Income Sorting Across Space: The Role of Amenities and Commuting Costs *

Carl Gaigné[†] Hans R.A. Koster[‡] Fabien Moizeau[§]
Jacques-François Thisse[¶]

November 1, 2019

Abstract

We study the sorting of skill/income-heterogeneous consumers within and between cities. We allow for non-homothetic preferences and locations that are differentiated by their accessibility to exogenous amenities and distance to employment centers, where production is subject to local externalities. The residential equilibrium is driven by the properties of an amenity-commuting aggregator obtained from the primitives of the model. Using the model's structure and estimated parameters based on micro-data of the Netherlands, we predict that exogenous amenities are a key driver of spatial sorting. Our general equilibrium counterfactual analysis shows that in the absence of amenities, the GDP increases by 10% because commutes are shorter. However, income segregation rises and 95% of consumers are worse-off.

Keywords: cities, social stratification, income, amenities, commuting

JEL classification: R14, R23, R53, Z13.

*We are grateful to N. Baum-Snow, J. Brinkman, P.-A. Chiappori, P.-P. Combes, M.S. Delgado, R.J.G.M. Florax (†), M. Fujita, J Hansbury, J.V. Henderson, M. Miyake, Y. Murata, H.G. Overman, P. Picard, G. Ponzetto, S. Proost, S. Riou, F. Robert-Nicoud, S.L. Ross, K. Schmidheiny, H. Takatsuka, M. Turner, P. Ushchev, and J.N. van Ommeren for insightful suggestions. We also thank participants to seminars at the LSE, the Higher School of Economics in Moscow, Purdue, the Workshop on Public Policies, Cities and Regions, the Utrecht School of Economics, Laval University, the 2017 Public Economic Theory conference, the summer school "Sustainable Cities", the 64th AFSE Congress, 32d EEA Congress, the Federal Reserve Bank in Philadelphia, and the 7th North American Meeting of the Urban Economics Association, for their comments. The authors acknowledge, respectively, financial supports from the Agence Nationale de la Recherche (ANR-12-INEG-0002), the Institut Universitaire de France, and the Russian Science Foundation under the grant N°18-18-00253.

[†]INRA, UMR1302, SMART-LERECO Rennes (France) and CREATE, Laval University, Quebec (Canada).

[‡]Department of Spatial Economics, Vrije Universiteit Amsterdam (The Netherlands), HSE University (Russian Federation), Tinbergen Institute, and CEPR.

[§]CREM (Condorcet Center), Université de Rennes 1.

[¶]CORE-UCLouvain (Belgium), HSE University (Russian Federation) and CEPR.

1 Introduction

Spatial segregation seems to generate negative and persistent effects on individual development and to threaten social cohesion (Chetty and Hendren, 2018; Bazzi *et al.*, 2019). This is why we find it important to study the various forces that underpin the sorting of heterogeneous households within and between cities. To this end, we develop a full-fledged general equilibrium model that determines the locations of skill-heterogeneous households and shows how their choices pin down the aggregate income. This in turn allows us to study the effects of exogenous amenities – such as natural and historic attributes of locations – on the spatial distribution of households and activities. Our approach tackles this problem from two complementary perspectives. First, we develop a new urban economic model in which skill-heterogeneous households choose where to live and where to work in a polycentric space-economy in the presence of agglomeration economies and heterogeneous commuting behavior. Second, we estimate this model by using various datasets from the Netherlands and undertake a counterfactual that highlights the impact of specific amenities on spatial sorting and the aggregate income.

The canonical monocentric city model leads to a fairly extreme prediction: households are sorted by increasing income order as the distance to the central business district rises (Fujita, 1989). One missing key explanation, at least for cities that have a long history, is the existence of *exogenous amenities*, such as historic buildings and architecture, scenic landscapes, river and sea proximity. That such amenities matter in residential choices has been well documented (Brueckner *et al.*, 1999; Glaeser *et al.*, 2001; Koster and Rouwendal, 2017; Lee and Lin, 2018). Furthermore, there are substantial differences in total factor productivity among employment centers (Hornbeck and Moretti, 2019). However, we do not know well how these forces interact to determine how households distribute themselves across urban areas. This paper attempts to fulfill this gap by proposing a new approach in which locations are distinguished by the distance to employment centers, which have an endogenous total factor productivity, and the accessibility to given and fixed amenities that vary across space.

What are our main contributions? First, we develop a quantitative urban sorting model based on the bid-rent function approach to show that, regardless of the functional form of the amenity, commuting time, and skill distributions, *the spatial sorting of heterogeneous households across a continuum of locations is imperfect.* In other words, a greater geographical distance between households no longer implies a wider income gap. Although our problem has the nature of a matching problem between landlords and households, matching theory cannot be applied here because a household's land consumption varies with both income and location while it is exogenous in matching theory (Chiappori, 2017). This is why we use the bid rent approach in this paper.

Second, we characterize the equilibrium skill mapping under Stone-Geary preferences, which are not homothetic, and a large number of potential employment centers. Homothetic preferences such as the Cobb-Douglas or the CES, which are the specifications most often used in the literature, must be ruled out because they generate multiplicity of equilibria when a continuum of households have different incomes (Gaigné *et al.*, 2017), while failing to capture the fact that housing expenditure shares decline with income (Albouy *et al.*, 2016). We show that the equilibrium skill mapping reflects the behavior of a location-quality index defined over the location set. This index is built from the primitives of the model and its value at any particular location is determined by households' commuting behavior. Furthermore, since the location-quality index is defined over a set that includes locations belonging to different cities, we account for the fact that amenity and productivity differences in urban areas are critical in choosing a residential place. It is worth stressing that our results are not specific to the Stone-Geary preferences. They hold true for other non-homothetic preferences; what changes is the functional form of the location-quality index.

The upshot is that the bliss point is the global maximizer of the location-quality index, thus implying that this location is occupied by the high-skilled because they propose the highest bid. As one moves away from this location along all admissible directions, households are sorted by decreasing skills until a local minimizer of the location-quality index is reached where low-skilled households are located. Around this minimizer, household skill starts rising. As a result, *households get more exposure to and have more contacts with other social groups when the number of turning points of the location-quality index rises.*

Third, production, which operates through employment centers whose size and productivity are endogenous, allows the determination of households' incomes according to their skills and residential choices. To make our setting consistent with real life, we account for the fact that individuals' commuting behavior is gravitational in nature. In line with the empirical literature on agglomeration economies (Combes and Gobillon, 2015), we also recognize that workers' productivity depends on the density of jobs in their working places, which are characterized by an endogenous and specific total factor productivity. The heterogeneity of space implies that workers having the same skills may end up earning different incomes though each skill-group enjoys the same utility level.

Fourth, we show that a spatial equilibrium always exists. It is well known that settings involving agglomeration economies are often plagued with the existence of multiple equilibria (Duranton and Puga, 2004). However, we show that the spatial equilibrium is unique if agglomeration economies are not too strong (Allen and Arkolakis, 2014). From the empirical viewpoint, this assumption does not seem to be an issue because our calculations show that the equilibrium is unique for the estimated values of the elasticity of agglomeration economies.

Fifth, our model is flexible enough to determine analytically the equilibrium outcome when skills and the location-quality index are Fréchet-distributed. We use this location-quality index to predict the equilibrium skill mapping and test these predictions by estimating recursively the main parameters of a discretized version of the theoretical model. We first estimate a commuting gravity equation and, then, the income mapping, which is shown to be directly related to the skill mapping. This provides the necessary information to estimate the elasticity of agglomeration economies.

For the empirical analysis, we use rich Dutch microdata for more than 10 million households covering the years 2010 to 2015 on incomes, residential and job locations at the household level, employment accessibility, as well as land values and amenities at each location. The choice of the Netherlands is motivated by *(i)* the availability of these disaggregated data and *(ii)* the fact that the public services that underpin social cohesion (e.g., education and health) are centrally financed and/or administered (Ritzen *et al.*, 1997). As a result, competition between jurisdictions supplying schools that characterizes many U.S. metropolitan areas is much less of an issue. The Netherlands is one of the countries with the highest population densities in the world (if we disregard city states). Today, with a population density of 407.4 pop/sq km, the Netherlands is almost as dense as the San Francisco Bay area whose area is similar to that of the Netherlands. This is an important feature in settings where density economies matter. It is also one of the richest, with a GDP per capita higher than the UK, Germany and Japan. Moreover, Dutch cities were established long ago and are known to offer a high quality of life, which is at least partly due to the presence of exogenous amenities. Despite being a small country, the Netherlands hosts no less than 8 UNESCO world heritage sites, which is almost as much as London and Paris together, while it hosts $61,908$ listed buildings, which is more than three times the number of listed buildings in Greater London.

To measure the level of exogenous amenities, we use a proxy suggested by Ahlfeldt (2013) and Saiz *et al.* (2018): the number of outside geocoded pictures taken by residents at a certain location. One key advantage of this index is that it lets consumers choose the aesthetic quality of buildings and locations they like best by "voting with their clicks" (Carlino and Saiz, 2019). This allows us to move beyond the approach of defining amenities implicitly, as in Ahlfeldt *et al.* (2015) and Albouy (2016). We show the robustness of our results by using an alternative proxy for amenities based on Lee and Lin (2018).

Admittedly, households also care about the proximity to private facilities such as shops, restaurants and theaters, which may be disproportionately located in upscale neighborhoods where many pictures are taken. In addition, since there is no proxy that perfectly captures the full amenity potential at a certain location, amenities are measured with error. Employment accessibility is also likely to be endogenous due to correlation with unobservable household characteristics and agglomeration economies – the latter being more prevalent in dense areas where commutes are shorter. We address the endogenous nature of amenities and accessibility in our econometric analysis in several ways, e.g., by obtaining Oster's (2019) bias-adjusted estimates and by constructing historic instruments. Since the strategy of using instruments based on historic data raises several issues, we devote considerable attention to the validity of such an identification strategy.

We first report results from reduced-form regressions. The results unambiguously suggest that both amenities and commuting costs are important in determining the spatial income distribution. We find that *doubling the amenity level attracts households whose incomes are* 2.3% *higher*, while *doubling accessibility attracts households whose incomes are* 3.8% *higher.* Hence, the impacts of amenities and accessibility have a similar order of magnitude. We then estimate the structural parameters of the model, which enables us to undertake counterfactual experiments.

Since our main focus is on the effects of exogenous amenities, we consider a scenario where these amenities are set to the lowest value observed in the sample. This mimics many U.S. cities, where households focus only on commuting. As a result, commutes are shorter and the overall output increases by 10.6%. This is because the Dutch high-skilled workers who reside in high amenity places move into the most productive locations. The aggregate real income rises by 7.3%, while the aggregate land rent decreases by 0.6%. Such numbers could suggest that the demand for amenities reduces the overall productivity of the Dutch economy. However, this argument ignores the fact that consumers, despite lower incomes, do value historic and natural amenities since 95% of the households lose utility when amenities are set to their lowest value. Furthermore, a flat distribution of exogenous amenities has strong repercussions for the spatial distribution of skills, hence of incomes: the correlation between the values observed in the data and in the counterfactual is only 0.556. Hence, we may conclude that *amenities are a key determinant of the skill-based sorting of households within and between cities.* In addition, we have constructed a measure of income mixing, i.e., the standard deviation of skills in adjacent neighborhoods, to see how the counterfactual scenario affects income mixing within the Netherlands. More specifically, a uniform amenity distribution implies that income mixing is reduced by approximately two-thirds. Hence, *there is substantially more spatial segregation when amenities are absent.* This confirms the armchair evidence that European, especially Dutch, cities are more socially mixed than American cities. Last, because our experiment is about consumer amenities, we do not observe substantial differences in the spatial distribution of production in a situation where exogenous amenities are absent.

Related literature. Suggesting the complexity of the issue, only a handful of papers in urban economics have studied the social stratification of cities with heterogeneous households. Beckmann (1969) was the first attempt to take into account a continuum of heterogeneous households in the monocentric city. Unfortunately, the assignment approach used by Beckmann was flawed (Montesano, 1972). Recent surveys, such as Duranton and Puga (2015) and Behrens and Robert-Nicoud (2015), highlight the various difficulties associated with the spatial assignment of heterogeneous agents and express some skepticism about the ability of the bid-rent approach to deal with heterogeneous households and a continuum of locations.

Diamond (2016) studied how local wages, urban costs and employment respond to local labor shocks. However, this author focuses on workers' locational choices between cities and disregard workers' residential choices within cities. Using a dynamic setting, Lee and Lin (2018) showed that richer households are anchored in neighborhoods with better natural amenities. We differ from them in at least one fundamental aspect: in their setting people are assumed to work where they live. In our setting, households are free to choose where to live and where to work, while accounting explicitly for commuting costs between the residence and the workplace.

In an important paper, Ahlfeldt *et al.* (2015) highlight the role of amenities, agglomeration economies and commuting in residential location choices in their study of the internal structure of Berlin. Our paper differs from theirs in several fundamental aspects. First, these authors assume an open city model in which the total city population is endogenous while households enjoy the same exogenous utility level. In contrast, we work with a model in which the utility level is endogenous. Second, they assume homogeneous individuals, whereas we consider skill-heterogeneous workers, so that the endogenous utility level varies across skills. Third, Ahlfeldt *et al.* do not provide any properties of the spatial equilibrium. This should not come as a surprise as characterizing the equilibrium outcome is problematic under a finite location set. Indeed, one cannot use the tools of analytical calculus. By contrast, by working with a continuum of locations, we are able to show that residential choices are driven by a location-quality index that allows us to pin down households' location choices.

These differences imply that the estimation procedures used in the two papers also differ along several lines. For example, Ahlfeldt *et al.* (2015) find that the elasticity of amenities with respect to residential density is 0.15, which is quite high. This is so mainly because amenities are measured as 'structural residuals', meaning that it is unclear what these amenities actually capture (e.g., they may capture housing characteristics or sorting on unobserved household characteristics). In our paper, we define amenities explicitly and strive to show that amenities and employment accessibility have a causal and significant impact on the spatial equilibrium. Whereas Ahlfeldt *et al.* obtain structural residuals proxying for amenities and local productivity by using Cobb-Douglas utility and production functions, we use the income mapping to recover preferences for amenities using Stone-Geary preferences and a CES production function. Closer to us, Tsivanidis (2019) considers two skill-groups and a closed-city setting. Like us, he uses Stone-Geary preferences because the observed Engel curves are nonlinear. However, his work differs from ours along several lines. In particular, Tsivanidis focuses on the intra-urban impact of a major transportation infrastructure and does not aim to develop a setting that yields theory results. In this respect, his work remains in the spirit of Ahlfeldt *et al.* (2015).

The remainder of the paper is organized as follows. We provide a detailed description of our model in Section 2. Section 3 characterizes the equilibrium skill mapping for general distribution functions. We also determine analytically the equilibrium skill and income mappings when skills and the location-quality index are Fréchet-distributed. Data are discussed in Section 4. In Section 5, we provide reduced-form evidence of the impact of amenities and accessibility to jobs on sorting by incomes. In Section 6, we outline the procedure to identify the model's parameters and present the results of our counterfactual analysis. Section 7 concludes.

2 The model and preliminary results

2.1 The economy

The economy involves a unit mass of skill-heterogeneous households. A household is characterized by her skill $s \in \mathbb{R}_+$ and is endowed with one unit of s-labor. The skill c.d.f. $F(s)$ is continuously differentiable on $\mathbb{R}_+$ and its density is denoted $f(s)$. Like most many recent contributions in urban economics, we treat the skill distribution as a given. Each household has one unit of time that she divides between commuting and producing. Her allocation of time is determined by the residence and working place she chooses.

The economy involves two normal consumption goods: *(i)* land h, which is a proxy for housing, and *(ii)* a homogeneous final consumption good q. Shipping the final good within the city is costless. Therefore, its price is the same across city locations. This good is used as the numéraire. The land density at each location of the network is 1 while the opportunity cost of land is given by the constant $R_0 \geq 0$.

The map formed by streets, roads, highways, and railway junctions (in a city, region or country) is modeled by means of a topological network. A topological arc, denoted a_z, is the image in $\mathbb{R}^2$ of a compact interval of $\mathbb{R}$ by a continuous one-to-one mapping. Clearly, any arc linking two distinct locations contains a continuum of locations. A *topological network* $N = \cup_{z=1}^{Z} a_z$ is defined as the union of a finite number $Z \geq 1$ of topological arcs. Each arc has a finite length. Furthermore, N is such that for any two points x_1 and x_2 belonging to N there is at least one concatenation of arcs and subarcs of N that links these two points. The distance $d(x_1, x_2)$ between x_1 and x_2 is given by the length of the shortest path that connects these locations. Clearly, $d(\cdot)$ is a metric defined on N. The endpoints of the arcs are called vertices. We assume that these vertices are not colinear, so that (N, d) is *not* a one-dimensional metric space. An example of transportation networks similar to ours can be found in Allen and Arkolakis (2014). In what follows, we assume that all functions are differentiable along each arc of the network N, except may at the vertices.

2.2 Consumption

Households share the same utility function. Since households prefer more amenities than less, we consider a preference structure similar to the one used in models of vertical product differentiation:

$$U(q, h; b) = b \cdot u(q, h),$$

where b denotes the amenity level, q the costlessly traded numéraire, and h the land consumption. Hence, the utility derived from consuming amenities rises with income. Let $b(x) > 0$ be a given function whose value expresses the amenity level (or, equivalently, an aggregator of distinct amenities) available at $x \in N$, which are exogenous and intrinsic to a location. In the featureless city of urban economics, $b(x)$ is constant across locations. In this paper, $b(x)$ varies with x.

We have seen that homothetic preferences must be ruled out to study the impact of skill heterogeneity on residential choices. A well-known example of non-homothetic utility is Stone-Geary's:

$$u(q, h) = q^{1-\mu} \cdot (h - \overline{h})^{\mu}, \tag{1}$$

where $0 < \mu < 1$ and $\overline{h} > 0$ the minimum amount of floor-space in which to live, which is supposed to be sufficiently low for the equilibrium consumption of the numéraire to be positive.

A s-household's residing at x faces the following budget constraint:

$$y(s) = q + R(x)h,$$

where $y(s)$ the income of a s-household and $R(x)$ the land rent at x, which are both specified below. In line with the literature, we assume that the land rent is paid to absentee landlords (Fujita, 1989).

Maximizing (1) with respect to q and h subject to the budget constraint leads to the linear expenditure system:

$$q^*(x, y(s)) = (1 - \mu)[y(s) - R(x)\overline{h}], \tag{2}$$

$$h^*(x, y(s)) = (1 - \mu)\overline{h} + \mu\frac{y(s)}{R(x)}, \tag{3}$$

which shows that the land demand at any location x increases less than proportionally with income.

The corresponding indirect utility $V(R(x), y(s))$ is given by the following expression:

$$V(R(x), y(s)) = (1-\mu)^{1-\mu}\mu^{\mu}b(x)\left[y(s) - R(x)\overline{h}\right] R(x)^{-\mu}.$$

2.3 Production and income

The final sector, which operates under constant returns and perfect competition, supplies the numéraire by combining horizontally differentiated intermediate inputs produced by workers in employment locations. Each worker produces a single intermediate input and each intermediate input is produced by a single worker. The production function of the final sector is given by

$$Y = \left\{ \int_0^1 \left[z\left(\varphi\right)\right]^{(\sigma-1)/\sigma} \mathrm{d}\varphi \right\}^{\sigma/(\sigma-1)},$$

where $z\left(\varphi\right)$ denotes the quantity of input φ while $\sigma > 1$ is the elasticity of substitution between intermediate inputs. Shipping these inputs across space is costless.

There is a given and finite number of employment locations $i = 1, ..., n \in N$. When a s-worker produces the intermediate input φ at i, her output is given by $z\left(\varphi\right) = A_i \ell_i s$ units of input φ, where $A_i > 0$ is the total factor productivity of location i and ℓ_i is the worker's labor time at i. Let $g_i(s) \geq 0$ be the endogenous density of s-households working at i (see (10) for a formal definition). Since the intermediate inputs are shipped at no cost, the total output of the economy is given by a nested CES (Bénabou, 1996):

$$Y = \left\{ \sum_{i=1}^{n} \left[A_i \left[\int_0^{\infty} (\ell_i s)^{(\sigma-1)/\sigma} g_i(s) \mathrm{d}s \right]^{\sigma/(\sigma-1)} \right]^{(\sigma-1)/\sigma} \right\}^{\sigma/(\sigma-1)}. \tag{4}$$

In other words, the aggregate output may be viewed as a CES-sum of what is accomplished in each employment location. In doing so, we account for the direct interdependence between employment centers, each one providing a particular range of intermediate inputs according to its skill composition.

Since we already account for the heterogeneity of the labor force at i in (4), we assume that the total factor productivity A_i of the location i depends only on its size:

$$A_i = \mathbb{A}_i L_i^{\delta} \geq 0, \qquad i = 1, .., n \tag{5}$$

where $\mathbb{A}_i > 0$ is an exogenous location-specific shifter, $L_i \geq 0$ the employment level at i, and $\delta > 0$ the elasticity of agglomeration economies with respect to L_i at location i. We treat the vector $\mathbf{L} \equiv (L_1, ..., L_n)$, with $L_1 \geq 0, ..., L_n \geq 0$ and $\Sigma_{i=1}^{n} L_i = 1$, as given and will determine the equilibrium values $L_1^*, .., L_n^*$ in Proposition 1.

Denote by $\ell_i(x)$ the labor time of a household residing at x and working at i. Since the final sector is competitive, it follows from (4) that the income earned by a s-household residing at x and working at i is given by

$$
\begin{aligned}
A_i \ell_i(x)\, s \frac{\partial Y}{\partial [A_i \ell_i(x)\, s]} &= [A_i \ell_i(x)\, s]^{(\sigma-1)/\sigma} Y^{1/\sigma} \\
&\equiv \omega(s) t_i(x)
\end{aligned}
$$

where

$$
\omega(s) \equiv s^{(\sigma-1)/\sigma} Y^{1/\sigma} \tag{6}
$$

is a s-household's skill-specific component of the household's income, which also depends on the overall productivity Y of the city, while

$$
t_i(x) \equiv [A_i(L_i)\ell_i(x)]^{(\sigma-1)/\sigma}
$$

is the commuting component of the income for any given A_i. For any A_i and A_j, we assume that $t_i(x) = t_j(x)$ for $j \neq i$ has a finite number of solutions.

2.4 Workplace choice

A s-household is characterized by an intrinsic income $\omega(s)$ that depends on her skill s and the total output Y, while her income also depends on her residential location x and workplace i. The s-households located at x have idiosyncratic reasons for working in different employment locations. In line with discrete choice theory, we assume that a s-household's income is random and given by $\omega(s)t_i(x)\nu_{kxi}$, where the ν_{kxi} are i.i.d. shocks on commuting which are specific to the individual k and locations x and i. These shocks capture households' idiosyncrasies.

The effect of uncertainty on location and consumption decisions depends on the timing of uncertainty resolution and on the flexibility that allows a household to revise her decision in response to information. We assume here a timing that endows households with the possibility to adjust their workplace and total consumption conditional upon their residential choices. Before observing their actual income, households choose their residential locations x at the spatial equilibrium associated with the distribution of expected incomes. Once households are located, they are able to observe their actual incomes. The households then choose the workplaces that give them the highest incomes, as well as the corresponding consumption of land and numéraire. Since households are heterogeneous in commuting, those who choose the same residential location x need not earn the same income and consume the same commodity bundle.

Since households anticipate they will choose the best workplace after the resolution of uncertainty, the expected indirect utility of a household at x is defined as follows:

$$\mathbb{E}\left[V(R(x),\omega(s)t_i(x)\nu_i)\right]=\mathbb{E}\left\{(1-\mu)^{1-\mu}\mu^{\mu}b(x)\left[\max_{i=1,..,n}\omega(s)t_i(x)\nu_i-R(x)\overline{h}\right]R(x)^{-\mu}\right\}. \tag{7}$$

When the ν_i are i.i.d. according to a Fréchet c.d.f. $I(z)=\exp(-K_i z^{-\varepsilon})$ where the shape parameter ε is an inverse measure of the dispersion of idiosyncratic tastes, which is assumed to be the same across employment locations, and K_i is the scale parameter of the employment location i. We show in Proposition 1 that the equilibrium outcome is such that the households who reside at x share the same skill $s(x)$. Since the land rent $R(x)$ is given to a s-household located at x, maximizing (7) amounts to maximizing her *expected income* given by

$$\overline{y}(s(x),x)\equiv\omega(s(x))t(x), \tag{8}$$

where

$$t(x)\equiv\mathbb{E}\left[\max_{i=1,..,n}t_i(x)\nu_i\right]=\Gamma\left(\frac{\varepsilon-1}{\varepsilon}\right)\left[\sum_{i=1}^{n}K_i t_i^{\varepsilon}(x)\right]^{1/\varepsilon}$$

and $\Gamma(\cdot)$ is the gamma function (McFadden, 1974). It follows from (8) that the expected income of a s-household located at x is strictly increasing in s. Furthermore, $\overline{y}(s,x)$ also depends on the locational choices made by all types of households through the total factor productivity of the employment locations captured in t_i. As a result, individual incomes are determined at the market outcome.

