Structural Change, Land Use and Urban Expansion: Online Appendix

Nicolas Coeurdacier, Florian Oswald and Marc Teignier

September 28, 2021

Contents

A Introduction and Notation **2**

B Data **2**
B.1 Agricultural Land Use . . . 2
B.2 Sectoral employment . . . 6
B.3 Sectoral National Accounts and Prices . . . 7
B.4 Sectoral Productivities . . . 10
B.5 Consumption expenditures . . . 12
B.6 Land and Housing Wealth . . . 14
B.7 Urban Area and Population Measurement - Manual and GHLS Data . . . 15
B.7.1 Manual Urban Area Measurements 1870 and 1950 . . . 15
B.7.2 Manual Population Measurements 1870 and 1950 . . . 18
B.7.3 Automatic Area and Population Measurement via GHSL . . . 23
B.7.4 Density Measurement Results . . . 25
B.7.5 Consistency of Area Measures across Methods and Sources . . . 29
B.7.6 GHSL `cutoff` Parameter Sensitivity . . . 29
B.8 CORINE Land Cover (CLC) 2018 Data . . . 33
B.9 Individual Commuting Data . . . 35
B.9.1 ENL: Enquête National du Logement . . . 35
B.9.2 DADS: Déclaration annuelle des données sociales . . . 38
B.9.3 Historical commuting speed in Paris . . . 42

C Quantitative Model, Numerical Solution and Estimation **49**
C.1 Extensions to the baseline theory . . . 49
C.1.1 CES production in the rural sector . . . 49
C.1.2 Commuting Costs . . . 50
C.1.3 Location-specific housing supply conditions . . . 51
C.1.4 Single City On Circular Surface . . . 53
C.1.5 Dynamic Model . . . 53
C.1.6 Numerical Integration over Circular City . . . 55
C.2 Numerical Model Solution and Parameter Estimation . . . 55
C.2.1 Timing and Productivity Series Smoothing . . . 56

C.2.2 Single City Model Solution . . . 57
C.2.3 Starting Values . . . 57
C.2.4 Estimation Procedure . . . 58
C.2.5 Estimation Results . . . 58

D Sensitivity Analysis and Extensions **60**
D.1 Estimation Results with $\epsilon(0) = 2.5$ . . . 60
D.2 Sensitivity with constant $\epsilon = 3$ . . . 60
D.3 Sensitivity w.r.t the elasticity of substitution between land and labor . . . 62
D.4 Agglomeration and Congestion . . . 63
D.5 Commuting Distance and Residence Location . . . 66
D.6 Multiple City Model . . . 68
D.6.1 Multiple City Model: Setup . . . 68
D.6.2 Multiple City Model: Numerical Solution . . . 70
D.6.3 Multiple City Model: Results . . . 70

A Introduction and Notation

This is the online appendix for *Structural Change, Land Use and Urban Expansion* by Coeurdacier, Oswald and Teignier. In this document we number sections alphabetically (A, B, ...) and equations with roman numbers (I, II, ...). Standard latin numbering (1, 2, ...) refers to the main text.

B Data

B.1 Agricultural Land Use

Data sources and definitions. Data for the land used in agriculture are available in various secondary sources based on the French Agricultural Statistics (Statistique Agricole). We checked the consistency of the measures across the different sources.

The variable of interest is the area of land used for agriculture (SAU, for 'Surface Agricole Utilisée'). It is important to note that it includes land that is cultivated but excludes all land that is not (woods and forests, rocky land unfit for agriculture, mountains, swamps...).

Post World War 2 (WW2), data for the SAU are provided by the Ministry of Agriculture (data available in Desriers (2007) until 2000 by decade and available on annual basis since 2000 on the website of the Ministry (Agreste)).

Before WW2, agricultural statistics on land use are also available but on a very irregular basis.[1]

[1] In the 19th century, starting 1840, France aimed at organizing every decade a detailed data collection of agricultural statistics (Agricultural Census, 'Statistique Agricole'). See for instance description in Fléchey (1898) and Augé-Laribé (1945). A comparison across years during the 19th century is available in the report of the 1892 Census. Before 1840, Lavoisier provides the first measure of land use in France, in 1790, as described in.

Through a search across various sources, we compute a measure for the SAU from the first Agricultural Census in 1840 until today. It is worth noting that one must be cautious with such a measure before WW2 in the earlier periods. While it is quite clear that the share of land used in agriculture fell over the whole period, the variations throughout the 19th century (before the 1882 Census) must be taken with caution.

The main difficulty is to make the data presented in various sources comparable across years. First, woods and forest, accounting for 15-20% of French land in the 19th century (and more than 25% today) were initially included in the cultivated agricultural land. We made sure to exclude them from the SAU consistently over the whole time period considered. A second difficulty arises because the French territory varied since 1790: some variations being due to measurement, some due to the loss (or addition) of some parts of France — loss of Alsace and Moselle after the war of 1870 until 1918 and addition of Savoye and Comté de Nice in 1860 (see discussion in Augé-Laribé (1945)). This makes the across-time comparison difficult, even though we show our measure of the SAU as a share of the French territory at the time. A third difficulty for the early periods (before 1882), detailed below, regards the treatment of pasture and grazing fields in a consistent way across years.

Period 1945-2015. Let us start with the most recent period where the data are arguably of better quality and coherent across time and then present our measures going further back in time. Since 1945, the land used in agriculture has clearly been clearly falling over the period 1950-2015: while land used for agriculture accounted for 62% of total French land post-WW2, this numbers falls to 52% in 2015.

Interwar Period. In between the world wars, we could find measures for the years 1929 and 1937. Two slightly different measures are available for 1929: one in Toutain (1993) and one in Mauco (1937). We take the average between the two, a SAU of 34 483 thousands of ha in 1929. A measure, very similar to 1929, is available in Augé-Laribé (1945) for 1937: 34 207 thousands of ha and 33 285 if one excludes Alsace-Moselle for comparison with earlier periods. This corresponds to about 62% of the French territory.[2]

Nineteenth century. Before World War 1, we have measures in 1882 and 1892 (Mauguin (1890), Fléchey (1898), Hitier (1899) and Toutain (1993) for further details). Both measures are consistent across sources, including the main results of the 1892 Agricultural Census as a more primary source.[3]

[2] Mauco (1937) compares to the 1892 value and find very similar numbers than ours once woods are excluded from his measurement. Augé-Laribé (1945) compares to the 1882 value and the measure given for 1882 is also consistent with our data.

[3] Statistique Agricole de la France: Résultats généraux de l'Enquête Décennale de 1892. The online archives are available at: https://gallica.bnf.fr/ark:/12148/bpt6k855121k/f1.item

This gives an SAU of 34 882 thousands of ha in 1882 and 34 720 in 1892—slightly higher than the values in between the wars despite a smaller French territory. Figure I provides the details of the measurement for the 1892 Agricultural Census.[4]

The measurement in 1840 constitutes our first observation. However, in the 1840 data, an important difficulty is the treatment of meadows, pasture and grazing fields (prés, herbages, pâturages,...). These should be included in the SAU to the extent that the land is used for agricultural purposes (feeding cattle). As grazing fields and meadows account for a large share of French agricultural land (up to 11% in 1892), their inclusion (or not) in the cultivated part of agricultural land (SAU) matters. However, in 1840, a significant share of grazing fields ('pâturages', 'pâtis communaux/vaines pâtures') is excluded from the SAU. The non-cultivated part of agricultural land thus appears to be a much larger measured area than in all subsequent years.[5] As discussed in the results of the 1892 Agricultural Census, comparison across years is difficult due to the reallocation of grazing fields into the cultivated part of French land over the period 1840-1880. This reallocation is quite artificial—mostly a statistical artefact coming from the earlier exclusion of common pasture. Excluding entirely the measured non-cultivated part from the SAU in 1840 gives thus a lower bound, while including it entirely to account for all grazing fields gives an upper bound. To solve this issue, Toutain (1993) provides an estimate of agricultural land in 1840, in between these two values, of 35 500 thousands of ha. While this is just a matter of definition and any solution is somehow arbitrary, we proceed in a similar fashion as Toutain (1993) and assume that the grazing fields later reallocated in the cultivated part are part of the SAU in 1840. This gives a land use in agriculture of 35 497 thousands of ha in 1840—a very similar number to Toutain (1993). Proceeding exactly in the same way for the year 1862 gives an SAU of 36 088 ha—a higher value but for a larger territory. Both values correspond to about two thirds of French land used in agriculture.

The measured cultivated agricultural land (as a share of French territory) over the period 1840-2015 is summarized in Figure II.

Pre-1800. Lastly, Lavoisier provided in 1790 the very first measure of French agricultural land before the creation of the Agricultural Census. Comparison of Lavoisier's measurement with the later 'Statistiques Agricoles' is however difficult. Like for the later measurements, a large fraction of land ('vaines patûres') includes grazing fields as well as rocky land and moor unfit for agriculture (see

[4]Comparison of land use as a share of total French land across the 19th century is also available in the report of the 1892 Census.

[5]As shown in Figure I, in 1892, the non-cultivated part includes moor and rocky land arguably unfit for agriculture, accounting for about 11% of French land. The corresponding non-cultivated part in 1840 accounts for 17% of French land as it includes a significant share of grazing fields.

RÉSUMÉ DES CULTURES.

