Aggregate Fluctuations from Firm Comovement

Nahyeon Bak* Daisoon Kim†

May 2026

Abstract

This paper demonstrates that correlated idiosyncratic fluctuations among firms are an important source of macroeconomic volatility. Standard methods based on cross-sectional demeaning obscure this channel by mechanically driving average pairwise correlations to zero asymptotically. We develop a nonparametric bounds approach quantifying the contribution of cross-firm comovement within industries directly from firm-level data. Applied to the United States economy, we find that such clustered comovement explains 10–15% of GDP volatility in normal times, 15–30% in downturns, and up to 45% following the Great Recession, while also helping account for the Great Moderation and its undoing.

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 Eric Swanson (Editor), anonymous referee, 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.

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 large firms can translate into significant aggregate fluctuations.

In this paper, we uncover another micro-level source of aggregate volatility: within-cluster firm comovement, which we term *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-sized 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—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 nonparametric 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 45% following turbulent periods such as the Great Recession. Third, their time path closely tracks major shifts in economic volatility: their subsequent decline following the Great Moderation coincides with falling GDP volatility, while their resurgence after 2000 parallels the concurrent rise in aggregate 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 role of microeconomic shocks and firm heterogeneity 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 literature, our paper introduces a complementary microeconomic channel—clustered origins—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 fluctuations, 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 works examine 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 Giovanni and Levchenko, 2012; Gaubert and Itskhoki, 2021). Departing from these works, we focus on within-cluster pairwise correlations among firms, offering a novel lens to capture correlated idiosyncratic fluctuations at the firm level.

Our contribution also extends the literature on the industrial and sectoral origins of macroeconomic volatility, tracing back to Long and Plosser (1983). This 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., 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 aligns closely with 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), consistent with theoretical arguments 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 organized as follows. Section 2 introduces the statistical framework. Section 3 demonstrates the pitfalls of cross-sectional demeaning. Section 4 develops the empirical methodology. Section 5 presents the main empirical findings. 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) is expressed 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, capturing 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, $\rho_{\mathrm{FF},ii't} \equiv \mathsf{corr}(\varepsilon_{\mathrm{F},it}, \varepsilon_{\mathrm{F},i't})$, to be nonzero for $i \neq i'$. This correlation can generate aggregate-level fluctuations and foster comovements ($\mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})$) 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 satisfies $\sum_i w_{it} = 1$.[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*

[2] These pairwise correlations prevent the average idiosyncratic component from vanishing, contributing to notable short-term departures of the cross-sectional mean from zero, i.e., $N_t^{-1}\sum_i \varepsilon_{\mathrm{F},it} \neq 0$. However, since correlations affect second moments rather than first moments, they leave the long-run level unaffected: $\mathbb{E}[\varepsilon_{\mathrm{F},it}] = 0$ and $\lim_{T\to\infty} T^{-1}\sum_t \varepsilon_{\mathrm{F},it} = 0$.

[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.

synchronized idiosyncratic movements to aggregate volatility through covariance of firm-specific components.

A Simple Example of Clustered Origins. To illustrate the significance of clustered origins, consider a simplified setting with identical variance and covariance of idiosyncratic components across firms. This assumption allows 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 origin from correlated idiosyncratic movements. The first terms is macro origins. The remaining terms are micro origins that represent 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. Specifically, the second term represents granular origins, which, Lucas (1977) argued, should vanish through diversification when $h_t \to 0$ 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: First, positive correlations across firms ($\rho_{\mathrm{F},t} > 0$) amplify aggregate volatility. Second, clustered origins and granular origins respond in opposite directions to market concentration: as HHI falls, the granular channel weakens while the clustered channel strengthens. 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 an HHI 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 clustered origins range from 68% to 342% of granular origins, according to equation (5). A 1.46% correlation coefficient results in an

[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).

equal contribution of idiosyncratic comovements and granularity to aggregate volatility. These calculations highlight that clustered origins can be one of the key drivers 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. A standard approach in the business-cycle literature approximates the common component using the cross-sectional mean and defines the idiosyncratic component as the deviation from that mean. Specifically, we define the pseudo common and idiosyncratic components as

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

where $\overline{\hat{y}}_t = N_t{}^{-1} \sum_i \hat{y}_{it}$ represents the cross-sectional sample mean.

3.1. Simulation Design and Results

We conduct Monte Carlo simulations to illustrate how standard demeaning obscures 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.

Two key findings emerge. First, aggregate fluctuations differ markedly between true and pseudo (demeaned) components. Figure 1 contrasts these patterns across two economies in Panels (a) and (b). While true idiosyncratic components (orange solid line) exhibit larger aggregate fluctuations when pairwise correlations are positive, pseudo components (blue dashed line) remain at zero by construction in both cases. Second,

