UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO SÓCIO-ECONÔMICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Leon Cardoso Esquierro

Grain by Grain: countries, cities, firms

Florianópolis

2024

Leon Cardoso Esquierro

Grain by Grain: countries, cities, firms

Tese submetida ao Programa de Pós-Graduação em Economia da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Doutor em Economia

Orientador: Prof. Dr. Eraldo Sergio Barbosa Da Silva

Florianópolis

2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Esquierro, Leon Cardoso
Grain by Grain : Countries, Cities, Firms / Leon Cardoso Esquierro ; orientador, Eraldo Sérgio Barbosa da Silva, 2024.
96 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Socioeconômico, Programa de Pós-Graduação em Economia, Florianópolis, 2024.

Inclui referências.

1. Economia. 2. Econofísica. 3. Granularidade. I. da Silva, Eraldo Sérgio Barbosa. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Economia. III. Título.

Leon Cardoso Esquierro

Grain by Grain: countries, cities, firms

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 15 de fevereiro de 2024 pela banca examinadora composta pelos seguintes membros:

Prof. Dr. Raul Yukihiro Matsushita
Universidade de Brasília

Prof. Dr. Guilherme De Oliveira
Universidade Federal de Santa Catarina

Prof. Dra. Eva Yamila Da Silva Catela
Universidade Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Economia.

ICP-Edu
Documento assinado digitalmente
Francis Carlo Petterini Lourenco
Data: 13/03/2024 20:08:37-0300
CPF: ***.714.300-**
Verifique as assinaturas em https://v.ufsc.br

Coordenação do Programa de Pós-Graduação: Prof. Dr. Francis Petterini

GOV.BR
Documento assinado digitalmente
ERALDO SERGIO BARBOSA DA SILVA
Data: 10/03/2024 11:32:29-0300
CPF: ***.930.814-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Eraldo Sergio Barbosa Da Silva
Orientador

Florianópolis, 2024.

AGRADECIMENTOS

Agradeço aos meus pais (Lilia e José Carlos) e à minha companheira de estudos Loren.

Agradeço aos colegas, funcionários e professores do programa. Em especial, agradeço ao professor Sérgio por ter me orientado nesta tese e, mais do que isso, pela paciência e atenção que teve comigo.

Agradeço à Bárbara pelo apoio e companheirismo ao longo da elaboração desta tese.

Agradeço ao contribuinte fiscal brasileiro que, por meio do CAPES, me concedeu auxílio financeiro ao longo da jornada do doutorado.

RESUMO

O elo que une os três ensaios desta tese é temático e metodológico. Os ensaios têm como tema principal a granularidade e seguem o mesmo itinerário metodológico: primeiro, testamos, utilizando o método de Gabaix e Ibragimov (2011), a hipótese de que os dados analisados estão distribuídos segundo uma lei de potência; segundo, testamos a “hipótese granular” de Gabaix (2011); terceiro, calcularemos o “tamanho granular” (em número de agentes em relação ao fenômeno analisado) utilizando o mecanismo de Blanco-Arroyo et al. (2018). No primeiro ensaio, testamos a hipótese de granularidade para os spillovers da inflação internacional, utilizando dados anuais de exportações e de inflação para 138 países, de 1991 a 2020. Descobrimos que os volumes de exportação entre países não são distribuídos de forma gaussiana, mas seguem uma lei de potência. Constatamos também que os países com maior peso relativo no comércio internacional determinam uma parcela dos spillovers internacionais superior à sua participação no comércio internacional. Além disso, oito grandes grãos são responsáveis pela maior parte dos spillovers de inflação. No segundo ensaio estendemos o conceito de granularidade das empresas para as cidades, que se refere à coexistência de algumas cidades grandes e numerosas cidades pequenas. Argumentamos que a granularidade e as leis de potência estão inter-relacionadas e levantamos a hipótese de que as grandes cidades desempenham um papel significativo no ciclo econômico além da sua dimensão relativa. Nosso estudo sobre dados de cidades americanas e brasileiras de 2003 a 2019 não descartam a hipótese granular. Descobrimos que o tamanho granular da cidade nos Estados Unidos é de três áreas metropolitanas. Se redefinirmos as cidades como condados, o tamanho granular é de cinco condados. No Brasil, o tamanho granular equivale a três municípios. No terceiro ensaio, exploramos o impacto das grandes empresas nos ciclos de contratação e demissão no mercado de trabalho brasileiro. As principais conclusões incluem a observação de que as empresas apresentam uma distribuição de lei de potência para a dimensão da sua força de trabalho, com os choques idiossincráticos das grandes empresas influenciando significativamente os ciclos de contratação e demissões. Em particular, o setor de serviços desempenha um papel substancial na explicação destes ciclos, enquanto a indústria de transformação tem um poder explicativo limitado. Determinamos que o tamanho granular do mercado de trabalho brasileiro é de 15 empresas envolvidas em serviços públicos, e que as empresas privadas têm um impacto relativamente pequeno nos ciclos de contratação e demissão. Cada ensaio demonstra a importância da granularidade em diferentes esferas econômicas, desde o comércio internacional e a dinâmica urbana até aos padrões de emprego, mostrando a importância das grandes entidades na definição dos resultados económicos.

Palavras-chave: Granularidade; Ciclos Econômicos.

ABSTRACT

This thesis comprises three essays linked by the theme of granularity and a consistent methodology. The essays explore the distribution and impact of granular structures in different contexts using specific analytical methods. The key steps across the essays include testing for power law distribution with the Gabaix and Ibragimov (2011) method, examining Gabaix's (2011) granular hypothesis, and calculating granular size based on the approach by Blanco-Arroyo et al. (2018). The first essay investigates the granularity of international inflation spillovers across 138 countries from 1991 to 2020, using export and inflation data. It finds that exports follow a power law rather than a Gaussian distribution, with major traders disproportionately influencing inflation spillovers. Eight major "grains" (countries) are primarily responsible for these spillovers. The second essay extends the concept of granularity to cities, exploring the existence of a few large cities amidst many small ones and their disproportionate economic influence. Analyzing data from American and Brazilian cities (2003-2019), it cannot reject the granular hypothesis. In the U.S., the granular size is three metropolitan areas or five counties. In Brazil, it is three municipalities. The third essay examines the role of large companies in Brazil's hiring and firing cycles, finding that workforce size distributions follow a power law. Large companies, particularly in the services sector, significantly impact employment cycles, with 15 public service companies constituting the Brazilian labor market's granular size. Private companies have a minor effect. Each essay demonstrates the significance of granularity in different economic spheres, from international trade and urban dynamics to employment patterns, showcasing the importance of large entities in shaping economic outcomes.

Keywords: Granularity; Business Cycles.

RESUMO EXPANDIDO

O elo que une os três ensaios desta tese é temático e metodológico. Os ensaios têm como tema principal a granularidade e seguem o mesmo itinerário metodológico: primeiro, testamos, utilizando o método de Gabaix e Ibragimov (2011), a hipótese de que os dados analisados estão distribuídos segundo uma lei de potência; segundo, testamos a "hipótese granular" de Gabaix (2011); terceiro, calcularemos o "tamanho granular" (em número de agentes em relação ao fenômeno analisado) utilizando o mecanismo de Blanco-Arroyo et al. (2018). No primeiro ensaio, testamos a hipótese de granularidade para os spillovers da inflação internacional, utilizando dados anuais de exportações e de inflação para 138 países, de 1991 a 2020. Descobrimos que os volumes de exportação entre países não são distribuídos de forma gaussiana, mas seguem uma lei de potência. Constatamos também que os países com maior peso relativo no comércio internacional determinam uma parcela dos spillovers internacionais superior à sua participação no comércio internacional. Além disso, oito grandes grãos são responsáveis pela maior parte dos spillovers de inflação. Adicionalmente, neste ensaio testamos a possibilidade de que alguns grupos de produtos (alimentos, combustíveis e manufaturados) seriam maiores condutores dos vazamentos de inflação. Também avaliamos com destaque a importância da Rússia nos vazamentos de inflação global, assunto em destaque por causa da Guerra da Ucrânia. De modo geral, os resultados sugerem que os alimentos explicam uma parte maior dos "spillovers". No segundo ensaio estendemos o conceito de granularidade das empresas para as cidades, que se refere à coexistência de algumas cidades grandes e numerosas cidades pequenas. Argumentamos que a granularidade e as leis de potência estão inter-relacionadas e levantamos a hipótese de que as grandes cidades desempenham um papel significativo no ciclo econômico além da sua dimensão relativa. Nosso estudo sobre dados de cidades americanas e brasileiras de 2003 a 2019 não descarta a hipótese granular. Descobrimos que o tamanho granular nos EstadosUnidos é de três áreas metropolitanas. Se redefinirmos as cidades como condados, o tamanho granular é de cinco condados. No Brasil, o tamanho granular equivale a três municípios. Para o Brasil, também constatamos que as regiões metropolitanas parecem explicar uma fração pequena dos ciclos econômicos de modo que rejeitamos a hipótese granular neste contexto. Além da estimação por Mínimos Quadrados Ordinários (MQO), também calculamos os expoentes de Pareto por Máxima Verossimilhança (MV), nossa constatação foi que as estimativas de MV são uma fração estável das de MQO, de modo que, exceto para fins de previsão, as estimativas parecem ser equivalentes. Noterceiro ensaio, exploramos o impacto das grandes empresas nos ciclos de contratação e demissão no mercado de trabalho brasileiro. As principais conclusões incluem a observaçãode que as empresas apresentam uma distribuição de lei de potência para a dimensão da sua força de trabalho, com os choques idiossincráticos das grandes empresas influenciandosignificativamente os ciclos de contratação e demissões. Em particular, o setor de serviçosdesempenha um papel substancial na explicação destes ciclos, enquanto a indústria detransformação tem um poder explicativo limitado. Determinamos que o tamanho granular do mercado de trabalho brasileiro é de 15 empresas envolvidas em serviços públicos, e que asempresas privadas têm um impacto relativamente pequeno nos ciclos de contratação edemissão. Cada ensaio demonstra a importância da granularidade em diferentes esferaseconômicas, desde o comércio internacional e a dinâmica urbana até aos padrões de emprego, mostrando a importância das grandes entidades na definição dos resultados económicos.

Palavras-chave: Granularidade; Ciclos Econômicos.

Summary

1. Introduction, research problem, and objectives

The three essays in this thesis are united by their thematic and methodological focus on granularity. They follow a similar approach, starting with testing the hypothesis that the data conforms to a power law using the method by Gabaix and Ibragimov (2011). Then, they examine Gabaix's (2011) granular hypothesis, followed by calculating the "granular size" using the mechanism defined by Blanco-Arroyo et al. (2018).

The overall objective of these essays is to test the granular hypothesis in non-traditional contexts. The first essay explores granularity in inflation spillovers, while the second essay investigates the role of large cities in their country's economic cycles. Finally, the third essay evaluates the impact of large companies on labor market dynamics, specifically with regards to hiring and firing.

Common to all these cases is the underlying phenomenon and research problem: the presence of exceptionally large agents that seem to have an outsized influence on the overall fluctuations, surpassing what their own size would suggest. This occurs because there are often a small number of very large agents and a large number of very small agents. The compensatory effect of the small agents' idiosyncratic shocks, as predicted by the Law of Large Numbers, allows the variations of the few large agents to explain a significant portion of the aggregate oscillation as they have a greater likelihood of being "uncompensated."

Previous research on granularity has primarily focused on the size of companies and their impact on the economic cycle. Key contributions include Gabaix's (2011) seminal article testing the granular hypothesis for large American companies. Similar findings have been observed by Silva and Da Silva (2020) for Brazil, Ebeke and Eklou (2017) for Europe, Blanco-Arroyo and Alfarano (2017) for Spain, Fornaro and Luomaranta (2018) for Finland, and Miranda-Pinto and Shen (2019) for Australia.

The research agenda on granularity has also expanded to cover new topics, with companies remaining the central actors. Kovalenko et al. (2021) examined granularity in the German labor market, while Dosi et al. (2019) explored the granularity of company investments (referred to as the granularity of demand shocks). Additionally, Giovanni et al. (2020) analyzed the granularity of external shocks in France. However, in all three cases, the focus remains on the behavior of exceptionally large firms and the influence of their size on the market or specific contexts such as labor markets, investments, and external shocks.

In this thesis, the focus extends beyond exceptionally large agents being solely companies (except for the third essay). The first essay examines the granularity in inflation spillovers using countries as the grains, while the second essay analyzes the fluctuations in economic cycles within specific locations, such as cities, in Brazil and the United States. Finally, the third essay investigates how large companies influence hiring and firing shocks in the Brazilian economy.

By testing the specific hypotheses in each essay, the aim is to broaden the understanding of the "granular hypothesis" and its applicability as a tool for studying economic phenomena in various contexts.

The first essay constructs a granular residual by considering each country's role in international trade as a relative weight. The hypothesis is that large countries export inflation through trade, and therefore, the variations in global average inflation can be explained by the idiosyncratic shocks of these prominent players.

In the second essay, the focus shifts to analyzing the geographical locations within countries where economic cycles occur. The hypothesis is that idiosyncratic shocks occurring in large human agglomerations, like counties, metropolitan regions, and municipalities, can explain a significant portion of the respective country's economic cycles.

The third essay investigates the influence of large companies on hiring and firing shocks in the Brazilian economy. The policy implication of this research suggests that addressing periods of high unemployment in Brazil may be more effectively achieved through investing in public services rather than fiscal stimulus for manufacturing.

Gabaix (2016) defines power laws as mathematical relationships of the form $Y = aX^b$, where X and Y are variables, and "a" and "b" are parameters. Power laws can arise in various domains, such as the size of cities, metabolic rates of animals, and daily stock return rates.

Power laws in economics have been observed in income distribution, as discovered by Pareto (1896), and in word frequency distribution, as identified by Zipf (1949). These power laws can arise through random growth models or through optimization and matching, where small differences in talent or skill lead to significant differences in outcomes.

Granularity, as discussed by Gabaix (2016), is an application of power laws that provides valuable insights into fluctuations in GDP and stock markets. It highlights the notion that economic cycles are often driven not by small diffuse shocks impacting all firms.

2. Essay 1: Granular inflation spillovers

Abstract
We test the granularity hypothesis to international inflation spillovers using annual exports and inflation data for 138 countries from 1991 to 2020. We find export volumes across countries are not Gaussian-distributed but follow a power law. This finding means the largest countries disproportionately impact world inflation. First, we quantify the power law for the right tail of the export volume distribution and discuss its implications. Then, we compute the granular residual, a measure of shocks to the largest countries. We find that countries with higher relative weight in international trade determine a portion of international spillovers greater than their trade share. Moreover, eight big grains are responsible for the bulk of inflation spillovers. The policy implication is that other countries' central banks should closely monitor the eight big grains when conducting their domestic monetary policy.

2.1. Introduction

Business cycle shocks can transmit through inflation spillovers, impacting various countries. For instance, inflation in the Eurozone affects US inflation expectations, making term-structure-based models ineffective in predicting long-term inflation (Ciccareli and Garcia, 2015). Istiak et al. (2021) find that Japan and the United States are the leading "exporters" of inflation among the G7 countries, while Baurle et al. (2021) highlight heterogeneous inflation spillovers for Switzerland. Balcilar et al. (2020) emphasize the importance of inflation spillovers to commodity prices in Nigeria. Additionally, Halka and Szafranek (2016) demonstrate that small European economies, such as Poland, are "net importers" of inflation from the Eurozone. These inflation spillovers indicate the need for central banks to monitor foreign inflation to inform domestic monetary policy.

To better understand international inflation spillovers, we propose considering "granularity." According to the granular hypothesis of Gabaix (2011), countries with significant international trade linkages play a substantial role in determining inflation spillovers beyond their trade participation. These countries, referred to as "big grains," often comprise large firms driving most of a country's exports (Freund and Pierola, 2015). The big grains, representing a significant share of economic activity, maintain direct trade connections with foreign countries (Di Giovanni et al., 2017, 2018). Consequently, by analyzing inflation spillovers, we can assume that if there is granularity in goods exports (Di Giovanni Levchenko, 2012), there is also granularity in inflation spillovers.

Our hypothesis suggests that the big grains have a greater influence on inflation spillovers than their relative trade participation. As these large countries possess market power and dominate production, they can determine prices in international trade, leading to their domestic inflation affecting smaller countries with limited capacity to influence global prices. The literature supports this notion, highlighting how shocks at the industry level can propagate globally through the input-output network, ultimately impacting inflation (Auer et al., 2019). Policymakers should consider the impact of important export items on inflation spillovers (Chandrarin et al., 2022). Importantly, we assume that the big grains' inflation is exogenous compared to other countries, meaning they influence global inflation but are not influenced by it.

Different transmission channels play a crucial role in the spillover of inflation, and the type of shocks that drive inflation in other countries affects how these spillovers occur (Baurle et al., 2021). The literature shows that there is significant variation in the extent to which international price fluctuations affect domestic inflation, depending on the source of the fluctuations.

In many models of international trade, heterogeneous firms are often represented as varying points on a continuum. However, studies have shown that when using whole numbers to represent firms in these models, random shocks at the individual firm level can impact aggregate variables (Eaton et al., 2012). As large firms play a significant role in international trade, it is worthwhile to consider sectors with a limited number of firms and examine comparative advantage across sectors with a more detailed analysis. Research suggests that shocks at both the aggregate and individual firm levels can affect comparative advantage, with granularity accounting for approximately 20% of the variation in export intensity across sectors (Gaubert and Itskhoki, 2021).

Furthermore, the standard "gravity equations" used to explain trade flows between countries consider factors like GDP, geographical distance, and tariffs. However, these equations often overlook the empirical reality that a small number of large firms dominate world trade. Neglecting granularity can lead to significant inaccuracies in estimating trade patterns (Breinlich et al., 2020).

To test the idea of granularity in inflation spillovers, we first examine whether export volumes among countries follow a power law, similar to the distribution of firm sizes (Da Silva et al., 2018). This serves as the empirical foundation for the granular hypothesis. Next, we calculate a measure called the "country granular residual" based on Gabaix's concept of granular residual at the firm level. This measure captures shocks experienced by the largest countries. Finally, we use the methodology developed by Blanco-Arroyo et al. (2018) to identify the optimal number of countries that contribute to inflation spillovers, which we refer to as the "country granular number."

2.2. Materials and methods

Data

We used annual export volumes and the annual rate of change of the consumer price index from the World Bank's World Development Indicators database from 1991 to 2020. Exports are expressed in 2015 US dollars. We excluded those countries with missing export data for more than 50% of the years in the sample. For those remaining countries with missing data, we filled the missing slots with the average from antecedent and subsequent data points. Moreover, we dropped from the sample the countries with no inflation data. We ended up with data for 138 countries. The dataset used is available at Figshare.

Power law in export volumes distribution

The granular hypothesis challenges the notion that shocks to individual firms diversify away by the law of large numbers and thus do not affect the business cycle (Gabaix, 2011). The economy is granular, not smooth. Empirical support to granularity comes from heavy-tailed firm size distributions because this means a few disproportionately large firms, implying that firm-level shocks do not cancel out. Similarly, we evaluate whether export volumes are normally distributed. If not, we compute the power law for the right tail of the export volumes distribution.

Inflation spillovers do not matter if export volumes are normally distributed. Positive inflationary shocks in some countries would be offset by negative shocks in others. By contrast, if export volumes follow a power-law distribution, the largest countries impact more world inflation.

We assess Gaussianity using the Kolmogorov-Smirnov test for 2019, considering n = 138 countries. In addition, to get an OLS estimation of the tail exponent, we use the rank – ½ method of Gabaix and Ibragimov (2011). Thus, we take

$$\ln(\text{rank}_i - \tfrac{1}{2}) = a - \alpha \ln \frac{\text{exports}_i}{\text{exports}_m}, \quad (1)$$

where exports_i is the sum of export volumes of the country i from 1991 to 2020, rank_i ranks exports_i from highest to lowest, and exports_m is the lowest bound we used as a cutoff to analyze the tail. Because few distributions follow a power law over their entire range, a power law occurs between a minimum cutoff and a maximum threshold. This is why we say a distribution has a power-law tail (Newman, 2005). The α in equation (1) is the tail index (Pareto exponent) tracking the heaviness of the right tail, with smaller values pointing to heavier tails (Jenkins, 2017).

A critical feature of such heavy-tailed distributed data is that aggregate fluctuations are not proportional to $1/\sqrt{\ }$, as expected if the data were Gaussian-distributed (Gabaix, 2011). Here, we relate this fact to foreign shocks canceling out. However, because foreign shocks are granular fluctuations, they are proportional to $1/\ln n$ (Gabaix, 2011). As a result, shocks to the largest countries are not offset, and international inflation spillovers are significant.

The country granular residual

At the firm level, the granular residual measures the shocks to the largest companies. One critical implication is that if we regress a country's growth rate on the granular residual of its largest companies, the adjusted R^2 will be higher than the percentage of participation of those companies in GDP[1]. This suggests productivity shocks to large companies explain a significant portion of the business cycle (Gabaix, 2011).

Similar to Gabaix's firm granular residual, we define the granular residual for *K* countries as

$$\Gamma_t = \sum_{i=1}^{K} \frac{\text{exports}_{i,t}}{\text{world exports}_t}(\pi_{i,t} - \pi_{w,t}), \tag{2}$$

where $\pi_{i,t}$ is the country *i*'s inflation rate in period *t*, and $\pi_{w,t}$ is the average world inflation in *t*. For the sake of simplicity, we do not consider the effects of exchange rate and interest rates on inflation spillovers; that is, we assume the only mechanism by which inflation is spilled from one country to others is through the international flow of goods. However, we appreciate the effects of relaxing this assumption after finishing our analysis.

In line with the main implication of the firm granular residual, we hypothesize that the countries with higher relative weight in international trade are responsible for a portion of inflation spillovers larger than their participation in international trade.

[1] See appendix (section 7 of the thesis) for insights about de granular residual.

Following the OLS methodology in Gabaix (2011), we regress $\pi_{w,t}$ on Γ_t and compare the resulting R^2 with a country's trade share.

Table A1 in the Appendix shows the ranking of the top 50 countries according to their share in international trade averaged between 1991 and 2020. These figures provide the coefficient values used as weights to compute the granular residual in equation (2). Listing the top 50 countries gives us a leeway to calculate the granular number next.

The country granular number

A country's granular residual contribution to inflation spillovers can be under- or overestimated if we do not calibrate using the optimal number of countries K^* in equation (2).

Following Blanco-Arroyo et al. (2018), we evaluate the explanatory power of the country granular residual by comparing a curve with weights (equation (2)) with another of equal weights after making $\text{exports}_{i,t} / \text{world exports}_t = 1$ in equation (2). The "granular curve" of function $C(L)$ is

$$C(L) = \frac{1}{Q}\sum_{K=1}^{Q} \mathrm{R}^2(K,L) ,$$

(3)

where Q is an arbitrary number of countries. The idea is to examine the sensitivity of R^2 to a sequential exclusion of the largest countries, which means increasing L. So, we consider the explanatory power curve $\mathrm{R}^2(K,L)$ as a function of an increasing number of countries K and for different values of the largest countries L. We want to see how the curve $\mathrm{R}^2(K,L)$ performs depending on the number L of highest-ranked countries that are removed from the sample and replaced by the $Q + 1, \ldots, Q + L$ following countries. Moreover, we run the same number Q of regressions with the granular residual Γ_t as the explanatory variable for each L. And $C(L)$ refers to the average R^2 in every L for Q regressions performed.

In turn, the equal-weight curve quantifies the shocks' contribution to inflation spillovers from equal-size countries, which are expected to be negligible for that matter.

Finally, we aim to see the transition from the "granular curve" $C(L)$ to the equal-weight curve as we progressively remove the L largest countries from the granular residual. The country granular number K^* corresponds to the L where the curve $C(L)$ first intersects the equal-weight curve.