A. — SITUATION EN 1892.

1. TERRITOIRE.

Nous donnons ci-après, par grandes catégories, la répartition du territoire de la France, telle qu'elle résulte des relevés opérés en 1892 :

CATÉGORIES DU TERRITOIRE.			SUPERFICIES.	RÉPARTITION et PROPORTION.
			hectares.	p. 100.
1° TERRITOIRE AGRICOLE.				
Superficie cultivée.	Terres labourables.	Céréales	14,827,085	28.06
		Grains autres que les céréales	319,705	0.60
		Pommes de terre	1,474,144	2.68
		Autres tubercules et racines pour l'alimentation humaine	128,238	0.24
		Cultures industrielles	531,508	1.00
		Cultures fourragères (1)	4,736,394	9.08
		Jardins potagers et maraîchers	386,827	0.73
		Jachères	3,367,518	6.37
		Terres labourables	25,771,419	48.76
	Vignes		1,800,489	3.40
	Prés naturels		4,402,836	8.33
	Herbages paturés (2)		1,810,608	3.42
	Bois et forêts		9,521,568	18.03
	Cultures arborescentes, etc.		934,800	1.76
		Cultures permanentes non assolées.	18,470,301	34.94
		Totaux de la superficie cultivée	44,241,720	83.70
Superficie non cultivée.	Landes, pâtis, bruyères		3,898,530	7.37
	Terrains rocheux et montagneux, incultes		1,972,994	3.73
	Terrains marécageux		316,373	0.60
	Tourbières		38,292	0.07
		Totaux de la superficie non cultivée.	6,226,189	11.77
		TOTAUX DU TERRITOIRE AGRICOLE	50,467,909	95.47
		2° TERRITOIRE NON AGRICOLE	2,389,290	4.53
		Totaux généraux du Territoire	52,857,199	100.00

(1) Non compris les cultures dérobées.
(2) Y compris les herbages alpestres.

Figure I: Land Use in the 1892 Recensement Agricole.

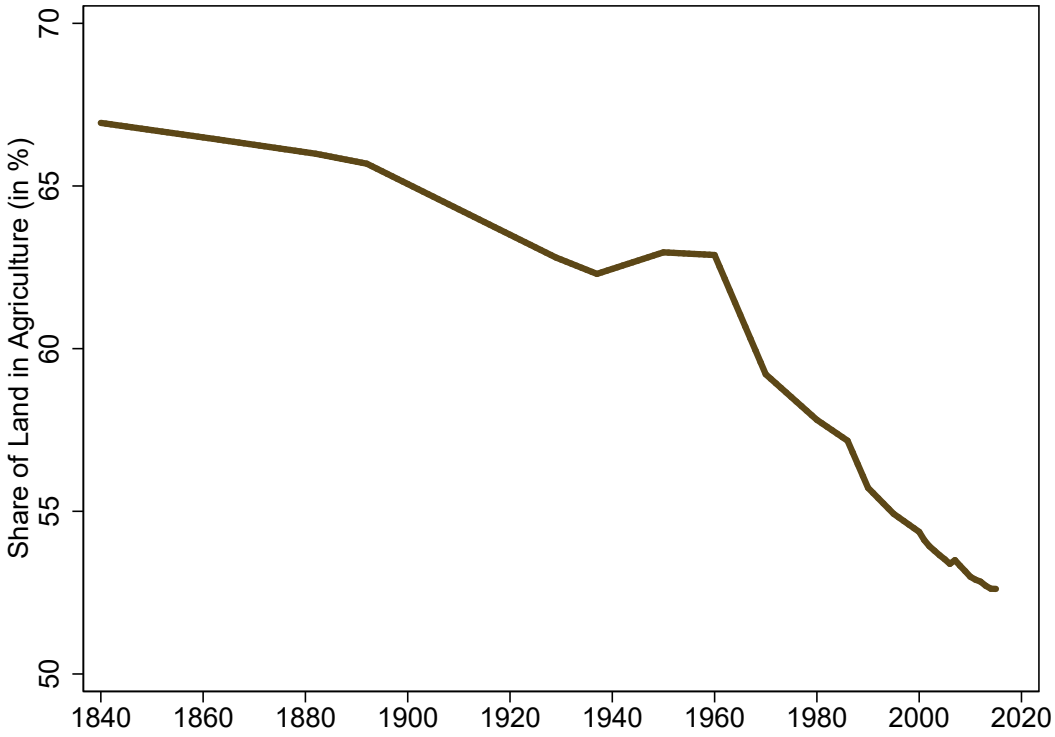


Figure II: Shares of Land used in Agriculture (1840-2015).

Mauguin (1890) for an attempt to compare with the 1882 Census). Excluding woods but including the 'vaines patûres' (common pasture) in 1790 gives a surface of almost 40 000 thousands of ha. Excluding all the 'vaines patûres' provides a lower bound of about 31 000 thousands of ha. This gives a reasonable but fairly wide bracket for the total land used for agriculture. Assuming that the non-cultivated part due to rocky land is comparable to the later measures gives a SAU in 1790 around 34 000 thousands of ha—comparable to the later years (on a smaller territory)—about 65% of French land measured at the time. While this measure should be taken with great caution, it is nevertheless comforting that we find a value in same ballpark as our first measure in 1840 using the Agricultural Census.

B.2 Sectoral employment

Sources. Data on employment are available in three different sources covering different time periods: Marchand and Thélot (1991) ('Deux siècles de travail en France') for the period 1806-1990; Herrendorf et al. (2014) for the period 1856-2006; OECD for the period 1950-2018. When overlapping, the different sources are largely consistent with each other.[6] We use the three sources allowing to span the entire 1806-2018 period. For the pre-WW2 period, data available in Marchand and Thélot (1991)

[6] Marchand and Thélot (1991) gives a slightly lower share of employment in agriculture in the first half of the 20th century relative to Herrendorf et al. (2014). Our results do not depend on the use of one series or the other.

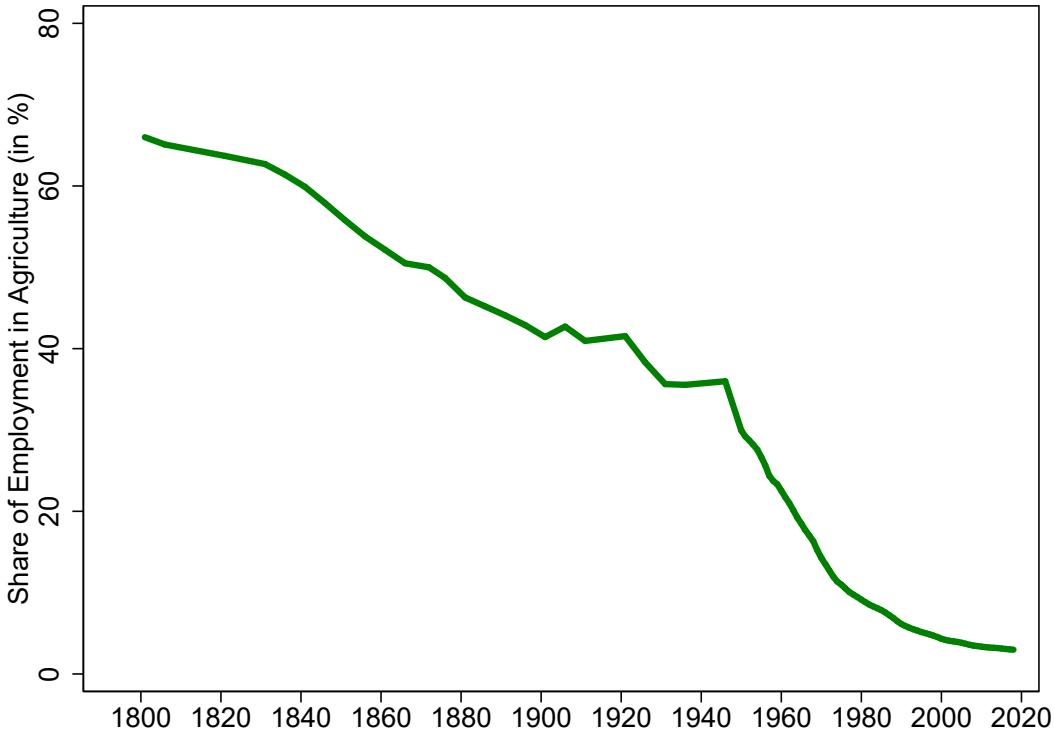


Figure III: Shares of Employment in Agriculture (1806-2018).

and Herrendorf et al. (2014) are on an irregular basis—typically one or two observations per decade (corresponding to Census years). Annual data are available from 1950 onwards.

Over the nineteenth century (until 1901), we use the data from Marchand and Thélot (1991) as the series goes further back in time. Over the period 1901-1950, we use the data from Herrendorf et al. (2014). Over the period 1950-2018, we use data provided by the OECD on an annual basis, where the measure of employment is expressed in full-time equivalent.

Share of employment in agriculture. This gives the share of employment in agriculture over the entire period (1806-2018) in Figure III. Data are linearly interpolated in between two values when data are not available on an annual basis (pre-1950). It starts with about 2/3 of the employment in agriculture in 1806 and falls progressively to 3% in 2018. One can notice the acceleration in the process of reallocation post WW2. In the matter of three decades, the employment share in agriculture went from 36% in 1946 to 10% in 1976.

B.3 Sectoral National Accounts and Prices

Sources. Data on value added at the sectoral level together with aggregate value added (GDP) at current prices are available in two different sources covering different time periods. Historical national accounts from Toutain and Marczewski (1987) are used to cover the period 1815-1938.

