Global Research Unit
Working Paper #2020-014

From the Regional Economy to the Macroeconomy

Santiago Pinto, Federal Reserve Bank of Richmond
Pierre-Daniel G. Sarte, Federal Reserve Bank of Richmond

From the Regional Economy to the Macroeconomy[ϒ]

Santiago Pinto & Pierre-Daniel G. Sarte,[κ]

Abstract

This chapter describes how two fields that traditionally evolved mostly separately, regional economics and macroeconomics, have increasingly come together over the past decade and a half to yield new insights into the relevance of regional forces for the macroeconomy. This chapter gives an overview to the basic question: why should macroeconomists care about the spatial allocation of economic activity or spatial models? There are no simple spatial aggregation theorems that give rise to an aggregate production function, and this chapter describes the variety of ways in which granular considerations and shocks that are regional in nature shape aggregate outcomes and motivate a need for policy. The macroeconomics literature is increasingly heading in the direction of unpacking the exact nature of granular forces in a way that leaves the representative agent and firm framework with aggregate shocks as an early and poor approximation to how actual economies work.

JEL classification number: R12, R38, R30, R5, E13

Keywords: Agglomeration, Externalities, Aggregation, Region, Nation

ϒ Acknowledgement: The views expressed herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

κ Correspondence: Both authors are from the Federal Reserve Bank of Richmond, USA.

1 INTRODUCTION

This chapter focuses on how and why the spatial distribution of economic activity is relevant for aggregate outcomes. In other words, topics that have traditionally been studied by regional or urban economists are not without implications for questions and concerns pertitent to macroeconomists. Economic activity is not uniformly distributed across regions. The composition of economic activity also varies across space. Some industries locate and operate in certain locations, and then trade with other industries located elsewhere. The nature of trade, and factors that influence it such as the distribution of trade costs and productivity across space, then affects aggregate allocations in a way that, in the last decade, has started to be fully fleshed out.

The spatial concentration of economic activity produces a number of benefits. Places with higher production and population densities exhibit faster growth in productivity and per capita GDP, resulting in higher wages. These places are also wellsprings of innovation, accounting for a disproportionate share of new patents.

The literature in urban and regional economics gives particular importance to the advantages of the spatial concentration of economic activity. These advantages are typically referred to as “agglomeration economies". While these advantages are certainly relevant at the local level, are they quantitatively important at the aggregate level? Do granular shocks, originated at the local or regional level, simply level out and vanish at the aggregate level, or do they propagate and amplify affecting the overall economy in a meaningful way?

Recent research has focused on understanding not only how the spatial characteristics of an economy determine the economic performance of regions, but also how they explain fluctuations in aggregate outcomes. In general, this research concludes that the effects observed at the aggregate level are not independent of the allocation of resources across space. The transmission and propagation of local and regional shocks, and their economic importance, critically depend on the spatial distribution of resources. To the extent that the existence of spatial spillovers, externalities, and mobility frictions lead to a sub-optimal allocation of factors of productions across locations, it becomes relevant to evaluate the ability of certain policy interventions to correct for those distortions.

The next sections review recent work that examines the connection between the regional allocation of resources and aggregate outcomes, and highlights directions for future research.

2 WHY SPACE MATTERS

Consider an economy in which locations are completely homogeneous, transportation is costly, and there are no economies of scale. Then, in a competitive equilibrium, each location is self-sufficient, i.e, regions do not trade (in equilibrium, there are no shipments of goods across regions). The latter is known as the “spatial impossibility theorem" (Starrett (1978)).

However, economic activity tends to be concentrated in a limited number of locations. Regions do not necessarily produce every good and service, so they trade. The clustering of productive activities, firms, and workers give rise to agglomeration economies, which enhances the locational advantages of a specific region. The spatial allocation of economic activity determines how much and the type of trade takes place across regions (Rossi-Hansberg (2005)).

Locations are heterogeneous, differing in several dimensions, such as endowments, geography, and accessibility, and may have different comparative advantages (Behrens and Robert-Nicoud (2015)). The location of economic activity is the outcome of the interaction of various forces. Individuals and firms are also heterogeneous. They sort themselves across space and decide where to locate driven by their preferences and productive abilities. These decisions are, however, constrained by spatial frictions, including mobility and transportation costs. As people decide where to reside and firms where to establish their operations, they affect the outcomes of other people and firms.

2.1 A simple conceptual framework

Understanding the relative importance of regional attributes and why some locations have certain attributes are key to explain the observed variation of spatial outcomes. What explains location choices of mobile factors? Are places attracting people by offering high wages or cheap housing or good weather? Why do firms stay in places where they must pay high wages? If all factors are completely mobile, why do some regions perform systematically better than others in terms of economic development and wealth? Moreover, the degree of factor mobility (households, firms) constrains the design of regional policy.

In this section, we briefly describe a simple theoretical framework used by regional economists, to explain spatial outcomes. Such framework, founded on the seminal work by Roback (1982), is built on three key no arbitrage conditions:

(1) Individuals must be indifferent across space: Flows of wages + amenities - housing costs is equal in every location

(2) Firms must be indifferent over space and over hiring new workers: Differences in wages must be offset by differences in productivity

(3) Developers must be indifferent about developing or not land or about building or not new units: Housing prices cannot rise too far above the total costs of construction

Different versions and extensions of the basic model developed by Roback (1982) are commonly used as the main analytical tool to predict how wages and land (or housing) prices adjust to differences in amenities across regions. The model has also been used to construct indexes of quality of life and quality of business environment. The main underlying assumption is that firms and households move across cities to attain the highest possible profit (firms) and utility (households). In equilibrium, generally referred to as a spatial equilibrium, there are no incentives to move, i.e., profits and utility levels should be equalized across space. This means that if a region has amenities that makes it a nice place to live, it will attract households until higher housing prices, lower wages, or a combination of both eliminate the incentives to move. A location that offers a good

business environment will attract firms into the region until a combination of higher land rents and higher wages makes it no longer desirable to move into that region. In other words, land or housing prices and wages vary to compensate consumers and firms for the interregional differences in the “quality" of locations.

