

# VU Research Portal

## Agglomeration and Human Capital Spillovers

Verstraten, P.W.A.H.

2019

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Verstraten, P. W. A. H. (2019). *Agglomeration and Human Capital Spillovers: Identification and Spatial Scope*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# Chapter 1

## Introduction

Neoclassical growth theories state that long-run economic growth rates are determined by population growth and technological progress. The latter component of economic growth has long been unexplained, and has therefore been considered to be an exogenous factor. During the second half of the previous century, economists attempted to do a better job in explaining long-run economic growth, which has led to the development of endogenous growth theories. The studies of Arrow (1962) and Lucas (1988) were pivotal in this respect. In explaining technological change, they stressed the importance of human capital accumulation, for instance due to schooling, learning by doing and the sharing of skills and knowledge. Lucas also noticed that spatial proximity to other individuals is indispensable for these mechanisms to fulfill their full potential. This is corroborated by the study of Jacobs (1969), who underlined the importance of cities for generating creative ideas. More recently, Desmet and Rossi-Hansberg (2014) showed that some important (spatial) economic developments can be explained by incorporating space into endogenous growth theory. Hence, when trying to explain economic growth, it is important to understand how economic activity is organized across geographic space.

### 1.1 The economics of cities

Distance plays a central role in urban economics. Von Thünen (1826) was a pioneer in this respect, as he was the first to analyze the role of distance in shaping land use

around a single town. To this end, he imagined an *isolated state*, consisting of a large town located at the center of a fertile plain without any objects that can be considered either an obstacle or aid for transportation. In reality, the concept of distance is of course much more complex. Mountains and (navigable) rivers, for instance, exert a significant impact on effective distance, as opposed to physical distance. Furthermore, it is evident that transportation costs can be reduced by the construction of transport infrastructure, such as roads and canals. The large town at the center of von Thünen's plain also raises the question as to why production is concentrated in one single town, rather than dispersed across space. What centripetal forces prevent a city from falling apart due to, for instance, traffic congestion and pollution?

The potential of cities to efficiently organize a society was already understood by the ancient Greeks. Plato, for instance, argued that the optimal city size would consist of 5,040 citizens, households or land pieces. A useful property of the number 5,040 is that it has many divisors, which makes it an attractive number to subdivide a city into smaller parts. This would ease tax collection, as well as preparations for warfare. Jowett (1871, p. 112) even claimed that "Plato ... really seems to have supposed that the well-being of the city depended almost as much on the number 5,040 as on justice and moderation." The argument made by Plato is related to one of the most apparent benefits of urban areas: sharing indivisible facilities. It is, however, difficult to justify the existence of today's most populous cities on the basis of indivisibilities alone. Economists have therefore proposed other mechanisms that give rise to increasing returns to urban scale.

Smith (1776) was the first modern economist to provide an economically motivated rationale for the existence of cities. With his famous pin factory example, he demonstrated the gains from labor specialization. The main idea of this argument is that a society becomes more productive when individuals trade their specialized skills instead of aiming to master all crafts. Somewhat later, von Thünen (1826) extended Smith's understanding of the urban economy by noting that large cities also offer advantages related to central governance and the provision of social and cultural amenities. The study of Marshall (1890), however, has been even more influential in urban economics. Marshall, whose insights are still relevant today, distinguished

between three kinds of agglomeration benefits: a constant supply of skilled labor, a wide variety of specialized intermediate inputs, and knowledge spillovers.

The theoretical underpinnings of the agglomeration spillovers have been further developed over the course of the 20<sup>th</sup> century. The most widely cited source on this topic is Duranton and Puga (2004), who significantly improve on Marshall's understanding of the urban economy by building theoretical models for the various mechanisms of agglomeration economies. They distinguish between three categories of agglomeration spillovers based on sharing, matching and learning mechanisms. The first mechanism includes the gains from sharing indivisible facilities and labor specialization. The sharing mechanism is also related to the benefits from sharing a large labor market, which makes it easier for firms to deal with idiosyncratic shocks, and sharing a large availability of specialized intermediate goods, which fosters productivity when firms have love of variety preferences. The matching spillover arises because large labor markets improve the expected quality of matches between employers and employees, allowing firms to save on costs for hiring and training. Finally, the learning mechanisms predict that cities are beneficial for the development of new cutting-edge technologies, knowledge diffusion and knowledge accumulation.

