PhD-FDEF-2020-04
The Faculty of Law, Economics and Finance

DISSERTATION

Defense held on 04 February 2020 in Luxembourg
To obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN SCIENCES ÉCONOMIQUES

by

TRAN Thi Thu Huyen
Born on 16 June 1990 in Nam Dinh (VIETNAM)

ESSAYS ON URBAN GREEN AMENITY AND CITY STRUCTURE

Dissertation defense committee

Prof. Dr. Pierre M. Picard (supervisor, Université du Luxembourg)
Prof. Dr. Geoffrey Caruso (president of the jury, Université du Luxembourg / LISER)
Assoc. Prof. Dr. Antonio Cosma (vice president of the jury, Université du Luxembourg)
Prof. Dr. Jos van Ommeren (Vrij Universiteit Amsterdam)
Prof. Dr. Elisabet Viladecans Marsal (Universitat de Barcelona)

Preface

I could not have completed this PhD dissertation without the support of many colleagues, friends and family.

Pursuing a PhD in urban economics is not the path I envisioned following five years ago, when I first came to the Toulouse School of Economics for the master's program. I became enthusiastic about urban economics when I did an internship at the Transport Policies and Development Division at UN ESCAP in 2014. I researched the impact of transport networks on economic activities, mostly in Southeast Asian countries. My experience at UN ESCAP inspired me about cities, geographies, transport and amenities. My PhD journey officially began four years ago when I joined the SOS-BUGS project on 'Spatial Optima and the Social Benefits of Urban Green Space' at the University of Luxembourg.

I would first like to thank my primary supervisor, Prof. Dr. Pierre M. Picard, for his guidance, support and encouragement for these past four years. Our discussions and work together have taught me a great deal about research approaches, and his efforts will undoubtedly shape the approaches I take to future projects. Thanks for recognizing the potential in me, encouraging and guiding me while still providing me with enough freedom to try on my own (and often make necessary mistakes).

I would also like to thank my second supervisor, Prof. Dr. Geoffrey Caruso. Thanks to his generosity, I had the opportunity to interact and connect with researchers and colleagues at Urban Planning and Geographical Centres for a semester. I learned a great deal from his useful comments and advice on my research. His vast knowledge, broad interests, and kindness will continue to inspire me in the future.

I am genuinely grateful for my third promoter, Assoc. Prof. Dr. Luisito Bertinelli. Although we worked on relatively few projects together, I learned much from his helpful comments, attention to detail and chiefly from his availability to discuss and resolve any issue that arose.

I was also fortunate to have been able to visit the Spatial Economic Department at the Vrije Universiteit (VU) Amsterdam in 2017 for four months. Prof. Dr. Jos van Ommeren, thanks for allowing me to spend some time at VU. Our discussions always made me consider matters from different perspectives and significantly improved the research. I genuinely appreciate his guidance and valuable remarks.

I am truly grateful to my committee members, Prof. Dr. Elisabet Viladecans Marsal and Assoc. Prof. Dr. Antonio Cosma, for their reading of the thesis and helpful comments. I am particularly thankful to Dr. Evgenii Monastyrenko, who generously read and gave considerable feedback on multiple drafts of the chapters in my dissertation.

The research center in economic and management (CREA) at the University of Luxembourg has been a great place to conduct research and to grow professionally. I thank all CREA members who have contributed to such a great working environment. I am also thankful to all my colleagues for creating a pleasant and friendly atmosphere and the interesting talks over coffee breaks and beers at the doctoral school events. I would particularly like to thank my office-mate, Ka Kit Iong, for easing my arrival at the university, for tolerating the noise in our office, and for interesting political discussions. His advice and conversions made the entire experience much better. I am also thankful to Marina LeGrand, Elisa Gesellchen-Ferreira Da Silva, and Roswitha Glorieux from the secretary office for providing me with such great support over the past few years.

Finally, I would like to thank my family. To my parents and my two little sisters, Trang and Thuy, thanks for the love, encouragement and constant support. Without them, this work would not have been possible.

Tran Thi Thu Huyen
Luxembourg, December 15, 2019

Abstract

Sustainable urban living is a growing issue in many regions worldwide. The debates over compact city policies and the provision of green areas within urban landscapes are increasing day-by-day as the number of urban dwellers has soared in recent decades. This dissertation focuses on studying the optimal provision of green urban areas and the welfare effects of a substantial change in green provision policies in the presence of other types of land uses and adverse shocks. It comprises the following four papers (chapters).

Green Urban Areas (joint with Pierre M. Picard). This paper studies the size and location of urban green areas across city spaces. Urban green areas offer amenities that affect residential choices, land consumption and land rent. This paper discusses the socially optimal sizes and locations of urban green areas within a city and their decentralized allocation through land markets. The main result is that the share of land dedicated to urban green areas is a concave function of the distance to the city center. This result is confirmed by the empirical study of urban structures in the 305 largest EU cities.

Welfare Evaluation of Green Urban Areas (joint with Pierre M. Picard). This paper quantifies the impacts of urban green areas on city structures under competitive land markets for more than 300 of the most populous cities in Europe. We first introduce the amenity of local and land-intensive green areas within an urban model. We utilize two different classes of preferences: standard Cobb-Douglas preference and hyperbolic preference. The theoretical equilibrium is then used to instruct empirical regressions for residential land uses. Notably, we calibrate the model using data on detailed residential land uses, green urban areas and population density. We back up the parameters and run counterfactual simulations under the scenario that urban planners reduce green areas by half. We consider the general equilibrium effects in households' willingness to pay for such a policy and the changes in the land market for both closed and open cities with two land conversion policies. We show that the gross benefits of urban green areas are substantial and that households are required to compensate up to $6.4 - 9\%$ of their yearly income on average to reach the same utility level before the changes. However, the net cost of urban green areas when we allow for land conversion into new residential land is negative, estimated at approximately 4% of annual household

income.

Geographical Stratification of Green Urban Areas (joint with Pierre M. Picard). This paper studies the provision of urban green areas in cities when residents have preferences for the size of and access to those areas. At the optimum, the number of urban green spaces is a non-monotone function of distance to the city centre, while the sizes and distances to other urban green areas increase as one moves to the urban fringe. This paper empirically investigates those properties for the 300 largest European cities by using the GMES Urban Atlas database (European Environmental Agency). The empirical analysis confirms the non-monotone relationship between the number of urban green spaces and the distance to the city centre. The distance between two parks also increases as one moves toward the urban fringe. Finally, richer cities are associated with a denser network of urban green areas.

Urban Floods, Amenity Land Uses and City Structure. This paper investigates the effects of concurrent negative shocks in the form of urban flood risks on the amenity land use planning and city structures. We examine both the first-best and second-best land use policy when parts of the cities are in flood-prone areas. We find a non-monotone relationship of public mitigation infrastructures and amenity provision with the level of flood risks. The outcomes of the first-best optimization are decentralized through market equilibrium through locational taxation. However, an income tax used to finance the flood prevention would distort the first-best allocation and induce more housing developments with lower flood mitigation efforts per unit of land in flood-prone areas compared to the number in the first-best scenario. For the empirical part, we use multiple return periods of flood hazard maps in Europe together with the GMES Urban Atlas to examine the actual land use in 305 of the most populated urban areas in Europe. To our knowledge, this is the first paper that discusses both theoretically and empirically the urban land uses under flood risk across all of Europe.

Contents

1 Introduction **1**

2 Green Urban Areas **5**
2.1 Introduction 5
2.2 Theoretical model 7
2.2.1 Social optimum 9
2.2.2 Comparative statics 11
2.2.3 City border and population 12
2.2.4 Land regulation 13
2.2.5 Competitive land market equilibrium 14
2.2.6 Specific preferences 15
2.3 Empirical analysis 19
2.3.1 Data 19
2.3.2 Urban green area profiles 25
2.4 Conclusion 31

3 Welfare Evaluation of Green Urban Areas **41**
3.1 Introduction 41
3.2 Literature review 43
3.3 Theoretical model 45
3.3.1 Demand side 47
3.3.2 Competitive land market equilibrium 47
3.3.3 Specific preferences 49
3.4 Empirical analysis and welfare valuation 50
3.4.1 Data 51
3.4.2 Estimation and valuation with Cobb-Douglas preferences 54
3.4.3 Estimation and valuation with hyperbolic preferences 65
3.5 Conclusion 71

4 Geographical Stratification of Green Urban Areas **85**
4.1 Introduction 85
4.2 The Model 89

4.2.1 Household preferences . . . 89
4.2.2 Optimal green urban spaces . . . 91
4.2.3 Cobb-Douglas preferences . . . 92
4.3 Empirical analysis . . . 95
4.3.1 Data . . . 95
4.3.2 Descriptive statistics . . . 100
4.3.3 Number of green urban spaces . . . 102
4.3.4 Size of green urban spaces . . . 105
4.3.5 Distance to other green urban spaces . . . 106
4.3.6 Population density . . . 110
4.4 Conclusion . . . 114

5 Urban Floods, Amenity Land and City Structure 127
5.1 Introduction . . . 127
5.2 Theoretical model . . . 130
5.2.1 Model settings . . . 130
5.2.2 First-best world . . . 132
5.2.3 Second-best world . . . 138
5.3 Empirical analysis . . . 143
5.3.1 Data . . . 143
5.3.2 Population density in floodplain areas . . . 150
5.3.3 Housing construction density . . . 153
5.3.4 Flood prevention infrastructure systems . . . 156
5.3.5 Land use for local public goods within flooded zone . . . 159
5.4 Conclusion . . . 161

Bibliography 173

Chapter 1

Introduction

In recent centuries, there has been a massive increase in the number of urban dwellers. Their share has grown from approximately less than 8 percent in 1800 to more than 50 percent in recent years (Glaeser, 2011; Glaeser, 2012; UN DESA, 2018). The emergence of large metropolises such as Tokyo, New York, London, Mexico City and Mumbai has become a more familiar scene. There are many reasons for this phenomenon: the agglomeration benefits of firm clusters (Glaeser and Gottlieb, 2009; Ellison et al., 2010), thick and dynamic labor markets (Fallick et al., 2006), social interaction benefits (Mossay and Picard, 2011; Picard and Tabuchi, 2013) or the provision of numerous consumption amenities due to the availability of a sufficient mass of customers (Glaeser et al., 2001; Rappaport, 2008). Due to this increasingly urbanized pattern, it is predicted that the population share living in urban areas will increase to 70%. If this trend is irreversible, then the next question should concern how we can make such urbanization work. How can we improve urban quality of life and make city life sustainable?

One of the main components that make cities attractive to their residents is their public park and garden systems. Green urban areas from a small community garden to famous areas such as 'Jardin du Luxembourg' in Paris not only shape the face of the city but are a quintessential aspect of the quality of life for local inhabitants. They offer places for local recreation, beautiful views, cleaner air and many other advantages. Recent research has validated the connection between urban parks and the well-being of the city's inhabitants. Bowler et al. (2010) and Manes et al. (2012) find a strong influence of urban natural ecosystems in mitigating air and noise pollution and urban heat island effects. Heidt and Neef (2008) suggest economic benefits by showing a significant increase in nearby property values. Access to nearby urban parks helps reduce stress, improve psychological well-being, and increase physical activity (Colding and Barthel, 2013; Saldivar-Tanaka and Krasny, 2004; White et al., 2013).

Although green urban areas might seem to be meagre in comparison with other natural ecosystems such as wetlands or forests, the value of the environmental, recreational, and other services they offer is likely to be disproportionally high due to their

strategic locations. The benefits are at the local level and right at urban dwellers' doorsteps. However, the provision of urban parks often entails high opportunity costs because it demands land at locations where it is the most precious and scarce. Recent studies report that some land development restrictions to preserve open spaces can distort land use allocations and reduce welfare (Cheshire and Sheppard, 2002; Turner et al., 2014). To correctly identify optimal policies, it is necessary to quantify the trade-offs between urban green areas and other types of land use as well as the welfare effects of current green areas.

This dissertation includes four research papers that use theoretical, empirical and simulation approaches to, first, explore the principles of optimal urban green provisions. The theoretical models are then applied to real-life data on city land uses to verify the predictions of the models and to estimate the welfare effects of a substantial change in urban green areas.

Chapter two of the dissertation studies the optimal level of small and scattered green urban areas in urban spaces in continuous fashion. Because green urban areas are land-intensive and offer highly localized amenities, we find that the optimal share of surface area devoted to green urban areas is a concave and non-monotonic function of the distance to the city center. On the one hand, the opportunity cost of land is too expensive in the city center for the planner to situate many green areas there. On the other hand, residents are too sparse at the city edges to give planners the incentives to invest in such highly localized amenities. The main contribution of the paper is to confirm our results using the very detailed geographical database of the European Environment Agency's Urban Atlas data, which describes the land use of 305 EU cities with more than 100,000 inhabitants. These data report land use and cover all of Europe using harmonized earth observations (EOs). The data represent a unique source of reliable and comparable European urban planning data. To the best of our knowledge, this is the first paper that uses both a theoretical model and empirical estimation of European urban land uses. This result is robust to many variations of the land use specifications, city structure specifications, and city and country characteristics.

The third chapter further utilizes the approach employed in the second chapter in a setting with competitive land markets to estimate the welfare effects of a hypothetically large change in green urban areas in European cities. We employ two different household preferences: Cobb-Douglas and hyperbolic linear utility. The marginal benefit of urban green areas varies across locations within cities concerning their existing coverage in residential areas, the population density and the opportunity cost of land at each location. The market equilibrium results in optimal residential land use at each location across urban areas with a specified level of urban green areas. We then calibrate the equilibrium of residential land using the data on detailed residential land cover, green urban areas and population density. To control for the potential endogeneity of

urban green provision from urban authorities, we employ the green areas in the past as an instrument for the current green areas. We store the parameters and run counterfactual simulations in a scenario in which urban planners reduce green areas by half. By using this approach, we can account for the new equilibrium in the housing market and households' choices over housing consumption. We consider both the case of open cities with free migration and closed cities with no migration. We show that the gross benefits of urban green areas are substantial, and households require compensation of up to $6.4 - 9\%$ of their yearly income on average to reach the same utility level before the changes. However, the net cost of urban green areas when we allow for land conversion into new residential land is negative, estimated at approximately 4% of annual household income.

In the fourth chapter, we examine the optimal geographical configurations of urban parks in discrete fashion. We discuss the optimal numbers, sizes and distances of small and scattered parks across different city locations. At the optimum, the number of urban green spaces is a non-monotonic function of distance to the city center, while the sizes and distances to other urban green areas increase as one moves to the urban fringe. This paper empirically investigates those properties for the 300 largest European cities by using the GMES Urban Atlas database. Chapter two uses the same database to study the trade-off between residential space and green urban areas, and we highlight a non-linear relationship between the land share of green urban areas and the distance to the city center. By contrast, this chapter disentangles the effect of the latter on the size, number and dispersion of parks. We find a non-monotonic relationship between the number of urban green spaces and the distance to the city center and a dispersed pattern in the optimal distance between two parks as one moves toward the urban fringe. Furthermore, wealthier cities are associated with a denser network of urban green areas.

The last chapter additionally investigates the negative shock of urban river flood risks to amenity land use planning and city structures. We examine both the first-best and second-best land use policy when parts of the cities are in flood-prone areas. We find a non-monotonic relationship between amenity land and flood risk. Under the first-best scenario, locations with higher flood risks in dense areas are more likely to have higher amounts of amenity land, while the opposite is true in less dense areas. The pattern is similar under the second-best scenario; however, an income tax used to finance flood prevention would distort the first-best allocation and induce more people to locate in the flood-prone area compared to the first-best. In the empirical analysis, we use the 500-year flood hazard maps in Europe together with the GMES Urban Atlas to examine the actual land use in the 305 most populated urban areas in Europe. To the best of our knowledge, this is the first paper that discusses both theoretically and empirically urban land uses under increased flood risk scenarios across Europe.

The findings presented in this dissertation contribute to our understanding of the optimal land use planning and welfare effects of land-intensive public goods in the context of a non-uniform population distribution. It also incorporates the adverse risk of urban river floods into green provision and land use planning. The policies on green areas and urban flood risk are commonly mentioned and debated in many regions worldwide. This discussion is also particularly relevant for the design and implementation of future urban areas and policies. A better understanding of the potential channels in which green areas and urban river flooding affect people's choices and locations can be used to examine future policy changes.

Chapter 2

Green Urban Areas

2.1 Introduction

Urban green areas play a crucial role in the debate on sustainable cities. They are an important part of any urban area whose quality and quantity are prime concerns for environmental sustainability. Recent research has confirmed the relationship between urban parks and the well-being of city's residents. Brack (2002) and Strohback et al. (2012) find a strong influence of urban natural ecosystems in reducing air and noise pollution and CO_2 absorption in Australia and Germany. Heidt and Neef (2008) find economic benefits by showing a significant increase in the nearby property values. Access to nearby urban parks helps reduce stress and improve psychological well-being[1] and increasing physical activities.[2]

Green urban areas are generally small, scattered and close to each other. As shown in Table 1, on average, residents find a green urban area at less than about 350m from their residences (about 6 minute walk) while more than 90% of green urban areas are smaller than squares of about 300m length (i.e. surface smaller than 100,000 m^2). Finally the numbers of green urban areas with each city are roughly proportionate to the city populations.

In this paper, we study the optimal level of small and scattered green urban areas in urban spaces. Because green urban areas are land-intensive and offer very localized amenities, we find that the relationship between the share of surface devoted to green urban areas is a concave function of the distance to the city center. On the one hand, the opportunity cost of land is too expensive in the city center for the planner to implement many green areas there. On the other hand, residents are too sparse at the city edges to give planners the incentives to invest in such very localized amenities. The main contribution of the paper is to confirm our results using the very detailed geographical

[1] See Berman et al. (2008), Hedblom et al. (2019), IFPRA (2013), Tzoulas et al. (2007), and van den Bosch and Sang (2017)

[2] See Cohen at al. (2007, 2016) and Evenson et al. (2013)

Table 2.1: Distance and Surface of Urban Green Areas (EU)

	Small Cities	Medium Cities	Big Cities
Number of inhabitants (million)	$[0.1, .5)$	$[0.5, 1)$	$[1, \infty)$
Closest distance between 2 urban green areas (m)	336	325	350
(std)	(100)	(62)	(70)
Number of urban green areas within city border	96	384	784
(std)	(74)	(334)	(513)
Average share of urban green areas with surface < 0.1 km^2	94.1%	92.2%	90.6%

Note: Authors' computation using GMES Urban Atlas database with the 305 largest European cities. Adjacent green urban spaces are merged and counted for a single park. Distance between two parks is the distance between the centroids of each park.

database of the European Environment Agency's Urban Atlas data, which describes the land use of 305 EU cities with more than 100,000 inhabitants. These data report land use and cover across Europe using harmonized Earth Observations (EOs), which are combined with Eurostat Urban Audit statistical data. The data represent a unique source of reliable and comparable European urban planning data. As far as we know, this is the first paper that uses both a theoretical model and empirical estimation of European urban land use. This result is robust to many variations of the land use specifications, city structure specifications, and city and country characteristics.

Our contribution relates to several strands of the economics literature. First, green urban areas share the nature of a local public good. Since Tiebout (1956), economists have discussed the issue of resident mobility and democratic decision over local public goods (voting with feet). Fujita (1986), Cremer et al. (1986) and Sakashita (1987) discuss the problem of the optimal location of local public goods and find that local public goods should spread to equidistant locations. Fujita (1989) studies the optimal allocation of "neighbourhood goods" (localized and congestible public good) and their decentralization through Pigouvian taxes. As such goods do not use land as input, the balance between amenity and land use cannot be discussed. Berliant et al. (2006) endogenize the public good provision and location in cities where households have inelastic land use. Optimal public good providers are found again to be equidistant and to serve basins of residents of the same size. Yet, in contrast to this paper, those studies are conducted under the assumption of no land use in the production of local public goods and/or no endogenous choice of residential land plots. Because green urban areas are rather land-intensive, it is important to study how land use affects public goods. Furthermore, the distribution of residents is not uniform across urban landscapes and not exogenous to the local amenities given by green urban areas. This paper focuses on the relationships between the endogenous distribution of residents and green areas.

Finally, this paper links to the urban economics literature regarding the effect of open spaces on urban form. Wu and Plantinga (2003) investigate the effect of an open space on the surrounding urban structure. They, however, treat the location and size of

this open space as exogenous. Warziniack (2010) considers voting on the location of a single open space when the geographical distribution of households is exogenous. Lee and Fujita (1997) and Yang and Fujita (1999) examine the effect of a greenbelt, which has an exogenous location. Yang and Fujita (1999) consider the effect of open spaces at the neighborhood level and conclude that the equilibrium open space provision is uniform across the distance to the city center. Such results contrast with our empirical analysis that shows that the share of green urban areas is not constant across the city space. Our model with endogenous locations and choices of residential space enables an explanation of this pattern.

Parallel to our question is the issue of unoccupied urban spaces, which are often seen as green areas. In contrast to green urban areas, open spaces are not maintained for human activities. Unoccupied land has been primarily justified by the leapfrogging effect. Capoza and Helsley (1990) and followers root this effect in the commitment of building decisions and the resulting option value of urban land. Turner (2005) explains unoccupied land by the negative externalities of dwellings in their direct neighborhoods. Walsh (2007) discusses and estimates the protection and regulation of open spaces in Wake County (California, USA), which expands the discussion beyond monocentric city frameworks. Caruso et al. (2007) simulate market equilibria with discrete house slots and a fixed housing consumption, which lead to open spaces. In contrast, our paper discusses a continuous model where households decide their locations and slots and where open green areas are costly and planned as in many EU cities. Urban green areas, such as parks or trees planted in rows, have maintenance and land opportunity costs that are incurred by society.

The organization of this paper is as follows: Section 2 presents the theoretical model and discusses the social optimal allocation of green urban areas and the decentralization through the land market. Section 3 is devoted to our empirical approach and results. We first provide evidence on the concave shape of the share of green urban areas, then estimate residents' land choice, and finally quantify the economic benefits of green urban areas. The last section concludes the paper.

2.2 Theoretical model

We consider a circular monocentric city hosting a central business district (CBD) and a mass N of individuals. We denote by $b \in \mathbb{R}_0^+$ the distance between the CBD and the city border. The population density is defined as the number of individuals in a unit of area at distance r from the CBD and is denoted by the function $n : [0, b] \to \mathbb{R}^+$, which varies across the city.

In this model, we focus on small green urban areas that spread across the city and are closely accessible to the local community around its location.[3] Green urban areas

[3]Because of proximity of small parks (see Table 2.1), residents' travel costs to parks can be neglected

provide quick and frequent access to greenery, quiet, children's parks, socialization areas, etc. We consider the few blocks in the vicinity of a green urban area as our unit of area or patch and model the urban area in a continuous fashion. In a unit of area at distance r from the CBD, green urban areas offers a service $x : [0, b] \to [0, \overline{x}]$, $\overline{x} \in \mathbb{R}_0^+$, to the local community living in the vicinity. This service brings a level of amenity $a = \alpha x(r)$, although it necessitates the use of a fraction of land $\beta x(r)$ and maintenance costs $\gamma x(r)$. The parameters $\alpha, \beta, \gamma \in \mathbb{R}^+$ distinguish the amenity, land use and maintenance factors that affect green urban areas. Hence, the fraction of land used for residential purposes is given by $1 - \beta x$, and the maximum service level $\overline{x}$ is bounded by $1/\beta$. We assume absentee landlords, and the outside opportunity value of land is given by the agricultural land rent $R_A \in \mathbb{R}^+$. For simplicity, we consider that rural areas beyond the city border consist of private properties that do not provide green urban area service for city dwellers (e.g., private crop fields, fenced areas, etc.). We denote the land supply at distance r from the CBD by $\ell : [0, b] \to \mathbb{R}^+$ (e.g., $\ell = 2\pi r$ if the city lies in a plain disk). In summary, land at distance r from the CBD includes a surface $\beta x(r)\ell(r)$ of maintained green urban area and a residential area $[1 - \beta x(r)]\,\ell(r)$, and it hosts $n(r)\ell(r)$ residents who all benefit from the green urban area amenity $\alpha x(r)$.

Individuals consume a quantity z of nonhousing composite goods and a quantity s of residential space, while they benefit from the amenity a of a green urban area. They are endowed with the utility function $U(z, s, a)$, which is assumed to be concave and increasing for each variable. We assume that demands for nonhousing composite goods, residential space and amenity are gross substitutes such that U has negative second derivatives and positive cross derivatives.[4] As individuals are homogeneous, they work and earn the same income $w \in \mathbb{R}^+$ in the CBD. Workers incur a total commuting cost $t : [0, b] \to \mathbb{R}^+$ with $t(0) = 0$ and $\mathrm{d}t/\mathrm{d}r > 0$. The price of composite good z is normalized to 1 without loss of generality. From this point on and whenever there is no confusion, we dispense the functions a, ℓ, n, s, t, x, z and R with reference to distance r.

We first study the social optimal allocation and then the land market equilibrium.

---

here as a first approximation. This contrasts with the literature on the location of scarce public goods where the travel to public facilities must be considered (see Cremer et al., 1986 and followers).

[4]We do not formally disentangle the land space used for house and private garden. First, our data on urban land cover does not permit to identify each separate item. Second, the empirical literature is ambiguous about the substitution effect of private gardens on the use of parks (also called "compensation effect"). For instance, Talen (2010) and Caruso *et al.* (2018) find no significant relationship between the ownership of a private garden and the frequency of park visits. Finally, note that it can be shown that our theoretical results are unaltered if individuals have utility function $U(z, S, a)$ over a house service function S that is a homothetic consumption bundle $S(s_g, s_h)$ of land consumptions for house and private garden s_h and s_g. In this case, private gardens substitute for parks but the elasticity of substitution between private gardens and parks is the same as the one between houses and parks.

2.2.1 Social optimum

In an ideal world, green urban areas and residential structures should be combined to balance their social benefits and costs. Analysis of the social optimal structure of residential and green urban areas provides urban planners with viable directions for urban planning. Towards this aim, we assume a benevolent social planner who controls residential and green urban plots across the city.

As in Herbert and Steven (1960), we assume that the planner desires to set the same utility target $u \in \mathbb{R}$ for all urban residents.[5] She (the planner) minimizes the cost in the city

$$C = \int_0^b (tn + zn + R_A + \gamma x)\ \ell \mathrm{d}r,$$

subject to the target constraint $U(z, s, \alpha x) = u$ and land use constraint $sn = 1 - \beta x$. The total city population results from the accumulation of population density across the city: $N = \int_0^b n\ \ell \mathrm{d}r$. The planner chooses the profiles of consumption (z, s) and spatial allocations (n, x) as well as the border b. Since wages w are exogenous, this is equivalent to the maximization of total surplus $\mathcal{S} = wN - C$. After substitution of the population and land use constraints, this provides

$$\mathcal{S} = \int_0^b \left[\frac{w - t - z}{s} (1 - \beta x) - R_A - \gamma x \right] \ell \mathrm{d}r. \tag{2.1}$$

The planner then chooses the variables (z, s, x, b) that maximize $\mathcal{S}$ s.t. $U \geq u$.

The optimal consumptions (z, s) are given by the pointwise maximization of (2.1), which is equivalent to the set the maximum of the residential land value:

$$V \equiv \max_{z,s} \frac{w - t - z}{s} \text{ s.t. } U(z, s, a) \geq u. \tag{2.2}$$

Since the objective function in this expression decreases with z and U increases with it, the constraint is binding. We denote the consumption $\widetilde{z}(s, a, u)$ as the solution of $U(z, s, a) = u$. Because the utility function increases for all variables, we obtain $\widetilde{z}_s = -U_s/U_z < 0$, $\widetilde{z}_a = -U_a/U_z < 0$ and $\widetilde{z}_u = 1/U_z > 0$, while concavity of utility yields $\widetilde{z}_{ss} > 0$, where the subscripts denote partial derivatives. Denoting an individual's net income (net of commuting cost) as

$$y \equiv w - t,$$

the problem simplifies to

$$V = \max_s \frac{y - \widetilde{z}(s, a, u)}{s}.$$

The optimal use of residential space is given by the solution of the following first-order condition:

$$\widetilde{z}(s, a, u) - s\widetilde{z}_s(s, a, u) = y.$$

[5] This assumption avoids Mirrlees's discussion of the *unequal treatment of equals,* has a close link to competitive land equilibrium and yields first and second welfare theorems (Fujita 1989).

Because $\widetilde{z} - s\widetilde{z}_s > 0$ and $(\partial/\partial s)(\widetilde{z} - s\widetilde{z}_s) = -s\widetilde{z}_{ss} < 0$, this condition determines the optimal residential space $\widehat{s}(y, a, u)$. We denote the optimal consumption of commodity goods by the function $\widehat{z}(y, a, u) = \widetilde{z}[\widehat{s}(y, a, u), a, u]$ and the optimal bid land value as $V(y, a, u)$. Ceteris paribus, $\widehat{s}$ increases with lower y, and since $\widetilde{z}_s < 0$, $\widehat{z}$ increases with higher y. By the envelop theorem, the residential land value V rises with increasing y and a and decreasing u.

The planner's problem can then be rewritten as

$$\max_{x,b} \mathcal{S} = \int_0^b \left[V(y, \alpha x, u)\,(1 - \beta x) - R_A - \gamma x\right] \ell \mathrm{d}r. \tag{2.3}$$

Pointwise differentiation w.r.t. x provides the necessary condition for green urban area service,

$$\alpha V_a\,(1 - \beta x) - \beta V - \gamma = 0, \tag{2.4}$$

where V is evaluated at $(y, \alpha x, u)$. Using the land use constraint, we obtain the following optimality condition:

$$\alpha s n V_a = \beta V + \gamma. \tag{2.5}$$

This condition expresses the planner's balance between the benefit of green area amenities (LHS) and the costs of green urban land and its maintenance (RHS). Let $x^*(y, u)$ be the optimal profile of the green urban area service. Note that x^* never reaches its upper bound $\overline{x} = 1/\beta$. If it did, the population density n would fall to zero, and green urban areas would lead to maintenance and land costs but no amenities (zero LHS in (2.5)). It can nevertheless be that $x^* = 0$ if the LHS is smaller than the RHS for all $x \in [0, 1/\beta)$. In summary, $x^* \in [0, 1/\beta)$. For the sake of conciseness, we assume in this section that the second-order condition holds and concentrate our discussion on interior solutions.[6]

Interestingly, expression (2.5) can be recovered as Samuelson's optimality condition of public goods after some mathematical transformations (see Appendix A):

$$\alpha \frac{U_a}{U_z} n = \beta \frac{U_s}{U_z} + \gamma \tag{2.6}$$

where U_a/U_z is the marginal rate of substitution between commodities and amenities and U_s/U_z that between commodities and residential spaces. The Samuelson's optimality condition states that the sum of marginal rates of substitution for green area amenities equates the maintenance cost γ plus the marginal rate of substitution for residential land. This last element is novel in the context of public goods theory. In our context, it applies at the level of the patch because externalities are localized at this level. A green urban area is a local public good because its amenities equally benefit the residents localized in its patch, which has a population density n. The same local park indeed serves many residents. Green urban areas differ from usual (spaceless)

[6]The second-order condition is given by $\alpha^2 V_{aa}\,(1 - \beta x) - (1 + \alpha)\,\beta V_a < 0$. This condition holds provided that $\widehat{V}_{aa} < 0$.

local public goods in their land intensity. This has an impact on locations with lower usage of residential spaces because low space consumption is usually associated with higher marginal rates of substitution for space U_s/U_z. To our knowledge, this tradeoff has not been highlighted in the literature.

Fixing the variables z, s and n, we observe that a lower amenity parameter α, higher land use parameter β and higher maintenance cost parameter γ entice the planner to reduce the green urban area service x and also its land area αx. We also distinguish between the effects of population density and use of space. On the one hand, areas with low population density should accommodate smaller shares of green urban areas because they benefit fewer people. On the other hand, areas with small residential plots imply high marginal rates of substitution for space and should also be provisioned with smaller shares of green urban areas. In general, population density is low at the city edges, and residential plots are small near CBDs. Hence, *the planner is enticed to set smaller shares of green urban areas at both the city edges and CBDs and while it implements larger shares in intermediate locations*, which is the property that we will explore in the empirical section.

2.2.2 Comparative statics

Comparative statics on the service of green urban areas can be obtained by totally differentiating (2.4). Noting that $V_a > 0$, it is easy to see that

$$\frac{\mathrm{d}x^*}{\mathrm{d}\beta}, \frac{\mathrm{d}x^*}{\mathrm{d}\gamma} < 0,$$

such that land and maintenance costs have negative impacts on the service of green urban areas to residents. Other comparative statics are ambiguous as

$$\begin{aligned}
\frac{\mathrm{d}x^*}{\mathrm{d}\alpha} > 0 &\iff \left(\left(1+\alpha V_{aa}\right)\left(1-\beta x^*\right)-\beta V_a\right)x^* > 0,\\
\frac{\mathrm{d}x^*}{\mathrm{d}y} < 0 &\iff \alpha V_{ay}\left(1-\beta x^*\right)-\beta V_y < 0,\\
\frac{\mathrm{d}x^*}{\mathrm{d}u} < 0 &\iff \alpha V_{au}\left(1-\beta x^*\right)-\beta V_u < 0.
\end{aligned}$$

First, the effect of the amenity parameter is ambiguous. It can be shown from the first expression that when green urban area services provide very few amenities and use small land pieces ($\alpha, \beta \to 0$), the optimal service x^* and surface βx^* rise with the amenity parameter α. In fact, at very low levels, a higher α gives the planner incentives to raise service x^* because the parameter raises the effectiveness of the service. However, at high levels, a higher parameter α substitutes for service level x and entices the planner to reduce it.

Net income y is given by a worker's wage minus his/her commuting cost, which increases with distance to the CBD. Therefore, comparative statics on y highlight the

effect of distance to the CBD. The effect of net income can be deduced from the second expression above as follows:

$$\frac{\mathrm{d}x^*}{\mathrm{d}y} > 0 \iff \alpha n^* \frac{\mathrm{d}}{\mathrm{d}\ln y}\left(\frac{U_a}{U_z}\right) - \alpha n^* \frac{U_a}{U_z}\frac{\mathrm{d}\ln \widehat{s}}{\mathrm{d}\ln y} - \beta \frac{\mathrm{d}}{\mathrm{d}\ln y}\left(\frac{U_s}{U_z}\right) > 0. \tag{2.7}$$

The effect of higher net income on green area services depends on three factors: first, on the income elasticity of demand for residential spaces $\mathrm{d}\ln\widehat{s}/\mathrm{d}\ln y$; second, on the marginal rates of substitution between commodities and green urban areas U_a/U_z; and finally on the reaction of those marginal rates to increases in income $(\mathrm{d}/\mathrm{d}\ln y)\,(U_a/U_z)$ and $(\mathrm{d}/\mathrm{d}\ln y)\,(U_s/U_z)$. The latter reactions are related to the Engel curves in spaces (z,s) and (z,a). It can be shown that the marginal rate of substitution rises (falls) with income in those spaces if the Engel curves rise and bend upward (downward). In other words, higher income raises more demand for residential spaces s and amenities a than demand for commodities z. As a result, the effect of net income on green urban area services depends on the balance between the income effects on the demands for amenities and space. If income effects are identical (as under the Cobb-Douglas preferences), the marginal rates of substitution are invariant to income, and the above inequality holds for all net incomes. The optimal green urban area service then rises with net income. As a result, this optimal service also rises with wages and falls with distance from the CBD. In the end, the relative importance of each income effect is still an empirical issue for which we have found no information in the literature.

A similar comparative exercise can be performed on the impact of utility target u. One simply substitutes y for u in condition (2.7). The effect of target utility therefore depends on how it affects the use of residential space and the above marginal rates of substitution.

2.2.3 City border and population

Finally, the planner sets the city border so that the first-order condition w.r.t. b,

$$V(y,\alpha x^*,u)\,(1-\beta x^*) - \gamma x^* = R_A, \tag{2.8}$$

holds, where y and x^* are evaluated at $r = b$. Since x^* maximizes the LHS, the latter should be no smaller than $V(y,0,u)$. A sufficient condition for an optimal border $b^*(u)$ is that the LHS lies above zero at $r = 0$ and decreases for increasing r. That is,

$$V(w,0,u) \geq R_A, \tag{2.9}$$

$$\left(-V_y\frac{\mathrm{d}t}{\mathrm{d}r} + \alpha V_a\right)(1-\beta x) < \beta V + \gamma \quad \text{for } r \in [0,b^*], \tag{2.10}$$

where the second line is evaluated at $(y,\alpha x^*,u)$. We assume that these conditions hold. Comparative statics can be obtained by totally differentiating (2.8). Since (2.8) decreases with x, the optimal border b^* increases with the parameters that increase the value of

the LHS of (2.8) and reduce its RHS. Recalling that $V_y, V_a > 0 > V_u$ and $y = w - t$, it follows that

$$\frac{\mathrm{d}b^*}{\mathrm{d}\beta}, \frac{\mathrm{d}b^*}{\mathrm{d}\gamma}, \frac{\mathrm{d}b^*}{\mathrm{d}u}, \frac{\mathrm{d}b^*}{\mathrm{d}R_A} < 0 < \frac{\mathrm{d}b^*}{\mathrm{d}\alpha}, \frac{\mathrm{d}b^*}{\mathrm{d}w}. \tag{2.11}$$

Hence, As in the literature, cities also expand with higher wages and *cities spread when green urban areas provide higher amenities, use smaller pieces of land and require lower maintenance costs.* They shrink with higher agricultural rent and utility level.

The city population is given by $N = \int_0^{b^*(u)} \left[1 - \beta x^*\left(y, u\right)\right] / \widehat{s}(y, \alpha x^*, u)\ \ell \mathrm{d}r$, with $y = w - t$. Finally, the total surplus is given by

$$\mathcal{S}^*(u) = \int_0^{b^*(u)} \left[V(y, \alpha x^*, u)\left(1 - \beta x^*\left(y, u\right)\right) - R_A - \gamma x^*\left(y, u\right)\right] \ell \mathrm{d}r. \tag{2.12}$$

Therefore, by the envelop theorem, the change in surplus is given by $\mathcal{S}_u^*(u) = \int_0^{b^*(u)} V_u$ $\left(1 - \beta x^*\right)\ \ell \mathrm{d}r$, which is negative since $V_u < 0$. Higher utility targets u reduce the city surplus.

2.2.4 Land regulation

We are now equipped to discuss the impact of migration restriction and land regulation in cities. In practice, utility targets are determined by city planners (city officials and representatives) through their land regulation and migration policies. On the one hand, a city planner may opt for unrestricted migration so that u is determined by the outside utility level, e.g., $\overline{u} \in \mathbb{R}$. The population densities and levels adapt to migration pressure, and the city generates a surplus $\mathcal{S}^*(\overline{u})$, which we assume to be positive (otherwise, the planner has no incentive to create the city without external funding). On the other hand, a city planner may opt to restrict land use and population as to maximize incumbent residents' utility. Then, he/she targets the highest possible utility subject that is compatible with a positive surplus: $\mathcal{S}^*(u) \geq 0$. As the surplus decreases with increasing u, the highest utility, say u^*, is reached for a zero surplus: $\mathcal{S}^*(u^*) = 0$. In this case, green urban areas can be self-financed by land value. Indeed, $\mathcal{S}^*(u^*) = 0$ can be written as

$$\int_0^{b^*(u^*)} \left(1 - \beta x^*\right)\left(V - R_A\right)\ \ell \mathrm{d}r = \int_0^{b^*(u)} x^*\left(\gamma + \beta R_A\right)\ \ell \mathrm{d}r,$$

which shows an exact balance between the aggregate differential residential land value $(V - R_A)$ and the land and maintenance costs of green urban areas (the functions V and x^* being evaluated at $(y, \alpha x^*, u^*)$). This balance is a reminiscence of the Henry George theorem, by which a confiscatory tax on land would by itself finance a city's public goods, provided that the city reaches the size that maximizes residents' utility (see e.g. Arnott and Stiglitz, 1979). This paper adds two new elements to the standard

version of this theorem: *optimal green urban areas, first, are very localized public goods that provide unequal amenities and, second, require uneven land areas through the city*.[7]

It must be emphasized that self-financing takes place at the city level and not at the patch level. There indeed exist *cross-subsidies across urban dwellers*. Indeed, at any distance r from the CBD, residents create a value V of their residential land, while they generate a green urban area maintenance cost γx and a land opportunity cost R_A on the patch. This approach yields a resident's net value equal to $S \equiv V(1-\beta x) - R_A - \gamma x$, which is the integrand of (2.3). Using the first-order condition (2.4), one readily obtains that $\mathrm{d}S/\mathrm{d}r = V_y \mathrm{d}y/\mathrm{d}r = -V_y \mathrm{d}t/\mathrm{d}r < 0$. Hence, a resident's net value falls with increasing distance from the CBD. When the city management maximizes residents' utility and imposes a zero surplus such that $\mathcal{S}^* = \int_0^{b^*} S\,\ell \mathrm{d}r = 0$, it is clear that residents close to the city center bring positive net values, while those away from it bring negative values. Hence, the central population subsidizes the green urban areas at urban edges.

2.2.5 Competitive land market equilibrium

In most modern cities, residents freely choose their residential locations and spaces. They make their decisions according to the land rent values signaled in the urban land market. Given that urban green areas generate externalities to residents, it is important to highlight the conditions under which the competitive land market replicates the social optimum discussed in the previous sub-sections. We here discuss the equilibrium allocations in a competitive land market for an exogenous profile of green urban services $x : [0, b] \to [0, 1/\beta)$ and amenities $a = \alpha x$. A household's budget constraint is given by $z + sR + t \leq w$, where $R : [0, b] \to \mathbb{R}^+$ is the land rent function of distance to the CBD.

In a competitive land market equilibrium, each land slot is awarded to the highest bidder, and individuals have no incentives to relocate within and out of a city. Therefore, they reach the same utility level u^e, where the superscript e refers to the equilibrium value. Households bid up to $(w - z - t)/s$ for each unit of residential space. Their bid rent $\psi : [0, b] \to \mathbb{R}^+$ is a function of distance r from the CBD such that

$$\psi = \max_{s,z} \frac{y - z}{s} \quad \text{s.t.} \quad U(z, s, a) \geq u^e, \tag{2.13}$$

where net income $y = w - t$ is a function of distance to the CBD. As individuals compete for land, they raise their bids to make their participation constraint binding and obtain the equilibrium utility level u^e. Note that (2.13) is equivalent to the social optimal consumption choice (2.2). Therefore, households' optimal consumptions are given

[7] Arnott and Stiglitz (1979) show that the Henry George theorem holds in the presence of a "green belt" in a model where the latter is located at the city border, provides a global pure public good to all residents, uses no land and requires no maintenance. In this paper green urban areas spread across the city, provide localized public goods and are intensive in land and maintenance.

by the functions $\widehat{s}(y,a,u^e)$ and $\widehat{z}(y,a,u^e)$, and the bid rent, by $\widehat{\psi}(y,a,u^e) = V(y,a,u^e)$. The bid rent inherits the properties of V. That is, $\widehat{\psi}_y, \widehat{\psi}_a > 0$, while $\widehat{\psi}_u < 0$.

A competitive land market equilibrium is defined as the set of functions (z,s,R,n) and scalars (b,N,u^e) satisfying the following four conditions. First, individuals choose their optimal consumptions: $z = \widehat{z}(y,a,u^e)$ and $s = \widehat{s}(y,a,u^e)$. Second, land is allocated to the highest bidder: $R = \max\{\widehat{\psi}(y,a,u^e), R_A\}$, with $R = \widehat{\psi}(y,a,u^e)$ if $n > 0$, and $R = R_A$ if $n = 0$. Third, the land market clears everywhere: $n\widehat{s}(y,a,u^e) = (1-g)$ if $n > 0$. Finally, the total population conforms to its density: $N = \int_0^b n\, 2\pi r \mathrm{d}r$. Here, N is taken as exogenous in a closed city model, while u^e is exogenous in an open city.

Within a city, equilibrium land rents are given by the winning bids such that $R = \widehat{\psi}(y,a,u^e)$. Since bid rents ψ increase with net income y and amenities a, the equilibrium land rent R falls with distance from the CBD but rises with the proportion of green urban area. Importantly, *at the social optimal amenity a and utility level u, consumptions in a competitive land market match the social optimal ones exactly, while land bid rents ψ and land rents R match social land values V.* This finding is reminiscent of the social optimum property in Herbert and Steven's (1960) model, where competitive land market equilibria are socially optimal. The land market is then allowed to decentralize the choices of land and commodity consumption. However, this applies only if green urban areas are optimally set in our framework.

The equilibrium population density is equal to $n^e = (1-g)/\widehat{s}(y,a,u^e) \geq 0$, while the equilibrium population aggregates the population density across the urban area as

$$N^e = \int_0^{b^e} \frac{1-\beta x}{\widehat{s}(y,a,u^e)}\, 2\pi r \mathrm{d}r.$$

The following proposition summarizes the above discussion.

Proposition 1 *The land and commodity consumption in the competitive land market equilibrium and social optimum coincide if green urban areas are provided at the socially optimal levels and if the planned and equilibrium utility levels u and u^e match.*

A competitive land market is a powerful mechanism to decentralize consumption decisions. It can be checked that land tax does not affect goods and land consumption choices; thus, land taxes may be used by city planners to finance green urban areas.

To obtain more analytical results, we focus on narrower classes of preferences.

2.2.6 Specific preferences

In this subsection we present three classes of preferences for which the optimal share of green urban space is a concave function of the distance to city center and the optimal population density is a decreasing function of it. For two classes, the optimal share of green urban area can be bell-shaped.

2.2.6.1 Hyperbolic preferences

We here assume that individuals are endowed with the utility function $U(z,s,a) = z - \theta/(2s) + a$ where preferences for residential space are represented by an hyperbolic function parametrized by θ.[8] Accordingly, we get the consumptions $\widetilde{z} = u - a + \theta/(2s)$, $\widehat{s} = \theta/(y+a-u)$, $\widehat{z} = (y-a+u)/2$ and land value $V = (y+a-u)^2/(2\theta)$ (see mathematical details in Appendix A). We focus on the case of positive consumption so that $y + a - u > 0$. Then, the marginal benefit of larger green urban service (LHS of 2.4) is proportionate to

$$\frac{\alpha}{\theta}(y+\alpha x-u)(1-\beta x) - \frac{\beta}{2\theta}(y+\alpha x-u)^2 - \gamma.$$

This expression is nil at an interior solution $x^* > 0$ and negative at the corner solution $x^* = 0$. The optimal green urban area service x^* can be shown to be equal to the interior solution

$$x_+^* = \frac{1}{\beta} - \frac{2}{3\alpha}\left(y - u + \frac{\alpha}{\beta}\right) + \frac{1}{3\alpha}\sqrt{\left(y-u+\frac{\alpha}{\beta}\right)^2 - \frac{6\theta\gamma}{\beta}},$$

if $y > \sqrt{6\theta\gamma/\beta} - \alpha/\beta + u$ and to the zero corner solution otherwise. It can readily be shown that x_+^* is a concave function of y and reaches a maximum at the net income $y^{\max} = \sqrt{8\theta\gamma/\beta} - \alpha/\beta + u$. Hence, if net incomes are larger and then respectively lower respectively than $y^{\max}$ at the city center and border, the green urban area increases and then respectively decreases with larger net incomes. Since y falls with distance to city center, the share of green urban area increases and then decreases with distance from the city center.

The optimal population density is given by

$$n_+^* = \frac{\beta}{9\theta\alpha}\left(y - u + \frac{\alpha}{\beta} + \sqrt{\left(y-u+\frac{\alpha}{\beta}\right)^2 - 6\theta\frac{\gamma}{\beta}}\right)\sqrt{\left(y-u+\frac{\alpha}{\beta}\right)^2 - \frac{6\theta\gamma}{\beta}}$$

for the interior solution. It is simply equal to $(y-u)/\theta$ at the corner solution. In both cases it increases with net income y and therefore decreases with distance from city center.

The reason for the concavity in green urban area density then lies in the balance between opportunity cost of land and density of residents as shown in the Samuelson condition (2.6). In the vicinity of the CBD, the social value of land is high because of shorter commutes. This entices the planner to increase population density at the expense of access to green urban areas. By contrast, the social value of land is low at the vicinity of the city border because of longer commutes and entices the planner to compensate individuals with larger residential plots and lower population density. In those low density locations, green urban areas yield smaller social benefit and the

[8] This yields a demand for residential space that has a price elasticity between one and zero (see Mossay and Picard 2011).

planner reduces their surfaces. So, it is the high land social value that refrains the planner to organize large green urban areas about the CBD and it is the low population density that refrains her to maintain large green urban areas at the city fringe.

Proposition 2 *Under hyperbolic preferences, population density falls with distance from CBD while the share of green urban area is a concave function of distance from CBD. The share of green urban area peaks at a location that moves away from CBD as cities are richer and less regulated.*

2.2.6.2 Cobb-Douglas preferences

We here assume the Cobb-Douglas utility $U = z^{1-\phi-\varphi}s^{\phi}e^{a\varphi}$ with $\phi, \varphi, (1-\phi-\varphi) \in (0,1)$. We compute $\tilde{z} = (us^{-\phi}e^{-a\varphi})^{\frac{1}{1-\phi-\varphi}}$, which gives

$$\widehat{z} = \frac{1-\varphi-\phi}{1-\varphi}\, y, \tag{2.14}$$

$$\widehat{s} = \left(\frac{1-\varphi}{1-\varphi-\phi}\right)^{\frac{1-\varphi-\phi}{\phi}} \left(uy^{-(1-\varphi-\phi)}e^{-a\varphi}\right)^{\frac{1}{\phi}}, \tag{2.15}$$

$$V = \kappa^{-1}\left(uy^{-(1-\varphi)}e^{-a\varphi}\right)^{-\frac{1}{\phi}} \tag{2.16}$$

where $\kappa = (1-\varphi)^{\frac{1-\varphi}{\phi}}(1-\varphi-\phi)^{-\frac{1-\varphi-\phi}{\phi}}$. Condition (2.6) becomes

$$e^{\frac{\alpha\varphi}{\phi}x}(\alpha\varphi - \beta\phi - \beta\alpha\varphi x) = \kappa\gamma y^{-\frac{1-\varphi}{\phi}}u^{\frac{1}{\phi}}. \tag{2.17}$$

As shown in Appendix A, there exists a unique interior optimal service level $x^* > 0$ if the green area amenities per surface unit are sufficiently large $\alpha/\beta > \phi/\varphi$, the maintenance cost γ is sufficiently low and the net income y is sufficiently high. Otherwise, there is a corner solution $x^* = 0$.

It is also shown that for any interior solution x^*,

$$\frac{\mathrm{d}x^*}{\mathrm{d}y} > 0 \quad \text{and} \quad \frac{\mathrm{d}^2x^*}{\mathrm{d}y^2} < 0.$$

Hence, since $y = w - t(r)$, the optimal share of green urban areas $g^* = \beta x^*$ increases with wage w and decreases with distance from the CBD r. The optimal share is also a concave function of wage and distance from the CBD. Finally, since $\widehat{s}$ falls in y and a and because $a = \alpha x^*$ increases with y, $\widehat{s}(y, \alpha x^*, u)$ also falls with the latter. Then, the population density increases with higher wages and falls with longer distances to the CBD.

Proposition 3 *Under the Cobb-Douglas preferences, population density is a decreasing function of distance from the CBD, while the share of green urban areas is a decreasing and concave function of this distance.*

2.2.6.3 Stone-Geary preferences

In this section we extend the Cobb-Douglas preferences with a Stone-Geary utility component for residential space such that $U(z,s,a) = e^{\varphi a} z^{1-\phi-\varphi}(s-s_0)^{\phi}$ with $\phi, \varphi \in (0,1)$ and $0 < \phi + \varphi < 1$. In this case, $s_0 > 0$ is a minimum housing consumption that raises the marginal utility for residential space disproportionally more at low residential consumptions. Under this specification, the elasticity of substitution between residential space and either composite good z or amenity a falls as the individual is given a residential surface closer to s_0. As we will see bellow, the planner is then unable to increase population density to high levels in city centers and devote much less land for green urban areas there.

The composite good consumption that gives the utility level u is given by $\tilde{z} = \left[ue^{-\varphi a}(s-s_0)^{-\phi}\right]^{\frac{1}{1-\phi-\varphi}}$. The residential land value $V = \max_s \left[y - \tilde{z}(u,a,s)\right]/s$ yields the optimal residential space given by the condition (2.6)

$$(s-s_0)^{-\frac{1-\varphi}{1-\phi-\varphi}}\left(\frac{1-\varphi}{1-\phi-\varphi}s - s_0\right) = yu^{-\frac{1}{1-\phi-\varphi}}e^{a\frac{\varphi}{1-\phi-\varphi}}$$

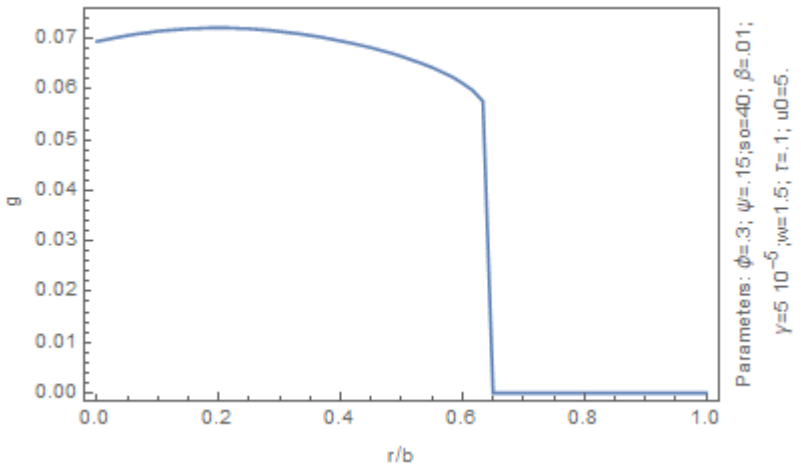
which is a transcendental equation that nevertheless accepts a unique solution $\hat{s} > s_0$. It can be shown that $\mathrm{d}\hat{s}/\mathrm{d}y < 0$, $\mathrm{d}\hat{s}/\mathrm{d}a < 0$ and $\mathrm{d}\hat{s}/\mathrm{d}u > 0$. Hence, since y falls with larger commuting cost t and distance from center r, the residential land consumption decreases with smaller distance to city center. However land consumption is bounded from below by s_0. Setting $\widehat{z} = \tilde{z}(u,a,\hat{s})$ and using $a = \alpha x$, the optimal green urban area service x is given by

$$\frac{\alpha\varphi}{1-\phi-\varphi}\widehat{z}\frac{(1-\beta x)}{\hat{s}} - \frac{\beta\phi}{1-\phi-\varphi}\frac{\widehat{z}}{\hat{s}-s_0} - \gamma = 0$$

An interior maximum for x^* exists when the LHS of this expression falls to a zero value. A corner solution $x^* = 0$ takes place when the LHS of this expression is negative for all $x \in [0, 1/\beta)$. The optimal green urban area service can only be found numerically. Figure 1 displays an example of a concave profile of x^*: as one moves away from the city center, x^* first increases and then decreases. Note that under this specification, x^* abruptly falls to zero at some distance from the city center. For higher opportunity costs of land R_A, this profile is truncated from the right so that the profile is concave and increasing. Our numerical exploration found only concave profiles for sufficiently high s_0. For sufficiently small s_0, the profile resembles to the decreasing and concave functions as the ones found under Cobb-Douglass preferences.

Proposition 4 *Under Stone-Geary preferences, the optimal residential space increases with distance from CBD. Numerical exercises show that the share of green urban area is a concave and bell-shaped function of distance from CBD for sufficiently high* s_0.

Figure 1: Share of green urban space under Stone Geary preferences



The preferences reviewed in the last three subsections show that green urban space profiles are generally concave function of distance from the center. The profile can be bell-shaped if the elasticity of substitution between residential space and either the composite good or amenity decreases with smaller residential space. The following section shows that such theoretical results are not inconsistent with the data on European cities.

2.3 Empirical analysis

In this section, we compare the model prediction to the actual green urban patterns in European cities.

2.3.1 Data

In this paper, we use the dataset on urban land use from the Urban Atlas 2006, implemented by the Global Monitoring for Environment and Security (GMES) service and provided by the European Environment Agency (EEA), for the time period 2005-2007. The dataset offers a high-resolution map of land use in European urban areas, containing information derived from Earth observations and backed by other reference data, such as navigation data and topographic maps. The Urban Atlas uses Earth observation satellite images with 2.5 m spatial resolution.[9] According to the GMES, the dataset covers the functional urban areas (FUAs) of the EU cities with at least 100,000 inhabi-

[9]GMES maps have a 100 times higher resolution than traditional maps in the Corine Land Cover inventory produced since 1990.

tants.[10] FUAs include land with both commuting distance and time below the critical levels defined by Eurostat.[11] The dataset includes all capital cities and covers nearly 300 of the most populous towns and cities in Europe (EU 27).[12] Figure 2 displays the urban areas covered by these cities.

The Urban Atlas provides a classification of city zones that allows for a comparison of the density of residential areas, commercial and industrial zones and extent of green areas. In this paper, we use the data on "green urban areas" (class 14100), which are defined as artificial nonagricultural vegetated areas. They consist of areas with planted vegetation that is regularly worked and/or strongly influenced by humans. More precisely, first, green urban areas include public green areas used predominantly for recreational use (gardens, zoos, parks, castle parks, cemeteries, etc.). Second, suburban natural areas that have become and are managed as urban parks are included as green urban areas. Finally, green urban areas also include forest and green areas that extend from the surroundings into urban areas with at least two sides being bordered by urban areas and structures and containing visible traces of recreational use. Importantly, for our study, green urban areas do not include private gardens within housing areas, buildings within parks, such as castles or museums, patches of natural vegetation or agricultural areas enclosed by built-up areas without being managed as green urban areas. It must be noted that green urban areas belong to the Urban Atlas' class of "artificial surfaces", which include all nonagricultural land devoted to human activities.[13] This class is distinguished from the agricultural, seminatural areas and wetlands, forest areas and water areas devoted to nonurban activities.

We select the (oldest) town hall locations as the CBDs. Then, we create a set of annuli (rings) around each CBD at 100 m intervals. We define the "annulus land area" as the intersection of the annulus and the land within the urban zone area reported by the GMES. This area includes artificial surfaces, agriculture, seminatural areas, wetlands and forest but does not include water areas because those seas and oceans are not appropriate for potential human dwellings. We then compute the share of green urban area as the ratio of the surface of green urban area to the total land in the annulus land area for each annulus . Figure 3 displays the annuli and the land use of green urban areas (green color) for Dublin.

Whereas urban theoretical models usually assume a neat frontier between residen-

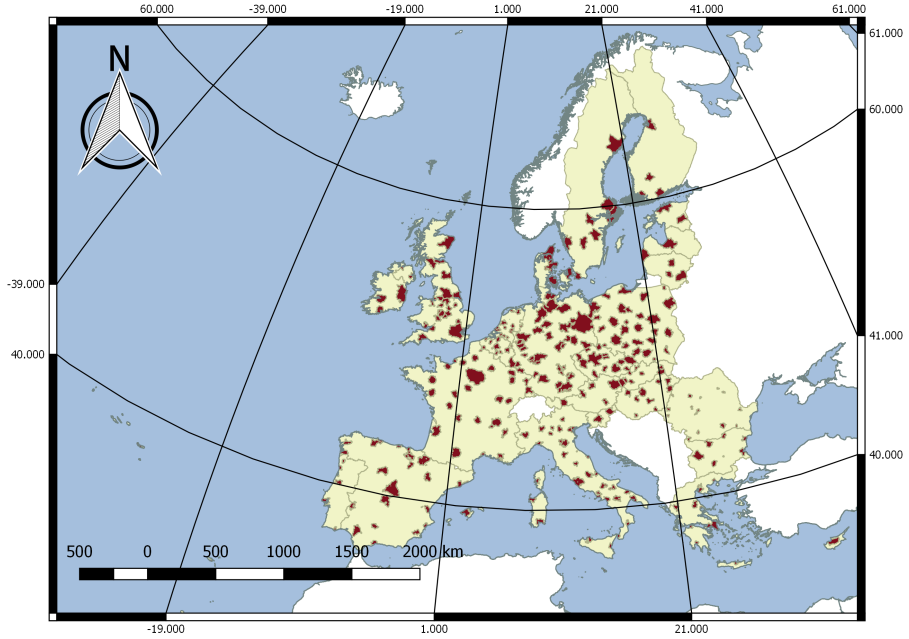
[10] See the definition in the Urban Audit database and European Environmental Agency, GMES Urban Atlas Metadata. Link: `https://land.copernicus.eu/local/urban-atlas` (accessed on Jan 25, 2018).