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

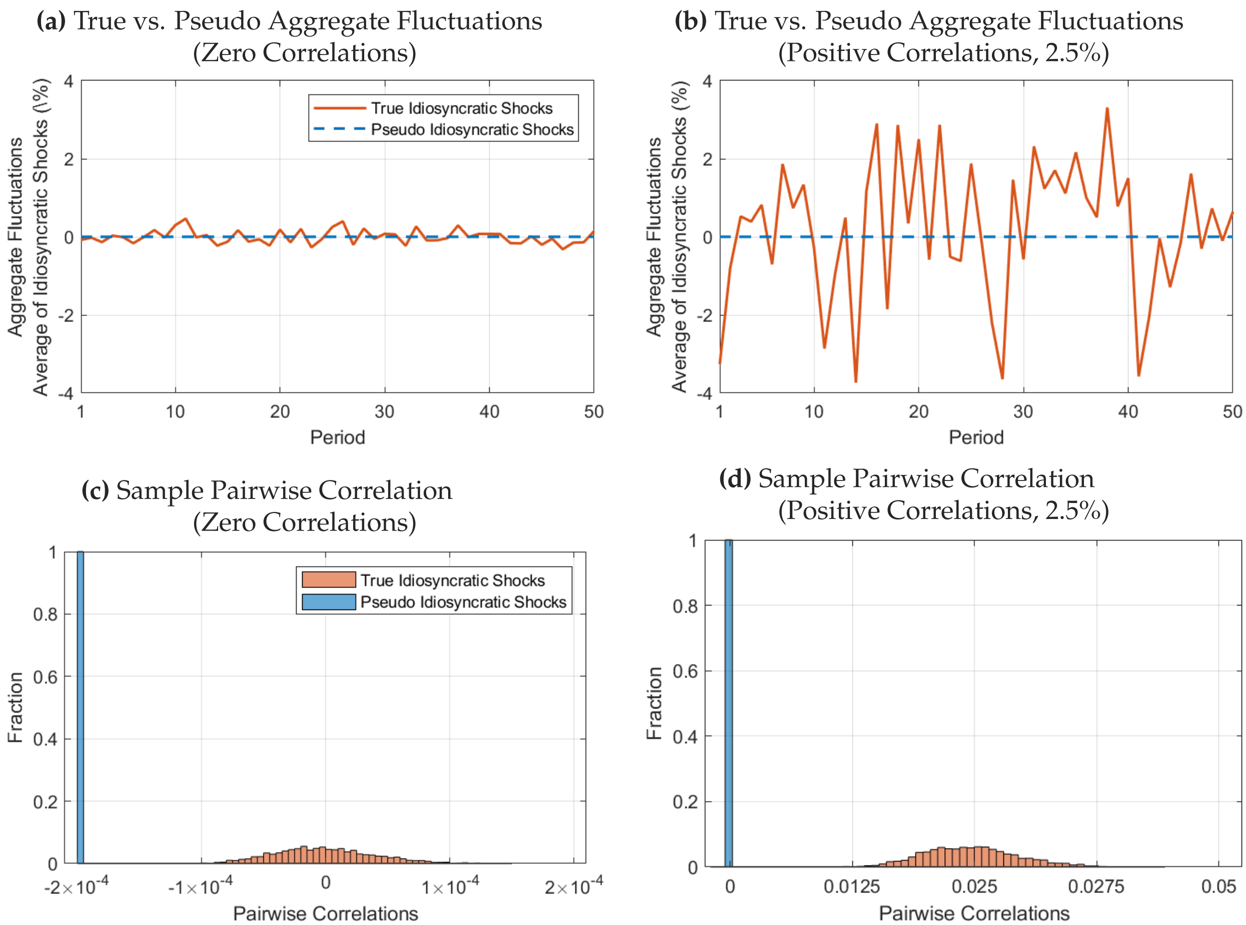


Figure 1: Aggregate Fluctuations and Sample Pairwise Correlations: True vs. Pseudo Components
Notes: The figures (a) and (b) plot the aggregate fluctuations based on the average true and pseudo idiosyncratic components in each period (orange solid line, $N_t^{-1}\sum_{i=1}^{5,000}\varepsilon_{\mathrm{F},it}$,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

demeaning masks true 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 ($\approx -2\times 10^{-4}$) in both cases, regardless of the underlying correlation structure.[6] Taken together, these results show that demeaning mechanically eliminates evidence of correlated idiosyncratic movements, potentially leading researchers to understate an important source of aggregate volatility.

[6]We discuss the slightly negative correlations of the pseudo components and show that they are asymptotically zero under identical variance–covariance structures ($-1/(N_t-1)$ in equation 16).

3.2. Non-Negligible Difference Between True and Pseudo Variables

When true idiosyncratic components exhibit correlation, pseudo variables derived from demeaning procedures are inadequate proxies for the true variables. We demonstrate this formally 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 cross-sectional average of the true idiosyncratic components. This equality reveals that non-zero pairwise comovements among idiosyncratic components generate systematic differences between true and pseudo measures.

To formalize this insight, we introduce the average of idiosyncratic variances defined as $\overline{\sigma}^2_{\mathrm{F},t} = N_t^{-1} \sum_i \sigma^2_{\mathrm{F},it}$. Also, $\overline{\overline{\mathrm{cov}}}_{\mathrm{FF},t} = N_t^{-1} \sum_i \overline{\mathrm{cov}}_{\mathrm{FF},it}$ denotes the average covariance of idiosyncratic components, where $\overline{\mathrm{cov}}_{\mathrm{FF},it} = (N_t - 1)^{-1} \sum_{i' \neq i} \rho_{\mathrm{FF},ii't} \sigma_{\mathrm{F},it} \sigma_{\mathrm{F},i't}$ is 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 following holds:*

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

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-hand side of equation (8) vanishes as the number of firms increases under finite variances, the second term remains non-negligible when average pairwise covariances are non-zero. Consequently, the pseudo variables are inconsistent proxies for the true components whenever $\overline{\overline{\mathrm{cov}}}_{\mathrm{FF},t} \neq 0$. 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 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 also 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 ($N_t^{-1}\sum_i \mathsf{cov}(e_{\text{F},it}, e_{\text{A},t}) = 0$), individual firms can show nonzero covariance when their covariances differ from the average ($\overline{\text{cov}}_{\text{FF},it} \neq \overline{\overline{\text{cov}}}_{\text{FF},t}$), 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:

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

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

This result shows that empirical methods relying on demeaned firm fluctuations mechanically erases evidence of within-industry comovement. Researchers relying on such methods may falsely conclude that idiosyncratic shocks are uncorrelated, thereby overlooking a meaningful source of aggregate fluctuations

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

Most existing business cycle studies assume zero or negligible cross-firm correlations. We show that cross-sectional demeaning is defensible only under a restrictive special case: when true idiosyncratic components have identical variance and covariance across firms.

Suppose 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 common component volatility and understate 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 the number of firms increases.

Specifically, the correlation structure simplifies 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, and the cross-firm correlation of pseudo idiosyncratic components converges to zero as the number of firms increases.

Under identical volatility and covariance structures, cross-sectional demeaning reasonably separates common and idiosyncratic shocks, yielding 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, this assumption rarely holds in practice. Extensive empirical evidence documents substantial heterogeneity in firm-level second moments of key economic variables, including productivity, output, and financial performance (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). In the presence of heterogeneous variance and covariance structures, cross-sectional demeaning does not accurately recover the common and idiosyncratic components, necessitating the alternative approach developed in the following section.

[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., with a 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.

4. Identifying Micro Origins of Macro Fluctuations

How can we quantify the micro origins from data? While firm-level business-cycle fluctuations ($\hat{y}_{it}$) are directly observable, their underlying common and idiosyncratic components ($\varepsilon_{\mathrm{A},t}$ and $\varepsilon_{\mathrm{F},it}$) are not. As a result, a point value decomposition is impossible without additional information or structural assumptions.

Rather than imposing additional structural assumptions for point identification, we develop a bounds approach that exploits observable moments to establish ranges for these unobservable 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 bound the clustered and granular origins within each cluster. 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 does not exceed $\tilde{\sigma}^2_{\mathrm{A},t}$.*

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

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, the 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:*

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

$$\sum_i w_{it}^2 \mathsf{var}(\hat{y}_{it}) - h_t^2\tilde{\sigma}^2_{\mathrm{A},t} \leq \Gamma_t \leq \sum_i w_{it}^2 \mathsf{var}(\hat{y}_{it}) \tag{24}$$

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

4.2. Aggregation: The Whole Economy's Micro Origins

To measure micro 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}). Then, the business cycle component of the whole economy's GDP, denoted by $\widehat{\mathrm{GDP}}_t$, 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 gross output to aggregate value-added (GDP), which is widely used in business cycle research to account for both direct and indirect supply chain impacts (e.g., Domar, 1961; Hulten, 1978).

The variance decomposition of aggregate GDP fluctuations is

$$\begin{aligned}
\mathsf{var}(\widehat{\mathrm{GDP}}_t) &= d_t^2 \sum_{s \in S} w_{st}^2 \mathsf{var}(\hat{Y}_{st}) + d_t^2 \sum_{s \in S} w_{st} \sum_{s' \in S \setminus \{s\}} w_{s't}\mathsf{cov}(\hat{Y}_{st}, \hat{Y}_{s't}) &(25)\\
&= d_t^2 \sum_{s \in S} w_{st}^2 [\sigma^2_{\mathrm{A},st} + \chi_{st} + \Gamma_{st}] + \mathrm{BIO}_t, &(26)
\end{aligned}$$

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 component has been widely studied, this paper primarily focuses on the clustered origins related 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

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 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 BEA deflator availability. Table A1 lists these clusters, and Appendix C provides details on the data construction and measurement methodology.

Business Cycle Components and Volatility Measurements. We estimate the following 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, but the regression excludes time and industry-time fixed effects.[8] Following Castro, Clementi and Lee (2015), we control for firm age ($t - \text{first year}_i$) and size (log employment, $\ln \text{emp}_{it}$), as prior literature documents their negative correlation with firm growth and volatility.[9] All coefficients are estimated separately for each cluster s, allowing for heterogeneity in firm dynamics across industries.