The learning mechanisms also act as a bridge between the literatures on agglomeration and human capital spillovers. The former category hypothesizes that knowledge spillovers thrive in dense urban environments (e.g., Jacobs, 1969; Eaton and Eckstein, 1997; Glaeser, 1999), while the latter emphasizes the importance of a well-educated workforce (e.g., Lucas, 1988). Knowledge spillovers are, however, not the only mechanism that causes the social return of education to exceed the private return. Education spillovers may also arise because skill and physical capital are complements in production. In such a case, the privately optimal amount of investment in physical capital depends on the expected quality of the future workforce, and the optimal investment in education increases in the anticipated amount of physical capital. As a consequence, an increase in the human capital stock gives rise to productivity spillovers because search frictions in the labor market make

that all workers – also those that did not increase their education level – produce more capital intensively (Acemoglu, 1996).<sup>1</sup>

## 1.2 The empirical literature

Weber (1899) – who reported an urban-rural wage differential of 50 percent in Germany – was among the first to provide empirical evidence for the existence of spatial wage disparities. Most of the empirical literature on this topic, however, has been developed only very recently. In particular, the increasing availability of microdata on individual earnings over the past few decades has been a huge boon for empirical research in labor and urban economics. Heuermann et al. (2010) observe that micro-evidence on agglomeration and human capital spillovers is divided into two separate strands. Studies on agglomeration spillovers predominantly focus on unraveling the urban wage premium, while the literature on human capital spillovers concentrates on estimating the external return to education. They also note that both literatures, despite their strong interrelations, seem to develop quite independent from each other. In what follows, I briefly discuss these two empirical literatures.<sup>2</sup>

### 1.2.1 Agglomeration spillovers and the urban wage premium

The urban wage premium is widely regarded as one of the most evident indicators for the existence of agglomeration spillovers. After all, with perfectly competitive firms we expect workers in tradable sectors to receive a wage equal to the marginal productivity of labor. Although not all workers are employed in tradable sectors, Moretti (2011) notes that places with higher nominal wages must be more productive on average, as long as traded goods are produced everywhere and workers are able to move between the traded and non-traded sectors.

One of the main empirical challenges is to identify which part of the urban wage premium stems from agglomeration spillovers and which part is merely the result of spatial differences in the quality of labor. The fact that high-skilled workers self-select themselves into large cities was already recognized by Marshall (1890, p.

---

<sup>1</sup> Non-market spillovers of education are, for instance, the negative effect on crime rates (e.g., Lochner, 2011) and enhanced democratic political participation (e.g., Friedman, 1962).

<sup>2</sup> While this review is not exhaustive, it provides an overview of the empirical studies and challenges that are most relevant for this thesis.

199), who wrote that “the large towns, and especially London, absorb the very best blood from all the rest of England; the most enterprising, the most highly gifted, those with the highest physique and the strongest characters go there to find scope for their abilities.” Others have rationalized spatial sorting on the basis of increased returns to education in cities (Costa and Kahn, 2000) and the availability of urban amenities (Van Duijn and Rouwendal, 2013).

Attempts to identify and quantify agglomeration spillovers heavily rely on microdata because it enables the econometrician to clean the individual wages from (un)observed characteristics of workers. To this end, the empirical literature has used either cross-sectional wage data to account for observed worker characteristics (e.g., Wheeler, 2001; Groot et al., 2014) or a wage panel to follow workers over time and thereby also account for time-invariant (unobserved) worker characteristics (e.g., Combes et al., 2008; Mion and Naticchioni, 2009). The former category of studies generally reports higher estimates of the wage-agglomeration elasticity, possibly because they run the risk of omitting some important worker characteristics.

Another source of concern is related to endogeneity in the wage-agglomeration relationship. For instance, the direction of causality may not run from city size to productivity, but in the opposite direction. Also, spatial wage disparities may be the result of local endowments or variations in capital intensity, for which it is difficult to adequately control. Researchers often aim to address this issue by employing historical instruments, such as long-lagged population levels, or geological instruments, such as mean land elevation, landslide hazard or the presence of sedimentary rock. However, Combes et al. (2010a), who have used both types of instruments, conclude that the endogeneity problem is relatively small. Other studies also find that simultaneity and omitted variables are only a second-order identification problem (e.g., Groot et al., 2014; De la Roca and Puga, 2017).

