AGGLOMERATION, CONCENTRATION, AND MISALLOCATION

AHMET YUSUF AYDIN

BOĞAZİÇİ UNIVERSITY

2018

AGGLOMERATION, CONCENTRATION, AND MISALLOCATION

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Economics

by

Ahmet Yusuf Aydın

Boğaziçi University

2018

DECLARATION OF ORIGINALITY

I, Ahmet Yusuf AYDIN, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature...

Date ....20.07.2018....................................

ABSTRACT

Agglomeration, Concentration, and Misallocation

Misallocation may cause extensive productivity and output loss in sectors and countries. I use firm-level data for Spain to investigate the misallocation in Spain with a focus on the effect of agglomeration on allocative efficiency. I discover a negative relationship between agglomeration and allocative efficiency despite the higher productivity in denser areas. I explain the negative effect of agglomeration on allocative efficiency with the higher price elasticity of substitution due to higher competition in agglomerated regions. High elasticity in large cities amplifies the effect of existing idiosyncratic distortions on allocative efficiency. I test this argument using a monte carlo simulation of output loss due to misallocation. Finally, I document the low concentration and high competition in agglomerated regions with Herfindahl-Hirschman Index.

ÖZET

Toplaşma, Yoğunlaşma ve Kaynak Dağılımında Bozukluk

Kaynak dağılımında bozukluk, sektörlerde ve ülke genelinde büyük üretkenlik ve üretim kayıplarına sebep olabilir. Toplaşmanın kaynak dağılımı etkinliği üzerindeki etkisine odaklanarak İspanya'daki kaynak dağılımındaki bozukluğu incelemek adına firma seviyesinde İspanya verisi kullandım. Yoğun bölgelerdeki nispeten yüksek üretkenliğe rağmen toplaşma ve kaynak dağılımı arasında olumsuz bir ilişki tespit ettim. Toplaşmanın kaynak dağılımı etkinliği üzerindeki olumsuz etkisini, toplaşmış bölgelerdeki daha yüksek rekabet dolayısıyla ikamenin daha yüksek fiyat elastikiyeti göstermesi ile açıkladım. Büyük şehirlerdeki yüksek elastikiyet, var olan firmaya özgü bozulmaların kaynak dağılımı etkinliği üzerindeki etkisini büyütür. Bu argümanı kaynak dağılımındaki bozukluktan kaynaklanan üretim kaybının monte carlo simulasyonunu kullanarak test ettim. Son olarak, Herfindahl-Hirschman Endeksi ile toplaşmış bölgelerde düşük yoğunlaşma ve yüksek rekabet gerçekleştiğini gösterdim.

ACKNOWLEDGMENTS

This study is a product of a long and instructive process of hardworking and academic progress. I would like to thank my thesis advisor, Assist. Prof. Malik Çürük, for his academic supervision to help me to improve this study and his valuable guidance to keep me on track. He has been always motivating me faithfully to progress and be passionate on my studies.

I also would like to thank other members of thesis committee, Assoc. Prof. Tolga Umut Kuzubaş and Assoc. Prof. Türkmen Göksel, for their helpful comments and contributions on my thesis. I would like to express my sincere gratitude to Assoc. Prof. Murat Koyuncu who guided and helped me throughout my graduate studies.

I am grateful to TÜBİTAK BİDEB which provided me financial support for my graduate studies.

Besides all these, I would like to express my deepest gratitude to my family for their immeasurable support. I am essentially indebted to my parents for all opportunities they provided me to come this point of my life. My wife, Ümran Gülbin, is the one who makes my life better with her love and self-sacrificing efforts.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Why do some countries with similar economic structures and growth potentials experience different growth paths? Why does a particular country follow different growth trends in different time intervals? What are the preventer factors of economies that keep the economic performance below the potential of a country? These are the key questions that draw attention of many researchers of economics studies. One crucial answer to those questions stems from a stylized fact of economies that is resource misallocation. Inefficient allocation of production factors between sectors and firms cause productivity and total output to stay under its potential.

Misallocation is a widely studied phenomenon in economics literature during the last few decades. Misallocation of inputs as a crucial handicap on economic performance has been verified in many studies. Allocative efficiency of production factors across sectors or plants with different productivity levels is one of the important determinants of productivity both at firm level and aggregate level. Therefore, beyond the resource and technological qualifications of a country, allocative efficiency of factors would carry one country to its optimal growth level, while misallocation of factors would restrict growth rates under its production possibilities frontier (Hsieh and Klenow, 2009). Therefore, the evolution of productivity for an economy is determined not only by technological change or capital accumulation but also by labor and capital movements between firms and plants. In this context, this study investigates the extent of misallocation within manufacturing sectors and contribution of allocative efficiency to overall productivity and production performance.

Furthermore, I investigate the relationship between agglomeration and misallocation. In this regard, I approach the misallocation of resources at regional level by analyzing and comparing misallocation levels at regions in a country. The general perception in economics is that agglomeration is improving the economic performance in denser areas both at firm level and industry level. Therefore, I specifically examine the effect of agglomeration on allocative efficiency within industries and aggregate productivity.

First, I implement Hsieh and Klenow (2009) methodology to calculate the misallocation levels with a detailed firm level data for Spain and France obtained from AMADEUS database for the range of 2006 and 2015. They measure the misallocation level across firms with the dispersion of revenue-based total factor productivity of firms within each industry. I also explain the negative effect of misallocation on the industry-level productivity by following this model. When we compare the overall misallocation in Spain and France, we observe that misallocation is a more serious problem for Spian compared to France. Therefore, I choose to work with Spanish data for further analysis in this study.

Secondly, I confirm the main argument of agglomeration literature by presenting positive relationship between productivity and agglomeration in Spain. However, the results of misallocation analysis in regional level are interesting in this regard. Using Spanish firm-level data, I document that firm level allocative efficiency is worse in denser areas in spite of higher total productivity. I consider the intense competition responsible for the inefficient allocation of resources in agglomerated regions through increasing the price elasticity of substitution between products of firms in the same industries.

In order to demonstrate this relationship numerically, I run a simulation of output performance compared to the case of efficient allocation with firm productivity and idiosyncratic distortion distributions for a set of different elasticity values. The simulation results suggest that the output loss due to misallocation is increasing with elasticity. Then, I also show that the competition is higher in agglomerated regions by using Herfindahl-Hirschman Index which is a widely used concentration index. I expect larger elasticity values in agglomerated regions since competition is one of the determinants of price elasticity between firms. Therefore, I conclude that the reason behind high misallocation in agglomerated regions may be the intense competition and high elasticity of substitution in those regions.

The rest of this paper is organized as follows: Chapter 2 covers the literature and points out the contributions of this study. Chapter 3 describes the data and its quality. Chapter 4 introduces the theoretical framework and basic results for misallocation and its effect on productivity. I present productivity and misallocation at regional level in Spian at Chapter 5, and discuss the effect of elasticity on misallocation via the simulation results and concentration analysis. Chapter 6 concludes the paper.

CHAPTER 2

LITERATURE REVIEW AND CONTRIBUTIONS

The importance of allocative efficiency on the economic performance of industries and countries is a widely studied and emphasized phenomena in the last century. Restuccia and Rogerson (2013), Hopenhayn (2014), Restuccia and Rogerson (2017), Bartelsman and Wolf (2017) provide the most recent survey articles and cover the prominent studies on misallocation literature.

