}}
\]

where \( j(i) \) denotes the industry to which firm \( i \) belongs to.

\[^{29}\text{This is potentially important as this correction will eliminate any variation in expenditure shares that comes from variation in} \]
obtained industry-by-industry (3 digit NAICS), are shown in table 10. The first two columns present the median of the mark-up distribution obtained for each industry. Under decreasing returns to scale, the methodology using the ratio of revenues over total costs overestimates mark-ups. The GMM-IV estimation procedure seems to be a decent job of correcting this bias as I find that, in general, \( \text{med}(\hat{\mu}_{it}^{CRS}) \geq \text{med}(\hat{\mu}_{it}^{GMM}) \). Median level of mark-ups vary substantially across industries and the average value of these medians is equal to 1.XXXX. These values seem to be consistent with estimates using Slovenian (de Loecker and Warzynski (2012)), Chilean (Lamorgese, Linarello and Warzynski (2014)) and Indian (de Loecker et al. (2016)) data. While there is a significant amount of variation in mark-ups between industries, the within-industry variation is also substantial. The third column displays the standard deviation of the mark-up distribution within each industry: the mean of this standard deviation across industries amounts to X.XXXX. Lastly, columns 4 and 5 show the estimates of the output elasticity with respect to intermediate inputs. As can be seen from table 10, the OLS estimator tends to slightly underestimate \( \beta_m \) highlighting the transmission bias in production function estimation.

Table XI. Estimation results of relationship between firm-level mark-ups and size.

<table>
<thead>
<tr>
<th>SIZE</th>
<th>( \hat{\mu}^{CRS} )</th>
<th>( \hat{\mu}^{CRS} )</th>
<th>( \hat{\mu}^{GMM} )</th>
<th>( \hat{\mu}^{GMM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
<td>XXXX</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>( Y )</td>
<td>( Y )</td>
<td>( Y )</td>
<td>( Y )</td>
</tr>
<tr>
<td>INDUSTRY-YEAR</td>
<td>( N )</td>
<td>( Y )</td>
<td>( N )</td>
<td>( Y )</td>
</tr>
<tr>
<td>PLANT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^k \) The estimation sample covers the U.S. manufacturing industries (NAICS 31 - 33) over 1976 - 2009. The independent variable is always the natural log of mark-ups. In columns 2 and 3, mark-ups are constructed by using the ratio of revenues over total costs. Columns 3 and 4 employ mark-ups as constructed by the GMM-IV methodology contained in the text. Standard errors are clustered at the industry (3-digit NAICS) level.

While these estimates are interesting in their own right, I am primarily focused on the relationship between firm-level mark-ups and size. To assess this relationship, I perform the following regression in the pooled sample:

\[
\tilde{\mu}_{it} = \gamma_0 + \gamma_1 \ln S_{it} + \Gamma'_{it} \sigma + \nu_{it}
\]

where a firm’s size is reflected by its total number of employees and \( \Gamma_{it} \) contains a full set of industry-year interactions to control for industry-specific aggregate trends in mark-ups. I run this regression specification for both measures of mark-ups (either \( \tilde{\mu}_{it} = \tilde{\mu}_{it}^{CRS} \) or \( \tilde{\mu}_{it} = \tilde{\mu}_{it}^{GMM} \)). The results are summarized in table 11.

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 as mentioned by Klette and Griliches (1996).
This table depicts a clear picture: firm-level mark-ups increase in size regardless of the way mark-ups are measured. Most importantly, this relationship even seems to hold whenever I restrict myself to variation in mark-ups within plants. The mark-up elasticity with respect to size varies from X.XXXX to X.XXXX depending on whether I control for plant fixed effects.

**Growth.** While the size-variance relationship has not received much attention in the literature on deviations from the strong version of Gibrat’s Law, the case is completely different for firm-level growth. Many studies find that smaller firms grow faster than larger firms. However, the relationship between firm-level growth and size is more nuanced than just being mere negative. Stylized fact 8 in Klette and Kortum (2004) also states that “among larger firms, growth rates are unrelated to past growth or to firm size”. 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, but a more thorough empirical investigation using the LBD can be found in Appendix B.4.

### 5.2 Identification

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.

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 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 mark-ups. 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 mark-ups across industries. More importantly, mark-ups 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.

