AIFE Working Paper
No. 07-2026

Aggregate Fluctuations from Firm Comovement

Nahyeon Bak and Daisoon Kim

May 2026

JEL classification: E23, E32.

Keywords: Business cycles, idiosyncratic comovements, pairwise correlation.

ANDERSEN INSTITUTE
for Finance & Economics

Andersen Institute for Finance & Economics working papers are written by Andersen Institute economists and associated contributors.

The views expressed within are those of the authors and not necessarily those of the Andersen Institute.

This publication is available on the Andersen Institute website (www.andersoninstitute.org).

ANDERSEN INSTITUTE™
for Finance & Economics

Aggregate Fluctuations from Firm Comovement

Nahyeon Bak*   Daisoon Kim†

May 1, 2026

Abstract

This paper shows that correlated idiosyncratic fluctuations among firms within industries are an important source of macroeconomic volatility. Standard methods, such as cross-sectional demeaning, obscure this channel by mechanically driving average pairwise correlations toward zero asymptotically. We develop a non-parametric bounds approach that quantifies the contribution of such clustered comovement directly from firm-level data. Applied to U.S. firms, we find that within-industry comovement explains 10–15% of GDP volatility in normal times, 15–30% in downturns, and up to 40% following the Great Recession, helping account for the Great Moderation. These findings highlight the importance of cross-firm interdependencies in business cycle fluctuations.

JEL Classification: E23, E32.
Keywords: Business cycles, idiosyncratic comovements, pairwise correlation.

*Amazon.com
E-mail: nahyeonbak@gmail.com URL: https://nahyeonbak-netizen.github.io
†Andersen Institute of Finance and Economics; NC State University
E-mail: kim.daisoon.kr@gmail.com URL: https://daisoonkim.github.io

The authors would like to thank Matteo Cacciatore, Yongsung Chang, Hyunbae Chun, Andrew Hessler, Jay Hyun, Ilze Kalnina, Aeimit Lakdawala, Yoonsoo Lee, Toshihiko Mukoyama, Rory Mullen, Jose-Victor Rios-Rull, Dongho Song, Nicolas Vincent, Yang Yu, Minchul Yum, as well as seminar participants at the Wake Forest University, Shanghai Jiao Tong University, HEC Montreal, UC Riverside, University of Manitoba, Korea Development Institute, 2023 I-85 Macroeconomics Workshop, 2023 Midwest Econometrics Meeting, Society for Nonlinear Dynamics and Econometrics Symposium in 2022 and 2023, 2022 Korean Bureau of Economic Research and Innovation Summer Institute, 2022 Asian Meeting of the Econometric Society, 2022 North American Summer Meetings of the Econometric Society, 2021 Korean Economic Review International Conference, and Recent Development in Quantitative Macroeconomics Conference for useful comments. Any remaining errors are our own.

The views expressed within are those of the authors and not necessarily those of the Andersen Institute of Finance and Economics, Amazon.com, or their subsidiaries.

1. Introduction

One of the central questions in macroeconomics is what drives business-cycle fluctuations. The canonical representative-agent framework attributes aggregate volatility primarily to economy-wide or sector-wide shocks, assuming that firm-specific disturbances wash out through the law of large numbers. Yet recent evidence from firm-level data reveals a much more heterogeneous economy. The firm-size distribution is highly skewed, with a handful of very large firms exerting outsize influence on aggregate outcomes. This observation gave rise to the influential *granular* view (Jovanovic, 1987; Gabaix, 2011), which shows that idiosyncratic shocks to outsized firms can translate into significant aggregate fluctuations.

In this paper, we uncover another micro-level source of aggregate volatility: *clustered origins*. This arises when firms experience correlated, though not identical, shocks within groups such as industries. They differ from common macro-shocks, since they do not affect all firms equally. Instead, they reflect a form of pairwise comovement or synchronization within groups. Because these shocks are correlated, they do not diversify. As a result, even moderate-size firms can collectively generate substantial volatility at the aggregate level.

Why has this channel been underappreciated? A key reason is methodological. Commonly used empirical approaches—most notably cross-sectional demeaning designed—to isolate firm-specific (idiosyncratic) components mechanically impose an average pairwise correlation across firms toward zero asymptotically. This removes the imprint of clustered comovement by construction, rather than because it is absent. As a result, attenuation bias arises and the importance of firm-level interdependence is understated. Our first contribution is to formalize and demonstrate this pitfall.

To address this challenge, we develop a simple and robust non-parametric decomposition that separates aggregate fluctuations into three components: (i) *clustered origins*—the contribution of within-industry comovement of firm-specific shocks, (ii) *granular origins*—the importance of idiosyncratic shocks to large firms, and (iii) *macro origins*—the role of genuinely common (industry- and economy-wide) shocks. Our method relies only on basic variance-covariance properties (such as nonnegativity and Cauchy-Schwarz bounds), observable firm- and industry-level moments, and natural economic weights like sales or

Domar weights. By construction, it requires no restrictive structural assumptions. The approach delivers robust upper and lower bounds for each source of volatility and is valid under heterogeneity and misspecification.

We apply this framework to U.S. publicly traded firms (Compustat, 1975–2023) and obtain three main results. First, clustered origins are important even during tranquil periods, accounting for roughly 10–15% of GDP variance. Second, they are countercyclical: their share of aggregate volatility rises to 15-30% in downturns and peaks near 40% following turbulent periods such as the Great Recession. Third, their time path closely tracks major shifts in economic volatility: their decline during the Great Recession coincides with falling GDP volatility, while their resurgence after 2000 parallels the post-crisis rise in volatility. These results are robust across alternative measures of firm activity (sales, productivity, proxies for value-added), filtering procedures, Domar adjustments, and sectoral subsamples. Although industries differ in the extent of within-industry comovement, no single sector drives the aggregate patterns.

Taken together, our evidence shows that the business cycle is shaped not only by aggregate shocks or the idiosyncratic fortunes of a few large firms, but also by correlated fluctuations across many moderate-sized firms within industries. By documenting the empirical importance and cyclical properties of clustered origins, we provide a new microeconomic foundation for aggregate volatility. This highlights the broader need for macroeconomic analysis to account for firm networks and interdependencies, which can amplify and transmit shocks across the economy.

Contribution to the literature. This paper contributes to several key strands of research on business cycle fluctuations by uncovering a novel microeconomic source of aggregate volatility. First, it challenges the widespread assumption that idiosyncratic firm-level shocks are uncorrelated, extending recent advances in the heterogeneous-firms literature that question conventional distributional assumptions and highlight complex dependence structures (e.g., Guvenen, Karahan, Ozkan and Song, 2021; Sterk, Sedlacek and Pugsley, 2021; Forneron, 2023; Jaimovich, Terry and Vincent, 2023).

A substantial body of work emphasizes the crucial role of micro-level factors in shaping aggregate volatility. Foundational contributions examine how firm-level shocks and heterogeneity aggregate up to influence macro volatility (e.g., Comin and Philippon, 2005; Comin and Mulani, 2006; Gabaix, 2011; Carvalho and Gabaix, 2013). Within this framework, our paper introduces a complementary microeconomic channel—clustered firm comovement—that enriches the well-established granularity framework. Originating from Jovanovic (1987) and formalized by Gabaix (2011), granularity captures how idiosyncratic shocks to large firms disproportionately affect aggregate fluctuations.[1]

Further advancing this line of inquiry, several studies explore the network origins of aggregate shocks, incorporating supply chains and input-output linkages that amplify firm-level volatility through interconnections (e.g., Acemoglu, Carvalho, Ozdaglar and Tahbaz-Salehi, 2012; Carvalho, 2014; Oberfield, 2018; Herskovic, Kelly, Lustig and Van Nieuwerburgh, 2020). Additionally, related research analyzes cross-sector correlations and the hierarchical transmission of shocks across levels of economic aggregation (e.g., Carvalho and Gabaix, 2013; di Giovanni, Levchenko and Mejean, 2014), as well as the micro-level sources of volatility in open economies (e.g., di Givonanni and Levchenko, 2012; di Giovanni, Levchenko and Mejean, 2014; Gaubert and Itoskhoki, 2021). Departing from these works, we focalize on within-cluster pairwise correlations among firms, offering a novel lens to capture correlated idiosyncratic fluctuations at a granular level.

Our contribution also deepens the literature on the industrial and sectoral origins of macroeconomic volatility, tracing back to Long and Plosser (1983). This foundational research shows that interconnected supply chains and production networks limit the diversification of industry-specific shocks, thereby propagating sectoral disturbances into aggregate fluctuations (e.g., Bak, Chen, Scheinkman and Woodford, 1993; Horvath, 1998; Horvath and Verbrugge, 1999; Dupor, 1999; Foerster, Sarte and Watson, 2011; Atalay, 2017). While those studies predominantly focus on between-industry correlations, we complement their insights by uncovering significant micro-level within-industry comovement among firms.

Finally, our empirical findings shed light on the dynamics of U.S. GDP volatility over

[1]For comprehensive treatments of granularity, including applications to financial sectors, see Buch and Neugebauer (2011), Amiti and Weinstein (2018), Bremus, Buch, Russ and Schnitzer (2018), Kim, Park and So (2025), among others.

recent decades. We document a pronounced U-shaped pattern in aggregate volatility—marked by a decline from the early 1980s to the early 2000s, followed by a resurgence thereafter—which we partially attribute to correlated firm-level shocks. This pattern ties closely to the literature on the Great Moderation and its aftermath (e.g., Kim and Nelson, 1999; Stock and Watson, 2002; Comin and Mulani, 2006; Davis, Haltiwanger, Jarmin, Miranda, Foote and Nagypal, 2006), supporting theoretical arguments like those in Carvalho and Gabaix (2013) that emphasize the amplified role of firm- and industry-level volatility in shaping macroeconomic stability and its reversal.

The remainder of the paper is structured as follows. Section 2 introduces the statistical framework. Section 3 shows the pitfalls of cross-sectional demeaning. Section 4 develops the empirical method. Section 5 presents empirical results. Section 6 concludes.

2. Framework and Motivation

This section introduces a simple statistical framework to illustrate how correlated idiosyncratic firm behaviors can generate aggregate fluctuations.

Framework. Consider a cluster (industry) with N_t number of firms, where firm i's variable (e.g., sales, output, employment, productivity) in log form and $\hat{y}_{it}$ denotes its business cycle component. Each firm's fluctuations comprise two uncorrelated random variables with zero mean:

$$\hat{y}_{it} = \varepsilon_{\mathrm{A},t} + \varepsilon_{\mathrm{F},it}, \tag{1}$$

where $\varepsilon_{\mathrm{A},t}$ is a (true) common component affecting all firms in the cluster, which includes both industry-specific factors and the responses of individual industries to economy-wide macroeconomic factors. The other term, $\varepsilon_{\mathrm{F},it}$, is a firm-specific (true) idiosyncratic component. Their standard deviations are denoted by $\sigma_{\mathrm{A},t}$ and $\sigma_{\mathrm{F},it}$, respectively.

Our key innovation is to relax the conventional assumption that idiosyncratic components are uncorrelated across firms. We allow the pairwise correlation, $\rho_{\mathrm{FF},ii't} \equiv \mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})$, to be nonzero for $i \neq i'$. This correlation can drive aggregate-level movements and foster comovements across firms within the cluster.[2] This motivates our use of 'cluster' rather than 'industry' terminology to emphasize these within-industry correlations. Sections 2 and 3 focus on within-cluster fluctuations, Section 4 will extend the analysis to the multi-cluster economy.

The Aggregate Business Cycle Fluctuations. Aggregate volatility arises from three sources: common fluctuations, idiosyncratic fluctuations, and pairwise comovements of idiosyncratic fluctuations. The aggregate business cycle component is:

$$\hat{Y}_t = \sum_i w_{it}\hat{y}_{it}, \tag{2}$$

where w_{it} represents the share of firm size within the cluster, and it satisfies the condition $\sum_i w_{it} = 1$.[3] The variance of this aggregate business cycle component decomposes into:

$$\sigma^2_{\hat{\mathrm{Y}},t} = \sigma^2_{\mathrm{A},t} + \sum_i w^2_{it}\sigma^2_{\mathrm{F},it} + \sum_i w_{it}\sum_{i'\neq i} w_{i't}\rho_{\mathrm{FF},ii't}\sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't}, \tag{3}$$

where the second and third terms represent the micro origins of macro fluctuations through the volatility and comovements of idiosyncratic fluctuations, respectively. To clarify the economic interpretation, we define three key terms:

Definition 1 (Macro origins) $\sigma^2_{\mathrm{A},t}$: *The variance of the common component affecting all firms.*

Definition 2 (Granular origins) $\Gamma_t = \sum_i w^2_{it}\sigma^2_{\mathrm{F},it}$: *The contribution from individual firms' idiosyncratic components, weighted by squared firm shares (Gabaix, 2011).*

Definition 3 (Clustered origins) $\chi_t = \sum_i w_{it}\sum_{i'\neq i} w_{i't}\rho_{\mathrm{FF},ii't}\sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't}$: *The contribution of synchronized idiosyncratic movements to aggregate volatility through covariance of firm-specific components.*

[2] These pairwise correlations can result in notable short-term departures from the long-run level, i.e., $N_t^{-1}\sum_i \hat{y}_{it} \neq 0$ or $\sum_i w_{it}\hat{y}_{it} \neq 0$. However, they do not affect the long-run level, i.e., $\mathbb{E}[\hat{y}_{it}] = 0$.

[3] In addition to size weights, business cycle research often uses Domar weights to account for input-output linkages (e.g., Domar, 1961; Hulten, 1978). For simplicity, we do not use them here; Section 4 does so.