They are directly available at the Groningen Growth and Development Centre (Historical National Accounts Database, http://www.ggdc.net/).

Post WW2, INSEE provides sectoral value added at current prices for the period 1949-2019. For both series, we use agricultural value added and aggregate GDP at current prices. Using both sources covers the entire period 1815-2019. The series are interrupted at war times: observations are missing for the periods 1914-1919 and 1939-1948.

Toutain and Marczewski (1987) also provides volume indices for GDP in agriculture and for aggregate GDP over the period 1815-1982 (also available Groningen Growth and Development Centre). The series for agricultural volumes is extended in Toutain (1993) until 1990. Together with the value added at current prices, these series will be used to compute an agricultural price deflator and a GDP deflator.

Sources for sectoral prices. Data on agricultural producer prices are available over the period 1815-2019 using two different data sources: one derived from the national accounts in value added and volume from Toutain (1987, 1993) and one from INSEE post-1949.

Using Toutain (1993), we compute a price index of agricultural goods using the value added in agriculture divided by the production volume index in agriculture (period 1815-1990). Post WW2, INSEE directly provides a producer price index for agricultural goods (Indice des prix agricoles à la production, IPPAP)—the series can be retropolated back to 1949.[7] These two series will be used to construct a price index for agriculture goods over the period 1815-2019 (with interruptions at war times). Similarly, a GDP deflator over the period 1815-1960 can be computed using GDP at current prices and a GDP volume index from Toutain and Marczewski (1987). Post-1960, we use the GDP deflator from the World Bank.[8] The price index for agricultural products and the GDP-deflator are both normalized to 100 in 1949.

Relative price for agricultural goods. Using the computed historical time-series for the agricultural producer price index and the GDP-deflator, one can take the ratio of the two series to shed some lights on the evolution of the relative prices of agricultural goods. The series for the relative price based on Toutain production data (solid green) over the period 1815-1990 and the INSEE producer price (solid black) starting 1949 are shown in Figure IV. While the relative price of agricultural goods appears fairly stable until 1910, it exhibits later a clear downward trend over the twentieth century.

[7]The IPPAP series is the 'Base 2000 rétropolée' available in Insee Méthodes 114 (INSEE (2006)). Until 1970, the retropolated series from INSEE excludes fruits and vegetables. The series including fruits and vegetables and the one excluding them are almost identical when both are available.

[8]We checked consistency with the consumer price index available over the period 1820-2015 (INSEE). Inflation is very similar in both series.

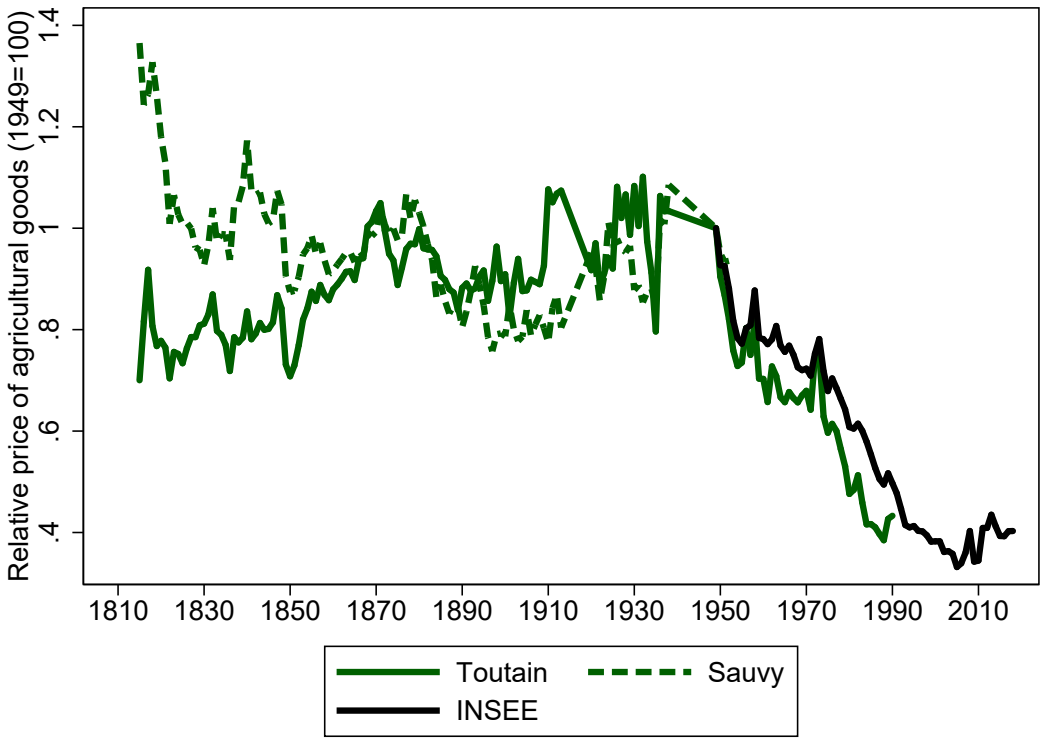


Figure IV: Relative prices of agricultural goods, 1949=100 (1815-2019).

Both series show a similar trend post WW2.

Our baseline price index of agricultural goods (denoted P_{agri}) uses the series computed using the national accounts of Toutain prior to WW2 (1815-1938) and the agricultural producer prices by INSEE post WW2 (1949-2019). The two series are linked by the same normalization to 100 in 1949. The final series for P_{agri} is only interrupted during the wars.

The model counterpart of our data is the relative price of rural/agricultural goods over the price of urban/non-agricultural goods. The latter is not observed but can be backed out using the GDP-deflator. Let us denote $P_{agri,t}$ the price index for agricultural goods at date t, $P_{non-agri,t}$ the price index for non-agricultural goods, and $P_{GDP,t}$ the GDP-deflator. The GDP-deflator can be written as

$$\frac{1}{P_{GDP,t}} = \frac{s_{agr,t}}{P_{agr,t}} + \frac{1 - s_{agr,t}}{P_{non-agr,t}}, \tag{I}$$

where $s_{agr,t}$ is the share in value-added of agricultural goods computed using historical national accounts. Since we observe in the data all the variables but $P_{non-agri,t}$, we can invert Eq. I to back out a price index for non-agricultural goods (urban goods including manufacturing and services),

$$P_{non-agri,t} = \left(\frac{1}{P_{GDP,t}} \frac{1}{1 - s_{agr,t}} - \frac{1}{P_{agr,t}} \frac{s_{agr,t}}{1 - s_{agr,t}} \right)^{-1}.$$

We are now equipped with a price index for agricultural goods, non-agricultural goods, and a GDP deflator over the period 1815-2019.

Sensitivity analysis for the price of agricultural goods. Before WW2, the Statistique Generale de France (the predecessor of INSEE), in particular thanks to the work of Alfred Sauvy, provides an alternative series for the price of agricultural goods: 'indice des prix de gros agricoles' which is constituted by a basket of 19 raw agricultural commodities (food related).[9] The series is retropolated back to 1810 by A. Sauvy (see Sauvy (1952)). This data includes some foreign commodities (e.g. English and US corn prices) and is in part computed using customs price data. For this reason, we use the price of agricultural goods computed using production data of Toutain pre WW2 as baseline. This said, the 'indice des prix de gros agricoles' still contains useful information regarding the price of agricultural goods in France before WW2. Comparison with the price computed using production data from Toutain indicates that the two series exhibit very similar patterns starting 1850. Prior to this date, the 'indice des prix de gros agricoles' from Sauvy (1952) exhibits a significant downward trend, while our baseline from Toutain stays roughly stable (see Figure IV).[10] Our baseline price series for agricultural goods uses the series based on Toutain for the period pre WW2. However, results are robust using data from Sauvy since our quantitative estimation starts in 1840 and both series roughly coincide over this time period.

B.4 Sectoral Productivities

Equipped with sectoral value added at current prices, sectoral price indices, sectoral employment and land use data, one can back out the sectoral productivities (in the agricultural and non-agricultural sector) that are the counterpart of the model (the θs) up to a constant of normalization. Our measure of land use in agriculture necessary to estimate rural productivity starts in 1840. Thus, we compute sectoral productivities for the period post 1840 and focus on the period 1840 until today for the quantitative analysis.

Urban Productivity. Let us start with the urban/non-agricultural sector. According to the model production function, $\theta_u = \frac{Y_u}{L_u}$. We observe the value added in the non-agricultural sector at current prices. Deflating this series by the constructed price index for non-agricultural goods gives Y_u. Dividing the latter variable by employment in the non-agricultural sector, $L_{non-agri,t}$, allows us to back out the empirical counterpart of $\theta_{u,t}$,

$$\theta_{u,t} = \frac{VA_{non-agri,t}}{P_{non-agri,t} L_{non-agri,t}}.$$

Due to the mere presence of a price index, this series is defined up to a multiplicative constant.

[9] Details about the index can be found in the 'Etudes spéciales' of the 'Bulletin de la Statistique générale de la France' in 1911. Available online at: https://gallica.bnf.fr/ark:/12148/bpt6k96205098/f73.image

[10] We also compare those series with the relative price of corn. While significantly more volatile, the latter is also fairly consistent with the other series. A period of volatile relative corn price but fairly constant on average until the early 20th century, followed by a downward trend. The downward trend is however more pronounced.