The probability that a household living at x chooses to work at i is given by the gravity equation:

$$\pi_i(x)=\frac{K_i\left[t_i(x)\right]^{\varepsilon}}{\sum_{j=1}^{n}K_j\left[t_j(x))\right]^{\varepsilon}}>0 \qquad \text{for all } x\in N. \tag{9}$$

Thus, households residing at the same location work in different employment locations.

Because the s-households may be distributed over several residential locations, we denote by $\zeta(x,s)\in[0,1]$ the share of s-households who reside at x. Therefore, we have:

$$g_i(s)=\textstyle\int_N\pi_i(x)\zeta(x,s)f(s)\mathrm{d}x. \tag{10}$$

Using this expression, (4) becomes:

$$Y=\left[\sum_{i=1}^{n}\int_0^{\infty}\int_N\left[A_i\ell_i\left(x\right)s\right]^{(\sigma-1)/\sigma}\pi_i(x)\zeta(x,s)f(s)\mathrm{d}x\mathrm{d}s\right]^{\sigma/(\sigma-1)}. \tag{11}$$

2.5 The spatial equilibrium

Given an amenity function $b(x)$, a given mass of heterogeneous households choose where to live and where to work in the city, how much land and how much of the composite good to consume. The s-households may be distributed over several locations. The land market clearing condition holds if $s(x)$ satisfies the following condition:

$$|\zeta(x,s)f(s)h(x,s)\mathrm{d}s| = \mathrm{d}x. \tag{12}$$

In other words, the amount of land available between any x and $x + \mathrm{d}x > x$ and the area occupied by the households whose skill varies from s to $s + |\mathrm{d}s|$ are the same. Since $s(x)$ need not be monotone, the land market clearing condition is expressed in absolute value.

A *spatial equilibrium* is defined by the following vector:

$$(s^*(x), \zeta^*(x, s^*(x)), Y^*, R^*(x), h^*(x, s^*(x)), q^*(x, s^*(x)), L_1^*, ..., L_n^*)$$

with $x \in N$, which is such that

$$b(x) \cdot u[q^*(x, s^*(x)), h^*(x, s^*(x))] \geq b(y) \cdot u[q^*(y, s^*(x)), h^*(y, s^*(x))] \qquad \text{for all } x \in N$$

holds under the budget constraints, the population constraint and (12).

If the inequality is strict for all $y \neq x$, then all $s^*(x)$-households are located at x ($\zeta^*(x, s^*(x)) = 1$). Otherwise, there exist at least two locations x_1 and x_2 such that the $s^*(x)$-households are indifferent between the locations x_1 and x_2. Thus, we have $0 < \zeta^*(\cdot, s^*(x)) < 1$ at x_1 and x_2, while the sum of the shares is equal to 1. In this case, we say that there is *spatial splitting* of identical households.

In our setting, heterogeneous households enjoy different equilibrium utility levels. This is to be contrasted with Ahlfeldt *et al.* (2015) who assume that households share the same expected utility level, which is the exogenous reservation utility that prevails in the rest of the economy.

3 The sorting of skills

3.1 The location-quality index

The *bid rent* $\Psi(x, \overline{y}(s,x), U)$ of a household whose expected income is $\overline{y}(s,x)$ is the highest amount she is willing to pay for one unit of land at x when her utility level is given and equal to U. In other words, the bid rent function is defined as follows:

$$\Psi(x, \overline{y}(s,x), U) \equiv \max_{q,h} \left\{ \frac{\overline{y}(s,x) - q}{h} \middle| \text{s.t. } b(x) \cdot u(q,h) = U \right\}$$
$$= \max_{h} \frac{\overline{y}(s,x) - Q(h, U/b(x))}{h}, \tag{13}$$

where $Q(h, U/b(x))$ is the unique solution to $b(x) \cdot u(q,h) = U$ because u is strictly increasing in h and indifference curves do not cut the axes.

Since households treat the utility level as given, applying the first-order condition to (13) yields the equation:

$$Q(h, U/b(x)) - hQ_h(h, U/b(x)) - \overline{y}(s,x) = 0 \tag{14}$$

whose solution, denoted $H(\overline{y}(s,x), U/b(x))$, is the quantity of land consumed by a s-household at x if her bid rent is equal to the land rent.[1] The solution $H(\cdot)$ is called the *bid-max lot size* (Fujita, 1989). In Appendix A.2, we show that this solution is unique.

The budget constraint implies that the bid rent function may be rewritten as follows:

$$\Psi(x, \overline{y}(s,x), U) \equiv \frac{\overline{y}(s,x) - Q(\overline{y}(s,x), U/b(x))}{H(\overline{y}(s,x), U/b(x))}. \tag{15}$$

Land at x is allocated to the highest bidder. Therefore, if the type $s^*(x)$ is located at x, then $s^*(x)$ must solve the equation:

$$\Psi_s = \frac{\partial \Psi(x, \overline{y}, U)}{\partial \overline{y}} \cdot \frac{\partial \overline{y}(s,x)}{\partial s} = 0, \tag{16}$$

with

$$\frac{\partial \overline{y}(s,x)}{\partial s} = \omega_s t(x) = \frac{\sigma - 1}{\sigma} s^{-1/\sigma} Y^{1/\sigma} t(x),$$

where we have used (6) and (8). The second-order condition implies $\Psi_{ss} < 0$. Applying the implicit function theorem shows that s_x^* and Ψ_{sx} have the same sign. Therefore, we know how the skill mapping varies when the sign of Ψ_{sx} is determined.

Defining

$$B(x) \equiv \frac{b_x(x)}{b(x)} \qquad T(x) \equiv -\frac{t_x(x)}{t(x)},$$

we show in Appendix A.1 that

$$\Psi_{sx} = \frac{t(x)\overline{h}}{H^2} \cdot [B(x) - (1-\mu)T(x)] \cdot \frac{\partial \overline{y}(s,x)}{\partial s}. \tag{17}$$

[1] For any function $f(y,z)$, let f_y (resp., f_{yz}) be the partial (cross-) derivative of f with respect to y (resp., y and z).

As will become clear after Proposition 1, under Stone-Geary preferences, it is possible to subsume the amenity and commuting effects at x into a single scalar that has the nature of a *location-quality index*. This scalar is given by

$$\Delta(x) \equiv b(x)[t(x)]^{1-\mu}. \tag{18}$$

As $\mathbf{L}$ is given, the function $\Delta(x)$ at x is well defined. Since $\partial \overline{y}(s,x)/\partial s > 0$, differentiating (18) shows that $\Delta_x(x)$, $B(x) - (1-\mu)T(x)$ and Ψ_{sx} have the same sign. Therefore, Ψ_{sx} changes sign at any extrema of the location-quality index. Furthermore, the higher μ, the stronger the preference for land. Therefore, as the intensity of preference for land increases, commuting matters less than the accessibility to amenities. We assume without much loss of generality that $b(x)$ and $t(x)$ are such that $\Delta(x)$ is never flat on a positive measure interval.

Although we assume Stone-Geary preferences, our results hold true whenever the location-quality index $\Delta(x)$ is a function of $b(x)$ and $t(x)$ which is independent of s. To illustrate, consider $u(q,h) = q^{\rho_1} + h^{\rho_2}$ with $0 < \rho_i < 1$ and $\rho_1 \neq \rho_2$. The elasticity of substitution between land and the numéraire is variable and equal to $1/(1 - \delta_1\rho_1 - \delta_2\rho_2)$ where δ_i is the expenditure share on good $i = 1, 2$. When $\rho_1 > \rho_2$, i.e., the composite good matters more than land, it can be shown that the above preferences generate the index $\Delta(x) \equiv [b(x)]^{1/\rho_1}\, t(x)$, which is similar to (18).

3.2 The equilibrium skill mapping

Our objective is now to determine the equilibrium skill mapping that specifies which s-households are located at x. The next proposition shows that skills are distributed across N according to the values of the location-quality index. Conditional on $\mathbf{L}$, we rank the values of $\Delta(x)$ by increasing order and denote by $G(\Delta)$ be the corresponding c.d.f. defined over $\mathbb{R}_+$.

The following proposition is proved in Appendix A.2.

Proposition 1. *Assume Stone-Geary preferences. Then, (i) each location hosts at most one household type; (ii) there exists a spatial equilibrium and this equilibrium is unique when the elasticity of* (5) *with respect to employment is not too large; (iii) the equilibrium skill mapping $s^*(x)$ and the equilibrium location-quality index $\Delta^*(x)$ vary together with x. Furthermore, denoting by $G(\cdot)$ the distribution of the values of $\Delta^*(x)$, the equilibrium skill mapping is given by*

$$s^*(x) = F^{-1}[G(\Delta^*(x))]. \tag{19}$$

In Appendix A.2, we also show that the equilibrium utility level satisfies the Spence-Mirrlees condition, thus implying the existence of a positive assortative matching between skills and the values of the location-quality index. In this case, there is a unique one-to-one and increasing relationship between s and Δ (Chiappori, 2017). Hence, the highest skilled locate where the equilibrium location-quality index Δ^* reaches its maximum. As Δ^* starts decreasing with x, the skill level of the corresponding residents also decreases. The lowest skilled reside at a global minimizer of the equilibrium location-quality index. Around this location, the skill level rises together with Δ^*. As a result, the skill sorting does not translate into spatial sorting because the function $\Delta^*(x)$ is in general not monotonic in x. In other words, we have:

$$\frac{\partial}{\partial x}\frac{\mathrm{d}U^*}{\mathrm{d}s} \gtrless 0.$$

For example, in a monocentric city, a wider income gap is no longer matched with a greater distance between two households.

The skill sorting generates a specific output level Y^*. When $b(x)$ is constant, a household chooses the location that maximizes her expected productivity. In this case, the sorting of skills leads to the highest expected total output. By contrast, an uneven distribution of exogenous amenities fosters a lower expected total output because *historic and natural amenities are likely to attract the most skilled people away from the places where they are the most productive.* In this event, the drop in the consumption of private goods is the counterpart of a higher level of local amenity.

3.3 From theory to data

To estimate the model, we need an explicit form of the skill-specific mapping $s^*(x) = F^{-1}[G(\Delta^*(x))]$. For this, we must consider specific distributions F and G. Earning distributions are skewed to the right and the Fréchet distribution is a good candidate to capture this. Equally important, the Fréchet distribution leads to an analytical solution of our model. In what follows, we assume that the variable s is drawn from a Fréchet distribution to the power $(\sigma - 1)/\sigma$ with the shape parameter $\gamma_s > 0$ and the scale parameter $K_s > 0$: $F(z) = \exp(-K_s z^{-\gamma s(\sigma-1)/\sigma})$ over $\mathbb{R}_+$ with density

$$f(s) = K_s\gamma_s\frac{\sigma-1}{\sigma}[\exp(-K_s s^{-\gamma_s(\sigma-1)/\sigma})]s^{-[\gamma_s(\sigma-1)+\sigma]/\sigma}.$$

An increase in γ_s leads to less income inequality.

It is analytically convenient to assume that the values of $\Delta^*(x) = b(x)\,[t^*(x)]^{1-\mu}$ are also drawn from a Fréchet distribution with the c.d.f. $G(z) = \exp(-K_\Delta z^{-\gamma_\Delta})$ over $\mathbb{R}_+$ and density $g(z)$. This holds if $b(x)$ and $t^*(x)$ are Fréchet-distributed.

Using (19), the mapping $s^*(x)$ can then be retrieved from the condition:

$$\int_{s^*}^{\infty} f(z)\mathrm{d}z = 1 - \exp(-K_s(s^*)^{-\gamma_s(\sigma-1)/\sigma}) = \int_{\Delta^*}^{\infty} g(\zeta)\mathrm{d}\zeta = 1 - \exp(-K_\Delta(\Delta^*)^{-\gamma_\Delta}),$$

which is the counterpart in the Δ^*-space of the land market clearing condition (12). It follows from Proposition 1 that households ranked by decreasing incomes are assigned to locations having a decreasing location-quality index.

Set $\gamma \equiv \gamma_\Delta/\gamma_s$ and $K \equiv K_s/K_\Delta$. Solving the above equation yields the equilibrium skill mapping:

$$s^*(x) = \left\{K^{1/\gamma_s} \left[\Delta^*(x)\right]^\gamma\right\}^{\sigma/(\sigma-1)}. \tag{20}$$

We show in Appendix A.2 that $\zeta^*(x, s^*(x))$ is uniquely determined for any x. Therefore, the equilibrium output is given by

$$(Y^*)^{(\sigma-1)/\sigma} = \sum_{i=1}^{n} \int_N \left[A_i(L_i^*)\ell_i(x)\, s^*(x)\right]^{(\sigma-1)/\sigma} \pi_i^*(x)\zeta^*(x, s^*(x)) f(s^*(x))\mathrm{d}x,$$

where $\pi_i^*(x)$ is obtained by replacing $t_i(x)$ by $t_i^*(x)$ in $\pi_i(x)$.

Since a continuous distribution of skills is not directly observed in the data, we estimate the *income mapping* instead. As $\omega = s^{(\sigma-1)/\sigma}Y^{1/\sigma}$, the equilibrium income mapping is thus given by

$$\overline{y}(s^*(x), x) = t^*(x)\omega(s^*(x)) = K^{1/\gamma_s}\left[\Delta^*(x)\right]^\gamma t^*(x)(Y^*)^{1/\sigma}. \tag{21}$$

Last, we show in Appendix A.3 that the equilibrium land rent at x is given by

$$R^*(x) = \mu(1-\mu)^{\frac{1-\mu}{\mu}} k^{-\frac{1}{\mu}} t^*(x)\left[\Delta^*(x)\right]^{\frac{1}{\mu}} \left[\frac{\mu t^*(x)}{R^*(x)} + \frac{(1-\mu)\overline{h}}{\omega(s^*(x))}\right]^{\frac{1}{(1-\mu)\mu\gamma}}, \tag{22}$$

where k is a positive constant.

Therefore, *the land rent is a priori neither monotonic nor the mirror image of the spatial income distribution*. In short, the interaction between amenities, commuting and income sorting may give rise to a variety of land rent profiles, which are not driven by the location-quality index alone.

4 Data and descriptives

4.1 Datasets

We have gained access to various nationwide non-public microdata from *Statistics Netherlands* between 2010 and 2015. Unlike the United States or the United Kingdom, the Netherlands does not undertake censuses to register their population, but the register is constantly updated when people move or when there are changes in the household composition. The first dataset we use is the *Sociaal Statistisch Bestand* (*SSB*), which provides basic information on demographic characteristics, such as age, country of birth, marital status and gender. We only keep people that could be part of the working population, that is, those who are between 18 and 65 years and aggregate these data to the household level. Importantly, the *SSB* data enable us to determine where households reside, up to the postcode level. Hence, space is discrete in the plane.

The data on yearly income of households is obtained from the *Integraal Huishoudens Inkomen* panel dataset. These data are based on the tax register, which provides information on taxable income, tax paid, as well as payments to or benefits from property rents or dividends. The income data also provide information on whether households are homeowners or renters. Public housing is rent controlled and there are often long waiting lists for public housing. So, households are not entirely free to choose their utility-maximizing location. Therefore, we will focus on owner-occupied housing, which means that we keep about 70% of the data.[2]

To estimate the commuting time for each household, we use the tax register information, which provides information on individual jobs and the number of hours worked in each firm for each year. Using data on location information on each establishment from *ABR Regio* and network travel time from *SpinLab* we calculate for each household the average commuting time. More information on how we calculate the commuting time between locations is provided in Appendix B.1.

Information on land values and lot sizes is not directly available. As is common practice, we infer them from data on housing transactions, provided by *Dutch Association of Real Estate Agents* (*NVM*). The methodology used to calculate land values and lot sizes is described in Appendix B.2. The *NVM* data contains information on the large majority (about 75%) of owner-occupied house transactions between 2000 and 2015. We know the transaction price, the lot size, inside floor space size (both in m^2), the exact address, and a wide range of housing attributes such as house type, number of rooms, construction year, garden, state of maintenance, and whether a house is equipped with central heating.[3] We also know whether the house is a listed building.

We are interested in the impact of amenities on income sorting and land prices. We proxy the amenity level by the picture density in a neighborhood. More specifically, we gather data from Eric Fisher's *Geotagger's World Atlas*, which contain all geocoded pictures on the website *Flickr*. The idea is that locations with an abundant supply of aesthetic amenities will have a high picture density. We show in Appendix B.6 that there is a strong positive correlation between picture density and historic amenities or geographical variables, such as access to open water or open space. There are, however, several issues with using geocoded pictures as a proxy for amenities.[4]

First, to avoid the possibility of inaccurate geocoding, we keep only one geocoded picture per

[2] We furthermore obtain information on the educational level of adults in the household. This is available for only 75% of the population, but our main specifications will not use these data, so this appears not to be an issue.

[3] We exclude transactions with prices that are above €1 million or below €25,000 and have a price per square meter which is above €5,000 or below €500. We furthermore leave out transactions that refer to properties that are larger than $250m^2$of inside floor space, are smaller than $25m^2$, or have lot sizes above $5000m^2$. These selections consist of less than one percent of the data and do not influence our results.

[4] Ahlfeldt (2013) shows that in Berlin and London the picture density is strongly correlated to the number of restaurants, music nodes, historic amenities and architectural sites, as well as parks and water bodies.

location defined by its geographical coordinates.[5] This reduces the number of pictures by about 50%. Second, one may argue that the patterns of pictures taken by tourists and residents may be very different. Since we have information on users' identifiers, we can distinguish between residents' and tourists' pictures by keeping users who take pictures for at least 6 consecutive months between 2004 and 2015 in the Randstad. It seems unlikely that tourists stay for 6 consecutive months in the area. Note that the correlation between residents' and tourists' pictures is 0.653, which is rather low. Third, many recorded pictures may not be related to amenities but to ordinary events in daily life occurring inside the house. Hence, we only keep pictures that are taken *outside* buildings, using information on all the buildings in the Netherlands from the *GKN* dataset, which comprises information on the universe of buildings. Furthermore, if pictures are not related to amenities, one would expect almost a one-to-one relationship with population density. However, if we calculate the population density in the same way as we calculate the amenity level, the correlation is only 0.223. Last, we recognize that people who take pictures may belong to a specific socio-demographic group (e.g., young people with a smartphone) by including demographic controls and using instrumental variables.

Though imperfect, we believe that the picture density is probably the best proxy available for the relative importance of urban amenities at a certain location because it captures both the heterogeneity in aesthetic quality of buildings and residents' perceived quality of a certain location. Nevertheless, we test the robustness of our results using an alternative hedonic amenity index in the spirit of Lee and Lin (2018) (see Appendix B.3 for more details). The hedonic index aggregates the average impact of several proxies of amenities, such as the locations of historic buildings, proximity to open space and water bodies, by testing their joint impact on house prices. We also construct historic instruments. Knol *et al.* (2004) have scanned and digitized maps of land use in 1900 into 50 by 50 meter grids and classified each grid into 10 categories, including built-up areas, water, sand, and forest. We aggregate these 10 categories into 3 categories: built-up areas, open space, and water bodies and calculate the share of the area used for each type in each neighborhood. We further gather data from the 1909 census on occupations and employment in each municipality. Those ones were much smaller than current ones and about 4 times the size of the current neighborhoods. For each occupation we obtain the required skill level. This enables us to calculate the share of households who are medium and high-skilled. We gather additional data on the railway network in 1900 and the stations which by then existed (see Appendix B.4 for more information), enabling us to calculate employment accessibility in 1909. To show robustness, similar instruments based on land use in 1832 obtained from *HISGIS* and *NLGIS* are constructed. *HISGIS* provides information on the exact space occupied by buildings. The cadastral income was

[5] In continuous space, the probability that several pictures are taken at *exactly* the same location is zero. Hence, observing multiple pictures at the same location is likely caused by inaccurate geocoding.

used to determine the property tax and reflected the land value at that time. A disadvantage of the *HISGIS* is that it is only available for parts of the Netherlands, thereby reducing the number of observations by about 50%. Additional information on the road network in 1821 is obtained from Levkovich *et al.* (2017).

4.2 Descriptive statistics

Figure 1.A provides a map of the Netherlands, the study area, where we indicate the most important cities. The conurbation formed by the four largest cities, i.e., Amsterdam, Rotterdam, The Hague, and Utrecht is known as the Randstad, which has a population of about 7.1 million. Figure 1.B displays the commuting pattern across neighborhoods and shows that the Dutch urban structure is really polycentric as many commuting flows occur between different cities. This underlines the need for a model that allows for location choices in the whole country. Figure 1.C is a map of the most important roads and railways that form the transportation network in the Netherlands.

[Figure 1 about here]

We report descriptive statistics of the $10,213,524$ households of our sample in Table 1. The average (median) yearly income is €$91,535$ (€$86,732$). Incomes are approximately Fréchet distributed (see Appendix B.5).[6] The average land price in the sample is €$1,312$, but there are stark spatial differences. For example, in the capital Amsterdam, it is €$3,046$, while in the rural province of Friesland it is only €716. As expected, the correlation between the estimated land price and lot size is negative ($\rho = -0.245$). The average lot size is 364m^2. However, in Amsterdam it is only 253m^2, which corresponds to the higher land values in this city. About 15% of households occupy apartments and the correlation between occupying an apartment and the land price is positive ($\rho = 0.153$).

[Table 1 about here]

We use the *neighborhood* definition proposed by Statistics Netherlands, so that we have $4,033$ neighborhoods, which define from now on to be the location set. The picture density, i.e., the proxy for amenities, range from 0 to 231 pictures per hectare. Only 0.2% of the households live in neighborhoods that do not have any pictures. We will disregard those households. The average picture density in Amsterdam (22.7) is much higher than in Rotterdam (9.63), The Hague (6.17), and Utrecht (7.66). Recall that we only use pictures *outside* a building taken by *residents* in determining the amenity index. It appears that 80% of the pictures are taken outside a building

[6] We report maps and histograms of income and land prices in Appendix B.5.

while about 60% of the pictures are taken by local residents. Going back to Table 1, we see that the average commuting time is 26 minutes, which is very close to statistics provided by other sources (Department of Transport, Communications and Public Works, 2010). The unconditional correlation of picture density with the income is close to zero ($\rho = 0.0533$), but this is not very informative as we do not control for household characteristics. The correlation of the amenity index with land prices is substantially higher ($\rho = 0.431$). Finally, *households that have a short commute do not seem to live in high amenity locations* as the correlation between the amenity level and commuting time is low ($\rho = -0.0454$).

The descriptive of the historic instruments that we use are described in Table B.6 of Appendix B.4.

5 Reduced-form income mapping

5.1 Econometric framework and identification

Before developing the structural estimation of the parameters of the model, we consider the *income mapping*, which plays a key role in our model. We first provide *reduced-form* evidence that sorting by incomes is indeed related to our proxy for amenities and accessibility to jobs – the variables that constitute the location-quality index (see (21)). Set

$$\log \tilde{y}_{ki}(x) = \alpha_1 \log \tilde{b}(x) + \alpha_2 \log \tilde{a}(x) + \alpha_3 C_k + \Omega_i + \xi_{ki}(x), \tag{23}$$

where $\tilde{y}_{ki}(x)$ is the observed income net of commuting of household k living at x and working in i; $\tilde{b}(x)$ is the density of geocoded pictures – our proxy for amenities, $\tilde{a}(x)$ is a proxy for employment accessibility, C_k are household characteristics, Ω_i are workplace fixed effects, and $\xi_{ki}(x)$ is an error term. The parameters α_1, α_2, α_3 and Ω_i are estimated. For the moment, we proxy $\tilde{a}(x)$ by:

$$\tilde{a}(x) = \sum_{i=1}^{I} F(\tau_i(x)) n_i.$$

In other words, at location x we weight the number of jobs n_i at i by the share of people whose commute is at most equal to $\tau_i(x)$.