A caveat is needed here. To avoid running more than 1,000 regressions, we simplified computation by considering $Q = 40$. Then, we ran the regressions for the L countries in fives, $L = 5$, $L = 10$..., until the curves with and without weights intersect. We also ran the regressions for K in fives. Finally, whenever we find a value for which $C(L)$ is lower than the R^2 for equal weights, we consider the intermediate values of L to find the granular number. This procedure will be clarified next.

Finally, the underlying empirical model is

$$\pi_t = \beta_1 + \beta_2 \Gamma_t + \varepsilon_t, \quad (4)$$

where π_t is the GDP-weighted averaged world inflation at time t, Γ_t is the granular residual, ε_t is the estimated error, β_1 is the mean inflation rate, and β_2 captures the sensitivity of world inflation to the granular residual. The parameters β_1 and β_2 are estimated using ordinary least squares in accordance with Gabaix (2011), where the adjusted R^2 estimate quantifies the explanatory power of the granular residual.

2.3 Results

We can reject the null hypothesis that export volumes are normally distributed because the calculated Komolgorov-Smirnov coefficient (0.4330) is greater than its tabulated values for the 1, 5, and 10% significance levels: 0.1387, 0.1157, and 0.1038, respectively. Because $n = 138 > 40$ we employ the values estimated by $1.63/\sqrt{138}$ for 1%, $1.36/\sqrt{138}$ for 5%, and $1.22/\sqrt{138}$ for 10%.

After dismissing Gaussianity, we quantify the tail exponents for different cutoffs using equation (1). The high R^2 values in Table 1 suggest we cannot ignore the power-law distribution. (Throughout this essay, R^2 values refer to adjusted R^2.) The smaller the cutoffs, the higher the R^2 values. Moreover, as the cutoffs decrease, the right tail becomes heavier (α becomes smaller). One can go further and calculate the optimal cutoff (as in Clauset et al., 2009), but high R^2 values suffice to settle the issue.

Table 1: Estimation of Equation (1)

Cutoff, %	*n*	*a*	*S.E.*	*α*	*S.E.*	*R2*
10	14	2.68	0.0957	-1.66	0.1185	0.938
20	28	3.59	0.0699	-1.39	0.0628	0.948
30	42	4.10	0.0841	-1.06	0.0546	0.901
40	56	4.35	0.0778	-0.87	0.0424	0.886
50	70	4.73	0.0784	-0.76	0.0342	0.876
60	84	4.80	0.0686	-0.68	0.0283	0.873
70	98	4.97	0.0659	-0.61	0.0241	0.876
80	112	5.19	0.0663	-0.55	0.0211	0.859
90	126	5.40	0.0686	-0.50	0.0188	0.847
100	138	6.15	0.1064	-0.41	0.0182	0.786

Note: The a and α estimates are all significant at 1%.

Furthermore, a back-of-the-envelope calculation allows us to appreciate a critical property of the power-law distribution mentioned earlier. Comparing shocks proportional to $1/\ln n$ (power law) and to $1/\sqrt{n}$ (Gaussian), we can see that inflation spillovers loom larger for the power-law case, regardless of cutoffs. Indeed, for the 10% cutoff, $1/\ln 14$ = 0.38, and $1/\sqrt{14}$ = 0.20; thus, the spillovers are 1.4 times larger for the power-law case. The same holds for the entire sample (100% cutoff). In particular, $1/\ln 138$ = 0.20, and $1/\sqrt{138}$ = 0.085, and thus the spillovers are 2.4 times larger.

Table 2 shows the granular residuals calculated after considering the K top-ranked countries from Table A1. We see we cannot reject the granular hypothesis. For example, if K = 5, R^2 is 0.9755 in the fourth column, which means the top five countries are responsible for 97% of inflation spillovers, even though these countries' relative share in the total exports of goods is only 34%.

Table 2: Estimation of Equation (2)

	Considering $\frac{\text{exports}_{i,t}}{\text{world exports}_t}$			Making $\frac{\text{exports}_{i,t}}{\text{world exports}_t} = 1$			*Share in international trade, %*
K	$\bar{\Gamma}_t$	Intercept	R^2	$\bar{\Gamma}_t$	Intercept	R^2	
5	-3.1675	*-0.2088*	0.9755	-0.3309	*1.0249*	0.7205	34.42
10	-3.0538	1.0333	0.8579	-0.1617	*0.7209*	0.8167	50.04
15	-2.6104	0.8886	0.8673	-0.1109	*0.5514*	0.8620	60.38
20	*0.1783*	4.5188	0.0027	0.0056	4.3293	0.0893	68.29
25	0.1530	4.1904	0.2325	0.0022	4.0209	0.3038	73.93
30	0.1520	4.2016	0.2249	0.0022	4.0337	0.2944	78.69
35	0.0830	4.1608	0.1948	0.0008	4.0895	0.1994	81.73
40	0.0832	4.1584	0.1953	0.0008	4.0950	0.1982	84.02

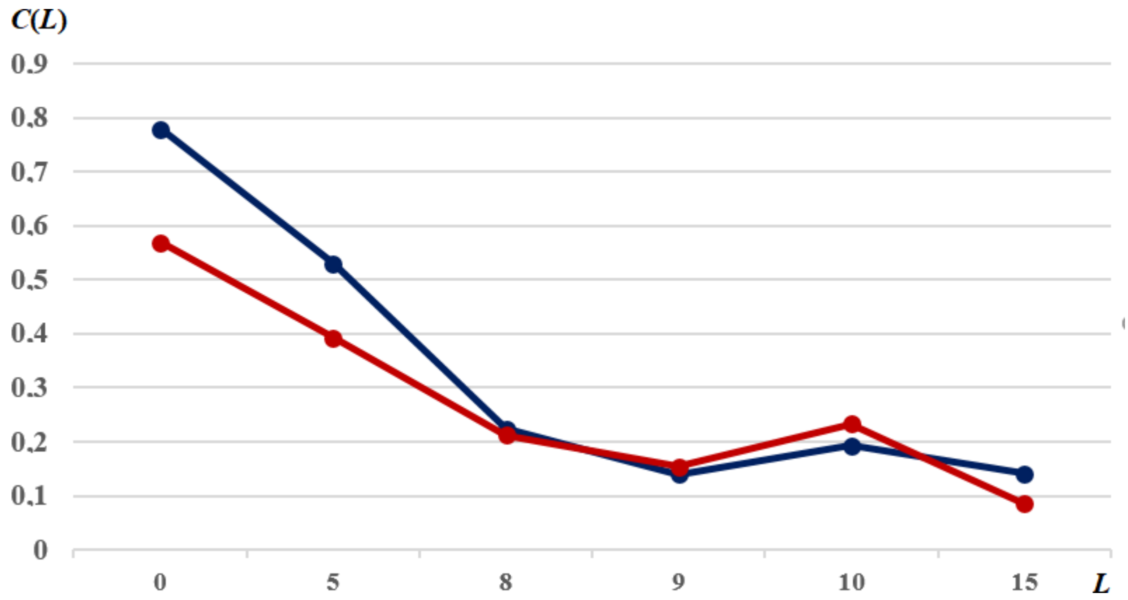
Note: the values in italics are nonsignificant at 1%.

Roughly, both the intercept and the estimated granular residual tend to grow as K increases in the second and third columns of Table 2. We explain this trend by the fact that the countries with more significant participation in international trade also present inflation rates lower than the global average, leading to a negative granular residual. This picture reverts as we add more countries (as we increase K).

The fifth to seventh columns show the results of the regressions after making $\text{exports}_{i,t} / \text{world exports}_t = 1$ in equation (2). We see the values of R^2 are reduced for K = 5, for example, which means the big grains' share in global trade adds explanatory power to the regressions. However, for $K > 15$, these effects vanish, suggesting that trade share does not explain the whole story. Here, we should consider the fact that outliers such as Russia, Brazil, Turkey, Ukraine, Kazakhstan, Iraq, Romania, Belarus, Bulgaria, Lithuania, Croatia, and Latvia – that experienced very high inflation rates in the sample period – distort the explanatory power for $K > 15$.

In light of these results, we make an educated guess that the granular number $K^* < 15$, and we should compute this number by leaving out the outliers. Figure 1 shows the granular curve $C(L)$ and the equal-weight curve.

Figure 1. Granular Curve (Blue) and Equal-weight curve (red)



Because we consider the L values in fives, a first inspection suggests the granular curve first intersects the equal-weight curve before $L = 10$. Then, computing the intermediate values, we find the curves intersect between $L = 8$ and $L = 9$, so the country granular number is $K^* = L = 8$. Therefore, the countries impacting world inflation more than their relative size in global trade are the United States, China, Germany, the United Kingdom, Japan, France, Italy, and the Netherlands. These big grains are responsible for the bulk of international inflation spillovers.

2.4 Discussion

Running a regression with these eight countries using equation (2) yields an R^2 equal to 0.87, while the relative participation of these countries in international trade is 44%. However, if we run a regression with the same eight countries but considering the coefficient $\text{exports}_{i,t} / \text{world exports}_t = 1$ in equation (2), R^2 becomes 0.84. Because these values are very close, we conclude that inflation spillovers may have an unexplained component related to exchange rates and interest rates.

Note that of the eight big grains, four belong to the Eurozone, and the UK was a member of the European Union for the entire sample period. An interesting exercise is to aggregate such countries in a Eurozone, and thus the new big grains become the United States, China, Japan, and the Eurozone. With $K = 4$, R^2 drops to 0.41, and with the coefficient equal to one in equation (2), R^2 is 0.50. Thus, the big five European grains play a larger role than the aggregate Eurozone. This fact occurs because we dilute the impact of the largest countries when aggregating and, therefore, R^2 decreases.

As the euro only started to circulate in 1999, it is interesting to observe our results for the period 1999-2020 as well. Now, R^2 is 0.52, and with the coefficient equal to one, R^2 is at 0.68. This finding reinforces that Germany, France, the UK, Italy, and the Netherlands contribute more to changes in world inflation than the aggregate Eurozone.

Our hypothesis that big grains can influence global inflation but cannot be affected by it is granularity-specific. Other transmission mechanisms undoubtedly play a role, and global inflation spillover may have an impact on large economies. There are two competing hypotheses about the impact of global inflation on domestic inflation, according to the literature. First, price fluctuations in the energy and commodity markets drive the global component of inflation (Bianchi and Civelli, 2015; Mikolajun and Lodge, 2016; Kamber and Wong, 2020; Attinasi and Balatti, 2021). These effects, however, are thought to be minor and transient. The other hypothesis is related to the slope of the Phillips curve, as globalization reduces the sensitivity of inflation to domestic slack (Ciccarelli and Mojon, 2010; Auer et al., 2017; Forbes, 2019). Because the Eurozone is a net energy importer, recent inflation reflects the rise in energy prices caused by Russia's invasion of Ukraine. As a result, it is considered "imported inflation," and monetary policy has no control over it. However, energy is a minor concern in the United States, and price pressures are primarily due to domestic factors. Used vehicles, for example, do to US inflation what energy does to Eurozone inflation (Schnabel, 2022). As a result, the United States is currently not importing global inflation.

We exclude Russia from the sample because the country experienced significant inflation during the 1990s. This raises the question of whether we subjected the data to significant "winsorizing" because the largest outlier was missed. The model would lose explanatory power if we included Russia. After all, this occurs in Gabaix's original model, which leads to a paradox because granular shocks appear to be important only if the shocks themselves are not heavy-tailed (Dosi et al., 2019). However, Russia is unimportant in our case.

Of course, Russia's invasion of Ukraine elevated Russia to the forefront of global inflation. Considering Russia would also reduce the granular residual's explanatory power in terms of adjusted R^2. However, if we reintroduce Russia using only data from 2000 to 2020, the adjusted R^2 increases from 0.7888 to 0.8115. As a result, it contributes little explanatory power to the granular residual. Furthermore, Russia ranks only 16th in terms of international market share from 1991 to 2020. Russia is unlikely to be a big grain.

Russia is primarily a commodity exporter and thus a price taker. Using World Bank data, we find an R^2 of 0.0917, 0.0069, and 0.0287 when we regress the average price of crude oil, natural gas, and wheat on Russian inflation as an independent variable, indicating that Russia's domestic prices have little impact on international prices of these commodities. As a result, rather than a spillover of Russian inflation via commodities, the invasion of Ukraine impacted global inflation through restrictions on commodity supply in international markets.

How do our results fit into the literature on inflation spillovers? First, they agree with those of Halka and Szafranek (2015) because small European economies import inflation from Eurozone countries. Furthermore, we add that more than the Eurozone as a whole, inflation is mainly exported by Germany, the United Kingdom, France, Italy, and the Netherlands, since when we replace these countries with an aggregate Eurozone, the granular residual loses its explanatory power. This conclusion also extends to the results in Ciccareli and Garcia (2015).

Baurle et al. (2021) argue that around 50% of price changes in Switzerland come from external inflationary shocks. However, as foreign monetary policy tends to be relatively more restrictive, such effects are amplified by the exchange rate depreciation of the Swiss franc. Although we have not directly studied the effects of the exchange rate on inflation spillovers, note that the magnitude of the impact of the spillovers they find is similar to that we see for high K.

Istiak et al. (2021) studied the inflation spillover in the G7 countries and concluded that Japan is the primary transmitter of inflation, followed by the United States. Our results are in line with theirs. However, Canada is not a big grain. Moreover, our study also suggests the United States is the top transmitter. However, those authors credit Abenomics the precedence of Japan. Because we do not consider the monetary policy as a determinant of inflation spillovers, we cannot settle this issue. Despite this limitation, we can confidently argue for the importance of granularity (an economic concept) over the G7 (a political convention) in studying inflation spillovers.

In line with all this previous literature, our results advise central banks to monitor foreign variables while conducting monetary policy. But, more importantly, we further pinpoint the big grains as deserving special attention.

2.5 Conclusion

Foreign shocks are essentially granular fluctuations. We can better understand international inflation spillovers considering the granularity hypothesis because countries responsible for the bulk of international trade linkages determine a portion of inflation spillovers larger than their size in global trade. These countries are the big grains. If there is granularity in the exports of goods, then there is granularity in inflation spillovers.

We consider annual exports and inflation data for 138 countries from 1991 to 2020. We first find export volumes across countries are not Gaussian-distributed but follow a power law. Because exports follow a power-law distribution, the largest countries disproportionately impact world inflation. This finding paves the way to the plausibility of our hypothesis of granularity in inflation spillovers because these do not matter if export volumes are normally distributed. Next, we quantify the power law for the right tail of the export volumes distribution and discuss its implications.

Then, we compute the granular residual, a measure of shocks to the largest countries. We find that countries with greater relative weight in international trade determine a portion of international spillovers larger than their trade share.

Finally, we address the problem of precisely identifying the number of countries most responsible for transmitting inflation. We find eight countries impact world inflation more than their relative size in global trade, namely the United States, China, Germany, the United Kingdom, Japan, France, Italy, and the Netherlands. These big grains account for the lion's share of inflation spillovers. The policy implication is that other countries' central banks should closely monitor these eight big grains when conducting their domestic monetary policy.

Appendix 2.1

Table A1. Ranking of the top 50 countries according to their average international trade share, 2020.

Ranking	Country	Trade share, %
1	USA	10.80
2	China	8.32
3	Germany	7.03
4	United Kingdom	4.38
5	Japan	3.90
6	France	3.86
7	Italy	3.25

8	Netherlands	3.02
9	Canada	2.92
10	Hong Kong	2.57
11	South Korea	2.19
12	Singapore	2.15
13	Switzerland	2.08
14	Spain	1.98
15	Belgium	1.93
16	Russia	1.91
17	Mexico	1.75
18	India	1.59
19	Saudi Arabia	1.34
20	Ireland	1.32
21	Australia	1.20
22	Thailand	1.16
23	Brazil	1.10
24	Sweden	1.09
25	Poland	1.09
26	Malaysia	1.05
27	Austria	1.02
28	Norway	0.98
29	Turkey	0.89
30	Denmark	0.83
31	Indonesia	0.81
32	Vietnam	0.60
33	Czechia	0.54
34	Luxemburg	0.54
35	Iran	0.54
36	South Africa	0.50
37	Finland	0.49
38	Israel	0.47
39	Ukraine	0.45
40	Hungary	0.39
41	Portugal	0.38
42	Philippines	0.37
43	Chile	0.36
44	Slovakia	0.32
45	Greece	0.32
46	Kazakhstan	0.31
47	Romania	0.31
48	Iraq	0.30
49	Algerie	0.26
50	New Zealand	0.24

Appendix 2.2

We used the percentage of a country's exports that went to international markets as a weight. We looked at all types of exports. In this Appendix 2, we will test the granular hypothesis by comparing exports in three important sectors: food, fuel, and manufacturing. High-tech exports are also important, but we do not include them because the known data series for this sector begins in 2007, and data for many countries are also lacking.

Table A2 displays the descriptive statistics for the data, which considers mean values for the countries from 1991 to 2020. In addition to high standard deviations and significant asymmetries, the variables, excluding global inflation, also exhibit excess kurtoses. This is predicted given that the variables obey power laws. Moreover, note that the average annual inflation of countries (28%) is higher than the average annual inflation of the world (4.5%). This is due to the fact that in average country inflation, the weight given to each country is equal, whereas, in global inflation, the weight given to each country in each year is proportional to the size of its economy. Therefore, larger countries with lower inflation carry more weight.

Table A2. Descriptive statistics

	Country inflation	*World inflation*	*Total exports*	*Food exports*	*Fuel exports*	*Manufacturing exports*
Sample size	138	30	138	163	156	119
Mean	28.07	4.48	1.01×10^{11}	5.51×10^{9}	9.22×10^{9}	6.26×10^{10}
S.D.	120.07	2.51	2.28×10^{11}	1.26×10^{11}	2.18×10^{10}	1.53×10^{11}
Asymmetry	9.50	0.98	4.31	4.21	4.17	4.15
Kurtosis	100.10	-0.13	22.04	22.19	21.14	19.56

Table A3 shows that the ranking of countries according to their share of the international food market differs from that of trade shares (Table A1 in Appendix 1). Note that due to Russia's relatively small market share, the invasion of Ukraine cannot specifically affect the global food market. Furthermore, without ignoring outliers for the entire sample period 1991-2020, Table A4 indicates that we cannot reject the granular hypothesis when constructing the granular residual weighted by international food market participation for $K = 1, 2, 3$, and 4.

Table A3. Top 30 countries according to their international food market share, 2020.

Ranking	*Country*	*Food market share, %*
1	United States	11.43
2	Netherlands	7.37
3	France	6.97

4	Germany	5.78
5	Brazil	4.15
6	Belgium	3.81
7	Canada	3.57
8	China	3.50
9	Spain	3.47
10	Italy	3.25
11	United Kingdom	2.95
12	Australia	2.35
13	Thailand	2.14
14	Denmark	1.98
15	Indonesia	1.72
16	Mexico	1.63
17	India	1.56
18	Malaysia	1.56
19	New Zealand	1.46
20	Ireland	1.26
21	Poland	1.20
22	Turkey	1.04
23	Hong Kong	1.03
24	Chile	1.01
25	Vietnam	0.98
26	Austria	0.80
27	Norway	0.78
28	Ukraine	0.74
29	Singapore	0.72
30	Russia	0.71

Table A4. Testing the granular hypothesis for food exports.

	Including outliers, 1991-2020			*Including outliers, 1996-2020*				*Without outliers, 1991-2020*			
K	$\bar{\Gamma}_t$	Intercept	R^2	$\bar{\Gamma}_t$	Intercept	R^2	Food market share, %	$\bar{\Gamma}_t$	Intercept	R^2	Food market share, %
1	-8.65	2.19	0.8487	-8.70	2.16	0.6310	11.43	-8.65	2.19	0.8487	11.43
2	-5.37	2.03	0.9054	-5.76	1.92	0.7832	18.80	-5.37	2.03	0.9054	18.80
3	-3.67	1.94	0.9122	-4.12	1.76	0.8022	25.76	-3.67	1.94	0.9122	25.76
4	-3.19	1.77	0.9229	-3.42	1.63	0.8315	31.55	-3.19	1.77	0.9229	31.55
5	0.09	4.06	0.2915	-2.98	2.24	0.8483	35.70	-2.83	1.78	0.9221	35.36
6	0.09	4.08	0.2871	-2.73	2.18	0.8427	39.51	-2.60	1.77	0.9246	38.92
7	0.09	4.08	0.2834	-2.49	2.14	0.8477	43.08	-2.76	1.55	0.9293	42.42
8	0.09	4.09	0.2836	-2.38	2.10	0.8317	46.58	-2.64	1.52	0.9210	45.90
9	0.09	4.09	0.2818	-2.24	2.07	0.8168	50.05	-2.53	1.46	0.9201	49.15
10	0.09	4.10	0.2798	-2.14	2.01	0.8147	53.30	-2.36	1.49	0.9205	52.05
15	0.09	4.12	0.2709	-2.01	1.99	0.8564	64.40	-2.13	1.75	0.9344	61.88
20	0.09	4.11	0.2690	-1.93	2.14	0.8702	71.87	-2.03	1.83	0.9281	68.92
25	0.09	4.08	0.2807	-1.90	2.61	0.7502	77.14	-2.02	1.85	0.9226	73.53
30	0.08	4.07	0.2660	-1.81	2.83	0.7408	80.90	-1.99	1.86	0.9098	76.58

Note: all the coefficients are statistically significant at 1%.

For $K = 5$, three phenomena are observed. First, the adjusted R^2 falls from 0.9229 to 0.2915, where it remains for subsequent K. Second, the intercept increases from approximately 2 to approximately 4. Third, the estimated granular residual becomes positive. As previously mentioned, these three indicators suggest that the inclusion of the fifth country (Brazil) biases the granular residual. Brazil experienced high inflation rates at the start of the sample period, 1991-1994. Due to the fact that Brazil's share of the food market is relatively greater than its share of total exports, the size of the granular residual is relatively very large. For example, the aggregate residual for 1991 is -1.91, which is the largest residual in absolute value. Ukraine, Russia, and Turkey have residuals greater than one in only a few of the sample years. Considering this, we have two options. The initial step is to eliminate Brazil, Ukraine, Russia, and Turkey from the sample. The alternative is to use the sample only from 1996 onwards, which excludes the years with high inflation for these outliers. Table A4 also shows that the adjusted R^2 and coefficient estimates are more stable if we replace Brazil, Ukraine, Russia, and Turkey with South Africa, Hungary, Greece, and Sweden, which are ranked 31 to 34. As a result, we do not reject the granular hypothesis because R^2 exceeds the share of the international food market. Furthermore, we cannot reject the granular hypothesis with data from 1996 to 2020, as evidenced by the fact that the adjusted R^2 is greater than the proportion of the international food market up to K = 20. For this time sample, Brazil has neither a significant impact on R^2 nor a significant impact on the estimated coefficients, indicating that the data have been adequately filtered. However, the results are worse, in terms of a lower R^2, than we discovered when we eliminated the outliers with high inflation.

Thus, when we remove outliers and restrict the number of years, the results are superior to what we observed when we did not take such precautions. We also find that the results based on participation in the international food market are superior to those based on exports in general, suggesting that inflation spillovers are transmitted more easily via food exports. The results are also significantly greater than those obtained with weighted regressions, indicating that weights provide additional information about the dynamics of global inflation.

Now we move on to the fuel markets. Table A5 displays the top 30 countries by fuel market share.

Table A5. Top 30 countries according to their international fuel market share, 2020.