A Simple Example of Clustered Origins. To illustrate the economic significance of clustered origins, consider a simplified setting with identical firms. Under this assumption, both variance and covariance of idiosyncratic components are uniform across firms, allowing us to rewrite equation (3) as:

$$\sigma^2_{\hat{Y},t} = \sigma^2_{\mathrm{A},t} + h_t^2 \sigma^2_{\mathrm{F},t} + (1 - h_t^2)\rho_{\mathrm{F},t}\sigma^2_{\mathrm{F},t}, \tag{4}$$

where $h_t = (\sum_i w_{it}^2)^{1/2}$ represents the Herfindahl-Hirschman index (HHI) within the cluster. This decomposition reveals three distinct sources of aggregate volatility: common shocks, granular origin from firm size dispersion, and clustered origins from correlated idiosyncratic movements. It is important to note that this term signifies micro origins in the form of a convex combination of variance and covariance ($\sigma^2_{\mathrm{F},t}$ and $\rho_{\mathrm{F},t}\sigma^2_{\mathrm{F},t}$) with weight h_t^2. The second term represents granular origins, which, Lucas (1977) argued, should vanish through diversification as $h_t \to 0$ as $N_t \to \infty$. However, Gabaix (2011) showed that with the fat-tailed firm size distribution, the HHI remains bounded away from zero even as the number of firms increases, preserving the granular channel. Our focus is on the third term—clustered origins. Two key insights emerge: (1) positive correlations across firms ($\rho_{\mathrm{F},t} > 0$) amplify aggregate volatility, (2) The impact of clustered origins decreases with market concentration (HHI), unlike granular origins. To quantify the relative importance of these channels, consider the ratio of clustered to granular origins:

$$\frac{\chi_t}{\Gamma_t} = \left(\frac{1}{h_t^2} - 1\right)\rho_{\mathrm{F},t}. \tag{5}$$

For instance, assume a size distribution with a Herfindahl–Hirschman index of $h_t = 0.12$, as demonstrated in Gabaix (2011).[4] In this context, even small positive pairwise correlations of idiosyncratic fluctuations, ranging from 1% to 5%, imply that the range of clustered origins is between 68% and 342% of granular origins, according to equation (5). A 1.46% correlation coefficient results in an equal contribution of idiosyncratic comovements across firms and granularity to aggregate volatility. These calculations highlight that

[4] In our U.S. public firm-level data, h_t is around 0.085. In that case, the correlation 1% implies that the clustered origins are around 137% of granular origins in equation (5).

clustered origins can be a first-order driver of aggregate business cycle fluctuations, even with seemingly small pairwise correlations.

3. The Problem with Demeaning

In equation (1), a firm's business cycle component is directly observable, but its true common and idiosyncratic components remain hidden. Many prior studies on business cycles utilize a common technique where they approximate the common component using the cross-sectional sample mean and estimate the idiosyncratic components by calculating the deviations from this mean. To be more precise, we can define pseudo common and idiosyncratic components using the cross-sectional sample mean and the deviations from it ($e_{\mathrm{A},t}$ and $e_{\mathrm{F},it}$):

$$\hat{y}_{it} = \left(\overline{\hat{y}}_t\right) + \left(\hat{y}_{it} - \overline{\hat{y}}_t\right) \equiv e_{\mathrm{A},t} + e_{\mathrm{F},it}, \tag{6}$$

where $\overline{\hat{y}}_t = {N_t}^{-1} \sum_i \hat{y}_{it}$ represents the cross-sectional sample mean. This approach allows us to estimate these latent common and idiosyncratic components.

3.1. Simulation Design and Results

We conduct Monte Carlo simulations to illustrate how standard demeaning hides important dynamics when idiosyncratic shocks are correlated. We simulate 5,000 firms over 50 periods, drawing firm-specific shocks from a multivariate normal distribution with zero mean, 12% standard deviation, and either zero or 2.5% pairwise correlation. Firms have uniform weights, and no common component is included.[5] We repeat this simulation 3,000 times to obtain robust sample statistics for both true and pseudo-idiosyncratic components.

The results show two key findings. First, aggregate fluctuations differ markedly between true and pseudo (demeaned) components. Figure 1 contrasts these patterns in two economies through Panels (a) and (b). While true idiosyncratic components (orange solid line) show larger aggregate fluctuations with positive correlation, pseudo components (blue dashed line) remain at zero by construction in both cases. Second, demeaning masks

[5] For detailed results and additional simulations with varying weights, please refer to Appendix A.

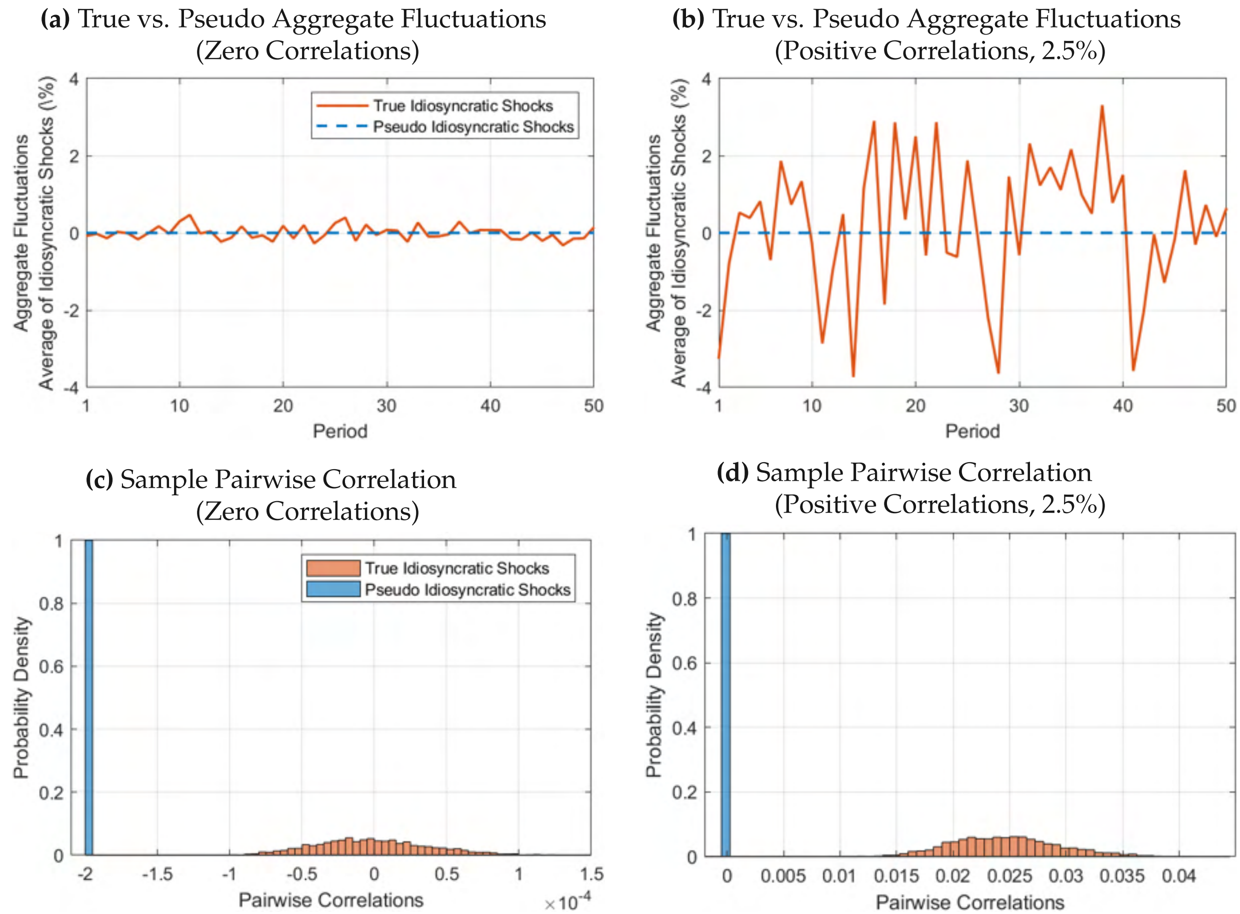

Figure 1: Aggregate Fluctuations and Sample Pairwise Correlations: True vs. Pseudo Components
Notes: The figures (a) and (b) plot the aggregate fluctuations based on the average true pseudo idiosyncratic components in each period (orange solid line, $N_t^{-1}\sum_{i=1}^{5,000}\varepsilon_{\mathrm{F},it}$, and blue dashed line, $N_t^{-1}\sum_{i=1}^{5,000} e_{\mathrm{F},it}$, respectively. The figures (c) and (d) plot histograms for sample correlations of true and pseudo variables (orange bars, $\mathrm{corr}(\varepsilon_{\mathrm{F},it},\varepsilon_{\mathrm{F},i't})$, and blue bars, $\mathrm{corr}(e_{\mathrm{F},it},e_{\mathrm{F},i't})$, respectively) from 3,000 simulations.

actual correlation patterns. Panels (c) and (d) present the distribution of pairwise correlations from our 3,000 simulations. While the true components retain their imposed correlations (0% and 2.5%), the pseudo components always exhibit near-zero correlations ($\approx -2\times 10^{-4}$), regardless of the underlying correlation structure.[6] These results demonstrate that demeaning mechanically eliminates evidence of correlated idiosyncratic movements, potentially leading researchers to overlook an important source of aggregate fluctuations.

[6]We discuss the slightly negative correlations of the pseudo components and show that they are asymptotically zero under identical variance–covariance structures (equation 16).

3.2. Non-Negligible Difference Between True and Pseudo Variables

When true idiosyncratic components exhibit correlation, pseudo variables derived from demeaning procedures fail to serve as adequate proxies for the true variables. This inadequacy can be formally demonstrated by examining the absolute difference between true and pseudo components:

$$\left|e_{\mathrm{A},t} - \varepsilon_{\mathrm{A},t}\right| = \left|e_{\mathrm{F},it} - \varepsilon_{\mathrm{F},it}\right| = |\overline{\varepsilon}_{\mathrm{F},t}|, \tag{7}$$

where $\overline{\varepsilon}_{\mathrm{F},t} = N_t^{-1}\sum_i \varepsilon_{\mathrm{F},it}$ represents the average of the true idiosyncratic components. This equality reveals that pairwise comovements among idiosyncratic components generate systematic differences between true and pseudo measures.

To formalize this insight, we define two key statistical measures. First, let $\overline{\sigma}^2_{\mathrm{F},t} = N_t^{-1}\sum_i \sigma^2_{\mathrm{F},it}$. It represents the average of idiosyncratic variances. Second, let $\overline{\overline{\mathrm{cov}}}_{\mathrm{FF},t} = N_t^{-1}\sum_i \overline{\mathrm{cov}}_{\mathrm{FF},it}$. It denotes the average covariance of idiosyncratic components, where $\overline{\mathrm{cov}}_{\mathrm{FF},it} = (N_t - 1)^{-1}\sum_{i'\neq i} \rho_{\mathrm{FF},ii't}\sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't}$ It represents the average covariance of firm i with all other firms. These definitions allow us to establish the following result:

Lemma 1 *The difference between true and pseudo components fails to converge to zero in mean square as the number of firms approaches infinity. For any firm i, the equation is as follows:*

$$\mathbb{E}\left[|e_{\mathrm{A},t} - \varepsilon_{\mathrm{A},t}|^2\right] = \mathbb{E}\left[|e_{\mathrm{F},it} - \varepsilon_{\mathrm{F},it}|^2\right] = \frac{1}{N_t}\overline{\sigma}^2_{\mathrm{F},t} + \left(1 - \frac{1}{N_t}\right)\overline{\overline{\mathrm{cov}}}_{\mathrm{FF},t}. \tag{8}$$

Proof. The result follows directly from equation (7), with $\mathbb{E}\left[|\overline{\varepsilon}_{\mathrm{F},t}|^2\right] = \mathsf{var}(\overline{\varepsilon}_{\mathrm{F},t})$. ∎

This lemma has important implications for business cycle analysis. While the first term on the right side of equation (8) vanishes with a large number of firms and with finite variances. However, the second term remains non-negligible when average pairwise covariances are non-zero. Consequently, the squared difference between true and pseudo components does not converge to zero in mean square. This renders the conventional approach of using sample means and deviations inadequate for business cycle research.

3.3. Inconsistent Estimator of Variance and Covariance

Building on our previous results, we now demonstrate that pseudo variables yield inconsistent estimators of variance and covariance when significant comovements exist ($\overline{\overline{\text{cov}}}_{\text{FF},t} \neq 0$). The pseudo components systematically misrepresent the variance of both common and idiosyncratic components as follows.

$$\mathsf{var}(e_{\text{A},t}) = \sigma^2_{\text{A},t} + \frac{1}{N_t}\overline{\sigma}^2_{\text{F},t} + \left(1 - \frac{1}{N_t}\right)\overline{\overline{\text{cov}}}_{\text{FF},t} \tag{9}$$

$$\begin{aligned}\mathsf{var}(e_{\text{F},it}) = {} & \left(1 - \frac{1}{N_t}\right)\left(\sigma^2_{\text{F},it} - \overline{\text{cov}}_{\text{FF},it}\right) \\ & - \left[\frac{1}{N_t}\left(\sigma^2_{\text{F},it} - \overline{\sigma}^2_{\text{F},t}\right) + \left(1 - \frac{1}{N_t}\right)\left(\overline{\text{cov}}_{\text{FF},it} - \overline{\overline{\text{cov}}}_{\text{FF},t}\right)\right]\end{aligned} \tag{10}$$

These equations reveal two important patterns. First, when firms show no correlation with others ($\overline{\text{cov}}_{\text{FF},it} = 0$), pseudo variables provide consistent estimates. However, if firm-level idiosyncratic components are positively correlated, the pseudo (demeaned) variables used in traditional analyses will systematically overestimate common component volatility while underestimating idiosyncratic volatility. This mismeasurement can mask important sources of business cycle fluctuations.

More fundamentally, pseudo variables misrepresent comovements. The covariance between the pseudo common component and firm i's idiosyncratic component is:

$$\mathsf{cov}(e_{\text{F},it}, e_{\text{A},t}) = \frac{1}{N_t}\left(\sigma^2_{\text{F},it} - \overline{\sigma}^2_{\text{F},t}\right) + \left(1 - \frac{1}{N_t}\right)\left(\overline{\text{cov}}_{\text{FF},it} - \overline{\overline{\text{cov}}}_{\text{FF},t}\right). \tag{11}$$

While this averages to zero across firms ($N_t^{-1}\sum_i \mathsf{cov}(e_{\text{F},it}, e_{\text{A},t}) = 0$), individual firms can show nonzero covariance when their covariances differ from the average ($\overline{\text{cov}}_{\text{FF},it} - \overline{\overline{\text{cov}}}_{\text{FF},t}$), even though true common and idiosyncratic components are uncorrelated by definition.