There are several issues when using (23) to identify the causal impact of $\tilde{b}(x)$ and $\tilde{a}(x)$ on sorting on the basis of income. First, regarding accessibility $\tilde{a}(x)$, a reason for a bias is that labor markets may not be fully competitive as households may bargain over to get an income compensation for living further away. Hence, observed incomes $\tilde{y}_{ki}(x)$ may be higher when people live further away. Note that about 15% of the costs of a longer commute is paid by the employer (Mulalic *et al.,* 2013).

Second, a more general concern about α_1 and α_2 as measures of the impacts of amenities and accessibility on the spatial income distribution is that there is an omitted variable bias due to sorting, heterogeneity in preferences for housing quality, agglomeration economies, and unobserved spatial features. More specifically, households may not only sort on the basis of income, but also on the basis of other household characteristics. Households with children, for example, may aim to locate in neighborhoods with a large amount of green space. The variables $\tilde{b}(x)$ and $\tilde{a}(x)$ could also be correlated with unobserved housing attributes because households with different incomes may have different preferences for housing quality, such as the age of the housing stock (Brueckner and Rosenthal, 2009). For example, a large share of the housing stock in the city center of Amsterdam takes the form of apartments. This may imply that the affluent are not willing to locate there because they eschew apartment living (Glaeser *et al.*, 2008).

Third, there may be reverse causality between $\tilde{y}_{ki}(x)$ and $\tilde{b}(x)$ and between $\tilde{y}_{ki}(x)$ and $\tilde{a}(x)$. For example, the provision of amenities may be a direct result of the presence of high-income households. Indeed, anecdotal evidence suggests that cultural and leisure services are often abundantly available in upscale neighborhoods (Glaeser *et al.*, 2001). Similarly, high income neighborhoods may attract employers that are in need of specialized and highly educated labor. Last, since we do not observe the 'exact' amenity level, there may be a measurement error in $\tilde{b}(x)$, which may lead to a downward bias of α_1 when the error is random.[7]

The first step to mitigate the biases associated with these concerns is first to 'purge' *household*, *job* and *housing* characteristics, C_k, from neighborhood characteristics. For example, C_k captures the members of the households who work full-time or part-time, the size of the household and the age of the adults, while housing attributes are, for example, housing type and construction year. This approach reduces the likelihood that we measure sorting on the basis of household characteristics other than incomes. Furthermore, since we also include workplace fixed effects Ω_i, we control for productivity differences (e.g., due to agglomeration economies) at the workplace.

Working with an endless string of controls will not fully address the endogeneity concerns raised above. Unfortunately, our data do not allow us to exploit quasi-experimental or temporal variation in $\tilde{b}(x)$ and $\tilde{a}(x)$. Therefore, to investigate the importance of omitted variable bias we analyze coefficient movements after including controls. Oster (2019) shows that coefficient movements together with changes in the R^2 can be used to estimate biased-corrected coefficients. We will outline this procedure and discuss the results in detail in Appendix B.6.

Omitted variable bias is not the only endogeneity issue. Our proxies may also suffer from

[7] As suggested by the literature on local public goods, there might be reverse causality, meaning that the location of local public goods and jobs is determined by the spatial income distribution. To a large extent, this is because the institutional context that prevails in the U.S. implies that the quality of schools and other neighborhood characteristics are often determined by the average income in the neighborhood (Bayer *et al.*, 2007). This is to be contrasted with what we observe in many other countries where local public goods such as schools are provided by centralized bodies.

measurement error and reverse causality. We will, therefore, rely on instrumental variables. Our first set of specifications uses contemporary instruments, while our second set of specifications appeals to historic instruments. Regarding contemporary instruments for amenities, we use a set of observed, arguably exogenous, proxies for amenities, such as the listed building density, the share of the neighborhood x that is in a historic district, as well as the share of built-up areas and water bodies. By using other proxies for amenities, the measurement error of $\tilde{b}(x)$ is likely to be mitigated. One may argue that the contemporary instruments do not convincingly address the issue of unobserved locational and household characteristics that may be correlated with $\tilde{b}(x)$. Moreover, they do not address the potential endogeneity of accessibility $\tilde{a}(x)$.

Alternatively, we exploit the fact that $\tilde{b}(x)$ and $\tilde{a}(x)$ are autocorrelated. First, land use in 1900 is used as an instrument. We expect aesthetic amenities to be positively correlated to the share of built-up area in 1900. For example, the historic city center of Amsterdam has many buildings that have been built before 1900, which are now listed buildings. Furthermore, we also expect water bodies available in 1900 to be correlated to current water bodies, which are often considered as an amenity. As an instrument for commuting time, we count the total number of households $E_{x,1909}$ in 1909 within a commuting distance by using the railway network in 1900:

$$a_{1909}(x) = \sum_{i=1}^{n} F(\tau_i(x)) n_{i,1909}, \tag{24}$$

where $\tau_i(x)$ is the commuting time between x and employment location $i = 1, ..., n$, while $F(\tau_i(x))$ is the share of people who commute at most τ minutes in the sample (see Appendix B.1). Hence, $F(\tau_i(x))$ represents the aggregate cumulative distribution of commuting times, while $n_{i,1909}$ is the total employment at i in 1909. Because of temporal autocorrelation, we expect that a better employment accessibility in 1909 also implies a better employment accessibility today.

Historic instruments can be criticized because of the (strong) identifying assumption that past unobserved locational features are correlated to current unobserved locational endowments. However, these instruments are more likely to be valid in the context of income sorting because the patterns of income sorting within each city have considerably changed throughout the last century. Around 1900, open water and densely built-up areas were not necessarily considered as amenities. For example, the canals in Amsterdam were essentially open sewers (Geels, 2006). Therefore, locations near a canal often repelled high-income households who located in lush areas just outside the city. It was also before the time when cars became the dominant mode of transport. People around 1900 often walked to their working place, so that commuting distances were short. However, the rich could afford to live outside the city and take the train to their workplace. The cities in 1900 were not yet influenced by (endogenous) planning regulations, as the first comprehensive city plans date from the 1930s.

Still, one may be concerned that the measure of amenities is itself determined by the wealth of individuals who locate there. The reason is that unobservables that determine the concentration of wealthy individuals in the past also determine the locations of landmarks today, and thus determine where pictures are taken. Moreover, one may argue that historic employment accessibility, which is correlated to current employment accessibility, makes it easier to find jobs for all household members, and thus increases household income due to better matching, rather than shorter commutes. We address these concerns in several ways.

1. We go back further in time as it is less likely that unobserved characteristics of a location or building in the past are correlated with those in present time. We exploit land use data from the census in 1832. We use municipal populations in 1832 and calculate the travel time of population within commuting distance using information on the road network of 1821. We further control for the share of buildings, the share of built-up area, and the share of water bodies within neighborhood as instruments. Moreover, using data on the Cadastral Income, we can control for the value of land at that time. If rich households sort themselves into the most attractive locations of the past, we expect to see a positive correlation with the Cadastral Income in 1832.

2. We estimate specifications where we control for the current share of built-up areas and population density. Locations that were attractive in the past attracted people and consequently have a high share of built-up area in 1900. The share of built-up areas in 1900 is likely to be correlated to the current population density and to shares of built-up areas nowadays. By controlling for the current share of built-up areas and population density we mitigate the issue that our proxy for amenity just captures contemporary population density, rather than a higher amenity level because of the historic buildings.

3. We gather data from the 1909 census on occupations and skills in each municipality. We then control in various ways for the average skill level of households in 1909 as a proxy for the income in the past. Controlling for the skill level should also address the issue that employment density in 1909 may be correlated to better matching opportunities. Since this proxy may be imperfect, we also use the share of Protestants in 1899 at the municipality level as another proxy for income/skill. Indeed, at that time Protestants had a higher education level and were wealthier.

4. We also consider another instrument for employment accessibility. From the 1899 census, we gather data on the share of locally born people (i.e., within the same municipality). If the (lack of) mobility of households is correlated over time, the share of locally born people should be correlated positively to current commuting times because immobile households have to commute on average longer to their jobs.

5. Finally, we estimate specifications where we exclusively focus on areas of reclaimed land since 1900. These are areas that are reclaimed from the sea (about 5% of the land) just before and after World War II. As these reclaimed locations are otherwise identical, and as no one was living in those locations at that time, we address reverse causality.

5.2 Reduced-form results

Table 2 reports the baseline reduced-form results of the income mapping. Column (1) shows a simple regression of log income on log amenities and log accessibility, while we only control for demographic characteristics and year fixed effects. This shows that more amenities and accessibility are associated with higher incomes. Doubling amenities implies an increase in income of $(\log 2 - \log 1) \times 0.0215 = 1.5\%$. Doubling of accessibility attracts households whose incomes are 6.9% higher. In column (2), we add a wider array of controls related to housing quality and job characteristics. Although the R^2 increases by almost 50%, the coefficients related to amenities and accessibility are hardly affected. This suggests that amenities are not so much correlated to building quality. In column (3), we include workplace fixed effects to control for agglomeration economies in the workplace and identify the 'pure' accessibility effect. We observe that the coefficient is somewhat lower. A 100% increase in amenities now attracts households whose incomes are 1.2% higher. The coefficient related to employment accessibility is hardly affected.

Despite the inclusion of controls and workplace fixed effects, one may argue that we do not convincingly address omitted variable bias. We deal with this issue by estimating bias-corrected regressions following Oster (2019) in Appendix B.6. We show that when we choose the appropriate maximum attainable R^2 (as only part of the variation in incomes can be explained by variables varying at the neighborhood level), the estimates are very close to the OLS estimates. This strongly suggests that omitted variable bias is not a major issue.

In column (4) we aim to address potential measurement error in the picture density as a proxy for amenities by instrumenting for it with observed proxies for amenities (e.g., nearby historic buildings or share water bodies). The first-stage results in Appendix B.6 show the expected signs: there is a higher picture density in built-up areas, in areas with more water bodies (e.g., the Amsterdam canal district), and where there are many historic buildings.[8] The contemporary instruments are strong instruments for amenities. The second-stage coefficient related to amenities in column (4), Table 2, is essentially identical, suggesting that measurement error is not a main concern.

[Table 2 about here]

Yet, amenities and accessibility may be endogenous due to reverse causality. The use of

[8] Since we have more instruments than endogenous variables, one might object that two-stage least squares estimates are biased (Angrist and Pischke, 2009). Hence, we also have experimented with other estimators that are (approximately) median unbiased, such as LIML or GMM estimators. The results are virtually identical. For this reason, we do not report them in the paper.

contemporary instruments may only partly address this issue. This is why we instrument amenities with historic variables in column (5). The instruments are the shares of water bodies and of built-up area in 1900 within a neighborhood x. In Appendix B.6, we report the corresponding first-stage results. The share of built-up area, the share of water bodies in 1900 are strongly and positively correlated to the current amenity level. Going back to Table 2, the coefficient of amenities is now somewhat higher: doubling amenities attracts households whose incomes are 2.3% higher. In column (6) we also instrument for employment accessibility with the number of households within commuting distance in 1909 using the railway network in 1900. The number of people reachable within commuting distance is positively correlated to current accessibility; the elasticity is 0.42. Overall, the Kleibergen-Paap F-statistic is above the rule-of-thumb value of 10 in all specifications, suggesting that the instruments are sufficiently strong.

The second-stage results reported in column (6), Table 2, reveal that when we instrument amenities and commuting times there is *a positive effect of picture density and accessibility on incomes*. This specification is our preferred specification. Doubling amenities attracts households whose incomes are 2.3% higher. Doubling accessibility leads to households whose incomes are 3.8% higher. The impacts of accessibility and amenities are thus similar.

Alternative proxies for amenities and effects on land prices. One may worried that our results hinge on the particular choice of the amenity index. We therefore consider three alternative proxies for amenities. Following Lee and Lin (2018), we construct an aggregate hedonic amenity index that describes the amenity provision at every location using house prices. The procedure is described in Appendix B.3. To make the results comparable, we rescale the hedonic amenity index in such a way that the standard deviation of the log of the hedonic amenity index is the same as that of the log of the picture index. In column (1), Panel A of Table 3, we re-estimate our preferred specification with historic instruments. It appears that the amenity elasticity is essentially the same as the estimates obtained by using the picture index. We also gather data on 'places of interest' from the augmented reality game *Pokémon Go* as another proxy for amenities (see Appendix B.3 for detail). Our results show that the density of Pokéstops is positively associated with incomes: doubling the Pokéstop density attracts households whose incomes are 2.2% higher. The commuting time elasticity is very much the same compared to the baseline specification. In column (3) we use the share of land available in a neighborhood that is part of an officially designated historic district. Using historic instruments, we find a strong and statistically significant effect on household incomes: a 10% increase in the share of land that is part of a historic district attracts households whose incomes are 3% higher.

[Table 3 about here]

In Panel B of Table 3, we investigate the reduced-form impacts of amenities and commuting times on land prices. In our setup the signs of the effects of amenities and accessibility on land prices and incomes are the same (although magnitudes may differ). Therefore, we now estimate the effects of amenities and commuting time on land prices. We start in column (4), Table 3, with a simple OLS specification including amenities and accessibility, while controlling for households, job and housing characteristics. This leads to a strong positive effect of amenities on land prices: doubling amenities implies a land price increase of 8.7%, while doubling accessibility leads to land prices that are 22.4% higher. When we control for workplace fixed effects, the coefficients are hardly affected. In the final column we instrument for amenities and accessibility with historic instruments from around 1900. The effect of accessibility becomes somewhat lower, while the effect of picture density becomes about twice as strong. Hence, the reduced-form effects on land prices do indeed have the same signs as the effects on income, but are stronger in magnitude.

Other sensitivity checks. Appendix B.8 shows that our results still hold for a wide range of alternative robustness checks and sample selections. To the extent one is still worried that endogeneity plagues our estimates, we strongly advise the reader to consult Appendix B.8. More specifically, we show that our results hold if we *(i)* only focus on the urban area of the Randstad or close to city centers, *(ii)* use data from 1832 to construct instruments for amenities and employment accessibility, *(iii)* control for current land use and population density, *(iv)* control for sorting based on skills in 1909, *(v)* use alternative (historic) instruments and *(vi)* only use observations in land that is reclaimed from the sea.

Further robustness analyses minimize any measurement error regarding accessibility and workplace productivity, by running specifications where we only keep households *(i)* with a single job, *(ii)* with a single job in a single-plant firm, and *(iii)* households with a company car so that it is likely that those households actually use car for commuting. We further test whether our results change when using the share of highly educated adults in the household, which is a more direct way to estimate the (reduced-form) skills mapping. We find very similar effects, both in terms of sign and magnitude, which confirms that looking at income or skill levels is more or less equivalent. We also use commuting time by rail instead of commuting time over the road. Overall, *the impact of amenities and commuting time on income sorting choice is robust.*

6 Structural estimation

6.1 Estimation and identification

We define the amenity function as follows:

$$b(x) \equiv \left[\tilde{b}(x)\right]^{\beta}.$$

In other words, amenities are related to picture density ($\tilde{b}(x)$) where β is the elasticity of preferences for amenities. Furthermore, letting $\ell_i(x) \equiv [\tau_i(x)]^{-\kappa}$ with κ the elasticity of commuting time, the labor supply is given by

$$t_i(x)\nu_{ki}(x) = \{A_i[\tau_i(x)]^{-\kappa}\}^{(\sigma-1)/\sigma}\nu_{ki}(x),$$

where $\nu_{ki}(x)$ are i.i.d. idiosyncratic shocks on commuting times drawn from a Fréchet distribution with a shape parameter $\varepsilon > 1$ and scale parameter $K_i > 0$.

We use the structure of the model to identify its parameters $\{\beta, \kappa, \mu, \varepsilon, \delta, \gamma, \gamma_S, \gamma_\Delta\}$. In this way we are able to calculate the counterfactual income mappings and land rents.

Our model has a recursive structure. Hence, estimation of the parameters consists of estimating a number of standard regression equations. However, only the first step – estimating the commuting gravity equation – is the same as in Ahlfeldt *et al.* To be precise, the gravity equation identifies the commuting time elasticity $\varkappa = \kappa\varepsilon$. Since our model differs considerably from Ahlfeldt *et al.* (2015), it comes as no surprise that the remaining steps used to recover the model parameters are substantially different.

In the second step, using actual data on incomes, we can recover commuting heterogeneity ε. Third, using information on land rents and lot sizes, which we observe for a subset of the data, we recover preferences for land μ. The fourth step uses the income mapping to identify the preferences for amenities β and the relative heterogeneity of the location quality index γ. This enables us to identify the location quality index up to a multiplication constant. The remaining heterogeneity parameters (γ_Δ and γ_S) are identified in the fifth step. In the final step, we estimate the agglomeration elasticity δ.

In line with spatial quantitative equilibrium models, we fix base parameters $\{\sigma, \bar{h}\}$.[9] Furthermore, we choose $\sigma = 4$ which is in line with the literature (Dustman *et al.*, 2009). We set $\overline{h} = 25\text{m}^2$, which corresponds to the minimum lot size in the sample, and we use a discount rate of 3.5% to go from land prices to land rents (see Koster and Pinchbeck, 2019). In what follows, we discuss the moment conditions, the identifying assumptions in each step, and the estimation

[9] Note that we identify everything up to a multiplication constant. Therefore, the scale parameters of the Fréchet distribution are not strictly identified.

procedure.

6.1.1 Estimating the gravity equation

It follows from (9) that the probability that a household living in x chooses to work in i is equal to:

$$\pi_i(x) = \frac{K_i(\omega^*(x)t_i(x))^{\varepsilon}}{\sum_{j=1}^{n} K_j(\omega^*(x)t_j(x))^{\varepsilon}} = \frac{K_i A_i^{\frac{\varepsilon(\sigma-1)}{\sigma}} [\tau_i(x)]^{-\frac{\varepsilon\kappa(\sigma-1)}{\sigma}}}{\sum_{j=1}^{n} K_j A_j^{\frac{\varepsilon(\sigma-1)}{\sigma}} [\tau_j(x)]^{-\frac{\varepsilon\kappa(\sigma-1)}{\sigma}}}. \tag{25}$$

In line with Ahlfeldt *et al.* (2015), we first recover an estimate for $\varkappa \equiv -\varepsilon\kappa\,(\sigma-1)\,/\sigma$ by estimating a log gravity model with residence and workplace fixed effects, which absorb K_i and A_i. The first moment condition is given by:

$$\mathbb{E}[\log \pi_i(x) - \varkappa \log \tau_i(x) - \tilde{\Upsilon}(x) - \tilde{\Omega}_i] = 0. \tag{26}$$

By including residence fixed effects $\tilde{\Upsilon}(x)$ and workplace fixed effects $\tilde{\Omega}_i$, we mitigate the endogeneity issues associated with $\tau_i(x)$. We then use Poisson Pseudo-Maximum Likelihood methods to deal with the zeroes. One remaining issue is the reverse causality between flows and travel times. Indeed, at locations where there is more demand for travel, better transport infrastructure is likely to be provided, which in turn leads to a shorter travel time. We address this issue by instrumenting $\log \tau_i(x)$ with the log of Euclidian distance between two locations. We use a control function approach where the first stage residual is inserted as a control function in the second stage.

6.1.2 Commuting heterogeneity

The next step is to recover ε from the data. In the spirit of Ahlfeldt *et al.* (2015), we choose to minimize the squared differences between variances within neighborhoods x of adjusted labor supply and labor supply observed in the data. More specifically, let $\tilde{y}_{ki}(x) \equiv \omega(x)t_i(x)\nu_{ki}(x)$ be the observed income in the data of a household k located in neighborhood x and working in neighborhood i. We observe income conditional on labor supply in the data. For example, if someone has a longer commute and therefore supplies less labor, we observe a lower income net of commuting costs. More specifically, we use the observed income and control for household characteristics and *location pair fixed effects* at the level of the neighborhood. We then recover location-specific income $\log \widehat{y}_i(x)$ by taking the estimated values of the location pair fixed effects. Let $\tilde{t}_i(x) \equiv K_i[t_i(x)]^{\varepsilon}$ be the transformed labor supply, obtained from (26). Note that $\sigma^2_{\log(\omega(x)\tilde{t}_i(x))|x} = \sigma^2_{\log \tilde{t}_i(x)|x}$ because $\omega(x)$ does not vary within the neighborhood. Hence, the relationship between the variance within neighborhood x of the log transformed incomes $\sigma^2_{\log \tilde{t}_i(x)|x}$ and the variance of the log of location-specific income, incomes $\sigma^2_{\log \widehat{y}_i(x)|x}$ is given by $\sigma^2_{\log \tilde{t}_i(x)|x} = \varepsilon^2 \sigma^2_{\log \widehat{y}_i(x)|x}$. This enables us to recover ε from the second moment condition:

$$\mathbb{E}[\sigma^2_{\log \tilde{t}_i(x)|x} - \varepsilon^2 \sigma^2_{\log \widehat{y}_i(x)|x}] = 0. \tag{27}$$

Since the above specification is linear in parameters, we can just use linear regression techniques to obtain ε.

6.1.3 Preferences for land

In the third step we use information on land prices $R(x)$ and lot sizes $h(x)$ for a subset of the sample. Moreover, we use $\widehat{y}_i(x)$ – the estimated location-specific income – from the previous step. Rewriting (3), we derive the third moment condition to determine μ:

$$\mathbb{E}\left[R(x) - \frac{\mu \widehat{y}_i(x)}{h(x) - (1-\mu)\overline{h}}\right] = 0. \tag{28}$$

Since this equation is non-linear, we use nonlinear least squares to obtain an estimate for μ.

6.1.4 Estimating the income mapping

The income mapping plays a central role in the structural estimation. Recall that the income mapping (21) is derived from the skill mapping (20). The household k who locates at x is given by

$$s(x) = \left[K^{1/\gamma_S}\left(\tilde{\Delta}(x)\right)^{\gamma}\right]^{\sigma/(\sigma-1)}$$

with $K \equiv K_S/K_\Delta$ and

$$\tilde{\Delta}(x) \equiv b(x)\left[\mathbb{E}\left(\max_{i=1,..,n} t_i(x)\nu_{ki}(x)\right)\right]^{1-\hat{\mu}} = [\tilde{b}(x)]^{\beta}\,[\tilde{a}(x)]^{\frac{1-\hat{\mu}}{\hat{\varepsilon}}}, \tag{29}$$

because the maximum of Fréchet variables is a Fréchet variable, while the employment accessibility $\tilde{a}(x)$ is defined as follows:

$$\tilde{a}(x) = \left[\Gamma\left(\frac{\hat{\varepsilon}-1}{\hat{\varepsilon}}\right)\right]^{\hat{\varepsilon}} \cdot \sum_{i=1}^{n} \tilde{t}_i(x) = \left[\Gamma\left(\frac{\hat{\varepsilon}-1}{\hat{\varepsilon}}\right)\right]^{\hat{\varepsilon}} \cdot \sum_{i=1}^{n} K_i A_i^{\frac{\hat{\varepsilon}(\sigma-1)}{\sigma}} \cdot [\tau_i(x)]^{-\frac{\hat{\varepsilon}\hat{\kappa}(\sigma-1)}{\sigma}},$$

where $\tilde{t}_i(x)$ is obtained from the gravity equation and $\Gamma(\cdot)$ is the gamma function. Hence, in contrast to the reduced-form specifications where we choose a somewhat arbitrary functional form for accessibility, we use here an accessibility measure that is dictated by the model.

Recall that the observed income in the data is $\tilde{y}_{ki}(x) \equiv \omega(x)t_i(x)\nu_{ki}(x)$. Since $t_i(x) \equiv [A_i(\tau_i(x))^{-\hat{\kappa}}]^{(\sigma-1)/\sigma}$ and $\omega(x) = [s(x)]^{\frac{\sigma-1}{\sigma}} Y^{1/\sigma}$, the expected income of a household k residing at x and working in neighborhood i can be rewritten as follows:

$$\tilde{y}_{ki}(x) = K^{1/\gamma_S} [\tilde{b}(x)]^{\beta\gamma} [\tilde{a}(x)]^{\frac{(1-\hat{\mu})\gamma}{\hat{\varepsilon}}} [t_i(x)\nu_{ki}(x)](Y^*)^{1/\sigma}.$$

Therefore,

$$\log \tilde{y}_{ki}(x) + \hat{\kappa}\frac{\sigma-1}{\sigma}\log\tau_i(x) = \Upsilon(x) + \Omega_i + \nu_{ki}(x), \tag{30}$$

where $\Upsilon(x)$ are residence fixed effects and Ω_i workplace fixed effects.

