

TI 2019-023/VIII
Tinbergen Institute Discussion Paper

The Welfare Effects of Greenbelt Policy: Evidence from England

Revision: January 2023

Hans R. A. Koster¹

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

The Welfare Effects of Greenbelt Policy: Evidence from England*

Hans R.A. Koster[†]

Abstract — I measure the economic effects of *greenbelts* that prohibit new construction beyond a predefined urban fringe and therefore act as urban growth boundaries. I focus on England, where 13% of the land is designated as greenbelt land. I provide reduced-form evidence and estimate a quantitative equilibrium model that includes amenities, housing supply, a traffic congestion externality, agglomeration forces, productivity, and household location choices. Greenbelt policy generates positive amenity effects, but also strongly reduces housing supply. I find that greenbelts increase welfare because amenity effects are sufficiently strong. At the same time, however, greenbelts decrease housing affordability by limiting housing supply.

Keywords — housing, supply constraints, land use regulation, greenbelts, urban growth boundary, open space.

JEL codes — R52, R13, R30, R41, H41

*This paper has received the Sakıp Sabancı International Research Award 2020. Mohammad Saeed Zabihidan provided useful research assistance in the early stages of this project. Paul Cheshire and Jos van Ommeren are thanked for their particularly helpful discussions. I also thank the editor Barbara Petrongolo, four anonymous referees, Gabriel Ahlfeldt, Federico Curci, Robert Elliott, Sofia Franco, Steve Gibbons, Stephan Heblich, Christian Hilber, Matthias Kalkuhl, Joris Kligen, Camille Landais, Michael Pflüger, Tobias Seidel, Saeed Tajrishy, Matthew Turner, Erik Verhoef, Xiao Yu, as well as seminar participants in Birmingham, Düsseldorf, London, Lugano, Madrid, Neuchâtel, Potsdam, Philadelphia, St. Petersburg, Vancouver, and Würzburg for their useful input. Data on historic transportation networks is kindly provided by Miquel-Àngel Garcia-López, Ilias Pasidis, and Elisabet Viladecans-Marsal.

[†]Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105 1081 HV Amsterdam, email: h.koster@vu.nl. Hans is also research fellow with the Tinbergen Institute and the Centre for Economic Policy Research.

1 Introduction

In most countries, urban growth leads to increasing pressure on developable land in and around cities. Many cities regulate urban development by imposing a range of constraints on *e.g.* building height or type of land use. Local governments also frequently restrict the expansion of urban areas to prevent urban sprawl. These *urban growth boundaries* or *greenbelts* reduce land available for development at the urban fringe. Many U.S. cities, such as Portland (OR), Miami, Minneapolis Saint-Paul, and San Jose (CA), have urban growth boundaries (UGBs), and similar restrictions can be found in many other countries (*e.g.* Austria, Canada, China, France, Germany, Iran, the Netherlands, New Zealand, Norway and South Korea). I focus on England, where greenbelts are important as they cover approximately 13% of the total area and surround most larger cities.

Land-use regulation does not necessarily lead to welfare losses, because constraints may reduce negative land-use externalities and frictions associated with development. Greenbelt policy indeed intends to protect agricultural land and secure amenity benefits from open space (Brueckner, 2001). At the same time, when regulatory constraints are too strict, regulation may lead to substantial economic losses. Thus, economists have argued that greenbelt policy should be relaxed to mitigate the ‘housing affordability crisis’ because restrictions on housing supply lead to potentially strong price increases (Cheshire, 2014; Economist, 2017). Despite the potentially large impacts of greenbelts on the growth of cities and housing markets, to the best of my knowledge, no study has yet attempted to evaluate the welfare effects of greenbelt policy.

This study therefore seeks to measure the effects of greenbelt policy on the spatial distribution of economic activity. To do this, I first use data on the location of all dwellings in England. It appears that 37% of dwellings lie within 1km of greenbelt land while 2% are on greenbelt land. I further exploit data on more than 10 million housing transactions between 1995 and 2017. Given information on the exact location, I can identify for each dwelling the share of greenbelt land in its vicinity. Furthermore, I use data at the Middle-layer Super Output Area (MSOA) level (which has on average a working population of around 3,800) on commuting flows and the locations of individuals and their workplaces.

The first aim of this paper is to identify the reduced-form *supply* and *amenity* effects of greenbelt

land. The supply effect captures the reduction in the supply of housing in greenbelts. The amenity effect refers to the increased attractiveness of a location due to better access to green space. Estimates of these effects are potentially biased, as greenbelts are located on the outskirts of cities, where land is usually cheaper. Hence, houses close to greenbelts tend to be larger and may have gardens. To address omitted variable bias, I pursue two identification strategies. First, I exploit the fact that greenbelt boundaries in England have hardly changed since their imposition in the 1970s. I gather data on approved and proposed greenbelt land in 1973 and explain variation in prices and densities in those areas. Alternatively, as concerns may arise that the selection process of greenbelts is correlated to unobserved locational endowments, I only select areas close to greenbelt boundaries and apply spatial differencing. In the spirit of [Turner et al. \(2014\)](#), I improve on this strategy by only including data on one side of the border so as to avoid discrete differences in housing and neighbourhood characteristics at the greenbelt border. The local *supply* and *amenity* effects of greenbelts are only one part of the story. The substantial reduction in the supply of developable land close to cities will also influence the spatial equilibrium, resulting in a different spatial distribution of production, implying *agglomeration* effects. These effects are potentially relevant because residents have to commute to work. Moreover, workers are more productive when they are in the vicinity of other workers ([Ciccone and Hall, 1996](#); [Ciccone, 2002](#); [Arzaghi and Henderson, 2008](#); [Combes et al., 2008](#); [Melo et al., 2009](#)) and because traffic congestion arises when people cluster together ([Combes et al., 2019](#); [Proost and Thisse, 2019](#)).

To model the complex interactions of *amenity*, *supply*, and *agglomeration* effects, I set up a quantitative spatial general equilibrium model, following [Ahlfeldt et al. \(2015\)](#). In this model, residents and firms compete for floor space, while workers benefit from each other's presence due to agglomeration economies. Residents commute to workplaces, so a greater concentration of firms typically implies higher commuting costs. I extend the model of [Ahlfeldt et al. \(2015\)](#) in three directions:

1. I embed land-use restrictions in the model, as greenbelt land reduces the available land available for development at certain locations;
2. I allow for greenbelts to generate a higher amenity level; hence, I explicitly specify the

amenity residual in [Ahlfeldt et al. \(2015\)](#) and estimate the spatial decay of these amenities;

3. I allow for mode choice and endogenous travel times of road travel, *i.e.* for traffic congestion close to the workplace. Several papers show that UGBs affect the congestion level in a city and therefore have welfare implications through the potential reduction of congestion externalities (see [Kanemoto, 1977](#); [Arnott, 1979](#); [Pines and Sadka, 1985](#); [Anas and Rhee, 2007](#); [Brueckner, 2007](#)).

Using the recursive structure of the model, together with the identification strategies to identify the amenity and supply effects outlined above, I *estimate* (rather than calibrate) the structural parameters of the model.¹

In contrast to [Ahlfeldt et al. \(2015\)](#), I do not choose but rather estimate the share of construction cost spend on land inputs, which appears to be important. To identify this parameter, I follow [Combes et al. \(2016\)](#) in using the first-order condition for profit maximisation with respect to building capital and relying on variation in systematic determinants of demand for real estate across space. To identify the parameters related to agglomeration and congestion forces, I first propose a standard identification strategy using historic instruments. Alternatively, I use spatial instruments using exogenous characteristics of faraway locations, following [Bayer and Timmins \(2007\)](#). Given that one may question these instruments' ability to identify causal spillover effects, I also provide analyses in which the instruments are only 'plausibly exogenous' ([Conley et al., 2012](#)).

I show that greenbelt policy has a small *positive* welfare effect given the structural parameters obtained in the preferred specifications. The income reduction if greenbelts were to be removed is 1.4%, which amounts to approximately £9 billion a year. Whether residents benefit from greenbelts critically depends on two parameters. First, it depends on the elasticity of building production with respect to non-land inputs. When this exceeds around 0.55, I find positive effects of greenbelts. For high values of this elasticity, it is cheaper to transform land into buildings, meaning the reduction in land due to greenbelt policy is less costly. Second, I show that the benefits of greenbelts for residents dissipate once greenbelt amenities cease to exist. This shows that greenbelt amenities are key in understanding why welfare effects of greenbelts can

¹When using calibration, parameters are chosen based on other studies and research contexts. By contrast, when estimating parameters, current data are used and empirical strategies are devised to obtain the model's parameters.

be positive. Interestingly, agglomeration economies do not seem to be crucial, which is in line with [Kline and Moretti \(2014\)](#), who argue that agglomeration economies are a localised market failure that cancels out in the aggregate. Furthermore, I show that greenbelts lead to higher real estate prices: in cities like London, Birmingham, and Manchester, property prices will decrease by 5-20% when greenbelts were to be removed. Hence, although greenbelts increase overall welfare, they inevitably and strongly reduce housing affordability.

The welfare measure used in this paper arguably does not include all possible general equilibrium effects. Still, to the extent I do not include potential other benefits of greenbelts (*e.g.* reductions in pollution, or city-wide amenity increases) that spill over to people living in cities, the estimate of welfare gains of greenbelts can be considered as an underestimate.

Related literature. Most previous studies on the effects of land-use regulation concentrate on housing supply restrictions and indicate that supply constraints are associated with increasing housing costs, a strong reduction in new construction, and rapid price growth ([Mayer and Somerville, 2000](#); [Glaeser et al., 2005](#); [Green et al., 2005](#); [Ihlanfeldt, 2007](#)). This effect is particularly pronounced for cities in England, in which land-use regulation is highly restrictive ([Hilber and Vermeulen, 2016](#)).² Other evidence for England by [Cheshire et al. \(2018\)](#) shows that land-use restrictions may also lead to higher vacancy rates and longer commutes. [Glaeser and Ward \(2009\)](#) find that local constraints in Boston (*i.e.* within a city) do not increase the price of land because of close substitutes. At the same time, they find that density levels are too low from a welfare perspective. [Koster et al. \(2012\)](#) find that the costs of regulation for homeowners or developers (so-called ‘own-lot effects’) may be substantial (up to 10% of the housing value). [Turner et al. \(2014\)](#) evaluate the own-lot and amenity effects of land-use regulation in the U.S. Own-lot effects appear to be substantial, but they do not find evidence for amenity effects, which suggests that land-use regulation has negative welfare consequences in that context.³ [Harari \(2020\)](#) shows that for India the shape of cities matters. Less compact cities, which imply longer within-city travelling distances, are associated with a lower quality of life. In India, land-use

²An older paper by [Cheshire and Sheppard \(2002\)](#) attempts to evaluate the welfare effects of land-use regulation for the city of Reading by relying on a monocentric city model. They find that expanding the city will yield positive welfare effects. However, the model is subject to several strong functional form assumptions and does not include commuting, congestion, and agglomeration economies.

³[Turner et al. \(2014\)](#), meanwhile, do not present a general equilibrium analysis of supply restrictions, because they focus on the effects of areas close to municipal borders. By contrast, I focus on a situation where land-use regulation is ubiquitous and where general equilibrium effects are expected to be sizeable.

regulation typically promotes less compact developments and therefore further reduces urban accessibility and welfare.

This paper also relates to research that measures the local benefits of open space. Some studies have specifically focused on reduced-form house price impacts of greenbelt land. An early study by [Correll et al. \(1978\)](#), for instance, reports that properties near greenbelt land are generally more expensive, but the reported effects are unlikely to be causal. [Jun \(2006\)](#) shows no evidence of a significant difference between housing prices inside and outside the UGB, confirming that these areas are part of a single housing market. By contrast, [Grimes and Liang \(2009\)](#) find that land just inside the UGB in Auckland, New Zealand, is valued at approximately 10 times that of land just outside the boundary, which is the result of better redevelopment opportunities within the urban limit. While these reduced-form estimates show that UGBs are relevant, they do not provide clear implications as to whether greenbelts increase or decrease welfare. Other studies focus explicitly on the measurement of the benefits of open space. [Bolitzer and Netusil \(2000\)](#), for example, find that living close to green space increases property values by maximally 5%, while [Irwin \(2002\)](#) finds that a hectare of farmland in the vicinity increases property values by 0.75%. [Anderson and West \(2006\)](#), however, find that the value of proximity to open space is higher in dense neighbourhoods. [Geoghegan \(2002\)](#) shows that open space labelled as ‘permanent’ increases nearby residential land values over three times as much as an equivalent amount of ‘developable’ open space. This may explain why I find somewhat strong effects of greenbelt land on house prices, as greenbelt land is non-developable. I note that none of the papers provides data on general equilibrium effects, such as the effects of commuting and housing supply.

Finally, this paper contributes to a mostly theoretical literature on the effects of urban growth boundaries on commuting – more specifically, whether UGBs can be a second-best policy to reduce congestion externalities. Early papers by [Kanemoto \(1977\)](#); [Arnott \(1979\)](#) and [Pines and Sadka \(1985\)](#) show that a not-too-stringent UGB is a second-best policy for congestion tolls when traffic congestion is unpriced. However, these papers assume that all jobs are exogenously located in one urban centre. [Anas and Rhee \(2007\)](#) show that with cross-commuting, boundaries of any stringency can be inefficient even when tolls shrink cities, as boundaries do little to reduce inefficient commuting from the suburb to the city centre. [Brueckner \(2007\)](#) corroborate

this conclusion, finding that greenbelts may not be a useful instrument for addressing the distortions caused by unpriced traffic congestion.

The plan for the remainder of the paper is as follows. In Section 2 I explain how greenbelts were designated, introduce the datasets, and provide descriptives. In Section 3, I provide reduced-form evidence for the effects of greenbelt land on dwelling density and house prices. I also provide a wealth of robustness checks. Section 4 then outlines the quantitative model, and in Section 5 I report the estimated structural parameters and discuss the different counterfactual analyses. Finally, Section 6 concludes.

2 Context, data, and descriptives

2.1 Greenbelts in England

There is a long-standing tradition in England of restricting urban growth. In the 1920s, proposals were put forward by the London Society and the Campaign to Protect Rural England (CPRE) to prevent development in a continuous belt within 2km of London. Then, in the 1947 Town and Country Planning Act, local authorities (LAs) were for the first time allowed to take planning decisions and to incorporate greenbelt proposals in their development plans.

In 1955, Duncan Sandy, who was then the Minister of Housing, encouraged local authorities around the country to consider protecting land around cities through the formal designation of well-defined greenbelts. In a statement in the House of Commons, he wrote the following:

“I am convinced that for the well-being of our people and for the preservation of the countryside, we have a clear duty to do all we can to prevent the further unrestricted sprawl of the great cities. The Development Plans submitted by the local planning authorities for the Home Counties provide for a Green Belt, some 7 to 10 miles deep, all around the built-up area of Greater London. [...]. No further urban expansion is to be allowed within this belt. [...] I am accordingly asking all planning authorities to submit to me proposals for the creation of clearly defined Green Belts, wherever this is appropriate.”

Greenbelts were introduced in the two decades after 1955 around almost all of England’s big cities (London, Birmingham, Liverpool, and Manchester) but also around smaller cities (e.g. Bournemouth, York, Oxford, and Cambridge). Most cities that put forward proposals

to designate greenbelt land had at least a population of 100,000 inhabitants at that time and therefore qualified as ‘*large urban areas*’. Most proposals were submitted in the late 60s and early 70s, while the final approval and exact demarcation of the greenbelt borders took place in the early 80s.⁴ Since the official approval of greenbelts in the early 1980s, no new greenbelts have been introduced and the total amount of greenbelt land has essentially not changed in the last 35 years.⁵ Currently, greenbelts cover approximately 13% of *all* land in England (for comparison, built-up land covers around 10%) and should, according to the National Planning Policy Framework in 2012, offer appreciable amenities to the urban population by improving access to the open countryside, by providing opportunities for outdoor sport and recreation, and by retaining attractive landscapes close to urban areas. Land-use data from [Open Street Map](#) indicate that roughly one-third of greenbelt land is used for agriculture, while only around 5% of the land is classified as parks or recreation grounds.⁶ Furthermore, greenbelts were not introduced to address the problems of inner cities (such as traffic congestion and walkability), but were instead introduced to protect the countryside and, in later years, to improve access to green space for people living in urban areas. Hence, although urban planners may apply other types of regulations (such as floor-area restrictions, historic building protection *etc.*), this is not part of the greenbelt policy per se.

In Figure 1, I show the 14 greenbelts in England as well as national parks and Areas of Outstanding Natural Beauty (AONBs), which do not overlap. As can be seen, greenbelts are about 10-20km deep, as suggested by Duncan Sandy, and unmistakably surround the large cities such as London, Birmingham, and Manchester. But, at the same time, Figure 1 also shows that greenbelt land is very patchy. Because towns and settlements existed before greenbelt policy was implemented, one can observe developments in greenbelt areas and on greenbelt land. In the identification strategy employed later in the paper, I employ a border-discontinuity

⁴Greater London was the first urban area to discuss the implementation of a greenbelt and did so before World War II. However, it was only in the late 1950s that greenbelt land around London was officially approved. The greenbelt area increased fivefold in 1965, while in 1971, the government decided to extend the Metropolitan Green Belt to include almost all of Hertfordshire. The Bristol and Bath Green Belt was adopted locally in 1957 and approved in 1966. The North-West greenbelt around Manchester, Liverpool, and Leeds was considered since the early 1960s, but it was approximately twenty years before it received official approval. The same holds for the greenbelt around York, which was formally created in 1980 after decades of being a local policy.

⁵For example, the total hectares of greenbelt land in 1997 was 1,652,310, while it was 1,638,610 in 2013, a change of less than 1%, which may as well be due to measurement error.

⁶In the empirical analysis, I will therefore include specifications where I distinguish between different types of greenbelt land.

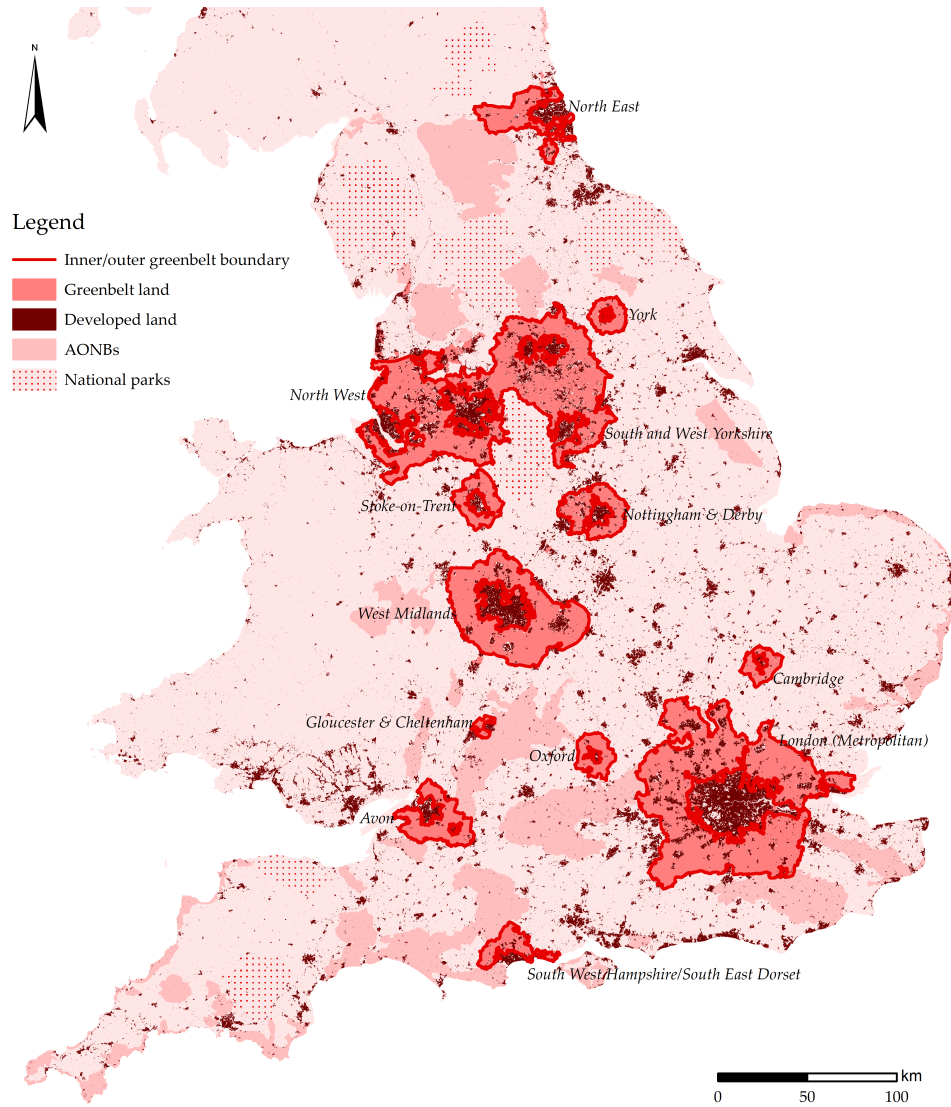


FIGURE 1 – GREENBELTS IN ENGLAND

design based on *inner* and *outer* greenbelt borders. I define those by looking at the border of development with greenbelt land surrounding *the most important cities*. Hence, inner greenbelt boundaries only surround the most important cities and not the small towns and villages that are fully surrounded by greenbelt land. I refer to locations as being in ‘a greenbelt area’ when they are in between the inner and outer greenbelt boundaries, although not all land in those greenbelt areas is necessarily designated as greenbelt land (see for an illustration Appendix A.1).

In the empirical analysis, I will also use the information on considered and approved greenbelt land in earlier times, *i.e.* in 1973. Appendix A.2 provides further details.

2.2 Data

2.2.1 Micro-data

I make use of four datasets on England. The first dataset pertains to the number of dwellings per postcode from the [Office of National Statistics \(ONS\)](#), which is based on the 2011 [Census](#). A postcode is very small and contains around 13 households.

Information on greenbelts in 2012 is obtained from the [Department of Communities and Local Governments \(DCLG\)](#). Each local authority digitised land-use information and [DCLG](#) merged these separate datasets. I aim to identify amenity and supply effects using the boundaries of greenbelts, as these capture the actual urban containment boundaries. Hence, I determine for each greenbelt the inner and outer boundaries and calculate the distance of the centroid of each postcode to the nearest inner or outer boundary of a greenbelt. Furthermore, for each postcode, I calculate the share of greenbelt land. Because postcodes are so small, this is typically zero or one. Plus, to calculate the effects of greenbelt amenities, I calculate the share of greenbelt land within 500m distance bands.

The third dataset contains the universe of housing transactions from England from the [Land Registry](#) between 1995 and 2017. These data provide information on the transaction price, the housing type, the date of the transaction, and the ownership structure (leasehold or freehold), as well as the location at the postcode level. A disadvantage of the [Land Registry](#) data is the limited amount of information on housing characteristics. For example, the size of the property is missing.

Therefore, I merge each [Land Registry](#) transaction to [Energy Performance Certificates \(EPCs\)](#). Since 2007, an [EPC](#) has been required whenever a home is constructed or sold. [EPCs](#) have been issued since October 1, 2008, and provide information regarding the energy performance of buildings and their characteristics, which are obtained by means of a physical inspection of the interior and exterior of the home by an independent assessor. Most notably, this provides me with the floor area of the property, the number of rooms, and the energy efficiency.

My merging strategy is to sequentially match individual sales to the [EPC](#) data using the full

address or a subset of the address and the date of the sale and certificate.⁷ Around one-third of the sales in the `Land Registry` remains unmatched, so I drop them from the analysis.⁸ I also drop some outliers and transactions that are matched to multiple `EPCs` (around 15%).⁹

Because greenbelt boundaries have hardly changed between 1995 and 2017, matching greenbelt data from 2012 to transactions in the past will imply little measurement error. Moreover, although `EPCs` are available from 2008 onwards, I still match transactions from the `Land Registry` from before 2008 to `EPC`, thereby assuming that housing characteristics do not change. Hence, I use the full temporal extent of the data (1995-2017).¹⁰ This leaves us with 10,070,791 sales.

2.2.2 MSOA data

For the structural model (to be introduced later), I use data at the Middle-Layer Super Output Area from the 2011 `Census`. To obtain floor space prices, I use the above-discussed sales data from the `Land Registry` and `EPCs` and regress log prices per square metre on housing characteristics and MSOA fixed effects. I obtain floor space prices by taking the exponent of the estimated fixed effects.¹¹ In the structural estimation, I normalise floor space prices to have a geometric mean of 1.

Furthermore, for each MSOA, I calculate the share of greenbelt land as well as the share in 1973 greenbelts. For each centroid of an MSOA, I calculate the distance to the nearest inner or outer greenbelt boundary.

⁷Specifically, I first match a sale to an `EPC` using the primary address object name, secondary address object name, street name, and postcode. I then keep the certificate that is closest in days to the sale. I repeat this exercise for unmatched properties but allow the address identifiers to be different. The final round of matching matches on the full postcode.

⁸The matching is more challenging for flats that often share an address, implying that the proportion of flats is reduced from 23% in the `Land Registry` sample to 3% in the final sample. My analysis therefore mainly focuses on single-family homes. One may be concerned that this implies sample selection bias. However, if I calculate floor space prices at the MSOA prices using the full `Land Registry` dataset and using the matched dataset, the correlation is 0.988. Moreover, I show in Appendix B.2.7 that the results are essentially the same if I only include apartments.

⁹More specifically, I drop transactions with prices that are above £1.5 million or below £15,000 and properties that are smaller than 25m² or larger than 250m². Together this applies to less than 0.5% of the data.

¹⁰To ensure that measurement error in greenbelt land does not bias my results, I also estimate regressions where I only include observations from the closest Census year 2011, which does not change the results (see Appendix B.2.7 for more details).

¹¹The results of this regression are available upon request. Housing characteristics included in the regression are the log of house size, house type (*i.e.* terraced, semi-detached or detached), a dummy indicating whether the house is newly built, the number of rooms, the number of habitable rooms, the floor level (if it is an apartment), the height of the average floors, the number of floors of the building, whether the property has a fireplace, and the quality of the windows, the roof, walls, and overall energy efficiency. I also include a dummy indicating whether the property is freehold. All the coefficients have the expected signs and magnitudes.

From the 2011 [Census](#), I obtain commuting flows between each of the 6,791 MSOAs, which means that I have 46,117,681 cells containing information on the number of workers commuting from home to work and by what mode. Using this information, I also calculate the total number of workers and residents by mode in each area. From [OS](#) I obtain information on the road network in 2012. That is, I keep motorways, A-roads, and B-roads for which I assume free-flow travel speeds of 110, 80, and 50km/h.¹² Using these network data, I calculate free-flow travel times between each MSA pair. To obtain actual travel times, I obtain data on average speeds on major roads from the [Department of Transport](#) at the county level in 2015. Because counties are much smaller in urban areas, this will provide a reasonable proxy for actual speeds. I then match each road to the county in which it is located and calculate actual travel times between each MSA pair. For travel times by rail, I use data on the railway network from [OS Meridian](#). Data on London's tube network is from [Transport for London](#). The average speed on railways is from [Cartmell \(2016\)](#) and is assumed to be 57km/h, while the average speed on tube lines from [Transport for London](#) is 33km/h.¹³ To calculate travel times for other modes (including walking, cycling, and working from home), I use the road network, but I assume a speed on the network of 10km per hour.

In Appendix [A.4](#), I provide some more details on commuting flows and show that calculated travel times are very highly correlated to actual travel times (obtained for a small subset of the data).

To construct instruments for agglomeration (to be discussed later on), I gather data on the historic population at the parish level for 1931. There were 11,450 parishes in 1931 (these are usually considerably smaller than MSOAs and parliamentary constituencies). For each MSA and constituency, I calculate the share in each parish. I then assume that the population is uniformly distributed within each parish by multiplying the share of each MSA/constituency in each parish by the parish population.

I further use information on historic travel times to calculate historic travel times between MSA pairs. I obtain data on railway networks from [Garcia-López et al. \(2021\)](#) on England's railway network in 1870. Assuming a speed of 50km/h, I calculate travel time in minutes between each

¹²The average speed on motorways is obtained from [statista.com](#).

¹³One may argue that I underestimate the speed of intercity services, say, between London and Birmingham. However, for most commutes, only local trains are used that have lower average speeds.

MSOA. I also gather data on developed land and soil characteristics. I use information from the [British Geological Survey](#) on the depth of bedrock and sediment thickness. Data on elevation are obtained from the [Surfzone DEM](#) model. From [Nomis](#), I obtain the share of construction workers.

2.3 Descriptive statistics

Panel A in Table 1 reports descriptive statistics for the housing transactions data. On average, 3.6% of the transactions are in a greenbelt. I show that the average price per m² of floor space is £1753 and the average floor size is 87m². In greenbelts, this is respectively £2057 and 91m². Hence, houses in greenbelts are, on average, similar in size to other homes but slightly more expensive. The highest share of the properties is taken up by terraced houses (41%), while the shares of flats, semi-detached, or detached properties are considerably lower. The average distance to the nearest greenbelt boundary is 16.4km. Still, approximately 37% (27%) of all dwellings in England lie within 1km (500m) of greenbelt land, so many properties will be potentially affected by greenbelt amenities. 2% of the dwellings are located on greenbelt land, which is because these properties were built before greenbelts were assigned or exceptions were granted.

I further report descriptive statistics at the postcode level in Panel B in Table 1. On average, 6% of the land is greenbelt land. I observe on average 17 (the median is 13) dwellings in a postcode. The median size of a postcode is only 0.93 hectares.¹⁴ In the sample, 72% of land in postcodes is developed, which is much higher than the overall share of developed land (8.7% in England), because postcodes in urban areas are much smaller and therefore overrepresented. The share of developed land in greenbelts, meanwhile, is only 5.6%, but definitely not zero. Hence, I observe properties that are on designated greenbelt land.

I also report descriptive statistics for MSOAs in Panel C in Table 1. England's total working population is 25,087,843. The average population density is 15.6 persons per hectare. There is a very high correlation of 0.980 to dwelling density. The floor space price is, on average, £2154, but there is considerable variation.¹⁵ The floor space price is also strongly positively correlated with density; the correlations with population density and worker density are, respectively,

¹⁴Postcodes are on average 10 hectares due to a few very large postcodes.

¹⁵The highest floor space can be found in London in the borough of Kensington-Chelsea.

TABLE 1 – KEY DESCRIPTIVE STATISTICS FOR MICRO-DATA

| PANEL A: Descriptives for house prices | (1) mean | (2) sd | (3) min | (4) max |
|---|-------------|-----------|------------|------------|
| Price per m ² | 1,754 | 1,268 | 100 | 10,000 |
| Size of the property in m ² | 87.40 | 31.62 | 25 | 250 |
| Share land in greenbelt | 0.0361 | 0.1635 | 0 | 1 |
| Share greenbelt land <500m | 0.0853 | 0.1903 | 0 | 1 |
| Distance to nearest greenbelt boundary (km) | 16.43 | 29.25 | 0 | 296.8 |
| Housing type – flat | 0.0318 | 0.1756 | 0 | 1 |
| Housing type – terraced | 0.4127 | 0.4923 | 0 | 1 |
| Share of developed land | 0.8327 | 0.3044 | 0 | 1 |
| Distance to the nearest city centre (km) | 35.80 | 33.49 | 0.0802 | 313.7 |
| PANEL B: Descriptives for postcode data | (1) mean | (2) sd | (3) min | (4) max |
| Number of dwellings | 16.56 | 14.96 | 0 | 646 |
| Area size of postcode | 10.12 | 50.69 | 0.0010 | 7,827 |
| Share greenbelt land | 0.0624 | 0.2256 | 0 | 1 |
| Distance to nearest greenbelt boundary (km) | 18.61 | 32.30 | 0 | 298.6 |
| Share of developed land | 0.7218 | 0.3969 | 0 | 1 |
| Share parks in postcode | 0.0349 | 0.1254 | 0 | 1 |
| Distance to the nearest city centre (km) | 37.50 | 36.20 | 0 | 316.2 |
| PANEL C: Descriptives for MSOAs | (1) mean | (2) sd | (3) min | (4) max |
| Population density (per ha) | 15.59 | 16.96 | 0.0226 | 157.6 |
| Worker density (per ha) | 14.99 | 44.85 | 0.0151 | 1,384 |
| Dwelling density (per ha) | 13.44 | 14.30 | 0.0253 | 133.6 |
| Floor space price (£ per m ²) | 2,154 | 1,198 | 606.0 | 12,336 |
| Share greenbelt land | 0.1520 | 0.2715 | 0 | 1 |
| Share land in 1973 greenbelt | 0.0802 | 0.2237 | 0 | 1 |
| Distance to greenbelt boundary (km) | 15.94 | 29.13 | 0.0003 | 295.3 |
| Distance to the nearest city centre (km) | 34.03 | 33.41 | 0.0716 | 312.0 |
| Population density in 1931 (per ha) | 20.36 | 39.51 | 0 | 354.7 |

Notes: The number of observations for house prices is 10,070,791. For the postcode data, the number of observations is 1,309,635. The number of MSOAs is 6,791.