Ranking	*Country*	*Fuel market share, %*
1	Saudi Arabia	10.95
2	Russia	9.87
3	Canada	4.90
4	Norway	4.58
5	United States	4.43
6	Iran	3.54
7	Netherlands	3.12
8	Nigeria	3.06
9	United Kingdom	2.88
10	Algeria	2.67
11	Australia	2.45

12	Singapore	2.33
13	Indonesia	2.31
14	Iraq	2.04
15	Mexico	2.03
16	Belgium	1.64
17	Germany	1.43
18	China	1.39
19	South Korea	1.34
20	Oman	1.33
21	Kazakhstan	1.26
22	France	1.25
23	India	1.03
24	Colombia	0.87
25	Italy	0.86
26	Spain	0.68
27	Brazil	0.58
28	Bahrain	0.54
29	Japan	0.49
30	Sweden	0.48

Only about 50% of the global market for fuel is dominated by the top ten exporters; this concentration rate is comparable to that of the global food market. We do not reject the granular hypothesis when using the international fuel market share, as shown in Table A6. When we weigh the granular residual based on food share, the results for the adjusted R^2 and the estimated coefficients are less stable, but the estimated explanatory power is higher. Russia, Iran, Nigeria, Iraq, Kazakhstan, Brazil, and Ecuador are outliers because between 1991 and 2020, their cumulative inflation rates exceeded 500%. We cannot rule out the granular hypothesis after removing these outliers from the sample because the adjusted R^2 is higher than the shares for all K analyzed. The regression results worsened when the first five years of the sample were excluded, when inflation was high in the outliers. In addition, for no K, the estimated granular residual is statistically different from zero. Overall, fuels offer a lower adjusted R^2 than total exports or food granular residuals, which suggests that fuels do not spill over inflation.

Table A6. Testing the granular hypothesis for fuel exports.

	Including outliers, 1991-2020			*Including outliers, 1996-2020*				*Without outliers, 1991-2020*			
K	$\bar{\Gamma}_t$	Intercept	R^2	$\bar{\Gamma}_t$	Intercept	R^2	Fuel market share, %	$\bar{\Gamma}_t$	Intercept	R^2	Fuel market share, %
1	-39.28	3.47	0.3816	-4.25	3.22	0.0151	10.95	-4.25	3.22	0.2301	10.95
2	1.47	3.99	0.2358	0.11	3.99	-0.0179	20.82	-4.142	2.72	0.4968	15.84
3	1.45	4.01	0.2283	0.11	4.01	-0.0381	25.72	-3.59	2.49	0.6324	20.42
4	1.44	4.02	0.2195	0.11	4.03	-0.0432	30.30	-3.44	2.31	0.7569	24.86
5	1.41	4.05	0.2158	0.11	4.04	-0.0412	34.73	-3.16	2.21	0.2807	27.97
6	1.44	3.96	0.2299	0.11	3.98	-0.0409	38.27	-2.86	2.22	0.4116	30.85
7	1.42	3.91	0.2245	0.11	3.99	-0.0435	41.39	-3.48	2.12	0.4997	33.53
8	1.41	3.92	0.2356	0.11	3.95	-0.0434	44.45	-3.23	2.11	0.1946	35.97
9	1.40	3.94	0.2318	0.11	3.95	-0.0425	47.33	-3.03	2.05	0.2700	38.30
10	1.41	3.92	0.2409	0.11	3.94	-0.0419	50.00	-2.86	2.47	0.3191	40.61
15	1.42	3.88	0.2454	0.11	3.91	-0.0430	61.15	-2.79	2.54	0.1513	48.43
20	1.43	3.87	0.2407	0.11	3.93	-0.0432	68.28	-2.56	2.65	0.240	53.78
25	1.42	3.87	0.2587	0.11	3.90	-0.0418	73.55	-2.45	2.52	0.2281	56.44
30	1.33	3.88	0.2683	0.10	3.90	-0.0374	76.33	-2.43	2.60	02269	58.41

Note: for the period 1991-2020, all coefficients are statistically significant at 1%; however, no estimated granular residual is statistically different from 0 at 10% for the period 1996-2020.

Regarding manufacturing, Table A7 shows the top 30 countries in terms of global manufacturing market share, and Table A8 shows the results of regressions that used manufacturing participation as weights in the construction of the granular residuals. Note that these are similar to those obtained when using food, despite the fact that the major countries in both cases are different. Furthermore, except for *K* = 30, where the adjusted R^2 is close to that obtained in the other *K*, the results obtained using only the sample after 1996 and for the case without outliers are quite similar.

Table A7. The top 30 countries according to their global manufacturing market share, 2020

Ranking	*Country*	*Manufacturing market share, %*
1	Germany	7.64
2	United States	7.15
3	China	6.98
4	Japan	5.41
5	France	3.47
6	Italy	3.07
7	United Kingdom	2.76
8	Belgium	2.41
9	South Korea	2.38
10	Netherlands	2.29
11	Hong Kong	2.14
12	Canada	1.79
13	Singapore	1.57
14	Mexico	1.57
15	Spain	1.32
16	Switzerland	1.23
17	Sweden	0.94
18	Malaysia	0.92
19	Austria	0.89
20	Thailand	0.82
21	Ireland	0.75
22	India	0.74
23	Poland	0.66
24	Czechia	0.63
25	Turkey	0.52
26	Denmark	0.49
27	Brazil	0.47

28	Finland	0.46
29	Hungary	0.42
30	Indonesia	0.42

Table A8. Testing the granular hypothesis for manufacturing exports.

	Including outliers, 1991-2020			
K	$\bar{\Gamma}_t$	Intercept	R^2	Manufacturing market share, %
1	-14.09	1.41	0.8521	7.64
2	-6.89	1.79	0.8812	14.79
3	-6.81	1.55	0.7644	21.77
4	-4.62	1.32	0.8703	27.18
5	-3.89	1.36	0.8820	30.65
6	-3.62	1.33	0.8887	33.72
7	-3.28	1.40	0.8852	36.49
8	-2.99	1.45	0.8908	38.89
9	-2.96	1.40	0.8908	41.27
10	-2.80	1.40	0.9054	43.56
15	-2.60	1.39	0.8963	51.94
20	-2.38	1.34	0.8929	56.72
25	-2.53	1.47	0.8995	60.01
30	0.33	4.53	0.0782	62.26

Note: all the coefficients are statistically significant at 1%.

As an exercise, we estimate the empirical model using the share of global GDP from 1991 to 2020 as a weight for the granular residual. Table A9 displays the results. Given the preceding discussion, it is not surprising that we cannot reject the granular hypothesis in Table A9 when GDP is used as a weight in constructing the granular residual. Because countries are distributed by a power law when ordered by GDP, the largest countries are expected to account for the largest share of changes in global inflation, which is greater than their relative size. This is because the variations of many small countries tend to cancel each other out and converge towards the average, whereas the variations of large "grains," in much smaller quantities, are not "compensated" by the law of large numbers. This means that even if we rejected the granular hypothesis using shares of global GDP, we could not rule out the possibility of inter-country inflationary spillovers. After all, the estimated correlations between the shares of various types of exports and the share of global GDP are strong (Table A10). The correlations by country were calculated in Table A10 for the intersection of the sets of countries for which we had GDP data and, concurrently, exports of the products in question. As a result, the number of countries n varies in each correlation calculation. Because the residual is calculated based on the ranking in each market rather than the fixed countries, we considered (and confirmed) the possibility that the correlation between the shares in the order of the countries was greater than for each specific country.

Table A9. Testing the granular hypothesis for world GDP share.

	Including outliers, 1991-2020			*Including outliers, 1996-2020*				*Without outliers, 1991-2020*			
K	$\bar{\Gamma}_t$	Intercept	R^2	$\bar{\Gamma}_t$	Intercept	R^2	World GDP share, %	$\bar{\Gamma}_t$	Intercept	R^2	World GDP share, %
1	-4.44	2.03	0.8395	-4.11	2.09	0.6177	25.64	-4.44	2.03	0.8395	25.65
2	-2.55	1.78	0.8653	-2.85	1.54	0.7298	35.12	-2.55	1,78	0.8653	36.34
3	-2.65	1.57	0.8647	-2.52	1.62	0.6898	43.71	-2.65	1.57	0.8647	43.82
4	-2.28	1.56	0.8734	-2.21	1.56	0.7230	49.30	-2.28	1.56	0.8734	49.82
5	-2.13	1.55	0.8757	-2.07	1.54	0.7451	53.66	-2.13	1.55	0.8757	54.19
6	-1.94	1.53	0.8852	-1.92	1.50	0.7619	57.72	-1.94	1.53	0.8852	58.44
7	-1.87	1.49	0.8912	-1.85	1.46	0.7772	61.00	-1.87	1.49	0.8912	61.92
8	0.16	4.26	0.2309	-1.82	1.61	0.7938	63.39	-1.80	1.50	0.8951	66.41
9	0.16	4.27	0.2264	-1.76	1.61	0.8033	65.60	-1.77	1.48	0.8957	68.42
10	0.16	4.28	0.2238	-1.72	1.60	0.8043	67.74	-1.74	1.63	0.9016	70.39
15	0.13	4.19	0.2337	-1.82	1.89	0.8572	76.47	-1.69	1.64	0.9137	79.25
20	0.13	4.18	0.2348	-1.81	2.14	0.8523	81.24	-1.62	1.68	0.9170	83.85
25	0.13	4.18	0.2346	-1.76	2.11	0.8562	84.51	-1.62	1.67	0.9109	86.50
30	0.13	4.18	0.2329	-1.74	2.09	0.8569	86.81	-1.61	1.71	0.9080	88.08

Note: all the coefficients are statistically significant at 1 percent.

Table A10. Correlations.

Correlations	*By country*	*Between the same places in the rankings*	*n*
	World GDP share		
Total exports share	0.8645	0.9165	128
Food share	0.7466	0.9177	113
Fuel share	0.3027	0.9208	111
Manufacturing share	0.8038	0.8672	110

Finally, Table A11 compares the various means of weighing the granular residual. The adjusted R^2 in the column Total exports share decreases significantly between $K = 1$ and $K = 2$ (when China is added). As a result, the results of the weights without China are shown in the following column. As we can see, except when we use the share of the international fuel market as weights, the explanatory power of exports is greater than the share of GDP for several K. This provides evidence that using exports adds explanatory power to the granular residual. Therefore, we cannot dismiss the hypothesis that there are granular inflation spillovers and that the transmission mechanism occurs, at least in part, through the international market. Although we cannot reject the granular hypothesis for fuel share, we can see that for any K, the adjusted R^2 exceeds the GDP share, implying that inflation is not transmitted through the fuel market. In contrast, the food share results are always higher than the GDP share results, indicating that inflation migrates from one country to another via this market. The results vary when weighted by manufacturing market participation. Six of the 14 K have results that exceed the GDP share, while eight have results that are lower, implying that, to a lesser extent than for food, we cannot rule out the possibility that the manufacturing market transmits some of the inflation.

Table A11. Adjusted R^2 for alternative ways of weighing the granular residual, data without outliers.

K	GDP share	Trade share	Trade share without China	Food share	Fuel share	Manufacturing share
1	0.8395	0.8406	0.8406	0.8487	0.3816	0.8521
2	0.8653	0.3642	0.8863	0.9054	0.5736	0.8812
3	0.8647	0.6007	0.8865	0.9122	0.6890	0.7644
4	0.8734	0.7214	0.8821	0.9229	0.7406	0.8703
5	0.8757	0.7880	0.9039	0.9221	0.7875	0.8820
6	0.8852	0.8408	0.9004	0.9246	0.7935	0.8887
7	0.8912	0.8537	0.9086	0.9293	0.7377	0.8852
8	0.8951	0.8733	0.9042	0.9210	0.7650	0.8908
9	0.8957	0.8840	0.8962	0.9201	0.7734	0.8984
10	0.9016	0.8579	0.8934	0.9205	0.7454	0.9054
15	0.9137	0.8673	0.8773	0.9344	0.6900	0.8963
20	0.9170	0.8784	0.8874	0.9281	0.6932	0.8929
25	0.9109	0.6610	0.8937	0.9226	0.7212	0.8995
30	0.9080	0.6823	0.8906	0.9098	0.7168	0.8931

3. Essay 2: Granular cities

Abstract

We propose extending the concept of granularity from firms to cities, where it refers to the coexistence of a few large and numerous small cities. City size distribution follows Zipf's law, a power law. We argue that granularity and power laws are interrelated, and hypothesize that large cities play a significant role in the business cycle beyond their relative size. Our study on American and Brazilian cities' data from 2003 to 2019 supports this granular hypothesis. We find that the granular city size for the United States is three metropolitan areas. If we redefine cities as counties, the granular size is five counties. In Brazil, the granular size equates to three municipalities. This essay contribution to the literature is to highlight a spatial component of granularity not considered so far.

3.1 Introduction

Cities are considered one of humanity's greatest inventions as they bring people closer, facilitating connections and the exchange of information, leading to the emergence of new ideas and innovations (Glaeser, 2011). Our hypothesis is that large cities play a crucial role in the business cycle beyond their relative size. To test this hypothesis, we apply the concept of "granularity," originally used in the context of firms (Gabaix, 2011).

Granularity within firms refers to the coexistence of a few large firms alongside numerous smaller ones. An economy is considered "granular" due to this diversity; if all firms were the same size, it would be "smooth." The "granular residual" is the cumulative effect of individual firm-specific shocks, weighted by their respective sizes. This term emphasizes the portion of the business cycle that cannot be explained by macroeconomic shocks alone and underscores the significance of microeconomic shocks unique to each firm. The granular residual captures the impact of shocks on the largest firms, accurately assessed by considering the "granular size."

The presence of significant "grains" necessitates a heavy right tail in the size distribution of firms. This power law tail enables these large grains to impact the business cycle in ways that a continuum of equally sized firms cannot. When the firm size distribution follows a heavy-tailed pattern, idiosyncratic shocks affecting the largest firms should not average out at the aggregate level. Instead, they are expected to influence GDP dynamics. Similarly, in cities, we observe a similar pattern where a few cities are significantly larger than the majority of smaller cities (Zipf's law). Consequently, idiosyncratic shocks to large cities may not be fully compensated by opposing shocks to smaller cities. If all firms were the same size and reacted similarly to shocks, the granular residual would be zero. However, if large firms are more susceptible to shocks, the granular residual could be substantial.

Zipf's law (Zipf, 1949) characterizes a power law distribution of city populations, where population is inversely related to rank. Although Zipf's law predates and is separate from the concept of granularity, both approaches acknowledge hierarchical patterns in urban systems. We highlight that granularity and Zipf's law are closely related, as the heavy-tailed distribution inherent in hierarchical quantities allows for the presence of exceptionally large units.

Growth rates exhibit correlation with individual unit shocks (firms or cities) (Gabaix, 2011; Dosi et al., 2019). Hence, these shocks can be viewed as growth rates for cities. The majority of the business cycle can be explained by growth, wherein recessions occur when significant units experience below-average growth, while booms witness the opposite. Consequently, idiosyncratic shocks serve as growth rates for cities. The level of integration of a city into the national economy determines the extent of spillover effects on other cities within the country. This integration, in turn, enhances the explanatory power of the granular residual.

Various factors contribute to differential growth rates of cities compared to the national average, such as the discovery of natural resources, the emergence of innovation hubs, changes in building codes, the arrival of multinational corporations, or substantial state investments. The key notion is that greater integration of a city into the national economy amplifies spillover effects, consequently intensifying the explanatory power of the granular residual.

The city-firm analogy finds justification in the fact that despite advancements in technology facilitating information transmission, direct human contact remains highly efficient due to centuries of human development. This efficiency enhances the benefits of human agglomeration in confined geographical areas, making new information technologies complementary to physical proximity. Consequently, the largest and most innovative companies tend to locate near dense populations, capitalizing on the knowledge flows and opportunities for economic growth facilitated by larger cities. This positive feedback process reinforces the emergence of exceptionally large cities (Glaeser, 2011).

The rationale behind Zipf's law for cities, which states that city sizes (in terms of population) follow a power law with an exponent of one, can thus be justified. However, alternative interpretations exist (Rauch, 2014), as Zipf's law lacks a sufficient theoretical explanation and is not accounted for by standard urban system models (Krugman, 1996). A random growth model may adequately explain it (Simon, 1955), or it might not require any theoretical justification at all (Mandelbrot, 1961). It is plausible that city location is fundamentally random, with city size unaffected by interactions with other cities (Ioannides and Overman, 2004). A meta-analysis of 515 estimates from 29 studies indicates that the Zipf coefficient exceeds one and that cities are more evenly distributed than suggested by Zipf's law (Nitsch, 2005). Chauvin et al. (2017) provide a concise overview of the ongoing debate surrounding Zipf's law.

Previous authors, including Zipf (1949), Krugman (1996), Dobkins and Ioannides (1998), and Gabaix (1999), have shown the presence of Zipf's law in various U.S. databases. Zipf's analysis focused on the top 100 metropolitan areas in the United States in 1940, calculating a power law with a slope of approximately one (-0.9835 ± 0.0625) based on the rank-frequency distribution. We replicate Zipf's law using a more recent dataset that includes counties and metropolitan areas in the United States. While Rozman (1990) shows Zipf's law for China in the mid-1800s, we were unable to obtain the specific data required to calculate the granular residual for recent China data. However, we do utilize data from Brazil, an emerging country. Gabaix (1999) attributes Zipf's law to the growth processes observed in cities in the upper tail (referred to as Gibrat's law) and the diminishing decline of shocks with size beyond a certain threshold. Although Giesen and Sudekum (2011) confirm this hypothesis for German cities, it fails to hold for a broader sample of U.S. cities from 1900 to 1990 (Black and Henderson, 2003). Nevertheless, we argue that Gabaix's insight emphasizes the interrelation of power law and granularity, motivating our investigation into granular cities.

Considering the insights from Glaeser (2011) and Gabaix (1999, 2011), it is plausible to hypothesize that the business cycle primarily occurs in large cities. If it were the opposite, small cities would have a more significant impact on the business cycle, resulting in symmetric cycles of rise and fall. Our hypothesis carries a crucial implication: economic growth spreads from large cities to smaller ones. Gabaix and Koijen (2020) propose that evaluating the granular hypothesis also allows us to test various types of spillovers. When large cities foster innovative developments linked not only to the business cycle but also long-term economic growth (Glaeser, 2011), such growth is expected to disseminate to smaller cities, strongly influencing a country's production fluctuations. Metropolitan areas often encompass multiple municipalities, making it challenging to separate the economic growth of specific municipalities from their surroundings. Hence, we examine counties and metropolitan areas, rather than focusing solely on municipalities, to account for these complexities.

Studying the business cycle at the city level involves examining economic fluctuations and patterns within specific urban areas. The concept of granularity aligns well with this endeavor, offering a comprehensive understanding of localized impacts and economic dynamics. While the business cycle is typically analyzed at the national or regional level, studying it at the city level provides a more detailed perspective. It allows exploration of the relationship between the city-level business cycle and broader macroeconomic factors, including national economic conditions, policies, international trade, and global trends. Such analysis holds practical implications for policymakers, urban planners, and businesses. It aids in identifying economic strengths and vulnerabilities, designing targeted policies to stimulate growth or address downturns, and assessing regional disparities within a country's economy.

To ensure the generalizability of our findings, we analyze data from both a developed country (the United States) and an emerging country (Brazil). We test the granular hypothesis using data from counties, metropolitan areas, and municipalities in these countries spanning from 2002 to 2019. Our hypothesis is stated as:

Hypothesis. Larger cities in the United States and Brazil explain a greater proportion of the business cycle than their relative size.

If this hypothesis holds true, it implies the existence of granular cities, similar to granular firms discussed in the literature, which explain the business cycle in both the U.S. (Gabaix, 2011) and Brazil (Silva and Da Silva, 2020). We use Blanco-Arroyo et al. (2018)'s methodology to compute the granular size of American and Brazilian cities.

The essay's structure is as follows: 1) We replicate Zipf's law for our dataset using Gabaix and Ibragimov's (2011) methodology. 2) We test the granular hypothesis, examining whether large cities contribute more to the business cycle relative to their size. 3) We calculate the granular size of cities. 4) We compare our findings between the two countries and with the existing literature on firms.

3.2 Materials and methods

Cities can be defined in various ways. A narrow definition sees cities as municipalities with specific boundaries and governing bodies. A broader definition includes metropolitan areas, which encompass a central city, suburbs, and interconnected urban and suburban regions. This definition acknowledges that cities extend beyond a single administrative unit and considers economic and social relationships between the center city and its surroundings. Legally defined cities such as counties or municipalities are typically not ideal for economic analysis unless there is a compelling reason. In this case, the availability of data justifies including counties and municipalities in our study. Despite this limitation, we can still obtain significant findings for metropolitan areas, especially for the United States.

To compute the granular residual, population and economic output data are essential. In the United States, we have access to output data for counties and metropolitan areas, so we consider both. The Census Bureau provides GDP data for counties and metropolitan areas, whereas the Brazilian Institute of Geography and Statistics releases GDP data for municipalities. New York City's five boroughs (Bronx, Kings, New York, Queens, and Richmond) correspond to five counties. On the other hand, Los Angeles County, the largest in the U.S., comprises 88 municipalities. With the municipality-level data available, we can calculate population and GDP for 82 Brazilian metropolitan areas. Therefore, we analyze data in four ways: 1) U.S. counties, 2) U.S. metropolitan areas, 3) Brazilian municipalities, and 4) Brazilian metropolitan areas. The examination covers population and economic output for both countries, spanning from 2002 to 2019. You can find the datasets on Figshare.

Table A1 in the Appendix presents the most populous counties in the United States. Over the period of 2002 to 2019, Los Angeles County had a population more than ten times larger than DuPage County, IL, which ranked 51st. Similarly, Table A2 in the Appendix displays the 51 largest metropolitan areas in terms of population during the same period. The largest metropolitan area, New York-Northern New Jersey-Long Island, was approximately 18 times larger than the 51st, Rochester. Moving on to Table A3, it showcases the largest Brazilian municipalities based on average population between 2002 and 2019. Out of around 5570 Brazilian municipalities, Sao Paulo, the largest one, accounted for nearly 5% of the national population and was almost 27 times the size of Santos, the 51st largest municipality. Lastly, Table A4 reveals the 51 largest Brazilian metropolitan areas out of a total of 82. The contrast is even more significant for metropolitan areas. Greater Sao Paulo, the largest metropolitan area, had a population approximately 56 times larger than Toledo, the 51st on the list.

In the Appendix, Tables A5 and A6 present descriptive statistics for four subsets of data. These statistics cover the 200 largest American counties and Brazilian municipalities with over 100,000 average inhabitants between 2002 and 2019, as well as all metropolitan areas in the United States and Brazil. The statistics exclude extreme idiosyncratic shocks in terms of growth rates but retain population outliers. We found that outliers greatly influenced the results when calculating growth rates, specifically impacting asymmetry, kurtosis, and the Jarque-Bera test. To address this, we removed growth rate outliers with absolute z-scores exceeding 2 from the sample. As a result, Tables A5 and A6 display population and growth rate descriptive statistics based on samples of varying sizes.

The Jarque-Bera test assesses the Gaussianity of the test statistic, which follows a chi-square distribution with two degrees of freedom. The critical value at a 5% significance level is 5.99. Across the four datasets, we reject the null hypothesis that cities are normally distributed when ordered by population. This aligns with the expected outcome if cities adhere to Zipf's law. Notably, growth rates are normally distributed for American counties, metropolitan areas, and Brazilian municipalities. Therefore, we anticipate the granular hypothesis to hold for these subsamples. However, the normality hypothesis for growth rates in Brazilian metropolitan areas is rejected, indicating a deviation from the granular hypothesis. In these cases, the estimated values in the Jarque-Bera test exceed the critical value of 5.99. These findings reflect the requirement for population to follow a power law in order to test the granular hypothesis, while growth rates are not necessarily characterized by heavy tails (Dosi et al., 2019).

