The Granular Nature of Emerging Market Economies: The Case of Kazakhstan[1] [2]

Jozef Konings[a,c]

Galiya Sagyndykova[b]

Venkat Subramanian[a]

Astrid Volckaert[c]

[a] Nazarbayev University Graduate School of Business, Nur-Sultan, Kazakhstan

[b] Department of Economics, School of Sciences and Humanities, Nazarbayev University, Nur-Sultan, Kazakhstan

[c] VIVES, Faculty of Economics and Business, KU Leuven, Leuven, Belgium

Abstract

This paper analyzes the granularity hypothesis in a large emerging economy, Kazakhstan. We use a new longitudinal dataset at the firm level and at quarterly frequency between 2012 and 2018 to document the size distribution of firms and to provide evidence that it follows a power law. We find that the largest 30 firms explain nearly 80 percent of the growth in aggregate total factor productivity. This confirms earlier research for the U.S. and other developed countries. However, the granular nature of the Kazakh economy is even more outspoken than in other countries. Thus idiosyncratic shocks and the way they ripple through the production network matter to understand changes in aggregate productivity growth. Moreover, since these granular firms are concentrated in the oil industry it exposes the vulnerability of the economy more to unexpected shocks in one industry in particular.

Keywords: granularity, firm heterogeneity, aggregate fluctuations, Total Factor Productivity, transitional economies

JEL classification numbers: D24, E23, E32, L16, L25, P27

[1] This research was partially financed by Nazarbayev University (Funder Project Reference: 16133385) and the KU Leuven Methuselah Grant (METH/15/004).

[2] Declarations of interest: none

1. Introduction

This paper analyzes the importance of idiosyncratic firm-specific shocks in explaining macroeconomic fluctuations. Gabaix (2011) showed for the U.S. that idiosyncratic firm-level shocks can explain an important part of aggregate business cycles, which goes against the mainstream macroeconomic view that firm-specific shocks average out in the aggregate. The latter, however, does no longer hold when the firm size distribution is fat-tailed. Relatively few large firms tend to contribute more to macro fluctuations than the large majority of small and medium sized enterprises, which populate the economy.[3] Since the economy may be dominated by large firms, idiosyncratic shocks to these firms can lead to substantial aggregate shocks. For instance, a shock to Wall-Mart in the U.S. is likely to have ripple down effects through its supply chain and its network of supermarkets spread throughout the U.S., which would be different than when each of these supermarkets were independent stores.

While evidence increasingly shows that the production network in developed market economies tends to be complex and dominated by a few 'superstar' firms (Bernard et al., 2019), little is known about the granular nature of emerging economies. This is of particular interest as emerging economies increasingly contribute to global GDP growth and take a greater share of the global economy, from 20 percent in 2000 to over 40 percent in 2020 (IMF, 2021). In addition, most emerging economies are characterized by a surge in de novo private enterprises and entrepreneurship, whilst large state owned firms are being split up and privatized (see Konings, Van Cayseele and Warzynski, 2005; De Loecker and Konings, 2006). In other cases, state owned enterprises turn into quasi-public sector firms, and tend to be nurtured as national champions, often becoming system relevant for the country's economy. For instance, the chaebols of Korea contribute upwards of 50 percent of GDP. In other Asian emerging economies, a few firms account for between 25 percent to 55 percent of stock market capitalization (Belenzon et al., 2012).

This paper contributes to this small, but growing literature by analyzing the granular nature of a relatively large emerging and transitional economy, Kazakhstan, the second largest economy of the former Soviet Union. With a GDP per capita of USD 9,800 in 2019, it transitioned to the upper-middle-income category in less than 20 years (World Bank, 2021). The contribution of governmentally owned firms to GDP steadily decreased over

[3] In 2015, enterprises employing fewer than 250 persons represented 99 percent of all enterprises in the EU, according to Eurostat (2021). See also OECD (2021) for detailed data by country.

this period (to 15 percent in 2019), with 84 percent of this contribution being made by firms with more than 250 employees, mainly concentrated in oil and gas, mining, and telecommunication industries.[4]

The granular hypothesis has been investigated for developed economies and confirmed the hypothesis that a few large firms have a disproportionate effect on macroeconomic cycles. Ebeke and Eklou (2017) looked at the effects of idiosyncratic productivity shocks on aggregate macroeconomic fluctuations in 8 European countries and found that firm-level productivity shocks explained around 40 percent of fluctuations in GDP growth over the period 2000-2013. Blanco-Arroyo, Ruiz-Buforn, Vidal-Tomás and Alfarano (2018) showed for Spain that 45 percent of the variation in GDP growth could be explained by idiosyncratic shocks happening to the (450) largest firms. Fornaro and Luomaranta (2018) confirmed that the Finnish economy was also granular, with only the top 57 firms accounting for about one third (30.8 percent) of the monthly fluctuations in Finnish GDP growth over the period from 1999 to 2013. Miranda-Pinto and Shen (2019) investigated the correlation between the microeconomic shocks and GDP growth over the period from 2000 to 2018 in Australia and also confirmed that idiosyncratic shocks to the largest 100 non-financial firms explained from 20 to 40 percent of fluctuations in the aggregate business cycle. Lee (2015) found that 18 percent of variation in Korean GDP growth over the period from 1981 to 2011 was attributable to firm-level shocks happening to the 20 largest companies in the Korean economy. Interestingly, in a recent study that replicated Gabaix's (2011) empirical approach, Wagner and Weche (2020) used data on the top 100 German companies and found that the German economy was not granular. That is to say, idiosyncratic shocks to the performance of these top companies were not significantly associated with fluctuations in the GDP per head growth rate in Germany.[5] This contrasts most empirical evidence supporting the granular hypothesis.

The granular hypothesis has also been applied to aggregate sales instead of GDP. For instance, Di Giovanni, Levchenko and Mejean (2014) analyzed French firms from 1990 to 2007 and found that firm-specific shocks accounted for 80 percent of the volatility in aggregate sales growth, indicating that the dynamics of large firms had a tremendous impact on the aggregate performance of the whole economy. Using a similar analytical approach to that of Di Giovanni et al., Czinkán (2017) found that firm-level shocks

[4] According to the Kazakh government press release (https://government.kz/ru/news/)

[5] This might be partly explained by the relatively high diversification of the German economy characterized by a favorable environment for 'hidden-champions', SME's that are global market leaders in niche sectors (Bleuel, 2018).

contributed to about 55.5 percent of aggregate sales fluctuations over the period from 2000 to 2013 in Hungary, relatively less compared to 80 percent of the aggregate sales volatility explained by idiosyncratic shocks in France. Friberg & Sanctuary (2016) tested the same methodology in the Swedish context and reported that firm-specific shocks were responsible for 52 percent of the variation in aggregate sales over 1997-2008. Popova's (2019) analysis revealed that Russia also showed characteristics of a granular economy. She analyzed Russian firms over 1999-2017 and found that idiosyncratic shocks to the largest firms were accountable for around 75 percent of variation in aggregate sales growth, similar to results observed in the French data (Di Giovanni et al., 2014).

Some studies checked if other measures such as TFP exhibit signs of granularity. Gnocato and Rondinelli (2018) analyzed the Italian business cycle over 1999-2014 and reported that firm-specific productivity shocks accounted for 30 to 40 percent of aggregate TFP fluctuations, despite the fact that the Italian economy has been dominated by many small rather than a few large firms. Papa (2019) tested the granular hypothesis in the Irish economy, which had been rated as the most concentrated market, according to the Herfindahl-Hirschman index, compared to other major economies such as France, Germany, Italy, Denmark, Netherlands, and Japan. The authors found that around one third (32.7 percent) of the growth in aggregate multifactor productivity over a 15 year time period (2000-2014) in Ireland could be explained by idiosyncratic productivity shocks to the 5 largest firms. The United Kingdom's economy was also found to be granular by Dacic & Melolinna (2019), who reported that firm-specific shocks to the top 100 firms accounted for about 30 percent of the aggregate productivity fluctuations over the period from 1988 to 2016.

The literature on granularity is also closely related to the literature on 'superstar firms'. Recent work by Autor et al. (2020) and Chen et al. (2022) tries to identify the mechanisms behind the rise of these top firms looking at the role of technological changes and innovation.

In this paper, we focus on the granularity of aggregate total factor productivity of the Kazakh economy. This paper contributes to the discussion on granularity in three important ways. First, it is one of the few papers to test the granularity hypothesis and the first to quantify the number of granular firms in one of the largest emerging economies. While reforms have taken place in the last 25 years, the economy continues to be heavily dependent on the public and resource sectors. Second, we focus on Total Factor Productivity (TFP) rather than simply labor productivity and estimate TFP through two distinct methodologies: (i) based on a control function approach as in Ackerberg, Caves and Frazer (ACF, 2015) and (ii) based on the Törnqvist index. Unlike labor productivity, TFP accounts for the capital intensity of firms as well, which is of particular importance given the share of capital intensive industries that are present in the economy. Third, the analysis is done using unique confidential firm-level data at quarterly frequency. This data is collected by the Bureau of National Statistics (BNS) of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. The BNS conducts an annual survey, which is compulsory for all firms with more than 100 employees. The detailed quarterly financial data help to deepen the analysis of the existence of heterogeneous effects. The dataset also includes information on the type of firm allowing us to characterize the firm in further detail and differentiate between state owned firms, private firms and foreign firms, amongst others. Using additional firm-level information on investment and intangible assets, we shed a light on the importance of innovation for granular firms.

The remainder of this paper is organized as follows: Section 2 explains the methodology used to estimate productivity and compose the Granular Residual. Section 3 describes the dataset with special attention to the sectoral decomposition and firm heterogeneity. Section 4 presents the empirical results including robustness checks and section 5 provides a discussion and highlights the conclusions and policy recommendations that it implies for the institutions of transitional economies.

2. Methodology

We follow closely the approach proposed in Gabaix (2011). In a first step, we identify idiosyncratic shocks that occur in firms. Where Gabaix and many others identify idiosyncratic shocks as changes in labor productivity, we employ Total Factor Productivity (TFP) instead. We measure TFP using a control function approach as in Ackerberg, Caves and Frazer (ACF, 2015), and for robustness, we consider a number of alternative measures, which we explain in further detail at the end of this section. Given the large role that the public sector plays in the GDP of Kazakhstan, we argue that aggregate TFP is a better measure to capture the overall firm performance of the Kazakh economy. Furthermore, the evolution of TFP typically closely tracks GDP growth, as it reflects also technological progress. Next, we construct the '*Granular Residual*' based on the weighted shocks occurring to the largest firms. Finally, we regress the aggregate TFP growth on this Granular Residual and its lags. As a robustness check, we also use aggregate value added, investment and intangible assets as the dependent variable. The R^2 of this regression gives us the explanatory power of the Granular Residual (based on the idiosyncratic shock) on aggregate growth. In other words, this reflects how much of the aggregate fluctuations are explained by shocks to large firms.

2.1. Estimating TFP

We calculate TFP based on the control function approach of Ackerberg, Caves & Frazer (ACF, 2015). The starting point for TFP is the rationale that firms with a higher productivity will have a higher output for the same type and level of inputs. We can therefore write a classical Cobb Douglas value added production function in logs, explicitly including the productivity variable as follows:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \varepsilon_{it} \tag{1}$$

Where the subscript i denotes the firm and t time. y stands for log output (value added), k for log capital, l for log labor, ω for log productivity and ε is the error term. The main difficulty to correctly estimate productivity lies in the fact that productivity is also correlated with labor and capital, which generates a simultaneity bias and renders the β-coefficients inconsistent.

This concern can be overcome using the control function (CF) approach whereby we proxy the productivity as a function of observables. Olley and Pakes (1996) use investment as a proxy, Levinsohn and Petrin (2003) use intermediate inputs (electricity, fuel, materials) as a proxy. Both methods assume labor inputs to be non-dynamic. ACF (2015) further improves this methodology by allowing labor inputs to be dynamic. Firm's intermediate input demand can be written as a function of capital, labor and productivity, that is:

$$m_{it} = \tilde{f}_t\ (k_{it}, l_{it}, \omega_{it}) \tag{2}$$

and assuming that this function is strictly increasing in ω (strict monotonicity), we can invert the demand function so it becomes a function of observables:

$$\omega_{it} = \tilde{f}_t^{-1} \ (k_{it}, \ l_{it}, \ m_{it}) \tag{3}$$

This can be substituted again into the production function:

$$\begin{aligned} y_{it} &= \beta_0 + \ \beta_k k_{it} + \ \beta_l l_{it} + \tilde{f}_t^{-1} \ (k_{it}, \ l_{it}, \ m_{it}) + \varepsilon_{it} \\ &= \ \widetilde{\Phi}_t \ (k_{it}, \ l_{it}, \ m_{it}) + \varepsilon_{it} \end{aligned} \tag{4}$$

In a first stage, we can find an estimate for $\widetilde{\Phi}_t \ (k_{it}, \ l_{it}, \ m_{it})$. We can then use this estimate in a second stage to estimate the coefficients. Filling these estimated coefficients into the production function gives us a measure for the log productivity, the exponent of which gives us the TFP that we are interested in.

We take value added to measure the output, the number of workers to denote labor, and fixed assets as a measure of capital. We use energy costs as a measure of intermediate inputs. Energy costs are thus the intermediate input demand m_{it}, which we assume to be a function of capital, labor and productivity, as mentioned above.

We estimate TFP separately for 5 different sectors, based on their main activity (NACE code). This allows us to estimate more segmented β-coefficients. The result is normalized by dividing TFP of the firm by the average TFP of the respective sector. TFP growth is measured based on the year-on-year lag in order to account for seasonality in the quarterly data. We also winsorize TFP growth between -1 and 1, to avoid unrealistically high or low values of TFP growth.

We then compute aggregate TFP growth, by weighing individual TFP by the average sales (fixed weight) of each firm, and summing across all firms in each quarter. As a robustness, we also use value added and labor instead of sales as a weighting factor. As the dependent variable, we calculate the aggregate year-on-year TFP growth.

2.2. Composing the Granular Residual

As in Gabaix (2011), we construct the '*Granular Residual*' which reflects the idiosyncratic shocks in the economy. First, we calculate the firm-specific contribution to the Granular Residual. This is defined as the demeaned and weighted firm-level TFP growth ($g_{i,t}$). We demean with the mean TFP growth of the top 1000 largest firms ($\overline{g}_t^Q$), that is Q=1000. The weighting factor ($W_{i,t}$) is the ratio of the lagged sales of the firm over the total lagged sales.[6] As a robustness check, we also use value added and labor instead of sales.

$$W_{i,t} = \frac{Sales_{\ i,t-4}}{Sales_{\ t-4}} \tag{5}$$

The *Granular Residual*, Γ_t, is the sum of these firm-specific contributions, summed up over K number of largest firms. Firms are ranked based on their year-on-year lagged sales. Using the lagged sales ensures that large firms experiencing a negative shock remain in the sample of large firms.

$$\Gamma_t = \sum_{i=0}^{K} W_{i,t}\,(g_{i,t} - \overline{g}_t^Q) \tag{6}$$

In the next step, we regress aggregate TFP growth, Y_t, on the Granular Residual and its two lags. The resulting adjusted R^2 gives us the explanatory power that these variables (which represent the idiosyncratic shocks of large firms) have in explaining the aggregate TFP growth.

$$Y_t = \alpha + \sum_{i=0}^{2} \beta_i \Gamma_{t-i} + \varepsilon_t \tag{7}$$

To test for robustness, we vary the sample size when calculating the Granular Residual. So, we sum the firm contributions over increasing values of k firms. This allows us to plot the R^2 for varying sample sizes. In addition, we repeat the exercise by dropping the largest firms from the sample.