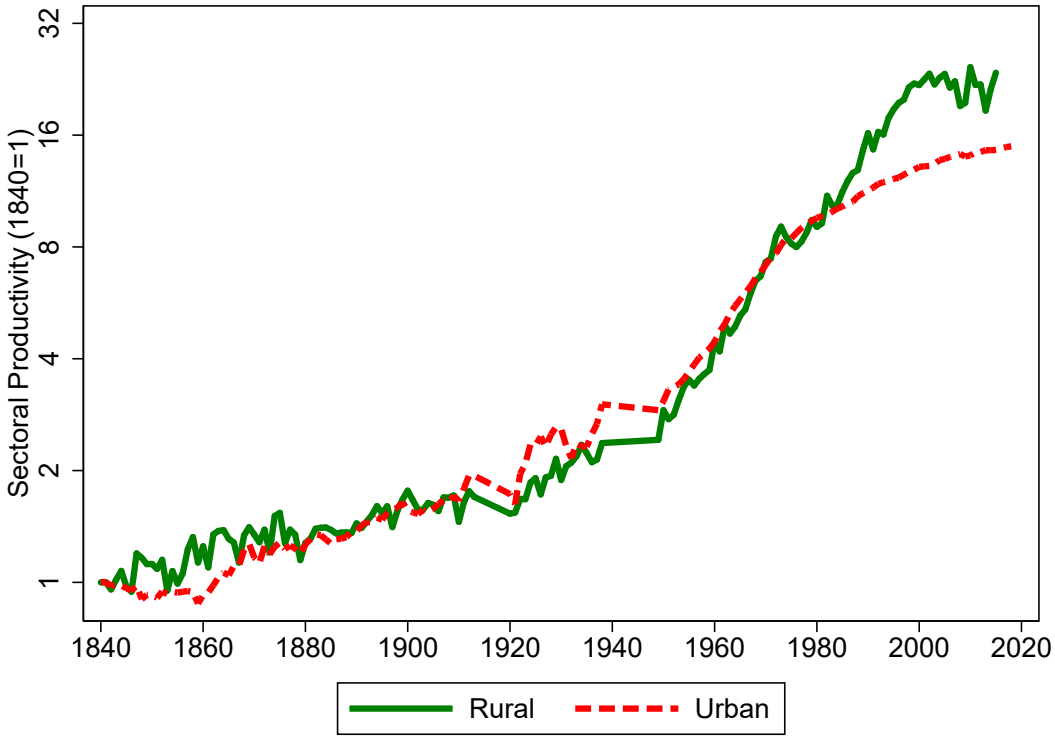


Figure V: Rural and Urban Productivity, 1840=1 (1840-2019).

We normalize $\theta_{u,t}$ to unity in the first period considered (1840). This gives the time-series for $\theta_{u,t}$ plotted in Figure V (dashed red line). This will be our baseline exogenous urban/non-agricultural productivity series. It is important to note that the measured urban labor productivity includes technological advances in the non-agricultural sector but also factor accumulation rising labor productivity (physical and human capital accumulation).

Rural Productivity. We proceed in a similar fashion to compute the model's counterpart of the rural productivity, $\theta_{r,t}$, with one important difference: the agricultural output per worker in the rural sector depends also on the land per worker available for agriculture,

$$\frac{Y_r}{L_r} = \theta_r \left(\alpha + (1-\alpha)(\frac{S_r}{L_r})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} = \theta_r F\left(\frac{S_r}{L_r}\right). \tag{II}$$

Thanks to the data on land use in agriculture, one can back out from the data the land per worker in agriculture at each date: it is simply the cultivated area (SAU) divided by employment in agriculture, $\frac{S_r}{L_r} = \frac{SAU}{L_{agri}}$. Using Eq. II, one can compute the rural productivity parameter, $\theta_{r,t}$, at each date,

$$\theta_{r,t} = \frac{VA_{non-agri,t}}{P_{non-agri,t} L_{non-agri,t}} \frac{1}{F(\frac{SAU_t}{L_{agri,t}})}.$$

With a unitary elasticity of substitution between land and labor ($\sigma = 1$), this gives,

$$\theta_{r,t} = \frac{VA_{agri,t}}{P_{agri,t} L_{agri,t}} \left(\frac{SAU_t}{L_{agri,t}}\right)^{\alpha-1}.$$

Due to the mere presence of a price index, this series is defined up to a multiplicative constant. Like $\theta_{u,t}$, we normalize $\theta_{r,t}$ to unity in the first period (1840). This gives the time-series for $\theta_{r,t}$ plotted in Figure V (solid green line). This will be our baseline exogenous rural/agricultural productivity shifters.

Comments. Comparing urban and rural productivity, one notices the important common component: this can be due to technological advances benefiting both sectors but also to physical and human capital accumulation, which increase labor productivity across the board. Focusing on the more sectoral specific component, it is visible that non-agricultural productivity grew faster from the late 19th century until WW2. Post WW2, agricultural productivity starts growing at a faster speed, catching-up with the non-agricultural one and eventually outpacing it. This is consistent with Bairoch's view that starting with the agricultural crisis in late nineteenth century, technological progress in the French agriculture is slow and delayed relative to other countries, before catching up post WW2. The period 1945-1985 period is more broadly characterized by a very fast technological progress in agriculture across developed countries (see Bairoch (1989)). A productivity slowdown is later observed in both sectors.

B.5 Consumption expenditures

Sources. Data on consumption expenditures are available using two different data sources. Pierre Villa provided data on consumption expenditures across 24 different categories of goods for the period 1896-1939.[11] INSEE provides data over the period 1959-2017 on personal consumption expenditures ('Consommation effective des ménages par fonction aux prix courants') across 12 broad categories (food, drinks, clothing, housing, transportation,...) and about 100 narrower categories. INSEE Data are from the Comptes nationaux (Base 2014).[12]

Expenditure shares. We compute expenditure shares on three broad categories: food/drinks, housing and the remaining goods. The expenditure share outside food, drinks and housing includes manufacturing goods and services. The expenditure share on food/drinks is computed by adding all the good categories corresponding to food and drinks consumption divided by aggregate household expenditures (for the pre and post WW2 data). However, it excludes consumption in restaurants that will enter the remaining category (urban goods). The housing expenditure shares include housing

[11] Data are publicly available thanks to the CEPII. For details and documentation, see http://gesd.free.fr/villadoc.pdf. See also Villa (1993).

[12] Over the period 1950-1958, the CREDOC was providing data on consumption expenditures across broad categories for French households. These data have not been made compatible with the INSEE data post-1959, when INSEE revised the methodology. Investigating data in reports by CREDOC provides some additional insights on consumption expenditure shares in the 1950s across broad categories. As expected, these shares are in between the ones computed using the data from Villa right before WW2 and the later national accounts data of INSEE.

related expenses: rents (effective and imputed), energy expenditures, some housing services (garbage, cleaning, repair, ...) but also housing equipment (furniture, tableware, household appliances...).[13]

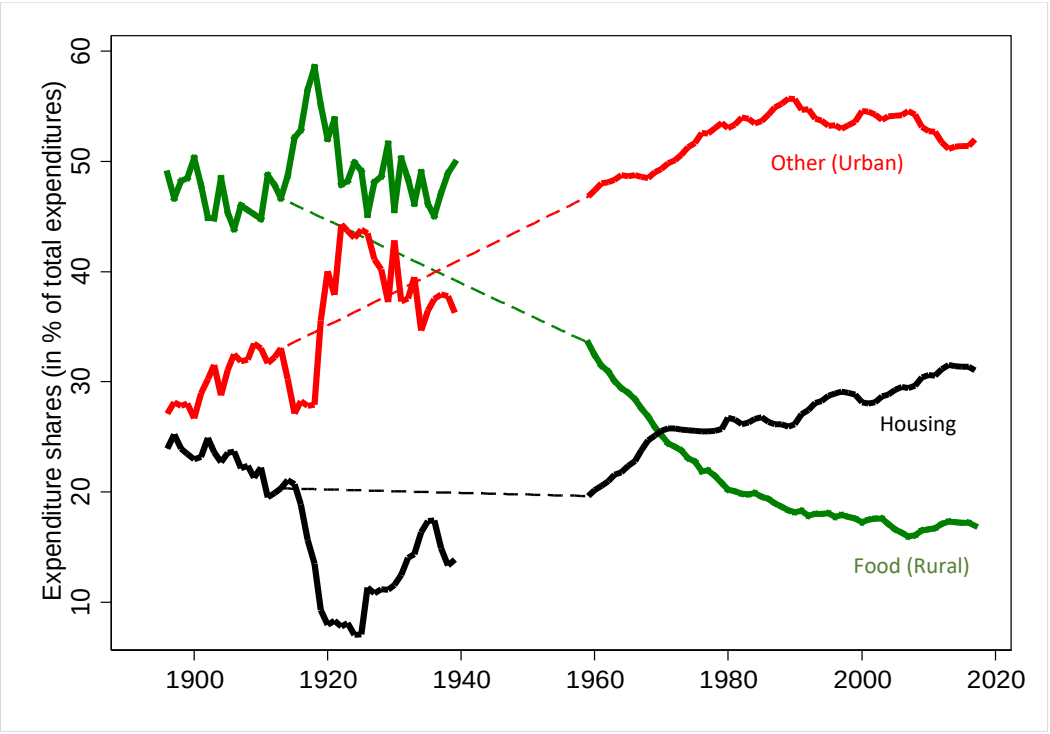


Figure VI: Spending Shares for Rural, Urban and Housing goods.

Notes: The observations around WW2 missing due to difficulties in data collection.