0.331 and 0.489.

The share of greenbelt land in an MSOA is, on average, 0.152. This is higher than for postcodes because postcodes are much smaller in cities. The correlation of historic population density (in 1931) with current densities is relatively high: it is 0.691 for current population density and 0.378 for current employment density. One may be surprised that the population density in 1931 was higher than the current population density. However, the current population (and employment) density only includes the working population, while the population in 1931 refers to the full population.

3 Reduced-form results

In this Section, I aim to show that greenbelt land has two major direct effects. First, it reduces the amount of land available for development and therefore leads to lower local densities. Second, it creates an amenity effect, leading to higher house prices within close vicinity of greenbelt land. I close this section by conducting a sensitivity analysis of these effects.

3.1 Supply effects and housing density

3.1.1 Methodology

First, I am interested in determining to what extent greenbelts limit development – in other words, to what extent is the greenbelt policy binding? Let us define h_i as the number of dwellings in postcode i and g_i as the share of greenbelt land in the postcode. Note that h_i is a non-negative count variable. As the size of (postcode) areas differ, I control for the size of the area L_i , so the effect of g_i can be interpreted as the effect on *housing density*. I use Poisson-Pseudo Maximum Likelihood to estimate:

$$h_i = \exp(\phi_1 g_i + \phi_2 \log L_i + f(x_i) + \eta_{i \in \mathcal{A}}), \quad (1)$$

where $f(\cdot)$ captures a third-order polynomial of the distance to the city centre, x_i . ϕ_1 is the coefficient of interest, ϕ_2 is another coefficients to be estimated, and $\eta_{i \in \mathcal{A}}$ denote local authority fixed effects.

A concern with the above specification is that greenbelts are not randomly distributed over space, as greenbelt land can be found at the outskirts of cities. I then consider two identification strategies to identify the causal effect of greenbelt land. For the first identification strategy, I only keep postcodes on approved and proposed greenbelt land in 1973. These areas are likely similar in terms of unobservables. In the regression analyses, I therefore estimate regressions where I only keep postcodes that are in approved and proposed greenbelts in 1973.

As a second identification strategy, I rely on a boundary-discontinuity design. That is, I only include postcodes within a distance b to an inner or outer greenbelt border (*e.g.* within 1km or even within 100m in some sensitivity analyses), while controlling flexibly for the *distance to the*

TABLE 2 – SUPPLY EFFECTS OF GREENBELTS: EFFECTS ON DWELLINGS
(Dependent variable: the number of dwellings in a postcode)

| | (1) Poisson | (2) Poisson | (3) Poisson | (4) Poisson | (5) Poisson | (6) Poisson |
|--------------------------------------|------------------------|---------------------------------|------------------------|--------------------------|--------------------------|---------------------------|
| | | + Controls and fixed effects | Greenbelts in 1973 | Greenbelt border <2km | Greenbelt border <1km | Inside greenbelt areas |
| Share greenbelt land | -0.4661*** (0.0159) | -0.6535*** (0.0142) | -1.1719*** (0.0255) | -1.0024*** (0.0177) | -0.9721*** (0.0185) | -1.0046*** (0.0349) |
| Area size of postcode (<i>log</i>) | 0.0486*** (0.0034) | 0.1105*** (0.0026) | 0.2726*** (0.0070) | 0.2271*** (0.0046) | 0.2265*** (0.0050) | 0.1723*** (0.0096) |
| Distance to the city centre | No | Yes | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | No | Yes | Yes | Yes |
| Local authority fixed effects | No | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 1,309,635 | 1,309,635 | 250,671 | 392,526 | 255,860 | 33,923 |
| Bandwidth (<i>km</i>) | ∞ | ∞ | ∞ | 2 | 1 | 1 |
| Pseudo- R^2 | 0.0123 | 0.0764 | 0.146 | 0.115 | 0.115 | 0.167 |

Notes: Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (3) includes observations in areas that are in greenbelts that were approved or considered in 1973. In columns (4) I include transactions that are within 2km of a greenbelt border, which is reduced to 1km in columns (5) and (6). In column (6) I only include postcodes in greenbelt areas. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

nearest greenbelt boundary on both sides of the border, captured by d_i^- and d_i^+ :

$$h_i = \exp(\phi_1 g_i + \phi_2 \log L_i + f(x_i) + \phi_4 d_i^- + \phi_5 (d_i^-)^2 + \phi_6 d_i^+ + \phi_7 (d_i^+)^2 + \eta_{i \in \mathcal{A}}), \quad \text{if } d_i < b. \quad (2)$$

This strategy addresses the issue of greenbelt borders being near the urban fringe (where commutes are longer and density is generally lower). Moreover, to address the potential issue that supply effects are still capturing the decrease in density when moving beyond the greenbelt boundary, I also estimate specifications where I only include postcodes in greenbelt areas (*i.e.* that are in between the inner and outer greenbelt borders).

3.1.2 Results

The results are reported in Table 2. In column (1) I include all postcodes and only control for postcode area size. It is shown that when the share of greenbelt land in a postcode is higher, the number of dwellings is substantially lower. The coefficient implies that when the whole postcode area is on greenbelt land, the number of dwellings changes by $e^{-0.466} - 1 = -37\%$. When I control for distance to the city centre and add local authority fixed effects, the reduction in dwellings is 48% (column (2)).

Column (3) further improves on identification by only including postcodes in 1973 green-

belts. The effect is then considerably stronger (-69%). Columns (4) and (5) rely on a border-discontinuity approach based on inner and outer greenbelt borders. I find that postcodes in dwellings have approximately 60% fewer dwellings and that the result is robust to the bandwidth choice. In the preferred specification in column (6), I only include postcodes that are in greenbelt areas. In this way, I address the issue that the supply effect captures a density gradient that is decreasing in distance to the city centre. The coefficient is essentially the same as in the previous specifications.

These reduced-form results confirm that the density of development is strongly affected by greenbelt policy with estimates that vary between 37% and 70%. However, the reduction is far from 100%; hence, recall that there are still (residential) buildings on greenbelt land, albeit in a much lower density.¹⁶ In Appendix B.2.1, I further investigate whether greenbelt policy also affects densities further away from greenbelt land, *e.g.* though the housing types provided close to greenbelt borders, or because developers build in higher densities in city centres. I do not find evidence for this.

3.2 *Amenity effects and house prices*

3.2.1 **Methodology**

I estimate the reduced-form local amenity effects of greenbelt policy using information on house prices. Let p_{it} be the house price in postcode i in year t and \tilde{g}_i be the share of greenbelt land within 500m. One may argue that the amount of greenbelt land in the vicinity is correlated to housing attributes; houses with particular characteristics may be predominantly located in greenbelts. For example, because of historic city limits, properties in greenbelts may be mostly detached, while houses outside greenbelts may appear more often in the form of apartments or terraced housing. To mitigate this problem, I include (time-invariant) housing characteristics, such as the log of house size and house type, denoted by c_i .

To control for unobservable locational attributes and for aggregate housing supply effects, I include local authority \mathcal{A} fixed effects $\rho_{i \in \mathcal{A}}$. These fixed effects aim to capture time-invariant unobserved characteristics that could be correlated to the share of greenbelt land, such as the

¹⁶The more convincing identification strategies (columns (3)-(6)) show larger supply effects. In these specifications, I use a control group of areas that are similar and geographically close. The effect is thus stronger because I compare areas that are only part of cities or towns to areas that are restricted by greenbelts. This helps to identify arguably the ‘true’ supply effect of greenbelts.

overall accessibility of the area and the provision of public goods. Moreover, they absorb price-increasing effects due to a limited supply of land in housing markets with an abundance of greenbelt land. I further control flexibly for distance to the nearest city centre of a city with at least 100,000 inhabitants. Hence:

$$\log p_{it} = \zeta_1 \tilde{g}_i + f(x_i) + \rho_{i \in \mathcal{A}} + \rho_t + \epsilon_{it}, \quad (3)$$

where ρ_t are year fixed effects and ϵ_{it} is an error term. To further address omitted variable bias I employ the same identification strategies as applied to measure the supply effect.

One may be concerned that unobserved housing quality is discontinuous at both sides of the border such that \tilde{g}_i does not capture greenbelt amenities but instead captures a difference in housing quality (*e.g.* that properties are more often detached and larger inside greenbelts). I then exploit an identification strategy similar to [Turner et al. \(2014\)](#). First, I only include areas within a distance b of the inner or outer greenbelt border and control for a second-order polynomial of the distance to the nearest greenbelt boundary on both sides:

$$\log p_{it} = \zeta_1 \tilde{g}_i + f(x_i) + \zeta_2 d_i^- + \zeta_3 (d_i^-)^2 + \zeta_4 d_i^+ + \zeta_5 (d_i^+)^2 + \rho_{i \in \mathcal{A}} + \rho_t + \epsilon_{it}, \quad \text{if } d_i < b. \quad (4)$$

Next, I refine this approach by including observations on either side of the border. The effect of greenbelts can still be identified because properties on one side of the border have different shares of greenbelt land in the vicinity. In this way, I control for the issue that unobserved housing quality may be discontinuous at the greenbelt border. I will address omitted variable bias further in [Appendix B.2.2](#) by obtaining [Oster's \(2019\)](#) bias-adjusted estimates.

Moreover, one may object that the assumption of no impact of greenbelt land beyond 500m is arbitrary. I therefore calculate the share of greenbelt land within 500m distance bands to test for the spatial extent of the greenbelt amenity effect.

3.2.2 Results

In [Table 3](#), I report the reduced-form amenity effects of greenbelts by looking at house prices. In column (1) I estimate a naive specification of having greenbelt land in the vicinity on house prices. I find that there is a strong effect of greenbelt land: a 10-percentage-point increase in the

TABLE 3 – AMENITY EFFECTS OF GREENBELTS: EFFECTS ON HOUSE PRICES
(Dependent variable: the log of house price per m²)

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) OLS | (6) OLS | (7) OLS |
|-------------------------------|-----------------------|---------------------------------|-----------------------|--------------------------|--------------------------|---------------------------|--------------------------|
| | | + Controls and fixed effects | Greenbelts in 1973 | Greenbelt border <2km | Greenbelt border <1km | Outside greenbelt area | Inside greenbelt area |
| Share greenbelt land 0-500m | 0.2351*** (0.0194) | 0.2175*** (0.0097) | 0.1619*** (0.0123) | 0.1233*** (0.0145) | 0.1239*** (0.0184) | 0.0641*** (0.0237) | 0.1777*** (0.0239) |
| Housing attributes | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to the city centre | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | No | Yes | Yes | Yes | Yes |
| Local authority fixed effects | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 10,070,791 | 10,070,791 | 1,952,693 | 3,331,788 | 2,172,516 | 1,988,253 | 184,259 |
| Bandwidth (km) | ∞ | ∞ | ∞ | 2 | 1 | 1 | 1 |
| R ² | 0.3778 | 0.7763 | 0.7911 | 0.7688 | 0.7682 | 0.7699 | 0.7821 |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (3) includes observations in areas that are in greenbelts that were approved or considered in 1973. In column (4) I include transactions that are within 2km of a greenbelt boundary respectively, while this is reduced to 1km in columns (5)-(7). Columns (6) and (7) include properties outside and inside greenbelt areas, respectively. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

share of greenbelt land within 500m increases prices by 2.4%. This is in line with papers that find an amenity effect of open space: greenbelts ensure that houses are closer to open space, which in turn generates positive benefits (see *e.g.* Irwin, 2002; Anderson and West, 2006; Brander and Koetse, 2011).

In column (2) I include a wide range of housing attributes, I control flexibly for the distance to the nearest city centre, and, importantly, I include local authority fixed effects. The latter implies that I identify the amenity effect *within* housing markets. This has limited repercussions for the effect I find, as the coefficient is very similar to the previous specification. Column (3) uses information on proposed and approved greenbelts. The impact of greenbelt land is slightly lower: a 10-percentage-point increase in the share of greenbelt land in the vicinity increases prices by 1.6%.

In the remainder of Table 3, I focus on observations close to inner or outer greenbelt borders. In column (4), I include observations within 2km of a greenbelt border. This implies that I still include approximately one-third of the total number of observations. The coefficient is, again, very similar. Reducing the threshold distance to just 1km does not materially change the results

either (see column (5)).

One may object against these results that housing quality may be discontinuous at the greenbelt border. Although controlling for observed housing quality alleviates these concerns, the possibility remains that unobserved housing quality changes discontinuously at the border. In columns (6) and (7), I therefore include houses on either side of the border. It appears that the greenbelt amenity effect is somewhat lower, albeit highly statistically significant, when I only include properties outside greenbelt areas. The coefficient indicates that a 10-percentage-point increase in the share of greenbelt land within 500m increases prices by 0.6%. The effect is considerably stronger when focusing on properties located inside greenbelt areas and more in line with previous specifications. A potential reason for this is that the effect of greenbelt land is somewhat non-linear. In Appendix B.2.3, I show that the effect of greenbelt land on prices is indeed increasing in the share of greenbelt land in the vicinity. The implication here is that a small amount of greenbelt land will hardly generate an amenity effect, while only beyond a certain share the effect is measurable and positive. Hence, because the share of greenbelt land *inside* greenbelt areas within 500m is considerably higher (on average 61% vs. 38%), this at least partly explains why the effect inside greenbelt areas is stronger.

Hence, these reduced-form regressions show a strong positive amenity effect of greenbelt land. One may wonder whether the assumption that the effect of greenbelt land is zero beyond 500m is valid. I therefore explicitly test for the spatial decay of greenbelt amenities by calculating the share of greenbelt land within 500m distance bands. I report the results in Table 4.

In the specification in column (1), where I include all observations, I find that the main effect is by far the largest and the most precise within 500m of the property. Still, coefficients are positive and (marginally) statistically significant between 1 and 2km. Column (2) is a more convincing specification in that it restricts the sample to observations in 1973 greenbelts. Only the coefficient within 500m is now strong and statistically significant at the 1% level. Alternatively, in column (3), I adopt the boundary-discontinuity approach and include observations within 4km of the greenbelt border, which confirms the previous result that the effect is only strong and statistically significant within 500m. Column (4) further reduces the distance to the nearest greenbelt border to 2km.¹⁷ The finding that the amenity effect of greenbelts is local is in line with a literature

¹⁷I also consider including observations on either side of the border, as in columns (6) and (7) of Table 3. However,

TABLE 4 – AMENITY EFFECTS OF GREENBELTS: SPATIAL DECAY
(Dependent variable: the log of house price per m²)

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS |
|---------------------------------|-----------------------|-----------------------|--------------------------|--------------------------|
| | All obs. | Greenbelts in 1973 | Greenbelt border <4km | Greenbelt border <2km |
| Share greenbelt land 0-500m | 0.1346*** (0.0101) | 0.1216*** (0.0118) | 0.1205*** (0.0114) | 0.1136*** (0.0137) |
| Share greenbelt land 500-1000m | -0.0009 (0.0159) | -0.0020 (0.0204) | -0.0070 (0.0180) | -0.0123 (0.0200) |
| Share greenbelt land 1000-1500m | 0.0565*** (0.0211) | 0.0467 (0.0286) | 0.0361 (0.0234) | 0.0277 (0.0251) |
| Share greenbelt land 1500-2000m | 0.0483* (0.0253) | 0.0215 (0.0355) | 0.0526* (0.0279) | 0.0345 (0.0310) |
| Share greenbelt land 2000-2500m | 0.0453 (0.0302) | 0.0780* (0.0421) | 0.0507 (0.0331) | 0.0509 (0.0368) |
| Share greenbelt land 2500-3000m | -0.0229 (0.0343) | -0.0042 (0.0497) | -0.0183 (0.0375) | -0.0329 (0.0421) |
| Share greenbelt land 3000-3500m | 0.0377 (0.0378) | 0.0629 (0.0526) | 0.0402 (0.0417) | 0.0443 (0.0486) |
| Share greenbelt land 3500-4000m | -0.0366 (0.0336) | -0.0059 (0.0455) | -0.0276 (0.0390) | -0.0314 (0.0441) |
| Housing attributes | Yes | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 10,070,791 | 1,952,693 | 4,710,723 | 3,331,788 |
| Bandwidth (km) | ∞ | ∞ | 4 | 2 |
| R ² | 0.7770 | 0.7924 | 0.7700 | 0.7690 |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (3) includes observations in areas that are in greenbelts that were approved or considered in 1973. In column (4) I include transactions that are within 2km of a greenbelt boundary respectively. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

showing that amenity effects of open space are very local (Bolitzer and Netusil, 2000; Anderson and West, 2006). I further study the spatial decay of greenbelt amenities in the structural model.

3.3 Sensitivity

In Appendix B.2, I investigate the robustness of the reduced-form effects. First, I test in Appendix B.2.1 whether the supply effects extend beyond the borders of greenbelts. For example, density close to greenbelts may be lower, while in city centres a few kilometres away the density may

coefficients are then too imprecise to yield much information.

be higher due to greenbelts. I show that the supply effect only pertains to one's own postcode.

Second, to the extent one is still worried that houses on greenbelt land are different in unobservable housing characteristics from other properties, I obtain in Appendix B.2.2 bias-adjusted estimates using Oster's (2019) GMM approach. This approach exploits coefficient movements after the inclusion of controls together with the variance of the added control variables. I show that for the preferred identification strategies (*e.g.* selecting observations in 1973 greenbelt areas or properties close to greenbelt borders), the bias-adjusted estimates are very close to the estimates shown in Table 3.

Third, as noted earlier, I test in Appendix B.2.3 for the non-linearity of the greenbelt amenity effect, where I find that the effect of greenbelt land on prices is increasing in the share of greenbelt land.

Fourth, in Appendix B.2.4, I categorise greenbelt land in 6 land-use types (parks and recreation areas, golf courses, forests, farms, meadows, and other) using data from OpenStreetMap. I find that parks and recreation areas do not have a statistically significant effect. This may be because recreation areas like football fields or cricket grounds are not particularly aesthetically pleasing and even may generate negative externalities. By contrast, golf courses and open land generally command a (strong) positive price premium. This is in line with existing evidence showing that open land is valued more highly by residents than *e.g.* forested land (Irwin 2002, Montgomery 2015, pp. 114-115).

Fifth, one may be concerned that the effects of greenbelts partly capture a sorting effect, which would mean that residents with price-increasing characteristics end up in the greenbelt (see Bayer et al., 2007). To test for sorting effects and other omitted variables, in Appendix B.2.5 I include output area (OA) fixed effects, which are *very small areas* and the lowest geographical level at which Census estimates are provided (the median size of an OA is just 6.6ha). I find that this does not materially change the supply and amenity effects of greenbelts, which strongly suggests that what I capture here is a direct effect rather than a sorting effect due to greenbelts.

Sixth, in Appendix B.2.6 I test for the sensitivity of the results by only including observations within just 100m of an inner or outer greenbelt border. This does not significantly affect my results, although coefficients are, unsurprisingly, less precise.

Seventh, as the structural estimation results will mostly rely on [Census](#) data from 2011 and the greenbelt data from 2012, I also re-estimate the preferred specifications including data only for 2012 in [Appendix B.2.7](#), leading to essentially the same results. Because one may be concerned that flats are under-represented in the data, I further estimate regressions where I only include flats, which leads to very similar results.

Eighth, one may be concerned that the spatial distribution of sales is not random and correlated to the share of greenbelt land. For example, particular houses in greenbelts may stay in the family or are passed through inheritance, and hence are not traded on the market. This would lead to a lower number of sales per dwelling in postcodes in greenbelts. In [Appendix B.2.8](#), I do not find that this is an issue.

Finally, one of the other effects on greenbelts is that they ‘absorb’ air pollution from cities and therefore lead to lower pollution levels in cities with a greenbelt. I test this by exploiting data on concentrations of particulate matter (PM_{10}) and nitrogen oxides (NO_x). These are pollutants commonly associated with the concentration of human activities. Nitrogen oxides are particularly associated with (gas) heating, traffic, and electricity production. I show in [Appendix B.3](#) that areas with more greenbelt land do seem to have lower pollution levels. However, I find that the effect is confined to one’s own MSOA. This implies that pollution reductions of greenbelts are relatively local and that generally, throughout the city, pollution levels are not lower.

4 A spatial general equilibrium model

4.1 Introduction

In this section, I introduce a structural model to analyse the general equilibrium effects of greenbelt policy. I extend [Ahlfeldt et al.’s \(2015\)](#) model in three ways. First, I embed land-use restrictions in the model, as greenbelt land reduces the density for development at certain locations (see [Section 3.1](#)). Second, I allow for greenbelts to generate a higher amenity level (see [Section 3.2](#)); hence, I explicitly specify the amenity residual in [Ahlfeldt et al. \(2015\)](#) and allow for the spatial decay of those amenities. Third, I allow for mode choice and endogenous travel times of road travel. Furthermore, I estimate, rather than choose, the share of land in construction costs, which is critical in determining the effect of greenbelts on floor space supply

and prices.

In Section 4.2, I first introduce how workers choose residential, workplace locations and transport modes as a function of amenities, floor space prices, and commuting costs. Then, in Section 4.3, I proceed to discuss production at each location, which is a function of agglomeration economies. Third, in Section 4.4, I discuss land use and construction, followed by the modelling of traffic congestion in Section 4.5. I elaborate on welfare issues in Section 4.6. The estimation of the model's parameter will be discussed in Section 4.7.

4.2 Workers and greenbelt amenities

There are $i = 1, \dots, \mathcal{L}$ locations, each with land area L_i and $m = \{\tau d, \tau w, oth\}$ are transport modes. I distinguish between travel by road (τd), rail (τw), and other modes (oth). In what follows, only road traffic is subject to traffic congestion. Land may be used for residential purposes and/or production. A worker z that lives in i and commutes to j by mode m has preferences over a consumption good c_{ijmz} and residential floor space ℓ_{ijmz} . The worker also has an idiosyncratic preference for pair ijm , denoted by ξ_{ijmz} .¹⁸ After deciding to locate in England, but before determining where exactly, the idiosyncratic component of utility is revealed to the worker. I assume that utility is Cobb-Douglas:

$$U_{ijmz} = \Psi_i \left(\frac{c_{ijmz}}{\beta} \right)^\beta \left(\frac{\ell_{ijmz}}{1-\beta} \right)^{1-\beta} \xi_{ijmz}, \quad (5)$$

where Ψ_i is the given amenity level of a location and preferences for the consumption good $0 < \beta < 1$. The idiosyncratic component is drawn from a Fréchet distribution:

$$F(\xi_{ijmz}) = e^{-\bar{\nu}_{im}\bar{\nu}_{jm}\xi_{ijmz}^{-\varepsilon}}, \quad \bar{\nu}_{im}, \bar{\nu}_{jm} > 0, \varepsilon > 1. \quad (6)$$

The location-mode-specific scale parameters $\bar{\nu}_{im}$ and $\bar{\nu}_{jm}$ determine the average utility of living in i and working in j given transport mode m , respectively, and ε governs the amount of commuting heterogeneity. Note that a higher value of ε implies a smaller dispersion of utilities.

Workers earn a wage w_j at their workplace j . They must commute to work, which implies a loss

¹⁸Other than heterogeneity through idiosyncratic preferences, I abstract from heterogeneity in *e.g.* income, such as in [Gaigné et al. \(2021\)](#) or [Tsivanidis \(2020\)](#). Unfortunately, I lack good micro-data on incomes. Moreover, allowing for heterogeneous workers would render the model much more difficult and probably obscure the key forces at work.

in their net wage. More specifically, the workers' budget constraint is given by $e^{-\kappa_m \tau_{ijm}} w_j = p_i \ell_{ijmz} + c_{ijmz}$, where $e^{-\kappa_m \tau_{ijm}}$ represents iceberg commuting costs, τ_{ijm} is the travel time between location i and j by mode m , and p_i is the price per unit of floor space.¹⁹

The indirect utility is then given by $u_{ijmz} = \Psi_i e^{-\kappa_m \tau_{ijm}} w_j p_i^{\beta-1} \xi_{ijmz}$. Given that ξ_{ijmz} is Fréchet distributed, I can determine the probability that a worker chooses to reside in i and work in j and use transport mode m :

$$\pi_{ijm} = \frac{\bar{\nu}_{im} \bar{v}_{jm} \left(\frac{\Psi_i e^{-\kappa_m \tau_{ijm}} w_j}{p_i^{1-\beta}} \right)^\varepsilon}{\sum_{r=1}^{\mathcal{L}} \sum_{s=1}^{\mathcal{L}} \sum_{m=1}^3 \bar{\nu}_{rm} \bar{v}_{sm} \left(\frac{\Psi_r e^{-\kappa_m \tau_{rsm}} w_s}{p_r^{1-\beta}} \right)^\varepsilon}. \quad (7)$$

I define 'transformed' wages as $\omega_{jm} = \bar{v}_{jm} w_j^\varepsilon$. 'Transformed' amenities are given by $a_{im} = \bar{\nu}_{im} \Psi_i^\varepsilon p_i^{\varepsilon(\beta-1)}$. The probability that a worker is employed in j , *conditional* on living in i and taking mode m , is given by:

$$\pi_{ijm|i} = \frac{e^{-\kappa_m \varepsilon \tau_{ijm}} \omega_{jm}}{\sum_{s=1}^{\mathcal{L}} e^{-\kappa_m \varepsilon \tau_{ism}} \omega_{sm}}. \quad (8)$$

Similarly, the probability that someone lives at i , *conditional* on working in j and taking mode m , is given by:

$$\pi_{ijm|j} = \frac{e^{-\kappa_m \varepsilon \tau_{ijm}} a_{im}}{\sum_{r=1}^{\mathcal{L}} e^{-\kappa_m \varepsilon \tau_{irm}} a_{rm}}. \quad (9)$$

Note that I observe the number of workers in the data, H_{Mjm} , as well as the number of residents living at i and using mode m , H_{Rim} . Given the above probability, I can define the *commuting market-clearing conditions*:

$$\begin{aligned} H_{Mjm} &= \sum_{i=1}^{\mathcal{L}} \pi_{ijm|i} H_{Rim}, \\ H_{Rim} &= \sum_{j=1}^{\mathcal{L}} \pi_{ijm|j} H_{Mjm}. \end{aligned} \quad (10)$$

The first condition implies that the number of workers in j is the sum of the residential population multiplied by the probability that they commute to j . The second condition ensures that

¹⁹One may argue that commuting costs contain a time and pecuniary component. I do not dispute this, but allowing for this type of commuting costs goes beyond the scope of this paper. Importantly, given the value of time obtained in empirical studies (see [Small and Verhoef, 2007](#)), it appears that time costs far exceed the pecuniary component of commuting costs.

the number of residents at i equals the sum over the workplace population multiplied by the probability that they commute from j .

The total residential floor consumption F_{Ri} at i is obtained by summing the floor space demand over all workers in a location:

$$F_{Ri} = \sum_{m=1}^3 \frac{(1-\beta)\mathbb{E}[\tilde{w}_{im}]}{p_i} H_{Rim}, \quad (11)$$

where $\mathbb{E}[\tilde{w}_{im}] = \sum_{j=1}^{\mathcal{L}} \pi_{ijm|im} e^{-\kappa_m \tau_{ijm}} w_j$ denotes the expected wage net of commuting at location i for transport mode m .

I assume that workers obtain an expected utility equal to a reservation utility \bar{u} which is the same for everyone. Moving is costless and population mobility implies the following:

$$\bar{u} = \mathbb{E}[u] = \Gamma\left(\frac{\varepsilon-1}{\varepsilon}\right) \left(\sum_{i=1}^{\mathcal{L}} \sum_{j=1}^{\mathcal{L}} \sum_{m=1}^3 \bar{\nu}_{im} \bar{\nu}_{jm} \left(\frac{\Psi_i e^{-\kappa_m \tau_{ijm}} w_j}{p_i^{1-\beta}} \right)^{\varepsilon} \right)^{\frac{1}{\varepsilon}}, \quad (12)$$

where $\Gamma(\cdot)$ is the Gamma function.

Given the reduced-form results, I expect a higher amenity level in areas that are closer to or inside greenbelts. Moreover, residents may dislike areas with high densities (*e.g.* because in dense areas buildings appear more often in the form of apartments, see [Glaeser et al., 2008](#)). Let $\hat{\Psi}_i$ denote the amenity value recovered from the data (I explain how Ψ_i is recovered in Section 4.7). Then:

$$\hat{\Psi}_i = \check{\Psi}_i e^{\zeta \tilde{g}_i(\theta)} \left(\frac{\sum_{m=1}^3 H_{Rim}}{L_i} \right)^{\psi}, \quad (13)$$

where $\check{\Psi}_i$ is an amenity constant and

$$\tilde{g}_i(\theta) = \frac{\sum_{j=1}^{\mathcal{L}} e^{-\theta d_{ij}} G_j}{\sum_{j=1}^{\mathcal{L}} e^{-\theta d_{ij}} L_j}. \quad (14)$$

Hence, $\tilde{g}_i(\theta)$ is a spatially weighted share of greenbelt land around i as G_j is the total amount of greenbelt land at j and recall that L_j is the total amount of land at j . Furthermore, $\zeta > 0$ indicates the amenity effect of greenbelt land on the amenity level Ψ_i , $\theta > 0$ indicates the decay

of this amenity effect, and ψ indicates how much residents value residential density.²⁰

4.3 Production and agglomeration economies

I now turn to production. A single final good is produced in a perfectly competitive market with constant returns to scale and sold to the wider economy without costs. Cobb-Douglas production in location j is given by:

$$Y_j = \sum_{m=1}^3 \Omega_j (\varsigma H_{Mjm})^\alpha (F_{Mjm})^{1-\alpha}, \quad (15)$$

where Ω_j denotes the final goods productivity at j , given mode m , F_{Mj} is the amount of floor space consumed by production, and ς is a scale parameter.²¹ Recall that the costs of labour are given by the wages w_j , while prices p_i are the costs for using floor space. Assuming profit maximisation implies that floor space consumption at a certain location j is then given by:

$$F_{Mj} = \sum_{m=1}^3 \left(\frac{\omega_{jm}}{\alpha \Omega_j \bar{v}_{jm}^{\frac{1}{\epsilon}}}} \right)^{\frac{1}{1-\alpha}} \varsigma H_{Mjm}. \quad (16)$$

Given the optimal use of labour, the final goods productivity of a location j , Ω_j can then be recovered up to a normalisation constant:²²

$$\Omega_j = \frac{\bar{v}_{jm}^{-\frac{\alpha}{\epsilon}}}{(1-\alpha)^{1-\alpha} \alpha^\alpha} p_j^{1-\alpha} \omega_{jm}^{\frac{\alpha}{\epsilon}}, \quad \forall m. \quad (17)$$

As there is a large literature showing that firms are more productive in the vicinity of others (see *e.g.* Combes et al., 2008; Melo et al., 2009), I assume that productivity is dependent on employment density. Using (17) and data on floor space prices and transformed wages, I recover

²⁰I may also allow for the decay of residential density, as in Ahlfeldt et al. (2015). In the current paper, it appears that this decay parameter is difficult to estimate but is, if anything, very high. In other words, density effects only pertain to one's own MSOA, which is in line with the strong decay found in Ahlfeldt et al. (2015).

²¹More specifically, I choose ς in such a way that $(\sum_{j=1}^{\mathcal{L}} \sum_{m=1}^3 F_{Mjm}) / (\sum_{j=1}^{\mathcal{L}} \sum_{m=1}^3 F_{Rjm})$ is equal to the overall ratio of commercial floor space to residential floor space in the data. Workers need relatively less office space than residential space so $\varsigma < 1$.

²²The optimal use of floor space is given by $F_{jm}^* = ((1-\alpha)Y_{jm})/p_j$, while the optimal use of labour is $H_{jm}^* = (\alpha Y_{jm})/\varsigma w_j$. I then plug F_{jm}^* and H_{jm}^* into Y_j to obtain Ω_j . As in Ahlfeldt et al. (2015), Ω_j is identified up to a normalisation constant. I normalise $\bar{v}_{j,\tau d} = 1$, while I choose values of $\bar{v}_{j,\tau w}$ and $\bar{v}_{j,\tau ofl}$ to ensure that $\Omega_j = \Omega_{j,\tau d} = \Omega_{j,\tau w} = \Omega_{j,\tau ofl}$.

$\hat{\Omega}_j$ from the data. Then:

$$\hat{\Omega}_j = \check{\Omega}_j \left(\frac{\sum_{i=1}^{\mathcal{L}} \sum_{m=1}^3 e^{-\delta \tau_{ij, \cdot d}} H_{Mim}}{\sum_{i=1}^{\mathcal{L}} e^{-\delta \tau_{ij, \cdot d}} L_i} \right)^{\gamma} \equiv \check{\Omega}_j A_{Mj}^{\gamma}, \quad (18)$$

where $\check{\Omega}_j$ denotes the constant exogenous productivity of a location j , γ is the agglomeration elasticity, A_{Mj} measures employment density, and δ captures the spatial decay of agglomeration economies.²³ Hence, productivity is positively affected by employment density if $\gamma > 0$.