In the robustness section, we examine a number of modifications to this methodology. In particular, we focus on modifying the calculation of the β-coefficients to calculate TFP. Next, we specify different weights used to calculate the aggregate TFP and the Granular Residual. We also use a different TFP estimation technique (Törnqvist index instead of ACF control function approach). Last but not least, we change the dependent variable from

[6] Note that due to the use of the quarterly data we use a 4 period lag.

aggregate TFP growth to aggregate growth in value added, investment and intangible assets.

3. Data

The unique data we use in this paper is obtained from the Bureau of National Statistics (BNS) of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. The BNS collects a longitudinal report from all firms with more than 100 employees, with the exception of banks and organizations of education, healthcare, insurance, pension funds, and government institutions. This data is appropriate for the analysis of granularity, since we focus on large firms.

The dataset contains quarterly firm-level data including the number of workers, as well as a wide range of financial variables such as sales, energy cost and fixed assets.[7] Information on the sector (NACE code) and ownership structure (such as whether the firm is a private firm, a joint venture or foreign owned) is also available. The dataset runs from the first quarter of 2012 to the last quarter of 2018. As standard in this literature, we exclude firms from the energy sector (NACE 35), the financial sector (NACE 64-66) and quaternary sector (NACE 84 and above) from the analysis. However, we keep oil extracting firms in the sample, as they play a key role in the economy of Kazakhstan.

We calculate the cost as the sum of the reported costs of raw materials, components, fuel, energy, other materials and third parties. We calculate the value added as the sales minus these costs and calculate the investment as the quarterly increase in fixed assets plus the depreciation.[8] We drop firms with missing sales or lagged sales as the ranking of the firms is based on the lagged sales data. We also drop firms with missing or zero values for the number of workers, energy or fixed assets, and negative or zero values of value added. We then estimate TFP for each firm in each quarter based on the methodology described earlier. We use the top 1000 firms (ranked on lagged sales) of this final sample to assess the granularity.

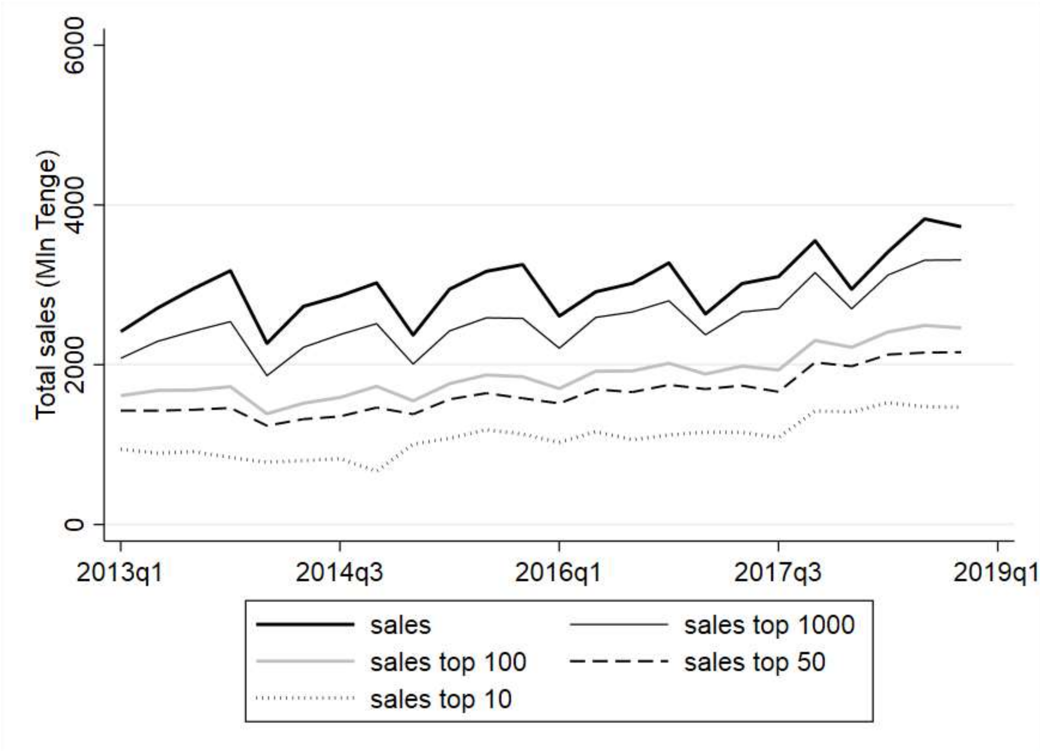
3.1. Summary statistics

Error! Reference source not found. presents the sample. There are 284 workers on average in each firm. An average firm's sales in a quarter is 1,133,000 Tenge, with total

[7] All financial data are deflated with a CPI deflator from Thomas Reuters Eikon database
[8] Depreciation is measured as 20% of the lagged fixed assets (lagged by 1 quarter)

fixed assets in the amount of 3.6 Mln. Tenge, and the cost of energy around 26,000 Tenge.[9] Moreover, we can see that an average firm from the largest 10 firms has over 6 thousand employees, which is significantly more than an average firm from the top 50 or top 100 firms. The same significant difference is observed in sales, value added, and the amount of energy and fixed assets. The means of all variables decrease over the top 1000 firms.

Error! Reference source not found. plots the total sales by quarter for the full sample (black line), as well as those for the top 1000, top 100, top 50 and top10 firms. The figure clearly shows that the top 1000 firms represent a significant share of the total sales and trending very similar to the overall sales, with the top 100 firms accounting for half of the overall sales. These numbers show that the firm size distribution is heavily skewed, with few large firms and many small firms. This is confirmed when we plot the Lorenz curve (see Appendix 2), which shows that 50 percent of the firms in the database only add up to 3 percent of sales, with 5 percent of the largest firms representing 73 percent of sales. Using the Lorenz curve, we also plot the non-negative investments, which are even more skewed than the one for sales (the top 5% largest firms represent 85% of total investments).



[9] 1 USD = 149.11 Tenge in 2012, 1 USD = 382.75 Tenge in 2019

Figure 1: Overview of the total sales (Mln. Tenge) of firms in the database

Error! Reference source not found. also indicates that total sales demonstrate a strong seasonality with sales steadily increasing throughout the year, reaching the highest sales in the fourth quarter followed by a steep decline in the first quarter of the next year. To account for this seasonality, we measure lagged variables as year-on-year lags.

This dominance of a few firms in sales is reflected in the skewed size distribution of firms. **Error! Reference source not found.** describes the skewness and kurtosis when all firms in the sample are plotted according to their size (based on log sales). The results are provided for each quarter of the year 2016. This year is randomly chosen, though is representative of the other years in the sample. The results for all quarters (2012-2018) can be found in the Appendix 3. A normal distribution is characterized by a skewness of 0 and a kurtosis of 3. We find that the skewness lies around 0.4, varying from 0.198 to 0.420 for the different quarters of 2016, indicating the distribution is slightly skewed. For each quarter, we find excess kurtosis (>3), in this case the distribution is called leptokurtic, which is an indication of a non-normal distribution.

Leptokurtic distributions are characterized by a higher peak, thinner 'shoulders' and fatter tails. This can be seen from the Kernel density plots shown in Figure 2, where we plot the firm size distribution (based on log sales) for each quarter of 2016, focusing on the top 1000 firms. The kernel density for all firms in the sample can be found in Appendix 4, which shows a similar pattern.

We see that in the segment of the very large firms (log sales between 16 and 18), there are more firms than what would be expected under a normal distribution. At the same time, we see an underrepresentation of large firms in the range of log sales smaller than 16 but larger than 14.

We also plot the quantiles of log sales against the quantiles of the normal distribution for each quarter of 2016 (for the sake of brevity, we add the graphs in Appendix 4). Also here, we see that deviation from the normal distribution occurs only in the tails of the distribution.

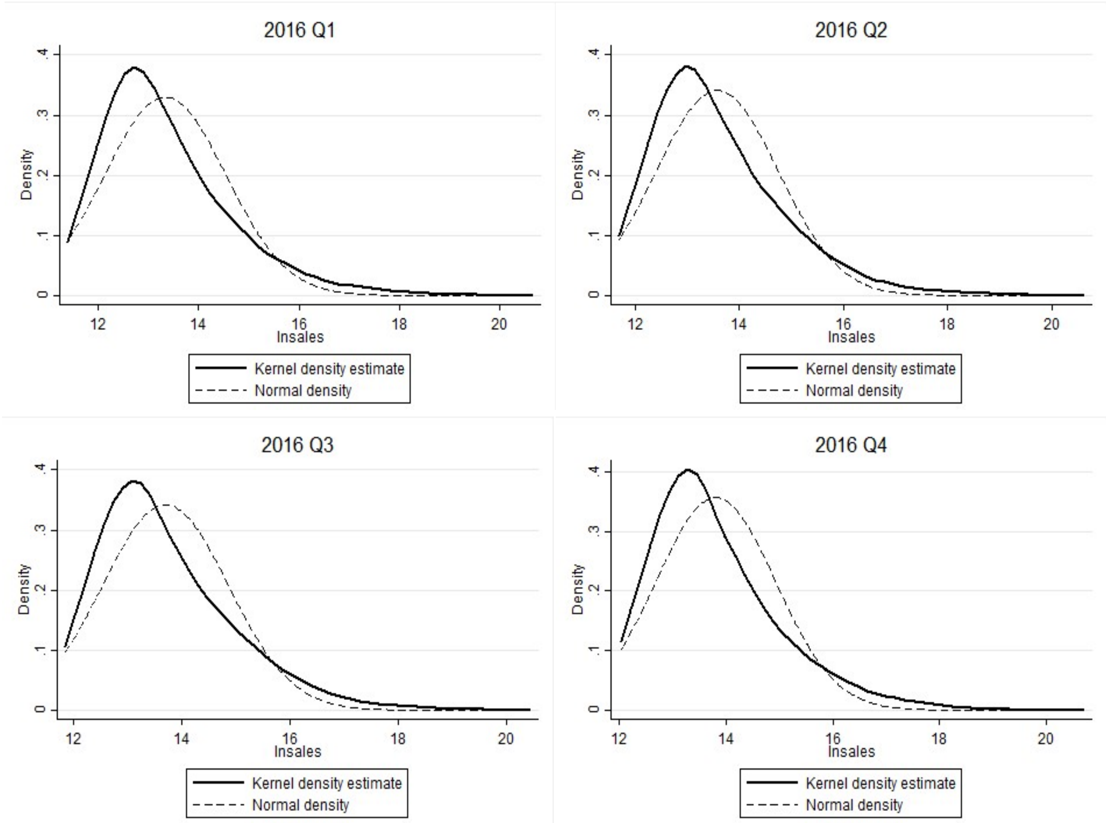


Figure 2: Kernel density plots of log sales for top 1000 firms, for each quarter of 2016

3.2. Sectoral decomposition and firm heterogeneity of the top 1000 firms

We now focus on the top 1000 firms, and decompose these firms in 5 different sectors based on their field of activity (given by the respective NACE code). The first sector combines agriculture and mining; the second focuses on the manufacturing industry; the third on the construction and utilities sector (water and waste). The fourth sector combines wholesale, retail and logistics. The last sector combines all the other services. As mentioned before, the energy, financial and quaternary sector are excluded from the scope. Further details on which NACE 2-digit codes belong to which sector can be found in Appendix 1.

Almost one third of the top 1000 firms are active in the manufacturing sector (NACE 10-33), as can be seen in Figure 3. In terms of sales, the largest sector is the agriculture and mining sector (see Figure 4). This sector is driven by the oil sector (NACE 06). Despite strong fluctuations per quarter, the mining and manufacturing sectors show an overall increase in sales over the years. The representation of the other sectors in the top 1000 is relatively stable of the years, both in number of firms and total sales.