The covariance between the pseudo idiosyncratic components of firm i and i' is:

$$\begin{aligned}\mathsf{cov}(e_{\text{F},it}, e_{\text{F},i't}) = {} & \rho_{\text{FF},ii't}\sigma_{\text{F},it}\sigma_{\text{F},i't} - \frac{1}{2}\left(1 - \frac{1}{N_t}\right)\left(\overline{\text{cov}}_{\text{FF},it} + \overline{\text{cov}}_{\text{FF},i't}\right) \\ & - \frac{1}{2N_t}\left(\sigma^2_{\text{F},it} + \sigma^2_{\text{F},i't}\right) - \frac{1}{2}\mathsf{cov}(e_{\text{F},it} + e_{\text{F},i't}, e_{\text{A},t}).\end{aligned} \tag{12}$$

This leads to our key result:

Proposition 1 *The cross-sectional average of pairwise covariances of pseudo idiosyncratic components (cross-sectionally demeaned fluctuations) converges to zero as the number of firms goes to infinity.*

Proof. Combining equation (11) into equation (12) yields the following equation:

$$\frac{1}{N_t}\sum_i \frac{1}{N_t-1}\sum_{i'\neq i} \mathsf{cov}(e_{\mathrm{F},it}, e_{\mathrm{F},i't}) = \frac{1}{N_t}\left(\overline{\overline{\mathrm{cov}}}_{\mathrm{FF},t} - \overline{\sigma}^2_{\mathrm{F},t}\right). \tag{13}$$

This value converges to zero as $N_t \to \infty$ ■

This result shows that empirical methods relying on demeaned firm fluctuations will mechanically erase evidence of within-industry comovement. Researchers may falsely conclude that idiosyncratic shocks are uncorrelated when, in reality, substantial micro-level synchronization exists and meaningfully contributes to aggregate fluctuations.

3.4. When Can Demeaning Be Justified? Identical Variance-Covariance

While our previous results highlight the problems with demeaning in the presence of correlated idiosyncratic fluctuations, most existing business cycle studies have assumed zero or negligible cross-firm correlations. This section examines the special case where such an assumption can be justified: when true idiosyncratic components have identical variance and covariance across firms. However, as we will show in Section 5, this assumption fails empirically in U.S. data.

Consider the case where all firms have identical variance and covariance for their idiosyncratic components, i.e., $\sigma^2_{\mathrm{F},it} = \sigma^2_{\mathrm{F},t}$ and $\rho_{\mathrm{FF},ii't} = \rho_{\mathrm{F},t}$ for all $i \neq i'$. Under these conditions, equations (9) and (10) simplify to:

$$\mathsf{var}(e_{\mathrm{A},t}) = \sigma^2_{\mathrm{A},t} + \frac{1}{N_t}\sigma^2_{\mathrm{F},t} + \left(1 - \frac{1}{N_t}\right)\rho_{\mathrm{F},t}\sigma^2_{\mathrm{F},t}, \tag{14}$$

$$\mathsf{var}(e_{\mathrm{F},it}) = \left(1 - \frac{1}{N_t}\right)(1-\rho_{\mathrm{F},t})\sigma^2_{\mathrm{F},t}. \tag{15}$$

While pseudo variables still overstate (understate) common (idiosyncratic) component volatility when true correlation is positive, they maintain two useful properties. First, pseudo common and idiosyncratic components are uncorrelated. Second, cross-firm correlation of pseudo idiosyncratic components becomes negligible as firm count increases.

Specifically, the correlation structure simplifies to as follows.[7]

$$\mathsf{corr}(e_{\mathrm{F},it}, e_{\mathrm{A},t}) = 0 \quad \text{and} \quad \mathsf{corr}(e_{\mathrm{F},it}, e_{\mathrm{F},i't}) = -\frac{1}{N_t - 1} \tag{16}$$

These correlations are independent of true idiosyncratic correlations, suggesting that demeaned fluctuations become asymptotically uncorrelated across firms.

In cases where firms have nearly identical volatility and covariance patterns, cross-sectional demeaning does a reasonable job separating common and idiosyncratic shocks with the following aggregate volatility decomposition:

$$\sigma^2_{\hat{Y},t} = \mathsf{var}(e_{\mathrm{A},t}) + h_t^2 \mathsf{var}(e_{\mathrm{F},it}) - \frac{1 - h_t^2}{N_t - 1} \mathsf{var}(e_{\mathrm{F},it}). \tag{17}$$

However, the homogeneous variance-covariance assumption rarely holds in real-world data making traditional approaches unsuitable for accurate micro-macro decompositions. Extensive empirical evidence contradicts the assumption of identical variance and covariance in firm-level outcomes. Studies examining idiosyncratic productivity, output, and financial performance consistently find substantial heterogeneity (e.g., Stanley, Amaral, Buldyrev, Havlin, Leschhorn, Maass, Salinger and Stanley, 1996; Xu and Malkiel, 2003; Comin and Philippon, 2005; Comin and Mulani, 2006; Chun, Kim, Morck and Yeung, 2008; Castro, Clementi and Lee, 2015; Tweedle, 2018; Sterk, Sedlacek and Pugsley, 2021; Kalnina and Tewou, 2025). Our analysis in Section 5 reinforces these findings, documenting significant variation in business cycle properties across U.S. firms, even within industries.

With heterogeneous variance and covariance structures, cross-sectional demeaning

[7] Note that the results in equation (16) hold when the pseudo factors are constructed from the unweighted mean. If the weights are unequal, i.e., consider weighted mean, then the pseudo common and idiosyncratic components are correlated, while the pairwise correlation of pseudo idiosyncratic components is independent of the true idiosyncratic components' pairwise correlation. See Appendix B for the related results with the weighted mean.

fails to properly identify common and idiosyncratic components. This invalidates the use of pseudo idiosyncratic fluctuations and necessitates alternative methods to recover true components and their moments for understanding business cycle origins.

4. Identifying Micro Origins of Macro Fluctuations

How can we quantify the micro origins from data? While we can directly observe a firm's business cycle component ($\hat{y}_{it}$), its underlying common and idiosyncratic components ($\varepsilon_{\mathrm{A},t}$ and $\varepsilon_{\mathrm{F},it}$) are unobservable, making direct point value decomposition impossible without additional information or structural assumptions.

Rather than imposing additional structural assumptions to attempt a precise point estimation of these unobservable components, we develop a bounds approach that exploits observable moments to establish ranges for the components of interest. We first establish upper and lower bounds for each firm's common component variance using the variance and covariance of observed $\{\hat{y}_{it}\}_{i=1}^{N_t}$ to determine the range of cluster and granular effects within clusters. We then extend this analysis to the entire U.S. economy across all clusters.

4.1. Within-Cluster Micro Origins

Within a cluster, the observed firm-level fluctuations ($\{\hat{y}_{it}\}_{i=1}^{N_t}$) provide two key moments:

$$\mathsf{var}(\hat{y}_{it}) = \sigma^2_{\mathrm{A},t} + \sigma^2_{\mathrm{F},it} \tag{18}$$

$$\mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't}) = \sigma^2_{\mathrm{A},t} + \rho_{\mathrm{FF},ii't}\sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't} \tag{19}$$

Using them, we can reformulate the clustered and granular origins from equation (3):

$$\chi_t = \sum_i w_{it} \sum_{i' \neq i} w_{i't} \mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't}) - (1 - h_t^2)\sigma^2_{\mathrm{A},t} \tag{20}$$

$$\Gamma_t = \sum_i w_{it}^2 \mathsf{var}(\hat{y}_{it}) - h_t^2 \sigma^2_{\mathrm{A},t} \tag{21}$$

While all terms except $\sigma^2_{\mathrm{A},t}$ are observable, we can bound this parameter:

Proposition 2 *In a cluster, the variance of the common component should not exceed $\sigma^{*2}_{\mathrm{A},t}$.*

$$0 \leq \sigma^2_{\mathrm{A},t} \leq \sigma^{*2}_{\mathrm{A},t} = \min_{i,i'} \left\{ \mathsf{var}(\hat{y}_{it}),\ \left[1 + \mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})\right] \mathsf{sd}(\hat{y}_{it}) \mathsf{sd}(\hat{y}_{i't}) \right\} \tag{22}$$

Proof. First, non-negative variance implies $\mathsf{var}(\hat{y}_{it}) \geq \sigma^2_{\mathrm{A},t}$ in equation (18). Thus, we obtain $\min_i\{\mathsf{var}(\hat{y}_{it})\} \geq \sigma^2_{\mathrm{A},t}$. Second, Cauchy–Schwarz inequality yields $\rho_{\mathrm{FF},ii't}\sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't} \geq -\sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't}$. From equation (19), we obtain $\mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't}) + \sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't} \geq \sigma^2_{\mathrm{A},t}$. Because non-negative $\sigma^2_{\mathrm{A},t}$ implies $\mathsf{var}(\hat{y}_{it}) \geq \sigma^2_{\mathrm{F},it}$ for all i, we obtain that $[1+\mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})]\mathsf{sd}(\hat{y}_{it})\mathsf{sd}(\hat{y}_{i't}) \geq \sigma^2_{\mathrm{A},t}$ for any i and i'. Thus, $\min_{i,i'}\{[1 + \mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})]\mathsf{sd}(\hat{y}_{it})\mathsf{sd}(\hat{y}_{i't})\} \geq \sigma^2_{\mathrm{A},t}$. Hence, we obtain equation (22). ■

Using the range of $\sigma^2_{\mathrm{A},t}$, we can derive ranges for the clustered and granular origins.

Corollary 1 *The clustered and granular origins are bounded as follows:*

$$\sum_i w_{it} \sum_{i' \neq i} w_{i't} \mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't}) - (1 - h^2_t)\sigma^{*2}_{\mathrm{A},t} \leq \chi_t \leq \sum_i w_{it} \sum_{i' \neq i} w_{i't} \mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't}) \tag{23}$$

$$\sum_i w^2_{it} \mathsf{var}(\hat{y}_{it}) - h^2_t \sigma^{*2}_{\mathrm{A},t} \leq \Gamma_t \leq \sum_i w^2_{it} \mathsf{var}(\hat{y}_{it}) \tag{24}$$

Proof. This is directly from Equations (20) and (21) with Proposition 2. ■

4.2. Aggregation: The Whole Economy's Micro Origins

To measure micro (clustered and granular) origins in aggregate GDP fluctuations, we aggregate across clusters as follows. Let $i \in I_{st} \subset I_t = \cup_{s' \in S} I_{s't}$ index firms, where $s \in S$ indexes clusters. Define $w_{it} = w_{st} w_{sit}$ as firm i's economy-wide share, decomposed into its within-cluster share (w_{sit}) and its cluster's share (w_{st}).

The business cycle component of the whole economy's GDP, denoted as $\widehat{\mathrm{GDP}}_t$, is computed as $\widehat{\mathrm{GDP}}_t = d_t \sum_{i \in I_t} w_{it}\hat{y}_{it}$ and equivalently expressed as $\widehat{\mathrm{GDP}}_t = d_t \sum_{s \in S} w_{st}\hat{Y}_{st}$, where $\hat{Y}_{st}$ is the cluster-level aggregate business cycle component and d_t is the Domar weight adjustment defined as the ratio of a gross output to aggregate value-added (GDP).[8]

[8]The business cycle research often uses the Domar weight to account for both direct and indirect supply chain impacts (e.g., Domar, 1961; Hulten, 1978).

This yields the variance decomposition, indicating the whole economy's volatility:

$$\mathsf{var}(\widehat{\mathrm{GDP}}_t) = d_t^2 \sum_{s \in S} w_{st}^2 \mathsf{var}(\hat{Y}_{st}) + d_t^2 \sum_{s \in S} w_{st} \sum_{s' \in S \setminus \{s\}} w_{s't} \mathsf{cov}(\hat{Y}_{st}, \hat{Y}_{s't}) \tag{25}$$

$$= d_t^2 \sum_{s \in S} w_{st}^2 [\sigma_{\mathrm{A},st}^2 + \chi_{st} + \Gamma_{st}] + \mathrm{BIO}_t, \tag{26}$$

where BIO_t denotes the between-industry origins. The granular and clustered origins of cluster s are $\Gamma_{st} = \sum_{i \in I_{st}} w_{sit}^2 \sigma_{\mathrm{F},it}^2$ and $\chi_{st} = \sum_{i \in I_{st}} w_{sit} \sum_{i' \in I_{st} \setminus i} w_{si't} \rho_{\mathrm{FF},ii't} \sigma_{\mathrm{F},it} \sigma_{\mathrm{F},i't}$.

Finally, we compute the micro origins of the whole economy as follows:

Definition 4 (Granular origins in the whole economy) $\Gamma_t = d_t^2 \sum_{s \in S} w_{st}^2 \Gamma_{st}$.

Definition 5 (Clustered origins in the whole economy) $\chi_t = d_t^2 \sum_{s \in S} w_{st}^2 \chi_{st}$.

The between-industry origins are defined as:

$$\mathrm{BIO}_t = d_t^2 \sum_{s \in S} w_{st} \sum_{s' \in S \setminus \{s\}} w_{s't} \left[\mathsf{cov}(\varepsilon_{\mathrm{A},st}, \varepsilon_{\mathrm{A},s't}) + \sum_{i \in I_{st}} w_{sit} \sum_{i' \in I_{s't}} w_{si't} \mathsf{cov}(\varepsilon_{\mathrm{F},it}, \varepsilon_{\mathrm{F},i't}) \right]. \tag{27}$$

This term includes both correlated industry-specific fluctuations across industries (clusters) and cross-cluster firm correlation. While the first part in the parenthesis is a well-documented aspect, this paper primarily focuses on the clustered origins, which pertain to cross-firm correlation within a cluster, leaving the issues related to BIO_t for future research.

5. The Evolution of Clustered Origins in the U.S. Economy

This section quantifies clustered origins of aggregated volatility in the U.S. economy, using our empirical framework from Section 4.

5.1. Data and Summary Statistics

Our analysis uses annual firm-level sales and employment data from the Compustat North America Fundamental Annuals database for the period 1975–2023. Because the COVID-19 pandemic caused significant disruptions and structural changes, our baseline analysis

Table 1: Summary Statistics

Variable	Full sample 1980–2013	Subperiod sample 1980–1985	Subperiod sample 1986–2000	Subperiod sample 2001–2013	Additional 2014–2018
Panel A. Within-firm volatility: standard deviation, $\mathsf{var}(\hat{y}_{it})$					
Mean	0.184	0.160	0.187	0.191	0.179
Standard deviation	0.208	0.153	0.212	0.223	0.219
Quantile 10%	0.056	0.055	0.054	0.058	0.051
50%	0.125	0.119	0.128	0.126	0.116
90%	0.345	0.284	0.352	0.364	0.340
Observations (firms)	76,230	12,814	32,474	30,942	10,494
Panel B. Within-cluster comovement: pairwise correlation: $\mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})$					
Mean	0.139	0.124	0.092	0.179	0.175
Standard deviation	0.354	0.354	0.341	0.353	0.362
Quantile 10%	−0.352	−0.352	−0.371	−0.311	−0.332
50%	0.152	0.128	0.101	0.203	0.200
90%	0.599	0.596	0.544	0.635	0.642
Observations (pairs)	8,004,848	1,086,406	3,093,306	5,033,342	1,208,206

Notes: The table reports summary statistics for firm-level volatility (standard deviation of $\hat{y}_{it}$) and within-cluster pairwise comovement (correlation of $\hat{y}_{it}$ and $\hat{y}_{i't}$). Cyclical components $\hat{y}_{it}$ are residuals from the regression of log real sales in equation (28). All moments are computed over an 11-year rolling window ($[t-5, t+5]$). Pairwise correlations are calculated only for firms within the same cluster (among 53 clusters). The standard deviations are winsorized at the top 1%.

focuses on 1975–2018. We convert nominal values to real terms using industry-specific price deflators from the U.S. Bureau of Economic Analysis (BEA). We group firms into 53 industry clusters, with grouping determined by the availability of the BEA deflator availability. Table A1 lists these clusters, and Appendix C provides the details on the data construction and measurement methodology.