4.4 Land, construction, and greenbelts

Greenbelts limit the available supply of land for development. Let Λ_i be the amount of developed land in i and L_i the total land area in location i . In line with the reduced form estimations in Section 3.1, I assume that the relationship between the share of built-up land and greenbelts is given by:

$$\frac{\Lambda_i}{L_i} = \Phi_i e^{\phi g_i}, \quad (19)$$

where Φ_i is a location-specific constant and $g_i = G_i/L_i$ is the share of greenbelt land at i .

I assume that floor space F_i is supplied in a competitive construction market that uses developed land Λ_i and capital K_i as inputs, with land prices P_i and the rental costs of capital denoted by r . I use a standard Cobb-Douglas production function. Land market-clearing then implies that:

$$F_i = F_{Mi} + F_{Ri} = \Upsilon_i K_i^{\mu} \Lambda_i^{1-\mu}, \quad (20)$$

The first-order condition for optimal capital use is then:

$$K_i^* = \left(\frac{\mu \Upsilon_i p_i}{r} \right)^{\frac{1}{1-\mu}} \Lambda_i, \quad (21)$$

which I plug in (20) to obtain:

$$\frac{F_i}{\Lambda_i} = \Upsilon_i^{\frac{1}{1-\mu}} \left(\frac{\mu p_i}{r} \right)^{\frac{\mu}{1-\mu}}. \quad (22)$$

²³In a previous version of this paper, I also allowed for the direct effects of greenbelts on productivity. I did not find clear evidence for a ‘productive amenity’ effect of greenbelts. More importantly, the results of the counterfactuals are hardly affected by the inclusion of direct effects of greenbelts, while they complicate the estimation. These results are therefore available upon request.

By plugging the first-order condition (21) into (20), I can solve for innate supply conditions Υ_i :

$$\hat{\Upsilon}_i = \left(\frac{r}{\mu p_i} \right)^\mu \left(\frac{F_i}{\Lambda_i} \right)^{1-\mu}. \quad (23)$$

4.5 Traffic congestion

In the model of Ahlfeldt et al. (2015), travel times between two locations are exogenous. However, according to a large literature on transportation, it is clear that travel times on the road are endogenous because of congestion. Traffic congestion mostly occurs in peak hours when commuting to work (Vickrey, 1969; Peer et al., 2015; Proost and Thisse, 2019). Workplaces tend to be more spatially concentrated than residences, meaning that congestion tends to occur mostly at the workplace. I therefore focus on congestion costs at the *workplace* when workers choose to travel by car, defined as $m = \tau \ell$.²⁴

Pigou (1932) assumed a relationship between traffic flow and travel speed, but what matters here is the relationship between (traffic) density and travel time (Small and Verhoef, 2007). I assume a generalisation of the relationship between density and travel time proposed by Underwood (1961) and for which there is ample empirical support (see e.g. Daganzo et al., 2011; Geroliminis and Daganzo, 2008; Adler et al., 2021; Russo et al., 2021):

$$\tau_{ij,\tau\ell} = \tau_{ij,\tau\ell}^f T_i \check{T}_j e^{\lambda \mathcal{D}_{Mj}}, \quad (24)$$

where τ_{ij}^f is the free-flow travel time between i and j , T_i and \check{T}_j are location-specific constants, and $\lambda > 0$ is the congestion elasticity. Importantly, the density of traffic is given by:

$$\mathcal{D}_{Mj} = \frac{\sum_{s=1}^{\mathcal{L}} e^{-\kappa \tau_{sj,\tau\ell}} H_{Ms,\tau\ell}}{\sum_{s=1}^{\mathcal{L}} e^{-\kappa \tau_{sj,\tau\ell}} \mathcal{R}_s}, \quad (25)$$

where \mathcal{R}_s is the total length of roads at s . Hence, I divide the spatially weighted employment by car in a commuting area by the total available roads to obtain the traffic density of workers. Equation (24) is, for example, in line with traffic congestion models, where congestion is modelled as a bathtub that is filled up by traffic (Arnott and Rowse, 2013; Fosgerau, 2015).

²⁴I will show in Appendix C.4 that congestion at the residential location indeed matters less.

4.6 Welfare

To evaluate whether greenbelt policy improves welfare, I will analyse the change in expected utility once greenbelt land is removed. The change in expected utility can be interpreted as the change in income that is necessary to obtain the same utility as in the baseline situation. The equivalent income increase that is necessary to make residents in the original scenario have the same utility as in the new scenario can be defined as (using (12)):

$$\left(\sum_{i=1}^{\mathcal{L}} \sum_{j=1}^{\mathcal{L}} \sum_{m=1}^3 \bar{v}_{im0} \bar{v}_{jm0} \left(\frac{\Psi_{i0} e^{-\kappa \tau_{ijm0}} w_{j0} \Delta \bar{w}}{p_{i0}^{1-\beta}} \right)^{\varepsilon} \right)^{\frac{1}{\varepsilon}} = \left(\sum_{i=1}^{\mathcal{L}} \sum_{j=1}^{\mathcal{L}} \sum_{m=1}^3 \bar{v}_{im1} \bar{v}_{jm1} \left(\frac{\Psi_{i1} e^{-\kappa \tau_{ijm1}} w_{j1}}{p_{i1}^{1-\beta}} \right)^{\varepsilon} \right)^{\frac{1}{\varepsilon}},$$

$$\Delta \bar{w} = \frac{\bar{u}_1}{\bar{u}_0},$$
(26)

Recall that w_{j0} and w_{j1} are wages in the baseline and counterfactual scenario without greenbelt land, respectively. Further, be reminded that τ_{ijm} are mode-specific travel times, Ψ_i are amenity levels, while p_i are floor-space prices. $\Delta \bar{w}$ is the income increase that compensates for the utility differential, to which I refer to as the change in *equivalent income*.

Moreover, as the workers in my model are renters, I should also take into account the change in the land rents for absentee landlords. Given that construction firms make zero profits, I have:

$$P_i = \frac{F_i p_i}{\Lambda_i} - r^{\frac{\mu}{\mu-1}} (\mu p_i \Upsilon_i)^{\frac{1}{1-\mu}}. \quad (27)$$

Then, the change in aggregate land prices is given by $(P_{i1} \Lambda_{i1}) / (P_{i0} \Lambda_{i0}) - 1$.

4.7 Model estimation

4.7.1 Set up

I use the recursive structure of the model to solve for the parameters of interest $\{\kappa_{\tau d}, \kappa_{\tau w}, \kappa_{\tau h}, \varepsilon, \lambda, \phi, \zeta, \theta, \psi, \gamma, \delta\}$. I assume that the expenditure on labour costs equals $\alpha = 0.75$, which is in line with the long-run average in the UK (Batini et al., 2000). The share of household expenditure on floor space is assumed to be $1 - \beta = 0.25$ (Department for Communities and Local Government, 2013). I set the parameter $\varsigma = 0.1435$ based on data on floor spaces from

ONS and DCLG (see Department for Communities and Local Government, 2018).²⁵ I estimate the model at the MSOA level.

Note that in what follows I use practically the same identification strategies to identify the parameters of interest as those outlined in the reduced-form regressions (see Section 3) but the dependent variables are now different and based on the model-implied location fundamentals (such as $\hat{\Psi}_i$ and $\hat{\Omega}_j$).

4.7.2 The gravity equation and wages

I obtain information on bilateral commuting pairs by travel mode from the Census. In order to estimate the model, I only keep commuting pairs for which the free-flow travel time is less than 120 minutes (one-way), as there are few people (around 1%) commuting more than this.²⁶

Let us now define $\varkappa_m \equiv \kappa_m \varepsilon$. I estimate the following gravity equation:

$$\pi_{ijm} \bar{H} = e^{-\varkappa_m \tau_{ijm} + \tilde{\nu}_{im} + \tilde{\nu}_{jm} + \epsilon_{ijm}}, \quad (28)$$

where \bar{H} is England's total population, \varkappa_m are the mode-specific commuting travel time elasticities, $\tilde{\nu}_{im}$ is a residential location-by-mode fixed effect absorbing $\{\Psi_i, p_i, \bar{\nu}_{im}\}$, and $\tilde{\nu}_{jm}$ is a workplace-by-mode fixed effect absorbing $\{\bar{\nu}_{jm}, w_j\}$ (see equation (7)). Because the dependent variable $\pi_{ijm} \bar{H}$ has many zeroes, I estimate equation (28) using a Poisson model with two-way fixed effects. In Appendix C.1, I consider different specifications to obtain \varkappa_m . Most importantly, I consider the issue that travel times might be endogenous: between locations where there is a higher commuting flow, it is more likely that new transport infrastructure is provided, leading in turn to lower travel times. Moreover, the measuring of travel times may be vulnerable to error. Travel times by road are also endogenous because a higher flow may imply that congestion may be more severe, which in turn increases travel times by road. I show in Appendix C.1 that when instrumenting travel times with the Euclidean distance, endogeneity hardly matters.

²⁵Note that ς does not capture the share of commercial space to total real estate space, but it is a parameter that ensures that the ratio of commercial floor space to residential floor space matches the data. In 2008, the total commercial and industrial space was 562 million m², while the total residential floor space was 22.2 million dwellings \times 92m² = 2.042 billion m². Hence, the ratio of commercial to residential floor space was 0.2745. In the baseline scenario, the total amount of commercial floor space predicted by the model is 1543, while the predicted amount of residential floor space is 805. Hence, to adjust the amount of commercial floor space to match the ratio to residential floor space, I set $\varsigma = 0.1435$.

²⁶To be more specific, I exclude only 1% of commuters by road, 1.1% of commuters by rail, and 1.2% of commuters by other modes.

Using data on the (working) population H_{Rim} , and the number of workers H_{Mjm} , and the estimated parameters $\hat{\kappa}_m$, I can recover transformed wages ω_{jm} at each location j for each mode m by solving this equation:

$$H_{Mjm} - \sum_{i=1}^{\mathcal{L}} \pi_{ijm|im} H_{Rim} = 0. \quad (29)$$

Similarly, I obtain amenities up to a constant by solving $H_{Rim} - \sum_{j=1}^{\mathcal{L}} \pi_{ijm|jm} H_{Mjm} = 0$, which produces an estimate for transformed amenities \hat{a}_{im} . Given (9), I recover amenity level, Ψ_i :

$$\hat{\Psi}_i = \left(\frac{\hat{a}_{im}}{\bar{\nu}_{im}} \right)^{\frac{1}{\varepsilon}} p_i^{1-\beta}, \quad \forall m \in j. \quad (30)$$

4.7.3 Commuting heterogeneity

I recover ε by using information on the distribution of estimated household incomes in England.²⁷ Following Ahlfeldt et al. (2015), I choose ε in such a way that it minimises the squared differences between the variances within local authority areas of log-transformed wages in the model and log wages in the data. I estimate:

$$\sigma_{\log w_{i \in z}}^2 = \left(\frac{1}{\varepsilon} \right)^2 \sigma_{\log \hat{\omega}_{im \in z}}^2 + \epsilon_{im}. \quad (31)$$

Using $\hat{\varepsilon}$, I obtain $\hat{\kappa}_m = \hat{\kappa}_m / \hat{\varepsilon}$.

4.7.4 Effects of greenbelt policy on land and construction

By log-linearising equation (19), I estimate the impact of greenbelts on the land available for development:

$$\log \frac{\Lambda_i}{L_i} = \phi g_i + \eta_{i \in \mathcal{G}} + \epsilon_i, \quad (32)$$

where ϕ captures the parameter of interest and $\eta_{i \in \mathcal{G}}$ are greenbelt-by-government region fixed effects.²⁸ Note that greenbelt policy is expected to lead to less land being available for development (as shown in the reduced-form results, Section 3.1); hence, ϕ is expected to be negative. I also use alternative identification strategies to mitigate endogeneity issues. More specifically, I use (i) only areas for which the share in proposed or approved greenbelt land in 1973 exceeds

²⁷I obtain data on estimated household incomes using MSOA by the Office of National Statistics from 2011.

²⁸There are nine government regions in England. When intersecting those with 12 greenbelts, I end up with 32 areas.

90%, and (ii) I focus on MSOAs that are within 1km of an inner or outer greenbelt boundary.

The next step is to recover the share of construction costs used for capital, as denoted by μ . I can calculate floor space consumption (see (11) and (16)) using the estimated parameters $\{\hat{\kappa}, \hat{\varepsilon}\}$; implied transformed wages; data on floor space prices; number of residents and workers; productivities; and amenities. By log-linearising (22), I estimate:

$$\log \frac{F_i}{\Lambda_i} = \frac{\mu}{1 - \mu} \log p_i + \iota s_i + \eta_{i \in \mathcal{Q}} + \epsilon_i, \quad (33)$$

where s_i denotes supply conditions, such as distance to bedrock and sediment thickness. Hence, I recover μ from a standard linear regression. However, I face endogeneity issues because the supply of floor space may also impact the floor space price (Combes et al., 2016). As is shown in Appendix C.2.1, μ even has the wrong sign if I estimate the above equation using OLS.

Following Combes et al. (2016), I therefore propose to instrument for p_i using the *distance to the city centre* as an instrument for floor space prices. According to the monocentric city model (Alonso, 1964; Mills, 1967; Muth, 1969), demand for housing closer to the city centre is higher due to shorter commutes, not because of differences in supply conditions. Still, one may be concerned that unobserved supply conditions are correlated to the distance to the centre. It has been argued that construction costs are related to supply conditions and therefore the locations of city centres may be related to, say, the depth to bedrock (Rosenthal and Strange, 2008; Barr et al., 2011; Holl, 2019). I therefore include greenbelt-by-government region effects and variables capturing supply conditions, such as the elevation and share of the workforce in construction (which may affect the wages for construction workers). To the extent one is concerned that the instrument is not valid, in Appendix C.2.2 I consider Conley et al.'s (2012) methodology to relax the assumption of strictly exogenous instruments. This enables me to construct bounds on the estimated μ if part of the effect of distance to the city centre is related to supply conditions. In this Appendix, I further consider alternative instruments based on temperature as a proxy for amenities and therefore demand for housing (see Glaeser et al., 2001).

Furthermore, one may argue that μ is heterogeneous across locations and dependent on the share of greenbelt land in i . To investigate this further, I estimate a semi-parametric version of (33) in Appendix C.2.3, but find little variation in μ . Given that allowing for variation in the

share of capital costs in construction will considerably complicate the estimation procedure, I will assume μ to be identical across locations for the remainder of this paper.

4.7.5 Greenbelt amenities

Armed with estimates for amenities (up to a normalisation constant) based on equation (30), I identify the impact of greenbelts and residential density on amenities using:

$$\log \hat{\Psi}_i = \zeta \left(\frac{\sum_{j=1}^{\mathcal{L}} e^{-\theta d_{ij}} G_j}{\sum_{j=1}^{\mathcal{L}} e^{-\theta d_{ij}} L_j} \right) + \psi \log \left(\frac{H_{Ri}}{L_i} \right) + \eta_{i \in \mathcal{G}} + \epsilon_i. \quad (34)$$

By adopting non-linear least squares and by using the above-discussed identification strategies, I identify ζ , the decay of greenbelt amenities θ , as well as the effect of residential density ψ , on the implied amenity level.

A concern with (34) is that H_{Ri}/L_i may be endogenous. It may well be that amenities attract more people, which in turn affects amenities. Also, H_{Ri}/L_i may be correlated to unobserved endowments. I follow two approaches in order to address this. First, I use the familiar approach to use population density in 1931 (H_{1931i}/L_i) as an instrument. The idea here is that past unobservables are uncorrelated to current unobservables (see [Ciccone and Hall, 1996](#); [Ciccone, 2002](#); [Combes et al., 2008](#); [Melo et al., 2009](#)). Moreover, this allows me to address any reverse causality.

Still, the identifying assumption that the unobserved reasons why people cluster are uncorrelated over 80 years may fail to be fully convincing. I therefore also consider an alternative identification strategy: I use the share of greenbelt land between 10 and 25km. Following [Bayer et al. \(2007\)](#) and [Bayer and Timmins \(2007\)](#), the identifying assumption is therefore that the share of greenbelt land far away does not generate direct utility effects other than via the effects on the spatial distribution of residents.

Because (34) is non-linear, I adopt a control function approach (see [Cameron and Trivedi, 2005](#)) where the first-stage errors are included as a third-order polynomial in equation (34).

4.7.6 Productivity and agglomeration economies

I also recover productivity $\hat{\Omega}_i$ using equation (17). I then estimate the following:

$$\log \hat{\Omega}_i = \gamma \log A_{Mi}(\delta) + \eta_{i \in \mathcal{G}} + \epsilon_i. \quad (35)$$

The above equation is non-linear in parameters γ and δ . Recall from equation (18) that A_{Mi} is a function of $H_{Mj} = \sum_{m=1}^{\mathcal{L}} H_{Mjm}$, $\forall j$. I therefore sum H_{Mj} using 2-minute travel time ‘doughnuts’ and estimate the above equation using non-linear least squares.²⁹

A concern with the above equation is that H_{Mj} is endogenous. For example, δ may capture the effect of unobserved locational endowments rather than agglomeration economies. I therefore once more use the standard strategy to instrument for spatial externalities. That is, I use the population in 1931 and estimate the following first stage:

$$\log A_{Mi}(\delta) = \tilde{\gamma} \log \left(\frac{\sum_{j=1}^{\mathcal{L}} e^{-\tilde{\delta} \tilde{\tau}_{1870ij}} H_{1931j}}{\sum_{j=1}^{\mathcal{L}} e^{-\tilde{\delta} \tilde{\tau}_{1870ij}} L_j} \right) + \tilde{\eta}_{i \in \mathcal{G}} + \tilde{\epsilon}_i. \quad (36)$$

where H_{1931j} is the population in the same MSOA in 1931 and $\tilde{\tau}_{1870ij}$ is the travel time over the railway network of 1870 (which was the closest year from which I could retrieve infrastructure data). I then obtain residuals $\tilde{\epsilon}_i$. To determine γ and δ simultaneously, I then use a control function approach where I include a third-order polynomial of $\tilde{\epsilon}_i$ in (35).

As an alternative instrument, I once more use the share of greenbelt land between 10 and 25km. Thus, the identifying assumption is that the share of greenbelt land far away does not generate direct productivity effects other than via the effects on the spatial equilibrium of workers choosing workplaces. Hence, the alternative first stage is given by:

$$\log A_{Mi}(\delta) = \check{\gamma}_M \frac{\sum_{j=1}^{\mathcal{L}} G_j I_{10 < d_{ij} \leq 25}}{\sum_{j=1}^{\mathcal{L}} L_j I_{10 < d_{ij} \leq 25}} + \check{\eta}_{i \in \mathcal{G}} + \check{\epsilon}_i. \quad (37)$$

Again, I plug in a third-order polynomial of the residuals $\check{\epsilon}_i$ in equation (35) to obtain values for the parameters of interest.

²⁹More specifically, for each location i , I calculate the travel time to all j . I then generate variables that sum workers for a given travel time ring, implying that I will include 60 variables; the total number of workers between 0-2 minute travelling, 2-4 minutes travelling, etc.

4.7.7 Traffic congestion elasticity

To obtain the congestion elasticity λ , I first log-linearise equation (24) and estimate:

$$\log \frac{\tau_{ij}^f}{\tau_{ij}} = \log T_{Ri} + \log T_{Mj} + \epsilon_{ij}. \quad (38)$$

Recall that τ_{ij}^f is the free-flow travel time between i and j . I estimate the above equation using a regression with two-way fixed effects. In the second step, I recover the workplace fixed effects ($\log T_{Mj}$), resulting in the following equation:

$$\log \hat{T}_{Mi} = \lambda \mathcal{D}_{Mi} + \eta_{i \in G} + \epsilon_i, \quad (39)$$

where $\eta_{i \in G}$ are greenbelt-by-government fixed effects, and I use the estimated parameter $\hat{\kappa}_{\mathcal{D}}$ to calculate traffic density \mathcal{D}_{Mi} . I estimate this equation using OLS.³⁰

One may be concerned that \mathcal{D}_{Mi} is endogenous because of reverse causality; that is, short travel times may attract residents and workers that are interested in shorter commutes, thus leading to a higher density. I follow a similar strategy as when measuring agglomeration economies: I instrument traffic densities with the population density in 1931. Alternatively, I use the weighted share of greenbelt land between 10 and 25km as an instrument for traffic density at i .

4.7.8 Standard errors

I obtain standard errors by bootstrapping this whole procedure 250 times. More specifically, I begin by randomly selecting \mathcal{L} MSOAs (with replacement) and, given this set of locations, estimate each of the consecutive steps. In this way, I take into account that errors are correlated between different equations.

5 Model parameters and counterfactuals

5.1 Structural parameters

Table 5 reports the estimated model parameters for different specifications. I first discuss the estimated parameters related to commuting, then the parameters related to greenbelt restrictions

³⁰I show in Appendix C.4 that my results are robust to omitting home location fixed effects.

and amenities. Finally, I discuss the parameters capturing spillovers.

5.1.1 Commuting parameters

In columns (1)-(3) in Table 5, I do not instrument for travel time. I find a commuting semi-elasticity for road travel $\varkappa_{i,d} = \kappa_{i,d}\varepsilon = 0.0686$. This is somewhat lower than Ahlfeldt et al. (2015) because I use the actual travel time between i and j , rather than the free-flow travel time. I show in Appendix C.1 that commuting time elasticity is approximately twice as strong when using free-flow travel times. In Appendix C.3, I show that around 50% of utility ‘melts’ away with a commute of 10 minutes by car and only 2% remains for a one-hour commute. This indicates that workers strongly dislike commuting. For travel time by train, meanwhile, I find a semi-elasticity of 0.043, which is somewhat lower than for car travel. An explanation for this may be that the opportunity costs when travelling by train may be lower, because people can engage in other activities while travelling (see Ohmori and Harata, 2009). The travel time elasticity for other modes is low ($\varkappa_{i,d} = 0.0121$). I calculated travel time whilst assuming a speed of 10km over the network. However, when workers use other travel modes (such as travelling by bus or by bicycle), the travel time elasticity will be underestimated. In columns (4)-(9), I instrument for travel times by Euclidean distance. This hardly influences the estimated parameters.

The commuting heterogeneity parameter $\hat{\varepsilon} = 5.527$ is very similar to that reported by Ahlfeldt et al. (2015) and within the range provided by Eaton and Kortum (2002).

5.1.2 Greenbelt parameters

Let us now move to the parameters capturing greenbelt restrictions and amenities, which are of primary interest.

When identifying the effect of greenbelt restrictions, I use different identification strategies. I have obtained $\hat{\phi}$ using a regression of the log of the share of the built-up area on the share of greenbelt land (see equation (32)). The reduction in land used for development due to greenbelts is very large: column (1) suggests that a 10-percentage-point increase in greenbelt land leads to a decrease in the share of built-up land of $e^{-1.160 \times 0.1} - 1 = 11\%$. This effect is similar in terms of magnitude when compared to the reduced-form results on dwelling density (see Section 3.1). The effects are stronger in areas that are in 1973 greenbelts (column (2)). If I keep only the areas within 1km of an inner or outer greenbelt boundary in column (3), the reduction in the share of

TABLE 5 – STRUCTURAL PARAMETERS

| | Baseline | | | Historic instrument | | | Spatial instrument | | |
|--|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | All areas | Greenbelts in 1973 | Greenbelt border <1km | All areas | Greenbelts in 1973 | Greenbelt border <1km | All areas | Greenbelts in 1973 | Greenbelt border <1km |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| <i>Commuting parameters:</i> | | | | | | | | | |
| Elasticity of commuting time by road, $\hat{\alpha}_{rd}$ | 0.0686*** (0.0001) | 0.0686*** (0.0001) | 0.0686*** (0.0001) | 0.0670*** (0.0001) | 0.0670*** (0.0001) | 0.0670*** (0.0001) | 0.0670*** (0.0001) | 0.0670*** (0.0001) | 0.0670*** (0.0001) |
| Elasticity of commuting time by rail, $\hat{\alpha}_{rw}$ | 0.0428*** (0.0001) | 0.0428*** (0.0001) | 0.0428*** (0.0001) | 0.0445*** (0.0001) | 0.0445*** (0.0001) | 0.0445*** (0.0001) | 0.0445*** (0.0001) | 0.0445*** (0.0001) | 0.0445*** (0.0001) |
| Elasticity of commuting time by other mode, $\hat{\alpha}_{oth}$ | 0.0121*** (0.0000) | 0.0121*** (0.0000) | 0.0121*** (0.0000) | 0.0110*** (0.0007) | 0.0110*** (0.0007) | 0.0110*** (0.0007) | 0.0110*** (0.0007) | 0.0110*** (0.0007) | 0.0110*** (0.0007) |
| Commuting heterogeneity, $\hat{\epsilon}$ | 5.5274*** (0.0826) | 5.5274*** (0.0826) | 5.5274*** (0.0826) | 5.5274*** (0.0826) | 5.5274*** (0.0826) | 5.5274*** (0.0826) | 5.5274*** (0.0826) | 5.5274*** (0.0826) | 5.5274*** (0.0826) |
| <i>Greenbelt parameters:</i> | | | | | | | | | |
| Greenbelt restrictions, $\hat{\phi}$ | -1.1600*** (0.0371) | -2.2111*** (0.0433) | -1.7926*** (0.0477) | -1.1600*** (0.0371) | -2.2111*** (0.0433) | -1.7926*** (0.0477) | -1.1600*** (0.0371) | -2.2111*** (0.0433) | -1.7926*** (0.0477) |
| Share capital in construction costs, $\hat{\mu}$ | 0.7096*** (0.0149) | 0.5627*** (0.0816) | 0.7018*** (0.1073) | 0.7096*** (0.0149) | 0.5627*** (0.0816) | 0.7018*** (0.1073) | 0.7096*** (0.0149) | 0.5627*** (0.0816) | 0.7018*** (0.1073) |
| Greenbelt amenity effect, $\hat{\zeta}$ | 0.0672*** (0.0046) | 0.1530*** (0.0170) | 0.1384*** (0.0185) | 0.0830*** (0.0048) | 0.1547*** (0.0224) | 0.1563*** (0.0214) | 0.0631*** (0.0106) | 0.1519*** (0.0220) | 0.1661*** (0.0207) |
| Greenbelt amenity decay, $\hat{\theta}$ | 0.6534*** (0.1125) | 0.8232*** (0.1500) | 0.5517*** (0.1319) | 0.5165*** (0.1211) | 0.7648*** (0.1565) | 0.5341*** (0.1344) | 0.1486 (0.1487) | 0.7475*** (0.1550) | 0.6014*** (0.1324) |
| <i>Spillover parameters:</i> | | | | | | | | | |
| Household density elasticity, $\hat{\psi}$ | -0.0435*** (0.0010) | -0.0181*** (0.0033) | -0.0367*** (0.0035) | -0.0667*** (0.0021) | -0.0183*** (0.0057) | -0.0307*** (0.0126) | -0.0595*** (0.0068) | -0.0193*** (0.0056) | -0.0262*** (0.0067) |
| Agglomeration elasticity, $\hat{\gamma}$ | 0.0464*** (0.0035) | 0.0464*** (0.0035) | 0.0464*** (0.0035) | 0.0463*** (0.0032) | 0.0463*** (0.0032) | 0.0463*** (0.0032) | 0.0413*** (0.0024) | 0.0413*** (0.0024) | 0.0413*** (0.0024) |
| Agglomeration decay, $\hat{\delta}$ | 0.0306*** (0.0053) | 0.0306*** (0.0053) | 0.0306*** (0.0053) | 0.0302*** (0.0051) | 0.0302*** (0.0051) | 0.0302*** (0.0051) | 0.0221*** (0.0053) | 0.0221*** (0.0053) | 0.0221*** (0.0053) |
| Congestion elasticity, $\hat{\lambda}$ | 0.1006*** (0.0021) | 0.1006*** (0.0021) | 0.1006*** (0.0021) | 0.2235*** (0.0050) | 0.2235*** (0.0050) | 0.2235*** (0.0050) | 0.1370*** (0.0026) | 0.1370*** (0.0026) | 0.1370*** (0.0026) |
| Greenbelt-by-government region fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of areas | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 |
| Number of area pairs | 19,673,517 | 19,673,517 | 19,673,517 | 19,673,517 | 19,673,517 | 19,673,517 | 19,673,517 | 19,673,517 | 19,673,517 |

Notes: We estimate the parameters using data at the Mid-layer Super Output Area (MSOA). **Bold** indicates instrumented. To identify $\hat{\alpha}_{rd}$, $\hat{\alpha}_{rw}$, and $\hat{\alpha}_{oth}$, we use Euclidian distance as an instrument in columns (4)-(9). To identify $\hat{\mu}$, we use distance to the nearest city centre squared as an instrument in all columns. For $\hat{\lambda}$, $\hat{\psi}$, $\hat{\gamma}$ and $\hat{\delta}$, we use (the log of) population density in 1931 in columns (4)-(6), and the share of greenbelt land 10-25km in columns (7)-(9). Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

built-up land is 16.4% for a 10-percentage-point increase in greenbelt land.

I find in column (1) that the share of capital in construction costs $\hat{\mu} = 0.710$. This value is close to the value assumed in [Ahlfeldt et al. \(2015\)](#) and within the range provided by the literature, which is between 0.5 and 0.75. It is slightly lower, though, than the value of 0.80 estimated in [Combes et al. \(2016\)](#) for single-family houses in France. I believe it is reasonable that $\hat{\mu}$ is somewhat higher in England than in France, as England is known for its stringent land-use regulation and height restrictions ([Hilber and Vermeulen, 2016](#); [Dericks and Koster, 2021](#)). This implies that a higher share of the costs of construction will be paid to land.

To establish a valid estimate for μ , I use distance to the city centre as an instrument for floor space prices, which may be problematic if the distance to the city centre is correlated to supply conditions. In Appendix [C.2.1](#), I show that the effect is robust when I control for soil conditions and the share of construction workers in the area. Moreover, I relax the assumption of strict exogeneity of the distance to the city centre instrument in Appendix [C.2.2](#) so that part of the effect of distance to the city centre may relate to supply conditions. The results are robust.

I also test for heterogeneity in μ in Appendix [C.2.3](#) and investigate whether μ varies for different levels of greenbelt land. Only for a very high share of greenbelt land ($g_i > 0.95$) in one's own MSOA I do find that μ is somewhat lower. In line with this finding, when only including greenbelts in 1973 in column (2), the estimate is lower ($\hat{\mu} = 0.563$), albeit not significantly different from the previous estimate. In 1973 greenbelts, there are few large towns, while buildings are low and the share of greenbelt land is high. Hence, a higher share of construction costs will be devoted to land. However, one may therefore be concerned that μ is not representative of the whole of England. When only including areas within 1km of a greenbelt border in columns (3), I find an estimate of $\hat{\mu}$ that is again close to 0.70.

I find a statistically significant residential amenity effect, denoted by $\hat{\zeta}$, which is robust across specifications. The preferred estimate in column (3) suggests that amenities Ψ_i increase by 1.4% when the share of greenbelt land increases by 10 percentage points. The decay parameter $\hat{\theta}$ indicates a relatively strong decay of greenbelt amenities. In Appendix [C.3](#), I show that most of the amenity effect attenuates within 1km, which is in line with the reduced-form results. Although the effect attenuates quickly over distance, the estimates still imply that the majority

of England's residents benefit somewhat from greenbelt amenities.

5.1.3 Spillover parameters

I find that residential density has a negative effect. Doubling residential density implies a reduction in amenities, Ψ_i , of approximately 1.5-4.5%. These estimates are very similar if I use population density in 1931 as an instrument in columns (4)-(6) or use greenbelt land far away as an instrument in columns (7)-(9). I believe the negative coefficient is intuitive as residents may dislike living in higher densities, *e.g.* because people do not prefer to live in apartments (Glaeser et al., 2008). Moreover, negative externalities, such as air pollution and crime, may be higher in denser areas.

The estimates of agglomeration economies are in line with expectations. I find an agglomeration elasticity $\hat{\gamma}$ of 0.046. This is very close to the mean estimate (0.058) provided by Melo et al. (2009). When instrumenting for agglomeration economies in columns (4)-(9), the estimates do not materially change. The decay parameter $\hat{\delta}$ is 0.031, which implies that around 60% of the productive effect of density disappears after a 30-minute drive (see Appendix C.3). After a one-hour drive, 5% remains. The decay is weak compared to Arzaghi and Henderson (2008) and Ahlfeldt et al. (2015). This makes sense as these studies identify productivity externalities *within* cities, while this study takes into account the whole of England. Hence, one would expect the decay to be less strong because, for example, input-output linkages are usually less important over short distances within cities. Indeed, the estimate is comparably close to Rosenthal and Strange (2008), who also use between-city variation in densities.³¹

I further estimate the parameter capturing congestion externalities, $\hat{\lambda}$. The parameter indicates that travel times increase by $e^{0.1006} - 1 = 10.6\%$ if traffic density on the road \mathcal{D}_M increases by one standard deviation. In columns (4)-(6), I show that this coefficient is stronger when addressing endogeneity concerns. This makes sense when infrastructure is constructed in places where traffic density is higher, as this would imply shorter travel times. Consequently, I would expect to find a (strong) underestimate if one does not instrument for traffic density. With the historic instrument being based on the population in 1931, a one standard deviation increase in \mathcal{D}_M implies an increase in congestion of 25%. When using the spatial instrument based on the

³¹If I fit the exponential decay function through Rosenthal and Strange's preferred estimates, I obtain decay parameters ranging from 0.027-0.064.