---

30 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)).

31 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).
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 facts on size-varying price elasticities and mark-ups. 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 mark-up 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 mark-up over marginal costs which is contradictory with the evidence of section 5.1.

Therefore, I construct an analytically tractable framework in continuous time featuring random growth à la Luttmer (2007) and a Kimball aggregator. 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 mark-ups are increasing in firm size as consistent with the results in section 5.1.

5.3 Framework

**Consumers.** A representative agent has preferences over a consumption stream \( \{C_t\}_{t \geq 0} \) of a composite good. Utility is specified as:

\[
\int_0^{\infty} \exp(-\rho t)u(C_t)dt
\]

for \( \rho > 0 \). Labor \( L_t = L\exp(\eta t) \) for some \( \eta \geq 0 \) is supplied inelastically and compensated at the rate \( w_t \). The composite good is created by the costless aggregation of a countable number \( N \geq 1 \) of monopolistically competitive goods using the Kimball (1995) structure:

\[
\frac{1}{N} \sum_{i=1}^{N} \psi \left( \frac{Nc_{it}}{C_t} \right) = 1
\]

where \( \psi(1) = 1, \psi' > 0 \) and \( \psi'' < 0 \). This Kimball aggregator features a price elasticity that decreases in the relative quantity consumed of a monopolistically competitive good. 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 \( \theta > 1 \). In general, the composite good \( C \)
is only implicitly defined through the Kimball structure. For ease of exposition, I will drop time subscripts unless denoted otherwise.

There is no savings device in the economy, thus all income which consists of labor income and firm profits is spent on the purchase of consumption goods. The cost minimization of this static problem leads to the following first order conditions:

$$
\psi'(\frac{Nc_i}{C}) = \frac{p_i}{P} \sum_{j=1}^{N} \frac{c_i}{C} \psi'(\frac{Nc_j}{C})
$$

where $P$ denotes the price of the composite good. $\psi''(\cdot) < 0$ implies that $\varphi(\cdot) \equiv (\psi'(\cdot))^{-1}$ must exit. From the consumer’s viewpoint, $D \equiv \sum_{j=1}^{N} \frac{c_i}{C} \psi'(\frac{Nc_j}{C})$ is constant. Then, relative demand is equal to:

$$
\frac{c_i}{C} = \frac{1}{N} \varphi \left( \frac{p_i}{P} D \right)
$$

This relative demand function is well-behaved in the sense that it is continuous and downward sloping in $p_i$.

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:

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

where $\Gamma(u, z) = \int_{z}^{+\infty} s^{u-1} \exp(-s) ds$ is the incomplete gamma function. This specification has two advantages. First, the model is parsimoniously specified by only two parameters $\theta$ and $\varepsilon$. Second, the specification has a natural interpretation. In the limit of $\varepsilon \to 0$, the Kimball aggregator converges to the standard CES aggregator. As a result, the parameter $\varepsilon$ can be interpreted as the “deviation” from a CES specification with elasticity $\theta > 0$. Most importantly however, the price elasticity of demand for a specific variety is decreasing in its relative quantity. This result can be shown analytically as:

$$
\sigma(\varphi(\tilde{p})) \equiv - \frac{d \ln \varphi(\tilde{p})}{d \ln(\tilde{p})} = - \frac{\psi'(x)}{\psi''(x)x} = \frac{\theta x^{-\varepsilon/\theta}}{x - \varepsilon/\theta}
$$

where $\tilde{p}$ denotes the relative price $p/P$ of some individual variety. Clearly, the limit $\varepsilon \to 0$ implies that the price elasticity of demand is constant.

**FIRMS.** A firm is characterized by a constant returns to scale production function with labor $\ell$ only. Whenever its productivity equals $z$, it can transform $1/z$ units of labor into one unit of output. Thus, the production
function is characterized by \( y(\ell) = z\ell \). Its profit function is characterized by:

\[
\pi(z) = \frac{C}{N} \max_{p \geq 0} \left( p - \frac{w}{z} \right) \varphi \left( \frac{p}{P} D \right)
\]

Under the chosen Kimball aggregator of Klenow and Willis (2016), the relative demand function can be expressed in closed form:

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

A firm’s first order condition in \( p \) can be rearranged in the form of an endogenous mark-up function over marginal costs. This results in:

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

The CES case of \( \varepsilon \to 0 \) would lead to the constant mark-up over marginal costs in which \( p^* = \frac{\theta}{\theta - 1} w \). 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 the Lambert \( W \)-function. In particular, the Lambert \( W \)-function (or product logarithm) is defined as the inverse of the mapping \( x \mapsto x \cdot \exp\left( x \right) \) which has some properties that I can exploit later.\(^{32}\) As a result, it can be shown that a firm’s optimal price equals:

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

Following Luttmer (2007) and Arkolakis (2016), I assume that a firm’s log productivity \( s = \ln(z) \) evolves over time (or age) according to:

\[
\text{d}s_a = \mu \text{d}a + \sigma F \text{d}W_a
\]

where \( W_a \) is a standard Brownian motion.\(^{33}\) Using the expression for a firm’s optimal price, firm-level revenue as a function of log productivity \( s \) can be expressed as:

\[
\text{rev}(s) = p^* \varphi \left( \frac{p^*}{P} D \right) = w \times \frac{\theta}{\varepsilon} \frac{\exp(-s)}{W[\varTheta \exp(-s)]} \left( \theta - \varepsilon W[\varTheta \exp(-s)] \right)^{\theta/\varepsilon} \tag{B}
\]

where \( \varTheta = \frac{\theta}{\varepsilon} \frac{\theta}{\theta - 1} \frac{w}{P} D \). The model is closed in general equilibrium by clearing the labor market. In the following, I will assume that there exists some balanced growth path (BGP) in which the real wage \( \frac{w}{P} \) is

\(^{32}\) 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.

\(^{33}\) Calendar time \( t = t^b + a \) is defined as a firm’s time of entry (or birth) \( t^b \) plus its age \( a \).
constant and the wage rate \( w \) grows at some constant rate \( g_w \geq 0 \).\(^{34}\) This implies that \( \Theta \) is constant as well. With this assumption in hand, it is relatively straightforward to derive that firm size, defined as firm-level revenue, and productivity are positively related. Lemma 1 formalizes this one-to-one relationship.

**Lemma 1.** A firm’s size is increasing in its log level of 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 and 5.1. Equation B indicates that firm-level revenue is the product of a deterministic term, i.e. the wage rate \( w \), and a stochastic term. By defining \( g(a) = wa \) and \( h(s_a) = \frac{\theta}{\varepsilon W[\exp(-s)]} (\theta - \varepsilon W[\Theta \exp(-s)])^\theta/\varepsilon \), firm-level revenue can simply be expressed as \( \text{rev}(s_a) = g(a) \times h(s_a) \). Applying Itô’s lemma to this expression, it is immediate that:

\[
\frac{dr(s_a)}{r(s_a)} = g_w + \left[ \mu \frac{h'(s_a)}{h(s_a)} + \frac{\sigma_F^2}{2} \right] da + \sigma_F \frac{h'(s_a)}{h(s_a)} dW_a
\]

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 and growth are declining in firm size. This is formalized in proposition 1 and 2.

**Proposition 1.** Firm-level volatility \( \text{vol}(s_0) = \sqrt{V \left( \frac{dr(s)}{r(s)} \bigg|_{s=s_0} \right) / da} \) 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 \( \varepsilon \to 0 \).

*Proof.* See Appendix D.3. □

**Proposition 2.** Firm-level growth \( m(s_0) = \mathbb{E} \left( \frac{dr(s)}{r(s)} \bigg|_{s=s_0} \right) / da \) declines in firm size \( \text{rev}(s_0) \) whenever \( \theta > \varepsilon > 0 \). For \( \varepsilon \to 0 \), firm-level growth is constant as \( m(s_0) = g_w + \mu (\theta - 1) + \frac{\sigma_F^2}{2} (\theta - 1)^2 \) for all \( s_0 > 0 \).