To isolate the business cycle component of firms, we estimate a panel regression equation, where the residual, $\hat{y}_{it}$, represents the cyclical component.

$$y_{it} = \beta_s y_{it-1} + \psi_s^{\text{age}} \times \ln(t - \text{first year}_i) + \psi_s^{\text{emp}} \times \ln \text{emp}_{it} + \psi_s^{\text{time}} \times t + \delta_i + \hat{y}_{it}, \tag{28}$$

where y_{it} is the logarithm of real sales or labor productivity (real sales per employee) for firm i at time (year) t. The firm fixed effect, δ_i, absorbs time-invariant firm characteristics such as location and cohort. Following Castro, Clementi and Lee (2015), we control for firm age ($t -$ first year$_i$) and size (log employment, $\ln \text{emp}_{it}$), as prior literature documents their negative correlation with firm growth and volatility.[9] Importantly, all coefficients are estimated separately for each cluster s, allowing for heterogeneity in firm dynamics across industries.

Using these cyclical components, $\hat{y}_{it}$, we compute firm-specific variances, pairwise covariances, and correlations using an 11-year rolling window centered at each year $t \pm 5$ (i.e., over the interval $[t-5, t+5]$). We denote these time-varying moments as $\mathsf{var}(\hat{y}_{it})$, $\mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't})$, $\mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})$. As a robustness check, we confirm our results using simple growth rates (i.e., log difference, $\hat{y}_{it} = y_{it} - y_{it-1}$) instead of the regression residuals from equation (28). This alternative choice does not materially affect our main conclusions.

Table 1 reveals two significant trends in our data. First, average firm-level volatility increased during the Great Moderation period (1986–2000) compared to the early 1980s, aligning with findings by Comin and Philippon (2005) and Comin and Mulani (2006), despite a concurrent decline in aggregate volatility. Second, average within-cluster correlation shows a U-shaped pattern: declining from the early 1980s through the 1990s, then doubling after 2000. However, these observed patterns do not directly reflect purely idiosyncratic movements, as they combine both idiosyncratic and common cluster-level fluctuations (equations 18 and 19).

To examine the homogeneity assumption of the idiosyncratic variance-covariance discussed in Section 4, we analyze the cross-sectional distribution of calculated moments. If the idiosyncratic components were uniform within the clusters, we would expect uniform second moments observed ($\mathsf{var}(\hat{y}_{it})$ and $\mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't})$). Figure 2 challenges this assumption by showing histograms of firm standard deviations and pairwise correlations for 1995.[10] The wide distributions, distinctly non-zero centered, demonstrate significant heterogeneity in firm-level variances and comovements, even within the same industry. This crucial

[9] For seminal evidence on U.S. manufacturing, see Evans (1987) and Hall (1987). These controls, in conjunction with fixed effects and a linear time trend, also serve to mitigate potential biases from endogenous comovement between aggregate and firm-level productivity driven by mechanisms like technology adoption or vintage capital effects (e.g., Schaal and Taschereau-Dumouchel, 2018; Mullen, 2020; Fiori and Scoccianti, 2021).

[10] The patterns—significant cross-sectional variation—are unchanged when using within-cluster demeaned firm statistics (see Appendix Figure A4).

(a) Standard Deviation **(b)** Cross-Firm Correlation within Cluster

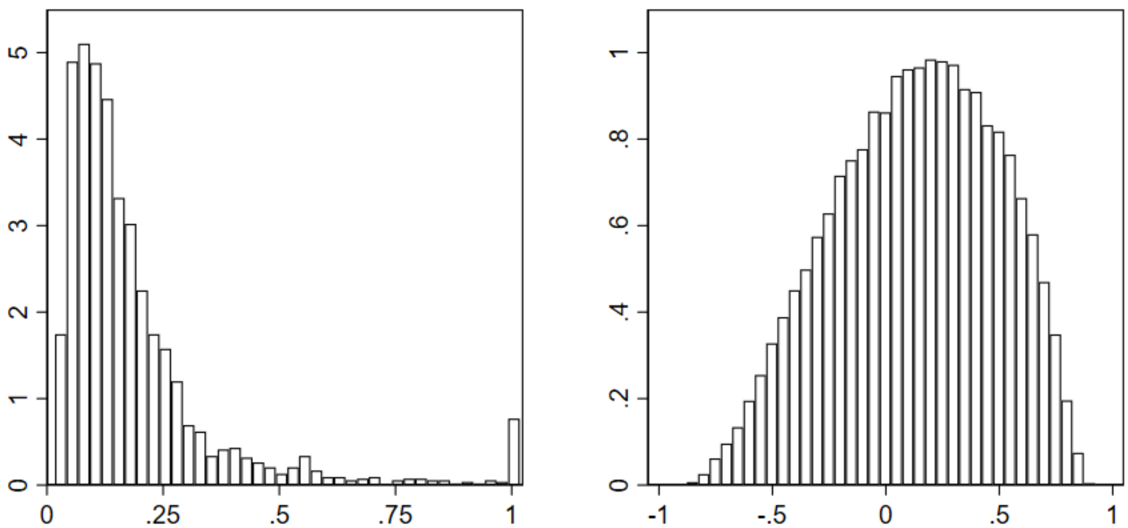

Figure 2: Histograms: Firm Volatility and Comovements

Notes: The figures plot histograms of firm volatility (Panel a) and pairwise correlations (Panel b) for 1995. Moments are first calculated for each firm (standard deviation, winsorized by one) and each intra-cluster firm pair (correlation) using data from the 1990–2000 window.

finding persists across different time periods.

5.2. Main Results: Clustered Origins of the U.S. Business Cycle

Our analysis reveals a significant and previously underappreciated driver of U.S. macroeconomic volatility: the synchronized, idiosyncratic fluctuations of firms within the same industry, which we term "clustered origins." We find a pronounced temporal pattern in the contribution of these clustered origins to aggregate volatility. Their impact fell during the stability of the Great Moderation and rose steeply during the post-2000 increase in macroeconomic volatility, closely tracking the overall business cycle. The influence of this firm-level comovement is particularly potent during turbulent periods, accounting for as much as 40% of GDP variance following the Great Recession, thereby highlighting a critical microeconomic channel for aggregate fluctuations.

Figure 3 illustrates the evolution of clustered origins in U.S. business cycles. (See Table 2 for its summary statistics.) Panel (a) compares overall U.S. real GDP volatility (black line)

Table 2: The Estimated Clustered Origins

Variable	Full sample	Subperiod sample			Additional
(Sample period average)	1980–2013	1980–1985	1986–2000	2001–2013	2014–2018
GDP volatility (std, %)	1.842	2.680	1.588	1.813	1.812
	(0.067)	(0.058)	(0.070)	(0.075)	(0.073)
Clustered origins (std, %)					
Upper-bound	0.789	1.209	0.596	0.849	0.852
	(0.037)	(0.059)	(0.018)	(0.036)	(0.045)
Lower-bound	0.613	0.887	0.437	0.711	0.738
	(0.040)	(0.062)	(0.019)	(0.039)	(0.046)
Ratio of Clustered origin variance to GDP variance (%)					
Upper-bound	19.57	20.64	15.15	24.24	23.90
Lower-bound	12.07	11.02	7.93	17.25	18.07

Notes: This table reports the contribution of within-cluster firm comovement (the 'clustered origin,' χ_t) to aggregate U.S. business cycle volatility. GDP volatility is the 11-year rolling standard deviation of the cyclical component of private GDP, estimated via an AR(1) process. Clustered origin contributions (std. dev.) are calculated as $100 \times \sqrt{|\chi_t|} \times \mathsf{sign}(\chi_t)$. The final two rows report these contributions as a percentage of total GDP variance ($100 \times \chi_t / \mathsf{var}(\widehat{\mathrm{GDP}}_t)$). Bootstrapped standard errors are provided in parentheses.

with the volatility of clustered origins (red area), measured in standard deviations. The GDP volatility follows a well-known pattern: a high level of around 2.7% in the early 1980s, followed by a sharp decline to around 1.6% during the Great Moderation, and a subsequent rise during the 2000s and 2010s due to the Great Recession and the Pandemic Crisis (Table 2).[11]

Our firm comovement volatility measure mirrors the overall pattern of U.S. aggregate volatility. The clustered component started high at 0.9–1.2 percentage points in the early 1980s, then declined sharply to approximately 0.4–0.6 percentage point in the mid-1990s, before rising again to over 1.0 percentage point following the Great Recession. These parallel movements reveal that aggregate volatility swings, including during the Great Moderation, significantly correspond to changes in within-industry firm coordination. This finding establishes firm comovement as a key force driving U.S. business cycle fluctuations.

Panel (b) of Figure 3 examines the time-varying importance of clustered origins in

[11] Aggregate GDP volatility is calculated as the standard deviation of the U.S. private economy's GDP business cycle components with a rolling window of ± 5 years, in which the business cycle components are from the residuals of AR(1) similarly to equation (28).

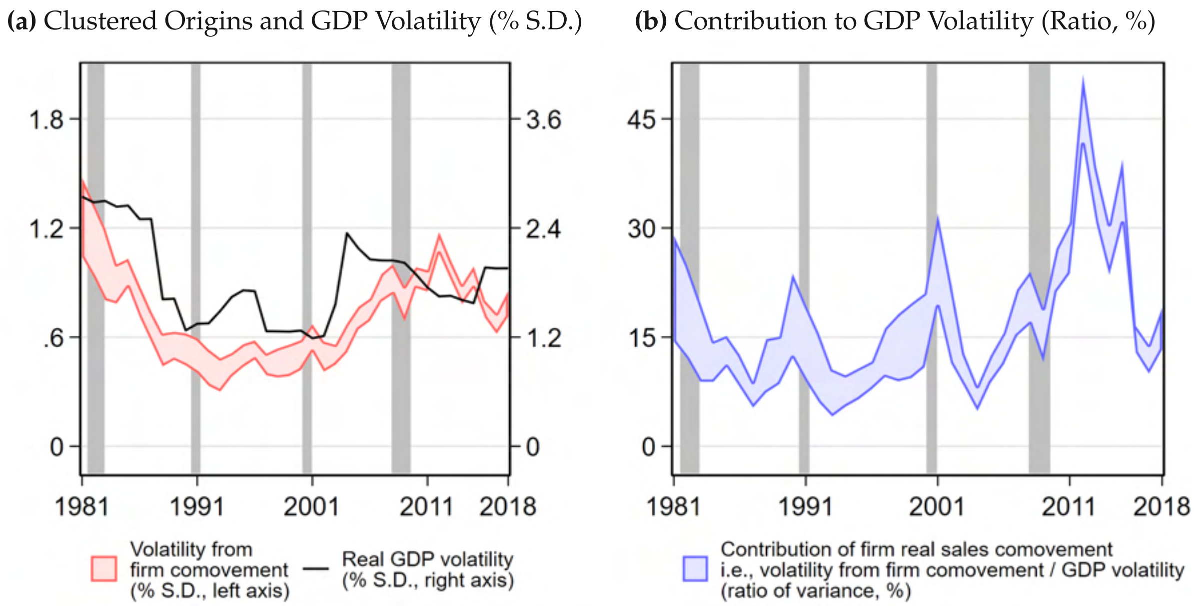

Figure 3: The Contribution of Firm Comovement to U.S. Business Cycles

Notes: The figure quantifies the contribution of within-cluster firm comovement (the 'clustered origin,' χ_t) to aggregate U.S. business cycle volatility. Panel (a) plots the standard deviation of aggregate real GDP growth (black line, right axis) against the contribution from clustered origins (red area, left axis). The clustered origin's contribution to standard deviation is calculated as $100 \times \sqrt{|\chi_t|} \times \mathsf{sign}(\chi_t)$. GDP volatility is the rolling 11-year standard deviation of the cyclical component of private GDP, estimated via an AR(1) process. Panel (b) plots the clustered origin's contribution as a percentage of total GDP variance, $100 \times \chi_t/\mathsf{var}(\widehat{\mathrm{GDP}}_t)$. Shaded gray areas denote NBER recession dates.

explaining GDP variance.[12] The share of total GDP variance attributable to clustered origins, $100 \times \chi_t/\mathsf{var}(\widehat{\mathrm{GDP}}_t)$, shows a distinct evolution, initially fluctuating between 5% and 25% during the 1980s and 1990s, before surging to over 40% of aggregate variance after the Great Recession. This time-series pattern indicates a fundamental transformation in the U.S. economy during last four decades. Moreover, this influence shows a countercyclical pattern. The contribution is roughly 10–15% in normal times but reaches its high levels during NBER-dated recessions (often rising to 15–30%), implying micro linkages are especially powerful during downturns.

Statistical Issues: Bootstrapping. While our nonparametric approach is tractable, it relies on extreme-order statistics that can be sensitive to finite-sample noise. As defined in

[12] Motivated from the aggregate GDP volatility decomposition in equation (26), we compute the ratio in terms of variance instead of standard deviation.