Measuring the level of misallocation among firms within industries has created a new strand of the literature in the 21st century after the widespread availability of detailed firm-level data in the last decades (Syverson, 2014). There are two major methods to measure the misallcation across firms in the literature. The first thriving methodology is introduced by Hsieh and Klenow (2009) (HK henceforth) which is based on the idea that marginal revenue products are equalized when the production factors are allocated optimally. Gopinath et al. (2017) used this method to measure capital and labor misallocation separately in manufacturing sectors and discuss the increasing capital misallocation in South Europe after the transition to Eurozone. Dias, Marques, and Richmond (2016) implement this model to Portuguese firm data for all sectors to examine the evolution of misallocation and productivity in the pre-crisis period. Ryzhenkov (2016) apply the model for Ukrainian manufacturing sector and compares the productivity loss in EU and US. The second prominent misallocation measure is firstly introduced by Olley and Pakes (1996) (OP henceforth) as the covariance between firm size and productivity. However, Bartelsman, Haltiwanger, and Scarpetta (2013) used this measure in a misallocation article first time. The main distinction of this method from the HK method is that OP do not assume euqalized marginal products in the efficient allocation which is the perfect correlation between the firm size and productivity. I prefer to use the HK method in this study. The contribution of this study to the misallocation literature is that I include regional dimension to HK method and compare misallocation in different regions of a country.

There are a couple of studies in the misallocation literature that incorporate agglomeration literature. Bin et al. (2018) utilize OP method to measure misallocation in regions of China and investigate the region-based determinants of allocative efficiency. Fontagne and Santoni (2018) investigate the labor allocation using Petrin and Sivadasan (2013) methodology which measures firm-specific misallocation with the gap between marginal product and marginal cost of labor. They propose the increasing labor productivity in denser areas is an indicator of better labor allocation across firms through better matching of employers and employees. However, their methodology does not sort out the allocation of resources between firms directly; indeed, their gap index is a measure of firm-specific performance of labor. On the other hand, HK methodology is a direct measure of misallocation within industries. Therefore, this study contributes to the literature of misallocation and agglomeration by introducing intra-industry misallocation and examining the effect of agglomeration on it. Furthermore, this study discovers the negative relationship between agglomeration and allocative efficiency.

In this regard, this study contributes to the agglomeration literature as well by introducing the result above. Combes et al. (2012), Duranton and Puga (2004), and many other studies in the agglomeration literature ascertain the high productivity performances of plants and workers in denser areas with the productivity enhancing effects of agglomeration. I also verify this situation in Spanish data; however, I document the counter effect of high misallocation in agglomerated regions on productivity and output.

CHAPTER 3

DATA

I use AMADEUS database provided by Bureau van Dijk which includes firm level financial data for European countries. The data involve balance sheet, profit & loss items, and general information for a high coverage of firms for 44 countries in Europe. The variables of gross output, material costs, employment, wage bill, fixed assets, and industry code allow us to compute total factor productivity for each firm and aggregate them to sector and industry level. Another advantage of this dataset is the location variables of each firm in the data. In this study, I use specified region variable[1] to be able to differentiate firms into regional level and analyze misallocation levels in different regions of a country.

I downloaded the data from Wharton Research Data Services (WRDS) for the period between 2005 and 2016. It is allowed to download the last 10 available years for each firm in this website. Moreover, there is a reporting lag of around 2 years for the firms. Therefore, the coverage is higher for this period when I download the data.

There are a lot of missing or misreported values in the data; hence, it is needed to clean the data to be able to make reliable analysis. I implement the relevant data management processes used by Kalemli-Ozcan et al. (2015) and Gopinath et al. (2017). I also drop the observations in 2016 and 2005 since those years include very small number of observations compared to other years. Therefore, the date range of this study covers the years between 2006 and 2015. I explained the data cleaning process in detail in Appendix A.

[1] Region variable corresponds to NUTS-3 regional division for most of the countries in the data.

The data include value added variable as a sum of *"Profit for period + Depreciation + Taxation + Interest paid + Cost of employees"*; however, the missing rate of this variable is high in the downloaded data. Therefore, I use an alternative measure[2] of value added which is obtained by the difference between *operating revenue* and *material costs* (Gopinath et al, 2017).

I restrict the scope of this study with manufacturing industries for two reasons. First, the production function estimates are more relevant for manufacturing firms; hence, misallocation measures in this study are more relevant for them. Secondly, the coverage rates are higher for manufacturing industries in this dataset.

3.1 Data coverage

After a detailed and rigorous data cleaning process, I end up with a highly representative dataset to make a reliable analysis. Furthermore, high coverage rates of this dataset is a superiority of AMADEUS over other firm level datasets. Table 1 and Table 2 documents the coverage rates of cleaned manufacturing data in employment, wage bill, and gross output relative to the aggregate values in Eurostat SBS database for Spain and France. "DATA" columns shows the coverage rates of data used in this study. Other columns are the coverage rates of the data used in Gopinath et al. (2017). Those rates are available until 2012 since their study covers the years until 2012. The coverage rates of our data are relatively low compared to their coverage rates for the first half of our sample since the only data source for this study is a single download from WRDS; however, they construct their data by merging different

[2]The correlation between newly constructed value added variable and the one in dataset is 0.92.

ORBIS[3] and AMADEUS vintages with different dates[4]. This merging procedure increases their coverage rates. Nevertheless, the coverage rates of our data reaches around the same rates with them in the second half of our sample period.

Table 1. Coverage of Spanish Data Relative to Eurostat

	Employment		Wage Bill		Gross Output	
year	Kalemli*	DATA***	Gopinath**	DATA***	Kalemli*	DATA***
2006	0.68	0.46	0.74	0.49	0.77	0.49
2007	0.67	0.50	0.74	0.53	0.77	0.56
2008	0.67	0.52	0.72	0.55	0.72	0.55
2009	0.71	0.56	0.72	0.58	0.75	0.61
2010	0.70	0.58	0.73	0.60	0.74	0.64
2011	0.72	0.62	0.74	0.64	0.75	0.65
2012	0.67	0.65	0.71	0.67	0.72	0.68
2013		0.68		0.70		0.71
2014		0.71		0.73		0.72
2015		0.68		0.70		0.65

*Coverage rates are from Table 6.6 in Kalemli-Ozcan et al. (2015).

**Coverage rates are from Table A.2 in Gopinath et al. (2017).

***Coverage rates of the data used in this study.

Table 2. Coverage of French Data Relative to Eurostat

	Employment		Wage Bill		Gross Output	
year	Kalemli*	DATA***	Gopinath**	DATA***	Kalemli*	DATA***
2006	0.60	0.42	0.72	0.53	0.80	0.57
2007	0.64	0.46	0.73	0.60	0.81	0.65
2008	0.69	0.44			0.85	0.66
2009	0.63	0.37	0.71	0.63	0.86	0.64
2010	0.67	0.43	0.73	0.66	0.83	0.70
2011	0.62	0.40	0.75	0.69	0.84	0.73
2012	0.64	0.34	0.73	0.72	0.82	0.75
2013		0.44		0.74		0.80
2014		0.52		0.72		0.80
2015		0.51		0.69		0.81

*Coverage rates are from Table 6.6 in Kalemli-Ozcan et al. (2015).