[11] See the definition in the Urban Audit in EEA, 2015, and the details in Appendix B.

[12] Austria, Belgium, France, Germany, Bulgaria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom.

[13] In addition to green urban areas, artificial surfaces include urban areas with dominant residential use, innercity areas with central business district and residential use, industrial, commercial, public, military and private units, transport units, mines, dump and construction sites, and sports and leisure facilities.

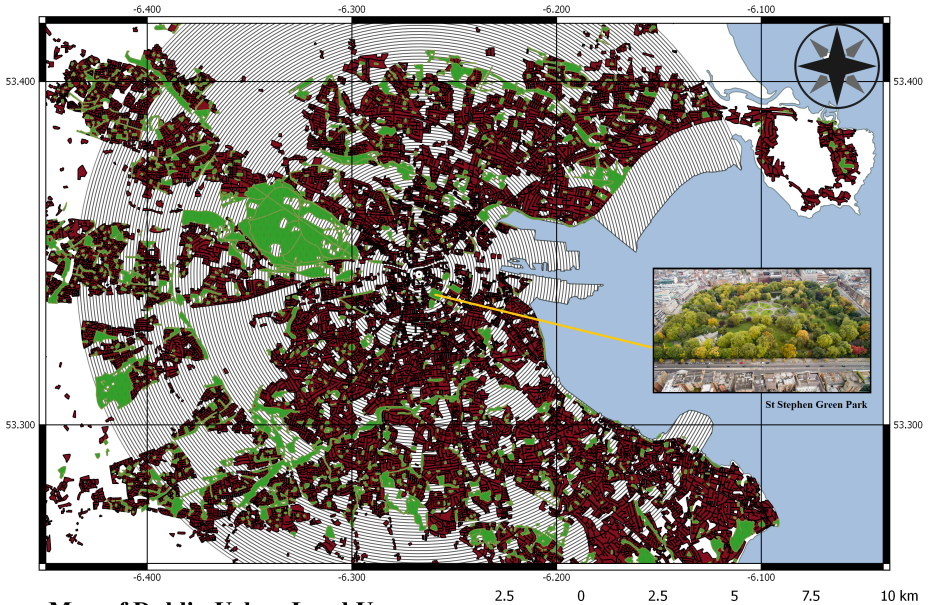
Figure 2: GMES 2006 maps



Maps of GMES 2006

Created using GMES 2006map boundary, EU map boundary data in Quantum GIS

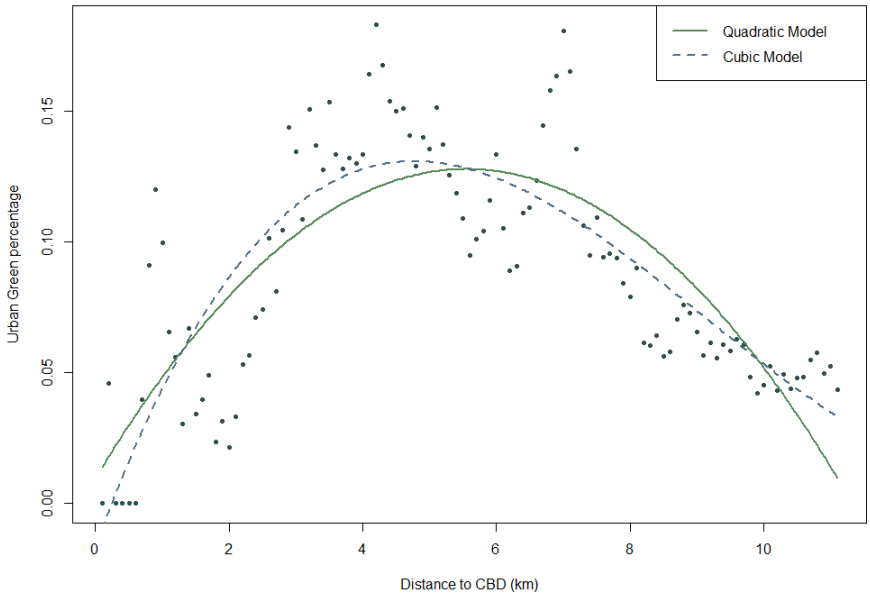
Figure 3: Dublin Land Use maps



Map of Dublin Urban Land Use

Created using GMES 2006 and Quantum GIS

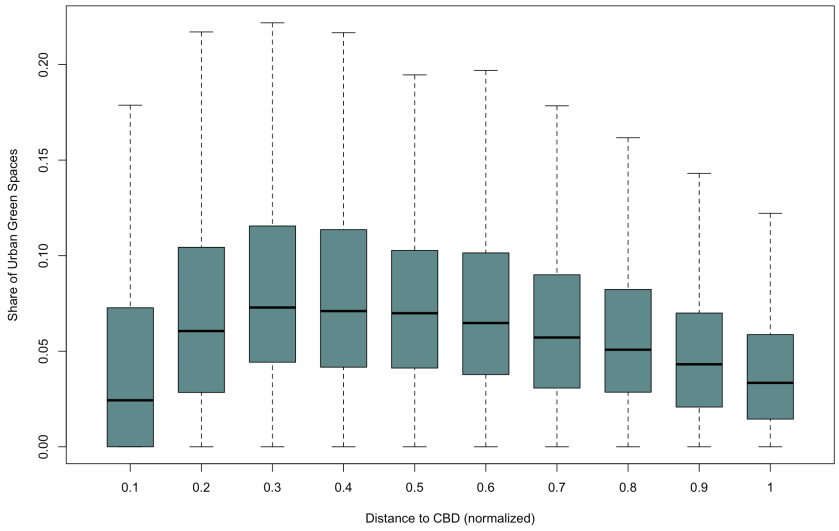
Figure 4: Dublin Green Urban Space



tial and nonresidential spaces, urban data do not provide a clear separation between residential locations and agricultural areas and forests. In this paper, we choose to fix the city borders to the annulus for which the ratio of residential surfaces over the annulus land area falls below 20%. Residential surfaces include urban areas with dominant residential use and innercity areas with central business district and residential use. They are shown in red in Figure 3. As shown in the sequel and Appendix B, the use of other thresholds does not lead to qualitative differences in our empirical results. We define the distance from the CBD, *dist*, as the distance from the CBD of the annulus and the relative distance, *rdist*, as this distance from the CBD divided by the distance between the CBD and the city border. Figure 4 plots the share of green urban areas in each annulus of Dublin city as a function of distance. It also includes the quadratic and cubic fit function. The non-monotone relationship is quite apparent. Figure 5 presents a box-plot of those shares as a function of the relative distance for all 305 cities. The non-monotone is also apparent although there is now a lot of heterogeneity related to country and city specificities.

We include several controls that do not depend on the relative distance to the CBD. A country dummy vector accounts for a country's specific urban regulations and wealth. We also divide the sample into three city groups: small cities, with a population below a half million; medium-sized cities, with a population between a half and one million; and large cities, with a population exceeding one million. The dummy vector *city_size* includes the fixed effect on each city group.

Figure 5: Green Urban Space in 305 EU Cities



In addition to the GMES, we use the population density from the European population grid.[14] We calculate the population mass at each distance to the city center and redistribute the population to the residential area in each annulus.[15] Because the Eurostat population grid does not cover Cyprus, we exclude Lefkosia, Cyrus. Other measures of cities' exogenous geographical characteristics are taken from the E-OBS database.[16] We control for these exogenous geographical characteristics because they may affect residential choices. We finally measure the city populations as the number (millions) of inhabitants living in the city and greater city (CGC) areas, as defined and reported in the Eurostat databases.[17] Table 2.2 presents the summary statistics.

On average, green urban areas account for nearly 6% of the total surface of city areas. Cities have a rather heterogeneous share of green urban areas. Figure 6 displays a histogram of the shares of green urban areas across the studied cities. There is a large dispersion in the average green urban areas across EU cities. In Figure 6, the city with lowest share of green urban area (0.62%) is Limerick, Ireland, and the one with the highest share (42.6%) is Karlovy Vary, Czech Republic. The latter is a spa resort city, which offers many green areas to its visitors. The former city includes few land surfaces classified as green urban areas because it also has many agricultural and

[14]For more information on the European population grid, please check the technical report of the GEOSTAT 1A project from Eurostat.

[15]Residential areas are called 'urban fabrics' in the GMES.

[16]The E-OBS database provides detailed data on the daily temperature, daily precipitation, sea level pressure and elevation across Europe. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (`http://ensembles-eu.metoffice.com`) and the data providers in the ECA&D project (`http://www.ecad.eu`).

[17]For more details, see metadata files for *urb_esms* in the Urban Audit database of the Eurostat website.

Table 2.2: Summary statistics

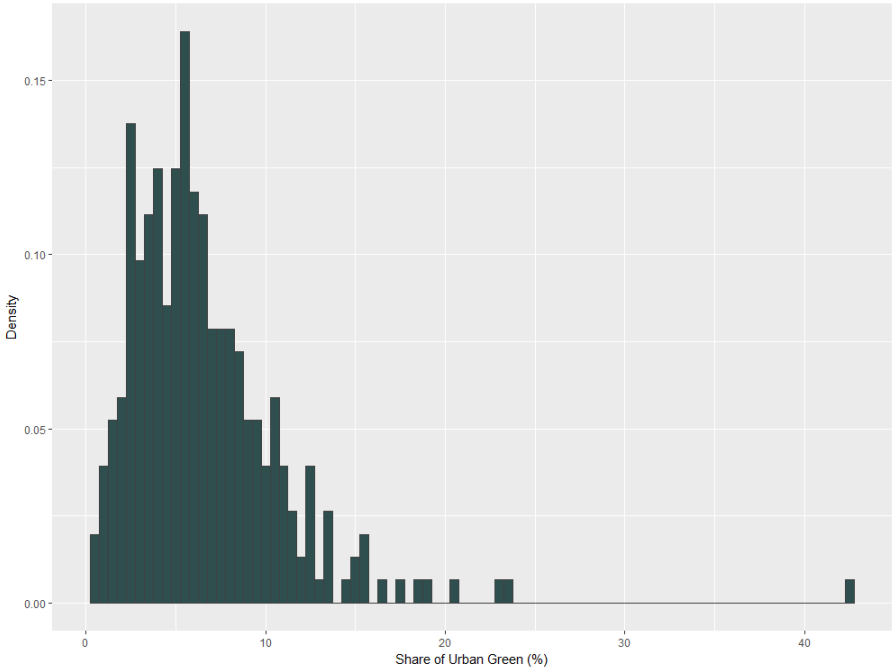
	average	sd	min	max	observations
City border (km)	4.3	3.2	1.0	24.0	305
City area (km^2)	84	174	1	1809	305
Number of annuli	43	32	10	240	305
Population in FUA (millions)	0.79	1.29	0.06	12.10	301
Population in CGC (millions): all cities	0.44	0.79	0.03	8.17	305
small cities	0.20	0.11	0.03	0.50	240
medium sized cities	0.64	0.13	0.51	0.98	36
large cities	2.25	1.66	1.01	8.17	29
Total share of green urban area in city (UGS) (%)	6.5	4.3	0.6	42.6	305
Highest share of UGS (%)	19.2	8.8	2.4	70.0	305
Highest share of UGS (%) (kernel smoothing)	9.5	5.0	0.4	42.2	305
Distance of highest share of UGS (km)	1.3	1.5	0.1	15.8	305
Distance of highest share of UGS (km) (kernel smoothing)	1.7	1.8	0.1	15.5	305
GDP per capita (€1000/hab.)	26.88	13.06	6.00	83.70	305
Household income (€1000/hab.)	15.46	5.63	3.70	30.90	304
City Geographical Controls					
Elevation (m)	212	210	−2	1,614	305
Average temperature at Jan 01 (^{o}C)	2.30	4.62	−8.48	15.57	305
Average temperature at July 01 (^{o}C)	19.18	12.27	12.27	28.70	305
Average daily precipitation (mm/day)	1.91	0.60	0.48	4.45	305
Share of Urban Green Land (%)	6.57	4.38	0.62	42.69	304
Share of Residential Land (%)	34.03	4.99	12.85	47.42	304

Note: The GMES database released on May 2010 reports only 301 FUAs for the time period 2005-2007. Cities without FUAs reporting are Wrexham and Derry (UK), and Gozo (Malta). Aix-en-Provence shares the same FUA as Marseille. We use the Nadaraya-Watson Gaussian Kernel to smooth variations of annuli values.[18] GDP per inhabitants and Household income are taken from Regional Economic Accounts from Eurostat at NUTS3 and NUTS2 level respectively. Note that in Eurostat database, household income level exists only at NUTS2 level. In eurostat database for household income at NUTS2, there is no data for Luxembourg (NUTS2 code LU00); therefore, there is only 304 cities instead of 305 cities. The number of inhabitants in each annuli is calculated based on Eurostat Population Grid. As Eurostat Population Grid 2006 does not cover Cyrus; hence, we also drop the city cy001l_lefkosia in our database. The total number of annuli are calculated for 303 cities excluding lu001l_luxembourg and cy001l_lefkosia. For city geographical controls, we take into account the average for period 1995-2010 for each city.

seminatural lands that can be used for urban green amenities. These outliers do not affect our results. Spots with the highest densities of urban green areas are located, on average, at $1.3 - 1.7$ kilometers from the CBD. City elevation also varies greatly, from two meters below sea level in Amsterdam, Netherlands, to $1,614$ meters above sea level in the mountainous city of Innsbruck, Austria. European cities belong to a mild climate zone, with temperatures varying between -8 and $+28$ degrees Celsius at the lowest and highest day temperature in winter and summer (measured on January 1 and July 1, respectively, for the period 1995-2000).[19] The average population density is approximately 4,400 inhabitants per square kilometer and ranges from 1,000 to 9,800 inhabitants per square kilometer. These numbers are reasonable because the database concentrates on the core of urban areas with no agricultural or seminatural land use.

[19] We use observations from the E-OBS databases from the EU-FP6 project (for details, see the references). Our samples do not contain some northern European cities in Iceland and Norway.

Figure 6: Average Share of Green Urban Space Distribution



2.3.2 Urban green area profiles

In this subsection, we compare our theoretical predictions with the empirical properties of green urban areas in EU cities. According to the theoretical model, the social optimal land share devoted to green urban areas is a concave and possibly nonmonotonic function of distance to the CBD; it first increases and then decreases as one moves away from the CBD. This pattern reflects the tradeoff between the high land values in the center, which make green urban areas too costly, and the too sparse population in the periphery, which associates green urban areas with too low social benefits. The aim of this subsection is to test the concavity of the land share of green areas in the studied European cities. In addition, we test whether this share is nil at the CBD. We propose the following reduced form model:

$$g_{ijc} = \eta_1 \, rdist_{ijc} + \eta_2 \, rdist_{ijc}^2 + \eta_3 \, rdist_{ijc}^3 + X_j + \epsilon_{ijc}, \tag{2.18}$$

where g_{ijc} is the land share of green urban areas in annulus i (ranked according to distance to the CBD) of city j in country c. We study both quadratic and cubic models, where the coefficient η_3 is constrained to zero in the first case. To allow for comparison across cities, we define the covariate of the relative distance of an annulus $rdist_{ijc}$ and add the city fixed effects X_j as controls. Urban green areas are likely to lie on adjacent annuli such that green urban area densities may not be independent observations, which biases the estimation. Additionally, more distant annuli aggregate more surface; thus, the estimation may suffer from heteroskedasticity. Hence, we report the Huber-

White heteroskedastic-consistent estimation of the standard residual errors. Table 2.3 presents the results from the regression of equation (2.18).

Table 2.3 reports a negative and significant correlation with the square of the distance to the CBD, which suggests that the hypothesis of a concave profile for the share of green urban areas g should not be rejected. Columns (1) to (3) report the results with and without country and city controls. As can be seen, these controls do not have a large effect on the amplitude and significance of results. Columns (4) to (6) present the results for the subsamples of small cities (population below a half million), medium-sized cities (population between a half and one million) and large cities (population exceeding one million), respectively. The signs are not altered, which corroborates the idea of concave profiles. A formal testing of concavity requires an examination of the p-value of a one-sided t-test of the respective coefficients. In Appendix B, we show that the p-values are very low; thus, the joint hypothesis of $\eta_1 > 0$ and $\eta_2 < 0$ cannot be rejected at the 99% confidence. We also show that the results are robust to various observation weightings, which suggests that misspecification issues can be excluded.

Columns (4) to (6) also show that the (absolute values of) amplitudes of the coefficients increase with city size. The shares of green urban areas reach higher levels in larger cities. Indeed, we can compute the average shares of green urban areas in the CBD ($rdist = 0$), at its peak location ($rdist = -\eta_1 / (2\eta_2)$) and at the city border ($rdist = 1$). The average shares of green urban areas in the CBD are given by the intercepts of the regression models, which are computed as the following averages of the city fixed effects in Columns (4) to (6): 0.067^{***}, 0.059^{***}, and 0.058^{***}. With this information, we deduce that, on average, the shares of green urban areas in small cities (Column (4)) rise from 6.7%, peak at 7.7% at $rdist = 33\%$ of the border and fall back to 3.6% at the city border. The shares in medium-sized cities (Column (5)) rise from 5.9%, peak at 9.0% at 46% of the border and fall back to 4.6%. Finally, large cities (Column (6)) have shares that rise from 5.8%, reach their peaks at 9.9% at 45% of the border and fall back to 3.7%. Finally, since the share of green urban areas over an entire city is given by $\int_0^1 \left(\eta_0 + \eta_1\xi + \eta_2\xi^2\right) \mathrm{d}\xi = \eta_0 + \frac{1}{2}\eta_1 + \frac{1}{3}\eta_2$, we can compute the average shares to be 6.7%, 7.7% and 8.1% for small, medium-sized and large cities, respectively.

Our model predicts a concave profile for the share of green urban areas rather than the quadratic profile in Columns (1) to (6). Columns (7) to (9) present the same regression analysis as in Columns (1) to (3) for a cubic regression model. One can observe that the coefficients are significantly different from zero. The results suggest a hump-shaped profile for the share of green urban areas, with the coefficients for the square of distance being significantly negative. In this cubic model, local concavity is given by the sign of the second derivative $g'' = 2\eta_2 + 6\eta_3 \times rdist$, with $rdist \in [0, 1]$. Since η_2 is negative in Table 2.3, the share of green urban areas is certainly a concave function in areas sufficiently close to the CBD. However, because η_3 is positive

Table 2.3: Profile of green urban spaces (EU27): baseline results

	Dependent variable: Share of Urban Green Space											
	Quadratic						Cubic					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance	0.102***	0.102***	0.102***	0.062***	0.134***	0.181***	0.297***	0.292***	0.294***	0.232***	0.299***	0.424***
	(0.016)	(0.016)	(0.016)	(0.018)	(0.040)	(0.037)	(0.038)	(0.038)	(0.039)	(0.047)	(0.082)	(0.089)
Distance_square	−0.127***	−0.127***	−0.127***	−0.093***	−0.147***	−0.202***	−0.598***	−0.588***	−0.591***	−0.499***	−0.550***	−0.802***
	(0.014)	(0.014)	(0.014)	(0.015)	(0.035)	(0.034)	(0.080)	(0.079)	(0.080)	(0.097)	(0.174)	(0.196)
Distance_cubic							0.307***	0.300***	0.302***	0.263***	0.264**	0.396***
							(0.048)	(0.047)	(0.048)	(0.058)	(0.108)	(0.120)
Country FE		✓						✓				
City FE			✓	✓	✓	✓			✓	✓	✓	✓
Sample of cities	All	All	All	Small	Medium	Large	All	All	All	Small	Medium	Large
Observations	13,091	13,091	13,091	7,982	2,106	3,003	13,091	13,091	13,091	7,982	2,106	3,003
Adj. R^2	0.046	0.148	0.369	0.367	0.338	0.383	0.056	0.158	0.379	0.374	0.346	0.402
df	13,088	13,063	12,785	7,741	2,068	2,972	13,087	13,062	12,784	7,740	2,067	2,971
F Stat.	79.63***	20.50***	26.13***	20.32***	30.05***	63.00***	54.35***	19.60***	27.13***	20.82***	30.37***	66.07***

Note: Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses. The row "df" reports the degree of freedom. The table shows results from the regression of share of green urban space within each annulus on the relative distance and its square to the city center using control variables and a constant. For comparison purpose, the relative distance of an annulus is given by its distance from city center divided by the distance of the farthest annulus in the city. Columns (2) to (3) sequentially includes controls on country and city fixed effects. Columns (4) to (6) restrict the sub-samples to small cities (population lower than half million), medium size cities (population lower between half and one million) and large cities population larger than one million). Same strategies are applied to cubic regressions (7) to (12).

in Table 2.3, concavity fails at greater distances from the CBD. Using the results in Column (9), the convexity coefficient is equal to $g''(0) = 2\eta_2 = -1.182$ in the CBD and $g''(1) = 2\eta_2 + 6\eta_3 = 0.630$ at the city edge. Hence, the profiles are increasing and concave near the CBD but convex at city edges. However, such convexity is not inconsistent with our theory because the share of green urban areas must have convex kinks when reaching zero. Finally, using Column (9), the share of green urban areas reaches a maximal value at 33% of the distance between the CBD and city border (solution of $g' = \eta_1 + 2\eta_2 \times rdist + 3\eta_3 \times rdist^2 = 0$).

We run a series of robustness checks, and the results are presented in Table 2.4. We study variations around the setup of Column (3) in Table 3, with both city and country fixed effects. First, as Figure 4 shows, there exists substantial serial correlation in the share of green urban areas, which questions the assumption of homoskedatisticity. Column (1) presents similar results under OLS without correction for heteroskedasticity and therefore suggests that heteroskedasiticy is not an important issue. Second, we check issues of truncation and size observation units. The surface areas of the annuli rise linearly with distance to their center. Therefore, the annuli about a CBD measure green urban areas within smaller surfaces and may have much more variability, as would be the case if an identically sized park were randomly dropped on the annuli. Furthermore, small-surface annuli are supposed to include the true city center but may miss this objective if true city centers are slightly away from the city hall locations used as the city centers. The observed share of green urban areas close to CBDs can then be more volatile and biased. To check for this issue, we aggregate the three most central 100-meter annuli into one larger central ring and the next two annuli in another ring, while we leave the other annuli intact. The results are presented in Column (2) and do not qualitatively differ from the baseline model. In Column (3), we also include the robustness check for urban green profiles, where all large parks are excluded. Since our model focuses on the effect of local urban green areas within a neighborhood, it might not apply to very large parks that have global effects on city inhabitants. Therefore, we further exclude all the parks exceeding one square kilometer,[20] and the regression results do not significantly different from our baseline regression.

Third, in Table 2.3, city borders have been defined by the locations where the share of residential space reaches a threshold of 20%. This definition resulted from our trade-off between theory and data. In theory, a city border is well defined and has a zero residential density, although it is not well defined in the data where the residential density never reaches zero. The use of a threshold that is too high certainly undershoots the actual distances between the CBD and city borders. Therefore, we extend the definition of a city border with a lower threshold of 15%. This extension is shown in Column (4), where the number of observations rises to 16,851 annuli. The results remain qual-

[20] We also checked other criteria for this size threshold for large parks in our model, such as the 99^{th} percentile level; however, the results were rather similar.

itatively the same. However, for this threshold value, the random variations in the share of residential spaces lead to fluctuations and downward biases in city border values. Column (5) displays the results obtained when the share of residential spaces is smoothed (with the same kernel smoothing as in Table 2.2). The number of observed annuli rises to 22,549. The results remain qualitatively the same, except for the coefficient of the linear distance term, which becomes not significant. This result reflects a decrease in the slope of the measured share of green urban areas and is explained by the fact that Column (5) includes new observations with no green urban areas. These new observations with zeros at far distance from the CBD reduce the slope of the share of green urban areas.

Fourth, the dataset may not match the monocentric city hypothesis of the theoretical model because it includes polycentric cities and contiguous cities. The next columns of Table 2.4 reduce this mismatch. Column (6) reports results for the set of 27 small and medium-sized EU cities with populations less than 1 million individuals (keeping fixed country effects). This approach eliminates the largest cities that are prone to host multiple subcenters. Column (7) focuses on monocentric cities using the OECD study on metropolitan urban polycentricity, keeps the 21 countries that are common to our GMES database and excludes reported polycentric cities (see Appendix B for details). Similarly, in our theory, cities are spatially separated, which is not always the case in the data. Columns (8) and (9) report results with the subsamples of cities that are at least 50 km and 100 km apart, respectively, keeping fixed country effects. The first distance usually corresponds to the extent of urban labor market areas. The second distance makes sure that daily commuting between cities is unattractive. As can be seen in Table 2.4, Columns (6) to (9) do not qualitatively deviate from the baseline results.

One may question to which land functionality green urban areas should be compared. Our theoretical model discusses the split between green urban areas and residential land. In the above empirical model, we have extended the areas for residential functionality to all human dwellers' functionalities. Accordingly, our above baseline empirical analysis used a measure of the share of green urban areas consisting of the ratio of the area of green urban areas as the numerator and the area of artificial surfaces, agriculture, seminatural areas, wetlands and forest as the denominator. Therefore, the denominator includes many potential land functions. The last columns of Table 2.4 present the results on alternative measures for this denominator, which increasingly narrow the comparison down to residential areas. Column (10) displays the share of green urban areas when we keep only the artificial surfaces in this denominator. This approach eliminates agricultural areas, wetlands and forests. Next, we compare green urban areas to the land used exclusively for human activities. Column (11) reports the results with the denominator measuring the land for residences, offices and green urban areas (i.e., urban fabrics, industrial, commercial, public, military and private units and green urban areas). This approach eliminates roads, railways, ports, airports, mines,

Table 2.4: Profile of green urban spaces: robustness

	Dependent variable: Share of Urban Green Space												
	OLS	Wider central annulli	Exclude big parks	Border 15%		Monocentric cities		Contiguity		Other measures			Medieval cities
				Raw	Smooth'd	<1Mo	OECD	50km	100km				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10	(11)	(12)	(13)
Distance	0.102***	0.068***	0.092***	0.065***	0.005	0.078***	0.102***	0.080***	0.055	0.108***	0.087***	0.126***	0.112***
	(0.006)	(0.019)	(0.015)	(0.015)	(0.014)	(0.017)	(0.029)	(0.021)	(0.043)	(0.020)	(0.027)	(0.030)	(0.016)
Distance_square	−0.127***	−0.100***	−0.120***	−0.116***	−0.078***	−0.105***	−0.128***	−0.112***	−0.089**	−0.105***	−0.094***	−0.111***	-0.129***
	(0.005)	(0.016)	(0.013)	(0.014)	(0.013)	(0.014)	(0.025)	(0.018)	(0.036)	(0.018)	(0.024)	(0.027)	(0.014)
Med.*Dist_sqr													0.002
													(0.008)
City FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sample of cities	All	All	All	All	All	Small+med.	EU21	50km	100km	All	All	All	All
Observations	13,091	12,179	13,091	16,819	22,507	10,088	5,094	8,634	2,826	13,091	13,091	13,091	13,045
Adj. R^2	0.369	0.435	0.349	0.399	0.442	0.363	0.300	0.387	0.301	0.340	0.306	0.326	0.368
df	12,785	11,873	12,785	16,513	22,201	9,811	5,011	8,425	2,758	12,785	12,785	12,785	12,740
F Stat.	26.13***	31.73***	22.42***	37.65***	59.54***	21.79***	27.58***	27.23***	19.18***	23.10***	19.89***	21.75***	26.03***

Note: Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses. The row "df" reports the degree of freedom. The table shows results from the regression of share of green urban space within each annulus on the relative distance and its square to the city center using control variables and a constant. The relative distance is normalized to one for farthest annulus for comparison purpose. Columns (1) includes simple OLS without corrections; other columns are controlled for heteroskedasticity. Column (2) merges the central annuli in wider areas. Column (3) excludes all big parks with the size above 1 kilometre square. Columns (4) and (5) define city borders with a lower threshold (15%) on the un-smoothed and smoothed profile of residential density. Columns (6) and (7) isolate monocentric cities respectively by using cities with less than one million inhabitants and by excluding the polycentric cities reported by OECD. Columns (8) and (9) exclude contiguous cities with inter-city distances lower than respectively 50 and 100 km. Column (10) measures the share of green urban space with respect to only artificial surfaces. Column (11) further excludes transport infrastructure, etc. Column (12) measures it with respect to only urban fabrics.# Colunm (13) includes a row with the effect of cities with medieval history on the square of distance. A city has a medieval history if it existed before year 1500 (Bairoch *et al.* 1988).#

construction sites, land with no use and sports and leisure facilities. Finally, Column (12) is even more restrictive by concentrating on only urban fabrics and green urban areas. The regression coefficients remain stable despite important variations in the definition of the share of green urban areas.

We also check the impact of the medieval history of each city on the share of green urban areas. One may argue that medieval cities included walls, moats and hunting places that have been preferably converted to parks by planners when the cities expanded in the 19th and 20th centuries. Accordingly green urban areas would be placed in annuli between the center and periphery of existing cities, which would explain the non-monotone relationship between the share of green urban space and distance to city center. To test this hypothesis, Column (13) shows the same regression as in the baseline regression (Table 2.3 Column (3)) with an additional cross-effect of a medieval history dummy on the square of the distance to center. Cities are defined as medieval if they are built before 1500. As it can be seen, this cross-effect is not statistically significant, indicating no difference between the groups of cities with medieval and non-medieval history.

Finally, we also run the regression (2.18) for each city in our sample and count the number of cities for which the concavity property holds. We observe only 4.59% (10.49%) of cities where we cannot reject convexity ($\eta_2 > 0$) for p-value < 0.01 (p-value < 0.1).

2.4 Conclusion

In this paper, we discuss the patterns of urban green areas in cities from theoretical and empirical perspectives. Urban green areas mainly include green areas maintained for recreational purposes by non-private human institutions (typically, municipalities). Green urban areas provide residents with amenities that have the property of local public goods and high land intensity. We find that the optimal provision of urban green areas is a nonmonotonic concave function of the distance to CBDs. It results from the balance between the higher opportunity cost of land near CBDs and the lower population density at city edges.

This property is confirmed by our study of the urban land use in the 305 most populous urban EU areas. We use detailed maps of urban land use from the European Environment Agency (GMES) to study the spatial configurations of urban green areas. Our study shows a concave and hump-shaped profile of urban green areas with respect to distance to the CBD. The result is robust to many variations in the land use specifications, city structure specifications, and city and country characteristics.

Appendix A: Proofs

Samuelson Rule We can transform expression (2.5) in the Samuelson's optimality condition of public goods. Plugging the optimal residential space condition $\widetilde{z} - s\widetilde{z}_s = y$ in the land value $V = (y - \widetilde{z})/s$ gives $V = -\widetilde{z}_s$. At the same time, applying the envelop theorem on V yields $V_a \equiv -\widetilde{z}_a/\widehat{s} > 0$. Finally, using $\widetilde{z}_s = -U_s/U_z$ and $\widetilde{z}_a = -U_a/U_z$, expression (2.5) can be written as (2.6).

Hyperbolic preferences We here assume that individuals are endowed with the utility function

$$U(z, s, a) = z - \theta/(2s) + a$$

where preferences for residential space are represented by an hyperbolic function parametrized by θ.[21] Accordingly, we get the consumptions $\widetilde{z} = u - a + \theta/(2s)$, $\widehat{s} = \theta/(y + a - u)$, $\widehat{z} = (y - a + u)/2$ and land value $V = (y + a - u)^2/(2\theta)$. We focus on the case of positive consumption so that $y + a - u > 0$. Then, condition for optimal green urban area service (2.4) simplifies to the identity

$$\frac{\alpha}{\theta}(y + \alpha x - u)(1 - \beta x) - \frac{\beta}{2\theta}(y + \alpha x - u)^2 - \gamma = 0, \tag{2.19}$$

where the LHS expresses the net benefit of green urban service and is a concave quadratic function of x. If the inequality

$$y - u + \frac{\alpha}{\beta} > \sqrt{\frac{6\theta\gamma}{\beta}} \tag{2.20}$$

does not hold, the LHS is negative so that there are no benefits from green urban areas for any net revenue y. If it holds, the identity accepts two solutions, but only the highest one, $x = x_+^*$, determines an interior maximum. Hence, the optimal interior solution for land use of green urban area is given by

$$g_+^* \equiv \beta x_+^* = 1 - \frac{2\beta}{3\alpha}\left(y - u + \frac{\alpha}{\beta}\right) + \frac{\beta}{3\alpha}\sqrt{\left(y - u + \frac{\alpha}{\beta}\right)^2 - \frac{6\theta\gamma}{\beta}},$$

Using the previous relationships, we can also infer the optimal interior solution for population density by

$$n_+^* = \frac{\beta}{9\theta\alpha}\left(y - u + \frac{\alpha}{\beta} + \sqrt{\left(y - u + \frac{\alpha}{\beta}\right)^2 - 6\theta\frac{\gamma}{\beta}}\right)\sqrt{\left(y - u + \frac{\alpha}{\beta}\right)^2 - \frac{6\theta\gamma}{\beta}}.$$

If (2.20) holds at the same time as $g_+^* > 0$ and $n_+^* > 0$, then the allocation $(g^*, n^*) = (g_+^*, n_+^*)$ is a socially optimal one. Otherwise, the allocation $(g^*, n^*) = (0, \max\{0, (y - u)/\theta\})$ is socially optimal.

[21] This yields a demand for residential space that has a price elasticity between one and zero (see Mossay and Picard 2011).

The objective of this paper is to understand the profile of the share of green urban area. We successively get the following comparative statics result:

$$\frac{\mathrm{d}g_+^*}{\mathrm{d}y} = \frac{\beta}{3\alpha}\left(\frac{1}{\sqrt{1-\frac{6\theta\gamma}{\beta\left(y-u+\frac{\alpha}{\beta}\right)^2}}}-2\right),$$
$$\frac{\mathrm{d}^2g_+^*}{\mathrm{d}^2y} = -\frac{2\gamma\theta}{\alpha\sqrt[3/2]{\left(y-u+\frac{\alpha}{\beta}\right)^2-\frac{6\theta\gamma}{\beta}}} < 0.$$

The share of green urban area g_+^* is therefore a concave function of $y-u$, and therefore, since $y = w - t$ and $\mathrm{d}t/\mathrm{d}r > 0$, it is a concave function of wage w and distance to CBD, r. Stronger land regulation reflected by higher u also have non monotone effect on green urban area. By contrast, it can readily be observed that the population density n_+^* rises with y so that it increases with wage w but falls with distance to the CBD r. The planner compensates longer commuting distances with larger residential space. The reason for the concavity in green urban area density then lies in the balance between opportunity cost of land and density of residents as shown in the Samuelson condition (2.6). In the vicinity of the CBD, the social value of land is high because of shorter commutes. This entices the planner to increase population density at the expense of access to green urban areas. By contrast, the social value of land is low at the vicinity of the city border because of longer commutes and entices the planner to compensate individuals with larger residential plots and lower population density. In those low density locations, green urban areas yield smaller social benefit and the planner reduces their surfaces. So, it is the high land social value that refrains the planner to organize large green urban areas about the CBD and it is the low population density that refrains her to maintain large green urban areas at the city fringe.

It can be further checked that the share of green urban area reaches the maximum value $1-\sqrt{2\theta\gamma\beta/\left(9\alpha^2\right)}$ for $y-u+\alpha/\beta = \sqrt{8\theta\gamma/\beta}$ (where $\mathrm{d}g_+^*/\mathrm{d}y = 0$). Hence, the share of green urban area reaches a peak that rises with higher amenity α, lower green urban area land use β, lower maintenance cost γ and lower preferences for residential space θ. All these properties are intuitive. The maximum value is however independent of wage w and regulation policies through changes in u. To know the peak location, we can use $y = w - t$ and write $t(r) = w - u + \alpha/\beta - \sqrt{8\theta\gamma/\beta}$. Since $\mathrm{d}t/\mathrm{d}r > 0$, we infer that the peak location shifts away from the CBD with higher amenity α, lower maintenance cost γ and weaker preference for residential space θ. It also shift away from CBD with higher wages w and weaker land regulation (smaller u).

To fix ideas, consider cities in the same country where labour mobility is high, urban land regulations are rather similar and residents' preferences are the same. Hence, cities differ only with respect to per capita incomes w, which are driven by first nature and second nature advantages, such as the presence of specific factors, harbour, steel in-

dustry, financial center, etcetera. Then, the location of peak in the share in green urban area increase with earnings and therefore city population sizes N. This reflects the fact that land rent becomes more expensive about the CBD and entices the planner to shift them away from it. Under these preferences, the amplitude of this peak is unrelated to earnings and city sizes. We will come back to this relationship in the empirical analysis.

Proposition 5 *Under hyperbolic preferences, population density falls with distance from CBD while the share of green urban area is a concave function of distance from CBD. The share of green urban area peaks at a location that moves away from CBD as cities are richer and less regulated.*

One can also check the following comparative statics on population density:

$$\frac{\mathrm{d}n_+^*}{\mathrm{d}\gamma}, \frac{\mathrm{d}n_+^*}{\mathrm{d}\theta}, \frac{\mathrm{d}n_+^*}{\mathrm{d}u} < 0 < \frac{\mathrm{d}n_+^*}{\mathrm{d}\alpha}, \frac{\mathrm{d}n_+^*}{\mathrm{d}y} \tag{2.21}$$

(see Appendix A). As standard in the urban economic literature, the population density falls with stronger preferences for residential space θ while, because $y = w - t(r)$, it also rises with larger income w but falls with distance from CBD. Population density also rises with larger green space amenity parameter α and lower maintenance cost γ. This is because green urban area services bring more utility to residents, who may be allocated to smaller residential plots. Also, since lower maintenance costs entice the planner to enlarge green urban areas, residents obtain higher amenity and can be offered smaller residences.

The optimal city border b^* is determined by the first order condition w.r.t. b (2.8) evaluated at $V(y, a, u) = (y + a - u)^2 / (2\theta)$ and $r = b^*$. It is instructive to first study the case where there is no green urban area at the city border so that $x^*(b) = 0$ and $n^*(b) = [y(b) - u] / \theta$. The above border condition simplifies to $y(b) - u = \sqrt{2\theta R_A}$, which determines the city border as the solution of $t(b^*) = w - u - \sqrt{2\theta R_A}$. Therefore, the city spreads with higher earnings and shrinks with higher agricultural land rents, preference for residential space and also with stronger land regulation (through lower u). The city border is however independent of the preferences and costs for green urban area. In the presence of green urban area at city border b ($g_+^*(b) > 0$), condition (2.8) does not accept a closed form solution and depends on preferences for green urban area and maintenance cost. We can totally differentiate (2.8) and get the following comparative statics (see Appendix A):

$$\frac{\mathrm{d}b^*}{\mathrm{d}\beta}, \frac{\mathrm{d}b^*}{\mathrm{d}\gamma}, \frac{\mathrm{d}b^*}{\mathrm{d}\theta}, \frac{\mathrm{d}b^*}{\mathrm{d}u}, \frac{\mathrm{d}b^*}{\mathrm{d}R_A} < 0 < \frac{\mathrm{d}b^*}{\mathrm{d}\alpha}, \frac{\mathrm{d}b^*}{\mathrm{d}w}. \tag{2.22}$$

This confirms the comparative statics (2.11) obtained under more general preferences. In addition, it is here shown that the optimal city shrinks with stronger preferences

for residential space θ. This is because the land value falls and lies above R_A within a smaller area when θ gets larger.

The optimal city population is given by $N^* = \int_0^{b^*} n^* \, \ell \mathrm{d}r$, which rises with larger city border b^* and higher population density n^*. Using (2.21) and (2.22), it readily comes that

$$\frac{\mathrm{d}N^*}{\mathrm{d}\gamma}, \frac{\mathrm{d}N^*}{\mathrm{d}\theta}, \frac{\mathrm{d}N^*}{\mathrm{d}u}, \frac{\mathrm{d}b^*}{\mathrm{d}R_A} < 0 < \frac{\mathrm{d}N^*}{\mathrm{d}\alpha}, \frac{\mathrm{d}N^*}{\mathrm{d}w}$$

In particular, city population rises with smaller maintenance cost and higher amenity parameter for green urban area.

Finally, we discuss the difference between green urban areas and local public goods in the case of hyperbolic preferences. Urban green space are local public goods as they provide the same amenity to the residents in its close vicinity and incur maintenance costs that are independent of the number of users. The difference lies in the land intensity of green urban areas. The urban economic literature presents local public good as spaceless amenities that do not use space (see Fujita Thisse, 2004, for synthesis). In the case of spaceless local public good ($\beta = g = 0$), land supply is fully supplied to residents so that the land market clears: $sn = 1$. Condition (2.19) simplifies to $\alpha (y + \alpha x - u) - \theta\gamma = 0$, in which the LHS measures the marginal surplus of local public good and rises in x. This implies that the surplus is convex in x. The planner must therefore choose between the minimum level $x = 0$ and the maximum level $x = x^o \equiv (y + u) / \alpha$ that is reached when individual is asked to consume no composite good, i.e. $\widehat{z} = (y - \alpha x + u) / 2 = 0$. Thus, because the marginal surplus increases with net income y, the surplus is a convex function of y and there exits a net income $\overline{y}$ above which the surplus is larger at $x = x^o$ than at $x = 0$. Hence, the planner sets the spaceless local public good to $x^* = x^o$ if $y \geq \overline{y}$ and to $x^* = 0$ otherwise. Population density is respectively given by $n^o = (y + \alpha x^o - u) / \theta = 2y/\theta$ and $(y - u) / \theta$. Since net income y falls with distance from CBD, the planner provides lower and lower levels of local public goods to locations farther from the center and none beyond the distance $\overline{y}$. Population density falls with distance from CBD. It falls half as fast beneath $\overline{y}$.

Proposition 6 *Under hyperbolic preferences, services offered by spaceless local public goods fall with distance from CBD and vanish at some distance from it.*

Proofs Comparative statics on population density n_+^*: Most comparative statics are trivial except the ones on α and β. Yet, one can show $\frac{\mathrm{d}\log n_+^*}{\mathrm{d}\alpha} > 0 \iff 1 + 6\theta\gamma\beta + (1 - \beta^2)\left(y - u + \frac{\alpha}{\beta}\right)^2 + 2\left(y - u + \frac{\alpha}{\beta}\right) > 0$. Comparative statics on β are unfortunately inconclusive.

To compute the comparative statics on the city border b^*, let us denote (**??**) as $F(x(b), \alpha, \beta, \gamma, \theta, R_A) = 0$. Hence, $\mathrm{d}b^*/\mathrm{d}\alpha = -(\partial F/\partial\alpha) / (\mathrm{d}F/\mathrm{d}b)$ has the sign of

$\partial F/\partial \alpha$ and the same argument applies for other parameters. At the optimal city border, we compute $\partial F/\partial \alpha = x^* V_a > 0$, $\partial F/\partial \beta = -x^* V < 0$, $\partial F/\partial \gamma = -x^* < 0$, $\partial F/\partial \theta = x^* \partial V/\partial \theta < 0$, $\partial F/\partial R_A = -1 < 0$, $\partial F/\partial w = V_y > 0$, $\partial F/\partial r = -V_y (1 - \alpha x^*) \, \mathrm{d}t/\mathrm{d}r < 0$ and $\partial F/\partial u = V_u (1 - \alpha x^*) < 0$.

Cobb Douglas preferences Under extreme value theorem, there always exists a global maximum for a continuous function on a compact set; therefore, there always exists solution for city planner as x is always within the domain $[0, \frac{1}{\beta}]$.

We denote the marginal welfare by $F(x) - G(y)$ where $F(x) = e^{Ax}(B - Cx)$ and $G(y) = \kappa\gamma y^{-\frac{1-\varphi}{\phi}} u^{\frac{1}{\phi}} > 0$ are the LHS and RHS of the FOC condition (2.17) while $x \in [0, 1/\beta]$, $A = \alpha\varphi/\phi > 0$, $B = \alpha\varphi - \beta\phi$, and $C = \alpha\beta\varphi > 0$. If $B \leq 0$, then $F(x) - G(y) < 0$ for $x \in [0, 1/\beta]$, so that the optimal service is the corner solution: $x^* = 0$. If $B > 0$ then, $F(-\infty) = 0$, $F(0) > 0$ and $F(\infty) = -\infty$ while $F(x)$ has a unique maximum at $\bar{x}$ such that $F(\bar{x}) \geq F(x)$ and $F(\bar{x}) > 0$. It can be checked that $F(1/\beta) < 0$. As a result $F(x)$ has a single root for $x < 1/\beta$. If $AB < C \iff \bar{x} < 0$, $F(x)$ is a decreasing function and there exists a single root x' for $F(x) - G(y) = 0$ iff $F(0) > G(y)$. The optimal service is then $x^* = x'$. If $AB > C \iff \bar{x} > 0$, $F(x)$ is a bell-shaped function and there exist two roots for $F(x) - G(y) = 0$ iff $F(\bar{x}) > G(y)$. The highest root x'' has $F'(x'') < 0$ and gives the optimal service: $x^* = x''$. To sum up, there exists an interior optimal service $x^* > 0$ if $F(\max(0, \bar{x})) > G(y)$. At the interior optimal service x^*, $F'(x^*) < 0$.

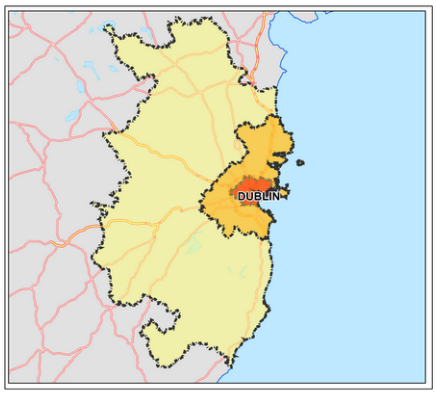
To get the comparative statics w.r.t. y on the interior optimum x^*, we denote $K(x, y) \equiv F(x) - G(y)$. So, $K_x(x^*, y) = F(x^*) < 0$, $K_{xy}(x^*, y) = 0$, $K_y(x^*, y) = -G'(y) > 0$, and $K_{yy}(x^*, y) = -G''(y) < 0$. Then, we get $\mathrm{d}x^*/\mathrm{d}y = -F_y/F_x > 0$ and $\mathrm{d}^2x^*/\mathrm{d}y^2 = -\left(K_{yy}K_x - K_y K_{xy}\right) / \left(K_x\right)^2 = -K_{yy}/K_x < 0$.

Appendix B: Data and robustness check for urban green space profiles

Definitions

We first summarize the definitions of urban area entities. Eurostat uses three level of spatial units that are based on clusters of high density grid cells and urban cores. Following Eurostat a *high density grid cells* are defined as population grid cells of one kilometer square with at least 1,500 inhabitants. A *cluster of high density cells* is a set of high density grid cells that are each surrounded by at least five other high density cells (over the eight surrounding cells). Clusters exclude the high density grid cells that are not connected or isolated. An *urban core* is a cluster of high density cells that totals at least 50,000 inhabitants. The first level of spatial unit is the *City*. It is related to an urban core and defined by the local administrative boundary so that more than 50% of inhabitants live inside the associated urban core.[22] The second level of spatial unit is the *Greater City*, which is created when the urban population resides far beyond the local administrative boundaries. Greater cities like Greater Manchester, Greater Nottingham and Greater Paris have been defined with alternative but close definitions. In most cases, a Greater City contains a single City. The *City, Greater City* (CGC) includes those two levels. A *Functional Urban Area* (FUA) combines the city area and its commuting zone, as defined in the EU-OECD functional urban area definition (OECD, 2013). A FUA includes the "working catchmen area" of a city and is defined as the collection of all surrounding municipalities with at least 15% of their employed residents working in the associated urban core. Figure B1 presents the three different levels of spatial units for Dublin.

Figure B1: Eurostat levels of spatial units (Source: Eurostats)



[22]There are some exception to this rule when the geography is disrupted by a river, a lake, fjords, or steep slopes etcetera, making it hard to recover the urban core. In this case, the City can be added to cover this urban centre.

Hypothesis testing of Table 2.3

Table B1 reports the p-value associated to Table 2.3 columns (1) to (6). Those values are extremely small, corroborating the strong non-monotone relationship.

Table B1: P-values for one-sided t-test for baseline models

	(1)	(2)	(3)	(4)	(5)	(6)
$H_0 : \beta_0 > 0$	1.23×10^{-58}	1.39×10^{-6}	2.01×10^{-160}	2.00×10^{-24}	3.55×10^{-10}	1.28×10^{-24}
$H_0 : \beta_1 > 0$	4.16×10^{-11}	4.99×10^{-11}	7.88×10^{-11}	2.29×10^{-4}	4.31×10^{-4}	4.90×10^{-7}
$H_0 : \beta_2 < 0$	5.14×10^{-20}	5.60×10^{-20}	1.31×10^{-19}	5.11×10^{-10}	1.55×10^{-5}	2.68×10^{-9}

Note: The covariance matrix for residuals used for hypothesis testing is clustered at city level and heteroskedastic-robust.

Polycentric cities

In the GMES EU27 database, some cities develop in a polycentric way and host several urban cores that are physically separated but economically connected. Column (6) in Table 3 excludes EU polycentric cities using an OECD study on polycentric cities. In the latter, OECD assigns a polycentric status to a FUA using the percentage of residential population commuting from one urban core to another. A polycentric city includes two or more urban cores, which are connected and attract at least 15% of each other's population as workforce (OECD, 2013).

Furthermore, the column (6) of Table 2.4 focuses on the 21 countries that are common to the 27 countries in GMES EU27 database and the 23 countries in the OECD study. This excludes six EU27 countries (Bungaria, Lithuania, Latvia, Malta, Romania and Cyrus) and 2 OECD countries (Switzerland and Norway). Table B2 reports the population statistics of monocentric and polycentric cities for those 21 countries. There are 18 FUAs with two urban centres, and 6 FUAs with more than two urban cores (Barcelona, Paris, Lyon, Amsterdam, Stockholm and London). We will exclude all those FUAs in our sample. Finally, since the OECD study concentrates on cities with more 500,000 inhabitants, we keep only the cities with same sizes in the robustness of column (6) of Table 2.4. This excludes the GMES 27 'small' sized cities that we used in other robustness analyses.

Table B3 provides further robustness analysis on the city border determination. It reports regressions coefficient when the city border is defined by various shares of residential density: 15%, 20%, 25% and 30%. Coefficient signs are unaffected by those definitions in both the quadratic and cubic models.

Table B4 presents regressions with alternative weighting schemes for observations. Larger cities are less numerous but include larger populations and larger numbers of annuli, which could influence the results. Table B4 reports the regression coefficients corresponding to Table 2.3 with weights proportional to the city population (columns

Table B2: Descriptive statistics for EU metropolitan forms in OECD database

	average	sd	min	max	obs.
Monocentric Cities					
Population of metro area (thousands)	1235	890	445	4399	87
Population of city area (thousands)	772	633	90	3467	87
Duocentric Cities					
Population of metro area (thousands)	1713	1609	561	7079	18
Population of city area (thousands)	1187	1274	314	5264	18
Polycentric Cities (with more than three centers)					
Population of metro area (thousands)	5786	5030	1096	12401	6
Population of city area (thousands)	4487	4105	1331	9942	6

Note: This table includes 21 countries: All EU27 except Bungaria, Lithuania, Latvia, Malta, Romania and Cyrus. Metro population is computed from Census 2010 according to OECD metropolitan boundary maps in 2001. The metropolitan population is very similar to the FUA population provided in Urban Audit database of Eurostat. For more details, see OECD Metropolitan Explorer database, version June 2016 (OECD, 2016).

Table B3: Profile of green urban areas (EU27): Robustness with different levels of cut-off

	Dependent variable: Share of Green Urban Area							
	Quadratic				Cubic			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Border cut-off	15%	20%	25%	30%	15%	20%	25%	30%
Dist.	0.065***	0.102***	0.119***	0.111***	0.303***	0.294***	0.264***	0.260***
	(0.015)	(0.016)	(0.018)	(0.020)	(0.037)	(0.039)	(0.041)	(0.045)
Dist. square	−0.116***	−0.127***	−0.123***	−0.098***	− 0.697***	−0.591***	−0.470***	−0.449***
	(0.014)	(0.014)	(0.016)	(0.018)	(0.079)	(0.080)	(0.085)	(0.093)
Dist. cubic					0.380***	0.302***	0.225***	0.226***
					(0.048)	(0.048)	(0.052)	(0.056)
Constant	0.099***	0.100***	0.098***	0.093***	0.078***	0.082***	0.085***	0.080***
	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
City FE	✓	✓	✓	✓	✓	✓	✓	✓
City Sample	All	All	All	All	All	All	All	All
Obs.	16,819	13,091	10,188	8,156	16,819	13,091	10,188	8,156
Adj. R^2	0.399	0.369	0.344	0.327	0.416	0.379	0.349	0.331
df	16,513	12,785	9,882	7,850	16,512	12,784	9,881	7,849
F Statistic	37.65***	26.13***	18.50***	13.97***	40.09***	27.13***	18.84***	14.21***

Note: This table shows results from the regression of share of green urban area within each annulus on the relative distance and its square to the city center using control variables and a constant. The relative distance is normalized to one for farthest annulus for comparison. All columns include controls for country and city size. The percentage of cut-off points is defined as the annuli with percentage of urban fabric areas over the total area of the city within the annuli smaller than the percentage of the cut-off points indicated. Standard errors are clustered at city level and reported in parentheses. The row "df" reports the degree of freedom. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01.

Table B4: Profile of green urban areas (EU27): Robustness with weighted OLS

	Dependent variable: Share of Urban Green Space					
	Quadratic			Cubic		
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	0.166***	0.074***	0.070***	0.344***	0.267***	0.278***
	(0.011)	(0.016)	(0.019)	(0.029)	(0.040)	(0.047)
Distance_square	−0.187***	−0.065***	−0.057***	−0.625***	−0.523***	−0.551***
	(0.009)	(0.016)	(0.019)	(0.061)	(0.094)	(0.111)
Distance_cubic				0.288***	0.295***	0.318***
				(0.038)	(0.063)	(0.074)
Constant	0.084***	0.079***	0.079***	0.068***	0.062***	0.060***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
City Size FE	✓	✓	✓	✓	✓	✓
City Sample	All	All	All	All	All	All
Weighting	City Pop.	Pop. Dens. 1	Pop. Dens. 2	City Pop.	Pop. Dens. 1	Pop. Dens. 2
Observations	13,117	13,117	13,117	13,117	13,117	13,117
Adj. R^2	NA	0.111	0.115	NA	0.118	0.123
df	13086	13086	13086	13085	13085	13085
F Stat.	69.30***	27.99***	27.47***	68.17***	28.77***	28.20***

Note: We include different weighting strategy. Weights are proportional to the city population (columns (1) and (4)). Columns (2) and (5) include population density measured by dividing total city population with area of whole city. Columns (3) and (6) measure city population by dividing city population to areas of artificial urban fabric, which mostly used for residential purpose. Standard errors are clustered at city level and reported in parentheses. The row "df" reports the degree of freedom. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01.

(1) and (4)) and the two different measures of city population density (columns (2)-(3) and (5)-(6)). Coefficients are invariant to weighting specifications, which suggests low risk of misspecification.

Chapter 3

Welfare Evaluation of Green Urban Areas

3.1 Introduction

Urban green area is one of the keys for building resilient and sustainable urban systems. These areas provide many valuable local services to urban dwellers such as mitigating air pollution and urban heat islands (Bowler et al., 2010; Manes et al., 2012), supporting ecological learning (Barthel et al., 2010), and enhancing human health (White et al., 2013). They are also an essential block of urban social life as they help to bond urban social ties, providing a place for recreational opportunities and other social benefits (Colding and Barthel, 2013; Saldivar-Tanaka and Krasny, 2004; White et al., 2013). It is no doubt that high-quality urban green spaces can bring many relevant services to urban areas, and their locations constitute a crucial component of urban planning in the quest for sustainable cities and communities (Tyväinen et al., 2005). This fact poses a critical question for city planners: how do the residents value them? Such information is vital to design more desirable residential communities and to thrive by limiting development.

There are three main approaches to evaluating the value of green and open areas. The first approach is the contingent evaluation (CV) method, which requires households to directly specify their willingness to pay (WTP) for the preservation or improvement of specific urban green areas. However, this method depends on hypothetical scenarios, and it does not reflect the actual behaviors of all households. The second approach is the hedonic pricing (HP) method, which reveals the marginal WTP of households through the premium that they give to housing prices through market transactions. The HP method adequately reflects the actual evaluation of households; however, it does not measure the nonmarginal changes for green areas. It also suffers from the 'Tiebout bias' (see Kuminoff et al., 2013), as households can move to other districts in response to a policy change. The third approach is the equilibrium sorting

(ES) method, which takes into account the sorting processes of households in the market by simulating a new equilibrium. The gains from the ES method come at a cost, as it is necessary to specify the preference structure and the choice sets available for consumers. Additionally, the housing supply is treated as fixed to allow for simple calibration (Freeman et al., 2014; Kuminoff et al., 2013). A detailed survey of the literature is reported in Section 2.

In this paper, we utilize a different approach. We first introduce the urban green areas inside an urban land market model and formalize the channel in which the green areas enter households' decisions and land prices. We employ two different households' preferences: Cobb-Douglas and hyperbolic linear utility. The marginal benefit of urban green areas varies across locations within cities concerning their existing covers in the residential areas, the population density and the opportunity cost of land at each location. The market equilibrium results in the optimal residential land use at each location across urban areas with a specified level of urban green areas. We then calibrate the equilibrium of residential lands using the data on detailed residential land covers, green urban areas and population density. To control for the potential endogeneity of urban green provision from urban authorities, we employ the green areas in the past as an instrument for the current green areas. We back up the parameters and run counterfactual simulations under the scenario that urban planners reduce green areas by half. By using this approach, we can account for the new equilibrium in the housing market and households' choices over housing consumption. We consider both the case of open cities with free migration and closed cities with no migration.

We consider only small and scattered green urban areas in this paper. It is because 95% of urban parks have a surface lower than 0.1 square kilometer, and the average distance between the two parks is lower than 400 meters in our sample of 305 of the most populated cities in Europe. We estimate residential land use and take the estimated parameters to study the value of green urban areas implied by our theoretical urban land market model. For the parameter estimation, we use the geographical database of the European Environment Agency's Urban Atlas data, which describes the land use of 305 EU cities with more than 100,000 inhabitants. These data report land use and cover across Europe using harmonized Earth observations (EOs), which are combined with Eurostat Urban Audit statistical data. The data represent a unique source of reliable and comparable European urban planning data. To our best knowledge, this is the first paper that uses the geographical land covers in both a theoretical model and empirical estimation of European urban land uses.

Through counterfactual exercises where 50% of urban green areas are removed, we show that open cities lose more than 6% of their population if those areas are left unused. The total loss for landlords is approximately €150 million per city under the Cobb-Douglas preference and more than €300 million in the case of the hyperbolic preference if green areas are not converted into residential land. Converting those ur-

ban green areas into residential land, however, would increase residential surfaces and increase the total housing market value by approximately €50 million in the Cobb-Douglas preference and a loss of less than €10 million in the hyperbolic preference. In closed cities, residents lose utility, which can be restored by a subsidy of nearly 6-9% of their net annual income (it is equivalent to a negative willingness to pay from households). Those results are robust for both types of preferences. This exercise, based on a dramatic decline in green urban areas, suggests that those areas provide highly valuable amenities to residents. However, when we allow for the land conversion of the erased green areas into new residential covers, the net costs of the green areas are up to 4.2% of annual household income, suggesting that the land opportunity costs of urban green areas are substantial. Our results are in line with those of Cheshire and Sheppard (2002), who also find that the net cost of restrictive land use planning can reach 3.9% of the household's income.

Our contributions are twofold. First, we utilize new datasets, including the geographical land covers, and population grids to estimate the parameters of our theoretical models. We also account for the potential endogeneity in green provision by using the green areas from the past to instrument for the current level of green provision. With these data, we account for the actual residential choices of households in reality instead of hypothetical questions as in the CV method. Second, the use of an urban model together with two different land conversion policies allows us to account for the change in residential land supply and enables us to compute the general equilibrium willingness to pay of households for a significant change in green provision. This approach, compared with the HP model, addresses the nonmarginal changes in green areas, and it also relaxes the restrictions of fixed housing supply in the sorting equilibrium method. Our tractable and clean models and counterfactual exercises also come with the cost of imposing a specific preference for households. We address this issue partially by considering two types of utility functions: Cobb-Douglas and hyperbolic. Second, we only consider one class of household for each city. The extension for multiple types of households in this model and the calibration could be a helpful direction for future research.