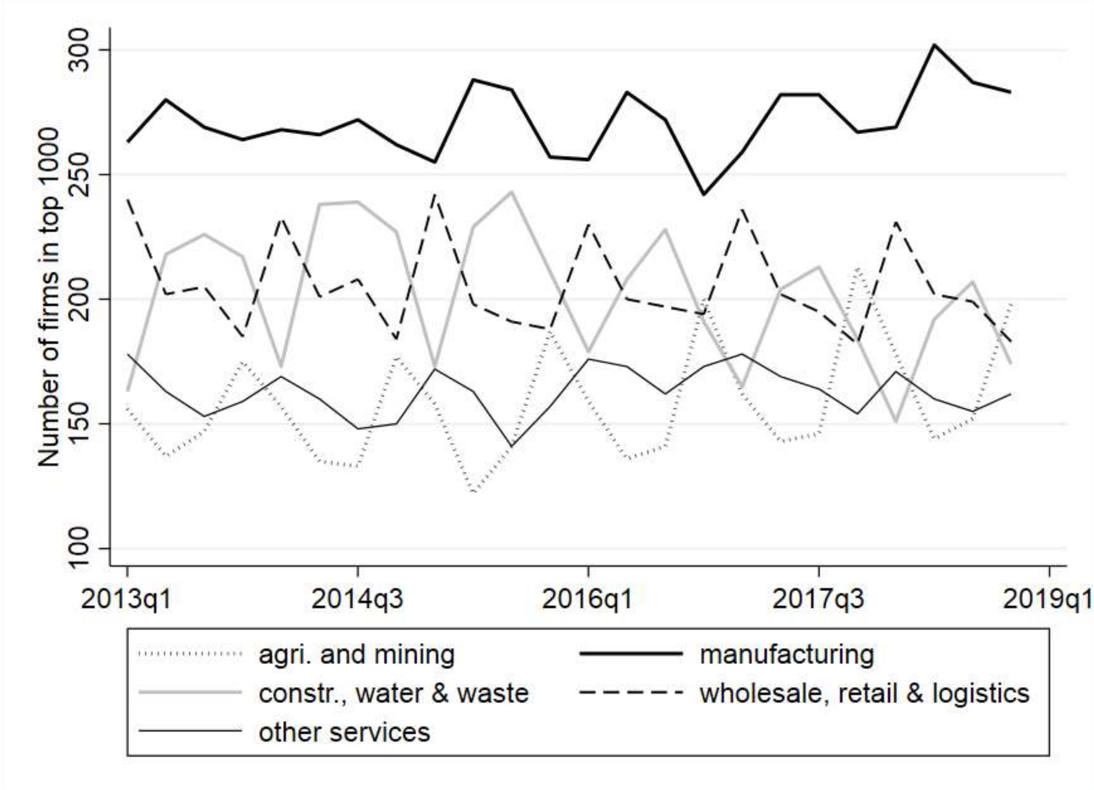


Figure 3: Number of firms in top 1000, by sector

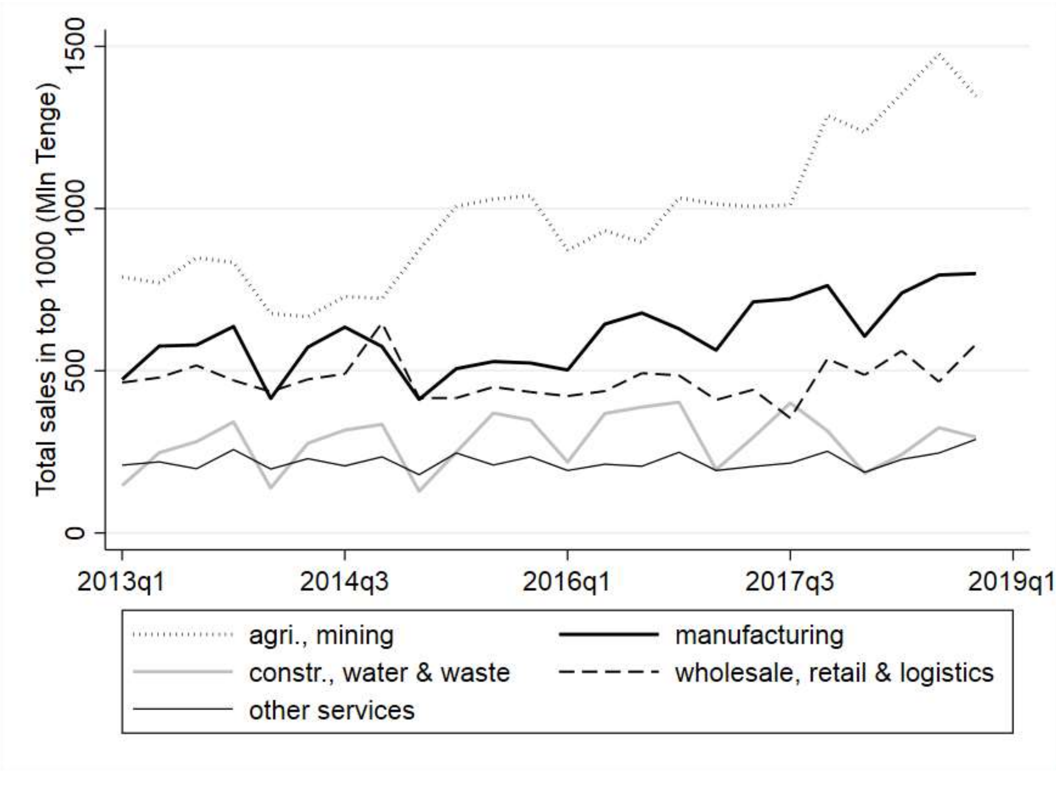


Figure 4: Total sales (Mln. Tenge) of firms in top 1000, by sector

In Figure 5, we plot the sales of each of the top 1000 firms by NACE 2-digit sector for the first quarter of 2016. The results are robust to selecting a different quarter or year. This figure clearly demonstrates that one firm in the oil sector (NACE 06) is disproportionately large compared to the other large firms. Its sales are almost three times larger than that of the second largest firm. We also repeat this exercise but excluding this largest firm. We can see that the distribution of the largest firms is not equal across all sectors. A number of large firms can be found in a few sectors, notably the oil sector (NACE 06), the metal industry (NACE 24), land transport (NACE 49) and telecommunications (NACE 61).

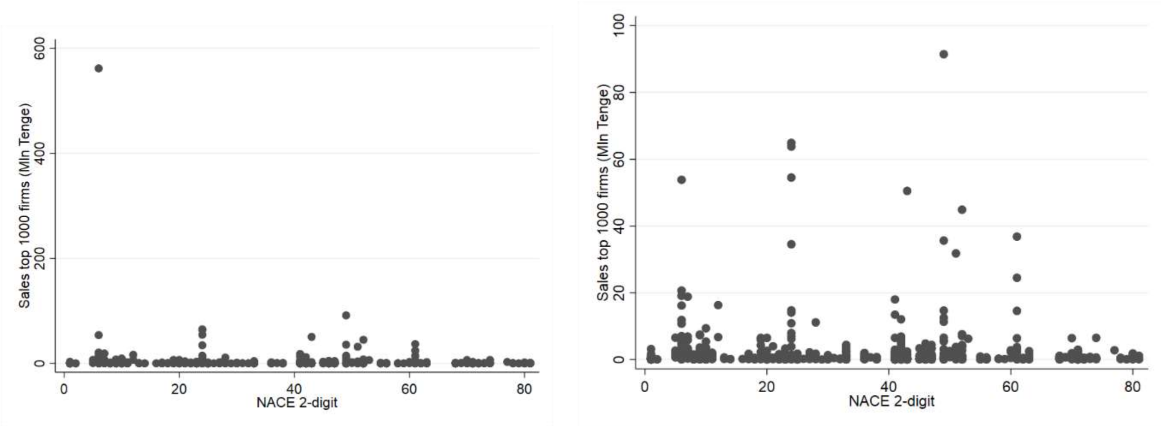


Figure 5: Sales per firm in top 1000, by NACE code, in year 2016 quarter 1, LHS: including all firms in top 1000, RHS: excluding the largest firm (in NACE 06)

4. Results

4.1. Normal distribution versus power law

The firm size distribution is an important aspect to assess whether an economy in general may exhibit granular properties. The Central Limit Theorem stipulates that a probability distribution converges to a normal distribution as the sample size (n) goes to infinity. Gabaix (2011) demonstrates that in case of a distribution with thin tails, the convergence rate to a normal distribution is $1/\sqrt{n}$. However, in the case this distribution follows, for example, a power law (or in the extreme case a Zipf distribution) characterized by fat tails, then the rate of convergence is much smaller, at $1/\ln n$. This explains why even large sample sizes ($n = 10^6$) might not be sufficiently large for the distribution to converge to a normal distribution. The previous section already demonstrated that the firm size distribution in our sample deviates from the normal. Thus, the Central Limit Theorem breaks down in cases with fat tail distributions. One can no longer assume that shocks to large and small firms cancel each other out, thus allowing the concept of granularity to kick in.

In order to check whether the right-hand tail of the distribution follows a power law, we plot the rank of the 200 largest firms against the sales (see Figure 6, LHS). The highest rank (1) is given to the firm with the highest sales. In this way, the rank gives the probability that we will find a firm with a higher sales, plotted against the sales. Such a graph is called a Complementary Cumulative Distribution Function (CCDF) (Blanco-Arroyo & Alfarano (2017)). From this graph, we can clearly see that the largest firm is a definite outlier in terms of sales. We also see that the difference in sales between firms ranked 1 and 2 is larger than firms ranked 2 and 3, which in turn is larger than firms ranked 3 and 4 and so on.

In general, a power law follows the functional form:

$$P(x) \propto x^{-\alpha} \tag{8}$$

$P(x)$ denotes the probability to encounter x (given by the rank), x in our case is the sales of the firm. Therefore, if the tail of the distribution follows a power law, then plotting log rank against log sales should provide a straight line with the coefficient equal to α.[10] We can see from Figure 6, RHS, that this is indeed the case. This suggests that the largest firms in the sample follow a power law distribution and shocks to these large firms can persist in the aggregate.

To test this in a more precise manner, Clauset et al. (2009) propose a novel approach to estimate the coefficient (or tail index) by first estimating the lower bound in the distribution from which the power law starts. Following this approach, we are able to estimate the tail index for each quarter. The results by quarter are given in Appendix 5. We find a declining trend in the tail index as shown in Figure 7. This indicates that the tails are becoming fatter over time. In other words, the top firms are becoming more dominant over time. This is also visualized by comparing the CCDF plot of the 3rd quarter of the first and last year of the dataset (2012 compared to 2018), which demonstrates that the coefficient declines as the largest firm becomes even larger over time (see Appendix 5).

[10] As pointed out by Gabaix and Ibragimov (2011), rather than regressing ln rank on ln sales to obtain the coefficient, the results can be improved by regressing ln rank – ½ on ln sales.

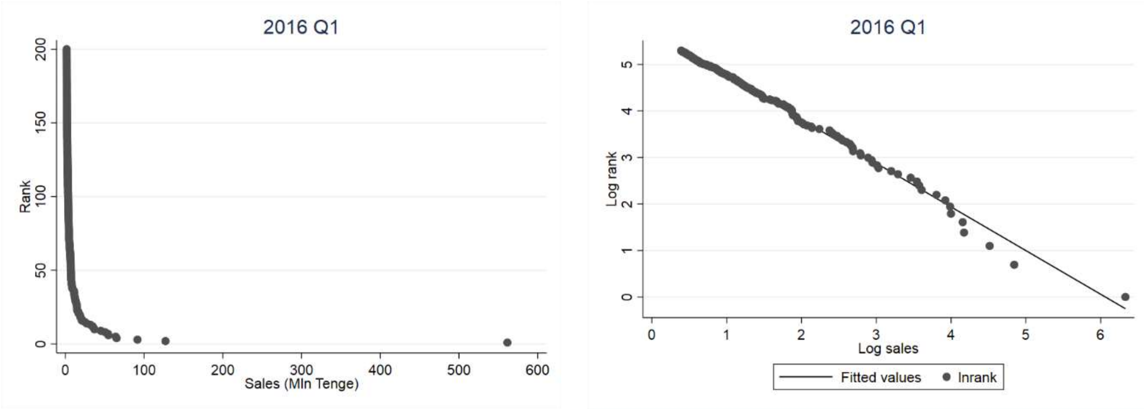


Figure 6: CCDF plotting rank on sales (LHS) and log rank on log sales (RHS), for the top 200 highest ranked firms in the first quarter of 2016

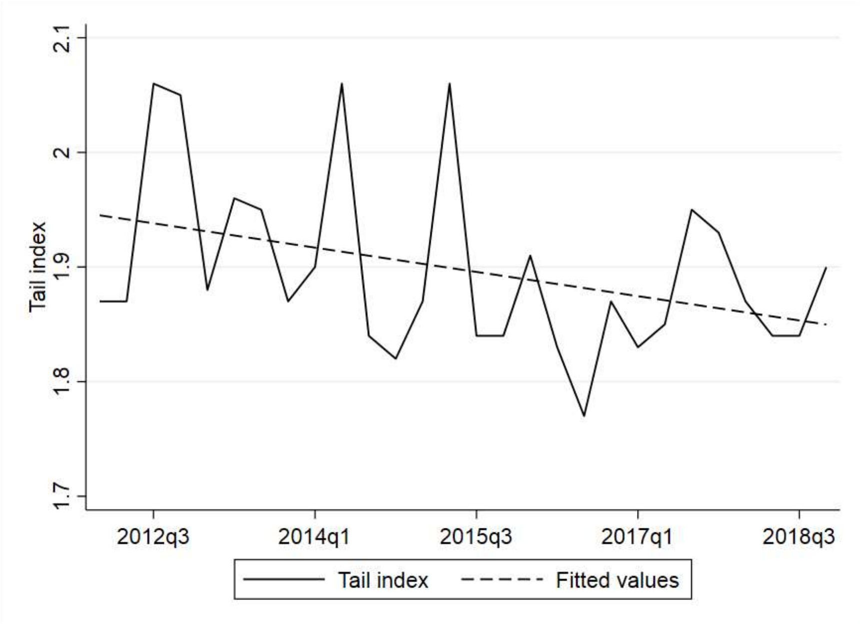


Figure 7 : Tail index by quarter, with a declining trend over time

4.2. Total factor productivity

As explained in the methodological part of this paper, we calculate the total factor productivity of each firm. We use the ACF control function approach, where we use energy cost as a proxy for intermediate inputs. The β-coefficients that we obtain are given in **Error! Reference source not found.**. We repeat this calculation for each of the 5 sectors that we have defined. Assuming a log linearized Cobb-Douglas production function, we can obtain a productivity measure for each firm using the respective sector specific coefficients.

As a robustness check, we extend the intermediate inputs from only energy cost to all intermediate costs (for which data are available): energy, fuel, raw materials, components, other materials and third parties. The results are comparable and given in the robustness section.

4.3. Testing Granularity

A granular economy is characterized by an extremely low number of large 'granular' firms that explain a non-trivial part of aggregate fluctuations. In line with Gabaix (2011), we measure this explanatory power with the R^2 that results from regressing the aggregate fluctuations on the Granular Residual and its lags (see the methodology section for further details). In line with Blanco-Arroyo (2018) we do not fix the sample size at 100, and assess the explanatory power for varying sample sizes.

Error! Reference source not found. shows the results when increasing the sample size from 1 to 50 firms.[11] The explanatory power of the single largest firm (K1) is already high, with an R^2 of 0.445. When we add more firms to the sample, the R^2 increases in general, reaching an R^2 of 0.780 when the 10 largest firms are included in the sample. The R^2 reaches a plateau at around 0.8. Once this plateau is reached, an increase in sample size has no further incremental effect on the R^2 value. It remains at the same level whether the sample size is 50, 100 or 200.

These results indicate that the Kazakh economy is granular with the idiosyncratic shocks in TFP growth to the top 30-50 firms explaining a substantial proportion of the fluctuations in aggregate TFP growth.

Figure 8 visualizes the changes in R^2 for increasing sample sizes, from 0 to 200, increasing by 1 firm at a time. Note that the values of the R^2 correspond to those given in **Error! Reference source not found.** for the line R^2, where no firms were dropped from the sample. The graph is typical for a granular economy: initially the R^2 responds to adding additional granular firms, and at a certain point, adding more firms no longer has incremental explanatory power, and the curve reaches a plateau.

[11] The regression results for different sample sizes can be found in Appendix 6.

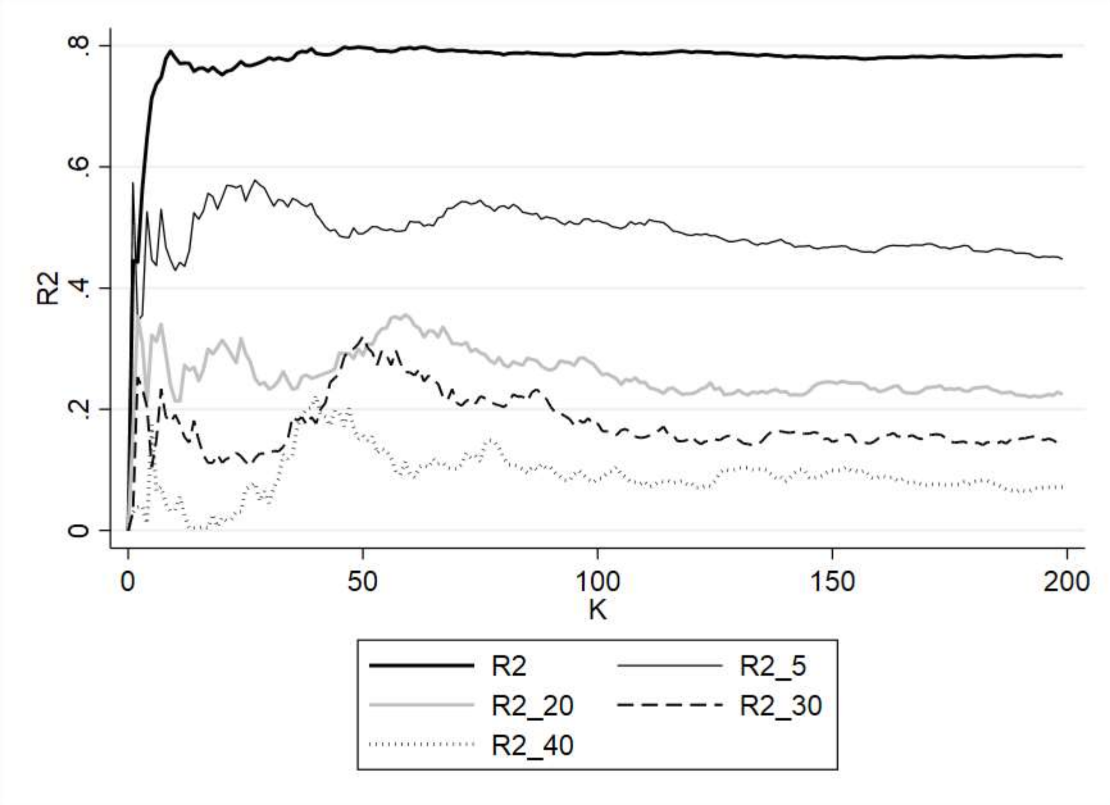


Figure 8: The explanatory power (R^2) for varying sample sizes when the largest firms are dropped

In line with the phenomenon on granularity, dropping the largest (granular) firms from the sample should reduce the overall R^2 until the point where the R^2 reaches a baseline, and dropping additional firms has little to no impact on the estimated R^2 (see also Blanco-Arroyo (2018)). The results obtained when analyzing the Kazakh economy confirm this phenomenon. Dropping the 5 largest firms from the sample (line R2_5) reduces the R^2 plateau to less than 0.5. Dropping the 20 largest firms reduces this plateau even further. When we drop the 40 largest firms from the sample, the remaining firms have almost no explanatory power left (R^2 close to 0). This confirms our earlier results that the top 30-50 firms in the Kazakh economy are indeed granular.