The literature on the urban wage premium also addresses the temporal scope of agglomeration economies by making a distinction between static and dynamic agglomeration spillovers (Rosenthal and Strange, 2004). The static agglomeration spillover is constrained in time, meaning that it only affects workers who are currently employed in a city, regardless of the worker’s employment history. This static spillover results in an urban wage-level premium, which is lost to the worker

when relocating to a more sparsely populated area. Dynamic agglomeration spillovers, on the other hand, accrue to workers over time and can (partially) be transferred to other areas. It is therefore speculated that the dynamic agglomeration spillover highlights the importance of cities in facilitating matching and learning.

Glaeser and Maré (2001) were the first to identify an urban wage-growth premium, and this finding was confirmed somewhat later by Wheeler (2006) and Yankow (2006). De la Roca and Puga (2017) emphasize the economic significance of rapid urban wage growth by showing that it can fully account for the wage gap that was traditionally thought to be the outcome of spatial sorting on unobserved worker characteristics. The dispute about the importance of spatial sorting is, however, not yet fully settled. In particular, D'Costa and Overman (2014) find that spatial sorting is an important source underlying the urban wage-growth premium.

It seems evident that agglomeration economies do not stop at administrative boundaries, but may stretch far beyond that. However, despite this intuitive logic, it has not been extensively studied. Fujita and Mori (2005) and Brakman et al. (2009) argue that spatial units are often considered to be 'floating islands' in urban economics. In order to shed light on the spatial scope of agglomeration spillovers, Rosenthal and Strange (2003) have proposed a concentric ring-based estimation strategy. This approach involves the calculation of agglomeration or human capital measures at various distance intervals (e.g., 0–5 kilometer, 5–10 kilometer, etc.), which can then be used to explain local economic phenomena. There are only a few studies that combine wage data with a concentric ring approach to estimate the spatial scope of agglomeration spillovers. Rice et al. (2006) and Rosenthal and Strange (2008) find that agglomeration spillovers can stretch across large distances: 80 minutes travel time and 80 kilometer straight line distance, respectively. A more narrow spatial scope of 12 kilometer is identified by Di Addario and Patacchini (2008). A common finding is that agglomeration spillovers attenuate rapidly with distance.

### *1.2.2 Human capital spillovers and the external return to education*

The private return to an additional year of education, which is estimated to be around 5–10 percent in terms of individual earnings (Card, 1999), falls short in rationalizing the strong relationship between cross-country GDP and schooling levels (Acemoglu and Angrist, 2000). Hence, economists have argued that there must exist substantial external returns to education. In order to identify these human capital spillovers, empirical studies have analyzed the relationship between productivity, as reflected by individual wages, and the regional human capital stock, which is proxied by schooling levels. The main objective of these studies is to evaluate whether the social return to education exceeds the private return.

Similar to the literature on the urban wage premium, this type of study is also worried about spatial sorting. Hence, to control for this, the studies have also used observed worker characteristics (e.g., Rauch, 1993; Acemoglu and Angrist, 2000) or worker fixed effects (e.g., Moretti, 2004a; Heuermann, 2011). These individual-specific control variables also enable these studies to distinguish between the private and social return to education, which is a necessary prerequisite to correctly identify the external return to education.

The problem of endogeneity plays a key role in this strand of the literature. For instance, the results could be driven by reverse causality, as high regional productivity levels may increase the educational attainment of its residents. Finding valid instruments for the average schooling level in a region, however, has proven to be very difficult. Almost all studies on the external return to education rely on policy interventions that have led to an exogenous variation in schooling levels across space. Acemoglu and Angrist (2000), for instance, used historic differences in state compulsory attendance laws and child labor laws, while Moretti (2004a) employed spatial variation in the presence of a land grant college. Muravyev (2008) followed a novel identification strategy by exploiting the abrupt end of communism in Russia. It is, however, difficult to replicate these studies in other settings due to the unique context of the instruments. Also, it remains unclear whether these policy interventions are truly orthogonal to productivity levels.