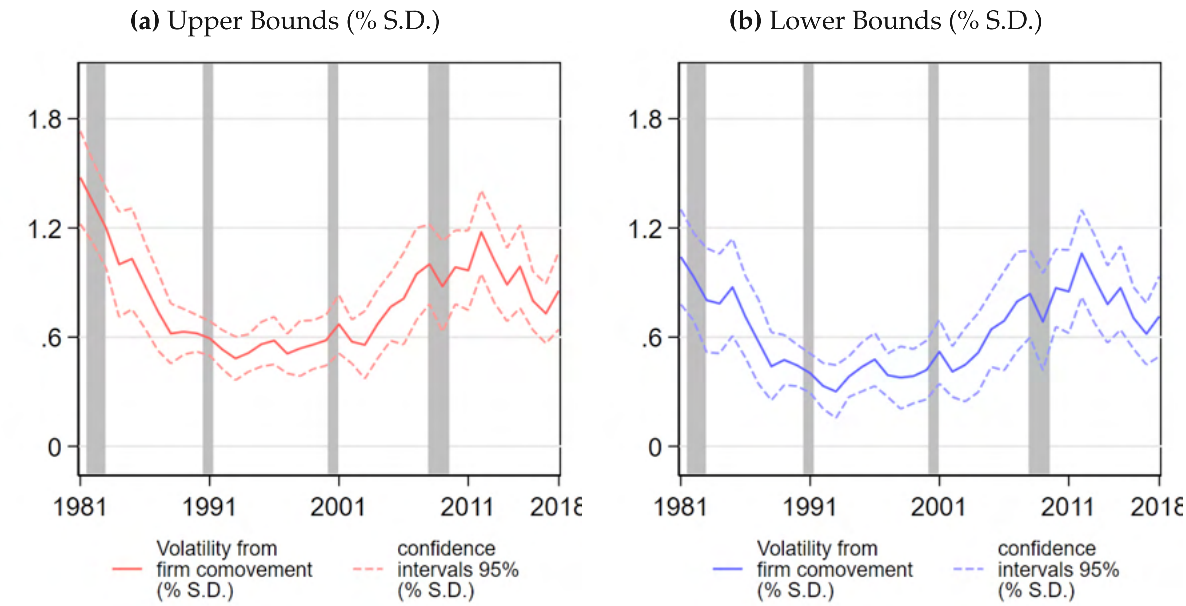

Figure 4: Clustered Origins with Confidence Intervals

Notes: The figure displays the statistical significance of the clustered origin contribution ($\chi_{\sqcup}$). Panels (a) and (b) plot the upper and lower bounds of this contribution (in standard deviation units) alongside 95% bootstrapped confidence intervals. Shaded gray areas denote NBER recession dates.

equations (22) and (23), the lower bound of the clustered origin depends on the minimum variance-covariance values within a cluster. Given our 11-year rolling windows, clusters often contain small samples where random sampling variability can produce firms with very low realized volatility. Because the bound is tied to the minimum estimated variance in the cluster, these low-variance realizations mechanically depress the estimated common component, even if the true underlying comovement is substantial.

To address this concern, we implement a cluster-based bootstrap by resampling firms with replacement within each cluster for every window. This allows us to construct standard errors and confidence intervals for both bounds. The results, reported in Table 2 and Figure 4, confirm that the time variation in clustered origins is statistically significant and robust to sampling noise. Notably, the upper-bound estimates during the Great Moderation (1986–2000) are statistically significantly lower than the lower-bound estimates of all other periods. As illustrated in Figure 4, the 95% confidence intervals remain tight throughout the sample, demonstrating that the observed U-shaped evolution of clustered origins is a structural feature of the data rather than an artifact of finite-sample variability.

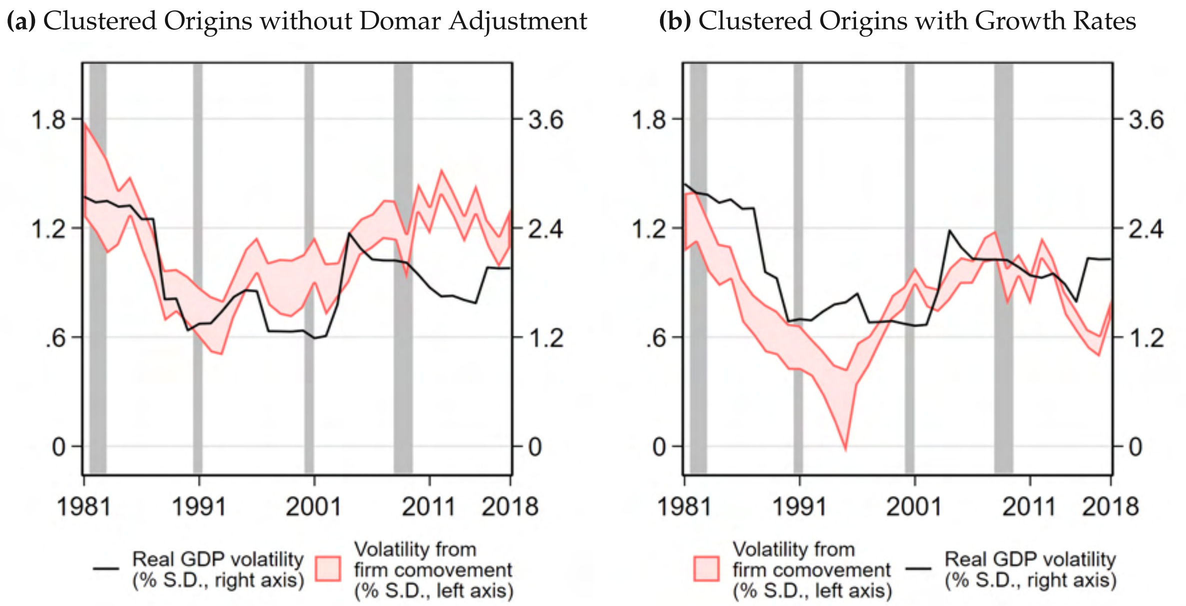

Figure 5: Robustness Check

Notes: Panel (a) recalculates the clustered origin contribution without applying the Domar adjustment, i.e., $d_t = 1$ in definition 5. Panel (b) recalculates the main result of Figure 3a using simple log-differences (growth rates) for both firm-level and aggregate GDP data instead of regression residuals.

5.3. Robustness Checks

Decomposing the Clustered Origin.. We examine whether our findings are driven by two key components from equation (26): firm-level comovements and the Domar adjustment term reflecting the economy's input-output structure. The Dormar adjustment varies between 50–80% over time, showing U-shaped pattern with low values during the Great Moderation and recent increases (Figure A3b). Panel (a) of Figure 5 shows that removing this time-varying adjustment still produces results that closely track aggregate GDP volatility. While the Domar adjustment affects absolute levels, the temporal pattern—declining during the Great Moderation and subsequent resurgence—stems primarily from changes in firm-level comovement. Additionally, market concentration (measured by HHI) shows only modest variations (7.5–9%) that do not correspond to the substantial changes in clustered origins. We plot the aggregate HHI in Figure A3a (see Appendix).

Using Growth Rates. We verify our results using simple growth rates instead of regression residuals. Panel (b) of Figure 5 demonstrates that using simple firm-level growth rates (log-differences) produces similar patterns. The close tracking between firm comovement volatility and aggregate GDP volatility persists regardless of measurement method, confirming our findings aren't artifacts of the filtering process. These checks validate the robustness of our main conclusion about the importance of firm comovement in driving aggregate volatility.

5.4. Further Discussions

Building on our findings that within-industry firm comovement substantially influences U.S. aggregate volatility, we explore three key extensions to deepen our understanding. Our investigation takes three distinct approaches. We first break down economy-wide results by industry groups to trace sectoral contributions to clustered origins over time. We then examine labor productivity alongside real sales to assess comovement in firm efficiency. Finally, we refine our measurement of idiosyncratic shocks beyond simple demeaning procedures. These extensions allow us to determine whether the U-shaped pattern of clustered origins is economy-wide or sector-specific, whether comovement extends beyond revenue to core firm efficiency, and how truly idiosyncratic shocks differ from synchronized forces. Through these analyses, we provide a more comprehensive understanding of how firm-specific and synchronized forces drive aggregate economic volatility.

Sectoral Decomposition. Is the U-shaped pattern of clustered origins a broad-based phenomenon or is it driven by a few specific sectors? Figure 6 reveals substantial sectoral heterogeneity in clustered origins. It plots the contribution of comovement to aggregate volatility for twelve major industry groups. (See Table A1 for the sector classification.) While economy-wide firm comovement generally shows positive influence on business cycles, individual sectors display varying patterns. The Information sector showed near-zero contribution in the recent decades, while Real Estate exhibited negative values in the 1990s, indicating offsetting firm-level fluctuations within these sectors. The Great Modera-

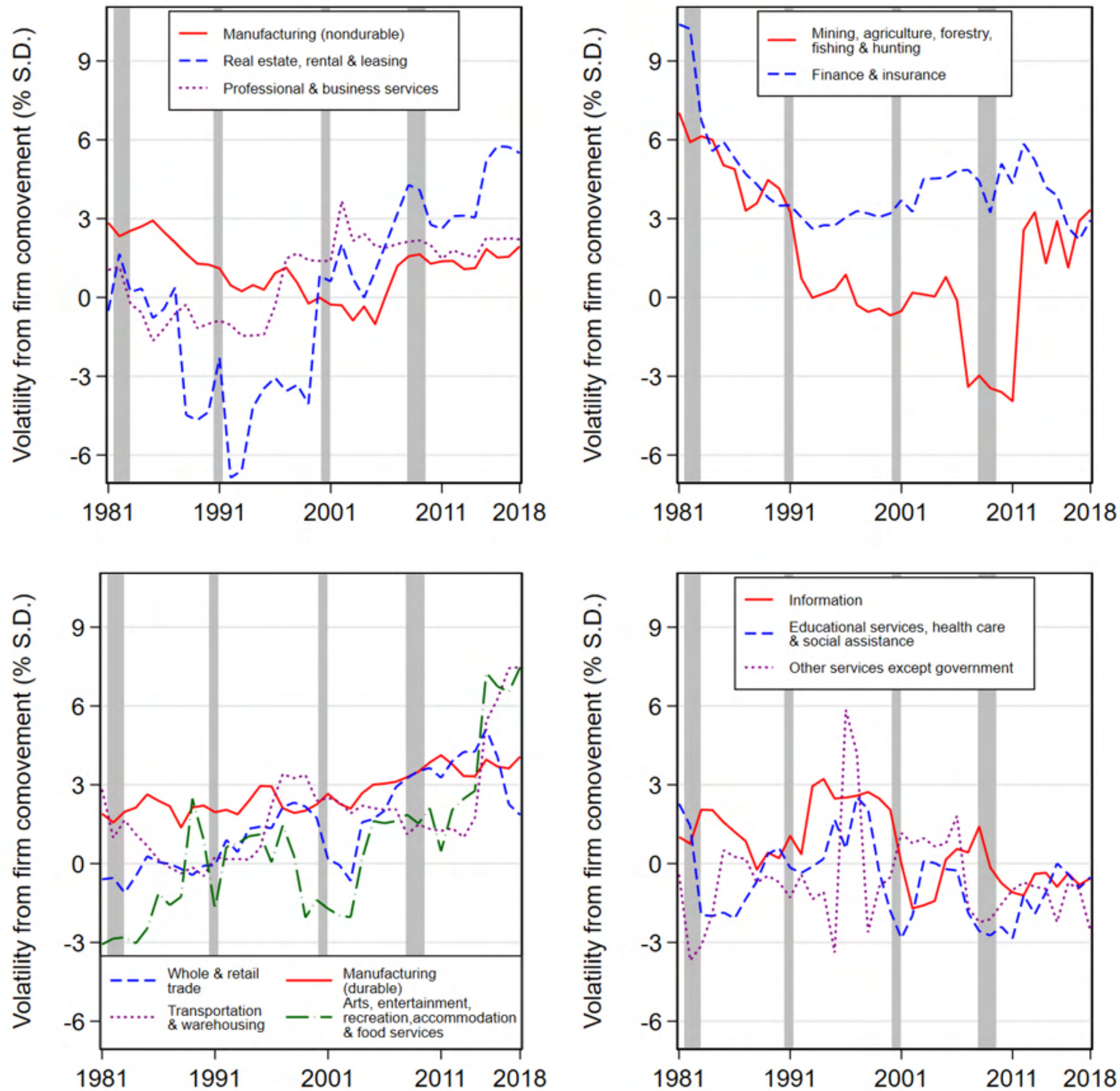

Figure 6: The Evolution of Clustered Origins by Sector

Notes: Each panel plots sector's (broadly defined 12 industry groups in Table A1) contribution of firm comovement within cluster to aggregate volatility, measured by the percent standard deviation. For clearer visualization, we plot the midpoint between the upper and lower bounds of clustered origins. Shaded areas denote NBER recession dates.

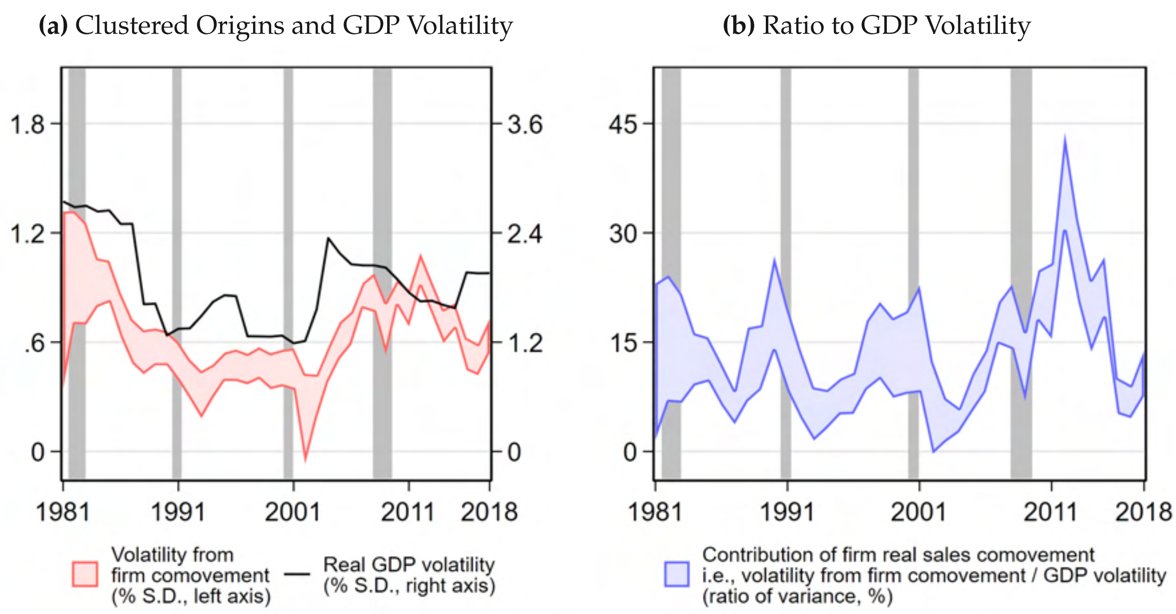

Figure 7: Clustered Origins with Labor Productivity

Notes: This figure replicates the main analysis from Figure 3 using firm-level labor productivity instead of real sales. Panel (a) plots volatility levels, and Panel (b) plots the share of aggregate variance explained by the clustered origins of labor productivity.

tion period (mid-1980s to early 1990s) demonstrated widespread decline in comovement across sectors. Finance & Insurance's contribution decreased from 9% to 3%, Real Estate, Rental & Leasing sector turned negative, and Mining & Agriculture approached zero. This broad-based decline was followed by a post-2000 resurgence, particularly evidence in Real Estate, Wholesale & Retail Trade, and Arts, Entertainment & Recreation sectors after the 2001 recession. These findings confirm that major shifts in U.S. economic stability reflect economy-wide changes rather than sector-specific phenomena, underscoring the broad-based nature of both the Great Moderation's stabilization and the subsequent increase in economic volatility.

