Business Groups and Aggregate Volatility in Portugal: A Micro-Based Factor Decomposition

Rodolfo Barbedo

Dissertation written under the supervision of professor Nuno Silva, with the collaboration of Banco de Portugal

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, 20/03/2026

Abstract

Title: Business Groups and Aggregate Volatility in Portugal: A Micro-Based Factor Decomposition

Author: Rodolfo Anselmo de Castro Barbedo

This dissertation measures aggregate sales growth volatility in the Portuguese non-financial corporate sector (2008–2022) from a micro-based portfolio perspective and asks whether ownership networks constitute a macro-relevant channel of dependence, consistent with the granular perspective of Gabaix (2011), whose underlying hypotheses are confirmed in the case of the Portuguese economy.

To evaluate this hypothesis, we combine firm-level sales-growth volatility estimates provided by Banco de Portugal and a factor model comprising aggregate, industry and municipality components to measure aggregate volatility. This estimate is compared with an alternative one where firms affiliated to business groups are consolidated in "Super-Firms" and an additional risk factor is added to capture their higher interdependence.

Results depend on the covariance estimator, but the central finding is robust: introducing business groups has a quantitatively small effect on reconstructed aggregate volatility and does not materially reduce the gap to the realized volatility of aggregate sales growth, despite clear evidence of meaningful within-group dependence at the entity level.

To interpret why micro relevance translates weakly into macro impact, we examine the overlap between the group factor and baseline industry–municipality dynamics and emphasize a mechanical constraint: the ownership channel carries limited aggregate weight in our filtered network. We also document that the group-factor impact varies over time, exploring a late-sample divergence.

Overall, ownership networks matter for consolidated entities' risk profiles, yet they do not emerge as the primary driver of aggregate volatility in Portugal.

Keywords: Aggregate Volatility, Business Groups, Granular Hypothesis, Factor Models, Portuguese Economy.

Resumo

Título: Grupos Económicos e Volatilidade Agregada em Portugal: Uma Decomposição de Fatores com Base Microeconómica

Autor: Rodolfo Anselmo de Castro Barbedo

Esta dissertação mede a volatilidade do crescimento das vendas agregadas no setor empresarial não financeiro em Portugal (2008–2022) numa perspetiva microeconómica de carteira e avalia se as redes de propriedade constituem um canal de dependência macroeconomicamente relevante, em linha com a perspetiva granular de Gabaix (2011), cujos pressupostos são confirmados para a economia portuguesa.

Para o testar, combinamos estimativas de volatilidade do crescimento das vendas ao nível da empresa, disponibilizadas pelo Banco de Portugal, com um modelo de fatores observáveis que inclui componentes agregados, setoriais e municipais, de modo a reconstruir a volatilidade agregada. Esta é comparada com uma especificação alternativa onde as empresas afiliadas a grupos são consolidadas em "Super-empresas" e é adicionado um fator de risco de grupo.

Os resultados variam consoante o estimador de covariâncias, mas a conclusão central é robusta: a introdução de grupos económicos tem um impacto reduzido no cálculo da volatilidade agregada reconstruída, apesar da clara evidência de dependência intra-grupo ao nível das entidades.

Para interpretar a fraca transmissão da relevância micro ao nível macro, analisamos a sobreposição entre o fator de grupo e as dinâmicas setoriais/municipais, enfatizando uma restrição mecânica: o canal de propriedade detém um peso agregado limitado na rede filtrada. Observamos também que o impacto do fator de grupo é variável no tempo, explorando uma divergência no final da amostra.

Em suma, embora as redes de propriedade influenciem o perfil de risco das entidades consolidadas, estas não emergem como o principal motor da volatilidade agregada em Portugal.

Palavras-chave: Volatilidade Agregada; Grupos Económicos; Hipótese Granular; Modelos de Fatores; Economia Portuguesa.

Contents

1. Introduction

Understanding the origins of aggregate economic fluctuations is a central question in economics. Traditional diversification logic implies that idiosyncratic firm-level shocks should largely average out in sufficiently large economies. The granular literature challenges this view by showing that when economic activity is highly concentrated, shocks to very large firms may remain visible at the macro level. In such settings, measuring aggregate risk requires understanding how firm-level volatility and cross-firm dependence map into aggregate outcomes.

This dissertation studies aggregate risk in Portugal through the volatility of firms' real sales growth, using an unbalanced panel of non-financial firms from Banco de Portugal over 2008–2022. The sample spans two major stress episodes—the Sovereign Debt Crisis and the COVID-19 pandemic—providing a natural environment to examine how micro-level disturbances translate into aggregate volatility.

The central research question is whether ownership structures, specifically business groups, constitute an economically important source of aggregate volatility. This hypothesis is especially relevant in Portugal, where business groups account for a large share of corporate activity and can create internal linkages across subsidiaries. If ownership networks are macro-relevant, consolidating affiliates into group "Super-Firms" and incorporating the group dimension into the dependence structure should materially change measured aggregate volatility.

To address this question, the thesis adopts a micro-based portfolio approach to aggregation. We combine observed firm-level volatility with an observable-factor framework that summarizes dependence through aggregate, industry, and municipality components. This structure yields a transparent decomposition of aggregate volatility into variance and covariance contributions. We then extend the framework to incorporate ownership linkages by consolidating affiliated firms into Super-Firms and introducing a group component in the dependence structure, allowing a direct comparison between the group channel and standard industry and location channels.

The dissertation delivers one main empirical message. While business-group affiliation is economically meaningful for the risk profile of consolidated entities, its incremental contribution to aggregate volatility is limited relative to standard aggregate, industry, and municipality channels. This result challenges the intuitive view that ownership networks act as macro-relevant transmission channels in concentrated economies.

The remainder of the thesis is organized as follows. Chapter 2 reviews the related literature. Chapter 3 describes the data and the construction of factors. Chapter 4 evaluates the structural conditions associated with granular effects. Chapter 5 presents the methodological framework for micro-based aggregation. Chapter 6 reports the main results and robustness checks, including the Super-Firm analysis.

2. Literature Review

This section reviews the literature underlying the measurement and aggregation of firm-level risk, starting with sales-growth volatility as a real-side risk proxy. It then covers dependence and high-dimensional covariance estimation, the granular hypothesis and concentration, sectoral and geographic sources of comovement, and the role of business groups, concluding with the integrated empirical strategy adopted in the thesis.

2.1. Sales Growth Volatility as a Measure of Risk

Volatility is the standard metric for assessing risk and uncertainty in economics. While the finance literature traditionally focuses on the volatility of asset returns (Markowitz, 1952), a growing body of work emphasizes the importance of real-side volatility, particularly sales and employment volatility, for understanding firm dynamics and business-cycle fluctuations. Comin and Mulani (2006), for example, document a divergence between firm-level and aggregate volatility, highlighting the importance of understanding how idiosyncratic firm-level variation maps into aggregate outcomes.

This thesis focuses on the volatility of sales growth as its primary measure of firm-level risk. Compared with asset prices, which may reflect valuation effects, market sentiment, or liquidity conditions, sales growth is more directly tied to firms' operating performance. High volatility in sales growth signals greater instability in firm outcomes and may affect investment, financing, and survival decisions (Bloom, 2009).

We deliberately do not focus on downside-risk measures such as Value at Risk (VaR). While such measures are central in banking regulation and portfolio risk management, our objective is different: we seek to characterize the overall contribution of firm-level fluctuations to aggregate volatility. In this context, both positive and negative firm-specific shocks may matter for aggregate dynamics, especially in concentrated economies where large firms can have non-negligible macroeconomic influence (Gabaix, 2011; di Giovanni et al., 2014).

2.2. Estimating Dependence in High-Dimensional Data

Once firm-level risk is measured, aggregation requires an estimate of the dependence structure across firms. If firm shocks were approximately independent, aggregation would substantially reduce volatility through diversification. But when firms are exposed to common factors or move together for structural reasons, this diversification logic weakens.

Empirically, estimating cross-firm dependence is difficult because firm-level datasets typically display a high-dimensional structure in which the cross-sectional dimension far exceeds the time dimension ($N \gg T$). In such settings, the sample covariance matrix becomes noisy, ill-conditioned, and often singular, making direct covariance estimation unreliable (Ledoit and Wolf, 2004).

For this reason, factor models have become a standard framework in both macroeconomics and finance (Stock and Watson, 2005). Following the logic of the Arbitrage Pricing Theory (Ross, 1976), factor-based approaches summarize dependence through a lower-dimensional structure, decomposing firm-level outcomes into a systematic component and an idiosyncratic residual.

These factors may be latent, as in Principal Component Analysis, or observable, such as industry, regional, or macroeconomic aggregates.

The empirical strategy adopted in this thesis follows the latter approach. Aggregate, industry and municipal factors are used to capture structured comovement across firms, while covariance regularization is employed as robustness tests to stabilize estimation in a short time sample.

2.3 The Granular Hypothesis and Aggregate Fluctuations

The idea that aggregate fluctuations may originate from microeconomic shocks has a long lineage. Jovanovic (1987) showed that firm-level disturbances need not vanish in the aggregate, although his mechanism relied on amplification forces that later proved difficult to reconcile with the data. A related strand of research shifted attention from firms to sectors. Long and Plosser (1983) and Horvath (1998, 2000) argued that, in the presence of input-output linkages, sectoral shocks can propagate and become a meaningful source of aggregate volatility. Dupor (1999), however, emphasized that in sufficiently disaggregated economies the Law of Large Numbers should still make such shocks wash out unless the cross-sectional structure is sufficiently asymmetric.

The granular literature resolves this tension by relaxing the implicit thin-tail assumption behind standard diversification logic. Axtell (2001) documents that the distribution of U.S. firm size follows a power law close to Zipf's Law, implying that economic activity is heavily concentrated in a small upper tail. Building on this empirical regularity, Gabaix (2011) shows that when firm size is heavy-tailed, idiosyncratic shocks to the largest firms do not diversify away at the standard $1/\sqrt{N}$rate. Instead, the largest "grains" of the economy can retain a non-negligible influence on aggregate fluctuations.

This mechanism has received empirical support in highly concentrated economies. Di Giovanni and Levchenko (2012), for example, show that large firms account for a substantial share of aggregate fluctuations in countries such as South Korea. Their framework also helps explain why smaller economies may be more volatile: if they host fewer firms, they are less diversified at the firm level, making shocks to dominant firms more visible in the aggregate. For the present thesis, this literature provides the key benchmark: if Portuguese firm size is sufficiently concentrated, then shocks at the top of the firm distribution, and the ownership structures organizing those firms, could, in principle, matter for aggregate volatility.

2.4 Sources of Comovement Beyond Firm Size

Firm size concentration alone does not determine aggregate volatility. A second crucial ingredient is the structure of dependence across firms, sectors, and locations. Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) show that in economies with asymmetric input-output networks, idiosyncratic shocks can propagate through production linkages and generate aggregate effects. In this setting, the rate at which volatility diversifies depends not only on the size distribution of firms, but also on the topology of the network connecting them.

Related work emphasizes that sectoral composition and sectoral covariance are quantitatively important. Carvalho and Gabaix (2013) show that aggregate volatility depends not only on the volatility of shocks themselves, but also on the relative weight of the sectors receiving those shocks. Using Domar weights, they demonstrate that changes in the sectoral composition of the economy can help explain major shifts in aggregate volatility. Complementing this view, Conley and Dupor (2003) show that covariance across sectors is itself structured: sectors that are economically "close" through input-output linkages tend to comove more strongly, so aggregate volatility depends not only on sector-specific shocks but also on the pattern of cross-sector synchronization.

To capture these interdependencies, aggregate volatility must be viewed as a combination of several layers of common and idiosyncratic variation. Campbell, Lettau, Malkiel, and Xu (2001) provide a seminal framework for this by separating returns into market, industry, and firm-specific components. While market volatility remains the most cyclical element, their findings document a trend where idiosyncratic volatility has increasingly become the largest contributor to total variability. This decomposition logic is essential for the present thesis, as it allows for the isolation of aggregate, sectoral, municipal, and ownership-related channels of comovement.

Beyond sectoral affiliation, geography and local institutions act as distinct factors of comovement. For instance, Hilary and Hui (2009) show that local norms (such as religious environment) influence corporate risk-taking, suggesting that firms in the same municipality share common behavioral drivers. Similarly, Onali and Mascia (2021) demonstrate that the benefits of diversification are heterogeneous: geographic diversification can reduce risk, whereas diversification across business segments often increases exposure to negative shocks. Finally, innovation dynamics also play a role; Comin and Mulani (2006) show that while R&D spending increases firm-level volatility, it is negatively associated with the correlation of growth between sectors, leading to a divergence in aggregate versus firm-level volatility trends.

These diverse exposures motivate our methodological choice of observable factors over latent-factor approaches (such as PCA). While latent factors summarize covariance parsimoniously, observable factors allow aggregate volatility to be attributed to interpretable economic sources (macro, sectoral and geographic), providing a structured baseline for the aggregation exercise developed in the empirical chapters.

2.5 Business Groups and Volatility: Mechanisms and Empirical Evidence

Business groups are networks of legally distinct firms connected through persistent formal and informal ties, often spanning multiple industries and activities (Khanna and Yafeh, 2007). Their potential relevance for aggregate volatility stems from two opposing mechanisms.

On the one hand, affiliation may stabilize firm outcomes. The classic "coinsurance" view (Lewellen, 1971) argues that internal capital markets allow groups to reallocate liquidity across affiliates, thereby smoothing shocks that would otherwise be amplified by financial frictions. Evidence from bank-based economies supports this mechanism: Santioni, Schiantarelli, and Strahan (2017) show that group affiliation improves survival during crises by easing financing constraints, while Matvos and Seru (2014) document that internal reallocation within conglomerates mitigates the effects of external financial dislocation. More broadly, affiliation has been associated with lower funding costs (Byun et al., 2013) and lower exposure to firm-specific shocks (Faccio, Morck, and Yavuz, 2019), suggesting that group structures may reduce idiosyncratic volatility.

On the other hand, business groups may also create opacity, agency distortions, and inefficient internal allocation. Morck, Wolfenzon, and Yeung (2005) argue that pyramidal and complex ownership structures can facilitate entrenchment and misallocation. Hong et al. (2017) and related work show that affiliation may reduce transparency and increase crash risk, while Khanna and Yafeh (2005) conclude that the volatility-reducing effects of groups are far from universal across countries. Taken together, the literature suggests that business groups can alter firm-level risk, but their net effect depends on institutional context and remains ultimately empirical.

For this thesis, the key unresolved question is not simply whether business groups affect affiliated firms, but whether this effect remains quantitatively important once risk is aggregated to the level of the economy. In a concentrated economy, business groups may appear to be natural candidates for macro-relevant "Super-Firms". Whether this intuition survives aggregation is the central issue examined in the empirical chapters.

2.6 Synthesis and Empirical Strategy

The literature suggests that aggregate volatility depends on both concentration in the upper tail of the firm-size distribution and the dependence structure linking firms across sectors and locations, with ownership networks as a potential additional layer. These channels are often studied separately.
This thesis brings them together in a unified empirical framework: using an observable-factor model to recover a structured covariance estimate, it evaluates whether business-group affiliation adds incremental explanatory power for aggregate volatility once standard aggregate, industry and municipality are accounted for.

3. Data

This section presents the dataset used in the empirical analysis and describes the construction of the variables and classifications underlying the factor model.

3.1 Data Analysis

The empirical analysis relies on a subset of the Banco de Portugal Central Balance Sheet Harmonized Panel (CBHP), covering the period from 2008 to 2022. The dataset constitutes an unbalanced panel containing approximately 1.3 million firm-year observations. It covers anonymized private non-financial corporations, identified by a unique tax identification number (NIPC), and includes detailed information on the fiscal year, firm size category, sector of activity, industry cluster, municipality cluster, and business-group affiliation. Crucially for our analysis, the dataset provides real sales, deflated using the Consumer Price Index, as well as sales growth rates and firm-level volatility measures computed following Silva (2025).

The sample is restricted to non-financial firms with annual sales and assets above 100 thousand euros (at 2021 constant prices). The unbalanced panel reflects natural entry and exit dynamics. On average, each firm is observed for 6.1 years, yielding an annual mean of approximately 71,449 active firms out of a total of 174,726 unique entities.

Despite these restrictions, the final dataset remains highly representative of the Portuguese non-financial corporate sector. On average, it covers 77% of total sales, 72% of employees, and 56% of total assets in the initial universe. The comparatively lower coverage for assets is largely explained by the exclusion of numerous real-estate firms, which typically hold substantial asset values while generating limited sales.

The construction of industry and municipal clusters follows the framework proposed in Silva (2025).

Industry clusters: The sectoral classification is based on the Portuguese Classification of Economic Activities (CAE-Rev.3) at the 5-digit level and then aggregated to ensure statistical density. To preserve representativeness, industries were grouped to meet a minimum threshold of approximately 50 observations per year, resulting in 317 clusters. The resulting distribution reflects the natural composition of the Portuguese economy, with Manufacturing, Construction, and Wholesale/Retail Trade jointly accounting for roughly 65% of all observations.

Municipality Clusters: Geographically, the starting point is the universe of 308 municipalities across mainland Portugal and the autonomous regions. The data exhibits significant spatial concentration: the districts of Lisbon, Porto, and Braga jointly account for 52% of all firm-year observations, whereas the ten least represented districts contain only 14%. To correct for this asymmetry and ensure sufficient density (min. 50 firms annually), smaller municipalities were merged based on district affiliation and GPS proximity (using the *scclust* algorithm). This resulted in a final map of 220 clusters, with a median composition of roughly 170 firms per year. Hereafter, these clusters are referred to simply as 'municipalities'.