Data on expenditure shares across these three broad categories are shown in Figure VI. Comparing the initial periods in the late nineteenth century to today gives the following broad facts: the food share went down from almost 50% of expenditures to 17%; the housing share increased slightly to 23% to 31%; the share of expenditure on other goods increased as a consequence from 27% to more than 50%. This reallocation of expenditures away from rural goods towards housing and urban goods fits well with the process of structural transformation.

Rent control and the housing expenditure share. An important issue is the significant and persistent dip of the housing expenditure share starting at WW1. This evolution is largely due to the presence of rent controls that were put in place at the beginning of WW1 in France. As the French government wanted families to be able to afford their home during the war, it decreed that rents would be blocked (in nominal terms). As inflation picked up, this generated a large fall in real housing rents. As rents were very cheap, it freed up resources for households that could be spent on other goods (rural and urban). This is immediately visible on Figure VI, where the share of expenditures

[13]We include housing equipment as (partly) furnished/equipped houses/flats are quite common—even in the early 20th century. Small furnished flats/bedrooms were very common in large cities in the interwar period ('garnis'). However, excluding the latter category from housing expenses does not affect our results.

on housing went down from 21% in 1914 to less than 10% at the end of the war in 1919—other expenditure shares increasing simultaneously. While the measure was meant to be temporary, rent control lasted effectively during the whole interwar period despite various modification in the laws. It was eventually profoundly reformed post WW2 in 1948.[14] The reform of 1948 led to a sluggish adjustment of rents and it took some further years before one can reasonably argue that the rent control put in place in 1914 starts playing a more minor role.[15] Given this, our aim is to match the long-run evolution of spending shares while abstracting from the fluctuations in between 1914 and 1959 (first year of observation in the series provided by INSEE), as illustrated by the dashed lines on Figure VI.

B.6 Land and Housing Wealth

Land and housing wealth data is from Piketty and Zucman (2014), which can be obtained in the World Inequality Database (`https://wid.world/fr/accueil/`).

The data provide the value of agricultural land (as a share of national income) and the value of housing (as a share of national income) in France, roughly every ten years since 1810. The value of housing incorporates the value of land used for housing as well as the value of the capital stock used for housing (buildings and structure). To confront the data to our model, one needs to separate the value of land from the value of capital. Data on the share of land in housing is only available since 1979 for France (also available in the World Inequality Database). Due to lack of historical data on the share of land in housing, we assume a constant share over the whole period and take the average for the period 1979-2019. We find an average of 0.32 over the period 1979-2019. The value of urban/housing land is thus computed as 32% of the total value of housing. Note that this value of 0.32 is consistent with Combes et al. (2021) which computes a land share in housing of 0.35. It is also consistent with the model's predictions given the calibrated supply elasticities of housing (the model gives an average value of 0.3 for this period).

[14] Rents did increase in real terms during the interwar period. However, regulations still significantly limited the rent increases. The reform of 1948 still kept some housing with regulated cheap rents. Rents could be changed for new renters. Few housing with very cheap rents under the special regime of 1948 still subsists.

[15] Data from CREDOC in the early 1950s suggests a fairly housing spending share at that time–around 15%.

B.7 Urban Area and Population Measurement - Manual and GHLS Data

As explained in Section 2.2 in the main text, we consider the 100 most populated cities in the census of 1876 as our sample. We constrain this list to contain only cities which are still independent entities nowadays.[16] With the master list of cities in place, we proceed as follows to obtain two measures for each city: the extent of urban area in square kilometers, and population count. Depending on the period, we consider different data sources. The earliest measure uses the Carte d'Etat Major for urban area (1866) and the census for urban population counts (1876), while the second measure uses 1950 aerial photographs and the 1954 population census. Due to the lack of other data sources, we regard 1866 and 1876 we well as 1950 and 1954 as the same points in time, and we will refer to 1870 and 1950 for simplicity. In subsequent years, the Global Human Settlement Layer (GHSL) provides built up area and population data for the 1975, 1990, 2000 and 2015.

B.7.1 Manual Urban Area Measurements 1870 and 1950

We rely on georeferenced maps provided via https://www.geoportail.gouv.fr to take area measures of our cities. This website is run by the Institut national de l'information géographique et forestière (IGN) and offers a large variety of map layers and measurement tools (distance, area, etc). We use the layer *Carte d'Etat Major 1820-1866* (EM henceforth), *Photographies aériennes 1950-1965* or *Cartes 1950* (depending on which allows better classification), as well as contemporary *Photographies aériennes* to cross-check our measures with what the satellite measures will produce for the later periods (see Section B.7.5).[17] We hence use the tools on geoportail.fr to delineate the urban area of the EM and 1950 aerial photos maps manually on screen, taking a screenshot of each measurement. The criteria to classify as urban area are such that they coincide with the criteria which will be applied to the automated satellite measures in later periods. We classify an area as part of a city, if:

1. We observe contiguous housing structure in a grid cell of 250m by 250m. This means disconnected municipalities or smaller parishes will not be considered as part of the main city. Given there are no grid cells displayed on the historical maps, the analyst has to be careful to consider different scales for different size cities in this step.

[16] This concerns Roubaix (today part of Lille), Versailles (Paris), Tourcoing (Lille), Saint-Denis (Paris), Levallois-Perret (Paris), Boulogne-Billancourt (Paris), Neuilly-sur-Seine (Paris), Clichy (Paris) and Saint-Germain-en-Laye (Paris). Our hand-collected data are published online at https://docs.google.com/spreadsheets/d/e/2PACX-1vS02WpT0e7YTiS6f-svIXR3sURjiMRw7kBgfH1XF8LRre_dhPD0Y8Oy67cU_L4Q2FHg0r711ffB3XYm/pubhtml?gid=0&single=true

[17] contemporary photographs are taken between 2016 and 2020: https://www.geoportail.gouv.fr/depot/fiches/photographies-aeriennes-RVB/geoportail_dates_des_prises_de_vues_aeriennes-RVB.pdf