Labor Productivity. We extend our analysis to labor productivity to examine whether firm comovement patterns reflect fundamental business cycle drivers beyond sales performance. Figure 7 reveals consistent patterns between labor productivity and our baseline sales findings. The figure shows that labor productivity results qualitatively mirror our sales findings, while their quantitatively smaller magnitudes suggest that idiosyncratic

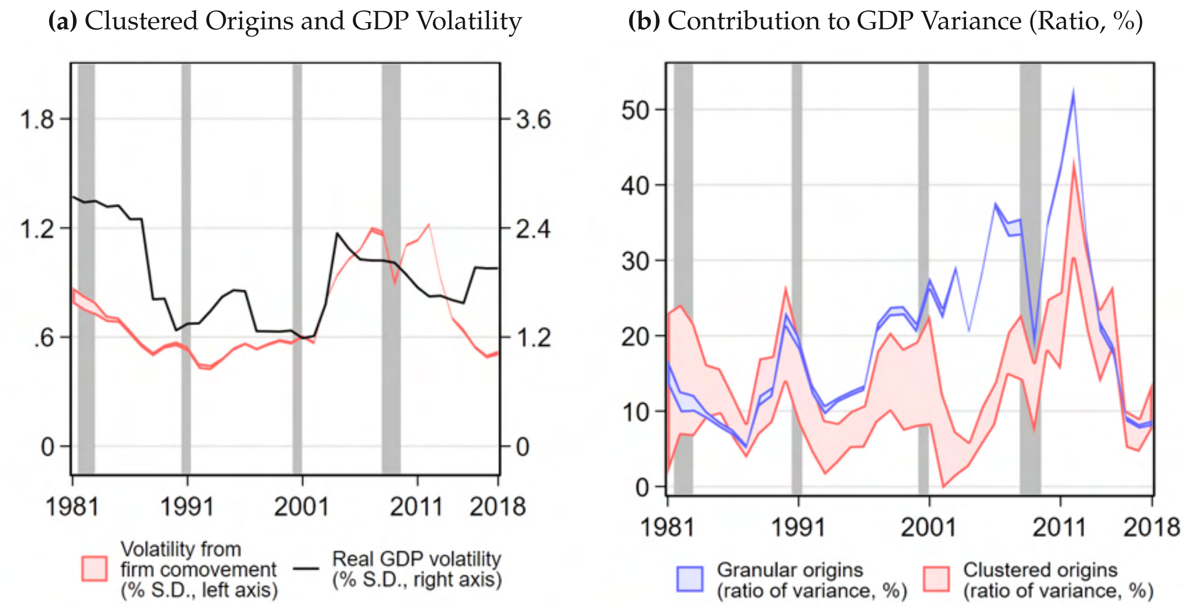

Figure 8: Granular Origins in U.S. Business Cycles

Notes: The figure quantifies the contribution of volatile idiosyncratic fluctuations (the 'granular origin,' Γ_t) to aggregate U.S. business cycle volatility. Panel (a) plots the standard deviation of aggregate real GDP growth (black line, right axis) against the contribution from granular origins (red area, left axis). In Panel (b), the blue and red areas plot granular and clustered origins' contribution as a percentage of total GDP variance ($100 \times \Gamma_t/\mathrm{var}(\widehat{\mathrm{GDP}}_t)$ and $100 \times \chi_t/\mathrm{var}(\widehat{\mathrm{GDP}}_t)$), respectively. Shaded gray areas denote NBER recession dates.

productivity fluctuations require additional channels to achieve their full macroeconomic impact. This pattern confirms that firm comovement represents a robust feature of the U.S. economy, reflecting fundamental changes in firm behavior beyond mere sales fluctuations. These findings emphasize the need to incorporate multiple endogenous comovement channels, such as network effects and linkages, into modern business cycle research to fully understand how firm-level productivity changes influence aggregate economic outcomes.

Granular Origins. Finally, we assess how our findings relate to the influential literature on 'granular origins,' which suggests that idiosyncratic fluctuations of large firms can drive aggregate fluctuations (e.g., Jovanovic, 1987; Gabaix, 2011; Carvalho and Gabaix, 2013, among many others).

Figure 8 shows both similarities and differences between granular (blue area) and clustered origins (red area). Both channels contributed 10–15% of GDP variance—through the 1980s and early 1990s, with their influence increasing significantly after the mid-1990s. However, their timing differs notably: granular origins peaked during the early 2000s dot-com bust, while clustered origins became more significant after the 2021 recession and during the Great Recession. Our methodological framework validates previous granular literature findings while identifies clustered comovement as a distinct source of aggregate fluctuations. The difference between our "true" granular measure and simpler de-meaned approaches proves minor due to the small squared HHI term. As shown in equation (21), this difference is proportional to $-h_t^2\sigma_{\mathrm{A},t}^2$. Because the squared HHI is quantitatively small, the adjustment is minor. Importantly, clustered origins show strong countercyclical behavior, consistently peaking during recessions, while granular origins display more complex dynamics not exclusively tied to economic downturns. These distinct patterns suggest that granular and clustered origins respond to different types of economic shocks, highlighting the need for economic models that incorporate both idiosyncratic firm risk and correlated industry-level dynamics.

6. Conclusion

This paper addresses a central question in macroeconomics: what are the micro-level origins of the business cycle? We argue that the prevailing focus on aggregate shocks and the idiosyncratic shocks of granular firms overlooks a critical channel—the correlated comovement of firms within industries. Standard empirical practices, such as cross-sectional demeaning, are ill-suited to detect this mechanism, as they mechanically strip these correlations from the data.

Our contribution is to develop a novel, non-parametric bounds method that quantifies the importance of these clustered origins without restrictive assumptions. Applying this framework to U.S. publicly traded firms, we show that clustered origins are a significant and dynamic driver of aggregate fluctuations. They account for 10–20% of volatility in normal periods, increase sharply during downturns, and reached nearly 40% during the Great Recession. This channel provides a new microfounded perspective on major macroeconomic episodes, including the Great Moderation and the subsequent resurgence of volatility.

These findings reshape our view of the business cycle. Aggregate volatility is not solely driven by common shocks or the fortunes of a few dominant firms, but also by the synchronized fluctuations of many firms. While our approach establishes the quantitative importance of clustered origins, it is descriptive rather than causal and remains agnostic about underlying drivers. Multiple mechanisms are likely candidates. Firm networks and input-output linkages (Oberfield, 2018; Bernard, Moxnes and Saito, 2019; Giroud and Mueller, 2019; Heise, 2024) can transmit shocks in a correlated manner. Earlier studies of industry-level interactions (Long and Plosser, 1983; Bak, Chen, Scheinkman and Woodford, 1993; Horvath, 1998; Horvath and Verbrugge, 1999; Dupor, 1999; Foerster, Sarte and Watson, 2011; Atalay, 2017) highlight channels that may operate at the firm level. Other forces, including technology spillovers, vintage capital dynamics, and adoption effects (Bloom, Schankerman and Van Reenen, 2013; Schaal and Taschereau-Dumouchel, 2018; Mullen, 2020; Fiori and Scoccianti, 2021), may also contribute to firm comovements within industries. Disentangling these forces is an important priority for future research. Doing so will help build a more complete microfounded theory of aggregate fluctuations and inform policy strategies aimed at enhancing macroeconomic stability.

References

ACEMOGLU, DARON, VASCO M. CARVALHO, ASUMAN OZDAGLAR, AND ALIREZA TAHBAZ-SALEHI (2012) "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80(5), 1977–2016.

AMITI, MARY, AND DAVID E. WEINSTEIN. (2018): "How Much Do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data," *Journal of Political Economy*, 126(2):525–587.

ATALAY, ENGHIN (2017): "How Important Are Sectoral Shocks?," *American Economic Journal: Macroeconomics*, 9(4), 254–80.

ATALAY, ENGHIN, ALI HORTAÇSU, JAMES ROBERTS, AND CHAD SYVERSON (2011): "Network Structure of Production," *Proceedings of the National Academy of Sciences*, 108(13), 5199–5202.

BAI, JUSHAN, AND SERENA NG (2013): "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics*, 176(1), 18–29.

BAK, PER, KAN CHEN, JOSE SCHEINKMAN, AND MICHAEL WOODFORD (1993): "Aggregate Fluctuations from Independent Sectoral Shocks: Self-Organized Criticality in a Model of Production and Inventory Dynamics," *Ricerche Economiche*, 47(1), 3–30.

BERNARD, ANDREW B. AND ANDREAS MOXNES AND YUKIKO U. SAITO (2019): "Production Networks, Geography, and Firm Performance," *Journal of Political Economy*, 127(2), 639–688.

BLOOM, NICHOLAS, MARK SCHANKERMAN, AND JOHN VAN REENEN (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81(4), 1347–1393.

BREMUS, FRANZISKA, CLAUDIA M. BUCH, KATHERYN N. RUSS, AND MONIKA SCHNITZER (2018): "Big Banks and Macroeconomic Outcomes: Theory and Cross-Country Evidence of Granularity," *Journal of Money, Credit and Banking*, 50(8), 1785–1825.

BUCH, CLAUDIA M., AND KATJA NEUGEBAUER (2011): "Bank-Specific Shocks and the Real Economy," *Journal of Banking & Finance*, 35(8), 2179–2187.

CARVALHO, VASCO M. (2014): "From Micro to Macro via Production Networks," *Journal of Economic Perspectives*, 28(4), 23–48.

CARVALHO, VASCO, AND XAVIER GABAIX (2013): "The Great Diversification and Its Undoing," *American Economic Review*, 103(5), 1697–1727.

CASTRO, RUI, GIAN LUCA CLEMENTI, AND YOONSOO LEE (2015): "Cross Sectoral Variation in the Volatility of Plant Level Idiosyncratic Shocks," *The Journal of Industrial Economics*, 63(1), 1–29.

CHUN, HYUNBAE, JUNG-WOOK KIM, RANDALL MORCK, AND BERNARD YEUNG (2008): "Creative Destruction and Firm-Specific Performance Heterogeneity," *Journal of Financial Economics*, 89(1), 109–135.

COMIN, DIEGO, AND SUNIL MULANI (2006): "Diverging Trends in Aggregate and Firm Volatility," *The Review of Economics and Statistics*, 88(2), 374–383.

COMIN, DIEGO, AND THOMAS PHILIPPON (2005): "The Rise in Firm-Level Volatility: Causes and Consequences," *NBER Macroeconomics Annual*, 20, 167–201.

DAVIS, STEVEN J., JOHN HALTIWANGER, RON JARMIN, JAVIER MIRANDA, CHRISTOPHER FOOTE, AND EVA NAGYPAL (2006): "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms," *NBER Macroeconomics Annual*, 21, 107–179.

DOMAR, EVSEY D. (1961): "On the Measurement of Technological Change," *The Economic Journal*, 71(284), 709–729.

DUPOR, BILL (1999): "Aggregation and Irrelevance in Multi-Sector Models," *Journal of Monetary Economics*, 43(2), 391–409.

EVANS, DAVID S. (1987): "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," *The Journal of Industrial Economics*, 35(4), 567–581.

FIORI, GIUSEPPE, AND FILIPPO SCOCCIANTI (2021): "Aggregate Dynamics and Microeconomic Heterogeneity: The Role of Vintage Technology," Unpublished Manuscript, North Carolina State University.

FOERSTER, ANDREW T., PIERRE-DANIEL G. SARTE, AND MARK W. WATSON (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119(1), 1–38.

FORNERON, JEAN-JACQUES (2023): "A Sieve-SMM Estimator for Dynamic Models," *Econometrica*, 91(3), 943–977.

GABAIX, XAVIER (2011): "The granular Origins of Aggregate Fluctuations," *Econometrica*, 79(3), 733–772.

GAUBERT, CECILE, AND OLEG ITSKHOKI (2021): "Granular Comparative Advantage." *Journal of Political Economy*, 129(3), 871–939.

GIROUD, XAVIER, AND HOLGER M. MUELLER (2019): "Firms' Internal Networks and Local Economic Shocks." *American Economic Review*, 109(10), 3617–49.

DI GIOVANNI, JULIAN AND ANDREI A. LEVCHENKO (2012): "Country Size, International Trade, and Aggregate Fluctuations," *Journal of Political Economy*, 120(6), 1083–1132.

DI GIOVANNI, JULIAN, ANDREI A. LEVCHENKO, AND ISABELLE MEJEAN (2014): "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 82(4), 1303–1340.

GUVENEN, FATIH, FATIH KARAHAN, SERDAR OZKAN, AND JAE SONG (2021): "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?," *Econometrica*, 89(5), 2303–2339.

HALL, BRONWYN H. (1987): "The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector," *The Journal of Industrial Economics*, 35(4), 583–606.

HERSKOVIC, BERNARD, BRYAN KELLY, HANNO LUSTIG, AND STIJN VAN NIEUWERBURGH (2020): "Firm Volatility in Granular Networks," *Journal of Political Economy*, 128(11), 4097–4162.

HEISE, SEBASTIAN (2024): "Firm-to-Firm Relationships and the Pass-Through of Shocks: Theory and Evidence." *The Review of Economics and Statistics*, Forthcoming.

HORVATH, MICHAEL (1998): "Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks," *Review of Economic Dynamics*, 1(4), 781–808.

HORVATH, MICHAEL AND RANDAL VERBRUGGE (1999): "Shocks and Sectoral Interactions: An Empirical Investigation," Unpublished Manuscript, Stanford University.

HULTEN, CHARLES R. (1978): "Growth Accounting with Intermediate Inputs," *The Review of Economic Studies*, 45(3), 511–518.

JAIMOVICH, NIR, STEPHEN J. TERRY, AND NICOLAS VINCENT (2023): "The Empirical Distribution of Firm Dynamics and Its Macro Implications." Working Paper 31337, National Bureau of Economic Research.

JOVANOVIC, BOYAN (1987): "Micro Shocks and Aggregate Risk," *The Quarterly Journal of Economics*, 102(2), 395–409.

KALNINA, ILZE AND KOKOUVI TEWOU (2025): "Cross-Sectional Dependence in Idiosyncratic Volatility," *Journal of Econometrics*, 249, No. 106003.