Similar to Blanco-Arroyo (2018), we also plot the cumulative curve. This graph is generated by taking the average of the R^2s we obtain over all sample sizes. In our case, we increase the sample sizes from 1 to 200, in steps of 5. We then repeat this exercise but gradually increase the number of largest firms dropped from the sample. In our case, we drop between 1 and 50 firms from the sample, 1 firm at a time. The number of firms dropped are denoted by 'L' on the X-axis of Figure 9. To illustrate, the R^2 for L=20 corresponds to the average R^2 based on 40 regressions, whereby the sample size increases each time by 5 firms, from 1 firm (with rank 21) to 200 firms (rank 221). Also, this cumulative curve demonstrates that the Kazakh economy has a limited number of granular firms, with the largest firms having most explanatory power. The most important decrease in explanatory power occurs in the top 30 firms.

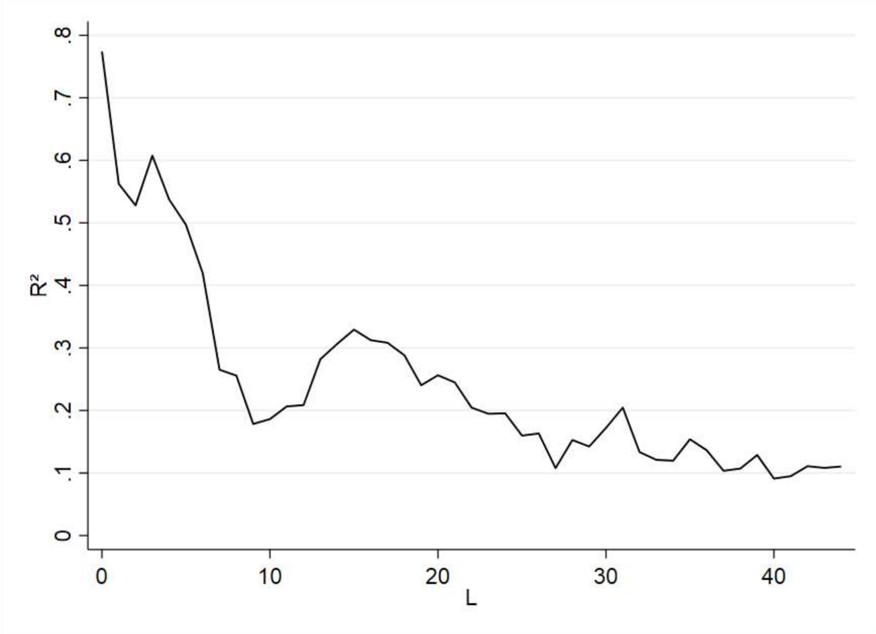


Figure 9: Cumulative curve (R^2 from increasing number of firms dropped)

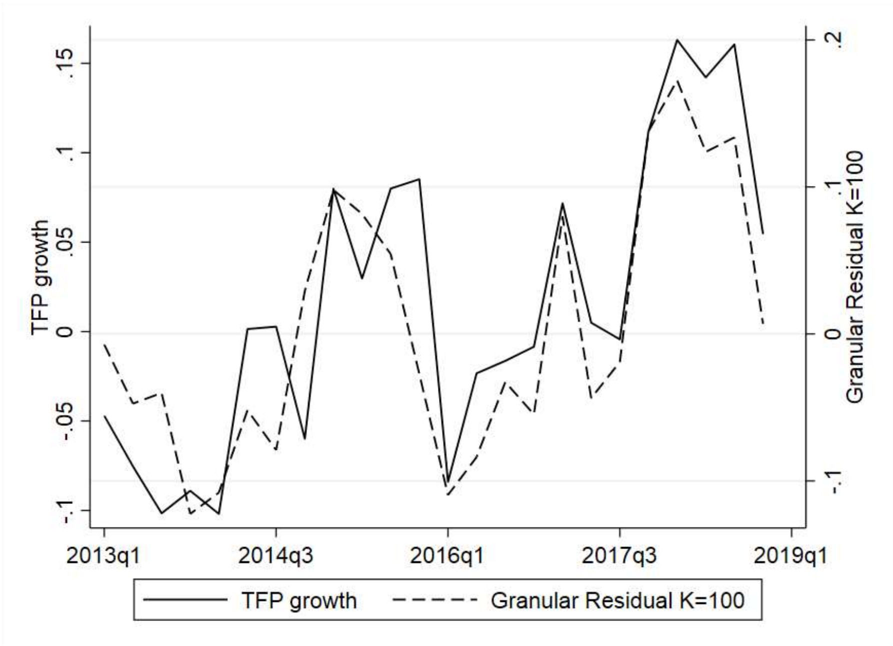


Figure 10: Comparing the evolution of aggregate TFP with the granular residual of the top 100 granular firms

To visualize the relationship with aggregate TFP growth, Figure 10 plots aggregate TFP growth for the period 2013-2018 (full line). In addition, we add the evolution of the Granular Residual, calculated for the top 100 highest ranked firms. We see a similar pattern in the evolution of TFP growth and the evolution of the Granular Residual. This correlation explains the relatively high R^2 value that we find in the regressions.

4.4 Robustness checks

In what follows, we show that the main conclusions still hold even when we modify parts of the methodology. First, we modify the control variable which is used to estimate TFP. Second, we modify the weights used to estimate the Granular Residual and the aggregate TFP. Third, we also estimate TFP using a different methodology, the Törnqvist index. Last but not least, we modify the dependent variable.

4.4.1 Modifying the control variable

Table 5 presents the β-coefficients when the control variable includes all intermediate inputs and not only energy. These intermediate inputs are calculated as the sum of energy, fuel, raw materials, components, other materials and third parties. These β-coefficients are in line with the results presented in Table 3. The choice of variable used to proxy intermediate inputs does not seem to affect the results.

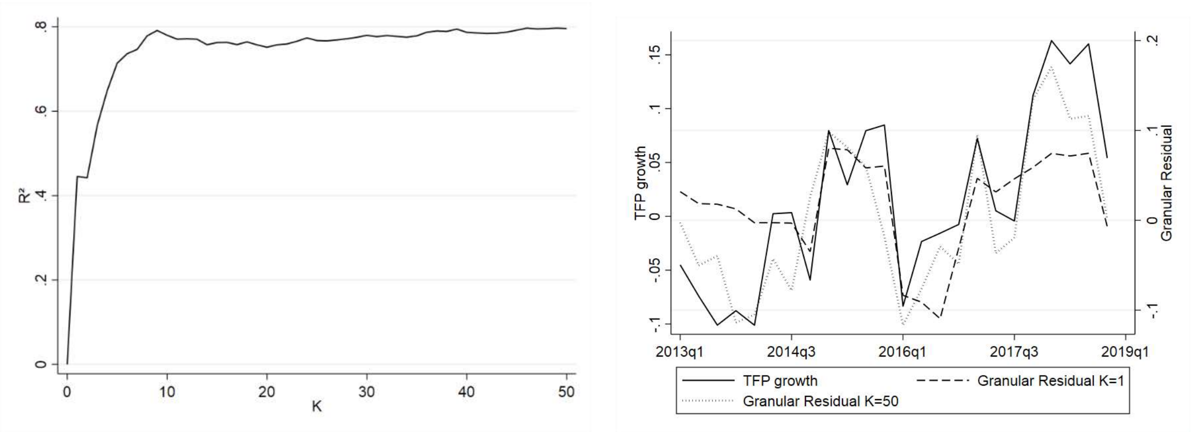


Figure 11: LHS: The R^2 value for varying sample sizes; RHS: aggregate TFP growth over time and the GR of the top 1 and top 50 firms over time.

Figure 11 plots the R^2 value for varying sample sizes (LHS) as well as the correlation between aggregate TFP growth and the Granular Residual of the largest firm and the top

50 largest firms (RHS). The Granular Residuals of the largest firms and aggregate TFP growth have largely similar fluctuations, which explains the high plateau of nearly 0.8.

4.4.2 Modifying the weighting factor

In the main results, both the TFP aggregate and the Granular Residual are weighted based on sales. In Figure 12 we use value added instead of sales for both weights, in Figure 13 we use labor instead of sales.

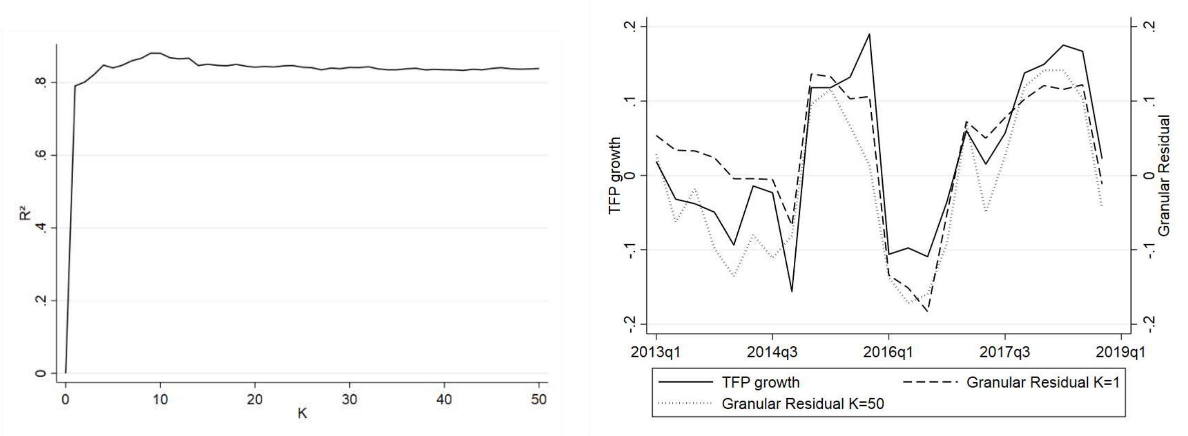


Figure 12: Aggregate TFP and Granular Residual weighted with value added instead of sales. LHS: The R^2 value for varying sample sizes; RHS: aggregate TFP growth over time and the GR of the top 1 and top 50 firms over time.

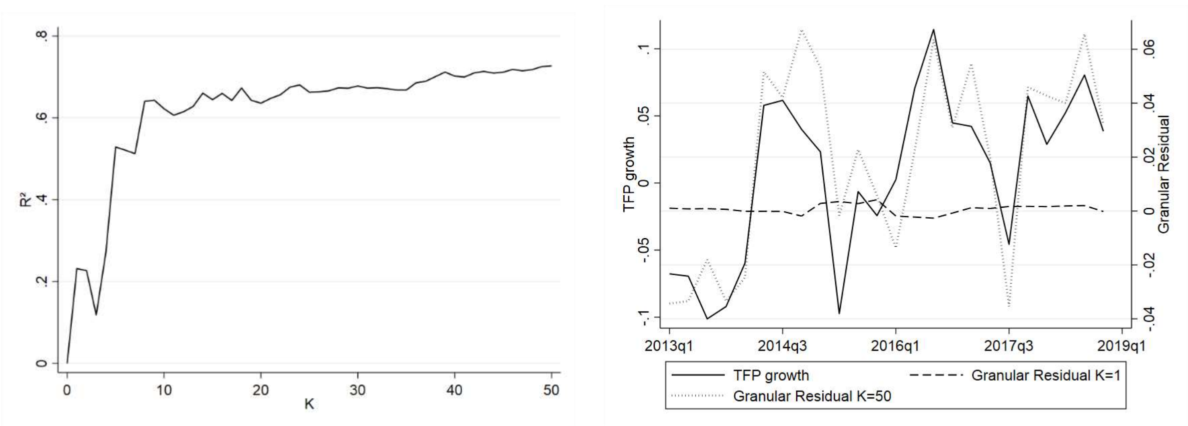


Figure 13: Aggregate TFP and Granular Residual weighted with labor instead of sales. LHS: The R^2 value for varying sample sizes; RHS: aggregate TFP growth over time and the GR of the top 1 and top 50 firms over time.

When using labor as a weight, we reach an even slightly higher plateau compared to using sales as a weight. When using labor as a weight, the GR of the largest firm is very stable over time. The R^2 values show more variation with increasing sample size, but still reaching a plateau at around 20-30 firms.

4.4.3 Törnqvist index

Next, we move away from estimating TFP using a control function approach and opt for an index approach. The Törnqvist index is an index for productivity, which is the result of an aggregate output index divided by an aggregate input index. The output index corresponds to the output (value added). The input index is a combination of labor and capital, each weighted with a share. The share of labor is based on the wages (payroll), the share of capital cost is estimated to be 1 - the share of wages. Figure 14 plots the R^2 when we use a Törnqvist index to calculate TFP rather than through a control function approach. In this case, we weight the aggregate TFP and the Granular Residual with value added instead of sales. We can see that the plateau at 0.9 is reached already after the first firm is included.

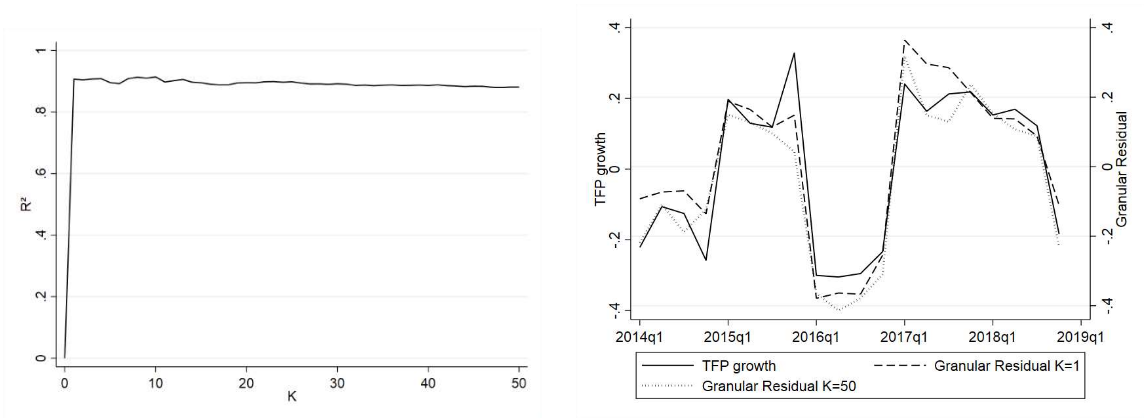


Figure 14: TFP based on Törnqvist index. Aggregate TFP and Granular Residual weighted with value added instead of sales. LHS: The R^2 value for varying sample sizes; RHS: aggregate TFP growth over time and the GR of the top 1 and top 50 firms over time.