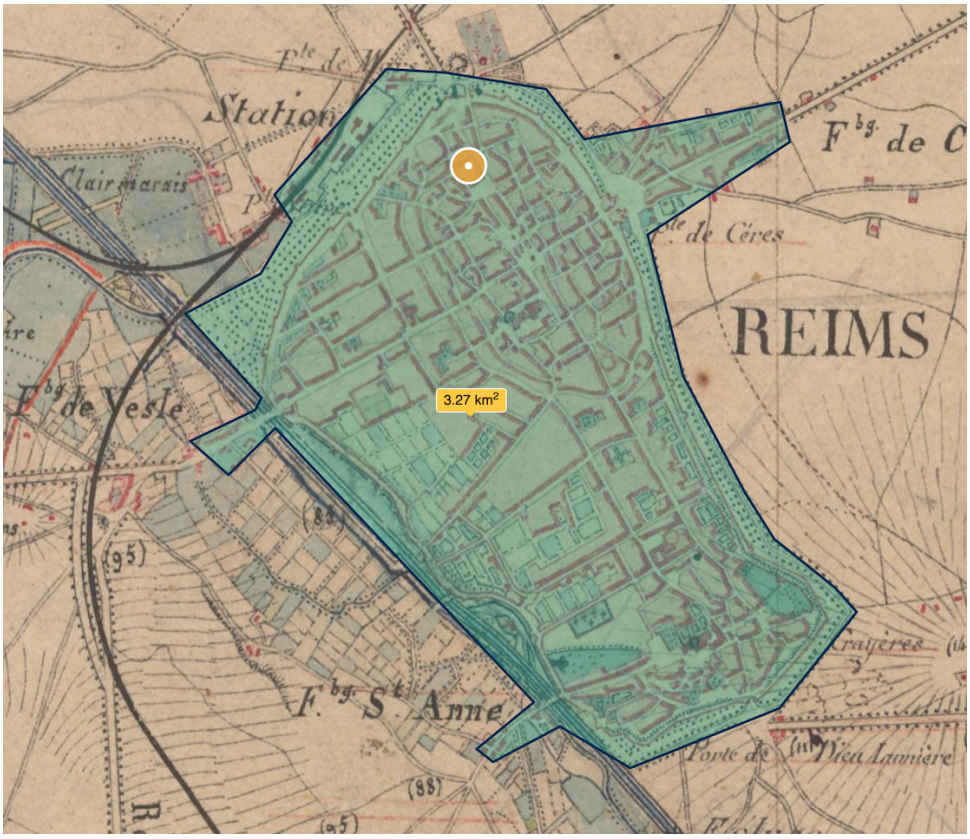


Figure VII: Area measurement of Reims using Etat Major map

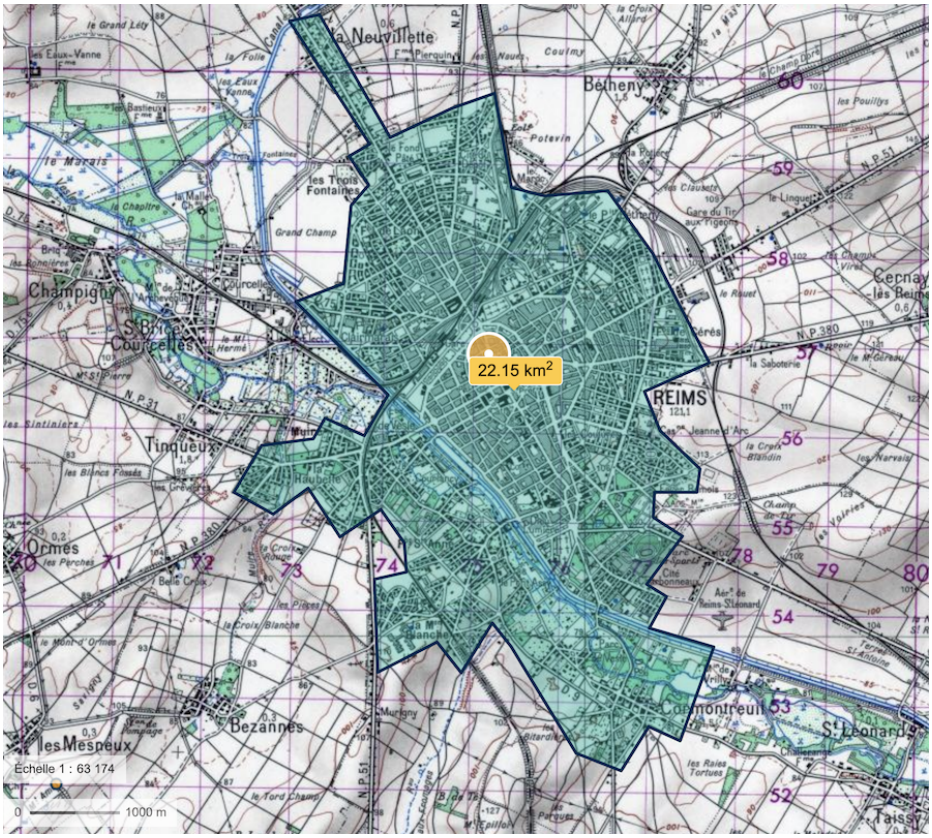


Figure VIII: Area measurement of Reims using 1950 map



Figure IX: Area measurement of Valence using Etat Major map

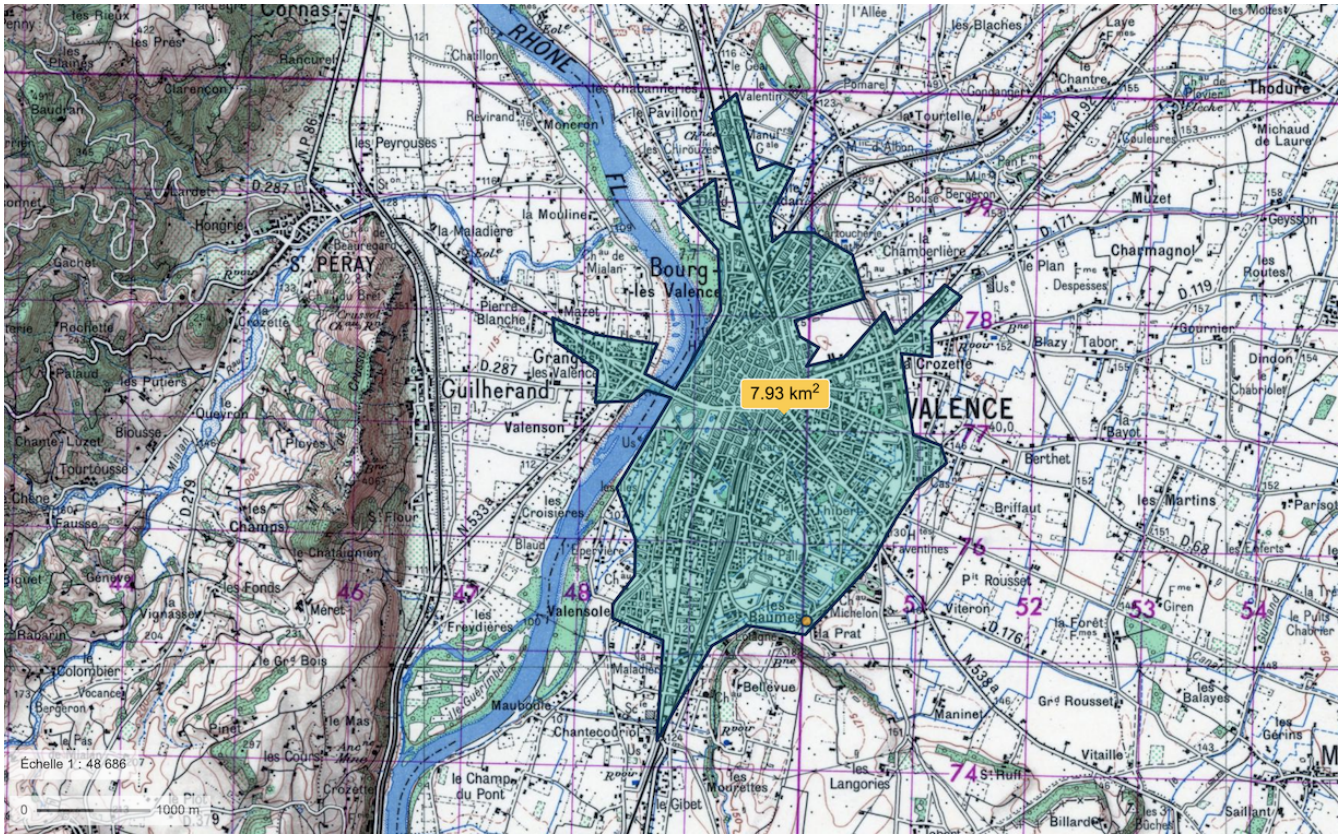


Figure X: Area measurement of Valence using 1950 map

2. Density of built-up environment: We try to enforce a 30% built-up threshold as in the automated approach in order to distinguish low-density suburbs from city proper. This means that areas which are contiguous to the main city, but significantly less dense because of interspersed gardens and rural land, will be excluded from the main city area.

We note that there is some degree of arbitrariness in this manual measurement procedure and that a fully automated approach like in Combes et al. (2021) is more rigorous. We verify that our approach is on average accurate and consistent with the automated procedure to be outlined below. We show some examples for Reims (figures VII, VIII) and Valence (figures IX, X).

B.7.2 Manual Population Measurements 1870 and 1950

In order to collect population counts for each city, we resort to the 1876 census as published by INSEE at https://www.insee.fr/fr/statistiques/3698339. This procedure is unambiguous, because all cities in our sample are contained within their administrative boundaries in the initial period.

The next data point for the 1950 cities is obtained from the census in 1954. Given the area measure obtained for 1950 as described in B.7.1, we verify for each city whether the total classified area falls within the administrative boundaries of the main city. If this is the case, we take the population measure directly from the census file, as before. If this is not the case (concerning in particular larger cities which incorporate surrounding villages by 1950, and in particular Paris), we carefully check which administrative areas (i.e. former independent villages) have now become part of our 1950 city area, and we sum the corresponding population counts for the concerned areas. The mapping of villages to cities is given in Table I and the one of Paris administrative areas is shown in Table II.

Table II: Paris 1950 Population Classification

CODGEO	REG	DEP	LIBGEO	year	population	date
75101	11	75	Paris 1er Arrondissement	1954	38926	1954-01-01
75102	11	75	Paris 2e Arrondissement	1954	43857	1954-01-01
75103	11	75	Paris 3e Arrondissement	1954	65312	1954-01-01
75104	11	75	Paris 4e Arrondissement	1954	66621	1954-01-01
75105	11	75	Paris 5e Arrondissement	1954	106443	1954-01-01
75106	11	75	Paris 6e Arrondissement	1954	88200	1954-01-01
75107	11	75	Paris 7e Arrondissement	1954	104412	1954-01-01
75108	11	75	Paris 8e Arrondissement	1954	80827	1954-01-01
75109	11	75	Paris 9e Arrondissement	1954	102287	1954-01-01
75110	11	75	Paris 10e Arrondissement	1954	129179	1954-01-01
75111	11	75	Paris 11e Arrondissement	1954	200440	1954-01-01
75112	11	75	Paris 12e Arrondissement	1954	158437	1954-01-01
75113	11	75	Paris 13e Arrondissement	1954	165620	1954-01-01