Zipf's law

The size distribution of cities, according to Zipf's law, follows a power law: the number of cities with populations greater than x is proportional to $1/x$ (Gabaix, 1999). To test for the presence of power laws in all subsets of population data, we obtain ordinary least squares (OLS) estimates of the tail exponents employing the rank – ½ method of Gabaix and Ibragimov (2011). Thus, we take

$$\ln(\text{rank}_i - \tfrac{1}{2}) = a - \alpha \ln \frac{\text{population}_i}{\text{population}_m}, \quad (1)$$

where population_i is the mean population for the regions in the samples from 2002 to 2019 for both countries, rank_i is rank from highest to lowest according to the size of population_i, and population_m is the lowest bound used as a cutoff to analyze the tail. A power law occurs between a minimum cutoff and a maximum threshold because few distributions follow a power law over their entire range. This is why a distribution is said to have a power-law tail (Newman, 2005). The α in equation (1) is the tail index (Pareto exponent) and measures the weight of the right tail, with smaller values indicating heavier tails (Jenkins, 2017). The aggregate fluctuations in heavy-tailed distributed data are not proportional to $\sqrt{\frac{1}{n}}$, as would be expected if the data were Gaussian-distributed. Granular shocks are proportional to $1/\ln n$ instead (Gabaix, 2011). As a result, shocks to large cities are not offset, and economic growth spillovers are substantial.

The city granular residual

The granular residual quantifies the impacts of shocks on major companies at the individual firm level. An important implication is that when regressing a country's growth rate on the granular residual of its largest companies, the adjusted R^2 will exceed the companies' share of GDP. This suggests that shocks, specifically those related to productivity, affecting large corporations play a substantial role in shaping the business cycle (Gabaix, 2011).

Similar to Gabaix's firm granular residual, we define the granular residual for *K* cities as

$$\Gamma_t = \sum_{i=1}^{K} \frac{\text{city population}_{i,t}}{\text{country population}_t} (g_{i,t} - G_t) , \quad (2)$$

where city $\text{population}_{i,t}$ is the population of city *i* at time *t*, country population_t is the country population at time *t*, $g_{i,t}$ is city *i*'s per capita output growth rate at time *t*, and G_t is the country's per capita GDP growth rate at time *t*. To compute Γ_t, we add up the residuals of the $i = 1, 2, \ldots, K$ cities. Then, we run an OLS regression of the country's per capita GDP growth rate on the granular residual Γ_t. The adjusted R^2 value is expected to be greater than the percentage of the country's population represented by such cities in the presence of granular effects. In this case, we do not reject our previously stated hypothesis.

The granular residual's contribution assesses the extent to which shocks impacting large cities influence a country's overall economic dynamics. It suggests that the effects of shocks experienced by large cities may not be entirely counteracted by opposing effects in smaller cities. This underscores the significance of comprehending the specific dynamics and characteristics of individual cities when explaining a nation's overall economic fluctuations. Essentially, the granular residual for cities implies that shocks originating from large cities have a greater influence than what can be accounted for by their relative size alone. This encompasses the integration of large cities into the national economy and the potential spillover effects on other cities within the country.

According to the standard method in the business cycles literature, after demonstrating how granular cities can explain the business cycle, the next step should be to evaluate the impact of a few cities on aggregate volatility (the second order moment). However, our analysis does not extend to this level of detail.

The link between city population size and economic output is well-supported in urban economics and economic geography. Key points include: 1) Larger cities can provide services like transportation and healthcare more efficiently, spreading costs over a larger population, which may reduce per-capita costs (economies of scale) (Glaeser, 2011). 2) The concentration of people and firms in big cities boosts productivity through knowledge spillovers, labor market pooling, and improved access to suppliers and customers, enhancing output and innovation (Duranton and Puga, 2004). 3) Larger cities attract a wide range of skills, helping firms find suitable employees and boosting productivity (Florida, 2019). Thus, improvements in technological efficiency and innovation contribute to increased labor productivity, the key measure we employed for calculating shocks. 4) This diversity in bigger cities drives innovation by bringing together varied perspectives and expertise (Jacobs, 1970). 5) Larger cities often see more development and investment in infrastructure, supporting business operations and contributing to higher economic output (Bettencourt et al., 2007).

Although there are reasons to link city population size with economic output, this relationship is intricate and shaped by factors like the city's industrial makeup, geographical position, policies, and history. Additionally, bigger cities often contend with issues such as higher living costs, traffic congestion, and environmental challenges. Nevertheless, empirical research consistently indicates a positive correlation between a city's size and its economic performance, with larger cities typically having higher per capita GDP than smaller cities or rural regions (Glaeser, 2011; Bettencourt et al., 2007; Rosenthal and Strange, 2004).

Replacing economic output with city population size in modeling is challenging. Specifically, in the granular equation (2), we need to factor in the Domar weight multiplied by idiosyncratic shocks. Originating from Hulten's theorem and explored in Gabaix (2011), this approach is based on an input-output economy framework. The Domar weight, defined as a firm's gross output relative to total GDP, indicates its contribution to overall output or productivity, typically in economies with intermediate inputs and networks. However, applying this concept to cities, where the links between intermediate goods among different cities are less clear, presents difficulties. In equation (2), the Domar weight should represent a city's gross output as a proportion of GDP, not its population share. While the $(g - G)$ component still quantifies the effect of a city's growth shock, it is important to note that unlike firms, which directly contribute to aggregate output, cities function as hubs that attract labor and businesses. The literature includes a few examples of studies related to this modeling task. Hsieh and Moretti (2019) used a spatial equilibrium model to measure the extent and aggregate costs of labor misallocation across U.S. cities. Duranton and Puga (2024) created a microfounded urban growth model in which human capital spillovers promote entrepreneurship and learning in heterogenous cities. They used this model to explore different hypothetical scenarios, quantitatively evaluating the impact of cities on economic growth and overall income.

The granular city size

To accurately assess a city's granular residual contribution, it is crucial to calibrate using the granular city size K^* in equation (2). We evaluate the explanatory power of the city granular residual by comparing a weighted curve (as defined in equation (2)) with an equal-weighted curve. The equal-weighted curve assumes all cities are of the same size and follows a methodology introduced by Blanco-Arroyo et al. (2018) for determining the granular size of firms. This ensures that the estimation of a city's granular residual is not underestimated or overestimated.

Therefore, to accurately analyze the granular residual contribution of cities, it is important to consider the concept of granular cities. This entails including the granular city size in the analysis, achieved through a calibration procedure. The calibration aims to strike a balance by including enough cities to capture meaningful variation while excluding cities that may introduce irrelevant information. Insufficient inclusion of cities may lead to an underestimation of the granular residual contribution, as significant variations in excluded cities are overlooked. Conversely, an excessive number of cities may result in an overestimation, as noise from less economically significant cities dilutes the overall signal. By calibrating the analysis, we identify a subset of cities that provide a representative sample, mitigating biases arising from too few or too many cities. This approach enables a more accurate evaluation of the explanatory power of the granular residual by comparing observed granular residuals to a hypothetical curve with equal weights assigned to all cities, treating them as if they were of equal size.

The "granular curve" of function $C(L)$ of average cumulative explanatory power is

(3)
$$C(L) = \frac{1}{Q}\sum_{K=1}^{Q} \mathrm{R}^2(K,L),$$

where Q is an arbitrary number of cities, and L is the number among the best-ranked cities that should be removed from the sample and replaced by the $Q + 1, \ldots, Q + L$ best-ranked. Thus, for each L, the same number of regressions is calculated, with the granular residual (with the weights attributable to population size) as the only explanatory variable. $C(L)$ indicates the average R^2, in each L, for Q regressions performed. Therefore, the idea is to examine the sensitivity of R^2 to a sequential exclusion of the largest cities, which means increasing L. So, we consider the explanatory power curve $\mathrm{R}^2 (K, L)$ as a function of an increasing number of cities K and for different values of the largest cities L. We want to see how the curve $\mathrm{R}^2 (K, L)$ performs depending on the number L of highest-ranked cities that are removed from the sample and replaced by the $Q + 1, \ldots, Q + L$ following cities. Moreover, we run the same number Q of regressions with the granular residual Γ_t as the explanatory variable for each L. And $C(L)$ refers to the average R^2 in every L for Q regressions performed.

In parallel, we must run the same regressions without weights, which means ignoring the population size weights. The granular number is obtained when the granular curve $C(L)$ equals the adjusted R^2 of the regressions without weights. In general, the granular size indicates the number of the largest grains that can be removed from the sample without affecting the average explanatory power (adjusted R^2) of the granular residual for the various possible $K \leq Q$.

In short, the equal-weight curve quantifies the contribution of shocks to the business cycle from cities as if they were all the same size, with negligible cities playing a larger role. As we gradually remove the L largest cities from the granular residual, we expect to see the transition from the granular curve $C(L)$ to the equal-weight curve. The number L where the granular curve $C(L)$ first intersects the equal-weight curve corresponds to the granular city size K^*.

We streamline the computation into three steps to avoid running several thousand regressions:

1. We define Q for each case as 1.5 the value of K from which the adjusted R^2/population size ratio is below 1, that is, $Q = 36$ for U.S. counties and $Q = 18$ for U.S. metropolitan areas, and $Q = 15$ for Brazilian municipalities.
2. We run regressions for the L in threes ($L = 3$, $L = 6$, etc.) until the curves with and without weights intersect. We also run regressions with the K in threes.
3. We compute $C(L)$ and the regressions for the intermediate values of L to find the granular size when $C(L)$ with weights is less than R^2 without weights.

The empirical model

The empirical model is then

$$g_{i,t} = \beta_1 + \beta_2\Gamma_t + \varepsilon_t\ , \tag{4}$$

where Γ_t is given by equation (2), β_1 and β_2 are the parameters estimated by Gabaix's (2011) OLS method, and ε_t is the estimated error. Parameter β_1 is the intercept, and parameter β_2 represents the country's average growth rate in response to the granular residual. The adjusted R^2 estimated for this empirical model, in turn, is a measure of the granular residual's explanatory power in relation to the business cycle.

3.3 Results

Power Law

Table 1 presents the estimation of equation (1) for American counties across different cutoffs from 2002 to 2019. The high adjusted R^2 values suggest that the population distribution of American counties may follow a power law or potentially Zipf's law. Notably, the α values in Tables 1-4 are approximately one, which supports the literature's conclusion that the Pareto exponent is greatly influenced by the definition of city and the sample size (Rosen and Resnick, 1980).

Table 1. Power law for the U.S. counties.

Cutoff, %	*n*	*Intercept*	α	R^2
1	31	3.46	−2.05	0.99
5	154	5.19	−1.74	0.98
10	308	5.96	−1.41	0.96
20	616	6.61	−1.17	0.96
30	924	7.03	−1.05	0.95
40	1232	7.27	−0.99	0.96
50	1539	7.51	−0.94	0.96
60	1847	7.73	−0.91	0.96
70	2155	7.92	−0.87	0.96
80	2463	8.15	−0.82	0.95
90	2771	8.44	−0.76	0.94
100	3079	10.69	−0.63	0.86

Note: Except for the 1% cutoff, where the intercept and slope are statistically significant at 5%, all estimated coefficients are statistically significant at 1%.

Based on the adjusted R^2 values from different cutoffs for U.S. metropolitan areas (Table 2), we find no evidence to reject the hypothesis that population distributions adhere to a power law. In most cases, the estimated Pareto exponent is near -1, suggesting that the distribution of metropolitan areas potentially follows Zipf's law, similar to Gabaix's (1999) sample.

Table 2. Power law for the U.S. metropolitan areas.

Cutoff, %	*n*	*Intercept*	α	R^2
1	4	1.43	−1.79	0.90
5	18	3.13	−1.89	0.96
10	36	3.69	−1.60	0.96
20	72	4.45	−1.25	0.94
30	107	4.80	−1.16	0.95
40	143	5.09	−1.11	0.96
50	179	5.33	−1.06	0.96
60	215	5.53	−1.02	0.96
70	251	5.70	−0.98	0.96
80	286	5.77	−0.96	0.96
90	322	5.93	−0.94	0.97
100	358	6.52	−0.91	0.96

Note: Except for the 1% cutoff, where the intercept and slope are statistically significant at 5%, all estimated coefficients are statistically significant at 1%.

In the Brazilian data, power laws are also observed. Table 3 indicates that the proximity of −1 for most cutoffs prevents us from dismissing the notion that municipalities sizes from 2002 to 2019 adhere to a power law. Specifically, the distribution potentially aligns with Zipf's law, supported by the high adjusted R^2 values.

Table 3. Power law for the Brazilian municipalities.

Cutoff, %	n	Intercept	a	R2
1	56	3.99	-1.36	0.99
5	279	5.68	-1.27	0.99
10	557	6.41	-1.21	0.99
20	1114	7.05	-1.17	0.99
30	1671	7.45	-1.15	0.99
40	2228	7.77	-1.13	0.99
50	2785	8.05	-1.11	0.99
60	3342	8.31	-1.08	0.99
70	3899	8.53	-1.03	0.98
80	4456	8.72	-0.98	0.97
90	5013	8.94	-0.92	0.96
100	5570	9.90	-0.84	0.93

Note: All estimated coefficients are statistically significant at 1%.

The presence of high adjusted R^2 values suggests that the distribution of Brazilian metropolitan areas may conform to a power law. Additionally, Pareto exponents close to −1 indicate the influence of Zipf's law (Table 4). These findings in Tables 3 and 4 align with previous literature (Chauvin et al., 2017).

Table 4. Power law for the Brazilian metropolitan areas.

Cutoff, %	*n*	*Intercept*	α	R^2
5	4	1.30	−1.13	0.89
10	8	1.71	−1.34	0.91
20	16	2.82	−1.39	0.95
30	25	3.31	−1.30	0.96
40	33	3.70	−1.20	0.96
50	41	3.89	−1.13	0.96
60	49	4.07	−1.08	0.96
70	57	4.28	−1.01	0.95
80	66	4.70	−0.95	0.94
90	74	4.81	−0.85	0.91
100	82	5.56	−0.70	0.84

Note: With the exception of the 5% cutoff, where estimates are statistically significant at 5%, all estimated coefficients are statistically significant at 1%.

In Appendix 3, following Clauset et al. (2009)'s method, we estimate Pareto exponents with maximum likelihood and compare power law to exponential distributions, both derived using maximum likelihood. While maximum likelihood estimates differ from those by ordinary least squares, their ratio is consistent across the four datasets. Except for Brazilian metropolitan areas, the power law distribution is more likely to describe the data than the exponential distribution.

Granular residual

We present the estimation of equation (4) for each of the four data subsets. Table 5 displays the results for the largest U.S. counties. In the range of K = 3 to K = 25, the R^2/population ratio is positive and greater than one, indicating the statistical significance of the estimated granular residual at a 5% level. Interestingly, the adjusted R^2 is consistently higher than the county population/country population ratio for K values from 3 to 25. Hence, we cannot dismiss the granular hypothesis, which suggests that economic growth spills over from large cities and they play a significant role in the business cycle. Notably, the highest R^2 value (0.53) is observed at K = 6, including counties such as Los Angeles, Cook, Harris, Maricopa, San Diego, and Orange, CA. However, when considering only the granular residual of Los Angeles County (combining 88 cities), the adjusted R^2 turns negative. To address this, we replace Los Angeles County with Maricopa County and rerun the regressions for the largest counties in each region, resulting in an adjusted R^2 of 0.34. Additionally, noteworthy contributions to the R^2/population ratio come from Cook, Harris, and Maricopa Counties, achieving an adjusted R^2 of 0.47 for 4.23% of the population, with an R^2/population ratio of 11.3, surpassing any other county.

Table 5. The granular residual of the U.S. counties

K	R^2	*County population/ Country population*	R^2*/population ratio*	*Growth rate*	*Intercept* β_1	*Estimated granular residual*	p-*value*
1	–0.06	0.03	–2.00	–	0.03	2.71	0.85
2	–0.05	0.04	–1.08	0.92	0.03	5.74	0.66
3	0.16	0.06	2.64	3.72	0.02	*17.88*	0.06
4	0.39	0.07	5.35	2.70	0.02	*21.61*	0.00
5	0.48	0.08	5.71	0.35	0.02	*22.17*	0.00
6	0.53	0.09	5.64	–0.06	0.02	*19.99*	0.00

7	0.35	0.10	3.44	–2.20	0.02	*17.57*	0.00
8	0.38	0.11	3.53	0.08	0.02	*15.99*	0.00
9	0.38	0.11	3.25	–0.28	0.02	*14.59*	0.00
10	0.27	0.12	2.19	–1.05	0.02	*12.33*	0.01
11	0.26	0.13	1.97	–0.21	0.02	*11.28*	0.02
12	0.25	0.13	1.85	–0.12	0.02	*10.51*	0.02
13	0.25	0.14	1.78	–0.07	0.02	*10.21*	0.02
14	0.32	0.15	2.12	0.34	0.02	*10.00*	0.01
15	0.34	0.15	2.17	0.04	0.02	*10.12*	0.00
16	0.29	0.16	1.78	–0.38	0.02	*9.02*	0.01
17	0.33	0.16	2.00	0.21	0.02	*9.51*	0.00
18	0.34	0.17	1.98	–0.01	0.02	*9.04*	0.00
19	0.32	0.17	1.80	–0.18	0.02	*8.65*	0.01
20	0.27	0.18	1.50	–0.30	0.02	*8.03*	0.01
21	0.25	0.18	1.35	–0.14	0.02	*7.53*	0.02
22	0.26	0.19	1.36	0.00	0.02	*7.90*	0.01
23	0.24	0.19	1.23	–0.12	0.02	*7.59*	0.02
24	0.20	0.20	1.00	–0.22	0.02	*7.27*	0.03
25	0.20	0.20	0.96	–0.04	0.02	*7.03*	0.04
26	0.15	0.21	0.73	–0.22	0.02	6.43	0.06
27	0.12	0.21	0.55	–0.18	0.02	5.96	0.09
28	0.13	0.22	0.60	0.04	0.02	5.86	0.08
29	0.14	0.22	0.62	0.02	0.02	5.94	0.07
30	0.12	0.23	0.52	–1.66	0.02	5.65	0.09
40	0.18	0.26	0.68	0.15	0.02	5.94	0.05
50	0.14	0.29	0.48	–0.19	0.02	4.94	0.07

Note: All intercept estimates are statistically significant at 1%. Values in italics are those that are 5% statistically significant.

In the context of U.S. metropolitan areas (Table 6), the granular hypothesis exhibits stronger explanatory power. In particular, for *K* values from 1 to 13, the R^2/population ratio exceeds one. The New York-Northern New Jersey-Long Island metropolitan area stands out with an impressive adjusted R^2 of 0.51, despite accounting for only 6.12% of the population. The impact of New York City is less evident in county-level analysis due to its division into five counties, with Queens ranking as the tenth largest. However, at K = 6 and K = 7, which correspond to the Philadelphia-Camden-Wilmington and Washington-Arlington-Alexandria metro areas, the R^2/population ratio increases significantly. Combining these three metro areas with the first New York-Northern New Jersey-Long Island area yields an astonishing adjusted R^2 of 0.65, representing 9.84% of the population. These findings suggest that the American Northeast plays a crucial role in explaining the U.S. business cycle and has a substantial growth spillover effect. In contrast, the inclusion of the second largest metro area, Los Angeles-Long Beach-Santa Ana, reduces the adjusted R^2 to 0.61, indicating that Los Angeles County may bias the granular residual and weaken its explanatory power. Comparing Tables 5 and 6, metropolitan areas demonstrate a stronger association with the granular residual than counties, potentially indicating a closer connection between the growth of U.S. central cities and their surroundings.

Table 6. The granular residual of the U.S. metropolitan areas.

K	R2	Metro area population/ Country population	R2/population ratio	Growth rate	Intercept β_1	Estimated granular residual	p-value
1	0.51	0.06	8.46	-	0.01	-11.50	0.00
2	0.51	0.10	4.99	-3.47	0.01	-10.19	0.00
3	0.49	0.13	3.73	-1.25	0.01	-8.75	0.00
4	0.46	0.15	3.02	-0.71	0.00	-8.26	0.00
5	0.41	0.17	2.40	-0.61	0.00	-7.38	0.00
6	0.50	0.19	2.59	0.19	0.00	-7.63	0.00
7	0.56	0.21	2.66	0.06	0.00	-7.38	0.00
8	0.45	0.22	1.98	-0.67	0.00	-6.73	0.00
9	0.39	0.24	1.61	-0.36	0.00	-6.04	0.00
10	0.41	0.26	1.59	-0.02	0.00	-5.69	0.00
11	0.41	0.27	1.48	-0.10	0.00	-5.27	0.00
12	0.33	0.28	1.15	-0.32	0.00	-4.71	0.00
13	0.26	0.30	0.87	-0.28	0.00	-4.44	0.01
14	0.21	0.31	0.67	-0.19	0.00	-4.01	0.03
15	0.20	0.32	0.63	-0.04	0.00	-3.94	0.03
16	0.21	0.33	0.63	-0.00	0.00	-3.90	0.03
17	0.20	0.34	0.59	-0.03	0.00	-3.83	0.03
18	0.19	0.35	0.53	-0.05	0.00	-3.66	0.04
19	0.22	0.36	0.61	0.07	0.00	-3.79	0.03
20	0.24	0.37	0.65	0.04	0.00	-3.81	0.02
21	0.25	0.38	0.65	-0.93	0.00	-3.73	0.02
22	0.27	0.39	0.70	0.04	0.00	-3.74	0.01
23	0.26	0.39	0.66	-0.03	0.00	-3.68	0.01
24	0.27	0.40	0.67	0.00	0.00	-3.66	0.01
25	0.24	0.41	0.58	-0.08	0.00	-3.47	0.02
30	0.22	0.44	0.51	-0.07	0.00	-3.16	0.02
40	0.20	0.50	0.41	-0.10	0.00	-2.76	0.03
50	0.24	0.54	0.44	0.03	0.00	-2.69	0.02
100	0.21	0.65	0.33	-0.11	0.00	-2.06	0.03
150	0.23	0.71	0.32	-0.00	0.00	-1.93	0.02
200	0.25	0.76	0.33	0.01	0.00	-1.90	0.02
250	0.25	0.79	0.32	-0.01	0.00	-1.83	0.02
300	0.26	0.81	0.33	0.00	-0.00	-1.80	0.01
357(all)	0.28	0.83	0.34	0.01	-0.00	-1.81	0.01

Note: All intercept estimates are statistically significant at 1%.

Analyzing data from Brazil between 2002 and 2019 (Table 7), the granular hypothesis remains unrefuted. Among the 281 municipalities with over 100,000 people on average during this period, the adjusted R^2 exceeds the population ratio for K values of 2, 3, 4, 5, 7, and 10. Consequently, we cannot dismiss the granular hypothesis when considering Brazilian municipalities, and there is a possibility of growth spillover from larger cities to the rest of the country. The next step involves identifying these granular municipalities.

To examine the correlation between municipalities within the same Brazilian geographic regions, we selected the five largest municipalities from each region: Curitiba

(South), Sao Paulo (Southeast), Brasilia (Midwest), Salvador (Northeast), and Manaus (North). Surprisingly, the adjusted R^2 of −0.01 was lower than all other cases in Table 7, indicating that the regional aspect does not significantly impact the granular effect of large municipalities. However, some municipalities, like Rio, Brasilia, Manaus, Recife, and Porto Alegre, displayed lower adjusted R^2 values (0.02) compared to others. In particular, including Sao Paulo and increasing *K* to 6 resulted in an impressive adjusted R^2 of 0.40 for a national population share of only 13%, underscoring Sao Paulo's crucial role in the Brazilian business cycle and growth spillover.

Table 7. The granular residual of the Brazilian municipalities.