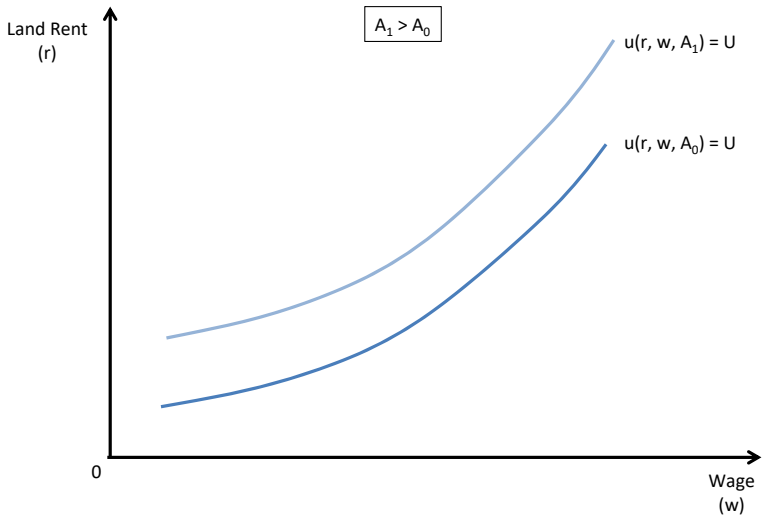
2.1.1 *The model*

Consider a system of regions, where each region i is characterized by a level of amenity A_i. Households (workers) and firms decide where to locate, and their decisions depend on wages earned and paid at each location (w), the cost of living at each location, driven mostly by land rents and housing prices (r), and the local amenity (A_i), which may affect households and firms differently.

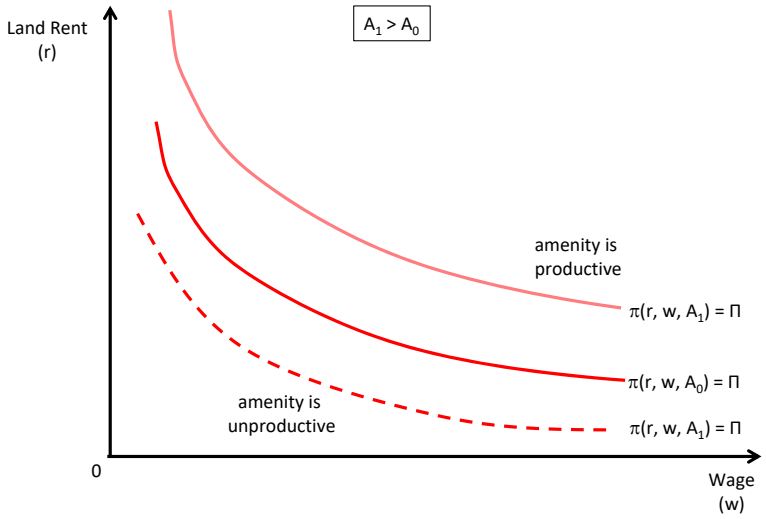
HOUSEHOLDS. The utility of a consumer in region i is described by $u(r, w, A_i)$. While a higher r reduces utility, higher levels of w and A_i increase utility. Consider a constant-utility curve for region i, defined as $u(r, w, A_i) = U$. This curve implicitly defines combinations of $\{r, w\}$ that offer the same utility U, given an amenity level A_i. As shown in Figure 1a, the curve depicts a positive relationship between r and w. If w increases and r remains constant, then utility increases. To restore utility level U, r should therefore increase. Moreover, consider two regions and suppose the amenity level in region 1 is higher than in region 0, i.e., $A_1 > A_0$. Then, the constant-utility curve for region 1 lies above the constant-utility curve for region 0. Consider a given combination of $\{\bar{w}, \bar{r}\}$ on the constant-utility curve of region 0. Since the utility increases as A_i rises, the utility evaluated at $\{\bar{w}, \bar{r}\}$ and A_1 will be higher than the utility evaluated at $\{\bar{w}, \bar{r}\}$ and A_0. To restore utility U, then r should increase and/or w should decrease.

FIRMS. The profit of a firm that operates in region i is given by $\pi(r, w, A_i)$. Both higher land rents r or wages w decrease profits. For firm, the amenity in region i can be productive (reduce production costs or increase productivity), in which case a higher A_i increases profits, or unproductive (increase production costs or reduce productivity), so that a higher A_i decreases profits. An isoprofit curve for region i is defined as combinations of r and w that gives the same profit Π for a fixed amenity level A_i. The curve, shown in Figure 1b, depicts an inverse relationship between r and w. If r increases, profits go down. To restore the previous level of profits Π, w should decrease.

As mentioned earlier, local amenities may affect firms and households differently. For instance, Suppose that amenities are productive (*unproductive*) for the firm. Then, if $A_1 > a_0$, the isoprofit curve for region 1 will lie above (*below*) the isoprofit curve of region 0.



(a) Households



(b) Firms

Figure 1: Households and firms: $A_1 > A_0$

SPATIAL EQUILIBRIUM. Suppose an economy with two regions, 0 and 1, with amenity levels A_0 and A_1. In a spatial equilibrium: (i) consumers should be equally well-off in all locations, so that $u(r, w, A_0) = u(r, w, A_1) = U$; and (ii) firms should have the same profits at all locations, so $\pi(r, w, A_0) = \pi(r, w, A_1) = \Pi$. In Figure 2, the equilibrium wage and land rent at a location with amenity A_i are represented by the $\{w_i, r_i\}$ (point E_i).

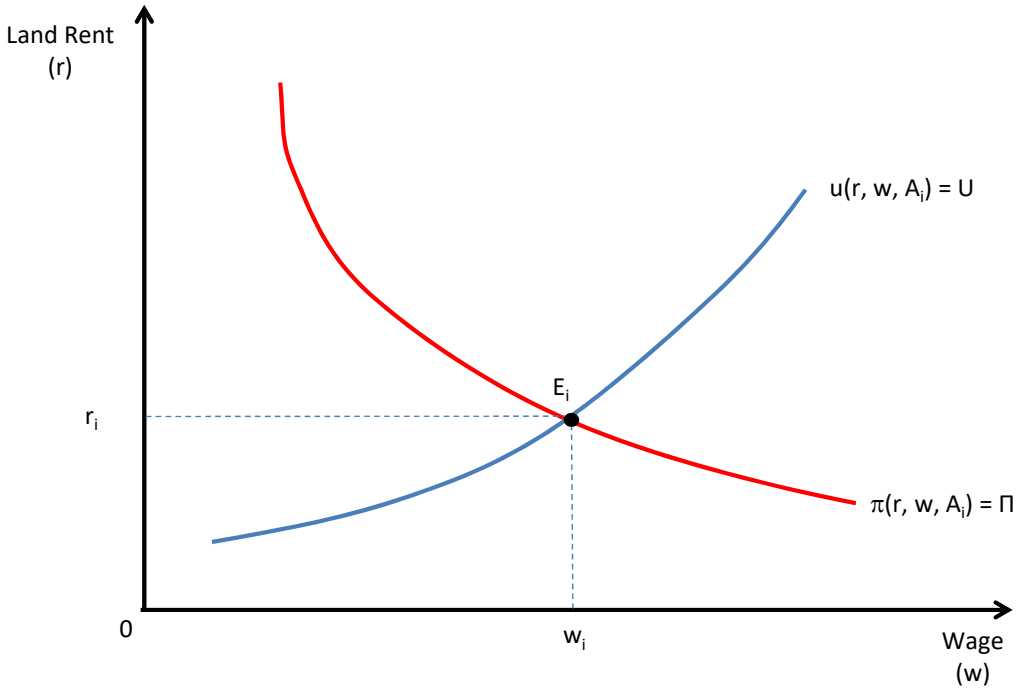
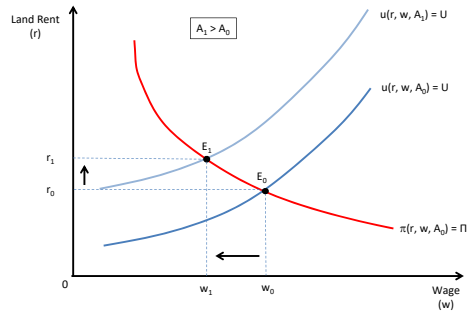