3.2 Literature review

There is a large and growing literature quantifying the value of open and green spaces in urban systems. The first branch of literature is to evaluate green areas through contingent valuation (CV). This approach designs surveys with hypothetical scenarios for environmental goods and asks survey respondents to specify their willingness to pay (WTP) for the preservation or improvement of specific urban green areas. Researchers aggregate those values and estimate the monetary value of this environmental asset. For Europe, Elsasser (1996) and Tyrväinen (2001) use the CV method for the city of

Hamburg in Germany and Joensuu and Salo in Finland. They find that the mean WTP was approximately €42 per person per annum in Germany[1] and €60-144 in Finland[2]. Schindler et al. (2018) find that the reported willingness to pay for green area access increases with household's average income. Even though the CV method provides valuable information on how much value each resident places on the usage of green areas, it is based on hypothetical scenarios and does not reflect their actual valuations and behaviors. The WTP measured in other studies revealed the preference approach, in that hedonic pricing or equilibrium sorting reports a much higher value. The WTP measured in our paper also has a higher value of approximately 6-9% of annual household income. Nevertheless, the WTP estimated under the Cobb-Douglas preference in our model also increases with income quantiles, which is similar to the pattern reported in Schindler et al. (2018).

The second approach is hedonic pricing (HP), which evaluates the value of green amenity through its capitalization in the housing market. The premise underlying this method is that if households enjoy green amenity in a location, they will pay a premium for it. Most of the studies use housing transaction data, and they find that the proximity to parks and open spaces induces an increase in home prices (Bolitzer and Netusil 2000; Geogrehan et al., 1997; Lutzenhiser and Netusil, 2001; Smith et al., 2002). Anderson and West (2006) further reveal that the value of proximity to open spaces is higher in dense and high-income areas. Luttik (2000) find that a pleasant view leads to an increase of 6-12% in house price in the Dutch land market, and Tyrväinen and Miettinen (2000) estimate a 5% increase in home price in Finland. Tyrväinen and Miettinen (2000) estimate that a one-hectare green area attributes to an increase of €11-15 million in the total value of surrounding houses in dense central districts and €4-5 million in single-home and less dense areas. In a competitive land market, this information helps us to infer the marginal willingness to pay for green amenity. Even though this information is undeniably valuable, it does not capture nonmarginal changes in urban green areas. A unique departure from this framework is Cheshire and Sheppard (2002). They use hedonic pricing methods to estimate the implicit price for accessible open spaces and then embed them into a monocentric model and calibrate the data for the city of Reading, UK. The use of an urban model allows them to incorporate the general equilibrium effects in their analysis. They estimate that the gross benefit of public open spaces is £2,424 per household per annum (in 1984 price). However, when accounting for the cost of restrictive residential lands[3], the net loss from planning could be as high as 3.9% of annual household income. In our paper, when urban planners remove half of the urban green areas, we find a WTP of €970 - 1300 per year per habitant to avoid such a

[1] in terms of the 1995 price, as cited by Tyrväinen et al. (2005), pg. 102.

[2] in terms of the 2000 price, also per person and per annum

[3] They compare between the current level of 37.9% of residential land to a less restrictive scenario of 42.5% of residential land

policy. However, when we account for the reconversion of those lands into residential uses, the net cost of these green areas is estimated at 3.2%-4.2% of annual household income, which is in line with the outcomes from Cheshire and Sheppard (2002).

The third approach is the equilibrium sorting (ES) approach, which seeks to evaluate the large changes in public policies. In contrast to the HP estimation, which assumes that a substantial change in amenity level does not affect household locational choice, the ES model seeks to characterize the household's sorting process in response to a policy change. It is then straightforward to measure both marginal and nonmarginal changes in urban amenity policies by simulating a new sorting equilibrium.[4] Wu and Cho (2003) apply this method to the environmental amenities in Portland, Oregon, and find that a 1.2% increase in the share of parks and open spaces is equivalent to a $100 increase in educational expenditure in household choices for communities. Walsh (2007) uses this sorting method to evaluate the open space protection policy in Wake County, North Carolina, and finds that an open space protection policy in denser areas can lead to a 11.4% increase in house price and approximately 8.14% in less dense areas. Our paper also reports a larger effect in denser areas near the city core and a smaller effect in land price near urban fringes. Klaiber and Paneuf (2010) use a similar approach to estimate household preference for open space, and they find that the general and partial equilibrium becomes diverse with the magnitude of the change in policy. However, this ES method is not without constraints. One constraint is that it is limited in choice sets for housing types due to the difficulties in increasing dimensions for calibration. Hence, most of the papers use the sorting process through the neighborhood instead of house types. They also have to specify the distribution of preference heterogeneities (Freeman et al., 2014). Second, the housing supply is treated as fixed to allow for simple calibration (Freeman et al., 2014; Kuminoff et al., 2013). In our paper, we allow for the land conversion policies, as discussed in the above section, which relaxes the housing supply constraints in a new equilibrium.

3.3 Theoretical model

The model considers a circular monocentric city. The economy consists of a mass N of individuals living within city boundary and a central business district (CBD) where all households commute to work. We denote $b \in \mathbb{R}_0^+$ as the distance between the CBD and the city border. The population density is defined as the number of individuals in a unit of area at distance r from the CBD and is denoted by the function $n : [0, b] \to \mathbb{R}^+$, which varies across the city.

In this model, we focus on small green urban areas that spread across the city and are closely accessible to the local community around its location.[5] Green urban areas

[4]For a comprehensive review, see Kuminoff et al. (2013).

[5]Because of proximity of small parks, residents' travel costs to parks can be neglected here as a first

provide quick and frequent access to greenery, quiet, children's parks, socialization areas, etc. We consider the few blocks in the vicinity of a green urban area as our unit of area or patch and model the urban area in a continuous fashion. In a unit of area at distance r from the CBD, green urban areas offers a service $x : [0, b] \to [0, \overline{x}]$, $\overline{x} \in \mathbb{R}_0^+$, to the local community living in the vicinity. This service brings a level of amenity $a = \alpha x(r)$, although it necessitates the use of a fraction of land $\beta x(r)$ and maintenance costs $\gamma x(r)$. The parameters $\alpha, \beta, \gamma \in \mathbb{R}^+$ distinguish the amenity, land use and maintenance factors that affect green urban areas. Hence, the fraction of land used for residential purposes is given by $1 - \beta x$, and the maximum service level $\overline{x}$ is bounded by $1/\beta$. We assume absentee landlords, and the outside opportunity value of land is given by the agricultural land rent $R_A \in \mathbb{R}^+$. For simplicity, we consider that rural areas beyond the city border consist of private properties that do not provide green urban area service for city dwellers (private crop fields, fenced areas, etc.). We denote the land supply at distance r from the CBD by $\ell : [0, b] \to \mathbb{R}^+$ (e.g., $\ell = 2\pi r$ if the city lies in a plain disk). In summary, land at distance r from the CBD includes a surface $\beta x(r)\ell(r)$ of maintained green urban area and a residential area $[1 - \beta x(r)]\,\ell(r)$, and it hosts $n(r)\ell(r)$ residents who all benefit from the green urban area amenity $\alpha x(r)$.

Individuals consume a quantity z of nonhousing composite goods and a quantity s of residential space, while they benefit from the amenity a of a green urban area. They are endowed with the utility function $U(z, s, a)$, which is assumed to be concave and increasing for each variable. We assume that demands for nonhousing composite goods, residential space and amenity are gross substitutes such that U has negative second derivatives and positive cross derivatives.[6] As individuals are homogeneous, they work and earn the same income $w \in \mathbb{R}^+$ in the CBD. Workers incur a total commuting cost $t : [0, b] \to \mathbb{R}^+$ with $t(0) = 0$ and $\mathrm{d}t/\mathrm{d}r > 0$. The price of composite good z is normalized to 1 without loss of generality. From this point on and whenever there is no confusion, we dispense the functions a, ℓ, n, s, t, x, z and R with reference to distance r.

approximation. This contrasts with the literature on the location of scarce public goods, where the travel to public facilities must be considered (see Cremer et al., 1986 and followers).

[6]We do not formally disentangle the land space used for houses and private gardens. First, our data on urban land cover do not permit to identify each separate item. Second, the empirical literature is ambiguous about the substitution effect of private gardens on the use of parks (also called the "compensation effect"). For instance, Talen (2010) and Caruso *et al.* (2018) find no significant relationship between the ownership of a private garden and the frequency of park visits. Finally, note that it can be shown that our theoretical results are unaltered if individuals have utility function $U(z, S, a)$ over a house service function S that is a homothetic consumption bundle $S(s_g, s_h)$ of land consumption for houses and private gardens s_h and s_g. In this case, private gardens substitute for parks, but the elasticity of substitution between private gardens and parks is the same as that between houses and parks.

3.3.1 Demand side

In most modern cities, residents freely choose their residential locations and spaces. They make their decisions according to the land rent values signaled in the urban land market. Given that urban green areas generate externalities to residents, it is important to highlight the conditions under which the competitive land market replicates the social optimum. We here discuss the equilibrium allocations in a competitive land market for an exogenous profile of green urban services $x : [0, b] \to [0, 1/\beta)$ and amenities $a = \alpha x$. A household's budget constraint is given by $z + Rs + t \leq w$, where $R : [0, b] \to \mathbb{R}^+$ is the land rent function of distance to the CBD. Households, therefore, face the following optimization problem.

$$\max_{s,z} \; U(z, s, a) \quad \text{s.t.} \quad w - t \geq z + Rs \tag{3.1}$$

In line with the literature, we assume that the land rent is paid to absentee landlords (Fujita (1989), Fujita and Thisse (2012)). Denote $y \equiv w - t$ as the disposable income after subtracting the commuting cost. The maximization problem (3.1) yields the demand for the composite good

$$\bar{z}(y, R, a) = y - R\bar{s}(R, y, a) \tag{3.2}$$

and the housing demand $\bar{s}(R, y, a)$, which is the unique solution of the following equation:

$$U_s(y - Rs, s, a) - RU_z(y - Rs, s, a) = 0. \tag{3.3}$$

We then obtain the indirect utility for each household

$$V(y, R, a) \equiv U(\bar{z}(R, y, a), \bar{s}(R, y, a), a) \tag{3.4}$$

However, housing price or land rent is rarely exogenous for households. The higher the desirability is of one location, the higher is the demand, and it leads to higher land rent. To investigate the equilibrium land price, we utilize the bid rent approach in the following section.

3.3.2 Competitive land market equilibrium

In a competitive land market equilibrium, each land slot is awarded to the highest bidder, and individuals have no incentives to relocate within and outside of a city. Therefore, they reach the same utility level u^e, where the superscript e refers to the equilibrium value. Households bid up to $(w - z - t)/s$ for each unit of residential space. Their bid rent $\psi : [0, b] \to \mathbb{R}^+$ is a function of distance r from the CBD such that

$$\psi = \max_{s,z} \frac{y - z}{s} \quad \text{s.t.} \quad U(z, s, a) \geq u^e. \tag{3.5}$$

As individuals compete for land, they increase their bids to make their participation constraint binding and obtain the equilibrium utility level u^e. To solve the problem,

we first denote $\tilde{z}(s, a, u^e)$ as the solution of $U(z, s, a) = u^e$. As household utility is an increasing and concave function of composite good z, it is apparent that $\tilde{z}$ exists and is unique. The bidding problem in equation (3.5) can be rewritten as follows

$$\psi = \max_s \frac{y - \tilde{z}(s, a, u^e)}{s} \tag{3.6}$$

The equilibrium housing slot size is given by the first-order condition

$$\tilde{z}(s, a, u^e) - s\tilde{z}(s, a, u^e) = 0 \tag{3.7}$$

We denote the solution for equation (3.7) as $\widehat{s}(y, a, u^e)$, which is also the households' optimal consumption.[7] Therefore, households' optimal consumption are given by the functions $\widehat{s}(y, a, u^e)$ and $\widehat{z}(y, a, u^e)$ as well as the bid rent $\widehat{\psi}(y, a, u^e)$. Ceteris paribus, we have the equilibrium housing slot $\widehat{s}$ that increases with distance to CBD (or lower y) and decreases with amenity level a. By the envelope theorem, the bid rent $\widehat{\psi}$ increases with disposable income y and amenity a and decreases with reservation utility u^e; that is, $\widehat{\psi}_y, \widehat{\psi}_a > 0$, while $\widehat{\psi}_u < 0$.

A competitive land market equilibrium is defined as the set of functions (z, s, R, n) and scalars (b, N, u^e) satisfying the following four conditions. First, individuals choose their optimal consumption: $z = \widehat{z}(y, a, u^e) \equiv \overline{z}(\widehat{\psi}(y, a, u^e), y, a)$ and $s = \widehat{s}(y, a, u^e) \equiv \overline{s}(\widehat{\psi}(y, a, u^e), y, a)$. Second, land is allocated to the highest bidder: $R = \max\{\widehat{\psi}(y, a, u^e), R_A\}$, with $R = \widehat{\psi}(y, a, u^e)$ if $n > 0$, and $R = R_A$ if $n = 0$. Third, the land market clears everywhere: $n\widehat{s}(y, a, u^e) = (1 - g)$ if $n > 0$. Finally, the total population conforms to its density: $N = \int_0^b n\, 2\pi r \mathrm{d}r$. Here, N is taken as exogenous in a closed city model, while u^e is exogenous in an open city. Within a city, equilibrium land rents are given by the winning bids such that $R = \widehat{\psi}(y, a, u^e)$. Since bid rents ψ increase with net income y and amenities a, the equilibrium land rent R falls with distance from the CBD but rises with the proportion of green urban area.

In equilibrium, land rents must exceed R_A for any location $r \in [0, b)$ and be equal to it at the equilibrium city border b^e. To simplify the exposition, we assume that $R^e(r)$ crosses R_A from above at $r = b^e$, which occurs if $\widehat{\psi}(y, a, u^e)$ lies above R_A in the CBD and falls in r. A sufficient condition is given by

$$\widehat{\psi}(w, 0, u^e) > R_A,$$
$$-\widehat{\psi}_y \frac{\mathrm{d}t}{\mathrm{d}r} + \alpha\widehat{\psi}_a \frac{\mathrm{d}x}{\mathrm{d}r} < 0.$$

After some reshuffling, this gives

$$w > \widehat{z}(w, 0, u^e) + R_A\widehat{s}(w, 0, u^e), \tag{3.8}$$
$$\frac{\mathrm{d}t}{\mathrm{d}r} > -\alpha\widetilde{z}_a \frac{\mathrm{d}x}{\mathrm{d}r}. \tag{3.9}$$

[7] We have the second-order condition $(\partial/\partial s)(\tilde{z} - s\tilde{z}_s) = -s\tilde{z}_{ss} < 0$ due to the concavity of the utility condition.

These sufficient conditions imply that urban productivity is sufficiently high for a city to exist in the absence of green urban areas and that green urban areas do not have excessively steep density profiles or do not yield too much spatial variation in amenities. Sufficiently high wages w and a low amenity parameter α guarantee these conditions. Under conditions (3.8) and (3.9), a spatial equilibrium exists. The equilibrium city border b^e is given by the unique solution of the land arbitrage condition: $R(b^e) = R_A$.

Proposition 1 *Suppose that conditions (3.8) and (3.9) hold. Then, a competitive land market equilibrium exists and is unique.*

The equilibrium population density is equal to $n^e = (1-g)/\widehat{s}(y, a, u^e) \geq 0$, while the equilibrium population aggregates the population density across the urban area as

$$N^e = \int_0^{b^e} \frac{1-\beta x}{\widehat{s}(y, a, u^e)} \, 2\pi r \mathrm{d}r = \int_0^{b^e} \frac{1-\beta x}{\overline{s}(\widehat{\psi}(y, a, u^e), y, a)} \, 2\pi r \mathrm{d}r.$$

To obtain more analytical results, we focus on narrower classes of preferences.

3.3.3 Specific preferences

In this subsection, we characterize the equilibrium land rent and population density for two classes of preferences: Cobb-Douglas and hyperbolic. Cobb-Douglas is a standard and classic homothetic preference, while the hyperbolic preference is a relatively less used preference and is nonhomothetic.

3.3.3.1 Cobb-Douglas preferences

We consider the following Cobb-Douglas utility

$$U(z, s, a) = z^{1-\phi-\varphi} s^{\phi} e^{\varphi a} \tag{3.10}$$

with $\phi, \varphi, (1-\phi-\varphi) \in (0, 1)$. Maximizing (3.10) with respect to z and s subject to the budget constraint $z + Rs \leq y$ leads to the linear expenditure system:

$$\overline{z} = \frac{1-\varphi-\phi}{1-\varphi} \, y, \tag{3.11}$$

$$\overline{s} = \frac{\phi}{1-\varphi} \frac{y}{R} \tag{3.12}$$

and the indirect utility function

$$V = (\phi\kappa)^{\phi} \, y^{1-\varphi} R^{-\phi} e^{\varphi a} \tag{3.13}$$

where $\kappa = (1-\varphi)^{-\frac{1-\varphi}{\phi}} \, (1-\varphi-\phi)^{\frac{1-\varphi-\phi}{\phi}}$.

Our objective now is to determine the land rent equilibrium at each location in the city area. Since the housing slot size is given for the highest bidder, what makes a

specific location attractive is both its closeness to the working place in the city center and the green amenity level. Solving the bidding problem in (3.5), we obtain the land rent at equilibrium:

$$\widehat{\psi} = \kappa \left(u^e y^{-(1-\varphi)} e^{-\varphi a} \right)^{-\frac{1}{\phi}}. \tag{3.14}$$

3.3.3.2 Hyperbolic preferences

In this section, we utilize a new class of model, which is not homothetic as in the Cobb-Douglas preference. We assume that individuals are endowed with the utility function as given below:

$$U(z, s, a) = z - \frac{\theta}{2s} + a \tag{3.15}$$

where preferences for residential space are represented by a hyperbolic function parametrized by θ.[8] Accordingly, we obtain the following expenditure system:

$$\bar{z} = y - \sqrt{\frac{\theta R}{2}} \tag{3.16}$$

$$\bar{s} = \sqrt{\frac{\theta}{2R}} \tag{3.17}$$

and the indirect utility function

$$V = y - \sqrt{2\theta R} + a. \tag{3.18}$$

As we can see from the previous part, the consumption of composite goods under Cobb-Douglas is constant and does not change with the new equilibrium of the housing market. However, with the hyperbolic function, the new equilibrium in the housing market will affect the consumption of the nonhousing good: higher land rent results in lower consumption of z. Solving the bidding problem in (3.5), we obtain the land rent at equilibrium for the hyperbolic preference as

$$\widehat{\psi} = \frac{(y + a - u^e)^2}{2\theta}. \tag{3.19}$$

The following section applies those theoretical results to the data on European cities.

3.4 Empirical analysis and welfare valuation

In this section, we estimate the relationship among population density, green urban spaces, distance to the city center and other controls, and we apply the estimates to our theoretical model. We then estimate the effect of altering the surfaces of green urban areas.

[8] This yields a demand for residential space that has a price elasticity between one and zero (see Mossay and Picard 2011).

3.4.1 Data

We utilize two main datasets in this paper. The first dataset is the population grid from Eurostat. The second dataset is the Urban Atlas implemented by the Global Monitoring for Environment and Security (GMES) service and provided by the European Environment Agency (EEA), for the time period 2005-2007. The Urban Atlas provides a classification of city zones that allows for a comparison of the density of residential areas, commercial and industrial zones and extent of green areas. The Urban Atlas uses Earth observation satellite images with 2.5 m spatial resolution.[9] According to the GMES, the dataset covers the functional urban areas (FUAs) of the EU cities with at least 100,000 inhabitants.[10] FUAs include land with both commuting distance and time below the critical levels defined by Eurostat.[11] The dataset includes all capital cities and covers nearly 300 of the most populous towns and cities in Europe (EU 27).[12]

In this paper, we use the data on "green urban areas" (class 14100), which are defined as artificial nonagricultural vegetated areas. They consist of areas with planted vegetation that is regularly worked and/or strongly influenced by humans. More precisely, first, green urban areas include public green areas used predominantly for recreational use (gardens, zoos, parks, castle parks, cemeteries, etc.). Second, suburban natural areas that have become and are managed as urban parks are included as green urban areas. Finally, green urban areas also include forest and green areas that extend from the surroundings into urban areas with at least two sides being bordered by urban areas and structures and containing visible traces of recreational use. Importantly, for our study, green urban areas do not include private gardens within housing areas, buildings within parks, such as castles or museums, patches of natural vegetation or agricultural areas enclosed by built-up areas without being managed as green urban areas. It must be noted that green urban areas belong to the Urban Atlas' class of "artificial surfaces", which includes all nonagricultural land devoted to human activities.[13] This class is distinguished from the agricultural, seminatural areas and wetlands, forest areas and water areas devoted to nonurban activities.

We select the (oldest) town hall locations as the CBDs. Then, we create a set of annuli (rings) around each CBD at 100 m intervals. We define the "annulus land area" as

[9] GMES maps have a 100-times higher resolution than that of traditional maps in the CORINE Land Cover inventory produced since 1990.

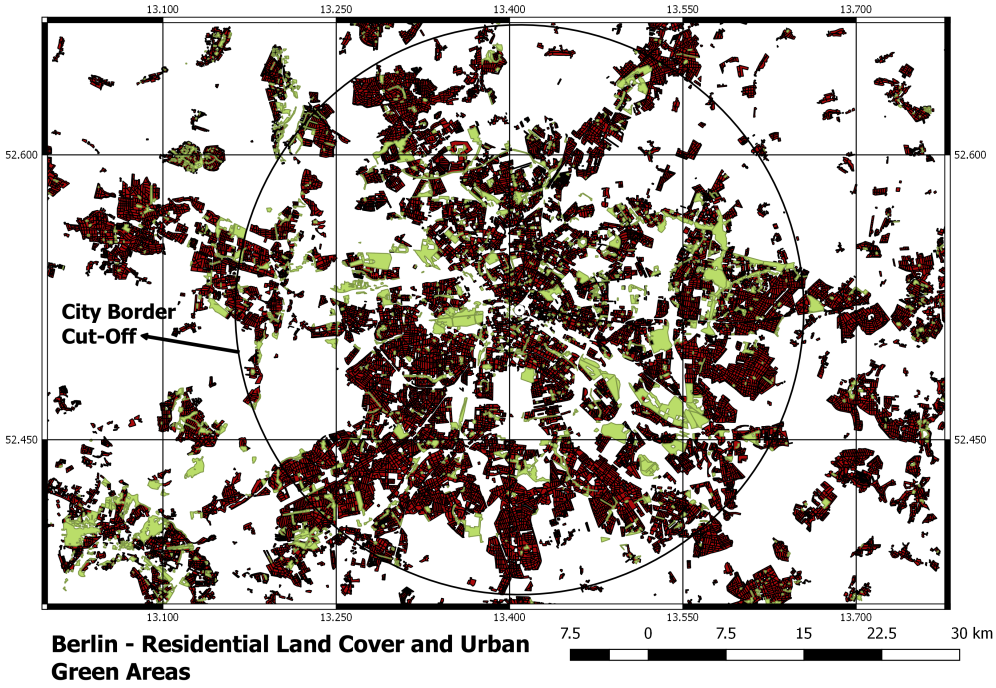
[10] See the definition in the Urban Audit database and European Environmental Agency, GMES Urban Atlas Metadata. Link: https://land.copernicus.eu/local/urban-atlas (accessed on Jan 25, 2018).

[11] See the definition in the Urban Audit in EEA, 2015, and the details in Appendix B.

[12] Austria, Belgium, France, Germany, Bulgaria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom.

[13] In addition to green urban areas, artificial surfaces include urban areas with dominant residential use, inner city areas with a central business district and residential use, industrial, commercial, public, military and private units, transport units, mines, dump and construction sites, and sports and leisure facilities.

Figure 1: Berlin Land Use maps



the intersection of the annulus and the land within the urban zone area reported by the GMES. This area includes artificial surfaces, agriculture, seminatural areas, wetlands and forest but does not include water areas because those seas and oceans are not appropriate for potential human dwellings. We then compute the share of green urban area as the ratio of the surface of green urban area to the total land in the annulus land area for each annulus. Figure 1 displays the land uses of green urban areas (green color) for Berlin. Whereas urban theoretical models usually assume a neat frontier between residential and nonresidential spaces, urban data do not provide a clear separation between residential locations and agricultural areas and forests. In this paper, we choose to fix the city borders to the annulus for which the ratio of residential surfaces over the annulus land area falls below 20%. Residential surfaces include urban areas with dominant residential use and inner city areas with a central business district and residential use. They are shown in red in Figure 1.[14] We define the distance from the CBD, *dist*.

In addition to the GMES, we use the population density from the European population grid.[16] We calculate the population mass at each distance to the city center and

[14]The use of other thresholds does not lead to qualitative differences in our empirical results.

[16]For more information on the European population grid, please check the technical report of the GEOSTAT 1A project from Eurostat. Link: `http://ec.europa.eu/eurostat/documents/4311134/4350174/ESSnet-project-GEOSTAT1A-final-report_0.pdf/fc048569-bc1c-4d99-9597-0ea0716efac3` (Accessed on May 30, 2018).

Table 3.1: Summary statistics

	average	sd	min	max	obs.
City border (km)	4.3	3.2	1.0	24.0	305
City area (km^2)	84	174	1	1809	305
Number of annuli	43	32	10	240	305
Population in FUA (millions)	0.79	1.29	0.06	12.10	301
Population in CGC (millions): all cities	0.44	0.79	0.03	8.17	305
GDP per capita (€1000/hab.)	26.88	13.06	6.00	83.70	305
Household income (€1000/hab.)	15.46	5.63	3.70	30.90	304
Density (hab./$100m^2$)	0.44	0.28	0.10	1.96	303
Residential Space ($100m^2$)	0.98	0.44	0.21	2.47	304
City Geographical Controls					
Elevation (m)	212	210	−2	1,614	305
Average temperature at Jan 01 (^oC)	2.30	4.62	−8.48	15.57	305
Average temperature at July 01 (^oC)	19.18	12.27	12.27	28.70	305
Average daily precipitation (mm/day)	1.91	0.60	0.48	4.45	305
Share of Urban Green Land (%)	6.57	4.38	0.62	42.69	304
Share of Residential Land (%)	34.03	4.99	12.85	47.42	304
Share of Industrial and Public Land (%)	16.61	6.02	2.26	47.57	304
Share of Sport and Leisure Land (%)	3.76	3.21	0.00	12.79	304
Share of Forest Land (%)	5.14	6.23	0.00	32.57	304
Share of Agricultural Land (%)	16.29	10.75	0.00	52.13	304
Share of Forest Land within 100m buffer (%)	1.43	2.04	0.00	13.42	304
Share of Agricultural Land within 100m buffer (%)	6.15	5.66	0.00	33.19	304

Note: The GMES database released on May 2010 reports only 301 FUAs for the time period 2005-2007. We use the Nadaraya-Watson Gaussian Kernel to smooth variations of annuli values.[15] GDP per inhabitants and Household income are taken from Regional Economic Accounts from Eurostat at NUTS3 and NUTS2 level respectively. Note that in Eurostat database, household income level exists only at NUTS2 level. In eurostat database for household income at NUTS2, there is no data for Luxembourg (NUTS2 code LU00); therefore, there is only 304 cities instead of 305 cities. The number of inhabitants in each annuli is calculated based on Eurostat Population Grid. As Eurostat Population Grid 2006 does not cover Cyrus; hence, we also drop the city cy0011_lefkosia in our database. The total number of annuli are calculated for 303 cities excluding lu0011_luxembourg and cy0011_lefkosia. For city geographical controls, we take into account the average for period 1995-2010 for each city.

redistribute the population to the residential area in each annulus.[17] Because the Eurostat population grid does not cover Cyprus, we exclude Lefkosia, Cyrus. For the income level of a city, we use the household net income at the NUTS2 level, as reported in the Eurostat's Regional Economic Accounts, which provides the finest detail on household net income. Our results are robust to the use of the city's per capita GDP at the NUTS3 level.[18] Other measures of cities' exogenous geographical characteristics are taken from the E-OBS database.[19] We control for these exogenous geographical characteristics because they may affect residential choices. We finally measure the city populations as the number (millions) of inhabitants living in the city and greater city (CGC) areas,

[17]Residential areas are called 'urban fabrics' in the GMES.

[18]See Supplementary Material.

[19]The E-OBS database provides detailed data on the daily temperature, daily precipitation, sea level pressure and elevation across Europe. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (`http://ensembles-eu.metoffice.com`) and the data providers in the ECA&D project (`http://www.ecad.eu`).

as defined and reported in the Eurostat databases.[20] Table 3.1 presents the summary statistics.

In this paper, we mostly use the household income that measures the *per capita* income net of all income taxes and at the NUTS2 level. Household incomes vary greatly across EU cities, from €3,700 per inhabitant to €30,900 per inhabitant. The average income is €15,460. Household income represents slightly more than one-half of the per capita production value (NUTS3), which reflects the high tax wedge between production cost and net income in the EU. The share of urban green, on average, accounts for approximately 6.5% of the total surface of city areas. Cities have a rather heterogeneous share of green urban areas. In our sample, the city with lowest share of green urban area (0.62%) is Limerick, Ireland, and the city with the highest share (42.6%) is Karlovy Vary, Czech Republic.[21] City elevation also varies greatly, from two meters below sea level in Amsterdam, the Netherlands, to 1,614 meters above sea level in the mountainous city of Innsbruck, Austria. European cities belong to a mild climate zone, with temperatures varying between -8 and $+28$ degrees Celsius at the lowest and highest day temperature in winter and summer (measured on January 1 and July 1, respectively, for the period 1995-2000).[22] The average population density is approximately 4,400 inhabitants per square kilometer and ranges from 1,000 to 9,800 inhabitants per square kilometer. These numbers are reasonable because the database concentrates on the core of urban areas with no agricultural or seminatural land use.

In the next section, we estimate residents' land use under the specification of Cobb-Douglas preferences. Those are convenient and popular in urban economics. We then present the same analysis for the hyperbolic preferences.

3.4.2 Estimation and valuation with Cobb-Douglas preferences

We suppose the presence of observable heterogeneity χ in the preference for land plots or specific characteristics of locations as well as unobservable heterogeneity or measurement errors ϵ. The utility function becomes

$$U(s, a, z, \chi, \epsilon) = \epsilon \chi z^{1-\phi-\varphi} s^{\phi} (e^{a})^{\varphi}.$$

We assume that the transport cost $t = w(1 - e^{-\tau(r)})$; thus, the net income is given by $y = w - t = we^{-\tau(r)}$, where $\tau(r)$ is a function of distance to the CBD. For simplicity, we assume the quadratic form $\tau(r) = \tau_1 \times r + \tau_2 \times r^2$. Green area amenities are given by $a = \alpha x$, which can be written as a function of green urban areas g as $a = \alpha g / \beta$,

[20] For more details, see metadata files for *urb_esms* in the Urban Audit database of the Eurostat website.

[21] The latter is a spa resort city, which offers many green areas to its visitors. The former city includes few land surfaces classified as green urban areas because it also has many agricultural and seminatural lands that can be used for urban green amenities. These outliers do not affect our results.

[22] We use observations from the E-OBS databases from the EU-FP6 project (for details, see the references). Our samples do not contain some northern European cities in Iceland and Norway.

where α and β are green amenity and land use intensity parameters, respectively. We can further standardize the amenity value $\alpha = 1$ and consider β as land use intensity to provide one unit for green amenities. We obtain the housing slot size in equilibrium as below:

$$\widehat{s} = \left(\frac{1-\varphi}{1-\varphi-\phi}\right)^{\frac{1-\varphi-\phi}{\phi}} \left(\frac{u}{\epsilon\chi}\left(we^{-\tau(r)}\right)^{-(1-\varphi-\phi)} e^{-\varphi\frac{g}{\beta}}\right)^{\frac{1}{\phi}} \tag{3.20}$$

Taking the natural logarithm of (3.20) and adding the heterogeneity and error terms, we obtain the following residential land use:

$$\begin{aligned}\ln\widehat{s} = \ln\left(\frac{1-\varphi-\phi}{1-\varphi}\right)^{-\frac{1-\varphi-\phi}{\phi}} + \frac{\varphi}{\phi}\ln\beta - \frac{1-\varphi-\phi}{\phi}\ln w \\ + \frac{1-\varphi-\phi}{\phi}\tau_1 r + \frac{1-\varphi-\phi}{\phi}\tau_2 r^2 - \frac{\varphi}{\phi\beta}g + \frac{1}{\phi}\ln\bar{u} - \frac{1}{\phi}\ln\chi - \frac{1}{\phi}\ln\epsilon.\end{aligned}$$

Accordingly, residents have larger land plots for cities with smaller incomes w, larger distances between residences and the CBD r, smaller green urban areas g, higher outside utility $\overline{u}$, and smaller observable characteristics χ.

3.4.2.1 Estimation

From these results, we build a regression model of residential land use

$$\ln(s_{ijc}) = \vartheta_0 + \vartheta_1 \ln w_{jc} + \vartheta_2 \text{dist}_{ijc} + \vartheta_3 \text{dist}_{ijc}^2 + \vartheta_4 g_{ijc} + \vartheta_{5c} I_c + \vartheta_6 X_{jc} + \vartheta_7 A_{ijc} + \varepsilon_{ijc}$$

for the observations of annulus i in city j of country c. We measure the city wage w_{jc} by the per capita household net wage in the NUTS2 areas[23] and the green urban areas g_{ijc} by the land share of green urban areas. Given the language, cultural and administrative barriers, we consider that individuals freely move across cities only within the same country. Thus, the country utility level is captured by the vector of country dummies I_c. Finally, vector X_{jc} controls for observed city characteristics, such as elevation, rainfall and temperature. Vector A_{ijc} controls for observed amenities in each annulus, such as the shares of sport leisure facilities and industrial lands and the shares of forest and agricultural lands within a 100 m distance from the residential areas.

A potential endogeneity issue arises because the choices for residents' land use and planers' green urban areas are intertwined. Indeed, urban planners are expected to organize green urban areas as a function of surrounding population densities and therefore residents' land use. To control for such a reverse causality, we use the historical

[23]In this text, city wages are measured by the incomes net of taxes at the NUTS2 level. Net incomes closely reflect the budget constraints faced by residents in their land use choices. However, NUTS2 encompasses larger areas than the cover of many cities, which may downward bias city income values. In Appendix C, we perform the same analysis with the production value at the NUTS3 level, which includes taxes. The results are similar except the values should be interpreted differently.

level of urban green areas as indicative of the current levels. The main idea behind using historical urban green area information is that once an urban green area is developed, it is rarely changed. In fact, many urban green areas in Europe were provided decades ago and have remained intact. Examples are Hyde Park in London, created in approximately the 16^{th} century by Henry VIII and originally intended for hunting, and the 'Jardin du Luxembourg', which was first built as a private garden of Queen Marie de Medici in the early 17^{th} century. Both private parks were later converted to public green areas by public authorities. Thus, we can mitigate the reverse causality using the data for old parks to predict the locations of current public green areas. Toward this aim, we use the CORINE Land Cover 1990 database, which unfortunately does not cover all GMES Urban Atlas countries. As a result, the regression results exclude cities in the UK, Sweden and Finland.[24] The details for the first-stage regression are reported in Appendix C, which confirm that the historical levels of urban green areas are a good predictor of the current levels.

The results are reported in Table 3.2. Columns (1) to (4) display OLS estimates without instrument variables. In all columns, the coefficient estimates are consistent with our model predictions: residents use larger land plots for smaller city income, larger distance between residences and the CBD and smaller green urban areas. The results are robust after controlling for country fixed effects, city geographical conditions, such as elevation, rainfall and temperature (see Columns (2) and (3)), and different types of amenities within annuli (see Column (4)). To account for the above endogeneity issue, Columns (5) to (8) display the same estimates after instrumenting the share of urban green areas with its historical value from the CORINE Land Cover 1990 database. The IV regression reports slightly stronger effects of urban green areas on the residential slot size than those of the OLS regressions, which is intuitive because the historical level of urban green areas was lower than the current level. We also apply the Wu-Hausman test for endogeneity (reported in Appendix A). The Wu-Hausman test coefficient is not significant at the 90% confidence level, which supports the alternative hypothesis of no endogeneity at the 90% confidence level and suggests that endogeneity may not be a critical issue in our analysis. Both the OLS and IV results show significant coefficients for the share of green urban area amenities g for approximately $2.24 - 2.25$ before including the control and $1.69 - 1.86$ after including all other controls. This finding implies that, ceteris paribus, residents in annuli with no green urban areas use 14% more

[24]CORINE Land Cover (CLC) 1990 does not cover the UK, Sweden and Finland. The database covers Austria, Belgium, Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, the Netherlands, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, and Turkey, out of which 23 countries are included in our data. For details, see CORINE Land Cover 1990 Metadata: `https://land.copernicus.eu/pan-european/corine-land-cover/clc-1990?tab=metadata` (Accessed May 02, 2018). To our knowledge, CLC 1990 is the oldest land use database that systematically covers all of European cities.

land than those residing in annuli with a 7% share of green urban area.[25] Population densities are reduced in the same proportions. According to this empirical estimation, green urban areas are an important factor explaining the use of residential land and population density.

3.4.2.2 Welfare valuation

In this subsection, we use the previous regression model to quantify the value of green urban areas with a focus on the general equilibrium effect. We recover all parameters of our theoretical model and run several counterfactual analyses. In particular, we build counterfactuals where half of the green urban areas are deleted in every annulus and are either left unused or converted to new residential land. We can then evaluate the changes in the residential land use and consumption of goods, population density, land rents and utility levels for each city. To express utility changes more intuitively, we measure the cost to residents by their incentives to leave the city and compensating variation wage (subsidy or tax) that they must receive to keep their utility levels. The difference between the status quo wage level and the new compensating variation income is precisely the willingness to pay for each household (Sieg et al. (2004)). By the same token, we discuss the distribution of the effect of green urban areas between cities and within them. We consider this analysis a useful exercise because it informs policy makers about the impact of urban green areas on city structures and sizes.

We recover the model parameters from the estimated coefficient of residents' land use using the values of ϑ_0, ϑ_1, ϑ_2, ϑ_3 and ϑ_4 from Column (8) in Table 3.2. Country utility levels are recovered from the parameters ϑ_{5c} and the constant term ϑ_0. Our baseline model and counterfactuals use the observed distance to the city center, city and country characteristics and local (nongreen) amenities. The baseline model simulates the variables under study using those estimated parameters and the observed characteristics (distance to CBD, wage, green urban areas, etc.). The counterfactual exercises investigate the impact of canceling 50% of the urban green areas in each annulus of each city, keeping the same observed characteristics. Counterfactual exercise 1 considers open cities where utility levels and unobserved heterogeneity are maintained. This helps us discuss a long-term and unregulated perspective, where urban planners do not impose restrictions on workers' mobility within and between cities. Counterfactual exercise 2 considers closed cities with exogenous city populations. The study of closed cities can be appropriate in evaluating policy changes that occur simultaneously in all cities, such as changes in EU policies.[26] Here, our aim is to discuss a midterm or regulated perspective, where urban planners are able to restrict workers' mobility between cities but allow residents' land use to change. To give a relevant measure of utility change,

[25] We compute $s_{ijc}/s_{i'jc} = e^{\vartheta_3(g_{ijc}-g'_{ijc})}$, with $g_{ijc} = 0$ and 0.07.

[26] Cheshire and Shepard (2002) also use the closed city model to analyze the welfare effects of policy changes in the UK.

Table 3.2: Residents' land use

	Dependent variable: Ln Residential Space							
	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-0.507^{***}	-1.448^{***}	-1.517^{**}	-0.896	-0.505^{***}	-1.420^{***}	-1.478^{**}	-0.868
	(0.143)	(0.356)	(0.631)	(0.556)	(0.145)	(0.360)	(0.635)	(0.560)
Ln Household Income	-0.074	-0.758^{***}	-0.784^{***}	-0.735^{***}	-0.074	-0.752^{***}	-0.778^{***}	-0.730^{***}
	(0.071)	(0.173)	(0.167)	(0.151)	(0.072)	(0.174)	(0.168)	(0.151)
Distance to CBD	1.502^{***}	1.412^{***}	1.455^{***}	0.885^{***}	1.497^{***}	1.415^{***}	1.448^{***}	0.879^{***}
	(0.237)	(0.220)	(0.212)	(0.213)	(0.237)	(0.221)	(0.212)	(0.213)
Distance to CBD square	-0.607^{***}	-0.549^{***}	-0.553^{***}	-0.314^{**}	-0.604^{***}	-0.556^{***}	-0.551^{***}	-0.312^{**}
	(0.130)	(0.130)	(0.133)	(0.129)	(0.130)	(0.130)	(0.133)	(0.127)
Share of Urban Green	-2.241^{***}	-2.052^{***}	-1.941^{***}	-1.695^{***}	-2.253^{***}	-2.274^{***}	-2.153^{***}	-1.866^{***}
	(0.400)	(0.254)	(0.251)	(0.214)	(0.431)	(0.285)	(0.273)	(0.237)
Country FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
City Geographical Controls	No	No	Yes	Yes	No	No	Yes	Yes
Annuli Amenity Controls	No	No	No	Yes	No	No	No	Yes
Sample	All	All	All	All	All	All	All	All
Observations	10,853	10,853	10,853	10,853	10,853	10,853	10,853	10,853
Adjusted R^2	0.207	0.468	0.483	0.579	0.197	0.464	0.480	0.576
Residual Std. Error	0.574	0.470	0.464	0.418	0.578	0.472	0.465	0.420
df	10,848	10,827	10,823	10,819	10,848	10,827	10,823	10,819
F Statistic	19.35^{***}	26.02^{***}	24.35^{***}	34.07^{***}	18.23^{***}	24.60^{***}	24.45^{***}	34.67^{***}

Note: Here, we use the Household income taken from Regional Economic Accounts from Eurostat at NUTS2 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €100,000. The distance to CBD is measured in 10 kilometres. The inverse of residential space is calculated by dividing the number of inhabitants in each annuli with annulus areas net of its urban green space. We exclude Cyrus and Luxembourg as the Eurostat population grid database does not cover Cyrus and the household income data at NUTS2 of Eurostat does not cover Luxembourg. United Kingdom and Finland are also excluded as they are not covered by Corine Land Cover 1990. City boundary is chosen at 20% cut-off point. For city control, we take into account the elevation, average rain fall, average temperature in Jan 01 and average temperature in July 01 for period 1995-2010. City amenity controls include the share of industrial, sport and leisure land use as well as the share of forest and agriculture land within 100 meters buffer from residential area. Standard errors are clustered at city level. The row "df" reports the degree of freedom. Significance levels are denoted by * for $p<0.1$, ** for $p<0.05$ and *** for $p<0.01$.

Table 3.3: Counterfactual analysis: open and closed cities

	Cities	Composite Goods (Z)	Housing Rent ($R \times s$)	Income (W)	Residential Area	Green Area (GA)
	number	(€1000)	(€1000)	(€1000)	(km^2)	(km^2)
Spatial heterogeneity	264	5.05	6.92	15.22	25.29	4.52
		(1.87)	(2.56)	(5.87)	(48.91)	(9.87)
No spatial heterogeneity	264	5.03	6.89	15.22	25.29	4.52
		(1.86)	(2.55)	(5.87)	(48.91)	(9.87)

Note: Household income is taken from Eurostat at NUTS2 level and is measured on purchasing power standard (PPS) at 1000€. More details on PPS measure, please check Eurostat technical documents. The standard deviation is reported in the parenthesis.

we also compute the *compensating variation wage* as the city wage that maintains the baseline utility level when we remove green urban areas and the corresponding willingness to pay. Counterfactual exercises 1 and 2 hinge on the assumption that empirical model residuals reflect land heterogeneity that is unobserved to the econometricians but observed and used by residents in their land plot size choices. Such heterogeneity is reported in the counterfactual results. This assumption may be strong, as it imposes strong information on behalf of residents. Therefore, we also take the opposite view in counterfactuals 3 and 4 where the residuals consist of measurement errors that can be observed neither by the econometricians nor the residents. The details of the counterfactual analysis are relegated to Appendix B.

The results of the baseline model are displayed in Table 3.3. The first column reports the number of cities in the baseline exercise. Every other column reports the average and standard deviation over the city averages imputed from the baseline model. The second and third columns show the consumption of composite goods and housing by households, while the fourth column displays the net income. The difference between this column and the sum of the two previous columns accounts for commuting costs. On average, individuals have €15,220 as disposable income and spend €6,920 for housing expenses, which accounts for approximately 45% of their net income. Such a value is slightly above the average housing costs in European cities, which are approximately a quarter of the household income for both European rural and urban areas. The literature reports a range between 18% and 32%, with higher levels for urban areas and renters (Fahey et al, 2004, Davis and Ortalo-Magne, 2011). Our model differs from this literature because we do not take into account housing furniture and maintenance (5% of housing costs in Eurostats, 2015), we consider city cores, which have more expensive housing locations, and finally, we do not model the construction process. The last two columns of Table 3.3 report the average residential area and green urban area across cities, the latter being approximately one-fifth of the former. The rows address the cases when we consider the regression model error as a spatial amenity for the residents (spatial heterogeneity) and when we do not (no heterogeneity). The difference between the two cases is not large.

Our main results are displayed in Table 3.4, which shows the baseline model (first row) and the counterfactual exercises (other rows) for cases with and without spatial heterogeneity. The table structure is the same as that of Table 3.3. Every column reports the average and standard deviation of the city averages over annuli imputed in the baseline and counterfactual exercises. Consider the first row, which presents our baseline model and permits comparison with the literature. The first five columns display the imputed residential surfaces, land rents on units of residential plots and green urban areas, population and relative utility. The average city size of 0.31 million inhabitants is consistent with the statistics that most European cities are medium-sized (European Commission and UN-Habitat (2017), Urban Audit, Eurostat). Residents' average use of space is approximately 95 m^2; the measure is reasonable given that we consider the core of the most populous cities in the EU, which are the densest areas of the most urbanized parts of the EU.[27] The land rent per square meter is 93.41 €/m^2/year on average for all cities. On average, the land values of green urban areas (102.32 €/m^2/year) are higher than residential land prices. These values are imputed from the residential land price associated with each annulus. Because urban green areas are concentrated at close and intermediate distances to CBDs, they are surrounded by more expensive residential land plots.

Consider, now, counterfactual exercise 1, where one removes 50% of the green areas in every annulus of open cities. In open city systems, the utility of city inhabitants is exogenous, but the change in green amenities affects the urban structures. Suppose, furthermore, the case where the removed land is not (re)used, as indicated in the second row of Table 3.4. To keep the same utility level, residents must compensate for the decrease in urban green amenities by larger residential plots, which implies that a share of the population must migrate out of the city. On average, city residents increase their land use from 95 to 100 m^2 (a rise of 5.3%), and the city population falls from 0.31 to 0.29 million (a loss of 6.5%). Land rents fall from 93.41 to 87.11 € per m^2 and year (a fall of 6.74%). We compute the total loss in the land market to be approximately €147 million for an average city.[28]

Suppose now that 50% of green urban areas are converted into residential land, as shown in the third row of Table 3.4. In an average city, there is a new land supply of 2.26 km^2 (half of 4.52 km^2) on top of the baseline residential land supply of 25.29 km^2, an increase of 8.9% (see Table 3.3). This increase is slightly more than the 5.3% space compensation that residents required without land conversion. As a result, the additional land supply attracts new city dwellers, and the city population rises to 0.32 million on average. Residential land rents rise slightly to 87.64 € per m^2 and year because the new

[27] They are more densely populated than US cities.

[28] To estimate the total land rent loss, we multiply the city residential area of each annulus with the per-square-meter land rent loss between the baseline and counterfactual models, aggregate over the city, and compute the average over all cities.

land is supplied at more central locations with higher values. We compute that, compared to the baseline model, the housing market increases its total value by nearly €55 million per year and city when we take into account converted areas.

Counterfactual exercise 2 allows us to discuss the impact of reducing urban green areas by half in closed cities where city planners prohibit migration. As predicted by our theoretical model, when there is no conversion of land, the utility of all residents decreases once we reduce the level of urban green amenities. Specifically, the average utility decreases from 1 to 0.94. Residential land rents decrease only by a small amount from 93.41 €/m^2/year to 93.36€/m^2/year, providing a total loss of €1.82 million per year. The decrease in utility combined with the change in land rent requires an increase in the baseline annual net income of €15,200 to the compensating-variation income of €16,650, an increase of €1,430 (9.4%). This is also the measure of our general equilibrium WTP as indicated in section 3.3 for canceling half of the urban green areas without land conversion in the city area. Multiplying this figure by the city population, we obtain a subsidy of €580 million for an average city. For reference, the general equilibrium WTP for closing all publicly accessible open spaces in Cheshire and Sheppard (2002) is £2,424 per household per year for the city of Reading, UK. Our average measure is not too far in magnitude given the 50% closing of green urban areas in our samples.

Suppose now that green urban areas are converted into residential land. Then, the residential land supply increases, land rents decreases, and residents can use more land to compensate for the loss of green area amenities. Specifically, the land rents drop by 9.6% to 84.47 €/m^2/year, and the total loss in housing market increases to €3.45 million per year. However, city residents enjoy both lower land rent and larger residential land plots, which increase their average utility level. They obtain a slightly higher average utility level (increase by 3%) and require a smaller compensating-variation wage of €14,570 per year to maintain their level of utility, which is equivalent to an income reduction of €650 per year. Those results are also in line with those of Cheshire and Sheppard (2002), who also found that the net effect of land use planning is negative and that allowing more residential developments is welfare improving.

The bottom panel of Table 3.4 replicates the above analysis when we replace the assumption of unobserved spatial heterogeneity by that of measurement errors. It can be seen that most effects are similar. The main differences lie in the level of averages and standard deviations of our variables of interest. On the one hand, the standard deviations are naturally smaller because residents are no longer assumed to consider spatial amenity variations in their choices. However, the reduction in the standard deviation is not drastic, which shows that the model is already well explained by the independent variables of the regression model. On the other hand, the assumption of measurement errors also alters the average values of imputed variables. For instance, in the baseline model of the bottom panel, the residential populations, land uses and land rents are smaller than in the baseline model of the top panel. This is because

Table 3.4: City structure under closed and open scenarios $g = 0.5 \times g_0$

		Cities (number)	s ($100m2$/hab)	Land Rent (€/m^2/y)	Green value (€/m^2/y)	Pop. (mil.)	$U/\bar{U}^b$	Comp. W (€1000)	Total loss in housing market (€ *mil.*)	Total loss in wage comp. (€ *mil.*)	Total loss in pop. 1,000*hab.*
Baseline 0 (spatial heterogeneity)		264	0.95	93.41	102.32	0.31	1.00				
			(0.44)	(68.95)	(75.71)	(0.69)	(0.36)				
$g = 0.5 \times g_0$											
Counterfactual 1: Open City											
Urban green conversion	No	264	1.00	87.11	92.73	0.29	1.00		146.61	0	20.38
			(0.46)	(63.11)	(66.70)	(0.64)	(0.36)		(407.11)		(52.53)
	Yes	264	1.00	87.64	92.73	0.32	1.00		−54.74	0	-8.24
			(0.46)	(63.46)	(66.70)	(0.71)	(0.36)		(126.62)		(18.39)
Counterfactual 2: Closed City											
Urban green conversion	No	264	0.95	93.36	99.38	0.29	0.94	16.65	1.82	580.56	0
			(0.44)	(69.00)	(72.98)	(0.64)	(0.34)	(6.46)	(16.97)	(1926.01)	
	Yes	264	1.03	84.47	89.26	0.32	1.03	14.57	3.45	−198.56	0
			(0.47)	(60.75)	(63.28)	(0.71)	(0.38)	(5.71)	(15.30)	(544.27)	
Baseline 0 (no spatial heterogeneity)		264	0.88	93.70	100.81	0.28	1.00				
			(0.34)	(59.34)	(61.55)	(0.50)	(0.36)				
$g = 0.5 \times g_0$											
Counterfactual 3: Open City											
Urban Green Conversion	No	264	0.93	87.37	91.53	0.29	1.00		130.20	0	18.13
			(0.35)	(54.04)	(54.44)	(0.64)	(0.36)		(311.97)		(41.07)
	Yes	264	0.93	87.60	91.53	0.32	1.00		−50.81	0	-7.58
			(0.35)	(54.03)	(54.44)	(0.71)	0.36)		(115.20)		(16.41)
Counterfactual 4: Closed City											
Urban Green Conversion	No	264	0.88	93.64	98.01	0.29	0.94	16.65	2.26	511.77	0
			(0.34)	(59.40)	(59.64)	(0.64)	(0.34)	(6.47)	(13.72)	(1483.03)	
	Yes	264	0.96	84.72	88.58	0.32	1.03	14.61	2.88	−183.74	0
			(0.36)	(51.84)	(52.31)	(0.71)	(0.37)	(5.70)	(11.47)	(484.95)	

Note: Household income is taken from Eurostat at NUTS2 level and is measured on purchasing power standard (PPS) at 1000€. More details on PPS measure, please check Eurostat technical documents. Population is calculated in million inhabitants. Utility is measured relatively to baseline average. Basically, we measure $\frac{U_{jc}}{\bar{U}^{baseline}} = \frac{U_{jc}}{\frac{1}{N}\sum_{jc} U_{jc}^{baseline}} = \frac{u_{jc}^{1/\hat{\phi}}}{\frac{1}{N}\sum_{jc}\left(u_c^{baseline}\right)^{1/\hat{\phi}}}$ where u_{jc} is calculated as in the above section. The standard deviation is reported in the parenthesis.

those variables are nonlinear functions of the error term ε under the Cobb-Douglas preferences.[29]

What type of cities are more sensitive to removing green urban areas? Where in the city are the changes more important? To answer these questions, we compare the impact of removing or converting green urban areas between cities of different incomes and population sizes as well as between within-city locations at different distances to the CBD. Toward this aim, Table 3.5 reports the baseline wage (first row), the changes in the compensating-variation wages to sustain constant utility (next four rows), the baseline land rent to landlords (sixth row), and the landlords' losses (last four rows) when we group cities by income quartiles (first four columns), by population size quartiles (next four columns) and by quartiles of relative distances to the CBD (last four columns). The positive changes in compensating-variation wages can be interpreted as subsidies required to maintain the residents at their baseline equilibrium levels. Table 3.5 presents the results for open and closed cities with and without the conversion of green urban areas to residential land. All figures are aggregated from the same counterfactual exercise with unobserved spatial heterogeneity and with the 50% reduction in the green urban areas presented in Table 3.3 and 3.4.

Let us consider first the case of open cites (counterfactual exercise 1). The reduction in green urban area amenities harms residents who partly leave the city. Because of free migration, city residents keep the same utility as in the countryside and ask for no compensation to stay in cities, which is why the second and third rows in Table 3.5 display a set of zeros for the change in compensating-variation wages. By contrast, landlords lose money. If urban green areas are not converted to residential land, they lose €3.27 and €8.22 per m^2 and year in cities belonging to the bottom and upper income quartiles, respectively (see seventh row). Similarly, they lose €5.21 and €7.90 per m^2 and year in cities belonging to the bottom and upper population quartiles, respectively. This result is explained by the fact that land value, city size and income are positively correlated. Landlords also lose €11.24 per m^2 and year in the central city quartile but only €3.89 per m^2 and year in the city periphery quartile, indicating that land rents decrease with distance from the CBD. This pattern remains approximately the same if urban green areas are converted to residential land (eighth row). In this case, the above figures decrease by approximately €1 per m^2 and year in the lowest income and population size cities but only slightly for those with the highest income and population size. The conversion of green urban areas mitigates the conclusions only to a small extent.

Let us now consider closed cities in which migration is restricted and half of the green urban areas is removed (conterfactual exercise 2). Suppose, initially, that there is no land conversion (fourth row in Table 3.5). To stay in the city, residents require an

[29]The expression of land use $\hat{s}(w, r, g, u, \epsilon)$ includes a term e^{ε} that is a convex function of ε. Similarly, the expression for the population density $1/\hat{s}(w, r, g, u, \varepsilon)$ includes a term $e^{-\varepsilon}$, which is also a convex function of ε.

Table 3.5: Counterfactual analysis: between- and within-city statistics

		Between cities								Within cities			
		City Income Quartiles				City Population Size Quartiles				Distance to CBD Quartiles			
		0-0.25	0.25-0.50	0.50-0.75	0.75-1	0-0.25	0.25-0.50	0.50-0.75	0.75-1	0-0.25	0.25-0.50	0.50-0.75	0.75-1
Baseline wage (€1000/hab/y)		8.19	12.27	17.71	22.93	12.36	14.97	15.85	17.72	16.71	16.62	16.64	16.54
		(2.03)	(1.23)	(1.16)	(2.23)	(4.47)	(5.91)	(6.19)	(5.57)	(5.92)	(5.91)	(5.92)	(5.92)
WTP (€1000/hab/y)													
Counterfactual 1: Open City													
Urban green conversion	No	0	0	0	0	0	0	0	0	0	0	0	0
	Yes	0	0	0	0	0	0	0	0	0	0	0	0
Counterfactual 2: Closed City													
Urban green conversion	No	−0.77	−1.18	−1.67	−2.09	−1.28	−1.37	−1.33	−1.71	−1.60	−1.85	−1.47	−1.12
		(0.54)	(1.12)	(0.97)	(0.92)	(1.29)	(0.89)	(0.78)	(1.09)	(1.75)	(1.53)	(1.27)	(1.31)
	Yes	0.36	0.66	0.71	0.90	0.83	0.64	0.51	0.64	0.29	0.43	0.61	0.92
		(0.41)	(0.84)	(0.77)	(0.54)	(1.04)	(0.60)	(0.42)	(0.49)	(0.83)	(0.67)	(0.63)	(1.03)
Land rent baseline (€/m^2/y)		49.73	94.59	104.09	126.23	73.45	85.55	100.08	114.56	176.38	114.34	80.06	63.85
		(25.29)	(56.73)	(93.81)	(57.47)	(43.00)	(52.39)	(73.99)	(90.54)	(131.70)	(85.43)	(66.05)	(63.93)
Landlord loss (€/m^2/y)													
Counterfactual 1: Open City													
Urban green conversion	No	3.27	6.30	7.49	8.22	5.21	5.53	6.54	7.90	11.24	8.69	5.23	3.89
		(2.95)	(6.21)	(9.40)	(7.31)	(5.86)	(4.79)	(8.09)	(8.80)	(16.55)	(10.29)	(7.55)	(9.75)
	Yes	2.97	5.35	6.79	8.09	4.01	5.02	6.22	7.83	11.24	8.69	5.23	3.89
		(2.69)	(5.20)	(8.57)	(8.41)	(4.43)	(4.35)	(7.41)	(9.58)	(16.55)	(10.29)	(7.55)	(9.75)
Counterfactual 2: Closed City													
Urban green conversion	No	0.05	0.08	0.04	0.01	0.01	0.03	0.06	0.09	0.08	1.27	−0.06	−0.45
		(0.08)	(0.17)	(0.21)	(0.43)	(0.12)	(0.18)	(0.16)	(0.43)	(14.95)	(7.05)	(3.34)	(5.14)
	Yes	4.69	9.04	10.51	11.67	8.43	7.87	8.83	10.73	15.47	11.71	7.49	5.92
		(5.15)	(10.02)	(13.61)	(10.99)	(10.58)	(7.24)	(8.83)	(10.73)	(19.39)	(13.30)	(10.42)	(13.85)

Note: Here, we report only the case with heterogeneity; for non-heterogeneity case (like in counterfactual exercises 3 and 4), the results are similar. The loss for landlord is accounted as the difference between new land rent equilibrium and land rent in baseline model 0. The standard deviations are reported in the parenthesis.

increase in compensating-variation wages (negative willingness to pay), or a subsidy, of €770 per year for the bottom city income quartile and €2,090 per year for the top quartile. This increase represents up to 9.1% and 9.6% of the baseline net incomes. These subsidies also increase with city population. One can check that larger cities require proportionally higher subsidies, which result from the higher losses incurred by the residents in larger cities. The subsidy is not monotonic with distance to the city center: it first increases from €1,600 to €1,850 per year when one moves from the first to the second distance quartiles and then drops to €1,120 for the last distance quartile. This pattern reflects the geographical distribution of the share of green urban areas. It can finally be seen that landlords are not substantially harmed by the reduction in green urban areas when cities are closed and land is not converted (ninth row).