K	R^2	*Metro area population/ Country population*	R^2*/population ratio*	*Growth rate*	*Intercept* β_1	*Estimated granular residual*	p-*value*
1	0.51	0.06	8.46	–	0.01	−11.50	0.00
2	0.51	0.10	4.99	−3.47	0.01	−10.19	0.00
3	0.49	0.13	3.73	−1.25	0.01	−8.75	0.00
4	0.46	0.15	3.02	−0.71	0.00	−8.26	0.00
5	0.41	0.17	2.40	−0.61	0.00	−7.38	0.00
6	0.50	0.19	2.59	0.19	0.00	−7.63	0.00
7	0.56	0.21	2.66	0.06	0.00	−7.38	0.00
8	0.45	0.22	1.98	−0.67	0.00	−6.73	0.00
9	0.39	0.24	1.61	−0.36	0.00	−6.04	0.00
10	0.41	0.26	1.59	−0.02	0.00	−5.69	0.00
11	0.41	0.27	1.48	−0.10	0.00	−5.27	0.00
12	0.33	0.28	1.15	−0.32	0.00	−4.71	0.00
13	0.26	0.30	0.87	−0.28	0.00	−4.44	0.01
14	0.21	0.31	0.67	−0.19	0.00	−4.01	0.03
15	0.20	0.32	0.63	−0.04	0.00	−3.94	0.03
16	0.21	0.33	0.63	−0.00	0.00	−3.90	0.03
17	0.20	0.34	0.59	−0.03	0.00	−3.83	0.03
18	0.19	0.35	0.53	−0.05	0.00	−3.66	0.04
19	0.22	0.36	0.61	0.07	0.00	−3.79	0.03
20	0.24	0.37	0.65	0.04	0.00	−3.81	0.02
21	0.25	0.38	0.65	−0.93	0.00	−3.73	0.02
22	0.27	0.39	0.70	0.04	0.00	−3.74	0.01
23	0.26	0.39	0.66	−0.03	0.00	−3.68	0.01
24	0.27	0.40	0.67	0.00	0.00	−3.66	0.01
25	0.24	0.41	0.58	−0.08	0.00	−3.47	0.02

30	0.22	0.44	0.51	−0.07	0.00	−3.16	0.02
40	0.20	0.50	0.41	−0.10	0.00	−2.76	0.03
50	0.24	0.54	0.44	0.03	0.00	−2.69	0.02
100	0.21	0.65	0.33	−0.11	0.00	−2.06	0.03
150	0.23	0.71	0.32	−0.00	0.00	−1.93	0.02
200	0.25	0.76	0.33	0.01	0.00	−1.90	0.02
250	0.25	0.79	0.32	−0.01	0.00	−1.83	0.02
300	0.26	0.81	0.33	0.00	−0.00	−1.80	0.01
357(all)	0.28	0.83	0.34	0.01	−0.00	−1.81	0.01

Note: All intercept estimates are statistically significant at 1%. Values in italics are statistically significant at 5%.

In Table 8, we shift our focus to Brazilian metropolitan areas and "urban agglomerations" as analysis units. These agglomerations represent extensions of central cities, encompassing neighboring municipalities with shared social and economic relations, urbanization, commuting patterns, and contiguity. However, the results indicate that none of the regressions' adjusted R^2 values surpass the percentage of the population residing in the metro area. As a consequence, we must reject the granular hypothesis for Brazilian metropolitan areas. Additionally, the β_2 value is not statistically significant at 10% for any *K*, emphasizing the limited relevance of the granular residual in explaining the business cycle within these metro areas.

Table 8. The granular residual of the Brazilian metropolitan areas.

K	R^2	*Metro area population/ Country population*	*R^2/population ratio*	*Growth rate*	*Intercept β_1*	*Estimated granular residual*	p-*value*
1	−0.05	0.10	−0.49	–	0.08	−1.86	0.64
2	0.13	0.16	0.78	1.27	0.08	−4.24	0.08
3	0.00	0.19	0.01	−0.77	0.08	−2.42	0.32
4	0.00	0.21	0.04	0.03	0.08	−2.28	0.30
5	0.03	0.23	0.14	0.10	0.08	−2.42	0.22
6	0.02	0.25	0.08	−0.06	0.08	−2.05	0.26
7	0.02	0.27	0.08	0.00	0.08	−2.00	0.25
8	0.04	0.29	0.14	0.05	0.08	−1.94	0.21
9	0.01	0.31	0.03	−0.10	0.08	−1.59	0.29
10	0.01	0.32	0.05	0.02	0.08	−1.60	0.27
11	0.01	0.34	0.04	−0.00	0.08	−1.55	0.27
12	0.02	0.35	0.07	0.02	0.08	−1.62	0.24
13	0.02	0.36	0.05	−0.01	0.08	−1.54	0.26
14	0.01	0.37	0.05	−0.00	0.08	−1.49	0.26
15	0.01	0.38	0.03	−0.01	0.08	−1.43	0.28
30	0.01	0.48	0.03	−0.00	0.08	−1.29	0.27
50	0.00	0.54	0.01	−0.02	0.08	−1.10	0.31
82	0.02	0.57	0.03	0.02	0.08	−1.24	0.26

Note: All intercept estimates are statistically significant at 1%.

It is worthwhile to make a comment on the finding of some negative R^2 values. In the U.S., negative adjusted R^2 values were observed for certain *K* values (number of largest cities included) at the county level, indicating weak explanatory power of the hypothesis for these areas. However, U.S. metropolitan areas generally showed positive, significant R^2 values, supporting the hypothesis in these regions. Brazilian municipalities presented mixed results. While some *K* values showed positive R^2 values, suggesting growth spillovers from larger cities, metropolitan areas did not support the granular hypothesis. No regression for these areas exceeded their population percentages in R^2 values, leading to the hypothesis's rejection.

The negative R^2 values in U.S. counties and Brazilian metropolitan areas imply the granular hypothesis's limitations for these specific analyses. This might be due to: 1) Varying levels of economic integration and dynamics across counties, municipalities, and metropolitan areas. Large cities and counties may exhibit unique economic patterns not captured by the hypothesis. 2) Large cities' distinct economic features and challenges, potentially leading to negative R^2 values if these specificities are not accounted for in the model. 3) Possible limitations of the methodology used to test the granular hypothesis, especially in capturing complex economic dynamics at the city level.

The difference in the applicability of the granular hypothesis between Brazilian municipalities and metropolitan areas might be due to the degree of aggregation. As aggregation increases, the granular residual tends to regress to the mean following the law of large numbers. In simpler terms, when calculating the residual for the entire metropolitan area, the impact of the central city's idiosyncratic shock gets diluted within the residual. To test the hypothesis, we used OLS regression, with the growth rate of metropolitan areas (excluding the central city) as the variable to be explained and the growth rate of the central city as the explanatory variable. Regression directly on the residual indicates a positive granular effect, given the larger population of the metropolitan area compared to the central city. As anticipated, the coefficient estimated for the growth rate is positive but less than one, showing a positive yet imperfect correlation, supporting regression to the mean. We excluded areas with central cities having less than 100,000 inhabitants on average, reducing observations from 82 to 53. Table 9 confirms that the hypothesis is not rejected. Generally, variations in central cities' per capita GDP growth rate have limited explanatory power for their periphery's growth rate (low R^2), as the β_2 value is as expected, positive and less than one.

Table 9. Regressing the growth rate of the largest municipalities in the periphery by the growth rate of their respective central cities, 2003-2019.

Central city	R^2	*Growth rate*	*p-value*
Average of the 53 metro areas	0.39	0.71	0.00
Sao Paulo	0.52	0.99	0.00
Rio	0.13	0.43	0.07
Salvador	−0.02	−0.94	0.43
Brasilia	0.25	0.77	0.01
Fortaleza	0.23	0.60	0.02
Belo Horizonte	0.24	1.05	0.02
Manaus	0.38	0.99	0.00
Curitiba	0.60	1.08	0.00
Recife	0.76	0.79	0.00
Porto Alegre	0.08	0.52	0.12

When calculating the residual solely for other municipalities in the metropolitan areas (excluding central cities), the adjusted R^2 significantly reduces, often becoming negative. This indicates that these outskirts contribute minimally to the business cycle and have little to no growth spillover effects on the rest of the country (Table 10). In particular, the outskirts of large central cities are themselves highly populous. For instance, the surroundings of Sao Paulo account for over 4% of the Brazilian population and, if treated as a single municipality, would be the country's second-largest.

Table 10. Regressions with the granular residual of metropolitan area municipalities, excluding the central cities.

K	R^2	*Intercept* β_1
1	0.00	0.08
2	−0.02	0.08
3	−0.05	0.08
4	−0.06	0.08
5	−0.06	0.08
6	−0.06	0.08
7	−0.06	0.08
8	−0.06	0.08
9	−0.06	0.08
10	−0.06	0.08
20	−0.05	0.08
30	−0.06	0.08
40	−0.05	0.08
53	−0.05	0.08

In Gabaix's (2011) methodology, the granular residual is a sum, preserving all data and potentially leading to biases and reduced explanatory power in the metro areas' granular residual. Consequently, the granular hypothesis is rejected for Brazilian metro areas, unlike large municipalities, which are central cities with a higher proportion of the national population. By adopting multiple regression methodology, we can discard irrelevant data, such as metro areas with a negative marginal effect on the R^2/population ratio. For instance, the exclusion of the Sao Paulo metropolitan area, which weakens the granular residual's explanatory power, is demonstrated in Table 11, which showcases metro areas where the R^2/population ratio consistently exceeds one in all specifications.

Table 11. Brazilian metropolitan areas' debiased granular residual.

K	R^2	*Metro area population/ Country population*	R^2/ *population ratio*	*Growth rate*	*Intercept* β_1	*Estimated granular residual*	p-*value*
Rio	0.19	0.06	3.04	–	0.08	−6.78	0.04
+ Porto Alegre	0.23	0.08	2.77	−0.26	0.08	−6.03	0.04
+ Brasilia	0.19	0.10	1.83	−0.94	0.08	−5.84	0.02
+ Fortaleza	0.21	0.12	1.73	−0.10	0.08	−5.27	0.03
+ Salvador	0.20	0.14	1.44	−0.29	0.08	−4.24	0.03
+ Campinas	0.21	0.15	1.36	−0.07	0.08	−4.12	0.03
+ Vale do Paraíba	0.23	0.17	1.36	−0.00	0.08	−4.10	0.02

Note: + means a metro area is added to the previous one.

Granular size

In Figure 1, the granular size for U.S. counties, based on equation (3), is illustrated. The granular (weighted) curve and equally weighted curve intersect at $L = 5$, determining the value $K^* = 5$. This refers to Los Angeles, Cook, Harris, Maricopa, and San Diego Counties, which represent about 8.4% of the U.S. population on average over the period but account for 48% of GDP growth (Table 5).

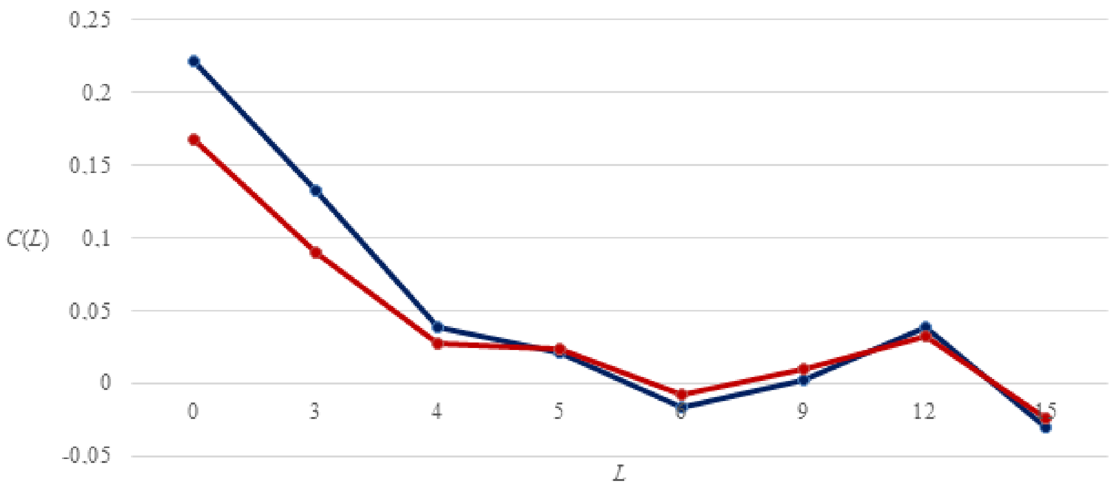


Figure 1. Granular size for the U.S. counties: granular curve (blue) and equal-weight curve (red), 2002-2019.

Equation (3) reveals that the granular size for U.S. metropolitan areas occurs at L = 3 (Figure 2). The value K^* = 3 corresponds to New York-Northern New Jersey-Long Island, Los Angeles-Long Beach-Santa Ana, and Chicago-Joliet-Naperville. These three metro areas represent around 13.4% of the American population but explain 49% of GDP growth (Table 6). As anticipated, due to higher population density in the largest metropolitan areas compared to counties, the granular size for metro areas is lower.

The top three metropolitan areas with the most significant granular effect are situated in different geographic regions of the United States. Interestingly, similar to the counties, none of these granular metro areas are capital cities (e.g., Albany, Springfield, and Sacramento). This aligns with the idea that there is a positive correlation between political power and human agglomerations, as proposed by Ades and Glaeser (1994).

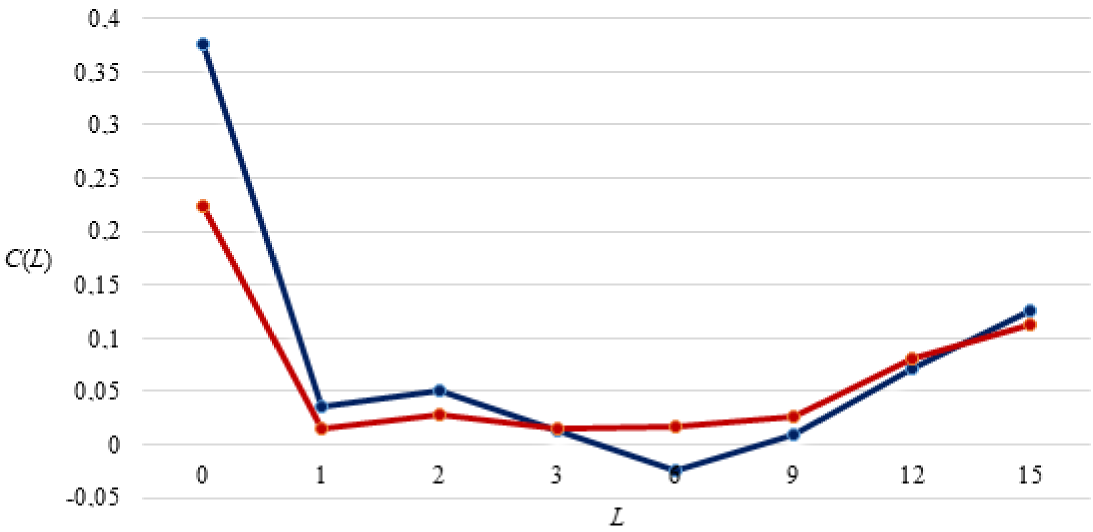


Figure 2. Granular size for the U.S. metro areas: granular curve (blue) and equal-weight curve (red), 2002-2019.

New York City, the largest municipality in the U.S., does not have a significant granular effect among counties, but it becomes relevant within the greater metropolitan area. The adjusted R^2 drops from nearly 0.4 with L = 0 (considering New York-Northern New Jersey-Long Island), to just below 0.05 with L = 1 (when this metro area is excluded). However, the increased importance of the greater Los Angeles area indicates that the growth of Los Angeles and its surroundings appears to be unrelated to that of the rest of the country.

When applying equation (3) to Brazilian municipalities, the granular and equal-weight curves intersect around L = 3 (Figure 3), implying a granular size K^* = 3. This indicates that idiosyncratic shocks in Sao Paulo, Rio, and Salvador significantly impact the economy due to their higher relative weight. Despite being just three municipalities, they represent 10% of the Brazilian population and explain 12% of GDP growth (Table 7). Notably, Sao Paulo alone contributes about 6% of the population and 13% of the GDP. When excluding the three largest grains, the average adjusted R^2 for K is no longer influenced by assigned weights (relative size in the national population).

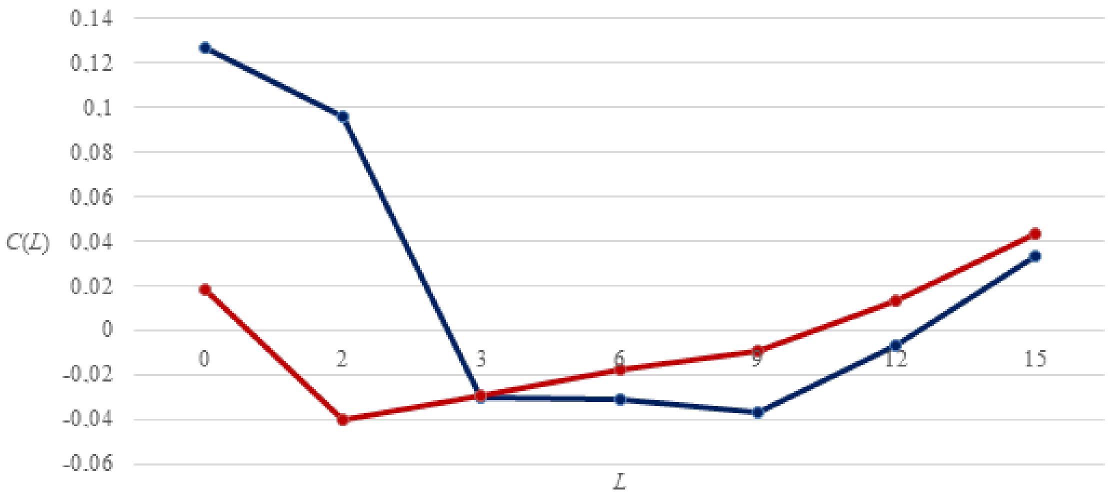


Figure 3. Granular size for Brazilian municipalities: granular curve (blue) and equal-weight curve (red), 2002-2019.

Table 12 displays the R^2 for each municipality's regressions using the granular residual, indicating significant variation in explanatory power. Fortaleza, Sao Bernardo do Campo, Joao Pessoa, Uberlandia, Ribeirao Preto, and Sorocaba have relatively high adjusted R^2 values. Apart from Fortaleza, these municipalities rank above 20, suggesting that as L increases, their relative weight in the granular residual composition grows, leading to higher explanatory power. This phenomenon applies to the equal-weight curve as well, as municipalities with strong explanatory power enter more regressions with increasing L. However, both the granular and equal-weight curves grow at similar rates, with the equal-weight curve slightly higher. This implies that the relative weights (percentage of the Brazilian population) of these municipalities do not contribute significantly to the granular residual's increased explanatory power. Instead, they exhibit a strong correlation with national growth, explaining the upward trajectory of the granular and equal-weight curves.

For Brazilian metropolitan areas, using equation (2) to calculate K^* is not applicable, as discussed earlier. Additionally, after debiasing the sample, determining K^* becomes unfeasible, as the debiasing process removes the largest grains. Without population shares, very small areas carry relatively higher weight, skewing the results. The equal-weight curve consistently exceeds the granular curve, mainly due to the high correlation of certain small metropolitan areas with national growth. This makes it impossible to ascertain the granular size, K^*, using the current methodology.

Table 12. Granular residual regressions for individual municipalities.

Rank	*Municipality*	R^2
1	Sao Paulo	0.05
2	Rio de Janeiro	–0.05
3	Salvador	0.05
4	Brasilia	0.06
5	Fortaleza	0.20
6	Belo Horizonte	–0.06
7	Manaus	–0.06
8	Curitiba	0.00
9	Recife	–0.00
10	Porto Alegre	–0.04
11	Belem	–0.03
12	Goiania	–0.05
13	Guarulhos	0.04
14	Campinas	–0.02
15	Sao Luis	–0.04
16	Sao Goncalo	0.09
17	Maceio	–0.04
18	Duque de Caxias	0.02
19	Natal	–0.01
20	Nova Iguacu	0.05
21	Teresina	–0.06
22	Campo Grande	–0.01
23	Sao Bernardo do Campo	0.17
24	Joao Pessoa	0.21
25	Osasco	0.03
26	Santo Andre	–0.06
27	Jaboatao dos Guararapes	–0.06
28	Sao Jose dos Campos	0.03
29	Uberlandia	0.11
30	Contagem	–0.06
31	Ribeirao Preto	0.20
32	Sorocaba	0.22
33	Feira de Santana	0.09
34	Aracaju	–0.04
35	Cuiaba	–0.04
36	Juiz de Fora	–0.02
37	Joinville	–0.03
38	Londrina	–0.06
39	Ananindeua	–0.06

3.4 Discussion

Although the granular size for U.S. counties (K^* = 5) is higher than for Brazilian municipalities (K^* = 3), the proportion of the population in the U.S. case is lower. This means granular municipalities in Brazil are larger than granular counties in the United States. Even the largest American municipality, New York City, and the largest county, Los Angeles County, only represent approximately 3% of the total U.S. population. In contrast, Sao Paulo accounts for about 6% of the Brazilian population.

Zipf's law, being a statistical phenomenon with no underlying causes (Mandelbrot, 1961), could account for random differences between the two countries. However, Ades and Glaeser (1994) present a compelling causal argument linking political factors to urban concentration, not vice versa. They argue that in more authoritarian countries with less economic freedom, population tends to concentrate around political poles like national or state capitals. This political power concentration also corresponds to income concentration, attracting the poorer populations to these large centers and leading to higher overall population concentrations.

Applying this causal narrative to Brazil, we find two granular municipalities in the Southeast, situated 450 kilometers apart. In contrast, only the Northeast lacks granular counties in the United States, possibly due to historical reasons. Economic Freedom Index rankings for 2022 place the United States 25th and Brazil 133rd, suggesting a higher concentration of the Brazilian population near centers of power. Remarkably, all three granular municipalities in Brazil are state capitals, and even a relatively new municipality like Brasilia, founded in 1960, already houses about 1.5% of the Brazilian population, surpassing Washington, D.C., established in 1791, which has only around 0.2% of the U.S. population. This points to a higher spatial concentration of economic activity in Brazil compared to the United States, potentially linked to political factors. The granular size serves as a measure of concentration, with fewer grains indicating greater economic importance in a region.

Moreover, the Sun Belt accommodates four out of the five granular counties in the United States, supporting the hypothesis that the warmest U.S. regions in January act as significant population magnets, experiencing faster growth rates than the national economy (Glaeser and Gottlieb, 2009). This population analysis also clarifies why U.S. cities have a greater impact on the business cycle than Brazilian cities, considering their relative sizes.

The findings align with previous literature on granular firms (Gabaix, 2011). In particular, granular city size is much smaller than firm city size, with the largest cities contributing more to GDP than the largest companies. For instance, in Brazil, Sao Paulo accounts for 6% of the population and 13% of GDP, while the largest company, Petrobras, only contributes 4% of GDP (Silva and Da Silva, 2018). Handling city granularity is akin to dealing with mega-grains. Interestingly, while granular firms in emerging economies have a greater impact on the business cycle (Grigoli et al., 2023), the opposite is true for granular cities. Our results show that large city grains explain a smaller percentage of the business cycle in Brazil (an emerging economy) compared to the United States.

We believe that the impact of granular cities on the business cycle varies between emerging and developed economies due to several factors: 1) Economic activities in emerging economies are often concentrated in a few large firms, significantly affecting the business cycle through their substantial GDP and employment share. In contrast, these activities are more dispersed across cities, reducing their influence on the business cycle compared to developed economies, where cities have more economic concentration. 2) Developed economies have urban economies diversified across multiple industries and advanced services, amplifying large cities' impact on the national business cycle. Emerging economies, with urban economies centered around fewer sectors or dominated by a single firm, experience less influence from their large cities. 3) Cities in developed economies, better integrated into the global economy with advanced infrastructure, significantly impact the business cycle through trade, finance, and technology. In emerging economies, less integration, both domestically and internationally, limits cities' influence. 4) Developed economies usually have more stable and effective governance, enhancing cities' role in the business cycle. In contrast, emerging economies may struggle with less efficient urban governance and infrastructure deficits, limiting their cities' economic impact. In summary, the differing impacts of cities on the business cycle in emerging versus developed economies may stem from economic diversification, structural urban differences, levels of integration and connectivity, and governance effectiveness.