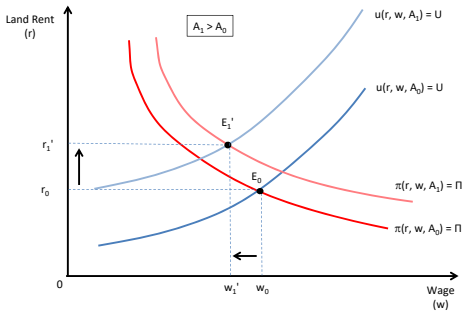
Figure 2: Spatial equilibrium

How can this framework be used to explain the spatial variation in land rents and wages? Consider two regions 0, 1. Suppose that region 1 is always more attractive than region 0 for consumers, so that $A_1 > A_0$. However, suppose A_i may have different effects on the costs or productivity of firms that produce in 1: (a) A_i may not have an impact, (b) A_i may only have a weak impact, or (c) A_i may have strong impact on local productivity. The three cases are shown in Figure 3. From the figures we conclude that land rents are unambiguously higher in region 1 than in region 0. However, wages can either be higher or lower. For instance, if the effect of the amenity on firm productivity is sufficiently small in magnitude (represented by cases (a) and (b)), then a higher level of A_i in region 1 would lead to lower wages at that location. If the effect is sufficiently large (such as in case (c)), then wages may also be higher in region 1.

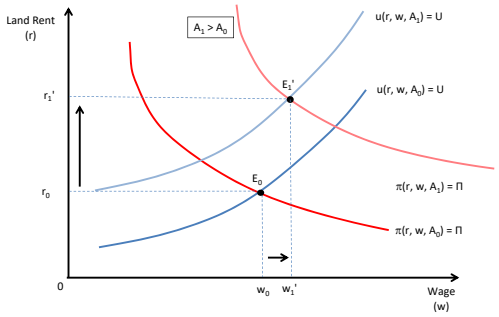
The shift from A_0 to A_1 can also be interpreted as a positive productivity shock that only affects firms that produce in a given region (in other words, suppose the constant-utility curve does not shift). The positive shock will then tend to increase land rents, wages, or a combination of both. The precise outcome will depend on the curvature of the constant-utility curve. For example, if this curve is vertical, then the local productivity shock will be fully capitalized into land values, while if the curve is completely horizontal, it would only increase nominal wages. In general, the shock will tend to increase both land rents and wages.



(a) Amenity has no impact on firms



(b) Weak production amenity



(c) Strong production amenity

Figure 3: Spatial equilibrium: Comparing two regions ($A_1 > A_0$)

Several factors explain how much wages and land rents would vary across regions or how they would change when a region experiences a productivity shock. These factors include mobility costs (for instance, mobility costs may be different for households with different skill levels, firms that operate in certain sectors may need to locate close to natural resources or other key inputs, making them less spatially mobile), the elasticity of housing supply (determined by both geographical constraints and local regulations), local taxes, and the quality of locally provided goods and services, and the quality of the transportation infrastructure, among others.

The main takeaway from this framework is that to the extent that factors of production are spatially mobile, in equilibrium land rents and wages will reflect the region's attributes.

EXTENSIONS. The basic Roback model has been extended in several directions to examine and understand the role played by several factors in determining the spatial distribution of factors of production. Modern approaches in regional economics, for instance, build on extended versions of the Roback model to quantify the impact of a variety of policy interventions. Some of the extensions assume that workers are heterogeneous in their skills, the type of work they are best suited for, or in their mobility costs. Firms may be heterogeneous as well: some of them may be more productive than others. Amenities, both in consumption and production, can be assumed to be, at least in part, endogenous.

For instance, A_i may depend on the population size, the demographic composition, or the number and types of firms operating in each region. We will review some of this literature in the following sections.

2.2 Implications of regional considerations for macroeconomics

Macroeconomists have traditionally looked to aggregate shocks as sources of aggregate economic fluctuation or trends. More recently, however, they are have come to increasingly rely on detailed and disaggregated data capturing granular and idiosyncratic shocks as sources of, and to characterize, aggregate changes.

The work by Gabaix (2011) was among the first to establish the importance of granularity to understanding both aggregate fluctuations and aggregate trends (the "granular hypothesis"). The paper claims that when the distribution of firm sizes in an economy has a "fat tail", idiosyncratic shocks affecting large firms do not simply level out and vanish in the aggregate. In fact, these shocks has the potential of generating considerable aggregate fluctuations.

The propagation of granular shocks crucially depends on how the economy is organized. Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) show that granular shocks may trigger aggregate fluctuations when sectors are interrelated through input and output linkages. A shock to a specific sector may be amplified as it triggers a chain of reaction to other sectors through the input-output network.

Giroud and Mueller (2019) uses U.S. establishment level data to examine how demand shocks propagate across regions through the firms' internal network. Their analysis uses the regional variation in housing prices observed during the Great Recession as drivers of negative local consumer demand shock. They find that the shock does not only adversely affect establishment-level employment in the local non-tradeable sector (restaurants, retailers, grocery stores), but it also negatively affects employment at locations where the parent firms operate.

Recent developments in quantitative models that combine calibration and structural estimation techniques with granular data are increasingly used in several fields in economics. This research approach has been widely used in international trade. Costinot and Rodríguez-Clare (2014) offers a thorough review of the work that relies on micro-founded gravity models to quantify different counterfactuals, such as the aggregate trade liberalization.

A similar literature has been emerging in urban and regional economics with the development of a wide of quantitative spatial models. Many of these models build on the theoretical framework developed by Roback (1982) summarized in the previous section. While the literature is broad, most of the research focuses on examining the broader implications of shocks and policy interventions that take place at the local or regional level, such as changes in amenities, productivity, or the expansion of the transportation network. Redding and Rossi-Hansberg (2017) thoroughly reviews recent work in these areas.[1]

1 Also see Gilles Duranton and Strange (2015) for a survey on different applications of structural estimation to urban equilibrium.