**Coverage rates are from Table A.2 in Gopinath et al. (2017).

***Coverage rates of the data used in this study.

Note: Wage bill values for France in 2008 is missing in Eurostat SBS database.

[3]ORBIS is the global version of AMADEUS which provides firm-level data covering around 100+ countries since 2005.

[4]They use ORBIS disk 2005, ORBIS disk 2009, ORBIS disk 2013, AMADEUS online 2010 from WRDS and AMADEUS disk 2014.

3.2 Data representativeness

Another advantage of AMADEUS database is that it is highly representative by including large number of firms from all size classes. It includes data for small firms and from each consecutive year which are the main handicap of other firm-level databases covering only large listed firms like Compustat and Worldscope. Table 3 presents the share of total manufacturing economic activity by size classes in comparison with the data used in Gopinath et al. (2017) and Eurostat SBS data for 2006. Our data represent each size class with a close size distribution to Eurostat SBS data.

Table 3. Size Distribution of Manufacturing Industry in 2006

SPAIN				FRANCE			
	Employment				Employment		
Employees	DATA*	Gopinath**	Eurostat	Employees	DATA*	Gopinath**	Eurostat
1-19	0.24	0.24	0.31	1-19	0.09	0.10	0.19
20-249	0.48	0.50	0.43	20-249	0.33	0.35	0.34
250+	0.29	0.26	0.26	250+	0.58	0.56	0.47
missing	5%			missing	55%		
	Wage Bill				Wage Bill		
Employees	DATA*	Gopinath**	Eurostat	Employees	DATA*	Gopinath**	Eurostat
1-19	0.18	0.19	0.20	1-19	0.07	0.08	0.14
20-249	0.45	0.47	0.43	20-249	0.28	0.30	0.31
250+	0.37	0.34	0.37	250+	0.65	0.61	0.55
missing	2%			missing	32%		
	Gross Output				Gross Output		
Employees	DATA*	Gopinath**	Eurostat	Employees	DATA*	Gopinath**	Eurostat
1-19	0.12	0.14	0.14	1-19	0.04	0.05	0.09
20-249	0.40	0.42	0.38	20-249	0.23	0.23	0.27
250+	0.49	0.45	0.49	250+	0.72	0.72	0.64
missing	1%			missing	27%		

*Size distribution of the data used in this study.

**Size distribution of the data used in Gopinath et al. (2017).

Note: Share of firms with missing employment for each variable is indicated. The distribution rates are accounted for the firms with reported number of employees.

CHAPTER 4

MISALLOCATION AND PRODUCTIVITY LOSS

There are several kinds of misallocation measures in the literature to quantify the level of misallocation between firms within industries. The first prominent measure of misallocation is introduced by Hsieh and Klenow (2009) which exploits the idea of idiosyncratic distortions faced by firms in an industry. Those idiosyncratic distortions; such as tax, credit constraints, size dependant regulations etc.; would cause differentiated marginal products of inputs for different firms in the same industry. Therefore, this would create a TFP loss and hence output loss in that industry, and also in aggregate economic performance, compared to the optimal resource allocation outcome. Gopinath et al. (2017) and many others[5] implement this method to measure the level of misallocation across firms within industries.

Hsieh and Klenow (2009) use a model of monopolistic competition with heterogeneous firms to quantify the differentiating effect of idiosyncratic distortions on the marginal products of capital and labor; and the resulting TFP loss for the Indian and Chinese economy. Gopinath et al. (2017) use the same model for the Spanish firms in comparison with some other European firms to illustrate the significant capital misallocation and TFP loss in South Europe. In this study, I utilize their model to focus on the misallocation between individual firms within each 4-digit manufacturing sector for the case of Spain and France.

[5] e.g. Bento and Restuccia (2017), Dias et al. (2016), Restuccia (2018), Ryzhenkov (2016)

4.1 Theoretical framework

The model assumes a CES aggregation of firm outputs for total industry output where s is 4-digit NACE Revision 2 industry code, t is year, and N_{st} is the number of firms in industry s at time t.

$$Y_{st} = \left[\sum_{i=1}^{N_{st}} D_{ist} (y_{ist})^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{1}$$

Here, y_{ist} is real output and D_{ist} is the demand shifter for firm i's variety. I take the elasticity of substitution between varieties[6] $\varepsilon = 3$ as in Hsieh and Klenow (2009) and Gopinath et al. (2017). For the price of variety i p_{ist} and price of industry output P_{st}, isoelastic demand for each firm's output is obtained by:

$$y_{ist} = (p_{ist}/P_{st})^{-\varepsilon} (D_{ist})^{\varepsilon} Y_{st} \tag{2}$$

Each individual manufacturing firm produces with a Cobb-Douglas production function:

$$y_{ist} = A_{ist} k_{ist}^{\alpha} \ell_{ist}^{1-\alpha} \tag{3}$$

where k_{ist} is capital stock, ℓ_{ist} labor, A_{ist} is physical productivity, and α is capital share. I assume that $\alpha = 0.35$ and homogeneous across industries. This assumption will not affect the dispersion measures since they are calculated at within-industry level and aggregated with industry shares (Gopinath et al, 2017). I calculate real output by deflating nominal value added ($p_{ist} y_{ist}$) with output price deflators at 2-digit industry level[7] from Eurostat[8]. I use wage bill as a measure of labor instead of employment level in order to capture the quality of workforce[9]. I deflate labor with the

[6] I discuss the crucial effect of elasticity on the level of misallocation in section 5.2.

[7] 2-digit level industry classifications of NACE Revision 2 are provided in Appendix B.

[8] Eurostat provides sector specific total output price index for European firms.

[9] In addition to this fact, data coverage for the wage bill variable is higher compared to employ-

same 2-digit level deflators. I measure capital stock as the sum of tangible and intangible fixed assets, and deflate it with the price of investment goods obtained from WDI[10].

Each firm maximizes its profit by choosing its price level, capital, and labor:

$$\max_{p_{ist},k_{ist},\ell_{ist}} \Pi_{ist} = (1-\tau^{y}_{ist})p(y_{ist})y_{ist} - (1+\tau^{k}_{ist})R_{st}k_{ist} - w_{st}\ell_{ist} \tag{4}$$

where $R_{st} = r_t + \delta_{st}$, r_t denotes real interest rate, δ_{st} denotes the depreciation rate, w_{st} denotes wage rate. Each firm faces with idiosyncratic output distortions τ^{y}_{ist} and capital distortions τ^{k}_{ist} relative to labor. We can regard output distortions as the factors that affect output levels of firms apart from their inputs and technology such as transportation costs or size dependant restrictions. On the other hand, it is enough to implement τ^{k}_{ist} as the distortions that affect the marginal product of capital relative to labor to capture the input distortions such as credit constraints. Both τ^{y}_{ist} and τ^{k}_{ist} are idiosyncratic and this variety in distortions is the key factor that create the misallocation in this model. If each firm within an industry would face the same level of distortions, there would be no variation in the marginal revenue products within industries.