Finally, suppose that the green urban areas are converted to residential land (fifth row), which increases the residential land supply and compensates residents for the lack of green area amenities. The negative changes in the compensating-variation wages indicate that residents are better off in this situation. Low-income cities would accept lower compensating-variation wages and would therefore pay a tax of €360 per year in the lowest city income quartiles and €900 per year in the highest. As shown in the table, this benefit is larger for peripheral residents. Finally, landlords are negatively affected by the additional supply of residential land (see tenth row). They are more impacted in the richest and the largest cities and at the most central locations.

3.4.3 Estimation and valuation with hyperbolic preferences

Similar to the case of Cobb-Douglas preference, we suppose the presence of observable heterogeneity χ in the preference for land plots or specific characteristics of locations as well as unobservable heterogeneity or measurement errors ϵ. The hyperbolic utility function becomes

$$U(s,a,z,\chi,\epsilon) = z - \frac{\theta}{2s} + a + \chi + \epsilon$$

Accordingly, we obtain the consumption $\widehat{s} = \theta/\left(y + a + \chi + \epsilon - u\right), \widehat{z} = \left(y - a - \chi - \epsilon + u\right)/2$ and land value $\Psi = \left(y + a + \chi + \epsilon - u\right)^2/(2\theta)$. We focus on the case of positive consumption so that $y - a - \chi - \epsilon + u > 0$.

We assume that the transport cost is given by $t = w \times \tau(r)$, where $\tau(r)$ is a function of distance to the CBD. For simplicity, we assume the quadratic form $\tau(r) = \tau_1 \times r + \tau_2 \times r^2$. Green area amenities are given by $a = \alpha x$, which can be written as a function of green urban areas g as $a = \alpha g/\beta$, where α and β are green amenity and land use intensity parameters, respectively. We can further standardize the amenity value $\alpha = 1$ and consider β as land use intensity to provide one unit for green amenities. We obtain the inverse of housing slot size in equilibrium as given below:

$$\frac{1}{\widehat{s}} = \frac{1}{\theta}\left(w - w \times \tau(r) + \frac{\alpha}{\beta}g - u + \chi + \epsilon\right) \tag{3.21}$$

3.4.3.1 Estimation

We run the following regression:

$$\frac{1}{s_{ijc}} = \vartheta_0 + \vartheta_1 w_{ic} + \vartheta_2 \left(r_{ijc} \times w_{ic}\right) + \vartheta_3 \left(r_{ijc}^2 \times w_{ic}\right) + \vartheta_4 g_{ijc} + \vartheta_5 I_c + \vartheta_6 X_{jc} + \vartheta_7 A_{ijc} + \epsilon_{ijc}$$

where i is the location of the annuli within the city border, j is the city and c is the country where the city belongs to. Here, we use household income from NUTS2 regions as the proxy for the wage level; r_{ijc} is the distance of annuli i of city j to its CBD. I_c, X_{jc} and A_{ijc} are the vectors of country dummies, city geographic controls and annuli amenity controls variables, respectively. We introduce the nonlinearity in transportation cost by adding the quadratic term as explained in Cobb-Douglas preference sections.

Table 3.6 reports the regression results from both OLS and IV regressions with different levels of controls. We utilize a similar IV strategy as in the Cobb-Douglas case and the first stage regression, as well as the Wu-Hausman test for endogeneity, as reported in Appendix A. Columns (1) to (4) display OLS estimates without instrument variables. In all columns, the coefficient estimates are consistent with our model predictions: residents use larger land plots for smaller city income, a larger distance between residences and the CBD and smaller green urban areas. The results are robust after controlling for country fixed effects, city geographical conditions, such as elevation, rainfall and temperature (see Columns (2) and (3)) and different types of amenities within annuli (see Column (4)). After controlling for all other amenities, the square of the distance to the CBD loses its significance, suggesting the linearity in transportation cost. Columns (5) to (8) display the same estimates after instrumenting the share of urban green areas with its historical value from the CORINE Land Cover 1990 database. The IV regression reports a similar magnitude of the effects of urban green areas on the residential slot size. The Wu-Hausman test coefficient (reported in Appendix A) is not significant at the 90% confidence level, which supports the alternative hypothesis of no endogeneity at the 90% confidence level and implies that endogeneity may not be a severe issue in our analysis. Both the OLS and IV results show significant coefficients for the share of green urban area amenities g for approximately $2.9 - 3.2$ before including the control and $2.4 - 2.7$ after including all other controls. This finding implies that, ceteris paribus, residents in annuli with no green urban areas use approximately $19 - 22\%$ more land than those residing in annuli with a 7% share of green urban area.[30] Those changes are slightly higher than those in the Cobb-Douglas preference (approximately 14%).

3.4.3.2 Welfare valuation

In this section, we examine the changes in city structure and land rent when we cancel 50% of green urban space within the city boundary. We proceed with similar steps as

[30] We compute $\frac{1}{s_{ijc}} - \frac{1}{s'_{ijc}} = \vartheta_4(g_{ijc} - g'_{ijc})$, with $g_{ijc} = 0.07$ and $g'_{ijc} = 0$ and on average s_{ijc} is at 0.95 ($100\ m^2/hab.$).

Table 3.6: Hyperbolic residents' land use regression results

	Dependent variable: Inverse of space							
	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Household Income	3.149***	9.908***	10.557***	9.145***	3.171***	9.842***	10.507***	9.096***
	(0.813)	(2.011)	(1.968)	(1.785)	(0.822)	(2.030)	(1.975)	(1.791)
Distance to CBD × Income	−10.202***	−9.302***	−9.716***	−5.323***	−10.200***	−9.263***	−9.693***	−5.302***
	(1.783)	(1.713)	(1.714)	(1.626)	(1.780)	(1.707)	(1.704)	(1.609)
Distance to CBD square × Income	4.344***	3.645***	3.718***	1.771	4.333***	3.635***	3.716***	1.767
	(1.162)	(1.346)	(1.414)	(1.288)	(1.142)	(1.333)	(1.400)	(1.264)
Share of Urban Green	3.190***	2.930***	2.749***	2.463***	2.930***	3.228***	3.034***	2.727***
	(0.701)	(0.405)	(0.414)	(0.384)	(0.735)	(0.467)	(0.461)	(0.435)
Country FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
City Geographical Controls	No	No	Yes	Yes	No	No	Yes	Yes
Annuli Amenity Controls	No	No	No	Yes	No	No	No	Yes
Sample	All	All	All	All	All	All	All	All
Observations	10,853	10,853	10,853	10,853	10,853	10,853	10,853	10,853
Adjusted R^2	0.131	0.455	0.474	0.550	0.118	0.452	0.471	0.548
Residual Std. Error	0.946	0.749	0.736	0.680	0.953	0.752	0.738	0.682
df	10,848	10,827	10,823	10,819	10,848	10,827	10,823	10,819
F Statistic	33.38***	24.61***	23.96***	36.54***	37.55***	24.63***	23.34***	36.37***

Note: Here, we use the Household income taken from Regional Economic Accounts from Eurostat at NUTS2 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €100,000. The distance to CBD is measured in 10 kilometres. The inverse of space is calculated by divided number of inhabitants in each annuli with the areas within the annuli minus the areas using as urban green (in 100 meters). We exclude Cyrus and Luxembourg as the Eurostat population grid database does not cover Cyrus and the household income data at NUTS2 of Eurostat does not cover Luxembourg. United Kingdom and Finland are also excluded as they are not covered by Corine Land Cover 1990. City boundary is chosen at 20% cut-off point. For city control, we take into account the elevation, average rain fall, average temperature in Jan 01 and average temperature in July 01 for period 1995-2010. City amenity controls include the share of industrial, sport and leisure land use as well as the share of forest and agriculture land within 100 meters buffer from residential area. Standard errors are clustered at city level. The row "df" reports the degree of freedom. Significance levels are denoted by * for $p<0.1$, ** for $p<0.05$ and *** for $p<0.01$.

in the Cobb-Douglas preference. The first results of the baseline model with the hyperbolic preference are displayed in Table 3.7, and the main results for the counterfactual exercises are displayed in Table 3.8. On average, individuals have €15,220 as disposable income and spend €8,120 for housing expenses, which is slightly more expensive than the case of the Cobb-Douglas preference and that in reality. The average city size, average residential space and average population are similar to the case of the Cobb-Douglas preference. The only difference is in the computed land rent. The land rent per square meter is 143 €/m^2/year on average for all cities. These values are higher than in the case of Cobb-Douglas utility, which exhibits a higher income elasticity of demand residential lands in the hyperbolic preference compared to those in the Cobb-Douglas preference.

Table 3.7: Hyperbolic counterfactual analysis: open and closed cities

	Cities number	Composite Goods (Z) (€1000)	Housing Rent ($R \times s$) (€1000)	Income (W) (€1000)	Residential Area (km^2)	Green Area (GA) (km^2)
With heterogeneity	264	5.22	8.12	15.22	25.29	4.52
		(6.86)	(4.60)	(5.87)	(48.91)	(9.87)
Without heterogeneity	264	5.01	8.32	15.22	26.55	4.52
		(6.01)	(3.64)	(5.87)	(48.91)	(9.87)

Note: The standard deviation is reported in the parenthesis. Household income is taken from Eurostat at NUTS2 level and is measured on purchasing power standard (PPS) at 1000€.

Examine again the counterfactual exercise 1, where one eliminates 50% of the green areas in every annulus of open cities. On average, city residents increase their residential land use by approximately 10%, and the city population falls by approximately 9.5%. Those results are congruent under both hyperbolic and Cobb-Douglas preferences with a slightly higher magnitude in the hyperbolic preference. Nevertheless, the land values decline at higher magnitude under the hyperbolic preference, which is at 9% (from 143 to 129 € per m^2 per year). This situation results in a larger loss in the housing market of roughly €322 million for an average city. First, it is because of the higher sensitivity in the bid rent in the case of the hyperbolic; we have a higher magnitude for the change in residential land uses and larger falls in land price (9% instead of 6% in Cobb-Douglas). Second, in the case of the hyperbolic preference, the consumption of nonhousing goods also changed under new equilibrium, as shown in subsection 3.3.2, leading to a smaller budget for housing consumption under the new equilibrium. In the case of land conversion into residential land use[31], the residential land rents rise slightly to 131 € per m^2 per year because the new land is supplied at more central locations with higher land values. We compute that, compared to the baseline model, the housing market increases its total value by approximately €22 million per year and

[31]In an average city, there is a new land supply of 2.26 km^2 (half of 4.52 km^2) on top of the baseline residential land supply of 25.29 km^2, an increase of 8.9%.

city when we take into account converted areas. In contrast, Cobb-Douglas preference results in an increase in the total value of the housing market by €54 million.

In the counterfactual exercise 2 with closed cities, the average utility shrinks from 1 to 0.84 with no conversion policy. Total residential land rents decline by a smaller amount of €19 million per year. This reduction in utility combined with the change in land rent requires an increase in the baseline annual net income of €970 (6.4%) per year per household, which is the WTP measure of an average household. This measure is slightly lower than that of the case of Cobb-Douglas at 9.4%. Suppose now that green urban areas are converted into residential land. Then, the residential land supply increases, land rents decrease, and residents can use more land to compensate for the loss of green area amenities. Specifically, the land rents drop by 17% to 118 €/m^2/year, and the total loss in the housing market rises to €236 million per year. Nevertheless, city residents enjoy both lower land rent and larger residential land plots, which enhance their average utility level. They reach a slightly higher average utility level (jump by 8%) and need a smaller compensating-variation wage of €14,730 per year to maintain their level of utility, which is equivalent to an income reduction of €490 per year, or equivalent to 3.2% of their annual income. Those results are also in line with the Cobb-Douglas preference, which also found that the net effect of land use planning is negative once we allow for land conversion into residential land.

To summarize, the effects of population reduction under open cities, the measures of compensating wage variations, and the corresponding WTP for an average household are robust under two preference specifications. The land prices computed are costlier under the hyperbolic preference than those under the Cobb-Douglas preference.

Table 3.9 summarizes the results across cities with different sizes and income levels, as well as at different distances to the CBD within-cities for the hyperbolic preference. In the case of open city systems, the decline in green urban area amenities significantly harms residents who partly leave the city, which is robust in both utility functions. If urban green areas are not converted to residential land, the landlords lose €12 and €16 per m^2 and year in cities belonging to the bottom- and upper-income quartiles, respectively (see the seventh row). Furthermore, they lose €12 and €17 per m^2 and year in cities belonging to the bottom and upper population quartiles, respectively. The pattern is similar to the case of Cobb-Douglas. However, the absolute increase is higher as land demand is more sensitive in the hyperbolic preference. Landlords also lose €21 per m^2 and year in the central city quartile but only €9 per m^2 and year in the city periphery quartile, indicating that land rents decrease with distance from the CBD. This pattern remains approximately the same if urban green areas are converted to residential land (eighth row), but with much smaller magnitude.

In the case of closed cities, residents require an increase in compensating-variation wages (negative willingness to pay), or a subsidy, of €930 per year for the bottom city income quartile and €980 per year for the top quartile to have the same level of utility as

Table 3.8: City structure with Hyperbolic preference under closed and open scenarios $g = 0.5 \times g_0$

		Cities (number)	s ($100m2$/hab)	Land Rent (€/m^2/y)	Green value (€/m^2/y)	Pop. (mil.)	$U/\bar{U}^b$	Comp. W (€1000)	Total loss in housing market (€ *mil.*)	Total loss in wage comp. (€ *mil.*)	Total loss in population $1,000hab.$
Baseline 0 (with heterogeneity)		264	0.96	143.14	172.33	0.31	1.00				
			(0.45)	(200.59)	(237.25)	(0.68)	(1.22)				
$g = 0.5 \times g_0$											
Counterfactual 1: Open City											
Urban green conversion	No	264	1.06	129.08	150.71	0.28	1.00		322.08	0	21.55
			(0.53)	(189.71)	(219.11)	(0.64)	(1.22)		(835.63)		(47.88)
	Yes	264	1.05	131.74	150.71	0.31	1.00		22.99	0	−6.60
			(0.52)	(193.30)	(219.11)	(0.70)	(1.22)		(479.24)		(34.20)
Counterfactual 2: Closed City											
Urban green conversion	No	264	0.96	142.38	161.94	0.31	0.84	16.19	19.45	315.31	0
			(0.45)	(200.15)	(230.77)	(0.68)	(1.22)	(5.93)	(72.94)	(845.57)	
	Yes	264	1.05	118.70	136.14	0.31	1.08	14.73	236.78	−202.02	0
			(0.48)	(161.82)	(182.87)	(0.68)	(1.13)	(6.07)	(850.99)	(1130.31)	

Note: Household income is taken from Eurostat at NUTS2 level and is measured on purchasing power standard (PPS) at 1000€. Population is calculated in million inhabitants. The standard deviation is reported in the parenthesis.

Table 3.9: Quantiles analysis: between- and within-city statistics (Hyperbolic preference)

		Between cities								Within cities			
		City Income Quartiles				City Size Quartiles				Distance to CBD Quartiles			
		0-0.25	0.25-0.50	0.50-0.75	0.75-1	0-0.25	0.25-0.50	0.50-0.75	0.75-1	0-0.25	0.25-0.50	0.50-0.75	0.75-1
WTP (€1000/hab/y)													
Urban green conversion	No	−0.93	−0.99	−0.99	−0.98	−1.05	−0.94	−0.89	−0.99	−1.06	−1.19	−0.95	−0.72
		(0.53)	(0.91)	(0.58)	(0.44)	(0.96)	(0.48)	(0.41)	(0.57)	(1.04)	(0.85)	(0.73)	(0.73)
	Yes	0.35	0.84	0.47	0.31	0.60	0.28	0.47	0.61	0.71	0.41	0.23	0.48
		(0.84)	(1.84)	(1.51)	(0.97)	(1.95)	(0.83)	(1.15)	(1.27)	(2.39)	(1.53)	(1.08)	(2.20)
Landlord loss (€/m^2/y)													
Counterfactual 1: Open City													
Urban green conversion	No	12.36	16.57	14.25	13.19	12.95	11.78	14.37	17.13	21.23	18.75	12.21	9.33
		(9.19)	(16.71)	(16.12)	(14.49)	(8.09)	(9.64)	(14.42)	(17.52)	(29.09)	(21.70)	(16.19)	(21.23)
	Yes	10.60	11.50	11.09	12.50	7.09	9.45	12.83	16.22	21.23	18.75	12.21	9.33
		(7.17)	(19.21)	(13.90)	(18.36)	(18.59)	(8.82)	(11.14)	(18.92)	(29.09)	(21.70)	(16.19)	(21.23)
Counterfactual 2: Closed City													
Urban green conversion	No	0.67	1.28	0.73	0.38	0.80	0.61	0.89	0.75	2.99	4.10	0.42	−1.14
		(1.12)	(2.61)	(1.56)	(1.02)	(2.67)	(1.03)	(1.51)	(1.23)	(26.17)	(15.22)	(7.37)	(11.38)
	Yes	17.42	31.53	28.52	20.40	19.14	16.29	26.59	35.73	30.84	29.92	21.85	19.69
		(22.00)	(44.78)	(57.64)	(50.00)	(35.36)	(23.67)	(45.85)	(65.69)	(53.56)	(53.71)	(47.24)	(56.86)

Note: Here, we report only the case with heterogeneity; for non-heterogeneity case (like in counterfactual exercises 3 and 4), the results are similar. The loss for landlord is accounted as the difference between new land rent equilibrium and land rent in baseline model 0. The standard deviations are reported in the parenthesis.

before. There is a much smaller difference in WTP between different income and population quantiles in the hyperbolic preference compared to those in the Cobb-Douglas preference. The reason is because the difference in the measures of the general equilibrium willingness to pay in the case of Cobb-Douglas is proportional to income level, whereas is not proportional in the case of the hyperbolic preference (for details, see Appendix B). The subsidy in the case of the hyperbolic preference is not monotonic with distance to the city center, which is in line with the Cobb-Douglas preference; it first increases from €1,060 to €1,190 per year when one moves from the first to the second distance quartiles and then drops to €720 for the last distance quartile. This pattern reflects the geographical distribution of the share of green urban areas. It can finally be seen that landlords are not substantially harmed by the reduction in green urban areas when cities are closed, and the land is not converted (ninth row), which is robust under both preferences.

3.5 Conclusion

This paper studies the effects of land-intensive public green areas on city structures under competitive land markets. We employ monocentric urban land market frameworks with two different classes of utility functions—Cobb-Douglas and hyperbolic—to formalize the effects in residential land uses and households' willingness to pay (WTP) from a substantial change in green urban areas. The model emphasizes the importance of incorporating the adjustments in housing markets under a nonmarginal change in urban green areas in measuring households' WTP.

We first compute the equilibrium in residential land uses in the theoretical model and utilize the geographical land covers combined with actual population density to estimate the model parameters. The level of green areas in the past is used to instrument for the current level to control for the potential endogeneity in green provision. To quantify the value of green urban areas, we present a set of counterfactual exercises, where half of the green urban areas are removed. In case of the Cobb-Douglas preference, we estimate that, on average, open cities lose more than 6.5% of their population and that landlords lose €147 million in each city and year if the green urban areas are not converted into residential land. Such a loss in population is robust for both the hyperbolic and Cobb-Douglas preferences, while the impact for the land market is higher in the hyperbolic case. If the erased green areas are converted into new residential land, the increase in total residential lands is sufficient to compensate locals with additional residential space and to attract new city dwellers. Compared to our baseline model, for the Cobb-Douglas preference, the housing market increases its total value by nearly €55 million per year and city.

In closed cities where the green urban areas are not converted, the willingness to pay to avoid such a policy is measured at €1,430 per person and year. The estimated

results for the hyperbolic preference are similar at €1,110 per person per year. These amounts are significantly large at 6% of annual household income for the hyperbolic preference and up to 9% in case of the Cobb-Douglas preference. However, when we allow for the land conversion, the net costs of the green areas are up to 4.2% of annual household income for both preferences, implying that the opportunity costs of urban green areas are also substantial.

While accounting for the general equilibrium effects of the land market equilibrium, our approach has obvious limitations. The first limitation is its reliance on the monocentric urban model. This limitation can be easily changed by including more than one business center in theoretical model[32]; however, in the empirical parts, a polycentric city analysis requires more intensive data on the location of job centers for every city. As most of our city samples are monocentric cities[33], we think that monocentric urban areas are a reasonable simplification with tractable results. The second limitation is the heterogeneity in households' preference for green amenity. We here consider only one class of households; therefore, our study is valid for a representative household. The effects of green amenity change in urban areas are calculated based on the long-run change in the land market. In the context of a durable housing structure, the adjustments of residential land use might take a considerable time. The dynamic effects in the housing market relating to the change in land-intensive amenity change are open for future research.

[32]One approach could be as in Lucas and Rossi-Hansberg (2002).

[33]The OECD study on polycentric metropolitan cities only reports 18 European cities with two urban centers and only 6 with more than two urban cores (Barcelona, Paris, Lyon, Amsterdam, Stockholm and London) among our sample of more than 300 cities.

Appendix A: First stage regression and Wu-Hausmann test for IV regressions

In this section, we report the first stage regression and the Wu-Hausman test for IV regression. We use the historical level of urban green spaces (land use code 141) in Corine Land Cover in 1990 as our instrument variable. To our knowledge, Corin Land Cover (CLC) was the first systematized the land use over whole Europe, and its earliest version was in 1990. However, there are two issues with CLC 1990. First, CLC 1990 did not cover UK, Sweden and Finland as those three countries only appeared in later version of Corine Land Cover in 2000 onward. Therefore, we need to drop the city samples which belong to those three above countries. Second, as CLC cover not only urban area but also the rural and all lands in Europe. Hence, its resolution is much less precise than GMES Urban Atlas that covers only urban areas. There are also discrepancies in these two databases. To decrease the discrepancies, we use the land cover in CLC 2006 and GMES Urban Atlas 2006 and correct for the discrepancies between these two sets. We assume that the changes between Corine 2006 and Corine 1990 is the evolution of urban green, while the difference between Corine 2006 and GMES 2006 are just discrepancies in measurement. We adjust the Corine 1990 with these measurement errors before using it in the first stage regression.

Cobb-Douglass preference

As showed in Table B1, the coefficient between adjusted urban green in Corine 1990 is a very good predictors for the current level of urban green. The slope is highly significant and is around 0.8. R^2 is at $0.76 - 0.78$. The Wu-Hausman test coefficient for endogeneity ($\hat{v}_{error}$) is not significant at 90% of confidence level, meaning that we can confirm the alternative hypothesis of no endogeneity at 90% confidence level, which further implies that endogeneity may not be a critical issue in this analysis.

Table B1: Cobb-Douglas - First stage regression and Wu-Hausman test for endogeneity

First stage regression				
	Dependent variable:			
	Share of green area in GMES 2006			
	(1)	(2)	(3)	(4)
Adjusted Share of green area in Corine 1990	0.802***	0.794***	0.792***	0.791***
	(0.027)	(0.027)	(0.027)	(0.027)
Adjusted R^2	0.760	0.782	0.784	0.785
Wu-Hausman test for endogeneity				
	Dependent variable: Ln Residential Space			
	(1)	(2)	(3)	(4)
$\hat{v}_{error}$	-0.421	0.603	0.581	0.429
	(0.885)	(0.472)	(0.450)	(0.429)
Country FE	No	Yes	Yes	Yes
City Geographical Controls	No	No	Yes	Yes
Annuli Amenity Controls	No	No	No	Yes
Sample	All	All	All	All
Observations	10,853	10,853	10,853	10,853

Note: We use the following procedure to test for endogeneity. First Stage: $g_{ijc} = \alpha g_{Corine90} + \vartheta' Z + v_{ijc}$. Second Stage: $\ln s_{ijc} = \vartheta Z + \vartheta_4 g_{ijc} + \vartheta_{err} \hat{v}_{ijc} + \varepsilon_{ijc}$ where Z is the vector $(\ 1 \quad w \quad dist \quad I_c \quad X_{jc} \quad A_{ijc}\ )$ and $\hat{v}_{ijc}$ is the residuals from first stage regression. For city control, we take into account the elevation, average rain fall, average temperature in Jan 01 and average temperature in July 01 for period 1995-2010. The observations are all annuli of all cities covered by both GMES and Corine Land Cover 1990. Other variables are those from original regression (GDP per capita at purchasing power standard and distance to CBD). Regression (1) to (4) are corresponding to IV regression (5) to (8) in Table 5 in the main text respectively. The row "df" reports the degree of freedom. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01.

Hyperbolic preference

As showed in Table B2, the coefficient between adjusted urban green in Corine 1990 is a very good predictors for the current level of urban green, which is similar to the Cobb-Douglas preference. The Wu-Hausman test coefficient for endogeneity ($\hat{v}_{error}$) is not significant at 90% of confidence level, meaning that we can confirm the alternative hypothesis of no endogeneity at 90% confidence level, which further implies that endogeneity may not be a critical issue in this analysis.

Table B2: Hyperbolic - First stage regression and Wu-Hausman test for endogeneity

First stage regression				
	Dependent variable:			
	Share of green area in GMES 2006			
	(1)	(2)	(3)	(4)
Adjusted Share of green in 1990	0.804***	0.796***	0.793***	0.792***
	(0.027)	(0.027)	(0.027)	(0.027)
Adjusted R^2	0.758	0.781	0.783	0.784
Wu-Hausman test for endogeneity				
	Dependent variable: Inverse of Space			
	(1)	(2)	(3)	(4)
$\hat{v}_{error}$	2.007	−0.626	−0.598	−0.526
	(1.904)	(0.844)	(0.776)	(0.720)
Country FE	No	Yes	Yes	Yes
City Geographical Controls	No	No	Yes	Yes
Annuli Amenity Controls	No	No	No	Yes
Sample	All	All	All	All
Observations	10,853	10,853	10,853	10,853

Note: We using the following procedure to test for endogeneity. First Stage: $g_{ijc} = \alpha g_{Corine90} + \vartheta' Z + v_{ijc}$; Second Stage: $\frac{1}{s_{ijc}} = \vartheta Z + \vartheta_4 g_{ijc} + \vartheta_{err}\hat{v}_{ijc} + \varepsilon_{ijc}$where Z is the vector ($1 \quad w \quad dist \quad I_c \quad X_{jc} \quad A_{ijc}$) and $\hat{v}_{ijc}$ is the residuals from first stage regression. For city control, we take into account the elevation, average rain fall, average temperature in Jan 01 and average temperature in July 01 for period 1995-2010. The observations are all annulus from all cities covered by both GMES and Corine Land Cover 1990. Other variables are those from original regression (GDP per capita at purchasing power standard and distance to CBD). Regression (1) to (4) are corresponding to IV regression (5) to (8) in Table 5 in the main text respectively. The row "df" reports the degree of freedom. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01.

Appendix B: Counterfactural analysis

B1: Willingness to pay computation

Cobb Douglas preference

One notes that within city boundary, we have $R = \widehat{\psi}$, and replace it into the indirect utility in equation (3.13), we obtain the general willingness to pay to keep the same level of utility by using the indirect utility equation $V(y - WTP, a^1, R^1) = V(y, a^0, R^0)$ in case of closed cities. We have:

$$WTP = y\left(1 - e^{\frac{\varphi}{1-\varphi}\left(a^0 - a^1\right)}\left(\frac{R^0}{R^1}\right)^{-\frac{\phi}{1-\varphi}}\right). \tag{3.22}$$

Using the equilibrium bid rent function in (3.14), we can recompute the general equilibrium willingness to pay as:

$$WTP = y\left(1 - \left(\frac{u^e}{u^1}\right)^{\frac{1}{1-\varphi}}\right) \tag{3.23}$$

where the new equilibrium utility u^1 is determined by the population constraint condition (as we are in closed cities), and it depends on whether the new policy involved the conversion of urban green land into residential land. Suppose, the new green amenity level a^1 required a smaller fraction of land devoted for urban green $g^1 < g^0$, and the unused land $(g^0 - g^1)\ell$ gets converted into residential land use, we can compute the new equilibrium utility as below:

$$u^1_{\text{conversion}} = u^e \left(\frac{\int_0^{b^e} \frac{1-g^1}{\left(y^{-(1-\phi-\varphi)} e^{-\varphi a^1}\right)^{\frac{1}{\phi}}} 2\pi r \mathrm{d}r}{\int_0^{b^e} \frac{1-g^0}{\left(y^{-(1-\phi-\varphi)} e^{-\varphi a^0}\right)^{\frac{1}{\phi}}} 2\pi r \mathrm{d}r} \right)^{\phi}. \tag{3.24}$$

And if there is no conversion policy, the new utility equilibrium is

$$u^1_{\text{no conversion}} = u^e \left(\frac{\int_0^{b^e} \frac{1-g^0}{\left(y^{-(1-\phi-\varphi)} e^{-\varphi a^1}\right)^{\frac{1}{\phi}}} 2\pi r \mathrm{d}r}{\int_0^{b^e} \frac{1-g^0}{\left(y^{-(1-\phi-\varphi)} e^{-\varphi a^0}\right)^{\frac{1}{\phi}}} 2\pi r \mathrm{d}r} \right)^{\phi}. \tag{3.25}$$

Hyperbolic preference

Using similar approach as in above section, we use the equilibrium land rent $R = \widehat{\psi}$, and replace them into the indirect utility in equation (3.18) to obtain the willingness to pay as below:

$$WTP = \left(a^1 - a^0\right) - \sqrt{2\theta}\left(\sqrt{R^1} - \sqrt{R^0}\right). \tag{3.26}$$

Using the equilibrium bid rent function in (3.19), we can recompute the general equilibrium willingness to pay as:

$$WTP = u^1 - u^e \tag{3.27}$$

where the new equilibrium utility u^1 is determined by the population constraint condition (as we are in closed cities), and by land conversion policy. We can compute the new equilibrium utility as below:

$$u^1_{\text{conversion}} = u^e \left(\frac{\int_0^{b^e} (1-g^0) 2\pi r \mathrm{d}r}{\int_0^{b^e} (1-g^1) 2\pi r \mathrm{d}r} \right) + \left(\frac{\int_0^{b^e} (1-g^1)(y+a^1) 2\pi r \mathrm{d}r - \int_0^{b^e} (1-g^0)(y+a^0) 2\pi r \mathrm{d}r}{\int_0^{b^e} (1-g^1) 2\pi r \mathrm{d}r} \right). \tag{3.28}$$

And if there is no conversion policy, the new utility equilibrium is

$$u^1_{\text{no conversion}} = u^e + \left(\frac{\int_0^{b^e} (1-g^0)(a^1 - a^0) 2\pi r \mathrm{d}r}{\int_0^{b^e} (1-g^1) 2\pi r \mathrm{d}r} \right). \tag{3.29}$$

B2: Parameters recovered

Cobb-Douglas preference

We recover the model parameters from the estimated coefficient of residents' land use using the values of ϑ_0, ϑ_1, ϑ_2, ϑ_3 and ϑ_4 from Column (8) in Table 5. Country utility levels are recovered from the parameters ϑ_{5c} and the constant term ϑ_0.[34]. From the theoretical model, we recover the residential space, composite goods and the residential land rent as

$$\begin{aligned}
\hat{s}(w,r,g,u,\varepsilon) &= (we^{-\hat{\tau}_1 r-\hat{\tau}_2 r^2})^{\hat{\vartheta}_1}\ u\ e^{g\hat{\vartheta}_4+\hat{\vartheta}_6 X_{jc}+\hat{\vartheta}_7 A_{ijc}+\hat{\varepsilon}_{ijc}} \\
\hat{z}(w,r) &= we^{-\hat{\tau}_1 r-\hat{\tau}_2 r^2}\left(\frac{-\hat{\vartheta}_1}{1-\hat{\vartheta}_1}\right) \\
\hat{R}(w,r,g,u,\varepsilon) &= \frac{we^{-\hat{\tau}_1 r-\hat{\tau}_2 r^2}-\hat{z}(w,r)}{\hat{s}(w,r,g,u,X,\epsilon)}
\end{aligned}$$

where we use $\hat{\tau}_1 = \hat{\vartheta_2}/\hat{\vartheta}_1$, $\hat{\tau}_2 = \hat{\vartheta_3}/\hat{\vartheta}_1$, and $(1-\hat{\varphi}-\hat{\phi})/(1-\hat{\varphi}) = -\hat{\vartheta_1}/(1-\hat{\vartheta}_1)$.

Hyperbolic preference

Similarly, we have the following functions for residential space s, composite good z and the residential land rent R for the Hyperbolic preference.

$$\begin{aligned}
\hat{s}(w,r,g,u,\varepsilon) &= \frac{\hat{\theta}}{w-w(\hat{\tau}_1 r+\hat{\tau}_2 r^2)+\frac{g}{\hat{\beta}}-u_c+\hat{\vartheta}_5 X_{jc}+\hat{\vartheta}_6 A_{ijc}+\hat{\varepsilon}_{ijc}} \\
\hat{z}(w,r,g,u,\varepsilon) &= \frac{w-w(\hat{\tau}_1 r+\hat{\tau}_2 r^2)-\frac{g}{\hat{\beta}}+u_c-\hat{\vartheta}_5 X_{jc}-\hat{\vartheta}_6 A_{ijc}-\hat{\varepsilon}_{ijc}}{2} \\
\hat{R}(w,r,g,u,\varepsilon) &= \frac{w-w(\hat{\tau}_1 r+\hat{\tau}_2 r^2)-\hat{z}(w,r,g,u,\varepsilon)}{\hat{s}(w,r,g,u,\varepsilon)}
\end{aligned}$$

where $\hat{\theta} = 1/\hat{\vartheta}_1$, $\hat{\tau}_1 = -\hat{\vartheta_2}/\hat{\vartheta}_1$, $\hat{\tau}_2 = -\hat{\vartheta_3}/\hat{\vartheta}_1$, and $\hat{\beta} = \hat{\vartheta}_1/\hat{\vartheta_4}$.

B3: Counterfactural exercises – detailed steps

We first define the baseline model and the counterfactual exercises. Both baseline and counterfactuals use the observed distance to the city center r_{ijc} and amenities X_{jc} and A_{jc}. The baseline model includes the observed city wage w_{jc}, the share of green

[34] As we drop Austria in the country dummies, we have $u_{at}^{1/\phi} = \vartheta_0$, and all other countries as $u_c^{1/\phi} = \vartheta_{5c} + \vartheta_0$

urban areas g_{ijc}, the estimated values of country utility $\widehat{u}_c$ and the unobserved heterogeneity or measurement error $\widehat{\varepsilon}_{ijc}$. Formally, we set the baseline model values to $\hat{s}^0_{ijc} = \hat{s}(w_{jc}, r_{ijc}, g_{ijc}, \widehat{u}_c, \widehat{\varepsilon}_{ijc})$, $\hat{z}^0_{ijc} = \hat{z}(w_{jc}, r_{ijc})$ and $\hat{R}^0_{ijc} = \hat{R}(w_{jc}, r_{ijc}, g_{ijc}, \widehat{u}_c, \widehat{\varepsilon}_{ijc})$.

We now investigate the impact of canceling 50% of the urban green areas in each annulus. In counterfactual exercise 1, we consider open cities where utility levels and unobserved heterogeneity are maintained at the estimated levels $\widehat{u}_c$ and $\widehat{\varepsilon}_{ijc}$. This consideration helps us discuss a long-term and unregulated perspective, where urban planners do not impose restrictions on workers' mobility within and between cities. We then remove half of the green urban areas by setting $g'_{ijc} = 0.5 \times g_{ijc}$. In both cases, we set $\hat{s}^1_{ijc} = \hat{s}(w_{jc}, r_{ijc}, g'_{ijc}, \widehat{u}_c, \widehat{\varepsilon}_{ijc})$ and $\hat{z}^1 = \hat{z}(w_{jc}, r_{ijc})$, while $\hat{R}^1 = \hat{R}(w_{jc}, r_{ijc}, g'_{ijc}, \widehat{u}_c, \widehat{\varepsilon}_{ijc})$. Residents' land use should increase ($\hat{s}^1_{ijc} > \hat{s}^0_{ijc}$) because residents require compensation for the reduction of green area amenities. If green urban areas are left with no use, the total available space remains constant and is given by $\sum_{ijc} \left(1 - g_{ijc}\right) \ell_{ijc}$, where ℓ_{ijc} is the land surface of annulus i in city j and country c. Since resident's land use increases, cities host fewer residents. If green urban areas are converted in residential land, city populations may grow if the new supply of land, $g'_{ijc}\ell_{ijc}$, is larger than the increase in residents' land demand from $\hat{s}^0_{ijc}$ to $\hat{s}^1_{ijc}$. More formally, population grows if $\sum_{ijc} \left(1 - g_{ijc}\right) \ell_{ijc}/s^0_{ijc} < \sum_{ijc} \left(1 - g'_{ijc}\right) \ell_{ijc}/s^1_{ijc}$.

In counterfactual exercise 2, we consider closed cities with exogenous city populations. The study of closed cities can be appropriate in evaluating policy changes that occur simultaneously in all cities, such as changes in EU policies.[35] Here, our aim is to discuss a midterm or regulated perspective, where urban planners are able to restrict workers' mobility between cities but allow residents' land use to change. We again remove half of the green urban areas ($g'_{ijc} = 0.5 \times g_{ijc}$). We set $\hat{s}^2_{ijc} = \hat{s}(w_{jc}, r_{ijc}, g'_{ijc}, u^2_{jc}, \widehat{\epsilon}_{ijc})$, $\hat{z}^2_{ijc} = \hat{z}(w_{jc}, r_{ijc})$ and $\hat{R_{ijc}}^2 = \hat{R}(w_{jc}, r_{ijc}, g'_{ijc}, u^2_{jc}, \widehat{\epsilon}_{ijc})$, where u^2_{jc} is the counterfactual city utility level. In the absence of the conversion of green urban areas to residential plots, we set the city utility level u^2_{jc} such that the city population spreads over the baseline residential area; that is, we impose that each u^2_{jc} solves the population identity $\sum_i (1 - g_{ijc}) l_{ijc}/\hat{s}^0_{ijc} = \sum_i (1 - g_{ijc}) l_{ijc}/\hat{s}^2_{ijc}$. In the case of land conversion, we set u^2_{jc} such that the city population spreads over the new residential land supply. Then, u^2_{jc} solves the population identity $\sum_i (1 - g_{ijc})\ell_{ijc}/\hat{s}^0_{ijc} = \sum_i (1 - g'_{ijc})\ell_{ijc}/\hat{s}^2_{ijc}$.[36] However, although utility levels are important concepts in welfare analysis, they are difficult to interpret quantitatively. Therefore, we also compute the *compensating variation wage* w^2_{jc} as the city wage that maintains the baseline utility level when we remove green urban areas,

[35] Cheshire and Shepard (2002) also use the closed city model to analyze the welfare effects of policy changes in the UK.

[36] This approach yields the counterfactual utility levels $\left(u^2_{jc}\right)^{1/\phi} = (u_c)^{1/\phi} \frac{\left(\sum_i (1-g_{ijc})\ell_{ijc}/s^2_{ijc}\right)}{\left(\sum_i (1-g_{ijc})\ell_{ijc}/s_{ijc}\right)}$ and $\left(u^2_{jc}\right)^{1/\phi} = (u_c)^{1/\phi} \frac{\left(\sum_i (1-g'_{ijc})\ell_{ijc}s^2_{ijc}\right)}{\left(\sum_i (1-g_{ijc})\ell_{ijc}/s_{ijc}\right)}$.

which is equivalent to setting the wage w_{jc}^2 such that the above population identities hold with $\hat{s}(w_{jc}^2, r_{ijc}, g'_{ijc}, \widehat{u}_{jc}, \widehat{\varepsilon}_{ijc})$. Under the Cobb-Douglas preferences, this assumption simplifies to the compensating variation wage $w_{jc}^2 = w_{jc}(u_{jc}^2/u_{jc}^0)^{\frac{1}{\vartheta_1\hat{\phi}}}$.

The above two counterfactual exercises hinge on the assumption that empirical model residuals ε_{ijc} reflect land heterogeneity that is unobserved to the econometricians but observed and used by residents in their land plot size choices. Such heterogeneity is reported in the counterfactual results. This assumption may be strong, as it imposes strong information on behalf of residents. Therefore, we take the opposite view and assume that the residuals ε_{ijc} consist of measurement errors that can be observed neither by the econometricians nor the residents. In that case, residents do not base their decisions on ε_{ijc}, and we set $\varepsilon_{ijc} = 0$ in counterfactual exercises 3 and 4.

Appendix C: Robustness check for Cobb-Douglas preference with NUTS3 incomes

In this section, we report the results of estimation of the structural model using GDP per capita at NUTS3 level as the proxy for city wage. The results are similar to those in Table 3.2 in the main text.

Table C1: Regression Results using GDP per capita at NUTS3 as proxy for city's wage level

	Dependent variable: Ln Residential Space							
	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-0.686^{***}	-0.899^{***}	-0.834^{*}	-0.230	-0.682^{***}	-0.880^{***}	-0.808^{*}	-0.213
	(0.119)	(0.256)	(0.467)	(0.425)	(0.119)	(0.262)	(0.473)	(0.430)
Ln Household Income	-0.210^{***}	-0.614^{***}	-0.608^{***}	-0.542^{***}	-0.211^{***}	-0.610^{***}	-0.604^{***}	-0.540^{***}
	(0.077)	(0.103)	(0.098)	(0.090)	(0.077)	(0.103)	(0.098)	(0.090)
Distance to CBD	1.596^{***}	1.482^{***}	1.514^{***}	1.004^{***}	1.591^{***}	1.475^{***}	1.509^{***}	0.998^{***}
	(0.239)	(0.220)	(0.212)	(0.219)	(0.240)	(0.221)	(0.213)	(0.220)
Distance to CBD square	-0.610^{***}	-0.515^{***}	-0.518^{***}	-0.322^{**}	-0.608^{***}	-0.512^{***}	-0.517^{***}	-0.320^{**}
	(0.134)	(0.146)	(0.149)	(0.142)	(0.133)	(0.146)	(0.148)	(0.140)
Share of Urban Green	-2.045^{***}	-1.773^{***}	-1.681^{***}	-1.487^{***}	-2.079^{***}	-1.948^{***}	-1.851^{***}	-1.611^{***}
	(0.386)	(0.245)	(0.241)	(0.203)	(0.413)	(0.274)	(0.262)	(0.228)
Country FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
City Geographical Controls	No	No	Yes	Yes	No	No	Yes	Yes
Annuli Amenity Controls	No	No	No	Yes	No	No	No	Yes
Sample	All	All	All	All	All	All	All	All
Observations	10,853	10,853	10,853	10,853	10,853	10,853	10,853	10,853
Adjusted R^2	0.227	0.513	0.525	0.607	0.218	0.510	0.522	0.604
Residual Std. Error	0.567	0.450	0.445	0.404	0.570	0.451	0.446	0.406
df	10,848	10,827	10,823	10,819	10,848	10,827	10,823	10,819
F Statistic	19.83^{***}	32.52^{***}	33.78^{***}	46.95^{***}	18.44^{***}	31.44^{***}	32.74^{***}	47.94^{***}

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €100,000. The distance to CBD is measured in 10 kilometres. The inverse of space is calculated by divided number of inhabitants in each annuli with the areas within the annuli minus the areas using as urban green (in 100 meters). We exclude Cyrus and Luxembourg as the Eurostat population grid database does not cover Cyrus and the household income data at NUTS2 of Eurostat does not cover Luxembourg. United Kingdom and Finland are also excluded as they are not covered by Corine Land Cover 1990. City boundary is chosen at 20% cut-off point. For city control, we take into account the elevation, average rain fall, average temperature in Jan 01 and average temperature in July 01 for period 1995-2010. City amenity controls include the share of industrial, sport and leisure land use as well as the share of forest and agriculture land within 100 meters buffer from residential area. Standard errors are clustered at city level. The row "df" reports the degree of freedom. Significance levels are denoted by * for $p<0.1$, ** for $p<0.05$ and *** for $p<0.01$.

Table C2: First stage regression and Wu-Hausman test for endogeneity

First stage regression				
	Dependent variable:			
	Share of green area in GMES 2006			
	(1)	(2)	(3)	(4)
Adjusted Share of green area in Corine 1990	0.797***	0.793***	0.790***	0.789***
	(0.028)	(0.027)	(0.027)	(0.027)
Adjusted R^2	0.759	0.781	0.784	0.784
Wu-Hausman test for endogeneity				
	Dependent variable: Ln Residential Space			
	(1)	(2)	(3)	(4)
$\hat{v}_{error}$	−0.330	0.413	0.406	0.241
	(0.911)	(0.455)	(0.438)	(0.407)
Country FE	No	Yes	Yes	Yes
City Geographical Controls	No	No	Yes	Yes
Annuli Amenity Controls	No	No	No	Yes
Sample	All	All	All	All
Observations	10,853	10,853	10,853	10,853

Note: We using the following procedure to test for endogeneity. First Stage: $g_{ijc} = \alpha g_{Corine90} + \vartheta' Z + v_{ijc}$; Second Stage: $\ln s_{ijc} = \vartheta Z + \vartheta_3 g_{ijc} + \vartheta_{err}\hat{v}_{ijc} + \varepsilon_{ijc}$where Z is the vector $(\ 1 \quad w \quad dist \quad I_c \quad X_{jc}\ )$ and $\hat{v}_{ijc}$ is the residuals from first stage regression. For city control, we take into account the elevation, average rain fall, average temperature in Jan 01 and average temperature in July 01 for period 1995-2010. The observations are all annulus from all cities covered by both GMES and Corine Land Cover 1990. Other variables are those from original regression (GDP per capita at purchasing power standard and distance to CBD). Regression (1) to (4) are corresponding to IV regression (5) to (8) in Table A1 respectively. The row "df" reports the degree of freedom. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01.

Table C3: Counterfactual analysis: open and closed cities

	Cities number	Composite Goods (Z) (€1000)	Housing Rent ($R \times s$) (€1000)	Income (W) (€1000)	Residential Area (km^2)	Green Area (GA) (km^2)
Spatial heterogeneity	264	6.39	11.84	26.55	25.29	4.52
		(3.01)	(5.58)	(13.51)	(48.91)	(9.87)
No spatial heterogeneity	264	5.03	6.35	11.76	26.55	4.52
		(2.99)	(5.55)	(13.51)	(48.91)	(9.87)

Note: Household income is taken from Eurostat at NUTS2 level and is measured on purchasing power standard (PPS) at 1000€. More details on PPS measure, please check Eurostat technical documents. The standard deviation is reported in the parenthesis.

Table C4: City structure under closed and open scenarios $g = 0.5 \times g_0$

		Cities (number)	s ($100m2$/hab)	Land Rent (€/m^2/y)	Green value (€/m^2/y)	Pop. (mil.)	$U/\bar{U}^b$	Comp. W (€1000)	Total loss in housing market (€ *mil.*)	Total loss in wage comp. (€ *mil.*)	Total loss in population $1,000hab.$
Baseline 0 (with heterogeneity)		264	0.95	161.88	177.25	0.31	1.00				
			(0.44)	(127.49)	(140.65)	(0.69)	(0.33)				
$g = 0.5 \times g_0$											
Counterfactual 1: Open City											
Urban green conversion	No	264	1.00	152.18	162.30	0.29	1.00		228.36	0	17.62
			(0.46)	(118.02)	(124.64)	(0.65)	(0.33)		(678.27)		(45.39)
	Yes	264	0.99	153.16	162.30	0.32	1.00		−131.58	0	−11.27
			(0.46)	(119.03)	(124.64)	(0.71)	(0.33)		(311.15)		(25.71)
Counterfactual 2: Closed City											
Urban green conversion	No	264	0.95	161.78	172.54	0.31	0.95	29.56	4.50	1,403.63	0
			(0.44)	(127.61)	(135.29)	(0.69)	(0.31)	(15.28)	(43.70)	(5,650.12)	
	Yes	264	1.03	145.86	154.38	0.31	1.04	24.58	9.65	−726.84	0
			(0.47)	(111.58)	(115.64)	(0.69)	(0.34)	(12.62)	(56.06)	(2,420.57)	
Baseline 0 (without heterogeneity)		264	0.89	163.66	175.17	0.28	1.00				
			(0.37)	(122.22)	(127.79)	(0.53)	(0.33)				
$g = 0.5 \times g_0$											
Counterfactual 3: Open City											
Urban Green Conversion	No	264	0.94	153.83	160.68	0.27	1.00		204.58	0	15.83
			(0.38)	(112.88)	(114.27)	(0.50)	(0.33)		(505.92)		(36.13)
	Yes	264	0.94	154.12	160.68	0.29	1.00		−112.14	0	−10.32
			(0.38)	(112.85)	(114.27)	(0.55)	(0.33)		(271.64)		(22.17)
Counterfactual 4: Closed City											
Urban Green Conversion	No	264	0.89	163.54	170.64	0.28	0.95	16.65	6.01	1,243.83	0
			(0.37)	(122.30)	(123.46)	(0.53)	(0.31)	(6.47)	(33.94)	(4,367.02)	
	Yes	264	0.97	147.30	153.74	0.28	1.04	14.61	7.64	−670.88	0
			(0.39)	(106.00)	(107.66)	(0.53)	(0.34)	(5.70)	(35.75)	(2,123.28)	

Note: Household income is taken from Eurostat at NUTS2 level and is measured on purchasing power standard (PPS) at 1000€. More details on PPS measure, please check Eurostat technical documents. Population is calculated in million inhabitants. Utility is measured relatively with baseline average. Basically, we measure $\frac{U_{jc}}{\bar{U}^{baseline}} = \frac{U_{jc}}{\frac{1}{N}\sum_{jc} U_{jc}^{baseline}} = \frac{u_{jc}^{1/\phi}}{\frac{1}{N}\sum_{jc}\left(u_c^{baseline}\right)^{1/\phi}}$ where u_{jc} is calculated as in above section. The standard deviation is reported in the parenthesis.

Table C5: Quantiles analysis: between- and within-city statistics

		Between cities								Within cities			
		City Income Quartiles				City Size Quartiles				Distance to CBD Quartiles			
		0-0.25	0.25-0.50	0.50-0.75	0.75-1	0-0.25	0.25-0.50	0.50-0.75	0.75-1	0-0.25	0.25-0.50	0.50-0.75	0.75-1
Baseline wage (€1000/hab/y)		13.20	20.28	27.89	44.83	19.50	25.79	26.67	34.24	30.65	30.36	30.45	30.22
		(2.71)	(2.16)	(2.02)	(12.69)	(6.98)	(14.34)	(11.62)	(15.44)	(15.16)	(15.00)	(15.06)	(15.01)
Increase in comp.var.wage(€1000/hab/y)													
Counterfactual 1: Open City													
Urban green conversion	No	0	0	0	0	0	0	0	0	0	0	0	0
	Yes	0	0	0	0	0	0	0	0	0	0	0	0
Counterfactual 2: Closed City													
Urban green conversion	No	1.38	2.15	3.08	5.42	2.48	2.86	2.71	3.96	3.54	4.15	3.25	2.50
		(1.79)	(1.32)	(1.69)	(2.82)	(2.66)	(2.34)	(1.74)	(2.87)	(4.24)	(3.97)	(3.08)	(2.50)
	Yes	−0.83	−1.59	−2.06	−3.40	−2.12	−1.96	−1.60	−2.21	−1.20	−1.66	−1.94	−2.59
		(1.09)	(1.43)	(1.63)	(2.61)	(2.74)	(1.99)	(1.25)	(1.73)	(2.47)	(2.03)	(1.99)	(2.96)
Land rent baseline (€/m^2/y)		80.16	146.52	178.45	242.40	124.73	152.69	167.74	202.36	349.81	201.97	133.21	102.17
		(37.40)	(95.91)	(140.07)	(147.18)	(86.63)	(110.20)	(135.50)	(156.72)	(307.07)	(157.57)	(118.50)	(113.42)
Landlord loss (€/m^2/y)													
Counterfactual 1: Open City													
Urban green conversion	No	4.10	8.25	10.66	15.80	8.14	8.79	9.65	12.22	19.43	13.57	7.66	5.69
		(4.68)	(7.16)	(11.76)	(15.32)	(11.90)	(8.66)	(12.08)	(12.11)	(29.44)	(13.57)	(11.01)	(16.61)
	Yes	3.58	6.89	9.99	14.44	5.88	7.79	9.12	12.11	19.43	13.57	7.66	5.69
		(4.40)	(6.96)	(10.29)	(13.97)	(6.95)	(7.49)	(11.23)	(13.52)	(29.44)	(13.57)	(11.01)	(16.61)
Counterfactual 2: Closed City													
Urban green conversion	No	0.12	0.14	0.10	0.05	0.00	0.03	0.15	0.23	−0.16	1.99	−0.12	−0.45
		(0.17)	(0.23)	(0.36)	(1.14)	(0.40)	(0.51)	(0.40)	(0.95)	(26.40)	(10.97)	(4.81)	(9.05)
	Yes	6.28	13.39	18.19	26.24	15.44	14.53	15.09	19.03	30.99	20.90	12.81	10.08
		(7.50)	(11.91)	(21.07)	(30.11)	(26.76)	(15.05)	(19.48)	(20.69)	(37.99)	(23.90)	(18.68)	(27.27)

Note: Here, we report only the case with heterogeneity; for non-heterogeneity case (like in counterfactual exercises 3 and 4), the results are similar. The loss for landlord is accounted as the difference between new land rent equilibrium and land rent in baseline model 0. The standard deviations are reported in the parenthesis.

Chapter 4

Geographical Stratification of Green Urban Areas

> "The more successfully a city mingles everyday diversity of uses and users in its everyday streets, the more successfully, casually (and economically) its people thereby enliven and support well-located parks that can thus give back grace and delight to their neighbourhoods instead of vacuity." **Jane Jacobs**, *The Death and Life of Great American Cities, 1961.*

4.1 Introduction

The future of human development is inextricably tied to the future of our urban areas. Throughout history, cities have been at the centre of change, civilization and attraction, ranging from the ancient Athens to 'Renaissance' Paris to 'financial centre' London. The world has changed gradually from mostly rural toward more urbanized areas, with more than half of all the world's citizens currently living in an urban area as of 2017 (UN DESA, 2018), and in Europe, urban dwellers account for approximately 74% of all its population (Eurostat, 2016). In other words, when we discuss the future of urban areas, the discussion relates to the lives of the nearly three-quarters of the European population and one-half of all human beings who have chosen to live in urban areas.[1] If this trend is irreversible, then the next question should concern how we can make such urbanization work. How can we improve urban quality of life and make city life sustainable? Undoubtedly, one of the main components that make cities attractive to their residents are their public park and garden system.[2]

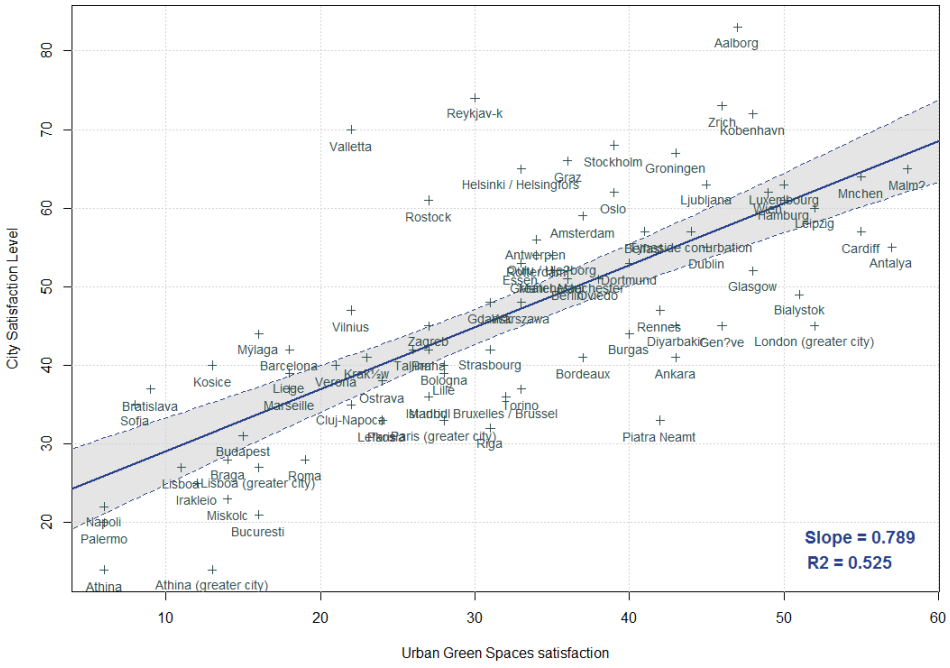
In the European context, green amenities are also a crucial factor in determining

[1] By 2030, this number is expected to increase to two-thirds of the world population.

[2] The benefits of urban green spaces have been well documented in many fields, such as sociology, psychology, economics, health, environment, etc. For a systematic review, see IFPRA (2013) and the EU DG-ENVI (2012) report.

the level of satisfaction with urban life. Since 2007, Eurostat has conducted a series of surveys in selected cities[3] in 28 Member States of the European Union to measure the perceptions of urban residents about their satisfaction with city life and other aspects such as housing, transportation, crime and green space access. Figure 1 depicts the city satisfaction level and the satisfaction in urban green systems from the European Urban Perception survey. The trend clearly shows that cities with well-maintained and efficiently accessed parks, as indicated by a high level of satisfaction with green amenities, are much more likely to have a higher level of life satisfaction. Well-located urban park systems play an essential role in establishing vibrant urban life, as Jane Jacobs (1961) noted. The value of green urban spaces in the everyday lives of urban citizens is, no doubt, a relevant indicator to measure the quality of life in a city.

Figure 1: Perceptions of Urban Green Areas and Life Satisfaction in the Europe Union



Note: Authors' own computation with the data from Eurostat Urban Perception Survey in 2012 for European cities. All measures are in percentage. Urban Green Space Satisfaction is the percentage of people strongly satisfied with current urban green space systems in their city. City Satisfaction Level is the percentage of people who strongly agreed that their city is satisfied to live.

In this paper, we study the provision of small urban green spaces in the context of EU cities. According to our dataset, 95% of urban green areas consists of small parks, which accounts for 50% of the total urban green surface in EU cities. Although small parks are valuable to residents for their close access, they are spread across the urban area and compete for the land allocated for residential purposes. To shed light on this

[3]Currently, 79 cities are used to implement the urban perception survey.

question, we first develop a theoretical framework that embeds two important features of urban green spaces: they are land intensive, and their sizes and access to them represent valuable amenities for residents. As most urban green areas are controlled by urban planning, we derive and discuss the optimal land allocation by a benevolent urban planner. The planner balances the space allocated to residents and urban green spaces. At the optimum, the number of urban green spaces is shown to be a non-monotone function of distance to the city centre. The sizes of urban green spaces and the distances to other urban green areas increase as one moves to the urban fringe. The model also allows us to compare cities with various productivities. It is shown that more productive cities include more numerous parks with smaller sizes and longer access distances. These properties are used to guide the empirical investigation on the 300 largest European cities by using the GMES Urban Atlas database (European Environmental Agency). The empirical analysis confirms the non-monotone relationship between the number of urban green spaces and the distance to the city centre. It also confirms that the distance between two parks increases as one moves toward the urban fringe. Finally, it corroborates that richer cities are associated with a denser network of urban green areas.

This paper contributes to the literature in public and urban economics. It adds to the public economics literature by investigating the provision of local public goods in the context of non-uniform population distribution and land-intensive public goods. Cremers et al. (1986), Fujita (1986), and Shakashita (1987) study the location of spaceless local public goods in cities with an exogenous uniform density distribution. Their main conclusion is that local public goods should be equidistant. In this paper, we consider the choices for residential space, park sizes and park locations. As a result, the residential density is not uniform, and the sizes and distances between parks change with their location relative to the city centre. Our paper shares a close resemblance with Berliant (2006), who considers the endogenous location of a local public good. However, as in the above contributions, the local public good in that study is a spaceless point with no land requirement. In contrast, this paper discusses the geographical allocation of small local public goods that require extensive land use and provide an amenity that decays in the distance travelled to them.