An analysis based on quantitative models offers several advantages over the traditional reduced-form approach. First, since this methodology is directly founded on theoretical models, it becomes easier to identify the causal effects of policies. Moreover, these models are constructed to explicitly account for the aggregation of granular shocks. Second, quantitative models can be used to construct counterfactual policy evaluation exercises, and characterize, quantify and compare outcomes under different scenarios. One advantage of this approach over the traditional reduced-form analysis is that it accounts for the general equilibrium of the policy interventions. Third, they can be used to perform a welfare analysis of regional policies, which takes into account their impact on the entire economy.

3 FROM THE REGION TO THE NATION

As mentioned earlier, some of the granularity is regional. Regional economics has increasingly gained importance in many other fields in economics. Different mechanisms explain how regional shocks propagate and gain importance in the aggregate. This section reviews some of the literature that quantifies the impact of some of these mechanisms and discusses their impact on regional and national economic performance.

3.1 Agglomeration and aggregate growth

The notion of agglomeration economies (AE) is quite broad and it encompasses a wide range of factors (Duranton and Puga (2004),). Agglomeration is generally introduced as a shifter of the production function in a Hicks neutral way. Specifically, establishment j's production function is given by $y_i = A_i f(x_i)$, where x_i is vector of inputs, including labor, capital (physical or human), land, and other materials, and $A_i = g(a_i)$ includes all external factors a_i that may affect the establishment's productivity. Models differ on how they specify the function $g(a_i)$. For instance, a_i may include variables that capture the size of the industry, the size of the city, or the skill-composition of the labor force.

A large body of research in urban and regional economics focuses on the impact of AE on city growth (see, for example, Rosenthal and Strange (2004), Combes and Gobillon (2015) for a comprehensive review of the literature). But what is the contribution of local agglomeration economies to aggregate economic growth? The literature that addresses this issue is more limited. Rossi-Hansberg and Wright (2007) develops an economic growth model in the context of cities. In their model, knowledge spillovers drive agglomeration effects. Specifically, the productivity of local firms depends on the total number of workers in the city, and on the total stock of human capital. Davis, Fisher, and Whited (2014) quantify the impact of local agglomeration on aggregate growth. The paper develops a dynamic spatial stochastic general equilibrium growth model. The notion of agglomeration used in their analysis assumes a very simple reduced form, based on previous work by Ciccone and Hall (1996). Specifically, output produced at location j is given by

$$y_i = A_i \ell_b^{(1-\phi)} k_b^{\alpha\phi} n^{(1-\alpha)\phi}, \quad \phi \in [0, 1], \tag{1}$$

where $\ell_{b,j}$ is finished land available for production at location j, $k_{b,j}$ is business capital, n_i is the number of workers, α the factor share, and ϕ represents a congestion effect in production. Total factor productivity (TFP) at location j is

$$A_i = (\tilde{A}_i z_i)^{(1-\alpha)\phi} a_i^{(\lambda-1)/\lambda}, \quad \lambda \geqslant 1, \tag{2}$$

where $\tilde{A}_i$ is a constant representing the aggregate level of technology, z_i is an exogenous productivity shock, and a_i is city j's output density, defined as total production of the local intermediate good per unit of finished land. The effect of a_i on TFP A_i is determined by λ. The quantitative exercise confirms the fact that λ is significantly greater than one, so agglomeration does affect TFP. Moreover, from their estimations it follows that local agglomeration forces increase the growth rate of aggregate consumption per capita in about 10 percent.

3.2 Aggregate effects of regional shocks

Granular shocks may have different aggregate effects depending on which regions and sectors are originally affected. To explain the transmission and propagation of shocks, Caliendo, Parro, Rossi-Hansberg, and Pierre-Daniel Sarte (2017) constructs a quantitatve spatial model which assumes that the sectoral composition differs across regions, regions are endowed with different stocks of immobile factors, and transportation costs across regions are determined by the geography. Using granular manufacturing data for the U.S., they estimate how regional and sectoral productivity shocks affect total factor productivity, GDP and employment at the regional, sectoral, and aggregate levels.

A recent strand of literature focuses on regional fiscal multipliers. This work uses cross-sectional variation as a strategy to identify how fiscal policy may affect the aggregate economy (see Nakamura and Steinsson (2018) for the more general question on identification in macroeonomics). While it is generally understood that the aggregate effect is not simply the sum of the regional effects, this approach has generally been used to assess and compare the effectiveness of different fiscal policies. To account for the aggregate effects of regional policies, it is necessary to develop a general equilibrium model. Nakamura and Steinsson (2014) Chodorow-Reich (2019), and Beraja, Hurst, and Ospina (2019), among several other papers, follow this kind of approach.

Local TFP shocks directly affects cities hit by the shock by increasing local employment, earnings and purchasing power of workers. However, as the spatial allocation of factors of production changes in response to the shock, other localities will be indirectly affected. The work by Hornbeck and Moretti (2019) quantifies the importance of the direct and indirect effects of local manufacturing TFP shocks. At the local level, they find that homeowners benefit from the positive shock. However, for renters, the increase in earnings is almost perfectly matched by an increase in the local cost of living. Their work also shows that local inequality declines after the local productivity growth in manufacturing, since earnings of local low-skilled workers increase more than those of high-skilled workers. The differential impact on the two groups is partly explained by the fact that low-skill workers tend to move less across space. It follows from their analysis that the indirect effects of local TFP shocks are also important. The local shock may have a small impact on other cities individually, but the aggregate effect is not negligible. In fact, their results

indicate that 38 percent on the increase in workers' purchasing power is explained by what happens in cities not directly affected by the shock.

3.3 Regional trends: Growth convergence (divergence) in the US

Several papers document a recent decline in regional income convergence. The observation holds for regions, states, and cities. The work by Ganong and Shoag (2017), for instance, studies regional convergence of per-capita incomes across US states. The authors document that from 1880 to 1980, incomes across states converged at a rate of 1.8% per year. Since 1980, however, this relationship has weakened dramatically, and during the period 1990-2010, the convergence rate declined to less than half the historical values. There was practically no convergence during the period before the Great Recession.

Using city-level (MSA) data, recent work by Giannone (2017) establishes that during the period 1940 and 1980, the wage gap between poorer and richer U.S. cities declined at an annual rate of roughly 1.4%. The data shows no further regional convergence after 1980, though. To explain for the lack of convergence, Giannone (2017) focuses on the differential behavior of workers with different skills. A closer look at the data reveals that while prior to 1980 the wage convergence for high and low-skilled workers was the same, after 1980 wages did not converge for the high-skilled, but continued to converge at 1.4% annually for low-skill workers.