Table II: Paris 1950 Population Classification *(continued)*

CODGEO	REG	DEP	LIBGEO	year	population	date
75114	11	75	Paris 14e Arrondissement	1954	181414	1954-01-01
75115	11	75	Paris 15e Arrondissement	1954	250124	1954-01-01
75116	11	75	Paris 16e Arrondissement	1954	214042	1954-01-01
75117	11	75	Paris 17e Arrondissement	1954	231987	1954-01-01
75118	11	75	Paris 18e Arrondissement	1954	266825	1954-01-01
75119	11	75	Paris 19e Arrondissement	1954	155028	1954-01-01
75120	11	75	Paris 20e Arrondissement	1954	200208	1954-01-01
93001	11	93	Aubervilliers	1954	58740	1954-01-01
93005	11	93	Aulnay-sous-Bois	1954	38534	1954-01-01
93006	11	93	Bagnolet	1954	26792	1954-01-01
93007	11	93	Le Blanc-Mesnil	1954	25363	1954-01-01
93008	11	93	Bobigny	1954	18521	1954-01-01
93010	11	93	Bondy	1954	22411	1954-01-01
93013	11	93	Le Bourget	1954	8432	1954-01-01
93014	11	93	Clichy-sous-Bois	1954	5105	1954-01-01
93015	11	93	Coubron	1954	1039	1954-01-01
93027	11	93	La Courneuve	1954	18349	1954-01-01
93029	11	93	Drancy	1954	50654	1954-01-01
93030	11	93	Dugny	1954	6932	1954-01-01
93032	11	93	Gagny	1954	17255	1954-01-01
93033	11	93	Gournay-sur-Marne	1954	2141	1954-01-01
93045	11	93	Les Lilas	1954	18590	1954-01-01
93046	11	93	Livry-Gargan	1954	25322	1954-01-01
93047	11	93	Montfermeil	1954	8271	1954-01-01
93048	11	93	Montreuil	1954	76239	1954-01-01
93049	11	93	Neuilly-Plaisance	1954	13211	1954-01-01
93050	11	93	Neuilly-sur-Marne	1954	12798	1954-01-01
93051	11	93	Noisy-le-Grand	1954	10398	1954-01-01
93053	11	93	Noisy-le-Sec	1954	22337	1954-01-01
93055	11	93	Pantin	1954	36963	1954-01-01
93057	11	93	Les Pavillons-sous-Bois	1954	16862	1954-01-01
93059	11	93	Pierrefitte-sur-Seine	1954	12867	1954-01-01
93061	11	93	Le Pré-Saint-Gervais	1954	15037	1954-01-01
93062	11	93	Le Raincy	1954	14242	1954-01-01
93063	11	93	Romainville	1954	19217	1954-01-01
93064	11	93	Rosny-sous-Bois	1954	16491	1954-01-01
93066	11	93	Saint-Denis	1954	80705	1954-01-01
93070	11	93	Saint-Ouen	1954	48112	1954-01-01
93071	11	93	Sevran	1954	12956	1954-01-01
93072	11	93	Stains	1954	19028	1954-01-01
93074	11	93	Vaujours	1954	3972	1954-01-01

Table II: Paris 1950 Population Classification *(continued)*

CODGEO	REG	DEP	LIBGEO	year	population	date
93077	11	93	Villemomble	1954	21522	1954-01-01
93078	11	93	Villepinte	1954	5503	1954-01-01
93079	11	93	Villetaneuse	1954	3937	1954-01-01
94001	11	94	Ablon-sur-Seine	1954	3220	1954-01-01
94002	11	94	Alfortville	1954	30195	1954-01-01
94003	11	94	Arcueil	1954	18067	1954-01-01
94015	11	94	Bry-sur-Marne	1954	6660	1954-01-01
94016	11	94	Cachan	1954	16965	1954-01-01
94017	11	94	Champigny-sur-Marne	1954	36903	1954-01-01
94018	11	94	Charenton-le-Pont	1954	22079	1954-01-01
94019	11	94	Chennevières-sur-Marne	1954	4032	1954-01-01
94021	11	94	Chevilly-Larue	1954	3861	1954-01-01
94022	11	94	Choisy-le-Roi	1954	32025	1954-01-01
94028	11	94	Créteil	1954	13793	1954-01-01
94033	11	94	Fontenay-sous-Bois	1954	36739	1954-01-01
94034	11	94	Fresnes	1954	7750	1954-01-01
94037	11	94	Gentilly	1954	17497	1954-01-01
94038	11	94	L'Haÿ-les-Roses	1954	10278	1954-01-01
94041	11	94	Ivry-sur-Seine	1954	48798	1954-01-01
94042	11	94	Joinville-le-Pont	1954	15657	1954-01-01
94043	11	94	Le Kremlin-Bicêtre	1954	15618	1954-01-01
94046	11	94	Maisons-Alfort	1954	40358	1954-01-01
94052	11	94	Nogent-sur-Marne	1954	23581	1954-01-01
94058	11	94	Le Perreux-sur-Marne	1954	26745	1954-01-01
94067	11	94	Saint-Mandé	1954	24522	1954-01-01
94068	11	94	Saint-Maur-des-Fossés	1954	64387	1954-01-01
94069	11	94	Saint-Maurice	1954	11134	1954-01-01
94073	11	94	Thiais	1954	10028	1954-01-01
94076	11	94	Villejuif	1954	29280	1954-01-01
94079	11	94	Villiers-sur-Marne	1954	9205	1954-01-01
94080	11	94	Vincennes	1954	50434	1954-01-01
94081	11	94	Vitry-sur-Seine	1954	51507	1954-01-01
92002	11	92	Antony	1954	24512	1954-01-01
92004	11	92	Asnières-sur-Seine	1954	77838	1954-01-01
92007	11	92	Bagneux	1954	13774	1954-01-01
92009	11	92	Bois-Colombes	1954	27899	1954-01-01
92012	11	92	Boulogne-Billancourt	1954	93998	1954-01-01
92014	11	92	Bourg-la-Reine	1954	11708	1954-01-01
92019	11	92	Châtenay-Malabry	1954	14269	1954-01-01
92020	11	92	Châtillon	1954	12526	1954-01-01
92022	11	92	Chaville	1954	14508	1954-01-01

Table II: Paris 1950 Population Classification *(continued)*

CODGEO	REG	DEP	LIBGEO	year	population	date
92023	11	92	Clamart	1954	37924	1954-01-01
92024	11	92	Clichy	1954	55591	1954-01-01
92025	11	92	Colombes	1954	67909	1954-01-01
92026	11	92	Courbevoie	1954	59730	1954-01-01
92032	11	92	Fontenay-aux-Roses	1954	8642	1954-01-01
92033	11	92	Garches	1954	10450	1954-01-01
92035	11	92	La Garenne-Colombes	1954	26753	1954-01-01
92036	11	92	Gennevilliers	1954	33137	1954-01-01
92040	11	92	Issy-les-Moulineaux	1954	47433	1954-01-01
92044	11	92	Levallois-Perret	1954	62871	1954-01-01
92046	11	92	Malakoff	1954	28876	1954-01-01
92048	11	92	Meudon	1954	24729	1954-01-01
92049	11	92	Montrouge	1954	36298	1954-01-01
92050	11	92	Nanterre	1954	53037	1954-01-01
92051	11	92	Neuilly-sur-Seine	1954	66095	1954-01-01
92060	11	92	Le Plessis-Robinson	1954	13147	1954-01-01
92062	11	92	Puteaux	1954	41097	1954-01-01
92063	11	92	Rueil-Malmaison	1954	32212	1954-01-01
92064	11	92	Saint-Cloud	1954	20668	1954-01-01
92071	11	92	Sceaux	1954	10601	1954-01-01
92073	11	92	Suresnes	1954	37149	1954-01-01
92075	11	92	Vanves	1954	21679	1954-01-01
92078	11	92	Villeneuve-la-Garenne	1954	4035	1954-01-01

Table I: France 1950 Population Classification. Cities containing more than one INSEE administrative area by 1950.

CODGEO	DEP	LIBGEO	components
02691	2	Saint-Quentin	Saint-Quentin, Harly , Gauchy
03185	3	Montluçon	Montluçon , Désertines
03190	3	Moulins	Moulins, Yzeure
14366	14	Lisieux	Lisieux , Saint-Désir
28085	28	Chartres	Chartres , Mainvilliers, Luisant
29151	29	Morlaix	Morlaix , Saint-Martin-des-Champs
33063	33	Bordeaux	Bordeaux , Talence , Bègles , Le Bouscat
36044	36	Châteauroux	Châteauroux, Déols
42207	42	Saint-Chamond	Saint-Chamond, L'Horme
43157	43	Le Puy-en-Velay	Le Puy-en-Velay , Vals-près-le-Puy
44109	44	Nantes	Nantes, Rezé
51108	51	Châlons-en-Champagne	Châlons-en-Champagne, Saint-Memmie
51454	51	Reims	Reims , Cormontreuil
57463	57	Metz	Metz , Montigny-lès-Metz , Longeville-lès-Metz
59122	59	Cambrai	Cambrai , Proville , Neuville-Saint-Rémy
59178	59	Douai	Douai, Dechy
59350	59	Lille	Lille , La Madeleine
59606	59	Valenciennes	Valenciennes , Marly , Saint-Saulve , La Sentinelle , Anzin , Trith-Saint-Léger , Beuvrages , Raismes , Bruay-sur-l'Escaut , Petite-Forêt , Aulnoy-lez-Valenciennes
62041	62	Arras	Arras , Achicourt
62160	62	Boulogne-sur-Mer	Boulogne-sur-Mer , Saint-Martin-Boulogne, Outreau , Le Portel
62193	62	Calais	Calais , Coulogne
63113	63	Clermont-Ferrand	Clermont-Ferrand, Chamalières
67482	67	Strasbourg	Strasbourg , Schiltigheim, Bischheim , Hoenheim
69123	69	Lyon	Lyon , Villeurbanne , Caluire-et-Cuire, Oullins
76231	76	Elbeuf	Elbeuf , Caudebec-lès-Elbeuf , Saint-Aubin-lès-Elbeuf
76351	76	Le Havre	Le Havre , Sainte-Adresse
83137	83	Toulon	Toulon , La Valette-du-Var

B.7.3 Automatic Area and Population Measurement via GHSL

For years 1975, 1990, 2000 and 2015 we can rely on satellite data provided by the Global Human Settlement Layer (GHSL) project. We use two products, the multitemporal built-up grid `GHS-BUILT` (see Corbane et al. (2018)) and the multitemporal population grid `GHS-POP`, see Schiavina et al. (2019). We first give a brief overview of the GHSL data, which is a global raster dataset to measure human activity over space and time (see Florczyk et al. (2019)).[18] Then we will outline our strategy to derive area and population measures for our 100 French cities.