TABLE 6 – RESULTS OF COUNTERFACTUAL ANALYSES

| | Scenario 1: 10% reduction in greenbelts | Scenario 2: No greenbelts | Scenario 3: Counterfactual greenbelts |
|--|--|------------------------------|--|
| | (1) | (2) | (3) |
| PANEL A: Baseline | | | |
| Change in expected utility (<i>in %</i>) | -0.248 | -1.375 | 0.319 |
| Change in overall land rents (<i>in %</i>) | 0.166 | -2.932 | 4.137 |
| Change in output (<i>in %</i>) | -0.151 | 1.051 | -1.103 |
| Change in average floor space price (<i>in %</i>) | -2.380 | -7.407 | 6.839 |
| Change in average wages net of commuting (<i>in %</i>) | -0.117 | -0.138 | -0.142 |
| Change in total developable land (<i>in %</i>) | 8.804 | 54.480 | -5.517 |
| Change in total floor space (<i>in %</i>) | 1.559 | 6.871 | -3.274 |
| PANEL B: No greenbelt amenities, $\zeta = 0$ | | | |
| Change in expected utility (<i>in %</i>) | 0.743 | 2.747 | -0.714 |
| Change in overall land rents (<i>in %</i>) | 4.060 | -0.318 | 7.370 |
| Change in output (<i>in %</i>) | 1.696 | 2.782 | 0.509 |
| Change in average floor space price (<i>in %</i>) | -1.619 | -6.618 | 7.408 |
| Change in average wages net of commuting (<i>in %</i>) | -0.060 | 0.036 | -0.115 |
| Change in total developable land (<i>in %</i>) | 8.804 | 54.480 | -5.517 |
| Change in total floor space (<i>in %</i>) | 2.952 | 8.054 | -2.197 |

Notes: I take the parameter values estimated in column (6), Table 5.

share of greenbelt land 10-25km, the impact is somewhere in between the baseline estimate and the estimate with population density in 1931 as the instrument. I provide additional sensitivity checks on $\hat{\lambda}$ in Appendix C.4.

5.2 Counterfactual analyses and welfare — aggregate effects

Armed with the estimated structural parameters, I calculate counterfactual scenarios to investigate changes in the provision of greenbelt land for workers and absentee landlords. I consider the coefficients reported in column (6) of Table 5 to be the preferred estimates. I will show that the results are similar when using parameters based on other specifications. The procedure to obtain the counterfactual outcomes is described in Appendix D.1.

I consider 3 experiments: I first determine inner greenbelt boundaries and shift the boundary approximately 800m outwards so that the total amount of greenbelt land is reduced by 10%. In the second experiment, I remove all greenbelt land. In the third experiment, I increase the size of greenbelts by constructing so-called counterfactual greenbelts around all major cities in England.

5.2.1 Counterfactual scenario 1

Let us first concentrate on a 10% reduction in greenbelt land. In Panel A, I show that the equivalent income decrease is 0.25%. If one multiplies this with the median gross earnings and the working population, this amounts to approximately a loss of £1.6 billion per year. There are two effects counteracting each other here. First, a reduction in greenbelt land implies a loss in the amenity value and therefore a loss in utility. At the same time, due to the easing of restrictions on greenbelt land, floor space availability increases by 1.6%. This, in turn, leads to lower floor space prices (−2.4%) and a higher utility. The fact that the overall utility of residents decreases implies that the amenity effect of greenbelt land dominates the supply effect, albeit marginally.

Note that the sign of the effect on aggregate land rents is a priori indeterminate. On the one hand, a reduction in greenbelt land reduces the price per unit of floor space, leading to lower land rents. But, on the other hand, more land becomes available for development. Because of lower prices, residents consume more floor space, which then increases land rents. The latter effect appears to dominate as total land rents increase, albeit with a very small amount (the increase is 0.17%). Given a discount rate of 1.8% (see [Bracke et al., 2018](#)), a total amount of developed land of 2km², and a median land price of £461 per m², the gain in land prices is equal to only £27 million per year.³² Hence, the losses for workers are considerably larger compared to the gains to absentee landlords. On top of that, homeowners only benefit slightly from higher land prices as only 5% of the land is owned by homeowners; most of the benefits of greenbelts accrue to other parties owning the land. [Shrubsole \(2019\)](#) shows that approximately 30% of England's land is owned by the aristocracy (which may even be an underestimate), 18% by corporations, 17% by oligarchs and bankers, 17% is unaccounted for, and 8.5% is owned by the public sector.

Finally, note that the sign of the effect of a reduction in greenbelt land on aggregate output is also a priori unclear. On the one hand, travel times become longer because development becomes less concentrated. Also, with lower densities, firms will benefit less from agglomeration economies. Both of these effects imply a reduction in output. On the other hand, floor space

³²I obtain data on estimated land prices for 326 local authorities from the *Department for Communities and Local Government* from 2014. I deflate land prices to 2011 values using the consumer price index. The estimated total land revenues at i are then given by $p_i F_i - r K_i$.

prices drop, which in turn may increase output. I find that the output decreases slightly by 0.15% so the latter effect does not dominate.

5.2.2 Counterfactual scenario 2

In the second counterfactual, where I consider removing all greenbelt land, the effects are somewhat amplified. The equivalent income increase of greenbelt policy is 1.4%. The total costs for workers if greenbelts are removed are roughly £9 billion per year. The overall land rents thus decrease by 2.9%, meaning the reduction in floor space prices dominates the increased availability of developed land. The reduction in land prices is therefore equivalent to a total loss in land prices of around £470 million per year.

Note further that the reduction in overall floor space prices is sizeable (7.4%). This number seems realistic as [Hilber and Vermeulen \(2016\)](#) provide back-of-the-envelope calculations showing that house prices would be reduced by 35% when *all* regulatory constraints are lifted.³³

What are the exact sources of the welfare benefits of greenbelts? In Panel B in Table 6, I turn off the amenity effect by restricting $\zeta = 0$. Thus, removing greenbelts implies an equivalent income *increase* to workers of 2.7%. Land revenues are now hardly affected, so without greenbelt amenities, removing greenbelts would generate positive welfare effects because of overall cheaper housing supply. I further test the sensitivity of the results with respect to the importance of greenbelt amenities in Figure 2a. I show that the change in land rents monotonically increases in ζ , while workers lose from the removal of greenbelts when $\zeta > 0.1$. In other words, when $\zeta > 0.1$, the amenity effect exceeds the supply effect of greenbelts.

Whether greenbelts increase welfare also depends critically on the value of μ . I find in Figure 2b positive utility effects of greenbelts for workers when μ exceeds 0.55. With a high μ , it is economical to build in higher densities in existing urban areas. Hence, the removal of greenbelts mostly leads to a loss in amenities, while the supply effect is limited. On the other hand, with a low μ , buildings are land-intensive and the removal of greenbelts will mean that the reduction in floor space prices due to increased supply is more substantial. With respect to land rents, a high μ means that the benefits of land are larger because building construction is capital intensive. In sum, when μ is higher, it is easier to build taller, and therefore revenues per unit of

³³Of course, greenbelt policy is an important example of a supply constraint, but it is definitely not the only one.

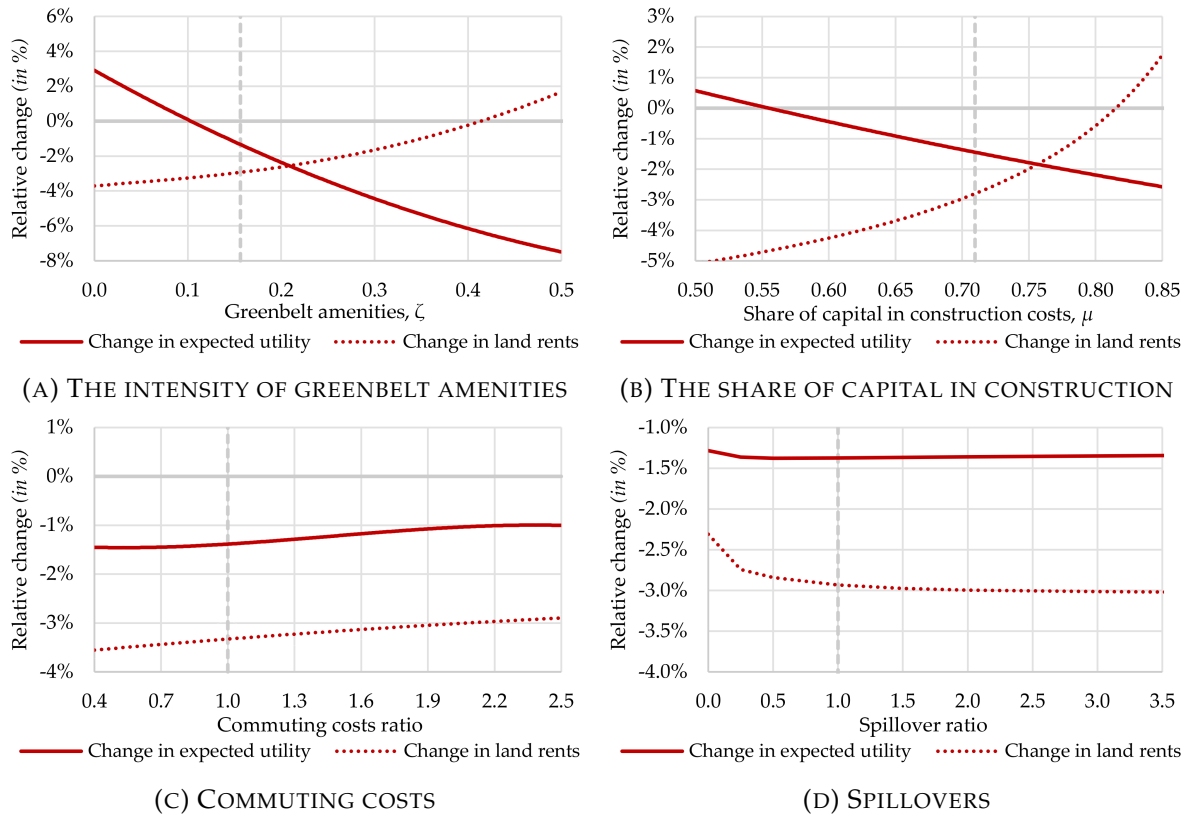


FIGURE 2 – NO GREENBELT SCENARIO FOR DIFFERENT PARAMETER VALUES

Notes: I estimate the effects using parameters estimated in column (6) in Table 5. The dashed lines indicate the baseline parameters. I consider the scenario where all greenbelts are removed. In Figure 2c, commuting cost parameters $\{\alpha_{rd}, \alpha_{rd}, \alpha_{rd}\}$ are multiplied by the commuting cost ratio. In Figure 2d, I multiply spillover parameters $\{\hat{\psi}, \hat{\gamma}, \hat{\lambda}\}$ with the spillover ratio.

land increase.

Figure 2c shows that the size of commuting costs does not have much of an influence on the aggregate welfare effects of greenbelts. I have studied the importance of spillovers in more detail in Figure 2d. I then show the welfare effects of removing greenbelts for lower and higher parameters capturing spillovers $\{\hat{\psi}, \hat{\gamma}, \hat{\lambda}\}$ than originally estimated. The welfare effects when greenbelts are removed for different values of spillovers are hardly affected by the size of the spillover parameters. Kline and Moretti (2014) show that agglomeration economies are a localised market failure that, given their stylised model, cancels out in the aggregate when the agglomeration elasticity is constant across space. Hence, agglomeration economies and congestion do not *exactly* cancel out in my model, which explains why there is a tiny change in welfare if spillovers are absent.

All in all, whether greenbelts generate positive welfare effects depends predominantly on the interplay between the intensity of greenbelt amenities, ζ , and how easy it is to build, μ . I observe

that with strong greenbelt amenities and a reasonably high share of capital in construction costs, greenbelts are welfare improving.

5.2.3 Counterfactual scenario 3

Let us now turn back to Table 6. The third counterfactual scenario describes what would happen if all major cities in England had greenbelts, which can be seen as a push towards more environmentally-friendly urban planning. I explain the procedure to determine these so-called counterfactual greenbelts in Appendix D.2. I observe in Table 6 that the equivalent income change is approximately 0.3%, so the overall effect is small. Hence, it is unlikely that adding more greenbelt land will lead to substantial positive welfare effects. Total land revenues increase by approximately 4.3% because land becomes more scarce and therefore much more expensive. Housing affordability thus strongly reduces as prices rise by 6.8%

I have explored the issue of increasing greenbelt land further by calculating utility and land rents for different scenarios. Using the counterfactual cities as shown in Appendix D.2, I consecutively increase the size of the greenbelt ‘doughnut’ around each city. Figure 3 shows that expected utility is monotonically increasing in the share of greenbelt land in England, although the utility gains are essentially flat beyond a share of greenbelt land that equals 30%. Land rents are also monotonically increasing in the total amount of greenbelt land. I caution, however, that for very high shares of greenbelt land, non-linear effects not taken into account in the current model may become relevant. For example, accommodating all workers and residents in currently existing cities may become prohibitively expensive.

In Appendix D.6, I provide a robustness analysis for these findings. First, I use the parameters identified when using all MSOAs (see column (1) in Table 5) or when using locations in 1973 greenbelts, which leads to similar results. Second, because one may be concerned about the identification strategies used to identify spillovers, I also estimate counterfactuals where I borrow parameters from the literature, leaving the results essentially unaffected.

5.3 Counterfactual analyses and welfare — local effects

Although aggregate effects may seem modest, local effects are large in relative terms. I plot the local effects of Scenario 2 (no greenbelts) in Figure 4.

In some locations, floor space prices drop substantially (Figure 4a). In former greenbelt areas,

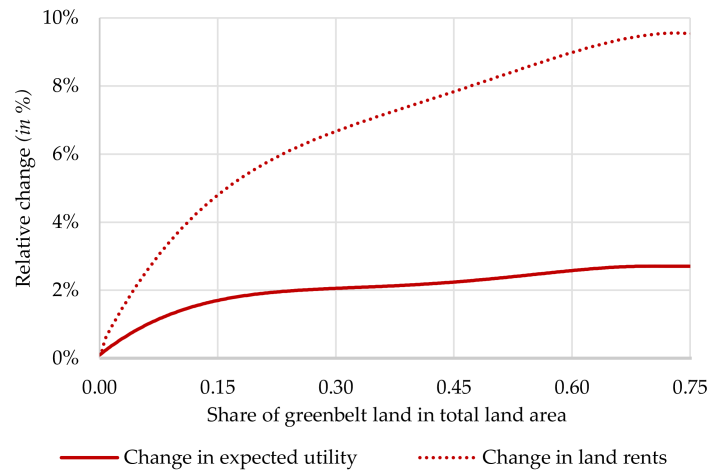


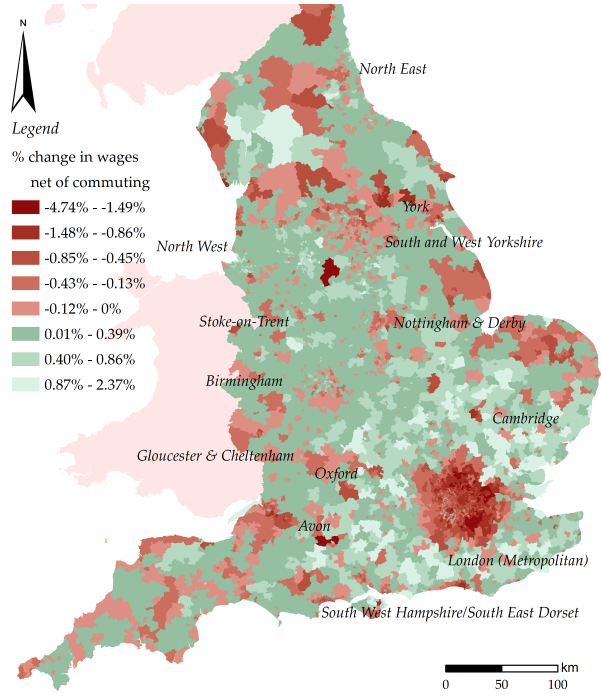
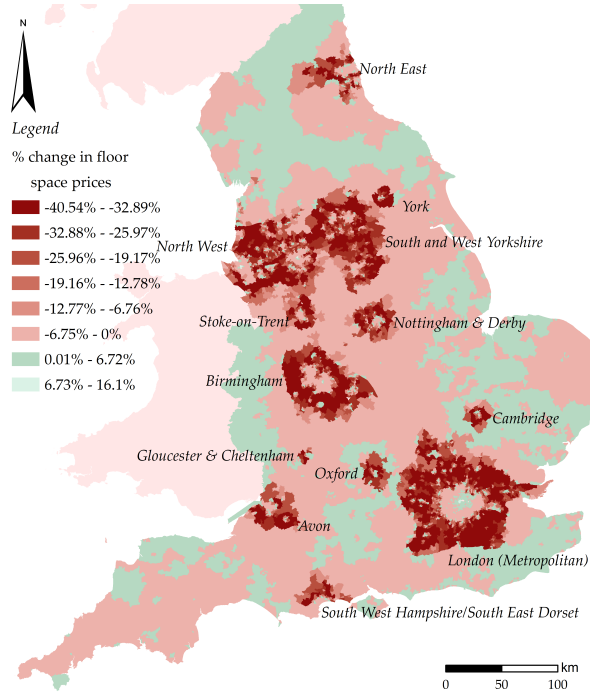
FIGURE 3 – INCREASING GREENBELT LAND

Notes: I estimate the effects using parameters estimated in column (6) in Table 5. Here, I calculate changes with respect to a situation where all greenbelts are removed.

this price drop can be as high as 40%. A key reason why prices drop in greenbelts is the loss in amenities. Outside greenbelts prices also drop substantially due to an increased supply of floor space. For example, in cities like London, Manchester, and Birmingham, prices drop by 5-20%. Given that England's cities are known to be expensive (Hilber and Vermeulen, 2016), removing greenbelt land seems to imply a substantial improvement in housing affordability. Wage effects are considerably smaller on average (see Figure 4b). Wages decrease in Greater London because of longer commutes, but the effects are small.

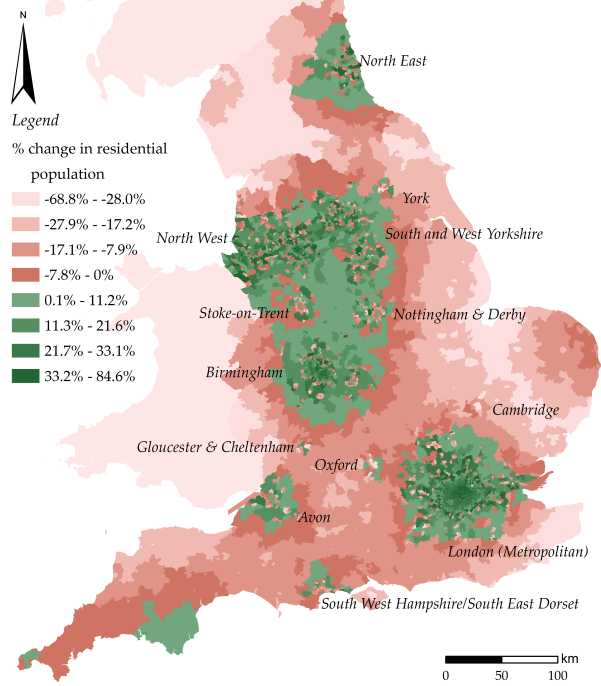
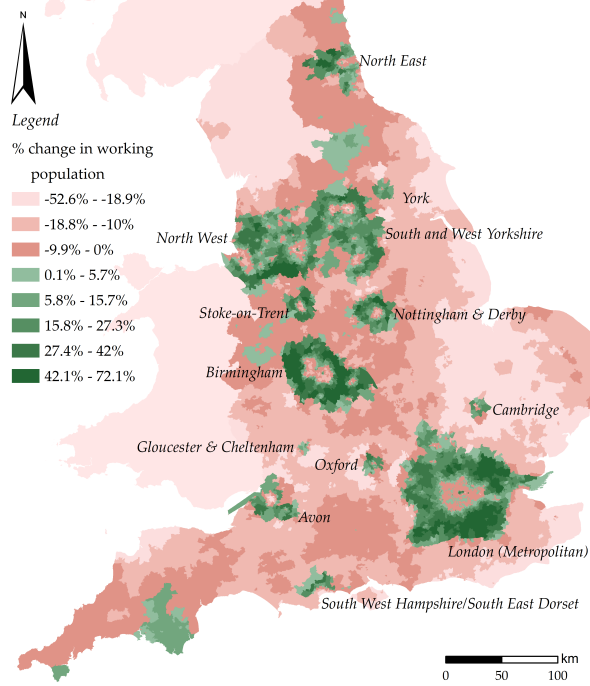
In Figure 4c I show that employment moves from city centres to former greenbelt areas. The relative increases in employment range from 5-70%. This may seem large, but note that these are relative differences. Because there were hardly any jobs in many former greenbelt areas in the first place, the relative increase in jobs is large. However, absolute changes are smaller. In Appendix D.3, it is shown that the absolute changes are relevant but are too small to fundamentally change England's urban structure.

For the residential population, the picture is somewhat similar (see Figure 4d). Because cities and their surrounding areas are now considerably cheaper, residents prefer to live in and around cities on former greenbelt land. At the same time, residents move into the inner cities of London and Birmingham. There are two reasons why people are doing this. First, given that production moves partly to former greenbelt areas, prices in city centres tend to be lower. This makes it attractive for people to move into city centres (which in turn increases floor space



(A) FLOOR SPACE PRICES, p_i

(B) WAGE NET OF COMMUTING COSTS, $E[\tilde{w}_i]$



(C) WORKERS, H_{Mi}

(D) RESIDENTIAL POPULATION, H_{Ri}

FIGURE 4 – COUNTERFACTUAL 2: NO GREENBELTS

Notes: I estimate the effects using parameters estimated in column (6) in Table 5. I consider the scenario where all greenbelts are removed.

prices again). Second, areas close to former greenbelt areas lose greenbelt amenities, which in relative terms makes areas of central cities more attractive. One may be concerned that it is very unlikely that city centres will accommodate many new residents because other restrictions, such as height restrictions, floor-area restrictions, and restrictions on new buildings in historic districts, may apply in central cities. However, when I plot the change in demand for floor space, it appears that the total amount of floor space is roughly constant in inner cities, so it is in actuality the space that lies empty due to firms moving to former greenbelt areas that is occupied by residents.

In Appendix D.6, I also consider the local effects of the other scenarios (*i.e.* the 10% reduction in greenbelts and the counterfactual greenbelts). The baseline mechanisms underlying the predictions discussed above hold up.

In Appendix D.6, I further calculate correlations between predicted percentage effects of specifications with different parameters. I show that the bivariate correlations with predicted local effects are mostly above 0.95; hence, the results are very similar. If I set amenities to zero, I find that correlations with the baseline counterfactual values are somewhat lower. This is not surprising given that greenbelt amenities play a pivotal role in explaining the local effects. Due to the fact that in the absence of greenbelt amenities, locations do not become less attractive when greenbelts are removed, I find that people move particularly to former greenbelt areas where there is a greater availability of cheap floor space. By contrast, the bivariate correlations with the predicted effects if I set spillovers to zero are very high, confirming that the main results can be explained mostly by the presence of greenbelt amenities and the share of capital in construction.

Interestingly, although commuting costs do not materially influence the aggregate effects, they are important for changes in the residential population. With low commuting costs, the effects of removing greenbelts are far more spread out, as residents can now commute long distances without incurring high costs. Given the current level of commuting costs, commuter towns like Milton Keynes, Reading, and Ashford are all relatively far away.³⁴ For lower commuting costs, commuter towns are more relevant but become partly obsolete when removing greenbelts

³⁴For example, to travel from Milton Keynes, Reading, and Ashford to the City of London takes, respectively, 54 minutes, 56 minutes, and 63 minutes.

because residents do not have to ‘leapfrog’ greenbelts anymore. This, in turn, will considerably shorten commuting trips. With high commuting costs, the changes in residential population are much more contained to the areas that were formerly designated as greenbelt land.

6 Conclusions

In this paper, I have investigated the economic effects of greenbelt policy that prohibit new construction beyond a predefined boundary. I focus on England, where 13% of the land area is designated greenbelt land. Greenbelts were constituted between the 1950s and 1970s – a time when cities were much smaller – and have hardly changed ever since.

Using reduced-form regressions, I first establish that the restrictions implied by greenbelts are binding: within greenbelts the density is more than 50% lower. Greenbelts also imply positive amenity effects as measured by house prices. These effects are relatively local and are mostly relevant within 500m. Still, given that approximately 27% of all dwellings in England lie within 500m of greenbelt land, many properties are affected by greenbelt amenities.

I proceed by setting up a general equilibrium model that takes into account the effects of greenbelts on amenity values and housing supply. Moreover, I take into account commuting, mode choice, congestion on the road, and agglomeration economies, all of which I show to be empirically relevant.

A counterfactual analysis then shows that greenbelt policy *increase* the utility of workers: when removing all greenbelts, the equivalent income decrease caused by greenbelts is roughly 1.5%, which amounts to £10 billion a year. On the other hand, monetary gains accrue to absentee landowners as total land revenues are around 1.3% higher. This amounts to £650 million per year. When greenbelts are removed, property prices inside greenbelts drop by 20% or more, but outside greenbelts prices also drop substantially due to an increased supply of floor space. For example, in cities like London, Manchester, and Birmingham, prices drop by approximately 5-20%. Given that England’s cities are known to be expensive, removing greenbelt land seems to imply a substantial improvement in housing affordability.

Hence, greenbelts seem to imply overall welfare gains but inevitably they worsen housing affordability. The finding of positive welfare effects critically depends on the strength of

greenbelt amenities, which may depend on the local context (such as the quality of provided open space and cultural traits). Only when greenbelt amenities are strong enough, greenbelts imply benefits to residents. Interestingly, agglomeration economies do not seem to play an important role, which is in line with [Kline and Moretti \(2014\)](#), who argue that agglomeration economies are a localised market failure that cancels out in the aggregate.

References

- Adler, M., Liberini, F., Russo, A. and Van Ommeren, J. N. (2021), 'The Congestion Relief Benefit of Public Transit: Evidence from Rome', *Journal of Economic Geography* **21**(3), 397–431.
- Ahlfeldt, G., Redding, S., Sturm, D. and Wolf, N. (2015), 'The Economics of Density: Evidence from the Berlin Wall', *Econometrica* **83**(6), 1217–2189.
- Alonso, W. (1964), *Location and Land Use. Toward a General Theory of Land Rent*, Harvard University Press, Cambridge, MA.
- Altonji, J. G., Elder, T. E. and Taber, C. R. (2005), 'Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools', *Journal of Political Economy* **113**(1), 151–184.
- Anas, A. and Rhee, H. J. (2007), 'When are Urban Growth Boundaries not Second-best Policies to Congestion Tolls?', *Journal of Urban Economics* **61**(2), 263–286.
- Anderson, S. and West, S. (2006), 'Open Space, Residential Property Values, and Spatial Context', *Regional Science and Urban Economics* **36**(6), 773–789.
- Arnott, R. (1979), 'Unpriced Transport Congestion', *Journal of Economic Theory* **21**, 294–316.
- Arnott, R. and Rowse, J. (2013), 'Curbside parking time limits', *Transportation Research Part A: Policy and Practice* **55**, 89–110.
- Arzaghi, M. and Henderson, J. (2008), 'Networking off Madison Avenue', *Review of Economic Studies* **75**(4), 1011–1038.
- Barr, J., Tassier, T. and Trendafilov, R. (2011), 'Depth to Bedrock and the Formation of the Manhattan Skyline, 1890-1915.', *Journal of Economic History* **70**(4), 1060–1077.
- Batini, N., Jackson, B. and Nickell, S. (2000), 'Inflation Dynamics and the Labour Share in the UK', *External MPC Unit Discussion Paper*.
- Bayer, P., Ferreira, F. and McMillan, R. (2007), 'A Unified Framework for Measuring Preferences for Schools and Neighborhoods', *Journal of Political Economy* **115**(4), 588–638.
- Bayer, P. and Timmins, C. (2007), 'Estimating Equilibrium Models of Sorting Across Locations', *The Economic Journal* **117**, 353–374.
- Blundell, R. and Powell, J. (2003), Endogeneity in Nonparametric and Semiparametric Regression Models, in M. Dewatripont, L. Hansen and S. Turnovsky, eds, 'Advances in Economics and Econometrics:

- Theory and Applications', Cambridge University Press, Cambridge.
- Bolitzer, B. and Netusil, N. (2000), 'The Impact of Open Spaces on Property Values in Portland, Oregon', *Journal of Environmental Management* **59**, 185–193.
- Bontemps, C., Simioni, M. and Surry, Y. (2008), 'Semiparametric Hedonic Price Models: Assessing the Effects of Agricultural Nonpoint Source Pollution', *Journal of Applied Econometrics* **23**(6), 825–842.
- Bracke, P., Pinchbeck, E. and Wyatt, J. (2018), 'The Time Value of Housing: Historical Evidence on Discount Rates', *The Economic Journal* **128**(613), 1820–1843.
- Brander, L. and Koetse, M. (2011), 'The Value of Urban Open Space: Meta-analyses of Contingent Valuation and Hedonic Pricing Results', *Journal of Environmental Management* **92**(10), 2763–2773.
- Brueckner, J. (2001), 'Urban Sprawl: Lessons from Urban Economics', *Brookings-Wharton Papers on Urban Affairs* **2001**(1), 65–97.
- Brueckner, J. K. (2007), 'Urban Growth Boundaries: An Effective Second-best Remedy for Unpriced Traffic Congestion?', *Journal of Housing Economics* **16**(3-4), 263–273.
- Cameron, A. and Trivedi, P. (2005), *Microeconometrics: Methods and Application*, Cambridge University Press, New York.
- Cartmell, J. (2016), Study on the Prices and Quality of Rail Passenger Services, Technical report, European Commission, Brussels.
- Cheshire, P. (2014), 'Turning Houses Into Gold', *CentrePiece Spring*, 14–18.
- Cheshire, P., Hilber, C. and Koster, H. (2018), 'Empty Homes and Longer Commutes: The Unintended Consequences of More Restrictive Planning', *Journal of Public Economics* **158**(2), 126–151.
- Cheshire, P. and Sheppard, S. (2002), 'The Welfare Economics of Land Use Planning', *Journal of Urban Economics* **52**(2), 242–269.
- Ciccone, A. (2002), 'Agglomeration Effects in Europe', *European Economic Review* **46**(2), 213–227.
- Ciccone, A. and Hall, R. (1996), 'Productivity and the Density of Economic Activity', *The American Economic Review* **86**(1), 54–70.
- Combes, P., Duranton, G. and Gobillon, L. (2008), 'Spatial Wage Disparities: Sorting Matters!', *Journal of Urban Economics* **63**(2), 723–742.
- Combes, P., Duranton, G. and Gobillon, L. (2016), 'The Production Function for Housing: Evidence from France', *GATE - Lyon Saint-Etienne, Working Paper* 1637.
- Combes, P. P., Duranton, G. and Gobillon, L. (2019), 'The Costs of Agglomeration: House and Land Prices in French Cities', *Review of Economic Studies* **86**, 1556–1589.
- Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012), 'Plausibly Exogenous', *Review of Economics and Statistics* **94**(1), 260–272.
- Correll, M., Lillydahl, H. and Singell, D. (1978), 'The Effects of Greenbelts on Residential Property Values: Some Findings on the Political Economy of Open Space', *Land Economics* **54**(2), 207–217.