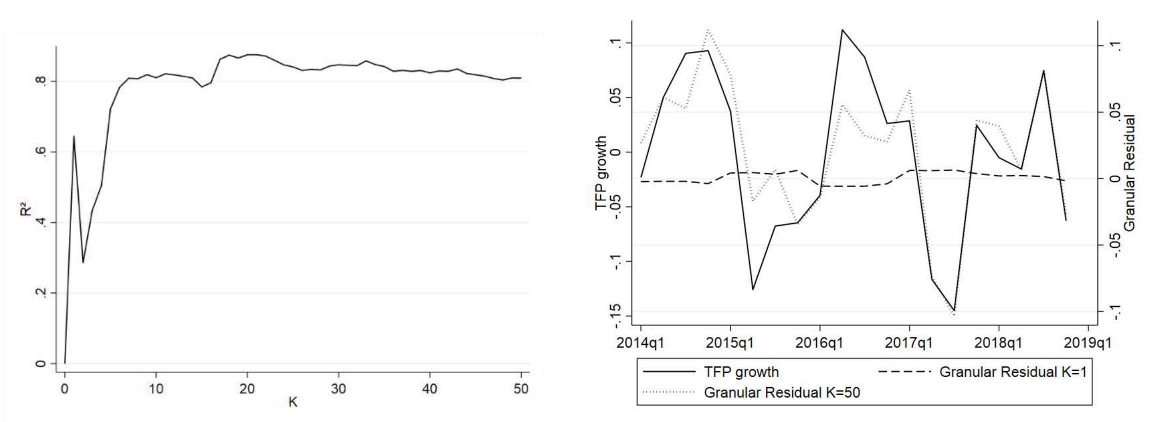


Figure 15: TFP based on Törnqvist index. Aggregate TFP and Granular Residual weighted with labor instead of sales. LHS: The R^2 value for varying sample sizes; RHS: aggregate TFP growth over time and the GR of the top 1 and top 50 firms over time.

Figure 15 also calculates TFP based on the Törnqvist index but uses labor as a weighting factor instead of value added in order to calculate the aggregate TFP and the Granular Residual. Similar to the case where TFP is estimated with a control function approach, we see that the GR of the largest firm is very stable over time.

In general, the results obtained when TFP is based on the Törnqvist index are very similar to the results based on the ACF method, both when the weighting factor is based on value added or labor.

4.4.4 Modifying the dependent variable

We also change the dependent variable. First, we use aggregate value added growth instead of aggregate TFP growth. Also here, we can show the Kazakh economy to be granular, with the single largest firm accounting for most impact. The results hold both when we estimate TFP using the control function approach (Figure 16) or the Törnqvist index (Figure 17). In both cases, the weighting factor used for the Granular Residual is value added instead of sales.

We see that with either method used to estimate TFP, we obtain similar results and that the granular hypothesis is confirmed even when we choose aggregate value added as dependent variable.

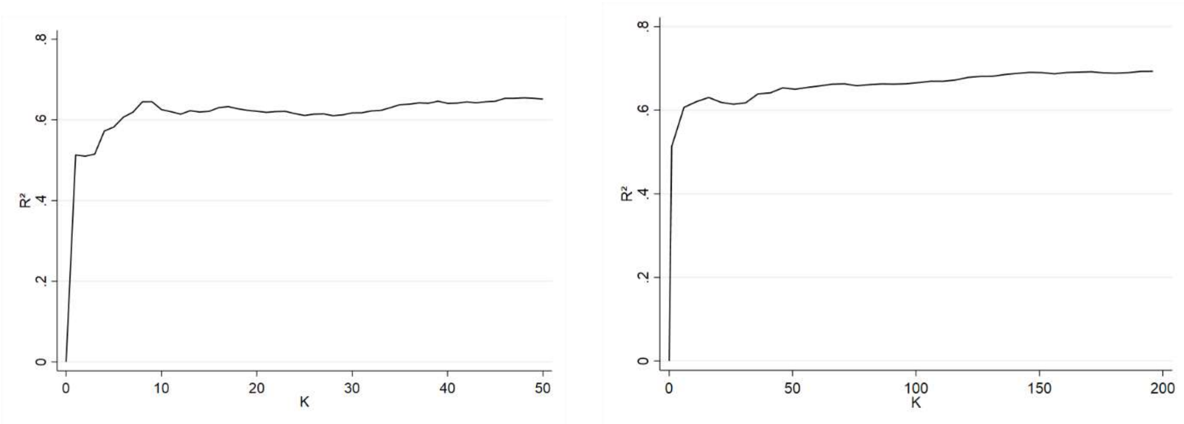


Figure 16: Aggregate added value growth as dependent variable. Granular Residual weighted with value added instead of sales. TFP estimated with control function approach. LHS: The R^2 value for varying sample sizes, in steps of 1 up to 50; RHS: The R^2 value for varying sample sizes, in steps of 5 up to 200.

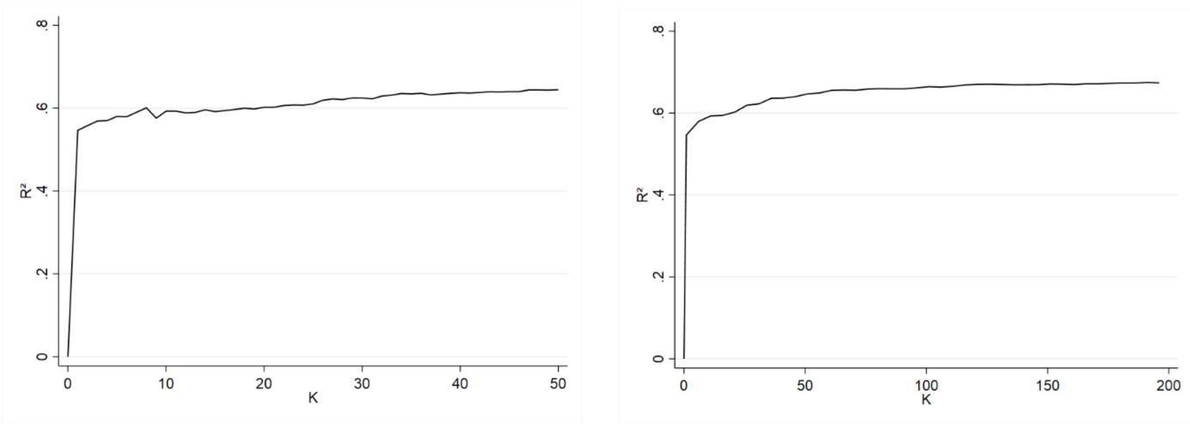


Figure 17: Aggregate added value growth as dependent variable. Granular Residual weighted with value added instead of sales. TFP estimated with Törnqvist index. LHS: The R^2 value for varying sample sizes, in steps of 1 up to 50; RHS: The R^2 value for varying sample sizes, in steps of 5 up to 200.

We then calculate again the cumulative curve (see Figure 18). For the regressions, we use aggregate value added growth. We estimate TFP (in order to obtain the Granular Residual) with the Törnqvist index and weight the Granular Residual with value added. As in the results section, we obtain the cumulative curve by dropping between 1 and 50 firms from the sample. Under these specifications we observe that most granularity resides with the single largest firm.

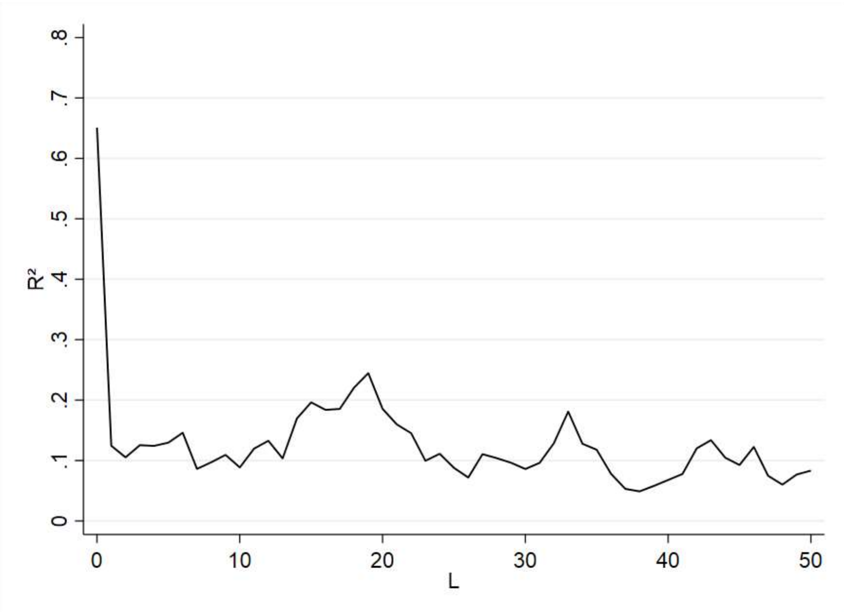


Figure 18: Aggregate added value growth as dependent variable. Cumulative curve.

As a final robustness check, we change the dependent variable to investment or intangible assets. As we can see in Figure 19 the granular firms are able to explain fluctuations in the growth patterns of these variables. Adding additional non-granular firms to the sample does not further increase the explanatory value.

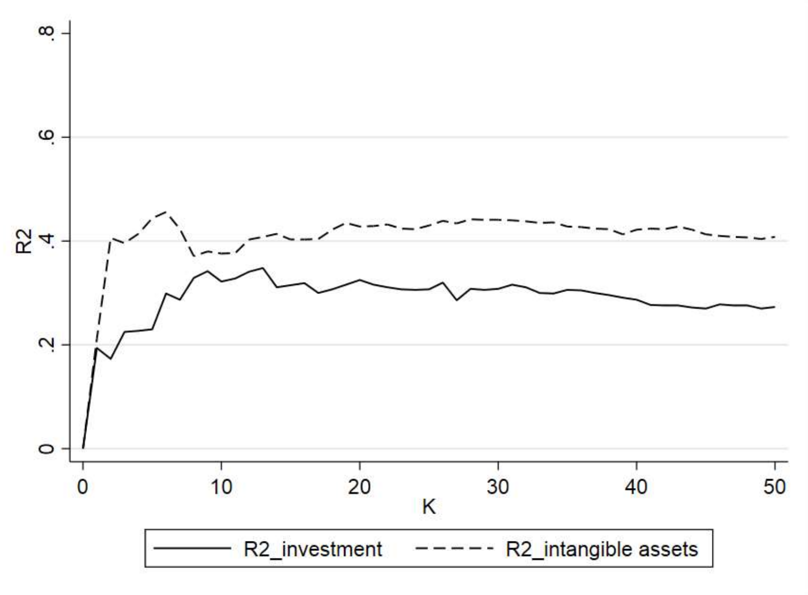


Figure 19: Aggregate investment growth as dependent variable (dashed line), aggregate growth in intangible assets as dependent variable. Granular Residual weighted with sales. TFP estimated with control function approach.

4.5 Characterizing the granular firms

Now that we have established that the top 30 largest firms are granular, we characterize these firms in terms of type and sector. It is important to note that these 30 firms are ranked based on their lagged sales. As the sales of firms varies from quarter to quarter, the composition of the 30 largest firms will vary in each quarter. The largest firm remains the same for the whole sample period (this firm outweighs all other firms as can be seen from Figure 5). This firm is active in the oil extraction industry, and its legal status is a joint venture with foreign participation.

Table 6 gives an overview of the percentage of the 30 highest ranked firms over the entire period, grouped by type. All firms are categorized in 1 out of 5 types, based on their legal status: (i) firms without any state or foreign participation, (ii) firms with state participation, (iii) firms that have a joint ownership with foreign participation, (iv) foreign owned firms and, (v) other type. We see that across all quarters, more than half of the firms are a joint ownership or foreign firms (56 percent). Around 40 percent of the top 30 firms in each quarter are Kazakh private firms, without state or foreign participation. Firms with state participation only represent about 3 percent of granular firms. We observe that state owned firms or firms with state participation are not driving the aggregate TFP growth. Appendix 7 provides more information on the distribution of the different types of firms with increasing sample size. We see that the share of foreign firms drops from 56 percent in the top 30 to 20 percent in the top 1000, which is still a relatively large fraction.

Figure 20 sorts all firms that rank amongst the top 30 firms in at least one quarter based on their legal status. Over time and in particular as of 2017, the share of foreign firms increases at the expense of the Kazakh firms.

Table 7 considers the sector and NACE field of activity, rather than the legal status. The full description of the NACE codes can be found in the Appendix 1. From this table, we can see that the oil industry (NACE 06) is the sector most represented amongst the granular firms, followed by NACE 24 (metal industry) and NACE 49 (land transport).