3.5 Conclusion

Large cities play a crucial role in the business cycle, as they are home to granular firms, which are primarily responsible for driving it. This study explores the concept of granularity, extending it from firms to cities, and investigates how granular cities influence the business cycle. Our contribution is to highlight a spatial component of granularity not considered so far. The analysis is based on data from 2003 to 2019, focusing on cities in the

United States and Brazil, where we observe that city size distributions adhere to Zipf's law. By computing the granular residual, we identify the granular size for these cities.

The granular size for counties in the U.S. is five, representing Los Angeles, Cook, Harris, Maricopa, and San Diego Counties. These five counties, comprising 8% of the U.S. population, contribute to 48% of GDP growth. Similarly, the granular size for metropolitan areas is three, including New York-Northern New Jersey-Long Island, Los Angeles-Long Beach-Santa Ana, and Chicago-Joliet-Naperville. These areas, accounting for 13% of the American population, explain 49% of GDP growth. In Brazil, the granular size is three, with municipalities representing 10% of the population and explaining 12% of GDP growth. Notably, Sao Paulo, accounts for 6% of the population and 13% of GDP. Therefore, we could not reject the hypothesis that cities in the United States and Brazil explain a greater proportion of the business cycle than their relative size.

Conventional analyses of the business cycle focus on national or regional levels, but examining it at the city-level offers deeper insights into local economic dynamics. This approach holds practical value for policymakers, urban planners, and businesses. The discovery that cities wield significant influence on the business cycle, beyond their size, has vital implications. It allows us to pinpoint cities' economic strengths and weaknesses, facilitating targeted policies for growth and resilience. Moreover, it aids in assessing regional imbalances, enabling more effective resource allocation to address disparities and promote balanced development.

Appendix 3.1

Table A1. The average population of the most populous U.S. counties from 2002 to 2019.

Rank	*County*	*Population*
1	Los Angeles	9,893,578
2	Cook	5,216,620
3	Harris	4,147,583
4	Maricopa	3,881,412
5	San Diego	3,118,644
6	Orange, CA	3,040,962
7	Miami-Dade	2,526,471
8	Kings	2,524,281
9	Dallas	2,416,129
10	Queens	2,242,563
11	Riverside	2,163,119
12	San Bernardino	2,031,639

13	King	1,976,515
14	Clark	1,939,196
15	Wayne	1,848,308
16	Tarrant	1,823,093
17	Broward	1,803,610
18	Santa Clara	1,800,432
19	Bexar	1,731,970
20	New York	1,601,516
21	Alameda	1,541,421
22	Philadelphia	1,533,884
23	Middlesex	1,524,906
24	Suffolk	1,483,503
25	Sacramento	1,430,306
26	Bronx	1,394,218
27	Palm Beach	1,345,861
28	Nassau	1,343,255
29	Cuyahoga	1,287,895
30	Hillsborough	1,259,067
31	Allegheny	1,230,892
32	Oakland	1,221,839
33	Franklin	1,187,835
34	Orange, FL	1,176,786
35	Hennepin	1,174,902
36	Contra Costa	1,063,088
37	Travis	1,051,423
38	Salt Lake	1,037,265
39	St Louis	1,001,204
40	Montgomery	979,754
41	Pima	973,391
42	Honolulu	952,012
43	Westchester	951,521
44	Milwaukee	945,2
45	Pinellas	934,915
46	Shelby	926,355
47	Fresno	926,164
48	Fulton	925,552
49	Erie	924,501
50	Fairfield	920,891
51	DuPage	920,738

Table A2. The average population of the most populous U.S. metropolitan areas from 2002 to 2019.

Rank	*Metropolitan area*	*Population*
1	New York-Northern New Jersey-Long Island	18,924,664
2	Los Angeles-Long Beach-Santa Ana	12,876,591
3	Chicago-Joliet-Naperville	9,403,216

4	Dallas-Fort Worth-Arlington	6,403,882
5	Houston-Sugar Land-Baytown	5,969,239
6	Philadelphia-Camden-Wilmington	5,941,397
7	Washington-Arlington-Alexandria	5,627,589
8	Miami-Fort Lauderdale-Pompano Beach	5,617,024
9	Atlanta-Sandy Springs-Marietta	5,227,443
10	Boston-Cambridge-Quincy	4,614,588
11	San Francisco-Oakland-Fremont	4,413,892
12	Detroit-Warren-Livonia	4,340,234
13	Phoenix-Mesa-Glendale	4,156,695
14	Riverside-San Bernardino-Ontario	4,110,535
15	Seattle-Tacoma-Bellevue	3,490,696
16	Minneapolis-St. Paul-Bloomington	3,313,846
17	San Diego-Carlsbad-San Marcos	3,107,209
18	Tampa-St. Petersburg-Clearwater	2,791,786
19	St. Louis	2,780,826
20	Baltimore-Towson	2,706,133
21	Denver-Aurora-Broomfield	2,577,039
22	Pittsburgh	2,368,081
23	Portland-Vancouver-Hillsboro	2,229,138
24	San Antonio-New Braunfels	2,150,906
25	Orlando-Kissimmee-Sanford	2,141,098
26	Sacramento-Arden Arcade-Roseville	2,126,362
27	Cincinnati-Middletown	2,120,806
28	Cleveland-Elyria-Mentor	2,086,535
29	Kansas City	2,014,011
30	Charlotte-Gastonia-Rock Hill	1,986,477
31	Las Vegas-Paradise	1,887,074
32	Columbus	1,873,479
33	San Jose-Sunnyvale-Santa Clara	1,867,850
34	Indianapolis-Carmel	1,812,350
35	Austin-Round Rock-San Marcos	1,740,875
36	Virginia Beach-Norfolk-Newport News	1,686,641
37	Nashville-Davidson-Murfreesboro-Franklin	1,625,721
38	Providence-New Bedford-Fall River	1,602,457
39	Milwaukee-Waukesha-West Allis	1,550,643
40	Jacksonville	1,343,183
41	Memphis	1,298,987
42	Oklahoma City	1,261,386
43	New Orleans-Metairie-Kenner	1,242,908
44	Louisville-Jefferson County	1,235,659
45	Richmond	1,214,209
46	Hartford-West Hartford-East Hartford	1,196,596
47	Buffalo-Niagara Falls	1,142,204
48	Raleigh-Cary	1,119,514
49	Salt Lake City	1,108,956

50	Birmingham-Hoover	1,086,659
51	Rochester	1,060,962

Table A3. The average population of the most populous Brazilian municipalities from 2002 to 2019.

Rank	*City*	*Population*
1	Sao Paulo	11,384,688
2	Rio de Janeiro	6,292,483
3	Salvador	2,780,167
4	Brasilia	2,612,834
5	Fortaleza	2,470,872
6	Belo Horizonte	2,421,608
7	Manaus	1,832,282
8	Curitiba	1,804,438
9	Recife	1,554,811
10	Porto Alegre	1,439,941
11	Belem	1,414,073
12	Goiania	1,314,843
13	Guarulhos	1,269,201
14	Campinas	1,097,047
15	Sao Luis	1,015,875
16	Sao Goncalo	998,588
17	Maceio	945,037
18	Duque de Caxias	860,655
19	Natal	816,39
20	Nova Iguacu	812,116
21	Teresina	810,6791
22	Campo Grande	792,935
23	Sao Bernardo do Campo	792,601
24	Joao Pessoa	721,941
25	Osasco	692,004
26	Santo Andre	686,246
27	Jaboatao dos Guararapes	660,396
28	Soo Jose dos Campos	639,016
29	Uberlandia	618,664
30	Contagem	616,581
31	Ribeirao Preto	605,768
32	Sorocaba	598,818
33	Feira de Santana	569,983
34	Aracaju	566,741
35	Cuiaba	555,092
36	Juiz de Fora	524,744
37	Joinville	520,189
38	Londrina	515,595
39	Ananindeua	486,143

40	Niteroi	485,986
41	Belford Roxo	481,33
42	Aparecida de Goiania	477,227
43	Sao Joao de Meriti	462,43
44	Campos dos Goytacazes	456,189
45	Caxias do Sul	437,092
46	Porto Velho	434,222
47	Vila Velha	428,659
48	Florianopolis	428,553
49	Serra	427,622
50	Maua	425,641
51	Santos	424,635

Table A4. The average population of the most populous Brazilian metropolitan areas (and "urban agglomerations") from 2002 to 2019.

Rank	*Metropolitan area*	*Population*
1	Sao Paulo	20,136,062
2	Rio de Janeiro	12,305,504
3	Belo Horizonte	5,593,040
4	Porto Alegre	4,147,724
5	Brasilia	4,015,091
6	Recife	3,832,479
7	Fortaleza	3,778,207
8	Salvador	3,708,305
9	Curitiba	3,345,439
10	Campinas	2,867,801
11	Belem	2,319,885
12	Vale do Paraiba-Litoral Norte	2,315,771
13	Manaus	2,252,979
14	Goiania	2,348,813
15	Sorocaba	1,916,248
16	Vitoria	1,752,142
17	Baixada Santista	1,705,743
18	Ribeirao Preto	1,531,931
19	Sao Luiz	1,489,629
20	Natal	1,441,807
21	Piracicaba	1,364,692
22	Norte-Nordeste Catarinense	1,245,245
23	Maceio	1,203,398
24	Joao Pessoa	1,166,235
25	Teresina	1,122,175
26	Florianopolis	1,032,434
27	Londrina	1,019,096
28	Vale do Rio Cuiaba	954,421
29	Feira de Santana	867,61

30	Aracaju	842,863
31	Serra Gaucha	740,592
32	Maringa	728,172
33	Vale do Aco	727,964
34	Jundiai	709,223
35	Petrolina-Juazeiro	707,27
36	Vale do Itajai	700,422
37	Franca	618,188
38	Campina Grande	610,77
39	Sul	593,69
40	Cariri	568,528
41	Agreste	558,205
42	Carbonifera	553,636
43	Foz do Itajai	540,842
44	Macapa	524,154
45	Contestado	507,898
46	Cascavel	497,542
47	Sobral	462,478
48	Porto Velho	456,288
49	Chapeco	439,07
50	Palmas	405,406
51	Toledo	359,618

Table A5. Descriptive statistics, United States.

	Counties		*Metropolitan areas*	
	Population	*Growth rate*	*Population*	*Growth rate*
n	200	188	357	346
Mean	873,035	0.03	722,190	0.01
Variance	8.68×10^{11}	4.04×10^{-5}	2.57×10^{13}	9.37×10^{-5}
Assimetry	5.72	0.14	6.59	2.79
Kurtosis	46.35	2.67	58.69	0.09
Jarque-Bera	16756.33	1.48	48735.59	1.19

Note: population entries are in thousands.

Table A6. Descriptive statistics, Brazil.

	Municipalities		*Metropolitan areas*	
	Population	*Growth rate*	*Population*	*Growth rate*
n	281	272	82	79
Mean	378,289	0.09	1,362,042	0.09
Variance	7.09×10^{11}	0.0002	7.18×10^{6}	0.0001
Assimetry	9.59	2.83	5.18	-0.412
Kurtosis	112.97	0.24	31.85	4.95
Jarque-Bera	145910.43	2.97	3211.96	14.85

Note: population entries are in thousands.

Appendix 3.2

In this analysis, we test the granular hypothesis by using GDP from the previous year ($t - 1$) as the share for constructing the current year's (t) granular residual. This approach aligns with Gabaix's (2011) methodology, allowing comparison with population-based share analsysis. Although there is a strong GDP-population correlation (about 0.98 in the U.S. and 0.96 in Brazil), the findings using GDP as a share show notable differences from those using population. To streamline our analysis, we focus on the U.S. metropolitan areas and Brazilian municipalities, where population share has shown the most significant results.

In Table A7, the results for U.S. metropolitan areas indicate an increase in explanatory power compared to Table 6. However, the significance of the estimated coefficient for the granular residual is generally lower. This implies that in the U.S., using population as a weighting factor for grain size suggests large grains' impact on GDP is not solely due to their larger GDP. Therefore, the population's role in influencing the impact of large grains on the business cycle is significant.

Table A7. Granular residual in U.S. metropolitan areas, calculated using GDP share.

n	R^2	*Intercept*	*p-value*	*Granular residual*	*p-value*
10	0.59	0.01	0.01	−4.39	0.00
20	0.41	0.02	0.01	−3.23	0.00
30	0.40	0.01	0.04	−2.99	0.00
40	0.34	0.01	0.05	−2.68	0.01
50	0.35	0.01	0.05	−2.44	0.01
60	0.38	0.01	0.08	−2.47	0.00
70	0.39	0.01	0.10	−2.43	0.00
80	0.40	0.01	0.12	−2.37	0.00
90	0.40	0.01	0.12	−2.31	0.00
100	0.39	0.01	0.12	−2.21	0.00

In Brazil, findings mirror those of the U.S. Table A8 shows that for larger sample sizes ($n \geq 10$), using last year's GDP share yields less explanatory power if compared with those in Table 7, with the granular residual's coefficient becoming statistically nonsignificant at 5%. This reinforces the idea that large cities influence the business cycle not just via their GDP, but also through their population size and the geographical distribution of economic activity.

Table A8. Granular residual in Brazilian municipalities, calculated using GDP share.

n	R^2	*Intercept*	*p-value*	*Granular residual*	*p-value*
5	0.23	0.09	0.00	−3.70	0.03
10	0.17	0.09	0.00	−3.02	0.05
15	0.14	0.09	0.00	−2.71	0.08
20	0.16	0.09	0.00	−2.49	0.07
25	0.11	0.09	0.00	−2.28	0.10
30	0.12	0.09	0.00	−2.35	0.10
35	0.10	0.09	0.00	−2.25	0.11
40	0.13	0.09	0.00	−2.46	0.09
45	0.02	0.09	0.00	−1.95	0.26
50	0.02	0.09	0.00	−1.94	0.27

In our analysis of both cases, using population to calculate the granular residual revealed significant insights. It showed that population density and economic activity have a broader impact on the business cycle than just GDP alone.

Appendix 3.3

In this appendix, we estimate Pareto exponents through maximum likelihood and compare the power law distribution to an exponential distribution, also derived using maximum likelihood. Following Clauset et al. (2009)'s methodology, we calculated parameter estimates using group population averages for the periods analyzed, as discussed in Section 2.

The Pareto exponent is estimated using maximum likelihood, expressed as:

$$\hat{\alpha} \simeq 1 + n\left(\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min} - \frac{1}{2}}\right)^{-1}.$$

Here, $\hat{\alpha}$ is the estimated Pareto coefficient, n represents the number of observations, $x_{\min}$ is the smallest observation's value, and x_i is the value of each observation. The estimated power law distribution function, denoted as $P(x_i) = Cx_i^{-\hat{\alpha}}$, αconstant $C = (\hat{\alpha} - 1)x_{\min}^{\hat{\alpha}-1}$.

The exponential distribution we consider is denoted by $P(x_i) = Ce^{-\hat{\lambda}x_i}$, which includes the normalization constant $C = \hat{\lambda}e^{\hat{\lambda}x_{\min}}$. Thus, the estimated parameter $\hat{\lambda}$ determined via maximum likelihood is represented as:

$$\hat{\lambda} = \frac{n}{\left(\sum_{i=1}^{n} x_i\right) - n x_{\min}}.$$

Table A9 displays the calculated Pareto exponents for each of the four data subsets.

Table A9. Estimates of Pareto exponents by maximum likelihood.

Dataset	*n*	α	*S.D.*
U.S counties	50	−2.88	0.29
	100	−2.81	0.19
	500	−2.03	0.05
	1000	−1.87	0.03
U.S. metro areas	50	−2.16	0.18
	100	−2.05	0.12
	357	−1.56	0.45
Brazilian municipalities	50	−2.42	0.22
	100	−2.38	0.15
	500	−2.17	0.05
	1000	−2.13	0.04
Brazilian metro areas	50	−1.87	0.14
	82	−1.33	0.06

Note: all exponents are significant at the 1% level.

Note that estimates obtained via maximum likelihood significantly differ from those calculated using ordinary least squares, as shown in Table A10. However, the ratio of ordinary least squares to maximum likelihood estimates remains stable and similar across the four datasets.

Table A10. Comparing Pareto exponent estimates: maximum likelihood vs. ordinary least squares.

Dataset	*n*	*OLS*	*ML*	*OLS/ML*
U.S counties	50	−1.90	−2.88	0.66
	100	−1.88	−2.81	0.67
	500	−1.24	−2.03	0.61
	1000	−1.03	−1.87	0.55
U.S. metro areas	50	−1.42	−2.16	0.66
	100	−1.17	−2.05	0.57
	357	−0.92	−1.56	0.59
Brazilian municipalities	50	−1.36	−2.42	0.56
	100	−1.37	−2.38	0.58
	500	−1.23	−2.17	0.57
	1000	−1.17	−2.13	0.55
Brazilian metro areas	50	−1.08	−1.87	0.57
	82	−0.71	−1.33	0.53

Note: all exponents are significant at the 1% level.

Now we test if the datasets follow an exponential distribution rather than a power law distribution. After confirming through the Jarque-Bera test that the data are not normally distributed, we consider the exponential distribution as an alternative. This choice is reasonable because the exponential distribution also accounts for large value occurrences. Parameter λ estimates are presented in Table A11.

Table A11. Maximum likelihood estimates of parameter λ in the exponential distribution.

Dataset	*n*	λ
U.S counties	50	1.09×10^{-6}
	100	1.48×10^{-6}
	500	2.91×10^{-6}
	1000	4.54×10^{-6}
U.S. metro areas	50	4.43×10^{-7}
	100	6.62×10^{-7}
	357	1.50×10^{-6}
Brazilian municipalities	50	1.23×10^{-6}
	100	1.90×10^{-6}
	500	5.41×10^{-6}
	1000	8.93×10^{-6}
Brazilian metro areas	50	0.0006
	82	0.0007

The likelihood ratio test compares how well two distributions fit a specific data set. Table A12 presents the results of this test for our power law and exponential distributions. It reveals that only Brazilian metropolitan areas do not follow a power law distribution, explaining our rejection of the granular hypothesis for this dataset.

Table A12. Likelihood ratio test *R* results: power law vs. exponential distribution.

Dataset	*n*	*R*	*S.D.*	*p-value*
U.S counties	*50*	*4.93*	*4.46*	*0.1357*
	100	11.81	7.74	0.0643
	500	68.48	27.36	0.0063
	1000	258.92	50.89	0.0000
U.S. metro areas	*50*	*0.88*	*4.44*	*0.4247*
	100	17.19	9.04	0.0288
	357	44.16	30.47	0.0750
Brazilian municipalities	50	14.45	7.66	0.0301
	100	32.19	14.84	0.0151
	500	209.32	60.11	0.0003
	1000	498.39	108.57	0.0000
Brazilian metro areas	*50*	*7.51*	*7.60*	*0.1611*
	82	*–23.25*	*13.41*	*0.9581*

Note: Italics indicate data sets where the exponential distribution is more likely than the power law to describe the data.

4. Essay 3: Is the Brazilian labor market granular?

Abstract
This study explores the impact of large firms, often referred to as “big grains,” on hiring and firing cycles in the Brazilian labor market. We found strong support for the granular hypothesis. Our methodology involved analyzing the power-law distribution, granular residuals, and the granular size of the labor market. Key findings include the observation that firms exhibit a power-law distribution based on their workforce size, with large companies’ idiosyncratic shocks significantly influencing hiring and firing cycles. In particular, the service sector plays a substantial role in explaining these cycles, while manufacturing has limited explanatory power. We determined that the granular size of the Brazilian labor market consists of 15 firms engaged in public services, and private companies have a relatively minor impact on hiring and firing cycles. The policy implication here is that addressing periods of high unemployment in Brazil may be more effectively achieved by investing in public services rather than providing fiscal stimulus for manufacturing. This study contributes to the global body of evidence on labor market granularity and is compared with the existing research focused on Germany. We find that the Brazilian labor market is less granular than the German one.

4.1 Introduction

In 2023, the Brazilian government established a dedicated ministry to enhance the role of micro and small businesses in job creation, recognizing that these businesses contribute around 80% of job opportunities in Brazil. The question arises: Do these small businesses, rather than large corporations, drive the hiring and firing cycles in the Brazilian labor market?

International evidence suggests that job turnover is primarily observed in large, well-established companies, which significantly affects fluctuations in unemployment rates (Davis et al., 1996). This phenomenon can be explained by the concept of "granularity" (Kovalenko et al., 2020).

The granular hypothesis acknowledges the coexistence of a few major companies, or "grains," with numerous smaller ones, challenging the assumption that individual shocks from firms are diluted by the law of large numbers (Gabaix, 2011). This aligns with the observation of skewed firm size distributions, where shocks from disproportionately large firms persist. The concept of "granular residual," which captures idiosyncratic shocks weighted by size, allows us to quantify its impact on aggregate quantities through statistical methods like regressions and R2 statistics. It is crucial to calibrate for the optimal number of firms to avoid underestimating or exaggerating the granular residual (Blanco-Arroyo et al., 2018). This study applies these ideas to investigate whether the Brazilian labor market exhibits granularity.

Figure 1 depicts the cyclical nature of the Brazilian labor market in terms of formal job growth from 1996 to 2019, deviating from a linear trend. This raises the question: What factors contribute to this phenomenon? We consider three possibilities: 1) Cycles solely result from aggregate shocks affecting the entire economy uniformly; 2) Cycles stem from idiosyncratic shocks to small businesses; and 3) Cycles arise from idiosyncratic shocks to major corporations.

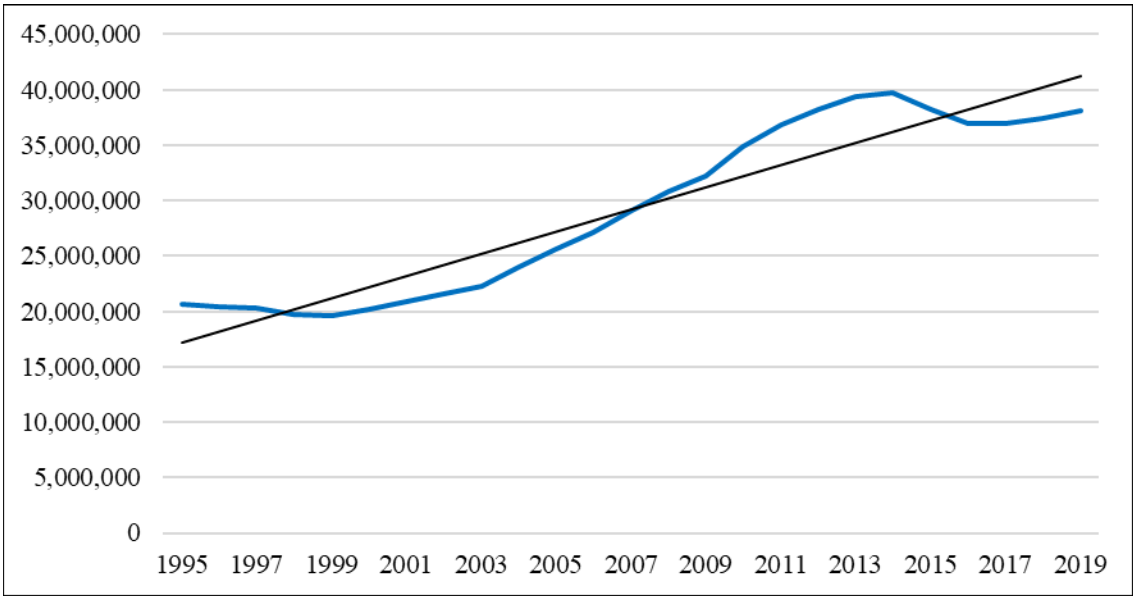


Figure 1. Formal job growth in Brazil in millions from 1996 to 2019.