To reduce administrative noise in cluster assignment, each firm is assigned a stable classification based on its modal affiliation. Ties are resolved via aggregation affinity, temporal continuity, or historical precedence. This substantially reduces classification instability: the number of firms exhibiting multiple industry classifications falls from 13,751 to 890, while firms with multiple municipality classifications decrease from 9,364 to 889. Filtering has a negligible impact on density, with only 5 cluster-year pairs falling slightly below the minimum threshold (by fewer than 5 firms).

The dataset is also highly skewed toward smaller firms. Using the standard European Commission criteria based on employment, turnover, and balance-sheet size, micro firms account for approximately 60% of observations and small firms for roughly one-third, as illustrated in Figure 1. Because the upper tail is comparatively sparse, medium and large firms are grouped into a single analytical category ("Medium/Large"). This size structure remains broadly stable over time, although the data show a visible contraction in the number of active firms around the sovereign-debt crisis, particularly among micro firms.

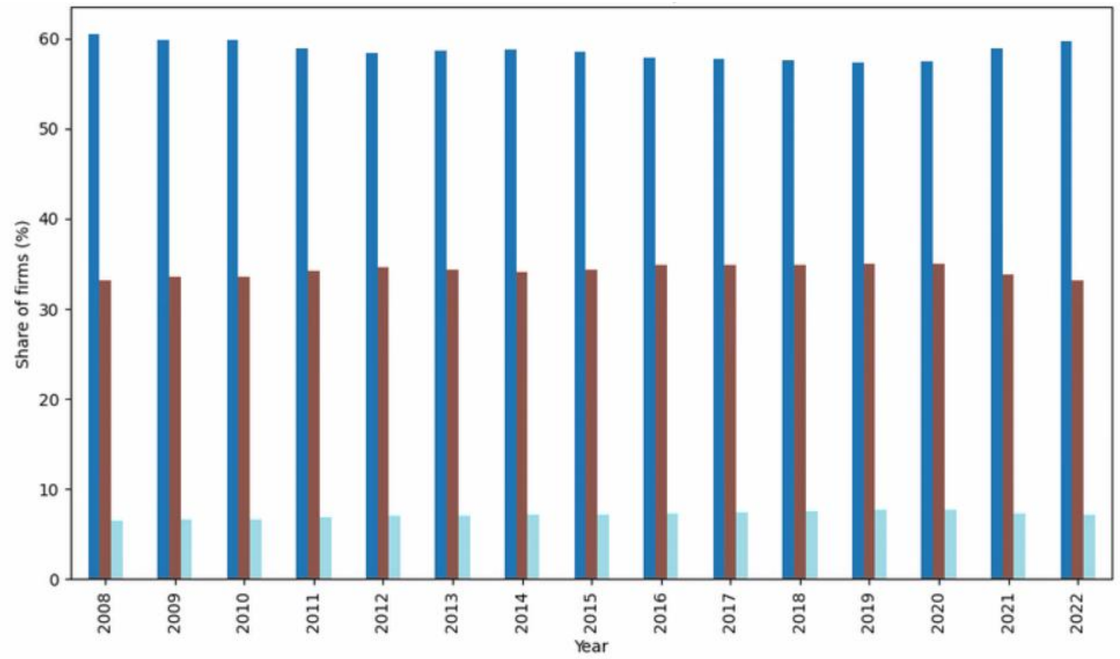


Figure 1 – Distribution of Firm Size per Year

A key input in our aggregation framework is firm-level sales-growth volatility. Table 1 reports annual descriptive statistics, based on Silva (2025).

The distribution is consistently right-skewed, with the median below the mean in all years, indicating that a relatively small subset of firms drives the upper tail. In the pooled sample, volatility exhibits positive skewness (1.87) and substantial excess kurtosis (4.89), consistent with heavy tails in firm-level risk. Average volatility is broadly stable over time but spikes during systemic stress episodes, notably the Sovereign Debt Crisis and the COVID-19 period, and dispersion increases in those years, pointing to heightened heterogeneity in firm responses.

Table 1 - Descriptive Statistics of Annual Firm-Level Volatility

Year	Mean	Median	Std. Deviation
2008	0.368	0.314	0.197
2009	0.385	0.333	0.196
2010	0.378	0.321	0.201
2011	0.389	0.331	0.200
2012	0.404	0.350	0.200
2013	0.387	0.330	0.200
2014	0.346	0.293	0.184
2015	0.341	0.288	0.182
2016	0.329	0.280	0.172
2017	0.321	0.273	0.168
2018	0.316	0.270	0.165
2019	0.319	0.273	0.166
2020	0.376	0.329	0.187
2021	0.371	0.311	0.211
2022	0.335	0.289	0.174
Total	0.358	0.305	0.190

Volatility also varies systematically across sectors and districts, reflecting differing exposure to cyclical and structural shocks. As reported in Appendix A.1, median volatility is highest in Construction and Real Estate, sectors traditionally sensitive to credit cycles, while Human Health and Social Work activities are among the most stable. Sectors such as Arts, Entertainment, and Recreation, heavily impacted by the pandemic, and Construction display the largest volatility swings.

Geographically, Madeira and Faro stand out with higher typical volatility and stronger swings, likely reflecting their higher exposure to tourism and external demand. In contrast, districts like Guarda remain consistently among the least volatile. These descriptive patterns motivate the subsequent analysis of sectoral and municipal dependence structures.

Finally, volatility exhibits a clear size gradient: average risk decreases from micro firms (0.379) to small (0.336) and medium/large entities (0.302). Medians consistently fall approximately 0.05 below these averages, reinforcing the structural right-skewness and heavy-tailed nature of the sample. Notably, the standard deviation of volatility remains relatively uniform across categories, indicating that the dispersion of these risk outcomes does not vary significantly with firm scale.

The final structural dimension of the dataset is business-group affiliation, which identifies networks of control linking legally distinct firms. According to Statistics Portugal (INE), these entities account for over 40% of total employment, ~60% of Gross Value Added (GVA), ~63% of turnover, and ~67% of Gross Operating Surplus, being crucial to understand the dynamics of Portuguese economy.

A key contribution of this study is the integration of ownership data provided by Banco de Portugal, based on the *Informação Empresarial Simplificada* (IES), specifically capturing the domestic network of control among Portuguese economic groups. Observed affiliation records are available from 2015 to 2022. Over this period, the number of affiliated firms rises from roughly 15,000 to 21,000 per year, and 25,734 distinct firms are observed as affiliated at least once. Due to data constraints, affiliations for 2008–2014 are backfilled using the earliest observed time-invariant group composition.

Although affiliated firms represent a minority of the sample (14.4% on average), they account for a disproportionately large share of economic activity, representing 63.1% of total real sales on average. Group affiliation is also strongly correlated with firm size: more than half of medium and large firms belong to a business group, compared with less than 20% of small firms and fewer than 10% of micro firms. This overlap is central to the empirical analysis, as it implies that ownership structures are disproportionately represented among the largest firms and therefore constitute a plausible channel through which micro-level risk may affect aggregate volatility.

These figures describe the raw ownership registry. In the empirical analysis, we restrict attention to stable and economically meaningful control links, so the effective group coverage used in the group-based construction is lower. The filtering criteria and their implications for coverage are described next.

3.2 Identification of Business Groups

A salient feature of the raw registry is overlapping affiliation. In our data, 34.4% of firms that are ever affiliated are associated with more than one business group. Among these multi-affiliated firms, the average number of groups is 1.94, and the 75th percentile is 2, indicating that most overlaps involve a small number of groups. At the same time, the distribution has extreme outliers: 38 firms are linked to more than 50 groups and the maximum observed count, an exceptional case, is 850.

Under the proposed methodology, firms are allocated to at most one business group. To identify meaningful ownership networks, we implement a filter assigning one stable group identity to each firm. This isolates persistent control linkages over transient changes, ensuring that the retained group structure reflects potentially relevant channels of coordination across legally distinct firms. A firm is classified as group-affiliated if a parent entity holds an equity stake of at least 10% for a minimum of three consecutive years.

Notably, 2,334 firms satisfy this equity criterion for at least two groups. When a firm meets the affiliation criteria for more than one group, overlapping memberships are resolved through a hierarchical two-stage procedure based on normalized z-scores, ensuring commensurability across the different assignment metrics.

Stage 1: Primary Allocation (Firm-Specific Intensity) The first stage calculates a score based exclusively on the years in which the firm successfully met the filtering criteria. This score prioritizes the intensity of the bilateral relationship between the firm and the potential group:

- **50%:** Average equity percentage held in the specific company.
- **40%:** Duration of the relationship (number of years meeting the filter criteria).

- **10%:** Average voting rights percentage in the specific company.

Stage 2: Secondary Tie-Breaker (Group Network Size) In the event of a tie in the primary allocation, a secondary score is computed considering the full sample history. This metric prioritizes groups with a larger number of constituents. By selecting groups with more firms, we ensure that the subsequent assignment of group-level attributes (industry and municipality) is based on a broader, more representative sample of entities rather than small, potentially erratic clusters:

- **50%:** Average network size (number of firms within the group).
- **50%:** Average equity percentage held across all firms in the group.

This methodology cannot be applied when business groups are composed of a very small number of firms. For this reason, a valid business group is defined as a network containing at least three distinct firms for at least three years. If a firm's primary assignment fails this criterion, we re-evaluate its alternative candidates. Firms that do not satisfy these criteria for any group are treated as independent entities throughout the dissertation. This procedure ensures that firms are definitively matched to the most prominent viable group entity available.

This algorithm identifies 1,128 distinct business groups, representing 3.3% of all entities. Though economically significant, affiliated firms do not span the full economy, numbering ~4,000 annually and representing 20.3%–27.1% of annual aggregate sales. As illustrated in Figure 2, which tracks the evolution of these weights over time, the presence of groups in the economy shows a high degree of constancy. Aside from a mechanical increase due to backfilling 2015 ownership data into the pre-2015 sample, the group footprint remains steady. This highlights a fundamental asymmetry: unlike industry and municipality dimensions which cover the full firm universe, the group dimension applies only to a subset of firms, implying that its total sales weight is mechanically below one.

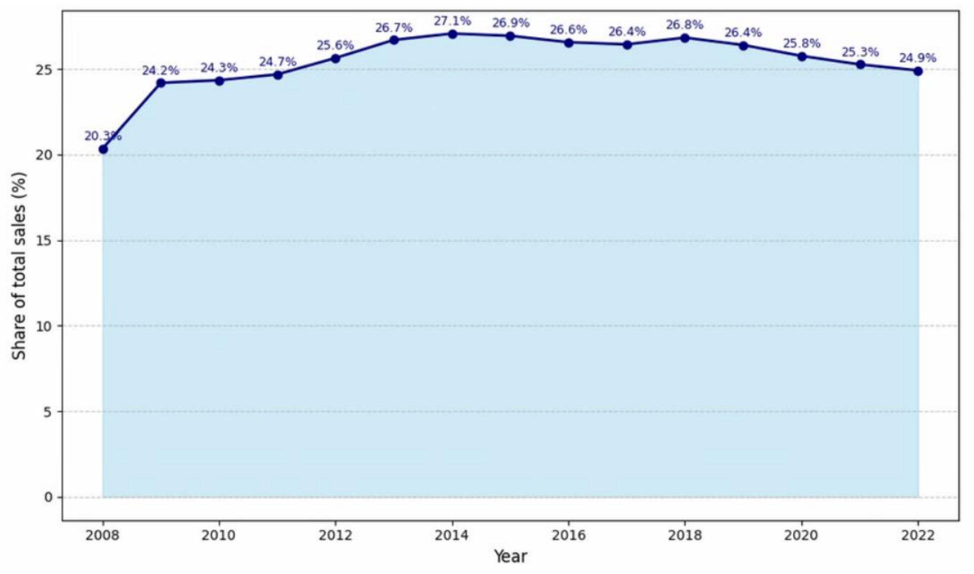


Figure 2 - Annual Evolution of the Group-Affiliated Share in Aggregate Sales

A related asymmetry concerns fragmentation. Cluster sales weights are computed relative to annual aggregate sales. Because groups are split across thousands of entities (versus 317 industry and 220 municipality clusters), the typical group carries a much smaller weight. Consistent with this, the mean cluster weight is 0.023% for groups, compared with 0.316% for industries and 0.455% for municipalities, while the median cluster weight is 0.0046% for groups versus 0.144% (industries) and 0.130% (municipalities). The upper tails differ sharply as well: the largest municipality reaches 22.45% in a year, compared with 6.44% for industries and 3.37% for groups.

This scale asymmetry provides a direct prior for the aggregate-volatility decomposition: because aggregate variance loads nonlinearly on weights, a dimension whose weight distribution is both incomplete (covering only part of total sales) and more fragmented (with a thinner upper tail), such as business groups in our sample, has mechanically limited scope to contribute as much to aggregate volatility as industries or municipalities, even if its shocks are economically meaningful within networks.

Although group-affiliated firms remain a minority of the population, they are disproportionately represented among the largest firms. However, concentration is not exclusive to groups: even among unaffiliated firms, individual entities can be macro-relevant. For example, the largest annual sales share attained by any unaffiliated firm is about 1.5%.

3.3 Model Factors

The construction of observable factors is a key step in the thesis, as it provides the baseline dependence structure against which the incremental contribution of business groups will later be assessed. The methodology is based on observable factors rather than latent statistical factors.

In the baseline specification, we consider $K = 538$ factors: 317 industry-cluster factors, 220 municipality-cluster factors, and one economy-wide factor. The aggregate factor is defined as the economy-wide median real sales growth rate in each year. Industry and municipality factors are constructed as deviations from the aggregate cycle: for each cluster, we compute the annual median sales growth rate within the cluster and subtract the economy-wide median.

This transformation serves two purposes. First, it isolates sectoral and municipal dynamics from the common macroeconomic component, so that these factors capture relative cluster-specific fluctuations rather than the national business cycle. Second, it mitigates collinearity in the regression framework by ensuring that lower-level factors are interpreted as deviations from the aggregate factor rather than as raw growth rates.

We use the median rather than the mean to construct these factors for two reasons. Firm-level data are characterized by substantial outliers, and the median provides a robust measure of typical growth within each cluster. Moreover, because the sample is dominated by micro and small firms, an unweighted statistic is preferable to a sales-weighted one for characterizing the central tendency of the cross-section.

Consistent with the same methodology, we later introduce a business-group factor, reaching $K = 539$. For each group-year, this factor is computed as the cross-sectional median sales growth rate of firms affiliated with the group, net of the aggregate factor. Group-year observations with fewer than two affiliated firms in the data are treated as missing (even if the group qualifies as a valid business group in the overall unbalanced sample).

3.3.1 Descriptive Analysis of the Standalone Factors

To assess the empirical properties of the factor structure used in the volatility decomposition, this subsection examines the distributional features of the estimated aggregate, industry, municipality, and group factors. To characterize the stochastic nature of the disaggregated shocks hitting the Portuguese economy, Table 2 presents the key statistical moments for these components.

Table 2 - Descriptive Statistics of the Estimated Common Factors

	Mean	Median	Std. Deviation	Skewness	Excess Kurtosis
Aggregate Factor	-0.012	0.007	0.053	-1.038	-0.099
Industry Factors	0.002 (0.02)	0.002 (0.02)	0.045 (0.02)	-0.069 (0.77)	0.253 (1.54)
Municipality Factors	0.001 (0.10)	0.001 (0.01)	0.022 (0.10)	0.100 (0.58)	-0.150 (1.12)
Group Factors	0.012 (0.13)	0.020 (0.10)	0.216 (0.21)	-0.023 (1.09)	1.159 (2.49)

Table 2 summarizes the distributional properties of the 539 observable factors used in the analysis. For each factor, we compute time-series moments over the pre-pandemic window and report the cross-factor average, with the cross-factor standard deviation in parentheses to highlight cluster heterogeneity.

Starting with central tendency, average factor means are close to zero in all dimensions, as industry, municipality, and group factors are defined as deviations from the aggregate. Medians slightly exceed means for the aggregate and group factors, reflecting more extreme negative realizations. Industry and municipality means and medians align, indicating typical deviations cluster around zero once the macro cycle is removed.

Dispersion differs sharply across dimensions. The aggregate factor retains full macro amplitude, with higher volatility than industry or municipality factors. Standard deviations show volatility heterogeneity is limited for municipalities but consistently highest for groups.

Higher moments reinforce these patterns. The aggregate factor is clearly left-skewed, reflecting large negative macro shocks in crisis years. Industry factors remain mildly left-skewed on average, while municipality factors appear comparatively well-behaved and right-skewed.

Notably, group factors exhibit the strongest tail behavior. They display markedly higher average excess kurtosis and much greater cross-group heterogeneity, suggesting that a specific subset of groups is susceptible to extreme realizations. While this should be interpreted with caution given the shorter time series for some groups (minimum $T = 5$), it confirms that ownership-based factors retain a distinctive asymmetric risk profile.

After reporting average moments across factors, we now examine how the factor series evolve over time and how their dynamics compare across dimensions.

Figure 3 plots the estimated Aggregate Factor against the official Portuguese Real GDP growth rate. The co-movement is striking, with the factor correctly capturing the two major recessions of the period: the Sovereign Debt Crisis (2011–2013) and the COVID-19 pandemic (2020-2021). However, a distinct feature is that the Aggregate Factor exhibits higher amplitude than GDP, its troughs are deeper and its peaks are higher.

This divergence is theoretically consistent with the composition of our sample. While GDP is a broad metric that includes the stabilizing effects of public sector expenditure and non-market services (which tend to be counter-cyclical or inelastic), our sample is composed exclusively of corporate entities. Consequently, the Aggregate Factor serves as a proxy for the pure private business cycle, which is inherently more volatile than the broader economy.