From these cyclical components, 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.

Summary Statistics. Table 1 reveals two notable patterns. 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 follows a U-shaped pattern: declining from the early 1980s through the 1990s, then doubling after 2000. However, these raw moments conflate idiosyncratic and common

[8]It is worth noting that our regression with fixed effects does not imply the cross-sectional demeaning process criticized in Section 2. Proposition 1 shows that time or industry-time fixed effects (δ_t or δ_{st}) should be excluded, as they mechanically distort pairwise comovement. However, the firm fixed effect implies that an individual firm's mean business-cycle component is zero over the sample period; that is, its long-run expectation equals zero. This does not imply that the cross-sectional mean of the cyclical component is zero in any given period.

[9]For 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., Mullen, 2020; Fiori and Scoccianti, 2021).

Table 1: Summary Statistics

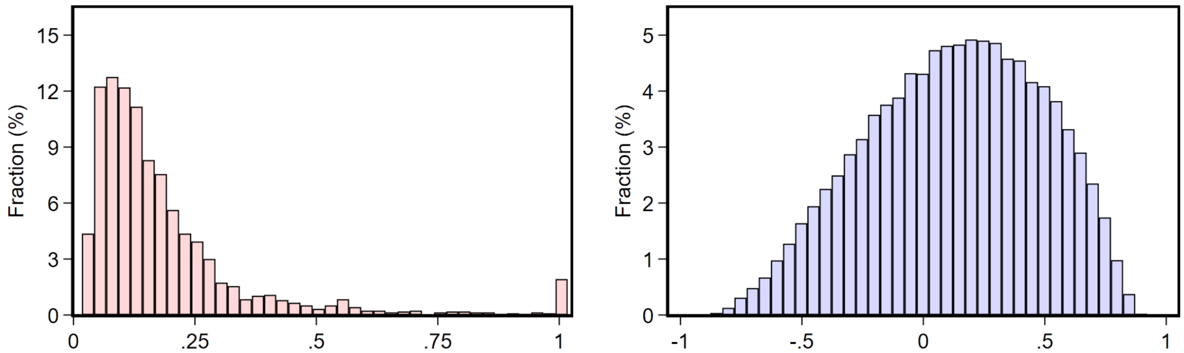
Variable	Full sample 1980–2013	Subperiod sample 1980–1985	Subperiod sample 1986–2000	Subperiod sample 2001–2013	Additional 2014–2018
Panel A. Within-firm volatility: standard deviation, $\mathsf{std}(\hat{y}_{it})$					
Mean	0.184	0.160	0.187	0.191	0.179
Standard deviation	0.208	0.153	0.212	0.223	0.219
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 ($[t-5, t+5]$). In Panel A, the standard deviations are winsorized at the top 1%. In Panel B, pairwise correlations are calculated only for firms within the same cluster (among 53 clusters).

cluster-level fluctuations (equations 18 and 19), so they do not isolate the synchronized component of firm behavior. The following analysis decomposes these moments to separately identify the contribution of within-cluster comovement to aggregate volatility.

To assess the homogeneous variance-covariance assumption discussed in Section 3, we examine the cross-sectional distribution of firm-level moments. If idiosyncratic fluctuations were characterized by a common variance-covariance structure within each cluster, we would expect relatively little dispersion in the observed second moments, $\mathsf{std}(\hat{y}_{it})$ and $\mathsf{corr}(\hat{y}_{it}, \hat{y}_{i't})$. Figure 2 provides strong evidence against this implication by plotting the distributions of firm-level standard deviations and pairwise correlations in 1995.[10] Both distributions exhibit substantial dispersion, indicating pronounced heterogeneity in firm-

[10]The same conclusion holds when moments are computed using within-cluster demeaned fluctuations; 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.

level volatility and comovement. Similar patterns emerge throughout the sample period.

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

We quantify the contribution of clustered origins to aggregate volatility in the U.S. economy using the framework developed in Section 4. The results point to a quantitatively important and previously underappreciated source of macroeconomic fluctuations: synchronized idiosyncratic movements among firms within the same industry. The estimated upper and lower bounds imply that within-cluster firm comovement accounts for roughly 10–20% of U.S. private GDP volatility over the 1980–2013 sample period (Table 2).

GDP Volatility and Macro Origins. Panel A of Table 2 documents the well-known U-shaped pattern of GDP volatility: roughly 2.7% in the early 1980s, declining sharply to approximately 1.5% during the Great Moderation (1986–2000), and rising again through the 2000s and 2010s.[11] Panel B shows that macro origins—aggregate and industry-level common shock volatilities—track only part of this pattern. They declined substantially

[11] Aggregate GDP volatility is measured as the standard deviation of the U.S. private economy's logarithmic GDP business-cycle component, computed using a rolling window of ±5 years. The business-cycle component is obtained from the residuals of an AR(2) process with a linear time trend. See Appendix A5 for the business-cycle components and alternative measures of aggregate GDP volatility from band- or high-pass filters.

Table 2: The Contribution of Firm Comovement to U.S. Business Cycles

Variable	Full sample	Subperiod sample			Additional
(Sample average)	1980–2013	1980–1985	1986–2000	2001–2013	2014–2018
Panel A. GDP volatility (std. dev., %): $\mathsf{std}(\widehat{\mathrm{GDP}}_t)$					
AR(2) residuals	1.79	2.66	1.53	1.69	1.92
AR(1) residuals	1.88	2.74	1.59	1.81	1.81
Growth rates	1.96	2.83	1.66	1.90	1.91
Panel B. Macro origins (std. dev., %): $d_t(\sum_{s\in S} w_{st}^2 \tilde{\sigma}_{\mathrm{A},st}^2)^{1/2}$					
Upper-bound	0.54	0.88	0.43	0.49	0.45
	(0.039)	(0.078)	(0.018)	(0.022)	(0.031)
Panel C. Clustered origins (std. dev., %): $\sqrt{\lvert\chi_t\rvert} \times \mathsf{sign}(\chi_t)$					
Upper-bound	0.81	1.24	0.60	0.85	0.85
Lower-bound	0.63	0.91	0.44	0.71	0.74
	(0.047)	(0.089)	(0.024)	(0.030)	(0.058)
Panel D. Ratio of clustered origin variance to GDP variance (%): $\chi_t/\mathsf{var}(\widehat{\mathrm{GDP}}_t)$					
Upper-bound	21.7	22.0	16.4	27.7	22.2
Lower-bound	13.4	11.9	8.6	19.7	16.8
	(1.89)	(2.81)	(1.17)	(1.86)	(3.23)

Notes: This table reports the contribution of within-cluster firm comovement (the 'clustered origin,' χ_t) to aggregate U.S. business cycle volatility. In Panel A, GDP volatility is the standard deviation of the cyclical component of logarithmic private GDP, estimated via an autoregressive regression with a linear time trend. In Panel B, macro origin upper bounds (std. dev.) are calculated as the contribution of the whole economy and industry common shocks in equation (26). Bootstrap standard errors for the sample period averages are reported in parentheses. In Panel C, clustered origins are reported in terms of standard deviations. In Panel D, the ratio measures the contributions of clustered origins as a percentage of total GDP variance. In Panels C and D, the standard errors of lower bounds in parentheses are based on the double bootstrap procedure for macro origin upper-bounds.

from 0.88% to 0.43% between 1980–1985 and 1986–2000, consistent with the Great Moderation. However, macro origins remained relatively stable after 2000 and do not account for the subsequent rebound in aggregate volatility, pointing to microeconomic sources as an increasingly important driver of business-cycle fluctuations.