This paper also extends the numerous urban economics models that investigate the effect of amenities on urban land use. For instance, Fujita (1989) studies the optimal allocation of *neighbourhood* goods (localized and congestionable public goods) and their decentralization through Pigouvian taxes. As such goods are spaceless, they do not use land as an input and allow immediate access, meaning that their geographical dispersion cannot be studied. Lee and Fujita (1997) consider the optimal location of a green belt and its effect on urban land rents. Lee (1998) considers the optimal size of a central park in a city. These two works consider the land intensity of urban green areas but focus on the provision of a single area. This contrasts with the (European) reality

of a system of many green spaces and public parks, which is our emphasis. Our paper also relates to the urban economics literature on *open spaces*. Open spaces differ from green urban areas to the extent that they are generally not organized and maintained by planners (or other agents) for recreational purposes. Turner (2005) discusses the residents' choice to locate close to such open spaces in a linear city model with discrete locations. In that work, residents have inelastic land consumption, and open spaces are exogenous land plots. Caruso et al. (2007) study the emergence of sprawl patterns where households balance the benefits from proximity to other humans and to open spaces (see also Caruso et al. 2005). Walsh (2006) investigates the impacts of open space protection and growth control on city patterns in discrete neighbourhood settings. Wu and Plantinga (2003) introduce an exogenous open space in a continuous two-dimensional model to discuss leap-frogging patterns. In contrast to the latter contribution, the present paper studies the endogenous locations and sizes of many small urban green spaces in a continuous two-dimensional space.

Although the literature offers theoretical guidance about the provision of green urban areas, the examination of its empirical reality appears to be scant. To the best of our knowledge, this paper is one of the first contributions that studies the allocation of urban green space with a high-resolution geographical database such as that provided in the GMES Urban Atlas database (European Environmental Agency). This database requires the assembly of a large set of land plots and the construction of the data to be exploited in regression analyses. In a companion paper, Picard and Tran (2019) use the same database and similar data processing to study the trade-off between residential space and green urban areas. They highlight a non-linear relationship between the land share of green urban areas and the distance to the city centre. By contrast, this paper disentangles the effect of the latter on the size, number and dispersion of parks.

The structure of the paper is as follows. Section 2 presents and discusses the theoretical model. It first highlights the general principles according to which the city planner chooses the locations and sizes of green urban spaces. It then focuses on Cobb-Douglas preferences to provide the closed-form solutions that are used for guidance in the empirical research. Section 3 presents the empirical study of green urban stratification across the largest European cities. The model predictions on the green urban areas and population densities are then tested. Section 4 concludes the paper. The appendices include additional material on the mathematical analysis and empirical robustness checks.

4.2 The Model

4.2.1 Household preferences

We consider a monocentric circular city with a central business district surrounded first by residences and then by agricultural areas. For simplicity, agricultural areas generate a land rent of R_A and consist of fenced private properties that provide no amenity to urban dwellers.

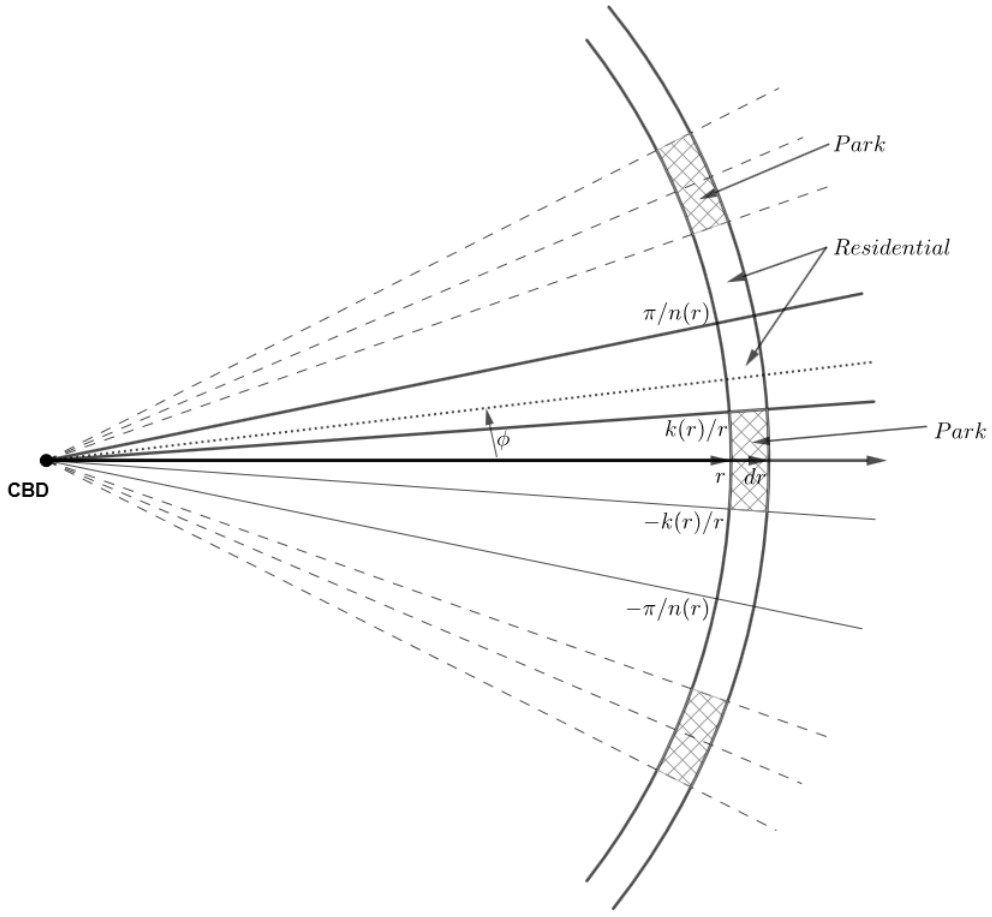
Households are homogeneous and consume an amount z of a composite good, use a land surface s for housing and enjoy a green amenity level a at their residential location. They are endowed with a utility function $U(z, s, a)$ that is increasing and quasi-concave with respect to z, s and a. Households earn the same wage w but differ according to their place of residence. Residence is measured in polar coordinates (r, ϕ), where r is the length of the ray to the centre and ϕ is the angle with the $x-$axis. The city extends to the border b, so that $(r, \phi) \in [0, b] \times [0, 2\pi]$. Workers commute on the ray from their residence to the business district and incur a commuting cost $t(r)$ with $t' > 0$.[4] We denote their net disposable income by $y(r) = w - t(r)$. We assume a positive net income on the whole city space: $y(r) > 0, r \in [0, b]$.

We focus on the green space amenity. Parks provide an amenity level that is increasing in their size k and decreasing in their distance to one's residence, d. Let $a(k, d)$ be the function representing this amenity ($a_k \equiv \partial a/\partial k > 0 > a_d \equiv \partial a/\partial d$). In addition to their land cost, parks incur a maintenance cost $C(k)$ that is an increasing and concave function of their size ($C' > 0$ and $C'' < 0$).

The analytical discussion of the allocation of residences and parks on a two-dimensional plane is a difficult task. In this paper, we simplify the analysis by distinguishing the commutes to work and the travels to parks on different space dimensions. Whereas commuting takes place on the ray passing through the business district, as is standard in the urban economic literature, traveling to a park is assumed to take place on the annulus (ring) on which a worker resides. Hence, residents residing at (r, ϕ) travel to the parks on the annulus of ray r and width dr or, more precisely, on the space $[r, r + \mathrm{d}r] \times [0, 2\pi]$. The amenity for a resident at (r, ϕ) is therefore given by $a[k(r), d(r, \phi)]$, where $d(r, \phi)$ is the distance travelled to the park on the annulus. Because of the symmetry of households and space, we can focus on parks that have the same size $k(r)$ and same distance between one another. Furthermore, as shown in Figure 2, symmetry allows us to concentrate on the park and residents on the *annulus strip* $[r, r + dr] \times [0, \pi/n(r)]$. There are $2n(r)$ such annulus strips at distance from the centre r. In this annulus strip, a resident at (r, ϕ) travels a distance to the closest park equal to $d(r, \phi) = r\phi$.

[4] The transportation cost can either be counted as a real cost or a time cost (in this case, $t(r)$ is a fraction of the wage).

Figure 2: Construction of Urban Green Areas on an Annulus



Note: The figure depicts the locations of residential and urban green areas in the theoretical model. Residents commute to work on the ray from the coordinate $(0, 0)$ to their residence at (r, ϕ). They travel to the closest park(s) on the annulus that includes their residence at distance r from city center. Park size is $k(r)$ and number of parks is $n(r)$.

4.2.2 Optimal green urban spaces

We consider a benevolent city planner who chooses the number $n(r)$, size $k(r)$, composite good consumption $z(r,\phi)$ and residential surface $s(r,\phi)$.[5] We consider an open city where, in the spirit of Herbert and Steven (1960), the planner keeps the residents' utility equal to the outside utility level u. The planner maximizes the surplus

$$\max_{\{s,z,n,k\}} \mathcal{S} = \int_0^b 2n(r) \left\{ \int_{k(r)/r}^{\pi/n(r)} \left[w - t(r) - z(r,\phi) \right] \frac{1}{s(r,\phi)} r \mathrm{d}\phi - \int_0^{\pi/n(r)} R_A r \mathrm{d}\phi - C\left[k(r)\right] \right\} \mathrm{d}r$$

subject to

$$U\left[z(r,\phi), s(r,\phi), a\left(k(r), r\phi\right)\right] = u$$

The term in curly brackets represents the net resource generated in the annulus strip $[r, r+dr] \times [0, \pi/n(r)]$. There are $2n(r)$ such annulus strips. The first term in curly brackets represents the net resources generated by a worker, namely, the value of her production minus her commuting cost and personal consumption of composite good. This is multiplied by the number of workers residing at (r,ϕ), which is given by the land supply, $r\mathrm{d}\phi$, divided by the worker's use of land, $s(r,\phi)$. Workers reside on land from the angle $k(r)/r$ to the angle $\pi/n(r)$. The second and third terms in curly brackets represent the land cost and the maintenance cost.

The consumption of the composite good is given by the utility constraint. Since U is an increasing function of z, the identity $U(z,s,a) = u$ can be inverted to give $\tilde{z}(s,a,u)$. Because U is an increasing function of all its arguments, it can be verified that $\partial\tilde{z}/\partial s < 0$, $\partial\tilde{z}/\partial a < 0$ and $\partial\tilde{z}/\partial u > 0$. Maximizing point-wise the surplus $\mathcal{S}$ with respect to $s(r,\phi)$ is equivalent to finding the function

$$V(r,a,u) = \max_s \frac{w - t(r) - \tilde{z}(s,a,u)}{s}$$

and applying it to $a = a\left(k(r), r\phi\right)$. This function represents the planner's value of land generated by workers. Using the envelope theorem, it can be verified that $V_r \equiv \partial V/\partial r < 0$, $V_a \equiv \partial V/\partial a > 0$ and $V_u \equiv \partial V/\partial u < 0$. The planner's problem then simplifies to

$$\max_{\{n,k\}} \mathcal{S} = \int_0^b 2n(r) \left\{ \int_{k(r)/r}^{\pi/n(r)} V\left[r, a\left(k(r), r\phi\right), u\right] r \mathrm{d}\phi - \int_0^{\pi/n(r)} R_A r \mathrm{d}\phi - C\left[k(r)\right] \right\} \mathrm{d}r$$

Note that at the optimum, parks cannot occupy the full surface of each annulus, i.e., $k(r)/r > \pi/n(r)$. Indeed, consider each annulus at a distance r from the business district. Otherwise, if $k(r)/r \leq \pi/n(r)$, the first integral in curly brackets would vanish to zero, meaning that parks would only be a net cost and that the planner would have an incentive to eliminate the park.

[5] The composite good consumption and residential surface $z(r,\phi)$ and $s(r,\phi)$ can be decentralized though a land market equilibrium.

The optimal size of parks is obtained by differentiating point-wise this expression by $k(r)$. The marginal welfare is given by

$$\mathcal{S}_k = \int_{k(r)/r}^{\pi/n(r)} V_a\left[r, a\left(k(r), r\phi\right), u\right] \; a_k\left[k(r), r\phi\right] \; r d\phi - V\left[r, a\left(k(r), k(r)\right), u\right] - C'\left[k(r)\right] \tag{4.1}$$

An increase in park size has three effects on welfare. First, it gives a higher amenity to residents and increases the value of land (first term). Second, it decreases the space available for residences and therefore social welfare by the residential value of the parcel of land (second term). Finally, it increases the cost of maintenance. The optimal size of parks requires that $\mathcal{S}_k$=0.

We now turn to the optimal number of parks. To ease the analysis, we avoid the problem of integers and treat the number of parks $n(r)$ as a function of real numbers (with $n(r) \geq 1$). The optimal number of parks is obtained by differentiating point-wise the planner's objective by $n(r)$. The marginal welfare is given by

$$\begin{aligned}\mathcal{S}_n = 2 \left\{ \int_{k(r)/r}^{\pi/n(r)} V\left[r, a\left(k(r), r\phi\right), u\right] r \mathrm{d}\phi - \int_0^{\pi/n(r)} R_A r \mathrm{d}\phi - C\left[k(r)\right] \right\} \\ - \frac{2\pi r}{n(r)} \left\{ V\left[r, a\left(k(r), r\pi/n(r)\right), u\right] - R_A \right\}\end{aligned} \tag{4.2}$$

An increase in the number of parks has two effects. It first adds a new annulus strip (first line). This generates a net surplus that balances the worker's net creation of resources, the cost of land and the maintenance cost. Second, it reduces the size of existing annulus strips (second line). There are $2n(r)$ annulus strips that are each reduced by a length $(\mathrm{d}/\mathrm{d}r)\left(r\pi/n(r)\right) = -r\pi/n^2(r)$. Multiplying those two figures yields the first term on the second line. The loss on those strips is proportional the land value created by workers V in addition to agricultural land value R_A (the second term on the second line). The optimal number of parks requires that $\mathcal{S}_n$=0.

Further analysis of the planner's problem requires characterizing the integrals in the expression of the above first-order conditions. This requires further specification of the functions for the utility, amenity and maintenance cost of parks.

4.2.3 Cobb-Douglas preferences

In this subsection, we assume that residents are endowed with the Cobb-Douglas utility function $U(z,s,a) = a^{\alpha} z^{1-\beta} s^{\beta}$, where α and β measure the preferences for amenities and residential space ($\alpha, \beta, \alpha+\beta \in (0,1)$). The amenity function is given by $a(k,d) = a_0 k^{\varphi} d^{-\delta}$, where a_0, φ, and δ are positive parameters. The maintenance cost is given by $C(k) = c_0 k^{\gamma}$, where c_0 and γ are positive parameters. To make the analysis more realistic, we restrict attention to the case in which the impact of park amenities α is small enough compared to the impact of residential space: $\alpha < \beta$. We also suppose that park amenities are proportional to their size but that residents are more sensitive to distance

to their park ($\delta > \varphi$). We finally further suppose that the maintenance cost is a concave function of park size, implying decreasing returns to scale in park maintenance ($\gamma < 1$). Increasing returns may be justified in several ways. The park infrastructure must be set up once and may not be proportional to surface (fences, paths, road, kiosk, restaurants, and so forth). Park security also has decreasing returns to scale. The second assumption entices the planner to reduce the distance between residents and parks, while the last assumption entices the planner to reduce the number of parks.

For the sake of exposition, we concentrate on the annulus at distance r and formulate the functions $n(r)$, $k(r)$ and $y(r)$ with reference to r. We compute

$$\tilde{z}(s,a,u) = s^{-\frac{\beta}{1-\beta}} a^{-\frac{\alpha}{1-\beta}} u^{\frac{1}{1-\beta}}$$
$$V(r,a,u) = \beta(1-\beta)^{\frac{1-\beta}{\beta}} y^{\frac{1}{\beta}} u^{-\frac{1}{\beta}} a^{\frac{\alpha}{\beta}}$$

where the optimal residential surface associated with V is $\hat{s} = (1-\beta)^{-\frac{1-\beta}{\beta}} y^{-\frac{1-\beta}{\beta}} a^{-\frac{\alpha}{\beta}} u^{\frac{1}{\beta}}$. The first-order conditions (4.1) and (4.2) simplify to

$$\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = 1 + \frac{\beta}{\varphi\alpha} - \frac{\delta}{\varphi} + \left(\frac{\beta}{\varphi\alpha} - \frac{\delta}{\varphi}\right) \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} \frac{c_0\gamma}{V(r,a_0,u)} \tag{4.3}$$

$$\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right) \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} \frac{c_0}{V(r,a_0,u)} \tag{4.4}$$

In Appendix A, we show that if

$$\zeta \equiv \frac{1 - \frac{\gamma\delta}{\varphi}}{\frac{\delta}{\varphi} - 1} > 0 \text{ and } \gamma < 1 + (\varphi - \delta)\frac{\alpha}{\beta} \tag{4.5}$$

the optimal park size and number of parks have the following interior and closed-form solution:

$$k^* = \left[\frac{\zeta c_0}{V(r,a_0,u)}\right]^{\frac{1}{(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta}}}$$
$$n^* = \frac{\kappa\pi r}{k^*}$$

where

$$\kappa \equiv \left(\frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right)\frac{1}{\zeta}\right)^{-\frac{1}{1-\delta\frac{\alpha}{\beta}}}$$

Since $(1-\gamma) + (\varphi-\delta)\alpha/\beta > 0$ by (4.5), the optimal park size increases in the maintenance cost c_0 and decreases in land value $V(r,a_0,u)$. Since $V_r < 0$, the optimal park size also increases in distance to the city centre r. Furthermore, *the number of parks* is proportional to r/k^*, which *is a non-monotone function of the distance to the CBD*: it first increases for small r and then decreases for sufficiently large r. For this interior solution, the share of park area on the annulus is given by

$$\frac{n^*k^*}{\pi r} = \left(\frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right)\frac{1}{\zeta}\right)^{-\frac{1}{1-\delta\frac{\alpha}{\beta}}} < 1$$

which is a constant independent of the distance to the city centre. The distance between two parks is given by

$$d^* = \frac{2\pi r}{n^*} = \frac{2}{\kappa}k^* = \frac{2}{\kappa}\left[\frac{\zeta c_0}{V\left(r, a_0, u\right)}\right]^{\frac{1}{(1-\gamma)+(1-\delta)\frac{\alpha}{\beta}}} \tag{4.6}$$

and is increasing in the distance from the city centre r. Finally, note that since V rises with w, richer cities include more numerous parks with smaller sizes and longer access distances.

Interior optimal solutions take place under conditions (4.5), which can be restated as $\delta > \varphi$ and $\gamma < \min\{\varphi/\delta, 1+ (\varphi - \delta)\,\alpha/\beta\}$. The first condition requires that the park amenity decreases proportionally faster with its access than its size. This entices the planner to set up more parks. The second condition requires sufficiently increasing returns to scale in maintenance, which entices the planner to set up fewer parks. Under those conditions, the planner balances the two effects by implementing a finite number and size of parks ($1 < n^* < \infty$ and $k^* > 0$). When those conditions are not satisfied, the planner implements corner solutions with either infinitely many small parks ($n^* \to \infty$, $k^* \to 0$) or a single large park ($n^* = 1$).

Proposition 1 *Assume Cobb-Douglas preferences and suppose that park amenities are proportional to their size but that residents are more sensitive to distance to their park ($\delta > \varphi$) and that maintenance costs have decreasing returns to scale in park maintenance ($\gamma < 1$). Then, under conditions (4.5) the optimal number of parks is a bell-shaped function of the distance to the CBD, while park size and the distance between parks increases with it. More productive cities include more numerous parks with smaller sizes and longer access distances.*

To further characterize our theoretical predictions, we finally assume an exponential commuting cost $t(r) = w(1 - e^{-tr})$. In this case, taking the natural logarithm after replacing the land value equation V, we obtain the following optimal number of parks:

$$\ln n = \ln\left[\kappa\pi\left(\frac{\beta(1-\beta)^{\frac{1-\beta}{\beta}}}{\zeta c_0}\right)^{\vartheta}\right] + \frac{\vartheta}{\beta}\ln w + \ln r - \vartheta t r + \vartheta \ln\left(u^{-\frac{1}{\beta}}a_0^{\frac{\alpha}{\beta}}\right) \tag{4.7}$$

where $\vartheta \equiv [(1-\gamma) + (\varphi - \delta)\alpha/\beta]^{-1}$. The non-monotonic relationship between the number of parks and distance to the city centre is reflected by the bundle of terms $\ln r - \vartheta t r$. There are two distinct channels through which the distance to the city centre affects the number of green urban spaces. The first term in $\ln r$ indicates the positive effect of an increase in land availability on the number of parks as one moves farther from the city core: more available land entices the planner to set a larger number of parks. The second term $(-\vartheta t r)$ shows the effect of the commuting cost and population density on the number of parks. As one moves away from the city centre, the commuting cost increases, and the attractiveness of land and therefore population density decrease. The

city planner then has less incentive to provide many parks. Those two effects are opposite in sign, and together form the non-monotone relationship between the number of green urban spaces and the distance r. At a small r, the effect of land availability dominates, leading to an increase in n, while at large r, the effect of population commuting and density dominates and induces a decrease in n.[6]

Similarly, the optimal park size is written as

$$\ln k = \vartheta \ln \left(\frac{\zeta c_0}{\beta(1-\beta)^{\frac{1-\beta}{\beta}}} \right) - \frac{\vartheta}{\beta} \ln w + \frac{\vartheta}{\beta} tr - \vartheta \ln \left(u^{-\frac{1}{\beta}} a_0^{\frac{\alpha}{\beta}} \right) \tag{4.8}$$

which increases with distance from the city centre and decreases with city per-capita income. The optimal distance between parks is given by

$$\ln d = \ln \left(\frac{2}{\kappa} \right) + \vartheta \ln \left(\frac{\zeta c_0}{\beta(1-\beta)^{\frac{1-\beta}{\beta}}} \right) - \frac{\vartheta}{\beta} \ln w + \frac{\vartheta}{\beta} tr - \vartheta \ln \left(u^{-\frac{1}{\beta}} a_0^{\frac{\alpha}{\beta}} \right) \tag{4.9}$$

which rises with distance from the city centre and falls with city wages. Finally, we can discuss the spatial pattern of population density for any given configuration of green urban space systems $\{k(r), d(r)\}$. The logarithm of population density is written as

$$\ln \lambda = -\ln \hat{s} = \ln \left[(1-\beta)^{\frac{1-\beta}{\beta}} a_0^{\frac{\alpha}{\beta}} \right] + \frac{1-\beta}{\beta} \ln w - \frac{1-\beta}{\beta} tr + \frac{\alpha\varphi}{\beta} \ln k - \frac{\alpha\delta}{\beta} \ln d - \frac{1}{\beta} \ln u \tag{4.10}$$

It increases with larger park sizes and smaller distance to parks because the these factors augment park amenities and attract more residents. Relationships (4.7) to (4.10) are used as guidance for the empirical analysis in the next section.

4.3 Empirical analysis

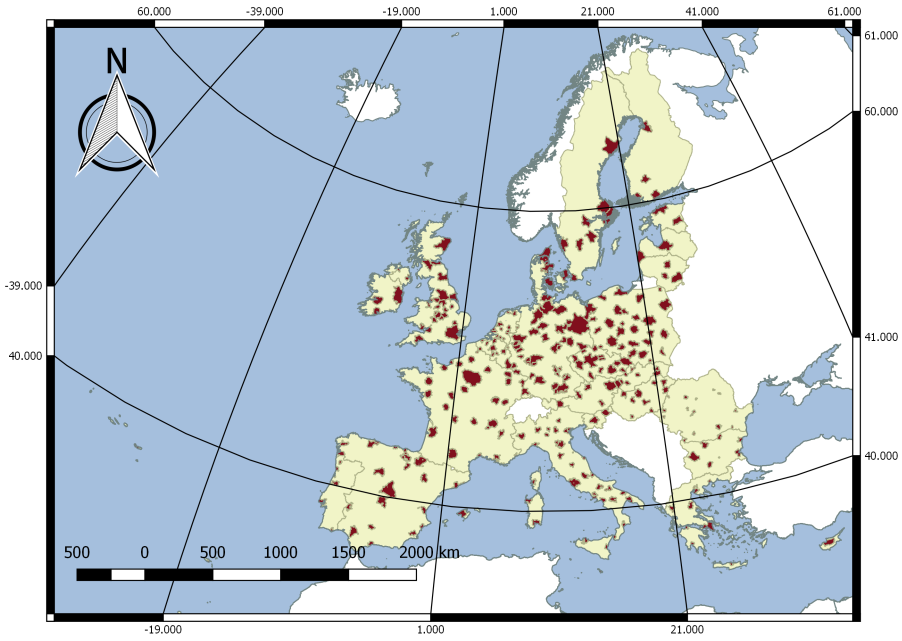
4.3.1 Data

To investigate the stratification of green urban spaces across cities in Europe, we use the Global Monitoring Environment Services (GMES) Urban Atlas from the European Environment Agency (EEA) for the period 2006-2008.

The GMES Urban Atlas provides reliable, inter-comparable, and high-resolution land use maps for 305 Large Urban Zones and their surroundings by collating satellite images with 2.5 m spatial resolution data. To the best of our knowledge, GMES Urban Atlas is the most detailed land use map for the urban areas in Europe, which provides

[6]Note that the above non-monotone relationship is not driven by the assumption of an exponential commuting cost. Indeed, under the assumption of a linear commuting cost, we may write the net income as $y = w - tr$, so that the second to fourth terms in (4.7) is written as $\frac{\vartheta}{\beta} \ln (w - tr) + \ln r$, which is also a concave function of r.

Figure 3: Maps of most populous urban centres in Europe 2006-2008



GMES Urban Atlas 2006 urban boundary, EU Coastal and country boundary were used to made this maps. GMES 2006 database covers 305 most populous urban centres in 27 EU member states including Austria, Belgium, Bulgaria, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and United Kingdom.

an advantage over the CORINE Land Cover maps commonly used in the literature.[7] The datasets cover all Functional Urban Areas (FUAs) with at least 100, 000 inhabitants[8] Figure 3 displays the urban areas covered by GMES Urban Atlas 2006.

The Urban Atlas provides a classification of city zones that allows for a comparison of the density of residential areas, commercial and industrial zones, and the extent of green areas. In this paper, we use the data on "green urban areas" (class 14100), which are defined as artificial non-agricultural vegetated areas. They consist of areas with planted vegetation that is regularly worked and/or strongly influenced by humans. Precisely, first, green urban areas include public green areas used predominantly for recreational use (gardens, zoos, parks, castle parks, cemeteries, etc.). Second, suburban natural areas that have become and are managed as urban parks are included as green urban areas. Finally, green urban areas also include forest and green areas that extend from the surroundings into urban areas with at least two sides being bordered by urban areas and structures and containing visible traces of recreational use. Important

[7] GMES maps have a 100 times higher resolution than CORINE Land Cover.

[8] See the definitions in the Urban Audit database and European Environmental Agency, GMES Urban Atlas metadata. Link `https://land.copernicus.eu/local/urban-atlas/urban-atlas-2006?tab=metadata` (accessed on June 03, 2019).

for our study, green urban areas do not include private gardens within housing areas, buildings within parks, such as castles or museums, patches of natural vegetation or agricultural areas enclosed by built-up areas without being managed as green urban areas. It must be noted that green urban areas belong to the Urban Atlas' class of "artificial surfaces", which include all non-agricultural land devoted to human activities.[9] This class is distinguished from agricultural, semi-natural areas and wetlands, forest areas and water areas devoted to non-urban activities. From this section on, we use the terms green urban spaces, urban parks, and parks interchangeably.

We select the (oldest) town hall locations as the CBDs.[10] Then, we create a set of circles around each CBD at 100 m intervals. For each circle, we create a buffer zone of 250 m, and we define the "annulus land area" as the intersection of the buffer zones and the land within the urban zone area reported by the GMES. This area includes artificial surfaces, agriculture, semi-natural areas, wetlands, and forest but does not include water areas because those seas and oceans are not appropriate for potential human dwellings. We then choose all the green areas located within the annulus and record their centroids, surface areas (sizes) and the closest distances to other green areas. Figure 4 displays the annuli and the land use of green urban areas (green colour) for Paris.

In general, urban models assume a clear separation between urban and non-urban zones. However, urban data do not provide a neat division between residential land use and agricultural and rural zones at each distance to the city core. Therefore, in this paper, we follow Picard and Tran (2019) and choose the city boundaries to the annulus at which the share of residential surfaces over the annulus land areas falls below 20%. Residential surfaces include urban areas with dominant residential use, inner-city area with the central business district and private buildings. Because the residential land ratio has high frequency variations around potential city borders, we employ the Nadaraya-Watson Gaussian Kernel[11] to smooth those variations before setting up the city borders. Alternative thresholds do not qualitatively alter the results below.

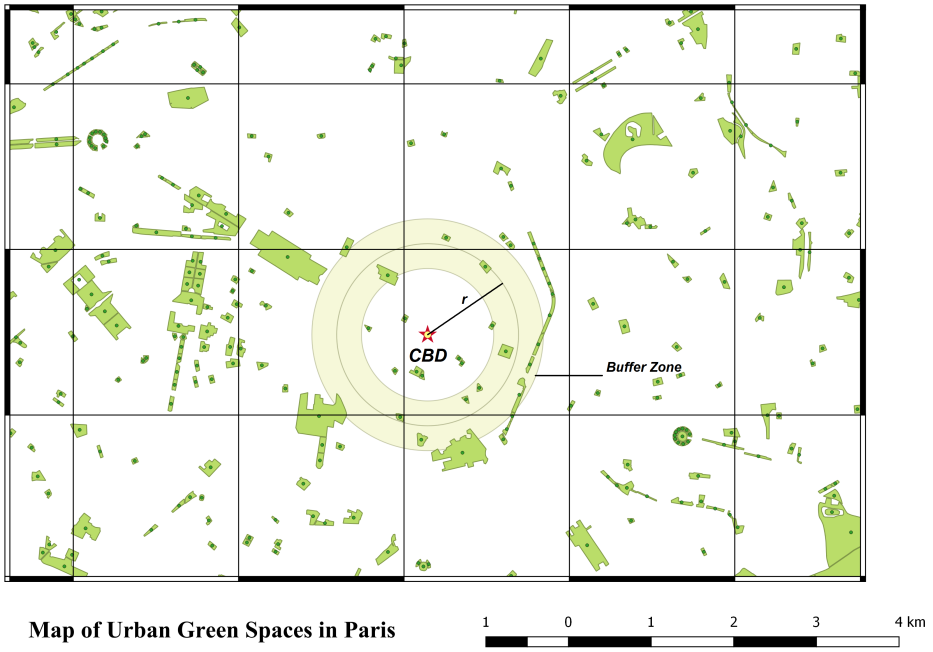
We measure the distance between two parks as the distance between two park centroids. One thing we want to avoid is the case in which a road separates a park into two separate polygons. We apply the following procedures to identify continuous green urban spaces. First, we compute the distance between every two pairs of green urban centroids. We further assume a circular shape for the park and derive its radius using

[9] In addition to green urban areas, artificial surfaces include urban areas with dominant residential use, inner-city areas with central business district and residential use, industrial, commercial, public, military and private units, transport units, mines, dumps, and construction sites, and sports and leisure facilities.

[10] The town hall is commonly used in the literature as the location of the CBD in European cities because most of the cities in Europe were formed many centuries ago.

[11] The Nadaraya-Watson estimate is given by $\left(\sum_{i=1}^{n} K(x-x_i)g_i\right)\left(\sum_{i=1}^{n} K(x-x_i)\right)$ where $K(x) = \frac{\exp(-x^2/(2h^2))}{\sqrt{(2\pi)h}}$. Here, we choose bandwidth $h = 2$ (km). For further details, see Hansen (2009).

Figure 4: Urban Green Areas in Paris



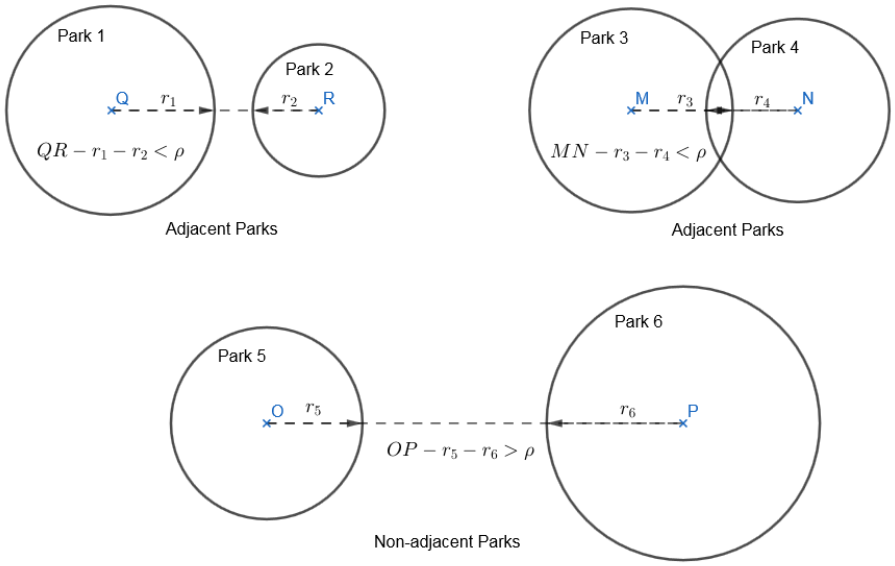
the circle area formula. Second, we obtain the distances between the two boundaries of green urban spaces by differencing the distance between two park centroids and the sum of two park radii. If this distance is less than $\rho = 5$ meters, we consider these two parks to be adjacent parks and combine them into one park.[12] The centroid of the new combined park polygon is recalculated as the middle point of the two old centroids, and the size becomes the sum of those two areas. We illustrate the criteria for adjacent parks in Figure 5 below. Once the distinguished polygon and centroid of each park are determined, we denote $d_{mm'}$ as the distance between two parks m and m': $d_{mm'} \equiv \sqrt{(x_m - x_{m'})^2 + (y_m - y_{m'})^2}$, where (x_m, y_m) is the coordinate of the centroid of park m. We compute the radius of each park and the distance to other parks as follows:

$$\begin{aligned}
\text{Park Radius} \quad & k_m = \sqrt{\frac{\text{surface of park } m}{\pi}} \\
\text{Closest Distance} \quad & d_m = \min_{m'} \; d_{mm'} \quad \forall m' \neq m \\
\text{CES Distance} \quad & d_m^{\psi} = \left(\frac{1}{N} \sum_{m'=1}^{N} \left(d_{mm'}\right)^{-\psi} \right)^{-1/\psi} \quad \text{where } \psi > 0
\end{aligned}$$

where N is the total number of parks located within urban areas. The value of ψ can

[12]The threshold for ρ here is equal to 5 meters, which approximates the width of the roads that cross parks. We considered alternative the thresholds such as 20, -5, and -20 meters. Our results are robust to the consideration of these thresholds.

Figure 5: Illustration of Merger of Adjacent Parks



change to any positive number, and note that when $\psi \longrightarrow \infty$, the CES distance d_m^{ψ} coincides with the closest distance $d_m = \min_{m'} d_{mm'} = d_m^{\infty}$.

In addition to GMES Urban Atlas, we use the population density from the European population grid.[13] We calculate the population mass at each distance to the city centre and redistribute the population to the residential area[14] in each annulus. As the Eurostat population grid does not cover Cyprus, we drop the city of Lefkosia in Cyprus. For the city income level, we utilize the GDP per capita measures in standard purchasing power at the NUTS3 level from the Eurostat's Regional Economic Accounts, which give the finest detail on household income and GDP per capita for Europe.[15] Other measures of cities' exogenous geographical characteristics are taken from the E-OBS database from the EU-FP6 project, which covers the entire range of daily temperature, daily precipitation, sea level pressure and the elevation across Europe at a 1 km square raster. We finally measure the city population as the number of inhabitants (in millions) living in the City and Greater City (CGC) defined and reported in the Eurostat database.[16]

[13]For further information on the European population grid, please check the technical report of the GEOSTAT 1A project from Eurostat. Link: `http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_GGS_v3.0.1.pdf` (accessed on June 09, 2019).

[14]This is the urban fabric category in GMES Urban Atlas.

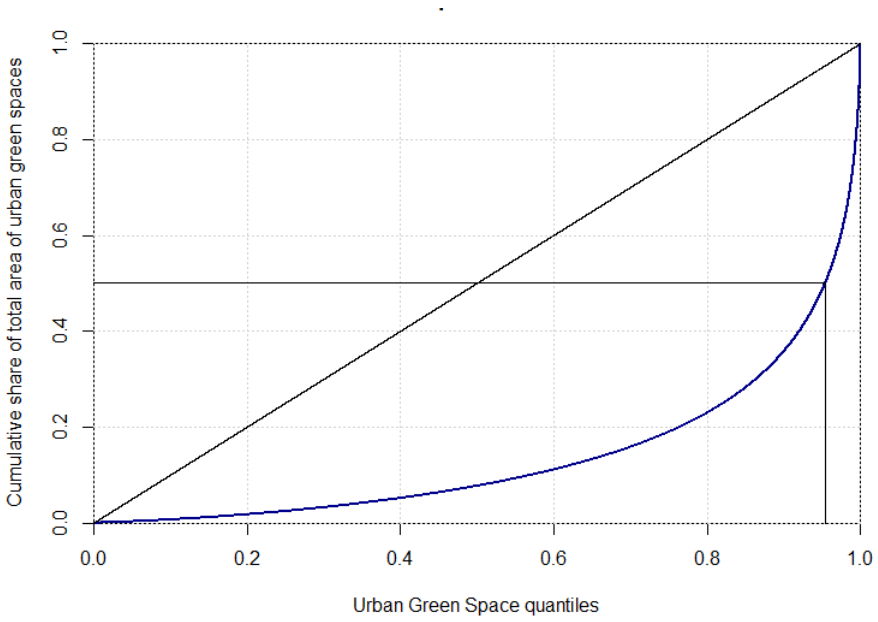
[15]If there exists more than one NUTS3 area for one annulus, we average the values. For the differences between NUTS2 and NUTS3 levels, see details at Eurostat, link `https://ec.europa.eu/eurostat/web/nuts/background` (accessed on July 03, 2019).

[16]For further details, see the metadata file for *urb_esms* on the Urban Audit database.

4.3.2 Descriptive statistics

Table 4.1 provides the summary statistics of green urban spaces in European cities. Our sample includes the 303 most populous cities in 27 countries in Europe.[17] On average, cities have approximately 196 green urban spaces, with a total area of 8.58 square kilometres per city. Catanzaro on the south coast of Italy is the city with the smallest number of green urban spaces, and London, U.K., is the city with the highest number of green urban spaces with more than 2000 parks. The city with the smallest total surface of green urban spaces is the city of Konin in Poland, while that with the highest surface is Paris, France, with nearly 300 square kilometres of green area. The average size of a green urban is approximately 3.42 hectares per park. The city with the smallest average park size is still Konin, and that with the largest average park size is Madrid, Spain, with 25 hectares per park on average. Figure 6 displays the size distribution of all green urban spaces across the 303 most populous cities in Europe. Note that the largest 5% of parks account for more than 50% of the total surface of green areas. This pattern is also similar within each urban area, with an even higher concentration in larger cities such as Madrid or Paris.

Figure 6: Urban Green Size Distribution within Urban Area



City size affects the surface and distribution of green urban spaces. To illustrate the difference in the park size distribution across cities, we calculate the Gini index of the size of green urban spaces in each city. A low level of the Gini index indicates homogeneity in park size, while a high coefficient indicates high heterogeneity in park

[17] We exclude Luxembourg and Cyprus as they are relatively small, and we do not have information at the NUTS3 level on household income and population grid.

size. The average Gini index is 0.61, with a standard deviation of 0.11, and the city with the highest level of the Gini index of 0.93 is Madrid, which has many large and small parks. Table 4.2 presents correlations between the total park surface and park distribution on the one hand and proxies for city size on the other hand (GDP per capita, border, and population). They show that more populous and richer cities are associated with larger park surfaces and greater park size heterogeneity. Such correlations are smaller if we exclude the first and last percentiles of park size. In this case, total green surfaces are almost proportional to population sizes (the coefficient is 1.098).

Table 4.1: Summary Statistics of Urban Green Spaces across European cities

A. City Statistics			
	Mean	Std. Dev.	Obs.
City Border (km)	5.68	3.98	303
Population (mil.)	0.45	0.80	303
GDP per capita (€1000/hab./y.)	26.75	12.88	303
Average Rainfall	1.91	0.59	303
Elevation (m)	212.19	211.16	303
Average Temperature in Jan 01 (°C)	2.28	4.61	303
Average Temperature in Jul 01 (°C)	19.16	3.36	303

B. Urban Green Space Statistics	Full UGS sample			Excluding big UGS sample		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Average Number	196.38	295.08	303	194.41	290.44	303
Average Size (ha)	3.42	1.92	303	2.88	1.17	303
Average Radius (m)	81.19	17.07	303	78.60	14.53	303
Total Green Area (km^2)	8.58	18.17	303	6.28	10.71	303
Gini Index	0.61	0.11	303	0.58	0.09	303
Average Closest Distance d (m)	336.85	94.59	302	336.42	94.66	302
Average 2^{nd} Closest Distance d (m)	496.55	148.55	302	496.47	148.82	302
Average CES Distance d^{ψ} (m)						
$\psi = 2$	1726.41	1001.42	302	1722.24	996.99	302
$\psi = 10$	520.67	144.46	302	519.63	144.07	302
$\psi = 50$	368.63	101.35	302	368.11	101.37	302

Note: City border is chosen at the distance where the artificial residential area is less than 20% of total land. The residential land share is smoothed using Gaussian Kernel before the cut-off. The threshold for big parks is set at 99% quantile, which is at 54.8 hectares, and is computed from whole samples of all parks within city border of all 303 cities. We excluded the city of Catanzaro, Italy, when calculating the average minimum distance to other parks as the sample size is too small in Catanzero.

In the theoretical model, we consider the parks located in the local area of patronizing residents but not the large parks that provide a global amenity to all residents. We leave the theoretical investigation of the location of large parks for future research. In this empirical analysis, we first consider the full sample of parks, but for robustness, we dedicate special attention to the parks in the lower 99th-percentile set of park size. As can be seen below, the results are qualitatively and quantitatively unaltered following this change in dataset.

In this paper, we use the GDP per capita at the NUTS3 level as a baseline proxy

Table 4.2: City Size and Urban Green Spaces across European Cities

Full Sample

	Dependent variable:					
	Gini Index			Ln Total Green Surface (km^2)		
	(1)	(2)	(3)	(4)	(5)	(6)
Ln City GDP per capita (€1000)	0.126***			2.337***		
	(0.024)			(0.246)		
Ln City Border (*km*)		0.125***			2.110***	
		(0.011)			(0.077)	
Ln City Population (*mil.*)			0.064***			1.247***
			(0.008)			(0.056)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	303	303	303	303	303	303
Adjusted R^2	0.139	0.350	0.251	0.384	0.780	0.706

Note: City GDP per capita is measured at NUTS3 level and at €1000 per inhabitant. The table presents the results of the regression of the log of total surfaces in each city and the Gini index of urban green sizes within each city. Data: GMES, European cities. Standard errors are reported in parenthesis and significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

for city income w. GDP per capita varies significantly across EU cities, from €6,000 to €83,700 per inhabitant, with an average of €26,750 measuring in terms of purchasing power standard.[18] GDP per capita represents more than 1.5 times the residents' net income, which reflects the tax wedge between production value and income. City elevation also varies vastly, from two meters below sea level in Amsterdam, the Netherlands, to 1,614 meters in the mountainous city of Innsbruck, Austria. The European continent belongs to a mild climate zone with temperature varying between -8 and $+28$ degrees Celsius for the lowest and highest day-time temperatures in winter and summer on average (measured on January 1 and July 1, respectively, for the period $1995 - 2006$).[19] The controls for climate capture the possible effects of residents' preference for nice weather (e.g., not too cold in winter and not too hot in summer; see Rappoport 2007).

4.3.3 Number of green urban spaces

In this section, we examine the patterns in the numbers of green urban spaces across locations within the city area. The theoretical model predicts that the number of green urban spaces is a non-monotone function of the distance to the city centre: it first increases at distances close to CBD and then decreases as one moves closer to the urban fringe. Using equation (4.7) above, we can test the following reduced-form regression

[18]The purchasing power standard unit is robust to the difference in the prices of goods and services across city and country boundaries. For further details, see Eurostat Statistics Explained Glossary, link `https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Purchasing_power_standard_(PPS)` (accessed on June 20, 2019).

[19]The samples in our data do not contain some Northern European cities in Iceland and Norway.

$$\ln\left(n_{ijc}\right) = \rho_0 + \rho_1 \ln w_{jc} + \rho_2 \ln dist_{ijc} + \rho_3 dist_{ijc} + \sum_{ijc} X_{ijc}^{(1)} + \sum_{jc} X_{jc}^{(2)} + \sum_{c} I_c + \epsilon_{ijc}$$

where (i, j, c) is an observation of annulus i in city j in region c, n_{ijc} is the number of parks for each annulus i, $dist_{ijc}$ the distance from the city centre and w_{jc} is city income per capita. $X_{ijc}^{(1)}$ are control variables for annuli (proximity of agricultural and forest land) and $X_{jc}^{(2)}$ for city observable characteristics (elevation, temperature, rain fall, latitude, etc.). I_c is a region dummy variable controlling for unobservable regional characteristics such as environmental policies or local regulation. Our theoretical model predicts that the number of parks first increases and then decreases as one moves farther from city centre; hence, the sign of ρ_2 is expected to be positive, while ρ_3 is expected to be negative. We also expect that the number of parks increases with city income w; hence, ρ_1 is expected to be positive.

The regression results are reported in Table 4.3. The standard errors are clustered and robust at the city level, which controls for the possible correlation between observations within the same urban area. Column (1) includes no controls, while Columns (2) and (3) successively introduce the city geographical controls and the annulus controls for nearby agriculture and forestry. Columns (4) and (5) add fixed effects at the country and regional level. Columns (6), (7), (9) and (10) restrict the data to monocentric cities. Columns (8), (9) and (10) exclude large parks by using the parks in the lower 99th percentiles of park sizes. Table 3 shows a positive and significant coefficient for $\hat{\rho}_2$ and a negative and significant coefficient for $\hat{\rho}_3$ across all kinds of controls and datasets. *This confirms the non-monotonic relationship between the number of parks and the distance to the city centre and the presence of the two above different channels of land availability and population density.* Moreover, as expected from the theoretical model, higher city income increases the number of parks. Note that the presence of agriculture and forest land controls diminishes the number of parks. This is to be expected, as those substitute for park amenities.

The magnitudes of $\hat{\rho}_2$ and $\hat{\rho}_3$ reflect the quantitative effects of the channels of land availability and population density. According to Column (7), a 1% increase in the distance to the city centre first leads to a 0.776% increase in the number of parks through the land availability channel, while a one-kilometre increase in distance is associated with a 7.9% decrease in the number of parks through the population density channel. To fix ideas, when one moves from 1 kilometre to 2 kilometres from the city centre, the first channel generates a 79% increase in the number of parks, while the second channel yields a 7.9% decrease, so the overall number of parks increases by 74%. Therefore, land availability matters more in locations close to the city centre. However, when one moves from 20 kilometres to 21 kilometres from city centre, the first channel generates a 3.8% increase in the number of parks, and the second channel yields a 7.9% decrease,

Table 4.3: Number of Urban Green Spaces across Locations within Urban Area (OLS)

	Dependent variable: Ln Number of Urban Green Spaces									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln GDP per capita	0.561***	0.478***	0.235***	0.229***	0.147	0.272**	−0.236	0.150	0.269*	−0.236
	(0.065)	(0.079)	(0.073)	(0.076)	(0.122)	(0.138)	(0.148)	(0.121)	(0.138)	(0.147)
Ln Distance to CBD	0.564***	0.564***	0.771***	0.774***	0.794***	0.783***	0.776***	0.788***	0.778***	0.771***
	(0.048)	(0.047)	(0.037)	(0.033)	(0.026)	(0.030)	(0.028)	(0.026)	(0.030)	(0.028)
Distance to CBD (km)	−0.044*	−0.044**	−0.044***	−0.036***	−0.046***	−0.051***	−0.079***	−0.046***	−0.050***	−0.078***
	(0.023)	(0.022)	(0.017)	(0.014)	(0.013)	(0.017)	(0.014)	(0.013)	(0.017)	(0.015)
Annuli controls										
Agriculture Land (%)			−5.228***	−5.295***	−6.119***	−6.034***	−5.401***	−6.095***	−6.021***	−5.394***
			(0.512)	(0.546)	(0.430)	(0.457)	(0.420)	(0.430)	(0.457)	(0.420)
Forest Land (%)			−7.236***	−7.114***	−4.975***	−4.647***	−2.840***	−4.958***	−4.642***	−2.848***
			(0.891)	(0.842)	(0.910)	(0.989)	(0.769)	(0.911)	(0.991)	(0.774)
City Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Monocentric cities	Monocentric cities	Full samples	Monocentric cities	Monocentric cities
UGS Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.
City Geo. Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Controls	No	No	No	Country	NUTS2	NUTS2	NUTS2 × City Size	NUTS2	NUTS2	NUTS2 × City Size
Observations	16,782	16,782	16,782	16,782	16,782	14,368	14,368	16,780	14,366	14,366
Adjusted R^2	0.350	0.374	0.486	0.547	0.675	0.645	0.679	0.671	0.642	0.676

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €1,000 per inhabitant. The distance to CBD is measured in kilometre. City geographical controls includes the average Elevation, Latitude, Longitude and the average and square of the average of the rain fall, the temperature in summer and winter time. The average rain fall, and average temperature in Jan 01 (winter time) and average temperature in July 01 (summer time) are measured for the period $1995 - 2010$. City boundary is chosen at 20% cut-off point of residential land use share (See details in Picard and Tran (2019)). The thresholds for urban green size at 99^{th} percentile is 54.82 hectares. Columns (6), (7), (9) and (10) isolate monocentric cities by excluding the polycentric cities reported by OECD. Columns (1) to (7) include all urban green spaces within urban boundary, while Columns (8) to (10) exclude all big parks above 99^{th} percentile. The Agriculture and Forest Land from the last two annuli control variables are measured within 100-meter buffer around the residential area. Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

so the overall number of parks falls by 4.1%. Thus, the non-monotone relationship is particularly marked in larger cities.

The regression analysis of the non-linear relationship $\ln r - \vartheta tr$ is driven by the theoretical model but is not commonplace in the empirical literature. A standard way to test for the concavity of the number of parks with distance to the CBD is to assume a quadratic function of distance. Appendix B shows that the quadratic regression also yields a concave relationship between the number of parks and the distance to the centre.

Finally, Table 4.3 also shows a positive and significant sign for the coefficient for the city income level of $\hat{\rho}_1$ in most cases. Using the regression results from Column (6), a 1% increase in income level is associated with a 0.27% increase in the number of parks after controlling for city observable characteristics and NUTS2 fixed effects. We also observe a significant and negative effect for the share of agricultural and forest land surrounding residences on the number of green urban spaces. Nearby agriculture and forest land can be seen as a substitute for the amenities of urban parks. A 1% increase in agriculture (resp. forest) land leads to a reduction of 6.0% (resp. 4.6%) in the number of parks.

4.3.4 Size of green urban spaces

In the theoretical model, the optimal park sizes rise with distance to the city centre and fall with city income. Using (4.8), we can run the following regression:

$$\ln k_{ijc} = \eta_0 + \eta_1 \ln w_{jc} + \eta_2 dist_{ijc} + \sum_{ijc} X_{ijc}^{(1)} + \sum_{jc} X_{jc}^{(2)} + \sum_{c} I_c + \epsilon_{ijc} \tag{4.11}$$

where k_{ijc} is the average size of the parks located at annulus i of city j of region c, $dist_{ijc}$ is the distance of the annulus to the CBD and $X_{ijc}^{(1)}$, $X_{jc}^{(2)}$ and I_c are the same control variables as in the previous subsection. From the theoretical model, we predict that green urban systems increase in size with distance to the city centre $\eta_2 > 0$ and decrease in size with city income $\eta_1 < 0$.

Table 4.4 reports the regression results with the same column structure as in Table 4.3. First, the effect of distance is positive and significant in all regressions whatever the controls and data set. This suggests that *the size of green urban spaces increases with distance to the city centre*. Notably, for a 1-kilometre increase in the distance to the CBD, the park radius increases by 2.3% (i.e., by approximately 5% of park area). When we restrict attention to monocentric cities and omit the large green urban spaces, the magnitude declines slightly to 2.0%. Second, the effect of city income w on park size is not significant, except when one restricts the dataset to monocentric cities. In that case, the magnitude of the effect is small; Column (6) suggests that a 1% increase in city income leads to a 0.15% increase in park radius (i.e., an approximately 0.3% increase in park

area). Finally, it is surprising to observe that the presence of nearby forestry does not affect park size. Nearby agricultural land also has no effect on monocentric cities.

4.3.5 Distance to other green urban spaces

According to our theoretical prediction, the distance between parks increases with the distance from the city centre and falls with city income. We run a similar regression as in the previous sections with the distance between two parks as the explanatory variable. Precisely, we run the following reduced-form regression:

$$\ln d_{ijc} = \mu_0 + \mu_1 \ln w_{jc} + \mu_2 dist_{ijc} + \sum_{ijc} X_{ijc}^{(1)} + \sum_{jc} X_{jc}^{(2)} + \sum_{c} I_c + \epsilon_{ijc} \tag{4.12}$$

where d_{ijc} is the average of the closest distance to other parks in annulus i of city j in region c. As above, we expect the sign of the coefficient for the distance to the CBD μ_2 to be negative.

Table 4.5 reports the regression results and has the same structure as Tables 4.3 and 4.4. On the one hand, the distance to the closest park also increases as one moves toward the urban fringe. Notably, a one-kilometre increase in distance to the CBD leads to a 2.2% increase in the distance to the closest park, as indicated in Column (5) with NUTS2 region fixed effects. This figure increases slightly when we restrict data to monocentric cities and further introduce the city size fixed effects. On the other hand, richer cities are associated with a denser network of green urban spaces, as shown in Columns (1) to (3). However, the effect becomes smaller and non-significant when we control for regional fixed effects and city size. Note that the distance between parks increases in the shares of nearby agricultural and forest land, as the latter is likely to provide amenities that compensate for longer trips to parks. For instance, a 1% increase in the land shares of nearby agriculture and forest leads to 1.8% and 1% increases, respectively, of the distance to the closest park.

It is interesting to note the close similarity between the coefficients on distance to the CBD in Tables 4.4 and 4.5. This suggests a strong consistency with the Cobb-Douglas model, where by (4.6) the effect on the distance to the closest park is proportional to that on park size: $\mathrm{d}\ln d/\mathrm{d}\ln r = \mathrm{d}\ln k/\mathrm{d}\ln r$. However, the consistency is weaker when we can compare the coefficients in Tables 4.3 and 4.5. By (4.6), the theoretical model gives the relationship $\mathrm{d}\ln d/\mathrm{d}\ln r = 1 - \mathrm{d}\ln n/\mathrm{d}\ln r$. Empirically, this means to compare the value 0.024 in Table 4.5 to the value $1 - 0.794$ obtained from Table 4.3, Column (5). Hence, the Cobb-Douglas model fits well the relationship between the distance to the closest park and park size but fits less well the relationship between the former and the number of parks.

How do those relationships hold when one considers not only the closest park but the parks surrounding a residence? To answer this question, we again use the CES distance d^{ψ} with various values of the parameter ψ. As described above, this index

Table 4.4: Sizes of Urban Green Spaces (OLS)

	Dependent variable: Ln Average Urban Green Space Radius									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln GDP per capita	0.031	−0.001	−0.019	0.002	0.047	0.151***	0.145**	0.049	0.129***	0.141***
	(0.024)	(0.025)	(0.024)	(0.029)	(0.056)	(0.044)	(0.057)	(0.047)	(0.040)	(0.051)
Distance to CBD (km)	0.022***	0.022***	0.026***	0.025***	0.023***	0.023***	0.021***	0.021***	0.020***	0.020***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Annuli controls										
Agriculture Land (%)			−0.460***	−0.513***	−0.397**	−0.176	−0.163	−0.371***	−0.188	−0.193
			(0.155)	(0.166)	(0.165)	(0.153)	(0.159)	(0.139)	(0.129)	(0.134)
Forest Land (%)			−0.370	−0.388	0.283	0.224	0.313	0.284	0.180	0.232
			(0.338)	(0.314)	(0.323)	(0.318)	(0.303)	(0.296)	(0.287)	(0.263)
City Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Monocentric cities	Monocentric cities	Full samples	Monocentric cities	Monocentric cities
UGS Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.
City Geo. Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Controls	No	No	No	Country	NUTS2	NUTS2	NUTS2 × City Size	NUTS2	NUTS2	NUTS2 × City Size
Observations	16,782	16,782	16,782	16,782	16,782	14,368	14,368	16,780	14,366	14,366
Adjusted R^2	0.070	0.158	0.163	0.202	0.411	0.430	0.435	0.413	0.429	0.435

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €1,000 per inhabitant. The distance to CBD is measured in kilometre. City geographical controls includes the average Elevation, Latitude, Longitude and the average and square of the average of the rain fall, the temperature in summer and winter time. The average rain fall, and average temperature in Jan 01 (winter time) and average temperature in July 01 (summer time) are measured for the period $1995 - 2010$. City boundary is chosen at 20% cut-off point of residential land use share (See details in Picard and Tran (2019)). The thresholds for urban green size at 99^{th} percentile is 54.82 hectares. Columns (6), (7), (9) and (10) isolate monocentric cities by excluding the polycentric cities reported by OECD. Columns (1) to (7) include all urban green spaces within urban boundary, while Columns (8) to (10) exclude all big parks above 99^{th} percentile. The Agriculture and Forest Land from the last two annuli control variables are measured within 100-meter buffer around the residential area. Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

Table 4.5: Closest Distance to Other Urban Green Spaces (OLS)

	Dependent variable: Ln Average Closest Distance to other parks *dmin*									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln GDP per capita	-0.124^{***}	-0.130^{***}	-0.068^{**}	-0.049	-0.057	-0.036	0.073	-0.055	-0.036	0.072
	(0.027)	(0.031)	(0.031)	(0.034)	(0.056)	(0.068)	(0.068)	(0.056)	(0.068)	(0.068)
Distance to CBD (km)	0.040^{***}	0.040^{***}	0.028^{***}	0.023^{***}	0.022^{***}	0.023^{***}	0.031^{***}	0.021^{***}	0.023^{***}	0.031^{***}
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)
Annuli controls										
Agriculture Land (%)			1.549^{***}	1.553^{***}	1.785^{***}	1.862^{***}	1.687^{***}	1.814^{***}	1.890^{***}	1.713^{***}
			(0.218)	(0.220)	(0.209)	(0.228)	(0.223)	(0.210)	(0.228)	(0.223)
Forest Land (%)			1.446^{***}	1.384^{***}	0.991^{**}	1.056^{***}	0.483	0.963^{**}	1.022^{**}	0.450
			(0.423)	(0.406)	(0.388)	(0.406)	(0.342)	(0.388)	(0.406)	(0.341)
City Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Monocentric cities	Monocentric cities	Full samples	Monocentric cities	Monocentric cities
UGS Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.
City Geo. Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Controls	No	No	No	Country	NUTS2	NUTS2	NUTS2 × City Size	NUTS2	NUTS2	NUTS2 × City Size
Observations	16,782	16,782	16,782	16,782	16,782	14,368	14,368	16,780	14,366	14,366
Adjusted R^2	0.164	0.184	0.236	0.309	0.452	0.454	0.478	0.453	0.455	0.479

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €1,000 per inhabitant. The distance to CBD is measured in kilometre. City geographical controls includes the average Elevation, Latitude, Longitude and the average and square of the average of the rain fall, the temperature in summer and winter time. The average rain fall, and average temperature in Jan 01 (winter time) and average temperature in July 01 (summer time) are measured for the period $1995 - 2010$. City boundary is chosen at 20% cut-off point of residential land use share (See details in Picard and Tran (2019)). The thresholds for urban green size at 99^{th} percentile is 54.82 hectares. Columns (6), (7), (9) and (10) isolate monocentric cities by excluding the polycentric cities reported by OECD. Columns (1) to (7) include all urban green spaces within urban boundary, while Columns (8) to (10) exclude all big parks above 99^{th} percentile. The Agriculture and Forest Land from the last two annuli control variables are measured within 100-meter buffer around the residential area. Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

Table 4.6: Closest Distance *dmin* and CES Distance d_ψ

	Dependent variable:						
	Ln Closest Dist.	Ln CES Distance d_ψ					
	dmin	$\psi = 2$	$\psi = 5$	$\psi = 10$	$\psi = 15$	$\psi = 20$	$\psi = 50$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln GDP per capita	−0.057	0.483***	0.205***	0.077	0.033	0.010	−0.030
	(0.056)	(0.086)	(0.062)	(0.055)	(0.055)	(0.055)	(0.055)
Distance to CBD (km)	0.022***	0.066***	0.041***	0.032***	0.028***	0.027***	0.024***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Annuli controls							
Agriculture Land (%)	1.785***	0.717***	1.236***	1.514***	1.607***	1.653***	1.733***
	(0.209)	(0.211)	(0.207)	(0.206)	(0.207)	(0.207)	(0.208)
Forest Land (%)	0.991**	−0.568	0.160	0.546	0.687*	0.760**	0.895**
	(0.388)	(0.392)	(0.328)	(0.346)	(0.358)	(0.364)	(0.378)
City Samples	Full	Full	Full	Full	Full	Full	Full
UGS Samples	Full	Full	Full	Full	Full	Full	Full
City Geo. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Controls	NUTS2	NUTS2	NUTS2	NUTS2	NUTS2	NUTS2	NUTS2
Observations	16,782	16,782	16,782	16,782	16,782	16,782	16,782
Adjusted R^2	0.445	0.843	0.611	0.499	0.472	0.461	0.448

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €1,000 per inhabitant. The distance to CBD is measured in kilometre. City geographical controls includes the average Elevation, Latitude and Longitude. The average rain fall, and average temperature in Jan 01 (winter time) and average temperature in July 01 (summer time) for period 1995 − 2010. City boundary is chosen at 20% cut-off point of residential land use share (See details in Picard and Tran (2019)). The table shows results from the regression of number of urban green spaces at each level of distance to city centre across 303 most populous cities in Europe. The Agriculture and Forest Land from the last two annuli control variables are measured within 100-meter buffer around the residential area. Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

includes all parks but assigns higher weights to closer parks as ψ rises. As ψ approaches ∞, the index tends to the value of the distance to closest park *dmin*. The first column in Table 6 reports the coefficient estimates for the measure *dmin* of the distance to the closest park measure *dmin*, while the next six columns report the estimates for the CES distance at various values of ψ. The second line of this table shows the coefficient estimates of the effect of the distance to the city centre. As one can see, the effect of distance to the CBD is positive and significant for all measures of distance between green urban spaces. The distance between parks increases with the distance from the city centre. This confirms and extends our theoretical prediction in the context where residents visit more than one park. Finally, it can be noted that the coefficient estimate on the distance to the city centre decreases in ψ. This stems from the fact that δ^ψ is an increasing function of distance to the city centre. To understand this, consider the situation in which parks are uniformly spread across the city space. Then, residents located at the city centre would have a lower average distance to parks than those dwelling at the city border. Such a difference diminishes with a higher value of ψ as more weight is placed on the (equally distant) closest parks.