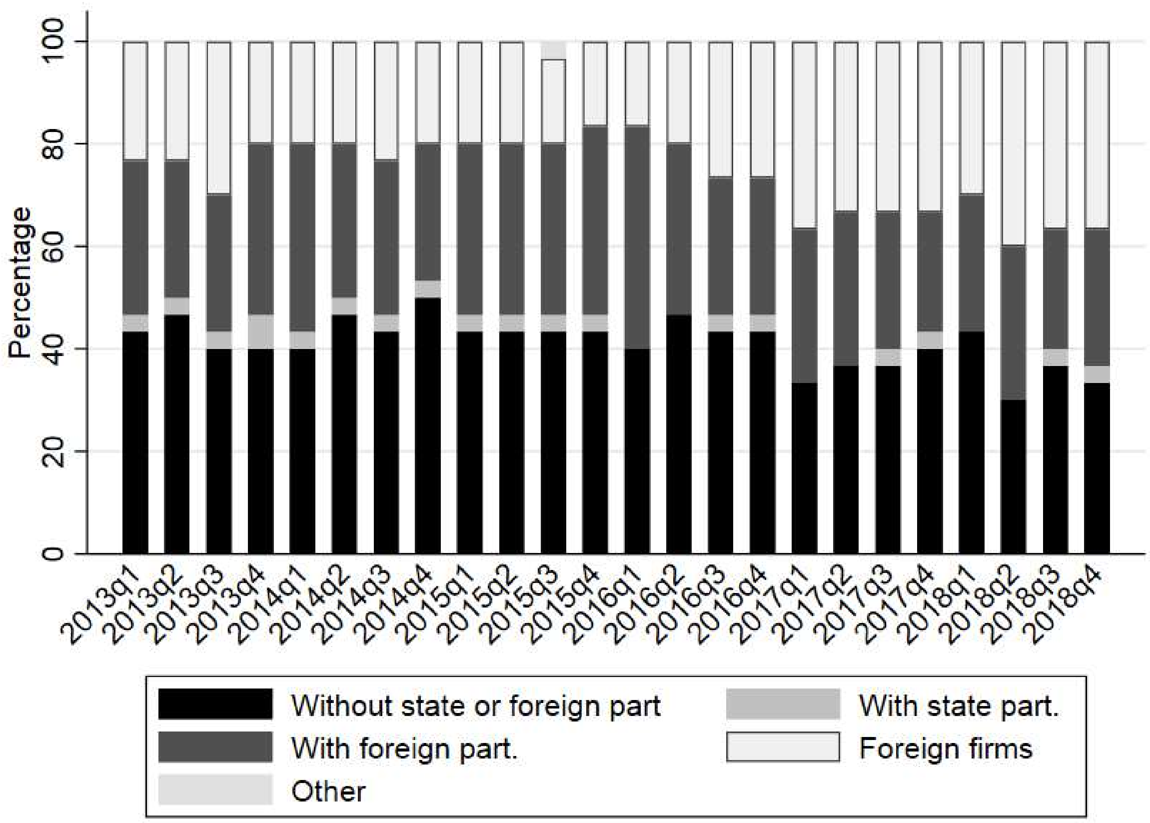


Figure 20: Top 30 granular firms by type and by quarter, in percentage

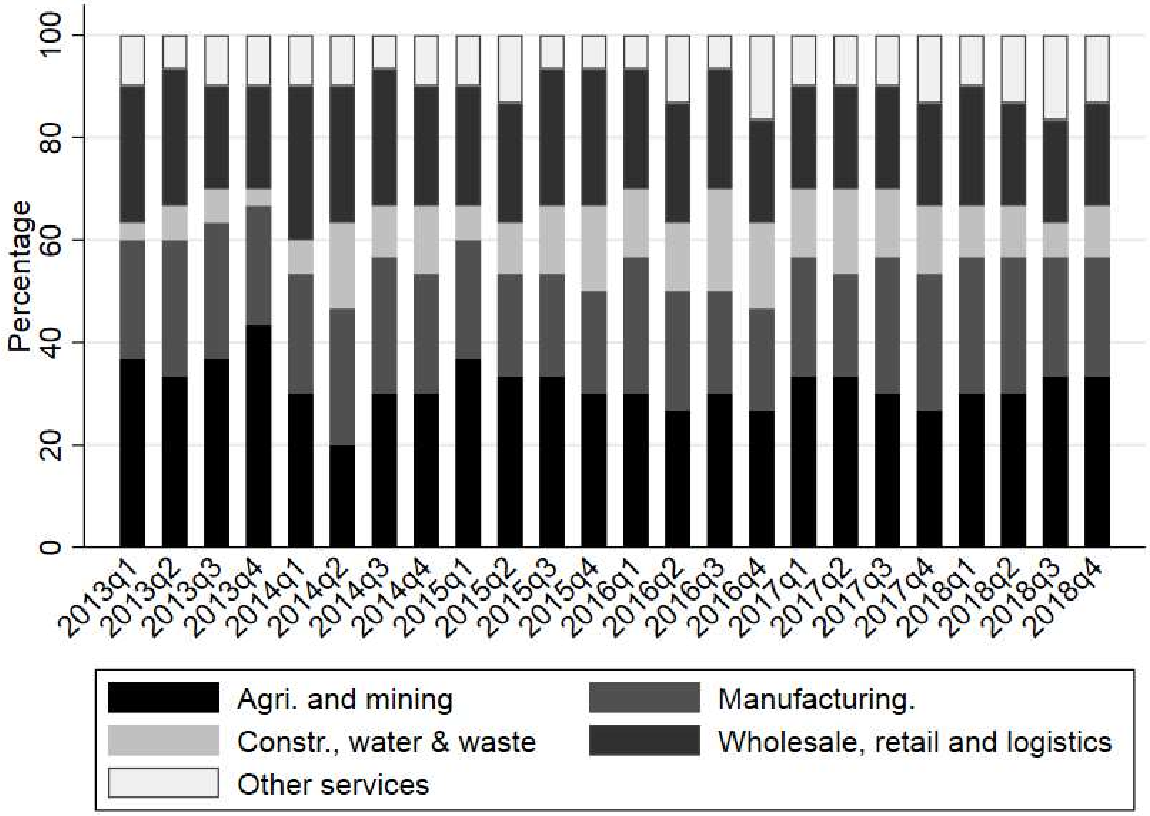


Figure 21: Top 30 granular firms by sector and by quarter, in percentage

Figure 21 plots the top 30 ranked firms based on their sector of activity. More than 50 percent of these firms are active in the agriculture, mining and manufacturing industry. The main NACE sectors in which granular firms are active are NACE 61: extraction of crude petroleum; NACE 244: manufacture of basic precious metals; and NACE 241: manufacture of basic iron and steel and of ferro-alloys. The importance of these sectors varies by quarter but remains relatively constant over time. Appendix 7 provides further details on the distribution of these sectors when we increase the sample size up to 1000 firms. Where the oil, metal and land transport sector account for 59.3 percent of the top 30 firms, they only account for 8.2 percent of the top 1000 firms.

Table 8 represents the percentage share when we consider together the NACE field of activity and the type (legal status) of the firm. From this table, we learn that the majority of the oil firms (NACE 06) have a joint ownership and there is only a limited number of foreign firms. With regards to the metal industry (NACE 24), we note that also here joint ownership is the most prominent type, but foreign firms occur almost twice as much as domestic firms. Nearly all firms in land transport (NACE 49) are firms without state or foreign participation.

5. Discussion and conclusion

This paper demonstrates that shocks affecting individual firms are important to understand macroeconomic fluctuations, confirming earlier results for the U.S. and for a number of European countries. We show that the 'granular hypothesis' also holds for an emerging and transitional economy such as Kazakhstan. In fact, the granularity is even more pronounced, with the 30 largest firms explaining nearly 80 percent of aggregate fluctuations in total factor productivity. Thus, a smaller group of large firms seem to dictate what is going on at the macroeconomic level. This compares to the top 450 firms identified in Spain.

In addition, state owned firms or mixed ownership firms do not play a major role in these top 30 firms. More than half of the granular firms are joint ventures with foreign participation or foreign owned firms. This immediately reveals that international linkages may also be an important factor to understand business cycles in Kazakhstan. In terms of sectors, granular firms are mainly operating in the oil and metal industry.

The higher level of granularity might also indicate certain specific features of such emerging economies. These economies may be less diversified, with more monopolistic and oligopolistic industry structures than in developed economies suggesting even more concentration.

As a consequence, we can also relate these findings to the type of capitalism that some emerging countries like Kazakhstan have adopted since the collapse of communism. In general, countries that emerged out of the communist bloc had two basic approaches to their subsequent economic model. One was a predominantly market oriented model that was accompanied by relatively rapid privatization and integration into the global economy and sometimes into larger economic blocs. The second was a more managed transition to a semi market-based economy where the state still played a major role, either directly through ownership or indirectly through policies that favored domestic players (or both). While former communist bloc countries in Eastern Europe followed the first model (Nölke, ten Brink, May and Claar, 2020; Feldman, 2006), countries such as Kazakhstan tended to develop their economic transition along the second model (Abilova and Subramanian, 2020).

Emerging countries in the Asian region typically followed a path of rapid industrialization, where the governments were focused on catching up with the developed economies. Initially focused on import substitution, these countries protected domestic markets from foreign competition and allowed domestic players to build scale and dominate the domestic market. This policy of import substitution was subsequently complemented with a policy of export promotion, which was later complemented by building global brands and multinational operations (Witt and Redding, 2013, 2014).

An important characteristic of the East Asian economic model was that the different Asian economies were not particularly resource rich and based their comparative advantages on low labor costs and productivity. However, countries such as Kazakhstan have used their resource endowments to pursue a different model of transitioning to a market-based economy. In particular, the Kazakh government initiated a policy of attracting FDI into resource intensive sectors that brought capital, technology and management practices into the economy. However, foreign firms were required to enter into joint ventures with Kazakh firms, fully owned by the state.

From our data, the largest firms, and therefore contributing to granularity, are from sectors that are dominated by joint ownership firms (NACE 06, and NACE 24) (see Table 8). This is largely a result of the policy of the Kazakh government, though it also shows the domination of a few sectors in the economy, with little meaningful diversification of the economy since the country became independent.

This research has a number of important implications. First, to understand macroeconomic performance in an emerging economy, such as Kazakhstan, it seems important to understand how shocks affect the top firms in the key sectors in the country, such as Shell or TCO (Tengiz Chevron). Second, shocks to large firms and in this case in particular to the oil industry, could have major ripple down effects to the rest of the economy. Third, we focused on the granularity of total factor productivity, but it is likely that granularity matters also for other key macroeconomic indicators, such as inflation, output and exchange rates.

The results can also inform policy makers in emerging economies such as Kazakhstan. It shows the importance of government institutions to monitor closely these granular firms as they have a direct impact on the aggregate fluctuations. At the same time, diversifying the economy away from certain sectors becomes a policy objective from a risk management perspective. Similarly, improving competitive conditions and lowering entry barriers into these sectors also becomes a policy objective to reduce the impact of the large firms to TFP and GDP.

6. References

Abilova, A. & Subramanian, V. (2020). ‘Emerging Market Multinationals – the case of Kazakhstan. Building Strategic Capabilities in Emerging Markets , pp. 186 – 206 Cambridge University press

Ackerberg, D.A., Caves K. & Frazer G. (2015). ‘Identification properties of recent production function estimators’, Econometrica, 83(6), pp. 2411-2451.

Autor, D., Dorn D., Katz L.F., Patterson C. & Van Reenen J. (2020). ‘The fall of the labor share and the rise of superstar firms’, Quarterly Journal of Economics, 135, pp. 645-709.

Belenzon, S., Berkovitz, T. & Rios, L. (2012). ‘Capital markets and firm organization: How financial development shapes European corporate groups’, Management Science, 59(6), pp. 1326-1343.

Bernard, A., Dhyne, E., Magerman, G., Manova, K. & Moxnes, A. (2019). ‘The Origins of Firm Heterogeneity: A Production Network Approach’, National Bureau of Economic Research (NBER) Working Paper No. w25441.
https://www.nber.org/papers/w25441

Blanco-Arroyo, O. & Alfarano, S. (2017). ‘Granularity of the Business Cycle Fluctuations: The Spanish Case’, Economia Coyuntural, Revista de temas de perspectivas y coyuntura, 2(1), pp. 31-58.

Blanco-Arroyo, O., Ruiz-Buforn, A., Vidal-Tomás, D. & Alfarano, S. (2018). ‘On the determination of the granular size of the economy’, Economics Letters, 173, pp. 35-38.

Bleuel, P. (2018). ‘Germany – A Fertile Ground For The Development Of Hidden Champions’, Working Papers hal-03198681, HAL. https://hal.archives-ouvertes.fr/hal-03198681/document

Chen, Y.-F., Hsu W.-T. & Peng S.-K. (2022). ‘Innovation, Firm Size Distribution, and Gains from Trade’, Theoretical Economics, forthcoming

Clauset, A., Shalizi C.R. & Newman M.E.J. (2009). ‘Power-law Distributions in Empirical Data’, SIAM Review, 51(4), pp.661-703.

Czinkán, N. (2017). ‘The Role of Individual Firms in Aggregate Fluctuations: Evidence from Hungary’, Financial and Economic Review, 16(2), pp. 40-63.

Dacic, N. & Melolinna, M. (2019). 'The empirics of granular origins: some challenges and solutions with an application to the UK. Bank of England', Staff working paper, No 842. https://www.bankofengland.co.uk/working-paper/2019/the-empirics-of-granular-origins-some-challenges-and-solutions-with-an-application-to-the-uk

De Loecker, J. & Konings, J. (2006). 'Creative Destruction and Productivity Growth: Evidence from Slovenia', European Journal of Political Economy, 22, pp. 388-408.

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

Ebeke, M.C.H. & Eklou, K.M. (2017). 'The Granular Origins of Macroeconomic Fluctuations in Europe', International Monetary Fund (IMF) Working Paper WP/17/229. https://www.elibrary.imf.org/view/journals/001/2017/229/article-A001-en.xml

Eurostat (2021). Statistics on small and medium-sized enterprises. Available at: https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Statistics_on_small_and_medium-sized_enterprises (Accessed on 07 September 2021)

Feldman, M. (2006). 'Emerging varieties of Capitalism in Transition Countries: Industrial Relations and Wage Bargaining in Estonia and Slovenia', Comparative Political Studies, 39 (7), pp. 829-854.

Fornaro, P. & Luomaranta, H. (2018). 'Aggregate fluctuations and the effect of large corporations: Evidence from Finnish monthly data', Economic Modelling, 70, pp. 245-258.

Friberg, R. & Sanctuary, M. (2016). 'The contribution of firm-level shocks to aggregate fluctuations: The case of Sweden', Economics Letters, 147, pp. 8-11.

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

Gabaix, X., & Ibragimov R. (2011). 'Rank - 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents', Journal of Business and Economic Statistics, 29, pp. 24-39.

Gnocato, N. & Rondinelli, C. (2018). 'Granular sources of the Italian business cycle. Bank of Italy Temi di Discussione', Working Paper No. 1190. https://www.bancaditalia.it/pubblicazioni/temi-discussione/2018/2018-1190/index.html?com.dotmarketing.htmlpage.language=1

International Monetary Fund (IMF) (2021). World Economic Outlook Database April 2021. Available at: https://www.imf.org/en/Publications/WEO/weo-database/2021/April (Accessed on 07 September 2021)

Konings, J., Van Cayseele, P. & Warzynski, F. (2005). 'The Effect of Privatization and Competitive Pressure on Firms' Market Power: Evidence from Emerging Markets', Review of Economics and Statistics, 87 (1), pp. 124-134.

Lee, M. (2015). 'Can Idiosyncratic Shocks to Firms Explain Macroeconomic Growth and Fluctuations in Korea', Journal of Economic Theory and Econometrics, 26(3), pp. 63-78.

Levinsohn, J. & Petrin, A. (2003). 'Estimating Production Functions Using Inputs to Control for Unobservables', The Review of Economic Studies, 70(2), pp. 317-341.

Miranda-Pinto, J. & Shen, Y. (2019). 'A Granular View of the Australian Business Cycle', Economic Record, 95(311), pp. 407-424.

Nölke, A., ten Brink, T., May, C. & Claar S. (2020). 'State-permeated Capitalism in Large Emerging Economies', Routledge: London

Olley, G. S. & Pakes, A. (1996). 'The Dynamics of Productivity in the Telecommunications Equipment Industry', Econometrica, 64(6), pp. 1263-1297.

Organization for Economic Co-operation and Development (OECD) (2021). Enterprises by business size (indicator). Available at: https://www.oecd-ilibrary.org/industry-and-services/enterprises-by-business-size/indicator/english_31d5eeaf-en (Accessed on 07 September 2021)

Papa, J. (2019). 'What is behind aggregate productivity growth in Ireland? A granular approach', Research paper, February 2019, Department of Business Enterprise and Innovation, Ireland. https://igees.gov.ie/wp-content/uploads/2019/03/Granularity-Productivity-Publication.pdf

Popova, S. (2019). 'Idiosyncratic shocks: estimation and the impact on aggregate fluctuations', Bank of Russia Working Paper Series wps46, Bank of Russia. https://www.cbr.ru/Content/Document/File/87574/wp_46_e.pdf

The World Bank (2021). GDP per capita, current US$, Kazakhstan. Available at: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=KZ (Accessed on 07 September 2021)

Wagner, J. & Weche, J.P. (2020). 'On the granularity of the German economy–first evidence from the top 100 companies panel database', Applied Economics Letters, 27(21), pp. 1768-1771.