Subperiod Evolution of Clustered Origins and the Great Moderation. Clustered origins closely track the U-shaped pattern of aggregate volatility documented in Panel A of Table 2. During the Great Moderation (1986–2000), the upper- and lower-bound range estimate averaged 0.44–0.60% (in standard deviation units), substantially below the bounds observed in both the preceding and subsequent periods. Over the full sample, clustered origins account for 13.4–21.7% of GDP variance, falling to 8.6–16.4% during the Great Moderation (Panels C and D).[12] These parallel movements indicate that shifts in aggregate volatility reflect, in part, changes in within-industry firm synchronization over the past four decades.

Countercyclicality of Clustered Origins and the Great Recession. The importance of clustered origins becomes particularly pronounced during periods of economic stress. In Figure 3, the contribution of clustered origins to GDP variance (Panel b) reveals a clear countercyclical pattern. During normal periods, clustered origins typically account for roughly 10–15% of aggregate variance. During NBER-dated recessions, however, their contribution frequently rises to 15–30%, suggesting that firm-level comovement becomes more important during downturns. Most notably, following the Great Recession, clustered origins accounted for more than 45% of GDP variance, indicating a high degree of synchronization among firms within industries. These findings highlight a potentially important microeconomic channel through which aggregate volatility is amplified during periods of economic turbulence.

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

[12] Motivated by the GDP volatility decomposition in equation (26), we report contributions in variance terms, i.e., $100 \times \chi_t / \mathsf{var}(\widehat{\mathrm{GDP}}_t)$.

(a) Clustered Origins and GDP Volatility **(b)** Contribution to GDP Variance

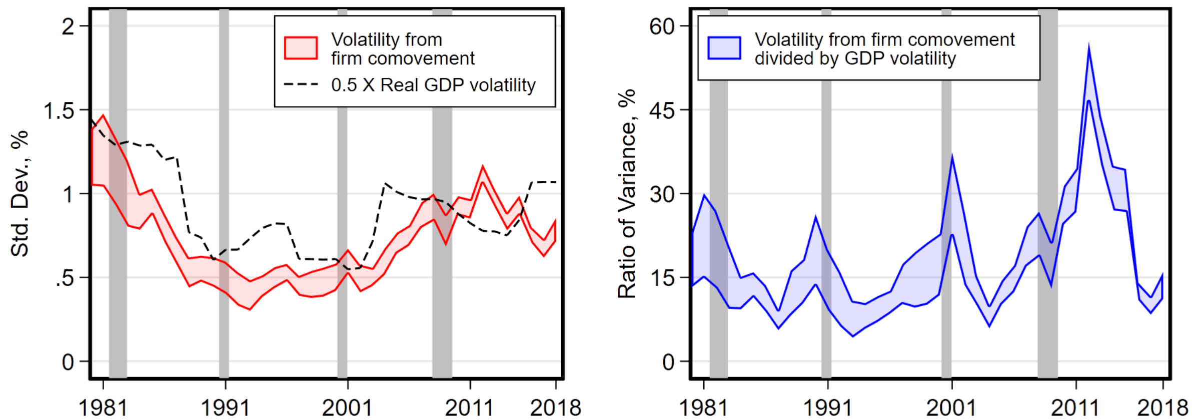


Figure 3: The Evolution of Clustered Origins

Notes: The time-series figure displays the contribution of within-cluster firm comovement (the "clustered origin," χ_t) to aggregate U.S. business-cycle volatility. Panel (a) plots the contribution from clustered origins (red area). The clustered origin's contribution to the standard deviation is calculated as $100 \times \sqrt{|\chi_t|} \times \mathsf{sign}(\chi_t)$. To compare it with the evolution of aggregate volatility, the black dashed line plots one-half of the standard deviation of aggregate real GDP growth, $0.5 \times 100 \times \mathsf{std}(\widehat{\mathrm{GDP}}_t)$. GDP volatility is measured as the 11-year rolling standard deviation of the cyclical component of private GDP, estimated using an AR(2) process. Panel (b) plots the clustered origin's contribution as a percentage of GDP variance, $100 \times \chi_t/\mathsf{var}(\widehat{\mathrm{GDP}}_t)$. Shaded gray areas denote NBER recession dates.

in equations (22) and (23), the upper bound of the macro origins and lower bound of the clustered origins depend on the minimum variance-covariance values within a cluster. Given our 11-year rolling windows, clusters often contain small samples in which 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 double-resampling bootstrap procedure that combines cluster and block resampling within each cluster for every window.[13] This procedure is valid in panel-data settings with cross-sectional and/or temporal heterogeneity,

[13]We first perform cross-sectional resampling (also referred to as cluster resampling), followed by block resampling of the time series within each cluster for every window. Based on the bootstrap samples, we construct each cluster's share, w^*_{st}, the corresponding Domar adjustment, d^*_t, and the upper bound on common-factor variance, $\tilde{\sigma}^{*2}_{\mathrm{A},st}$. We then construct the lower bound of clustered origins, $\sum_{i\in I_{st}} w_{sit} \sum_{i'\neq i\in I_{st}} w_{si't}\mathsf{cov}(\hat{y}_{it}, \hat{y}_{i't}) - (1-h^{*2}_t)\tilde{\sigma}^{*2}_{\mathrm{A},st}$. Aggregating these components using the bootstrapped weights (w^*_{st}, d^*_t), we finally calculate bootstrap standard errors and confidence intervals.

(a) **Upper-Bound of Macro Origins (std. dev., %)** (b) **Lower-Bound of Clustered Origins (std. dev., %)**

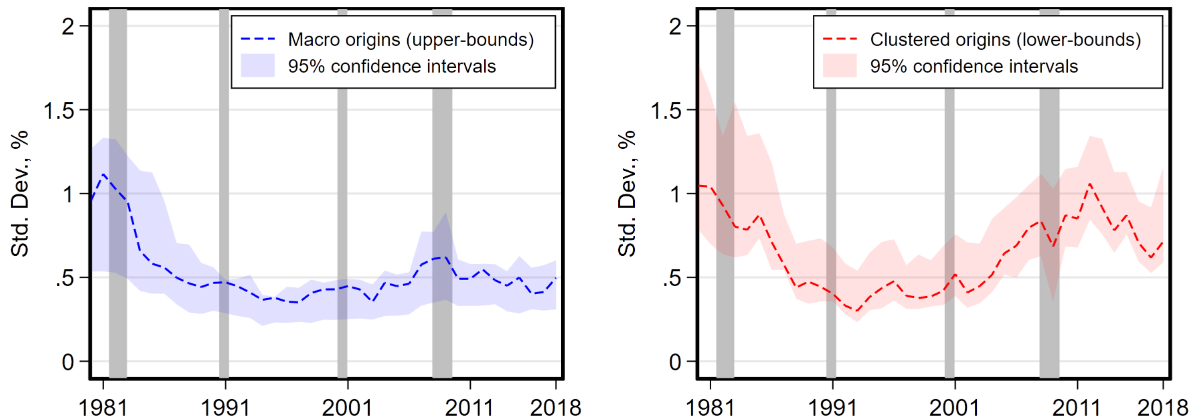


Figure 4: Macro and Clustered Origins with Confidence Intervals

Notes: The figure displays confidence intervals around the macro origin upper bound and the corresponding clustered origin lower bounds. Panels (a) and (b) plot these bounds (in standard deviation units) alongside 95% bootstrapped confidence intervals. Shaded gray areas denote NBER recession dates.

including environments with spatial dependence (Hounkannounon, 2008; Kapetanios, 2008). Using the bootstrap samples, we construct confidence intervals for cluster-level common-component variance bounds, as well as for the aggregate macro-origin upper bound and clustered-origin lower bound. Figure 4 shows that the nonlinear nature of the bounds induces some skewness in the bootstrap distributions. Nevertheless, the results confirm that the estimated contribution of clustered origins in Table 2 is statistically significant and that its time-series variation is robust to sampling uncertainty.