According to Hsieh and Moretti (2019), the spatial distribution of wages, and consequently, productivity differences across urban areas within the USA has been increasing. The latter suggests that the spatial allocation of labor is inefficient.

The lack of regional convergence is generally attributed to a variety of factors. For instance, the lower income convergence coincides with a decline in migration. There is substantial evidence suggesting that labor mobility in the U.S. is not as high as in the past (Molloy, Trezzi, Smith, and Wozniak (2016), Austin, Glaeser, and Summers (2018)).[2]

However, not only has labor mobility has been declining, it has become less directed toward high-income areas (Austin, Glaeser, and Summers (2018)). Ganong and Shoag (2017) relate the lack of convergence of state income per capita to the observed decline in directed migration. While before 1980, people were migrating from low- to high-income places, this pattern has declined over the last 30 years. Ganong and Shoag (2017) attribute such decline to changes that have been taking place in the housing markets, specifically at in high-income locations. The returns to migration and the resulting migration flows depend in part on housing prices and how they respond the increase in housing demand. While housing prices have always been higher in higher-income states, Ganong and Shoag (2017) claims that that housing supply has become more inelastic in high income locations due to more stringent land-use regulations. Such changes tend to affect disproportionately more low-skill workers than high-skill workers, reducing their incentives to move.

Using Public Use Microdata Areas (PUMAs) data Austin, Glaeser, and Summers (2018) show evidence on a number of stylized facts: (i) declining geographic mobility; (ii) increasingly inelastic housing supplies in high-income areas; (iii) declining income conver-

2 Other factors that may explain the lack of convergence across regions are reviewed in the next section.

gence; (iv) increased sorting by skills across space; and (v) persistent pockets of nonemployment. All this supports the conclusions from the previous papers.

4 WHAT EXPLAINS THE INEFFICIENT SPATIAL ALLOCATION OF RESOURCES?

The spatial misallocation of resources will have an impact on aggregate outcomes. The literature has considered several factors that may prevent poorer regions from integrating with the more prosperous parts of the national economy, and explain the observed inefficient spatial distribution of labor and firms. Some of those factors include spatial spillovers and externalities, land use regulations, and regional taxes.

4.1 Externalities and spatial spillovers

The spatial concentration of economic activity generates spillover effects. Such external effects lead to a sub-optimal spatial allocation of resources. Greenstone, Hornbeck, and Moretti (2010) quantify agglomeration spillovers by estimating the impact of the opening of a large manufacturing plant, the “million dollar plant" (MDP), on the total factor productivity (TFP) of incumbent plants in the same county. They compare incumbent plants in the county where the new plant chose to locate (winning county), with incumbent plants in the runner-up county (losing county). The main finding of their research is that five years after the new MDP opening, TFP of incumbent plants in winning counties is on average 12% higher than the TFP of incumbent plants in losing counties. The estimated productivity gains are, however, very heterogeneous, with some incumbent plants even showing a decline in TFP.

The work by Gaubert (2018) explains productivity differences between cities by disentangling agglomeration forces from firm sorting. The paper develops a quantitative spatial model that examines how much of the productivity advantage of a region is driven by the efficiency of the firms it attracts. The model examines the localization choices of heterogeneous firms in an environment where firm productivity is higher at places with larger populations, and characterizes the optimal spatial allocation. Two main conclusions emerge from the analysis. First, the equilibrium is sub-optimal in the sense that firms locate in cities that are too small. The first-best solution would entail taxing firms' wages in smaller cities and subsidize firms' wages in larger cities. Second, as city size increases in 1%, firm productivity increases by 4.2%, where approximately half of it is attributed to firm sorting (2.3%). The model is next used to perform a number of counterfactual exercises in order to evaluate the effectiveness of certain place-based policies. We will revisit the policy implications of this exercise in Section 6.

Fajgelbaum and Gaubert (2020) also examine the welfare impact of spatially targeted policies. The main difference with Gaubert (2018) is that the model includes spillover effects among heterogeneous workers taking place both through the productive and consumption process. The analysis, which focuses on the U.S. economy, proceeds in three steps. First, it evaluates the observed spatial distribution of economic activity and

compares it to the efficient allocation. Second, it examines the kind of policies (transfers) that would restore efficiency. Third, it characterizes the properties of the optimal spatial distribution and compares to the observed outcome. Specifically, it examines whether an optimal spatial distribution entails stronger or weaker spatial disparities and sorting by skill relative to the observed one. The main conclusion from their analysis is that the U.S. economy shows an excessive concentration of high-skill workers and wage inequality in larger cities relative to efficient outcome.

Rossi-Hansberg, Sarte, and Schwartzman (2019) focuses on the spillover effects among workers in different occupations, performing different types of tasks. Occupations are generally divided into those that require “cognitive" and “non-routine" tasks, or CNR occupations, and the rest, grouped into non-CNR occupations. Moreover, the spatial distribution of occupations is not uniform: Workers in CNR occupations tend to be concentrated in large cities, workers in non-CNR occupations locate in small, generally declining cities. The work by Rossi-Hansberg, Sarte, and Schwartzman (2019) presents evidence showing large productivity spillovers among CNR occupations, but none among non-CNR workers. Motivated by this evidence, the paper evaluates next the ability of certain spatially targeted policies to attain an optimal allocation resources across regions.

4.2 Land-use regulations (LURs)

Land and housing can be costly at certain locations or regions for a number of possible reasons. As described in Section 2.1, local amenities attract population and firms, raising the demand for land, and consequently, land prices in those areas. Prices could also be high if land supply is constrained by the geography. In some other areas, however, land prices are high because of stringent land-use regulations (LURs), such as zoning laws or minimum lot sizes. While the implementation of LURs could theoretically be justified on the basis that they intend to correct for market imperfections, many researchers have questioned their cost-effectiveness, though. Regardless of their merits, the use of LURs has become widespread and their intensity has been steadily increasing. Understanding the impact of LURs is important, but at the same time challenging. Due to the complexity and overlap of a large number of local rules in place, it is not easy to quantify their economic consequences.

Shifts in population from less-productive areas to more-productive ones are desirable since they would increase the overall well-being in a country. LURs, however, make it difficult for local housing markets to respond to growing demand. By reducing land availability and increasing land prices at certain locations, LURs make the process of moving to thriving regions more difficult, beyond the normal costs of changing residential locations. Workers facing these additional hurdles to moving may end up being trapped in less-productive areas. Otherwise productive labor migration is discouraged, generating a sub-optimal distribution of labor across the nation and an excessively large dispersion of wages across regions.