There is also disagreement about which type of education gives rise to productivity spillovers. Rauch (1993) and Acemoglu and Angrist (2000), for instance,

have used the average years of schooling, which is based on the complete education distribution. Krueger and Lindahl (2001), however, claim that productivity spillovers are more likely to stem from highly educated workers, which led other studies to use the regional share of highly educated workers as a proxy for the human capital stock (e.g., Moretti, 2004a; Heuermann, 2011). Yet, using the share of highly educated workers is not free of complications either because low and highly educated workers are thought to be imperfect substitutes in production. This implies that an increase in the share of highly educated workers will increase the productivity of low-educated workers, even in the absence of a spillover effect.

On balance, it can be concluded that the empirical literature on the external return to education is plagued by problems regarding endogeneity and imperfect substitution across workers. This may also explain why these studies have reported so many mixed results. Positive evidence was found by, for instance, Rauch (1993), Moretti (2004a) and Muravyev (2008), while other studies, such as Acemoglu and Angrist (2000), Rudd (2000) and Groot and De Groot (2014), found no evidence for the existence of human capital spillovers. Finally, it should be emphasized that all micro-based studies on this topic have focused on static human capital spillovers. Micro-evidence on the existence of dynamic human capital spillovers does, to my knowledge, not exist.

Empirical evidence indicates that human capital spillovers have a smaller spatial scope compared to the agglomeration spillovers. Fu (2007) and Rosenthal and Strange (2008), who employ a concentric ring-based approach to estimate the spatial scope of human capital spillovers, find a spatial extent of three and 40 kilometer, respectively, with a sharp distance decay effect after one and eight kilometer, respectively. The relatively small spatial scope of human capital spillovers is in line with the idea that the transmission of tacit knowledge is unlikely to stretch across large distances.

### **1.3 Contribution of this thesis**

This thesis makes some significant contributions to the empirical literature on agglomeration and human capital spillovers. The key asset of this research is the availability of a detailed wage panel that allows me to trace the earnings profiles of

workers over time, and thereby account for unobserved worker characteristics, such as intelligence and ambition. This aspect of the dataset is crucial for the proper identification of agglomeration and human capital spillovers because it enables me to deal with one of the most pressing endogeneity concerns in urban economics: non-random spatial sorting of labor. Any remaining endogeneity issues in the wage-agglomeration relationship are accounted for by using an extensive set of historical and geological instruments: pre-industrial population counts in 1840, geographic distance to railway stations in 1870, geographic distance to ancient Roman forts and the percentage of the area that has been drained since 1840.

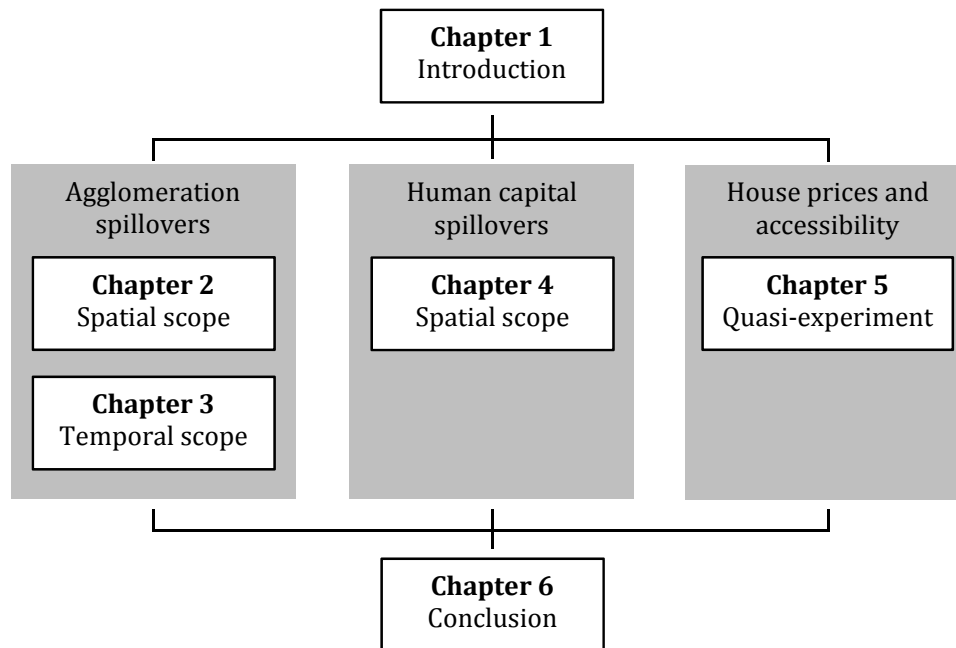
The empirical identification strategies in this thesis are built on two influential papers in urban economics. First, I follow the study of Combes et al. (2008) by adopting a two-stage estimation procedure to estimate a Mincerian wage equation. The important benefit of this estimation approach is that it offers an elegant solution to the problem of non-independent disturbances within spatial units. These disturbances may arise because individuals influence each other or because they are subject to the same local economic shocks. Ignoring this potential problem is detrimental for determining the statistical significance of the results, as it will generally lead to underestimated standard errors and, therefore, overestimated significance levels.