5.3. Robustness Checks

Decomposing the Clustered Origin. We examine whether the results are driven by two key components of equation (26): firm-level comovement and the Domar adjustment reflecting the economy's input-output structure. Panel (a) of Figure 5 shows that removing the time-varying Domar adjustment leaves the temporal pattern intact—declining during the Great Moderation and rebounding thereafter—indicating that the adjustment affects absolute levels but not the underlying dynamics.[14] Market concentration (measured

[14]The Domar adjustment varies between 50–80% over time, exhibiting a U-shaped pattern consistent with the evolution of clustered origins; see Figure A3b.

(a) Clustered Origins without Domar Adjustment **(b)** Clustered Origins with Growth Rates

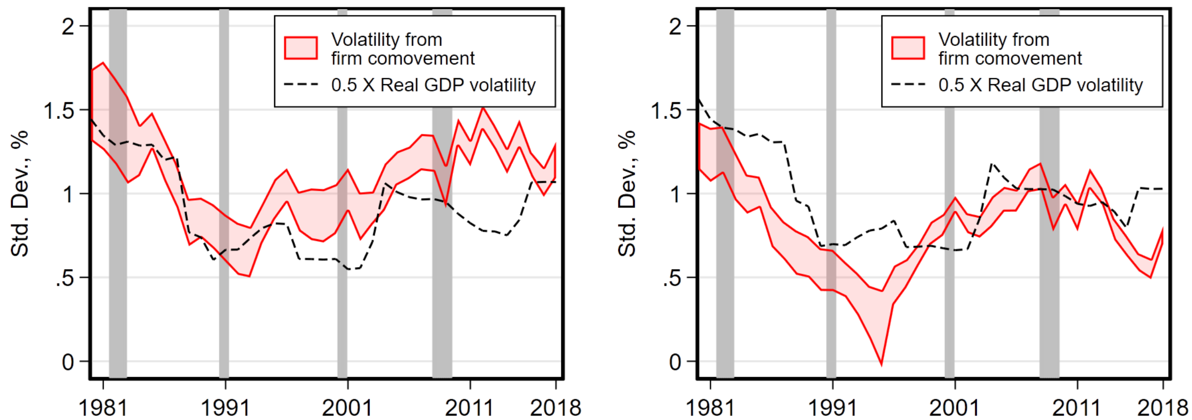


Figure 5: Robustness Check

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 using simple log-differences (growth rates) for both firm-level and aggregate GDP data instead of AR(2) regression residuals. Shaded gray areas denote NBER recession dates.

by HHI) shows only modest variation of 7.5–9% over the sample period (Appendix Figure A3a), suggesting that changes in industry structure are unlikely to explain the observed shifts in clustered origins.

Using Growth Rates. We verify the robustness of our findings using simple growth rates instead of regression residuals. Panel (b) of Figure 5 shows that firm-level growth rates (log-differences) produce patterns similar to those obtained under our baseline specification. In particular, the close association between clustered origins and aggregate GDP volatility persists under this alternative measure. The result indicates that clustered origins are a robust feature of U.S. business cycle dynamics rather than an artifact of a particular filtering or measurement procedure.

5.4. Further Discussions

Sectoral Decomposition. Is the aggregate pattern driven by a few sectors or does it reflect a broad-based phenomenon? To answer this question, we decompose clustered origins across twelve major industry groups. Figure 6 reveals substantial heterogeneity across sectors (see Table A1 for the classification). While clustered origins are generally positive at

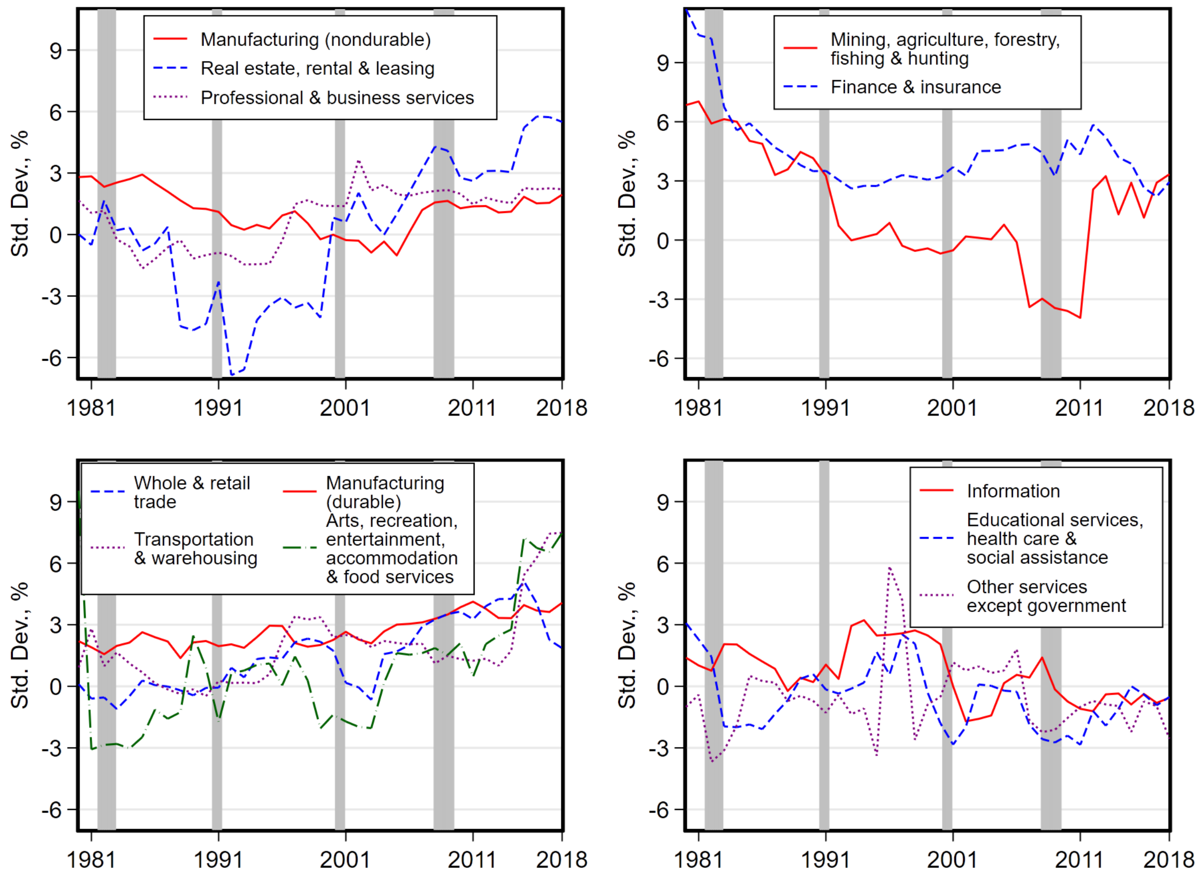


Figure 6: The Evolution of Clustered Origins by Sector

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.

the aggregate level, individual sectors display markedly different dynamics. For example, the Information sector contributes little to aggregate volatility in recent decades, while Real Estate exhibits negative contributions during parts of the 1990s, reflecting offsetting firm-level fluctuations.