KIM, DAISOON, JEE WON PARK AND INHWAN SO (2025): "Investment Giants in Emerging Markets," Working Paper 2025-9, Bank of Korea.

KIM, CHANG-JIN AND CHARLES R. NELSON (1999): "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics*, 81(4), 608–616.

LONG, JOHN B., AND CHARLES I. PLOSSER. (1983): "Real Business Cycles," *Journal of Political Economy*, 91(1), 39–69.

LUCAS, ROBERT E. (1977): "Understanding Business Cycles," *Carnegie-Rochester Conference Series on Public Policy*, 5, 7–29.

MULLEN, RORY (2020): "On Aggregate Fluctuations, Systemic Risk, and Covariance of Firm-Level Activity," Unpublished Manuscript, Warwick Business School.

OBERFIELD, EZRA (2018): "A Theory of Input-Output Architecture," *Econometrica*, 86(2), 559–589.

SCHAAL, EDOUARD, AND MATHIEU TASCHEREAU-DUMOUCHEL (2018): "Herding, Technology Adoption and Boom-Bust Cycles," 2018 Meeting Papers 111, Society for Economic Dynamics.

STANLEY, MICHAEL H. R., LUIS A. N. AMARAL, SERGEY V. BULDYREV, SHLOMO HAVLIN, HEIKO LESCHHORN, PHILIPP MAASS, MICHAEL A. SALINGER, AND H. EUGENE STANLEY (1996): "Scaling Behaviour in the Growth of Companies," *Nature*, 379(6568), 804–806.

STERK, VINCENT, PETR SEDLACEK, AND BENJAMIN PUGSLEY (2021): "The Nature of Firm Growth," *American Economic Review*, 111(2), 547–79.

STOCK, JAMES H. AND MARK W. WATSON (2002): "Has the Business Cycle Changed and Why?," *NBER Macroeconomics Annual*, 17, 159–218.

TWEEDLE, JESSE (2018): "Correlated Shocks within Firms," *Economics Letters*, 163, 95–97.

XU, YEXIAO, AND BURTON G. MALKIEL (2003): "Investigating the Behavior of Idiosyncratic Volatility," *The Journal of Business*, 76(4), 613–645.

APPENDIX

A. Simulation Exercise

First, we generate 5,000 firms' true idiosyncratic shocks ($\varepsilon_{\mathrm{F},it}$) during 50 periods which are randomly generated from a multi-normal distribution with mean zero, 12% standard deviation, and 2.5% correlation. Figure A1 reports results with a time-invariant size distribution with fat-tails where the distribution is generated from Pareto distribution with shape parameter 1.2 on support $[1, \infty)$, which yields approximately a 13% Herfindahl-Hirschman index.

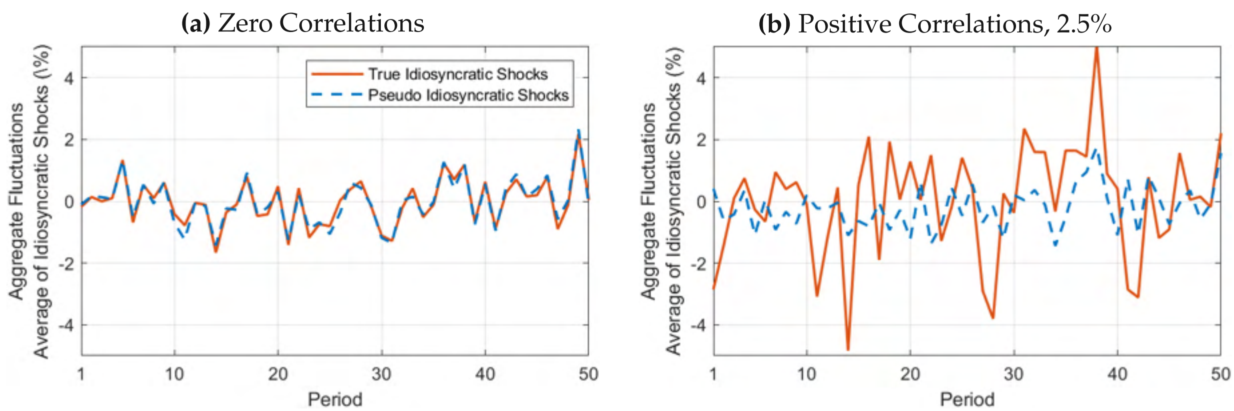

Figure A1: True vs.Pseudo Idiosyncratic Shocks and Their Aggregate Fluctuations with fat tails

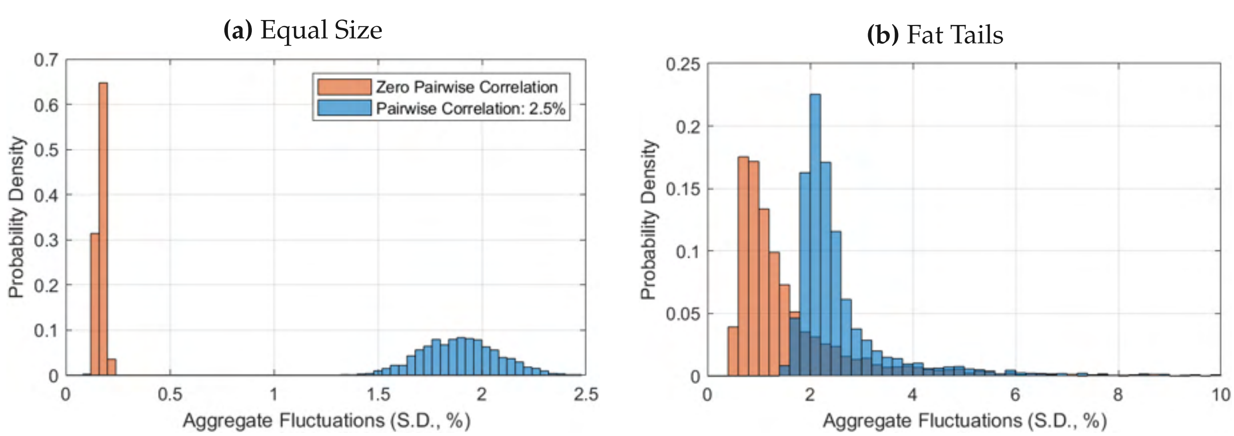

Figure A2: Sample Standard Deviations of Aggregate Fluctuations

Compared to Figure 1, Figure A1 shows larger aggregate fluctuations due to granularity. Correlations still generate additional aggregate fluctuations. The solid orange line of Panel B with correlations is more volatile than Panel A's solid orange line without correlations. However, the dashed blue lines constructed by pseudo shocks have indistinguishable volatilities between Panels A and B. Second, we redo the above exercise 3,000 times and calculate the sample statistics of true and pseudo idiosyncratic shocks. Figure A2 plots histograms for the aggregate volatility with and without unequal size distributions. Positive correlations lead to sizable aggregate fluctuations.

B. Arbitrary Weights in Section 4

Consider an arbitrary weight, $\{\mathrm{w}_{it}\}_i$, satisfying $\sum_{i'} \mathrm{w}_{i't} = 1$ and $\mathrm{w}_{it} \geq 0$. Define pseudo common and idiosyncratic factors based on the weighted mean, as follows.

$$e^{\mathrm{w}}_{\mathrm{A},t} = \sum_{i'} \mathrm{w}_{i't}\hat{y}_{i't} = \varepsilon_{\mathrm{A},t} + \sum_{i'} \mathrm{w}_{i't}\varepsilon_{\mathrm{F},i't} \tag{A1}$$

$$e^{\mathrm{w}}_{\mathrm{F},it} = \hat{y}_{it} - e_{\mathrm{A},t} = \varepsilon_{\mathrm{F},it} - \textstyle\sum_{i'} \mathrm{w}_{i't}\varepsilon_{\mathrm{F},i't} \tag{A2}$$

Then, the variance of pseudo idiosyncratic shocks is

$$\mathsf{var}(e^{\mathrm{w}}_{\mathrm{F},it}) = (1 - 2\mathrm{w}_{it} + \mathrm{m}^{\mathrm{w}}_2)(1 - \rho_{\mathrm{F},t})\sigma^2_{\mathrm{F},t} \tag{A3}$$

where $\mathrm{m}^{\mathrm{w}}_2 = \sum_{i'} \mathrm{w}^2_{i't} \in [N_t^{-1}, 1]$ measures how much equally weighted. Also, the pairwise correlation of pseudo idiosyncratic shocks are

$$\mathsf{corr}(e_{\mathrm{F},it}, e_{\mathrm{F},i't}) = -\frac{\mathrm{w}_{it} + \mathrm{w}_{i't} - \mathrm{m}^{\mathrm{w}}_2}{\sqrt{1 - 2\mathrm{w}_{it} + \mathrm{m}^{\mathrm{w}}_2}\sqrt{1 - 2\mathrm{w}_{i't} + \mathrm{m}^{\mathrm{w}}_2}}, \tag{A4}$$

which is independent of the pairwise correlation of true idiosyncratic shocks denoted by $\rho_{\mathrm{F},t}$. If the weights are unequal, the idiosyncratic shocks correlate with the pseudo common in contrast to the equal wights case with homogeneous variance and covariance.

$$\mathsf{corr}(e_{\mathrm{A},t}, e_{\mathrm{F},it}) = -\frac{\mathrm{w}_{it} - \mathrm{m}_2^{\mathrm{w}}}{\sqrt{\frac{\sigma_{\mathrm{A},t}^2/\sigma_{\mathrm{F},t}^2 + \rho_{\mathrm{F},t}}{1-\rho_{\mathrm{F},t}} + \mathrm{m}_2^{\mathrm{w}}}\sqrt{1 - 2\mathrm{w}_{it} + \mathrm{m}_2^{\mathrm{w}}}}, \tag{A5}$$

where the equal weights — $\forall i$, $\mathrm{w}_{it} = N_t^{-1}$ — lead the pseudo common and idiosyncratic shocks to be uncorrelated for all firms; $\forall i$, $\mathsf{corr}(e_{\mathrm{A},t}, e_{\mathrm{F},it}) = 0$.

C. Data and Measurements

[Step 1] We correct the industry-level deflators ($\mathtt{p}_{st}$) — Chain-Type Price Indexes for Gross Output by Industry [2012=100] — from the U.S. Bureau of Economic Analysis (BEA) database. Sales ($\mathtt{sale}_{it}$) and employments ($\mathtt{emp}_{it}$) are directly from the Compustat North America: Fundamental Annuals (1975–2018) databases.

[Step 2] We construct the sample as follows. First, we keep the following observations in the Compustat database.

- No major mergers flag: Comparability status ($\mathtt{compst}_{it}$) does not equal to AB.
- Country ISO 3 digit code ($\mathtt{loc}_{it}$): USA
- Currency ISO 3 digit code ($\mathtt{curcd}_{it}$): USD

Then, we exclude firms with the following criteria.

- Non-positive sales
- Non-positive employments
- Utilities sector (NAICS 22)
- Public administration sector (NAICS 91–92)

[Step 3] We merge the Compustat sample and the industry-level BEA deflator using Table A1. We calculate the logged labor productivity as real sales divided by employments ($\ln \mathtt{sale}_{it} - \ln \mathtt{p}_{st} - \ln \mathtt{emp}_{it}$) for firm i in industry s at t.

[Step 4] Since some clusters have low observations, we merge them. See Table A1 for the list of clusters.

D. Further Discussion: Factor Model Interpretation

In this section, we explore an alternative approach to address pairwise correlations in the framework discussed in the preceding sections by introducing a factor model. This model assumes that comovements among idiosyncratic shocks are driven by reactions to an underlying latent common factor.

We begin by decomposing the diosyncratic shock into the latent factor and pairwise uncorrelated shocks, i.e., $\varepsilon_{\mathrm{F},it} = \lambda_{it} f_t + u_{it}$ where $\varepsilon_{\mathrm{A},t}$, f_t, u_{it} and $u_{i't}$ are uncorrelated for any firm $i \neq i'$. In this factor model, it is expressed as:

$$\hat{y}_{it} = \varepsilon_{\mathrm{A},t} + \underbrace{\lambda_{it} f_t + u_{it}}_{=\varepsilon_{\mathrm{F},it}}. \tag{A6}$$

This factor model allows for variation in firms' responses to the latent common factor, and it provides an alternative representation of our business cycle components. The resulting variance and covariance for the business cycle components are expressed as:

$$\mathsf{var}(\hat{y}_{it}) = \sigma^2_{\mathrm{A},t} + \mathsf{var}(\lambda_{it} f_t + u_{it}) = \sigma^2_{\mathrm{A},t} + \lambda^2_{it}\sigma^2_{f,t} + \sigma^2_{u,it}, \tag{A7}$$

$$\mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't}) = \sigma^2_{\mathrm{A},t} + \mathsf{cov}(\lambda_{it} f_t, \lambda_{i't} f_t) = \sigma^2_{\mathrm{A},t} + \lambda_{it}\lambda_{i't}\sigma^2_{f,t}, \tag{A8}$$

where $\lambda^2_{it}\sigma^2_{f,t} + \sigma^2_{u,it}$ and $\lambda_{it}\lambda_{i't}\sigma^2_{f,t}$ are the corresponding part of $\sigma^2_{\mathrm{F},it}$ and $\rho_{\mathrm{FF},ii't}\sigma_{\mathrm{F},it}\sigma_{\mathrm{F},i't}$ previously seen in equations (18) and (19) within our benchmark framework.

Within this factor model, pairwise comovements among idiosyncratic shocks result from the latent factor. This means that differences between true and pseudo shocks do not necessarily converge to zero. In fact, this difference can be quantified as:

$$\mathbb{E}\left[|e_{\mathrm{A},t} - \varepsilon_{\mathrm{A},t}|^2\right] = \mathbb{E}\left[|e_{\mathrm{F},it} - \varepsilon_{\mathrm{F},it}|^2\right] = \frac{1}{N_t}\overline{\sigma}^2_{u,t} + \overline{\lambda}^2_t \sigma^2_{f,t},, \tag{A9}$$

where $\overline{\lambda}_t = N_t^{-1}\sum_i \lambda_{it}$. If $\overline{\lambda}_t$ equals zero, it implies negligible pairwise comovements, making $\overline{\text{cov}}_{\text{FF},it}$ and $\overline{\overline{\text{cov}}}_{\text{FF},t}$ approach zero, as discussed in section 3.[13] In light of these findings, it's essential to recognize that sample means and deviations are not ideal tools for studying business cycle fluctuations and their origins when significant pairwise comovements exist.