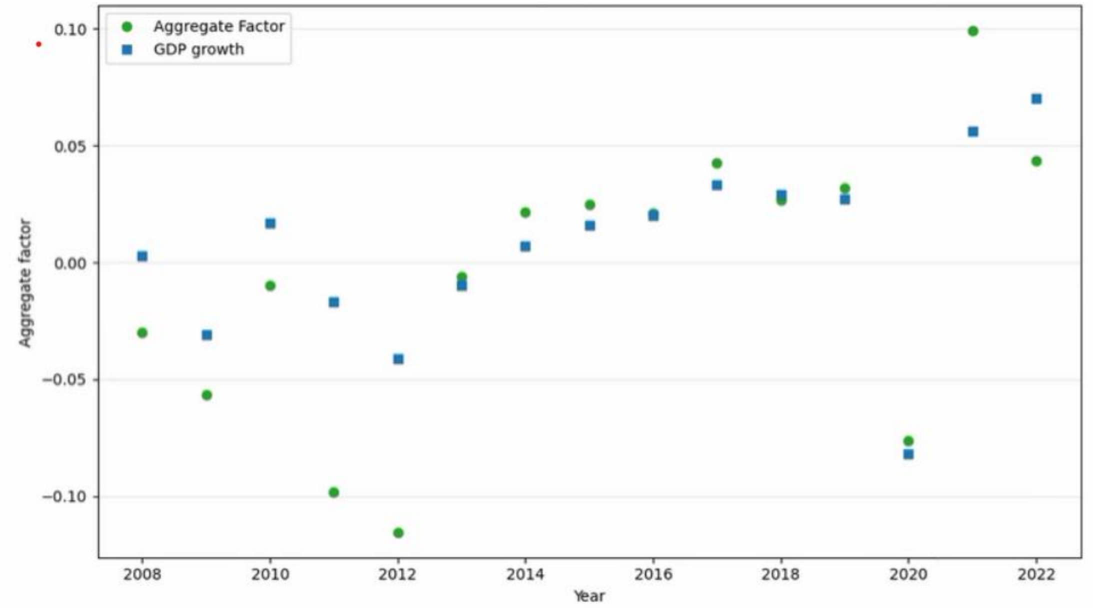


Figure 3 – Distribution of the Aggregate Factor per year

The industry and municipality factor families display markedly different distributional features. Industry factors are substantially more dispersed than municipality factors (about 2.7 times higher). Their combination of negative skewness and very high excess kurtosis (32) suggests that industry-specific deviations are characterized by occasional large downside realizations and heavier tails. By contrast, municipality factors are less dispersed and exhibit lower excess kurtosis, indicating comparatively milder tails and a distribution closer to Gaussian benchmarks.

Figures 4 and 5 display the boxplots of Industry and Municipality factors, respectively, for each year,

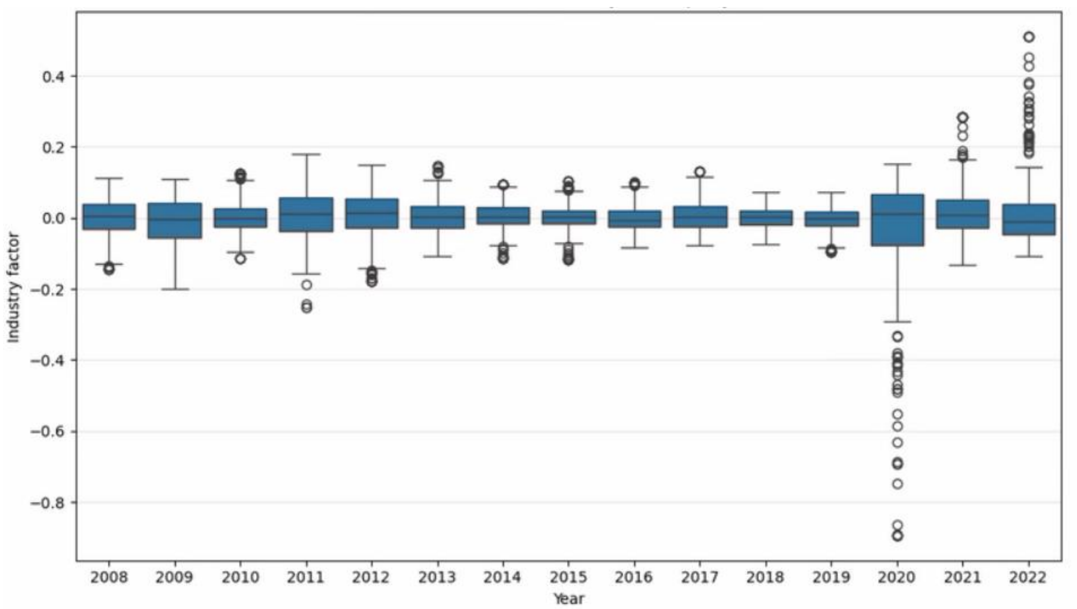


Figure 4 - Distribution of the Industry Factor per year

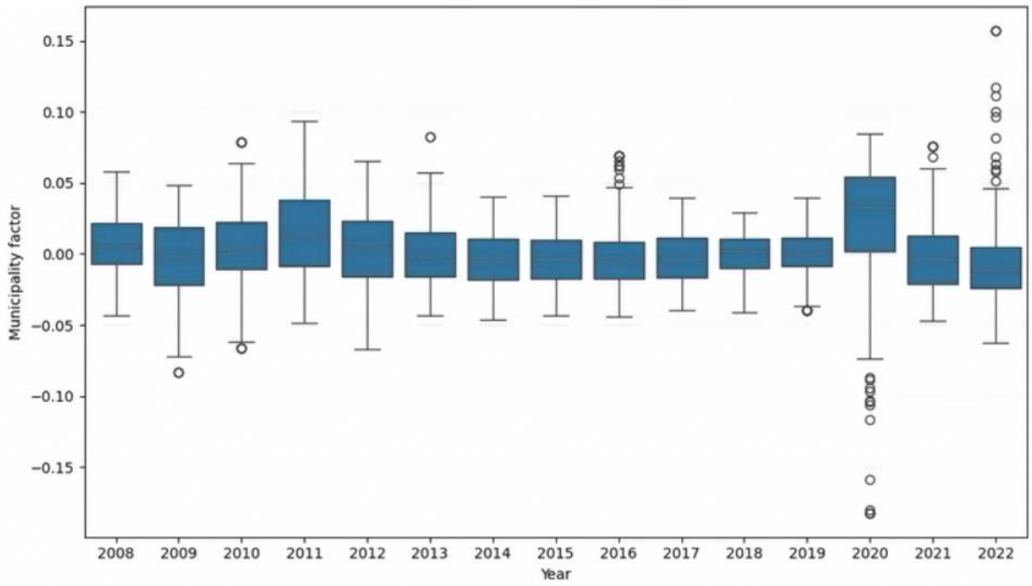


Figure 5 - Distribution of the Municipality Factor per year

These visualizations confirm the "fat tail" diagnosis:

1. **Stability vs. Stress:** In normal years, the interquartile range is relatively narrow. However, the dispersion explodes during the COVID-19 crisis (2020–2021).
2. **The Outlier Asymmetry:** Figure 4 (Industry) reveals massive outliers during the pandemic, far exceeding those in Figure 5 (Municipality). This confirms that the pandemic was not a uniform shock, but rather a highly asymmetric event that devastated specific industries (e.g., tourism, transport) while leaving others relatively unscathed.
3. **The Positive Median Nuance:** In both boxplots, the median shock during the COVID years is slightly positive or near zero. This seemingly counter-intuitive result, given the deep recession, is explained by the construction of the variables. Since these factors represent deviations from the aggregate (which itself collapsed, as shown in Figure 3), an industry that declined *less* than the average economy effectively records a positive relative shock. This highlights that while the "macro" environment was disastrous, the "micro" distribution was skewed: a few large industries suffered catastrophic losses (the negative outliers), dragging down the aggregate, while the median industry performed relatively better than the systemic collapse.

Table 2 indicates that group factors display comparatively strong tail behavior and high cross-group heterogeneity. The annual boxplots reinforce a different notion of “stability” than low volatility: the shape of the cross-group distribution is remarkably similar across years, typically centered close to zero but consistently wide, suggesting that group-specific deviations are persistent in dispersion rather than dominated by a few exceptional years. In other words, the group dimension does not exhibit large time-varying shifts in its overall distribution, even though individual groups can move around within that distribution over time. The main exception occurs in the first pandemic year, when the group-factor distribution becomes markedly left-skewed, indicating that group-affiliated firms experienced relatively weaker growth than the typical firm in that year.

3.3.2 Factor Inter-Dependence Structure

Given the pronounced heterogeneity and fat-tailed risks documented in the previous section, it is crucial to understand the degree of interdependence between factors. To assess whether ownership networks add genuinely new dependence, we must first understand how much comovement is already captured by the baseline factors.

We assess factor dependence using the pre-pandemic sample (2008–2019) to avoid spurious correlations driven by the systemic COVID-19 shock. As shown in the previous analysis, the pandemic induced a synchronized collapse across most industries and several municipalities. Including this period would artificially inflate measures of comovement, masking the structural independence we aim to test.

3.3.2.1 Among Industry Factors

Figure 6 displays pairwise Ledoit–Wolf correlations across industry factors, spanning –0.73 to 0.87, with a standard deviation of 0.19 and an average extremely close to zero (0.003).

The symmetric, zero-centered distribution indicates very weak systematic co-movement between industry-specific growth deviations. Rare higher correlations do not shift the pattern: most industry pairs behave independently. This suggests sectoral shocks in Portugal are heterogeneous, driven by industry-specific dynamics rather than systemic forces.

Before netting the aggregate factor, raw correlations ranged from -0.53 to 0.92 with a mean of 0.41 (lower than the median of 0.45), reflecting strong commonality driven by the business cycle.

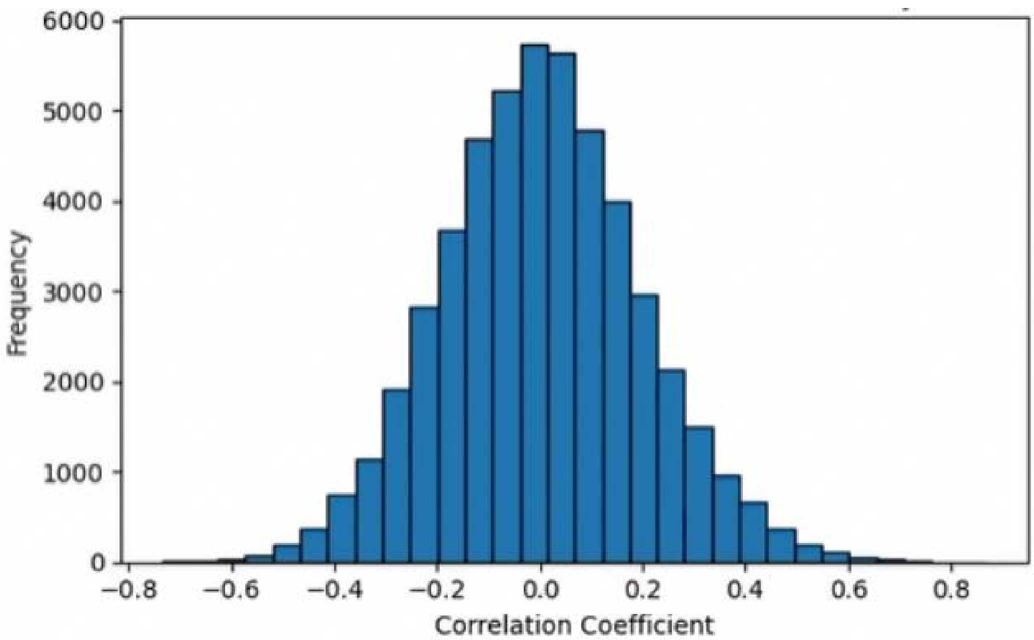


Figure 6 – Distribution of Pairwise Correlations Between Industry Factors

The 317 industries can be grouped into 19 sectors of activity. Figure 7 presents the heatmap of average pairwise correlations between the sector specific grouped by 19 economic sectors. The diagonal elements show the correlation between the industries in the same sector of activity. As expected, these are typically positive. Construction (06) exhibits the highest internal synchronization (0.41), suggesting strong exposure to common sector-specific factors. In contrast, Information and Communication (10) shows near-zero coherence. Notably, Real Estate (12) displays a slightly negative diagonal value, theoretically consistent with the substitution effect between sales and rental markets driven by shifting credit conditions.

Regarding off-diagonal interdependence, two distinct patterns emerge. First, a "Public Cluster" links Education (16) and Health (17) (correlation 0.26), likely reflecting shared sensitivity to state funding and fiscal cycles. Second, the Energy sector (04) illustrates a structural dichotomy. Although this sector aggregates distinct industrial activities, ranging from power generation to grid distribution and gas supply, its aggregate behavior correlates negatively with energy-intensive sectors like Construction (consistent with supply-shock dynamics). Conversely, it correlates positively with essential services such as Water and Transport, forming a defensive 'Utilities' block.

Despite these specific clusters, the broader network is remarkably sparse. The average off-diagonal correlation stands at a negligible 0.06, significantly lower than the within-sector average (diagonal) of 0.16. This contrast confirms that, after netting out the economy-wide factor, sectoral dynamics are largely independent, a structural fragmentation that mirrors the spatial decoupling observed in the municipal analysis.

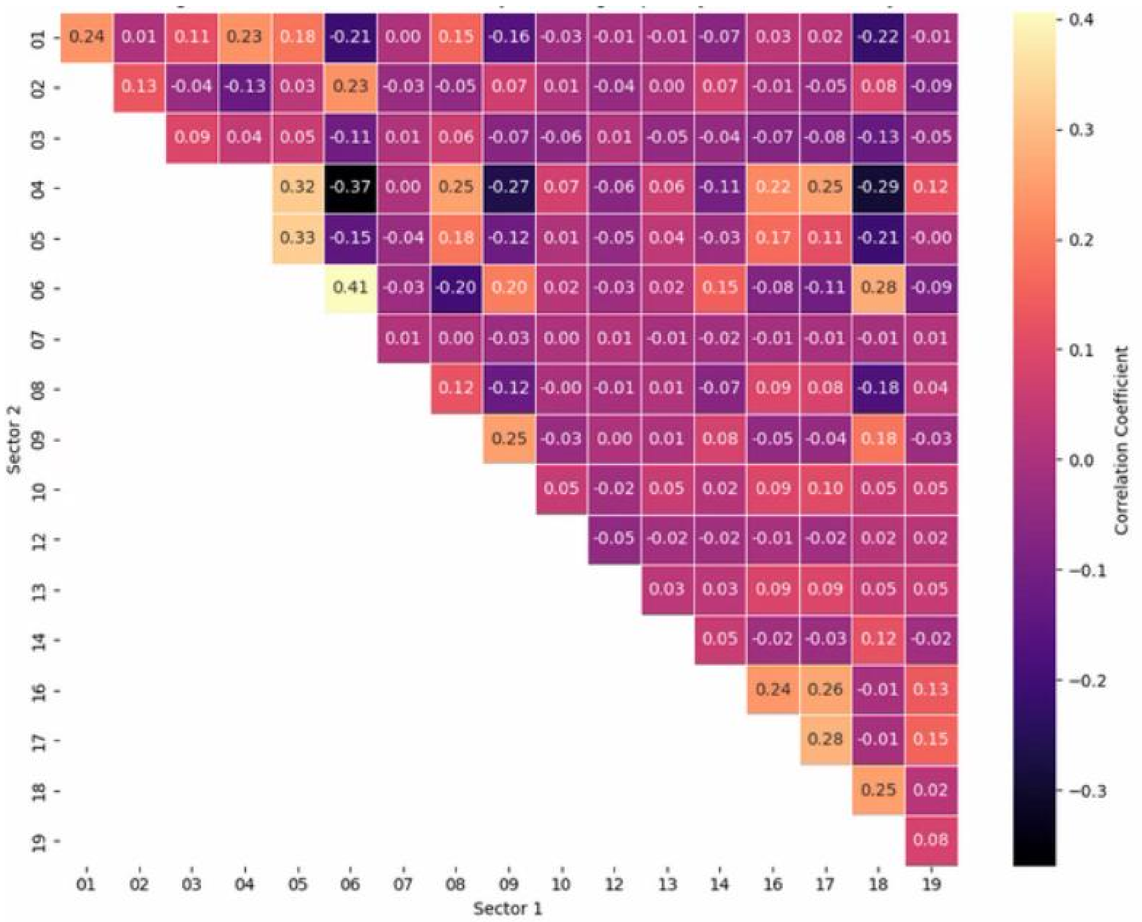


Figure 7 – Average Correlations between Industry Factors grouped by Sectors of Activity

3.3.2.2 Among Municipality Factors

For municipality clusters, the distribution of pairwise Ledoit–Wolf correlations is centred close to zero but slightly right-shifted, as shown in Figure 8. The range spans −0.67 to 0.70, with a small positive mean (≈0.019) and median (0.015).

This indicates that, on average, municipality-level growth deviations exhibit only weak co-movement across locations. In other words, after controlling for national trends, local shocks appear largely idiosyncratic to each municipality cluster rather than driven by a strong common spatial component. At the same time, Correlation dispersion (standard deviation ≈ 0.178) and heavy tails suggest heterogeneity: some municipalities move together (possibly due to shared regional exposure or supply-chain links), while others move in opposite directions, consistent with local specialisation and differing economic structures. Municipality correlations are marginally more positive than industry ones, implying slightly stronger residual local commonality.