The first-order conditions of the profit maximization problem with respect to labor and capital yields the following marginal revenue products for each firm:

$$MRPL_{ist} := \left(\frac{\varepsilon-1}{\varepsilon}\right)(1-\alpha)\left(\frac{p_{ist}y_{ist}}{\ell_{ist}}\right) = \left(\frac{1}{1-\tau^{y}_{ist}}\right)w_{st} \tag{5}$$

$$MRPK_{ist} := \left(\frac{\varepsilon-1}{\varepsilon}\right)(\alpha)\left(\frac{p_{ist}y_{ist}}{k_{ist}}\right) = \left(\frac{1+\tau^{k}_{ist}}{1-\tau^{y}_{ist}}\right)R_{st} \tag{6}$$

ment as discussed in section 3.

[10] I obtain capital deflator by dividing current and constant values of country specific gross capital formation from Word Development Indicators provided by World Bank.

The capital-labor ratio and the output price of firms are also obtained from the profit maximization problem:

$$\frac{k_{ist}}{\ell_{ist}} = \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{w_{st}}{R_{st}}\right)\left(\frac{1}{1+\tau_{ist}^{k}}\right) \tag{7}$$

$$p_{ist} = \left(\frac{\varepsilon-1}{\varepsilon}\right)\left(\frac{R_{st}}{\alpha}\right)^{\alpha}\left(\frac{w_{st}}{1-\alpha}\right)^{1-\alpha}\left(\frac{(1+\tau_{ist}^{k})^{\alpha}}{A_{ist}(1-\tau_{ist}^{y})}\right) \tag{8}$$

The model with distortions suggest that firms would equate their marginal revenue products to the input costs times the distortion wedges. Therefore, we would observe a variation in marginal revenue products of firms, as a result of idiosyncratic output and capital distortions, which is the reason behind the misallocation within industries.

Hereafter, we are able to define revenue-based total factor productivity as price time physical productivity in the following way (Foster, Haltiwanger, & Syverson, 2008; Hsieh & Klneow, 2009; Bartelsman et al., 2013):

$$TFPR_{ist} := p_{ist}A_{ist} = \frac{p_{ist}y_{ist}}{k_{ist}^{\alpha}\ell_{ist}^{1-\alpha}} = \left(\frac{\varepsilon-1}{\varepsilon}\right)\left(\frac{MRPK_{ist}}{\alpha}\right)^{\alpha}\left(\frac{MRPL_{ist}}{1-\alpha}\right)^{1-\alpha} \tag{9}$$

We can infer from the equations above that, if distortions are equalized across firms within industries, we would expect to have same MRPL, MPRK, and TFPR across the firms in each industry, hence no dispersion. This means resources are allocated optimally within each industry, that is more labor and capital are allocated to highly productive and highly demanded firms. If firms have idiosyncratic distortions, dispersion of MRPK, MRPL, and TFPR would cause a lower sectoral TFP and output since the firms with relatively higher distortions would be smaller than their optimal size. In other words, resource movements from low productive firms to higher productive firms within each industry would increase the total productivity and output.

4.2 Misallocation and TFP loss in data

Following this methodology, I calculate dispersion measures for Spain and France at the industry level as the standard deviation of log(TFPR). Then, I aggregate them with country specific time-invariant industry weights and normalize the initial values to 1. The evolution of aggregated manufacturing TFPR dispersion for Spain and France are presented in Figure 1.

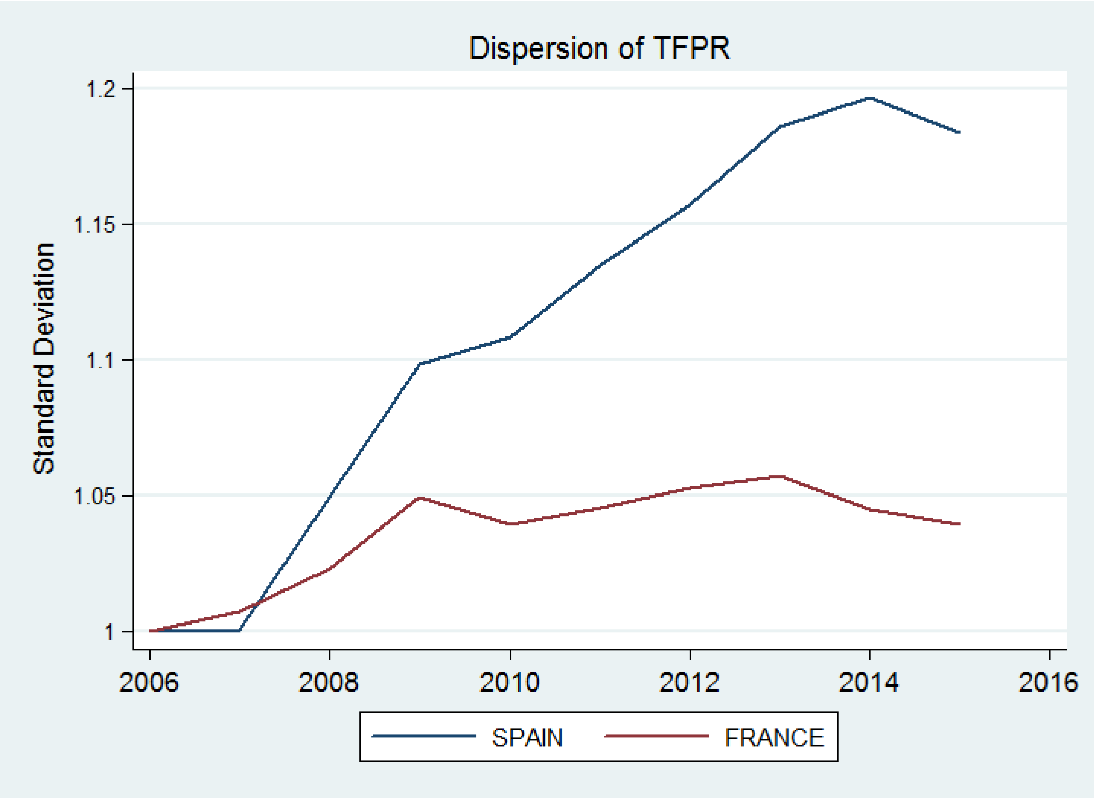


Figure 1. Evolution of TFPR dispersion

When we compare Spain and France, manufacturing TFPR dispersion of Spain has an increasing pattern in the sample period whereas that of France is relatively stable. Hence, Spain manufacturing industry suffers more seriously from the misallocation compared to France. Gopinath et al. (2017) account for this result with a model that exploits the size-dependant borrowing constraints. According to this model, capital flows into Spain manufacturing industry is not distributed optimally among firms and created a higher dispersion of MRPK by increasing the size of unproductive firms instead of productive ones. I also preferred to study with Spanish manufacturing data in this paper because of the severity of misallocation in Spain as documented in Figure 1 and Gopinath et al. (2017).