Similarly, the pseudo factor variance, based on a sample mean, does not provide a consistent estimator. The variances are given by:

$$\text{var}(e_{\text{A},t}) = \sigma^2_{\text{A},t} + \overline{\lambda}^2_t \sigma^2_{f,t} + \frac{1}{N_t}\overline{\sigma}^2_{u,t} \tag{A12}$$

$$\text{var}(e_{\text{F},it}) = \left(1 - \frac{1}{N_t}\right)\sigma^2_{u,it} + (\lambda_{it} - \overline{\lambda}_t)^2\sigma^2_{f,t} - \frac{1}{N_t}\big(\sigma^2_{u,it} - \overline{\sigma}^2_{u,t}\big), \tag{A13}$$

where $\overline{\sigma}^2_{u,t} = N_t^{-1}\sum_i \sigma^2_{u,it}$, and $\overline{\lambda}_t = N_t^{-1}\sum_i \lambda_{it}$ represent averages. These values are not suitable for estimating the true common and idiosyncratic shock variances ($\text{var}(\varepsilon_{\text{A},t})$ and $\text{var}(\varepsilon_{\text{F},it})$) unless pairwise comovements are nearly absent (i.e., when $\overline{\lambda}_t = 0$).

The covariances of pseudo common and idiosyncratic shocks also present challenges. Specifically:

$$\text{cov}(e_{\text{A},t}, e_{\text{F},it}) = (\lambda_{it} - \overline{\lambda}_t)\overline{\lambda}_t\sigma^2_{f,t} + \frac{1}{N_t}\big(\sigma^2_{u,it} - \overline{\sigma}^2_{u,t}\big). \tag{A14}$$

This covariance tends to zero as the number of firms approaches infinity in the scenario where $\overline{\lambda}_t = 0$. It also approaches uncorrelated status as the factor coefficients become uniform across firms ($\lambda_{it} = \overline{\lambda}_t$). However, in this homogeneous case, the pseudo factors represent well-defined macro and micro shocks.[14] It is important to note that an identical coefficient implies identical covariance between business cycle components, which contradicts the empirical data. Finally, it's worth noting that the pairwise covariance of pseudo

[13]Note that $\overline{\text{cov}}_{\text{FF},it} = (N_t-1)^{-1}\sum_{i'\neq i}\text{cov}(\varepsilon_{\text{F},it}, \varepsilon_{\text{F},i't}) = (N_t-1)^{-1}\sum_{i'\neq i}\text{cov}(\lambda_{it}f_t, \lambda_{i't}f_t)$ implies

$$\overline{\text{cov}}_{\text{FF},it} = \frac{1}{N_t-1}\big(-\lambda_{it} + \sum_{i'}\lambda_{i't}\big)\sigma^2_{f,t} = \frac{N_t}{N_t-1}\left(\overline{\lambda}_t - \frac{\lambda_{it}}{N_t}\right)\lambda_{it}\sigma^2_{f,t} \tag{A10}$$

$$\overline{\overline{\text{cov}}}_{\text{FF},t} = \frac{1}{N_t}\sum_i \overline{\text{cov}}_{\text{FF},it} = \frac{N_t}{N_t-1}\left(\overline{\lambda}^2_t - \frac{1}{N_t^2}\sum_i \lambda^2_{it}\right)\sigma^2_{f,t}. \tag{A11}$$

[14]Also, the model of $\hat{y}_{it} = \varepsilon_{\text{A},t} + \varepsilon_{\text{F},it}$ with $\varepsilon_{\text{F},it} = \overline{\lambda}_t f_t + u_{it}$ versus the model of $\hat{y}_{it} = \tilde{\varepsilon}_{\text{A},t} + \tilde{\varepsilon}_{\text{F},it}$ with $\tilde{\varepsilon}_{\text{A},t} = \varepsilon_{\text{A},t} + \overline{\lambda}_t f_t$ and $\tilde{\varepsilon}_{\text{F},it} = u_{it}$ cannot be identified.

idiosyncratic shocks tends to disregard the actual covariance:

$$\mathsf{cov}(e_{\mathrm{F},it}, e_{\mathrm{F},i't}) = (\lambda_{it} - \overline{\lambda}_t)(\lambda_{i't} - \overline{\lambda}_t)\sigma^2_{f,t} - \frac{1}{N_t}\overline{\sigma}^2_{u,t} - \frac{1}{N_t}\left[(\sigma^2_{u,it} - \overline{\sigma}^2_{u,t}) + (\sigma^2_{u,i't} - \overline{\sigma}^2_{u,t})\right] \tag{A15}$$

In summary, the conclusions drawn in the previous sections remain valid when considering pairwise comovements that arise from heterogeneous responses to an underlying common shock. Researchers should be cautious when using sample means and deviations and consider implementing zero normalization of average coefficients (as in Bai and Ng, 2013) when examining firms' interdependencies and networks.

E. Additional Figures and Tables

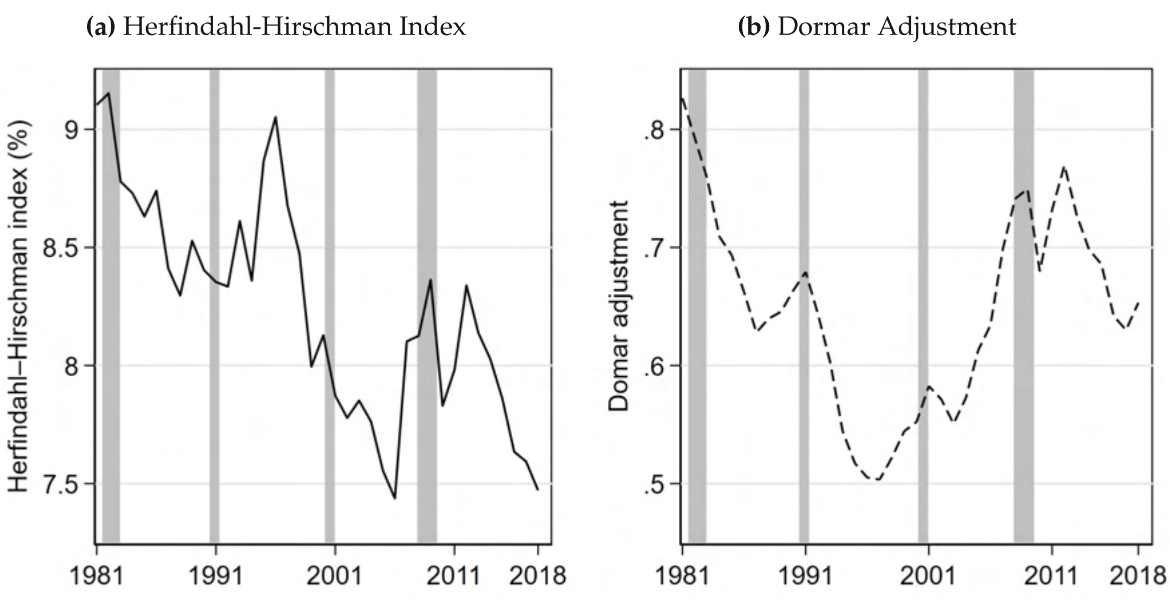

Figure A3: Herfindahl-Hirschman Index and Dormar Adjustment

(a) Standard Deviation **(b)** Cross-Firm Correlation within Cluster

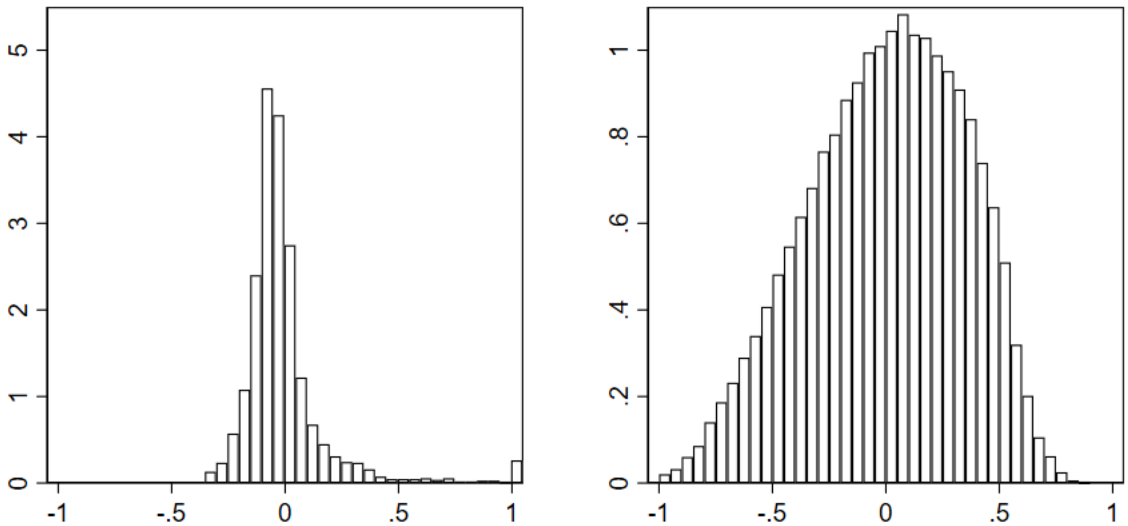

Figure A4: Histograms: Firm Volatility and Comovements

Notes: The figures plot histograms of within-cluster demeaned firm volatility (Panel a) and pairwise correlations (Panel b) for 1995. Moments are first calculated for each firm (standard deviation) and each intra-cluster firm pair (correlation) using data from the 1990–2000 window. We then demean these values by subtracting the corresponding cluster-specific mean. The plots visualize the remaining cross-sectional variation.

Table A1: List of Clusters

Cluster	BEA Industry	NAICS	
		1997	2017
1. Agriculture, forestry, fishing, and hunting	· Farms	111–2	111–2
	· Forestry, fishing, and related activities	113–5	113–5
2. Oil and gas extraction	· Oil and gas extraction	211	211
3. Mining, except oil and gas	· Mining, except oil and gas	212	212
4. Support activities for mining	· Support activities for mining	213	213
5. Construction	· Construction	230	230
6. Wood products	· Wood products	321	321
7. Nonmetallic mineral products	· Nonmetallic mineral products	327	327
8. Primary metals	· Primary metals	331	331
9. Fabricated metal products	· Fabricated metal products	332	332
10. Machinery	· Machinery	333	333
11. Computer and electronic products	· Computer and electronic products	334	334
12. Electrical equipment, appliances, and components	· Electrical equipment, appliances, and components	335	335
13. Motor vehicles, bodies and trailers, and parts, and Other transportation	· Motor vehicles, bodies and trailers, and parts	3361–6	3361–6
	· Other transportation equipment	3369	3369
14. Furniture and related products	Furniture and related products	337	337
15. Miscellaneous manufacturing	Miscellaneous manufacturing	339	339
16. Food and beverage and tobacco products	· Food and beverage and tobacco products	311–2	311–2
17. Textile mills and textile product mills	· Textile mills and textile product mills	313–4	313–4
18. Apparel and leather and allied products	· Apparel and leather and allied products	315–6	315–6
19. Paper products	· Paper products	322	322
20. Printing and related support activities	· Printing and related support activities	323	323
21. Petroleum and coal products	· Petroleum and coal products	324	324
22. Chemical products	· Chemical products	325	325
23. Plastics and rubber products	· Plastics and rubber products	326	326
24. Wholesale trade	· Wholesale trade	420	420

Continued on next page

Table A1 — continued from previous page

Cluster	BEA Industry	NAICS 1997	NAICS 2017
25. Retail trade	· Motor vehicle and parts dealers · Food and beverage stores · General merchandise stores · Other retail	441 445 452 442-4, 446-8, 451, 453-4	441 445 452 442-4, 446-8, 451, 453-4
26. Air transportation	· Air transportation	481	481
27. Rail transportation	· Rail transportation	482	482
28. Water transportation	· Water transportation	483	483
29. Truck transportation	· Truck transportation	484	484
30. Pipeline transportation	· Pipeline transportation	486	486
31. Other transportation (transit and ground) and support activities, and Warehousing and storage	· Transit and ground passenger transportation · Other transportation and support activities · Warehousing and storage	485 487-8, 491-2 493	485 487-8, 491-2 493
32. Publishing industries, except internet (includes software)	· Publishing industries, except internet (includes software)	511	511
33. Motion picture and sound recording industries	· Motion picture and sound recording industries	512	512
34. Broadcasting and telecommunications	· Broadcasting and telecommunications	513	515, 517
35. Data processing, internet publishing, and other information services	Data processing, internet publishing, and other information services	514	518–9
36. Federal Reserve banks, credit intermediation, and related activities	· Federal Reserve banks, credit intermediation, and related activities	521–2	521–2
37. Securities, commodity contracts, and investments	· Securities, commodity contracts, and investments	523	523
38. Insurance carriers and related activities	· Insurance carriers and related activities	524	524
39. Funds, trusts, and other financial vehicles	· Funds, trusts, and other financial vehicles	525	525
40. Real estate	· Real estate	531	531
41. Rental and leasing services and lessors of intangible assets	· Rental and leasing services and lessors of intangible assets	532–3	532–3
42. Computer systems design and related services	· Computer systems design and related services	5415	5415

Continued on next page

Table A1 — continued from previous page

Cluster	BEA Industry	NAICS	
		1997	2017
43. Legal services, and miscellaneous professional, scientific, and technical services	· Legal services	5411	5411
	· Miscellaneous professional, scientific, and technical services	5412–4, 5416–9	5412–4, 5416–9
44. Administrative and support services	· Administrative and support services	561	561
45. Waste management and remediation services	· Waste management and remediation services	562	562
46. Educational services	· Educational services	610	610
47. Ambulatory health care services	· Ambulatory health care services	621	621
48. Hospitals, Nursing and residential care facilities, and social assistance	· Hospitals	622	622
	· Nursing and residential care facilities	623	623
	· Social assistance	624	624
49. Arts, entertainment, and recreation	· Performing arts, spectator sports, museums, and related activities	711–2	711–2
	· Amusements, gambling, and recreation industries	713	713
50. Accommodation	· Accommodation	721	721
51. Food services and drinking places	· Food services and drinking places	722	722
52. Other services, except government	· Other services, except government	810	810

Table A2: List of Sectors

Sector	Clusters
Mining, agriculture, forestry, fishing, & hunting	1–5
Manufacturing (Durable)	6–15
Manufacturing (Nondurable)	16–23
Wholesale & retail trade	24–25
Transportation & warehousing	26–31
Information	32–35
Finance & insurance	36–39
Real estate, rental & leasing	40–41
Professional & business services	42–45
Educational services, health care & social assistance	46–48
Arts, entertainment, recreation, accommodation & food services	49–51
Other services, except government	52