Before netting the aggregate factor, raw pairwise correlations ranged from -0.12 to 0.91 (mean 0.70; median 0.71). This contrast underscores the national business cycle's dominance in driving raw local fluctuations.

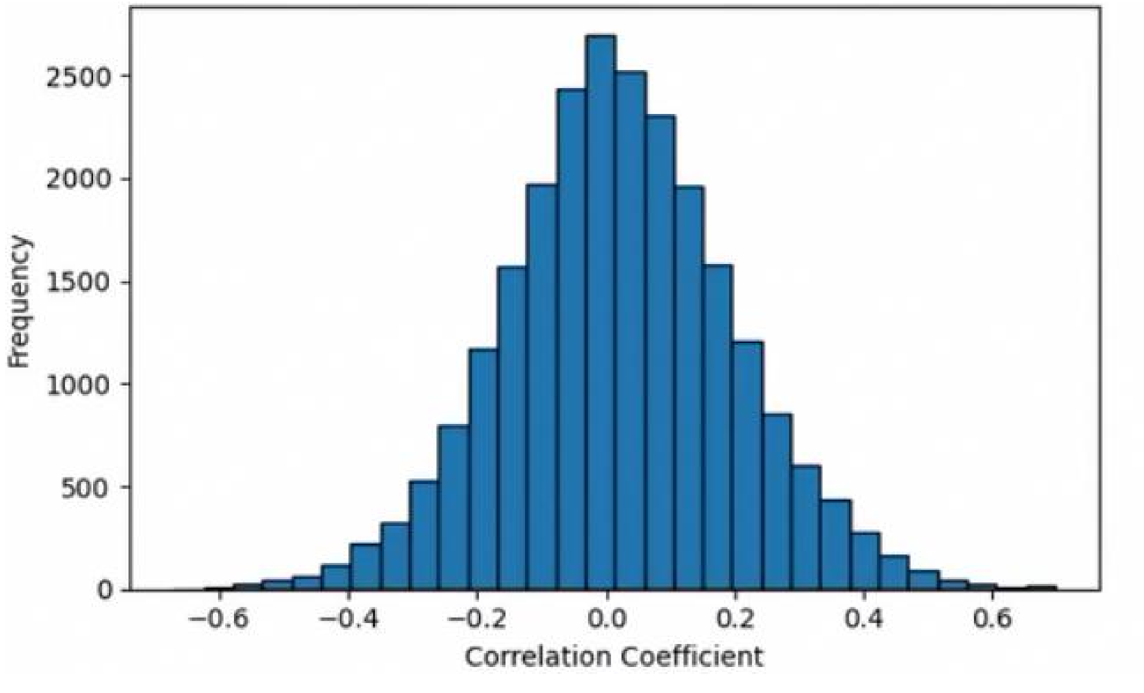


Figure 8 - Distribution of Pairwise Correlations Between Municipality Factors

Figure 9 displays the heatmap of average pairwise correlations between the 20 administrative districts.

A salient feature of the matrix is the distinct behavior of the 'Tourism Cluster' formed by the Faro district (08 - Algarve) and the Autonomous Region of Madeira (22). Both regions exhibit the highest internal synchronization (0.47 for Faro and 0.43 for Madeira), indicating an extremely high degree of coherence consistent with their specialized economic base and simultaneous exposure to external demand shocks.

They exhibit the highest mean absolute correlations in the sample, yet the lowest medians. This signals a polarized dynamic that validates the structural dichotomy of the Portuguese economy. Faro and Madeira are strongly synchronized with each other (positive cross-correlation of 0.32) but exhibit robust negative correlations with the industrial and inland districts (Guarda, Viseu). This confirms that the coastal tourism engines operate on a distinct business cycle that often diverges from the national industrial base.

In sharp contrast, the large districts of Setúbal, Lisbon, and Porto display the lowest mean absolute correlations and medians close to zero, supporting the diversification argument. As large, dense economic hubs hosting a heterogeneous mix of firms, these districts benefit from the Law of Large Numbers. Internal idiosyncratic shocks tend to average out, resulting in a local residual component that behaves as unrelated noise, orthogonal to the specific dynamics of smaller, more specialized districts.

Consistent with the sectoral analysis, the heatmap reveals a stark contrast between intra-district and inter-district dynamics. While the diagonal elements represent the full correlation of local shocks within their own district (averaging 0.15), the average off-diagonal correlation drops precipitously to near zero (0.02).

This quantitative gap implies that, generally, regional dynamics are idiosyncratic. However, the notable exceptions to this trend are likely driven by geographical proximity or shared supply chains. This suggests that while nationwide spillovers are limited, specific regional sub-clusters remain strongly interconnected.

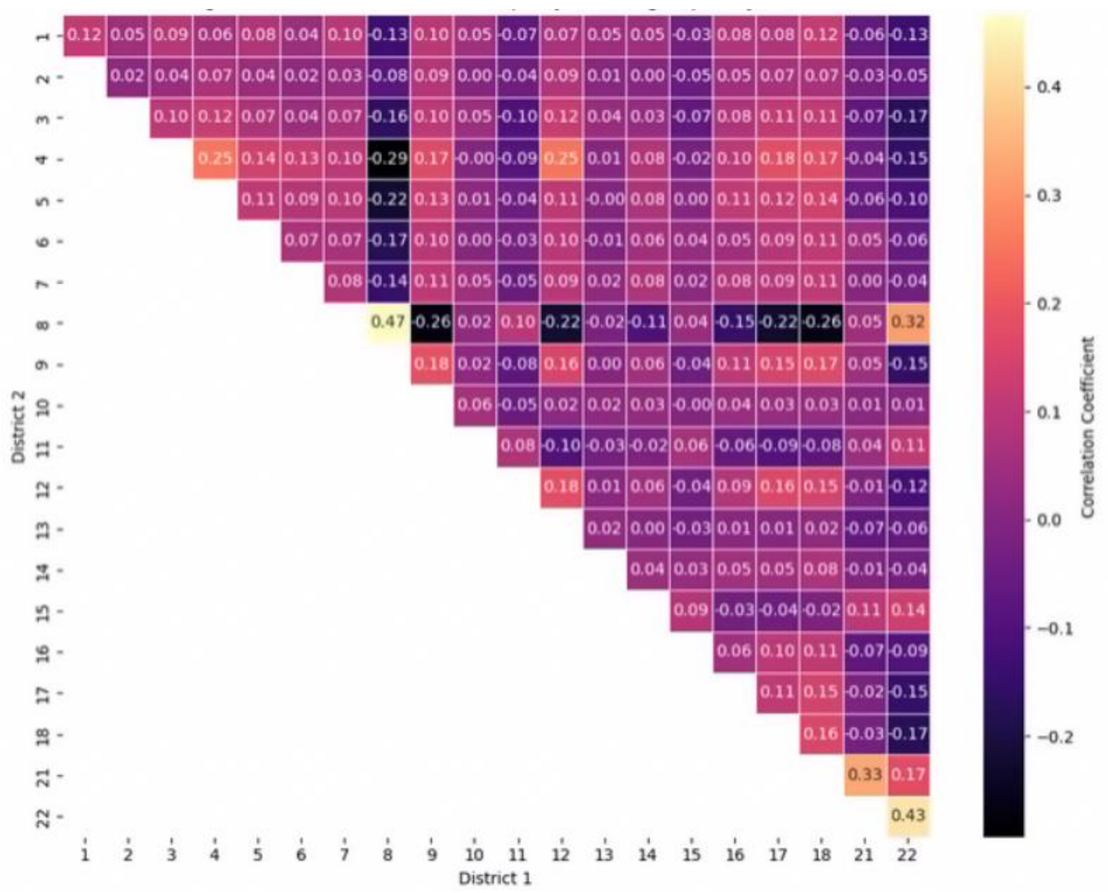


Figure 9 - Average Correlations between Municipality Factors grouped by District

3.3.2.3 Among Economic Group Factors

Empirically, group factors display weak cross-correlation once the aggregate factor is removed (mean ≈ 0.003, median ≈ 0.001, standard deviation ≈ 0.042). While correlations range from -0.45 to 0.82, the distribution is sharply concentrated around zero. We do not pursue a richer correlation mapping (e.g., histograms) for business groups because, in this thesis, the group factor is used to capture within-group dependence for the Super-Firm aggregation rather than to study between-group network comovement. Moreover, groups do not form a standardized economic partition like industries or municipalities, making cross-group correlation summaries less directly comparable across units. Nevertheless, the near-zero average correlation indicates that, net of the aggregate cycle, group-specific fluctuations are not organized around a strong common component across groups.

3.3.2.4 Between Categories of Risk Factors

We observe virtually no statistical dependence between the common macro component and the disaggregated factors, with mean coefficients of -0.025 for industries and -0.09 for municipalities. This orthogonality confirms that the economy-wide factor successfully absorbs the common business cycle, validating our identification strategy. The results ensure that the estimated factors isolate dynamics strictly inherent to specific industrial and regional clusters, independent of the aggregate trend.

The cross-sectional correlation structure between industry and municipality cluster factors also exhibits very weak co-movement. The distribution of pairwise Ledoit–Wolf correlations (Figure 10) ranges from –0.65 to 0.68, with a mean of 0.003 and a median close to zero. The standard deviation of 0.147 indicates a tighter concentration around zero than in the within-industry or within-municipality distributions. These results show that deviations in industry-specific sales growth are largely orthogonal to deviations in municipality-specific growth. In other words, local economic shocks do not systematically align with sectoral shocks. This weak association suggests that industrial structure and geographic location capture distinct dimensions of heterogeneity in firms' growth dynamics, reinforcing the importance of separating these two factors in the volatility decomposition.[1]

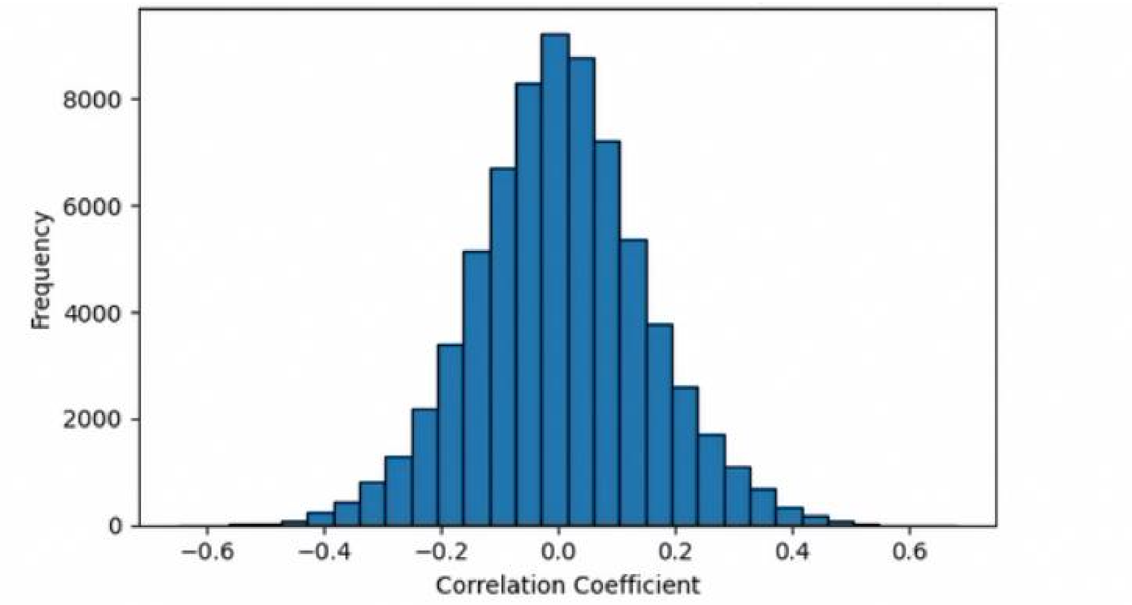


Figure 10 – Distribution of Pairwise Correlations between Industry and Municipality Factors

3.3.3 Assessing Factor Normality

Since the volatility framework used in this thesis relies on a covariance-based decomposition, it is essential to verify whether the estimated factors are sufficiently well-behaved for second-moment analysis to remain informative. To this end, we assess whether the extracted factor series display major departures from normality by applying the Shapiro–Wilk test at a 5% significance level, a procedure well-suited to our relatively short time samples ($T \approx 15$).

The results reveal a clear contrast between aggregate and disaggregated dynamics, with the COVID-19 shock playing a central role. For the aggregate factor, the null of normality is not rejected in the full sample. However, once the pandemic years are excluded, the aggregate factor fails the test, suggesting that the extreme dispersion associated with COVID masks an underlying non-normality in the aggregate business cycle.

For the disaggregated factors, the pattern is reversed. In the full sample, the null of normality is rejected for 20% of industry factors, 9% of municipality factors, and 20% of business-group

[1] Correlations between the group factor and the industry and municipality factors are explored in Section 6.2.4

factors. Excluding pandemic years reduces rejection rates (9% for industries, 5% for municipalities, 13% for groups), suggesting that non-normality is driven by the 2020–2021 disruption rather than persistent structural departures from Gaussian behavior.

Business-group factors remain somewhat less regular than the others even in the pre-pandemic sample. Their higher rejection rate appears to be associated with more pronounced negative skewness and excess kurtosis, consistent with occasional sharp downside adjustments, but also with some groups having a lower number of firms.

Still, the overall evidence suggests that outside extreme crisis periods the factor distributions are sufficiently well-behaved for variance-based analysis to remain informative. In that sense, the covariance framework adopted in the thesis remains an appropriate approximation, even if ownership-related factors retain a somewhat more asymmetric risk profile than industry and municipal factors.

4. Testing the Preconditions for Granularity

Before turning to factor-based measurement of aggregate volatility, it is useful to assess whether the Portuguese economy exhibits the structural environment in which firm-level shocks - and, by extension, ownership networks - could plausibly matter in the aggregate. In the granular literature, this requires at least two conditions: firm size must be sufficiently concentrated for the upper tail to carry substantial weight and volatility must not decline too quickly with size, otherwise shocks to large firms would be mechanically dampened by scale. This section evaluates both conditions and then examines whether business groups alter the size–volatility relationship.

4.1. Sales concentration: descriptive evidence

Figure 11 highlights the high degree of concentration in the Portuguese economy: the top 1% of firms account for roughly 50% of total sales, a dominance that remains substantial (about 42%) even when micro firms are excluded.

Figure 12 reports the evolution of standard inequality measures for the cross-sectional distribution of firm sales. All three indices also indicate a persistently high level of concentration throughout the sample, with the Dissimilarity index averaging approximately 0.68, the Gini coefficient approximately 0.82, and the Theil index approximately 2.3. Concentration rises mildly around 2012. While part of this pattern may reflect sample-composition changes, given the contraction in the number of active firms, it is also broadly consistent with the disruption associated with the Portuguese sovereign debt crisis.

Taken together, these descriptive facts suggest that economic weight is concentrated in a relatively small upper tail. The next subsection formalizes this observation by testing whether that upper tail follows a Pareto/Zipf pattern.

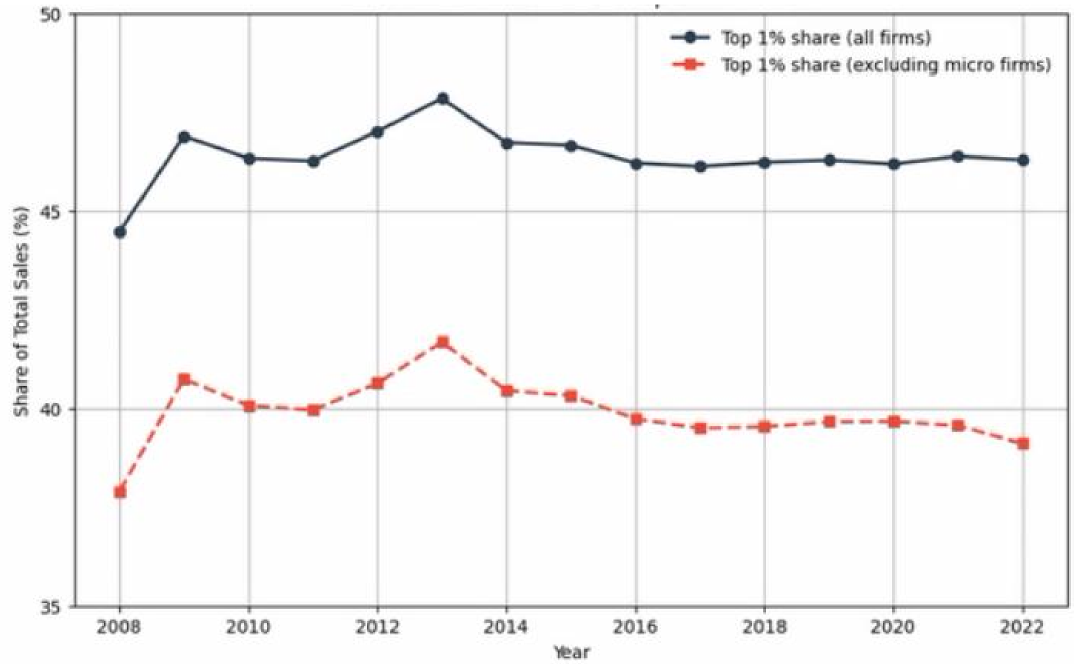


Figure 11 – Sales Concentration in the Top 1% of Firms

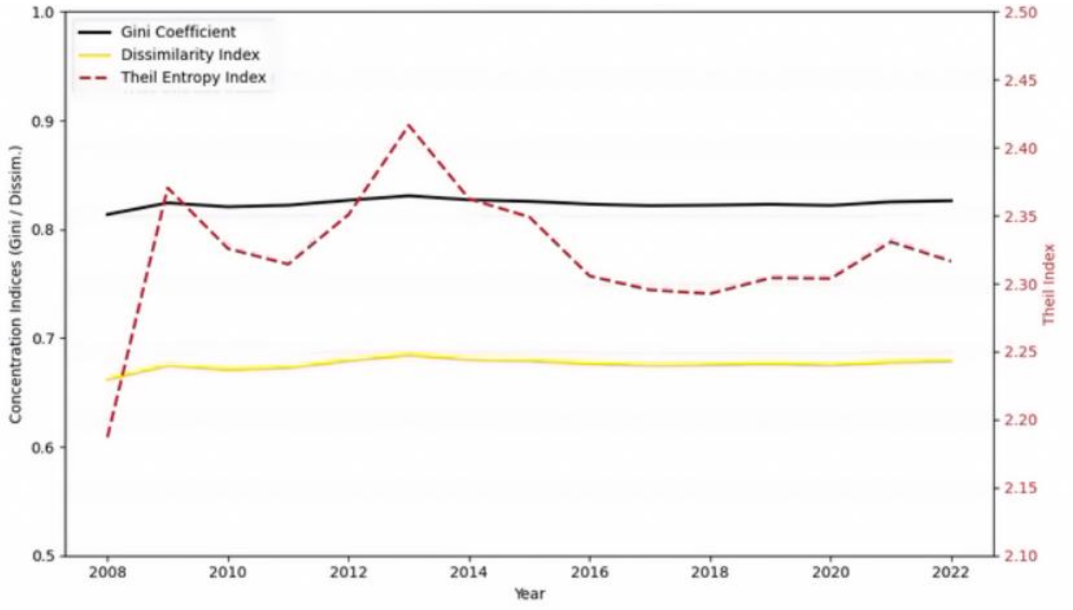


Figure 12 – Concentration of Real Sales

4.2. Evidence of Economic Concentration: Zipf's Law

To assess the degree of concentration, we conduct a two-stage empirical analysis of the firm-size distribution.