4.3.6 Population density

In this subsection, we test the relationship between population density and the system of green urban spaces. Theory tells us that population density increases with larger park sizes and smaller distance to parks, as the latter augment the park amenities and attract more residents. From (4.10), we obtain the following reduced-form regression model:

$$\ln \lambda_{ijc} = \theta_0 + \theta_1 \ln w_{jc} + \theta_2 dist_{ijc} + \theta_3 \ln k_{ijc} + \theta_4 \ln d_{ijc} + \sum_{ijc} X_{ijc}^{(1)} + \sum_{jc} X_{jc}^{(2)} + \sum_{c} I_c + \epsilon_{ijc} \tag{4.13}$$

where k_{ijc} and d_{ijc} are the average of park radius (or park size) and the distance to the closest park in annulus i of city j and region c. $X_{ijc}^{(1)}$, $X_{jc}^{(2)}$ and I_c are the same control variables as before. We expect that $\theta_3 > 0$ and $\theta_4 < 0$. In the theoretical model, we assume that park amenities rise with park size, but residents are more sensitive to the distance to their park ($\delta > \varphi$). If that statement is correct, we should expect the magnitude of the distance effect $|\theta_4|$ to be greater than that of park size $|\theta_3|$. The model also predicts that population density increases with income w and decreases with distance to the city centre; hence, we expect a negative sign for θ_2 and a positive sign for θ_1.

The regression results are reported in Table 4.7 with standard errors clustered at the city level. The effect of the distance to the city centre on population density is robust and significant with all control variables included. Notably, the results in Column (5), which include the whole sample of green urban spaces and all necessary controls and region fixed effects, show that a 1-kilometre increase in distance r to the CBD lessens the population density by 5.4%. This figure rises to 6.2% and 7.3% when we consider only the sub-sample of all monocentric cities[20] and add city size fixed effects as shown in Columns (6) and (7). Those numbers are robust when we exclude all large green urban spaces above the 99^{th} percentile. The coefficients for city income level w are also positive across all regressions with varying levels of controls. Specifically, a 1% increase in income boosts the population density by 0.15-0.35%, ceteris paribus. Overall, the signs for $\hat{\theta}_1$ and $\hat{\theta}_2$ match the prediction from the theoretical model and are robust to different levels of controls.

Table 4.7 also presents an interesting result about the effect of green urban systems on population density. Precisely, the effect of the shortest distance to a green urban space d is significant and economically relevant, whereas the effect of average park

[20] We exclude all polycentric cities reported in the OECD database. The list comprises Praha (CZ), Brno (CZ), Ostrava (CZ), Frankfurt am Main (DE), Stuttgart (DE), Madrid (ES), Barcelona (ES), Bilbao (ES), Paris (FR), Lyon (FR), Lille (FR), Rennes (FR), Napoli (IT), Amsterdam (NL), Eindhoven (NL), Lisboa (PT), Stockholm (SE), Malmo (SE), London (UK), Birmingham (UK), Glasgow (UK), Newcastle upon Tyne (UK) and Portsmouth (UK). Further details are available at `https://stats.oecd.org/Index.aspx?Datasetcode=CITIES`.

Table 4.7: Population Density and Urban Green Stratification (OLS)

	Dependent variable: Ln Population Density									
	Base Regressions					Robustness				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln GDP per capita	0.150**	0.330***	0.192***	0.334***	0.278***	0.349***	0.148	0.278***	0.348***	0.148
	(0.059)	(0.052)	(0.048)	(0.059)	(0.104)	(0.085)	(0.106)	(0.103)	(0.085)	(0.105)
Dist. to CBD (km)	−0.058***	−0.059***	−0.030***	−0.024***	−0.054***	−0.062***	−0.073***	−0.054***	−0.062***	−0.073***
	(0.009)	(0.009)	(0.008)	(0.006)	(0.007)	(0.009)	(0.010)	(0.007)	(0.009)	(0.010)
Ln UGS Radius k	−0.007	0.103**	−0.001	−0.037	−0.029	−0.036	−0.038	−0.032	−0.036	−0.034
	(0.044)	(0.043)	(0.031)	(0.026)	(0.026)	(0.027)	(0.026)	(0.029)	(0.029)	(0.027)
Ln Closest Dist. *dmin*	−0.432***	−0.411***	−0.217***	−0.143***	−0.136***	−0.140***	−0.103***	−0.138***	−0.143***	−0.106***
	(0.048)	(0.046)	(0.035)	(0.028)	(0.024)	(0.027)	(0.027)	(0.024)	(0.027)	(0.027)
Annuli controls										
Agriculture Land (%)			−3.908***	−4.315***	−2.948***	−2.833***	−2.637***	−2.943***	−2.826***	−2.631***
			(0.328)	(0.305)	(0.342)	(0.366)	(0.372)	(0.342)	(0.366)	(0.371)
Forest Land (%)			−5.592***	−4.640***	−4.398***	−3.935***	−3.551***	−4.400***	−3.939***	−3.556***
			(0.748)	(0.659)	(0.558)	(0.573)	(0.666)	(0.557)	(0.573)	(0.665)
City Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Monocentric cities	Monocentric cities	Full samples	Monocentric cities	Monocentric cities
UGS Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.
City Geo. Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Controls	No	No	No	Country	NUTS2	NUTS2	NUTS2 × City Size	NUTS2	NUTS2	NUTS2 × City Size
Observations	16,782	16,782	16,782	16,782	16,782	14,368	14,368	16,780	14,366	14,366
Adjusted R^2	0.290	0.404	0.552	0.679	0.787	0.783	0.794	0.787	0.783	0.794

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €1,000 per inhabitant. The distance to CBD is measured in kilometre. City geographical controls includes the average Elevation, Latitude, Longitude and the average and square of the average of the rain fall, the temperature in summer and winter time. The average rain fall, and average temperature in Jan 01 (winter time) and average temperature in July 01 (summer time) are measured for the period $1995 - 2010$. City boundary is chosen at 20% cut-off point of residential land use share (See details in Picard and Tran (2019)). The thresholds for urban green size at 99^{th} percentile is 54.82 hectares. Columns (6), (7), (9) and (10) isolate monocentric cities by excluding the polycentric cities reported by OECD. Columns (1) to (7) include all urban green spaces within urban boundary, while Columns (8) to (10) exclude all big parks above 99^{th} percentile. The Agriculture and Forest Land from the last two annuli control variables are measured within 100-meter buffer around the residential area. Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

size k is much smaller and, in addition, non-significant (see, for instance, Column (5)). Such figures maintain a similar magnitude and significance when we restrict the data to monocentric cities and add an additional control for city size. This confirms our theoretical assumption that *residents are more sensitive to park distance than park size.* It even suggests the stronger argument that *residents are not sensitive to park size* (i.e., $\varphi = 0$). Only park access seems to matter to them.

This analysis raises the question of whether residents are more sensitive to the closest park rather than to the set of parks surrounding their dwellings. To explore this issue, we re-run the regressions with the distance to the closest park *dmin* and the CES distances d^{ψ} with $\psi \in \{2, 5, 10, 15, 20, 50\}$. Intuitively, for small ψ, the CES distance d^{ψ} averages the distances to many far surrounding parks. For higher ψ, it places more weight on closer parks and therefore averages the distances to a subset of closer parks. For $\psi = \infty$, it returns the distance to the closest park. If the effect of distance becomes more significant when ψ increases, this means that individuals are more concerned with the parks closer to their residences. The results are displayed in Table 4.8. First, note that the effect on income and distance to the city centre are unaffected by the measure of distance to parks. The effect of city income w and distance to the CBD r remain significant with different measures of distance *dmin* and d^{ψ}. The coefficient on park size $\hat{\theta}_3$ remains non-significant and has a smaller magnitude than the coefficient on park distance $\hat{\theta}_4$. Finally, the significance level of $\hat{\theta}_4$ increases as ψ increases. Furthermore, the R^2 value is slightly higher for higher ψ. *Such results suggest that households care more about the nearest park to their homes than the distance to other nearby parks.*

Finally, the regression model (4.13) might be subject to endogeneity issues, as there could exist confounding omitted variables that affect both the dependent variable (population density) and the independent variables (park size and distance to closest park). Therefore, Table 4.7 may present biased estimates of the coefficients $\hat{\theta}_3$ and $\hat{\theta}_4$. To check for this omitted variable issue, we first extract the residuals of the density regression model (4.13) and then run the park size and distance-to-closest-park regression models (4.11) and (4.12) augmented with a linear term including the above residuals and having coefficients η_3 and μ_3. In the case of an omitted variable, the estimated coefficients $\widehat{\eta}_3$ and $\widehat{\mu}_3$ should be significantly different from zero (see Appendix C for further details). We also apply this procedure to the regression on the number of parks to ensure that the regression model (4.13) is also robust to the inclusion of this variable (see the estimated coefficient $\widehat{\rho}_4$).

The estimation results are reported in Table 4.9. The controls are the same as in previous tables. We report results for the samples with all cities and only monocentric cities. It can be observed that the estimated coefficients $\widehat{\eta}_3$, $\widehat{\mu}_3$ and $\widehat{\rho}_4$ are no longer significant once we introduce the city controls for geography and annulus controls (proximity of agriculture and forest land). In the presence of those controls, there is no suspicion of omitted variable problems. Moreover, the magnitude of those coeffi-

Table 4.8: Population Density and Urban Green Stratification with Closest Distance and CES Distance (OLS)

	Dependent variable: Ln Population Density						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln GDP per capita	0.278***	0.268***	0.302***	0.294***	0.290***	0.287***	0.282***
	(0.104)	(0.103)	(0.106)	(0.105)	(0.105)	(0.104)	(0.104)
Dist. to CBD (km)	−0.054***	−0.059***	−0.054***	−0.054***	−0.054***	−0.054***	−0.054***
	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
Ln UGS Radius k	−0.029	−0.056**	−0.039	−0.034	−0.032	−0.031	−0.030
	(0.026)	(0.027)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Ln Closest Dist. $dmin$	−0.136***						
	(0.024)						
Ln CES Dist. d_ψ							
$\psi = 2$		0.040					
		(0.038)					
$\psi = 5$			−0.078***				
			(0.027)				
$\psi = 10$				−0.109***			
				(0.025)			
$\psi = 15$					−0.118***		
					(0.025)		
$\psi = 20$						−0.123***	
						(0.025)	
$\psi = 50$							−0.131***
							(0.024)
City Samples	Full	Full	Full	Full	Full	Full	Full
UGS Samples	Full	Full	Full	Full	Full	Full	Full
City Geo. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Annuli controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Controls	NUTS2	NUTS2	NUTS2	NUTS2	NUTS2	NUTS2	NUTS2
Observations	16,782	16,782	16,782	16,782	16,782	16,782	16,782
Adjusted R^2	0.787	0.783	0.784	0.785	0.786	0.786	0.786

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €1,000 per inhabitant. The distance to CBD is measured in kilometre. City geographical controls includes the average Elevation, Latitude and Longitude. The average rain fall, and average temperature in Jan 01 (winter time) and average temperature in July 01 (summer time) for period $1995 - 2010$. City boundary is chosen at 20% cut-off point of residential land use share (See details in Picard and Tran (2019)). The Agriculture and Forest Land from the last two annuli control variables are measured within 100-meter buffer around the residential area. Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

Table 4.9: Check for Endogeneity: estimation of ρ_4, η_3, and μ_3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\rho}_4$	0.162***	0.259***	0.095	0.048	0.078	0.075	−0.003
	(0.055)	(0.057)	(0.059)	(0.052)	(0.052)	(0.061)	(0.057)
$\hat{\eta}_3$	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.019)	(0.019)	(0.018)	(0.019)	(0.022)	(0.023)	(0.023)
$\hat{\mu}_3$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.023)	(0.025)	(0.024)	(0.023)	(0.025)	(0.030)	(0.030)
City Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Monocentric cities	Monocentric cities
UGS Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples
City Geo. Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Annuli controls	No	No	Yes	Yes	Yes	Yes	Yes
FE Controls	No	No	No	Country	NUTS2	NUTS2	NUTS2 × City Size

Note: Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by *p<0.1, **p<0.05, ***p<0.01.

cients diminishes with the introduction of various fixed effects and the restriction of the data to monocentric cities. This point strengthens our belief in the absence of omitted variables.

4.4 Conclusion

Urban land is a scarce resource that is highly contested for different kinds of uses. In particular, green urban spaces compete with residential properties to which they bring positive amenities. Therefore, the choice of the location and size of green urban spaces is a challenging task for city governments. This paper sheds light on this choice from a theoretical and empirical perspective.

The paper shows that the optimal number of green urban spaces is not uniform across city locations. The number of parks first increases at small distances from the city centre and then decreases at large distances from the centre, forming a bell-shaped pattern in large cities. Two channels affect the number of green urban areas: on the one hand, at short distances from the city centre, the limited land availability constrains the creation of green urban areas; on the other hand, in the city suburbs, the low population density diminishes the social benefit and therefore the desirability of creating parks. By contrast, the optimal park sizes and distances to other parks are shown to be monotone increasing functions of the distance to the city centre. The main contribution of this paper is to empirically confirm those relationships with the high-resolution geographical database of GMES Urban Atlas. From this dataset, we find that the number of green urban areas is bell-shaped or concave in city size. Our empirical analysis also reveals that a 1-kilometre increase in distance to the city centre is associated with increases of 2.0% and 2.7% in average park size and average distance to the closest park, respectively.

While the empirical analysis fits well the properties of the distribution of green urban spaces with respect to their distance to the city centre, it presents a weaker match with respect to the productivity of cities. In theory, more productive cities are expected to include more numerous parks with smaller sizes and longer access distances. The empirical analysis finds support for the effect of higher productivity cities on the number of parks but no support for its effect on park sizes or the distance between parks.

In addition, population density is theoretically expected to rise with the presence of larger and closer parks because residents are expected to prefer larger and closer parks. The data suggest a strong effect of park access but no effect of park size on population density. This suggests that residents are much more sensitive to the proximity than the size of green urban areas.

Appendix A: Cobb-Douglas utility

In this appendix we prove (4.3) and (4.4). We recall that $U(z,s,a) = a^{\alpha} z^{1-\beta} s^{\beta}$, $a(k,d) = a_0 k^{\varphi} d^{-\delta}$ and $C(k) = c_0 k^{\gamma}$.

Under this setting we have $a_k = \frac{\varphi}{k} a$ and $\hat{V}_a(r,a,u) = \frac{\alpha}{\beta}\frac{1}{a}\hat{V}(r,a,u)$ while $\hat{V}(r,a(k,d),u) = \hat{V}\left(r,\beta k^{\varphi} d^{-\delta},u\right) = \hat{V}\left(r,\beta k^{\varphi}(r\phi)^{-\delta},u\right) = (\phi)^{-\delta\frac{\alpha}{\beta}}\left(k^{\varphi} r^{-\delta}\right)^{\frac{\alpha}{\beta}}\hat{V}(r,a_0,u)$ where ϕ is the angle of the polar coordinate of the individual at (r,ϕ). So, we also compute

$$\begin{aligned}\int_{k/r}^{\pi/n} \hat{V}\left[r,a\left(k,r\phi\right),u\right] r\mathrm{d}\phi &= \hat{V}\left(r,a_0 k^{\varphi} r^{-\delta},u\right) r\int_{k/r}^{\pi/n} (\phi)^{-\delta\frac{\alpha}{\beta}}\,\mathrm{d}\phi \\ &= r(k^{\varphi} r^{-\delta})^{\frac{\alpha}{\beta}}\hat{V}(r,a_0,u)\,\frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}}\end{aligned}$$

$$\begin{aligned}\int_{k/r}^{\pi/n} \hat{V}_a\left[r,a\left(k,r\phi\right),u\right] a_k\left[k,r\phi\right] r\mathrm{d}\phi &= \int_{k/r}^{\pi/n} \frac{\alpha}{\beta}\hat{V}\left[r,a\left(k,r\phi\right),u\right]\frac{\varphi}{k}\, r\mathrm{d}\phi \\ &= \frac{\alpha}{\beta}\frac{\varphi r}{k}(k^{\varphi} r^{-\delta})^{\frac{\alpha}{\beta}}\hat{V}(r,a_0,u)\,\frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}}\end{aligned}$$

First order conditions

We can check conditions (4.1) and (4.2). We successively get

$$\begin{aligned}
\mathcal{S}_k{=}0 &\iff \frac{\alpha}{\beta}\frac{\varphi r}{k}(k^{\varphi} r^{-\delta})^{\frac{\alpha}{\beta}}\hat{V}(r,a_0,u)\,\frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}} - \hat{V}\left[r,a\left(k,k\right),u\right] - c_0\gamma k^{\gamma-1} = 0 \\
&\iff \varphi\frac{\alpha}{\beta}\frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}}\left(\frac{k}{r}\right)^{\delta\frac{\alpha}{\beta}-1}(k^{\varphi-\delta})^{\frac{\alpha}{\beta}}\hat{V}(r,a_0,u) - (k^{\varphi-\delta})^{\frac{\alpha}{\beta}}\hat{V}(r,a_0,u) - c_0\gamma k^{\gamma-1} = 0 \\
&\iff \left[\varphi\frac{\alpha}{\beta}\frac{\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} - 1}{1-\delta\frac{\alpha}{\beta}} - 1\right] k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)} = \frac{c_0\gamma}{\hat{V}(r,a_0,u)} \\
&\iff \left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = \frac{1-\delta\frac{\alpha}{\beta}}{\varphi\frac{\alpha}{\beta}}\left(\frac{c_0\gamma}{\hat{V}(r,a_0,u)}\frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} + 1\right) + 1 \\
&\iff \left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = 1 + \frac{1-\delta\frac{\alpha}{\beta}}{\varphi\frac{\alpha}{\beta}} + \frac{1-\delta\frac{\alpha}{\beta}}{\varphi\frac{\alpha}{\beta}}\frac{c_0\gamma}{\hat{V}(r,a_0,u)}\frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} \\
&\iff \left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = 1 + \frac{1}{\varphi}\left(\frac{\beta}{\alpha}-\delta\right) + \frac{1}{\varphi}\left(\frac{\beta}{\alpha}-\delta\right)\frac{c_0\gamma}{\hat{V}(r,a_0,u)}\frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}}
\end{aligned}$$

and

$$
\begin{aligned}
\mathcal{S}_n = 0 \iff & 2\left\{\int_{k(r)/r}^{\pi/n(r)} \hat{V}\left[r, a\left(k(r), r\phi\right), u\right] r \mathrm{d}\phi - \int_0^{\pi/n(r)} R_A r \mathrm{d}\phi - C\left[k(r)\right]\right\} \\
& - \frac{2\pi r}{n(r)}\left\{\hat{V}\left[r, a\left(k(r), r\pi/n(r)\right), u\right] - R_A\right\} = 0 \\
\iff & 2\left\{r(k^{\varphi} r^{-\delta})^{\frac{\alpha}{\beta}} \hat{V}\left(r, a_0, u\right) \frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}} - R_A \frac{\pi r}{n} - c_0 k^{\gamma}\right\} - \\
& \frac{2\pi r}{n(r)}\left\{\hat{V}\left(r, a_0 k^{\varphi}\left(\frac{\pi r}{n}\right)^{-\delta}, u\right) - R_A\right\} = 0 \\
\iff & r(k^{\varphi} r^{-\delta})^{\frac{\alpha}{\beta}} \hat{V}\left(r, a_0, u\right) \frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}} - c_0 k^{\gamma} - \frac{\pi r}{n(r)} \hat{V}\left(r, a_0 k^{\varphi}\left(\frac{\pi r}{n}\right)^{-\delta}, u\right) = 0 \\
\iff & r\left(\frac{k}{r}\right)^{\delta\frac{\alpha}{\beta}} k^{(\varphi-\delta)\frac{\alpha}{\beta}} \hat{V}\left(r, a_0, u\right) \frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}} - c_0 k^{\gamma} \\
& - \frac{\pi r}{n}\left(\frac{nk}{\pi r}\right)^{\delta\frac{\alpha}{\beta}} k^{(\varphi-\delta)\frac{\alpha}{\beta}} \hat{V}\left(r, a_0, u\right) = 0 \\
\iff & \left[k\left(\frac{k}{r}\right)^{\delta\frac{\alpha}{\beta}-1} \frac{(\pi/n)^{1-\delta\frac{\alpha}{\beta}} - (k/r)^{1-\delta\frac{\alpha}{\beta}}}{1-\delta\frac{\alpha}{\beta}} - \left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} k\right] k^{(\varphi-\delta)\frac{\alpha}{\beta}} \hat{V}\left(r, a_0, u\right) - c_0 k^{\gamma} = 0 \\
\iff & k\left(\left[\frac{\delta\frac{\alpha}{\beta}\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} - 1}{1-\delta\frac{\alpha}{\beta}}\right] k^{(\varphi-\delta)\frac{\alpha}{\beta}} \hat{V}\left(r, a_0, u\right) - c_0 k^{\gamma-1}\right) = 0 \\
\iff & \left[\frac{\delta\frac{\alpha}{\beta}\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} - 1}{1-\delta\frac{\alpha}{\beta}}\right] k^{(\varphi-\delta)\frac{\alpha}{\beta}} \hat{V}\left(r, a_0, u\right) - c_0 k^{\gamma-1} = 0 \text{ if } k > 0 \\
\iff & \left[\frac{\delta\frac{\alpha}{\beta}\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} - 1}{1-\delta\frac{\alpha}{\beta}}\right] k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)} = \frac{c_0}{\hat{V}\left(r, a_0, u\right)} \text{ if } k > 0 \\
\iff & \delta\frac{\alpha}{\beta}\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = 1 + \left(1-\delta\frac{\alpha}{\beta}\right) \frac{c_0}{\hat{V}\left(r, a_0, u\right)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} \text{ if } k > 0 \\
\iff & \left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = \frac{1}{\delta\frac{\alpha}{\beta}} + \frac{\left(1-\delta\frac{\alpha}{\beta}\right)}{\delta\frac{\alpha}{\beta}} \frac{c_0}{\hat{V}\left(r, a_0, u\right)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} \text{ if } k > 0 \\
\iff & \left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right) \frac{c_0}{\hat{V}\left(r, a_0, u\right)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} \text{ if } k > 0
\end{aligned}
$$

and for $k > 0$. The last equality gives an inverse relationship between n and k.

Suppose for moment park amenities are proportional to park sizes and distances and maintenance cost is linear ($\varphi = \delta = \gamma = 1$) . Then, the two conditions become co-linear and give

$$
\frac{nk}{\pi r} = \left[\frac{\beta}{\alpha} + \left(\frac{\beta}{\alpha} - 1\right) \frac{c_0}{\hat{V}\left(r, a_0, u\right)}\right]^{-\frac{\beta}{\beta-\alpha}}
$$

which is smaller than 1 as $\beta > \alpha$. It's true as

$$\beta > \alpha \Longrightarrow \frac{\beta}{\alpha} + \left(\frac{\beta}{\alpha} - 1\right) \frac{c_0}{\hat{V}(r, a_0, u)} > 1$$

$$\beta > \alpha \Longrightarrow -\frac{\beta}{\beta - \alpha} < 0$$

$$\text{Then, } \left[\frac{\beta}{\alpha} + \left(\frac{\beta}{\alpha} - 1\right) \frac{c_0}{\hat{V}(r, a_0, u)}\right]^{-\frac{\beta}{\beta-\alpha}} < 1$$

Now, suppose that $\delta > \varphi$ and $\gamma < 1$, plugging condition (1) into condition (2), we have

$$1 + \frac{1}{\varphi}\left(\frac{\beta}{\alpha} - \delta\right) + \frac{1}{\varphi}\left(\frac{\beta}{\alpha} - \delta\right) \frac{c_0\gamma}{\hat{V}(r, a_0, u)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} = \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right) \frac{c_0}{\hat{V}(r, a_0, u)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}}$$

$$\iff \left(1 + \frac{1}{\varphi}\left(\frac{\beta}{\alpha} - \delta\right) - \frac{\beta}{\delta\alpha}\right) - \left[\left(\frac{\beta}{\delta\alpha} - 1\right) - \left(\frac{\beta}{\alpha} - \delta\right)\frac{\gamma}{\varphi}\right] \frac{c_0}{\hat{V}(r, a_0, u)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} = 0$$

$$\iff \left(1 + \frac{1}{\varphi}\left(\frac{\beta}{\alpha} - \delta\right) - \frac{\beta}{\delta\alpha}\right) = \left[\left(\frac{\beta}{\delta\alpha} - 1\right) - \left(\frac{\beta}{\alpha} - \delta\right)\frac{\gamma}{\varphi}\right] \frac{c_0}{\hat{V}(r, a_0, u)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}}$$

$$\iff \frac{c_0}{\hat{V}(r, a_0, u)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}} = \frac{\left(1 + \frac{1}{\varphi}\left(\frac{\beta}{\alpha} - \delta\right) - \frac{\beta}{\delta\alpha}\right)}{\left(\frac{\beta}{\delta\alpha} - 1\right) - \left(\frac{\beta}{\alpha} - \delta\right)\frac{\gamma}{\varphi}}$$

$$\iff k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)} = \frac{\left(\frac{\beta}{\delta\alpha} - 1\right) - \left(\frac{\beta}{\alpha} - \delta\right)\frac{\gamma}{\varphi}}{\left(1 + \frac{1}{\varphi}\left(\frac{\beta}{\alpha} - \delta\right) - \frac{\beta}{\delta\alpha}\right)} \frac{c_0}{\hat{V}(r, a_0, u)}$$

$$\iff k^{(\varphi-\gamma)+(1-\delta)\frac{\alpha}{\beta}} = \frac{\left(\frac{\beta}{\delta\alpha} - 1\right)\left(1 - \frac{\gamma\delta}{\varphi}\right)}{\left(\frac{\beta}{\delta\alpha} - 1\right)\left(\frac{\delta}{\varphi} - 1\right)} \frac{c_0}{\hat{V}(r, a_0, u)}$$

$$\iff k = \left(\frac{1 - \frac{\gamma\delta}{\varphi}}{\frac{\delta}{\varphi} - 1} \frac{c_0}{\hat{V}(r, a_0, u)}\right)^{\frac{1}{(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta}}} \text{ if } \left(1 - \delta\frac{\alpha}{\beta}\right) \neq 0$$

and

$$\left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right) \frac{c_0}{\hat{V}(r, a_0, u)} \frac{1}{k^{(\varphi-\delta)\frac{\alpha}{\beta}-(\gamma-1)}}$$

$$\iff \left(\frac{\pi r}{nk}\right)^{1-\delta\frac{\alpha}{\beta}} = \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right) \frac{\frac{\delta}{\varphi} - 1}{\frac{1}{\varphi} - \gamma\delta}$$

$$\iff n = \frac{\pi r}{k}\left(\frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1\right) \frac{\frac{\delta}{\varphi} - 1}{\frac{1}{\varphi} - \gamma\delta}\right)^{-\frac{1}{1-\delta\frac{\alpha}{\beta}}}$$

Therefore, when $\left(1 - \delta\frac{\alpha}{\beta}\right) \neq 0$ we have the interior solution for n and k as

$$k^* = \left(\frac{\zeta c_0}{\widehat{V}\left(r, a_0, u\right)} \right)^{\frac{1}{(1-\gamma)+(1-\delta)\frac{\alpha}{\beta}}}$$
$$n^* = \frac{\kappa \pi r}{k^*}$$

where

$$\zeta \equiv \frac{1 - \frac{\gamma\delta}{\varphi}}{\frac{\delta}{\varphi} - 1}$$
$$\kappa \equiv \left(\frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1 \right) \frac{1}{\zeta} \right)^{-\frac{1}{1-\delta\frac{\alpha}{\beta}}}$$

When $\left(1 - \delta\frac{\alpha}{\beta}\right) = 0$ we can rewrite the condition (1) and (2) as follows:

$$\mathcal{S}_k = 0 \iff \ln\left(\frac{\pi}{n}\right) - \ln\left(\frac{k}{r}\right) = \frac{\beta}{\alpha} + \frac{\beta}{\alpha}\frac{c_0\gamma}{\hat{V}\left(r, a_0, u\right)}\frac{1}{k^{\varphi\frac{\alpha}{\beta}-\gamma}}$$
$$\mathcal{S}_n = 0 \iff \ln\left(\frac{\pi}{n}\right) - \ln\left(\frac{k}{r}\right) = 1 + \frac{c_0}{\hat{V}\left(r, a_0, u\right)}\frac{1}{k^{\varphi\frac{\alpha}{\beta}-\gamma}}$$

if $k > 0$. Replacing condition (1) into condition (2), we have the solution for n and k

$$k^* = \left(\frac{\zeta c_0}{\hat{V}\left(r, a_0, u\right)} \right)^{\frac{1}{\frac{\varphi}{\delta}-\gamma}}$$
$$n^* = \frac{1}{\exp\left(1 + \frac{1}{\zeta}\right)}\frac{\pi r}{k^*}$$

Interior conditions

In order to have interior solutions, the two follow conditions need to satisfy

$$\text{(i)} \quad \zeta \equiv \frac{1 - \frac{\gamma\delta}{\varphi}}{\frac{\delta}{\varphi} - 1} > 0$$
$$\text{(ii)} \quad \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha} - 1 \right) \frac{1}{\zeta} > 0$$

$$\text{Condition (i)} \Longleftrightarrow \frac{1-\frac{\gamma\delta}{\varphi}}{\frac{\delta}{\varphi}-1} > 0 \Longleftrightarrow \gamma < \frac{\varphi}{\delta} \text{ as } \delta > \varphi$$

$$\begin{aligned}
\text{Condition (ii)} &\Longleftrightarrow \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} > 0 \\
&\Longleftrightarrow \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1-\frac{\gamma\delta}{\varphi}}{\frac{\delta}{\varphi}-1} > 0 \\
&\Longleftrightarrow \frac{\beta}{\delta\alpha}\left(1-\frac{\gamma\delta}{\varphi}\right) + \left(\frac{\beta}{\delta\alpha}-1\right)\left(\frac{\delta}{\varphi}-1\right) > 0 \text{ as } 1-\frac{\gamma\delta}{\varphi} > 0 \\
&\Longleftrightarrow \left(\frac{\beta}{\delta\alpha} - \gamma\frac{\beta}{\varphi\alpha}\right) + \left(\frac{\beta}{\varphi\alpha} - \frac{\delta}{\varphi} - \frac{\beta}{\delta\alpha} + 1\right) > 0 \\
&\Longleftrightarrow \frac{\beta}{\varphi\alpha} - \frac{\delta}{\varphi} + 1 > \gamma\frac{\beta}{\varphi\alpha} \\
&\Longleftrightarrow 1 - \frac{\delta\alpha}{\beta} + \frac{\varphi\alpha}{\beta} > \gamma \\
&\Longleftrightarrow 1 + (\varphi-\delta)\frac{\alpha}{\beta} > \gamma
\end{aligned}$$

It can be combined as $\gamma < \min\{\frac{\varphi}{\delta}; 1+(\varphi-\delta)\frac{\alpha}{\beta}\}$.

Now we need to prove that under interior condition above, the share of urban green in each annulus is smaller than one. More specifically, we need to prove that

$$\frac{n^*k^*}{\pi r} = \kappa \equiv \left(\frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta}\right)^{-\frac{1}{1-\delta\frac{\alpha}{\beta}}} < 1$$

We consider two case:

Case 1: $\delta < \frac{\beta}{\alpha}$ which is automatically satisfy condition (ii) as $\left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} > 0$

We have

$$\begin{aligned}
&\delta < \frac{\beta}{\alpha} \Longrightarrow 1-\delta\frac{\alpha}{\beta} > 0 \Longrightarrow -\frac{1}{1-\delta\frac{\alpha}{\beta}} < 0 \\
&\delta < \frac{\beta}{\alpha} \Longrightarrow \frac{\beta}{\delta\alpha} > 1 \text{ and } \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} > 0 \text{ as } \zeta > 0 \\
&\quad\Longrightarrow \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} > 1 \text{ and } -\frac{1}{1-\delta\frac{\alpha}{\beta}} < 0 \\
&\quad\Longrightarrow \kappa \equiv \left(\frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta}\right)^{-\frac{1}{1-\delta\frac{\alpha}{\beta}}} < 1
\end{aligned}$$

Case 2: $\delta > \frac{\beta}{\alpha}$

We have

$$\begin{aligned}
\delta > \frac{\beta}{\alpha} &\Longrightarrow 1-\delta\frac{\alpha}{\beta} < 0 \Longrightarrow -\frac{1}{1-\delta\frac{\alpha}{\beta}} > 0 \\
\delta > \frac{\beta}{\alpha} &\Longrightarrow \frac{\beta}{\delta\alpha} < 1 \text{ and } \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} < 0 \text{ as } \zeta > 0 \\
&\Longrightarrow \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} < 1 \\
\text{Condition (ii)} &\Longrightarrow \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} > 0 \\
&\Longrightarrow 0 < \frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta} < 1 \text{ and } -\frac{1}{1-\delta\frac{\alpha}{\beta}} > 0 \\
&\Longrightarrow \kappa \equiv \left(\frac{\beta}{\delta\alpha} + \left(\frac{\beta}{\delta\alpha}-1\right)\frac{1}{\zeta}\right)^{-\frac{1}{1-\delta\frac{\alpha}{\beta}}} < 1
\end{aligned}$$

Case 3: $\delta = \frac{\beta}{\alpha}$ we are back to the special case where:

$$\begin{aligned}
k^* &= \left(\frac{\zeta c_0}{\hat{V}(r, a_0, u)}\right)^{\frac{1}{\frac{\varphi}{\delta}-\gamma}} \\
n^* &= \frac{1}{\exp\left(1+\frac{1}{\zeta}\right)}\frac{\pi r}{k}
\end{aligned}$$

then the share of green in annulus at distance r is $\frac{nk}{\pi r} = \frac{1}{\exp\left(1+\frac{1}{\zeta}\right)} < 1$.

Comparative statics

Under interior condition (i) and (ii), we have

$$\begin{aligned}
k^* &= \left(\frac{\zeta c_0}{\hat{V}(r, a_0, u)}\right)^{\frac{1}{(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta}}} \\
n^* &= \frac{\kappa\pi r}{k^*}
\end{aligned}$$

We have

$$\begin{aligned}
k_r^* &= \frac{1}{(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta}}\left(\frac{\zeta c_0}{\hat{V}(r, a_0, u)}\right)^{\frac{1}{(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta}}-1}\frac{-\zeta c\hat{V}_r}{\left(\hat{V}(r, a_0, u)\right)^2} \\
&= \frac{1}{(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta}}\frac{-k^*\hat{V}_r}{\hat{V}(r, a_0, u)} > 0
\end{aligned}$$

As $(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta} > 0$ and $\hat{V}_r < 0$

Note that:

$$\begin{aligned}(1-\gamma)+(\varphi-\delta)\frac{\alpha}{\beta}>0 &\iff (1-\gamma)>-(\varphi-\delta)\frac{\alpha}{\beta}\\ &\iff \gamma<1+(\varphi-\delta)\frac{\alpha}{\beta}\\ &\iff \gamma<1+\varphi\frac{\alpha}{\beta}-\delta\frac{\alpha}{\beta}\end{aligned}$$

which is always true under Interior Condition (ii)

Appendix B: Quadratic regression robustness check

In this section, we use quadratic form of distance to city centre to test for the hypothesis that the number of urban green spaces is first increasing with small value of distance and then decreases as one moves away from the center. Specifically, we test the following reduced form regression

$$n_{ijc} = \alpha_1 dist_{ijc} + \alpha_2 dist_{ijc}^2 + \sum_{ijc} X_{ijc}^{(1)} + \sum_{jc} X_{jc}^{(2)} + \sum_{c} I_c + \epsilon_{ijc}$$

where (i, j, c) is an observation of annuli at annulus i of city j of region c; $X_{ijc}^{(1)}$ and $X_{jc}^{(2)}$ are control variables for annuli and city observable characteristics; I_c is region dummy variable controlled for unobservable characteristics like environmental policies or local regulation. Theoretical model predicts the number of parks first increases and then decreases as one move farther from city centre; hence, the sign of α_1 is expected to be positive while α_2 is expected to be negative.

The regression results are reported in Table B1. The standard errors are clustered at city level, which controls for the possible correlation between observations within one urban area. The regression shows a negative and significant coefficient for the square of the distance to CBD, suggesting the non-monotonic relationship between numbers of urban green spaces and the distance to CBD. It also shows a positive and significant coefficient for $\hat{\alpha_1}$. It means that the number of parks first increases and then decreases with distance to city centers. Column (1) indicates the results without controlling for city geographic and observable characteristics and without city fixed effect. As it can be seen, the magnitude increases when we introduce more controls, but the sign ad significant level are not affected. Column (3) to (10) further introduce the control for other type of open spaces like Agriculture and Forest Land available within 100-meter buffer from residential areas in each annuli. The magnitude for $\hat{\alpha_1}$ and $\hat{\alpha_2}$ increase nearly two folds from -0.083 to -0.153 for the square of the distance to city center $\hat{\alpha}_2$, and from 2.8 to 4.5 for distance to CBD $\hat{\alpha}_1$ with same level of significance. The higher presence of agriculture and forest land associate with lower number of urban green spaces. Specifically, a 1% increase in agriculture land associates with a reduction of 60-74 parks, and the magnitude for forest land is similar at 39-104 parks reduction for 1% increase in forest land within 100-meter buffer from residential area. Column (8) to (10) further exclude big parks at 99^{th} percentile threshold, and the sign and magnitude are still similar in those cases.

Table B1: Number of Urban Green Spaces (Quadratic OLS Regression)

	Dependent variable: Ln Number of Urban Green Spaces									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dist. to CBD (km)	3.190***	2.856***	4.551***	4.753***	4.800***	5.138***	4.581***	4.701***	5.043***	4.491***
	(0.437)	(0.400)	(0.509)	(0.470)	(0.412)	(0.434)	(0.471)	(0.411)	(0.431)	(0.467)
Dist. to CBD square	−0.098***	−0.083**	−0.152***	−0.154***	−0.188***	−0.216***	−0.210***	−0.183***	−0.212***	−0.206***
	(0.036)	(0.034)	(0.039)	(0.035)	(0.024)	(0.030)	(0.030)	(0.024)	(0.030)	(0.029)
GDP per capita		0.212***	0.112***	0.124***	−0.071	0.095	−0.101	−0.060	0.096	−0.101
		(0.039)	(0.040)	(0.041)	(0.053)	(0.087)	(0.108)	(0.051)	(0.086)	(0.106)
Annuli controls										
Agri. Land (%)			−74.079***	−74.436***	−75.091***	−71.006***	−60.495***	−73.499***	−69.756***	−59.356***
			(9.486)	(9.528)	(7.766)	(8.035)	(7.541)	(7.659)	(7.924)	(7.431)
Forest Land (%)			−104.275***	−95.416***	−61.227***	−71.406***	−39.950***	−60.126***	−70.414***	−39.194***
			(15.831)	(14.642)	(16.677)	(17.293)	(12.261)	(16.583)	(17.202)	(12.257)
City Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Monocentric cities	Monocentric cities	Full samples	Monocentric cities	Monocentric cities
UGS Samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Full samples	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.	Excld. UGS abv. 99^{th} per.
City Geo. Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Controls	No	No	No	Country	NUTS2	NUTS2	NUTS2 × City Size	NUTS2	NUTS2	NUTS2 × City Size
Observations	16,782	16,782	16,782	16,782	16,782	14,368	14,368	16,780	14,366	14,366
Adjusted R^2	0.239	0.325	0.421	0.486	0.671	0.615	0.656	0.665	0.611	0.653

Note: Here, we use the GDP per capita taken from Regional Economic Accounts from Eurostat at NUTS3 level with adjustment to purchasing power standard (PPS) as the proxy for city income level, and it is measured in €1,000 per inhabitant. The distance to CBD is measured in kilometre. City geographical controls includes the average Elevation, Latitude, Longitude and the average and square of the average of the rain fall, the temperature in summer and winter time. The average rain fall, and average temperature in Jan 01 (winter time) and average temperature in July 01 (summer time) are measured for the period $1995 - 2010$. City boundary is chosen at 20% cut-off point of residential land use share (See details in Picard and Tran (2019)). The thresholds for urban green size at 99^{th} percentile is 54.82 hectares. Columns (6), (7), (9) and (10) isolate monocentric cities by excluding the polycentric cities reported by OECD. Columns (1) to (7) include all urban green spaces within urban boundary, while Columns (8) to (10) exclude all big parks above 99^{th} percentile. The Agriculture and Forest Land from the last two annuli control variables are measured within 100-meter buffer around the residential area. Standard errors are clustered at city level and reported in parenthesis. The significance levels are denoted by $^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

Appendix C: Robustness Check - Test for endogeneity

Note that in the theoretical model, the urban planner makes decisions in two steps. First, given the configuration of urban green spaces, she assigns the consumption of composite goods, housing slot sizes, and at the same time, the population density. Second, when we know the change in population density and housing slot size at each configuration of urban green, she decides the optimal locations. We assume there is a heterogeneous variable that we could not observe from the data, but it affects the household decision. Specifically, we assume

$$U(z,s,a) = \xi a^{\alpha} z^{1-\beta} s^{\beta}$$

where ξ follow some distribution with mean one. Using the same procedure as in the Cobb-Douglas Utility section, we have

$$\lambda = \frac{1}{s} = \frac{\xi^{\frac{1}{\beta}}}{(1-\beta)^{-\frac{1-\beta}{\beta}} y^{-\frac{1-\beta}{\beta}} a^{-\frac{\alpha}{\beta}} u^{\frac{1}{\beta}}}$$

Here, we observe and infer λ, w, k, and d from our database, and from each observation of λ, we can infer ξ as the residuals (which we affects the population density but are not explained by urban green space systems and other observable controls). Basically, we have the log-linear equation and the reduced-form equation as follows:

$$\ln \lambda = \ln\left[(1-\beta)^{\frac{1-\beta}{\beta}} a_0^{\frac{\alpha}{\beta}}\right] + \frac{1-\beta}{\beta}\ln w - \frac{1-\beta}{\beta} tr + \frac{\alpha}{\beta}\ln k - \delta\frac{\alpha}{\beta}\ln d - \frac{1}{\beta}\ln u + \frac{1}{\beta}\ln \xi$$

$$\ln \lambda_{rfc} = \theta_0 + \theta_1 \ln w_{fc} + \theta_2 dist_{rfc} + \theta_3 \ln k_{rfc} + \theta_4 \ln d_{rfc} + \sum_{rfc} X^{(1)}_{rfc} + \sum_{fc} X^{(2)}_{fc} + \sum_{fc} I_{fc} + \epsilon_{rfc}$$

Then, we have

$$\hat{\epsilon}_{rfc} \equiv -\frac{1}{\beta}\ln \xi_{rfc} \Longrightarrow \hat{\xi}_{rfc} = e^{-\beta \hat{\epsilon}_{rfc}}$$

If the distribution of error terms approximate the normal distribution with mean zero, then we also have the distribution of ξ with a mean close to 1.

Now, we also have

$$k = \left[\frac{\zeta c_0}{V^*(r, a_0, u, \xi)}\right]^{\frac{1}{(1-\gamma)+(1-\delta)\frac{\alpha}{\beta}}}$$

$$\ln k = \frac{1}{(1-\gamma)+(1-\delta)\frac{\alpha}{\beta}}\left(\ln\left(\frac{\zeta c_0}{\beta(1-\beta)^{\frac{1-\beta}{\beta}}}\right) - \frac{1}{\beta}\ln w + \frac{1}{\beta} tr - \ln\left(u^{-\frac{1}{\beta}} a_0^{\frac{\alpha}{\beta}}\right) - \frac{1}{\beta}\ln \xi\right)$$

Using the value of $\hat{\hat{\xi}}_{rfc}$ above, we can add them into the regression for the size of urban green space and the distance *dmin* and d_ϕ. By that way, we can control for other types of controls that drive the population density but not from the green systems and controlled variables. As the decision of urban planner in our model channels through the land value and population density, if we can examine here, how other amenities or location characteristics affect the choice of sizes and locations of urban green that are not purely geographical (like distance to CBD or elevation).

Similarly, with the total number of parks

$$n^* = \kappa\pi r \left[\frac{V^*(r, a_0, u, \xi)}{\zeta c_0}\right]^{\frac{1}{(1-\gamma)+(1-\delta)\frac{\alpha}{\beta}}}$$

$$\ln n = \ln\left[\kappa\pi\left(\frac{\beta(1-\beta)^{\frac{1-\beta}{\beta}}}{\zeta c_0}\right)^{\vartheta}\right] + \ln r + \frac{\vartheta}{\beta}\ln w - \vartheta t r + \vartheta \ln\left(u^{-\frac{1}{\beta}} a_0^{\frac{\alpha}{\beta}}\right) + \frac{\vartheta}{\beta}\ln\xi$$

Hence, we run the following regressions

$$\ln n_{ijc} = \rho_0 + \rho_1 \ln w_{jc} + \rho_2 \ln dist_{ijc} + \rho_3 dist_{ijc} + \rho_4 \hat{\epsilon}_{ijc} + \sum_{ijc} X^{(1)}_{ijc} + \sum_{jc} X^{(2)}_{jc} + \sum_{c} I_c + \varepsilon_{ijc}$$

$$\ln k_{ijc} = \eta_0 + \eta_1 \ln w_{jc} + \eta_2 dist_{ijc} + \eta_3 \hat{\epsilon}_{ijc} + \sum_{ijc} X^{(1)}_{ijc} + \sum_{jc} X^{(2)}_{jc} + \sum_{c} I_c + \varepsilon_{ijc}$$

$$\ln d_{ijc} = \mu_0 + \mu_1 \ln w_{jc} + \mu_2 dist_{ijc} + \mu_3 \hat{\epsilon}_{ijc} + \sum_{ijc} X^{(1)}_{ijc} + \sum_{jc} X^{(2)}_{jc} + \sum_{c} I_c + \varepsilon_{ijc}$$

to estimate the effect of the unobserved factor in household decisions which affect both the population density and the choices for urban green spaces numbers, sizes and distances. If the estimated value of $\hat{\rho}_4$, $\hat{\eta}_3$, and $\hat{\mu}_3$ is not significant and close to zero, we can exclude a systematic issues in omitted unobserved variables that affect both population density and the location and size of urban green space systems at the same time.

The estimation results are indicated in Table 4.9 in the main text. We use both full samples of all annulus in all cities as well as annulus from only monocentric cities with different level of controls and fixed effects. Specifically, in Column (1) and (2), we use no control and use only control for city geographical variables, and we can see the estimated value for $\hat{\rho}_4$ is positive and significant suggesting that the annuli controls for agriculture and forest presence and the regional fixed effects are important factors in deciding the total number of parks in each annulus across locations and affect the population density as well. Once, we control for annuli factors like agriculture and forest shares, all variable $\hat{\rho}_4$, $\hat{\eta}_3$, and $\hat{\mu}_3$ are close to zero in magnitude and are not significant. The estimation remains the same with sub-samples of annulus within monocentric cities only.

Chapter 5

Urban Floods, Amenity Land and City Structure

"Nature no longer exists apart from humanity. Henceforth, the world we will inhabit is the one we have made." - **Jedediah Purdy (2015),** ***After Nature: A Politics for the Anthropocene***

5.1 Introduction

Cities or urban areas are the places that host half of the world population; in addition, this proportion will soon be two-thirds in the next thirty years (UN DESA, 2014). In some countries, urban areas are also estimated to generate approximately 70 to 80% of the gross domestic product (GDP) (Weiss, 2001, Dobbs et al., 2011). Meanwhile, the total build-up areas account for only 1-2% of the total land on earth. Such a high level of concentration of population and properties in urban areas is essential for innovation and growth. However, it also exposes cities to spatial vulnerabilities, which could make them susceptible to the higher risk of natural hazards than rural areas. If the urbanization process, together with its downside vulnerabilities, is unchangeable and inevitable, the next question should be how we manage to make the urban system work. The question regarding how to manage urban areas has been one of the most critical issues on the UN post-2015 development agenda (UN DESA, 2014).

From our point of view, flood hazard is one of the most recurring and costliest natural disasters. It is estimated that from 1985-2014, floods have caused a death toll of half a million people and approximately 500 billion in US dollars lost in property worldwide (Dartmouth Flood Observatory, 2014). In Europe from 1985 to 2008, there were 285 flood events in total, which on average lasted 8,5 days, killed approximately eight people, displaced approximately 11,000 people and damaged approximately 1 billion worth in US dollars per a flood event as indicated in Table 5.1. Despite this fact, flood-prone areas are more likely to be overpopulated due to the cost of constructing defen-

Table 5.1: General flood statistics in EU27 for 1985-2008

	obs.	mean	sd	median	min	max
Duration (days)	285	8.51	11.61	4.00	1.00	102.00
No. of people death (hab.)	285	8.05	15.67	2.00	0.00	148.00
No. of people displaced (thousand hab.)	185	11.09	35.38	1.00	0.00	294.10
Damages (millions USD)	100	1039.27	4223.08	32.79	0.00	36466.01
Affected Area (thousand km^2)	285	49.65	110.11	19.59	0.01	1164.00

Note: The data are taken from Dartmouth Flood Observatory (DFO) flood map for the period 1985-2008, and only the flood disasters happened within the administrative boundaries of EU27 are taken into account. The data for displaced population and damages for some flood events were not available.

sive infrastructure and because the costs of reconstruction are partially borne by the residents in safer places or nonresidents (Kydland and Prescott,1977; Turner, 2012). If this mechanism is precisely the case, it will create a considerable misallocation of the resources in the economy. A land-intensive amenity such as green urban areas typically involves a high opportunity cost as it requires the land in areas where it is most valuable. Therefore, a lower land price at a central area due to higher flood risk results in lower cost for building any land-intensive green area. This situation, in turn, attracts more residents and induces a higher level of residents susceptible to flood risks. The same mechanism also applies to the building of flood defense systems. Hence, an understanding of such trade-offs when planning for the provision level of public amenity and flood defense infrastructures is essential for urban planners.

Another important question is also why we should care about flood risk in urban areas and whether it it better to relocate the whole city to a safe area. In the European context, most of its old and most populous cities such as London, Paris, Dublin, Brussels or Berlin are close or next to networks of rivers, as they provided easy access to water and defense in the past. Even though the waterways have become increasingly less important in contemporary society, the substantial cost of moving and the long duration of housing infrastructure make the relocation of a whole city rigid and hard. Many studies also report that households positively value the proximity to waterways such as rivers or canals and the coasts (Orford, 2002; Luttik, 2000; Cho et al., 2006). Therefore, it is vital to incorporate the flood risk into urban land use planning, especially in cities that contain any areas susceptible to this type of risk.

In this paper, we focus chiefly on the impact of flood hazards on optimal urban amenity provision, flood defense systems and city structures under both first-best and second-best scenarios. Toward this aim, we first create a model of a two-dimensional city with an exogenous level of flood hazards. Households can decide on how much consumption of composite goods, the sizes of their housing floor service, and where to locate within an urban area. Under the first-best scenario, a city planner decides the residential area and amenity land, as well as the density of households at each location,

and how much to invest in the flood defense systems. Hence, in this case, the urban planner internalizes all cost borne by the flood hazard. Under this scenario, we find a nonmonotonic relationship between the level of flood risks and the optimal provision of public flood defenses and amenity land. In the second-best case, the city government collects taxes to cover the expenses of flood defense systems, as well as the provision of amenity land. It faces budget balance constraints. We find that a locational tax is efficient while an income tax system would result in over-concentration of population and housing construction in a flood-prone area.

In the second part of the paper, we examine the current state of land uses in more than 200 of the most populated European cities that are close to at least a river network and susceptible to some flood hazards. Those cities are a subset of the city samples of the Urban Atlas in Picard and Tran (2019).[1] We combine that information with the European 500-year and 100-year flood hazard maps provided by the European Environment Agency (EEA) and Eurostat population grid database. Both of these datasets cover the entirety of Europe. We first find that the population density and the fraction of dense residential construction are more likely to establish in a safer area than in a hazardous area, which confirms the prediction from our theoretical model. Second, the provision of amenity lands is nonmonotonic with the level of flood risks and depends on the relative location within the urban area. Specifically, the fraction of amenity land is more likely to be built in risky and dense area than in a risky but sparse area near the urban fringe.

The contribution of this paper to the literature is twofold. First, on the theoretical side, it contributes to the understanding of the optimal allocation of amenity and urban land use planning under an adverse and exogenous flood risk. Frame (1998) is the first paper that incorporates the flood risk and insurance into a residential development model. He finds a negative impact of flood risk on residential development under a competitive land market. Bakkensen and Barrage (2018) use a dynamic asset pricing model with heterogeneous beliefs of households to explain the ambiguous effect of flood risk on housing prices. However, neither the papers addresses the effects of flood risk on the choices of amenity land and flood defense systems. Picard and Tran (2019) are the closest to this paper; nevertheless, they do not incorporate the flood risk inside their model. To the best of our knowledge, this is the first paper that considers the effect of river flood risks on optimal amenity land provision and flood defense systems under both first-best and second-best frameworks.

The second contribution is to empirically measure the effects of river flood risk on city structures and amenity provision across cities in Europe. Most of the empirical pa-

[1]GMES Urban Atlas uses Earth observation satellite images with 2.5-m spatial resolution multispectral data, which is at a much higher resolution than that of the traditional CORINE land cover resolution. According to the European Environmental Agency, the current resolution of the GMES Urban Atlas is 100 times higher than that of the CORINE land cover.

pers in the literature use hedonic pricing to estimate the effects of flood hazards. Many have reported negative effects of floods following hurricanes and other catastrophic events (Hallstrom and Smith, 2005; Atreya et al., 2013; Bin and Landry, 2013; Zhang, 2016) as well as increases in flood insurance take-up rates (Gallagher, 2014). However, those effects are often short-lived and mostly vanish after five years. Kocornik-Mina et al. (2018) use the night light data to investigate the short-term effect of significant flooding events worldwide. They also find that the economic activities in urban areas were shortly reduced immediately after the flood but soon reverted to a reasonable level. One exception is Ortega and Sulleyman (2018), which find a persistent decrease in prices of flood-prone properties in New York City after Hurricane Sandy due to a more extensive and more detailed dataset at a precise postal-code. In this paper, we utilize a different dataset of geographical land covers to track different types of land uses at precise locations within an urban area. We also additionally combine those data with the Eurostat population grid and the detailed flood hazard maps with multiple return periods in Europe. These data at a high spatial resolution help us to examine the prediction from the theoretical model precisely. We confirm a persistent pattern of lower population density and lower residential construction in flood-prone areas. We also find a nonmonotone relationship between amenity and flood risk level.

The organization of this paper is as follows: Section 2 introduces the theoretical model and discusses the effects of flood risk on the city structures, as well as the optimal provision of flood prevention facilities and amenity areas. The second part of Section 2 provides the framework of second-best policies with two different types of taxes: locational and income tax. Section 3 is devoted to our empirical approach and results. We first provide evidence on the negative impact of flood risk on the population density and the residential construction cover. We, then estimate the amenity land choices with flood risk levels across locations within the city area. The last section concludes the paper.

5.2 Theoretical model

5.2.1 Model settings

5.2.1.1 Households and flood risk

We construct an urban spatial model to incorporate flood risk, commuting, housing cost and local public amenity provision, as well as flood defense allocation. To achieve this, we first follow the framework of Fujita (1989) and add an exogenous flood risk level inside the model together with flood prevention efforts. We consider a monocentric city where all households commute to a central business district (CBD) to work. We assume that households necessitate commuting cost t and receive an income of w.

They are homogeneous and have the same preference $U(z,h,a)$ represented over the nonhousing composite good z, housing floor service h and amenity level a. We assume U is increasing in each of the components, indicating the desirability of the additional consumption of housing and nonhousing goods, and a higher level of amenity such that $U_z, U_h, U_a > 0$. All goods are assumed to be normal goods with decreasing marginal utility level $U_{zz}, U_{hh}, U_{aa} < 0$. We also further assume global concavity for our utility function U.

We introduce the flood risk in our model in the following way. Households that inhabit a flood risk area face a financial loss once a flood event occurs. The severity of the loss is uncertain ex ante for households and is measured by a random variable $x \in [0,1]$. Once a flood event impacts households, households realize a financial loss, which increases with the severity level x. We denote $l(x)$ as the financial loss function per each square meter of housing floor for each household. We assume that the financial loss is increasing with the severity level x, such that $l_x > 0$. Households also decide to invest in some flood prevention effort e, which combined with the public prevention effort p can reduce the financial loss from a flood event. Denote $m(e,p) : R^2 \to [0,1]$ as the reduced fraction of financial loss stemming from the prevention efforts. We further assume an increasing and concave function of the prevention fraction with respect to e and p as well as

$$\lim_{e\to 0;p\to 0} m(e,p) = 0 \quad \text{and} \quad \lim_{e\to+\infty;p\to+\infty} m(e,p) = 1. \tag{5.1}$$

Specifically, given a level of flood protection $e > 0$ and $p > 0$, the realized financial loss from a flood event is $l(x)(1-m(e,p))$ per square meter of housing floor service. The cost of providing the flood prevention from households is $C^H(e)$ per square meter of housing floors. $C^H(e)$ is assumed to be an increasing and convex function $C^H_e > 0; C^H_{ee} > 0$. We can then formalize the budget constraint for households as $w - t = z + Rh + h\left(l(x)(1-m(e,p)) + C^H(e)\right)$. The price of numeraire composite good z is normalized to one, and the cost of renting a housing floor unit is R.