Dispersion of TFPR has a detractive effect on industry level TFP and output. In order to show that, Hsieh and Klenow (2009) and Gopinath et al. (2017) express industry level TFP using the definition of TFPR in equation (9) as the following[11]:

$$TFP_{st} = \frac{Y_{st}}{K_{st}^{\alpha} L_{st}^{1-\alpha}} = \frac{\overline{TFPR}_{st}}{P_{st}} = \left[\sum_{i}^{N_{st}} \left((D_{ist})^{\frac{\varepsilon}{\varepsilon-1}} A_{ist} \frac{\overline{TFPR}_{st}}{TFPR_{ist}} \right)^{\varepsilon-1} \right]^{\frac{1}{\varepsilon-1}} \tag{10}$$

where $K_{st} = \sum_i k_{ist}$, $L_{st} = \sum_i \ell_{ist}$, and $\overline{TFPR}_{st}$ is an industry-level TFPR measure.

If revenue-based total factor productivities were equalized across firms within industries by the equalization of marginal revenue products due to homogeneous distortions, i.e. $TFPR_{ist} = \overline{TFPR}_{st}$ for all i, industry level TFP would be efficient which is the case in which factors are allocated optimally. Hsieh and Klenow (2009) and Gopinath et al. (2017) exploits this situation to demonstrate the magnitude of aggregate TFP and output loss stem from the misallocation. As I explained the fact of high misallocation in Spanish manufacturing industry, Spain would suffer a significant TFP loss due to misallocation. In the following section, I investigate the extent of misallocation in Spain at regional level.

[11] The expression of industry level TFP is obtained using the definitions of industry price index P_{st} from equation (2), firms' prices $p_{ist} = TFPR_{ist}/A_{ist}$, and $\overline{TFPR}_{st} = P_{st}Y_{st}/K_{st}^{\alpha}L_{st}^{1-\alpha}$.

CHAPTER 5

MISALLOCATION IN AGGLOMERATED REGIONS

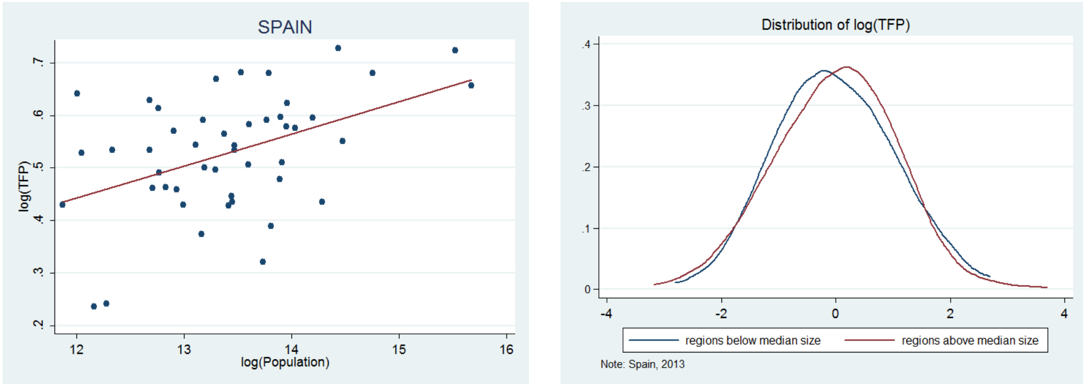
Large cities are considered to be more productive both at firm-level and aggregate level. Combes et al. (2012) and Fontagne and Santoni (2018) reveal that the productivity performance is increasing with the density of the NUTS 3 regions and employment areas in France. They explain this phenomena by the productivity enhancing effects of agglomeration factors which are basically sharing, matching, and learning with the classification of Duranton and Puga (2004). Therefore, the positive effect of agglomeration on economic performance by increasing the average productivity of firms is a verified argument in the agglomeration literature. However, the relation between agglomeration and misallocation needs to be discovered more by the researchers.

Panel a of Figure 2 depicts the positive relationship between population[12] and aggregate TFP for NUTS 3 regions of Spain[13] in our data. This relationship suggests that productivity is increasing 0.06% with 1% increase in population[14]. Panel b illustrates the distribution of aggregate log(TFP) for the regions below and above median population. The mean log(TFP) is 0.47 for the regions below the median size whereas the mean log(TFP) is 0.57 for the regions above the median size. Those graphs confirm the productivity advantage of large regions.

[12] I use 2011 population values.

[13] The number of NUTS 3 regions in Spain is 52.

[14] $R^2 = 0.20$



(a) Population and aggregate TFP (b) Distribution of aggregate TFP

Figure 2. Productivity and region size

5.1 Spatial misallocation in Spain

Now, I amplify the analysis of misallocation across firms within industries by including regional dimension to the Hsieh and Klenow (2009) model to investigate the effect of agglomeration on misallocation beyond the productivity. If we assume mobility of factors across regions in the short-run, we are able to implement the methodology above for the regions and compare the level of misallocation in different regions. I prefer to use 3-digit industry codes[15] instead of 4-digit when working at regional level since it would not be possible to calculate the dispersion measures in a large number of sectors with small number of observations.

Table 4 documents the cross-section regression results of the standard deviation of log(TFPR) on log(population) for the year 2011. Models 2 and 4 control for the mean log(TFPR) within region-sector groups. Controlling for the average productivity is important since our misallocation measure is a mean dependent measure and we expect higher standard deviation in larger regions because of the higher overall productivity as documented in figure 2. Regression results suggest that, even if we control for the average productivity, dispersion of TFPR is increasing with population.

[15]The number of 3-digit manufacturing industries in NACE Revision 2 classification is 94.

Table 4. Cross-Section Regression of Misallocation on Agglomeration

	(1) sdltfpr	(2) sdltfpr	(3) sdltfpr	(4) sdltfpr
lpop	0.0258***	0.0159***	0.0245***	0.0161***
	(6.47)	(3.46)	(6.66)	(3.46)
mltfpr		0.130***		0.170***
		(8.12)		(8.63)
_cons	0.126**	0.169***	0.143***	0.138**
	(2.32)	(2.69)	(2.80)	(2.18)
year	2011	2011	2011	2011
N	1635	1635	1635	1635
industry FE			yes	yes

t statistics in parentheses
Robust Standard Errors for all regressions
Clustered at region groups
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5. Panel Regression of Misallocation on Agglomeration

	(1) sdltfpr	(2) sdltfpr	(3) sdltfpr	(4) sdltfpr	(5) sdltfpr	(6) sdltfpr	(7) sdltfpr	(8) sdltfpr
lemp	0.0125***	0.0111***	0.0117***				0.00393***	
	(6.57)	(5.44)	(9.34)				(2.70)	
lnof				0.0178***	0.0162***	0.0147***		0.00569***
				(7.77)	(6.59)	(9.58)		(3.31)
mltfpr							0.144***	0.143***
							(22.89)	(23.18)
_cons	0.350***	0.364***	0.269***	0.349***	0.360***	0.283***	0.248***	0.248***
	(19.38)	(18.86)	(20.93)	(22.88)	(21.99)	(24.98)	(18.23)	(21.10)
N	16148	16148	16148	16148	16148	16148	16148	16148
industry FE		yes	yes		yes	yes	yes	yes
year FE			yes			yes	yes	yes

t statistics in parentheses
Robust Standard Errors for all regressions
Clustered at region-year groups
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel regression results in Table 5 with two alternative agglomeration measures, logarithm of employment and logarithm of number of firms at region level, are similar with the cross-section results above. Both employment and number of firms have significant positive effects on the dispersion of TFPR. Therefore, we observe higher misallocation in larger regions in spite of the productivity enhancing agglomeration forces.