Recent work by Hsieh and Moretti (2019) quantifies the aggregate economic importance of LURs. To the extent that these regulations induce a regional mismatch between workers and regions, they would entail lower aggregate production and welfare. Using data from 220 U.S. metropolitan areas, the paper finds that during the period 1964-2009 LURs

effectively decreased aggregate economic growth in 36 percent. Moreover, the authors state that LURs in exceptionally productive cities (such as New York City, San Francisco, and San Jose), are particularly responsible for curtailing aggregate economic growth in the U.S.

While rules and standards are necessary to generate the best possible urban life, there is always the risk of shifting toward an excessively regulated environment in which the cost of the regulations overshadows their intended objectives. The challenge is, of course, to determine what kind of minimal regulations would be necessary to ensure a pleasant and, at the same time, productive environment without imposing unwarranted costs on both the local and the aggregate economy.

4.3 Regional taxes

Regional taxes may also distort the spatial allocation of resources. In general, an heterogenous regional tax system will generate a distortion that in principle depends on the ability of factors of production to change their locations in response to tax changes. The work by Suárez Serrato and Zidar (2016) quantifies welfare effects of changes in state corporate taxes on workers, firms, and landowners. It is generally believed that corporate taxation in an open economy tends to hurt workers more because companies can be moved to places with lower tax pressure. In other words, firms are more mobile than workers. However, from their analysis it follows that firm owners bear approximately 40% of the tax burden, workers between 30 and 35%, and landowners between 25 and 30 percent.

A series of studies, including Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Desmet and Rossi-Hansberg (2013), Suárez Serrato and Zidar (2016), and Fajgelbaum, Morales, Suárez Serrato, and Zidar (2019) among others, focus on the benefits of tax harmonization. This research agrees with the conclusion that heterogeneity in regional taxes generates aggregate welfare losses. A shift towards a regime with a lower dispersion of regional taxes induces a reallocation of workers and firms and increase welfare. Specifically, the work by Fajgelbaum, Morales, Suárez Serrato, and Zidar (2019) finds that if state harmonize their tax system, aggregate welfare may increase by 0.6% if spending remains constant, and 1.2% if state spending endogenously responds to the tax changes.

5 TRANSPORTATION

The transportation system is a key determinant of the economic performance of regions. Translating the impact of transportation investment or improvements on regional growth is, however, particularly challenging. If, for instance, current transportation investment is driven by expected population growth, then the benefits of the investment will be confounded with other effects. Before reviewing recent work that have dealt with some of these challenges, we first briefly explain how regional economists think about the role of transportation in a spatial setting.

5.1 How do regional economists think about the role of transportation in the economy?

Accessibility, which is in part determined by the transportation system in place, affect the localization decisions of individuals and firms, and how land is used. Accessibility however, depends on where individuals and firms locate. In other words, transportation and land use interact and influence one another. Economists do not consider transportation in isolation but as one of the components of a more complex and interrelated system that includes cities and regions.

The role of transportation in the context a spatial equilibrium model is twofold. First, as mentioned earlier, one of the main reasons for the existence of cities refers to the advantage of carrying out economic activities in close proximity, or the presence of agglomeration economies. Transportation in this context plays a critical role, since lower transportation costs would allow a higher concentration of production, and larger benefits from agglomeration.

Second, local wages and housing prices adjust at every location so that households and firms do not have an incentive to move (as described in Section 2.1). In other words, wages and land prices will reflect regional differences making households and firms indifferent across locations. When choosing where to live, individuals consider a wide range of factors, including job opportunities, housing options, social networks, and commuting costs. Within a city, some people might choose to live far away from jobs, possibly accepting a costlier commute, because they would be compensated, in effect, by other factors such as lower housing costs. As a result, a trade-off between commuting costs and land prices emerges in equilibrium: At locations near employment centers, commuting costs are low and land prices are high; at more distant locations, commuting costs are higher and land prices are lower. The different levels of accessibility are explained, in part, by the quality of the local transportation system.

5.2 The economic importance of transportation

Research in urban transportation has mainly focused on the effects of transportation on job accessibility and local economic conditions. Estimating those effects is challenging, however, precisely because of the interdependence between transportation and land use explained earlier.

Duranton and Turner (2011) explore the relationship between transportation infrastructure and traffic congestion. Specifically, they estimate the effect of increasing highway capacity on congestion. Their main finding is that people actually drive more when the stock of roads in their city increases. In fact, they find a one-for-one relationship between the two. It follows from this research, then, that an increase in the provision of highways would not alleviate congestion. Their explanation of this outcome is that cities with better roads attract more people. The use of the roads would therefore increase until traffic congestion reaches its pre-existing levels.

Duranton and Turner (2012) examine the effect of increasing highway miles on employment growth in American cities; they find that a 10 percent increase in a city's initial

stock of highways caused about a 1.5 percent increase in its employment over a 20-year period.

Between 1950 and 1990, the aggregate population of central cities in the United States declined by 17 percent, despite the fact that population increased by 72 percent in metropolitan statistical areas (MSAs). This process is generally known as suburbanization. Baum-Snow (2007) investigates the extent to which this phenomenon is attributable to the expansion of the highway system, which eventually lead to lower commuting costs. The paper finds a positive relationship between roads and suburbanization. The results show that one additional "ray", or segment, of interstate highway originating from the city center leads to about a 9 percent decline in the central city population. It should be noted, however, that other factors occurring at that time were also inducing residents to move out of downtown areas. Some of these factors include an increase in income, a flight from blight due to crime, the degradation of housing stock, and changes in the school system.

U.S. cities show differing patterns of residential sorting by income. In most U.S. MSAs, the suburbs are of higher income status and the central cities are relatively poor. There are important exceptions, such as Chicago, Philadelphia, and others. To explain this kind of spatial sorting, the literature suggests a variety of different mechanisms. One such explanation focuses on the transportation mode choices made by households with different income levels. The work by Glaeser, Kahn, and Rappaport (2008), for instance, states that transport modes are key for explaining the central location of the poor. The reasons are twofold: First, the larger financial costs associated with owning a car may cause lower-income families to rely on other modes of transportation, such as public transit; and second, public transit is more accessible in central cities than in suburbs.

A different line of research that also focuses on job accessibility is related to the spatial mismatch hypothesis. The spatial mismatch hypothesis pioneered in 1968 by John Kain intends to explain an apparent spatial disconnection between jobs and workers' locations. The suburbanization of jobs observed during the 1960s and 1970s hurt the labor market prospects of minorities. African-American populations, largely concentrated in central urban areas, were unable to relocate closer to the jobs for different reasons. As a result, they end up experiencing either excessive commuting costs or higher and persistent unemployment levels. In Kain's view, the inability of minorities to move and follow the jobs to the suburbs was mainly due to racial discrimination in the suburban housing market.