Household location is indicated by a pair (r,ϕ) where r is the Euclidean distance from the CBD, and ϕ is the radian distance from the East axis. The space of households' locations is then given by the product of $\mathcal{R}_+ \times [0,2\pi]$, and is Lebesgue measurable. For each location (r,ϕ), the cumulative distribution of flood severity is $F(x;r,\phi)$. We also further assume that the flood risk distribution has the property of first-order stochastic dominance, meaning that for every pair of locations (r,ϕ) and (r',ϕ'), we can always compare the risk level between them. A location (r',ϕ') is said to be riskier than location (r,ϕ) if and only if $F(x;r',\phi') \leq F(x;r,\phi), \forall x \in [0,1]$. A location is safe when $F(0;r,\phi) = 1$.

5.2.1.2 Housing sector

Following Owens et al. (2017), we assume that the housing sector is developed by a large number of small developers with the free entry condition. Denote $q(r,\phi)$ as the total housing floor developed per unit of land at location (r,ϕ). Land developers have a convex variable cost $f(q) = \kappa q^{\alpha}$ with $\kappa > 0, \alpha > 1$, in addition to a land rent Ψ from absentee landlords.

Residential developers then choose the quantity housing floor q to maximize the total profit

$$\max_{q} \ \Pi = Rq - \kappa q^{\alpha} - \Psi. \tag{5.2}$$

The free entry condition states that

$$\Pi = 0 \Longleftrightarrow \Psi = Rq - \kappa q^{\alpha} \tag{5.3}$$

5.2.2 First-best world

Consider a benevolent urban planner whose objective is to maximize social welfare at the minimal cost. To avoid the unequal treatment of the equals as indicated in Mirrlees (1972), we assume that each consumer reaches a level of utility u, and the objective of the urban planner is to minimize the costs to attain that level of utility. There are two types of land uses that we consider in this model: residential and amenity land. We denote $g(r,\phi)$ as the fraction of land used for amenity at each location, and the rest $1 - g(r,\phi)$ is for the housing sector. For simplicity, it is further assumed that linear amenity service $a(r,\phi) = \theta g(r,\phi)$ with $\theta > 0$, and a linear maintenance cost $C^M(r,\phi) = cg(r,\phi)$. Additionally, the urban planner incurs a cost $C^U(p)$ per unit of land for a public flood prevention effort p. We assume that $C^U(p)$ is an increasing and convex function of p such that $C^U_p > 0$ and $C^U_{pp} > 0$.

Due to the uncertain severity of flood disaster events, the urban planner then chooses $z(r,\phi), h(r,\phi), n(r,\phi), e(r,\phi)$, $q(r,\phi)$, $p(r,\phi)$, and $g(r,\phi)$ to minimize the following expected costs:

$$\begin{aligned} EC = &\int_0^b \int_0^{2\pi} \int_0^1 \Big(z + t + h\Big(l(x)(1 - m(e,p)) + C^H(e)\Big)\Big)\, n\ell \mathrm{d}F(x;\phi,r)\mathrm{d}\phi\mathrm{d}r \\ &+ \int_0^b \int_0^{2\pi} \int_0^1 \Big(R_A + cg + \Big[C^U(p) + \kappa q^{\alpha}\Big](1 - g)\Big)\, \ell \mathrm{d}F(x;\phi,r)\mathrm{d}\phi\mathrm{d}r \end{aligned} \tag{5.4}$$

where $n(r,\phi), \ell(r,\phi)$ are the number of inhabitants and available land at location (r,ϕ) respectively, and b is the city boundary. The optimization is subject to the housing floor constraint $n(r,\phi)s(r,\phi) = (1 - g(r,\phi))q(r,\phi)\ \forall (r,\phi) \in [0,b] \times [0,2\pi]$, the population

constraint

$$\int_0^b \int_0^{2\pi} n(r,\phi)\mathrm{d}\phi\mathrm{d}r = N \tag{5.5}$$

and free mobility condition

$$U(z,h,a) \geq u \tag{5.6}$$

Applying the housing floor constraint condition, and since wages w are exogenous in the urban areas, we can transform the minimization of the expected social expense in (5.4) to the maximization of expected social surplus $ES = wN - EC$. After substitution of the population and land use constraints, this provides

$$ES = \max_{z,h,n,e,p,g} \int_0^b \int_0^{2\pi} \int_0^1 \left(\frac{w-t-z}{h} - l(x)(1-m(e,p)) - C^H(e) \right) q(1-g)\ell \mathrm{d}F(x;\phi,r)\mathrm{d}\phi\mathrm{d}r$$
$$- \int_0^b \int_0^{2\pi} \left(\left(\kappa q^{\alpha} + C^U(p) \right)(1-g) + R_A + cg \right) \ell \mathrm{d}\phi\mathrm{d}r \tag{5.7}$$

The social planner then chooses a bundle (z,h,n,e,p,g) to maximize the social surplus such that $U(z,h,a) \geq u$.

5.2.2.1 Housing demand and housing provision

We first denote the average financial loss from a flood risk event at a location (r,ϕ) with flood severity accumulative distribution $F(x;r,\phi)$ as

$$l(r,\phi) \equiv \int_0^1 l(x)\mathrm{d}F(x;\phi,r). \tag{5.8}$$

Since the urban planner is risk-neutral, the uncertainty can be eliminated in (5.7).

The optimization problem of the urban planner can be solved in three stages. First, the urban planner chooses the optimal consumption bundle (z,h) given by the point-wise first-order condition of the residential housing floor value.

$$V(h,u,a) \equiv \max_{s,z} \left[\frac{w-t-z}{h} - l(r,\phi)(1-m(e,p)) - C^H(e) \right] \text{ s.t } U(z,h,a) \geq u \tag{5.9}$$

We denote the solution of $U(z,h,a) = u$ as $\hat{z}(h,u,a)$. As $U(z,h,a)$ is an increasing function of z, it is easy to see that $\hat{z}(h,u,a)$ exists and is unique. Replacing it in equation (5.9), we can now take the first-order condition with respect to s, and equalize to zero, the optimal housing slot size needed to satisfy the following condition;

$$\hat{z} - h\hat{z}_h = y \tag{5.10}$$

where $y = w - t(r)$ is the disposable income after subtracting the commuting cost. Given the concavity of the utility function yielding $z_{hh} > 0$, the second-order conditions are satisfied at any location[2], which implies a unique solution to the above maximiza-

[2]We have the second order equal to $-hz_{hh} < 0$.

tion problem. Denote the solution from equation (5.10) as $\tilde{h}(y,u,a)$, and we obtain the optimal composite good consumption as $\tilde{z}(y,u,a)$ and the land value $\tilde{V}(y,u,a,e,p,l)$

$$\tilde{V}(y,u,a,e,p,l) = \frac{w-t-\tilde{z}(y,u,a)}{\tilde{h}(y,u,a)} - l(1-m(e,p)) - C^H(e)$$

where l is the average financial loss of flood risks at location (r,ϕ). One notices that the demands for composite good $\tilde{z}$ and housing floor size $\tilde{h}$ do not depend on the flood risk and protection effort, as the financial loss is linear with housing floor size. Using the envelop theorem, we get the change in land value with respect to the risk is

$$\tilde{V}_l = -(1-m(e,p)) < 0 \tag{5.11}$$

We, then, have that the housing value declines with the risk level, which is intuitive as the potential damage from flood risk makes the land less attractive to the city residents.

Proposition 1 *An increase in risk levels has a negative impact on housing value $\tilde{V}$, while there is no change in housing demand $\tilde{h}$ and demand for composite good $\tilde{z}$.*

The objective of the urban planner, now, can be rewritten as follows.

$$ES = \max_{e,p,q,g} \int_0^b \int_0^{2\pi} \Big([\tilde{V}(y,a,u,e,p,l)q - \kappa q^\alpha - C^U(p)](1-g) - R_A - cg \Big) \ell \mathrm{d}\phi \mathrm{d}r \tag{5.12}$$

The first-order condition w.r.t the total housing construction $q(r,\phi)$ requires

$$\tilde{V} = \alpha\kappa q^{\alpha-1} \Longleftrightarrow \tilde{q} = \left(\frac{V}{\alpha\kappa}\right)^{\frac{1}{\alpha-1}} \tag{5.13}$$

Taking derivative w.r.t. l on both sides of equation (5.13), we obtain

$$\tilde{q}_l = \frac{1}{\alpha-1}\left(\frac{\tilde{V}}{\alpha\kappa}\right)^{\frac{1}{\alpha-1}-1} \frac{V_l}{\alpha\kappa} < 0 \tag{5.14}$$

as $V_l < 0$ and $\alpha > 1$. It states that more housing floors are constructed at locations with higher housing value V and with lower flood risk. It is interesting to note that the housing construction decreases with flood risk regardless of the form of the financial loss function. From here, we obtain as well the population density at each developed location

$$\tilde{n} = \frac{\tilde{q}(1-g)}{\tilde{h}} \Longrightarrow \tilde{n}_l = \frac{1-g}{\tilde{h}}\tilde{q}_l < 0 \tag{5.15}$$

which is negative as $\tilde{q}_l < 0$, suggesting that the population density decreases with the level of flood severity.

One notes that we have a lower bound of zero surplus at any undeveloped area since the urban planner can choose not to lease the land from absentee landlords and

set $n(r,\phi) = 0$. Therefore, a location will be developed if and only if it incurs non-negative point-wise social surplus. In other words, the condition for a location gets developed is

$$[\tilde{V}\tilde{q} - \kappa\tilde{q}^{\alpha} - C^{U}(p)](1-g) - R_A - cg \geq 0 \tag{5.16}$$

$$\iff \tilde{V} \geq \underline{V}(R_A, p) \equiv \left(\frac{\alpha}{\alpha-1}(R_A + cg + C^{U}(p)(1-g))\right)^{\frac{\alpha-1}{\alpha}} (\alpha\kappa)^{\frac{1}{\alpha}} \tag{5.17}$$

From equation (5.11), we can see that the flood risk level reduces the bid of households, leading to lower housing floor value $\tilde{V}_l < 0$. If the effect of the flood risk is strong enough to reduce the bid rent V at some location to be lower than the threshold $\underline{V}$, then this location will be left undeveloped. This corner equilibrium can lead to the discontinuity in land use within the urban area. Specifically,

$$\tilde{V} \geq \underline{V}(R_A, p) \iff \underbrace{L(p,l) \equiv l(1-m(e,p)) + C^{H}(e)}_{\text{Loss due to flood hazard}} \leq \frac{y - \tilde{z}(y,a,u)}{\tilde{h}(y,a,u)} - \underline{V}(R_A, p) \tag{5.18}$$

$$\iff l \leq \underline{l}(p, R_A) \text{ where } \underline{l}(p, R_A) \equiv \frac{\frac{y-\tilde{z}}{\tilde{h}} - \underline{V}(R_A, p) - C^{H}(e)}{(1-m(e,p))} \tag{5.19}$$

Any location with average financial loss of flood risk l higher than $\underline{l}$ is not developed. One notices that $\underline{l}$ is decreasing with distance to the city center, meaning the restriction is stricter near urban fringe and with higher outside opportunity cost of land, all else being equal.

Proposition 2 *All locations with average flood damage lower than $\underline{l}(p, R_A)$ threshold are developed, and the housing construction level $\tilde{q}$ as well as population density in those areas are lower than those in safe areas.*

5.2.2.2 Optimal prevention efforts

First-order conditions with respect to the private $e(r,\phi)$ and public protection effort $p(r,\phi)$ from the social surplus function give us the following conditions:

$$V_e = 0 \tag{5.20}$$

$$V_p = \frac{C_p^{U}}{q} \tag{5.21}$$

The equation (5.20) states that the optimal level of private protection effort e is chosen such that its marginal benefit on housing value is equal to zero. Using the equation (5.9) in the previous section, we can rewrite the condition for optimal private prevention effort e as

$$l \times m_e(e,p) = C_e^{H}(e) \tag{5.22}$$

It means at the optimal e, the marginal cost of an additional prevention effort is equal to the marginal reduction in financial loss once the location is hit by a flood event.

The optimal level of public prevention effort p is such that the marginal benefit for each unit of housing value multiplied by all housing floors at one location is equal to the marginal cost C_p^U. To see how the decision differs compared to that of private prevention effort e, we utilize again equation (5.9) and rewrite condition (5.21) as follows:

$$l \times m_p(e,p) \times q = C_p^U(p) \tag{5.23}$$

Equation (5.23) states that the marginal cost of an additional public prevention effort p is equal to the total additional benefits (in the form of financial loss reduction) for all households located at the same location. Taking the derivative w.r.t to average flood loss l on both sides of equations (5.22) and (5.23), we have the following comparative statics characteristics of the optimal private and public prevention effort $\tilde{e}$, $\tilde{p}$ with respect to the average financial loss from flood risk l.

$$\tilde{e}_l = \frac{m_e}{-l \times m_{ee} + C_{ee}^h} > 0 \tag{5.24}$$

$$\tilde{p}_l = \frac{m_p + \frac{C_p^U}{\tilde{q}^2}\tilde{q}_l}{-l \times m_{pp} + C_{pp}^U} > 0 \iff m_p + \frac{C_p^U}{\tilde{q}^2}\tilde{q}_l > 0 \tag{5.25}$$

Interestingly, the private protection effort is an increasing function with respect to the flood risk $\tilde{e}_l > 0$, while the flood risk has an ambiguous effect on the optimal public protection effort $\tilde{p}$. There are two channels here that affect the provision of public protection p: the financial loss in the floodplain area and the household density (in terms of housing construction). First, a higher level of flood risk leads to a higher level of financial loss, which induces the urban planner to provide a higher level of p (*the first term* $m_p > 0$). Second, a higher risk results in a lower level of housing construction, leading to a lower housing floor available and lower population density (*the second term* $\tilde{q}_l < 0$); hence, the urban planner provides lower protection p.[3] One also notices a corner solution $\tilde{e} = \tilde{p} = 0$ at safe areas $l = 0$.

Proposition 3 *The optimal private protection e is an increasing function with the level of flood risk, while the effect of flood risk on the optimal public prevention choice p is ambiguous.*

$$\tilde{e}_l > 0$$

$$\tilde{p}_l \lesseqgtr 0 \iff m_p + \frac{C_p^U}{\tilde{q}^2}\tilde{q}_l \lesseqgtr 0$$

In the secured area where $l = 0$, the optimal protection efforts are equal to zero $e = p = 0$.

[3] As seen in condition (5.23), the marginal benefit of a public prevention p is increasing in total housing floor constructed q or population density n.

5.2.2.3 Optimal provision of amenity land

The pointwise first-order condition of the social surplus function in (5.12) with respect to g gives us the following condition for the optimal prevention efforts and amenity provision:

$$\theta q V_a(1-g) = [qV - \kappa q^{\alpha} - C^U(p)] + c \tag{5.26}$$

Equation (5.26) shows the principle to provide urban amenity, which balances the total benefits of an increase in the fraction of land used for amenity and the forgone cost of the parcel of land used for providing amenity instead of housing and the marginal maintenance cost c. Here, the urban planner also needs to take into account the expected loss in land value where she chooses to construct the amenity spaces. To see how the change in the risk level influences the amenity level, we take the total differentiation of equation (5.26) with respect to the risk level. We have

$$g_l = \frac{\tilde{q}_l \tilde{V}_a(1-g) - \tilde{q}\tilde{V}_l}{\tilde{q}(\tilde{V}_a - \theta \tilde{V}_{aa}(1-g))} \tag{5.27}$$

Applying the concavity condition of the utility function with respect to amenity service a, we have

$$sign(g_l) = sign\left(\chi q_l - V_l\right) \tag{5.28}$$

where $\chi \equiv \frac{V_a(1-g)}{q} > 0$. The sign of g_l is ambiguous as both q_l and V_l are negative. There are two forces at play here: first, the sign of g_l is in the opposite direction of V_l, showing that the hazardous area with less valuable land value yields an advantage in cost reduction. That is the reason why, ceteris paribus, when the risk level increases leading to a lower land price as $V_l < 0$, it is optimal for the planner to put more amenity land there as the cost of the land is lower. The second part of the equation relates to the population density or the housing construction in the hazardous area. As we noted from the previous section, $q_l < 0$, indicating a lower level of housing floors constructed in the risky area and therefore a lower number of inhabitants, which in turns induces the planner to provide less amenity. These two forces work in opposite directions, and the sign of g_l is dependent on which channel is dominating at which location.

Proposition 4 *Under the uncertainty of severity when a flood event hits in risky locations, the optimal provision of land-consuming amenity facilities g needs to satisfy*

$$\theta q V_a(1-g) = [qV - \kappa q^{\alpha} - C^U(p)] + c$$

The change in g with respect to the risk level is also ambiguous and depends on the change in land value and population density with risks:

$$g_l \lesseqgtr 0 \iff \chi q_l - V_l \lesseqgtr 0$$

To obtain a closed-form solution of the land use under the uncertainty of flood risk, we need the explicit form of the utility function. In the appendices, we consider some popular and common forms of the utility functions and derive how the change in risk level influences urban structure and the provision of endogenous amenity. In the next section, we consider how the urban structure and the provision of protection efforts and local amenity change when the urban planner faces the budget constraints and needs to work with a decentralized equilibrium market.

5.2.3 Second-best world

Suppose a competitive land market with absentee landlords and a budget-balanced city government. We investigate the outcome of the land market when the tax systems, amenity levels, distribution of flood severity and damage are known for all agents. We consider two types of tax systems: a locational tax and an income tax. We show that a locational tax[4] can decentralize the first-best outcomes; hence, it inherits all the properties of the first best and is analyzed in Appendix A. However, a locational tax might not be easy to apply since it requires the city government to collect all information on flood damage and flood risk at different locations across cities over a long period of times. An income tax system is much easier to collect, but as in our analysis, it deviates from the first-best outcomes. To see how different the city structure under the income tax, we construct our second-best city with an income tax system as follows.

In this section, we consider a tax rate γ on household income to finance the city government's public spending. The housing bid rent changes as households now face a lower level of disposable income after tax $(1-\gamma)w$. We first study housing demand from the household side, then move to the land developers' problem once the bid rent function is determined, and the decisions of the city government over the provision of public flood mitigation and amenity land.

Housing demand

The household chooses z, h, e. During a unique stage, the household finds z, s, e that

$$\max_{z,h,e} U(z,h,a) \text{ s.t. } z + t + Rh + h\int_0^1 \Big(l(x)(1-m(e,p)) + C^H(e)\Big)\,\mathrm{d}F(x;r,\phi) = (1-\gamma)w. \tag{5.29}$$

One can interpret this as the consumer considering her permanent income (she can borrow to smooth her consumption).

Denote again $l = l(r,\phi) \equiv \int_0^1 l(x)\mathrm{d}F(x;r,\phi)$. We will drop the location indicator (r,ϕ) where there is no confusion in the notation. Each household will bid for the

[4]We consider the case where a locational tax is levied on land developers for easy notation and calculation. The locational tax on households results in similar outcomes.

location of their preference, and the slot will be given to the highest bidder. Hence, we can write the bid rent function as follows The household program is given by equation (5.29). The bid rent function is given by

$$\mathrm{Y} = \max_{z,h,e} \left(\frac{(1-\gamma)w - t - z}{h} - l(1 - m(e,p)) - C^H(e) \right) \text{ s.t. } U(z,h,a) = u \tag{5.30}$$

We have utility function U is strictly increasing in z; therefore, we can write $\hat{z}(u,h,a)$ as the unique solution of equation $U(z,h,a) = u$, at each level of housing consumption and amenity level of indifferent curve. For general form of utility, we have $z_h = -U_h/U_z < 0$, $z_a = -U_a/U_z < 0$ as utility is increasing function in all z, h, a; therefore, the consumption of composite good along the indifferent curve is decreasing with the housing floor service h and amenity level a. Then, the bid rent function can be defined as follows:

$$\mathrm{Y} = \max_{h,e} \left(\frac{(1-\gamma)w - t - \hat{z}(u,h,a)}{h} - l(1 - m(e,p)) - C^H(e) \right) \tag{5.31}$$

We have the following first-order conditions:

$$\hat{z} - h\hat{z}_h = y - \varpi \tag{5.32}$$

$$l \times m_e(e,p) = C_e^H(e) \tag{5.33}$$

where $\varpi = \gamma w$ is the total amount of tax that each household contributes to the city government's budget. We get the solutions $\bar{\bar{h}}(y - \varpi, a, u)$, $\bar{\bar{z}}(y - \varpi, a, u)$, and $\bar{\bar{e}}(p,l)$, which gives the equilibrium housing floor area value

$$\bar{\bar{R}} = V(y - \varpi, a, u, p, l) \equiv \frac{y - \varpi - \bar{\bar{z}}}{\bar{\bar{h}}} - l(1 - m(\bar{\bar{e}}, p)) - C^H(\bar{\bar{e}}). \tag{5.34}$$

As the left-hand side (LHS) of the equation (5.32) is a decreasing function[5] for housing floor area h, the lower disposable income of the right-hand side (RHS) results in higher housing floor area consumption h. The new equilibrium housing bid rent R in (5.34) is lower for each location (r,ϕ) compared to the case of first-best city and of locational taxation, keeping the same level of p and l.

Housing supply

On the supply side, we assume the land is owned by absentee landlords, who have two options to either lend the piece of land for land developers to construct residential houses or to the agriculture sectors at R_A. At each location, a land developer pays a land price of Ψ to the landlords to build q units of housing floors that can be rented out for households at R.[6] The construction cost is κq^α with $\alpha > 1$ as in the previous section.

[5]This is due to the concavity of the utility function in h.

[6]As it is clear that the tax and land price are at each location (r,ϕ), so we drop the notation.

The residential developers, then, choose the quantity housing floor q to maximize the total profit

$$\max_{q} \ \Pi = \overline{\overline{R}}q - \kappa q^{\alpha} - \Psi$$

Taking the first-order condition w.r.t q, we have

$$\overline{\overline{q}} = \left(\frac{\overline{\overline{R}}}{\alpha\kappa}\right)^{\frac{1}{\alpha-1}}. \tag{5.35}$$

Additionally, the free entry condition dictates

$$\overline{\overline{\Psi}} = \frac{\alpha-1}{\alpha}\overline{\overline{R}}\left(\frac{\overline{\overline{R}}}{\alpha\kappa}\right)^{\frac{1}{\alpha-1}}. \tag{5.36}$$

The rent is at least equal to or higher than the outside opportunity cost of land which is the agricultural land rent R_A. Therefore, we can denote the actual land rent at location (r, ϕ) is as follows:

$$R^{L}_{(r,\phi)} = \max\{R_A; \overline{\overline{\Psi}}_{(r,\phi)}\}$$

Therefore, a location will be developed if and only if

$$\overline{\overline{\Psi}}_{(r,\phi)} \geq R_A \iff R \geq \underline{\underline{R}}(R_A) \equiv \left(\frac{\alpha}{\alpha-1}R_A\right)^{\frac{\alpha-1}{\alpha}}(\alpha\kappa)^{\frac{1}{\alpha}} \tag{5.37}$$

From the condition (5.17), one sees that the threshold level $\underline{\underline{R}}(R_A)$ is lower than that for first-best scenario and for locational taxation (see Appendix A). This means that some locations that are left undeveloped by land developers in the case of locational tax are now eligible for development. The corresponding flood damage threshold is also lower. Specifically, we have

$$l \leq \underline{\underline{l}}(p, R_A) \text{ where } \underline{\underline{l}}(p, R_A) \equiv \frac{\frac{y-\omega-\overline{\overline{z}}}{\overline{\overline{h}}} - \underline{\underline{R}}(R_A) - C^{H}(\overline{\overline{e}})}{(1 - m(\overline{\overline{e}}, p))} \tag{5.38}$$

City government

The city government's objective changes to

$$ES = \max_{p,g} \int_0^b \int_0^{2\pi} \left([V(y - \gamma w, a, u, p, l)\overline{\overline{q}} - \kappa \left(\overline{\overline{q}}\right)^{\alpha} - C^{U}(p)](1-g) - R_A - cg \right) \ell \mathrm{d}\phi \mathrm{d}r \tag{5.39}$$

such that $U(z, h, a) \geq u$, and

$$\gamma = \frac{\int_0^b \int_0^{2\pi} \left(C^{U}(p)(1-g) + cg\right) \ell \mathrm{d}\phi \mathrm{d}r}{w \times N}. \tag{5.40}$$

This equation (5.40) specifies the tax rate that balances the city government's budget. The higher the provision for flood prevention p, the higher the tax rate γ, and since the cost of flood protection is convex, we also have an increasing marginal tax rate with p. In another world, we have $\gamma_p > 0$ and $\gamma_{pp} > 0$. An increase in amenity land induces two opposite effects on γ. First, it increases the maintenance cost, hence resulting in a higher tax rate. Second, an increase in g in risky areas means a smaller fraction of land is used for residential purpose, leading to a smaller area for flood protection; therefore, it induces lower cost for flood mitigation facilities and a lower tax rate. As the latter effect is shut down in safe areas, an increase in amenity land in safe zones will lead to an increase in the tax rate. However, the effect is unclear for an increase of amenity land in risky zones as these two forces work in opposite directions.

Taking the first order conditions for the city government's maximization problem in (5.39) with respect to p, we have

$$\overline{\overline{q}} \times (V_p + V_\gamma \gamma_p) = C_p^U \tag{5.41}$$

where V_γ is the marginal change in land bid rent with income tax γ. It is apparent that the higher level of income tax, the lower the land bid rent. Therefore, we have $V_\gamma < 0$, which implies lower marginal benefits on the LHS of equation (5.41) compared to the condition with locational tax in (5.58). An increase in public protection effort p has two effects in the case of income tax. First, it lowers the average financial loss from flood hazard, resulting in higher bid rent from households (the first term of the LHS in equation (5.41)). At the same time, an increase in flood protection induces a higher tax rate, leading to less disposable income and lower bid rent (the second term). Therefore, this leads to a lower level of public flood protection in risky areas compared to the case of locational tax. The flood protection level in safe areas does not change and is equal to zero.[7]

The implication from equation (5.41) is surprising and not straightforward. The common insight is that income tax would lead to over-provision of flood defense systems since residents located in safer areas would partially finance the cost. Here, the public flood protection effort is lower in risky areas under an income tax mechanism. Additionally, it is apparent from condition (5.37) that the threshold for land development in the case of income tax is lower than in the case of locational tax. This condition further implies that more land in flood-prone areas could be converted into residential land in the case of an income tax system. Then, an income tax system would lead to an outcome in which more locations in floodplain areas are converted into residential areas while each unit of residential land in a risky zone has less investment in terms of public flood mitigation infrastructure.

[7]Note that we have $l = 0$ in safe areas; hence, we have $V_p = l \times m_p = 0$. If it is replaced in equation (5.41) and rearranged, we get $(1 - \lambda V_\gamma) C_p^U = 0$ where $\lambda = \frac{(1-g)\ell}{w \times N} > 0$. Hence, $C_p^U = 0 \Rightarrow p = 0$.

The first-order condition with respect to g gives us the following equation:

$$\bar{\bar{q}}(1-g)\left(\theta V_a + V_\gamma \gamma_g\right) = V\bar{\bar{q}} - \kappa\bar{\bar{q}}^{\alpha} - C^U(p) + c. \tag{5.42}$$

Similar to the choice of p above, we also have a secondary effect of an increment in amenity land on the tax rate and bid rent V (the second term on the LHS of equation (5.42)). As discussed above, in areas with no flood risk, we have $\gamma_g > 0$. Hence, an increase in amenity land in safe zones would lead only to a higher tax rate, and a higher tax rate imposes lower land bid rent as $V_\gamma < 0$. The total marginal benefits of an increase in g on land bid rent are reduced by the second term, leading to a lower level of amenities in safe areas compared to the level of g in the same area in the case of locational tax. In risky areas, the intuition is ambiguous as there are two effects of an increase in g in these areas on tax, as discussed above. If amenity maintenance costs dominate, an increase in g in risky areas has the same effect as in safe areas. A compelling case is when the decrease in the flooded protection zone dominates, which leads to an overall reduction in tax, hence making the second term in the LHS of equation (5.42) positive. This increases the marginal benefit of an increase in g in risky areas, resulting in a higher level of g in equilibrium compared to the case of locational tax.

Proposition 5 *Under locational taxation rule, the outcomes of the first-best optimization are decentralized through market equilibrium through locational taxation.*

Proof. See Appendix A ∎

Proposition 6 *Under the income tax system, more land in risky areas is converted into residential use, while population density and housing construction are lower in both safe and risky zones. The city government provides less flood mitigation efforts p per unit of land in risky areas compared to the first best and the locational tax and is more likely to provide more amenity land there compared to the optimal level in the first best case.*

The analysis of the two types of tax systems confirms Henry George's theorem, which states that locational tax is sufficient to finance local public goods and does the least harm as it does not distort the market. However, this type of locational tax would not be simple to implement in reality as it requires all necessary information about the flood risk of a property to be reflected in household decisions. However, it may be valid with large-scale events[8] as these are more likely to increase the perceived risk among residents. With smaller-scale flood events, the literature regularly reports only short-live effects (Atreya et al., 2013; Hallstrom and Smith, 2005; Kocornik-Mina, 2019; Zhang, 2016). The income tax system is more convenient to implement; however, it leads to more construction in frequently flooded zones since people in safer areas partially bear the costs. Combining income tax with a partial zoning policy for frequently flooded areas could be a good alternative to a locational tax.

[8] like the flooding after Hurricane Sandy, which affected property values in New York in the long-run (Ortega and Taspinar, 2018)

5.3 Empirical analysis

In this section, we test some predictions from the theoretical models to determine if they reflect the reality. We test the prediction from Proposition 2 that predicts a lower level of housing construction and a lower population density in risky areas, respectively. Regarding the level of flood mitigation facilities, p and the level of amenity land provision g, the theoretical model gives ambiguous predictions. We check both flood prevention facilities and the level of amenity land in both safe and risky zones to determine which channel dominates in reality.

5.3.1 Data

5.3.1.1 Main sample

Flood risk is typically characterized by two main parameters: the return period and the damage associated with a flood. The return period, or recurrence interval, indicates the frequency that a flood event will occur. For example, a 100-year return flood hazard describes the potential damage of a flood event that occurs once every 100 years, which is equivalent to a 1% probability of occurrence. Similarly, a 10-year flood hazard map shows the locations flooded once every 10 years. The second parameter is the potential damages or the magnitude of a flood (if it occurs) at a specific location. There are two main types of flood mapping: flood hazard maps and flood risk maps. According to Directive 2007/60/EC of the European Commission, flood hazard maps show the extent and inundation depth of a flooded area with different return periods.[9] Meanwhile, a flood risk map indicates the potential damage to the population, economic activities, and the environment (e.g., the pollution aftermath of a flood event) with different flood return periods. While the return period or the frequency of flooding is similar for both types of flood maps, the magnitudes are different. For example, the substantial inundation depth of a flood hazard in a rural or deserted area could cause almost no damage to economic activities and result in no human casualties.

In this paper, we utilize flood hazard maps from the Joint Research Centre (JRC) of the European Commission. Figure 1 is a 100-year return-period flood hazard map for Europe with different levels of the inundation depth. We employ a flood hazard map instead of a flood risk map because the flood hazard map is exogenous and relatively independent of human economic activities. In contrast, the flood risk map takes into account population density and economic activities when mapping and hence could cause endogeneity in our regression. Furthermore, the flood hazard map produced by the JRC covers all of Europe and is consistent across all European countries.[10] An ad-

[9] See more detail at `https://ec.europa.eu/environment/water/flood_risk/flood_atlas/` (accessed on Dec 03, 2019).

[10] For more detail about the methodology and quality controls, see Alfieri et al. (2014) and Alfieri et al. (2015).

Figure 1: EU - Flood Hazard Map (100-year return period)

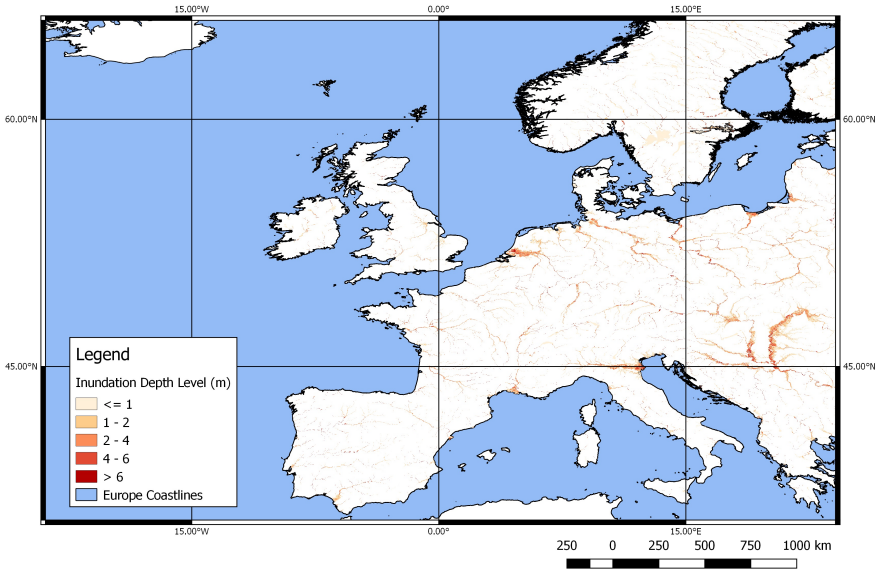
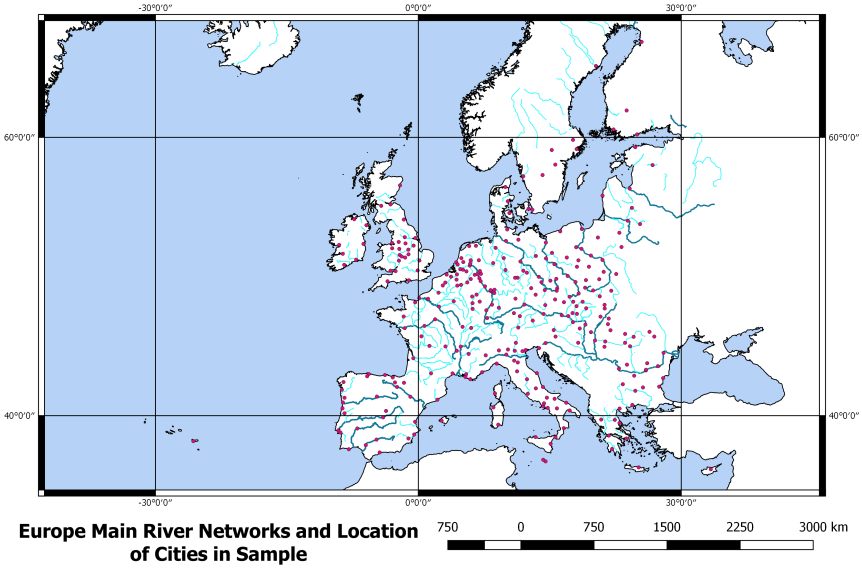


Figure 2: Europe Main River Networks and City Locations



Note: Main river networks combined of large rivers and other main rivers from WISE database. According to EEA, large rivers are rivers that have a catchment area large than 50,000 km2 (in dark color) and other main rivers are those that have a catchment of at least 5,000 km2 (in brighter color). Location of urban areas are taken as the location of the main communal hall in each city. The list of cities are contained most populous 305 cities in Europe from GMES Urban Atlas 2006 database.

ditional advantage is that the JRC flood hazard maps are high resolution, which allows us to compare the potential risk at different locations across urban areas. Flood maps like those provided by Dartmouth Flood Observatory (DFO, 2014)[11] are also useful as they outline the actual flooded areas in significant flood events; however, the coverage is broad and at low resolution. Further, it is impossible to derive the frequency of flooding from the DFO maps. Hence, this type of map cannot be used to investigate the effect of flood risk for small, local zones within urban areas such as those discussed in the theoretical part.

The JRC flood hazard maps include six return periods: 500-, 200-, 100-, 50-, 20-, and 10-year return periods. For each return period, the flood hazard maps indicate the inundation depth at each location using gridded cells. To implement our research design, we determine the flood exposure level as the average inundation depth at each $1km \times 1km$ grid cell for all return periods. This exposure level is used as a proxy for the magnitude of the potential damage that households face if they choose to locate there. We also calculate the flood frequency as the inverse of the return period (i.e., 1% probability for a once-in-100-years event) for each gridded cell in our database. Any cells within the safe area have zero probability of getting hit by a flood event, and no flood exposure is measured. We calculate this for all $1km \times 1km$ gridded cells in the 305 most populous cities in Europe. Among these cities, 221 have at least one grid measuring 1 km^2 that contains a positive flood risk level, which accounts for 72% of the cities in our sample. As shown in Figure 2, these cities are located close to main river networks. In this section, we restrict our samples to cities in which a significant part of the city faces flood risk. Therefore, we choose only cities with at least 20% of the city area under flood hazard, which results in a sample of 162 cities.

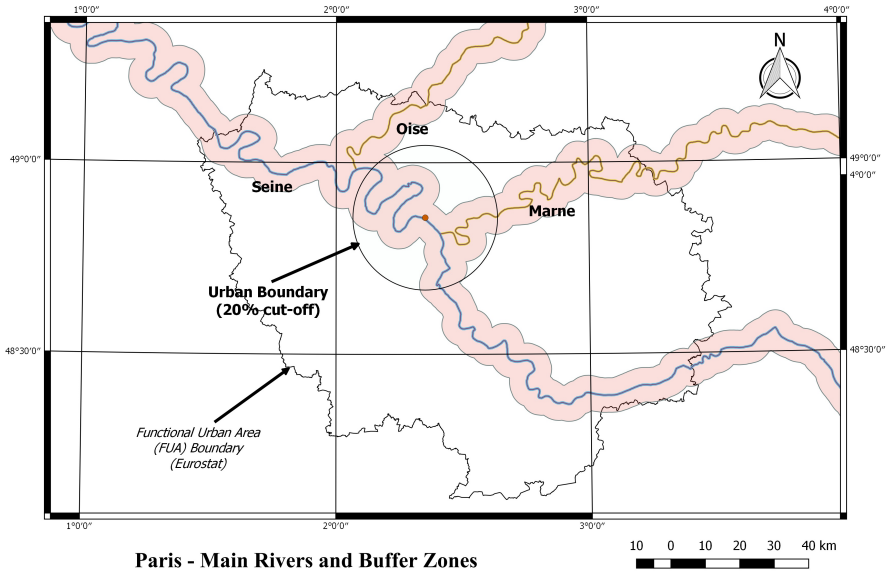
To investigate the land uses within urban areas, we follow Picard and Tran (2019) and use the GMES Urban Atlas. This dataset offers a high-resolution land use map of urban areas, with information derived from Earth observation (EO) and backed by other reference data such as commercial off-the-shelf (COTS) navigation data and topographic maps. The GMES Urban Atlas uses EO satellite images with 2.5-m spatial resolution and multi-spectral data, which offers much higher resolution compared to traditional CORINE land cover resolution [12]. It also covers 305 of the most populous towns and cities in Europe (EU 27); the data include all EU capitals plus a large sample of medium-sized cities. We focus on the 'urban fabric' classes, which specify different levels of continuity of residential areas[13], as well as different types of amenity land uses such as green urban areas (14100) and sports and leisure facilities (14200). To calculate population density, we combined these two datasets with the Eurostat population

[11] Kocornik-Mina et al. (2019) uses the DFO flood database.

[12] According to the European Environmental Agency, the current resolution of the GMES Urban Atlas is 100 times higher than that of CORINE land cover.

[13] the class codes are 11100, 11210, 11220, 11230, and 11240.

Figure 3: Paris Urban Boundaries and River Buffer Zones



grids to measure the density in locations across cities. We chose the city center (CBD) for each city at the location of the old town hall. The distance of each gridded cell to the city center is the Euclidean distance from the centroid of the grid to the CBD.

Most of the flood risk cities are close to river networks and experience moderate to high rainfall. As outlined in the previous section, the effects of location in a floodplain are difficult to measure due to the dual effect of water proximity. Many studies have reported a positive effect of being close to water (Luttik, 2000; Or, 2017, chap. 5). Therefore, in this section, we first restrict our sample to grids that are within a 10-kilometer buffer of main river networks[14]. Figure 3 illustrates the buffer zone from the main river network in an urban area for the city of Paris. By this restriction, we can control for river proximity and check, for the same distance from the CBD, and the same distance from the river, whether, how, and at which magnitude the flood hazard changes household location decisions and public provision of amenity lands. We also run a robustness check for the whole city sample and for different buffer zones around the river networks, the results of which are provided in the Appendix.

The summary statistics are indicated in Table 5.2. We define a dummy variable *Flood Exposure* that has a value of 1 if a grid has a positive probability of flooding and 0 otherwise. The average distance to the city center for the whole sample and for the flooded sample is not significantly different. A grid in a risky zone is, on average, 11.41 km away from the CBD, while the average for a whole city is 11.53 km. However,

[14]For maps and a definition of main river networks in Europe, please check the WISE database and Figure 4. We can adjust the buffer zone to 5, 7.5, and 12.5 kilometers as a robustness check. However, the results do not change significantly.

Table 5.2: Summary Statistics

	Panel A: Urban grids					
	Full River Sample			Flood Exposure = 1		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Population (thous. hab./km^2)	2.29	2.71	22239	2.28	2.47	9319
Distance to CBD (km)	11.53	8.60	22239	11.41	9.14	9319
Avg. Elevation (m)	118.81	110.99	22239	101.58	108.57	9319
Relative Avg. Elevation (m)	17.55	42.87	22239	-2.77	22.69	9319
Ruggedness	7.76	10.07	22239	7.15	10.92	9319
Distance to River (km)	3.80	2.74	22239	2.48	2.43	9319
Amenity Land (ha)	8.10	11.71	22239	8.50	11.40	9319
Commercial and Public Bld. (ha)	13.43	14.67	22239	14.44	14.94	9319
Flood Frequencies (per annum)	0.07	0.09	22239	0.16	0.06	9319
Flood Exposure	0.42	0.49	22239			
Avg. Inundation Depth Level (IDL) (m)				2.02	1.61	9319
	Panel B: Within 20% cut-off boundaries					
	Full River Sample			Flood Exposure = 1		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Population (thous. hab.)	3.17	3.36	11269	3.20	3.04	4516
Distance to CBD (km)	7.07	4.76	11269	6.05	4.54	4516
Avg. Elevation (m)	112.38	100.63	11269	90.90	97.59	4516
Relative Avg. Elevation (m)	19.22	45.67	11269	-3.82	20.14	4516
Ruggedness	8.01	10.76	11269	6.88	10.98	4516
Distance to River (km)	3.61	2.66	11269	2.21	2.24	4516
Amenity Land (ha)	10.08	13.37	11269	10.60	13.04	4516
Commercial and Public Bld. (ha)	14.75	15.22	11269	15.94	15.29	4516
Flood Frequencies (per annum)	0.07	0.09	11269	0.16	0.06	4516
Flood Exposure	0.40	0.49	11269			
Avg. Inundation Depth Level (m)				2.16	1.55	4516

Note: This table indicates the summary statistics from our samples of 1 km^2 grid cells. All grids are restricted to those within 10 kilometres from the main river networks. Urban cells are those that have at least 20% of land as artificial land use. Urban boundaries with 20% cut-off points are the distance to city center of the rings that have at least 20% of land are urban fabric or residential land use. See Paris map for the illustration. Relative Elevation is the difference between the grid's elevation and the elevation of the CBD of the same city. Ruggedness are measured as the standard deviation of the elevation within the grids. Amenity lands include green urban areas and the land used for sport and leisure activities.

flood zones are more likely to be at a lower elevation, less rugged, and closer to river networks. More precisely, a grid located in a flood-prone area is, on average, 1.3 km closer to a main river network. Flood zones are also more likely to be located at a lower elevation; the average elevation difference between a grid located in a safe area and one in a risky zone is 18 m. On average, a location in a flood zone has fewer residents and a higher amount of amenity land. Ruggedness index values are also lower in flood zones. Therefore, if the hypothesis is that it is easier to build a home in a less rugged area, this means that it is less expensive to build a home close to a river. Therefore, we would expect negative ruggedness values for areas with a high population density.

If we consider only grids that are at least 20% inside the cut-off boundary of an urban zone, the sample is reduced to nearly half of the whole sample. The average distance to the CBD is reduced to around 7 km instead of 11 km as before, and it is

6 km for grids in flood zones. Flood frequencies are calculated as the inverse of the return period. On average, a grid in a floodplain area has approximately 0.16 floods per annum. The frequency for a grid in the whole sample is approximately half, at 0.07. In addition, risk zones suffer higher inundation once a flood occurs. The average inundation depth for risk zones is 2.16 m while the average for the whole sample is 0.86 m. We later use the inundation level as a proxy for the potential damage once a location is hit by a flood event.

5.3.1.2 Supplemental data

Building heights

We combine our main datasets with building height data from the European Environmental Agency (EEA) under the framework of the Copernicus program. This dataset contains information on building block height at a spatial resolution of 10 m for the reference year of 2012.[15] The information on building height allows us to test the housing supply channel, as stated in Proposition 2. In particular, we use the data to obtain information about the vertical density of housing and to investigate if there are differences in housing structures in floodplain areas compared to those that are in safety zones. However, the data provided by EEA covers only the capital cities across EU28. To complement studies on housing with flood hazard, we utilize data on different categories of residential surfaces in the GMES Urban Atlas.

Table 5.3: Building Height Summary Statistics

	Panel A: Urban grids					
	Full River Sample			Flood Exposure = 1		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Avg. Building Height (m)	2.53	2.42	4391	2.47	2.41	1417
Avg. Height of Highest Building (m)	32.74	24.38	4391	34.10	27.52	1417
	Panel B: Within 20% cut-off boundaries					
	Full River Sample			Flood Exposure = 1		
Avg. Building Height (m)	2.85	2.56	3507	2.94	2.54	1074
Avg. Height of Highest Building (m)	35.54	26.24	3507	39.06	30.82	1074

Note: This table indicates the summary statistics of building heights from our samples of 1 km^2 grid cells. All grids are restricted to those within 10 kilometres from the main river networks. Urban cells are those that have at least 20% of land are artificial land use. Urban boundaries with 20% cut-off points are the distance to city center of the rings that have at least 20% of land are urban fabric or residential land use.

Table 5.3 reports the basic statistics for our building height database for all capital cities with at least 20% of land under flood risk. The average building height for all grids in the sample is 2.53 m, which is equivalent to a one-story building. The height

[15] Full details are at `https://land.copernicus.eu/local/urban-atlas/building-height-2012?tab=metadata` (accessed on Dec 06, 2019).

for flood exposed grids only is slightly lower at 2.47 m. The average height of the tallest building for all urban grids is 32.74 m, which approximately equal to a 10-story building. The height for floodplain grids is surprisingly a bit more at 34.10 m. When including only the grids within a circular urban border (which is defined as the border at which at least 20% of the land is residential), the number is higher on average. This indicates that the sample within the circular border is closer to the dense core area of the city. In our sample, the lowest point of a building block is 0 m, indicating a location at which there is no building, and the highest building block is 368 m, which is located in Berlin's city center.

Public flood prevention data

The second set of supplemental data that we use in this paper is for flood prevention infrastructure in France. We obtained the data from the French government on the *Georisques*[16]. The dataset contains information about the length and location of the flood prevention infrastructure currently in place in France's territories. It mostly contains information on dykes and pumping stations. In this paper, we consider only the infrastructure classified as "Ouvrage de protection". An illustration of the flood prevention infrastructure for the city of Paris is included in Appendix B. Currently, there is no systematic data on flood mitigation facilities for all cities in Europe.

Table 5.4: Flood Prevention Summary Statistics

	Full River Sample			Flood Exposure = 1		
	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Flood Prevention Infrastructures (m)						
A: All urban grids	50.83	267.71	6104	130.65	418.27	2279
B: Within 20% cut-off borders	72.91	325.59	3662	215.10	530.57	1232

Note: This table indicates the summary statistics of flood prevention infrastructures in France. All grids are restricted to those within 10 kilometres from the main river networks. Urban cells are those that have at least 20% of land are artificial land use. Urban boundaries with 20% cut-off points are the distance to city center of the rings that have at least 20% of land are urban fabric or residential land use.

Table 5.4 describes the summary statistics for flood prevention facilities measured as the length of the "Ouvrage de protection" for all the grids in our sample. The city sample includes all French cities in both the GMES Urban Atlas data in which at least 20% of the land is under flood hazard. The length for the grids within the floodplain areas is at least two to three times the average for the full samples. Another important detail is that the average flood prevention within the urban core (the 20% cut-off border) is higher than for the sample with all urban grids, indicating a possible different treatment in terms of the provision of flood facilities between the urban core and urban fringes.

[16]Link: `https://www.georisques.gouv.fr`.

In the next part, we study the effect of flood hazard on city structure. Specifically, we examine its effects on population density, level of housing construction, public flood prevention facilities, and local amenity land. We also compare the predictions from the theoretical part and the outcomes in reality.

5.3.2 Population density in floodplain areas

From the theoretical model, we predict that the population density decreases along with the level of flood risk. However, in reality, most risky areas are in close proximity to a water body such as a main river. For a simple check, we can see that grids that are close to a main river network exhibit a higher risk of flooding. It is well-documented in the literature that households gain some amenities by being close to a river or canal (e.g., ease of walking, view). Historically, being close to a river contributed to better defense and a better transportation network. Therefore, we follow the strategy of Berstein et al. (2019) and create buffer zones around river networks, restricting our grid samples to this zone to further control for the effect of river networks.

As discussed above, there are two main characteristics of flooding, namely frequency and magnitude. In this section, we examine the effects of both on population density. The reduced-form regressions of the effect of flood exposure on household location choice or population are specified below. Consider an observation (i, r, c) where i refers to 1 $km \times 1$ km cells, r is the distance to the city center, and c is the city to which the grids belong. Our estimation model is as follows:

$$\ln Pop_{irc} = \beta FloodD_{irc} + \theta Dist_{rc} + \sum_{c} \alpha_c I_c + \sum X_{irc} + \epsilon_{irc} \tag{5.43}$$

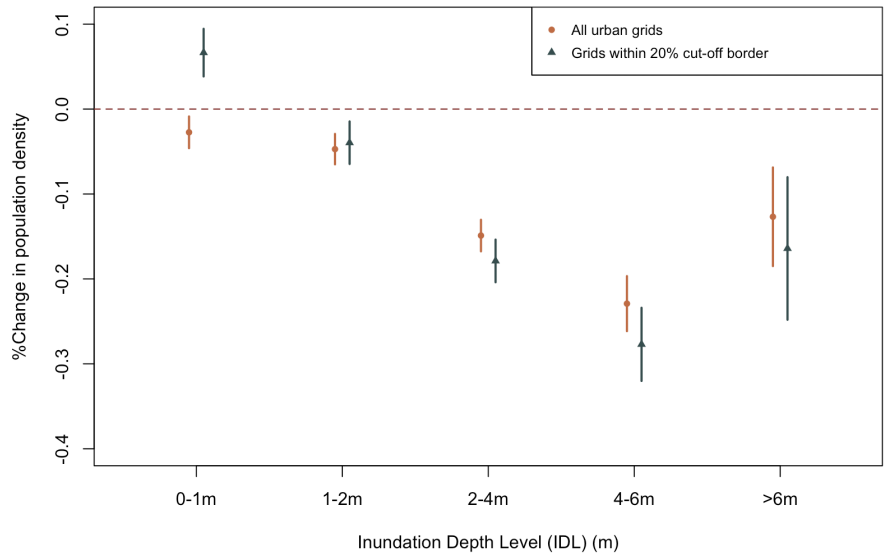
where $FloodD$ is our explanatory variable of interest. We consider three different measures for the $FloodD$ variable. First, we utilize the indicator *Flood Exposure*, which is equal to 1 if the grid has a positive risk of inundation and 0 otherwise. The second possible measure is flood frequency, as stated in a previous section. The third measure is the magnitude of flooding for each location, which is proxied by different inundation depth levels (IDL). We also use X_{irc} as a grid control. It absorbs the variation in population density due to the interaction between relative location and land characteristics such as ruggedness (lower cost of building) and elevation. Specifically, we control for distance to the river network[17], a grid's elevation above sea level, and its ruggedness. City fixed effects are also used to control for unobserved city-wide policies.

The baseline regression results reported in Table 5.5 provide initial evidence of the effect of flood exposure on population density. In Column (1), we simply run the regression of the natural logarithm of population density with only the controls for city fixed effect and flood exposure. The significantly positive coefficients for *Flood Exposure*

[17]We can also follow Bernstein et al. (2019) to introduce distance to river categories instead of continuous distance. In this way, we can avoid the assumption of linear effects.

indicate that flood-prone grids are associated with higher population density, which is consistent with Kocornik-Mina et al. (2019). The same regression with *Flood Frequencies* also gives similar results, as shown in Column (1) of the second panel of Table 5.5. These results are not surprising, given that, on average, the grids in floodplain areas are more proximate to river and canal networks. In Column 2, we further control for the geographical characteristics of each grid, such as the ruggedness and elevation level. The coefficients for *Flood Exposure* are still positive, but no longer significant after clustering the standard errors at the city level. The result in Column (3) shows that the correlation between flood exposure and the proximity to the water drives the positive relationship between flood exposure and population density. After controlling for the distance to river networks, and 2-km from the coast fixed effects, the relationship between population density and flood exposure becomes significantly negative. Any further controls for possible variables with possible influence on population density do not alter the sign and significance of the coefficients for flood exposure on population density.

Figure 4: Inundation Depth Level and population density



This figure denotes the relationship between the % change in population density of the grids within the zones with flood exposure (in comparison with unexposed grids), partitioned by different levels of inundation depth level. We categorize the flood hazard according to the exposure level from Very Low to Very High, which are defined as the average Inundation Depth Level (IDL) in all return periods. We set the IDL between 0 and 1 meter as very Low level and above 6 meters in Very High level.

The estimated coefficients for *Flood Exposure* of -0.085 and -0.082 in Columns (5) and (8), respectively, with the strictest controls, suggest that a grid in a flood-prone areas has 8.2% to 8.5% lower population relative to an unexposed grid in the same location in the city center and with similar characteristics. Similarly, the regression for

Table 5.5: Population density and flood hazard - Main regression results

	Dependent variable: Ln(Pop)							
Panel A:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Exposure	0.168***	0.009	−0.083**	−0.090***	−0.085***	−0.256***	−0.091***	−0.082***
	(0.033)	(0.035)	(0.037)	(0.028)	(0.028)	(0.048)	(0.030)	(0.029)
Dist. to CBD	Y	Y	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y	Y	Y
Geographical controls	N	Y	Y	Y	Y	Y	Y	Y
Dist. to river	N	N	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	N	N	N	Y	Y	N	Y	Y
Artificial Amenity Controls	N	N	N	N	Y	N	N	Y
River Buffer Zones	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km
Grid Sample	Urban grids	Urban grids	Urban grids	Urban grids	Urban grids	Within 20% cut-off border	Within 20% cut-off border	Within 20% cut-off border
Adjusted R^2	0.374	0.398	0.405	0.598	0.623	0.427	0.691	0.711
N	22,121	22,121	22,121	22,121	22,121	11,172	11,172	11,172
Panel B:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Frequencies	0.751***	−0.102	−0.633***	−0.636***	−0.569***	−1.964***	−0.887***	−0.790***
	(0.186)	(0.195)	(0.211)	(0.163)	(0.162)	(0.323)	(0.182)	(0.180)
Dist. to CBD	Y	Y	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y	Y	Y
Geographical controls	N	Y	Y	Y	Y	Y	Y	Y
Dist. to river	N	N	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	N	N	N	Y	Y	N	Y	Y
Artificial Amenity Controls	N	N	N	N	Y	N	N	Y
River Buffer Zones	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km
Grid Sample	Urban grids	Urban grids	Urban grids	Urban grids	Urban grids	Within 20% cut-off border	Within 20% cut-off border	Within 20% cut-off border
Adjusted R^2	0.373	0.398	0.406	0.599	0.624	0.432	0.693	0.712
N	22,121	22,121	22,121	22,121	22,121	11,172	11,172	11,172

Note: This table reports the OLS estimates where the dependent variable is the natural log of population count in $1k \times 1km$ grid. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (5) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (6) to (8), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

Flood Frequencies in Columns (5) and (8) in Panel B gives the coefficients of -0.569 and -0.790, respectively. This suggests that an increase of flood frequency of 0.01 (equivalent to one additional flood event in 100 years, or an increase of 1% in flood probability) leads to a reduction of 0.57% to 0.79% in population density. In other words, an increase of one additional flood event every 10 years is associated with a decrease population density of approximately 5.6% to 7.9%.

Figure 4 illustrates the effect of *Flood Exposure* on population density using a more comprehensive measure of flood exposure. In the figure, we estimate the regression (5.43) again, but we replace *FloodD* with a series of indicators of the IDL. This specification allows us to look at the non-linear relationship between IDL or potential flood damage and the population density. Across all interactions, we see a statistically significant and negative exposure level above 2 m in IDL. This means that all exposed IDLs above 1 m are related to a population decrease relative to those in safe areas. When we consider the sample with all urban grids, the population decrease begins even at a deficient IDLs. The greatest decline is at locations with an IDL above 2 m, with a magnitude between 10% and 30%. It is surprising to see a lower magnitude for the highest IDL (above 6 m), which could be explained by the lower initial settlements or zoning policies in highly inundated areas.

5.3.3 Housing construction density

Proposition 2 in our theoretical model predicts that the total floor area of housing constructed in floodplain areas is lower compared to the area in safety zones if everything else is equal. This is also the main channel in our model, which results in lower population density in flooded zones as the choice of housing floor area for each household does not change with flood hazard. In other words, lower population density is mainly due to lower housing supply in risky zones. Therefore, it is necessary to check if this mechanism matches what happens in reality. We thus run the following regression

$$HCL_{irc} = \beta_h FloodD_{irc} + \theta_h Dist_{rc} + \sum_c \alpha_c I_c + \sum X_{irc} + \epsilon_{irc} \tag{5.44}$$

where HCL_{irc} is the housing construction level of grid i at distance r from CBD of city c.

As discussed above, we use two types of measures for the HCL, namely average building height and residential land classification. An advantage of the building height database is that it precisely measures the vertical density of housing construction. However, the building height data does not cover all of Europe, only the capital cities. In contrast, the residential land classification data cover the most populous cities across EU27; however, it measures the horizontal density of sealed surfaces. Hence, we first run a simple regression with building height as the dependent variable and different classes of residential urban fabric as the explanatory variables. This regression reveals

the relationship between building height and different classes of residential land. In particular, we use all available classes for residential land, including continuous fabric (H11) with at least 80% of residential land as sealed surfaces (S.L.). The other classes for residential land range from dense, discontinuous urban fabric with S.L. between 50% and 80% (H21) to the very low density with S.L. below 10% (H24).

Table 5.6: Correlation between building height and different residential land classifications

	Dependent variable: Avg. Building Height					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of residential land categories						
H11 Continuous Fabric	10.638***					10.767***
(Sealing Surface (S.L.) > 80%)	(1.146)					(1.266)
H21 Discontinuous Dense Fabric		−0.044				1.306
(S.L. 50% - 80%)		(1.295)				(0.819)
H22 Discontinuous Medium			−5.051***			0.325
Density Fabric (S.L. 30% - 50%)			(1.301)			(0.521)
H23 Discontinuous Low Density				−7.785***		−1.374
Fabric (S.L. 10% - 30%)				(1.742)		(1.065)
H24 Discontinuous Very Low					−13.433***	−5.307***
Density Fabric (S.L. < 10%)					(0.704)	(0.862)
City FE	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.597	0.070	0.104	0.090	0.075	0.606
N	4,391	4,391	4,391	4,391	4,391	4,391

Note: Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

Table 5.6 describes the first regression results. The standard errors are all robust and clustered at the city level. Columns (1) through (5) show the results of the regressions of average building height for each class of residential land use, and Column (6) reports the regression results for all classifications together. Continuous residential urban fabric (H11) with S.L. above 80% has the highest explanation power with a significantly positive coefficient and adjusted R^2 equal to 0.597. The other classes of residential land have low explanation power with adjusted R^2 ranging from 0.070 to 0.104, which are just a sixth of the adjusted R^2 for the H11 class. The coefficient for the second residential fabric class (H21) is not significant, and the coefficients for the lower residential fabric classes (H22, H23, and H24) are all significantly negative. The results in Column (6) are similar when we rerun the regression with all residential urban fabric classes. These results suggest that building height is highly associated with dense, continuous urban areas. This is unsurprising as we mostly observe high-rise buildings in very dense core and central areas. Hence, we utilize both building height and the type of continuous urban fabric (H11) as proxies for housing construction density level (HCL).