First, to evaluate temporal stability, we estimate the Zipf (Pareto) exponent ζ_t separately for each year in the sample period (2008–2022). Following applied work on Portuguese firms (Cabral, Gouveia, and Manteu, 2020), we focus on the upper tail of the distribution and estimate a log rank–size regression with the Gabaix–Ibragimov (2011) rank correction ($\text{Rank}-1/2$), which reduces small-sample bias in OLS-based tail estimation. As a practical cut-off, we retain the top 1,000 firms by size in each year.

To assess robustness, we also compute the Hill tail-index estimator and report estimates from a standard log rank–size regression without the rank correction. Across specifications, the estimated Zipf exponent remains consistently heavy-tailed, fluctuating around 1.27.

This magnitude is economically meaningful because it lies in the range $1 \leq \zeta < 2$. For a Pareto tail in this interval, the second moment diverges asymptotically, implying that standard LLN/CLT-style diversification is substantially weakened: shocks to very large firms need not average out rapidly and may therefore retain a non-negligible imprint on aggregate fluctuations.

Second, to filter out transitory annual shocks and capture the long-run hierarchy of firm sizes, we construct a time-averaged size measure by defining each firm's "structural size" as its average real sales over the sample period. As illustrated in Figure 13, estimating the rank–size relationship on this smoothed cross-section yields a Zipf exponent of 1.18 with a high goodness-of-fit ($R^2 \approx 0.97$), close to the benchmark documented by Axtell (2001) for the U.S. economy.

The evidence indicates the Portuguese firm-size distribution is strongly heavy-tailed. This provides a first reason to take firm-level shocks seriously in aggregate volatility measurement: the upper tail is sufficiently concentrated for large entities to matter quantitatively.

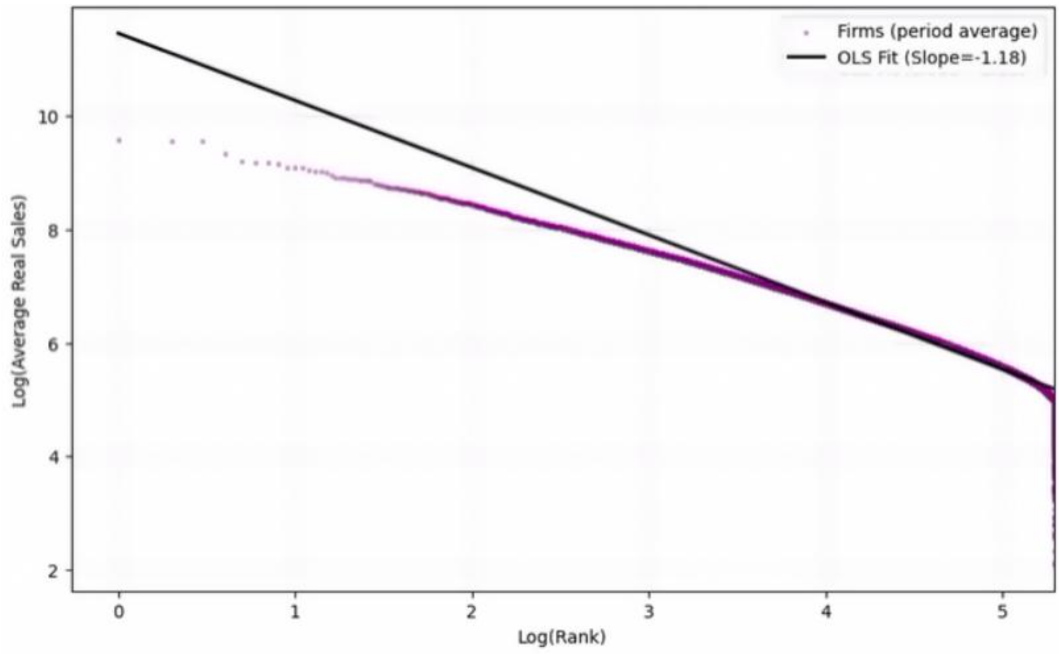


Figure 13 – Zipf's Law: Distribution of Average Real Sales

4.3 Scale Invariance of Firm Volatility

Concentration alone is not sufficient for firm-level shocks to matter in the aggregate. A second condition is that dominant firms should not become too stable as they grow. Following Gabaix (2011), we estimate the size–volatility scaling relationship $\sigma_i = kS_i^{-\alpha}$ through the log-linear specification:

$$\ln(\sigma_i) = c - \alpha \ln(S_i) + \epsilon_i.$$ [2]

where α is the volatility–size scaling exponent.

To reduce sensitivity to transitory shocks and outliers, firm size S_i is measured as median real sales over the sample period, while firm volatility σ_i is measured using the median annual sales-growth volatility estimates reported in Silva (2025).

While classical diversification implies $\alpha = 0.5$, Figure 14 reveals weak scaling ($\alpha \approx 0.078$). This is closer to the Gibrat-style benchmark ($\alpha = 0$, approximately size-independent volatility) and far below the full-diversification case $\alpha = 0.5$. and confirms that volatility decays slowly with scale. Quantitatively, a firm 100 times larger is only about $100^{-0.078} \approx 0.70$ as volatile.

[2] In Gabaix (2011), under Zipf-like size distributions and $\alpha \leq 1/2$, the same exponent governs the scaling of volatility with size at both firm and aggregate levels (Corollary 1), motivating the focus on α as a key parameter for potential macro relevance.

Although the fit is low ($R^2 \approx 0.05$), indicating substantial dispersion in volatility at a given size, the estimated slope implies that volatility decays only slowly with scale. A binscatter (100 quantiles) confirms that this pattern is not driven by outliers.

The main implication is that volatility is not mechanically dampened by size in Portugal: the upper tail is not only large, but remains sufficiently volatile for shocks to large firms to be potentially aggregate-relevant.

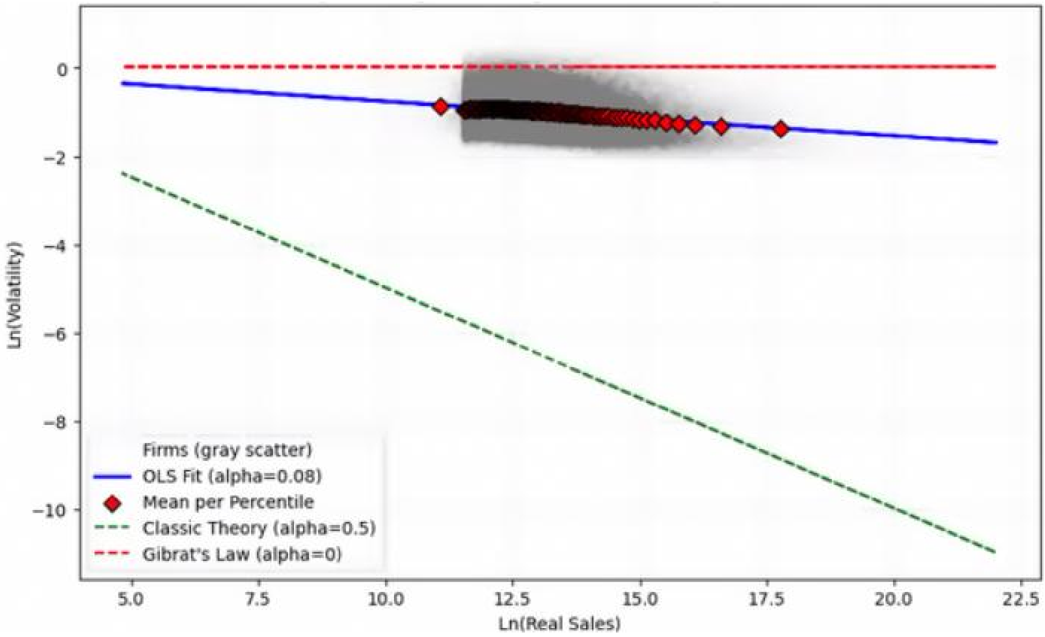


Figure 14 – Size–Volatility Scaling Relationship: Ln(Volatility) vs. Ln(Sales)

4.4 The Role of Business Groups in the Size-Volatility Relationship

Against the backdrop of strong concentration and weak volatility scaling, we next examine whether business groups display a distinct risk profile by treating each major group as a consolidated "Super-Firm".

We construct each group's scale by aggregating subsidiaries' real sales within each year and taking the median over the sample period. Group volatility is computed using a structural portfolio approach rather than short-run realized group time series. Crucially, instead of imposing a single intra-group correlation parameter, we allow within-group dependence to reflect the group's sectoral and municipal composition. We do so using a covariance-based aggregation logic that incorporates aggregate, industrial, geographic, and group components, formally introduced in the next chapter, yielding a structural, time-aggregated measure suitable for cross-sectional comparison.

We then map business groups into the same size–volatility diagram used for standalone firms. The benchmark for the Portuguese economy is weak scaling, with $\alpha \approx 0.075$. Relative to this benchmark, business groups display only a modest downward shift: the average log-residual is approximately -0.05, and 57.5% of groups lie below the market scaling line. Economically, this corresponds to roughly 5% lower volatility than standalone firms of comparable size ($e^{-0.05} \approx 0.95$).

Overall, once within-group covariance is structured by aggregate, sectoral and geographic dynamics, the remaining “ownership-only” stabilization in the size–volatility relationship is small.

This result previews the main aggregate finding in Chapter 6: even when group affiliation affects consolidated entities, its incremental contribution to aggregate volatility is limited.

4.5. Granular Signals and Their Limits

The evidence above suggests that Portugal satisfies the main structural conditions under which firm-level shocks could matter in the aggregate: firm size is highly concentrated and volatility declines only weakly with size.

To assess whether this mechanism appears directly in the data, we construct the granular residual of Gabaix (2011) based on the top K firms:

$$\Gamma_t = \sum_{i=1}^{K} \frac{S_{i,t-1}}{Y_{t-1}} \hat{\epsilon}_{it}$$

where $S_{i,t-1}/Y_{t-1}$ is the firm's lagged sales share in the relevant aggregate. The relevant aggregate measure can be GDP growth or sales growth. Firm-specific shocks are measured using regression residuals $\hat{\epsilon}_{it}$ from the specification introduced next, which nets out aggregate, industry, and municipal components.

A standard benchmark in the literature is to consider $K = 100$. In this case, the granular residual has limited explanatory power in our sample ($R^2 \approx 0.1, p \approx 0.3$) both for GDP growth and for aggregate sales growth. [3] Tightening the cutoff to $K = 50$, given that Portugal has a smaller firm universe than the U.S. economy studied by Gabaix and so avoiding diluting the signal of the largest “granular” firms, raises explanatory power to $R^2 = 0.16$ ($p = 0.18$), suggesting that any granular signal is concentrated at the extreme upper tail, albeit still statistically weak in this short time series.

Moreover, the time-series fit is uneven, weakening notably during 2010–2015, consistent with the idea that severe systemic episodes synchronize firms broadly and dampen the relative contribution of firm-specific shocks.

Overall, these results do not contradict concentration; rather, they indicate that concentration alone is insufficient for a simple granular-residual mapping to track aggregate volatility. This motivates the factor-based framework developed in the next chapter, which measures aggregate volatility through an explicit decomposition into a concentrated variance component and a structured covariance component.

[3] More can be read at Appendix A.2

5. Methodology

5.1. Model Specification and Factor Structure

We now formalize the framework that allows us to quantify how micro-level dependencies map into aggregate volatility.

To model the cross-section of firm-level growth rates we use an observable-factor representation:

$$X_t = \alpha + B_t F_t + \varepsilon_t + \text{Controls}, \qquad t = 1, \dots, T,$$

where $X_t \in \mathbb{R}^{N\times 1}$ collects firm-level sales growth rates $g_{i,t}$ in year t, $\alpha \in \mathbb{R}^{N\times 1}$ is a vector of firm fixed effects (time-invariant means), $F_t \in \mathbb{R}^{K\times 1}$ stacks the $K = 538$ observed common factors in year t, $B_t \in \mathbb{R}^{N\times K}$ is the exposure (loading) matrix capturing each firm's sensitivity to these factors, $\varepsilon_t \in \mathbb{R}^{N\times 1}$ collects residual firm-level shocks. Controls will be described below. Here, N denotes the number of firms and $T = 12$ the number of years in the estimation sample after excluding the pandemic period.

We assume $\varepsilon_{i,t}$ is conditionally mean-zero and orthogonal to the factor space. Firm-specific variability is allowed to be heteroskedastic and is proxied by the annual firm-level sales-growth volatilities reported in Silva (2025), rather than estimated from regression residuals. For tractability, the aggregation exercise uses the factor-implied covariance structure.

We estimate factor loadings by fixed-effects OLS using the constructed aggregate, industry, and municipality factors (and controls). In particular, to prevent the COVID-19 shock from contaminating the estimation of structural exposures, we include industry–year interaction controls for the pandemic window. The fitted component captures the variation aligned with the observable factor space, while the residuals collect firm-level movements not explained by these components. The baseline fixed-effects regression has $R^2 = 6\%$ (within).

In scalar form, the baseline specification is:

$$g_{i,t} = \alpha_i + \beta_e F_{e,t} + \beta_s F_{s(i),t} + \beta_m F_{m(i),t} + \varepsilon_{i,t} + \text{Controls},$$

where $F_{e,t}$ is an aggregate factor, and $F_{s(i),t}$ and $F_{m(i),t}$ are the industry- and municipality-specific factors associated with firm i's industry $s(i)$ and municipality $m(i)$. In the empirical implementation:

- Aggregate factor: economy-wide median real sales growth.
- Industry factor: industry-specific deviation from the aggregate trend.
- Municipality factor: municipality-specific deviation from the aggregate trend.

The estimated factor loadings ($\beta_e = 0.69$, $\beta_s = 0.94$, $\beta_m = 0.85$) confirm a positive and significant association between firm-level growth and the dynamics of their respective clusters. The results highlight that industry affiliation is the strongest driver of firm performance.

Regarding the COVID-19 period, the pandemic controls indicate a disruption in these standard transmission channels, with heterogeneous sectoral reactions driving the breakdown in the typical associations.

The factor matrix F is formed by stacking the time series of these factors. The exposure matrix B is structured by two identification restrictions:

1. **Sparsity:** firms load only on the industry and municipality factors that correspond to their own clusters; loadings are zero for all other clusters.
2. **Homogeneity within clusters:** exposures are assumed common within each dimension, matching the coefficients estimated in the fixed-effects regression.

5.2. Covariance Estimation and Aggregation Weights

Under the hypotheses established in Section 5.1, the factor structure presented is enough to fully describe the interdependence between the firms sales growth rate. Instead of estimating the pairwise correlation between firms, we only need to estimate the dependence structure between the factors, represented by the covariance matrix $\hat{\Sigma}_F$. The factor-implied covariance matrix for firm-level growth is then given by:

$$\mathrm{Cov}(X_t) = \Sigma_{\mathrm{X,t}} = B_t \hat{\Sigma}_F {B_t}^{\mathrm{T}} + \Sigma_{\varepsilon,\mathrm{t}}$$

The diagonal of $\mathrm{Cov}(X_t)$, denoted as $\sigma_{i,t}$, corresponds to the estimates provided by Silva (2025). Thus, we replace this diagonal with the observed total volatility $\sigma_{i,t}^2$, while keeping the off-diagonal covariances generated by the factor structure. This ensures that aggregate volatility reflects both (a) the empirical volatilities of firms and (b) the systematic comovement captured through Σ_F.

Given the high dimensionality of the model relative to the time dimension (T), a standard sample covariance estimation would be ill-conditioned. We therefore estimate $\hat{\Sigma}_F$ using the Ledoit–Wolf shrinkage estimator, which provides a well-conditioned covariance estimate with improved mean-squared-error properties in high-dimensional settings. Only the pre-pandemic period was used to avoid COVID-driven spurious comovement.