The spatial mismatch hypothesis motivated a large body of research on job accessibility and transportation. This literature has mainly examined how the lack of connection to job opportunities affects individuals' labor market prospects, especially for low-skilled workers and minorities. Research generally confirms the hypothesis. The main findings can be summarized as follows. First, the effect of spatial mismatch is stronger in large central urban areas, where low-skilled minorities tend to live. Jobs are generally located far away from central areas, and minorities face geographical barriers that prevent them from finding and keeping jobs. Second, the research establishes that better job accessibility significantly decreases the duration of joblessness among lower-paid displaced workers, the result being strongest for non-Hispanic, African-Americans, females, and older workers.

A corollary of these findings is that improving spatial access to jobs would lead to better labor market outcomes. Investing in transportation infrastructure and improving transportation services (increasing frequency, capacity, and so on) would increase connectivity between high-unemployment neighborhoods and locations with an abundance of jobs and help alleviate the negative consequences of the spatial mismatch.

5.3 Aggregate economic impact of transportation improvements

It is important, however, to distinguish the impact transportation has on economic activity and growth, from its effects on the spatial reorganization of existing activity (see Redding and Turner (2015) for a review of this discussion). As stated earlier, accessibility and localization choices are jointly determined in a spatial equilibrium model. Improvements in transportation infrastructure may simply induce a relocation of resources from one place to another and generate only localized benefits, if any.

Some recent work attempts to quantify the aggregate impact of investment in transportation. This work accounts for all the direct and indirect effects generated and propagated throughout the entire transportation network. The paper by Donaldson and Hornbeck (2016) focuses on the impact on the agricultural sector of expanding the U.S. railroad system in 1890. Their approach assumes that the expansion of the network increases market access, and its benefits will be capitalized into land values. The authors find that, after accounting for the general equilibrium effects, the removal of the railroad system would have entailed a 60 percent decrease in agricultural land values.

Fajgelbaum and Schaal (2017) study the properties of an optimal transportation network. They use their framework to study the welfare loss of road missallocation in Europe. They find that, on average, such missallocation reduces real consumption in two percent. Allen and Arkolakis (2019) develop a spatial general equilibrium model to evaluate how an improvement of the U.S. Highway System would affect aggregate welfare. They find that adding 10 lane-miles would generate large and heterogeneous effects across different highway segments ranging between $10 and $20 million for most of them. Moreover, the benefits are higher than construction and maintenance costs for all segments.

6 POLICY IMPLICATIONS

In light of the importance of the spatial distribution of resources in explaining aggregate economic fluctuations, and given the role of frictions and spatial externalities in determining where individuals and firms locate, is there a rationale, for policy interventions? What is the nature of these policies? To what extent to resources and policies be regionally targeted?

In the presence of agglomeration externalities, attracting firms may generate external productivity benefits for existing firms (as shown by Greenstone, Hornbeck, and Moretti (2010)). So, are spatially targeted publicly-financed subsidies to attract new firms efficiency-enhancing? From the locality's perspective, the subsidies may by designed to internalize externalities and increase efficiency. From an aggregate point of view, overall efficiency

gains depend on whether the benefits for the receiving location of attracting news firms are similar everywhere else.

The quantitative model developed by Gaubert (2018) is used to evaluate the aggregate impact of two types of place-based policies: a subsidy targeted to firms that locate in the smallest (and also less productive) cities of the country (local tax incentives); and the relaxation of local land-use regulation. The paper concludes that a subsidy targeted to smaller localities will benefit the local area, but will decrease aggregate TFP. In other words, subsidizing firms to locate in smaller cities may not be welfare enhancing. However, consistent with other recent research (for example, Hsieh and Moretti (2019)), an overall increase in the housing supply elasticity (through the relaxation of LURs), would lead to an aggregate increase in TFP.

A number of policy implications also emerge from the analysis performed by Fajgelbaum and Gaubert (2020) mentioned earlier. First, skill heterogeneity along with the presence of spillovers across different types of workers could justify the implementation of place-specific labor subsidies for each labor type. Second, the paper documents that the U.S. economy is characterized by an excessive concentration of high-skill workers and wage inequality in larger cities compared to the efficient allocation. Spatial efficiency would require a redistribution towards low-wage cities and weaker sorting by skill relative to the observed data, leading to lower wage inequality in larger cities. This policy would generate a greater mixing of high and low skill workers in low-wage cities, and generate large welfare gains.

7 DIRECTIONS FOR FUTURE RESEARCH

The literature offers different explanations on the possible channels through which granular shocks propagate, amplify, and affect the aggregate economy in non-trivial ways. Evidence seem to support some of those explanations. The previous discussion, however, reveals several important areas in the literature that deserve further study and that will drive much of the future research work on the aggregate effects of regional shocks. We focus below on four different areas of interest.

- Transportation investment: reorganization of economic activity vs. economic growth.

  Research has found that improvements in transportation system may increase local employment and local economic growth. But are the positive economic effects explained by the spatial reallocation of economic activity, at the expense of other regions, or do the transportation improvements increase overall productivity? As mentioned earlier, quantifying and distinguishing among the two effects is key in order to evaluate the aggregate implications of investing in transportation infrastructure.

- General equilibrium effects of agglomeration economies.

  Understanding the general equilibrium implications of agglomeration economies is still pending. Several of the policy interventions described earlier rely on the assumption of the existence of such external forces at the local level. While a wide range of papers show that agglomeration economies do play an important role in

explaining regional growth, a precise quantification of the aggregate effects of these forces is still necessary.

- General equilibrium effects of changes in housing policy.

  Another area that has received a lot of attention at the micro level is housing policy. However, the general equilibrium and aggregate implications of changes in the housing policy (see, for example, Davis and Van Nieuwerburgh (2015)), such as the elimination of mortgage interest tax deductions, or incentives for the development of affordable housing, are not completely understood.

- Localization decisions by both households and firms.

  The original Roback model described earlier in Section 2.1 highlights the importance of combining the simultaneous decisions of households (or workers) and firms to explain variations in prices across space. However, most of the recent quantitative models account for either workers' or firms' localization decisions. In order to understand historical trends and changes in regional economic activity, the quantitative analysis should consider both household and firm mobility.

REFERENCES

Acemoglu, Daron et al. (2012). "The network origins of aggregate fluctuations". In: *Econometrica* 80.5, pp. 1977–2016.

Allen, Treb and Costas Arkolakis (2019). *The welfare effects of transportation infrastructure improvements*. Tech. rep. National Bureau of Economic Research.

Austin, Benjamin, Edward Glaeser, and Lawrence H Summers (2018). "Saving the heartland: Place-based policies in 21st century America". In: *Brookings Papers on Economic Activity* 8.

Baum-Snow, Nathaniel (2007). "Did highways cause suburbanization?" In: *The quarterly journal of economics* 122.2, pp. 775–805.

Behrens, Kristian and Frédéric Robert-Nicoud (2015). "Agglomeration theory with heterogeneous agents". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 171–245.