- Daganzo, C. F., Gayah, V. V. and Gonzales, E. J. (2011), 'Macroscopic Relations of Urban Traffic Variables: Bifurcations, Multivaluedness and Instability.', *Transportation Research Part B: Methodological* **45**, 278–288.
- Department for Communities and Local Government (2013), English Housing Survey: Households 2011-12, Technical report.
- Department for Communities and Local Government (2018), English Housing Survey: Floor space in England – main report, Technical report.
- Dericks, G. and Koster, H. R. A. (2021), 'The Billion Pound Drop: the Blitz and Agglomeration Economies in London', *Journal of Economic Geography* (Forthcoming).
- Eaton, J. and Kortum, S. (2002), 'Technology, Geography, and Trade', *Econometrica* **70**(5), 1741–1779.
- Economist, T. (2017), 'Britain's Delusions About the Green Belt Cause Untold Misery', *The Economist*.
- Fosgerau, M. (2015), 'Congestion in the Bathtub', *Economics of Transportation* **4**(4), 241–255.
- Gaigné, C., Koster, H. R. A., Moizeau, F. and Thisse, J. (2021), 'Who Lives Where in the City? Amenities, Commuting and Income Sorting', *Journal of Urban Economics* **128**(103394).
- Garcia-López, M. A., Pasidis, I. N. and Viladecans-Marsal, E. (2021), 'Congestion in Highways when Tolls and Railroads Matter: Evidence from European cities', *Journal of Economic Geography* (Forthcoming).
- Geoghegan, J. (2002), 'The Value of Open Spaces in Residential Land Use', *Land Use Policy* **19**(1), 91–98.
- Geroliminis, N. and Daganzo, C. F. (2008), 'Existence of Urban-scale Macroscopic Fundamental Diagrams: Some Experimental Findings', *Transportation Research Part B: Methodological* **42**, 759–770.
- Glaeser, E., Gyourko, J. and Saks, R. (2005), 'Why have Housing Prices Gone up?', *American Economic Review: Papers and Proceedings* **95**(2), 329–333.
- Glaeser, E., Kahn, M. and Rappaport, J. (2008), 'Why do the Poor Live in Cities? The Role of Public Transportation', *Journal of Urban Economics* **63**(1), 1–24.
- Glaeser, E., Kolko, J. and Saiz, A. (2001), 'Consumer City', *Journal of Economic Geography* **1**(1), 27–50.
- Glaeser, E. and Ward, B. (2009), 'The Causes and Consequences of Land Use Regulation: Evidence from Greater Boston', *Journal of Urban Economics* **65**(3), 265–278.
- Green, R., Malpezzi, S. and Mayo, S. (2005), 'Metropolitan-Specific Estimates of the Price Elasticity of Supply of Housing, and their Sources', *American Economic Review* **95**(2), 334–339.
- Grimes, A. and Liang, Y. (2009), 'Spatial Determinants of Land Prices: Does Auckland's Metropolitan Urban Limit Have an Effect?', *Applied Spatial Analysis* **2**, 23–45.
- Harari, M. (2020), 'Cities in bad shape: Urban geometry in india', *American Economic Review* **110**, 2377–2411.
- Hilber, C. and Vermeulen, W. (2016), 'The Impact of Supply Constraints on House Prices in England', *Economic Journal* **126**(591), 358–405.
- Holl, A. (2019), 'Natural Geography and Patterns of Local Population Growth and Decline in Spain:

- 1960–2011', *Sustainability* **11**(18), 4979.
- Ihlanfeldt, K. (2007), 'The Effect of Land Use Regulation on Housing and Land Prices', *Journal of Urban Economics* **61**(3), 420–435.
- Irwin, E. (2002), 'The Effects of Open Space on Residential Property Values', *Land Economics* **78**(4), 465–480.
- Jun, M. J. (2006), 'The Effects of Portland's Urban Growth Boundary on Housing Prices', *Journal of the American Planning Association* **72**(2), 239–243.
- Kanemoto, Y. (1977), 'Cost-benefit Analysis and the Second-best Land Use for Transportation', *Journal of Urban Economics* **4**, 483–503.
- Kline, P. and Moretti, E. (2014), 'Local Economic Development, Agglomeration Economies, and the Big Push: 100 years of Evidence from the Tennessee Valley Authority.', *The Quarterly Journal of Economics* **129**(1), 275–331.
- Koster, H., Van Ommeren, J. and Rietveld, P. (2012), 'Bombs, Boundaries and Buildings: a Regression-discontinuity Approach to Measure Costs of Housing Supply Restrictions', *Regional Science and Urban Economics* **42**(4), 631–641.
- Mayer, C. and Somerville, C. (2000), 'Land use regulation and new construction', *Regional Science and Urban Economics* **30**(6), 639–662.
- Melo, P., Graham, D. and Noland, R. (2009), 'A meta-analysis of estimates of urban agglomeration economies', *Regional Science and Urban Economics* **39**(3), 332–342.
- Mills, E. (1967), 'An Aggregative Model of Resource Allocation in a Metropolitan Area', *American Economic Review* **57**(2), 197–210.
- Mills, E. S. and Lubuele, L. S. (1997), 'Inner Cities', *Journal of Economic Literature* **35**(2), 727–756.
- Montgomery, C. (2015), *Happy City*, Penguin Press, Vancouver.
- Muth, R. (1969), *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*, University of Chicago Press, Chicago.
- Ohmori, N. and Harata, N. (2009), 'How Different are Activities while Commuting by Train? A Case in Tokyo', *Tijdschrift voor Economische en Sociale Geografie (Journal of Economic and Human Geography)* **99**(5), 547–561.
- Oster, E. (2019), 'Unobservable Selection and Coefficient Stability: Theory and Evidence', *Journal of Business and Economic Statistics* **37**(2), 187–204.
- Pattey, E. and Qiu, G. (2012), 'Trends in primary particulate matter emissions from Canadian agriculture', *Journal of the Air & Waste Management Association* **62**(7), 737–747.
- Peer, S., Verhoef, E. T., Knockaert, J., Koster, P. R. and Tseng, Y. Y. (2015), 'Long-Run Versus Short-Run Perspectives On Consumer Scheduling: Evidence From A Revealed-Preference Experiment Among Peak-Hour Road Commuters', *International Economic Review* **56**(1), 303–323.
- Pigou, C. (1932), *The Economics of Welfare*, MacMillan, London.

- Pines, D. and Sadka, E. (1985), 'Zoning, First-best, Second-best and Third-best Criteria for Allocating Land to Roads', *Journal of Urban Economics* **17**(2), 167–183.
- Powe, N. A. and Willis, K. G. (2004), 'Mortality and Morbidity Benefits of Air Pollution (SO₂ and PM₁₀) Absorption Attributable to Woodland in Britain', *Journal of Environmental Management* **70**(2), 119–128.
- Proost, S. and Thisse, J. F. (2019), 'What Can be Learned from Spatial Economics', *Journal of Economic Literature* **57**(3), 575–643.
- Robinson, P. M. (1988), 'Root-N-consistent Semiparametric Regression', *Econometrica* **56**(4), 931–954.
- Rosenthal, S. S. and Strange, W. C. (2008), 'The Attenuation of Human Capital Spillovers', *Journal of Urban Economics* **64**(2), 373–389.
- Russo, A., Adler, M., Liberini, F. and Van Ommeren, J. N. (2021), 'Welfare Losses of Road Congestion: Evidence from Rome', *Regional Science and Urban Economics* **89**(103692).
- Shrubsole, G. (2019), *Who Owns England? How We Lost Our Green and Pleasant Land and How to Take it Back*, HarperCollins, London.
- Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.
- Small, K. A. and Verhoef, E. T. (2007), *The Economics of Urban Transportation*, Routledge, New York.
- Tsivanidis, N. (2020), 'Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio', *Working paper, UC Berkeley*.
- Turner, M., Haughwout, A. and Van der Klaauw, W. (2014), 'Land Use Regulation and Welfare', *Econometrica* **82**(4), 1341–1403.
- Underwood, R. (1961), Speed, Volume, and Density Relationship, in 'Quality and Theory of Traffic Flow. A Symposium, Yale University Bureau of Highway Traffic', pp. 141–188.
- Vickrey, W. (1969), 'Congestion Theory and Transport Investment', *American Economic Review* **59**(2), 251–260.
- Yang, J. and Jinxiang, Z. (2007), 'The Failure and Success of Greenbelt Program in Beijing', *Urban Forestry and Urban Greening* **6**(4), 287–296.

Appendix A. Data and descriptives

A.1 *An example of a greenbelt border*

In Figure A1, I show an enlarged version of greenbelt land in the north of the Metropolitan Greenbelt encompassing London. One may notice that greenbelt land is very patchy. Also, beyond London's suburbs one finds development preceding the designation of greenbelts. In the identification strategy employed in the reduced-form analyses, I employ a border-discontinuity design based on *inner* and *outer* greenbelt borders. I define these by looking at the border of development with greenbelt land surrounding *the most important cities* in the greenbelt land. Hence, inner greenbelt boundaries only surround the most important cities and not the small towns and villages that are fully surrounded by greenbelt land.

A.2 *Greenbelts in 1973*

In Figure A2, I show approved and proposed greenbelt land in 1973. The approved greenbelts surround London, Bristol, Leeds, Bradford, and Newcastle. Most of the approved and proposed greenbelts are very similar to the current ones, as can be seen in Figure 1. Although not officially approved until the early 1980s, stringent building regulations were typically already applied

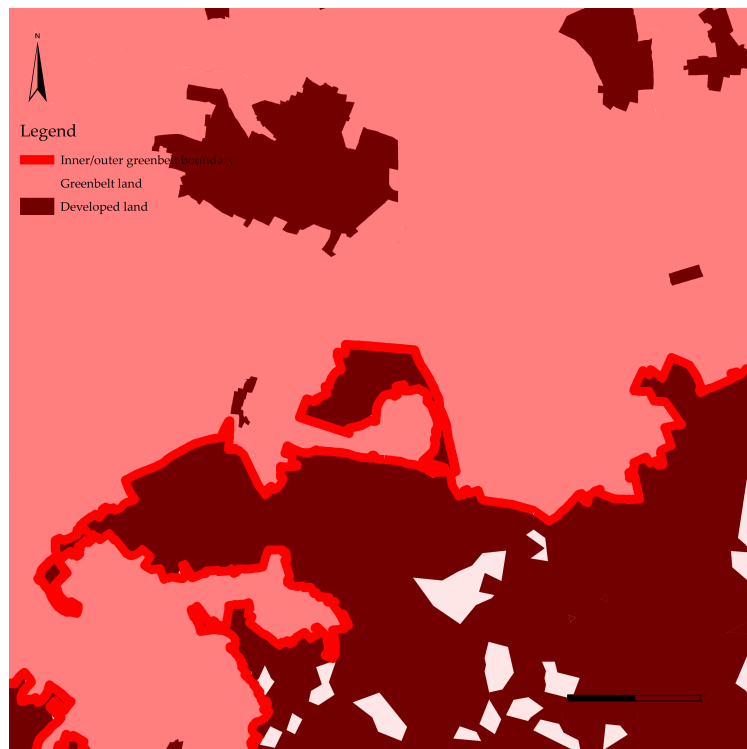


FIGURE A1 – GREENBELT BORDERS AND GREENBELT LAND: A CLOSER LOOK

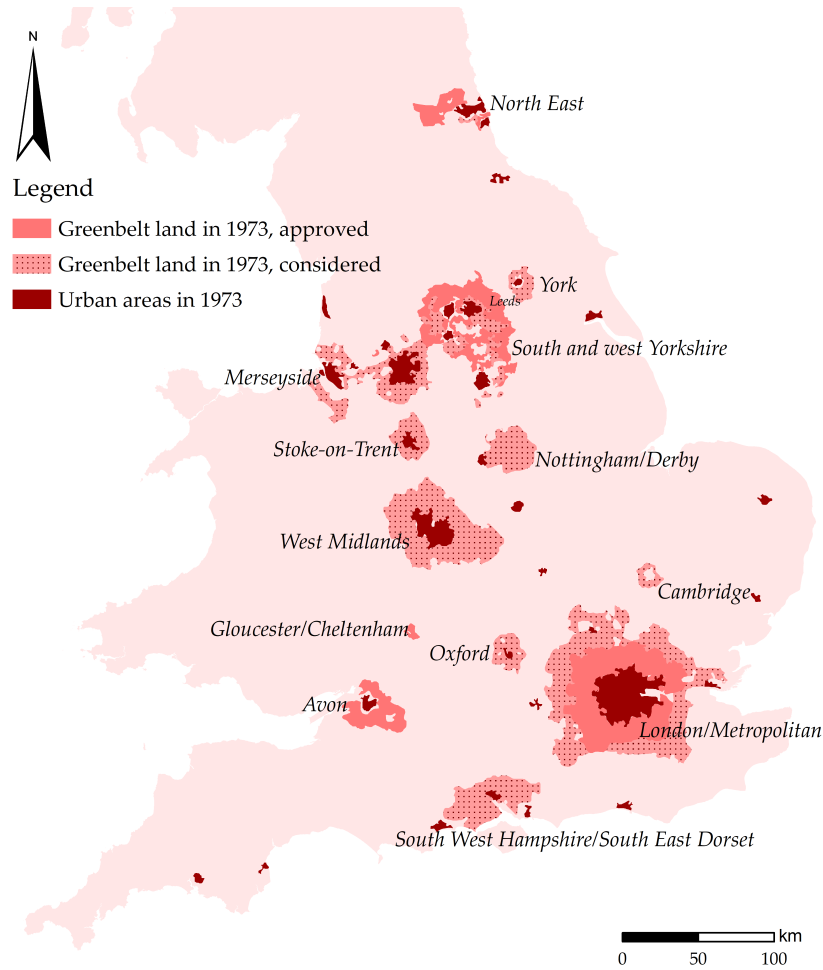


FIGURE A2 – CONSIDERED AND APPROVED GREENBELTS IN 1973

to the proposed greenbelt land. The correlation between the current share of greenbelt land and the share of approved and proposed greenbelt land in 1973 at the MSOA level is indeed reasonably high (0.622). The most notable exception of considered greenbelt land that has not been granted is a large area around Southampton. Part of that area is now the New Forest National Park.

A.3 Developed land

To estimate the impact of greenbelts on land, I use data from [Ordnance Survey \(OS\)](#) on exact polygons of all buildings in England. I plot these data in [Figure A3](#). Only 2% of the land is occupied by buildings, which is a lower bound of the total share of built-up land including roads, parking lots, etc. (around 10%). The share of developed land is considerably lower in greenbelts.

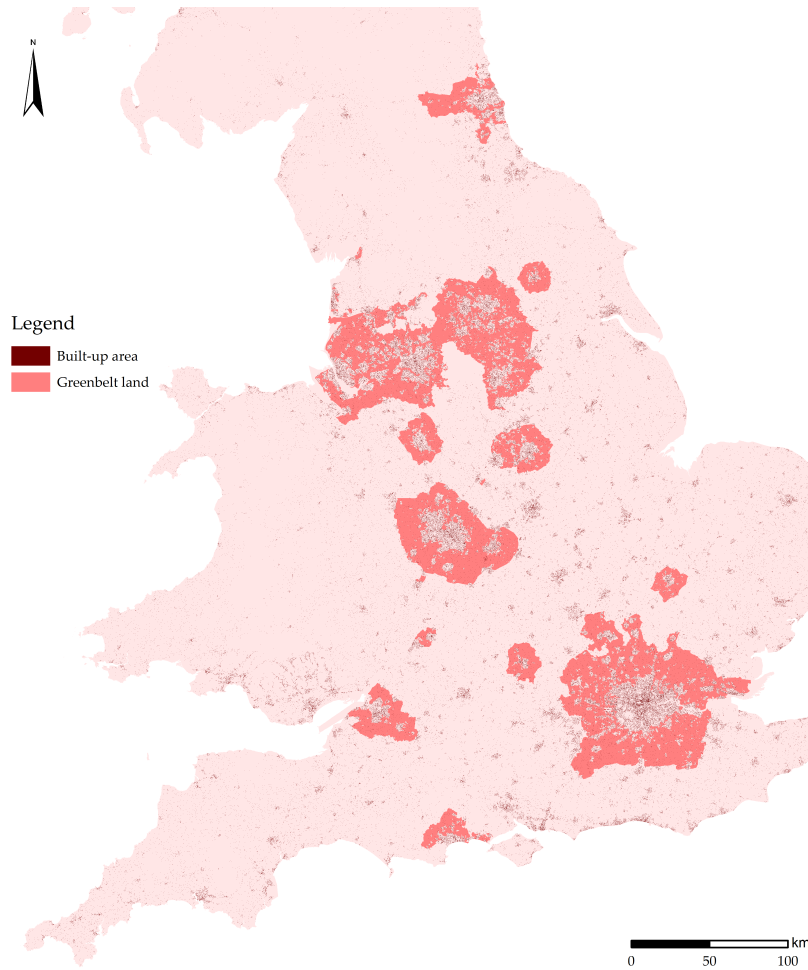


FIGURE A3 – BUILT-UP LAND IN GREAT BRITAIN

A.4 *Commuting flows and travel times*

Here I provide some additional information on the distributions of commuting flows from the 2011 [Census](#) per transport mode. First, I combine all commuting flows between all MSOAs and calculate the free-flow travel time. Approximately 10% of the flows are within MSOAs so that travel time is zero. One may further observe that flows above 90 minutes are rare. To be more specific, only 1.6% of the flows take longer than 90 minutes (one-way), while the average free-flow travel time is 15 minutes.

In the data, 66% of the flows are by road, 19% by rail, and 15% by other modes. For road travel I observe that roughly 7% of the flows occur within an MSOA. Around 4% of the flows take longer than 90 minutes, whereas for one-way commutes exceeding two hours this figure is 2.5%. The average commuting time is 30 minutes, which seems realistic.

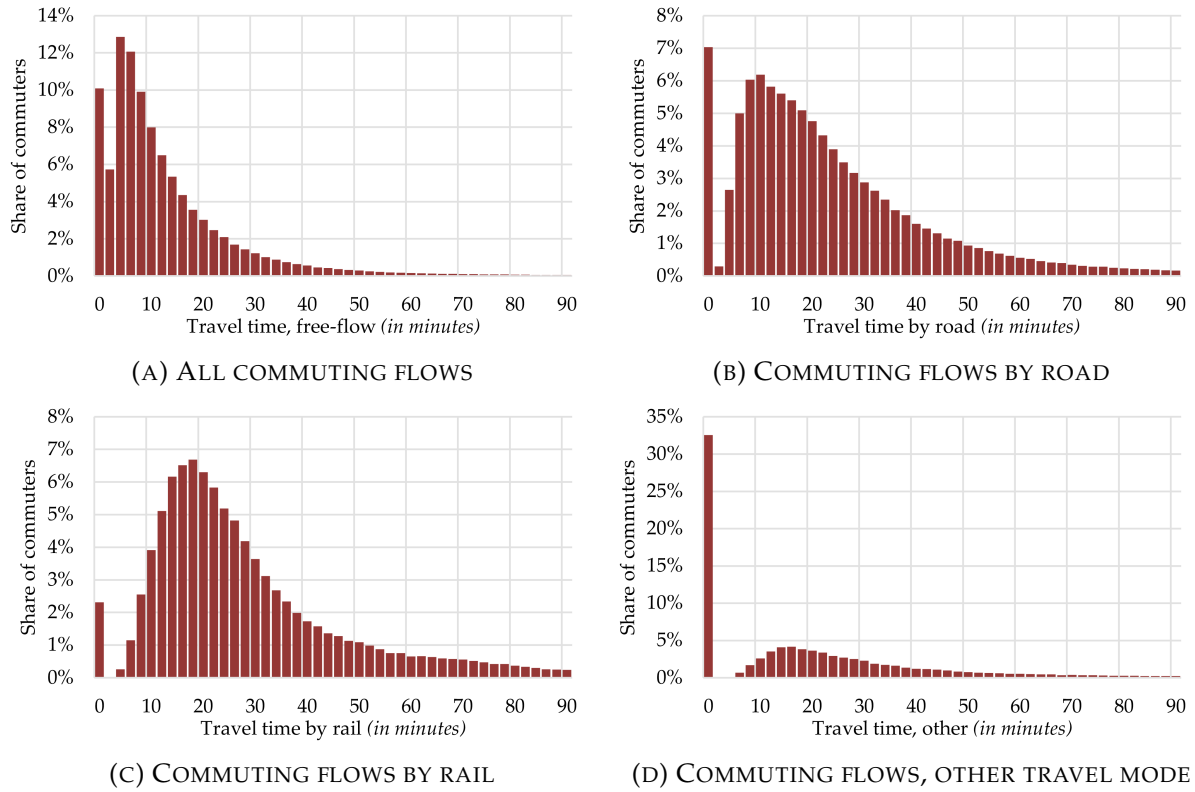


FIGURE A4 – HISTOGRAMS OF COMMUTING FLOWS

For travel time by train and metro, I observe that approximately 2.3% of the flows occur within an MSOA, which makes sense given that few people will take public transport for short-distance trips. 7.3% of the flows take longer than 90 minutes, whereas for one-way commutes exceeding two hours this figure is 4.7%. The average commuting time is 55 minutes, which also seems realistic.

Furthermore, I plot the histogram for other travel modes (Figure A4d). I calculate the travel time by assuming a speed of 10km over the road network. As ‘other modes’ include walking and cycling, 33% of the people commute within their own MSOA. Still, because people may use other modes of travel (such as working from home or taking a scooter), the average travel time is 58 minutes.

Note that, when estimating the gravity models, I exclude flows between locations that have a *free-flow* travel time exceeding 120 minutes. This implies that I exclude only 1% of commuters by road, 1.1% of commuters by rail, and 1.2% of commuters by other modes.

In addition, I make some 100+ observations of travel times between randomly chosen MSOA-

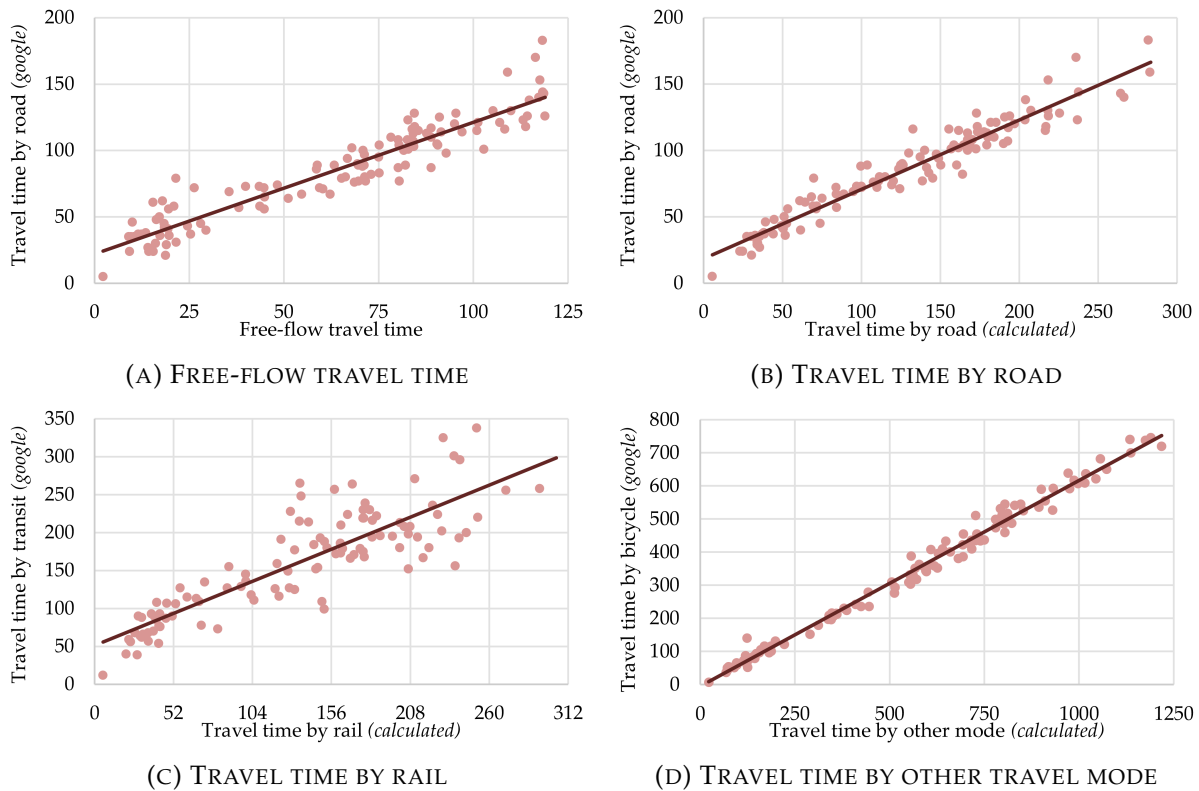


FIGURE A5 – CALCULATED AND ACTUAL TRAVEL TIMES

Notes: The travel times are in minutes. Actual travel times are from Google and gathered on a typical Wednesday around noon.

pairs that are within 120 minutes' drive (free-flow) of each other. Actual travel times are obtained from Google on a typical Wednesday around noon. Despite the seemingly low number of observations, the picture shown in Figure A5 is very clear: there is a very high correlation between calculated travel times and actual travel times.

More specifically, in Figure A5a, I show the correlation between the calculated free-flow travel time and actual travel time by road. The correlation is 0.940. However, the correlation is higher (0.965) if I consider the travel time by road using speeds provided by the Department of Transport. Especially for long trips, my calculated travel time seems to be somewhat higher, but for shorter trips they are very close to each other. Fortunately, most commuting trips are relatively short.³⁵

In Figure A5c, I show the correlation between calculated travel time by rail and actual travel time by public transport. It is shown that the correlation is somewhat lower (0.855), albeit still very high. The reason why the correlation is somewhat lower is that I do not sufficiently take

³⁵Note further that I obtained travel times at a time where little congestion is to be expected in the first place (*i.e.* around noon)

into account scheduling and interchanges in calculating the railway travel time.

Finally, the travel time by other travel modes and the travel time by bicycle are very highly correlated (0.994). The implied bicycle speed is slightly above 10km/h (which is the average speed assumed for other travel modes); hence, the calculated travel time is slightly above the travel time by bicycle.

Appendix B. Reduced-form effects

B.1 Housing characteristics and greenbelt land

First, I explore to what extent houses in and close to greenbelts are different. To this extent I use the data on *Energy Performance Certificates* before I merge the data to the housing transactions data from *Land Registry*, which produces almost 3 million more observations (around 28%). I then focus on three housing characteristics: a dummy indicating whether a property is in a flat, the house size, and the (predicted) energy consumption. I regress these housing characteristics on the share of greenbelt land and use the same identification strategies as before. I weigh the observations based on the inverse number of observations of the same property in the data. The results are reported in Table B1.

As expected, the share of flats is considerably lower in greenbelts. Column (1) suggests a reduction of 13 percentage points. Given a mean of 23%, this effect is considerable. In column (2), where I only include postcodes in greenbelts that have been proposed or assigned in 1973, I find a similar effect. However, when I only include postcodes within 1km of inner or outer greenbelt borders, the effect is just -5.1 percentage points.

In columns (4)-(6), I focus on the effects on house size. The coefficients in columns (4) and (5) imply that properties in greenbelts are around $e^{0.212} - 1 = 23.6\%$ larger, which is in line with the previous result showing that properties are less often (typically smaller) apartment units. Again, the effect of greenbelts on house size is considerably smaller when focusing on properties within 1km of greenbelt borders.

Finally, in columns (7)-(9) of Table B1, I test the impact on the predicted energy consumption of the property. I do not find economically meaningful effects (only in column (8) is the effect statistically significant).

All in all, given that housing characteristics may be different between properties on and outside greenbelt land, it is important to control for housing characteristics in the hedonic price regressions. Still, please recall that in column (2) of Table 3 I show that the effect of greenbelt land is hardly affected when including potentially relevant observable housing controls. To the extent one is still concerned that *unobservable* housing characteristics can explain part of the positive greenbelt effect, note that I also estimate specifications on *either side* of the inner or

TABLE B1 – HOUSING CHARACTERISTICS AND GREENBELTS

| | Dependent variable: property is in flat | | | Dependent variable: the log of house size | | | Dependent variable: the log of energy consumption | | |
|-------------------------------|---|------------------------|------------------------|---|-----------------------|-----------------------|---|-----------------------|-----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | WLS | WLS | WLS | WLS | WLS | WLS | WLS | WLS | WLS |
| | All postcodes | Greenbelts in 1973 | Greenbelt border <1km | All postcodes | Greenbelts in 1973 | Greenbelt border <1km | All postcodes | Greenbelts in 1973 | Greenbelt border <1km |
| Share greenbelt land | -0.1261*** (0.0046) | -0.1236*** (0.0062) | -0.0511*** (0.0062) | 0.2121*** (0.0046) | 0.2153*** (0.0059) | 0.1476*** (0.0060) | 0.0013 (0.0048) | 0.0183*** (0.0058) | 0.0088 (0.0086) |
| Distance to the city centre | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | No | No | Yes | No | No | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 12,938,943 | 2,391,690 | 2,707,959 | 12,938,943 | 2,391,690 | 2,707,959 | 12,935,543 | 2,391,150 | 2,707,542 |
| Bandwidth (km) | ∞ | ∞ | 1 | ∞ | ∞ | 1 | ∞ | ∞ | 1 |
| R ² | 0.1176 | 0.0619 | 0.0716 | 0.0489 | 0.0696 | 0.0512 | 0.0502 | 0.0532 | 0.0535 |

Notes: I weight the observations based on the inverse number of observations of the same property in the data. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Columns (2), (5) and (7) include observations in areas that are in greenbelts that were approved or considered in 1973. In columns (3), (6) and (9) I include transactions that are within 1km of a greenbelt boundary respectively. Standard errors are clustered at the MSOA level and in parentheses.

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

outer greenbelt borders (see columns (6) and (7) in Table 3).

B.2 Sensitivity analyses for reduced-form analyses

Here I report a wealth of additional sensitivity checks for the reduced-form analyses, whereby I focus on the effects of greenbelts on housing density and house prices. I report coefficients for each of the identification strategies: (i) include all observations, (ii) include only observations in greenbelts in 1973, and (iii) focus on the boundary-discontinuity design based on locations close to greenbelt borders.

B.2.1 Spatial extent of the greenbelt supply effect

I investigate whether greenbelt policy implies density effects beyond one's own postcode. It may, for example, be that overall city density is lower or cities surrounded by greenbelts may allow for denser construction in the interior of the city. I re-estimate equation (1) but now include the share of greenbelt land in 1000m distance bands. Column (1) in Table B2 shows that there is a strong drop in housing supply in one's own postcode, but effects beyond one's own postcode are small. For example, for the average share of greenbelt land between 1 and 2km, the rise in density is just $0.18 \times 0.0836 = 1.5\%$, which is negligible. Any statistically significant positive effects found in column (1) are not robust across specifications. If I keep areas in 1973 greenbelts, I again find a small effect between 1 and 2km, but this effect is considerably smaller when applying the boundary-discontinuity design in columns (3) and (4). Hence, the results in Table B2 provide evidence that greenbelts do not change the density beyond one's own postcode.

B.2.2 Omitted variable bias

I further aim to investigate whether omitted variable bias is important in measuring the amenity effect. It is then informative to use coefficient movements after the inclusion of control variables to investigate whether omitted variable bias is important. However, [Oster \(2019\)](#) argues that coefficient movements alone are not a sufficient statistic to calculate the bias but that the variance of the added control variables also matters. [Oster \(2019\)](#) then derives a GMM estimator to correct estimates for omitted variable bias under the assumption that the relationship between the variables of interest and unobservables can be recovered from the relationship between the variables of interest and the observable control variables. In the current context, this makes

TABLE B2 – DENSITY EFFECTS OF GREENBELTS
(Dependent variable: the number of dwellings in a postcode)

| | (1) Poisson | (2) Poisson | (3) Poisson | (4) Poisson |
|--------------------------------------|------------------------|------------------------|--------------------------|----------------------------|
| | All areas | Greenbelts in 1973 | Greenbelt border <5km | Greenbelt border <2.5km |
| Share greenbelt land | -0.7211*** (0.0132) | -1.1903*** (0.0233) | -0.9954*** (0.0163) | -1.0110*** (0.0171) |
| Share greenbelt land 0-1000m | 0.0157 (0.0180) | 0.0199 (0.0238) | -0.0088 (0.0211) | -0.0145 (0.0252) |
| Share greenbelt land 1000-2000m | 0.0836*** (0.0277) | 0.0634* (0.0381) | 0.0386 (0.0301) | 0.0413 (0.0323) |
| Share greenbelt land 2000-3000m | 0.0058 (0.0374) | -0.0603 (0.0558) | -0.0126 (0.0404) | 0.0449 (0.0443) |
| Share greenbelt land 3000-4000m | 0.0665 (0.0459) | 0.0727 (0.0651) | 0.1033** (0.0488) | 0.0841 (0.0548) |
| Share greenbelt land 4000-5000m | 0.0825** (0.0391) | -0.0043 (0.0582) | -0.0255 (0.0436) | -0.0018 (0.0485) |
| Area size of postcode (<i>log</i>) | 0.1124*** (0.0026) | 0.2732*** (0.0068) | 0.2147*** (0.0041) | 0.2256*** (0.0045) |
| Dwelling density | Yes | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 1,310,750 | 250,671 | 626,633 | 445,577 |
| Bandwidth (<i>km</i>) | ∞ | ∞ | 5 | 2.5 |
| R^2 | 0.0774 | 0.146 | 0.110 | 0.114 |

Notes: Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (2) includes observations in areas that are in greenbelts that were approved or considered in 1973. In columns (3) I include transactions that are within 2km of a greenbelt border, which is reduced to 1km in column (4). Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

sense because control variables that are added likely bear some relationship to unobservables. I add many housing controls to the regressions, which are likely to be correlated to unobservable housing characteristics that may change discontinuously between properties on greenbelt land and outside of greenbelts.

There are two key input parameters that must be determined. First, a parameter must be chosen that determines the relative degree of selection on observed and unobserved variables, which is denoted by Ξ . Despite this parameter being fundamentally unknown, Altonji et al. (2005) and Oster (2019) show that $\Xi = 1$ is a reasonable upper-bound value. Second, there is the maximum R^2 from a hypothetical regression of prices on the share of greenbelt land, observable and unobservable controls. $R^2_{\max} = 1$ is again a reasonable upper bound value.