*Proof.* See Appendix D.4. □

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 and 2 are only qualitative statements, so it is unclear whether the generated size-variance relationship is also a declining power law. However, figure 6 shows that the Kimball specification is flexible enough to generate the desired power law. The right picture of figure 6 shows that my framework is able to generate another qualitative feature on the relationship between firm-level growth and size. As documented in Klette and Kortum (2004), smaller firms grow faster than larger firms, but Gibrat’s Law cannot be rejected amongst the largest firms. Lastly,

\(^{34}\)Note that I also allow for a stationary case in which growth rates are equal to zero. It is possible to derive the results for propositions 1 to 3 without these assumptions. However, the analysis is considerably more convenient under this scenario.
Figure 6: The left figure shows the generated size-variance relationship 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. The figure on the right shows the relationship between firm-level growth and size under the same parameterization. Smaller firms grow at a higher rate than larger firms but, among the large firms, growth rates are approximately unrelated to firm size.

This random growth framework with a Kimball aggregator is able to generate a positive relationship between firm-level mark-ups and size. This is depicted in proposition 3.

**Proposition 3.** A firm’s mark-up (i.e. optimal price over marginal cost) declines in firm size $\text{rev}(s_0)$. Whenever $\varepsilon \to 0$, mark-ups 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 mark-up. 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 pricing effect.

### 5.4 Calibration

TBA

### 6 Concluding remarks

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. In particular, the contribution of the granular channel as proposed by Gabaix (2011) is substantially smaller than previously found. While the firm-specific component of aggregate volatility is relatively small, it is not negligible. In fact, the majority of the firm-specific component seems to be generated by input-output linkages as described in Acemoglu et al. (2012) which gives additional empirical support for the potential importance of buyer-supplier networks.

Furthermore, the seemingly surprising small role of the granular channel can be rationalized by the size-variance relationship alone: 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 aggregate volatility is strongly diminished. The log-linear relationship between 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 proposed a theory of firm dynamics that incorporates a demand specification with price elasticities that are declining in firm size. Most importantly, this tractable framework did 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 mark-ups being increasing in firm size. This finding is directly supported by the empirical findings from section 5.1 using data from the U.S. manufacturing sector.

My findings imply that deviations from Gibrat’s Law extend to a greater domain than just firm-level growth and volatility. Furthermore, I show that one of these deviations (i.e. size-variance relationship) has important aggregate implications in granular economies. The results of section 5.1 are furthermore strongly indicative of size-varying pass-through which could have important implications for the literature dealing with exchange rate pass-through. 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.

A U.S. Census Bureau data

A.1 Longitudinal Business Database (LBD)
A.2 Standard Statistical Establishment List (SSEL)
A.3 Longitudinal Firm Trade Transactions Database (LFTTD)
A.4 Annual Survey of Manufacturers/Census of Manufacturers (ASM/CM)
A.5 Homogeneous product-level data in CM
A.6 LBD-SSEL-LFTTD merging process
Robustness and estimation strategies

B.1 Variance decomposition

**INPUT-OUTPUT LINKAGES.** In section 3.3, I argue that the firm-level idiosyncratic component is dominated by the \( \text{LINK}_\tau \) component. Obviously, positive comovement between firms does not immediately imply the existence of input-output linkages. As a result, I provide suggestive evidence for the existence of these “network” effects. To do this, I correlate the \( \text{LINK}_\tau \) component at the industry pair level with a measure of input-output linkages. For the former, I define:

\[
\text{LINK}_{ij\tau} = \sum_{(f,n) \in I_i\tau} \sum_{(g,m) \in I_j\tau} w_{fn\tau-1}w_{gm\tau-1}\text{COV}(\varepsilon_{fnt}, \varepsilon_{gmt})
\]

where \( I_i\tau \) denotes the set of firms that are active in industry \( i \) at time \( \tau \). To create a measure of input-output linkages, I define a pair \((i,j)\)’s “mean intensity of input-output linkages” as suggested by di Giovanni, Levchenko and Mejean (2014):

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

where \( 1 - \lambda_{i\tau} \) denotes the ratio of spending on intermediate inputs over gross output in sector \( i \) at time \( \tau \). 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.\(^{35}\) Second, the measure can be constructed with data at the sector level. This is important as there is no extensive information available on firm-level interconnections in the U.S. economy.

To do this, I use public source 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 \( \tau \).

Lastly, 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 \( \text{II} \)) and gross output (code \( \text{GO} \)) can be taken directly from the tables. Aggregation to the required industry level as defined above is straightforward as the BEA uses \( \text{IO}_{\text{code}} \) in both data sets.