Despite this heterogeneity, the major shifts in clustered origins are broadly shared across sectors. During the Great Moderation, comovement declined in most industries, including Finance & Insurance, Real Estate, Rental & Leasing, and Mining & Agriculture. This decline was followed by a widespread resurgence after 2000, particularly in Real Estate, Wholesale & Retail Trade, and Arts, Entertainment & Recreation. These results suggest that the evolution of clustered origins reflects economy-wide changes in firm

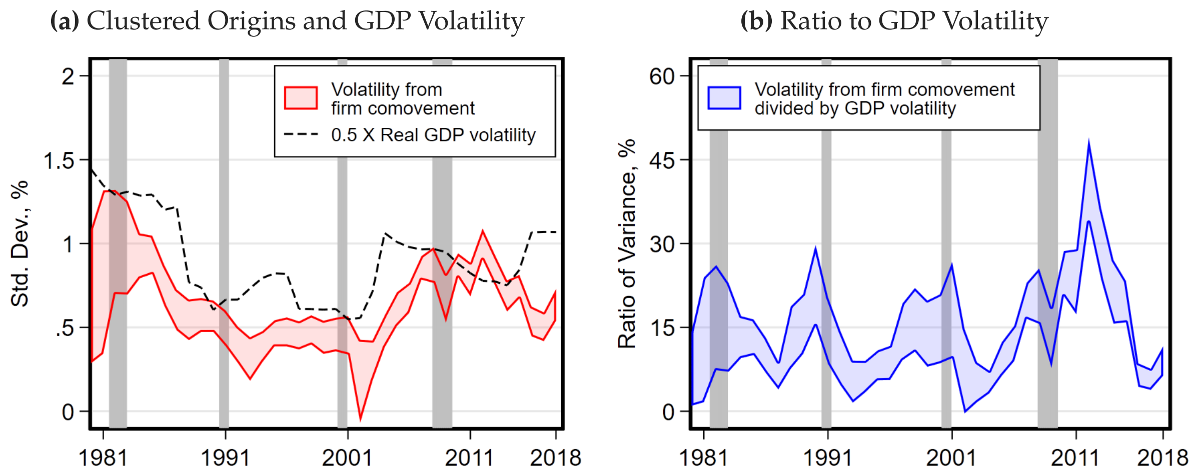


Figure 7: Clustered Origins with Labor Productivity

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.

synchronization rather than developments concentrated in a small number of sectors.

Labor Productivity. We extend the analysis to labor productivity to assess whether clustered origins reflect fundamental firm performance beyond sales performance. Labor productivity-based estimates closely replicate the baseline sales results: the contribution of productivity comovement declines during the Great Moderation and rises during the recessions of the 2000s (Panel a, Figure 7), and its share of aggregate variance increases from roughly 10% in the late 1990s to more than 25% during the Great Recession and subsequent recovery (Panel b). Although smaller in magnitude, these estimates confirm that clustered origins reflect broader firm-level synchronization rather than sales-specific dynamics.

Granular Origins. Finally, we compare clustered origins with the granular origins, which attributes aggregate fluctuations to idiosyncratic shocks of large firms (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 after the mid-1990s. Our granular measure closely

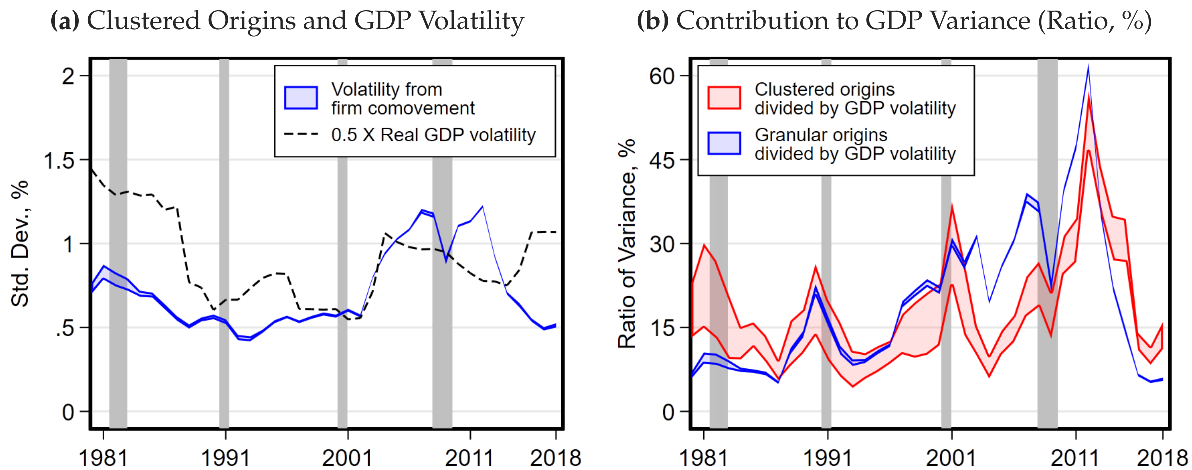


Figure 8: Granular Origins in U.S. Business Cycles

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 dashed line) against the contribution from granular origins (blue area). In Panel (b), the blue and red areas plot granular and clustered origins' contribution as a percentage of total GDP variance ($100 \times \Gamma_t/\text{var}(\widehat{\text{GDP}}_t)$ and $100 \times \chi_t/\text{var}(\widehat{\text{GDP}}_t)$), respectively. Shaded gray areas denote NBER recession dates.

replicates prior estimates: the correction term $-h_t^2\sigma_{\text{A},t}^2$ in equation (21) is quantitatively negligible given the small squared HHI.[15] The two channels nevertheless differ in their cyclical behavior: clustered origins peak consistently around recessions, while granular origins display different dynamics not exclusively tied to downturns. These distinct patterns suggest that the two channels respond to different types of economic shocks, pointing to the need for 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 within-industry comovement of firm idiosyncratic components. Standard empirical practices, such as cross-

[15]The difference between our measure and simpler demeaned approaches is proportional to $-h_t^2\sigma_{\text{A},t}^2$; because squared HHI is small, the adjustment is minor.

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, nonparametric 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, rise sharply during downturns, and reach nearly 45% 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 at play. 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; 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; Mullen, 2020; Fiori and Scoccianti, 2021), may also contribute to firm comovements within industries. Disentangling the relative importance of these mechanisms remains an open and important question. Such work 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.

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 Investment-Specific Technology," Unpublished Manuscript, Federal Reserve Board of Governors.

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.

HOUNKANNOUNON, BERTRAND (2008): "Bootstrap for Panel Data," Unpublished Manuscript, Universite de Montreal, CIREQ.

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.

KAPETANIOS, GEORG (2008): "A Bootstrap Procedure for Panel Data Sets with Many Cross-Sectional Units," *The Econometrics Journal*, 11(2), 377–395.

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.

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.