To aggregate firm-level dynamics into an economy-wide series, we define time-varying sales weights:

$$w_{i,t} = \frac{\mathrm{Sales}_{i,t}}{\sum_{j=1}^{N} \mathrm{Sales}_{j,t}}, W_t = (w_{1,t} \quad \cdots \quad w_{N,t})'.$$

This yields the sales-weighted (SW) aggregate. The equal-weighted (EW) counterpart sets $w_{i,t} = 1/N_t$ within each year.

The implied aggregate variance is then given by:

$$\sigma^2_{\text{Agg}}(t) = W_t' \, \hat{\Sigma}_{X,t} \, W_t = \underbrace{\sum_{i=1}^{N} w_{i,t}^2 \, \sigma_{i,t}^2}_{\text{Observed diagonal contribution}} + \underbrace{\text{OffDiag}(t)}_{\text{Factor-implied covariance contribution}} ,$$

where

$$\text{OffDiag}(t) = W_t' \, (B_t \hat{\Sigma}_F {B_t}^T) \, W_t - \sum_{i=1}^{N} w_{i,t}^2 \, \left(B_{t,i}^{\mathrm{T}} \hat{\Sigma}_F B_{t,i}\right).$$

This construction has two useful interpretations:

- The first term is the independence lower bound implied by the observed total firm variability, obtained by ignoring all cross-firm covariances.
- The second term adds the systemic covariance contribution generated by the factor structure.

Because factor covariances may be negative, $\text{OffDiag}(t)$ can in principle reduce aggregate variance in some years; we therefore work with $\sigma^2_{\text{Agg}}(t)$ directly and verify non-negativity of the resulting variance series in implementation.

5.3. The Business Group Hypothesis: A "Super-Firm" Approach

So far, the model incorporates only interdependencies generated by the three categories of factors: aggregate, industry and municipalities. Firms may be interdependent also because they belong to the same business group, the focus of this dissertation. To evaluate the contribution of this factor, we consolidate subsidiaries into business groups and treat each consolidated unit as a single "Super-Firm". The economy is therefore redefined as a collection of

$$(N_{\text{firms}} - N_{\text{subsidiaries}}) + N_{\text{groups}}$$

entities. Importantly, this redefinition includes both: (i) multi-firm business groups, and (ii) standalone firms, which can be viewed as degenerate Super-Firms consisting of a single firm.

Let g index Super-Firms. If g is a business group, let G_g denote its set of constituent firms. If g is a standalone firm, then G_g contains only that firm. Group-level (Super-Firm) sales are defined as:

$$\text{Sales}_{g,t} = \sum_{i \in \mathcal{G}_g} \text{Sales}_{i,t} \, .$$

The Super-Firm's macroeconomic weight is then:

$$W_{g,t} = \frac{\text{Sales}_{g,t}}{\sum_j \text{Sales}_{j,t}} .$$

5.3.1. Assigning industry and municipality to Super-Firms

To map Super-Firms into industry and municipal exposures, we use sales-share weights rather than a single dominant classification. This means that super-firms can be exposed to multiple industry and municipality factors. For each Super-Firm $g$and year t, we compute the distribution of $Sales_{g,t}$ across industries and municipalities. For simplicity, we retain only exposures accounting for at least 10% of group sales. We then renormalize these exposures to sum to one within each dimension. This preserves the interpretation that the total industry (municipality) exposure of each Super-Firm sums to $\beta_s(\beta_m)$, while allowing diversified Super-Firms to span multiple clusters.

5.3.2 Super-Firm individual volatility

To capture the possibility that ownership structure generates an additional common source of comovement beyond aggregate, industry, and municipal factors, we introduce a dedicated Business Group Factor $\left(F_g\right)$. We therefore re-estimate the regression framework including this factor, expanding the loading matrix to capture exposure to F_g for affiliated firms. The estimated group loading is $\beta_{\text{Group}} = 0.84$, statistically significant at the 1% level. Importantly, the aggregate, industry, and municipal betas remain stable and significant at the 1% level.

Having constructed $B_t$and estimated the factor covariance matrix, we compute each Super-Firm's time-varying volatility $\sigma_{g,t}$ using the same covariance-based aggregation logic introduced in Section 5.2, now applied within each Super-Firm's constituent set $G_g$and its relevant factor subset. This delivers a Super-Firm volatility series that already reflects within-entity dependence implied by the factor structure.

6. RESULTS

This chapter presents the main empirical results on aggregate volatility measurement and evaluates the incremental contribution of business-group (ownership) linkages. Throughout, we report results under two alternative estimators for the factor covariance matrix $\hat{\Sigma}_F$: the Ledoit–Wolf shrinkage estimator and the raw sample covariance matrix. While Ledoit–Wolf provides a more stable and conservative benchmark in a high-dimensional setting, the raw covariance allows for a more granular exploration of time-varying dynamics. By analyzing both, we ensure that our conclusions are robust to the specific treatment of the dependence structure and provide a comprehensive view of the potential impact of ownership networks.

The key conclusion is that business-group affiliation is economically meaningful for the risk profile of consolidated entities, yet its incremental contribution to aggregate volatility is quantitatively small and sensitive in sign to the covariance estimator.

6.1 Benchmarks and Baseline Reconstruction

We begin by benchmarking model-based aggregate volatility against two reference points.

Realized aggregate volatility: The unconditional realized volatility of sales-weighted aggregate sales growth over the pre-pandemic estimation window is approximately 5.7%[4]. This serves as the primary empirical benchmark for our model, as it is computed over the same sample window used to estimate the dependence structure, ensuring a direct and consistent comparison with the factor-based reconstructions. This choice is further supported by our use of sector-time controls to net out pandemic-driven disruptions.

Independence lower bound: As a theoretical lower bound, we compute the aggregate volatility implied by the observed firm-level volatility measures under the assumption of cross-firm independence. This yields an aggregate volatility of approximately 0.97%. The large gap between this lower bound and the realized benchmark indicates that aggregate volatility cannot be explained by firm-specific variance alone: interdependence must play a central role.

Using the factor structure based on the economy-wide, sectoral, and municipal components, we obtain a range of plausible reconstructions for aggregate volatility. The Ledoit–Wolf estimator yields an average volatility of 3.281%, serving as a stabilized benchmark. In contrast, the raw sample covariance matrix yields 3.616%, capturing a broader range of the underlying variation

Although estimators differ in levels, both reconstructions remain significantly above the independence lower bound yet below the realized benchmark, indicating that standard observable factors capture an important, but incomplete, share of aggregate volatility.

One interpretation of the remaining gap is state-dependent dependence: correlations may rise in adverse states, so a covariance structure estimated from average second moments can understate realized volatility during crisis episodes (see, e.g., Ang and Chen, 2002, for evidence

[4] The corresponding volatility over the full sample is 7.2%, reflecting the inclusion of the Covid-19 shock years

of asymmetric conditional correlations). To shed light on when the model is closer to the realized benchmark, and on how much of reconstructed volatility comes from own-variance versus covariance, we next examine its time profile and decomposition.

Figure 15 reports the time profile of reconstructed aggregate volatility, for the raw sample covariance estimator case, and its decomposition into three components: (i) total reconstructed volatility (the final aggregate series), (ii) an independence benchmark (the diagonal/own-variance contribution implied by observed firm-level volatility, abstracting from cross-firm covariances), and (iii) a covariance contribution (the interdependence component implied by factor-driven covariances). The covariance contribution dominates throughout the sample, indicating that Portuguese aggregate volatility is driven primarily by cross-firm co-movement rather than firm-specific risk alone.

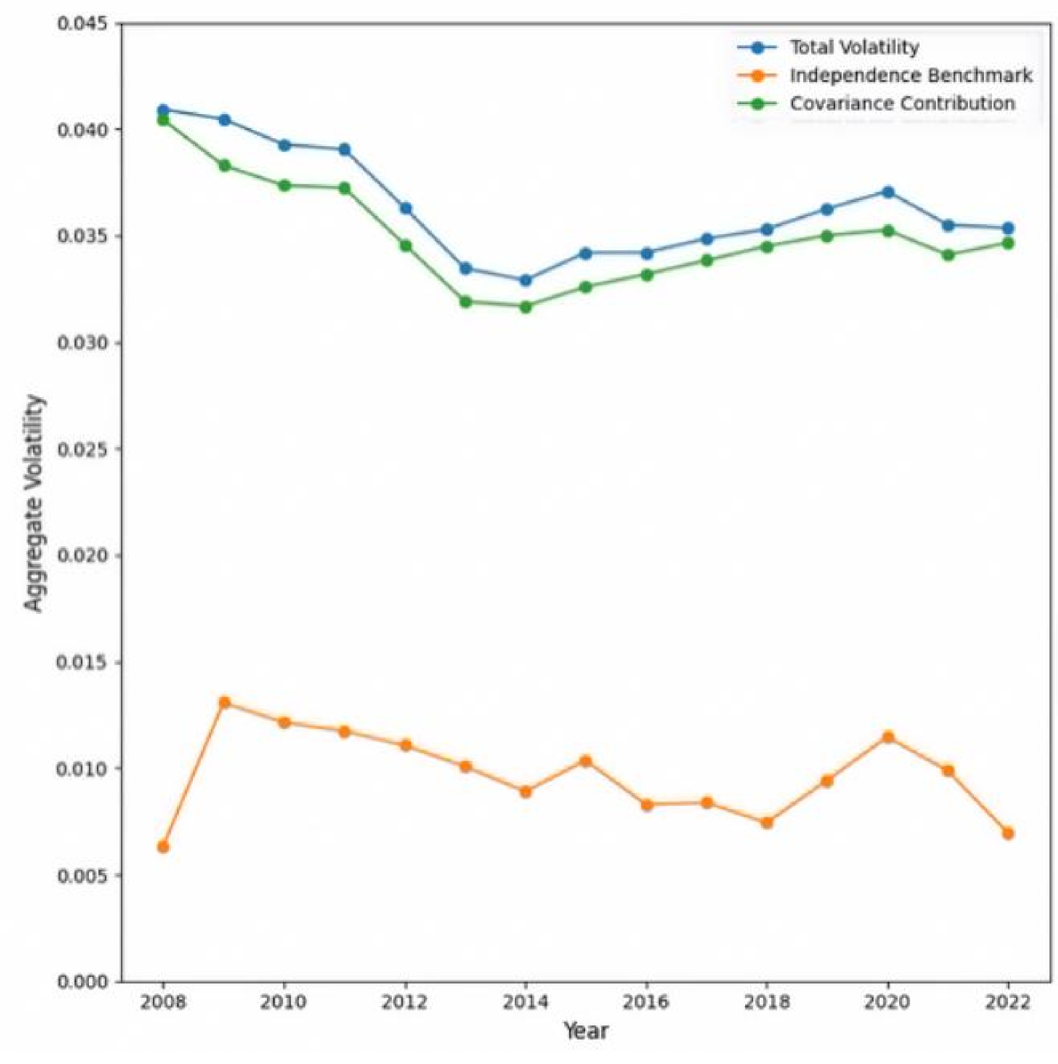


Figure 15 – Annual Evolution and Decomposition of Aggregate Volatility

Volatility in 2021–2022 should be interpreted with caution. To prevent the COVID-19 shock from mechanically dominating estimates of structural comovement, particularly the exposure matrix $B_t$and the factor covariance matrix $\hat{\Sigma}_F$, we estimate dependence using a pre-pandemic window and apply controls that net out pandemic-driven sector–time effects. Without this treatment, reconstructed volatility during the pandemic years would be substantially higher. Even so, the post-2019 period remains more volatile than the relatively stable inter-crisis years.

6.2 The Incremental Contribution of Business Groups

We next extend the framework by consolidating affiliated firms into business groups ("Super-Firms") and incorporating the group dimension in the dependence structure. This allows us to assess whether ownership-based linkages materially affect aggregate volatility measurement.

6.2.1 Increment from the Group Factor

Before assessing whether business groups move aggregate volatility, we quantify their effect at the entity level. We compare the Super-Firm volatility implied by the group specification with the counterfactual Super-Firm construction that excludes the group factor, holding the remaining factor structure fixed. This isolates the incremental within-network dependence captured by common ownership.

Introducing the group factor raises estimated Super-Firm idiosyncratic volatility relative to the specification without the group factor. Under Ledoit–Wolf, the increase is about 11% on average (median 5%). Under the raw sample covariance, the increase is larger, about 29% on average (median 15%), reflecting the greater sensitivity of within-group covariance to the estimator in a short time sample. The next subsections assess how much of this entity-level increment survives aggregation once Super-Firms are combined into economy-wide volatility.

6.2.2 Mechanical Effect: Independence Benchmark under Consolidation

A first effect appears even before considering firms interdependence. Aggregating subsidiaries into Super-Firms increases weight concentration, mechanically raising the independence benchmark. This would be the case even without the new business group factor because we are frontloading the independence created by the aggregate, industry and municipality factors in the case of these superfirms.

Under the independence hypothesis, aggregate volatility rises from 0.97% to 1.10% using the raw estimator, and to 1.00% under Ledoit-Wolf. This confirms that even after mitigating noise through shrinkage, consolidation still increases individual volatility.

6.2.3 Aggregate Effect: Baseline vs. Super-Firm Volatility

We then reintroduce factor-implied covariances between entities to obtain the full reconstructed aggregate volatility for an economy composed of standalone firms and Super-Firms and compare it with the baseline (firm-level) reconstruction.

Despite the sizeable entity-level increment documented in Section 6.2.1 and the increase in the independence benchmark under consolidation in Section 6.2.2, the incremental effect of introducing the group dimension on reconstructed aggregate volatility remains small. Under the raw sample covariance estimator, reconstructed aggregate volatility rises from 3.616% (baseline) to 3.653% (Super-Firms). Under Ledoit–Wolf shrinkage, reconstructed aggregate volatility increases only slightly from 3.281% to 3.283%. In both cases, the group specification shifts aggregate volatility by only a few basis points.

Importantly, these levels remain far below the realized benchmark of 5.7% discussed in Section 6.1, indicating that adding the group dimension does not materially close the gap to observed aggregate variation, even though common ownership is economically meaningful for the volatility of consolidated entities.

Figure 16 compares the baseline and Super-Firm specifications under the raw sample covariance estimator. The two reconstructed volatility series are nearly indistinguishable for most of the sample, reinforcing the conclusion that adding the group dimension has a negligible aggregate effect. The only visible divergence occurs in the final years.

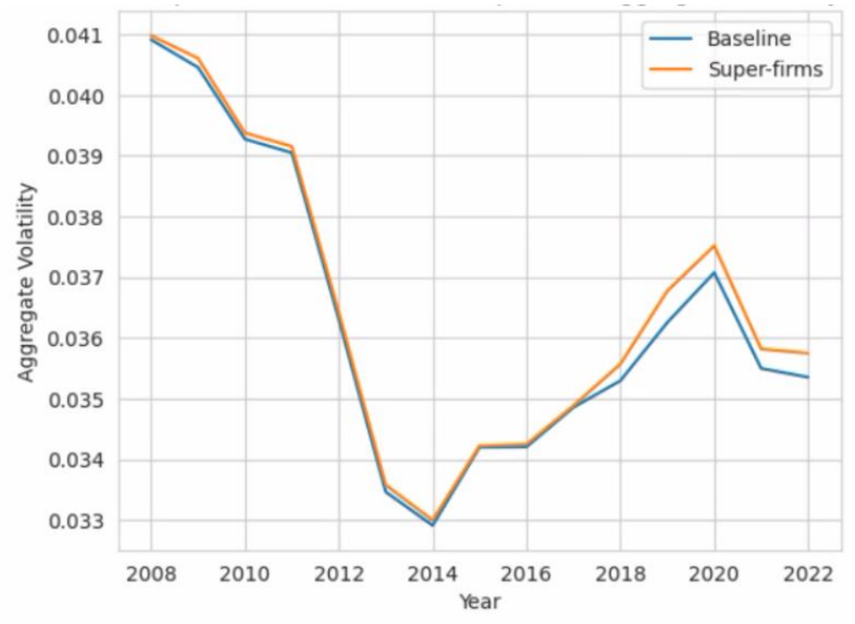


Figure 16 – Annual Evolution of Aggregate Volatility: Baseline vs. Super-Firm Specifications

To understand this late-sample nuance, we consider two non-mutually exclusive explanations.

First, a change in the economic footprint of group-affiliated activity in the effective (filtered) ownership network used for the Super-Firm construction. However, chapter 3 (Figure 2) shows that, while group coverage increases relative to the earliest years, it is broadly stable from the mid-sample onward, suggesting that the late divergence is unlikely to be driven purely by mechanical weight expansion.

Second, we examine time variation in the group-factor impact, measured as the relative change in Super-Firm volatility when including the group factor versus excluding it. As shown in Figure 17, the impact is higher in 2018–2019 and 2022, but it falls during the pandemic years (2020–2021), a pattern that may reflect the dominance of systemic shocks and/or the fact that $\hat{\Sigma}_F$is estimated using a pre-pandemic window. Figure 17 also indicates some non-negligible effects in the first three years of the sample in mean terms. However, this early impact is not mirrored in the median, suggesting that it is driven by a subset of groups rather than being broad-based. Moreover, those early years coincide with lower effective group weights and weaker group coverage (Chapter 3), which mechanically limits how much within-group effects can translate into the economy-wide aggregate. Consistent with this, Figure 16 shows little separation between the baseline and Super-Firm series in the early period, with the only visibly larger divergence emerging after 2018.