The first possibility suggests that in a market with evenly-sized companies, negative shocks to some firms would be offset by positive shocks in others. However, data from previous studies (Da Silva et al., 2018; Silva and Da Silva, 2020) indicates that Brazilian businesses follow a power-law distribution, where larger firms have a more significant impact on the business cycle. This supports the application of the granularity hypothesis in analyzing hiring and firing cycles in the Brazilian labor market, similar to the approach taken by Kovalenko et al. (2020) in their study of the German labor market.

In this study, the authors examine whether specific sectors, such as industrial and service sectors, have a greater influence on hiring and firing cycles compared to the overall labor market. Additionally, they analyze the role of the labor market in private enterprises separately from public companies. The essay is organized as follows: firstly, the data and methodology used are provided; descriptive statistics are then presented; the results are reported; comparisons are made with previous studies; and finally, the essay concludes with concluding comments.

4.2. Materials and methods

From 1996 to 2019, we obtained data on the total number of employees per company from the RAIS (Annual Social Information List) database and calculated the total numberof formal occupations in Brazil using the CAGED (General Register of Employed and Unemployed) database, both provided by the Brazilian Ministry of Labor. Unlike Kovalenko et al. (2021), we do not have quarterly statistics, so we do not need to accountfor seasonality. Our data is aggregated by CNPJ, the National Register of Legal Businesses, an identity number for businesses of taxation relevance, including firms, partnerships, and foundations.

We applied the Gabaix and Ibragimov (2011) method to estimate coefficients andtest if employee numbers follow a power law. We used ordinary least squares to derive estimates with this equation:

$$\ln\left(\text{rank}_i - \tfrac{1}{2}\right) = a + \alpha \ln \frac{\text{employees}_i}{\text{employees}_m}, \tag{1}$$

where rank_i represents the company's position in the ranking, sorted from largest to smallest based on employees_i (the number of employees at the company i) and employees_m (the smallest number of employees among the companies in the sample).

To investigate the granular hypothesis regarding the influence of large companies on hiring and firing cycles, we calculate the granular residual Γ_t:

$$\Gamma_t = \sum_{i=1}^{K} \frac{\text{employees}_{it}}{\text{total employment}_t}\left(g_{it} - g_t\right). \tag{2}$$

Here, employees_{it} represents the number of employees at a specific firm i at time t, $\text{total employment}_t$ is the total population employment at time t, g_{it} refers to the growth rate of the number of employees at the largest firms in t, and g_t is the labor market's overall formal employment growth rate in t. We then use ordinary least squares regressionto analyze the relationship between the growth rate of formal employment and the granular residual.

Using a nonoptimal number of firms K in equation (2) can lead to an inaccurate estimation of a firm's contribution to hiring and firing cycles. Blanco-Arroyo et al. (2018) suggest a method for finding the granular size K^*. This involves comparing the explanatory power of a firm's granular residual by evaluating a weighted curve (the same equation (2)) against another curve with identical weights after making equation (2). The function $C(L)$'s "granular curve" is

$$C(L) = \frac{1}{Q}\sum_{K=1}^{Q} \mathrm{R}^2(K, L) . \tag{3}$$

Here, Q is an arbitrary number of firms. Our aim is to assess how R^2 reacts to the gradual exclusion of the largest firms by increasing L. We want to see how the curve performs aswe replace the top L firms in the sample with the $Q + 1, \ldots, Q + L$ following firms. For each L value, we run Q regressions using the granular residual (the curve with weights) as the explanatory variable. $C(L)$ represents the average R^2 for these Q regressions.

Furthemore, the equal-weight curve estimates the impact of shocks from equally-sized firms, expected to be minor. We anticipate observing a shift from the granular curve $C(L)$ towards the equal-weight curve as we remove the L largest firms from the granular residual. The granular size K^* corresponds to the L value where the curve $C(L)$ intersects the equal-weight curve for the first time.

To streamline the computation process and reduce the number of regressions to under 1,000, we start by assuming $Q = 40$. Then, we run regressions for the subset of firms L at intervals of ten ($L = 0, 10, 20$, and so on) until we observe the intersection of curves with and without weights. For each L value, we run regressions with variable K, incrementing by twenty up to 160 (i.e., $K = 20, 40, ..., 160$). When we identify a value of$K$ where $C(L)$ falls below the R^2 value obtained without weights, we calculate $C(L)$ and run regressions for intermediate L values to pinpoint the granular size.

The empirical model is defined as follows:

$$G_t = \beta_1 + \beta_2 \Gamma_t + \varepsilon_t \ . \qquad (4)$$

Here, G_t is the growth rate of the formal jobs in time period t, β_1 and β_2 are parameters estimated using ordinary least squares, as describeb in Gabaix (2011), where β_1 represents the average value of the growth rate of the number of jobs relative to the granular residual, β_2 is the sensitivity of the growth rate to the granular residual, and ε_t is the estimated error. The adjusted R^2, calculated for this model, quantifies how well the granular residual explains hiring and firing cycles.

4.3. Descriptive statistics

Table 1 shows that as the sample size grows from 50 to 1,000 firms, the average numberof employees decreases over the period from 1996 to 2019. A positive asymmetry valuesuggests that the sample is skewed to the left, meaning that smaller enterprises are closer to the mean. Additionally, kurtosis values exceeding three indicate leptokurtosis.

Table 1. Companies' descriptive statistics in relation to the number of employees.

Number of firms	*Average number of employees*	*Standard deviation*	*Asymmetry*	*Kurtosis*	*Jarque-Bera test*
50	50,298	66,881	4.80	27.42	1,434.49
100	31,105	51,053	6.30	48.78	9,394.81
200	19,236	38,007	8.43	89.46	64,661.46
500	10,387	25,106	12.62	205.63	868,676.91
1,000	6,534	18,168	17.24	389.94	6,288,095.91

We consider the Jarque-Bera test to reject the null hypothesis that the distributionof firms' number of employees is Gaussian. In all cases, the test yields much higher values than the critical value of 5.99, as it follows a chi-square distribution with two degrees offreedom.

Table 2 provides descriptive statistics for employee growth rates. The aggregate growth rates for the largest 50, 100, 200, and 500 enterprises are lower than that for the 1,000 largest firms. This suggests that job growth during the period primarily occurred in smaller firms, as mentioned earlier.

Table 2. Companies' descriptive statistics in relation to the growth rates in the number of employees.

Number of firms	*Growth average*	*Standard deviation*	*Asymmetry*	*Kurtosis*	*Jarque-Bera test*

50	-0.0033	0.0333	0.08	3.38	0.17
100	0.0001	0.0310	0.35	3.70	0.94
200	0.0039	0.0294	0.35	3.53	0.72
500	0.0093	0.0279	0.22	2.87	0.21
1000	0.0296	0.0403	0.61	4.00	2.38

We found no evidence to reject the hypothesis that growth rates follow a normal distribution at the 5% significance level when we look at the kurtosis values. This aligns with Dosi et al.'s (2019) assertion that while firm size distribution may follow a power law, the growth rates they experience should still conform to a normal distribution.

4.4 Results

4.4.1 Power Law

To support the granular hypothesis, firm size (in terms of employee count) must follow a power-law distribution rather than a normal distribution (Gabaix, 2011). In a power-law distribution, larger firms are more abundant, leading to a slower tail decay. This implies that idiosyncratic shocks to these large grains play a more substantial role in explaining hiring and firing cycles, challenging the conventional idea that only aggregate shocks influence these cycles observed under a normal distribution.

The Jarque-Bera test in Table 1 indicates that the normal distribution hypothesis for employee numbers in firms can be rejected. Meanwhile, the high R^2 value in Table 3 from equation (1) suggests that we cannot dismiss the possibility that firms follow a power-law distribution.

Table 3. The power law for the firm size distribution in terms of employee count.

Number of firms	R^2	*Intercept*	*Pareto exponent*
1,000	0.99	6.92	-1.60

Note: The estimates are significant at 1%.

4.4.2 Granular residual

Following Kovalenko et al. (2021), we use the Akaike information criterion to identify the number of lags needed for testing the granular hypothesis, with a maximum lag duration of two years. This figure is derived from Kovalenko et al.'s selection of six lags, equivalent to a year and a half in their quarterly data. The results are presented in Table 4.

Table 4. The number of lags selected according to the Akaike information criterion.

	Number of firms			
Lags	50	100	200	500
0	-6.73	-6.74	-6.76	-6.72
1	-6.37	-6.42	-6.46	-6.45
2	-6.29	-6.33	-6.38	-6.37

When we exclude lags in the model, values decrease, suggesting that the impact of large grains on hiring and firing cycles is short-lived within the current period. Consequently, we employed the lag-free model for data analysis.

Since we employed annual data, we refrained from seasonally adjusting the quarterly results. Table 5 provides a summary of the results from equation (4), where only the granular residual was used as a regressor.

Table 5. Explanatory power of the granular residual.

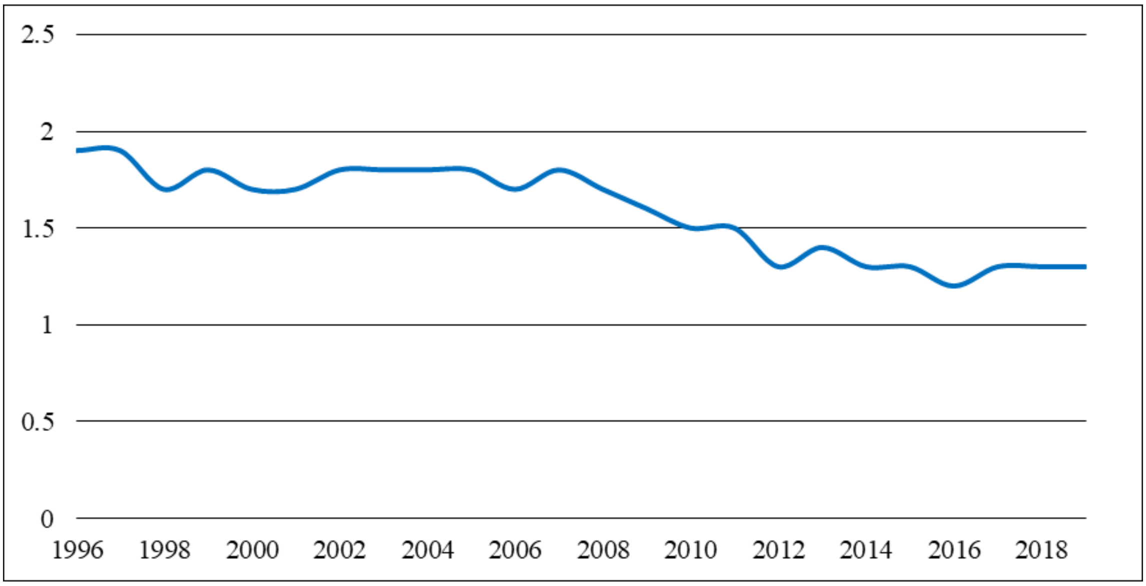
	All firms				*Manufactures*	*Services*
	Top 50	Top 100	Top 200	Top 500	Top 100	Top 100
Granular residual	*-7.07*	*-6.29*	*-5.69*	*-5.03*	2.64	*-4.17*
Intercept	0.0090	0.0092	0.0098	0.0110	*0.0289*	0.0166
F-statistic	22.14	21.48	20.44	22.18	0.05	10.20
Adjusted R^2	0.50	0.49	0.48	0.50	-0.04	0.29
Labor market, %	8.53	10.55	13.05	17.61	1.53	10.29

Notes: Italicized values indicate significance at the 1% level. When calculating the labor market percentage, the average from 1996 to 2019 is utilized.

The relatively high adjusted R^2 values support the validity of the granular hypothesis in the Brazilian labor market. In particular, the granular residual's explanatory power surpasses the labor market's firm representation percentage, highlighting the significant influence of idiosyncratic shocks from large firms on hiring and firing cycles.

One intriguing finding is that manufacturing does not seem to drive employment trends in the Brazilian labor market. This is indicated by the non-significant granular residual and a negative adjusted R^2. Several possible explanations include: 1) Manufacturing has a smaller share compared to other sectors; 2) The manufacturing workforce is shrinking (as shown in Figure 2); and 3) The largest employers in the Brazilian industrial sector are not big grains. For example, the top company that once ranked 20th in job creation in 1996 has now dropped to 68th place in 2019.

Figure 2. Manufacturing jobs as a percentage of total jobs.

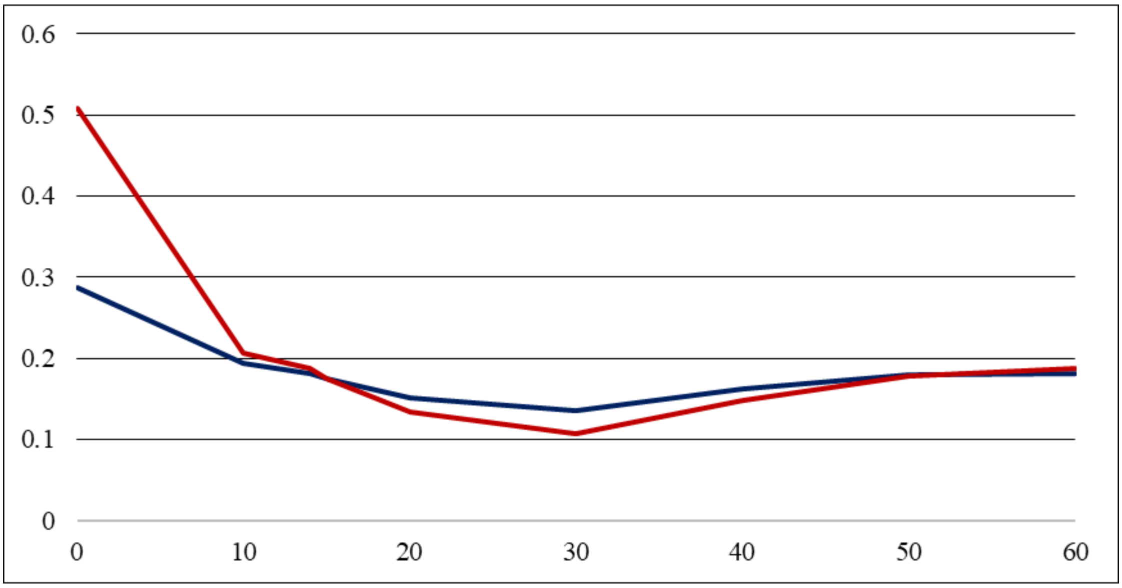


In Kovalenko et al. (2021), manufacturing's explanatory power in the German labor market is weaker compared to other specifications, and the granular hypothesis was not disproven. However, a notable contrast arises: the top 100 German manufacturers account for 14.7% of jobs, whereas in Brazil, they represent only 1.5% of formal job positions. This indicates that German manufacturers are relatively larger grains than their Brazilian counterparts. The largest German company employed approximately 62,000 individuals, whereas in Brazil, the largest company had 23,000 employees in 1997. It is important to note that our data are based on CNPJ registration, not establishments as in Kovalenko et al., suggesting that the actual difference may be even more substantial.

4.4.3 Granular size

Figure 3 shows that when K*=15, the wighted curve (equation (2)) crosses the equal-weight curve. The two curves move very near to each other after the crossing, indication that the explanatory power owing to the differing weights becomes gradually irrelevant.

Figure 3. The granular size of the Brazilian labor market.



Note: The weighted curve (equation (2)) is represented by the red line, and the equal-weight curve is represented by the blue line.

These 15 companies that define the granular size of the Brazilian job market are linked to the government. When we recalculate explanatory power using only private companies, the results are shown in Table 6.

Table 6. Explanatory power of the granular residual when only private companies are considered.

	Number of firms			
	Top 50	Top 100	Top 200	Top 300
Granular residual	19.43	*18.55*	12.62	8.17
Intercept	*0.04*	*0.04*	*0.04*	*0.03*
F-statistic	3.85	4.46	2.90	1.90
Adjusted R^2	0.11	0.14	0.08	0.04
Labor market, %	2	3	4	5

Notes: Italicized values indicate significance at the 5% level. When calculating the labor market percentage, the average from 1996 to 2019 is utilized.

The coefficient for the granular residual is only statistically significant when the number of firms equals 100. Furthermore, the explanatory power is significantly lower compared to using all companies (as shown in Table 5). This suggests that the hiring and firing cycles in the Brazilian labor market are minimally influenced by idiosyncratic shocks from Brazilian private companies.

This can be attributed to two factors. First, the private businesses in our sample are smaller in size compared to the public companies. The top CNPJ among public companies has over 400,000 employees, while among private enterprises, the leading CNPJ has approximately 20,000 employees. This difference is partially due to private corporations having multiple CNPJs for various activities or regions.

Second, the absolute value of the Pareto exponent for the distribution of private enterprises (2.4; Table 7) is notably higher than the exponent for the entire sample (1.6; Table 3). This is due to the smaller workforce in large private companies compared to the overall sample, resulting in larger grains being relatively smaller, a faster tail decay, and consequently, a decrease in explanatory power.

Table 7. The power law for the firm size distribution in terms of employee count when only private companies are considered.

Number of firms	R^2	*Intercept*	*Pareto exponent*
100	0.98	4.71	-2.96
400	0.98	6.15	-2.38

Note: The estimates are significant at 1%.

4.5. Discussion

The study conducted by Carlsson et al. (2021) emphasizes the significant role of idiosyncratic shocks in labor market cycles, specifically analyzing Swedish firms' reactions to various shocks and their impact on workforce dynamics. Their findings indicate that firms often adjust their workforce through hiring and firing in response to these shocks, with larger firms and persistent shocks leading to more substantial changes. However, unlike our analysis, their study did not specifically explore the granular hypothesis, which explains how shocks in larger companies contribute to overall firing and hiring cycles in the labor market.

Our findings in the Brazilian labor market align with the results of Kovalenko et al. (2020) in the German labor market, supporting the granular hypothesis. However, we observe that the explanatory power of the granular hypothesis diminishes when we focus solely on manufacturing or services. In particular, in our analysis, when we isolate manufacturing, the statistical significance of the granular residual is lost, suggesting that manufacturing does not account for hiring and firing cycles in Brazil. This may be attributed to the absence of large-scale grains within the Brazilian manufacturing sector.

It is important to note that our study reveals lower adjusted R2 values for Brazil compared to Kovalenko et al.'s findings, indicating that the Brazilian labor market is less granular than the German labor market. Additionally, our focus on companies by CNPJ, rather than establishments, may further accentuate the differences between the labor markets of the two countries.

One notable implication of our study is that manufacturing has a minimal impact on hiring and firing cycles in the Brazilian labor market, suggesting that job-protection policies should not primarily target the manufacturing sector. This finding aligns with a study by Geracy et al. (2019), which found that tax policies aimed at stimulating the economy and job creation had limited influence on the labor market in Brazil.

Furthermore, our research reveals that government-affiliated companies play a significant role in the Brazilian job market, indicating a dependency on government employment. This lower degree of labor market "fluidity" compared to economies driven by the private sector can hinder overall productivity growth and economic development (Davis and Haltiwanger, 2014). Additionally, the presence of numerous small and low-productivity firms can pose another obstacle to economic growth in Brazil (Firpo and Pieri, 2017).

4.6. Conclusion

According to our research on the Brazilian labor market, job turnover is mainly seen in large corporations, commonly known as big grains. This phenomenon can be measured in terms of granularity. Our study did not reject the granular hypothesis by showing that the labor market in Brazil is significantly affected by idiosyncratic shocks in big companies, leading to hiring and firing cycles. Additionally, the service sector has a more significant impact on these cycles compared to manufacturing. We found that the size of the labor market's granularity in Brazil is around 15 firms connected to public service provision. Furthermore, private companies have minimal influence on the hiring and firing cycles. Therefore, rather than providing fiscal stimulus for manufacturing, addressing high unemployment periods in Brazil can be more efficiently achieved by investing in public services (Author's last name, Year).

5. Final Considerations

In our research, we conducted three tests of Gabaix's (2011) granular hypothesis, and in each case, we did not find evidence to reject it.

In the first essay, we discovered that countries with a higher relative weight in international trade exert a larger impact on global repercussions than their trade share would suggest. Specifically, we identified eight major players, or "large grains," that account for most of the inflation spillovers. This implies that central banks in other countries should closely monitor these eight major players when formulating their monetary policies. In future research, it would be interesting to explore the hypothesis that economic growth from these "big grains" may be responsible for global economic cycles, and examine the granularity of global endogenous growth.

The second essay explores the relationship between granularity and power laws, and proposes that large cities play a significant role in the economic cycle beyond their population size. We found that in the United States, the granular city size consists of three metropolitan areas, while in Brazil, it is equivalent to three municipalities. This research emphasizes the spatial component of granularity that has not been previously considered. To extend this study, it would be beneficial to incorporate indicators of economic complexity to further understand the impact of large agglomerations on economic cycles.

Lastly, in the third essay, we focused on the distribution of power among companies based on their workforce size, observing that idiosyncratic shocks in large companies significantly influence hiring and firing cycles. The services sector was found to play a substantial role in explaining these cycles, whereas the manufacturing industry's explanatory power was limited. We determined that the granular size of the Brazilian labor market consists of 15 companies involved in public services, with private companies having a relatively smaller impact on hiring and firing patterns. This research could be extended to investigate the granular effects of large firms on technological innovation, such as through patent analysis.

6. Bibliographic References

Ades, A., and E. Glaeser (1994) Trade and circuses: Explaining urban giants. NBER Working Paper No. 4715. https://www.nber.org/system/files/working_papers/w4715/w4715.pdf

Attinasi, M.G., and M. Balatti (2021) Globalisation and its implications for inflation in advanced economies, ECB Economic Bulletin, Issue 4. https://www.ecb.europa.eu/pub/economic-bulletin/articles/2021/html/ecb.ebart202104_01~ae13f7fe4c.en.html

Auer, R., C. Borio, and A. Filardo (2017) The globalisation of inflation: The growing importance of global value chains, BIS Working Paper No. 602. https://www.bis.org/publ/work602.pdf

Auer, R.A., A.A. Levchenko, and P. Saure (2019) International inflation spillovers through input linkages, Review of Economics and Statistics 101 (3), 507-521. https://doi.org/10.1162/rest_a_00781

Balcilar, M., and F.V. Bekun (2020) Spillover dynamics across price inflation and selected agricultural commodity prices, Journal of Economic Structures 9, 2. https://doi.org/10.1186/s40008-020-0180-0

Baurle, G., M. Gubler, and D.R. Kanzig (2021) International inflation spillovers: The role of different shocks, International Journal of Central Banking, 191-230. https://www.ijcb.org/journal/ijcb21q1a6.pdf

Bettencourt, L.M.A., J. Lobo, D. Helbing, C. Kühnert, and G.B. West (2007) Growth, innovation, scaling, and the pace of life in cities, Proceedings of the National Academy of Sciences 104 (17), 7301-7306. https://doi.org/10.1073/pnas.0610172104

Bianchi, F., and A. Civelli (2015) Globalization and inflation: Evidence from a time-varying VAR, Review of Economic Dynamics 18 (2), 406-433. http://dx.doi.org/10.1016/j.red.2014.07.004

Black, D., and V. Henderson (2003) Urban evolution in the USA, Journal of Economic Geography 3 (4), 343-372. https://doi.org/10.1093/jeg/lbg017

Blanco-Arroyo, O., A. Ruiz-Buforn, D. Vidal-Tomas, and S. Alfarano (2018) On the determination of the granular size of the economy, Economics Letters 173, 35-38. http://dx.doi.org/10.1016/j.econlet.2018.08.020

Breinlich, H., H. Fadinger, V. Nocke, and N. Schutz (2020) Gravity with granularity, CEPR Discussion Papers No. 15374.