It can be counter intuitive to get a negative relationship between allocative efficiency and region size at first glance; however, there might be some negative forces that deteriorate the allocative efficiency related with agglomeration in spite of the higher average productivity in agglomerated regions. Indeed, regression results confirm this idea by documenting the increasing misallocation even if we control for the average productivity.

Now, we need to investigate the reason behind the higher misallocation in larger regions. One possible explanation might be the higher competition in agglomerated regions. Higher competition increases the elasticity of substitution between firms within industries by expediting demand shifts across varieties of a particular good. Therefore, easier movements of production factors between firms may worsen the allocative efficiency among firms within the same industry by amplifying the effect of idiosyncratic distortions. In the next section, I explain the importance of the elasticity of substitution on the factor misallocation.

5.2 The effect of elasticity of substitution on misallocation

Hsieh and Klenow (2009) emphasize a special case for the industry-level TFP expressed in equation (10) when the firm productivity and TFPR are jointly log-normally distributed and present the following closed-form expression for aggregate TFP:

$$\log(TFP_{st}) = \frac{1}{\varepsilon-1}\log\left(\sum_{i=1}^{N_{st}} D_{ist}^{\varepsilon} A_{ist}^{\varepsilon-1}\right) - \frac{\varepsilon}{2}\mathrm{var}(\log(TFPR_{ist})) \qquad (11)$$

Here, the variance of TFPR is a decreasing factor for the industry-level TFP. As I discussed in section 4, dispersion of TFPR across firms within the same industry, which is an indicator of misallocation, would decrease the overall economic performance compared to the first-best allocation case. Further to that, it is also important to point out the multiplier effect of ε, elasticity of substitution, on the dispersion of TFPR. In other words, the cost of misallocation on productivity is increasing with ε for a given firm productivity distribution. This relationship is documented by Epifani and Gancia (2011) in a different form as the effect of elasticity on misallocation and welfare equation with a specific model and data.

In order to clarify this relationship, I construct a ratio of industry-output with idiosyncratic distortions and output with efficient allocation using the model introduced in section 4. This ratio provides us the output loss due to misallocation. First, I obtain the labor share of each firm in industry by combining the demand equation in (2), capital to labor ratio in (7), and the firm price in (8).

$$\frac{\ell_{ist}}{L_{st}} = \frac{D_{ist}^{\varepsilon} A_{ist}^{\varepsilon-1}(1+\tau_{ist}^{k})^{\varepsilon(1-\alpha)}(1-\tau_{ist}^{y})^{\varepsilon}}{\sum_{j}^{N_{st}} D_{jst}^{\varepsilon} A_{jst}^{\varepsilon-1}(1+\tau_{jst}^{k})^{\varepsilon(1-\alpha)}(1-\tau_{jst}^{y})^{\varepsilon}} \qquad (12)$$

Then, I obtain the industry-level output with and without idiosyncratic distortions as the following:

$$Y_{st} = \left[\sum_{i}^{N_{st}} D_{ist} \left(A_{ist} \left[\frac{\alpha}{1-\alpha} \frac{w_{st}}{R_{st}} \frac{1}{1+\tau_{ist}^{k}} \right]^{\alpha} L_{st} \frac{D_{ist}^{\varepsilon} A_{ist}^{\varepsilon-1} (1+\tau_{ist}^{k})^{\varepsilon(1-\alpha)} (1-\tau_{ist}^{y})^{\varepsilon}}{\sum_{j}^{N_{st}} D_{jst}^{\varepsilon} A_{jst}^{\varepsilon-1} (1+\tau_{jst}^{k})^{\varepsilon(1-\alpha)} (1-\tau_{jst}^{y})^{\varepsilon}} \right)^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{13}$$

$$Y_{st}^{eff} = \left[\sum_{i}^{N_{st}} D_{ist} \left(A_{ist} \left[\frac{\alpha}{1-\alpha} \frac{w_{st}}{R_{st}} \right]^{\alpha} L_{st} \frac{D_{ist}^{\varepsilon} A_{ist}^{\varepsilon-1}}{\sum_{j}^{N_{st}} D_{jst}^{\varepsilon} A_{jst}^{\varepsilon-1}} \right)^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{14}$$

Dividing Y_{st} and Y_{st}^{eff} gives the following ratio which identifies the output with idiosyncratic distortions divided by the output with homogeneous distortions.

$$\frac{Y_{st}}{Y_{st}^{eff}} = \left[\frac{\sum_{i}^{N_{st}} z_{ist} \left(\left(\frac{1}{1+\tau_{ist}^{k}} \right)^{\alpha} \frac{\left(1+\tau_{ist}^{k}\right)^{\varepsilon(1-\alpha)} \left(1-\tau_{ist}^{y}\right)^{\varepsilon}}{\sum_{j}^{N_{st}} z_{jst} \left(1+\tau_{jst}^{k}\right)^{\varepsilon(1-\alpha)} \left(1-\tau_{jst}^{y}\right)^{\varepsilon}} \right)^{\frac{\varepsilon-1}{\varepsilon}}}{\sum_{i}^{N_{st}} z_{ist} \left(\frac{1}{\sum_{j}^{N_{st}} z_{jst}} \right)^{\frac{\varepsilon-1}{\varepsilon}}} \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{15}$$

where $z_{ist} = D_{ist}^{\varepsilon} A_{ist}^{\varepsilon-1}$ is a combination of demand shifter and firm productivity. I call z_{ist} as "firm productivity" henceforth in this paper.

This ratio quantifies the output loss due to misallocation in a monopolistic competition model with idiosyncratic output and capital distortions. Finally, I simulate this ratio with log-normally distributed firm productivity[16], z_{ist}, and normally distributed output and capital distortions[17] for four different elasticity values[18]. Then, I regress the simulated output ratios on the correlation between firm productivity and output distortions. Table 6 reports the simulated average output ratios and the coef-

[16] parameters of the lognormal distribution of z_{ist} is obtained by fitting the distribution in data.

[17] I use $\mu = 0.2$ and $\sigma = 0.2$ for this distribution.

[18] I choose those 4 values since the estimations in the literature for the elasticity of substitution for manufacturing industries are in the range of [1.5,10].

ficients of regressions. Figure 3 depicts the relationship between the output loss and correlation of firm productivity with output distortions for four different elasticity values. Those results confirm the negative relationship between the elasticity and allocative efficiency given productivity and distortion distribution.

Table 6. Simulation Results

	Y/Y^{eff}	β^*
$\varepsilon = 1.5$	0.8976	-0.0882
$\varepsilon = 3$	0.8615	-0.1766
$\varepsilon = 5$	0.8271	-0.2681
$\varepsilon = 10$	0.7809	-0.3532

*β is the regression coefficient of output ratio on the correlation of firm productivity and output distortions. Coefficients are significant at 0.01 significance level.