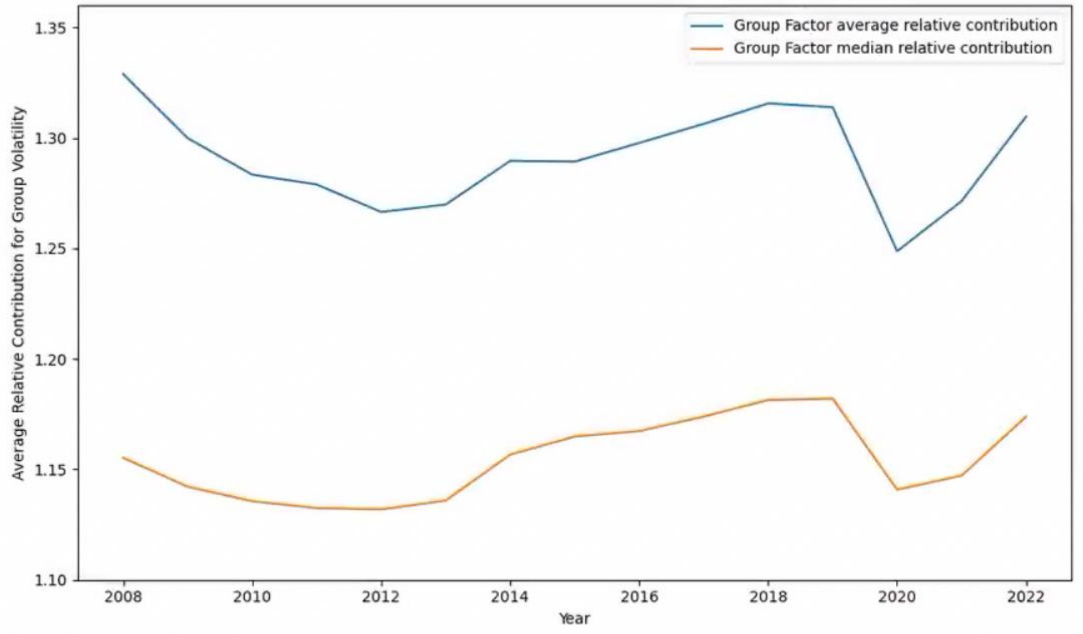


Figure 17 - Annual Evolution of the Relative Impact of the Group Factor on Individual Group Volatility

Importantly, this late-sample divergence is estimator-sensitive, which prevents it from being a primary conclusion of the analysis. Repeating Figure 16 under the Ledoit–Wolf shrinkage estimator yields baseline and Super-Firm series that are essentially superposed in all years. This suggests that the aggregate divergence observed under the raw estimator may be amplified by noise in a shorter time sample, which shrinkage effectively compresses.

This aggregate-level sensitivity **contrasts** with our micro-level findings: while the magnitude varies, the group factor's role in increasing individual volatility remains qualitatively robust across both estimators. Thus, ownership interdependence is a consistent feature of the micro-level data, but its translation into a late-sample aggregate jump is statistically fragile and fails to survive the shrinkage process.

We also explored whether the group-factor impact varies systematically across Super-Firm weight deciles. Under the raw sample covariance, the decile profiles do not deliver a robust additional pattern beyond the common drop in 2020–2021. Under Ledoit–Wolf, however, one regularity emerges: the lowest-weight decile exhibits a consistently smaller incremental effect. A plausible interpretation is that shrinkage attenuates noisier within-group covariance estimates more strongly for economically smaller groups, whose factor moments are less precisely estimated. Overall, these diagnostics suggest potential time variation in the group channel, but the lack of estimator robustness for the aggregate series implies that the evidence remains suggestive rather than conclusive.

6.2.4 Why Is the Aggregate Increment So Small?

The previous subsections establish a clear tension: group affiliation is economically meaningful at the entity level, yet the aggregate volatility barely moves when we consolidate firms into

Super-Firms and add the group dimension, despite a late-sample widening in the baseline vs. Super-Firm series. This subsection investigates why the micro-level effect translates so weakly into the aggregate.

It is important to note that the ownership channel is measured under a deliberately conservative identification strategy (Chapter 3): each firm is assigned at most one stable controlling group and minority stakes below the control threshold are excluded. This improves interpretability but narrows the measured ownership network and therefore mechanically limits how much the group channel can move aggregate volatility.

A second, purely mechanical constraint is scale and coverage. Unlike industry and municipality factors, which span the full firm universe by construction, the ownership channel covers only a subset of activity, 20.3% to 27.1% of total sales, and is split across thousands of entities. This implies that even economically meaningful within-group dependence has limited leverage on aggregate volatility.

There is one further mechanism that helps rationalize why Super-Firm volatility rises while aggregate volatility barely changes: the group channel partly overlaps with existing industry (and, to a lesser extent, municipality) structure, especially for economically large groups. To quantify this overlap, we construct within-group industry and municipality composites using only the clusters in which each group operates. These composites are sales-share weighted within the group (i.e., based on firms' sales shares inside the group, aggregated at the cluster level), so that the comparison emphasizes the group's economically dominant industries and municipalities. We then examine (i) the explanatory power (R-squared) of these composites for the group factor and (ii) the absolute correlation between the group factor and each composite. Consistent with this overlap, both the industry and municipality betas decrease by 0.04 in the expanded regression that includes the group factor, confirming that the group dimension absorbs variation previously attributed to these clusters.

Figure 18 reports these diagnostics by deciles of group economic weight. While decile medians are naturally noisy, the figure suggests that economically larger groups exhibit higher $R^2$and stronger alignment with the sector composite, whereas the municipal-alignment pattern is weaker.

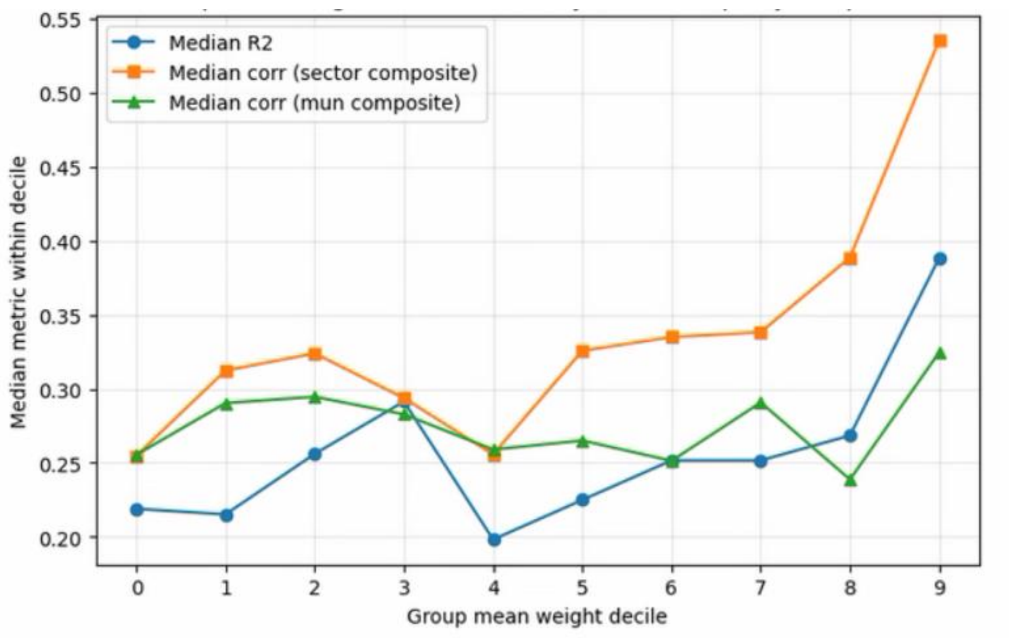


Figure 18 – Group Factor Alignment with Industry and Municipality composites

This visual impression is aligned with **Table 2**, that reports the distribution of the key group diagnostics (percentiles).

Table 2 - Business Group Factor Overlap with Sectoral and Municipal Factors

Percentile	P10	P25	P50	P75	P90
Group Sales Weight	0.001%	0.002%	0.005%	0.015%	0.042%
R^2	3.8%	10.8%	25.0%	45.3%	61.8%
Industry Alignment	0.063	0.165	0.332	0.555	0.725
Municipal Alignment	0.050	0.137	0.275	0.438	0.587

Complementing these descriptive patterns, non-parametric tests confirm the same message. Spearman rank correlations show that group weight is positively associated with R^2 ($\rho = 0.125$, $p = 3.13 \times 10^{-5}$) and even more strongly with industry alignment ($\rho = 0.181$, $p = 1.56 \times 10^{-9}$), while the association with municipal alignment is negligible ($\rho = -0.016$, $p = 0.60$). A top-decile split yields similar conclusions: the top 10% groups by weight have higher median R^2(0.388 vs. 0.235) and substantially stronger industry alignment (median $| corr |= 0.535$vs. 0.312), with Mann–Whitney differences statistically significant ($p = 0.0157$ and $p = 3.49 \times 10^{-10}$, respectively). Municipal alignment differs only modestly (0.325 vs. 0.269) and is not statistically significant ($p = 0.138$).

These diagnostics imply that for macro-relevant groups, the group factor overlaps mainly with industry dynamics, naturally limiting the incremental contribution of the group channel. Finally, as a note for future methodological improvements, the indication that factor correlation is stronger among larger entities suggests it would be useful to employ a sales-weighted covariance matrix. This refinement would ensure that correlations among the largest, macro-relevant firms are not underestimated in the aggregation process.

6.3 Sensitivity Analysis and Robustness

This section assesses how sensitive reconstructed volatility levels are to alternative assumptions, reporting both the baseline and Super-Firm specifications throughout. Additional diagnostics on the location and industry composition of large groups, including the "safe-haven headquarters" counterfactuals, are reported in Appendix A.3. We focus here on three implementation choices that could mechanically affect reconstructed volatility levels and the incremental group effect.

Weighting Scheme: Beyond the baseline sales-weighted (SW) results, we also evaluate an equal-weighted (EW) aggregation.

Under EW, the independence benchmark becomes negligible (≈0.137%), because equal weights remove the influence of extreme sales weights in the diagonal term $\sum_i w_i^2 \sigma_i^2$, mechanically increasing the relative importance of the covariance component.

Empirically, reconstructed volatility rises to 3.44% (Ledoit–Wolf) and 4.11% (raw sample covariance). This shows that aggregate volatility is sensitive to how weights interact with the covariance structure: equal weighting diversifies the diagonal component but can amplify the covariance contribution more than proportionally. A more detailed diagnostic is provided in Appendix A.4.

Note: equal-weighting is reported for the baseline firm universe. Applying equal weights to Super-Firms is less informative because consolidation changes the unit of observation and the interpretation of a "representative" exposure.

Cluster-assignment stability. In the baseline construction, each firm is assigned a stable industry and municipality identity to mitigate administrative noise from occasional reclassifications. As a robustness check, we remove this stabilization step and allow firms to switch clusters freely over time (including ties rather than forcing a unique modal cluster). Reconstructed volatility levels remain extremely similar. The largest deviation occurs under the raw sample covariance estimator in the Super-Firm specification, where reconstructed volatility changes by only about 0.015 percentage points. Importantly, this does not affect any qualitative conclusions, including the baseline vs. Super-Firm comparison and the interpretation of the small incremental group effect.

Group-identification threshold. A second strong assumption concerns how affiliates are consolidated into Super-Firms. We therefore relax the minimum equity threshold used to define an economically meaningful control link, lowering it from 10% to 5%. The resulting Super-Firm universe expands, but reconstructed aggregate volatility and the incremental group effect remain essentially unchanged, with differences again within the one-basis-point range (≤0.01 pp). Hence, the core findings are robust to reasonable alternative definitions of group membership.

7. Conclusion:

This dissertation shows that aggregate volatility in the Portuguese non-financial corporate sector (2008–2022) is driven primarily by cross-firm covariance rather than by a simple aggregation of firm-specific risk. The large gap between realized aggregate volatility and the independence benchmark indicates that co-movement is essential for explaining macro risk in firm-level data.

Portugal exhibits the main structural features associated with granular effects (high concentration in firm size and only weak volatility scaling), so that micro shocks in the upper tail may not wash out mechanically. In this setting, business groups are a natural candidate for a macro-relevant "Super-Firm" channel, since ownership networks concentrate activity and may transmit shocks across affiliates.

Within this environment, the thesis's central result is that ownership networks are economically meaningful yet only weakly macro-relevant. Consolidating affiliates into Super-Firms mechanically raises the independence benchmark through greater weight concentration. In addition, introducing a group factor increases Super-Firm idiosyncratic volatility, providing clear evidence that common ownership shapes risk within networks. However, the incremental effect on reconstructed aggregate volatility is small. These conclusions are robust to using either the raw sample covariance matrix or Ledoit–Wolf shrinkage for factor covariances, with differences mainly in levels rather than in the ranking of specifications.

In our implementation, the group factor primarily operates *within* groups, rather than creating an explicit between-group covariance channel. As a result, ownership networks matter for consolidated entities, yet they do not emerge as a primary driver of aggregate volatility at the economy level.

This gap between micro relevance and limited macro impact is explained by two reinforcing mechanisms. First, for economically large groups, the group factor largely overlaps with industry dynamics already captured by the baseline specification. Second, the group dimension is mechanically constrained by scale: group affiliation covers only a subset of firms, and the group universe is highly fragmented, implying a much thinner upper tail of weights than the industry or municipality dimensions.

Notably, the baseline vs. Super-Firm gap becomes slightly more visible in the last years of the sample under the raw covariance estimator, while remaining essentially negligible under Ledoit–Wolf shrinkage. Given the small magnitude of the divergence and the short post-2019 window, this evidence should be interpreted as suggestive rather than conclusive. A plausible interpretation is that the incremental relevance of ownership linkages may be time-varying and more detectable in periods where (i) the observed group registry is richer and less reliant on backfilled affiliations, and/or (ii) within-group comovement is stronger. At the same time, the estimator sensitivity underscores that this late-sample nuance is not robust enough to overturn the main conclusion that the ownership channel has a limited aggregate effect in this framework.

Future research could strengthen these conclusions by extending the time span or using higher-frequency data to increase statistical power and identify regime changes in exposures and dependence. It would also be valuable to model dependence between groups, using observable inter-group linkages (input–output relations, common bank exposure or trade-market overlap) that may generate systematic co-movement across large corporate networks. Another natural extension is to relax the single-group assignment and allow for multi-group membership. In the current framework, firms are mapped to at most one group to ensure stable "control" links, but some firms may plausibly be influenced by multiple owners or networks. A richer approach would treat group affiliation as a weighted exposure (e.g., based on equity stakes or voting rights), allowing firms to load on more than one group factor and capturing overlapping structures beyond strict consolidation. Furthermore, as explored in Chapter 6.2.4, employing a sales-weighted covariance matrix could help prevent underestimating comovement among the largest firms. Finally, explicitly estimating factor loadings and covariance structures for the COVID period, rather than relying on a pre-pandemic window with sector-time fixed effects, would allow a direct quantification of how crisis episodes reshape the covariance matrix and the relative importance of each channel, including ownership-based linkages.

A.1. Volatility measure across Sectors of Activity and Districts

Appendix A.1. - Sectoral and District Volatility Rankings: Top and Bottom Categories

Rank	Sector of Activity	Mean (Std. Deviation)	District	Mean (Std. Deviation)
1º	Real estate activities	0.561 (0.157)	Madeira	0.443 (0.143)
2º	Construction	0.554 (0.167)	Faro	0.432 (0.145)
3º	Arts, entertainment, recreational activities	0.475 (0.177)	Lisbon	0.429 (0.138)
4º	Information, communication activities	0.471 (0.135)	Setúbal	0.428 (0.134)
5º	Administrative and support service activities	0.452 (0.153)	Beja	0.412 (0.128)
…	…	…	…	…
18º	Manufacturing industries	0.372 (0.123)	Aveiro	0.377 (0.118)
19º	Education	0.371 (0.119)	Coimbra	0.375 (0.118)
20º	Wholesale and retail trade; repair of motor vehicles and motorcycles	0.352 (0.113)	Azores	0.371 (0.122)
21º	Transportation and storage	0.350 (0.121)	Guarda	0.361 (0.112)
22º	Human health and social support activities	0.307 (0.100)		

This table presents the average of median sales-growth volatility and the associated standard deviation (in parentheses) for the five highest and lowest-ranked sectors of activity and districts. Rankings are determined by the average volatility value.

A.2. Broken Chain of Transmission

To elucidate the sources of this low predictive power, we mapped the correlation structure connecting the Granular Residual, the Tradable and Non-Tradable sectors, and Real GDP. This analysis reveals a **"broken chain of transmission"** rooted in the structural dichotomy of a small open economy: while the granular giants are exposed to external cycles, domestic activity is driven by internal dynamics.[5]

First, the Granular Residual is strongly linked to the Tradable/Export Sector (correlation o 0.65). This is consistent with the idea that large firms are more exposed to external conditions rather than purely domestic idiosyncratic factors (as described by di Giovanni and Levchenko, 2012).

Crucially, however, the Tradable Sector's sales growth exhibits a very low correlation with Real GDP (0.15). This suggests a significant decoupling: the sales performance of exporters, driven by external demand, does not move in lockstep with the aggregate value added of the economy.

In contrast, the **Non-Tradable Sector** shows a significantly stronger correlation with GDP (0.34), indicating that aggregate fluctuations are more heavily influenced by domestic demand, services, and construction.

Consequently, Granular Residual fails to predict GDP not because it is 'noise,' but because it captures an 'External Cycle' diluted by the domestic economy's weight. As illustrated in Appendix A.2.1, which plots both series over the sample period, the granular signal and aggregate growth often diverge, though synchronization improves during the COVID-19 pandemic. To facilitate comparison, the granular residual is reported on a three-fold scale.