Second, in line with the study of Rosenthal and Strange (2003), this thesis makes use of Geographic Information System (GIS) tools to construct concentric ring variables. These variables, which measure employment and education levels at various distance intervals, allow me to estimate the spatial scope of agglomeration and human capital spillovers. This thesis stands apart from previous attempts to identify the spatial scope of agglomeration spillovers because the wage data contain an exceptionally high level of geographic detail: four-digit postal codes with a mean area of only nine km<sup>2</sup>. As a result, I am able to estimate a more detailed spatial decay effect compared to earlier studies.

Apart from the introduction and conclusion, this thesis consists of four self-contained empirical studies. Figure 1.1 presents a schematic outline of the thesis. The grey shaded boxes indicate the strand of the literature to which each study belongs. In

the remainder of this introduction, I will briefly discuss the content of the four empirical studies.

*Figure 1.1. Schematic outline of the thesis*



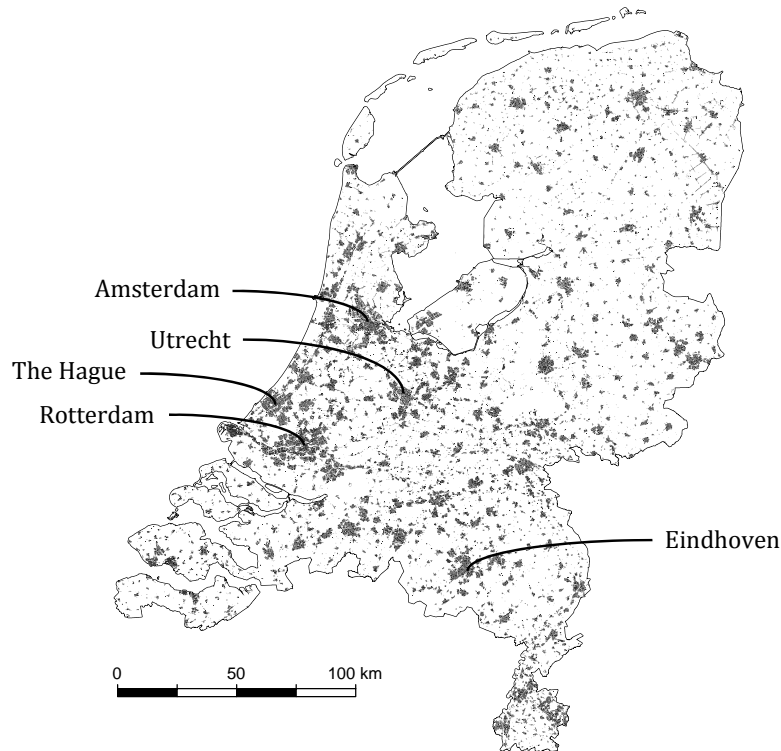
### **Geographical setting of the thesis**

All four empirical studies in this thesis are carried out in the Netherlands. The Dutch economy is characterized by a high average income and relatively low income inequality. Furthermore, the Netherlands is very open towards international trade and, as a founding member of the European Union, it has deep economic ties with other European countries. In 2017, the Netherlands was ranked fourth in the World Economic Forum's Global Competitiveness Index on account of its excellent infrastructure, technological innovations and highly educated workforce.

Around 17 million people live in the Netherlands, on a surface area of only 41,543 km<sup>2</sup>, making it one of the most densely populated countries in the world. The urban geography of the country, as illustrated in Figure 1.2, is characterized by a high degree of polycentricity (Brezzi and Veneri, 2015). The Netherlands mostly

consists of small to medium-sized cities, of which five qualify as a metropolitan area according to the OECD definition. Although the cities are relatively small in size, they are well-connected and situated on short distances from each other. The Netherlands's four largest cities – Amsterdam, Rotterdam, The Hague and Utrecht – are all located in the mid-western part of the country, which is widely seen as a megalopolis called the Randstad. The second-largest urban region in the Netherlands is the Brabantse Stedenrij, containing the country's fifth largest city, Eindhoven.