We first estimate the location and workplace fixed effects. Then it should hold that

$$\Upsilon(x) = \alpha_0 + \alpha_1 \log \tilde{b}(x) + \alpha_2 \log \tilde{a}(x) \qquad \text{and} \qquad \Omega_i = \frac{\sigma-1}{\sigma}\log A_i,$$

where $\alpha_1 \equiv \beta\gamma$, $\alpha_2 \equiv (1-\hat{\mu})\gamma/\hat{\varepsilon}$. Hence, $\gamma = \hat{\alpha}_2\hat{\varepsilon}/(1-\hat{\mu})$ and $\beta = \hat{\alpha}_1(1-\hat{\mu})/\hat{\alpha}_2\hat{\varepsilon}$.

Equipped with estimates for $\tilde{t}_i(x)$ (from the gravity equation), $\hat{\kappa}$, $\hat{\varepsilon}$, and $\hat{\mu}$, we can infer γ and β from $\Upsilon(x)$. Note further that Ω_i are workplace fixed effects, so that wage differences associated with workplace productivity differences A_i (e.g., due to agglomeration economies) are absorbed by the fixed effects. More specifically, we focus on the job within the household that generates the highest number of working hours and use a work-location fixed effect for each location pair. Hence, we compare households that work at the same location(s), but have different residential locations. Last, given Ω_i, we recover the adjusted workplace productivity $\tilde{A}_i$ (up to a constant):

$$\tilde{A}_i = e^{\frac{\sigma}{\sigma-1}\Omega_i}.$$

We estimate the income mapping in two stages. We define the fourth moment condition as follows:

$$\mathbb{E}[\log \tilde{y}_{ki}(x) + \hat{\kappa}\frac{\sigma-1}{\sigma}\log\tau_i(x) - \Upsilon(x) - \Omega_i - \alpha_3 C_k] = 0, \tag{31}$$

while the fifth moment condition is given by:

$$\mathbb{E}\left[\widehat{\Upsilon}(x) - \alpha_0 - \alpha_1 \log \tilde{b}(x) - \alpha_2 \log \tilde{a}(x)\right] = 0. \tag{32}$$

To obtain the causal parameters α_1 and α_2, and therefore causal estimates for β and γ, we face the same endogeneity issues as in the reduced-form specification (23). We refrain from repeating this discussion here. Like in the reduced-form analysis, we will rely on historic instruments to mitigate endogeneity issues, such as the presence of open space, water bodies and employment accessibility in 1909 using (24). For the income mapping (moment conditions 4 and 5), we also use linear regression techniques. When instrumenting for amenities and employment accessibility, we use two-stage least squares (2SLS) implying that we replace $\log \tilde{b}(x)$ and $\log \tilde{a}(x)$ by their fitted values obtained in the first stage.

6.1.5 Recovering the parameters of the Fréchet distributions

Using $\{\hat{\beta}, \hat{\kappa}, \hat{\mu}, \hat{\gamma}\}$, we may obtain the shape parameters of the location-quality index and the income mapping. First, we calculate the expected labor supply at each location by using (25). Using observed amenities, commuting distances, and the adjusted workplace productivity, we can recover the location-quality index (up to a multiplication constant) at each location:

$$\tilde{\Delta}(x) = [\tilde{b}(x)]^{\hat{\beta}} \cdot [\tilde{a}(x)]^{\frac{1-\hat{\mu}}{\hat{\varepsilon}}}.$$

Hence, the sixth moment condition may be written as follows:

$$\mathbb{E}\left[f_{\tilde{\Delta}}\left(\tilde{\Delta}(x)\right) - \frac{\gamma_{\tilde{\Delta}}}{\tilde{K}_{\tilde{\Delta}}} e^{-\left(\frac{\tilde{\Delta}(x)}{K_{\tilde{\Delta}}}\right)^{-\gamma_{\tilde{\Delta}}}} \left(\frac{\tilde{\Delta}(x)}{\tilde{K}_{\tilde{\Delta}}}\right)^{-(1+\gamma_{\tilde{\Delta}})} \right] = 0, \tag{33}$$

where $f_{\tilde{\Delta}}(\tilde{\Delta}(x))$ is the p.d.f. of the adjusted location-quality index. To obtain the Fréchet parameters (moment condition 6), we use Maximum Likelihood. From this, we obtain $\gamma_s = \hat{\gamma}_{\Delta}/\hat{\gamma}$.

6.1.6 Recovering the agglomeration elasticity

In the last step, we estimate the elasticity of agglomeration economies. We first determine the skill mapping for each location given the estimated parameters:

$$[s(x)]^{\frac{\sigma-1}{\sigma}} = K^{1/\hat{\gamma}_S} [\tilde{b}(x)]^{\hat{\beta}\hat{\gamma}} [\tilde{a}(x)]^{\frac{1-\hat{\mu}}{\hat{\varepsilon}}}.$$

Note that we identify $s(x)$ up to a multiplication constant. Hence, we set K_{Δ} in such a way that the geometric mean of $\widehat{s}(x)$ equals one and then fit a Fréchet distribution to $\widehat{s}(x)$ to obtain K_S.

Using (5), we assume that $A_i = \mathbb{A}_i \tilde{L}_i^{\delta}$, where $\tilde{L}_i$ is given by:

$$\tilde{L}_i = \sum_{x=1}^{N} \frac{\tilde{t}_{xi}}{\sum_{j=1}^{n} \tilde{t}_{xj}} \left[\hat{K}_S \hat{\gamma}_S \frac{\sigma - 1}{\sigma} e^{-\hat{K}_S [\widehat{s}(x)]^{-\hat{\gamma}_S(\sigma-1)/\sigma}} [\widehat{s}(x)]^{-[\hat{\gamma}_S(\sigma-1)+\sigma]/\sigma} \right].$$

The first term is the share of households living at x commuting to i and the bracketed term is the employment density at location x. The moment condition is then given by:

$$\mathbb{E}\left[\log \tilde{A}_i - \log \mathbb{A}_i - \delta \log \tilde{L}_i\right] = 0. \tag{34}$$

Once again, one may argue that $\tilde{L}_i$ is endogenous and correlated to unobserved locational characteristics. As discussed in the foregoing, we use the employment accessibility in 1909. Note that historic instruments are frequently used in the literature to address the endogeneity of employment density (Combes *et al.*, 2010). The identifying assumption permits that past unobservables that cause employment accessibility in 1909 are unrelated to current unobservables to give rise to $\tilde{L}_i$. We provided extensive support for this assumption in Section 5.2. Moreover, we test whether the results are different when using data on accessibility to population in 1832 as it is less likely that unobserved characteristics of a location in the past are correlated with those in present time.

Moment condition 7 can be estimated by OLS, but if we instrument for $\tilde{L}_i$, we replace $\tilde{L}_i$ by the fitted values obtained in a first stage.

6.2 Structural parameters

In Table 4, we report the results of the structural estimation. We obtain cluster-bootstrapped standard errors by first choosing a set of randomly drawn neighborhoods and then estimate the consecutive steps described above 250 times.

We find that a commuting time elasticity equal to $\kappa = 0.22$, which is higher than in the literature. However, one should keep in mind that we use the log of commuting time, so that this represents an elasticity rather than the semi-elasticity. In Appendix B.9, we show that when we instrument travel times with the Euclidian distance, the travel time elasticity is considerably lower, in line with the expectation that reverse causality would lead to an overestimate. Given that endogeneity is quite important, we consider this specification as the preferred one.[10] Commuting heterogeneity ε is about 2.73, which is somewhat lower than Ahlfeldt *et al.* (2015), but close to the value picked by Brinkman and Lee (2019). The estimate μ indicates the preferences for land. We find that $\mu = 0.0955$, which confirms that richer households spend less of their income on land (Albouy *et al.*, 2016). Note that $\hat{\mu}$ may seem low, but we only include payments to land, not to housing itself.

[Table 4 about here]

So far, all the estimated parameters are identical for different specifications because the historic instruments are used only in the later steps to identify preferences for amenities, accessibility, and agglomeration economies. The preference parameter β that indicates how households value amenities in column (1) is similar to the baseline reduced-form result. However, when we use instruments based on data from 1909 β is considerably larger. This is mainly because the relative location-quality heterogeneity parameter γ is about 50% smaller. The preference for amenities

[10] In Appendix B.9 we also show other specifications of the gravity model. We consider *(i)* to use commuting flows based on the two jobs that generate the most working hours, *(ii)* use travel time using railways, and *(iii)* only keep location pairs with a sufficiently high number of commuters. The results are very robust.

is not much affected if we use instruments based on data from 1832. The estimated elasticity of agglomeration economies is 0.0465 if we do not instrument, while it is higher when we use historic instruments (0.0745 and 0.0887 using instruments from 1900 and 1832, respectively). These estimates fall within the range provided by the literature. For example, Rosenthal and Strange (2004) suggest a range of 0.03-0.08. Our estimates are higher than those reported by Combes and Gobillon (2015) who study the elasticity of wages with respect to population density.

Overidentification checks. Our structural estimation procedure suggests natural overidentification checks that can be used to investigate whether our model is able to fit the data reasonably well. We do not expect to find a perfect fit because we consider only two determinants of location choices, while actual location choices are affected by more factors. First, our estimation procedure leads to an approximation for the employment level $\tilde{L}_i$ at each location i. If we compare the estimated $\tilde{L}_i$ to the observed employment level in each area, we find a correlation of 0.839, which is fairly high. One may be worried that this high correlation might be driven by a few locations that host many workers. This appears not to be an issue because the correlation between the log of estimated employment to the log of observed employment is equal to 0.907.

Another overidentification check involves the comparison of the ex-post estimated land rents (see (22)) to the observed land prices. Because land prices are not a direct input in our model, there is no pre-determined mechanical correlation between estimated and actual land prices.[11] We find a correlation between estimated and observed land prices of 0.643. When we correlate the log of estimated land prices to the log of observed land prices, we find a slightly higher correlation ($\rho = 0.718$). Given that we only include two determinants of locational choices, these correlations are quite high as and suggest that amenities and accessibility are very important determinants of locational choices.

6.3 Counterfactual

Given the estimated parameters, our model allows for the undertaking of counterfactual analyses. We describe the exact procedure to solve for the counterfactual values and derive the aggregate land rent and real income in Appendix B.9. Let us consider the scenario where we assume away amenities throughout the Netherlands; that is, we set the amenity level equal to the minimum value of amenities observed in the data.[12] The idea is to mimic U.S. cities where exogenous

[11] One may argue that land prices are used in the determination of μ. However, spatial differences in estimated land price is largely insensitive to the exact magnitude of μ and will hardly affect the correlation between observed and estimated land prices.

[12] The absolute amenity level makes no difference for the outcomes because we re-adjust the parameter K_Δ for the aggregate skill distribution to have a geometric mean equal to 1. Moreover, assuming an equal value for $\tilde{b}_x$ leads to the same result as when setting $\beta = 0$.

amenity levels are considerably lower than in the Netherlands. We estimate the outcomes for the baseline scenario and the counterfactual scenarios and report them in Table 5.

[Table 5 about here]

Since households do not care about amenities anymore, they live on average closer to their workplace and earn higher incomes. We find that the overall output increases by 10.6%, while the aggregate real income rise by 7.3%. The aggregate land rent decreases by 0.6% in the absence of amenities. We also construct a measure of income mixing, which is the standard deviation of skills in adjacent neighborhoods, to see how the counterfactual scenario affects income mixing within the Netherlands. A uniform amenity distribution implies *substantially less income mixing* as the standard deviation is much lower than the baseline estimate. More specifically, income mixing is reduced by about two-thirds, which is substantial. This confirms the anecdotal evidence that European, especially Dutch, cities are more socially mixed than American cities.

Furthermore, a priori one would expect households to be better off because the net income is significantly higher. However, things are not that simple. Even though households are able to consume more, they no longer enjoy the historic and natural amenities. As a result, the result is a priori ambiguous. Our analysis shows that no less than 95% of the whole population of households are worse-off, despite their higher income, than in the situation in which amenities are available.

Having a uniform distribution of amenities also has strong repercussions for the spatial distribution of skills, hence of incomes. Indeed, the correlation between the values observed in the data and in the counterfactual is 0.556. Hence, *amenities are a key determinant of the spatial sorting of households within and between cities*. We report maps in Figure 2. Figure 2.A shows the relative changes in skills at each location. In locations with high levels of exogenous amenities, such as the city center of Amsterdam, The Hague or Utrecht, we observe a relatively large decrease in skills, e.g., up to 20% in Amsterdam, thus confirming that high-skilled people value amenities more. The most skilled households in the Netherlands live in the city center of Amsterdam in the baseline scenario. However, this would change in the no amenity scenario. In Figure 2.B, where we zoom in on Amsterdam, we show that skills decrease by 10-20% in the historic city center of Amsterdam. High skilled households are found in the suburbs where there is an abundance of space. For the lower skilled households changes are less severe: they can mostly be found in the more sparsely populated northern provinces of the Netherlands in both scenarios.

[Figure 2 about here]

As shown in Figure 2.C, the land rent generally decreases in locations with initially high amenities. However, because consumers earn higher incomes they bid more for land, leading to increases in land rent in locations with a high employment accessibility. This is illustrated in Figure 2.D where we zoom in on Amsterdam. The city center, where amenities are now gone, witness decreases in land rents. By contrast, the port area, which has a good employment accessibility, experiences increases in land rents (up to 5.6%).

So far, we did not discuss the implications of the counterfactual for the spatial employment distribution. This distribution is hardly affected by changes in amenities that affect mainly residential choices. The impact on employment centers is second order because agglomeration economies are relatively weak (δ) as compared to the locational fundamentals $\mathbb{A}_i$, which are unaffected by changes in amenities. Production is, to a large extent, anchored in the same locations, thus reflecting the impact of history, like in Bleakley and Lin (2012). We have tested this contention by assuming unrealistically strong agglomeration economies. In this case, the spatial distribution of jobs changes considerably.

7 Concluding remarks

In this paper, we used a new setup in which any location is differentiated by two attributes, i.e., the benefit generated by the amenity field at this location and its distance to employment locations. The bid rent function of urban economics may be used to show that the uneven provision of exogenous amenities is sufficient to break down the perfect sorting of households across the space-economy. Under Stone-Geary preferences, there exists a location-quality index that blends amenities and commuting costs into a single aggregate whose behavior drives households's residential choices. Studying this index allows us to gain insights about how governments and urban planners can design policies whose aim is to redraw the social map of cities. For example, the higher the index of a particular location, the higher the income of households who choose to locate there. The relevance of exogenous amenities and commuting costs to explain the residential choices of heterogeneous consumers is confirmed by the empirical analysis of where both effects are found to be significant. More generally, policies that aim at a uniform distribution of public services push toward more spatial segregation as residential choices are mainly driven by commuting costs. Rather, if social mixing is a major policy objective, our results suggest that governments or related bodies should target specific neighborhoods where to build public facilities providing services to the population.

References

[1] Ahlfeldt, G.M. (2013) Urbanity. London School of Economics. Memo.

[2] Ahlfeldt, G., S.J. Redding, D. Sturm, and K. Wolf (2015) The economics of density: Evidence from the Berlin Wall. *Econometrica* 83: 2127-89.

[3] Albouy, D. (2016) What are cities worth? Land rents, local productivity, and the total value of amenities. *Review of Economics and Statistics* 98: 477-87.

[4] Albouy, D., G. Ehrlich and Y. Liu (2016) Housing demand, costs-of-living inequality, and the affordability crisis. NBER Working Paper 22816.

[5] Allen, T. and C. Arkolakis (2014) Trade and the topography of the spatial economy. *Quarterly Journal of Economics* 129: 1085-140.

[6] Altonji, J.G., T.E. Elder and C.R. Taber (2005) An evaluation of instrumental variables strategies for estimating effects of Catholic schooling. *Journal of Human Resources* 40: 791-821.

[7] Bazzi, S. A. Gaduh, A.D. Rothenberg, and M. Wong (2019) Unity in diversity? How intergroup contact can foster nation building. *American Economic Review* 109(11): 3978-4025.

[8] Beckmann, M.J. (1969) On the distribution of urban rent and residential density. *Journal of Economic Theory* 1: 60-7.

[9] Bénabou, R. (1996) Heterogeneity, stratification, and growth: Macroeconomic implications of communitystructure and school finance. *American Economic Review* 86: 584-609.

[10] Bleakley, H. and J. Lin (2012) Portage and path dependence. *Quarterly Journal of Economics* 127: 587-644.

[11] Brinkman, J. and J. Lin (2019) Freeway Revolts! Federal Reserve Bank of Philadelphia, working paper.

[12] Brueckner, J., J.-F. Thisse and Y. Zenou (1999) Why is Central Paris rich and Downtown Detroit poor? An amenity-based theory. *European Economic Review* 43: 91-107.

[13] Carlino, G.A. and A. Saiz (2019) Beautiful city: Leisure amenities and urban growth. *Journal of Regional Science* 59: 369-408.

[14] Chetty, R. and N. Hendren (2018) The effects of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Quarterly Journal of Economics* 133: 1107-62.

[15] Chiappori, P.-A. (2017) *Matching with Transfers. The Economics of Love and Marriage*. Princeton: Princeton University Press.

[16] Combes, P.P., G. Duranton and L. Gobillon (2008) Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63: 723-42.

[17] Combes, P.-P. and L. Gobillon (2015) The empirics of agglomeration economies. In: G. Duranton, J.V. Henderson and W.C. Strange (eds) *Handbook of Regional and Urban Economics. Volume 5*, 247-348. Amsterdam: Elsevier.

[18] Department of Transport, Communications and Public Works (2010) *Mobiliteitsbalans 2010*. The Hague: Kennisinstituut voor Mobiliteit.

[19] Diamond, R. (2016) The determinants and welfare implications of US workers diverging location choices by skill: 1980-2000. *American Economic Review* 106: 479-524.

[20] Duranton, G. and D. Puga (2004) Micro-foundations of urban increasing returns: Theory. In J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics. Volume 4*. Amsterdam: Elsevier, 2063-117.

[21] Duranton, G. and D. Puga (2015) Urban land use. In G. Duranton, V. Henderson and W. Strange (eds). *Handbook of Regional and Urban Economics, Volume 5*. Amsterdam: Elsevier, 2063-117.

[22] Dustmann C., J. Ludsteck, and U. Schönberg (2009) Revisiting the German wage structure. *Quarterly Journal of Economics* 124: 843-81.

[23] Fujita, M. (1989) *Urban Economic Theory*. Cambridge: Cambridge University Press.

[24] Gaigné, C., H. Koster, F. Moizeau, and J.-F. Thisse (2017) Amenities and the social structure of cities. CEPR Discussion Paper 11958.

[25] Geels, F. W. (2006) The hygienic transition from cesspools to sewer systems (1840–1930): The dynamics of regime transformation. *Research Policy* 35: 1069-82.

[26] Glaeser, E.L., M. Kahn and J. Rappaport (2008) Why do the poor live in cities? *Journal of Urban Economics* 63: 1-24.

[27] Glaeser, E.L., J. Kolko and A. Saiz (2001) Consumer city. *Journal of Economic Geography* 1: 27-50.

[28] Hornbeck, R. and E. Moretti (2019) Estimating who benefits from productivity growth: Direct and indirect effects of city manufacturing TFP growth on wages, rents and inequality, University of Chicago. Memo.

[29] Knol, W. C., H. Kramer and H. Gijsbertse (2004) *Historisch grondgebruik Nederland: Een landelijke reconstructie van het grondgebruik rond 1900*. Wageningen: Alterra.

[30] Koomen, E., J. Dekkers and T. Van Dijk (2008) Open-space preservation in the Netherlands: Planning, practice and prospects. *Land Use Policy* 25: 361-77.

[31] Koopmans, C., P. Rietveld and A. Huijg (2012) An accessibility approach to railways and municipal population growth, 1840–1930. *Journal of Transport Geography* 25: 98-104.

[32] Koster, H.R.A. and T. Pinchbeck (2019). How do households value the future? Evidence from property taxes. *CEP Discussion Paper* 1571.

[33] Koster, H.R.A. and J. Rouwendal (2017) Historic amenities and housing externalities. Evidence from the Netherlands. *Economic Journal* 127: F396-F420.

[34] Lee, S. and J. Lin (2018) Natural amenities, neighborhood dynamics, and persistence in the spatial distribution of income. *Review of Economic Studies* 85: 663-94.

[35] Levkovich, O., J. Rouwendal, and J.N. van Ommeren (2017) The impact of highways on population redistribution: The role of land development restrictions. *Tinbergen Institute Discussion Paper 17-109/VIII.*

[36] McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: P. Zarembka (ed.) *Frontiers in Econometrics*. New York, Academic Press, 105-42.

[37] Montesano, A. (1972) A restatement of Beckmann's model of distribution of urban rent and residential density. *Journal of Economic Theory* 4: 329-54.

[38] Mulalic, I., J.N. van Ommeren and N. Pilegaard (2013) Wages and commuting: Quasi-natural experiments' evidence from firms that relocate. *Economic Journal* 124: 1086-105.

[39] Oster, E. (2019) Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business and Economic Statistics* 37: 187-204.

[40] Ritzen, J.M.M., J. van Dommelen and F.J. de Vijlder (1997) School finance and school choice in the Netherlands. *Economics of Education Review* 16: 329-35.

[41] Rosenthal, S.S. and W.C. Strange (2004) Evidence on the nature and sources of agglomeration economies. In J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics. Volume 4*. Amsterdam: Elsevier, 2120-71.

[42] Rossi-Hansberg, E., P.D. Sarte and R. Owens III (2010) Housing externalities. *Journal of Political Economy* 118: 485-535.

[43] Saiz, A., A. Salazar and J. Bernard (2018) Crowdsourcing architectural beauty: Online photo frequency predicts building aesthetic ratings. *PLOS One* 13: e0194369.

[44] Tsivanidis, N. (2019) Evaluating the impact of urban transit infrastructure: Evidence from Bogotá's TransMilenio. UC Berkeley, memo.

Tables

Table 1 – Descriptive statistics

	(1) mean	(2) sd	(3) min	(4) max
Gross income *(in €)*	91,535	53,683	3,589	999,897
Land price *(€per m^2)*	1,312	752.2	0.00753	22,418
Lot size *(m^2)*	364.3	923.8	25	24,824
Pictures per ha	2.189	8.840	0	231.9
Hedonic amenity index	2.821	0.0915	2.723	3.885
Share historic district	0.0347	0.139	0	1
Listed building	0.0941	0.699	0	17.06
Share built-up land	0.449	0.298	0.000856	1
Share water	0.0496	0.0738	0	0.813
Commuting time in minutes	26.39	17.18	0	120.0
Employment accessibility	624,940	275,990	14,427	1.347e+06
Total hours worked in household	2,159	913.1	416.1	6,239
Household has company car	0.149	0.356	0	1
Works at single-establishment firm	0.443	0.497	0	1
Number of jobs in household	1.511	0.968	1	18
Person is male	0.521	0.215	0	1
Person is foreigner	0.0718	0.217	0	1
Age of person	41.99	9.008	18	64
Apartment	0.153	0.360	0	1
House built <1945	0.192	0.394	0	1

The number of observations is 10,213,540. For land price and lot size the number of observations is 2,196,280. Because of confidentiality restrictons the minimum and maximum values refer to the 0.01% and 99.99% percentile. This implies that we exclude the bottom and top 1,024 observations.