Table 5.7 reports the OLS estimates for *Flood Exposure* in regression (5.44) with building height and share of continuous residential fabric as proxies for HCL. Panel A in Table 5.7 shows the first results for the regression with building height levels and different

Table 5.7: Housing construction and flood hazard - regression results

Panel A:	*Dependent variable:* Building Height							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Exposure	−0.012	−0.298*	−0.377**	−0.322**	−0.262*	−0.342**	−0.328**	−0.249*
	(0.146)	(0.181)	(0.183)	(0.144)	(0.155)	(0.153)	(0.136)	(0.134)
Adjusted R^2	0.415	0.452	0.460	0.526	0.601	0.474	0.541	0.624
N	4,391	4,391	4,391	4,391	4,391	3,507	3,507	3,507
Panel B:	*Dependent variable:* Share of Continuous Residential Categories (S.L > 80%)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Exposure	0.002	−0.008*	−0.018***	−0.019***	−0.018***	−0.036***	−0.028***	−0.024***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.007)	(0.005)	(0.005)
Adjusted R^2	0.212	0.223	0.233	0.348	0.420	0.363	0.411	0.500
N	22,121	22,121	22,121	22,121	22,121	11,172	11,172	11,172
Dist. to CBD	Y	Y	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y	Y	Y
Geographical controls	N	Y	Y	Y	Y	Y	Y	Y
Dist. to river	N	N	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	N	N	N	Y	Y	N	Y	Y
Artificial Amenity Controls	N	N	N	N	Y	N	N	Y
River Buffer Zones	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km
Grid Sample	Urban grids	Urban grids	Urban grids	Urban grids	Urban grids	Within 20% cut-off border	Within 20% cut-off border	Within 20% cut-off border

Note: This table reports the OLS estimates where the dependent variable is the average building height and the share of continuous residential land (with sealing surface (S.L.) > 80%) as proxies for housing construction level (HCL) in $1km \times 1km$ grid. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (5) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (6) to (8), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

levels of control as in the previous section. The grid samples are restricted to capital cities with at least 20% of the city area under flood hazard. The regression results are similar to those of the population density regression. Column (1), with a simple control for the city fixed effects and distance to the CBD, suggests a negative but insignificant effect on building height. Once we further control for geography and proximity to water, the coefficient is significant and negative. Using the coefficients in Columns (5) and (8), we find a 0.25 to 0.26 m decrease in building height in grids exposed to flood risk relative to unexposed grids. This measure accounts for approximately 9% of the average building height (2.5 m, as reported in Table 5.3).

Panel B of Table 5.7 reports the regression (5.44) with the share of continuous and dense urban fabric as a proxy for HCL. The number of observations is considerably higher than in Panel A, which is reasonable as the building height database covers only part of the datasets in Panel B. The results are robust even with the change in measurement for HCL. A grid situated within a floodplain has a 1.8% to 2.4% lower share of the continuous and dense residential urban fabric, as indicated in Column (5) and Column (8), respectively. These effects are robust with different control levels and urban cut-offs. The results in this section provide initial evidence for our theoretical prediction in Proposition 2.

5.3.4 Flood prevention infrastructure systems

In this section, we investigate the effects of flood hazard on the provision of public flood prevention facilities. As stated in Proposition 4, the change in flood mitigation facilities is not linear to the level of potential flood damage. In fact, the first derivative of public flood protection effort p with respect to average financial loss from flooding is ambiguous and depends on whether the increased marginal cost or the reduced benefit coverage dominates. To test the effect of flood hazard on the public protection level, we utilize data from France, which provide the length and location of flood protection facilities as shown in the supplemental data section above. In particular, we run the following regressions:

$$FloodPrev_{irc} = \beta_f FloodD_{irc} + \sum_c \alpha_c I_c + \sum X_{irc} + \epsilon_{irc} \tag{5.45}$$

$$FloodPrev_{irc} = FloodD_i \times \sum_{k=1}^{4} (\beta_{fk} DistQuartile_{kc}) + \sum_c \alpha_c I_c + \sum X_{irc} + \epsilon_{irc} \tag{5.46}$$

where $FloodPrev_{irc}$ is the length of flood prevention facilities within a grid of 1 $km \times$ 1 km. The regression (5.45) initially examines the overall effect of flood exposure and flood frequency on the provision of flood mitigation facilities. The second regression in (5.46) further introduces the interaction between flood exposure and different quartiles of the distance to CBD. This allows us to look at the non-linear relationship between flood exposure at different locations in urban areas.

Table 5.8: Regression results: Flood prevention infrastructure and flood hazards

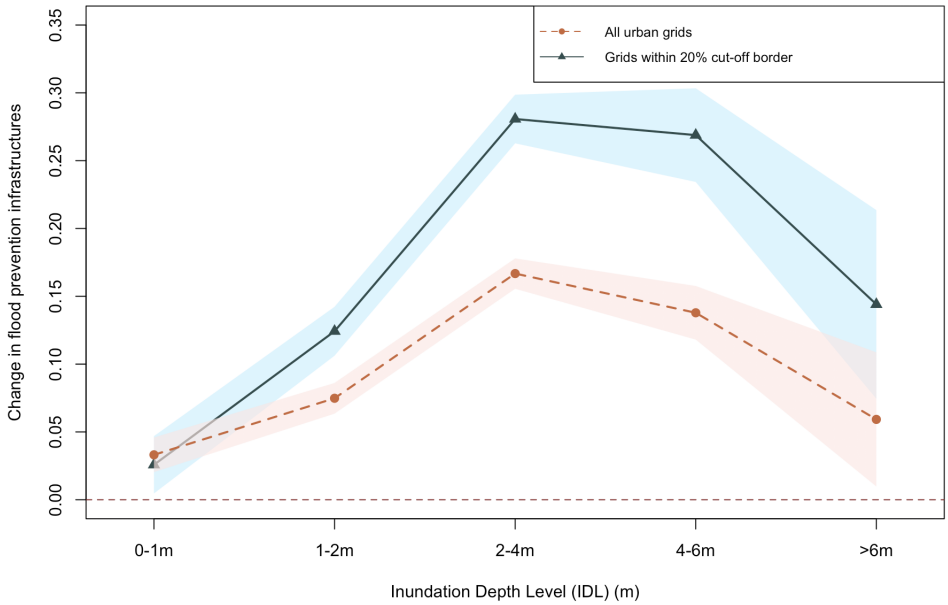
	Dependent variable: Flood Prevention Infrastructures Level (km)					
	(1)	(2)	(3)	(4)	(5)	(6)
Flood Exposure	0.100***			0.161***		
	(0.019)			(0.033)		
Flood Frequency		0.610***			1.014***	
		(0.120)			(0.198)	
Flood Exp. × Dist. to CBD			0.177***			0.290**
(1^{st} quartile)			(0.039)			(0.126)
Flood Exp. × Dist. to CBD			0.039**			0.184***
(2^{nd} quartile)			(0.019)			(0.054)
Flood Exp. × Dist. to CBD			0.034			0.152***
(3^{rd} quartile)			(0.023)			(0.047)
Flood Exp. × Dist. to CBD			0.014			0.099**
(4^{th} quartile)			(0.012)			(0.039)
Dist. to CBD	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y
Geographical controls	Y	Y	Y	Y	Y	Y
Dist. to river	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	Y	Y	Y	Y	Y	Y
Artificial Amenity Controls	Y	Y	Y	Y	Y	Y
River Buffer Zones	$<$ 10 km	$<$ 10 km	$<$ 10 km	$<$ 10 km	$<$ 10 km	$<$ 10 km
Grid Sample	Urban grids			Within 20% cut-off border		
Adjusted R^2	0.142	0.147	0.160	0.163	0.174	0.170
N	6,104	6,104	6,104	3,662	3,662	3,662

Note: This table reports the OLS estimates where the dependent variable is the flood prevention infrastructure within $1km \times 1km$ grids. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (3) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (4) to (6), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

Table 5.8 reports the regression results. We consider all the possible controls that we apply in the previous section. Columns (1) and (4) indicate the effect of *Flood Exposure* on the level of flood infrastructure. Specifically, for a grid located in a hazardous area, there is an approximately 0.1 to 0.16 km increase in the length of flood prevention infrastructure. The results in Columns (2) and (5) also show a significant and positive coefficient for flood frequency, suggesting that an increase in flood frequency leads to a higher level of flood protection, on average. In particular, an increase of one additional flood event every 10 years (equivalent to a 0.1 increase in annual flood frequency) leads to an increase of 0.06 to 0.1 km in the length of flood prevention facilities. Columns (3) and (4) show the estimation results of the regression (5.46). The regression results show positive coefficients for all interactions between *Flood Exposure* and different quartiles of distance to the CBD. Interestingly, the magnitude of the estimators decreases with higher distance quartiles. Specifically, using the results in Column (6), a grid located in

a floodplain zone in the city core (within the first distance quartile) has 0.29 km more flood prevention facilities. Meanwhile, a grid in a floodplain zone in the urban fringe (the last quartile) has only 0.099 km more of such facilities, which is just one-third the magnitude for the city core.

Figure 5: Flood prevention infrastructures and IDL exposure



This figure demonstrates the relationship between the change in flood prevention infrastructure (measured as kilometers of "Ouvrage de protection") relative to the unexposed location and the IDL exposure categories. These coefficients are based on the regression of the stock of the flood prevention infrastructures and five different categories of IDL from Very Low level (0-1m) and Very High level (above 6m) will all controls as in Column (5) of Table 5.8.

Figure 5 illustrates the effect of IDL exposure on the amount of flood prevention facilities. In the figure, we estimate equation (5.45) but substitute *Flood Exposure* with a series of IDL indicators. The regression with different indicators of IDL can reveal the non-linear relationship between the level of potential flood damage and the provision of flood defense systems. Across all coefficients in Figure 5, we see a positive and significant increase in the provision of flood defense systems with all IDL indicators relative to unexposed grids. However, the magnitude is not linear with IDL. In particular, we observe an inverted U-shape pattern for the change in flood defense systems and the level of potential flood damage, as shown by IDL indicators. Flood prevention facilities first increase with the IDL until it is equal to 4 m; after this point, it starts to decrease. This means that in our sample, the increase in flood protection coverage dominates up to an IDL of 4 m. After this point, the increase in the marginal cost of providing an additional unit of flood prevention dominates and leads to a decrease in flood prevention, even with a higher level of flood hazard or potential flood damage. Another expla-

nation probably is the zoning policies in highly inundated areas, resulting in lower population, housing construction and lower flood mitigation infrastructures.

5.3.5 Land use for local public goods within flooded zone

In this section, we study the effect of flood hazard on the provision of amenity land. The theoretical model shows an ambiguous relationship between the provision of amenity land and the level of flood hazard (see Proposition 4).

Table 5.9: Regression results: Amenity land uses and flood hazards

	Dependent variable: Share of Amenity Land Use					
	(1)	(2)	(3)	(4)	(5)	(6)
Flood Exposure	0.003*			0.006**		
	(0.002)			(0.003)		
Flood Frequency		0.030***			0.062***	
		(0.010)			(0.016)	
Flood Exp. × Dist. to CBD			0.005**			−0.007
(1^{st} quartile)			(0.002)			(0.007)
Flood Exp. × Dist. to CBD			−0.0004			0.015***
(2^{nd} quartile)			(0.003)			(0.004)
Flood Exp. × Dist. to CBD			0.002			0.005
(3^{rd} quartile)			(0.004)			(0.004)
Flood Exp. × Dist. to CBD			0.006			0.005
(4^{th} quartile)			(0.006)			(0.004)
Dist. to CBD	Y	Y	Y	Y	Y	Y
Dist. to CBD (sq)	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y
Geographical controls	Y	Y	Y	Y	Y	Y
Dist. to river	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	Y	Y	Y	Y	Y	Y
River Buffer Zones	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km	< 10 km
Grid Sample	Urban grids			Within 20% cut-off border		
Adjusted R^2	0.228	0.228	0.228	0.227	0.227	0.227
N	22,239	22,239	22,239	11,269	11,269	11,269

Note: This table reports the OLS estimates where the dependent variable is the share of amenity land within $1km \times 1km$ grids. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (3) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (4) to (6), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

We follow the same strategies and regressions as with flood prevention infrastructure in the previous section. In particular, we run the following regressions:

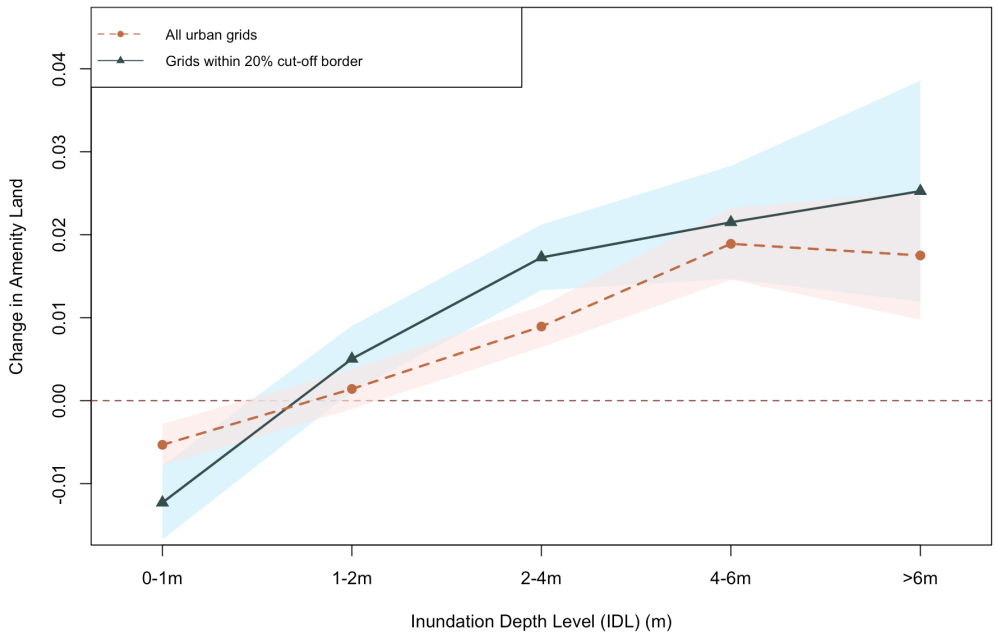
$$AmenityLand_{irc} = \beta_f FloodD_{irc} + \alpha_c X_c + \gamma_{irc} + \epsilon_{irc} \tag{5.47}$$

$$AmenityLand_{irc} = FloodD_i \times \sum_{k=1}^{4} (\beta_{fk} DistQuartile_{kc}) + \alpha_c X_c + \gamma_{irc} + \epsilon_{irc} \tag{5.48}$$

where $AmenityLand_{irc}$ is the share of urban green space and sport and leisure facility land uses for each grid in our samples. We also consider different quartile levels as in the previous section.

Table 5.9 shows the results. Columns (1) and (4) show a significant and positive coefficient for *Flood Exposure* with amenity land use. However, the magnitude is small. A grid located in a floodplain area induces a 0.3% to 0.6% increase in the amenity land. A similar regression with flood frequency level also shows a significant and positive sign. Specifically, one additional flood event every 10 years would increase the provision of amenity land in the same grid by 0.3% to 0.62% relative to unexposed grids. Surprisingly, the interaction between *Flood Exposure* and the distance quartiles do not result in a significant coefficient except for the first and second quartiles.

Figure 6: Amenity land uses and IDL exposure



This figure demonstrates the relationship between the change in amenity land uses relative to the unexposed location and the IDL exposure categories. These coefficients are based on the regression of the stock of the flood prevention infrastructures and five different categories of IDL from Very Low level (0-1m) and Very High level (above 6m).

Figure 6 shows the effect of different inundation levels on the provision of amenity land. We obtain the coefficients by rerunning regression (5.47) but replace the exposure indicators with IDL indicators. As shown in Figure 6, there is a mostly linear relationship between IDL and the provision of amenity land for both urban grids and those within a circular urban border. However, there is a switch in sign from negative to positive from very low IDL below 1 m and IDL above 1 m. This pattern could be related to the positive coefficient in population density for very low IDL, as shown in Figure 4.

The gradual increase in amenity land with respect to IDL suggests that the cost reduction effect dominates in equation (5.28) at a higher IDL. Only at very low IDL (below 1 m) does less housing construction dominate, leading to less amenity land at very low IDL.

5.4 Conclusion

To what extent do city structure and land use planning reflect climatic risks, specifically urban flood risks? This issue is of growing interest due not only to its policy importance (Anderson et al., 2018) but also to its relation to the central question regarding the increasing vulnerability of urban zones. Flooding has long been one of the costliest natural disasters worldwide, and risks are set to increase with the effects of climate change and rising sea levels over the coming decades.

In this paper, we study the effect of flood risks on urban structures and optimal land uses. In particular, we examine theoretically (i) whether population density, housing demand, and housing constructions concentrate disproportionately in flood-prone urban areas; (ii) whether higher risks lead to higher levels of private and public flood protection efforts; and (iii) whether the amenity provision is changed due to flood risk within the urban area. Our model anticipates a lower population density and a sparser residential construction area in a hazardous area. It also predicts a nonmonotonic relationship between the level of flood risks and the optimal provision of public flood defenses and amenity land. In the second-best case, the city government collects taxes to cover the expenses of flood defense systems as well as the provision of amenity land. It faces budget balance constraints. We find that a locational tax is efficient while an income tax system would result in over-concentration of population and housing construction in a flood-prone area.

The empirical studies across European cities confirm the persistent and negative impacts of flood risk on urban population density and residential land covers, which confirms the prediction from our theoretical model. Second, the provision of amenity lands is nonmonotonic with the level of flood risks and depends on the relative location within the urban area. Specifically, the fraction of amenity land is more likely to be built in a hazardous and compact area than in a risky but sparse area near the urban fringe. These results further contribute to our understanding of the effects of river flood risk on land use planning policies.

Cities are enormous and intricate economic engines, but ultimately, they are creatures of human free will and respond to people's desires for a livable environment. People can optimize the use of land within urban locations to help mitigate the drawbacks and embellish the strengths. We believe that the environments that we have lived in are those that we created.

Appendix A: Second-best world with locational tax system

Here, we assume that the burden of the locational tax $\tau(r,\phi)$ is on the land developers.[18] We first study the housing demand from households' side, then move to the land developers' problem once the bid-rent function is determined, and the decisions of the city government over the provision of public flood mitigation and amenity land.

Housing demand

The household chooses z,h,e. During a unique stage, the household finds z,s,e that

$$\max_{z,h,e} U(z,h,a) \text{ s.t. } z+t+Rh+h\int_0^1 \Big(l(x)(1-m(e,p))+C^H(e)\Big)\,\mathrm{d}F(x;r,\phi)=w. \tag{5.49}$$

One can interpret this as the consumer considering her permanent income (she can borrow to smooth her consumption).

Denote again $l = l(r,\phi) \equiv \int_0^1 l(x)\mathrm{d}F(x;r,\phi)$. We will drop the location indicator (r,ϕ) where there is no confusion in the notation. Each household will bid for the location of their preference, and the slot will be given to the highest bidder. Hence, we can write the bid rent function as follows The household program is given by equation (5.49). The bid rent function is given by

$$\mathrm{Y} = \max_{z,h,e}\left(\frac{w-t-z}{h} - l(1-m(e,p)) - C^H(e)\right) \text{ s.t. } U(z,h,a)=u \tag{5.50}$$

We have utility function U is strictly increasing in z; therefore, we can write $\hat{z}(u,h,a)$ as the unique solution of equation $U(z,h,a) = u$, at each level of housing consumption and amenity level of indifferent curve. For general form of utility, we have $z_h = -U_h/U_z < 0$, $z_a = -U_a/U_z < 0$ as utility is increasing function in all z, h, a; therefore, the consumption of composite good along the indifferent curve is decreasing with the housing floor service h and amenity level a. Then, the bid rent function can be defined as follows:

$$\mathrm{Y} = \max_{h,e}\left(\frac{w-t-\hat{z}(u,h,a)}{h} - l(1-m(e,p)) - C^H(e)\right) \tag{5.51}$$

The first-order conditions give

$$\hat{z} - h\hat{z}_h = y \tag{5.52}$$

$$l \times m_e(e,p) = C^H_e(e) \tag{5.53}$$

where $y = w - t$ is the disposable income for each household. We obtain $\overline{h}(y,a,u)$, $\overline{z}(y,a,u)$ and $\overline{e}(p,l)$, which gives the equilibrium housing floor value

$$R = V(y,a,u,p,l) \equiv \frac{y-\overline{z}}{\overline{h}} - l(1-m(\overline{e},p)) - C^H(\overline{e}). \tag{5.54}$$

[18] It is similar in the case of locational tax for households.

Housing supply

On the supply side, we assume the land is owned by absentee landlords, who have two options to either lend the piece of land for land developers to construct residential houses or to the agriculture sectors at R_A. At each location, a land developer pays a land price of Ψ to the landlords and a locational tax τ to build q units of housing floors that can be rented out for households at R.[19] The construction cost is κq^α with $\alpha > 1$ as in the previous section. The residential developers, then, choose the quantity housing floor q to maximize the total profit

$$\max_q \ \Pi = Rq - \kappa q^\alpha - \Psi - \tau$$

Taking the first-order condition w.r.t q, we have

$$\overline{q} = \left(\frac{R}{\alpha\kappa}\right)^{\frac{1}{\alpha-1}}$$

which is at the same level as in the first best case.

Additionally, the free entry condition dictates

$$\Pi = 0 \Longleftrightarrow \Psi = R\overline{q} - \kappa\overline{q}^\alpha - \tau = \frac{\alpha-1}{\alpha} R\left(\frac{R}{\alpha\kappa}\right)^{\frac{1}{\alpha-1}} - \tau.$$

The rent is at least equal to or higher than the outside opportunity cost of land which is the agricultural land rent R_A. Therefore, we can denote the actual land rent at location (r, ϕ) is as follows:

$$R^L_{(r,\phi)} = \max\{R_A; \Psi_{(r,\phi)}\}$$

Therefore, a location will be developed if and only if

$$\Psi_{(r,\phi)} \geq R_A \iff R \geq \underline{R}(R_A, \tau) \equiv \left(\frac{\alpha}{\alpha-1}(R_A+\tau)\right)^{\frac{\alpha-1}{\alpha}} (\alpha\kappa)^{\frac{1}{\alpha}} \tag{5.55}$$

From equation (5.54), we can see that the flood risk level reduces the bid rent of households, leading to lower house rent $R_l < 0$. If the effect of the flood risk is strong enough to reduce the bid rent of household R at some location to be lower than the threshold $\underline{R}$, then this location will be left undeveloped. This corner equilibrium can lead to the discontinuity in land use within the urban area. Specifically,

$$R \geq \underline{R}(R_A, \tau) \iff \underbrace{L(p,l) \equiv l(1 - m(\overline{e}, p)) + C^H(\overline{e})}_{\text{Loss due to flood hazard}} \leq \frac{y - \overline{z}(y,a,u)}{\overline{h}(y,a,u)} - \underline{R}(R_A, \tau)$$

$$\iff l \leq \underline{l}(p, R_A, \tau) \text{ where } \underline{l}(p, R_A, \tau) \equiv \frac{\frac{y-\overline{z}}{\overline{h}} - \underline{R}(R_A, \tau) - C^H(\overline{e})}{(1 - m(\overline{e}, p))}$$

Any location with average financial loss of flood risk l higher than $\underline{l}$ is not developed. One notices that $\underline{l}$ is decreasing with distance to the city center, R_A and the locational tax τ, meaning the restriction is stricter near urban fringe and with higher outside opportunity cost of land, all else being equal.

[19] As it is clear that the tax and land price are at each location (r, ϕ), so we drop the notation.

City government

The city government has three choices over $\tau(r,\phi)$, $p(r,\phi)$ and $g(r,\phi)$, and she chooses to minimize the total social cost while maintaining the same level of utility for every city inhabitant. We drop the location notation where it is not caused misunderstanding. The objective of the urban planner, now, can be written as follows.

$$ES = \max_{p,g} \int_0^b \int_0^{2\pi} \Big([V(y,a,u,p,l)\overline{q} - \kappa \overline{q}^{\alpha} - C^U(p)](1-g) - R_A - cg \Big) \ell \mathrm{d}\phi \mathrm{d}r \tag{5.56}$$

such that

$$U(z,h,a) \geq u.$$

and balanced budget at each location

$$\tau = C^U(p)(1-g) + cg \tag{5.57}$$

As the budget constraint is always binding at each location, and it is just an accounting problem, we can solve the city government problem first with the choice of public flood prevention p and amenity g. Taking the first order conditions of equation (5.56) w.r.t. p and g, we get

$$V_p \times \overline{q} = C_p^U \tag{5.58}$$

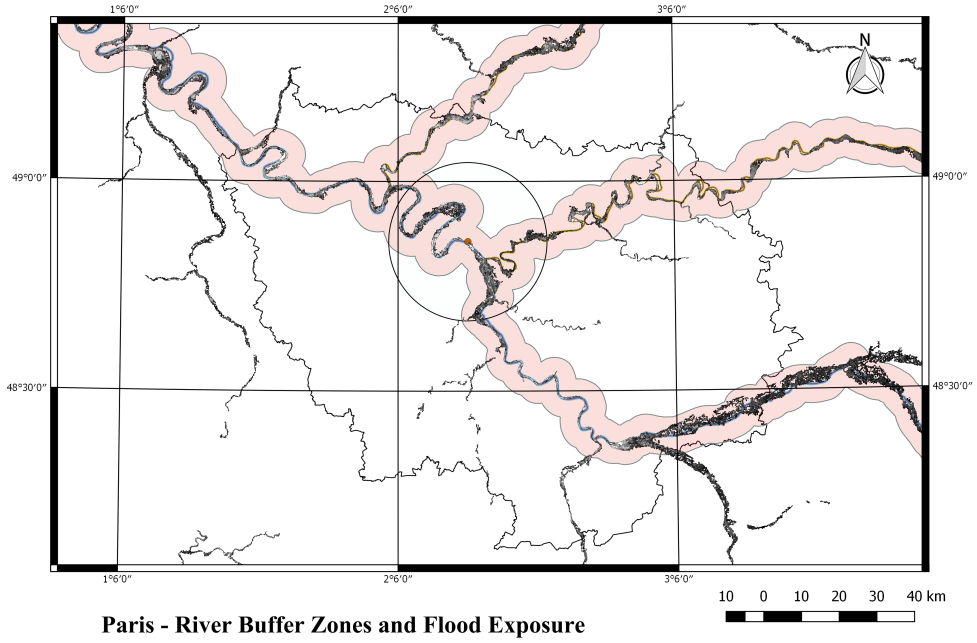
$$\theta \overline{q} V_a (1-g) = V\overline{q} - \kappa \overline{q}^{\alpha} - C^U(p) + c \tag{5.59}$$

which are identical to the first-best conditions. One also notice that in safe area, we have $V_p = l \times m_p = 0$, leading to zero public flood protection effort p; hence, the locational tax in safe zone only cover the cost for amenity provision $\tau_{safe} = cg$. Under locational taxation rule (5.57) and condition (5.55) binding for all locations within the urban area, the outcomes of the first-best optimization are decentralized through market equilibrium through locational taxation.

Appendix B: Flood risk and river networks

Figure B1: Flood Risk, River Networks and Land Use in Paris

(a) River Networks with Flood Hazard



(b) Continuous Urban Fabric, Urban Boundary and River Buffer Zones

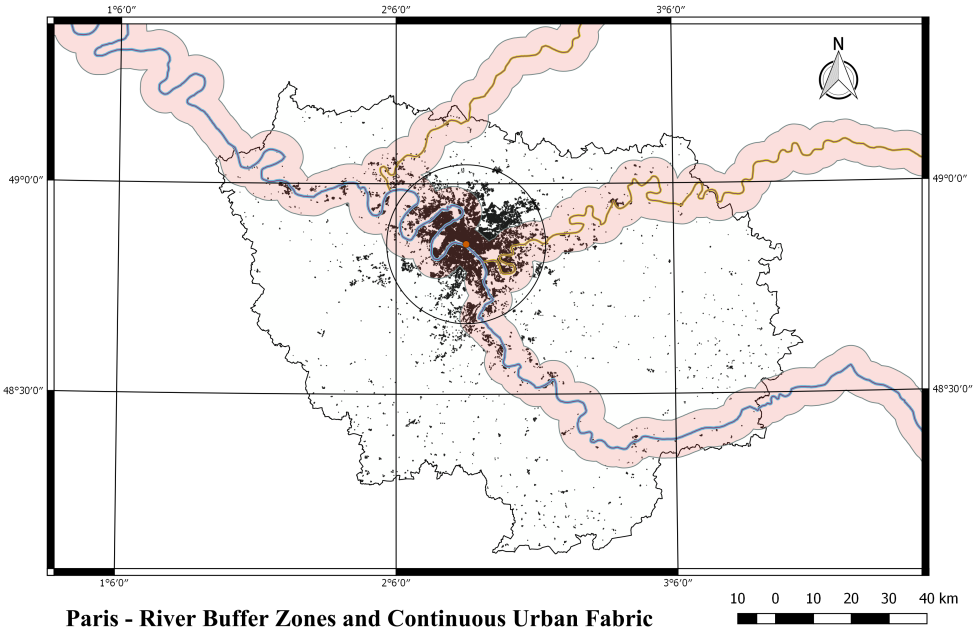
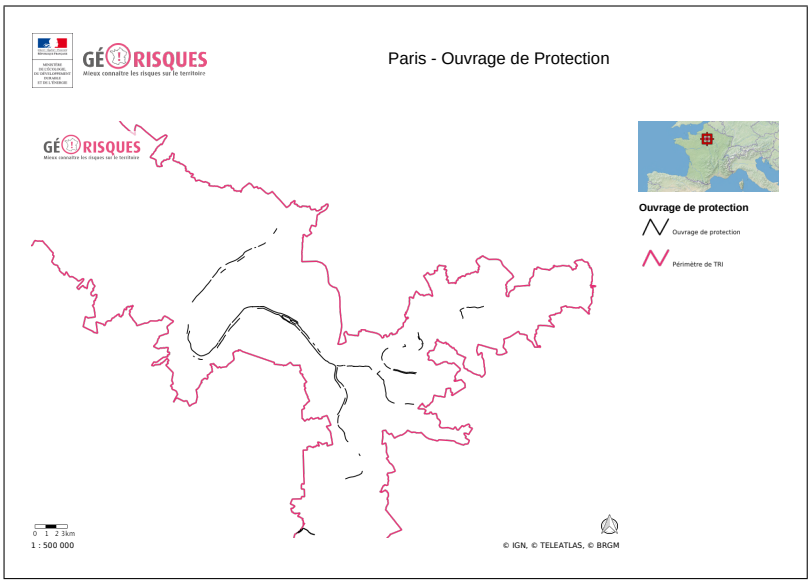
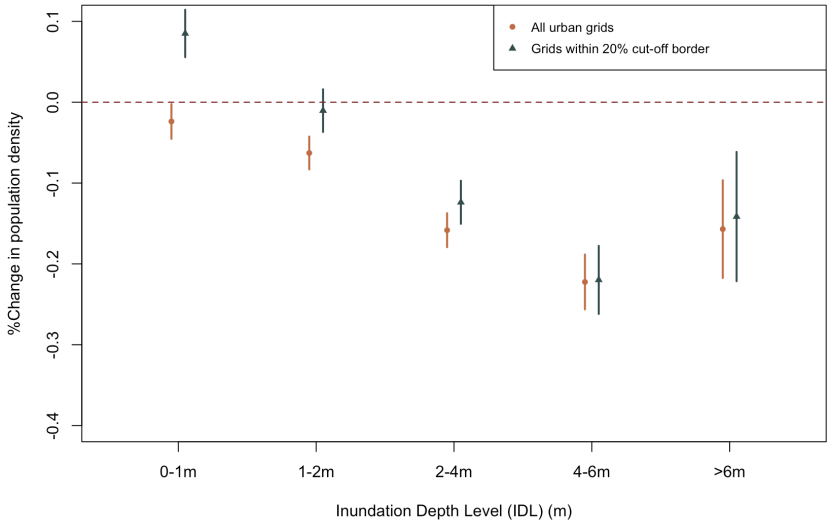


Figure B2: Paris - Flood Prevention Infrastructures



Appendix C: Robustness checks

Figure C1: Inundation Depth Level and population density (5 km buffer)



This figure denotes the relationship between the % change in population density of the grids within the zones with flood exposure (in comparison with unexposed grids), partitioned by different levels of inundation depth level. We categorize the flood hazard according to the exposure level from Very Low to Very High, which are defined as the average Inundation Depth Level (IDL) in all return periods. We set the IDL between 0 and 1 meter as very Low level and above 6 meters in Very High level.

Table C1: Population density and flood hazard - Robustness check

	Dependent variable: Ln(Pop)							
Panel A:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Exposure	0.077^{**}	-0.068^{*}	-0.127^{***}	-0.101^{***}	-0.091^{***}	-0.205^{***}	-0.055	-0.041
	(0.032)	(0.038)	(0.041)	(0.033)	(0.035)	(0.049)	(0.037)	(0.036)
Dist. to CBD	Y	Y	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y	Y	Y
Geographical controls	N	Y	Y	Y	Y	Y	Y	Y
Dist. to river	N	N	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	N	N	N	Y	Y	N	Y	Y
Artificial Amenity Controls	N	N	N	N	Y	N	N	Y
River Buffer Zones	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km
Grid Sample	Urban grids	Urban grids	Urban grids	Urban grids	Urban grids	Within 20% cut-off border	Within 20% cut-off border	Within 20% cut-off border
Adjusted R^2	0.399	0.421	0.424	0.612	0.638	0.457	0.701	0.726
N	15,014	15,014	15,014	15,014	15,014	7,964	7,964	7,964
Panel B:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Frequencies	0.229	-0.550^{***}	-0.929^{***}	-0.745^{***}	-0.639^{***}	-1.680^{***}	-0.702^{***}	-0.565^{***}
	(0.185)	(0.211)	(0.230)	(0.183)	(0.191)	(0.326)	(0.199)	(0.193)
Dist. to CBD	Y	Y	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y	Y	Y
Geographical controls	N	Y	Y	Y	Y	Y	Y	Y
Dist. to river	N	N	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	N	N	N	Y	Y	N	Y	Y
Artificial Amenity Controls	N	N	N	N	Y	N	N	Y
River Buffer Zones	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km
Grid Sample	Urban grids	Urban grids	Urban grids	Urban grids	Urban grids	Within 20% cut-off border	Within 20% cut-off border	Within 20% cut-off border
Adjusted R^2	0.399	0.422	0.425	0.612	0.639	0.461	0.702	0.727
N	15,014	15,014	15,014	15,014	15,014	7,964	7,964	7,964

Note: This table reports the OLS estimates where the dependent variable is the natural log of population count in $1k \times 1km$ grid. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (5) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (6) to (8), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for $p<0.1$, ** for $p<0.05$ and *** for $p<0.01$. Standard errors are clustered at city level and reported in parentheses.

Table 5.10: Table C2: Housing construction and flood hazard - Robustness check

Panel A:	*Dependent variable:* Building Height							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Exposure	−0.175	−0.517***	−0.586***	−0.524***	−0.464***	−0.552***	−0.496***	−0.434***
	(0.144)	(0.161)	(0.147)	(0.118)	(0.119)	(0.116)	(0.129)	(0.109)
Adjusted R^2	0.422	0.458	0.467	0.522	0.600	0.485	0.540	0.626
N	2,903	2,903	2,903	2,903	2,903	2,375	2,375	2,375
Panel B:	*Dependent variable:* Share of Continuous Residential Categories (S.L > 80%)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flood Exposure	−0.006	−0.015***	−0.025***	−0.023***	−0.021***	−0.036***	−0.029***	−0.024***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.007)	(0.006)	(0.006)
Adjusted R^2	0.228	0.238	0.246	0.352	0.429	0.386	0.425	0.516
N	15,093	15,093	15,093	15,093	15,093	8,026	8,026	8,026
Dist. to CBD	Y	Y	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y	Y	Y
Geographical controls	N	Y	Y	Y	Y	Y	Y	Y
Dist. to river	N	N	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	N	N	N	Y	Y	N	Y	Y
Artificial Amenity Controls	N	N	N	N	Y	N	N	Y
River Buffer Zones	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km
Grid Sample	Urban grids	Urban grids	Urban grids	Urban grids	Urban grids	Within 20% cut-off border	Within 20% cut-off border	Within 20% cut-off border

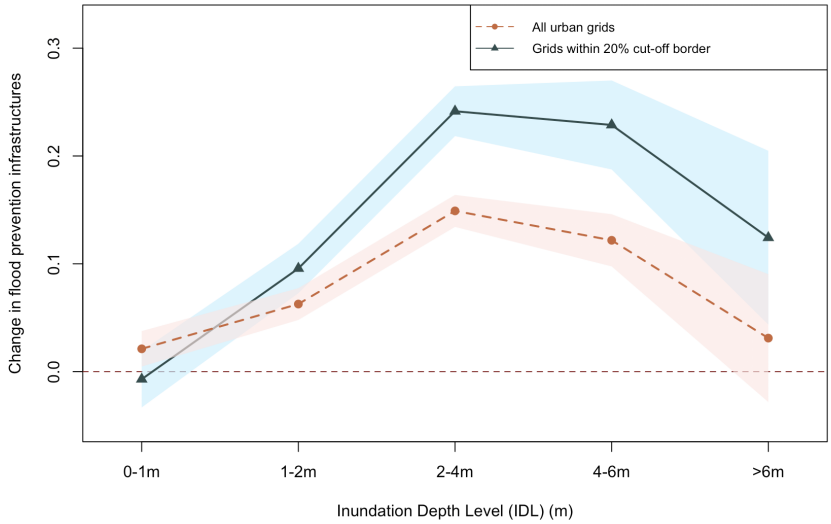
Note: This table reports the OLS estimates where the dependent variable is the natural log of population count in $1km \times 1km$ grid. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (5) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (6) to (8), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

Table C3: Flood prevention infrastructure and flood hazards - Robustness check

	Dependent variable: Flood Prevention Infrastructures Level (km)					
	(1)	(2)	(3)	(4)	(5)	(6)
Flood Exposure	0.082***			0.122***		
	(0.018)			(0.031)		
Flood Frequency		0.542***			0.856***	
		(0.114)			(0.175)	
Flood Exp. × Dist. to CBD			0.155***			0.214*
(1^{st} quartile)			(0.038)			(0.125)
Flood Exp. × Dist. to CBD			0.013			0.137***
(2^{nd} quartile)			(0.019)			(0.050)
Flood Exp. × Dist. to CBD			0.022			0.119**
(3^{rd} quartile)			(0.023)			(0.048)
Flood Exp. × Dist. to CBD			−0.006			0.076*
(4^{th} quartile)			(0.021)			(0.041)
Dist. to CBD	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y
Geographical controls	Y	Y	Y	Y	Y	Y
Dist. to river	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	Y	Y	Y	Y	Y	Y
Artificial Amenity Controls	Y	Y	Y	Y	Y	Y
River Buffer Zones	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km
Grid Sample	Urban grids			Within 20% cut-off border		
Adjusted R^2	0.159	0.164	0.176	0.180	0.190	0.183
N	4,390	4,390	4,390	2,636	2,636	2,636

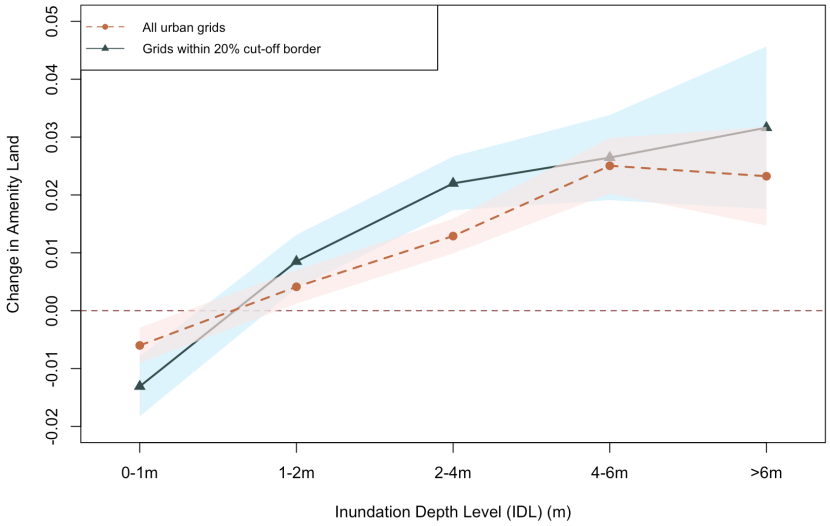
Note: This table reports the OLS estimates where the dependent variable is the flood prevention infrastructure within $1km \times 1km$ grids. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (3) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (4) to (6), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

Figure C2: Flood prevention infrastructures and IDL exposure (5km buffer)



This figure demonstrates the relationship between the change in flood prevention infrastructure (measured as kilometers of "Ouvrage de protection") relative to the unexposed location and the IDL exposure categories. These coefficients are based on the regression of the stock of the flood prevention infrastructures and five different categories of IDL from Very Low level (0-1m) and Very High level (above 6m).

Figure C3: Amenity land uses and IDL exposure (5km buffer)



This figure demonstrates the relationship between the change in amenity land uses relative to the unexposed location and the IDL exposure categories. These coefficients are based on the regression of the stock of the flood prevention infrastructures and five different categories of IDL from Very Low level (0-1m) and Very High level (above 6m).

Table C4: Amenity land uses and flood hazards - Robustness check

	Dependent variable: Amenity Land Use (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Flood Exposure	0.006**			0.009*		
	(0.003)			(0.005)		
Flood Frequency		0.051***			0.086***	
		(0.014)			(0.025)	
Flood Exp. × Dist. to CBD			0.009**			−0.003
(1^{st} quartile)			(0.003)			(0.009)
Flood Exp. × Dist. to CBD			−0.0003			0.021***
(2^{nd} quartile)			(0.003)			(0.006)
Flood Exp. × Dist. to CBD			0.005			0.006
(3^{rd} quartile)			(0.005)			(0.007)
Flood Exp. × Dist. to CBD			0.006			0.006
(4^{th} quartile)			(0.011)			(0.006)
Dist. to CBD	Y	Y	Y	Y	Y	Y
Dist. to CBD (sq)	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y
Geographical controls	Y	Y	Y	Y	Y	Y
Dist. to river	Y	Y	Y	Y	Y	Y
Natural Amenity Controls	Y	Y	Y	Y	Y	Y
River Buffer Zones	< 5 km	< 5 km	< 5 km	< 5 km	< 5 km	< 10 km
Grid Sample	Urban grids			Within 20% cut-off border		
Adjusted R^2	0.227	0.228	0.227	0.215	0.216	0.216
N	15,093	15,093	15,093	8,026	8,026	8,026

Note: This table reports the OLS estimates where the dependent variable is the flood prevention infrastructure within $1km \times 1km$ grids. The explanatory variable of interest is the Flood Exposure which equal to 1 if the cell is in floodplain area and 0 otherwise. We, then, further use the flood annual frequencies (which is the inverse of flood return periods). Column (1) to (3) includes all the grids with urban status equal to 1. The cells that have urban status equal to 1 are those with at least 20% of artificial land use within the grid. Column (4) to (6), we consider all cells within a circular urban boundary. Urban boundaries are defined as the circular cut-off at which the total urban fabric (or residential land use) are of at least 20% of total land within its rings. For more details, see Picard and Tran (2019). Other controls include the coastal grid indicator. Significance levels are denoted by * for p<0.1, ** for p<0.05 and *** for p<0.01. Standard errors are clustered at city level and reported in parentheses.

Bibliography

[1] Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P.D., Feyen, L. (2014). *Advances in pan-European flood hazard mapping*, Hydrol. Process., 28 (18), 4928-4937, doi:10.1002/hyp.9947

[2] Alonso W. (1964). Location and Land Use. Harvard Univ. Press, Cambridge, MA, 1964.

[3] Anderson S.E., T.L. Anderson, A.C. Hill, M.E. Kahn, H. Kunreuther, G.D. Libecap, H. Mantripragada, P. Merel, A. Plantinga, and V.K. Smith (2018). The Critical Role of Markets in Climate Change Adaptation. NBER Working Paper No. w24645.

[4] Anderson S.T. and S.E. West (2006). Open Space, residential property values, and spatial context. *Regional Science and Urban Economics* 36, 773-789.

[5] Arnott, R. J., and Stiglitz, J. E. (1979). Aggregate land rents, expenditure on public goods, and optimal city size. *The Quarterly Journal of Economics*, 93(4), 471-500.

[6] Atreya A., Ferreira S., and Kriesel (2013). Forgetting the Flood? An Analysis of the Flood Risk Discount over Time. *Land Economics*, Vol 89, No.4, Nov 2013, pp. 577-596.

[7] Bairoch P., Batou J. and Chèvre P. (1988). La Population des villes européennes de 800 à 1850 : banque de données et analyse sommaire des résultats. Publications d'histoire économique et sociale internationale. Ed. Droz. France.

[8] Bakkensen L. A., and Barrage L. (2018). Flood Risk Belief Heterogeneity and Cotal Home Price Dynamics: Going Under Water? NBER Working Paper 23854.

[9] Barthel S., C. Folke, J. Colding (2010). Social–ecological memory in urban gardens: retaining the capacity for management of ecosystem services. *Global Environmental Change* 20, 255-265.

[10] Bayer B. and Timmins C. (2005). Estimating equilibrium models of sorting across locations. *Economic Journal* 117(3). 353-374.

[11] Berman M.G., J. Jonides, and S. Kaplan (2008). The Cognitive Benefits of Interacting With Nature. *Psychological Science* 19, 1207. DOI: 10.1111/j.1467-9280.2008.02225.x

[12] Bernstein A., Gustafson M. T., and Lewis R. (2018). Disaster on the horizon: the price effect of sea level rise. *Journal of Financial Economics*, forthcoming.

[13] Bolitzer B. and N.R. Netusil (2000). The impact of open spaces on property values in Portland, Oregon. *Journal of Environmental Management* 59, 185-193.

[14] Bowler D.E., L. Buying-Ali, T.M. Knight, and A.S. Pullin (2010). Urban Greening to cool towns and cities: a systematic review of the empirical evidences. *Landscapes an Urban Planning* 97, 147-155.

[15] Brander L.M. and M.J Koetse (2011). The value of urban open space: Meta-analyses of contingent valuation and hedonic pricing results. *Journal of Environmental Management* 92(10), 2763-2773.

[16] Brack, C. L. (2002). Pollution mitigation and carbon sequestration by an urban forest. *Environmental Pollution*, 116, 195–200.

[17] Brakenridge, G.R. Global Active Archive of Large Flood Events, Dartmouth Flood Observatory, University of Colorado.

[18] Brueckner J. K. (1983). The Economics of Urban Yard Space: An 'Implicit-Market' Model for Housing Attributes. *Journal of Urban Economics* 13, 216-234 (1983).

[19] 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.

[20] Capozza, D. R., & Helsley, R. W. (1990). The stochastic city. *Journal of urban Economics*, 28(2), 187-203.

[21] Caruso G., D. Peeters, J. Cavailhes, and M. Rounsevell (2007). Spatial configuration in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics* 37(5), 542-567.

[22] Cheshire P. and S. Sheppard (2002). The welfare economics of land use planning. *Journal of Urban Economics* 52, 242-269.

[23] Cho S., J. Bowker, and W. Park (2006). Measuring the contribution of water and green space amenity to housing values: An application and comparison of spatially weighted hedonic models. *Journal of Agricultural and Resource Economics* 31(3), 485-507.

[24] Cohen D.A., T.L. MacKenzie, A. Sehgal, S. Williamson, D. Golinelli, and N. Lurie (2007). Contribution of Public Parks to Physical Activity. *American Journal of Public Health*, 97(3), pp. 509-514.

[25] Cohen D.A., B. Han, C.J. Nagel, P. Harnik, T.L. MacKenzie, K.R. Evenson, T. Marsh, S. Williamson, C. Vaughan, and S. Katta (2016). The First National Study of Neighborhood Parks: Implications for Physical Activity. *American Journal of Preventive Medicine*, 54(4), pp. 419-426.

[26] Colding J. and S. Barthel (2013). The potential of 'Urban Green Commons' in the resilience building of cities. *Ecological Economics* 86, 156-166.

[27] Cremer H., A. De Kerchove, and J-F. Thisse (1985). An economic theory of public facilities in space. *Mathematical Social Sciences* 9, 249-262.

[28] Cribari-Neto F. and Da Silva W.B. (2011). A New Heteroskedasticity-Consistent Covariance Matrix Estimator for the Linear Regression Model. *Advances in Statistical Analysis*, 95(2), 129-146.

[29] Davis, M.A, and Ortalo-Magne, F. (2011). Household Expenditures, wages, rents. *Review of Economic Dynamics* 14 (2), 248-261.

[30] Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F.A., Feyen, L. (2016a). Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources* 94, 87-102.

[31] Dottori, Francesco; Alfieri, Lorenzo; Salamon, Peter; Bianchi, Alessandra; Feyen, Luc; Lorini, Valerio (2016b): Flood hazard map for Europe - 100-year return period. European Commission, Joint Research Centre (JRC).

[32] E-OBS database (2017 release version). We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu).

[33] Ellison G., E.L. Glaeser, and W.R. Kerr (2010). What Causes Industry Agglomeration: Does Natural Advantage Explain Agglomeration? *American Economic Review* 100, 1195-1213.

[34] Elsasser P. (1996) Der Erholungwert des Waldes: Monetare Bewertung der Erholungsleistung ausgewählter Wälder in Deutchland (The recreational value of forests: Monetary valuation of the recreational performance of selected forests in Germany). Schriften zur Forstökonomie Bd. 11. Sauerländer, Frankfurt am Main, (in German).

[35] EEA (2009). WISE Large River and Large Lakes. Water Pattern Europe, scale 10 million, version 2, from EUROSTAT GISCO database; Water Framework Directive article 3 data on rivers and lakes from countries; Joint Research Centre catchment database CCM1.

[36] Epple D. and H. Sieg (1999). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy* 107 (4), 645-681.

[37] EU DG-ENVI (2012). The multi-functionality of green infrastructure. Science for Environmental Policy, In-depth Reports.

[38] Eurostat (2016). Urban Europe: Statistics on cities, towns and suburbs. 2016 Edition. ISBN 978-92-79-60139-2.

[39] Evenson K.R., F. Wen, D. Golinelli, D.A. Rodriguez, and D.A. Cohen (2013). Measurement Properties of a Park Use Questionnaire. *Environmental and Behavior*, 45(4), pp. 526-547.

[40] Fahey, T., Nolan, B., Maitre, B. (2004) Housing expenditures and income poverty in EU countries. *Journal of Social Policy* 33 (3), 437-454.

[41] Fallick B., C.A. Fleischmann, and J.B. Rebitzer (2006). Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster. *Review of Economics and Statistics* 88(3), 472-481.

[42] Frame, D.E. (1998). Housing, Natural Hazards, and Insurance. *Journal of Urban Economics* 44, 93-109 (1998).

[43] Freeman A.M., J.A. Herriges, and C.L. Kling (2014). The measurement of Environmental and Resource Values: Theory and Method. Third Edition. Resources for the Future - RFF Press, Routledge, NY.

[44] Fujita M. (1986). Optimal Location of public facilities - Area Dominance Approach. *Regional Science and Urban Economics* 16, 241-268.

[45] Fujita M. (1989). Urban Economic Theory: Land Use and City Size. Cambridge Univ. Press.

[46] Fujita M. and J-F. Thisse (2002). Economics of Agglomeration: Cities, Industrial Location, and Regional Growth. Cambridge Univ. Press.

[47] Gallagher, Justin (2014). Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States. *American Economic Journal: Applied Economics* 2014, 6(3): 206-233.

[48] Geoghegan J., L.A. Waigner, and N.E. Bockstael (1997). Spatial landscape indices in a hedonic framework: an ecological economics using GIS. *Ecological Economics* 23, 251-264.

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

[50] Glaeser E.L., and J. Gottlieb (2009). The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. *Journal of Economic Literature* 47(4), 983-1028.

[51] Glaeser E.L. (2011). Cities, Productivity and the Quality of Life. *Science* 333: 592-594

[52] Glaeser E.L. (2012). The Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier. *New York: The Penguin Press*.

[53] Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, M. New. 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res* (Atmospheres), 113, D20119, doi:10.1029/2008JD10201

[54] Hedblom M., B. Gunnarsson, B. Iravani, et al. (2019). Reduction of physiological stress by urban green space in a multisensory virtual experiment. *Nature, Scientific Reports* 9, 10113. `https://doi.org/10.1038/s41598-019-46099-7`

[55] Heidt V. and M. Neef (2008). Benefits of Urban Green Space for Improving Urban Climate. In: Carreiro M.M., Y.C. Song, J. Wu (eds) Ecology, Planning, and Management of Urban Forests, Springer, New York, NY.

[56] IFPRA (2013). Benefits of Urban Parks - A systematic review. Copenhagen, International Federation of Parks and Recreation Administration. 68p.

[57] Indaco Agustín, Francesc Ortega, and Süleyman Taspmar (2019). The Effects of Flood Insurance on Housing Markets. Working paper 2019. [Older version: IZA DP No. 11810].

[58] Kabisch N. and D. Haase (2014). Green justice or just green? Provision of green urban spaces in Berlin, Germany. *Landscape and Urban Planning* 122 (0), 129-139.

[59] Klaiber H. A. and D.J. Phaneuf (2010). Valuing open space in a residential sorting model of the Twin Cities. *Journal of Environmental Economics and Management* 60, 57-77.

[60] Kocornik-Mina A., McDermott, G. Michaels, F. Rauch (2019). Flooded Cities. *American Economic Journal: Applied Economics*, forthcoming.

[61] Kuminoff N.V., V:K: Smith, and C. Timmins (2013). The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets. *Journal of Economic Literature* 51(4), 1007-1062.

[62] Lee C. and M. Fujita (1997). Efficient configuration of a greenbelt: Theoretical modelling of greenbelt amenity. *Environment and Planning* A 29, 1999-2017.

[63] Lucas R.E. and E. Rossi-Hansberg (2002). On the Internal Structure of Cities. *Econometrica* 70, 1445-1476.

[64] Luechinger S. and Raschky P. A. (2009). Valuing flood disaster using the life satisfaction approach. *Journal of Public Economics* 93 (2009) 620-633.

[65] Luttik, J. (2000). The value of trees, water and open space as reflected by house prices in the Netherlands. *Landscape and Urban Planning* 48, 161-167.

[66] Lutzenhiser M. and N.R. Netusil (2001). The effect of open spaces on a home's sale price. *Contemporary Economic Policy* 19, 291-298.

[67] Manes F., G. Incerti, E. Salvatori, M. Vitale, M. Ricotta, and R. Constanza (2012). Urban ecosystem services: tree diversity and stability of tropospheric ozone removal. *Ecological Application* 22, 349-369.

[68] McFadden D. (1978). Modeling the choice of residential location. in: Krlqvist A., L. Lundqvist, and F. Snickars (Eds.) Spatial Interaction Theory and Planning Models. Amsterdam, North Holland, 1978.

[69] Mossay P. and P.M. Picard (2011). On spatial equilibria in a social interaction model. *Journal of Economic Theory*, 146(6), 2455-2477.

[70] OECD (2013). Definition of Functional Urban Areas (FUA) for the OECD metropolitan database. Link: https://www.oecd.org/cfe/regional-policy/Definition-of-Functional-Urban-Areas-for-the-OECD-metropolitan-database.pdf (acessed on Mar 01, 2018).

[71] Orford S. (2002). Valuing locational externalities: A GIS and multilevel modelling approach. *Environment and Planning B: Planning and Design* 29(1), 105-127.

[72] Ortega Francesc and Suleyman Taspinar (2018). Rising Sea Levels and Sinking Property Values: The Effects of Hurricane Sandy on New York's Housing Market. *Journal of Urban Economics* 106 (2018) 81-100.

[73] Owens, R., Rossi-Hansberg, E., Sarte, P. D. (2017). Rethinking Detroit. Federal Reserve Bank of Richmond Working Papers (2017), Vol. 17, Issues 04, 1-45.

[74] Picard P.M. and Tabuchi T. (2013). On microfoundations of the city. *Journal of Economic Theory* 148(6), 2561-2582.

[75] Picard P.M., and Tran, T.T.H (2019) Green Urban Spaces. CREA Discussion Papers 2019.

[76] Purdy J. (2015). After Nature - A politics for the Anthropocene. Harvard Univ. Press.

[77] Rappaport J. (2008). Consumption Amenities and City Population Density. *Regional Science and Urban Economics* 38(6), 533-552.

[78] Sakashita N. (1987). Optimum Location of Public Facilities under the influence of the land market. *Journal of Regional Science* 27.

[79] Saldivar-Tanaka L. and M. Krasny (2004). Culturing community development, neighborhood open space, and civic agriculture: the case of Latino community gardens in New York City. *Agriculture and Human Values* 21, 399-412.

[80] Schindler M., Le Texier M., and Caruso G. (2018). Spatial Sorting, attitudes and the use of green space in Brussels. *Urban Forestry & Urban Greening* 31, 169-184.

[81] Sieg H., V.K. Smith, H.S. Banzhaf, and R. Walsh (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International Economic Review* 45(4), 1047-1077.

[82] Smith S.K., C. Poulus, and H. Kim (2002). Treating open space as an urban amenity. *Resource and Energy Economics* 24, 107-129.

[83] Strohbach M.W., E. Arnold and D. Haase (2012). The carbon footprint of urban green space - A life cycle approach. *Landscape and Urban Planning* 104, 220-229.

[84] Tiebout C. M. (1956). A Pure Theory of Local Expenditures. *The Journal of Political Economy* 64(5), 416-424.

[85] Thisse J-F. and D. Wildasin (1992). Public facility location and urban spatial structure - Equilibrium and welfare analysis. *Journal of Public Economics*, 48, 83-118.

[86] Timmins C. (2007). If you can't take the heat, get out of Cerrado... recovering the equilibrium amenity cost of nonmarginal chimate change in Brazil. *Journal of Regional Science* 47(1), 1-25.

[87] Turnbull, G. K. (1991). The spatial demand for housing with uncertain quality. *Journal of Real Estate Finance and Economics*, 4, 5-17 (1991).

[88] Turner M. (2005). Landscape preferences and patterns of residential development. *Journal of Urban Economics* 57, 19-54.

[89] Turner M.A., A. Haughwout, and W. Van der Klaauw (2014). Land Use Regulation and Welfare. *Econometrica* 82(4), 1341-1403.

[90] Tyrväinen L. and A. Miettinen (2000). Property prices and urban forest amenities. *Journal of Environmental Economics and Management* 39(2), 205–223.

[91] Tyrväinen L. (2001). Use and valuation of urban forest amenities in Finland. *Journal of Environmental Management* 62, 75–92.

[92] Tyrväinen L., Pauleit S., Seeland K., de Vries S. (2005) Benefits and Uses of Urban Forests and Trees. In: Konijnendijk C., Nilsson K., Randrup T., Schipperijn J. (eds) Urban Forests and Trees. Springer, Berlin, Heidelberg.

[93] Tzoulas K., K. Korpela, S. Venn, V. Yli-Pelkonen, A. Kaźmierczak, J. Niemela, and P. James (2007). Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. *Landscape and Urban Planning* 81, 167-178. doi:10.1016/j.landurbplan.2007.02.001.

[94] UN DESA (2014). World Urbanization Prospects: The 2014 Revision.

[95] UN DESA (2018). World Urbanization Prospects: The 2018 Revision, custom data acquired via website. Accessed on June 19^{th}, 2018.

[96] van den Bosch M. and A.O. Sang (2017). Urban natural environments as nature-based solutions for improved public health - A systematic review of reviews. *Environmental Research* 158, 373-384.

[97] Walsh R. (2007). Endogenous open space amenities in a locational equilibrium. *Journal of Urban Economics* 61, 319-344.

[98] Warziniack T. (2010). Efficiency of public goods provision in space. *Ecological Economics* 69, 1723-1730.

[99] White H. (1980). A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. *Econometrica* 48, 817-838.

[100] White M.P., I. Alcock, B.W. Wheeler, and M.H. Depledge (2013). Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. *Psychological Science* 24, 920-928.

[101] Wu J. and S.H. Cho (2003). Estimating households' preferences for environmental amenities using equilibrium models of local jurisdictions. *Scottish Journal of Political Economy* 50(2), 189-206.

[102] Wu J. and A. Plantinga (2003). The influence of public open space on urban spatial structure. *Journal of Environmental Economics and Management* 46, 288-309.

[103] Zhang L. (2016). Flood hazards impact on neighborhood house price: A spatial quantile regression analysis. *Regional Science and Urban Economics* 60 (2016) 12-19.