*Figure 1.2. The urban geography of the Netherlands*



*Notes:* The figure is constructed using the 2012 land use database, provided by Statistics Netherlands (CBS). The shaded areas represent the location of built-up areas in the Netherlands.

Chapter 2 examines the spatial scope of agglomeration spillovers in the Netherlands. The results indicate that the spatial decay effect of agglomeration spillovers is much more complex than is often assumed in urban economics. Wages and agglomeration appear not to be significantly related on short distances (<5 kilometer), whereas they are strongly related on medium distances (5–10 kilometer).

This agglomeration spillover decays rapidly across geographic space, becoming statistically insignificant after 40–80 kilometer. The chapter offers several explanations for this non-monotonic spatial decay effect. The results, however, do not allow for the conclusion that nearby agglomeration is irrelevant for the wage formation because the estimates show that only large cities are able to benefit from agglomeration spillovers on further distances. Finally, the chapter also provides evidence for the existence of substantial border barriers by showing that foreign agglomerations have no significant effect on domestic wages.

The existence of an urban wage-growth premium is a well-established empirical fact, and many economists have rationalized its existence on the basis of learning and matching mechanisms. Chapter 3 quantifies the urban wage-growth premium in the Netherlands and examines whether it is really the result of an agglomeration spillover or merely driven by spatial sorting. To this end, we introduce controls for firm- and individual-specific wage-growth determining characteristics. The estimates show that the size of the urban wage-growth premium is drastically reduced when wage-growth controls are included to the model. Eventually, when controlling for individual-specific returns to experience, the urban wage-growth premium becomes even statistically insignificant. The chapter also examines heterogeneities across workers. Having controlled for wage-growth determining characteristics, it is shown that only young workers experience a significant urban wage-growth premium. This result is intuitive because young workers are generally more receptive to external influences. The urban wage-level premium is relevant to all types of workers, especially the highly educated.

Many studies have attempted to estimate the external return to (higher) education. Most of them employed spatial differences in education levels to identify an effect on individual wages, while others have used variation in schooling levels across economic sectors or firms. Chapter 4 estimates the external return to higher education across all these three work environments simultaneously. Understanding the exact scope at which the higher education spillover operates is important because it helps us to explain what mechanisms drive the productivity spillover. The results indicate that the scope of the higher education spillover is very limited. Most of the spillovers occur at the firm level, while a relatively small portion of the spillovers occurs only on

short distances from the work location and only within the own sector. These findings suggest that higher education spillovers mainly percolate through the exchange of skills and knowledge, as this mechanism is heavily dependent on face-to-face contact.

Finally, Chapter 5 provides relevant insights into the role of transport infrastructure in shaping the attractiveness of regions. To this end, we analyze the relationship between accessibility and house prices. This is an interesting perspective because the other chapters have relied on straight line distances rather than actual travel times. The empirical literature on the relationship between house prices and accessibility is plagued by a wide variety of endogeneity issues. Most importantly, investments in transport infrastructure seldom occur at random, which raises questions regarding reverse causality: did the improvement in accessibility trigger an increase in the attractiveness of the area, or the other way around?

Chapter 5 aims to break the vicious circle in the causal relationship between house prices and accessibility by studying a novel quasi-experiment in the Netherlands: the Westerscheldetunnel. We exploit the fact that the opening of the tunnel caused a major shift in accessibility for people and firms in the connected regions. Also, the location of the tunnel can be considered to be exogenous because it had to be situated at the middle of the Westerschelde estuary in order to allow for the abolishment of the eastern and western ferry services – a necessary prerequisite to finance the tunnel. The results indicate that a one percent improvement of accessibility results into a 0.8 percent increase in the price of housing. The chapter also provides evidence that people anticipated the opening of the tunnel: capitalization of the accessibility gains started more than one year before the opening of the tunnel. Interestingly, we also find that not all regions have benefited equally from the improved transport infrastructure. Resident heterogeneity in terms of the attained education level is the most plausible explanation for the regional differences.