Table 2 – Baseline reduced-form regression results

(Dependent variable: the log of household gross income)

		+ Housing and job controls	+ Workplace fixed effects	Contemporary Instruments	Historic instruments	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	2SLS	2SLS	2SLS
Pictures per ha *(log)*	0.0215***	0.0285***	0.0166***	**0.0168*****	**0.0333*****	**0.0382*****
	(0.0016)	(0.0013)	(0.0011)	(**0.0023**)	(**0.0037**)	(**0.0037**)
Employment accessibility *(log)*	0.0999***	0.0942***	0.0881***	0.0879***	0.0737***	**0.0526*****
	(0.0043)	(0.0040)	(0.0035)	(0.0038)	(0.0048)	(**0.0100**)
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Housing and job controls	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Workplace fixed effects	No	No	Yes	Yes	Yes	Yes
Number of observations	10,213,540	10,213,540	10,213,524	10,213,524	10,213,524	10,213,524
R^2	0.2041	0.2949	0.3316			
Kleibergen-Paap F-statistic				386.2	238.8	86.04

Notes: **Bold** indicates instrumented. Household controls include household size, mean age of adults, mean gender, household type (couple, single, kids), the share of the household that is foreign-born. Job controls are the total hours worked, whether the household has a company car, the share of full-time contracts, the share of permanent contracts. Housing controls include house type, height of the building, construction year dummies and whether a building is listed. Standard errors are clustered at the neighborhood level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 3 – Reduced form results: alternative proxies for amenities and land prices

(Dependent variable: the log of household gross income)

	Alternative proxies for amenities			Effects on land prices		
	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	OLS	OLS	2SLS
Pictures per ha *(log)*				0.1016***	0.0919***	**0.2061*****
				(0.0038)	(0.0034)	(**0.0138**)
Hedonic amenity index *(log, std)*	**0.0250*****					
	(**0.0028**)					
Pokéstops per ha *(log)*		**0.0396*****				
		(**0.0046**)				
Share historic district			0.2914***			
			(0.0309)			
Employment accessibility *(log)*	**0.0782*****	**0.0580*****	**0.0731*****	0.3586***	0.3343***	**0.2119*****
	(**0.0086**)	(**0.0113**)	(**0.0088**)	(0.0104)	(0.0091)	(**0.0323**)
Number of observations	10,233,115	9,839,819	10,236,308	2,196,280	2,196,280	2,196,280
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Housing and job controls	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Workplace fixed effects	Yes	Yes	Yes	No	Yes	Yes
R^2				0.5564	0.5891	
Kleibergen-Paap F-statistic	29.02	68.45	29.14			75.29

Notes: **Bold** indicates instrumented. Household controls include household size, mean age of adults, mean gender, household type (couple, single, kids), the share of the household that is foreign-born. Job controls are the total hours worked, whether the household has a company car, the share of full-time contracts, the share of permanent contracts. Housing controls include house type, height of the building, construction year dummies and whether a building is listed. Standard errors are clustered at the neighborhood level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4 – Structural estimation

	No instruments	1900 instruments	1832 instruments
	(1)	(2)	(3)
Commuting time elasticity, $\hat{\kappa}$	0.2210***	0.2210***	0.2210***
	(0.0048)	(0.0048)	(0.0048)
Commuting heterogeneity, $\hat{\varepsilon}$	2.7323***	2.7323***	2.7323***
	(0.0144)	(0.0144)	(0.0144)
Land preferences, $\hat{\mu}$	0.0955***	0.0955***	0.0955***
	(0.0003)	(0.0003)	(0.0003)
Amenity preferences, $\hat{\beta}$	0.0404**	0.1559**	0.1062***
	(0.0181)	(0.0712)	(0.0262)
Relative location quality heterogeneity, $\hat{\gamma}$	0.3142***	0.1421***	0.2493***
	(0.0124)	(0.0370)	(0.0408)
Agglomeration elasticity, $\hat{\delta}$	0.0465***	0.0887***	0.0745***
	(0.0016)	(0.0040)	(0.0034)
Location quality heterogeneity, $\hat{\gamma}_{\Delta}$	6.0911***	3.2809***	4.2179***
	(0.2162)	(0.5683)	(0.3531)
Skills heterogeneity, $\hat{\gamma}_s$	23.0909***	16.9213***	6.0911***
	(1.7681)	(1.0818)	(0.2162)
Fixed parameters:			
Minimum lot size, $\bar{h}$	25	25	25
Elasticity of substitution, σ	4	4	4
Number of areas	4,033	4,033	4,033
Number of area pairs	16,265,089	16,265,089	16,265,089

Notes: We estimate the parameters using data at neighborhood level. In column (2) we use as instruments the share of water bodies in 1900 in the neighborhood, the share of built-up land in 1900 in the neighborhood, the share of built-up land in 1900 <500m of the neighborhood, the share of built-up land in 1900 500-1000m, and employment accessibility in 1909. In column (3) we use as instruments the share of water bodies in 1832 in the neighborhood, the share of built-up land in 1832 in the neighborhood, the share of buildings in 1832 in the neighborhood, the share of buildings in 1832 <500m of the neighborhood, the share of buildings in 1832 500-1000m, and population accessibility in 1832. Standard errors are bootstrapped (250 replications) and clustered at the neighborhood level; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

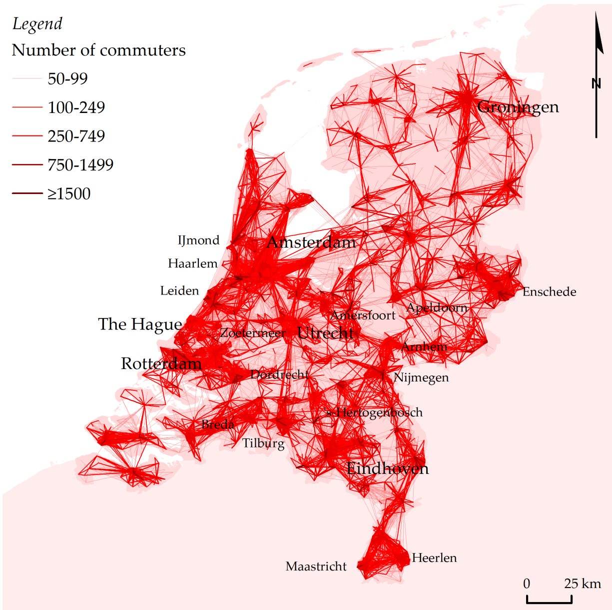
Table 5 – Counterfactual analysis

	Baseline scenario	No amenities scenario
	(1)	(2)
Total output	120,959	133,805
Aggregate land rents	437,039	434,299
Aggregate real income	11,821	12,680
Income mixing, $\bar{\sigma}_x$	0.0472	0.0148

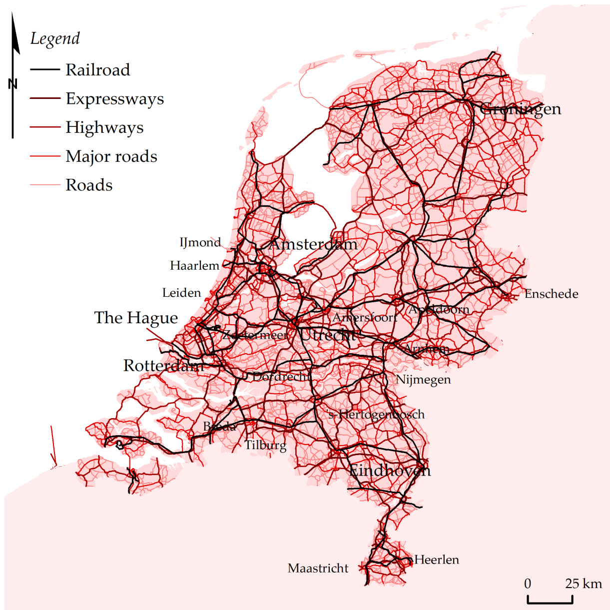
Notes: We calculate aggregate land rents as: $\sum_{x=1}^{\mathcal{L}^*} h_x^c R_x^c$ and aggregate net wages as : $\sum_{x=1}^{\mathcal{L}^*} (1/h_x^c)\omega_x^c t_x^c$. Hence, we weight aggregate labor income by the density in each location.

Figures

(A) Overview map

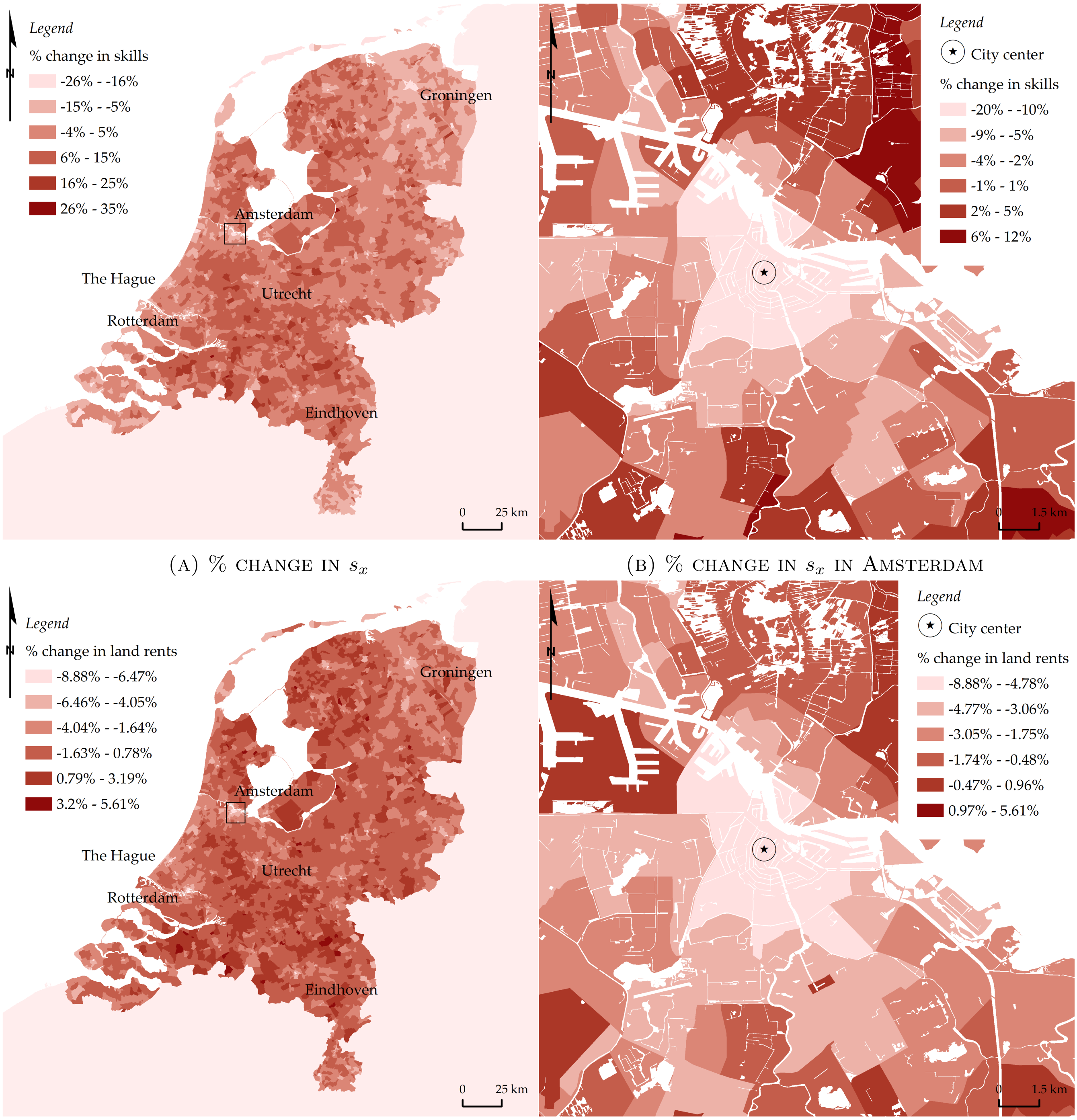


(B) Commuting networks



(C) Transport Network

Figure 1 – The Netherlands



(A) % CHANGE IN s_x

(B) % CHANGE IN s_x IN AMSTERDAM

(C) % CHANGE IN R_x

(D) % CHANGE IN R_x IN AMSTERDAM

FIGURE 2 – COUNTERFACTUAL ANALYSIS: NO AMENITIES

Online Appendix of "Income Sorting Across Space: The Role of Amenities and Commuting Costs"

Carl Gaigné* Hans R.A. Koster† Fabien Moizeau‡
Jacques-François Thisse§

Abstract

In this Appendix we provide derivations of bid rent, a proof of Proposition 1, the derivation of the land rent and the real wage under Stone-Geary preferences. We also provide more detailed information on the data and discuss a wide range of robustness checks for the reduced-form income mapping. Finally, we describe the procedure to obtain the counterfactual values.

Keywords: cities, social stratification, income, amenities, commuting

JEL classification: R14, R23, R53, Z13.

*INRA, UMR1302, SMART-LERECO Rennes (France) and CREATE, Laval University, Quebec (Canada).

†Department of Spatial Economics, Vrije Universiteit Amsterdam (The Netherlands), HSE University (Russian Federation), Tinbergen Institute, and CEPR.

‡CREM (Condorcet Center), Université de Rennes 1.

§CORE-UCLouvain (Belgium), HSE University (Russian Federation) and CEPR.

Appendix A

A.1 The bid rent

Since $\omega(s)$ is strictly increasing in s, we work with the variable $\omega(s)$ and write the equilibrium utility level as follows: $U^*(\omega(s))$. Differentiating (15) w.r.t. x and using (14), we obtain:

$$\Psi_x(x, \omega(s)t(x), U^*(\omega(s))) = \frac{\omega t}{H}\left(\frac{t_x}{t} - \frac{Q_b}{\omega t}b_x\right). \tag{A.1.1}$$

Differentiating (A.1.1) w.r.t. ω and rearranging terms yields the following expression:

$$\begin{aligned}\Psi_{\omega x}(x, \omega(s)t(x), U^*(\omega(s))) = \frac{t}{H}\Bigg\{&\frac{t_x}{t}\left[1 - \frac{\omega}{H}(H_\omega + H_U U^*_\omega)\right]\\ &+\frac{b_x}{t}\left[\frac{H_\omega + H_U U^*_\omega}{H}Q_b - (Q_{bH}(H_\omega + H_U U^*_\omega) + Q_{bU}U^*_\omega)\right]\Bigg\}.\end{aligned} \tag{A.1.2}$$

It is readily verified from (1) that

$$Q(h, U/b(x)) = \left[\frac{1}{(h-\overline{h})^\mu}\frac{U}{b}\right]^{\frac{1}{1-\mu}}, \tag{A.1.3}$$

while it follows from (A.1.3) that

$$\begin{aligned}Q_U &= \frac{1}{1-\mu}U^{\frac{1}{1-\mu}-1}\left[\frac{1}{b(h-\overline{h})^\mu}\right]^{\frac{1}{1-\mu}} = \frac{1}{(1-\mu)}\frac{Q}{U},\\ Q_{Ub} &= -\frac{Q_U}{(1-\mu)b},\\ Q_b &= -\frac{U}{b}Q_U,\\ Q_h &= -\frac{\mu}{1-\mu}\left[\frac{1}{(h-\overline{h})}\frac{U}{b}\right]^{\frac{1}{1-\mu}}\\ Q_{bH} &= \frac{U}{b}\frac{\mu}{1-\mu}(h-\overline{h})^{-1}Q_U.\end{aligned}$$

Plugging Q_b, Q_{bH} and Q_{bU} into (A.1.2) and rearranging terms leads to

$$\begin{aligned}\Psi_{\omega x}(x, \omega(s)t(x), U^*(\omega(s))) = \frac{t}{H}\Bigg\{&\frac{t_x}{t}\left[1 - \frac{\omega}{H}(H_\omega + H_U U^*_\omega)\right]\\ &+\frac{b_x}{b}\left[\frac{H_\omega + H_U U^*_\omega}{H}\left(-\frac{U}{t}Q_U\right)\left(\frac{h-(1-\mu)\overline{h}}{(1-\mu)(h-\overline{h})}\right)\right.\\ &\left.+\frac{Q_U}{(1-\mu)t}U^*_\omega\right]\Bigg\}.\end{aligned} \tag{A.1.4}$$

Plugging Q_h and Q in (14) and solving the corresponding equation yields

$$\frac{h-(1-\mu)\overline{h}}{(1-\mu)(h-\overline{h})} = \omega t \left[\frac{b}{U}(h-\overline{h})^{\mu}\right]^{\frac{1}{1-\mu}}. \tag{A.1.5}$$

Using Q_U, (A.1.5) may be rewritten as follows:

$$\left(-\frac{U}{t}Q_U\right)\left[\frac{h-(1-\mu)\overline{h}}{(1-\mu)(h-\overline{h})}\right] = -\frac{\omega}{1-\mu}. \tag{A.1.6}$$

Differentiating (15) with respect to ω and using (14), we obtain:

$$\Psi_\omega(x,\omega(s)t(x),U^*(\omega(s))) = \frac{t}{H}\left(1-\frac{Q_U}{t}U^*_\omega\right), \tag{A.1.7}$$

which is equal to 0 if and only if

$$U^*_\omega = \frac{t}{Q_U}. \tag{A.1.8}$$

Plugging (A.1.6) and (A.1.8) in (A.1.4) yields

$$\Psi_{\omega x}(x,\omega(s)t(x),U^*(\omega(s))) = \frac{t}{H}\cdot\left[1-\frac{\omega}{H}(H_\omega+H_U U^*_\omega)\right]\cdot\frac{1}{1-\mu}\cdot\left[(1-\mu)\frac{t_x}{t}+\frac{b_x}{b}\right]. \tag{A.1.9}$$

Applying the implicit function theorem to (A.1.5) yields

$$H_U = \frac{(h-(1-\mu)\overline{h})(h-\overline{h})}{U\mu h}$$

and

$$H_\omega = -\frac{t(1-\mu)^2}{\mu h}U^{-\frac{1}{1-\mu}}b^{\frac{1}{1-\mu}}(h-\overline{h})^{1+\frac{1}{1-\mu}}.$$

Given Q_U, (A.1.8) is equivalent to

$$U^*_\omega = t\cdot(1-\mu)\left[b\cdot(h-\overline{h})^{\mu}\right]^{\frac{1}{1-\mu}}(U^*(\omega))^{-\frac{\mu}{1-\mu}}. \tag{A.1.10}$$

Using the above three expressions, we obtain:

$$H_\omega + H_U U^*_\omega = t\cdot(1-\mu)(h-\overline{h})\left[\frac{b}{U^*(\omega)}(h-\overline{h})^{\mu}\right]^{\frac{1}{1-\mu}}.$$

Therefore, by implication of (A.1.5), we have:

$$1-\frac{\omega}{H}(H_\omega+H_U U^*_\omega) = \frac{(1-\mu)\overline{h}}{H}. \tag{A.1.11}$$

Substituting this expression into (A.1.9) yields the desired expression:

$$\Psi_{\omega x}(x, \omega(s)t(x), U^*(\omega(s))) = \frac{t}{H} \cdot \frac{\overline{h}}{H} \cdot [B(x) - (1-\mu)T(x)] \, .$$

A.2 Proof of Proposition 1

The proof involves six steps.

***(i)* The bid-max lot size.** From the definition of the location-quality index given by (18), (A.1.5) can be rewritten as follows:

$$\frac{H - (1-\mu)\overline{h}}{(1-\mu)(H - \overline{h})} = \omega \Delta^{\frac{1}{1-\mu}} \left[\frac{(H - \overline{h})^{\mu}}{U} \right]^{\frac{1}{1-\mu}} , \tag{A.2.1}$$

which implies $H(\omega(s)t(x), U/b(x)) \equiv H(\Delta(x), \omega(s), U)$ so that the bid-max lot size depends on $b(x)$ and $t(x)$ only through the location-quality index $\Delta(x)$.

The LHS of (A.2.1) is decreasing and tends to $+\infty$ when $H \to \overline{h}$ and to $1/(1-\mu) > 0$ when $H \to +\infty$. The RHS of (A.2.1) is increasing in H. It tends to 0 when $H \to \overline{h}$ and to $+\infty$ when $H \to +\infty$. Therefore, (A.2.1), equivalently (14), has a single solution $H(\omega t(x), U/b(x))$, which implies that the housing demand is uniquely determined.

Applying the implicit function theorem to (A.2.1) yields

$$\frac{\partial H}{\partial \Delta} = - \left[U^{\frac{1}{1-\mu}} (H - \overline{h})^{-\frac{1}{1-\mu} - 1} \frac{\mu H}{(1-\mu)} \right]^{-1} \omega \Delta^{\frac{\mu}{1-\mu}} < 0. \tag{A.2.2}$$

***(ii)* Equilibrium utility level.** Using the definition of the location-quality index, (A.1.10) implies that the equilibrium utility level is a solution to the differential equation in U^*:

$$U^*_{\omega} = \Delta^{\frac{1}{1-\mu}} (1-\mu)(H - \overline{h})^{\frac{\mu}{1-\mu}} \left(U^*(\omega)\right)^{-\frac{\mu}{1-\mu}} , \tag{A.2.3}$$

so that $U^*(\omega)$ depends on Δ only.

***(iii)* Supermodularity of the equilibrium utility level.** Differentiating (A.2.3) w.r.t. Δ, we obtain:

$$\frac{\partial}{\partial \Delta} \frac{\mathrm{d}U^*}{\mathrm{d}\omega} = \Delta^{\frac{\mu}{1-\mu}} (H - \overline{h})^{\frac{\mu}{1-\mu}} (U^*(\omega))^{-\frac{\mu}{1-\mu}} \cdot \left[1 + \mu \Delta (H - \overline{h})^{-1} \frac{\partial H}{\partial \Delta} \right] .$$

Using (A.2.2), this expression may be rewritten as follows:

$$\frac{\partial}{\partial \Delta} \frac{\mathrm{d}U^*}{\mathrm{d}\omega} = \Delta^{\frac{\mu}{1-\mu}} (H - \overline{h})^{\frac{\mu}{1-\mu}} (U^*(\omega))^{-\frac{\mu}{1-\mu}} \cdot \left[1 - (H - \overline{h})^{\frac{1}{1-\mu}} \frac{(1-\mu)\omega \Delta^{\frac{1}{1-\mu}}}{(U^*(\omega))^{\frac{1}{1-\mu}} H} \right] .$$

From (A.2.1), the expression in the bracketed term is equivalent to

$$1-(H-\overline{h})^{\frac{1}{1-\mu}}\frac{(1-\mu)\omega\Delta^{\frac{1}{1-\mu}}}{(U^{*}(\omega))^{\frac{1}{1-\mu}}H}=(1-\mu)\frac{\overline{h}}{h}>0.$$

Therefore,

$$\frac{\partial}{\partial\Delta}\frac{\mathrm{d}U^{*}}{\mathrm{d}s}=\frac{\partial}{\partial\Delta}\frac{\mathrm{d}U^{*}}{\mathrm{d}\omega}\frac{\mathrm{d}\omega}{\mathrm{d}s}>0.$$

The Spence-Mirrlees condition thus holds, which implies the existence of a positive assortative matching between skills and the values of the location-quality index. In other words, there is a unique one-to-one and increasing relationship between s and Δ (Chiappori, 2017). Regardless of the value of $Y>0$, households ordered by increasing skills must be assigned to locations endowed with rising values of the location-quality index. Since a single value of Δ is associated with x, a unique value of s must be associated with x. Therefore, the equilibrium conditions (16) has a unique solution, which means that $s^{*}(x)$ is a mapping.

Note that the supermodularity of $U^{*}(\omega)$ is equivalent to the inequality $\Psi_{\omega\Delta}>0$. Indeed, differentiating (A.1.7) w.r.t. Δ and using (A.1.8) yield:

$$\begin{aligned}\Psi_{\omega\Delta}(x,\omega(s)t(x),U^{*}(\omega(s)))|_{\Psi_{\omega}=0} &= \frac{t}{H}\left[\frac{\partial(t/Q_{U})/\partial\Delta}{U_{\omega}^{*}}\right]\\ &= \frac{t}{H}\frac{\partial U_{\omega}^{*}/\partial\Delta}{U_{\omega}^{*}}>0.\end{aligned}$$

***(iv)* Uniqueness of the equilibrium shares** $\zeta^{*}(x,s^{*}(x))$. The proof follows Montesano (1972). Assume that there are $m\geq 2$ points $x_1\neq x_2...\neq x_m$ exist such that $\Delta(x_1)=\Delta(x_j)$ for $j=2,...,m$. Using (12) and Step (i), we have:

$$|\zeta(x_j,s)f(s^{*}(x_j))H\left\{\Delta(x_j),\omega(s^{*}(x_j)),U^{*}\left[\omega(s^{*}(x_j))\right]\right\}\mathrm{d}s|=\mathrm{d}x \qquad j=1,...,m. \tag{A.2.4}$$

The supermodularity of U^{*} w.r.t. Δ implies $s^{*}(x_1)=s^{*}(x_j)$, so that $f(s^{*}(x_1))=f(s^{*}(x_j))$ for $j=2,...,m$. In other words,

$$H\left\{\Delta(x_1),\omega(s^{*}(x_1)),U^{*}\left[\omega(s^{*}(x_1))\right]\right\}=H\left\{\Delta(x_j),\omega(s^{*}(x_j)),U^{*}\left[\omega(s^{*}(x_j))\right]\right\}.$$

Hence, dividing relationships (A.2.4) between themselves leads to

$$\frac{\zeta(x_1,s^{*}(x_1))}{\zeta(x_k,s^{*}(x_k))}=1 \qquad k=2,...,m.$$

Since $\sum_{j=1}^{m}\zeta(x_j,s)=1$, we obtain:

$$\zeta^*(x_j,s^*(x_j))=\frac{1}{m} \qquad j=1,...,m.$$

Since $b(x)$ and $t^*(x)$ are never constant on a nonegligible subset of N, we may assume that there is an integer M such that $m\leq M$.