Beyond the economic decoupling, two other factors likely contribute to the weak statistical signal:

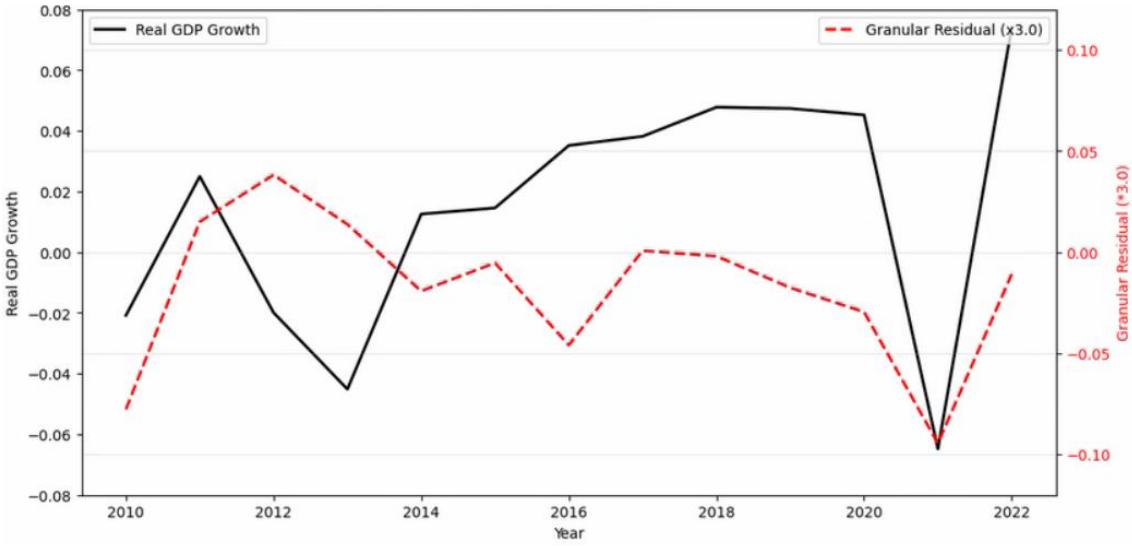
1. **Statistical Power:** The regression relies on a limited time series (14 years), reducing the power to reject the null hypothesis.
2. **Methodological "Over-cleaning":** In highly concentrated sectors (e.g., energy), standard de-meaning techniques may capture granular shocks as "sectoral" volatility, reducing the explanatory power of the residual.

This statistical disconnect highlights that the transmission mechanism is not direct but is instead mediated by complex interdependencies. Large firms are not isolated islands; their shocks spill over through sectoral supply chains, local labor markets, and ownership networks. Therefore, to uncover the true origins of fluctuations, we must move beyond simple residuals and employ

---

[5] The distinction between tradable and non-tradable sectors follows the methodology proposed by Amador and Soares (2012). Unlike the traditional classification based on Manufacturing versus Services, this approach relies on industry-specific export-to-sales ratios, providing a more accurate proxy for external exposure in the Portuguese economy.

the **Factor Model** detailed in the next section, which allows us to mathematically disentangle common shocks from true idiosyncratic risks and explicitly model these structural linkages.



Appendix A.2.1 – The Granular Residual vs. Real GDP Growth

A.3. "Safe Haven" Headquarters Counterfactual

We now investigate whether aggregate results are driven by the specific volatility environments in which the largest business groups are headquartered. We restrict attention to groups whose median weight exceeds 0.5%, yielding nine groups (ranging from 2 to 240 constituent firms). These groups display relatively balanced sectoral exposure, but a pronounced geographic pattern: four of the five largest groups are headquartered in municipalities with extremely low volatility (below the 4th percentile), effectively "safe havens," while a notable outlier is headquartered in a high-volatility municipality (82nd percentile).

To assess whether baseline volatility is merely a byproduct of these idiosyncratic environments, we conduct a counterfactual exercise in which we replace municipality- and industry-specific volatility terms (diagonal entries of the F matrix) with their median values. Under this combined counterfactual, aggregate volatility declines by roughly 0.2 percentage points. Although assigning median volatility to safe-haven locations mechanically increases local risk for those groups, this effect is more than offset by the compression of extreme sectoral (and location) volatility pockets elsewhere.

Extending the analysis to the top 10 municipalities by total weight (accounting for roughly 60% of sales) confirms that headquarters are concentrated in low-volatility environments (below the 15th percentile). Yet this geographic stability is insufficient to offset instability arising from sectoral exposure: several of the largest entities operate in sectors at very high volatility percentiles. Isolating the geographic channel reinforces this interpretation: fixing only municipality volatility increases aggregate volatility slightly, whereas aggregate volatility declines only when sectoral variance is also neutralized.

A.4. Measuring Synchronization through an Implied-Correlation Approach

Computing the full pairwise correlation matrix for the universe of firms ($N \approx 95{,}000$ per year) is computationally infeasible and further complicated by the strongly unbalanced panel. To characterize how dependence relates to firm size in a tractable way, we complement the main results with two diagnostics computed within firm-size bins: (i) an implied equicorrelation measure, and (ii) direct (sampled) pairwise correlations.

We adapt an "implied correlation" framework commonly used in finance[6] to the firm-level context, defining ρ_{impl}as the constant equicorrelation that rationalizes the observed equal-weighted variance, given the cross-sectional distribution of firm volatilities.

For a given size bin $b$with $N_b$firms, consider the equal-weighted aggregate growth rate within the bin and its realized variance $\sigma^2_{EW,b}$. Under an equicorrelation benchmark that allows heterogeneous volatilities σ_i, equal-weighted variance satisfies:

Allowing for heterogeneity in firm-level volatilities, equal-weighted aggregate variance can be written as:

$$\sigma^2_{EW,b} = \frac{1}{N_b^2} * \left[\sum_{i \in b} \sigma_i^2 + \rho * \sum_{i \neq j,\ i,j \in b} (\sigma\_i * \sigma\_j) \right],$$

where σ_i^2 denotes the time-series variance of firm i's growth rate. Solving for the constant equicorrelation that matches the realized $\sigma^2_{EW,b}$ yields:

$$\rho_{\text{impl,b}} \approx \frac{N_b * \sigma^2_{EW,b} - \bar{\sigma_b}^2}{N_b * \bar{\sigma}_b{}^2 - \bar{\sigma_b}^2},$$

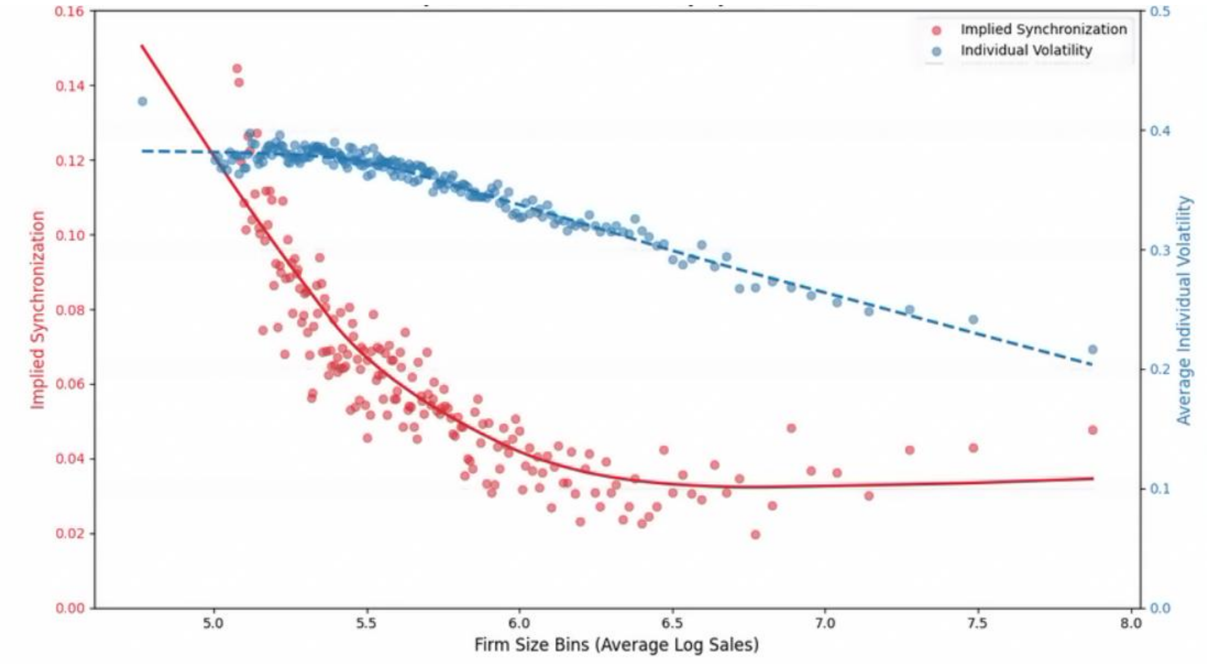
where $\bar{\sigma_b}^2 = \sum_{i \in b} \sigma_i^2$, $i \in b$, and $\bar{\sigma}_b = \frac{1}{N_b} \sum_{i \in b} \sigma_i$.

The key interpretation is that $\rho_{\text{impl},b}$ is not an estimate of the average pairwise correlation $\mathbb{E}[\rho_{ij}]$, but the effective equicorrelation required to match the observed equal-weighted variance. As a result, $\rho_{\text{impl},b}$can vary across size even when direct pairwise correlations do not display a clear size gradient.

Appendix A.4.1 reports $\rho_{\text{impl},b}$across 100 bins sorted by average sales. The implied equicorrelation is higher in the lower tail and declines with size. This pattern indicates that,

[6]See "Cboe® Implied Correlation® Index (COR3M)", for the original derivation of the equicorrelation parameter in the context of index options. Available at: https://cdn.cboe.com/resources/indices/documents/Implied_Correlation-WhitePaper-v1.0.5.pdf

under equal weighting, matching the observed volatility of the lower tail requires a larger effective dependence term than in the upper tail (peaking at 0.15).



Appendix A.4.1 – Synchronization and Volatility by Firm Size

Next, we estimate within-bin average correlations using random sampling. For each bin, we draw 5000 random firm pairs (i, j) and compute Pearson correlations of their sales-growth series over the pre-pandemic window, using only overlapping observations and imposing a minimum overlap requirement to avoid spurious estimates in sparse pairs. We then average these correlations within each bin.

This direct diagnostic does not show a clear monotonic size gradient in average pairwise correlations. Importantly, this is not a contradiction with the behavior of $\rho_{\mathrm{impl},b}$. A size gradient in $\rho_{\mathrm{impl},b}$ can arise even with broadly flat $\bar{\rho}_{ij}$, because $\rho_{\mathrm{impl},b}$ embeds how equal weighting interacts with (i) the level and dispersion of σ_i within the bin and (ii) the large number of cross-terms in the covariance component.

Taken together, these diagnostics support a careful interpretation of the EW > SW pattern. Equal weighting can raise reconstructed volatility without requiring systematically higher pairwise correlations among smaller firms. Instead, the result is consistent with the idea that the lower tail contributes disproportionately to the covariance term because the interaction of equal weights with the lower-tail volatility/exposure structure increases the effective importance of cross-firm covariance.

Finally, Appendix A.4.2 complements the dependence diagnostics by documenting the extreme concentration of the sales-weighted diagonal term $\sum_i w_i^2 \sigma_i^2$: the cumulative share rises rapidly in the upper tail, with the top 1,000 firms accounting for over 90% of the variance mass, so SW aggregation loads heavily on a relatively small set of large firms. This highlights why changing the weighting scheme can substantially shift the balance between diagonal and covariance contributions even if average correlations do not change materially with size.

Appendix A.4.2 – Cumulative Concentration of the Sales-Weighted Variance Component

This combination of results motivates a more structural interpretation of the EW>SW pattern. Because $\rho_{\text{impl},b}$ is pinned down by the total covariance term required to match the observed EW variance in each bin, a higher $\rho_{\text{impl},b}$ in the lower tail in the absence of a clear size gradient in sampled $\bar{\rho}_{ij}$ is consistent with smaller firms carrying a larger effective systematic component of volatility under equal weighting. One possibility is that smaller firms load more strongly on common shocks (higher effective factor exposure) or are more exposed to particularly volatile local/industry components, so that the $\sigma_i\sigma_j$ cross-terms contribute disproportionately even if average pairwise correlations are not higher. As an exploratory check, we re-estimate the baseline factor regression by size bins and find no stable monotonic pattern in estimated betas, suggesting that the mechanism may operate more through the volatility of the shocks faced by smaller firms.

Overall, EW>SW should therefore be interpreted as evidence that the covariance contribution becomes more important under equal weighting, but it does not by itself establish that smaller firms are more correlated in the usual pairwise sense. Future work could sharpen this distinction by allowing exposures and dependence to vary jointly with firm size and with the state of the economy (e.g., crisis vs. normal times), and by modeling nonlinear/tail dependence rather than relying only on average second moments.

Bibliography:

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77-91

Comin, D., & Mulani, S. (2006). Diverging Trends in Aggregate and Firm Volatility. *The Review of Economics and Statistics*, 88(2), 374-383.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, 77(3), 623–685.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3), 733-772

di Giovanni, J., Levchenko, A. A., & Mejean, I. (2014). Firms, destinations, and aggregate fluctuations. *Econometrica*, *82*(4), 1303-1340.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365-411.

Jovanovic, B. (1987). Micro shocks and aggregate risk. *The Quarterly Journal of Economics*, *102*(2), 395–409.

Long, J. B., & Plosser, C. I. (1983). Real business cycles. *Journal of Political Economy*, *91*(1), 39–69.

Horvath, M. (1998). Cyclicality and sectoral linkages: Aggregate fluctuations from independent sectoral shocks. *Review of Economic Dynamics*, *1*(4), 781–808.

Horvath, M. (2000). Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics*, *45*(1), 69–106.

Dupor, B. (1999). Aggregation and irrelevance in multi-sector models. *Journal of Monetary Economics*, *43*(2), 391–409.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, *293*(5536), 1818–1820.

di Giovanni, J., & Levchenko, A. A. (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy*, *120*(6), 1083–1132.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, *80*(5), 1977–2016.

Carvalho, V. M., & Gabaix, X. (2013). The great moderation and the US. *American Economic Review*, *103*(5), 1981–2027.

Conley, T. G., & Dupor, B. (2003). A Spatial Analysis of Sectoral Complementarity. *Journal of Political Economy*, 111(2), 311-352

Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance*, *56*(1), 1–43.

Hilary, G., & Hui, K. W. (2009). Does religion matter in corporate decision making in America? *Journal of Financial Economics*, *93*(3), 455–473.

Mascia, D. V., & Onali, E. (2021). Disentangling the sources of bank opacity. *Journal of Economic Behavior & Organization*, *184*, 550–580.

Comin, D., & Mulani, S. (2006). Diverging trends in aggregate and firm volatility. *The Review of Economics and Statistics*, *88*(2), 374–383.

Khanna, T., & Yafeh, Y. (2007). Business groups in emerging markets: Paragons or parasites? *Journal of Economic Literature*, *45*(2), 331–372.

Lewellen, W. G. (1971). A pure financial rationale for the conglomerate merger. *The Journal of Finance*, *26*(2), 521–537.

Santioni, R., Schiantarelli, F., & Strahan, P. E. (2017). *Internal capital markets in times of crisis: The benefit of group affiliation in Italy* (Working Paper No. 23234). National Bureau of Economic Research.

Matvos, G., & Seru, A. (2014). Resource allocation within firms and financial market dislocation: Evidence from diversified conglomerates. *The Review of Financial Studies*, *27*(4), 1143–1189.

Byun, H.-Y., Choi, S., Hwang, L.-S., & Kim, R. G. (2013). Business group affiliation, ownership structure, and the cost of debt. *Journal of Corporate Finance*, *23*, 311–331.

Faccio, M., Morck, R., & Yavuz, M. D. (2019). *Business groups and the incorporation of firm-specific shocks into stock prices* (Working Paper No. 25908). National Bureau of Economic Research.

Masulis, R. W., Pham, P. K., Zein, J., & Ang, A. E. S. (2022). Crises as opportunities for growth: The strategic value of business group affiliation. *Journal of Financial and Quantitative Analysis*, *58*(4), 1508–1546.

Naaraayanan, S. L., & Wolfenzon, D. (2023). *Business group spillovers* (Working Paper No. 31107). National Bureau of Economic Research.

Morck, R., Wolfenzon, D., & Yeung, B. (2005). Corporate governance, economic entrenchment, and growth. *Journal of Economic Literature*, *43*(3), 655–720.

Hong, H. A., Kim, J.-B., & Welker, M. (2017). Divergence of cash flow and voting rights, opacity, and stock price crash risk: Evidence from a quasi-natural experiment. *Review of Accounting Studies*, *22*(3), 1174–1205

Desai, C. A., & Savickas, R. (2010). On the causes of volatility effects of conglomerate breakups. *Journal of Corporate Finance*, *16*(4), 554–571.

Khanna, T., & Yafeh, Y. (2005). Business groups and risk sharing around the world. *The Journal of Business*, *78*(1), 301–340.

Silva, N. (2025). *On the measurement and forecasting of sales volatility: Is the quantile approach better?* [Working Paper]. [Banco de Portugal]

Amador, J., & Soares, A. C. (2018). Competition in the Portuguese economy: insights from a profit elasticity approach. *Empirica*, 45(2), 339-365.

Cabral, S., Gouveia, C. M., & Manteu, C. (2020). The granularity of Portuguese firm-level exports. *Banco de Portugal, Economic Studies*.

Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. Journal of Financial Economics, 63(3), 443–494.