Witt, M.A. & Redding, G. (2013). 'Asian business systems: Institutional comparison, clusters and implications for varieties of capitalism and business systems theory', Socio-Economic Review, 11(2), pp. 265-300.

Witt, M.A. & Redding, G. (2014). 'The Oxford Handbook of Asian Business Systems.' Oxford: Oxford University Press.

7. Tables

Table 1: Summary Statistics, 2012-2018

Variables	Mean of all firms	Mean of top 10 firms	Mean of top 50 firms	Mean of top 100 firms	Mean of top 1000 firms
Number of workers	284	6035	2789	1897	500
Sales (Mln. Tenge)	1.133	108.820	32.863	18.831	2.562
Value added (Mln. Tenge)	0.673	74.594	21.115	11.882	1.567
Energy (Mln. Tenge)	0.026	2.307	0.765	0.456	0.059
Fixed assets (Mln. Tenge)	3.581	276.937	109.792	64.219	8.338

Table 2: Skewness and kurtosis based on log sales for each quarter of 2016

Time	Number of firms	Skewness	Kurtosis	P-value for Shapiro-Wilk test	P-value for Shapiro-Francia test
2016q1	2146	0.420	4.284	0.000	0.000
2016q2	2122	0.412	4.087	0.000	0.000
2016q3	2108	0.394	4.030	0.000	0.000
2016q4	2244	0.198	4.617	0.000	0.000

Table 3: : β - coefficients obtained in TFP estimation (intermediate input based on energy input)

Variables	Agriculture & mining	Manufacturing	Construction, waste & water	Wholesale, retail & logistics	Other services
Log labor	0.594***	0.667***	0.757***	0.603***	0.654***
	(0.0899)	(1.05e-05)	(1.56e-05)	(4.90e-06)	(1.53e-05)
Log fixed assets	0.535***	0.363***	0.282***	0.329***	0.265***
	(0.107)	(1.14e-05)	(1.70e-05)	(1.05e-05)	(1.67e-05)
Observations	11,707	17,001	14,212	16,810	13,815

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Regression results for sample sizes between 1 and 50

Variables	K1	K5	K10	K20	K30	K40	K50
Γ_t	1.0383***	0.8502***	0.8139***	0.6597***	0.6564***	0.6130***	0.6304***
	(0.325)	(0.178)	(0.153)	(0.148)	(0.134)	(0.128)	(0.127)
Γ_{t-1}	-0.2586	0.0626	-0.0294	0.2242	0.2228	0.2854*	0.2770*
	(0.412)	(0.221)	(0.199)	(0.184)	(0.165)	(0.155)	(0.155)
Γ_{t-2}	0.2558	0.1114	0.2342	0.0900	0.0538	0.0147	0.0255
	(0.347)	(0.191)	(0.161)	(0.158)	(0.144)	(0.136)	(0.135)
Constant	0.0070	0.0140	0.0162*	0.0114	0.0117	0.0144	0.0163*
	(0.015)	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)
Observations	22	22	22	22	22	22	22
R-squared	0.445	0.714	0.780	0.752	0.780	0.788	0.797
Ad. R-squared	0.353	0.666	0.744	0.711	0.744	0.752	0.763

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: β - coefficients obtained in TFP estimation (intermediate input based on all costs)

Variables	Agriculture & mining	Manufacturing	Construction, waste & water	Wholesale, retail & logistics	Other services
Log labor	0.591***	0.658***	0.755***	0.580***	0.667***
	(0.0481)	(0.0226)	(0.143)	(4.90e-06)	(0.0312)
Log fixed assets	0.532***	0.354***	0.280***	0.306***	0.277***
	(0.0457)	(0.0268)	(0.0865)	(6.86e-06)	(0.0370)
Observations	11,707	17,001	14,212	16,810	13,815

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Percentage share of the top 30 highest ranked firms, by type

Type	share
without state or foreign participation	41%
with state participation	3%
joint ownership with foreign participation	30%
foreign firms	26%
other type	0.1%

Table 7: Percentage share of the top 30 highest ranked firms, by sector and NACE code

Sector	NACE	share

agriculture and mining	1	0.1%
agriculture and mining	6	22.9%
agriculture and mining	7	7.8%
agriculture and mining	9	0.7%
manufacturing	12	3.3%
manufacturing	19	0.7%
manufacturing	24	19.0%
manufacturing	29	0.3%
manufacturing	30	0.4%
construction, water and waste	41	3.5%
construction, water and waste	42	6.7%
construction, water and waste	43	1.1%
wholesale, retail and logistics	46	0.6%
wholesale, retail and logistics	47	0.3%
wholesale, retail and logistics	49	17.4%
wholesale, retail and logistics	51	3.2%
wholesale, retail and logistics	52	1.7%
other services	61	9.0%
other services	63	0.1%
other services	70	1.0%
other services	71	0.3%

Table 8: Percentage share of the top 30 highest ranked firms, by sector, NACE code and type

sector	NACE	type	share
agriculture and mining	1	other type	0.1%
agriculture and mining	6	foreign firms	3.9%
agriculture and mining	6	joint ownership	12.5%
agriculture and mining	6	without state or foreign participation	6.5%
agriculture and mining	7	foreign firms	2.9%
agriculture and mining	7	joint ownership	1.1%
agriculture and mining	7	without state or foreign participation	3.8%
agriculture and mining	9	foreign firms	0.3%
agriculture and mining	9	joint ownership	0.4%
manufacturing	12	foreign firms	3.3%
manufacturing	19	without state or foreign participation	0.7%
manufacturing	24	foreign firms	6.9%
manufacturing	24	joint ownership	8.5%
manufacturing	24	without state or foreign participation	3.6%
manufacturing	29	without state or foreign participation	0.3%
manufacturing	30	joint ownership	0.3%
manufacturing	30	without state or foreign participation	0.1%
construction, water and waste	41	foreign firms	2.2%
construction, water and waste	41	without state or foreign participation	1.3%
construction, water and waste	42	foreign firms	0.4%
construction, water and waste	42	joint ownership	0.1%
construction, water and waste	42	with state participation	1.5%
construction, water and waste	42	without state or foreign participation	4.6%

construction, water and waste	43	joint ownership	1.0%
construction, water and waste	43	without state or foreign participation	0.1%
wholesale, retail and logistics	46	foreign firms	0.1%
wholesale, retail and logistics	46	without state or foreign participation	0.4%
wholesale, retail and logistics	47	joint ownership	0.3%
wholesale, retail and logistics	49	joint ownership	0.7%
wholesale, retail and logistics	49	without state or foreign participation	16.7%
wholesale, retail and logistics	51	joint ownership	2.9%
wholesale, retail and logistics	51	without state or foreign participation	0.3%
wholesale, retail and logistics	52	without state or foreign participation	1.7%
other services	61	foreign firms	5.6%
other services	61	joint ownership	2.4%
other services	61	with state participation	1.0%
other services	61	without state or foreign participation	0.1%
other services	63	without state or foreign participation	0.1%
other services	70	with state participation	0.1%
other services	70	without state or foreign participation	0.8%
other services	71	foreign firms	0.3%

8. Appendices

Appendix 1: NACE 2-digit description of sectors

Table 9: Description of sector and NACE 2-digit code

Code	NACE 2-digit description	Sector
01	Crop and animal production, hunting and related service activities	Agriculture and mining
02	Forestry and logging	Agriculture and mining
03	Fishing and aquaculture	Agriculture and mining
05	Mining of coal and lignite	Agriculture and mining
06	Extraction of crude petroleum and natural gas	Agriculture and mining
07	Mining of metal ores	Agriculture and mining
08	Other mining and quarrying	Agriculture and mining
09	Mining support service activities	Agriculture and mining
10	Manufacture of food products	Manufacturing
11	Manufacture of beverages	Manufacturing
12	Manufacture of tobacco products	Manufacturing
13	Manufacture of textiles	Manufacturing
14	Manufacture of wearing apparel	Manufacturing
15	Manufacture of leather and related products	Manufacturing
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	Manufacturing
17	Manufacture of paper and paper products	Manufacturing
18	Printing and reproduction of recorded media	Manufacturing
19	Manufacture of coke and refined petroleum products	Manufacturing
20	Manufacture of chemicals and chemical products	Manufacturing
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	Manufacturing
22	Manufacture of rubber and plastic products	Manufacturing
23	Manufacture of other non-metallic mineral products	Manufacturing
24	Manufacture of basic metals	Manufacturing
25	Manufacture of fabricated metal products, except machinery and equipment	Manufacturing
26	Manufacture of computer, electronic and optical products	Manufacturing
27	Manufacture of electrical equipment	Manufacturing
28	Manufacture of machinery and equipment n.e.c.	Manufacturing
29	Manufacture of motor vehicles, trailers and semi-trailers	Manufacturing
30	Manufacture of other transport equipment	Manufacturing
31	Manufacture of furniture	Manufacturing
32	Other manufacturing	Manufacturing
33	Repair and installation of machinery and equipment	Manufacturing
36	Water collection, treatment and supply	Construction, water & waste
37	Sewerage	Construction, water & waste
38	Waste collection, treatment and disposal activities; materials recovery	Construction, water & waste
39	Remediation activities and other waste management services	Construction, water & waste

41	Construction of buildings	Construction, water & waste
42	Civil engineering	Construction, water & waste
43	Specialized construction activities	Construction, water & waste
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	Wholesale, retail & logistics
46	Wholesale trade, except of motor vehicles and motorcycles	Wholesale, retail & logistics
47	Retail trade, except of motor vehicles and motorcycles	Wholesale, retail & logistics
49	Land transport and transport via pipelines	Wholesale, retail & logistics
50	Water transport	Wholesale, retail & logistics
51	Air transport	Wholesale, retail & logistics
52	Warehousing and support activities for transportation	Wholesale, retail & logistics
53	Postal and courier activities	Wholesale, retail & logistics
55	Accommodation	Other services
56	Food and beverage service activities	Other services
58	Publishing activities	Other services
59	Motion picture, video and television program production, sound recording and music publishing activities	Other services
60	Programming and broadcasting activities	Other services
61	Telecommunications	Other services
62	Computer programming, consultancy and related activities	Other services
63	Information service activities	Other services
68	Real estate activities	Other services
69	Legal and accounting activities	Other services
70	Activities of head offices; management consultancy activities	Other services
71	Architectural and engineering activities; technical testing and analysis	Other services
72	Scientific research and development	Other services
73	Advertising and market research	Other services
74	Other professional, scientific and technical activities	Other services
75	Veterinary activities	Other services
77	Rental and leasing activities	Other services
78	Employment activities	Other services
79	Travel agency, tour operator reservation service and related activities	Other services
80	Security and investigation activities	Other services
81	Services to buildings and landscape activities	Other services

Appendix 2: Lorenz curve of firm size distribution

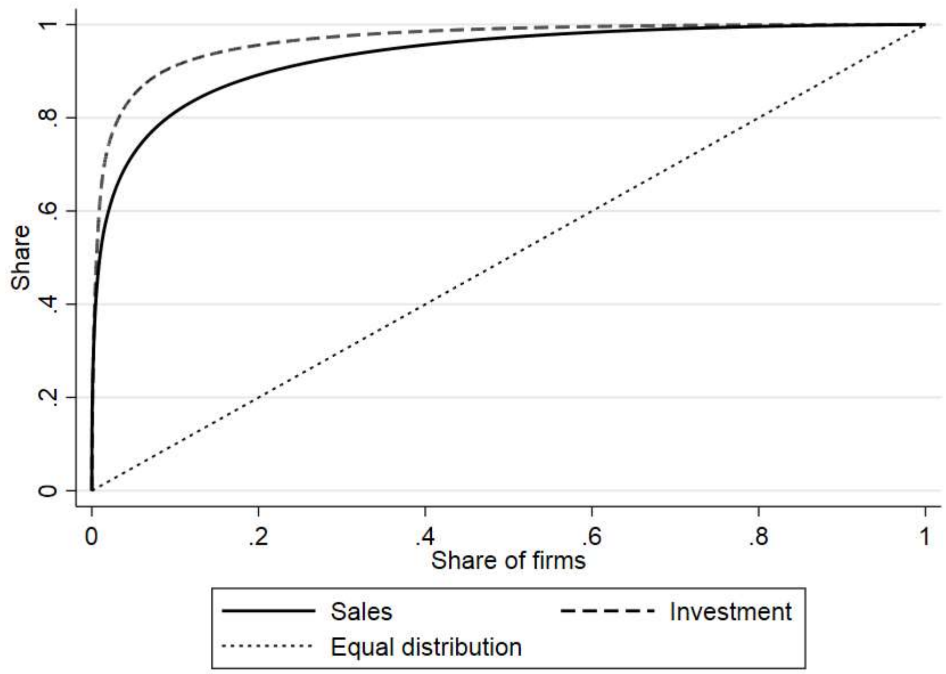


Figure 22: Lorenz curve of firm size distribution with regard to sales and investment (dotted line: equal distribution diagonal – solid line: distribution of sales, dashed line: distribution of non-negative investment. The distribution is based on all firms in the dataset, across all years and quarters)

Appendix 3: Skewness and kurtosis based on log sales for each quarter in the dataset

Table 10: Skewness and kurtosis based on log sales for each quarter

Time	Number of firms	Skewness	Kurtosis	P-value for Shapiro-Wilk test	P-value for Shapiro-Francia test
2012q1	3160	0.526	4.047	0.000	0.000
2012q2	3189	0.411	3.939	0.000	0.000
2012q3	3215	0.397	3.687	0.000	0.000
2012q4	3392	0.432	3.711	0.000	0.000
2013q1	2858	0.506	4.026	0.000	0.000
2013q2	2823	0.355	3.956	0.000	0.000
2013q3	2926	0.444	3.764	0.000	0.000
2013q4	3098	0.406	3.814	0.000	0.000
2014q1	3021	0.455	3.995	0.000	0.000
2014q2	3015	0.405	3.811	0.000	0.000
2014q3	3022	0.343	3.789	0.000	0.000
2014q4	3126	0.399	3.839	0.000	0.000
2015q1	3043	0.469	3.922	0.000	0.000
2015q2	3128	0.367	3.856	0.000	0.000
2015q3	3153	0.291	3.949	0.000	0.000
2015q4	3244	0.355	3.905	0.000	0.000
2016q1	2146	0.420	4.284	0.000	0.000
2016q2	2122	0.412	4.087	0.000	0.000
2016q3	2108	0.394	4.030	0.000	0.000
2016q4	2244	0.198	4.617	0.000	0.000
2017q1	1957	0.460	4.168	0.000	0.000
2017q2	1976	0.359	4.148	0.000	0.000
2017q3	2009	0.341	4.099	0.000	0.000
2017q4	2090	0.323	4.247	0.000	0.000
2018q1	1851	0.532	4.071	0.000	0.000
2018q2	1842	0.370	4.269	0.000	0.000
2018q3	1847	0.522	4.021	0.000	0.000
2018q4	1940	0.445	4.420	0.000	0.000