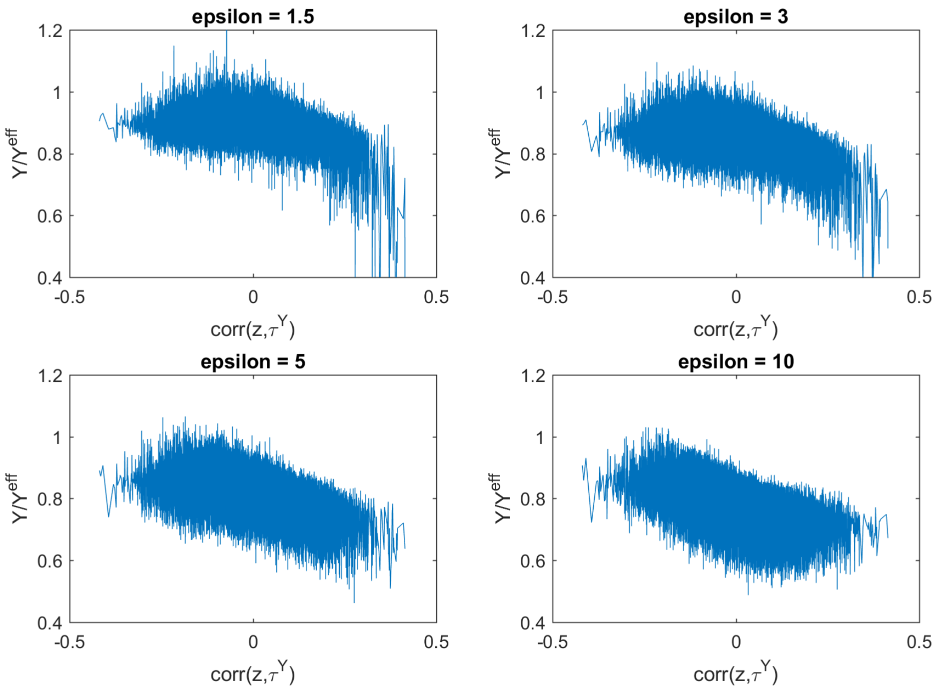


Figure 3. Simulation graphs

This simulation brings out two important implications about the role of the elasticity of substitution on misallocation. First, industry output relative to efficient level is decreasing with the elasticity of substitution. In other words, the output loss due to misallocation is higher when ε is larger. This result is consistent with the equation (11) and its interpretation provided above. When the elasticity increases, markups (the difference between marginal revenues and marginal costs) would decrease; hence firm-level efficiency would improve as shown in Fontagne and Santoni (2018). However, a rise in elasticity amplifies the negative effect of idiosyncratic distortions on allocative efficiency at the same time by increasing the similarity of product varieties and increasing the pass-through of consumer demand among firms. Therefore, we can observe higher levels of misallocation in agglomerated regions together with higher productivity levels due to agglomeration. In other words, those two forces may work reciprocatively on productivity; however, we are able to decompose the negative effect of misallocation on productivity with the ratio above.

Second implication is about the correlation between firm productivity and distortions. The negative relationship between this correlation and relative output ratio is intuitive. If we impose larger distortions to more productive firms, they will shrink more compared to less productive firms. Hence, this correlation is increasing the effect of distortions on allocative efficiency. Furthermore, the negative coefficient of the correlation term is increasing in absolute value with ε. The intuition is that the rise in elasticity amplifies also the negative effect of the correlation between firm productivity and distortions on the allocative efficiency. That is to say, as the elasticity increases, allocative efficiency responds more to the idiosyncratic distortions and hence we observe a higher output loss due to misallocation.

5.3 Concentration and competition

In this section, I try to test empirically the role of price elasticity of substitution on the observed higher misallocation in agglomerated regions. I account for this relationship with the high elasticity of substitution which stem from low concentration and high competition in denser areas. I use Herfindahl-Hirschman Index (HHI) to measure concentration and competition for each industry in each region in Spain. HHI is calculated as following:

$$hhi_{srt} = \sum_{i}^{N_{srt}} share_{isrt}^{2} \tag{16}$$

where i is firm index, N_{srt} is the number of firms operating in sector s in region r at year t, $share_{isrt}$ is the value added share of each firm in total industry-region-year value added. Low HHI means low concentration and high competition in that industry.

I regress HHI for each industry-region-year group on agglomeration measures. Table 7 presents the cross-section regression results of HHI on the logarithm of population for three different years, with and without industry dummies. Table 8 presents panel regression results with two alternative agglomeration measures. The coefficients of agglomeration measures are negative and significant at 0.01 significance level. Therefore, we can conclude that concentration is lower and competition is higher in industries located in agglomerated regions.

Competition is one of the factors that determine the elasticity of substitution between similar products in the same industry beyond the substitutability of goods in their nature. Therefore, we would expect larger elasticity of substitution when the

Table 7. Cross-Section Regression of Concentration on Agglomeration

	(1)	(2)	(3)	(4)	(5)	(6)
	hhi	hhi	hhi	hhi	hhi	hhi
lpop	-0.0271***	-0.0301***	-0.0375***	-0.0594***	-0.0634***	-0.0693***
	(-4.21)	(-5.54)	(-6.48)	(-7.73)	(-8.63)	(-9.16)
_cons	0.593***	0.642***	0.750***	1.035***	1.098***	1.186***
	(6.83)	(8.80)	(9.62)	(9.94)	(11.09)	(11.62)
year	2006	2010	2015	2006	2010	2015
N	1556	1625	1638	1556	1625	1638
industry FE				yes	yes	yes

t statistics in parentheses

Robust Standard Errors for all regressions

Clustered at region groups

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8. Panel Regression of Concentration on Agglomeration

	(1)	(2)	(3)	(4)	(5)	(6)
	hhi	hhi	hhi	hhi	hhi	hhi
lemp	-0.0237***	-0.0535***	-0.0534***			
	(-16.94)	(-40.79)	(-40.52)			
lnof				-0.0321***	-0.0691***	-0.0695***
				(-22.30)	(-55.83)	(-56.93)
_cons	0.462***	0.752***	0.661***	0.452***	0.708***	0.614***
	(34.17)	(57.93)	(44.24)	(45.29)	(78.68)	(57.80)
N	16148	16148	16148	16148	16148	16148
industry FE		yes	yes		yes	yes
year FE			yes			yes

t statistics in parentheses

Robust Standard Errors for all regressions

Clustered at region-year groups

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

competition is higher. For instance, elasticity is 0 for a monopoly and it goes to infinity for the industries with perfect competition. As we documented lower concentration and higher competition in agglomerated regions with HHI above, we would expect larger elasticities (ε) in agglomerated regions.

I documented the negative effect of high elasticity on allocative efficiency in section 5.2. Therefore, we can associate the high levels of misallocation in agglomerated regions presented in section 5.1 with the high levels of competition in those regions presented in this section.

CHAPTER 6

CONCLUSION

The importance of allocative efficiency on the economic performance of developing countries is a commonly held issue in economics literature. Furthermore, misallocation of production factors is a serious issue also for developed countries as discussed in the last decades. Misallocation may cause extensive output loss in many developed countries as well as developing countries. Gopinath et al. (2017) draw attention to the high capital misallocation and productivity loss in southern European countries, specifically in Spain, before and after the Euro crisis. Therefore, I use the firm-level data for Spain to investigate the misallocation in Spain with a focus on the effect of agglomeration on allocative efficiency. I discover a negative relationship between agglomeration and allocative efficiency in contradiction with the general perception and the higher productivity in denser areas.