TABLE B3 – AMENITY EFFECTS OF GREENBELTS:
BIAS-ADJUSTED ESTIMATES

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

| | (1) GMM | (2) GMM | (3) GMM |
|-------------------------------|-----------------------|-----------------------|--------------------------|
| | All areas | Greenbelts in 1973 | Greenbelt border <1km |
| Share greenbelt land 0-500m | 0.0682*** (0.0198) | 0.0987*** (0.0216) | 0.1091*** (0.0280) |
| Housing attributes | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes |
| Distance to the border | No | No | Yes |
| Local authority fixed effects | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes |
| Number of observations | 10,070,791 | 1,952,693 | 3,331,788 |
| Bandwidth (km) | ∞ | ∞ | 1 |
| Ξ | 1 | 1 | 1 |
| R^2_{\max} | 1 | 1 | 1 |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (2) includes observations in areas that are in greenbelts that were approved or considered in 1973. In column (3) I include transactions that are within 1km of a greenbelt boundary respectively. Bootstrapped standard errors (250 replications) are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

I then show the results of bias-adjusted estimates given $\Xi = R^2_{\max} = 1$ for the three identification strategies used: using all observations, using observations in 1973 greenbelts, and using observations within 1km of a greenbelt border.

Column (1) in Table B3 shows that the bias-adjusted estimate is considerably lower than the comparable estimate in column (2) of Table 3. A 10% increase in greenbelt land now seems to increase prices by just 0.68%. However, when I turn to the more believable identification strategy based on the boundary discontinuity, I find that the bias-adjusted estimates are much closer to the baseline estimates. In column (3), where I focus on properties within 1km of the border, I find a bias-adjusted estimate of 0.109, while the comparable estimate in column (5), Table 3 is 0.1239. Also note that, because R^2_{\max} and Ξ are upper-bound values, the differences between bias-adjusted estimates and baseline OLS estimates are likely even smaller. This provides additional evidence that the border-discontinuity approach successfully controls for

TABLE B4 – AMENITY EFFECTS OF GREENBELTS: NON-LINEARITY
(Dependent variable: the log of house price per m²)

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS |
|---|-----------------------|-----------------------|--------------------------|--------------------------|
| | All obs. | Greenbelts in 1973 | Greenbelt border <2km | Greenbelt border <1km |
| Share greenbelt land 0-500m × ($\tilde{g}_i \leq 0.2$) | 0.1346*** (0.0174) | 0.1033*** (0.0230) | 0.0764*** (0.0241) | 0.0700** (0.0289) |
| Share greenbelt land 0-500m × ($0.2 < \tilde{g}_i \leq 0.4$) | 0.1227*** (0.0120) | 0.1003*** (0.0150) | 0.0751*** (0.0172) | 0.0677*** (0.0219) |
| Share greenbelt land 0-500m × ($0.4 < \tilde{g}_i \leq 0.6$) | 0.1289*** (0.0097) | 0.1063*** (0.0123) | 0.0931*** (0.0137) | 0.0847*** (0.0174) |
| Share greenbelt land 0-500m × ($0.6 < \tilde{g}_i \leq 0.8$) | 0.1401*** (0.0088) | 0.1233*** (0.0111) | 0.1108*** (0.0123) | 0.1017*** (0.0155) |
| Share greenbelt land 0-500m × ($\tilde{g}_i > 0.8$) | 0.1821*** (0.0093) | 0.1646*** (0.0111) | 0.1574*** (0.0134) | 0.1451*** (0.0175) |
| Housing attributes | Yes | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 10,069,771 | 1,952,358 | 3,331,326 | 2,172,119 |
| Bandwidth (km) | ∞ | ∞ | 2 | 1 |
| R ² | 0.8721 | 0.8800 | 0.8732 | 0.8736 |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (2) includes observations in areas that are in greenbelts that were approved or considered in 1973. In columns (3) and (4) I include transactions that are within 2 and 1km of a greenbelt boundary, respectively. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

unobservable housing characteristics.

B.2.3 Non-linearity in the share of greenbelt land

Table B4 analyses non-linearity in the effects of greenbelt land on house prices. I interact the share of greenbelt land within 500m with 5 dummies capturing the share of greenbelt land within 500m in 5 discrete categories. The results indeed indicate that the marginal effect of greenbelt land is increasing. For example, in the preferred specification in column (4), when the share of greenbelt land is below 20%, a 10-percentage-point increase in greenbelt land increases prices by 0.7%. However, when the share of greenbelt land within 500m exceeds 80%, the effect more than doubles.

TABLE B5 – AMENITY EFFECTS OF GREENBELTS: LAND-USE TYPES
(Dependent variable: the log of house price per m²)

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) OLS |
|---|---------------------------------|-----------------------|--------------------------|---------------------------|--------------------------|
| | + Controls and fixed effects | Greenbelts in 1973 | Greenbelt border <2km | Outside greenbelt area | Inside greenbelt area |
| Share greenbelt land 0-500m – Golf courses | 0.5537*** (0.0606) | 0.4459*** (0.0692) | 0.4704*** (0.0707) | 0.4700*** (0.0841) | 0.4171*** (0.0974) |
| Share greenbelt land 0-500m – Parks and recreational grounds | -0.0495 (0.0456) | 0.1074* (0.0554) | -0.0542 (0.0524) | -0.0565 (0.0595) | -0.0287 (0.0966) |
| Share greenbelt land 0-500m – Farms | 0.2292*** (0.0204) | 0.1652*** (0.0239) | 0.1449*** (0.0279) | 0.1031** (0.0415) | 0.1830*** (0.0282) |
| Share greenbelt land 0-500m – Meadows | 0.2747*** (0.0515) | 0.2584*** (0.0515) | 0.1911*** (0.0551) | 0.0647 (0.0732) | 0.2012*** (0.0615) |
| Share greenbelt land 0-500m – Forests | 0.1189*** (0.0421) | 0.1327*** (0.0427) | 0.0128 (0.0547) | -0.1190 (0.0735) | 0.1947*** (0.0623) |
| Share greenbelt land 0-500m – Other | 0.2365*** (0.0157) | 0.1463*** (0.0205) | 0.1262*** (0.0202) | 0.0615** (0.0311) | 0.1232*** (0.0239) |
| Housing attributes | Yes | Yes | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | Yes | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 10,070,791 | 1,952,693 | 3,331,788 | 2,858,590 | 473,194 |
| Bandwidth (km) | ∞ | ∞ | 2 | 2 | 2 |
| R ² | 0.7765 | 0.7913 | 0.7691 | 0.7695 | 0.7897 |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (2) includes observations in areas that are in greenbelts that were approved or considered in 1973. In columns (3)-(5) I include transactions that are within 2km of a greenbelt boundary respectively. Columns (4) and (5) include properties outside and inside greenbelt areas, respectively. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

B.2.4 Types of greenbelt land

I start by reporting the amenity effects of greenbelts broken down into different land-use categories in Table B5. Using data from [OpenStreetMap](#), I distinguish between greenbelt land used for the following: golf courses; parks and recreational grounds; farms; meadows, grass, and heath; forests; and other. High-end recreation purposes, such as golfing, increase house prices substantially. A 10-percentage-point increase in greenbelt land used for golf courses increases house prices by 4-5%.

Perhaps surprisingly, the effects of greenbelt land classified as parks and recreation grounds are mostly statistically insignificant. My interpretation is that the demand for public parks in suburbs and beyond is lower than in inner cities because most properties are terraced, semi-

detached, or detached homes with private gardens. Hence, parks and recreation grounds may even imply negative external effects because of noise and visual pollution, particularly in the case of sports fields. Still, note that the relatively large confidence bands do not exclude positive effects.

The effects of greenbelt land classified as farms, meadows, and other are generally positive and are, in terms of magnitude, comparable to the baseline results. By contrast, the effect of forested greenbelt land is not very clear. For identification strategies based on all observations and 1973 greenbelts (columns (1) and (2), respectively), I find positive effects, whereas the effects are statistically insignificant for the boundary-discontinuity approach. On the other hand, if I only include properties inside greenbelt areas, I find a positive effect again. This may suggest that particularly for forested land, small patches are hardly price increasing. Only when properties are mostly surrounded by forested greenbelt land (as is more likely inside greenbelt areas) the effect of forested greenbelt land is potentially positive. Overall, my results are in line with existing evidence showing that open land is valued higher by residents than forested land (Irwin 2002, Montgomery 2015, pp. 114-115).

B.2.5 Detailed fixed effects

Table B6 investigates whether the detail of the fixed effects impacts the results. Instead of 326 local authority fixed effects, I include 159,867 output area fixed effects to capture sorting effects related to greenbelts and to further address omitted variable bias. Recall that output areas are the lowest geographical level at which Census estimates are provided; the median size of an OA is only 6.6ha. Fortunately, this leads to very similar results. In columns (1)-(3) I find that the effects on housing density are very similar, as the reduction in dwellings when a postcode is in a greenbelt is around 60%.

In columns (4)-(6) I find that increasing the share of greenbelt land by 10 percentage points increases prices by 1.5%, which is essentially the same as the baseline results reported in Table 3. Hence, this increases the belief that my identification strategies indeed identify a causal amenity effect of greenbelts, as unobserved locational endowments (here captured by the detailed fixed effects) seem to be uncorrelated to the share of greenbelt land within 500m.

TABLE B6 – SUPPLY AND AMENITY EFFECTS OF GREENBELTS: DETAILED FIXED EFFECTS

| | Dependent variable: the number of dwellings in a postcode | | | Dependent variable: the log of house price per m ² | | |
|--------------------------------------|---|------------------------|------------------------|---|-----------------------|-----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Poisson | Poisson | Poisson | OLS | OLS | OLS |
| | All obs. | Greenbelts in 1973 | Greenbelt border <1km | All obs. | Greenbelts in 1973 | Greenbelt border <1km |
| Share greenbelt land 0-500m | | | | 0.1544*** (0.0084) | 0.1427*** (0.0106) | 0.1244*** (0.0145) |
| Share greenbelt land | -0.9402*** (0.0112) | -1.2719*** (0.0182) | -1.1607*** (0.0198) | | | |
| Area size of postcode (<i>log</i>) | 0.3210*** (0.0024) | 0.4331*** (0.0050) | 0.4225*** (0.0047) | | | |
| Housing attributes | No | No | No | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | No | No | Yes |
| Output area fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | No | No | No | Yes | Yes | Yes |
| Number of observations | 1,309,635 | 249,542 | 253,914 | 10,069,771 | 1,952,358 | 2,172,119 |
| Bandwidth (<i>km</i>) | ∞ | ∞ | 1 | ∞ | ∞ | 1 |
| R^2 | | | | 0.8721 | 0.8800 | 0.8736 |
| Pseudo- R^2 | 0.355 | 0.374 | 0.362 | | | |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Columns (2) and (5) include observations in areas that are in greenbelts that were approved or considered in 1973. In columns (3) and (6) I include transactions that are within 1km of a greenbelt boundary. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.2.6 Reducing the bandwidth

I further investigate robustness by improving on the boundary-discontinuity design by restricting the bandwidth to just 100m on both sides of the greenbelt border. This reduces the number of observations by approximately 95%. Given that, locally, there may be some measurement error in the greenbelt border, I expect the results to be somewhat smaller. Indeed, columns (1) and (2) indicate that when a postcode is fully designated as greenbelt land, the reduction in housing density is around 50%. Column (3)-(5) indicate that prices increase by roughly 0.5-0.7% for every 10-percentage-point increase in the share of greenbelt land within 500m. However, particularly in column (5), the coefficient is too imprecise to draw strong conclusions. Still, I believe that these results confirm the presence of strong supply and amenity effects caused by greenbelt policy.

TABLE B7 – SUPPLY AND AMENITY EFFECTS OF GREENBELTS:
WITHIN 100M OF THE BORDER

| | Dependent variable: the number of dwellings in a postcode | | Dependent variable: the log of house price per m ² | | |
|--------------------------------------|--|--------------------------|--|---------------------------|--------------------------|
| | (1) Poisson | (2) Poisson | (3) OLS | (4) OLS | (5) OLS |
| | Greenbelt border <0.1km | Inside greenbelt area | Greenbelt border <0.1km | Outside greenbelt area | Inside greenbelt area |
| Share greenbelt land 0-500m | | | 0.0680*** (0.0212) | 0.0461** (0.0225) | 0.0708 (0.0537) |
| Share greenbelt land | -0.6198*** (0.0208) | -0.7603*** (0.0681) | | | |
| Area size of postcode (<i>log</i>) | 0.1569*** (0.0063) | 0.0665*** (0.0098) | | | |
| Housing attributes | No | No | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes | Yes |
| Distance to the border | Yes | Yes | Yes | Yes | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | No | No | Yes | Yes | Yes |
| Number of observations | 64,990 | 8,199 | 519,894 | 498,025 | 21,864 |
| Bandwidth (<i>km</i>) | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| <i>R</i> ² | | | 0.7706 | 0.7736 | 0.7525 |
| Pseudo- <i>R</i> ² | 0.0945 | 0.0831 | | | |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. I include transactions that are within 100m of a greenbelt boundary. Columns (2) and (5) include properties inside greenbelt areas, while column (4) includes observations outside greenbelt areas. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

B.2.7 Sample selections

In columns (1)-(3) of Table B8, I only include observations in 2011. As the structural estimation results will rely on *Census* data from 2011, I want to ensure that amenity effects are roughly the same over time. I find evidence for this: the results relying solely on housing transactions in 2011 are not materially different from the baseline reduced-form estimates on prices.

Columns (4)-(6) of Table B8 only include flats. As single-family homes are overrepresented in the data, it may be that the overall estimates are biased. However, I show that the results for apartments are essentially the same.

B.2.8 The spatial distribution of sales

One may be concerned that the spatial distribution of sales is not random and correlated to the share of greenbelt land. For example, houses in greenbelts in particular may stay in the family

TABLE B8 – AMENITY EFFECTS OF GREENBELTS: DATA FROM 2011 AND FLATS
(Dependent variable: the log of house price per m²)

| | Data from 2011 only | | | Only flats | | |
|-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | OLS | OLS | OLS | OLS | OLS | OLS |
| | All obs. | Greenbelts in 1973 | Greenbelt border <1km | All obs. | Greenbelts in 1973 | Greenbelt border <1km |
| Share greenbelt land 0-500m | 0.2072*** (0.0104) | 0.1604*** (0.0134) | 0.1602*** (0.0128) | 0.1674*** (0.0133) | 0.1549*** (0.0195) | 0.1570*** (0.0208) |
| Housing attributes | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | No | No | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 360,591 | 71,581 | 79,413 | 320,573 | 61,657 | 59,741 |
| Bandwidth (km) | ∞ | ∞ | 1 | ∞ | ∞ | 1 |
| R ² | 0.6535 | 0.6739 | 0.6265 | 0.6211 | 0.6183 | 0.6311 |

Notes: Housing attributes include the log of house size, housing type dummies (flat, terraced, semi-detached, detached), the number of rooms and the number of habitable rooms, an indicator for newly built properties, the floor level of the property, the height of the property, the number of stories of the building, whether the property has a fire place, whether the property is freehold and variables capturing the energy efficiency of windows, roof, walls. Distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Columns (2) and (5) include observations in areas that are in greenbelts that were approved or considered in 1973. In columns (3) and (6) I include transactions that are within 1km of a greenbelt boundary. Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

or are passed through inheritance, and hence are not traded on the market. This would lead to a lower number of sales per dwelling in postcodes in greenbelts.

I therefore calculate the total number of sales from the Land Registry between 1995 and 2017 in each postcode, which I then divide by the total number of dwellings in 2011. I then estimate a similar set of regressions as before but take the relative number of sales as the dependent variable. The results are reported in Table B9

Column (1) shows that there is indeed a small negative effect of the share of greenbelt land in a postcode on the relative number of sales. However, the effect is economically small (*i.e.* given a mean of 0.391, the effect is only 2.6%) and only significant at the 5% level. In column (2), where only postcodes in approved or proposed greenbelt land in 1973 are considered, the effect is essentially zero. In column (3) of Table B9, I exploit variation around greenbelt borders. The coefficient is again negative, but once I reduce the distance to greenbelt borders to 1km, the effect is again small and statistically insignificant. Hence, the issue of non-random sales does not seem to play a major role.

TABLE B9 – SUPPLY EFFECTS OF GREENBELTS: SALES PER DWELLING
(Dependent variable: number of transactions / number of dwellings)

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS |
|--------------------------------------|------------------------|------------------------|--------------------------|--------------------------|
| | All areas | Greenbelts in 1973 | Greenbelt border <2km | Greenbelt border <1km |
| Share greenbelt land | -0.0103** (0.0048) | -0.0035 (0.0082) | -0.0196** (0.0081) | -0.0086 (0.0091) |
| Area size of postcode (<i>log</i>) | -0.0624*** (0.0011) | -0.0503*** (0.0023) | -0.0473*** (0.0023) | -0.0465*** (0.0026) |
| Dwelling density | Yes | Yes | Yes | Yes |
| Distance to the city centre | Yes | Yes | Yes | Yes |
| Distance to the border | No | No | Yes | Yes |
| Local authority fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 1,205,114 | 233,504 | 363,456 | 238,391 |
| Bandwidth (<i>km</i>) | ∞ | ∞ | 2 | 1 |
| R^2 | 0.0613 | 0.0598 | 0.0451 | 0.0440 |

Notes: I calculate the total number of sales between 1995 and 2017 in each postcode. The number of dwellings is from 2011. Dwelling density refers a linear, squared and cubic term of the log of dwelling density. Similarly, distance to the city centre refers to a linear, squared and cubic term of distance to the nearest city centre. Distance to the border refers to a linear and squared term of distance to the nearest greenbelt border on either side of the greenbelt border. Column (2) includes observations in areas that are in greenbelts that were approved or considered in 1973. In columns (3) I include transactions that are within 2km of a greenbelt border, which is reduced to 1km in column (4). Standard errors are clustered at the MSOA level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.3 Greenbelts and air pollution

It has been argued that greenbelts may also reduce air pollution in cities (Yang and Jinxing, 2007). If this is indeed the case, it may be that amenity levels are also influenced further away from a greenbelt. I test this more explicitly by using data on two sources of pollution: particulate matter (PM₁₀) and nitrogen oxide (NO_x). I gather data from the UK National Atmospheric Emissions Inventory for 2016, which modelled air pollution in great detail using a 1km grid. Note that greenbelts can influence pollution in various ways. First, trees and wetlands can absorb PM₁₀ and NO_x (Powe and Willis, 2004). As greenbelts likely have more trees, this may lead to local reductions in PM₁₀. Second, greenbelts further imply that there is more land available for agricultural activities. The latter are associated with emissions of PM₁₀ and NO_x (Bontemps et al., 2008; Pattey and Qiu, 2012).

As a dependent variable, I take the log of one of the sources of pollution and estimate regressions

TABLE B10 – GREENBELTS AND POLLUTION: PARTICULATE MATTER

(Dependent variable: the log of PM_{10})

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) OLS |
|--|------------------------|---------------------------------|------------------------|--------------------------|--------------------------|
| | | + Controls and fixed effects | Greenbelts in 1973 | Greenbelt border <2km | Greenbelt border <1km |
| Share greenbelt land | -1.3581*** (0.0525) | -0.2277*** (0.0390) | -0.2814*** (0.0619) | -0.3631*** (0.0628) | -0.4214*** (0.0844) |
| Share greenbelt land 0-2500m | -1.0188*** (0.1192) | -0.0072 (0.0870) | 0.1245 (0.1325) | -0.0170 (0.1315) | 0.1357 (0.1752) |
| Share greenbelt land 2500-5000m | 1.6690*** (0.0662) | -0.0877 (0.0605) | -0.1958* (0.0999) | -0.0552 (0.0944) | -0.1273 (0.1308) |
| Dwellings per ha (<i>log</i>) | | 0.5713*** (0.0073) | 0.5390*** (0.0178) | 0.5299*** (0.0215) | 0.4942*** (0.0333) |
| Dwellings per ha 0-2500m (<i>log</i>) | | 0.1498*** (0.0180) | 0.1817*** (0.0323) | 0.1602*** (0.0356) | 0.1944*** (0.0476) |
| Dwellings per ha 2500-5000m (<i>log</i>) | | -0.0211 (0.0192) | -0.0517 (0.0326) | -0.0513 (0.0373) | -0.0801 (0.0530) |
| Distance to the city centre | No | Yes | Yes | Yes | Yes |
| Local authority fixed effects | No | Yes | Yes | Yes | Yes |
| Number of observations | 6,791 | 6,789 | 2,252 | 2,140 | 1,374 |
| R^2 | 0.1271 | 0.8644 | 0.8050 | 0.7809 | 0.7411 |

Notes: Distance to the city centre controls for a linear, squared and cubic term of distance to the nearest city centre. Column (3) includes observations in areas that are for more than 90% in proposed or approved greenbelts in 1973. In columns (4) and (5) I include transactions that are within 2km or 1km of a greenbelt boundary respectively. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

of the following form:

$$\log p_i = \chi_1 g_i + \sum_{\mathcal{B}} \chi_{\mathcal{B}}^g g_{i\mathcal{B}} + \chi_2 \log \left(\frac{d_i}{L_i} \right) + \sum_{\mathcal{B}} \chi_{\mathcal{B}}^d \left(\frac{d_{i\mathcal{B}}}{L_{i\mathcal{B}}} \right) + \chi_3 m_i + \overline{\omega}_{i \in \mathcal{A}} + \epsilon_i. \quad (\text{B.1})$$

where p_i denotes pollution and $\chi_1, \chi_{\mathcal{B}}^g \forall \mathcal{B}, \chi_2, \chi_{\mathcal{B}}^d \forall \mathcal{B}, \chi_3$, and $\overline{\omega}_{i \in \mathcal{A}}$ are parameters to be estimated. I include 2.5km distance bands capturing the share of greenbelt land, as pollution in a certain location may be influenced by other locations. I control for the density of dwellings in each band. I report results in Table B10 for particulate matter and in Table B11 for nitrogen oxide.

I show in column (1) of Table B10 that the share of greenbelt land in one's own MSOA is negatively associated with the concentration of particulate matter. Further away, I find a negative coefficient for greenbelt land between 0 and 2.5km and a positive coefficient for greenbelt land between 2.5 and 5km, which is relatively challenging to interpret. I therefore include dwelling density, a flexible function of distance to the nearest city centre, and local authority fixed effects in the next specification. Column (2) does still find an effect of greenbelt

TABLE B11 – GREENBELTS AND POLLUTION: NITROGEN OXIDE
(Dependent variable: the log of NO_x)

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) OLS |
|-----------------------------------|------------------------|---------------------------------|------------------------|--------------------------|--------------------------|
| | | + Controls and fixed effects | Greenbelts in 1973 | Greenbelt border <2km | Greenbelt border <1km |
| Share greenbelt land | -0.7323*** (0.0639) | -0.0457 (0.0609) | -0.3378*** (0.0948) | -0.4625*** (0.0991) | -0.5167*** (0.1349) |
| Share greenbelt land 0-2500m | -1.3465*** (0.1398) | 0.0696 (0.1358) | 0.5478*** (0.2098) | 0.1736 (0.2079) | 0.1857 (0.2897) |
| Share greenbelt land 2500-5000m | 1.6845*** (0.0757) | -0.1669* (0.0945) | -0.1738 (0.1529) | 0.0762 (0.1519) | 0.0673 (0.2114) |
| Dwellings per ha (log) | | 0.3751*** (0.0108) | 0.2450*** (0.0270) | 0.1691*** (0.0328) | 0.1217** (0.0508) |
| Dwellings per ha 0-2500m (log) | | 0.2902*** (0.0257) | 0.3682*** (0.0511) | 0.3494*** (0.0581) | 0.3460*** (0.0779) |
| Dwellings per ha 2500-5000m (log) | | -0.0319 (0.0269) | -0.0450 (0.0509) | -0.1311** (0.0606) | -0.1147 (0.0842) |
| Distance to the city centre | No | Yes | Yes | Yes | Yes |
| Local authority fixed effects | No | Yes | Yes | Yes | Yes |
| Number of observations | 6,791 | 6,789 | 2,252 | 2,140 | 1,374 |
| R ² | 0.0705 | 0.6957 | 0.5342 | 0.4867 | 0.4464 |

Notes: Distance to the city centre controls for a linear, squared and cubic term of distance to the nearest city centre. Column (3) includes observations in areas that are for more than 90% in proposed or approved greenbelts in 1973. In columns (4) and (5) I include transactions that are within 2km or 1km of a greenbelt boundary respectively. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

land on the concentration of particulate matter inside the MSOA. The coefficient implies that particulate matter is reduced by 2.3% for a 10-percentage-point increase in the share of greenbelt land. Outside one's own MSOA, I do not find statistically significant effects. Note that dwelling density, unsurprisingly, leads to an increase in PM₁₀ (the elasticity is around 0.5). Also, dwelling density outside the MSOA impacts pollution in one's own MSOA. These results are confirmed in column (3), where I focus on areas in greenbelts in 1973, except that I also observe effects beyond 2500m. This is likely a Type I error, as there are very few areas that are inside greenbelts in 1973 but are far away from current greenbelts. When I only exploit local variation in pollution in columns (4) and (5), I confirm that *inside* greenbelts the concentration of PM₁₀ is lower, but there is no effect of greenbelt land beyond one's own MSOA.

In Table B11 I report the results for nitrogen oxide. In areas with a high concentration of motor vehicle traffic, such as in and around large cities, nitrogen oxides emitted can be an important source of air pollution, as NO_x is mostly produced from the reaction between nitrogen and oxygen during the combustion of fuels. Column (1) shows that the concentration of nitrogen

oxides inside greenbelts is lower, but, again, I find both positive and negative coefficients beyond one's own MSOA. Including local authority fixed effects solves the issue (column (2)). Once more, in line with the results found for PM_{10} , I find that coefficients are insignificant beyond one's own MSOA, confirming the conclusion that, if anything, the impact of greenbelts on pollution is local and is therefore absorbed in revealed amenity levels.

Appendix C. Structural model

In this Appendix, I undertake a wide range of additional analyses to support the identification of structural parameters reported in Table 5. In Appendix C.1, I focus on the parameters estimated in the gravity model. In Appendix C.2, I demonstrate the robustness of the parameter that measures the capital cost share in construction. Appendix C.3 shows how quick net utility, productivity, and greenbelt amenities attenuate given the implied decay parameters, followed by commuting congestion models in Appendix C.4

C.1 Gravity models

In Table C1 I report the results of several commuting gravity models identifying the commuting time elasticity, ε_m .

I start with the baseline commuting gravity model with free-flow travel time on the road with residence and workplace location fixed effects. I estimate the model using Pseudo-Maximum Likelihood Poisson regressions. I find a (semi-)elasticity of -0.128 , indicating that when commuting time increases by 1 minute, the number of commuters changes by $e^{-0.128} - 1 = -12\%$. The elasticity is very similar to that indicated by Ahlfeldt et al. (2015), who use travel time by public transport and find an elasticity of approximately 0.1.

Note that in the main analyses I rely on actual travel time between i and j , rather than the free-flow travel time. I show in column (2) that this matters: the elasticity when using actual travel times is reduced by almost 50%. This is because travel times are actually longer, so people are less sensitive to longer commutes than free-flow travel times might suggest.

One may argue that travel times are endogenous. On the one hand, commuting times are longer on links where there are many commuters, leading to longer travel times. On the other hand, as busy links with lots of commuters may attract infrastructure investments, leading to lower travel times. As an instrument for commuting times, I use the Euclidean distance. Because a Poisson model is a non-linear model, I cannot employ 2SLS. Alternatively, I use a control function approach where the first-stage error is inserted as a control variable in the second stage. This leads to a very similar effect as that seen in column (3). The first-stage error appears to be statistically significant, suggesting that endogeneity is relevant. However, the travel time elasticity is hardly affected and is essentially the same as that in the baseline specification.

TABLE C1 – COMMUTING GRAVITY MODELS
(Dependent variable: the number of commuters between i and j)

| | By road | | | | By rail | | | | Other modes | |
|--|------------------------|------------------------|-------------------------------|-------------------------------|------------------------|-------------------------------|-------------------------------|------------------------|-------------------------------|-------------------------------|
| | Free flow | Baseline specification | +Euclidian distance | Travel time <60 minutes | Baseline specification | +Euclidian distance | Travel time <60 minutes | Baseline specification | +Euclidian distance | Travel time <60 minutes |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Commuting time (<i>minutes</i>) | | -0.0686*** (0.0001) | -0.0670*** (0.0001) | -0.0984*** (0.0003) | -0.0428*** (0.0001) | -0.0445*** (0.0001) | -0.0797*** (0.0004) | -0.0121*** (0.0000) | -0.0110*** (0.0007) | -0.0340*** (0.0021) |
| Commuting time, free-flow (<i>minutes</i>) | -0.1280*** (0.0004) | | | | | | | | | |
| First-stage error | | | -0.0100*** (0.0008) | 0.0457*** (0.0008) | | 0.0084*** (0.0004) | 0.0472*** (0.0006) | | -0.0326*** (0.0020) | -0.0099*** (0.0010) |
| Residence location fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Work location fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of areas | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 |
| Number of area pairs | 19,673,517 | 19,673,517 | 19,673,517 | 7,069,155 | 19,673,517 | 19,673,517 | 7,069,155 | 19,673,517 | 19,673,517 | 7,069,155 |

Notes: We estimate the parameters using data at the Mid-layer Super Output Area (MSOA). **Bold** indicates instrumented. I use Euclidian distance as an instrument in columns (3), (4), (6), (7), (9) and (10). Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

In column (4), I reduce the maximum commute in the data from 120 to 60 minutes free-flow travel time. This reduces the number of location pairs by almost 65%. The coefficient is slightly higher: when commuting time increases by 1 minute, the number of commuters decreases by 9.4%. The elasticity is slightly higher when restricting the commuting time because around 3% of the commuters that are not particularly sensitive to travel time are excluded.

I repeat the latter three specifications for travel by train and metro in columns (4)-(6) and other modes in columns (7)-(9) in Table C1. I find a lower semi-elasticity for travel time by rail, which is in line with the idea that commuters can engage in other activities (*e.g.* reading, working, listening to music) while travelling, making them less sensitive to travel time (see [Ohmori and Harata, 2009](#)). The semi-elasticity for other modes is low, which is possibly because I have to group transport modes that may be quite different (*e.g.* walking, working from home, and taking a scooter). I assume a speed on the network of 10km per hour, but this may be too low for some transport modes, leading to a relatively low semi-elasticity. Because the share of commuters using other modes is not large (around 15%), the size of the semi-elasticity does not materially influence the results.

C.2 *Capital cost share in construction*

C.2.1 **Sensitivity of μ**

The capital cost share in construction, μ , is obtained through a regression of total floor space consumption on floor space prices as stipulated in equation (33). Table C2 reports the results.

Column (1) shows a naive regression of the log of the ratio of built-up land over all land on floor space prices. Reverse causality thwarts a causal interpretation of this coefficient. I find a counter-intuitive *negative* μ , implying that lower floor space prices are associated with a higher share of built-up area (meaning more supply of floor space).

Column (2) then instruments for floor space by the distance to the city centre. The idea here is that closer to the city centre, commutes are shorter and agglomeration economies are more pronounced so that prices are higher, which makes the prospect of constructing taller buildings seem more attractive ([Alonso, 1964](#); [Muth, 1969](#); [Mills and Lubuele, 1997](#)). The first-stage results in Table C3 confirm this hypothesis: when the distance to the nearest city centre increases, the price decreases. Although the second-order term is positive, the marginal effect is always

TABLE C2 – SHARE OF CAPITAL IN CONSTRUCTION COSTS
(Dependent variable: the log of the ratio of built-up land over all land in i)

| | OLS | Instrument for prices | + soil, controls | Greenbelts in 1973 | Greenbelt border <1km | + county fixed effects | Temperature instruments |
|---|------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|----------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Share of construction costs in capital, μ | -0.4549*** (0.0430) | 0.6647*** (0.0158) | 0.7096*** (0.0149) | 0.5627*** (0.0816) | 0.7018*** (0.1073) | 0.7131*** (0.0281) | 0.7067 (0.4520) |
| Supply condition variables | No | No | Yes | Yes | Yes | Yes | Yes |
| County fixed effects | No | No | No | No | No | Yes | No |
| Greenbelt-by-government region fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of areas | 6,791 | 6,791 | 6,791 | 1,019 | 1,348 | 6,791 | 6,791 |
| R^2 | 0.373 | | | | | | |
| Kleibergen-Paap F-statistic | | 251.96 | 203.47 | 33.23 | 16.42 | 66.69 | 14.66 |

Notes: **Bold** indicates instrumented. The instrument for prices is the distance to the city centre and distance to the city centre squared. Soil condition variables include the log of elevation, the thickness of the sediment layer, the share of accessible bedrock, and the share of workers in the construction sector. We estimate the parameters using data at the Mid-layer Super Output Area (MSOA). Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

TABLE C3 – SHARE OF CAPITAL IN CONSTRUCTION, FIRST-STAGE RESULTS
(Dependent variable: the log of floor space price in i)

| | | + soil, controls | Greenbelts in 1973 | Greenbelt border <1km | + county fixed effects | Temperature instruments |
|--|--------------------------|--------------------------|------------------------|--------------------------|---------------------------|----------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Distance to the nearest city centre (in km) | -0.00525*** (0.00027) | -0.00453*** (0.00027) | -0.00592* (0.00360) | 0.00047 (0.00231) | -0.00437*** (0.00040) | |
| (Distance to the nearest city centre (in km)) ² | 0.00002*** (0.000001) | 0.00002*** (0.000001) | -0.00001 (0.000052) | -0.00008** (0.000038) | 0.00001*** (0.000002) | |
| January temperature (log) | | | | | | -0.20151*** (0.04947) |
| July temperature (log) | | | | | | 0.63370*** (0.16495) |
| Supply condition variables | No | Yes | Yes | Yes | Yes | Yes |
| County fixed effects | No | No | No | No | Yes | No |
| Greenbelt-by-government region fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of areas | 6,791 | 6,791 | 1,019 | 1,348 | 6,791 | 6,791 |
| R^2 | 0.687 | 0.708 | 0.753 | 0.690 | 0.817 | 0.693 |

Notes: Soil condition variables include the log of elevation, the thickness of the sediment layer, the share of accessible bedrock, and the share of workers in the construction sector. We estimate the parameters using data at the Mid-layer Super Output Area (MSOA). Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

negative for the distance to the city centre observed in the data. A 1km increase in distance to the city centre reduces prices by around 0.5%.