<u>*Appendix 4: Kernel density plot and quantile plot*</u>

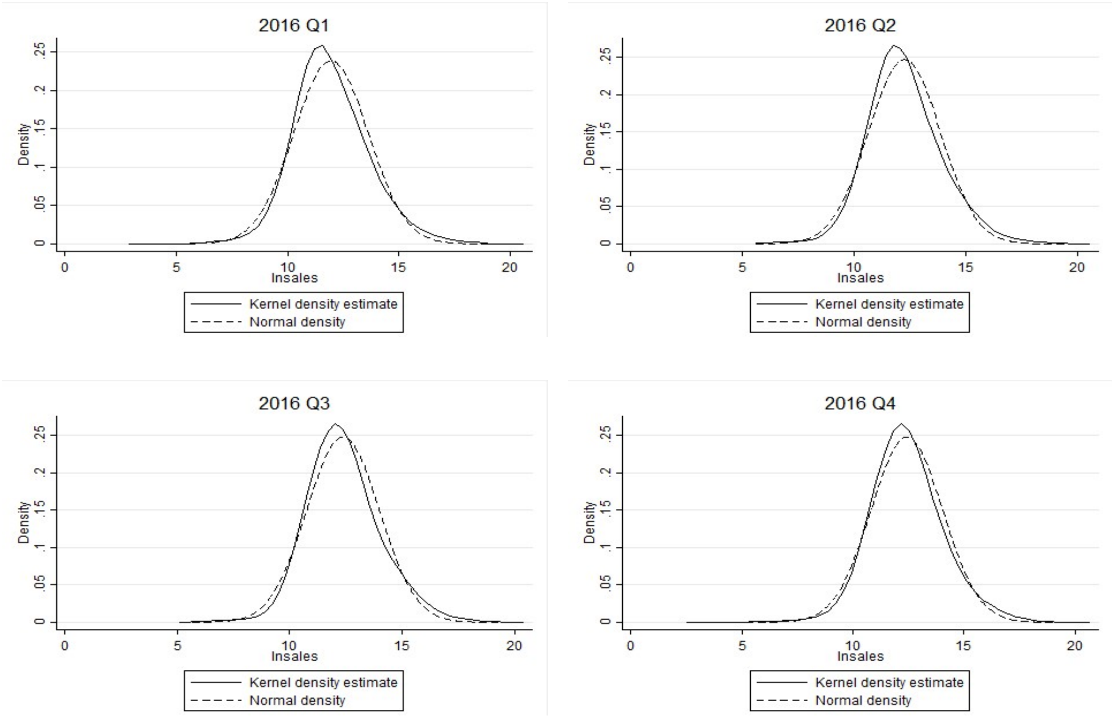


Figure 23: K-density plot of firm size distribution, for each quarter of 2016

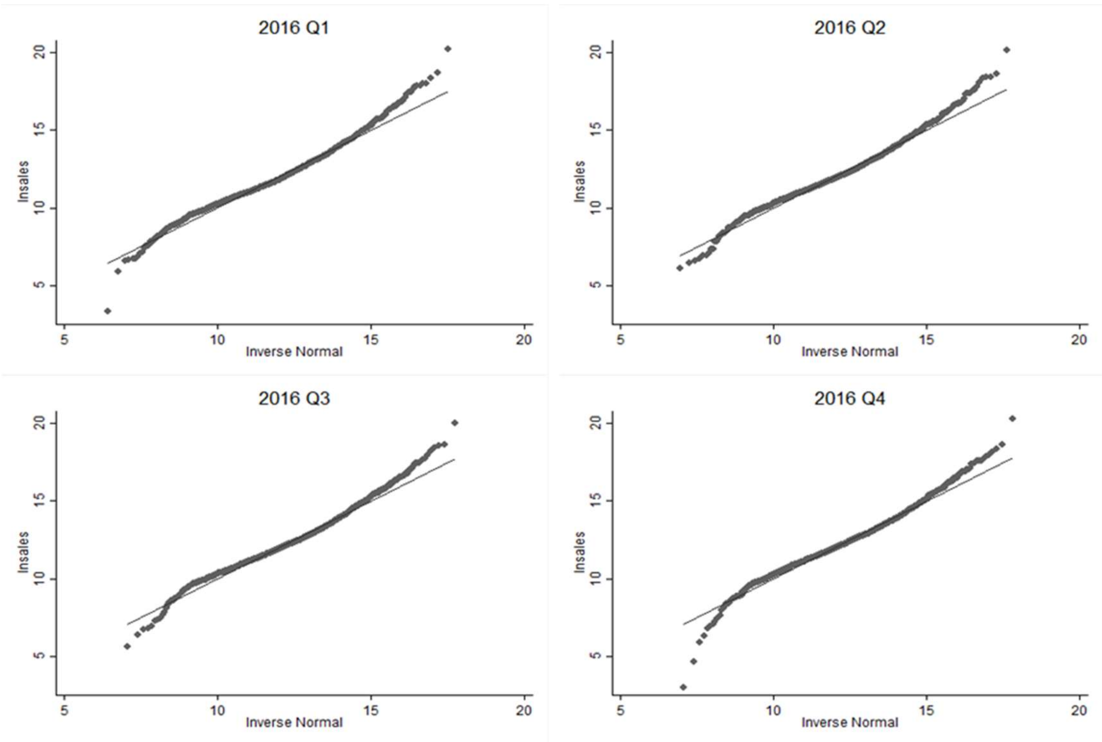
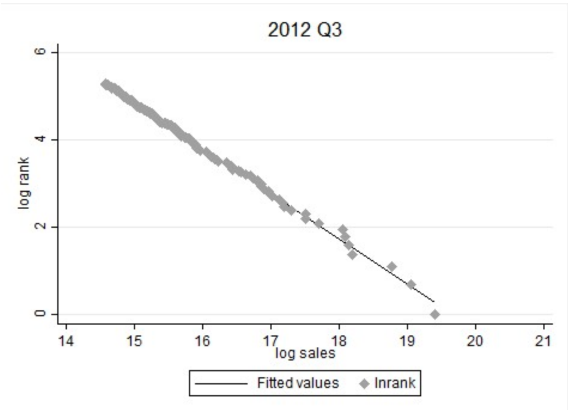


Figure 24: Quantile plot of firm size distribution, for each quarter of 2016

Appendix 5: Evolution of tail index and CCDF plot

Table 11: The lower bound and tail index by quarter

Time	Lower bound	Tail index
2012q1	33	1.87
2012q2	33	1.87
2012q3	126	2.06
2012q4	146	2.05
2013q1	49	1.88
2013q2	79	1.96
2013q3	90	1.95
2013q4	40	1.87
2014q1	44	1.9
2014q2	180	2.06
2014q3	29	1.84
2014q4	26	1.82
2015q1	33	1.87
2015q2	228	2.06
2015q3	28	1.84
2015q4	27	1.84
2016q1	101	1.91
2016q2	44	1.83
2016q3	31	1.77
2016q4	58	1.87
2017q1	45	1.83
2017q2	52	1.85
2017q3	120	1.95
2017q4	97	1.93
2018q1	63	1.87
2018q2	50	1.84
2018q3	61	1.84
2018q4	88	1.90



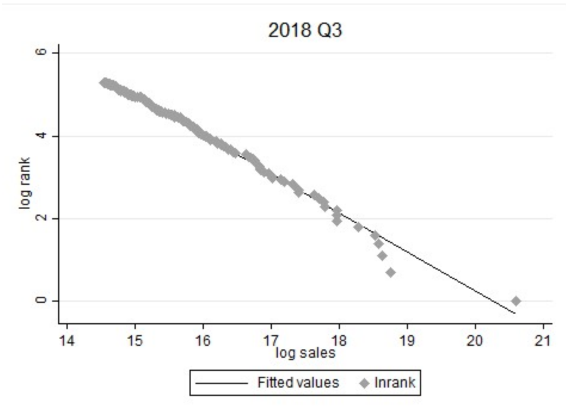


Figure 25: CCDF plot for the top 200 firms in the 3rd quarter of 2012 (LHS) versus the 3rd quarter of 2018 (RHS). In 2018 the largest firm became larger, making the slope flatter (the tail index is smaller).

Appendix 6: Regression results for different sample sizes

Table 12: Regression results for sample sizes between 1 and 60 – sample sizes 1 to 5

Variables	K1	K2	K3	K4	K5
Γ_t	1.0383***	0.8081**	1.0765***	0.8558***	0.8502***
	(0.325)	(0.294)	(0.283)	(0.192)	(0.178)
Γ_{t-1}	-0.2586	0.0133	-0.2266	0.0555	0.0626
	(0.412)	(0.377)	(0.362)	(0.231)	(0.221)
Γ_{t-2}	0.2558	0.0560	0.1676	0.0883	0.1114
	(0.347)	(0.311)	(0.293)	(0.204)	(0.191)
Constant	0.0070	0.0179	0.0169	0.0165	0.0140
	(0.015)	(0.014)	(0.013)	(0.011)	(0.010)
Observations	22	22	22	22	22
R-squared	0.445	0.443	0.565	0.649	0.714
Ad. R-squared	0.353	0.350	0.492	0.590	0.666

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Regression results for sample sizes between 1 and 60 – sample sizes 6 to 10

Variables	K6	K7	K8	K9	K10
Γ_t	0.8015***	0.7800***	0.8031***	0.8135***	0.8139***
	(0.171)	(0.165)	(0.153)	(0.148)	(0.153)
Γ_{t-1}	0.0826	0.0666	0.0422	-0.0400	-0.0294
	(0.222)	(0.216)	(0.199)	(0.194)	(0.199)
Γ_{t-2}	0.1446	0.1342	0.1839	0.2525	0.2342
	(0.188)	(0.178)	(0.166)	(0.159)	(0.161)
Constant	0.0142	0.0135	0.0124	0.0138	0.0162*
	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
Observations	22	22	22	22	22
R-squared	0.737	0.747	0.779	0.791	0.780
Ad. R-squared	0.693	0.705	0.742	0.756	0.744

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Regression results for sample sizes between 1 and 60 – sample sizes 15 to 35

Variables	K15	K20	K25	K30	K35
Γ_t	0.7282***	0.6597***	0.6566***	0.6564***	0.6279***
	(0.154)	(0.148)	(0.140)	(0.134)	(0.131)
Γ_{t-1}	0.1857	0.2242	0.2310	0.2228	0.2488
	(0.195)	(0.184)	(0.174)	(0.165)	(0.160)
Γ_{t-2}	0.0946	0.0900	0.0424	0.0538	0.0497
	(0.165)	(0.158)	(0.150)	(0.144)	(0.140)
Constant	0.0136	0.0114	0.0134	0.0117	0.0137
	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)
Observations	22	22	22	22	22
R-squared	0.763	0.752	0.768	0.780	0.779
Ad. R-squared	0.723	0.711	0.729	0.744	0.742

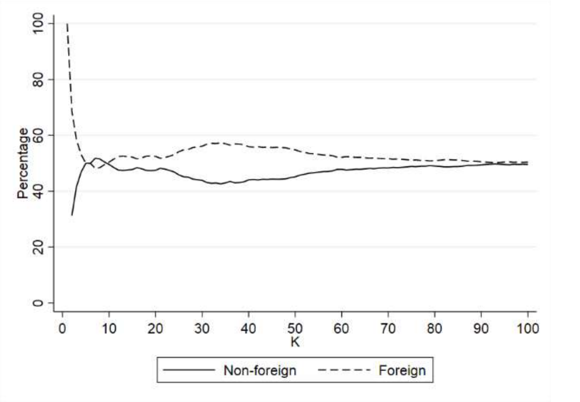
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: Regression results for sample sizes between 1 and 60 – sample sizes 40 to 60

Variables	K40	K45	K50	K55	K60
Γ_t	0.6130***	0.6243***	0.6304***	0.6228***	0.6264***
	(0.128)	(0.126)	(0.127)	(0.130)	(0.127)
Γ_{t-1}	0.2854*	0.2845*	0.2770*	0.2771*	0.2724*
	(0.155)	(0.153)	(0.155)	(0.159)	(0.154)
Γ_{t-2}	0.0147	0.0129	0.0255	0.0211	0.0199
	(0.136)	(0.134)	(0.135)	(0.137)	(0.134)
Constant	0.0144	0.0157*	0.0163*	0.0156*	0.0159*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Observations	22	22	22	22	22
R-squared	0.788	0.793	0.797	0.791	0.797
Ad. R-squared	0.752	0.758	0.763	0.757	0.763

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 7: Characterizing the top 1000 firms by type and sector



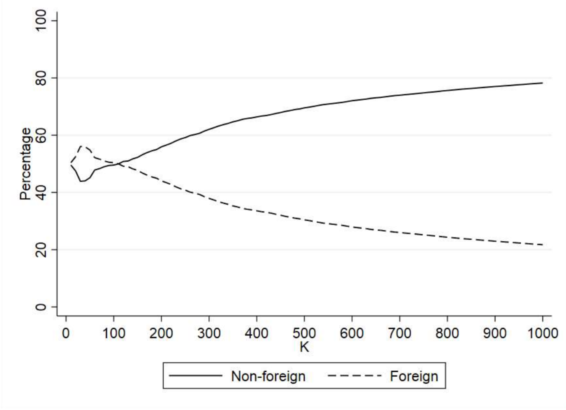
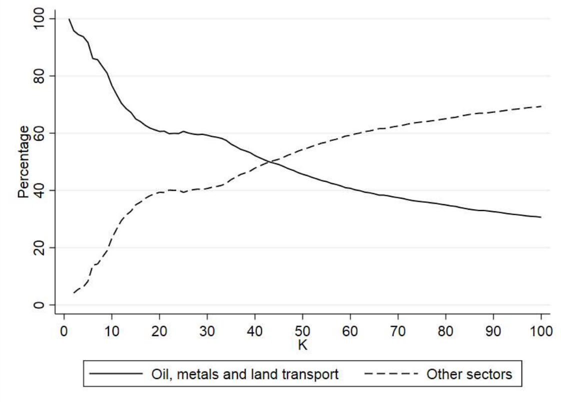


Figure 26: Share of foreign (foreign owned as well as joint ventures with foreign participation) compared to non-foreign firms (all other types of firms), when increasing the sample size from 1 to 100 in steps of 1 (LHS) or 10 to 1000 in steps of 10 (RHS). Foreign firms are the dominant type only in the top 100 largest firms.



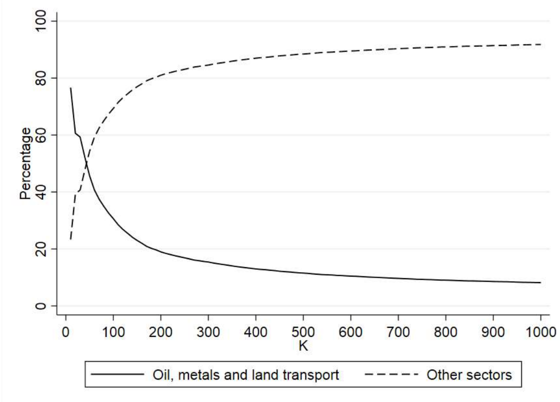


Figure 27: Share of firms active in NACE 06 (oil industry), NACE 24 (metal industry) and NACE 49 (land transport), when increasing the sample size from 1 to 100 in steps of 1 (LHS) or 10 to 1000 in steps of 10 (RHS). These 3 sectors are dominant only in the top 45 largest firms.