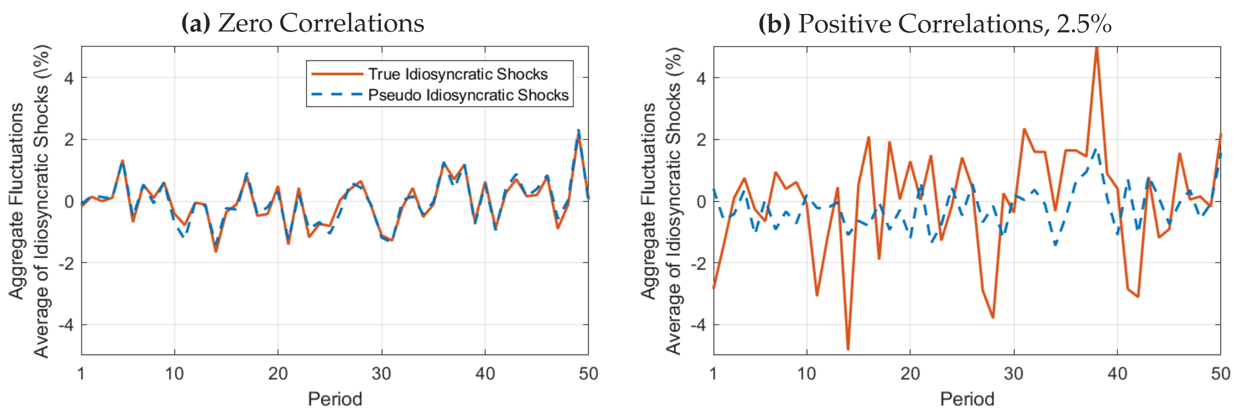


Figure A1: True vs.Pseudo Idiosyncratic Shocks and Their Aggregate Fluctuations with fat tails

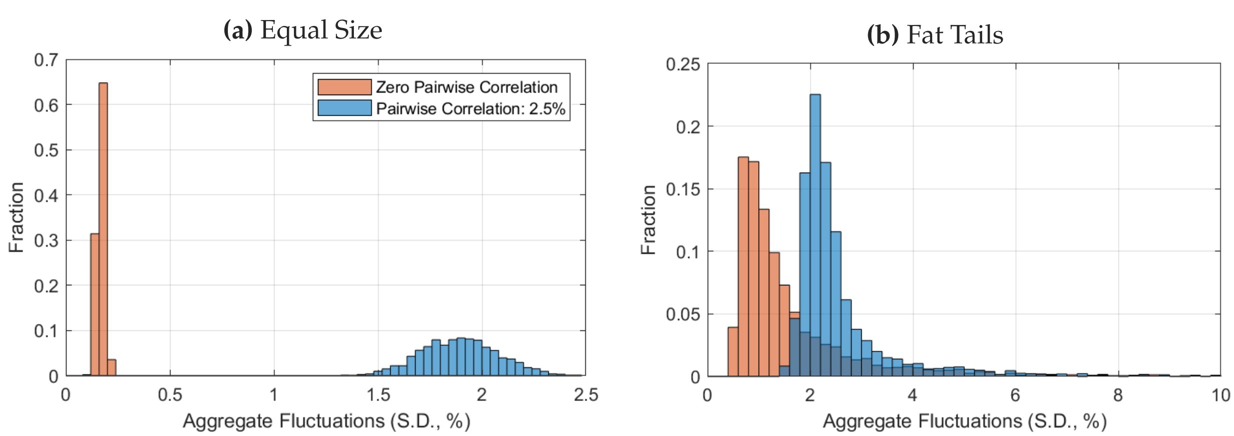


Figure A2: Sample Standard Deviations of Aggregate Fluctuations

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.

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

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

This section explores an alternative approach to address pairwise correlations in the framework discussed in the preceding sections by introducing a factor model with heterogeneous factor loadings.[16] In this framework, comovements among idiosyncratic shocks are driven by heterogeneous reactions to a latent common factor. We show that, under the standard factor-loading normalization, the estimated factor model faces the same problem as the demeaned process discussed earlier: it mechanically removes firm comovements and their contribution to aggregate volatility.

Framework. We begin by decomposing the idiosyncratic 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:

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

where heterogeneous factor loadings imply that pairwise correlations between $\varepsilon_{\mathrm{F},it}$ and $\varepsilon_{\mathrm{F},i't}$ vary across firms.

This representation provides an alternative formulation of the business-cycle components. The corresponding variance and covariance are:

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

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

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.

Pseudo Common and Idiosyncratic Components. Identification in latent factor models requires a normalization on factor loadings, typically imposing a zero cross-sectional

[16] Our main framework is more general and does not exclude this factor framework.

mean (see Bai and Ng, 2013). Under this normalization, equation (6) implies the following decomposition:

$$\hat{y}_{it} = \left(\varepsilon_{\mathrm{A},t} + \overline{\lambda}_t f_t\right) + \Big[\overbrace{\left(\lambda_{it} - \overline{\lambda}_t\right)}^{\tilde{\lambda}_{i,t}} f_t + u_{it}\Big] \tag{A9}$$

$$= \left(\overline{\hat{y}}_t\right) + \left[\hat{y}_{it} - \overline{\hat{y}}_t\right] \equiv e_{\mathrm{A},t} + e_{\mathrm{F},it}, \tag{A10}$$

where $\overline{\lambda}_t = N_t^{-1}\sum_i \lambda_{it}$ and $0 = N_t^{-1}\sum_i \tilde{\lambda}_{i,t}$. The conventional factor model thus estimates $(\tilde{\lambda}_{i,t} \equiv \lambda_{it} - \overline{\lambda}_t)$ rather than λ_{it}, linking the estimated idiosyncratic component to $e_{\mathrm{F},it}$.

The discrepancy between the true and pseudo components does not generally vanish asymptotically. Specifically,

$$\mathbb{E}\left[|e_{\mathrm{A},t} - \varepsilon_{\mathrm{A},t}|^2\right] = \mathbb{E}\left[|e_{\mathrm{F},it} - \varepsilon_{\mathrm{F},it}|^2\right] = \frac{1}{N_t}\overline{\sigma}^2_{u,t} + \overline{\lambda}_t^2 \sigma^2_{f,t}., \tag{A11}$$

The pseudo components also fail to consistently estimate the true variances:

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

$$\mathsf{var}(e_{\mathrm{F},it}) = \left(1 - \frac{1}{N_t}\right)\sigma^2_{u,it} + \tilde{\lambda}^2_{i,t}\sigma^2_{f,t} - \frac{1}{N_t}\left(\sigma^2_{u,it} - \overline{\sigma}^2_{u,t}\right), \tag{A13}$$

where $\overline{\sigma}^2_{u,t} = N_t^{-1}\sum_i \sigma^2_{u,it}$ is the average variance. These expressions generally do not recover the true common and idiosyncratic variances ($\mathsf{var}(\varepsilon_{\mathrm{A},t})$ and $\mathsf{var}(\varepsilon_{\mathrm{F},it})$) unless true pairwise comovements are nearly absent (i.e., when $\overline{\lambda}_t = 0$).

Firm Comovements. Finally, it's worth noting that the pairwise covariance of pseudo idiosyncratic shocks tends to disregard the actual covariance:

$$\mathsf{cov}(e_{\mathrm{F},it}, e_{\mathrm{F},i't}) = \tilde{\lambda}_{i,t}\tilde{\lambda}_{i',t}\sigma^2_{f,t} - \frac{1}{N_t}\overline{\sigma}^2_{u,t} - \frac{1}{N_t}\left[(\sigma^2_{u,it} - \overline{\sigma}^2_{u,t}) + (\sigma^2_{u,i't} - \overline{\sigma}^2_{u,t})\right], \tag{A14}$$

which cross-sectional average converges to zero as $N_t \to \infty$, implying asymptotically uncorrelated pseudo idiosyncratic shocks regardless of the true underlying comovement.