Going back to column (2) in Table C2, I find that $\hat{\mu} = 0.665$. This implies that floor space prices have a positive effect on the ratio of floor space supply to total developed land. The reason

for this is that when floor space prices increase, according to the first-order condition (21), the amount of capital should increase, which in turn increases the ratio of floor space to developed land.³⁶

When I control for potential supply conditions in column (3) $\hat{\mu}$ is somewhat higher ($\hat{\mu} = 0.710$). This value is close to the value assumed in Ahlfeldt et al. (2015) and within the range provided by the literature, which is between 0.5 and 0.75, but it is slightly lower than the value of 0.80 estimated in Combes et al. (2016) for single-family houses in France. It is not surprising that $\hat{\mu}$ is slightly higher in England than in France, as England is known for its stringent land-use regulation and height restrictions (Hilber and Vermeulen, 2016; Dericks and Koster, 2021). This will imply that a higher share of the costs of construction will be paid to land.

When I consider selections of samples, *e.g.* by selecting areas in greenbelts in 1973, or when areas are within 1km of a greenbelt border, the instrument is much less strong (see columns (4) and (5) in Table C2). When considering MSOAs in greenbelts in 1973, the estimate is slightly lower ($\hat{\mu} = 0.563$), albeit not statistically significantly different from the previous estimate. In 1973 greenbelts, there are few large towns and buildings are low. Hence, a higher share of construction costs will be devoted to land. When only considering areas within 1km of greenbelt borders in column (5), the estimate is essentially the same as the baseline estimate in column (3). In column (6) I include all areas but also include detailed county fixed effects, which hardly affects the results.

To the extent one is concerned that distance to the city centre is still correlated to supply conditions, *conditional on the inclusion of detailed location fixed effects and several variables capturing differences in supply conditions*, I also consider alternative instruments. More specifically, I calculate the average temperature in January and July as a potential demand shifter, following Glaeser et al. (2001). I expect that higher temperatures are considered as an amenity and are therefore associated with higher floor space prices. This is confirmed by the first-stage regression shown in columns (5) and (6) of Table C3, in which July temperatures in particular are correlated with higher prices. By contrast, January temperatures have a negative impact. This may be because mountainous and natural areas have lower January temperatures, while

³⁶The coefficient of floor space prices is 1.98. Hence, a 1% increase in floor space prices leads to an increase in the ratio of floor space to developed land in i of almost 2%.

still generating a high amenity level. More specifically, a 10% increase in July temperatures is associated with a price increase of approximately 6.3%. A 10% increase in January temperatures is associated with a price decrease of around 2.0%. Going back to column (6) in Table C2, I find $\hat{\mu} = 0.707$, which is virtually the same as the baseline estimate reported in column (3). However, the estimate is relatively imprecise, which is also because of a not-so-strong first stage.

C.2.2 Plausibly exogenous instruments

An alternative approach is to consider the distance to the city centre to be an imperfect, only ‘plausibly exogenous’, instrument. I consider Conley et al.’s (2012) methodology to relax the assumption of strictly exogenous instruments and construct bounds on the estimated μ if the instrument is only ‘plausibly exogenous’. Consider the following regression:

$$\log \frac{F_i}{\Lambda_i} = \frac{\mu}{1 - \mu} \log p_i + \iota_1 x_i + \iota_2 x_i^2 + \iota_3 s_i + \eta_{i \in \mathcal{G}} + \epsilon_i, \quad (\text{C.1})$$

where x_i is the distance to the city centre and ι_1 and ι_2 capture the direct impact of distance to the city centre, conditional on the effect of prices, supply conditions, and county and greenbelt fixed effects. In a standard IV-framework with exogenous instruments $\iota_1 = \iota_2 = 0$. However, if the distance to the city centre still captures favourable supply conditions, $\{\iota_1, \iota_2\} \neq 0$; hence, construction costs and therefore prices are *lower* closer to the city centre.

Of course, the actual values of ι_1 and ι_2 are unknown, but one may guess reasonable upper-bound values by performing a reduced regression of supply on the distance to the city centre, *while excluding prices*. I multiply ι_1 and ι_2 by Π , indicating the share of the effect that is assumed to be direct. I then consider two cases: (i) $\Pi = 0.1$ so that 10% of the impact of distance to the city centre on supply is direct; and (ii) I choose $\Pi = 0.25$ so 25% of the impact of distance to the city centre on supply is direct.

Table C4 reports the results. In Panel A I consider a direct impact of 10% using the reduced-form regression of the share of the log of the ratio of built-up land over all land. I find statistically significant point estimates of around 0.55-0.70 in all specifications. However, the bounds are only informative when I use all areas; in columns (3) and (4) the bounds are too imprecise to be meaningful. In the preferred specification in column (2), I find that $\hat{\mu}$ is between 0.58 and 0.71. The point estimate $\hat{\mu} = 0.687$ is very similar to the baseline estimate (*i.e.* $\hat{\mu} = 0.710$). Hence, the

TABLE C4 – SHARE OF CAPITAL IN CONSTRUCTION,
PLAUSIBLY EXOGENOUS INSTRUMENT
(Dependent variable: the log of the ratio of built-up land over all land in i)

| | | | + Supply controls | Greenbelts in 1973 | Greenbelt border <1km |
|--|-------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| | | (1) | (2) | (3) | (4) |
| <i>Panel A: $\Pi = 0.10$</i> | | | | | |
| Share of construction costs in capital | μ | 0.6408*** (0.0163) | 0.6874*** (0.0155) | 0.5366*** (0.0819) | 0.6793*** (0.1079) |
| | $\underline{\mu}$ | 0.5824*** (0.0176) | 0.6307*** (0.0170) | 0.1931 (0.2562) | 0.4072 (0.5461) |
| | $\bar{\mu}$ | 0.7066*** (0.0149) | 0.7493*** (0.0141) | 0.7040*** (0.0520) | 0.8078*** (0.0598) |
| Supply condition variables | | No | Yes | Yes | Yes |
| Greenbelt-by-government region fixed effects | | Yes | Yes | Yes | Yes |
| Number of areas | | 6,791 | 6,791 | 1,019 | 1,348 |
| <i>Panel B: $\Pi = 0.25$</i> | | | | | |
| Share of construction costs in capital | μ | 0.5979*** (0.0170) | 0.6470*** (0.0165) | 0.4911*** (0.0815) | 0.6383*** (0.1083) |
| | $\underline{\mu}$ | 0.4718*** (0.0222) | 0.5230*** (0.0236) | -0.3485 (10.1756) | 0.0336 (83.3610) |
| | $\bar{\mu}$ | 0.7271*** (0.0147) | 0.7681*** (0.0139) | 0.7452*** (0.0434) | 0.8358*** (0.0450) |
| Supply condition variables | | No | Yes | Yes | Yes |
| Greenbelt-by-government region fixed effects | | Yes | Yes | Yes | Yes |
| Number of areas | | 6,791 | 6,791 | 1,019 | 1,348 |

Notes: We report the point estimate (μ), the 5% lower bound ($\underline{\mu}$), as well as the 95% upper bound ($\bar{\mu}$). Soil condition variables include the log of elevation, the thickness of the sediment layer, the share of accessible bedrock, and the share of workers in the construction sector. We estimate the parameters using data at the Mid-layer Super Output Area (MSOA). Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

downward bias of μ due to the potential endogeneity of the instrument is limited.

Note that the bounds become only slightly wider when I consider a 25% direct impact of distance to the city centre on the log of the ratio of built-up land over all land in Panel B. In the preferred specification in column (2), the bounds imply that the estimate is between 0.523 and 0.768. When I allow for the specifications where I only use MSOAs in 1973 greenbelts or within 1km, the bounds are not particularly meaningful. Still, all in all, these results suggest that even if the instrument is only plausibly exogenous, I would still yield reasonably precise estimates for the preferred specification, as long as I use all areas to identify μ .

C.2.3 Heterogeneity in μ

Another issue is that the estimate of μ may be heterogeneous across locations and depend on the share of greenbelt land. On the one hand, the share of capital in construction costs may

be higher in areas with more greenbelt land because the limited availability of land forces developers to build taller. On the other hand, areas with lots of greenbelt land have lower densities, which possibly implies a lower μ .

To investigate this further, I re-estimate equation (33), but I allow for non-linear effects of floor space prices. That is, I estimate the following partially linear regression model:

$$\log \frac{F_i}{\Lambda_i} = f_{g_i}(\log p_i) + \iota_1 g_i + \iota_2 s_i + \eta_{i \in \mathcal{G}} + \epsilon_i. \quad (\text{C.2})$$

I specify $f_{g_i}(\cdot)$ by a locally linear function: $f_{g_i}(\cdot) = \mu_i / (1 - \mu_i) \log p_i$. Hence, each ‘local’ coefficient μ_i is dependent on the share of greenbelt land in i . For notational simplicity, let us ignore the control variables and the endogeneity of floor space prices for now. The estimator is then:

$$(\hat{\mu}_i) = \arg \min_{\mu_i} \sum_{i=1}^{\mathcal{L}} K\left(\frac{g_i - g_j}{h}\right) \times \left(\log \frac{F_i}{\Lambda_i} - \frac{\mu_i}{1 - \mu_i} p_i\right)^2. \quad (\text{C.3})$$

I specify $K(\cdot)$ to be a Gaussian kernel function:

$$K\left(\frac{g_i - g_j}{h}\right) = \frac{1}{h\sigma_g\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{g_i - g_j}{h\sigma_g}\right)^2}. \quad (\text{C.4})$$

where σ_g is the sample standard deviation of greenbelt land and h is the bandwidth. Hence, the kernel function determines the vector of weights for an MSOA i , which is between 0 and 1. It is 1 when another location j has the same share of greenbelt land. The bandwidth h determines how ‘smooth’ the function to be estimated is. I set the bandwidth as per [Silverman \(1986\)](#) so $h = 1.06\sigma_g\mathcal{L}^{-1/5}$.

I still have to deal with two issues in estimating (C.2): adding linear controls and addressing endogeneity.

To include supply conditions $\{g_i, s_i\}$ and greenbelt-by-government area dummies $\eta_{i \in \mathcal{G}}$, I apply [Robinson’s \(1988\)](#) procedure. This procedure separately regresses $\log F_i/\Lambda_i$ and the supply variables and greenbelt-by-government area dummies non-parametrically on the log of floor space prices using local linear regressions. I then obtain residuals for the dependent variable and s_i and greenbelt-by-government area dummies. The residuals of the dependent variable are then regressed on the dummy residuals and residuals of s_i and g_i using OLS, which identifies

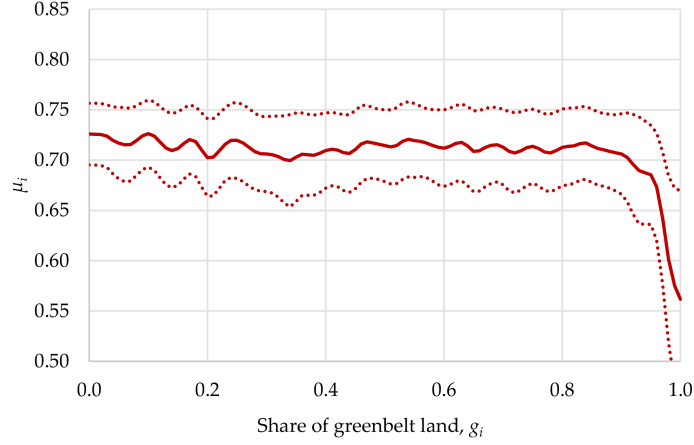


FIGURE C1 – THE CAPITAL COST SHARE AND GREENBELT LAND

Notes: The dotted lines denote 95% confidence bands.

ι and $\eta_{i \in \mathcal{G}}$. The final step is to replace the dependent variable $\log F_i/\Lambda_i$ in equation (C.2) by $\log F_i/\Lambda_i - \hat{\iota}_1 g_i - \hat{\iota}_2 s_i - \hat{\eta}_{i \in \mathcal{G}}$ and regress this residual non-parametrically on p_i to obtain the coefficients of interest.

The second issue is endogeneity. Because (C.2) is non-linear in p_i , one cannot use standard 2SLS. Instead, I adopt a control function approach (Blundell and Powell, 2003). The first stage then is given by:

$$\log p_i = \tilde{f}_{g_i}(x_i, x_i^2) + \tilde{\iota}_1 g_i + \tilde{\iota}_2 s_i + \tilde{\eta}_{i \in \mathcal{G}} + \tilde{\epsilon}_i, \quad (\text{C.5})$$

Next, I insert the first-stage residuals as a control function in the second stage:

$$\log \frac{F_i}{\Lambda_i} = f_{g_i}(\log p_i) + \iota_1 g_i + \iota_2 s_i + \eta_{i \in \mathcal{G}} + \check{f}(\tilde{\epsilon}_i) + \epsilon_i. \quad (\text{C.6})$$

I approximate $\check{f}(\cdot)$ using a third-order polynomial.³⁷

The results for μ_i are reported in Figure C1 when I include all areas (the results where I make selections are relatively imprecise). It can be clearly seen that there is very little variation in μ across the share of greenbelts, although the share of capital in construction costs seems to be slightly lower for very high shares of greenbelt land. However, given that heterogeneity is small, I will ignore heterogeneity in μ in the structural estimation because allowing for variation in the share of capital costs in construction will considerably complicate the estimation procedure, while it is unlikely to materially change the results.

³⁷Note that with $h \rightarrow \infty$ I obtain essentially the same estimate for μ as with 2SLS.

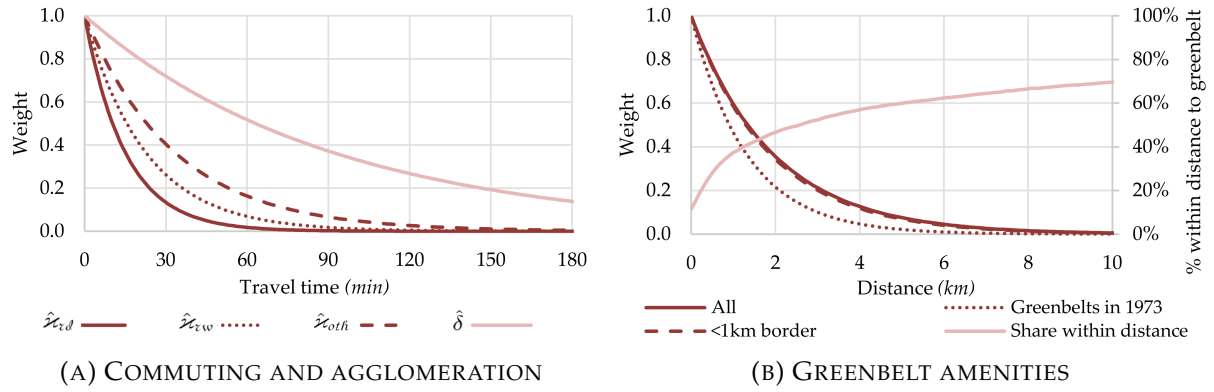


FIGURE C2 – ATTENUATION OF COMMUTING, AGGLOMERATION, AND GREENBELT AMENITIES

C.3 Attenuation

I report the attenuation of commuting and residential and production externalities in Figure C2a. It is shown that most of the utility disappears after a one-hour commute, which confirms the results of previous papers.

By contrast, I find less steep decay for agglomeration economies. Most of the externalities take place within 60 minutes of travelling. The reason for the less steep decay compared to, say, Arzaghi and Henderson (2008), Ahlfeldt et al. (2015) or Dericks and Koster (2021), is that I include data from all of England, instead of from just one city. By focusing on a whole country, interactions related to input and output sharing and labour market pooling are included, which are known to take place over longer distances.

In Figure C2b, I illustrate the decay of greenbelt amenities. In line with reduced-form evidence, greenbelt amenities decay quickly and are concentrated within 1km of the resident's location. In this figure, I also plot the share of residents within a distance of any patch of greenbelt land. It is shown that a large majority of residents are somehow affected by greenbelts, although the impact attenuates quickly over distance (*i.e.* the impact of greenbelts on amenities is given by the spatial weight, governed by θ , multiplied by the amenity impact, indicated by ζ , and the share of greenbelt land within a certain distance range (see equations (13) and (14))).

C.4 Travel time and congestion costs

Here I report alternative specifications used to estimate the traffic congestion elasticity. I measure traffic density in standard deviations. In column (1) I report the baseline specification. The coefficient indicates that a standard deviation increase in (workplace) traffic density leads

TABLE C5 – COMMUTING CONGESTION MODELS
(Dependent variable: the travel time by road between i and j)

| | Baseline specification | Population 1931 | Greenbelts 10-25km | Total population | Commuting <30 minutes | No residential fixed effects |
|--|---------------------------|------------------------------|------------------------------|-----------------------|--------------------------|---------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Traffic density, \mathcal{D}_M | 0.1006*** (0.0021) | 0.2235*** (0.0050) | 0.1370*** (0.0026) | | 0.1117*** (0.0079) | 0.1089*** (0.0024) |
| Total traffic density, $\mathcal{D}_M + \mathcal{D}_R$ | | | | 0.0948*** (0.0021) | | |
| Residence location fixed effects | Yes | Yes | Yes | Yes | Yes | No |
| Number of areas | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 | 6,701 |
| Number of area pairs | 19,673,517 | 19,673,517 | 19,673,517 | 19,673,517 | 550,788 | 19,673,517 |

Notes: We estimate the parameters using data at the Mid-layer Super Output Area (MSOA). **Bold** indicates instrumented. Standard errors are bootstrapped (250 replications) and in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

to a travel time that is $e^{0.1006} - 1 = 10.6\%$ higher. Hence, traffic density has a substantial impact on travel times.

One may argue that traffic density is endogenous, *e.g.* because infrastructure is constructed in places where traffic density is higher, leading to shorter travel times. In this case, one would find an underestimate of the estimated effect. Furthermore, actual travel times are used to determine the ‘relevant’ traffic density. In column (2) I instrument workplace traffic density by the population density in 1931. I observe that this matters for the estimated coefficient because I find a stronger effect of traffic density on travel times. When I instrument for traffic density with the share of greenbelt land between 10 and 25km I find in column (3) a coefficient that is somewhere in between the OLS estimate and the estimate where I use the 1931 population density instrument.

To calculate traffic density, I only take into account workplace density. One may be concerned that ignoring residential density might lead to incorrect estimates. Column (4), in which I calculate total traffic density by summing residential and workplace density, shows that this hardly matters. Column (5) explores whether the results change if I only consider commuting pairs that are within 30 minutes travelling, which strongly reduces the number of observations. I show that this does not materially influence the results. Furthermore, when I exclude residential fixed effects in the estimation in column (6), I also find a very similar congestion elasticity.

Appendix D. Counterfactual experiments

In Appendix D.1, I discuss the solution algorithm to obtain counterfactual outcomes, which is followed by a description of how ‘counterfactual’ greenbelts are defined in Appendix D.2. In Appendix D.3, I plot residential and employment density in the baseline scenario and in the scenario without greenbelts. Appendix D.4 shows the results for the first scenario where greenbelt land is reduced by 10%, followed by the third scenario in Appendix D.5 where the amount of greenbelt land is increased. Finally, I show robustness for the aggregate and local effects of various additional counterfactual analyses in Appendix D.6.

D.1 Solution algorithm

To solve for the new equilibrium in each of the counterfactual scenarios, I follow a similar procedure as described in the Supplement of Ahlfeldt et al. (2015). I first choose starting values for transformed wages, floor space prices, travel times, and initial population equal to the ones obtained in the baseline scenario:

$$\{\omega_{im\mathcal{S}} = \omega_{im0}, \omega_{jm\mathcal{S}} = \omega_{jm0}, p_{i\mathcal{S}} = p_{i0}, \tau_{ijm\mathcal{S}} = \tau_{ijm0}\}, \quad (\text{D.1})$$

where 0 refers to the baseline scenario. I also determine for each counterfactual scenario \mathcal{S} the counterfactual amount of developed land using:

$$\frac{\Lambda_{i\mathcal{S}}}{L_i} = \Phi_i e^{\phi g_{i\mathcal{S}}}, \quad (\text{D.2})$$

where $g_{i\mathcal{S}} = G_{i\mathcal{S}}/L_i$ is the counterfactual share of greenbelt land in each MSOA. Note Φ_i does not change for different scenarios.

Furthermore, using ancillary data on land prices and the zero-profit condition for building construction, I calibrate building capital costs using the following non-linear regression:

$$P_i = \frac{p_{i0} F_{i0}}{\Lambda_{i0}} - r^{\frac{\hat{\mu}}{\hat{\mu}-1}} (\hat{\mu} \Upsilon_i p_{i0})^{\frac{1}{1-\hat{\mu}}} \quad (\text{D.3})$$

I then take the following steps:

1. I determine the commuting probability:

$$\pi_{ijm\mathcal{S}} = \frac{\left(\frac{\Psi_{i\mathcal{S}} e^{-\hat{\kappa}_m \tau_{ijm\mathcal{S}}} w_{j\mathcal{S}}}{p_{i\mathcal{S}}^{1-\beta}} \right)^{\hat{\varepsilon}}}{\sum_{r=1}^{\mathcal{L}} \sum_{s=1}^{\mathcal{L}} \sum_{m=1}^3 \left(\frac{\Psi_{r\mathcal{S}} e^{-\hat{\kappa}_m \tau_{rsm\mathcal{S}}} w_{s\mathcal{S}}}{p_{r\mathcal{S}}^{1-\beta}} \right)^{\hat{\varepsilon}}}, \quad (\text{D.4})$$

as well as the conditional probabilities that a worker is employed in j , or is living in i :

$$\begin{aligned} \pi_{ijm|i\mathcal{S}} &= \frac{e^{-\hat{\kappa}_m \hat{\varepsilon} \tau_{ijm\mathcal{S}}} \omega_{j\mathcal{S}}}{\sum_{s=1}^{\mathcal{L}} e^{-\hat{\kappa}_m \hat{\varepsilon} \tau_{ism\mathcal{S}}} \omega_{s\mathcal{S}}}, \\ \pi_{ijm|j\mathcal{S}} &= \frac{e^{-\hat{\kappa}_m \hat{\varepsilon} \tau_{ijm\mathcal{S}}} a_{im\mathcal{S}}}{\sum_{r=1}^{\mathcal{L}} e^{-\hat{\kappa}_m \hat{\varepsilon} \tau_{irm\mathcal{S}}} a_{rm\mathcal{S}}}. \end{aligned} \quad (\text{D.5})$$

2. Based on the commuting probabilities, I determine the residential population and workers for each mode in each area:

$$\begin{aligned} H_{Mjm\mathcal{S}} &= \sum_{i=1}^{\mathcal{L}} \pi_{ijm|i\mathcal{S}} H_{Rim\mathcal{S}}, \\ H_{Rim\mathcal{S}} &= \sum_{j=1}^{\mathcal{L}} \pi_{ijm|j\mathcal{S}} H_{Mjm\mathcal{S}} \end{aligned} \quad (\text{D.6})$$

3. This provides us with the necessary information to determine amenities in a location:

$$\Psi_{i\mathcal{S}} = \check{\Psi}_i e^{-\hat{\zeta} \tilde{g}_{i\mathcal{S}}(\hat{\theta})}, \quad (\text{D.7})$$

and productivity:

$$\Omega_{i\mathcal{S}} = \check{\Omega}_i \left(\hat{\delta} \sum_{j=1}^{\mathcal{L}} \sum_{m=1}^3 e^{-\hat{\delta} \tau_{ij\mathcal{S},d}} H_{Mjm\mathcal{S}} \right)^{\hat{\gamma}}, \quad (\text{D.8})$$

where the exogenous location fundamentals $\check{\Psi}_i$ and $\check{\Omega}_i$ are kept fixed in each scenario.

4. I obtain land use in each MSOA:

$$\begin{aligned} F_{Hi\mathcal{S}} &= \sum_{m=1}^3 \frac{(1-\beta) \sum_{j=1}^{\mathcal{L}} \bar{v}_{jm}^{-\frac{1}{\hat{\varepsilon}}} \pi_{ijm|i\mathcal{S}} e^{-\hat{\kappa}_m \tau_{ijm\mathcal{S}}} \omega_{j\mathcal{S}}^{\frac{1}{\hat{\varepsilon}}}}{p_{i\mathcal{S}}} \varsigma H_{Rim\mathcal{S}}, \\ F_{Mj\mathcal{S}} &= \sum_{m=1}^3 \left(\frac{\omega_{im\mathcal{S}}^{\frac{1}{\hat{\varepsilon}}}}{\alpha \Omega_j \bar{v}_{jm}^{\frac{1}{\hat{\varepsilon}}}} \right)^{\frac{1}{1-\alpha}} \varsigma H_{Mim\mathcal{S}}. \end{aligned} \quad (\text{D.9})$$

I then define the share of commercial floor space use as:

$$\Theta_{i\mathcal{S}} = \frac{F_{Mj\mathcal{S}}}{F_{Mj\mathcal{S}} + F_{Hi\mathcal{S}}}. \quad (\text{D.10})$$

5. I update the use of building capital using the first-order condition for optimal use:

$$K_{i\mathcal{S}} = \left(\frac{\hat{\mu} \Upsilon_i p_{i\mathcal{S}}}{r} \right)^{\frac{1}{1-\hat{\mu}}} \Lambda_{i\mathcal{S}}, \quad (\text{D.11})$$

6. I then determine the output (up to a constant) in each location:

$$Y_{i\mathcal{S}} = \sum_{m=1}^3 v_{im}^{-\frac{\alpha}{\varepsilon}} \Omega_{i\mathcal{S}} H_{Mim\mathcal{S}}^{\alpha} (\Theta_{i\mathcal{S}} \Upsilon_i K_{i\mathcal{S}}^{\hat{\mu}} \Lambda_{i\mathcal{S}}^{1-\hat{\mu}})^{1-\alpha} \quad (\text{D.12})$$

7. The updated rents are given by:

$$p_{i\mathcal{S}} = \frac{(1-\alpha)Y_{i\mathcal{S}}}{\Theta_{i\mathcal{S}} \Upsilon_i K_{i\mathcal{S}}^{\hat{\mu}} \Lambda_{i\mathcal{S}}^{1-\hat{\mu}}} = \frac{(1-\beta) \sum_{m=1}^3 \sum_{j=1}^{\mathcal{L}} \bar{v}_{jm}^{-\frac{1}{\varepsilon}} \pi_{ijm|i\mathcal{S}} \omega_{im\mathcal{S}}^{1/\varepsilon} e^{-\hat{\kappa}_m \tau_{ijm\mathcal{S}}}}{(1-\Theta_{i\mathcal{S}}) \Upsilon_i K_{i\mathcal{S}}^{\hat{\mu}} \Lambda_{i\mathcal{S}}^{1-\hat{\mu}}} \quad (\text{D.13})$$

8. The updated transformed wages are given by:

$$\omega_{im\mathcal{S}} = \frac{\alpha Y_{im\mathcal{S}}}{H_{Mim\mathcal{S}}}. \quad (\text{D.14})$$

9. In the final step, I update travel times:

$$\tau_{ij\mathcal{S},\mathcal{L}} = \tau_{ij,\mathcal{L}}^f T_i \check{T}_j e^{\hat{\lambda} \mathcal{D}_{Mis}}, \quad (\text{D.15})$$

where traffic densities depend \mathcal{D}_{Mis} depend on counterfactual values of $H_{Mj\mathcal{S}}$, $\forall j$, I treat T_i and \check{T}_j as constants across different scenarios.

I repeat these 9 steps until the values for transformed wages and rents between the current and previous iteration converge. In practice, it appears that I need around 40 iterations to obtain the new equilibrium values. I also made sure that the above procedure recovers the original baseline values.

D.2 Counterfactual greenbelts

I aim to construct ‘counterfactual’ greenbelts. I gather data on the population in parishes in 1951, which was just after the 1947 Town and Country Planning Act was implemented, but before greenbelt policies were introduced.

To construct counterfactual greenbelts, I exploit data on the population in parishes from the 1951 [Census](#). Parishes are the lowest unit for which the data are available, and are rather small. The median size is 847ha. In line with the recommendation of Duncan Sandy, the Minister of Housing at that time, greenbelts were primarily implemented around larger cities of at least 100 thousand inhabitants. To identify ‘large urban areas’, I first select parishes with a population density of at least 10 people per hectare, which accounts for around 5% of parishes. I then amalgamate all those areas and keep 37 amalgamated urban areas that have a population of more than 100 thousand inhabitants. I then draw circles of 15km around each of these urban areas, in line with the suggestion of Duncan Sandy (*‘The Development Plans submitted by the local planning authorities for the Home Counties provide for a Green Belt, some 7 to 10 miles deep, [...]’*). The last step is to erase the areas in the counterfactual greenbelts with a density of at least 10 persons per hectare because that land has already been converted to built-up land and is therefore not part of a greenbelt.

In Figure [D1](#), I display a map showing the counterfactual greenbelts. At the MSOA level, the correlation between the share of greenbelt land and the share of counterfactual greenbelt land is reasonably high at around 0.458. In general, I can predict the location of most greenbelts relatively well, such as the greenbelts around London, Birmingham, Manchester, and Liverpool. For example, in MSOAs with a share of greenbelt land above 90% the share of counterfactual greenbelt land is approximately 0.9. Only the Cheltenham/Gloucester greenbelt, as well as the greenbelt around Cambridge, are not included in the counterfactual greenbelt sample because those cities had a population of fewer than 100,000 in 1951. On the other hand, reasonably large cities such as Leicester, Norwich, Middlesbrough, and Plymouth do not have greenbelts, although one would expect there to be a greenbelt around those cities based on the 1951 population distribution.