Beraja, Martin, Erik Hurst, and Juan Ospina (2019). "The aggregate implications of regional business cycles". In: *Econometrica* 87.6, pp. 1789–1833.

Caliendo, Lorenzo et al. (2017). "The impact of regional and sectoral productivity changes on the US economy". In: *The Review of Economic Studies* 85.4, pp. 2042–2096.

Chodorow-Reich, Gabriel (2019). "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" In: *American Economic Journal: Economic Policy* 11.2, pp. 1–34.

Ciccone, Antonio and Robert E. Hall (1996). "Productivity and the Density of Economic Activity". In: *The American Economic Review* 86.1, pp. 54–70.

Combes, Pierre-Philippe and Laurent Gobillon (2015). "The empirics of agglomeration economies". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 247–348.

Costinot, Arnaud and Andrés Rodríguez-Clare (2014). "Trade theory with numbers: Quantifying the consequences of globalization". In: *Handbook of international economics*. Vol. 4. Elsevier, pp. 197–261.

Davis, Morris A, Jonas DM Fisher, and Toni M Whited (2014). "Macroeconomic implications of agglomeration". In: *Econometrica* 82.2, pp. 731–764.

Davis, Morris A and Stijn Van Nieuwerburgh (2015). "Housing, finance, and the macroeconomy". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 753–811.

Desmet, Klaus and Esteban Rossi-Hansberg (2013). "Urban accounting and welfare". In: *American Economic Review* 103.6, pp. 2296–2327.

Donaldson, Dave and Richard Hornbeck (2016). "Railroads and American economic growth: A "market access" approach". In: *The Quarterly Journal of Economics* 131.2, pp. 799–858.

Duranton, Gilles and Diego Puga (2004). "Chapter 48: Micro-foundations of Urban Agglomeration Economies". In: *Handbook of Regional and Urban Economics*. Ed. by J. Vernon Henderson and Jacques-Francois Thisse. Vol. 4. Elsevier, pp. 2063–2117.

Duranton, Gilles and Matthew A Turner (2011). "The fundamental law of road congestion: Evidence from US cities". In: *American Economic Review* 101.6, pp. 2616–52.

– (2012). "Urban growth and transportation". In: *Review of Economic Studies* 79.4, pp. 1407–1440.

Fajgelbaum, Pablo D and Cecile Gaubert (2020). "Optimal Spatial Policies, Geography, and Sorting". In: *The Quarterly Journal of Economics* 135.2, pp. 959–1036.

Fajgelbaum, Pablo D et al. (2019). "State taxes and spatial misallocation". In: *The Review of Economic Studies* 86.1, pp. 333–376.

Fajgelbaum, Pablo D and Edouard Schaal (2017). *Optimal transport networks in spatial equilibrium*. Tech. rep. National Bureau of Economic Research.

Gabaix, Xavier (2011). "The Granular Origins of Aggregate Fluctuations". In: *Econometrica* 79.3, pp. 733–772.

Ganong, Peter and Daniel Shoag (2017). "Why has regional income convergence in the U.S. declined?" In: *Journal of Urban Economics* 102, pp. 76 –90.

Gaubert, Cecile (2018). "Firm sorting and agglomeration". In: *American Economic Review* 108.11, pp. 3117–53.

Giannone, Elisa (2017). "Skilled-biased technical change and regional convergence". In:

"Structural estimation in urban economics" (2015). In: *Handbook of Regional and Urban Economics*. Ed. by J. Vernon Henderson Gilles Duranton and William C. Strange. Vol. 5. Elsevier. Chap. 2, pp. 69 –114.

Giroud, Xavier and Holger M. Mueller (Oct. 2019). "Firms' Internal Networks and Local Economic Shocks". In: *American Economic Review* 109.10, pp. 3617–49.

Glaeser, Edward L, Matthew E Kahn, and Jordan Rappaport (2008). "Why do the poor live in cities? The role of public transportation". In: *Journal of Urban Economics* 63, pp. 1–24.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings". In: *Journal of Political Economy* 118.3, pp. 536–598.

Hornbeck, Richard and Enrico Moretti (2019). "Estimating Who Benefits from Productivity Growth: Direct and Indirect Effects of City Manufacturing TFP Growth on Wages, Rents, and Inequality". In:

Hsieh, Chang-Tai and Peter J Klenow (2009). "Misallocation and manufacturing TFP in China and India". In: *The Quarterly journal of economics* 124.4, pp. 1403–1448.

Hsieh, Chang-Tai and Enrico Moretti (2019). "Housing Constraints and Spatial Misallocation". In: *American Economic Journal: Macroeconomics* 11.2, pp. 1–39.

Molloy, Raven et al. (2016). "Understanding declining fluidity in the US labor market". In: *Brookings Papers on Economic Activity* 2016.1, pp. 183–259.

Nakamura, Emi and Jón Steinsson (2014). "Fiscal Stimulus in a Monetary Union: Evidence from US Regions". In: *The American Economic Review* 104.3, pp. 753–792.

– (2018). "Identification in Macroeconomics". In: *Journal of Economic Perspectives* 32.3, pp. 59–86.

Redding, Stephen J. and Esteban Rossi-Hansberg (2017). "Quantitative Spatial Economics". In: *Annual Review of Economics* 9.1, pp. 21–58. eprint: https://doi.org/10.1146/annurev-economics-063016-103713.

Redding, Stephen J and Matthew A Turner (2015). "Transportation costs and the spatial organization of economic activity". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 1339–1398.

Restuccia, Diego and Richard Rogerson (2008). "Policy distortions and aggregate productivity with heterogeneous establishments". In: *Review of Economic dynamics* 11.4, pp. 707–720.

Roback, Jennifer (1982). "Wages, rents, and the quality of life". In: *Journal of political Economy* 90.6, pp. 1257–1278.

Rosenthal, Stuart S and William C Strange (2004). "Evidence on the nature and sources of agglomeration economies". In: *Handbook of regional and urban economics*. Vol. 4. Elsevier, pp. 2119–2171.

Rossi-Hansberg, Esteban (2005). "A spatial theory of trade". In: *American Economic Review* 95.5, pp. 1464–1491.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman (2019). *Cognitive Hubs and Spatial Redistribution*. Tech. rep. National Bureau of Economic Research.

Rossi-Hansberg, Esteban and Mark LJ Wright (2007). "Urban structure and growth". In: *The Review of Economic Studies* 74.2, pp. 597–624.

Starrett, David (1978). "Market allocations of location choice in a model with free mobility". In: *Journal of Economic Theory* 17.1, pp. 21–37.

Suárez Serrato, Juan Carlos and Owen Zidar (2016). "Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms". In: *American Economic Review* 106.9, pp. 2582–2624.