I explain the negative effect of agglomeration on allocative efficiency with the higher price elasticity of substitution due to higher competition in agglomerated regions. High elasticity in large cities amplifies the effect of existing idiosyncratic distortions on allocative efficiency. I test this argument using a monte carlo simulation of output loss due to misallocation. Finally, I document the low concentration and high competition in agglomerated regions with Herfindahl-Hirschman Index.

APPENDIX A

DATA CLEANING

I implement the relevant data management processes used by Kalemli-Ozcan et al. (2015) and Gopinath et al. (2017).

There is no exact year variable for each observation in the data; however, “fiscal year end date (CLOSDATE)” and “year part of CLOSDATE” are reported for each observation. I construct YEAR variable by determining June 1st as threshold for fiscal year end date. I assign those observations with fiscal year end date before June 1st to the previous year and the current year for later closing firms. Then, I obtain an unbalanced panel data for the period between 2005-2016; however, we drop observations for 2005 and 2016 since those years have very low coverage compared to other years and including them may cause misrepresentation for those years. I have very low number of firms in 2016 because of the reporting lag, that is the data for 2016 has not been fully updated yet.

I first implement the following basic data cleaning steps:

1. I drop consolidated accounts and keep only unconsolidated accounts.
2. Delete the observations with just a name or missing BvD ID or BvD Account number.
3. There are duplicating firm-year observations in the data. Drop those duplicates.
4. Drop observations with missing fiscal year end date.
5. Drop observations with missing NACE industry code.

6. Drop company-years with missing information on total assets and operating revenue and sales and employment (simultaneously).

7. Drop the entire company (all years) if total assets is negative in any year.

8. Drop the entire company if sales are negative in any year.

9. Drop the entire company if Tangible Fixed Assets (buildings, machinery, etc) is negative in any year.

10. Drop observations with mising, zero or negative operating revenue.

11. Drop observations with missing, zero or negative material cost.

12. Drop observations with mising, zero or negative total assets.

13. Drop observations with missing, zero or negative wage bill.

14. Drop observations with currencies other than euro. I drop those observations instead of converting to euro since there are negligible amount of them.

Then, in order to check the internal consistency of balance sheet information we exclude from the analysis extreme values by dropping 0.1 and 99.9 percentiles for the following ratios:

1. (Intangible + Tangible + Other Fixed Assets) / (Fixed Assets)

2. (Stocks + Debtors + Other Current Assets) / (Current Fixed Assets = Total Current Assets)

3. (Fixed Assets + Current Assets) / (Total Assets)

4. (Capital + Other Shareholder Funds) / (Total Shareholder Funds)

5. (Long Term Debt + Other Non-Current Liabilities) / (Total Non-current Liabilities)

6. (Loans + Creditors + Other Current Liabilities) / (Total Current Liabilities)

7. (Non-current Liabilities + Current Liabilities + Shareholder Funds) / (Total Shareholder Funds Liabilities)

Then we construct the value added variable as the difference between *operating revenue* and *material costs* (Gopinath et al, 2017), and drop the negative values for value added. Then we calculate ratio of wage bill to value added and drop the extreme values in 1 and 99 percentiles. I also drop the observations with ratio higher than 1.1.

I obtain capital stock by the total value of tangible and intangible fixed assets. For the quality of this measure, we drop observations with negative intangible fixed assests, zero or missing tangible fixed assets, and higher values for tangible fixed assets than total assets. I also drop observations with negative depreciation.

I also construct capital-labor ratio as capital stock divided by wage bill, and drop 0.1 and 99.9 percentiles for this ratio.

Finally, we winsorize value added, fixed assets, wage bill, operating revenues, material costs, and capital stock at 1 and 99 percent. I also winsorize calculated values of MRPK and MRPL at 0.1 and 99.9 percent, and TFPR at 1 and 99 percent before calculating dispersion measures to make them less sensitive to outliers.

APPENDIX B

MANUFACTURING SECTORS

	2-digit Manufacturing Sectors (NACE Revision 2)
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and of products of wood and cork, except furniture, etc.
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
22	Manufacture of rubber and plastic products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals
25	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment n.e.c.
29	Manufacture of motor vehicles, trailers and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment

REFERENCES

Bartelsman, E., Haltiwanger, J., & Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *The American Economic Review, 103*(1), 305-334.

Bartelsman, E. J. & Wolf, Z. (2017). *Measuring productivity dispersion* (Tinbergen Institute Discussion Papers 17-033/VI). Retrieved from https://papers.tinbergen.nl/17033.pdf

Bento, P. & Restuccia, D. (2016). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics, 9*(3), 267-303.

Bin, P., Chen, X., Fracasso, A., & Tomasi, C. (2018). Resource allocation and productivity across provinces in China. *International Review of Economics and Finance.* doi: 10.1016/j.iref.2018.02.015

Combes, P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica, 80*(6), 2543-2594.

Dias, D., Marques, C., & Richmond, C. (2016). Misallocation and productivity in the lead up to the Eurozone crisis. *Journal of Macroeconomics, 49,* 46-70.

Duranton, G. & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, J. V. & Thisse, J. F. (Eds.), *Handbook of Regional and Urban Economics, 4*, (pp. 2063-2117). The Netherlands: Elsevier.

Epifani, P. & Gancia, G. (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics, 83*(1), 1-13.

Fontagné, L. & Santoni, G. (2018). Agglomeration economies and firm-level labor misallocation. *Journal of Economic Geography*. doi:10.1093/jeg/lby007

Foster, L., Haltiwanger, J., & Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?. *American Economic Review, 98*(1), 394-425.

Gopinath, G., Kalemli-Ozcan, S., Karabarbounis, L., & Villegas-Sanchez, C. (2017). Capital allocation and productivity in south Europe. *Quarterly Journal of Economics, 132*(4), 1915–1967.

Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics, Annual Reviews, 6*(1), 735-770.

Hsieh, C. & Klenow, P. (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics, 124*(4), 1403-1448.

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., & Yesiltas, S. (2015). *How to construct nationally representative firm level data from the ORBIS global database* (NBER Working Papers 21558). Retrieved from http://www.nber.org/papers/w21558.pdf

Olley, G., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*(6), 1263-1297.

Petrin, A. & Sivadasan, J. (2013). Estimating lost output from allocative inefficiency, with an application to Chile and firing costs. *Review of Economics and Statistics, 95*, 286-301.

Restuccia, D. (2018). *Misallocation and aggregate productivity across time and space* (Working Papers tecipa-596). Retrieved from https://www.economics.utoronto.ca/public/workingPapers/tecipa-596.pdf

Restuccia, D. & Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics,Elsevier for the Society for Economic Dynamics, 16*(1), 1-10.

Restuccia, D. & Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives, American Economic Association, 31*(3), 151-174.

Ryzhenkov, M. (2016). Resource misallocation and manufacturing productivity: The case of Ukraine. *Journal of Comparative Economics, 44*(1), 41-55.

Syverson, C. (2014). The importance of measuring dispersion in firm-level outcomes. *IZA World of Labor 2014: 53*. doi: 10.15185/izawol.53