(v) **Existence of a spatial equilibrium.** Since $t(x)$ is the mean of the upper envelop of n continuous functions of L_i, the mapping $s^*(x,\mathbf{L})$ is continuous in $\mathbf{L}$. Let $N_m\subset N$ be the set of locations, which may be empty or negligible, such that $\zeta(x,s^*(x,\mathbf{L}))=1/m$ for $x\in N_m$ and $\cup_{m=1}^{M}N_m=N$ for $m=1,,..,M$ (up to a negligible set).

Let Z be the number of arcs a_z of the network N. The set N_m is the union of a finite number of subarcs; a subarc of a_z links the points $\alpha_z^m\in a_z$ and $\beta_z^m\in a_z$. For notational simplicity, we assume that each arc a_z includes at most one subarc (α_z^m,β_z^m) in N_m (otherwise a third summation over the subarcs of a_z included in N_m is needed).

Hence, the function

$$\begin{aligned}\mathbb{F}_i(\mathbf{L})&\equiv L_i^\delta, \qquad i=1,...,n\\&=\left\{\sum_{m=1}^{M}\frac{1}{m}\sum_{z=1}^{Z}\int_{\alpha_z^m}^{\beta_z^m}f[s^*(x;\mathbf{L})]\frac{K_i[t_i(x,L_i)]^\varepsilon}{\sum_{j=1}^{n}K_j\left[t_j(x,L_j)\right)]^\varepsilon}\mathrm{d}x\right\}^\delta\end{aligned} \tag{A.2.5}$$

is also continuous in $\mathbf{L}$.

Since $\mathbb{F}(\mathbf{L})\equiv(\mathbb{F}_1(\mathbf{L}),...,\mathbb{F}_n(\mathbf{L}))$ is a continuous mapping from the simplex

$$\mathbf{S}_n\equiv\{\mathbf{L};L_1\geq 0,..,L_n\geq 0 \text{ and } \Sigma_{i=1}^{n}L_i=1\}$$

into itself, $\mathbb{F}(\mathbf{L})$ has at least one fixed point $\mathbf{L}^*$.

(vi) **Uniqueness.** We determine a sufficient condition for the spatial equilibrium to be unique. It is well known that uniqueness holds when the function $\mathbb{F}(\mathbf{L})$ given by (A.2.5) is a contraction. This is so when the matrix norm $||\cdot||_\infty$ of the Jacobian $J(\mathbb{F})$ of $\mathbb{F}(\mathbf{L})$ is smaller than 1. The function $\mathbb{F}(\mathbf{L})$ is differentiable everywhere but over the negligible set $\mathbf{I}$ of $\mathbf{L}\subset\mathbf{S}_m$ such that $t_i(x)=t_j(x)$ or $L_i^\delta/L_j^\delta=\mathbb{A}_j\ell_j(x)/\mathbb{A}_i\ell_i(x)$ for $i\neq j$ since the function $\ell_i(x)$ is strictly increasing in the distance $d(x,i)$ for $i=1,...,n$.

Differentiating (A.2.5) with respect to L_k yields the following expression defined on the interior of $\mathbf{S}_n - \mathbf{I}$:

$$\begin{aligned}\frac{\partial \mathbb{F}_i(\mathbf{L})}{\partial L_k} &= \delta \left\{ \sum_{m=1}^{M} \frac{1}{m} \sum_{z=1}^{Z} \int_{\alpha_z^m}^{\beta_z^m} f[s^*(x;\mathbf{L})] \frac{K_i[t_i(x,L_i)]^\varepsilon}{\sum_{j=1}^n K_j\left[t_j(x,L_j))\right]^\varepsilon} \mathrm{d}x \right\}^{\delta-1} \\ &\cdot \left\{ \sum_{m=1}^{M} \frac{1}{m} \sum_{z=1}^{Z} f[s^*(\beta_z^m(\mathbf{L});\mathbf{L})] \frac{K_i[t_i(\beta_z^m(\mathbf{L}),L_i)]^\varepsilon}{\sum_{j=1}^n K_j\left[t_j(\beta_z^m(\mathbf{L}),L_j)\right]^\varepsilon} \cdot \frac{\partial \beta_z^m(\mathbf{L})}{\partial L_k} \right. \\ &- \sum_{m=1}^{M} \frac{1}{m} \sum_{z=1}^{Z} f[s^*(\alpha_z^m(\mathbf{L});\mathbf{L})] \frac{K_i[t_i(\alpha_z^m(\mathbf{L}),L_i)]^\varepsilon}{\sum_{j=1}^n K_j\left[t_j(\alpha_z^m(\mathbf{L}),L_j)\right]^\varepsilon} \cdot \frac{\partial \alpha_z^m(\mathbf{L})}{\partial L_k} \\ &\left. + \sum_{m=1}^{M} \frac{1}{m} \sum_{z=1}^{Z} \int_{\alpha_z^m}^{\beta_z^m} \frac{\partial}{\partial L_k} \left[f(s^*(x;\mathbf{L})) \frac{K_i[t_i(x,L_i)]^\varepsilon}{\sum_{j=1}^n K_j\left[t_j(x,L_j))\right]^\varepsilon} \right] \mathrm{d}x \right\}, \qquad i,k = 1,...,n.\end{aligned}$$

Since all the terms in the right-hand side of this expression are continuous on $\mathbf{S}_n - \mathbf{I}$, $\partial \mathbb{F}_i(\mathbf{L})/\partial L_k$ has a supremum $C_{ik} \neq 0$. Therefore, we have:

$$\sum_{k=1}^{n} \left| \frac{\partial \mathbb{F}_i(\mathbf{L})}{\partial L_k} \right| < \delta \sum_{k=1}^{n} |C_{ik}| < 1,$$

where the second inequality holds for all $0 < \delta < \delta_i \equiv 1/(\Sigma_k |C_{ik}|)$. Let $\delta_{\min}$ be the minimum of δ_i over $i = 1,...,n$. If $\delta < \delta_{\min}$, $||J(\mathbb{F})||_\infty$ is smaller than 1. In other words, when $\delta > 0$ is small enough, $\mathbb{F}(\mathbf{L})$ is a contraction.

A.3 The equilibrium land rent under Fréchet distributions

Using (14), we may rewrite (15) as follows:

$$\Psi(x, \omega t, U) = -Q_H(H, U/b(x)).$$

Using Q_H leads to

$$\Psi(x, \omega t, U) = \frac{\mu}{1-\mu}(H - \overline{h})^{\frac{-1}{1-\mu}} \left[\frac{U}{b}\right]^{\frac{1}{1-\mu}}. \tag{A.3.1}$$

Rearranging terms in (3) yields:

$$H - \overline{h} = \mu \left[\frac{\omega t}{\Psi(x,\omega,U)} - \overline{h} \right] \tag{A.3.2}$$

and plugging the above expression into (A.3.1) leads to

$$\Psi(x, \omega, U) = \mu^{-\frac{\mu}{1-\mu}} (1-\mu)^{-1} \left[\frac{\omega t}{\Psi(x,\omega,U)} - \overline{h} \right]^{\frac{-1}{1-\mu}} \left[\frac{U(\omega)}{b} \right]^{\frac{1}{1-\mu}}.$$

Dividing this expression by $t(x)$ and setting $\Phi \equiv \Psi/t$, we get

$$\Phi = \mu^{-\frac{\mu}{1-\mu}}(1-\mu)^{-1}\left(\frac{\omega}{\Phi}-\overline{h}\right)^{-\frac{1}{1-\mu}}[U(\omega)]^{\frac{1}{1-\mu}}(\Delta^*)^{\frac{-1}{1-\mu}}.$$

Rearranging terms, this expression becomes:

$$\Phi = \mu(1-\mu)^{\frac{1-\mu}{\mu}}\left(\omega-\Phi\overline{h}\right)^{\frac{1}{\mu}}[U(\omega)]^{-\frac{1}{\mu}}(\Delta^*)^{\frac{1}{\mu}}. \tag{A.3.3}$$

Applying the first-order condition to Φ yields the following differential equation in ω:

$$U^*_\omega(\omega) = \frac{1}{\omega-\Phi\overline{h}}U^*(\omega).$$

Let

$$U^*(\omega) = \left(\omega-\Phi\overline{h}\right)X(\omega) \tag{A.3.4}$$

be a solution to the above differential equation where $X(\omega)$ is determined below. Differentiating (A.3.4) with respect to ω, we obtain

$$U_\omega(\omega) = \left[\frac{1}{\omega-\Phi\overline{h}} - \frac{\overline{h}}{\omega-\Phi\overline{h}}\Phi_\omega + \frac{X_\omega(\omega)}{X(\omega)}\right]U(\omega).$$

Totally differentiating Φ leads to

$$\Phi_\omega \equiv \frac{\mathrm{d}\Phi}{\mathrm{d}\omega} = \frac{\partial\Phi}{\partial\omega} + \Phi_\Delta\Delta_\omega = \Phi_\Delta\Delta_\omega. \tag{A.3.5}$$

Differentiating (A.3.3) with respect to Δ yields:

$$\Phi_\Delta = \Phi\left[\frac{1}{\mu}(\Delta^*)^{-1} - \frac{1}{\mu}\Phi_\Delta\overline{h}\left(\omega-\Phi\overline{h}\right)^{-1}\right],$$

whose solution in Φ_Δ is

$$\Phi_\Delta = \frac{1}{\Delta^*}\frac{\Phi}{\mu}\left[\frac{\mu(\omega-\Phi\overline{h})}{\mu(\omega-\Phi\overline{h})+\overline{h}\Phi}\right].$$

Therefore, we may rewrite (A.3.2) as follows:

$$H\Phi = \mu(\omega-\Phi\overline{h}) + \overline{h}\Phi. \tag{A.3.6}$$

Plugging (A.3.6) into Φ_Δ leads to

$$\Phi_\Delta = \frac{\omega-\Phi\overline{h}}{\Delta^* H}.$$

Using Φ_ω and Δ_ω, (A.3.5) becomes:

$$\Phi_\omega = \Phi_\Delta \Delta^*_\omega = \frac{1}{\gamma}\frac{\omega - \Phi\overline{h}}{\omega H} = \frac{1}{\gamma\mu}\frac{(H-\overline{h})\Phi}{\omega H} > 0.$$

Since $U_\omega(\omega)/U(\omega)$ is equal to $1/(\omega - \Phi\overline{h})$ in equilibrium, it must be that

$$\frac{X_\omega(\omega)}{X(\omega)} = \frac{\overline{h}}{\omega - \Phi\overline{h}}\Phi_\omega = \frac{\overline{h}}{\omega - \Phi\overline{h}}\frac{1}{\gamma\mu}\frac{(H-\overline{h})\Phi}{\omega H}.$$

Therefore, using (A.3.6) leads to the following differential equation in ω:

$$X_\omega(\omega) = \frac{1}{\gamma}\frac{\overline{h}}{\omega H}X(\omega),$$

whose solution is

$$X(\omega) = k\left(\frac{\omega}{H}\right)^{\frac{\beta}{1-\mu}}, \tag{A.3.7}$$

where $k > 0$ is the constant of integration. Indeed, differentiating the above equation with respect to ω leads to

$$X_\omega(\omega) = \frac{1}{(1-\mu)\gamma}\frac{H - \omega(H_\omega + H^*_U U_\omega)}{H^2}\frac{H}{\omega}X(\omega).$$

Using (A.1.11), we obtain:

$$X_\omega(\omega) = \frac{1}{(1-\mu)\gamma}\frac{(1-\mu)\overline{h}}{H}\frac{1}{\omega}X(\omega) = \frac{1}{\gamma}\frac{\overline{h}}{\omega H}X(\omega).$$

Substituting (A.3.7) into (A.3.4) yields:

$$U(\omega) = \left(\omega - \Phi\overline{h}\right)k\left(\frac{\omega}{H}\right)^{\frac{1}{(1-\mu)\gamma}}.$$

Plugging this expression into (A.3.3) and rearranging terms, we obtain the following implicit solution for the equilibrium land rent:

$$R^*(x) = \mu(1-\mu)^{\frac{1-\mu}{\mu}}k^{-\frac{1}{\mu}}t(x)(\Delta^*(x))^{\frac{1}{\mu}}\left[\frac{\mu t(x)}{R^*(x)} + \frac{(1-\mu)\overline{h}}{K^{1/\gamma_s}\left[\Delta^*(x)\right]^\gamma (Y^*)^{1/\sigma}}\right]^{\frac{1}{(1-\mu)\mu\gamma}}. \tag{A.3.8}$$

Since the RHS of (A.3.8) is strictly decreasing and tends to 0 (∞) when $R(x) \to \infty$ (0), (A.3.8) has a unique solution in $R^*(x)$.

The lowest income in the sample, denoted by $\underline{\omega}$, is strictly positive. It follows from (21) that the lowest location-quality index associated with the poorest household is given by

$$\underline{\Delta} = \left(\frac{K_S}{K_\Delta}\right)^{-\gamma/\gamma_S} (Y^*)^{-\gamma/\sigma} (\underline{\omega})^{1/\gamma} > 0.$$

The constant k may be obtained by evaluating $R^*(x)$ at the least enjoyable location $\underline{x}$ where $\Delta^*(x)$ reached its minimum $\underline{\Delta}$. We assume that $\underline{x}$ is unique. Furthermore, the land rent at $\underline{x}$ is equal to the opportunity cost of land, R_0. Therefore, it is readily verified that k is given by

$$k^{-\frac{1}{\mu}} = R_0 \mu^{-1} (1-\mu)^{-\frac{1-\mu}{\mu}} \left[t(\underline{x})\right]^{-1} \underline{\Delta}^{-\frac{1}{\mu}} \left[\frac{\mu t(\underline{x})}{R_0} + \frac{(1-\mu)\overline{h}}{\underline{\omega}}\right]^{\frac{-1}{(1-\mu)\mu\gamma}}.$$

Plugging this expression into (A.3.8) yields the equilibrium land rent at x:

$$R^*(x) = R_0 \frac{t(x)}{t(\underline{x})} \left[\frac{\Delta^*(x)}{\underline{\Delta}}\right]^{\frac{1}{\mu}} \left[\frac{\mu \frac{t(x)}{R^*(x)} + (1-\mu)\frac{\overline{h}}{\omega^*(x)}}{\mu \frac{t(\underline{x})}{R_0} + (1-\mu)\frac{\overline{h}}{\underline{\omega}}}\right]^{\frac{1}{(1-\mu)\mu\gamma}}.$$

Note that this expression captures several effects: the commuting costs at x and $\underline{x}$, the location-quality index at x and $\underline{x}$, and the mapping $\omega^*(x)$.

A.4 The real wage under Stone-Geary preferences

With a Stone-Geary utility function, we have $U = b \cdot u(q, h)$

$$u = (1-\mu)^{-(1-\mu)} \mu^{-\mu} q^{1-\mu} \left(h - \overline{h}\right)^{\mu} \tag{A.4.1}$$

and the budget constraint is given by $q + Rh = \omega t$. The price index under Stone-Geary preferences is given by

$$P = R^\mu \frac{\omega t}{\omega t - R\overline{h}}.$$

Proof. Inserting the equilibrium consumption of numéraire and housing in (A.4.1) yields the indirect utility of consumption:

$$\begin{aligned} u^* &= (\omega t - R\overline{h}) R^{-\mu} \\ &= \omega t R^{-\mu} \frac{\omega t - R\overline{h}}{\omega t}. \end{aligned}$$

Hence, total expenditures are given by

$$\omega t = u^* R^\mu \frac{\omega t}{\omega t - R\overline{h}}$$

so that the price index is

$$P = R^{\mu} \frac{\omega t}{\omega t - R\overline{h}}.$$

Because the Stone-Geary utility function is non homogeneous, the price index P depends on income and varies across individuals.■

Hence, the real wage is given by

$$\frac{\omega t}{P} = \frac{\omega t - R\overline{h}}{R^{\mu}}.$$

Appendix B

In this appendix, we first pay attention to the construction of the various datasets. In Appendix B.1 we elaborate on how we calculate network distances and show the relationship with Euclidian distance. Appendix B.2 continues by explaining how we measure land prices and lot sizes for all locations. This is followed in B.3 by more information on our proxies for amenities: the picture index and the construction of the hedonic amenity index. In Appendix B.4 we introduce the historical data based on 1900 land use maps and the 1832 Census. Appendix B.5 reports distributions of the variables of interest.

The second part of this appendix reports various additional econometric results. First, we report bias-corrected estimates using Oster's (2019) methodology in Appendix B.6. Second, first-stage results in Appendix B.7. We undertake additional robustness checks in Appendix B.8. Appendix B.9 discusses the outcomes of alternative specifications of the gravity model. In Appendix B.10 we outline the procedure to solve for counterfactual outcomes of the model.

B.1 Commuting and travel times

To estimate the commuting time for each household, we use the tax register information, which provides information on individual jobs and the number of hours worked in each firm for each year. From the *ABR Regio* dataset, we get information on all firms which provide information on each establishment in the Netherlands, such as its exact location, the industrial sector, and the *estimated* number of employees in each establishment. To avoid miscoding and to exclude employment agencies (where people do not actually work), we exclude firms with more than 10 thousand employees. Since we do not know the exact establishment, only the firm, people work for, we assume that they work at the nearest establishment of the firm. This assumption may be problematic for firms having a large number of establishments (e.g., supermarkets or large banks). Therefore, we keep only firms with a maximum of 15 establishments throughout the Netherlands. As many such firms have establishments in different cities, it is reasonable to assume that people work in the nearest establishment.[13] Overall, we are left with 95% of firms.

We first calculate the commuting time from each home location x to each job location i for each year. Then, we determine the commuting time of each household by computing the average commuting time of each adult household member weighted by the number of hours (s)he worked. To calculate the travel time (as well as the time to travel to amenities) we obtain information on the street network from *SpinLab*, which provides information on average free-flow speeds per short road segment (the median length of a segment is 96m), which are usually lower than the speed limit.

The dataset from *SpinLab* provides information on actual free-flow driving speeds for every major street in the Netherlands. The actual speeds are usually well below the free-flow driving speeds, due to traffic lights, roundabouts and intersections. For each neighborhood we calculate the straight-line distance to the nearest access points on the network and then calculate the network distance. The median distance from an observation in the dataset to the nearest access point of the network is 195m (the average is 326m). We assume that the average speed to get to the nearest access points is 10km/h. This is the speed based on the Euclidian distance; in reality the distance to get to the network will be higher because streets are usually curved. Hence, the assumption of 10km/h seems reasonable as the minimum speed on roads in the network is 20km/h. Furthermore, because of the dominance of the bicycle, this would be close to the average cycling speed. Using these information, we calculate the total driving time, which is the sum of the driving time to access the network, the network driving time and the time it takes from the network to arrive at the destination. Alternatively, we calculate for each location pair the Euclidian distance and assume again an average speed of 10km/h.

We also calculate the travel time using the train, using a similar approach. The median distance of each centroid to the nearest station is 5.25km. We then choose the minimum of the travel time over the road, using the train or taking the Euclidian travel time.

[Figure B.1 about here]

The correlation between travel time and Euclidian distance is modest ($\rho = 0.643$). For short distances (< 10km) the correlation is, however, much higher ($\rho = 0.862$). We plot the relationship between distance and travel time in Figure B.1.A. This relationship is monotonic. Figure B.1.B shows the share of commuting people who travel at most τ minutes, which we use to calculate employment accessibility in 1900.

[13] Alternatively, we could consider a distance-decay average of distances to the firm's establishments. Instead, we test robustness by keeping households which have only one working-member who works during the whole year in a single-establishment firm leading to nearly identical results.

B.2 Land prices and lot sizes

Information on land values and lot sizes is not directly available but may be inferred from data on home sales. We use information on home sales from *NVM* (The Dutch Association of Realtors), which comprises the large majority (about 75%) of owner-occupied house transactions between 2003 and 2017. We know the transaction price, the lot size, inside floor space size (both in m^2), the exact address, and a wide range of housing attributes such as house type, number of rooms, construction year, garden, state of maintenance, and whether a house is equipped with central heating. We make some selections to make sure that our results are not driven by outliers. First, we exclude transactions with prices that are above €1 million or below €25,000 and have a price per square meter which is above €5,000 or below €500. We also leave out transactions that refer to properties that are larger than 250m^2 or smaller than 25m^2, or have lot sizes above 5000m^2. These selections consist of less than 1% of the data and do not influence our results. We follow a similar procedure as Rossi-Hansberg *et al.* (2010), implying that we can only use information on residential properties with land. We are left with 1,337,445 housing transactions.

Let $\mathcal{P}(x)$ denote the house price in year y, $H(\tilde{x})$ the observed lot size and $C(\tilde{x})$ the housing characteristics of property $\tilde{x}$. The log land rent $R(x)$ is equal to the fixed effects at the level of the postcode (about 15-20 addresses), which we denote by $\varsigma(x)$, while $\vartheta(y)$ denote year y fixed effects. For each city, we estimate:

$$\log \frac{\mathcal{P}(\tilde{x}, y)}{H(\tilde{x}, y)} = \eta_1 C(\tilde{x}, y) + \varsigma(x) + \vartheta(y) + \epsilon(\tilde{x}, y), \tag{C.2.1}$$

where $\epsilon(\tilde{x}, y)$ is an identically and independently distributed error term that is assumed to be uncorrelated to land rents and housing characteristics, while η_1 are parameters to be estimated. As $\log R_x$ are given by the very local fixed effects, we do not impose any structure on how land rents R_x vary across locations. For about 80% of the data we do not observe land prices directly, because either there were no multiple sales in our study period or because there is no owner-occupied housing in the respective postcode. We therefore also estimate the above equation with neighborhood fixed effects instead.

[Tables B.1 and B.2 about here]

Descriptive statistics for the housing sample are reported in Table B.1. Coefficients η_1 related to the housing attributes are reported in Table B.2. It appears that the house price per square meter of land is generally a bit lower when the property is larger. However, the house price per square meter of land of properties that are (semi-)detached is generally higher. Furthermore, when the maintenance state of a property is good, prices are about $502/1269 = 40\%$ higher. When a property has central heating, the price per square meter is about 5.1% higher. The dummies related to the construction decades show the expected signs. Properties constructed after World War II until 1970 generally have lower prices because this is a period associated with a lower building quality. Lot sizes are inversely related to pattern of land prices ($\rho = -0.245$). In other words, more expensive locations generally have smaller lots, which makes sense.

B.3 Amenities

Hedonic amenity index. We test whether our results are robust to using an alternative hedonic amenity index, rather than relying on geocoded pictures. Following Lee and Lin (2018), we construct an aggregate amenity index that describes the amenity level in every neighborhood x.[14] We will make a distinction between *historic* amenities and *natural* amenities.

Let $\mathcal{A}(\tilde{x})$ be a set of variables that describe amenities of property $\tilde{x}$ (so the location is more detailed than the neighborhood x). For example, we calculate the share of historic districts, the number of listed buildings, water bodies and open space within 500m of each property. Let $\mathcal{P}(\tilde{x}, y)$ the house price, while $C(\tilde{x}, y)$ are housing characteristics at location $\tilde{x}$, and $\vartheta(y)$ are year y fixed effects. We also include neighborhood fixed effects $\varsigma(x)$, so we identify the effects of amenities on prices *within* neighborhoods. We then estimate:

$$\log \mathcal{P}(\tilde{x}, y) = \eta_0 \mathcal{A}(\tilde{x}) + \eta_1 C(\tilde{x}, y) + \vartheta(y) + \varsigma(x) + \epsilon(\tilde{x}, y), \tag{B.3.1}$$

where η_0 and η_1 are parameters to be estimated and $\epsilon(\tilde{x}, y)$ is an identically and independently distributed error term. We then use $\widehat{\eta}_0$ and $\mathcal{A}(\tilde{x})$ to predict the amenity level in each location x in the Randstad:

$$\tilde{b}(x) = \frac{1}{N(x)} \sum_{\tilde{x}=1}^{N_x} \widehat{\eta}_0 \mathcal{A}(\tilde{x}), \tag{B.3.2}$$

where $\tilde{b}_x$ is the (alternative) amenity value at x and $N(x)$ are the number of observations in neighborhood x. Hence, we take the mean amenity value within neighborhoods x.

We use data on the universe of housing transactions in the Netherlands between 2010 and 2015 from the *NVM*. Additional descriptive statistics of the *NVM* data are reported in Table B.3. We have $695,709$ observations and the average house price is €229 thousand.