The conclusions from the previous sections therefore continue to hold when pairwise comovements arise from heterogeneous exposure to a common latent factor. Researchers should be cautious when using sample demeaning or zero-normalization restrictions (as in Bai and Ng, 2013), as these procedures can mechanically eliminate firm interdependencies and understate their contribution to aggregate volatility.

E. Additional Figures and Tables

(a) Herfindahl-Hirschman Index

(b) Domar Adjustment

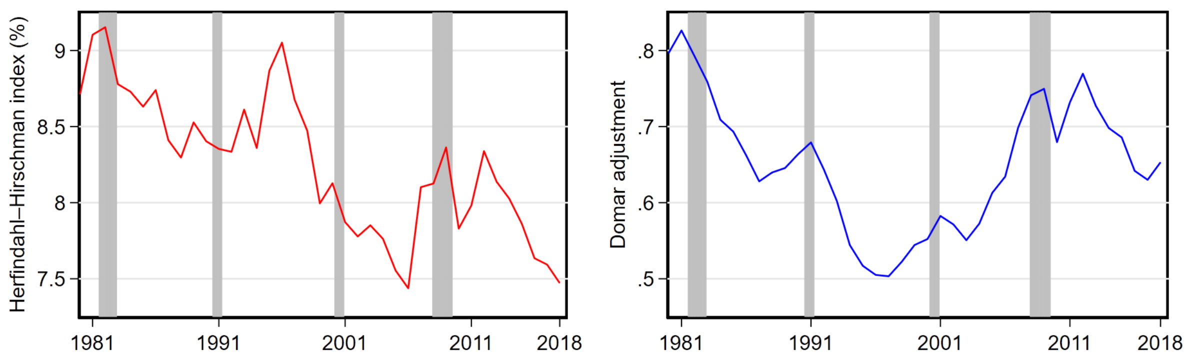


Figure A3: Herfindahl-Hirschman Index and Domar Adjustment

(a) Standard Deviation

(b) Cross-Firm Correlation within Cluster

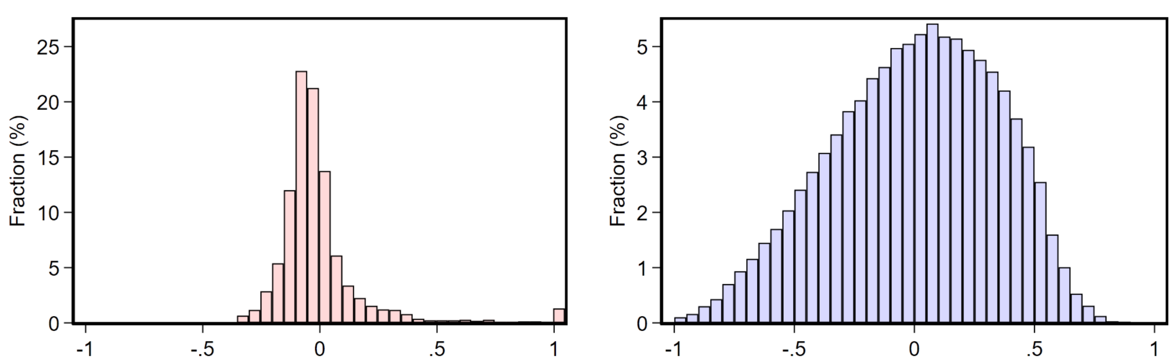
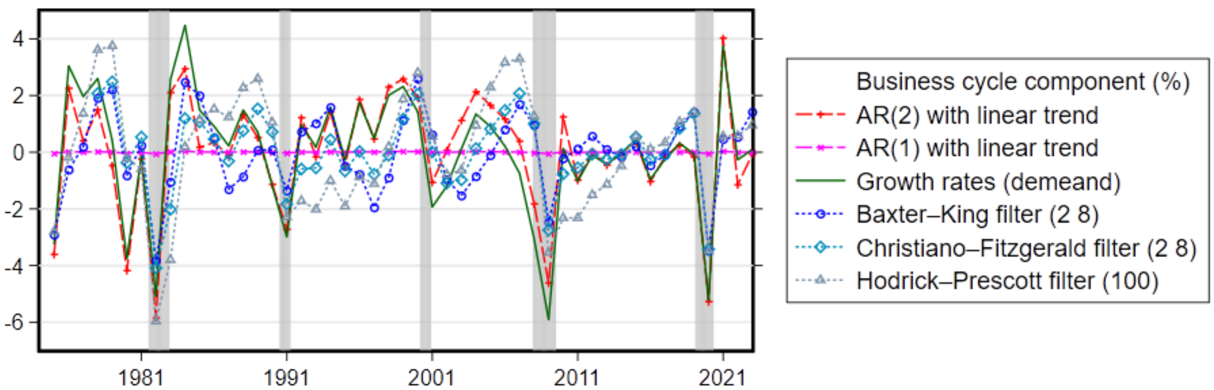


Figure A4: Histograms: Firm Volatility and Comovements (Demeaned Across Clusters)

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.

(a) GDP Business Cycle Components (%)



(b) GDP Volatility (% std. dev.)

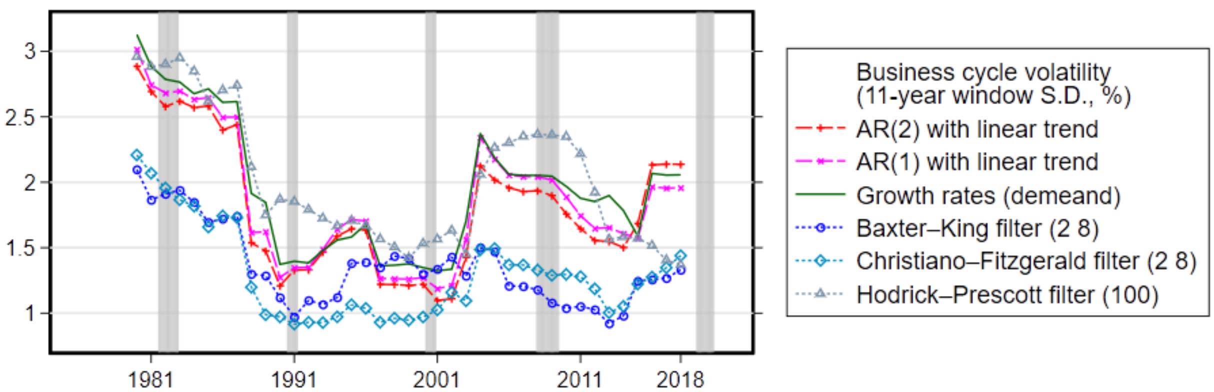


Figure A5: U.S. Business Cycle Fluctuations

Notes: The figure plots the aggregate U.S. business-cycle component and volatility under various specifications. Panel (a) plots the business-cycle component of annual U.S. private GDP. Panel (b) reports the standard deviation computed over an 11-year rolling window.

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	441	441
	· Food and beverage stores	445	445
	· General merchandise stores	452	452
	· Other retail	442-4, 446-8, 451, 453-4	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	485	485
	· Other transportation and support activities	487-8, 491-2	487-8, 491-2
	· Warehousing and storage	493	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