\(^{35}\)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.\(^2\)

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

\(^2\)This 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 (ICodes 42 and 44RT), Finance, Insurance and Real Estate (ICodes 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 (ICodes GFG, GFE, GSLG and GLE).

**ADDITIONAL ROBUSTNESS SPECIFICATIONS.**
B.2 Size-variance relationship

ROBUSTNESS TESTS FOR KOREN-TENREYRO PROCEDURE.

Table B.2.1. Robustness specifications for the size-variance relationship.3

<table>
<thead>
<tr>
<th>ESTABLISHMENTS</th>
<th>FULL</th>
<th>FULL</th>
<th>SINGLE-UNIT</th>
<th>SINGLE-UNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
</tr>
<tr>
<td>(d_{it})</td>
<td>0.XXX</td>
<td>0.XXX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRODUCTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
<td>-0.XXX</td>
</tr>
<tr>
<td>(d_{it})</td>
<td>0.XXX</td>
<td>0.XXX</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fixed effects

<table>
<thead>
<tr>
<th>FIRM</th>
<th>N</th>
<th>Y</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
</table>

ESTIMATION SAMPLE

3Firm size is defined as employment count. Standard errors are clustered at the industry (3-digit NAICS) level. Each row with “SIZE” displays the estimate \(\hat{\alpha}\) whenever the regression specification controls for output diversification (proxied by number of establishments within a firm) or product diversification with an indicator variable. These results can be found under the columns “FULL”. The estimates \(\hat{\alpha}\) for the subsample of firms with only one establishment or product can be found under the columns “SINGLE-UNIT”.

TESTING FOR STRUCTURAL BREAK IN THE SIZE-VARIANCE RELATIONSHIP.

Table B.2.2. Rejecting structural breaks in the size-variance relationship.4

<table>
<thead>
<tr>
<th>(H_0: \alpha_b = 0) vs (H_1: \alpha_b \neq 0)</th>
<th>(F)-STATISTIC</th>
<th>(p)-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_{it} = 1) if (S_{it} &gt; 1000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d_{it} = 1) if (S_{it} &gt; 5000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d_{it} = 1) if (S_{it} &gt; 10000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fixed effects

<table>
<thead>
<tr>
<th>FIRM</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
</table>

ESTIMATION SAMPLE

4Firm size is defined as employment count. Standard errors are clustered at the industry (3-digit NAICS) level. Each row displays the \(F\)-statistic and \(p\)-value of the Chow-test in which I test for a structural break at some known point. The table displays these statistics for three potential breaking points, i.e. firm size \(S_{it}\) equal to 1000, 5000 or 10000 employees.

STANLEY-SUTTON PROCEDURE FOR SIZE-VARIANCE RELATIONSHIP.
### B.3 Size-varying price elasticities

Table B.3. Estimation results of size-varying price elasticities.\(^5\)

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>IV</th>
<th>IV - SMALL</th>
<th>IV - LARGE</th>
<th>IV - d(_{it}^P)</th>
<th>IV - d(_{it}^{TE})</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BOXES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p_{it})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p_{it} \times d_{it})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BREAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p_{it})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p_{it} \times d_{it})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ICE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p_{it})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p_{it} \times d_{it})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>YEAR</strong></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

\(^{5}\text{The first two columns estimate the restricted regression specification in which } \sigma_2 = \sigma_3 = 0. \text{ 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 \text{ but the sample is restricted to small/medium and large firms respectively. In the last two columns, the results of the unrestricted IV regression are displayed. The column “IV - d\(_{it}^P\)” displays the estimated coefficients when the interaction dummy equals 1 whenever a plant’s size exceeds the 75\(^{th}\) percentile. Similarly, the column “IV - d\(_{it}^{TE}\)” displays the estimated coefficients whenever the dummy threshold is set to 50 employees. Standard errors are clustered at the plant level.}\)

### B.4 Deviation from Gibrat’s Law: firm-level growth and size

### B.5 Mark-up estimation: de Loecker and Warzynski (2012)

In this section, I will provide more details on the mark-up 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 \(Q_{it} = Q_{it}(X_{it}, K_{it}, \omega_{it})\) where \(X_{it} = (X_{1it}, \ldots, X_{Vit})\) is a vector of inputs free of adjustment costs, \(K_{it}\) denotes capital and \(\omega_{it}\) is a plant’s technology level. Following de Loecker and Warzynski (2012), I will impose the following assumption.