[Tables B.3 and B.4 about here]

In Table B.4 we report the results of the regression of equation (B.3.1). We first investigate the

[14] Albouy (2016) uses information on wages and housing costs to infer the level of amenities for U.S. cities. However, his approach is not applicable here because we are also interested in *intra*-city variation in amenities. Using Albouy's proxy for amenities could capture the sorting of rich households in certain locations, but this is exactly the relationship we aim to test.

impact of listed buildings. It can be seen that the share of historic districts leads to higher price. A 10 percentage point increase in the share of land designated as historic district increases prices by 1.8%. Listed buildings do have a small additional effect of 0.5% per listed building. In column (2) we investigate the impact of water bodies and open space. For a 10 percentage point increase in water bodies, prices rise by 3%. Moreover, a 10 percentage point increase in open space implies a price increases of 0.6%, so this effect is considerably smaller. When we put historic amenities and natural amenities together, the coefficients are essentially unaffected. We consider this as the preferred specification. In the last specification we investigate whether the results change when we include endogenous amenities, such as shops, cafés, and leisure establishments. This appears not to be the case. Only hotels restaurants and cafés have a statistically significant impact on prices, which suggests that exogenous amenities related to the built environment and land use are more important than endogenous amenities.

Pokémon amenity index. Pokémon was a hugely popular game in 2017. The game could be played at certain places of interest, the so-called 'Pokéstops'.[15] The locations of Pokéstops were determined in the geolocation game by *Ingress*. The developers then chose some of the first portals based on sites with historic or cultural significance, such as The Washington Monument, Big Ben, or museums. Other locations were chosen based on geotagged photos from Goggle. Many more portals were submitted as suggestions by *Ingress* players. There were approximately 15 million player-submitted portal locations, 5 million of which have been approved. In other words, these Pokéstops are not randomly located across space and signify locational attractiveness. We construct the Pokémon Go amenity index by using the density of Pokéstops in a neighborhood.

B.4 Historic data

To instrument current amenity levels and commuting time we use information on land use, the railway network and amenities in 1900. For the 1900 land use maps, Knol *et al.* (2004) have scanned and digitized maps into 50 by 50 meter grids and classified these grids into 10 categories, including built-up areas, water, sand and forest. We aggregate these 10 categories into built-up, open space and water bodies. Knol *et al.* document large changes in land use across the Netherlands from 1900 to 2000. For example, the total land used for buildings has increased more than fivefold. On the other hand, the amount of open space has decreased by about 10%. We also use information on municipal population in 1900 from *NLGIS*. Municipalities were much smaller at that time and about the size of a large neighborhood nowadays. We impute the local population distribution using the location of buildings and assuming that the population per building is the

[15] Another type of locations that are used in the game are so-called 'Gyms'. The latter types are unfortunately less useful, as these are almost uniformly distributed within urban areas in gardens, open spaces and public squares.

same within each municipality. We further use information on railway stations from Koopmans *et al.* (2012). We enrich these data by adding missing stations from various sources on the internet and create a network with travel times. To approximate the speed, we fit a regression of the length of (current) railway segments between stations on current travel time on the railway network. Based on historic sources, it appears that the average speed is about 50% of what it is currently (about 70km/h).

We show a map of land use and the railway network for the Netherlands in 1900 in Figure B.2. In Panel A it is shown that cities like Amsterdam, Rotterdam, The Hague, and Utrecht were already large by 1900. It can also be seen that some areas that have been reclaimed from the sea (e.g., to the northeast of Amsterdam) did not exist in 1900. The Panel B of Figure B.2 shows the railway network. In particular, places around Amsterdam and Utrecht have a high accessibility. The first railway line in the Netherlands was opened in 1839 between Amsterdam and Haarlem, soon followed by the openings of many other lines.

[Figure B.2 about here]

We use data composed by *HISGIS*, which has compiled and digitized data from the first Dutch census in 1832. This dataset provides information on the land use of each parcel in the current inner cities of Amsterdam, Rotterdam, Leiden, Delft, Hoorn, as well as for the province of Utrecht, Drenthe, Groningen, Friesland, Overijssel, Gelderland, and parts of Noord-Brabant. The *HISGIS* data also provide information on the cadastral income for about one-third of the observations, which was used to determine the tax at that time and is a proxy for land values. In Panel A of Figure B.3 we show that the study area is much smaller and excludes the city of The Hague. Hence, the results using data from 1832 is only based on a subsample of the population. We rely on municipal population data from *NLGIS* to calculate the accessibility in 1832. We assume that population is uniformly distributed within the municipality. Rail networks did not exist yet, so in order to calculate the population that could be reached within commuting time, we use information on the road network from 1821 obtained from Levkovich *et al.* (2017). Panel B of Figure B.3 shows the network back then.

[Figure B.3 about here]

In Table B.5 we provide descriptives for all instruments. The average share of built-up area in 1900 was 4.3%, while it was 4.2% in 1832. However, this figure is a bit misleading because for 1832 we have more data near urban areas. On average about 38 thousand jobs and 89 thousand people could be reached within commuting distance in 1900. Not surprisingly, this was much lower (40 thousand) in 1832.

[Table B.5 about here]

B.5 Other descriptive statistics

In Figure B.4 we report the distributions of the log of income and the log of land price. The distributions of land prices is somewhat positively skewed.

[Figure B.4 about here]

In Figure B.5 we show maps of income and land price distributions across the Netherlands. As expected, land prices are generally higher in cities. The pattern for incomes is less clear, but generally speaking we find that wealthier households locate close to or in cities.

[Figure B.5 about here]

B.6 Bias-corrected estimates

Many non-experimental papers use coefficient movements after the inclusion of control variables to investigate whether omitted variable bias is important. Oster (2019) argues that coefficient movements alone are not a sufficient statistic to calculate bias. Instead, she argues that whether omitted variable bias is a concern depends on the variance of the added control variables, as well as coefficient movements. In other words, changes in the coefficient(s) of interest after adding controls should be scaled by the change in the R^2. Oster (2019) then derives an estimator to correct estimates for omitted variable bias under the assumption that the relationship between the variables of interest and unobservables can be recovered from the relationship between the variables of interest and observables. In our context, this assumption makes sense as control variables that are added bear some potential relationship to unobservables. In our case, we add many housing controls as well as workplace fixed effects, which are likely to have at least some correlation to unobservables.

Oster (2019) then derives a GMM estimator to derive bias-corrected estimates of the impact of amenities and employment accessibility on incomes. There are two key input parameters that have to be determined. First, there is the maximum R^2 from a hypothetical regression of income on amenities, accessibility and controls, which we denote as $\bar{R}$. Given that our variables are neighborhood-specific, rather than household-specific variables, $\bar{R}$ is likely to be much smaller than 1. Second, a parameter must be chosen that determines the relative degree of selection on observed and unobserved variables, which we denote by ϖ. Although this parameter is fundamentally unknown, Altonji *et al.* (2005) and Oster (2019) show that $\varpi = 1$ is a reasonable (upper-bound) value. Oster (2019) then shows that

$$\alpha_1^* \approx \hat{\alpha}_1 - \varpi\left[\mathring{\alpha}_1 - \tilde{\alpha}_1\right]\frac{\bar{R}^2 - \hat{R}^2}{\hat{R}^2 - \mathring{R}_1^2} \quad \text{and} \quad \alpha_2^* \approx \hat{\alpha}_2 - \varpi\left[\mathring{\alpha}_2 - \hat{\alpha}_2\right]\frac{\bar{R}^2 - \hat{R}^2}{\hat{R}^2 - \mathring{R}_2^2}, \tag{B.3.2}$$

where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are parameter estimates obtained from a regression with controls (say household, job and housing controls, as well as workplace fixed effects), and $\hat{R}^2$ is the corresponding R^2. $\mathring{\alpha}_1$ and $\mathring{\alpha}_2$ are parameter estimates obtained from a regression without controls and $\mathring{R}_1^2$ and $\mathring{R}_2^2$ are the corresponding R^2s. Hence, this equation provides a simple way to evaluate robustness of the results. We then report bootstrapped bias-corrected estimates in Table B.6 of the coefficients of interest. We replicate the first three specifications reported in Table 3.

[Table B.6 about here]

In columns (1)-(3) of Table B.6, we naively assume that in theory we can fully explain variation in wages, so that $\bar{R} = 1$. Given this assumption, we find in column (1) – where we only include household controls and year fixed effects – that the effect of amenities is about 10 times as strong, and the effect of employment accessibility is about twice as strong as in the corresponding OLS specification. The effect of amenities becomes even stronger once we add housing and job controls in column (2) and is again comparable to column (1) once we add workplace fixed effects. This may lead to the conclusion that the OLS estimates are not robust and subject to omitted variable bias. However, the assumption that $\bar{R} = 1$ is likely to be wrong because the dependent variable is a variable measured at the micro-level (the household), while amenities and employment accessibility are measured at the neighborhood level. Hence, the maximum attainable $\bar{R}$ when omitted variables are important is likely substantially lower. To determine $\bar{R}$ we therefore run a regression of income on household, job and housing controls, as well as residence, workplace and year fixed effects. This leads to an R^2 of 0.357 , which is considerably smaller than 1. Moreover, it is around the value of $\bar{R} = 1.3\hat{R}^2$, which is supported by experimental data (Oster, 2019).

Columns (4)-(6) then show that the effect of amenities and employment accessibility are very similar to the OLS estimates. In our preferred specification, we find an elasticity of 0.0239, which is close to 0.0166 found in the OLS specifications. For employment accessibility we find an elasticity of 0.0936, which is essentially the same as 0.0881 reported in the corresponding OLS specification.

In other words, these results strongly suggest that omitted variable bias is not a major issue. Having said this, Oster's (2019) methodology does not account for measurement error in amenities or employment accessibility or reverse causality. It is therefore still important to apply our instrumental variables strategy.

B.7 First-stage results

We report first-stage estimates in Table B.7. In column (1) we use contemporary instruments for amenities. We show that current proxies for amenities are strongly positively correlated to picture density. For example, when the number of listed buildings per hectare increases by 1, picture density increases by 12.7%. Also the share of a neighborhood designated as historic district is positively correlated to the picture density. Furthermore, we find positive correlations with the share of built-up land and water bodies located in the neighborhood. Hence, picture density seems a meaningful proxy for amenities.

[Table B.7 about here]

In column (2) we use historic instruments. This means that we use the share of built-up land in 1900 and share of water in 1900 as instruments for picture density. We find strong positive effects of the share of built-up land in 1900 on picture density. This effect is about twice as strong as the share of contemporary built-up land, likely because the share of built-up land in 1900 is positively correlated to the current intensity of historic amenities.

Column (3) also includes the instruments for employment accessibility: the share of built-up land in 1900 within 500m, the share of built-up land in 1900 between 500 and 1000m and, most importantly, employment accessibility, This leaves the effects of the share of built-up land in 1900 in the own neighborhood almost unaffected.

In column (4) we take employment accessibility as dependent variable. The instruments for accessibility are relevant. We find a strong positive effect of the share of built-up land in 1900 between 500 and 1000m on accessibility, which makes sense. Also employment accessibility in 1909 has a strong positive effect on current employment accessibility. More specifically, doubling employment accessibility in 1909 is associated with an increase in current employment accessibility of 29%.

B.8 Sensitivity checks for the reduced-form income mapping

Identification revisited. We consider additional robustness analyses in Table B.8 that should increase confidence in the validity of our identification strategy. First, we show that our results are similar once we focus solely on urban areas. In column (1) we only include observations in the Randstad, i.e., the main polycentric metropolitan area in the Netherlands. This reduces the total number of observations by more than 50%. However, our results are similar, in particular for amenities. For employment accessibility we find that the coefficient is somewhat stronger, which may be due to traffic congestion in some parts of the Randstad (e.g., around Amsterdam and Rotterdam), which would imply that the commuting time elasticity is underestimated. In column

(2) we exclusively focus on observations close to city centers. That is, we only include locations within 15km of the center of an urban area with at least $100,000$ inhabitants. The coefficients are very similar, but for employment accessibility it becomes somewhat imprecise.

In column (3) we go back further in time and use instruments from 1832. This reduces the number of observations considerably because the 1832 data is not available for whole of the Netherlands. The Kleibergen-Paap F-statistic in column (3) is lower, which is not too surprising as going back further in time implies that correlations between instruments and endogenous variables become less strong. We find an effect for accessibility that is about twice as strong as when using instruments from 1900. In column (4) we use the information on the cadastral income, a proxy for the land value in 1832. This is missing in two thirds of the cases, so our number of observations drop further to about 1.8 million observations. Again, we find that the effect of amenities is very much comparable to the baseline specification. The effect of commuting time is even somewhat stronger. Interestingly, the effect of cadastral income is negative. A 10% decrease in the cadastral income in 1832 attracts households whose incomes are 0.03% higher, meaning that the effect is small. This is in line with anecdotal evidence that amenities in the past are essentially uncorrelated, or even negatively correlated, to current amenities.

[Table B.8 about here]

In column (5), Table B.8, we estimate specifications where we again use instruments from 1900, but control for the current share of built-up areas and population density to make sure that our amenity proxy is not just capturing population density or built-up land. We find very similar effects for amenities and accessibility.

One may be more worried that the concentrations of high-income households are autocorrelated so that our instruments are correlated to the concentrations of high-income households in 1909. To investigate whether this is an issue, we calculate the share of medium and high-skilled households in 1909. Municipalities then were much smaller, so this is a rather fine-grained measure of skill sorting across space. We also gather data on the share of Protestants in each municipality in 1899 and control for population accessibility in 1900. Including those measures does not impact our coefficients at all. Note that locations of high-skilled and medium-skilled households in 1909 are correlated to the locations of lower incomes nowadays, which suggests that the determinants of residential choices in the two periods are fairly different. This also confirms the negative association of Cadastral Incomes in 1832 to current incomes. Also, conditional on employment accessibility, population accessibility in 1900 is negatively correlated to current incomes. In column (7), we further study the sensitivity of our results by choosing another instrument for accessibility. We use the share of the population in 1909 born in the same municipality. If mobility of households is correlated over time, the share of locally born people should be negatively correlated to current accessibility, as the areas that host a high number of jobs (so have a better accessibility) are expected to attract workers from other places. Indeed, we find that the share of locally born people in 1909 is negatively associated with current employment accessibility. The Kleibergen-Paap F-statistic again indicates that these are strong instruments. We find a similar coefficient related to employment accessibility.

If one is still worried that household income sorting is autocorrelated, in column (8) we only include neighborhoods on reclaimed land. The Netherlands is well known for its large-scale projects that reclaim land from the sea. We consider the three main projects (Wieringermeer, Noordoostpolder, Oostelijk, and Zuidelijk Flevoland) that occurred between 1930 and 1968, but permission by the government to reclaim those areas was already given in 1930. Most of the land was intended for agriculture, but a few small settlements were planned on the newly reclaimed land. Moreover, Lelystad was planned to be the largest city, but nowadays Almere is by far the largest city in the area. In other words, the plans differ considerably from the current spatial economic distribution. Since only a small share of the population lives in those areas, we only keep about 2.5% of the observations.

We then instrument for amenities with the share of planned built-up and green areas in column (8). We observe that the impact of amenities is slightly lower, but, given the standard error, the effect is not statistically significantly different from that of the baseline estimate. The coefficient of employment accessibility is very similar to the baseline estimate, albeit imprecise. When we also instrument for employment accessibility with the planned accessibility in column (9), the point estimates are again similar, but we now have weak instruments leading to imprecise coefficients. In sum, we address reverse causality as no one was living in those locations at that time, and thus income was zero.

Other sensitivity checks Table B.9 reports the results of additional robustness checks. Our dataset contains observations on households. When calculating the commuting elasticity and when including workplace fixed effects, we focus on the job that generates the most working hours. This may be problematic when more people are employed in the household that work in different location. In column (1) we therefore only include households that are associated with one job. This does not lead to material differences in outcomes. When calculating the commuting time, we calculate the commuting time to the nearest plant of a firm, if it has multiple establishments. We test whether this introduces error by only including households that are associated with one job in a single plant firm in column (2). In this way we address any measurement error in commuting time. Again, the estimates are very similar.

[Table B.9 about here]

Our measures of commuting time rely on the minimum of travel time on the road and rail. However, in almost all cases travel time over the road is shorter. To make sure that households actually consider this travel time, we only keep households having a company car in column (3). This does not materially change the results. Column (4) replaces the dependent variable income by the share of adults in the household that have a college degree or more. We find very similar effects. For example, when the picture density doubles this increases the share of highly educated households by 3.3 percentage points. Conversely, doubling commuting times decreases the share of highly educated households by 20.3 percentage points.

Column (5) tests whether the results are robust when using commuting time by rail instead of commuting time over the road or rail. The results are comparable. Overall, *the impact of amenities and commuting time on income sorting choice is robust.*

B.9 The gravity model

In Table B.10 we report the results for the travel time elasticity. In column (1) we only include location pairs that are within 60 minutes drive from each other. Thus, we drop 77% of the data and we are left with 3.8 million residence-workplace pairs (note that many of those pairs have zero commuters so that more than 90% of the commutes are within 60 minutes). The estimated elasticity is -0.732, thus implying that doubling the commuting time reduces the probability that someone commutes between x and i is reduced by about 50%. In column (2) we address the potential endogeneity of travel times. That is, locations that attract many commuters may invite transport investments, thus leading to lower travel times. We instrument travel times with the Euclidian distance. Unsurprisingly, this is a very strong instrument. We do include the first-stage residual in the second stage as a control function. As one may observe, the first-stage residual is highly statistically significant, strongly suggesting that endogeneity is an issue. The travel time elasticity is now somewhat lower (-0.549), in line with the expectation that reverse causality would lead to an overestimate. Given that endogeneity is quite important, we consider this specification as the preferred one.

[Table B.10 about here]

In previous specifications we focus on commuting flows based on the job that generates the most hours in the household. In column (3), as a sensitivity check, we consider the two jobs that generate the most hours (if applicable). This hardly impacts the results. Column (4) investigates what happens if we use the railway travel time instead of travel time over the road. We show that this leads to similar estimates, although the elasticity is somewhat smaller. Rather than making a selection on maximum commuting time, we can also select locations with a sufficient number of commutes. In column (5) we include location pairs that have at least 25 commuters, including about 60% of the commutes. This leads to very similar results.

B.10 Counterfactual analysis

We outline the procedure for the counterfactual analysis discussed in Section 6.

1. The first step is to determine the location-specific scale parameters K_i, and productivity endowments $\mathbb{A}_i$. We set $\nu_{kxi} = 1$ and use the estimated $\tilde{\Omega}_i$ and Ω_i to obtain

$$K_i = e^{\tilde{\Omega}_i - \hat{\varepsilon}\Omega_i} \qquad \mathbb{A}_i = e^{\frac{\sigma}{\sigma-1}\Omega_i} / \widehat{L}_i^{\hat{\delta}}.$$

2. We build the values for commuting times τ_{xi}, exogenous amenities $\tilde{b}_x$ and productivity endowments A_i. If values do not vary for the specific scenario under consideration, we take the values from the data. Moreover, we set the starting values for L_i equal to the estimated value from the data and the initial value for the parameters $\hat{\gamma}$, $\hat{\gamma}_\Delta$, K_Δ, K_S to the values obtained in the structural estimation. We treat the parameters $\hat{\kappa}$, $\hat{\varepsilon}$, $\hat{\mu}$, $\hat{\beta}$, and $\hat{\delta}$ as given and obtain them from the structural estimation results.

3. We calculate labor productivity $t_{xi} = \left[\mathbb{A}_i L_i^{\hat{\delta}} \tau_{xi}^{-\hat{\kappa}}\right]^{(\sigma-1)/\sigma}$ for each location pair (xi), as well as the accessibility $\tilde{a}_x = \sum_{i=1}^{n} \tilde{t}_{xi} = \sum_{i=1}^{n} K_i t_{xi}^{\hat{\varepsilon}}$ of location i.

4. We calculate the location-quality indices:

$$\Delta_x = (\tilde{b}_x)^{\hat{\beta}} \left[\Gamma\left(\frac{\hat{\varepsilon}-1}{\hat{\varepsilon}}\right)(\tilde{a}_x)^{\frac{1}{\hat{\varepsilon}}}\right]^{1-\hat{\mu}}.$$

5. We fit a Fréchet distribution to Δ_x to obtain the adjusted values of the shape parameter $\hat{\gamma}_\Delta$. Since the aggregate skill distribution is given, it must be that $\hat{\gamma} = \hat{\gamma}_\Delta / \hat{\gamma}_S$.

6. We determine the skill mapping $s_x = \left[(K_S/K_\Delta)^{1/\hat{\gamma}_S} (\Delta_x)^{\hat{\gamma}}\right]^{\sigma/(\sigma-1)}$ and re-adjust K_Δ for the geometric mean of s_x to remain equal to 1. Hence, K_S, $\hat{\gamma}_S$ and the geometric mean should not change in the counterfactual.

7. We calculate total counterfactual labor supply in each employment location i. We have:

$$L_i = \sum_{x=1}^{N} \frac{\tilde{t}_{xi}}{\sum_{j=1}^{I} \tilde{t}_{xj}} f(\hat{s}_x).$$

where

$$f(\hat{s}_x) = \frac{\sigma-1}{\sigma} \hat{K}_S \hat{\gamma}_S e^{-\hat{K}_S (\hat{s}_x)^{-\hat{\gamma}_S(\sigma-1)/\sigma}} (s_x)^{-[\hat{\gamma}_S(\sigma-1)+\sigma]/\sigma},$$

is the skill density. Since L_i is an input to Step 3, we repeat steps (3)-(7) until L_i converges, which is usually within 10 iterations.

8. We now have all the information to solve for the total output in the city:

$$Y = \left[\sum_{i=1}^{n} \sum_{x=1}^{N} \frac{K_i \left[t_i(x)\right]^{\hat{\varepsilon}}}{\sum_{j=1}^{n} K_j \left[t_j(x)\right]^{\hat{\varepsilon}}} t_{xi} \left(\hat{s}_x\right)^{\frac{\sigma-1}{\sigma}} f(\hat{s}_x) \right]^{\frac{\sigma}{(\sigma-1)}}.$$

9. We also determine the income mapping $\omega_x t_{xi} = \left(K_S / \hat{K}_\Delta\right)^{1/\hat{\gamma}_S} \left(\Delta_x\right)^{\hat{\gamma}} (Y)^{1/\sigma} (t_{xi})$, which enables us to determine the land rent at each location x:

$$R_x = R_0 \frac{t_x}{t_{\underline{x}}} \left(\frac{\Delta_x}{\Delta_{\underline{x}}}\right)^{\frac{1}{\mu}} \left[\frac{\hat{\mu}\frac{t_x}{R_x} + (1-\hat{\mu})\frac{\overline{h}}{\omega_x}}{\hat{\mu}\frac{t_{\underline{x}}}{R_0} + (1-\hat{\mu})\frac{\overline{h}}{\omega_{\underline{x}}}} \right]^{\frac{1}{(1-\hat{\mu})\hat{\mu}\hat{\gamma}}},$$

where $\underline{x}$ is the location where the poorest household (with the lowest ω_x) lives, while R_0 is the agricultural land rent. We do not have good data on agricultural land prices. In any case, these will be not very useful as agricultural land prices in the Netherlands are highly regulated. We therefore set R_0 equal to the 5th percentile value of the observed land rents in our data. We use a standard Newton-Raphson procedure to determine the solution R_x.

10. We find consumption level of the composite good $q_x = (1-\hat{\mu})(\omega_x t_x - R_x \overline{h})$ and the housing consumption $h_x = (1-\hat{\mu})\overline{h} + \hat{\mu}\omega_x t_x / R_x$, which is identified up to a multiplication constant, so that the utility level is given by $u_x = (q_x)^{1-\hat{\mu}} (h_x - \overline{h})^{\hat{\mu}}$. This enables us to determine the aggregate land rent and aggregate real income:

$$ALR = \sum_{x=1}^{N} h_x R_x, \qquad \text{and} \qquad ARI = \sum_{x=1}^{N} \frac{1}{h_x} \frac{\omega_x t_x - R_x \overline{h}}{(R_x)^{\hat{\mu}}}.$$

where $1/h_x$ is the density of households in neighborhood x while, as shown in Appendix A.4, the individual real income is equal to $\left(\omega_x t_x - R_x \overline{h}\right) (R_x)^{-\hat{\mu}}$.