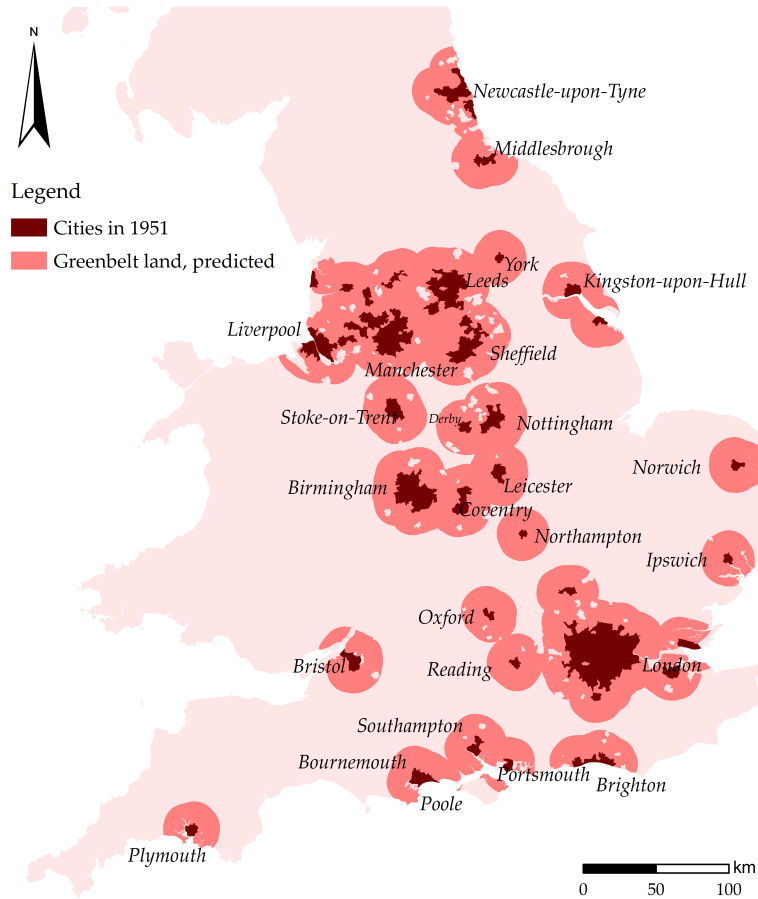


FIGURE D1 – COUNTERFACTUAL GREENBELTS

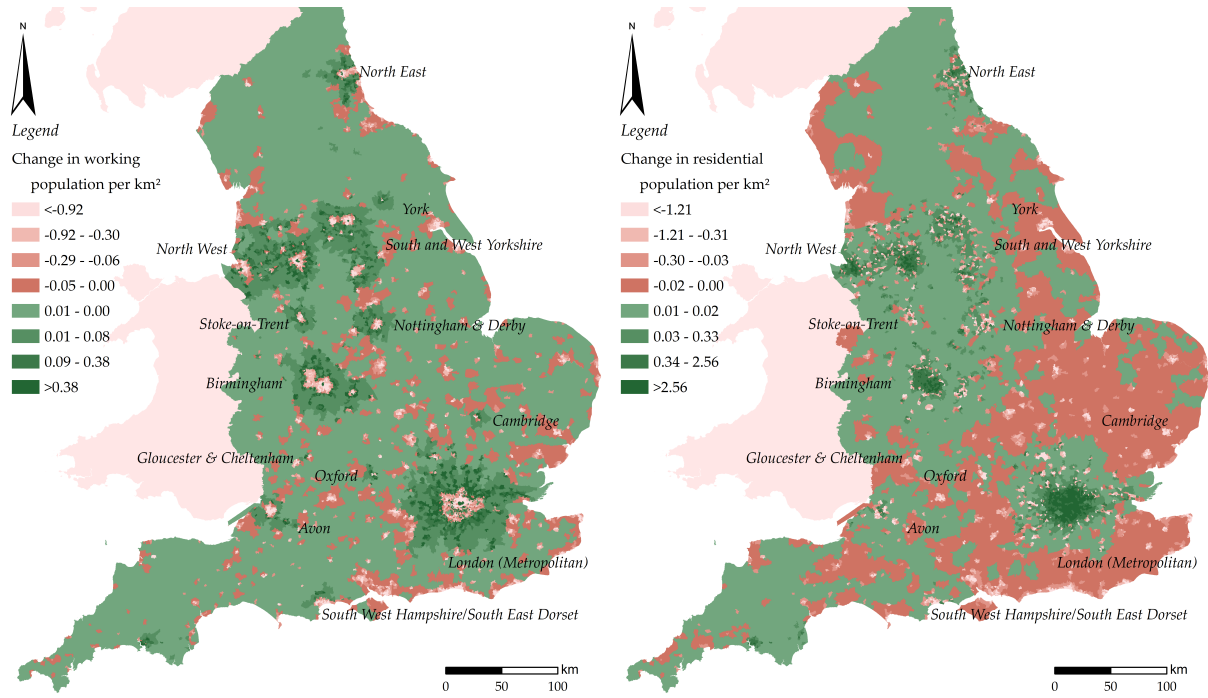
D.3 Local effects – absolute differences

In Figure 4, I display large relative changes in employment and population density if greenbelts were to be removed. However, as the relative changes depend on the employment or population in the baseline scenario, one may wonder to what extent England’s spatial structure drastically changes due to the removal of greenbelts.

In Figure D2, I plot absolute changes in employment and residential density. Although former greenbelt areas attract jobs and residents, the absolute changes are too small to fundamentally change England’s urban structure.

D.4 Local effects – 10% reduction in greenbelt land

Here I provide additional results for the local effects of the different scenarios. First, in Figure D3, I show the results for the first counterfactual experiment in which I reduce greenbelt land by 10%. Due to a loss in amenities, one observes a drop in prices in the areas that are now



(A) ABSOLUTE CHANGE IN H_{Mi}

(B) ABSOLUTE CHANGE IN H_{Ri}

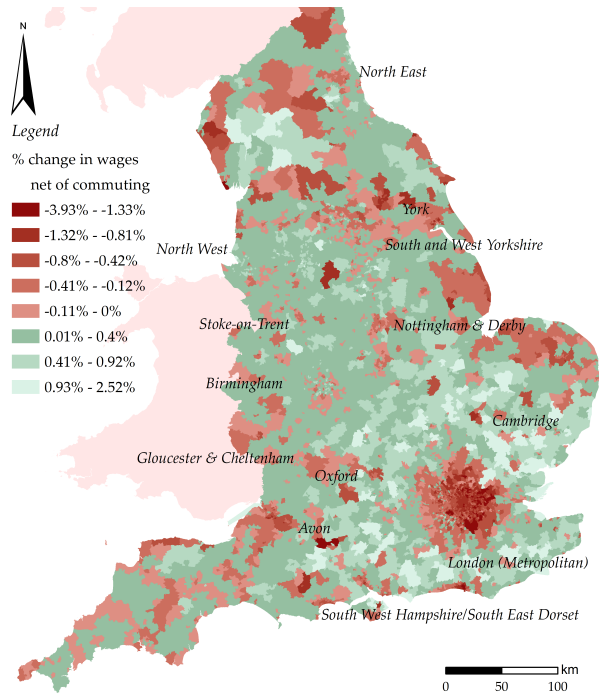
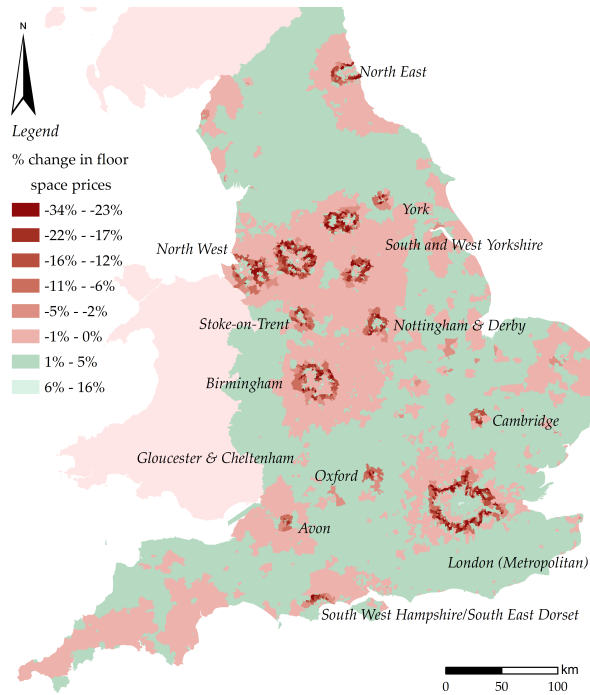
FIGURE D2 – ABSOLUTE CHANGES IN DENSITY WITH AND WITHOUT GREENBELTS

Notes: I estimate the effects using parameters estimated in column (6) in Table 5. I consider the scenario where all greenbelts are removed.

available for development (see Figure D3a); the effects can be large and up to 34%. Wage effects, reported in Figure D3b, are not so clear. In line with the maps reported earlier, I observe small wage decreases in and around London. The latter may be a result of firms moving into former greenbelt areas (Figure D3c). Density increases may be large and up to 50%. I further find that people move into former greenbelt areas as well as inner cities. The latter offer high amenity levels, which are at the same time unaffected by the removal of part of greenbelts.

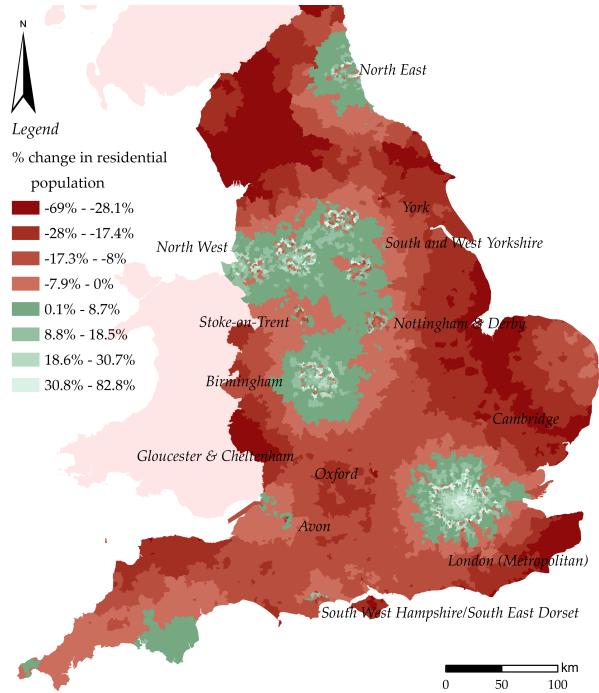
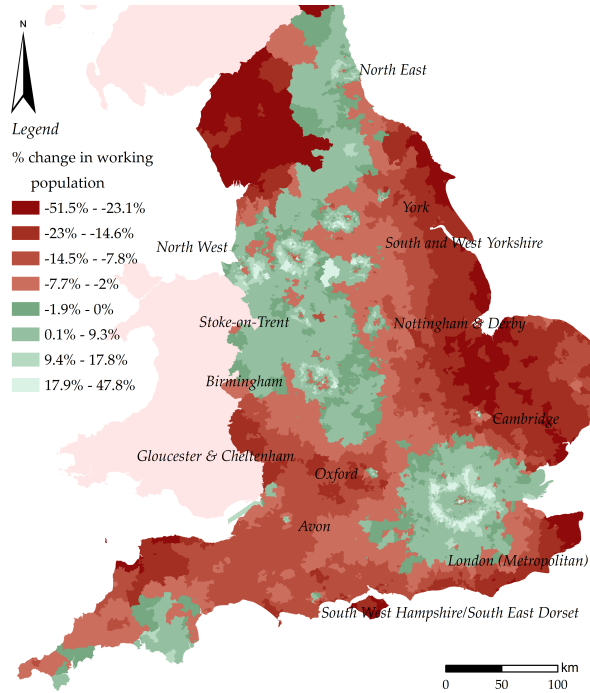
D.5 Local effects – counterfactual greenbelts

In Figure D4, I study the local effects of the third counterfactual experiment. Generally, one observes price increases in areas that are now designated as greenbelts, while price reductions in greenbelts that are not defined as ‘counterfactual’ greenbelts (such as Avon, Cambridge, and Gloucester and Cheltenham) (Figure D4a). One further observed changes in working population, where London in particular seems to benefit from the counterfactual greenbelt policy because the counterfactual greenbelt is actually somewhat smaller compared to the existing greenbelt (Figure D4c). This is confirmed in Figure D4d, where I observe that residents move to Greater London.



(A) FLOOR SPACE PRICES, p_i

(B) WAGE NET OF COMMUTING COSTS, $\mathbb{E}[\tilde{w}_i]$

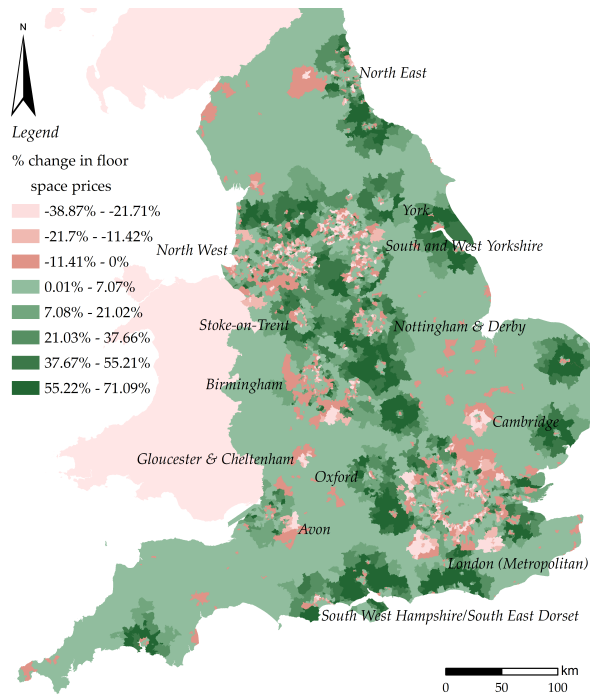


(C) WORKERS, H_{Mi}

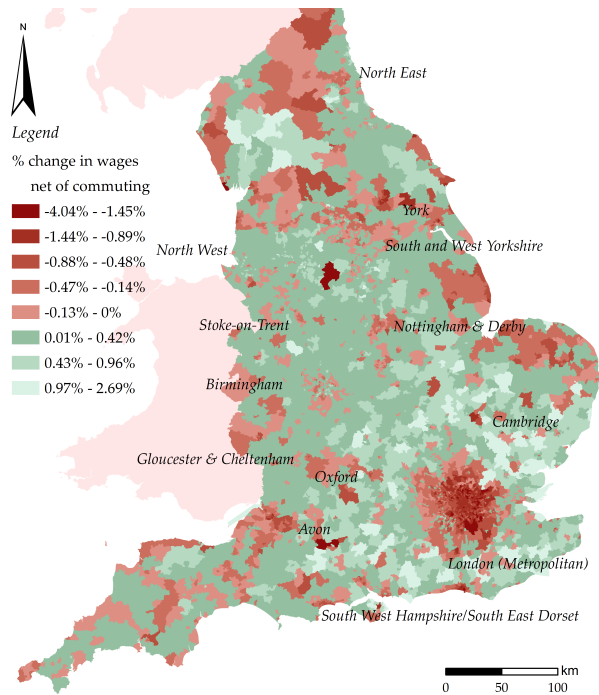
(D) RESIDENTIAL POPULATION, H_{Ri}

FIGURE D3 – COUNTERFACTUAL 1: 10% REDUCTION IN GREENBELT LAND

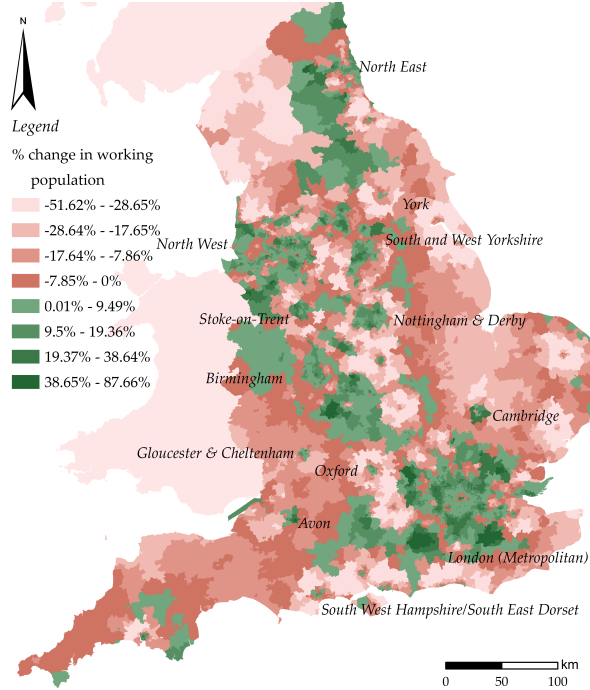
Notes: I estimate the effects using parameters estimated in column (6) in Table 5. I consider the scenario where 10% of greenbelt land is removed from the areas that are closest to the inner greenbelt border.



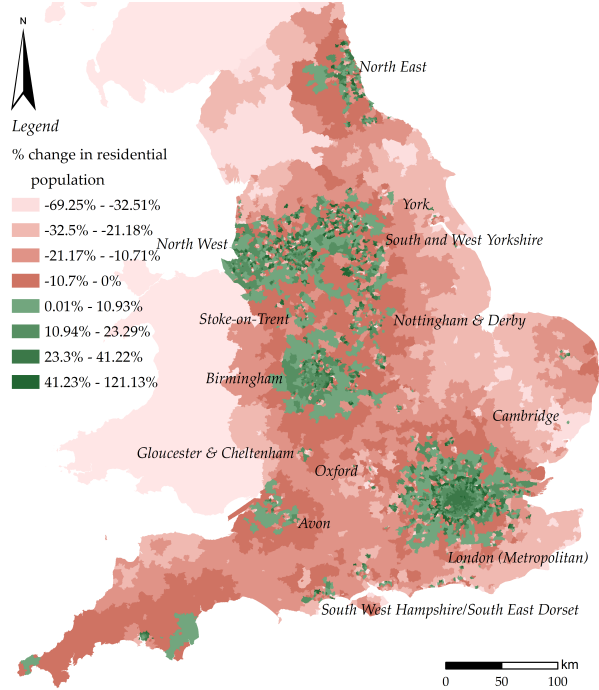
(A) FLOOR SPACE PRICES, p_i



(B) WAGE NET OF COMMUTING COSTS, $E[\tilde{w}_i]$



(C) WORKERS, H_{Mi}



(D) RESIDENTIAL POPULATION, H_{Ri}

FIGURE D4 – COUNTERFACTUAL 3: COUNTERFACTUAL GREENBELTS

Notes: I estimate the effects using parameters estimated in column (6) in Table 5. I consider the scenario where counterfactual greenbelt land is designated to be in line with Figure D1.

TABLE D1 – ROBUSTNESS OF COUNTERFACTUAL ANALYSES

| | Scenario 1: 10% reduction in greenbelts | Scenario 2: No greenbelts | Scenario 3: Counterfactual greenbelts |
|--|--|------------------------------|--|
| | (1) | (2) | (3) |
| PANEL A: Identification of parameters using all MSOAs | | | |
| Change in expected utility (<i>in %</i>) | -0.016 | -0.486 | 0.066 |
| Change in overall land rents (<i>in %</i>) | 0.953 | -1.247 | 3.448 |
| Change in output (<i>in %</i>) | -0.287 | 0.389 | -1.043 |
| Change in average floor space price (<i>in %</i>) | -1.698 | -5.084 | 3.597 |
| Change in average wages net of commuting (<i>in %</i>) | -0.111 | -0.112 | -0.131 |
| Change in total developable land (<i>in %</i>) | 4.925 | 26.953 | -7.291 |
| Change in total floor space (<i>in %</i>) | 0.910 | 3.933 | -2.536 |
| PANEL B: Greenbelts in 1973 | | | |
| Change in expected utility (<i>in %</i>) | -0.029 | -0.596 | 0.062 |
| Change in overall land rents (<i>in %</i>) | -0.181 | -4.192 | 4.561 |
| Change in output (<i>in %</i>) | -0.049 | 1.432 | -1.303 |
| Change in average floor space price (<i>in %</i>) | -2.780 | -8.753 | 9.314 |
| Change in average wages net of commuting (<i>in %</i>) | -0.111 | -0.104 | -0.147 |
| Change in total developable land (<i>in %</i>) | 12.031 | 81.466 | -2.275 |
| Change in total floor space (<i>in %</i>) | 2.071 | 9.004 | -3.654 |
| PANEL C: Spillovers from literature, $\{\lambda = 0, \psi = 0, \gamma = 0.02, \delta = 0.04\}$ | | | |
| Change in output (<i>in %</i>) | 1.704 | 2.991 | 0.650 |
| Change in expected utility (<i>in %</i>) | -0.386 | -1.534 | 0.215 |
| Change in overall land rents (<i>in %</i>) | 5.222 | 2.393 | 8.936 |
| Change in average floor space price (<i>in %</i>) | -1.686 | -6.684 | 7.298 |
| Change in average wages net of commuting (<i>in %</i>) | -0.277 | -0.287 | -0.323 |
| Change in total developable land (<i>in %</i>) | 8.804 | 54.480 | -5.517 |
| Change in total floor space (<i>in %</i>) | 2.825 | 8.330 | -1.922 |

Notes: In Panel A, I take the parameter values estimated in column (4), Table 5. In Panel B, I take the parameter values estimated in column (5), Table 5. Panel C takes the parameters estimated in column (6), Table 5, but adjusts the parameters related to spillovers.

D.6 Robustness

D.6.1 Aggregate effects

Here I provide several sensitivity analyses on the aggregate effects of the counterfactuals. Table D1 reports the results.

In Panel A I use parameters when using all observations to identify the model's parameters, implying that I use the parameters reported in column (4) in Table 5. The findings confirm that greenbelt land generally increases welfare, as removing greenbelt land generally harms workers' welfare, despite an overall increase in land rents. Interestingly, I do not find that increasing greenbelt land (column (3)) does benefit workers materially because floor space will inevitably become more expensive.

TABLE D2 – CORRELATIONS BETWEEN CHANGES IN COUNTERFACTUAL VALUES

| | <i>All areas</i> | <i>Greenbelts in 1973</i> | <i>No amenities</i> | <i>No spillovers</i> | <i>Low commuting costs</i> | <i>High commuting costs</i> |
|----------------------------------|----------------------|-------------------------------|-------------------------|--------------------------|--------------------------------|---------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Change in working population | 0.9556 | 0.9949 | 0.6869 | 0.9387 | 0.7488 | 0.7664 |
| Change in residential population | 0.9742 | 0.9798 | 0.7018 | 0.9485 | 0.0335 | 0.5035 |
| Change in prices | 0.9993 | 0.9998 | 0.9985 | 0.9994 | 0.9959 | 0.9972 |
| Changes in net wages | 0.9880 | 0.9944 | 0.8077 | 0.9301 | 0.5552 | 0.0636 |

Notes: I report Pearson correlation coefficients of the changes in the respective variables with respect to the no greenbelt scenario of which the results are reported in Figure 4. Recall that the baseline coefficients are used from column (6), Table 5. In column (1), I take the parameter values estimated in column (4), Table 5. In column (2), I take the parameter values estimated in column (5), Table 5. Columns (3) and (4) take the parameters estimated in column (6), Table 5, but adjusts the parameters related to amenities or spillovers.

In Panel B of Table D1, I employ parameters identified using observations in 1973 greenbelts (see column (5), Table 5), but choose $\mu = 0.7$, because the value estimated in this specification is unlikely to apply to the whole of England. In doing that, I again find that workers' utility is affected negatively by the removal of greenbelts.

Panel C indicates robustness with respect to the parameters of agglomeration economies. Because one may be concerned that the estimation of spillover parameters is plagued by endogeneity, I borrow those from the literature implying that $\gamma = 0.02$ and $\delta = 0.04$ (following Combes et al., 2008; Rosenthal and Strange, 2008). Moreover, I set $\lambda = \psi = 0$. It is then shown that this hardly affects the results.

D.6.2 Local effects

In Table D2, I show that the predicted local effects of greenbelts are similar across different specifications. To calculate the bivariate correlations between the predicted percentage effects of different specifications, I take the values observed in Figure 4 as the baseline. Hence, I estimate parameters using areas within 1km of greenbelt borders and consider the 'no greenbelt' scenario. I show that local patterns are very similar if I estimate parameters using all MSOAs, as correlations are all above 0.95.

When using the greenbelts in 1973 to estimate the parameters in column (2), I set $\mu = 0.7$, as the value that is estimated (*i.e.* $\hat{\mu} = 0.56$) is unlikely to apply to the whole of England. The bivariate correlations with predicted local effects are again very high.

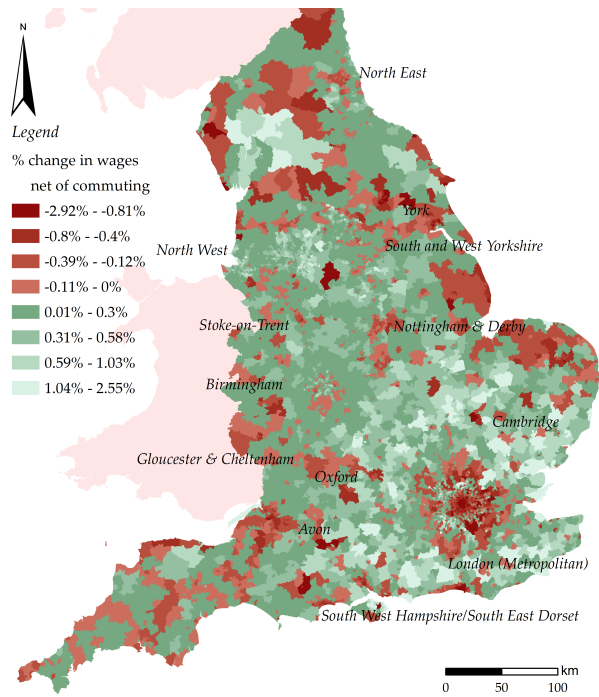
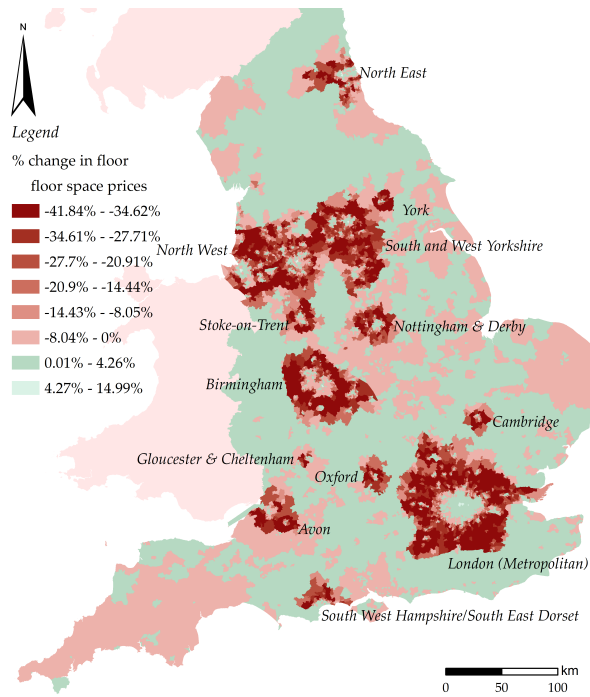
If I set amenities to zero, I find that correlations with the baseline counterfactual values are considerably lower (column (3)). This is not surprising given that greenbelt amenities play

a pivotal role in explaining the local effects. I show in Figure D5 the local effects of the ‘no greenbelt’ scenario but assuming that $\zeta = 0$. Because in the absence of greenbelt amenities, locations do not become less attractive when greenbelts are abandoned, I find that people move most notably to former greenbelt areas where there is a more abundant availability of cheap floor space. Together with the finding that without greenbelt amenities overall welfare will *increase* when removing greenbelts, greenbelt amenities play a pivotal role in understanding why greenbelts generate positive welfare effects.

The bivariate correlations with the scenario without spillovers are also very high (see column (4), Table D2), confirming that the main results can be explained mostly by the presence of greenbelt amenities and the parameter capturing the share of capital in construction.

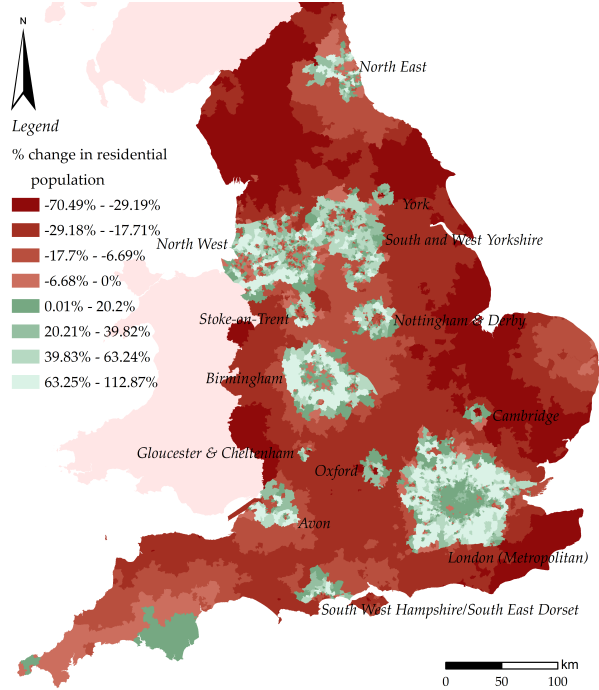
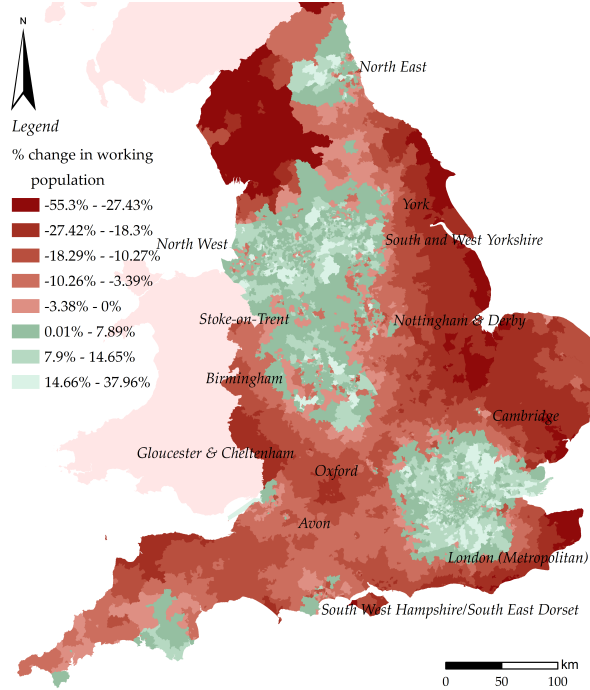
In columns (5) and (6), I change commuting costs. Earlier, in Figure 2c, I showed that aggregate effects were hardly influenced by commuting costs. However, the local effects of removing greenbelts are quite different, as can be seen from the low correlations. In particular, the pattern of changes in the residential population H_{Ri} is quite different. To investigate this further, I plot changes in the residential population for low and high commuting costs in Figure D6.

Figure D6a shows that with low commuting costs, the effects of removing greenbelts are much more spread out, as residents can now commute long distances without incurring high costs. Particularly striking is the exodus of London. With low commuting costs, residents can avoid high floor space prices by living much further away from work. With high commuting costs, the changes in residential population are much more contained to the areas that were formerly designated as greenbelt land (see Figure D6b).



(A) FLOOR SPACE PRICES, p_i

(B) WAGE NET OF COMMUTING COSTS, $E[\tilde{w}_i]$

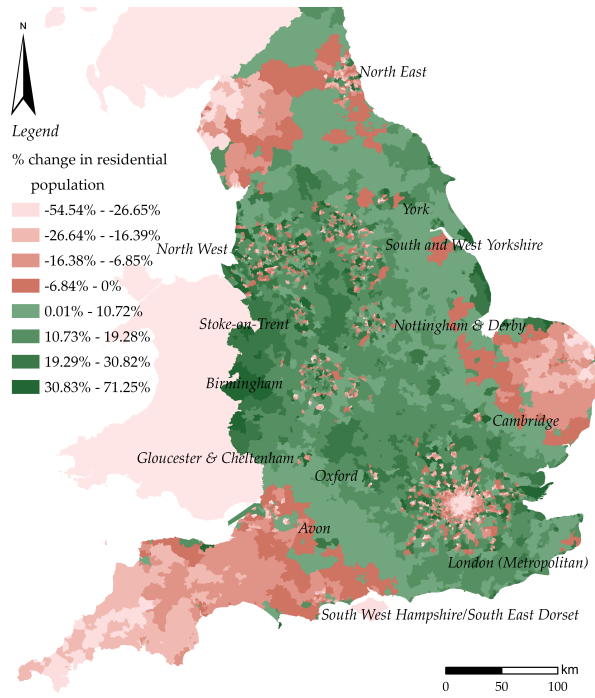


(C) WORKERS, H_{Mi}

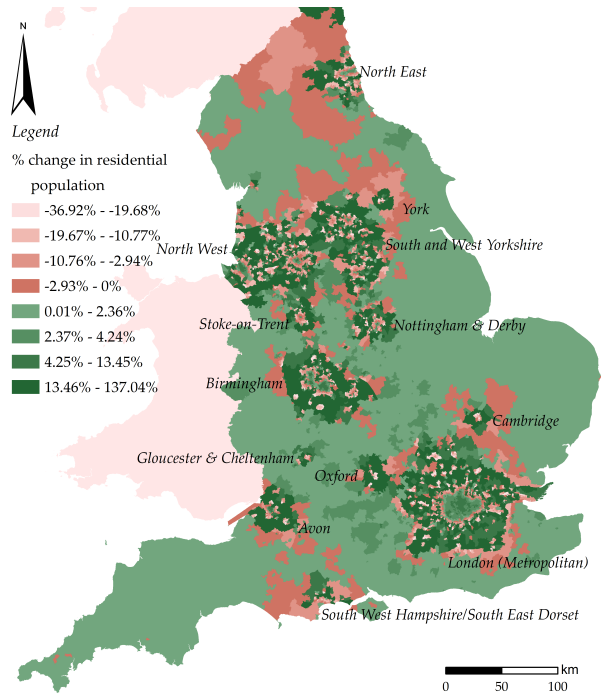
(D) RESIDENTIAL POPULATION, H_{Ri}

FIGURE D5 – COUNTERFACTUAL 2: NO GREENBELTS WITH $\zeta = 0$

Notes: I estimate the effects using parameters estimated in column (6) in Table 5, but where the greenbelt amenity effect is put to zero so $\zeta = 0$. I consider the scenario where all greenbelts are removed.



(A) LOW COMMUTING COSTS, H_{Ri}



(B) HIGH COMMUTING COSTS, H_{Ri}

FIGURE D6 – COUNTERFACTUAL 2: NO GREENBELTS WITH LOW AND HIGH COMMUTING COSTS

Notes: I estimate the effects using parameters estimated in column (6) in Table 5, but I respectively divide or multiply the commuting parameters $\{\kappa_{id}, \kappa_{iw}, \kappa_{ofh}\}$ by 2.5 for low and high commuting costs. I consider the scenario where all greenbelts are removed.