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

**B2.** Productivity \(\omega_{it}\) 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 \( \theta_{it}^X = \frac{\partial Q_{it}(\cdot)}{\partial X_{it}} X_{it} \) is the output elasticity with respect to input \( X \), \( \mu_{it} \) is the plant-specific time-varying mark-up and \( \alpha_{it}^X = \frac{p_{it}^X X_{it}}{r_{it} Q_{it}} \) is 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, mark-ups \( \mu_{it} \) are easily obtained through the identity:

\[
\mu_{it} = \frac{\theta_{it}^X}{\alpha_{it}^X}
\]

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

\[
y_{it} = f(x_{it}, k_{it}; \beta) + \omega_{it} + \epsilon_{it}
\]

Obviously, plant-level productivities \( \omega_{it} \) 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:

\[
m_{it} = m_t(k_{it}, \omega_{it}, p_{it})
\]

where capital \( k_{it} \) and productivity \( \omega_{it} \) are state variables at the time of input choice(s). Furthermore, the vector \( p_{it} \) 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 \( \omega_{it} = h_t(m_{it}, k_{it}, p_{it}) \). As a result, production \( y_{it} \) can be written in terms of observables only:

\[
y_{it} = f(x_{it}, k_{it}; \beta) + h_t(m_{it}, k_{it}, p_{it}) = \phi_t(x_{it}, k_{it}, m_{it}, p_{it}) + \epsilon_{it}
\]

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 Ackerberg, Caves and Frazer (2015).

**Step 1. Non-parametric estimation of \( \phi_{it} \) and \( \epsilon_{it} \).**

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

\[
y_{it} = \sum_{j=1}^{M} \ell_{it}^j + m_{it}^j + k_{it}^j + \epsilon_{it}^j + \sum_{j'=1}^{N} \ell_{it}^{j'} m_{it}^{j'} + \ell_{it}^{j'} k_{it}^{j'} + \ell_{it}^{j'} \epsilon_{it}^{j'} + k_{it}^{j'} m_{it}^{j'} + k_{it}^{j'} \epsilon_{it}^{j'} + m_{it}^{j'} \epsilon_{it}^{j'} + \ell_{it}^{j} m_{it}^{j} k_{it}^{j} \epsilon_{it}^{j} + \epsilon_{it}^{j}
\]

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 \( \xi_{it} \) to productivity \( \omega_{it} \).**

Second, I assume that productivity \( \omega_{it} \) 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:

\[
\omega_{it}(\beta) = \tilde{\omega}_{it} - \beta_k \ell_{it} - \beta_k k_{it} - \beta_m m_{it} - \beta_\epsilon \epsilon_{it}
\]
where $\beta \equiv (\beta_k, \beta_m, \beta_e, \beta_{\ell_k}, \beta_{\ell_m})'$. Then, I approximate $g_t(.)$ with a third order polynomial in its argument:

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

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

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

where $\hat{\rho}(\beta) = (\hat{\rho}_0, \hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3)'$ is the OLS estimator of $\Omega_{it-1}(\beta) = (1, \omega_{it-1}(\beta), \omega_{it-1}^2(\beta), \omega_{it-1}^3(\beta))'$ on $\omega_{it}(\beta)$.

**Step 3. GMM-IV estimation of $\beta$.**

I assume that capital $k_{it}$ is decided one period ahead, thus it is orthogonal to the innovation $\xi_{it}(\beta)$. Furthermore, lagged labor $\ell_{it-1}$ is used to instrument for $\ell_{it}$ as current period labor is decided after the realization of the innovation $\xi_{it}(\beta)$. As a result, I expect $\mathbb{E}(\ell_{it}\xi_{it}) \neq 0$. However, I require that $\ell_{it}$ and $\ell_{it-1}$ to be correlated with each other to satisfy the rank conditions. This is the case whenever wages are correlated over time. A similar argument is made for intermediate inputs $m_{it}$ and energy inputs $e_{it}$. Define the instrument vector as $z_{it} = (\ell_{it-1}, k_{it}, m_{it-1}, e_{it-1})'$, then the parameters $\beta$ are identified through the following $K = 4$ moment conditions:

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

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

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

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

$$\hat{\mu}_{it} = \hat{\beta}_k \left( \frac{\nu m_{it} + g o_{it}}{g o_{it} / \ell_{it}} \right)^{-1}$$

where $\nu m_{it}$ and $g o_{it} = tv s_{it} + (f ie_{it} - f i b_{it}) + (w i b_{it} - w i b_{it})$ 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 mark-up results are robust to whenever I do not include inventories.