Appendix tables

Table B.1 – Descriptives for *NVM* dataset

	(1) mean	(2) sd	(3) min	(4) max
Price *(in € per m^2)*	1,269	927.2	25	25,000
Lot size *(in m^2)*	445.7	1,189	25	25,000
Size of property *(in m^2)*	132.4	45.16	26	538
Number of rooms	4.944	1.363	0	25
Terraced property	0.417	0.493	0	1
Semi-detached property	0.370	0.483	0	1
Detached property	0.189	0.392	0	1
Private parking space	0.454	0.498	0	1
Garage	0.394	0.489	0	1
Garden	0.966	0.182	0	1
Number of bathrooms	0.929	0.483	0	8
Number of kitchens	0.677	0.484	0	5
Number of balconies	0.132	0.354	0	4
Number of roof terraces	0.0674	0.257	0	3
Number of floors	2.717	0.636	1	13
Internal office space	0.00444	0.0665	0	1
Maintenance score of the outside	0.758	0.131	0	1
Maintenance score of the inside	0.753	0.143	0	1
Number of types of insulation	2.381	1.831	0	5
Central heating	0.920	0.271	0	1
Listed building	0.00652	0.0805	0	1
Newly built property	0.0417	0.200	0	1
Construction year	1,967	34.95	1,362	2,017
Year of observation	2,011	4.389	2,004	2,017

Notes: The number of observations is 1,337,495. Because of confidentiality restrictions the minimum and maximum values refer to the 0.01% and 99.99% percentile. This implies that we exclude the bottom and top observations

Table B.2 – Estimating land prices and lot sizes

(Dependent variable: the log of land price per m^2)

	(1)
Rooms	-6.1664***
	(0.4506)
Terraced property	702.4875***
	(6.5087)
Semi-detached property	510.0447***
	(6.5516)
Detached property	360.7740***
	(6.7580)
Private parking space	-56.3558***
	(1.9988)
Garage	-42.8166***
	(2.0556)
Garden	47.5907***
	(2.8356)
Number of bathrooms	17.3274***
	(0.9885)
Number of kitchens	-7.2575***
	(1.0818)
Number of balconies	47.8147***
	(1.5204)
Number of roof terraces	109.0801***
	(1.8878)
Number of floors	94.9407***
	(1.0148)
(Internal) office space	-55.3454***
	(6.3595)
Maintenance score of the outside	29.5137***
	(6.3366)
Maintenance score of the inside	501.7345***
	(5.8143)
Number of types of insulation	8.3945***
	(0.3138)
Central heating	65.8404***
	(1.7719)
Listed building	27.9334***
	(6.2691)
Newly built property	-13.3758***
	(4.3108)
3^{th}-order polynomial of property size	Yes
Construction decade dummies	Yes
Year fixed effects	Yes
Postcode fixed effects	Yes
Observations	1,280,031
R^2	0.8295

Notes: Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.3 – Other descriptive statistics for *NVM* data

	(1) mean	(2) sd	(3) min	(4) max
Price of home *(in €)*	229,238	116,074	25,000	1,000,000
Share land in historic district <500m	0.0695	0.192	0	1
Listed buildings <500m	0.179	0.894	0	19.53
Share water bodies <500m	0.0411	0.0713	0	0.920
Share open space <500m	0.244	0.217	0	1
Shops, <500m	0.254	0.394	0	4.711
Hotels, restaurants, cafés <500m	0.159	0.364	0	7.983
Leisure establishments <500m	0.0127	0.0215	0	0.318

The number of observations is 695,709. Because of confidentiality restrictions the minimum and maximum values refer to the 0.01% and 99.99% percentile. This implies that we exclude the bottom and top 70 observations.

Table B.4 – Determining the hedonic amenity index

(Dependent variable: the log of house price per m^2)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Share land in historic district <500m	0.1796***		0.1710***	0.1695***
	(0.0210)		(0.0204)	(0.0209)
Listed buildings <500m	0.0047**		0.0052**	-0.0043
	(0.0024)		(0.0024)	(0.0029)
Share water bodies <500m		0.3014***	0.2824***	0.2869***
		(0.0255)	(0.0253)	(0.0251)
Share open space <500m		0.0604***	0.0636***	0.0690***
		(0.0084)	(0.0084)	(0.0085)
Shops <500m				-0.0084
				(0.0074)
Hotels, restaurants, cafés <500m				0.0423***
				(0.0118)
Cultural establishments <500m				0.0480
				(0.0640)
Leisure establishments <500m				0.0232
				(0.0730)
Housing controls	Yes	Yes	Yes	Yes
Neighborhood fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Number of observations	695,709	695,709	695,709	695,709
R^2	0.8206	0.8207	0.8217	0.8219

Notes: Housing controls include house type, house size, whether the property has a garage, garden and/or central heating, the number of layers of insulation, the maintenance quality, the number of rooms, construction year dummies and whether a building is listed. Standard errors are clustered at the neighborhood level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table B.5 – Descriptive statistics for historic data

	(1) mean	(2) sd	(3) min	(4) max
Employment accessibility in 1909	38,029	23,884	1,494	163,349
Share of high-skilled workers in 1909	0.0298	0.0285	0	0.197
Share of medium-skilled workers in 1909	0.216	0.128	0.00386	0.688
Population accessibility in 1900	89,184	62,641	3,008	422,544
Share built-up land in 1900	0.0432	0.103	0	0.930
Share water in 1900	0.0591	0.175	0	1
Share locals in 1899	0.643	0.102	0.217	0.950
Share protestants in 1899	0.518	0.337	0	0.998
Population accessibility in 1832	40,389	20,970	1,986	135,168
Cadastral income in 1832 per ha	603.6	2,235	0	61,866
Share buildings in 1832	0.00726	0.0338	0	0.412
Share built-up land in 1832	0.0416	0.0890	0	1
Share water in 1832	0.120	0.264	0	1

The number of observations is 10,213,524. For the 1832 data it is 5,556,498. Because of confidentiality restrictions the minimum and maximum values refer to the 0.01% and 99.99% percentile. This implies that we exclude the bottom and top 1,024 observations

Table B.6 – Bias corrected estimates

	$\bar{R}^2 = 1.000$, $\varpi = 1.000$			$\bar{R}^2 = 0.357$, $\varpi = 1.000$		
		+ Controls	+ Workplace f.e.		+ Controls	+ Workplace f.e.
	(1) GMM	(2) GMM	(3) GMM	(4) GMM	(5) GMM	(6) GMM
Pictures per ha *(log)*	0.2735***	0.5509***	0.2105***	0.0326***	0.0345***	0.0239***
	(0.0495)	(0.0802)	(0.0162)	(0.0017)	(0.0011)	(0.0009)
Employment accessibility *(log)*	0.1900***	0.1900***	0.0778***	0.2377***	0.1044***	0.0936***
	(0.0647)	(0.0647)	(0.0285)	(0.0154)	(0.0054)	(0.0045)
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Housing and job controls	No	Yes	Yes	No	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Workplace fixed effects	No	No	Yes	No	No	Yes
$\bar{R}$	1.000	1.000	1.000	0.357	0.357	0.357
ϖ	1.000	1.000	1.000	1.000	1.000	1.000
Number of observations	10,213,540	10,213,540	10,213,540	10,213,540	10,213,540	10,213,540

Notes: Household controls include household size, mean age of adults, mean gender, household type (couple, single, kids), the share of the household that is foreign-born. Job controls are the total hours worked, whether the household has a company car, the share of full-time contracts, the share of permanent contracts. Housing controls include house type, height of the building, construction year dummies and whether a building is listed. Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

TABLE B.7 – FIRST-STAGE REGRESSION RESULTS

	Dep.var. Pictures per ha (log)			Dep.var. Employment accessibility (log)
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Listed buildings per ha	0.1269***			
	(0.0346)			
Share historic district	2.2677***			
	(0.1894)			
Share built-up land	2.4775***			
	(0.0847)			
Share water	2.5560***			
	(0.3015)			
Share built-up land in 1900		5.4119***	4.4326***	-0.2133***
		(0.2521)	(0.2908)	(0.0400)
Share water in 1900		0.6163***	0.6727***	0.0211
		(0.1520)	(0.1525)	(0.0253)
Share built-up land in 1900, 0-500m			0.2646	0.0612
			(1.1085)	(0.1824)
Share built-up land in 1900, 500-1000m			5.8759***	0.3781**
			(1.3125)	(0.1928)
Employment accessibility *(log)*	0.3448***	0.7939***		
	(0.0420)	(0.0443)		
Employment accessibility in 1909 *(log)*			0.3150***	0.4204***
			(0.0458)	(0.0112)
Household controls	Yes	Yes	Yes	Yes
Housing and job controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Workplace fixed effects	Yes	Yes	Yes	Yes
Observations	10,213,524	10,213,524	10,213,524	10,236,308
R^2	0.6046	0.5036	0.4989	0.7875

Notes:Household controls include household size, mean age of adults, mean gender, household type (couple, single, kids), the share of the household that is foreign-born. Job controls are the total hours worked, whether the household has a company car, the share of full-time contracts, the share of permanent contracts. Housing controls include house type, height of the building, construction year dummies and whether a building is listed. Standard errors are clustered at the neighborhood level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table B.8 – Reduced form results: identification

(Dependent variable: the log of household gross income)

	Only Randstad	City center <15km	1832 instruments		Control for current land use	1909 skills	Other instrument	Only obs. on reclaimed land	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Pictures per ha *(log)*	**0.0382*****	**0.0374*****	**0.0375*****	**0.0491*****	**0.0494*****	**0.0483*****	**0.0501*****	**0.0221****	**0.0201**
	(**0.0048**)	(**0.0046**)	(**0.0048**)	(**0.0061**)	(**0.0050**)	(**0.0053**)	(**0.0044**)	(**0.0101**)	(**0.0153**)
Employment accessibility *(log)*	**0.154**4***	**0.1647*****	**0.1134*****	**0.1503*****	**0.0597*****	**0.1081*****	**0.2170*****	0.0413	**0.0453**
	(**0.032**3)	(**0.0501**)	(**0.0160**)	(**0.0319**)	(**0.0100**)	(**0.0352**)	(**0.0645**)	(0.0362)	(**0.1039**)
Cadastral income in 1832 per ha *(log)*				-0.0050**					
				(0.0023)					
Share built-up land					-0.0936***				
					(0.0138)				
Population per ha *(log)*					-0.0074**				
					(0.0033)				
Population accessibility in 1900 *(log)*						-0.0264*	-0.0681***		
						(0.0150)	(0.0249)		
Share of medium-skilled workers in 1909						-0.1630***	-0.1860***		
						(0.0211)	(0.0248)		
Share of high-skilled workers in 1909						-0.1378	-0.1645		
						(0.1081)	(0.1053)		
Share protestants in 1899						-0.0169***	-0.0133*		
						(0.0061)	(0.0072)		
Household controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing and job controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Workplace fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4,340,639	6,023,886	5,549,488	1,782,784	10,213,325	9,778,046	9,778,046	270,106	270,106
Kleibergen-Paap F-statistic	70.51	34.43	22.73	33.87	61.92	21.16	15.79	9.468	0.804

Notes: **Bold** indicates instrumented. Household controls include household size, mean age of adults, mean gender, household type (couple, single, kids), the share of the household that is foreign-born. Job controls are the total hours worked, whether the household has a company car, the share of full-time contracts, the share of permanent contracts. Housing controls include house type, height of the building, construction year dummies and whether a building is listed. Standard errors are clustered at the neighborhood level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table B.9 – Sensitivity analysis for reduced form regressions

	One job households	+ Single plant firm	Company car	Education level	Commuting by rail
	(1)	(2)	(3)	(4)	(5)
	2SLS	2SLS	2SLS	2SLS	2SLS
Pictures per ha *(log)*	**0.0401*****	**0.0367*****	**0.0325*****	**0.0540*****	**0.0382*****
	(**0.0038**)	(**0.0038**)	(**0.0040**)	(**0.0032**)	(**0.0039**)
Employment accessibility *(log)*	**0.0551*****	**0.0455*****	**0.0708*****	**0.0333*****	**0.0269*****
	(**0.0111**)	(**0.0134**)	(**0.0112**)	(**0.0099**)	(**0.0062**)
Household controls	Yes	Yes	Yes	Yes	Yes
Housing and job controls	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Workplace fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	6,706,524	3,532,906	1,523,567	7,626,355	10,213,524
Kleibergen-Paap F-statistic	85.60	88.36	77.53	82.87	80.51

Notes: **Bold** indicates instrumented. Household controls include household size, mean age of adults, mean gender, household type (couple, single, kids), the share of the household that is foreign-born. Job controls are the total hours worked, whether the household has a company car, the share of full-time contracts, the share of permanent contracts. Housing controls include house type, height of the building, construction year dummies and whether a building is listed. Standard errors are clustered at the neighborhood level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

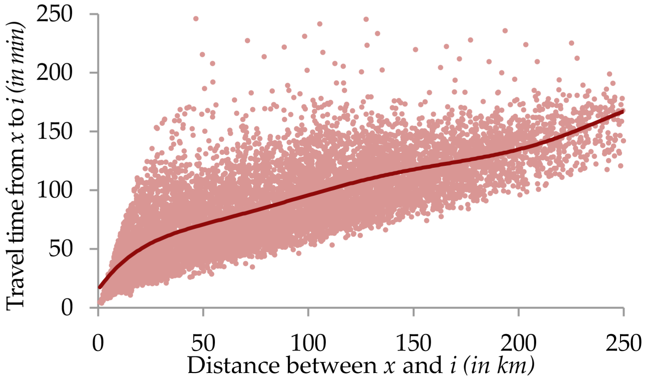
Table B.10 – Regression results of gravity model

(Dependent variable: the number of commuters)

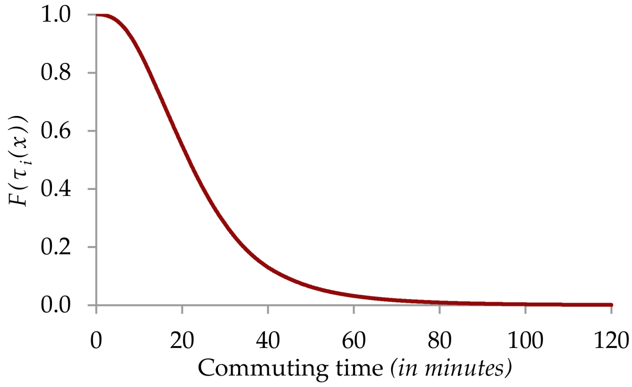
	Baseline Poisson	Control function	Flows based on two jobs	Travel time by train	Flow >25
	(1)	(2)	(3)	(4)	(5)
	Poisson	Poisson-CF	Poisson-CF	Poisson-CF	Poisson-CF
Commuting time elasticity, $\hat{\varkappa}$	-0.7318***	-0.5485***	-0.5703***	-0.3393***	-0.5215***
	(0.0139)	(0.0122)	(0.0111)	(0.0080)	(0.0086)
First-stage error		-0.2378***	-0.2079***	0.3402***	0.4154***
		(0.0653)	(0.0475)	(0.0217)	(0.0207)
Residence location fixed effects	Yes	Yes	Yes	Yes	Yes
Workplace location fixed effects	Yes	Yes	Yes	Yes	Yes
Number of area pairs	3,904,262	3,904,262	3,904,262	3,904,262	66,147

Notes: We use commuting flows between neighborhoods based on the job that generates the most working hours. In columns (2)-(5) we use as instrument the euclidian distance between two neighbourhoods as instrument. In column (3) we derive the commuting flow based on the two jobs that generate the most working hours in the household. Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

Appendix figures

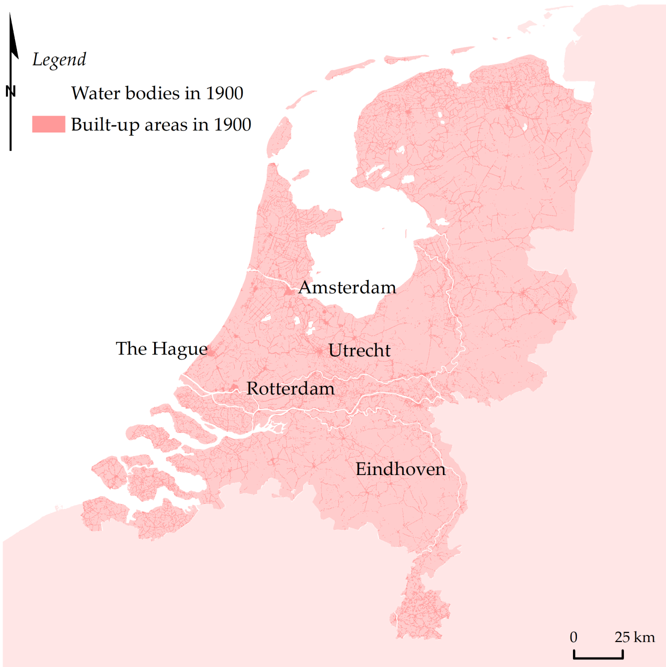


(A) Distance and travel time

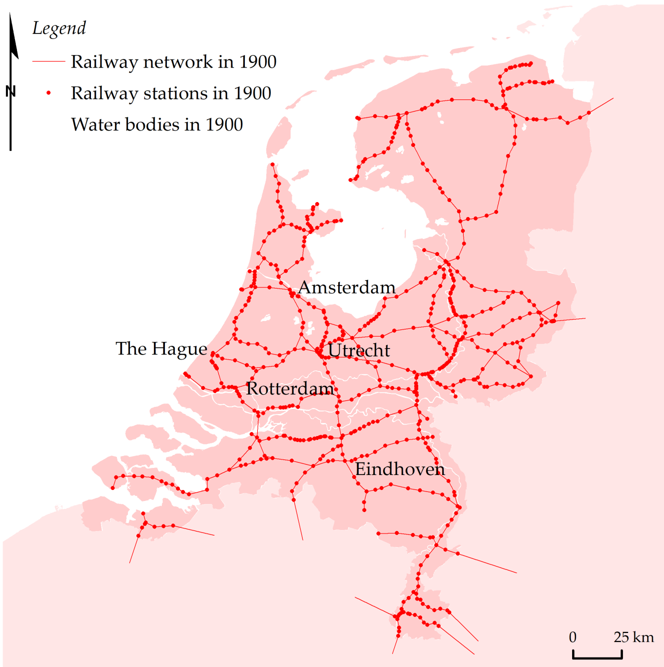


(B) Commuting time distribution

Figure B.1 – Calculation of travel time and speed

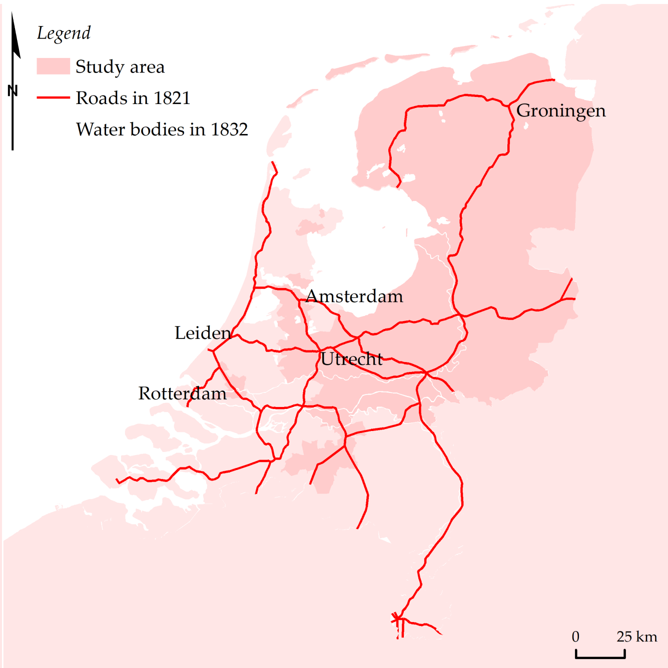


(A) Built-up land



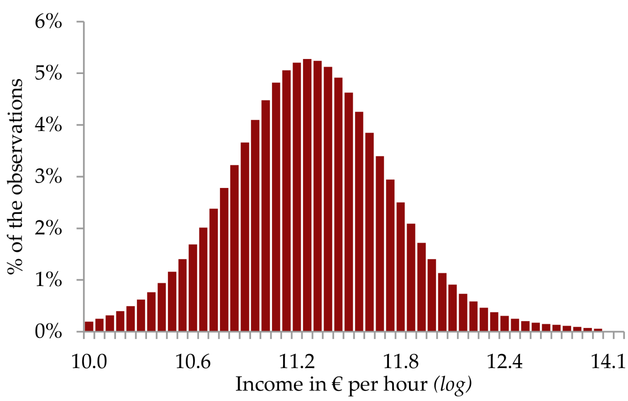
(B) The railway network and accessibility

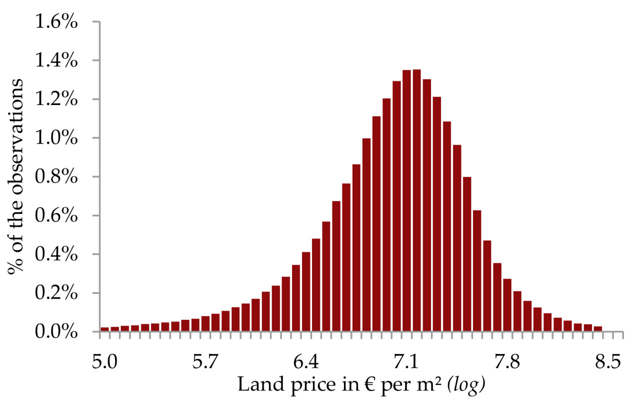
Figure B.2 – Historic data from 1900



(A) Built-up land (B) The road network and accessibility

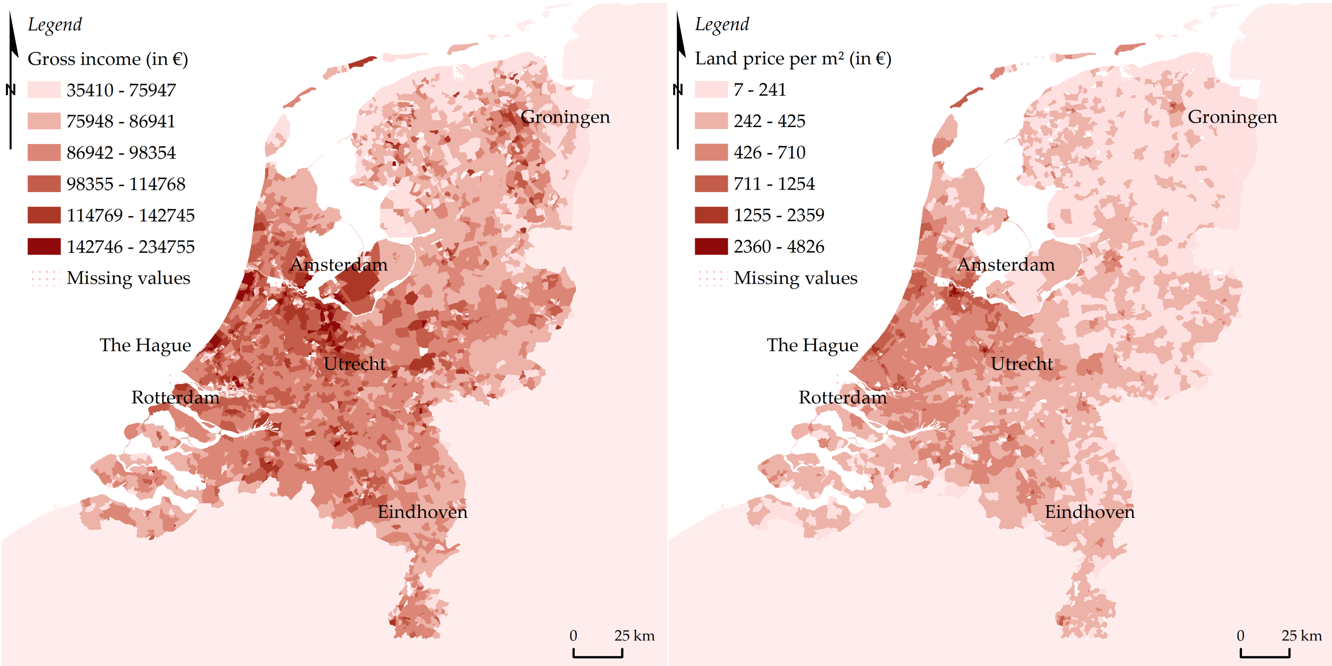
Figure B.3 – Historic data from 1832





(A) Incomes (B) Land prices

Figure B.4 – Histograms for the variables of interest



(A) Average gross income *(in €)*　　(B) Average land prices per m^2 *(in €)*

Figure B.5 – Spatial distribution of variables of interest