Carlsson, M., Messina, J., Nordström Skans, O. (2021) Firm-level shocks and labour flows, Economic Journal 131 (634), 598-623. https://doi.org/10.1093/ej/ueaa087

Chandrarin, G., K. Sohag, D.S. Cahyaningsih, D. Yuniawan, and H. Herdhayinta (2022) The response of exchange rate to coal price, palm oil price, and inflation in Indonesia: Tail dependence analysis, Resources Policy 77, 102750. https://doi.org/10.1016/j.resourpol.2022.102750

Chauvin, J.P., E. Glaeser, Y. Ma, and K. Tobio (2017) What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States, Journal of Urban Economics 98 (3), 17-49. https://doi.org/10.1016/j.jue.2016.05.003

Ciccarelli, M., and J.A. García (2015) International spillovers in inflation expectations, ECB Working Paper, No. 1857. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1857.en.pdf

Ciccarelli, M., and B. Mojon (2010) Global inflation, Review of Economics and Statistics 92 (3), 524-535. https://www.jstor.org/stable/27867556

Clauset, A., C.R. Shaliza, and M.E.J. Newman (2009) Power-law distributions in empirical data, SIAM Review 51 (4), 661-703. https://doi.org/10.1137/070710111

Davis, S.J., John Haltiwanger, J., Schuh, S. (1996) Small business and job creation: Dissecting the myth and reassessing the facts, Small Business Economics 8 (4), 297-315. https://www.jstor.org/stable/40228707

Davis, S.J., Haltiwanger, J. (2014) Labor market fluidity and economic performance, NBER Working Papers 20479.https://doi.org/10.3386/w20479

Da Silva, S., Matsushita, R., Giglio, R., Massena, G. (2018) Granularity of the top 1,000 Brazilian companies, Physica A 512 (C), 68-73. http://dx.doi.org/10.1016/j.physa.2018.08.027

Di Giovanni, J., and A.A. Levchenko (2012) Country size, international trade, and aggregate fluctuations in granular economies, Journal of Political Economy 120 (6), 1083-1132. https://doi.org/10.1086/669161

Di Giovanni, J., A.A. Levchenko, and I. Mejean (2017) Large firms and international business cycle comovement, American Economic Review 107 (5), 598-602. http://dx.doi.org/10.1257/aer.p20171006

Di Giovanni, J., A.A. Levchenko, and I. Mejean (2018) The micro origins of international business cycle comovement, American Economic Review 108 (1), 82-108. http://dx.doi.org/10.1257/aer.20160091

Di Giovanni, J., A.A. Levchenko, and I. Mejean (2021) Foreign shocks as granular fluctuations, NBER Working Paper No. 28123. http://dx.doi.org/10.3386/w28123

Dobkins, L., and Y. Ioannides (1998) Dynamic evolution of the U.S. city size distribution, in J.M. Huriot and J.F. Thisse (eds.), The Economics of Cities, New York: Cambridge University Press.

Dosi, G., M. Napoletano, A. Roventini, and T. Treibich (2019) Debunking the granular origins of aggregate fluctuations: From real business cycles back to Keynes, Journal of Evolutionary Economics 29 (1), 67-90.https://doi.org/10.1007/s00191-018-0590-4

Duranton, G., and D. Puga (2004) Micro-foundations of urban agglomeration economies, in Henderson, J.V., and J.F. Thisse (eds.), Handbook of Regional and Urban Economics, Vol. 4, New York: North Holland, pp. 2063-2117. https://doi.org/10.1016/S1574-0080(04)80005-1

Duranton, G., and D. Puga (2024) Urban growth and its aggregate implications, Econometrica, forthcoming. https://doi.org/10.1109/5.771073

Eaton, J., S.S. Kortum, and S. Sotelo (2012) International trade: Linking micro and macro, NBER Working Paper No. 17864. http://dx.doi.org/10.3386/w17864

Forbes, K.J. (2019) Has globalization changed the inflation process? BIS Working Paper No. 791. https://www.bis.org/publ/work791.pdf

Firpo, S., Pieri, R. (2017) Structural change, productivity growth, and trade policy in Brazil. In Structural change, fundamentals, and growth: A framework and case studies. McMillan, Margaret S.; Rodrik, Dani; and Sepúlveda, Claudia (Eds.). Chapter 7. Pp. 267-292. Washington, D.C.: International Food Policy Research Institue. http://dx.doi.org/10.2499/9780896292147_ch7

Florida, R. (2019) The Rise of the Creative Class. New York: Basic Books.

Freund, C., and M.D. Pierola (2015) Export superstars, Review of Economics and Statistics 97 (5), 1023-1032. http://dx.doi.org/10.1162/REST_a_00511

Gabaix, X. (1999) Zipf's law for cities: An explanation, Quarterly Journal of Economics 114 (3), 739-767. https://doi.org/10.1162/003355399556133

Gabaix, X. (2011) The granular origins of aggregate fluctuations, Econometrica 79 (3), 733-772. https://doi.org/10.3982/ECTA8769

Gabaix, X., and R. Ibragimov (2011) Rank – ½: A simple way to improve the OLS estimation of tail exponents, Journal of Business & Economic Statistics 29 (1), 24-39. https://doi.org/10.1198/jbes.2009.06157

Gabaix, X., and R.S.J. Koijen (2020) Granular instrumental variables, NBER Working Paper No. 28204. https://www.nber.org/system/files/working_papers/w28204/w28204.pdf

Gaubert, C., and O. Itskhoki (2021) Granular comparative advantage, Journal of Political Economy 129 (3), 871-939. https://doi.org/10.1086/712444

Geracy, I., Corseuil, C., Silveira, F. (2019) Desonerações do imposto sobre produtos industrializados e seus impactos sobre o mercado de trabalho. Texto para Discussão 2515, IPEA. https://repositorio.ipea.gov.br/handle/11058/9377

Giesen, K., and J. Sudekum (2011) Zipf's law for cities in the regions and the country, Journal of Economic Geography 11 (4), 667-686. https://doi.org/10.1093/jeg/lbq019

Glaeser, E. (2011) Triumph of the City: How Our Greatest Invention Makes Us Richer, Greener, Smarter, Healthier, and Happier. New York: Penguin Books.

Glaeser, E.L., and J.D. Gottlieb (2009) The wealth of cities: Agglomeration economies and spatial equilibrium in the United States, Journal of Economic Literature 47 (4), 983-1028. https://doi.org/10.1257/jel.47.4.983

Grigoli, F., E. Luttini, and D. Sandri (2023) Idiosyncratic shocks and aggregate fluctuations in an emerging market, Journal of Development Economics 160 (1), 102949. https://doi.org/10.1016/j.jdeveco.2022.102949

Halka, A. and K. Szafranek (2016) Whose inflation is it anyway? Inflation spillovers between the euro area and small open economies, Eastern European Economics 54 (2), 109-132. http://dx.doi.org/10.1080/00128775.2015.1126788

Hsieh, C.T., and E. Moretti (2019) Housing constraints and spatial misallocation, American Economic Journal: Macroeconomics 11 (2), 1-39. https://doi.org/10.1257/mac.20170388

Ioannides, Y.M., and H.G. Overman (2004) Spatial evolution of the US urban system, Journal of Economic Geography 4 (2), 131-156. https://doi.org/10.1093/jeg/4.2.131

Istiak K., A.K. Tiwari, H. Husain, and K. Sohag (2021) The spillover of inflation among the G7 countries, Journal of Risk and Financial Management 14 (8), 392. https://doi.org/10.3390/jrfm14080392

Jacobs, J. (1970) The Economy of Cities. New York: Vintage.

Jenkins, S.P. (2017) Pareto models, top incomes and recent trends in UK income inequality, Economica 84 (334), 261-289.https://doi.org/10.1111/ecca.12217

Kamber, G., and B. Wong (2020) Global factors and trend inflation, Journal of International Economics 122, 103265. https://doi.org/10.1016/j.jinteco.2019.103265

Kovalenko, T., Schnabel, C., Stuber, H. (2022) Is the German labour marketgranular? Applied Economics Letters 29 (1), 41-48. https://doi.org/10.1080/13504851.2020.1855300

Krugman, P. (1996) Confronting the mystery of urban hierarchy, Journal of the Japanese and International Economies 10 (4), 399-418. https://doi.org/10.1006/jjie.1996.0023

Mandelbrot, B. (1961) Final note on a class of skew distribution functions: Analysis and critique of a model due to H. A. Simon, Information and Control 4 (2-3), 198-216. https://doi.org/10.1016/S0019-9958(61)80008-9

Mikolajun, I. and D. Lodge (2016) Advanced economy inflation: The role of global factors, ECB Working Paper No. 1948. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1948.en.pdf

MIRANDA-PINTO, Jorge; SHEN, Yuanting. A granular view of the Australian business cycle. Economic Record, v. 95, n. 311, p.407-424, 2019.

Newman, M.E.J. (2005) Power laws, Pareto distributions and Zipf's law, Contemporary Physics 46 (5), 323-351. http://dx.doi.org/10.1080/00107510500052444

Nitsch, V. (2005) Zipf zipped, Journal of Urban Economics 57 (1), 86-100. https://doi.org/10.1016/j.jue.2004.09.002

Rauch, F. (2014) Cities as spatial clusters, Journal of Economic Geography 14 (4), 759-773. https://doi.org/10.1093/jeg/lbt034

Rosen, K.T., and M. Resnick (1980) The size distribution of cities: An examination of the Pareto law and primacy, Journal of Urban Economics 8 (2), 165-186. https://doi.org/10.1016/0094-1190(80)90043-1

Rosenthal, S.S., and W.C. Strange (2004) Evidence on the nature and sources of agglomeration economies, in Henderson, J.V., and J.F. Thisse (eds.), Handbook of Regional and Urban Economics, Vol. 4. New York: North Holland, pp. 2119-2171. https://doi.org/10.1016/S1574-0080(04)80006-3

Rozman, G. (1990) East Asian urbanization in the nineteenth century, in Van der Woude, A., A. Hayami, and J. de Vries (eds.), Urbanization in History. New York: Oxford University Press, pp. 61-73.

Schnabel, I. (2022) The globalisation of inflation, European Central Bank, Speech made on May 11, 2022, in Vienna. https://www.ecb.europa.eu/press/key/date/2022/html/ecb.sp220511_1~e9ba02e127.en.html

Silva, M., and S. Da Silva (2020) The Brazilian granular business cycle, Economics Bulletin 40 (1), 463-472. http://www.accessecon.com/Pubs/EB/2020/Volume40/EB-20-V40-I1-P40.pdf

Simon, H.A. (1955) On a class of skew distribution functions, Biometrika 42 (3-4), 425-440. https://doi.org/10.1093/biomet/42.3-4.425

WAGNER, Joachim. The German manufacturing sector is a granular economy. Applied Economics Letters, v. 19, n. 17, p.1663-1665, 2012.

Zipf, G. (1949) Human Behavior and the Principle of Last Effort. Cambridge: Addison-Wesley.

7. Appendix

This appendix aims to use simplified models to gain insights into the relationship between grains and their shocks, as well as the explanatory power of the granular residue in relation to economic cycles.

In Case 1, we make the following assumptions:

1. The economy as a whole is initially normalized to 1.
2. There is a large company that holds $(1-\alpha)\%$ of the market, where $0<\alpha<1$.
3. There are n small and equal companies that each hold α/n of the market.
4. The large company's production is determined by the following equation:

$$y_{g,t} = (1-\alpha) + (1-\alpha)\,\theta_t + (1-\alpha)\varepsilon_t\,,$$

where:

θ_t = is the aggregate shock in percentage terms with expectation 0 and variance σ_θ^2.

ε_t = is the idiosyncratic shock in percentage terms in the large firm with expectation 0 and variance σ_ε^2

Like this,

$$E(y_{g,t}) = (1-\alpha)\;;$$
$$Var\,(y_{g,t}) = (1-\alpha)^2\,\sigma_\theta^2 + (1-\alpha)^2\,\sigma_\varepsilon^2 + 2(1-\alpha).Cov(\theta_t,\,\varepsilon_t)\,.$$

Small companies:

$$y_{p,t} = \frac{\alpha}{n} + \frac{\alpha}{n}\theta + \frac{\alpha}{n}\gamma_{i,t}\,,$$

where:

$\gamma_{i,t}$ = is the idiosyncratic shock in percentage terms in small firm i with mean 0 and variance (the same for all companies) σ_γ^2 .

The aggregate of the economy (y_t) is given by the sum of the production of the large company with the sum of the production of the n small companies, therefore we have:

$$y_t = 1 + \theta_t + (1- \alpha)\, \varepsilon_t + (\alpha/n) \,.\, \sum_{i=1}^{n} \gamma i,t \,.$$

Since the expectation of γ_i is zero, we will assume that n is large enough for us to consider $\sum_{i=1}^{n} \gamma it$ equal to zero (as argued in Lucas (1977) and Gabaix (2011)) , so we will have that the expected output of the economy will be E(yt) = 1.

And the variance:

$$\mathrm{Var}(y_t) = \sigma_\theta^2 + (1\text{-}\alpha)\sigma_\varepsilon^2 + 2(1\text{-}\alpha)\, \mathrm{Cov}(\theta_t, \varepsilon_t).$$

When we test the granular hypothesis, we generally use the "granular residual" (GR) as an explanatory variable. For the case with only one "large grain" that we are analyzing, it would be calculated as:

GR = (share).(large firm growth rate – economy growth rate).

Assuming that the initial values for the large firm are (1-α) and for the economy as a whole are 1, we have:

$$\text{Large company growth rate} = ((1\text{-}\alpha) + (1\text{-}\alpha)\,\theta_t + (1\text{-}\alpha)\,\varepsilon_t - (1\text{-}\alpha))/(\,(1\text{-}\alpha)) = \theta_t + \varepsilon_t \;;$$

$$\text{Economy growth rate} = (1 + \theta_t + (1\text{-}\alpha)\,\varepsilon_t - 1)/1 = \theta_t + (1\text{-}\alpha)\,\varepsilon_t \,.$$

Replacing the growth rate of the large company and the growth rate of the economy in the granular residual (for the large company) we have:

$$GR = (1\text{-}\alpha)(\,\theta_t + \varepsilon_t - \theta_t + (1\text{-}\alpha)\,\varepsilon_t) = (1\text{-}\alpha)\alpha\varepsilon_t \,.$$

In testing the granular hypothesis, we performed Ordinary Least Squares (OLS) regression of y_t in relation to GR and analyzed the explanatory power (R^2). Note that R^2 is given by the square of the correlation coefficient between y_t and GR, in this case we have:

$$R^2 = \frac{(Cov(\theta t,\varepsilon t))^2 + 2.(1-\alpha)\sigma^2{}_\varepsilon Cov(\theta t,\varepsilon t) + (1-\alpha)^2\sigma^4{}_\varepsilon}{\sigma^2{}_\theta\, \sigma^2{}_\varepsilon + 2(1-\alpha)\sigma^2{}_\varepsilon Cov(\theta t,\varepsilon t) + (1-\alpha)^2\sigma^4{}_\varepsilon} \,,$$

in which we observed that $\frac{\partial R^2}{\partial Cov(\theta t, st)} > 0$; $\frac{\partial R^2}{\partial \sigma^2{}_s} < 0; as\ (Cov(\theta t, \varepsilon t))^2 < \sigma^2{}_\theta\, \sigma^2{}_s$, $\frac{\partial R^2}{\partial (1-\alpha)} > 0;\ \ and\ \ \frac{\partial R^2}{\partial \sigma^2{}_\theta} < 0$.

From the analysis of R^2 we can infer that:

i) $\frac{\partial R^2}{\partial Cov(\theta t, st)} > 0$, therefore the higher $Cov(\theta_t, \varepsilon_t)$, the greater the explanatory power of the Granular Residual. In this way, when we add a grain to the Granular Residual whose errors are more correlated with θ_t, the tendency is for the explanatory power of the residue to increase. On the other hand, if $Cov(\theta_t, \varepsilon_t)$ decreases, it is possible (ignoring changes in $\sigma^2{}_s$ and if $Cov(\theta_t\ \varepsilon_t)$ decreases, dominate the effect of the increase in (1-α)) in which If R^2 decreases, in this case, we would be "contaminating" the residue with a grain that removes explanatory power.

ii) $\frac{\partial R^2}{\partial \sigma^2{}_s} < 0$, therefore, if we add grains with a variance greater than that of the residual to the residual, it is possible that the explanatory power will decrease. In some circumstances in which we are faced with grains whose growth rates fluctuated sharply (for example, in the essay on the granularity of inflation spillovers, countries that were in acute inflationary processes and had an abrupt stabilization, such as Brazil or Russia, had a disproportionately large granular residue for some years, which would imply a shock with very high variance) we observed that the addition of these grains in the aggregate residue ended up reducing its explanatory power.

iii) Notice that $(Cov(\theta t, \varepsilon t))^2 \leq \sigma^2{}_\theta\, \sigma^2{}_s$, then $(Cov(\theta t, \varepsilon t))^2/\sigma^2{}_\theta\, \sigma^2{}_s)$ would be the square of the correlation coefficient between θ_t and εt and this is, by construction, between 0 and 1. This causes the denominator in R^2 to be greater than the numerator and consequently $\frac{\partial R^2}{\partial (1-\alpha)} > 0$. Note that the "maximum" impact of increasing share occurs when $Cov\ (\theta_t, \varepsilon_t) = 0$, which indicates that when the aggregate and idiosyncratic shocks are less correlated, the granular size is probably greater in a given scenario, as share gains "protagonism" in determining R^2.

iv) $\frac{\partial R^2}{\partial \sigma^2{}_\theta} < 0$, indicates that the greater the variance of the aggregate shock, the lower the explanatory power of the granular residual. This makes sense, because the greater the variance of the aggregate shock, the smaller the relative size of the idiosyncratic shock and consequently its explanatory power for variations in the growth rate of y_t.

v) If we imagine "perfectly idiosyncratic" shocks, that is, θ_t and ε_t are independentes (Cov (θ_t, ε_t) = 0), and ε_t i.i.d. and use the reference value R^2> large firm share = (1-α) we have $\sigma^2{}_s > (\frac{\sigma^2{}_\theta}{\alpha(1-\alpha)})$. If α=0,5 (which maximizes α(1- α)), the explanatory power of the granular residue will be greater than the share only if $\sigma^2{}_s > 4\,\sigma^2{}_\theta$, a result with a strong indication that large grains explain a large part of economic cycles.

vi) if only the aggregate shocks mattered ($\sigma^2{}_s = 0$) which would imply that Cov (θ_t, ε_t) = 0 (since ε_t would be constant) the estimated R^2 would be 0. Therefore, in order not to reject the Granular Hypothesis (i.e., that idiosyncratic shocks are important to explain cycles) is sufficient for R^2 to be sufficiently greater than zero.

Case 2

Assumptions:

- two large companies (y_{g1} and y_{g2}) of the same size that together represent (1-α) the market:

$$y_{g1t} = \frac{(1-\alpha)}{2} + \left(\frac{1-\alpha}{2}\right)\,\theta t + \frac{(1-\alpha)}{2}\,\varepsilon_{1t}$$

and

$$y_{g2t} = \frac{(1-\alpha)}{2} + \left(\frac{1-\alpha}{2}\right)\,\theta t + \frac{(1-\alpha)}{2}\,\varepsilon_{2t}\,.$$

Both with "perfectly idiosyncratic" shocks (Cov (θ_t,ε_{1t}) = Cov (θ_t,ε_{2t})=0), but with shocks that are not independent of each other (Cov (ε_{1t},ε_{2t}) ≠ 0). We consider that the two companies may have idiosyncratic shocks of different variances, that is, Var(ε_{1t}) ≠ Var (ε_{2t}).

Small businesses are the same as in Case 1. The economy as a whole is given by:

$$y_t = 1 + \theta_t + ((1-\alpha)/2)(\varepsilon_{1t} + \varepsilon_{2t}) + (\alpha/n)\,.\,\textstyle\sum_{i=1}^{n} \gamma i, t,$$

again, we will consider that n is large enough so that we can disregard shocks from small companies.

Doing the necessary algebra, we conclude that the explanatory power of the granular residual calculated using the two large firms in relation to yt is given by:

$$R^2 = (\frac{(1-\alpha)^2\left(\sigma_1^2+\sigma_2^2+2Cov\ (\varepsilon 1t,\varepsilon 2t)\right)}{(4\ \sigma^2{}_\theta\ +(1-\alpha)^2\left(\sigma_1^2+\sigma_2^2+2Cov\ (\varepsilon 1t,\varepsilon 2t)\right)} ,$$

where $\sigma_1^2 = Var(\varepsilon_{1t})$ and $\sigma_2^2 = Var\ (\varepsilon_{2t})$.

If $\sigma_1^2 = \sigma_2^2 = \sigma^2{}_s$, R^2 becomes:

$$R^2(\sigma^2{}_\varepsilon) = (\frac{(1-\alpha)^2\left(\sigma^2{}_\varepsilon+Cov\ (\varepsilon 1t,\varepsilon 2t)\right)}{(2\ \sigma^2{}_\theta\ +(1-\alpha)^2(\sigma^2{}_\varepsilon+Cov\ (\varepsilon 1t,\varepsilon 2t))} .$$

Note that, if $\sigma_1^2 = \sigma^2{}_s > \sigma_2^2$ (and multiplying $R^2(\ \sigma^2{}_s))$ by 2/2), then:

$$(\frac{(1-\alpha)^2\left(2\sigma^2{}_\varepsilon+2Cov\ (\varepsilon 1t,\varepsilon 2t)\right)}{(4\sigma^2{}_\theta\ +(1-\alpha)^2(2\sigma^2{}_\varepsilon+2Cov\ (\varepsilon 1t,\varepsilon 2t))} > (\frac{(1-\alpha)^2\left(\sigma_1^2+\sigma_2^2+2Cov\ (\varepsilon 1t,\varepsilon 2t)\right)}{(4\ \sigma^2{}_\theta\ +(1-\alpha)^2\left(\sigma_1^2+\sigma_2^2+2Cov\ (\varepsilon 1t,\varepsilon 2t)\right)} .$$

Based on our observations, we have come to realize the following insights:

i) R^2 is positively correlated with $Cov(\varepsilon_{1t}, \varepsilon_{2t})$. This means that adding grains to the granular residual whose idiosyncratic shocks are positively correlated with the existing shocks in the residue (but not correlated with the aggregate shock) increases the explanatory power of the residue in relation to y_t.

ii) If the added grain has a lower variance than the variance of the granular residue, it has the potential to reduce the explanatory power of the residue. However, when another grain is added, it also increases the "share" represented by the granular residue. The net effect on the explanatory power depends on the magnitudes involved.

Conclusion

When a q+1-th grain is added to a granular residue calculated from q grains, several phenomena occur with R2:

i) The share of the Granular Residual will necessarily increase, which can enhance the explanatory power.

ii) However, if the added grain has a lower variance than the residual and its idiosyncratic shocks are poorly correlated with the existing shocks in the residue, or it is less correlated with θ, then R^2 tends to decrease. This negative effect can outweigh the positive effect of increasing the share, resulting in a decrease in explanatory power. We refer to this as the contamination of the Granular Residue by the additional grain.

iii) Furthermore, if the added grain has a negative $Cov(\varepsilon_{1t}, \varepsilon_{2t})$ (a negative covariance between idiosyncratic shocks) and is of significant magnitude, it can abruptly reduce explanatory power. In cases where some grains exhibit aggressive oscillations and are somewhat disconnected from other agents, the increase in their influence tends to bring R^2 close to zero. This phenomenon was observed in the context of studying inflation, where countries with very high inflation rates had to be excluded from the analysis, such as Russia and Brazil.