**C Accounting framework for granular economy**

**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 $\theta$. Labor endowment $L$ is supplied inelastically and compensated at the wage rate $w > 0$. Total expenditure $X$ comprises of labor income and received profits $\hat{\pi}$ from

---

36 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 $\beta$ is substantially increased as $\beta = (\beta_k, \beta_m, \beta_e, \beta_{\ell_k}, \beta_{\ell_m}, \beta_{\ell_e})'$ whenever I adopt a polynomial approximation of degree $M = 3$.

37 By construction, the number of parameters in $\beta$ is equal to the amount of identifying moments. This case of “just identification” renders the specification of a weighting matrix useless.
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 \( P = \left( \sum_{k=1}^{N} \frac{p(k)}{P} \right)^{1/(1-\theta)} \) 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 \( \iota \) is a Cobb-Douglas composite of labor \( \ell \) 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 \( M(k) = \left( \sum_{k'=1}^{N} \frac{m(k,k')}{\beta} \right)^{\frac{\beta}{\beta-1}} \) 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 \( \beta \to 1 \), these linkages are absent. By standard arguments, the cost of one input bundle \( \iota \) is \( c = \theta P^{1-\beta} \). A firm’s production function is linear in this input bundle, thus its optimal price is characterized by the standard constant CES mark-up over marginal costs which in turn results in the following equilibrium supplied quantity:

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

**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.\(^{38}\) Thus, there exists a cut-off value \( \bar{\pi} \) such that firms only incur the fixed cost of production \( f_P \) whenever \( a \leq \bar{\pi} \). 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 \( P \).

\(^{38}\)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:

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

As \( \theta > 1 \), there exists a cut-off \( \bar{\pi} \) such that \( \bar{\pi}(a) \geq 0 \) if and only if \( a \leq \bar{\pi} \).
ASSUMPTION 2. The marginal firm treats \( X \) and \( P \) as non-stochastic.

These assumptions deliver substantial computational and analytical simplifications: solving the equilibrium is reduced to a static problem as will become clear.\(^{39}\) The number of potential entrants \( \bar{N} \) will be pinned down by the free entry condition in which the \textit{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:

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

Then, I can solve for a \textit{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 \( k \in \{1, 2, \ldots, N\} \) and imposing free entry which brings net aggregate profits to zero such that \( \beta X = wL \).

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 conditional standard deviation of a firm’s idiosyncratic shock is decreasing in its size at the rate \( \alpha \), 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 \( \zeta = \frac{\varphi}{\varphi - 1} \) and aggregate volatility is given by:

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

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

\(^{39}\)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 mark-up over marginal cost would not be the firm’s optimal pricing function. However, di Giovanni and Levchenko (2012) suggest that deviations of fully flexible mark-ups from constant CES mark-ups 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 mark-up 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 \textit{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.
negligible. This ensures that assumption A.2 is satisfied. The parameters $\alpha$ 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.$^6$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Target</th>
<th>Model</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varphi$</td>
<td>5.3</td>
<td>1.06</td>
<td>1.06</td>
<td>Pareto tail (Luttmer, 2007)</td>
</tr>
<tr>
<td>$f_e$</td>
<td>1.9189</td>
<td>0.1091</td>
<td>0.1091</td>
<td>Entry rate (BDS, 1977 - 2012)</td>
</tr>
<tr>
<td>$f_P$</td>
<td>1.0546</td>
<td>5.0309 · 10^6</td>
<td>5.0309 · 10^6</td>
<td>Firm count (BDS, 2012)</td>
</tr>
</tbody>
</table>

$^6$The parameter $\varphi$ 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.

D Theoretical framework

D.1 Properties of the Lambert $W$-function

D.2 Proof of lemma 1

D.3 Proof of proposition 1

D.4 Proof of proposition 2

D.5 Proof of proposition 3
References