`GHS-BUILT` Area Classification. We rely on the multitemporal (years 1975, 1990, 2000, 2015) grid `GHS_BUILT_LDSMT_GLOBE_R2018A` which uses satellite imagery of various Landsat generations. The methodology to classify a certain pixel as built-up or not is described in Corbane et al. (2019). The task at hand is a classical supervised learning, or classification, task, whereby an automated procedure learns from a labeled dataset (the training dataset) how to label new and unseen data. The method used here is called *Symbolic Machine Learning* (SML), and it outperforms other methods such as Maximum Likelihood, Logistic Regression, Linear Discriminant Analysis, Naive Bayes, Decision Tree, Random Forest and Support Vector Machine both in terms of accuracy and in terms computing cost. We refer to Corbane et al. (2019) for greater details concerning accuracy assessment. We end up using the 250m resolution data in Mollweide projection, where a grid cell is characterized by a numeric (Float32) value in $[0, 100]$ representing the percentage of area in the cell which is *built up*. Finally, note that

> the concept of "built-up area" applied in the GHSL is compliant with the definition of the "building" abstraction in the Infrastructure for Spatial Information in Europe (INSPIRE). The "built-up area" as defined in the GHSL framework is "the union of all the satellite data samples that corresponds to a roofed construction above ground which is intended or used for the shelter of humans, animals, things, the production of economic goods or the delivery of services". (Corbane et al. (2019) page 141)

`GHS-POP` Population Grid. We use the product `GHS_POP_MT_GLOBE_R2019A` in this part. For later periods (after 2000), `GHS-POP` uses the Gridded Population of the World (v4.10) dataset produced by CIESIN/SEDAC. For the earlier years 1975 and 1990 it takes as input the `GHS-BUILT` grid and disaggregates population data from census enumerations according to a simple model. The disaggregation starts from knowledge of population counts in certain census areas, and then uses the building density from `GHS-BUILT` to distribute the census population into `GHS-POP` cells which constitute the

[18] https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2019.pdf

concerned census area. We use again the 250m resolution in Mollweide projection, where a grid cell is characterized by a numeric value $[0, \infty]$ representing population count – notice that given the fixed geography (a box 250m by 250m), the measure is synonymous for *population density* in this instance. For more details on the generation of `GHS-POP` data please refer to Freire et al. (2016).

`GHSL` Measurement Procedure. We first describe the exact data products we use, and then how we process them in order to obtain area and population measurements for all grid cells which are part of our list of 100 French cities. We begin by downloading the data via https://ghsl.jrc.ec.europa.eu/download.php?ds=bu, selecting the tiles covering continental France (tiles `18_3` and `17_3`). The precise data versions we use are as follows:

`GHS-POP` GHS_POP_E1975_GLOBE_R2019A_54009_250_V1_0_18_3 and GHS_POP_E1975_GLOBE_R2019A_54009_250_V1_0_17_3

`GHS-BUILT` ...

`year < 2015` GHS_BUILT_LDS1975_GLOBE_R2018A_54009_250_V2_0_17_3 and GHS_BUILT_LDS1975_GLOBE_R2018A_54009_250_V2_0_18_3

`year == 2015` GHS_BUILT_LDS2014_GLOBE_R2018A_54009_250_V2_0_18_3 and GHS_BUILT_LDS2014_GLOBE_R2018A_54009_250_V2_0_17_3

We proceed as follows with the data:

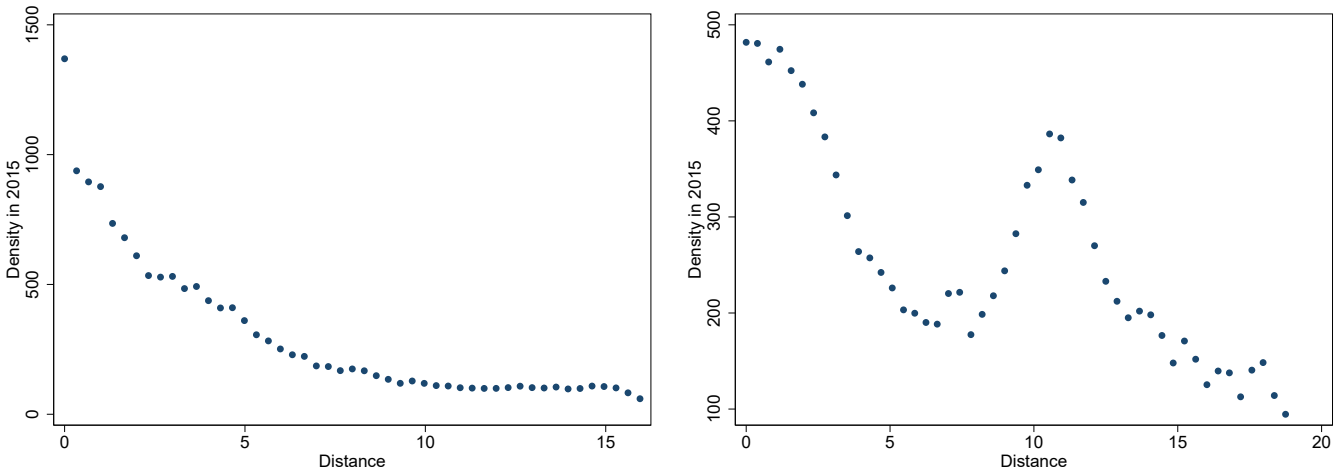
1. Read results of manual measurement (see B.7.1) to obtain list of cities and historical measures.
2. Crop GHS rasters to bounding boxes containing cities.
3. For each GHS-year, measure area from `GHS-BUILT` and population from `GHS-POP`. We delineate city extent based exclusively on `GHS-BUILT`, as follows:
   (a) Classify all grid cells with built-up proportion greater than threshold `cutoff` as *urban*. The baseline value for this parameter is 30%, and we present sensitivity analysis below in Section B.7.6.
   (b) For larger cities we have to decide what the *main* city is, as there may be disconnected parts of urbanized area outside the main city's boundary. We select the largest connected set of grid cells, where connection is established via *queen's case* directional movement (i.e. connected in any direction).
   (c) We count the so-classified grid cells of `GHS-BUILT` in order to obtain total urban area, and we sum the corresponding cells of `GHS-POP` in order to get urban population.

B.7.4 Density Measurement Results

Built-up and Density Measures. We show example output for built-up area classifications for two cities in figures XII and XIII. Example measures of resulting urban densities can be seen in figures XV for the top 5 cities as well as in Figure XIV for the entire distribution.

Within-city Density Gradients. For each grid cell of our GHSL data (2015), we define its distance from the center of the corresponding city, where the center is defined as location of the townhall by the the French National Institute of Geography (IGN). We cut each city into 50 bins of distance of equal size from the center and measure the average density across cells in each bin of distance. Thus, for each city i, we compute the density $D_{i,\ell}$ at distance ℓ from the city center. The set of distances ℓ varies across cities, as they are of different size.

Figure XIa illustrates the negative relationship between density and distance for the monocentric city of Lyon. Note that this relationship is quite different in a polycentric city such as Lille as shown in Figure XIb.



(a) Lyon (b) Lille

Figure XI: Density gradients.

Notes: Figures show the density (number of residents per 250m by 250m) at a given distance (in km) from the city center. GHSL data 2015.

We define the density gradient of city i, b_i, as the absolute value of slope coefficient of $\ln(D_{i,\ell})$ on ℓ,

$$D_{i,\ell} \approx a_i \exp(-b_i \ell), \tag{III}$$

where a_i and b_i are both positive. We use such a functional form as it fits very well the data for all French cities—apart few polycentric ones such a Lille. This gradient can be computed for every city in our sample.

The unweighted mean of gradients is equal to 0.200, while the population-weighted mean is equal to 0.121. This reflects the lower values of the gradient in larger cities as they are more likely to be polycentric. These two values provide reasonable bounds for average value of the gradient across cities.

Our sample contains few large polycentric cities (Lille, Nice, Paris, Saint-Etienne, Toulon and Toulouse) where density as a function of distance is clearly non-monotonic. One way to deal with the issue is to compute the the population-weighted mean of gradients, excluding large polycentric cities. This gives a value of 0.176 for the average gradient. Another way to deal with large polycentric cities is to adjust the gradients for those cities by cutting the city at a given threshold of distance, abstracting from the rise in density further away from the center. If we compute the gradient within the first 10kms of distance from the center for those cities, we obtain a population-weighted gradient of 0.146. A slightly higher value of 0.152 is obtained if the gradient is computed only on the central part of the most polycentric ones (below 6kms distance from the center). If we consider only the first 10kms from the center for all cities in the sample, we get a gradient of 0.154.

Thus, according to our empirical estimates, we find a density gradient ranging from 0.14 to 0.18 for the average city in our sample and the value of 0.15 constitutes our baseline estimate. Note that beyond the value for this average density gradient, our empirical investigation also shows that the exponential shape of Eq. III provides a very accurate description of the density data within cities.

From data to model's counterpart. The estimated density gradients correspond to a distance expressed in kms and the value is sensitive to the unit. Thus, one need to convert this value in a unit that can be confronted to our model. To do so, we measure the average radius of cities in our sample as the population-weighted mean of the largest distance bin in each of the 100 cities. This gives a mean radius of 21.43 kms. One can cut this radius into 20 equal size bins of 1.072 kms. Converting the density gradient of 0.15 in bin-unit gives a gradient of 0.16. This value can be compared to the gradient obtained in our model if we cut the radius ϕ of the city into 20 bins of equal distance increment.

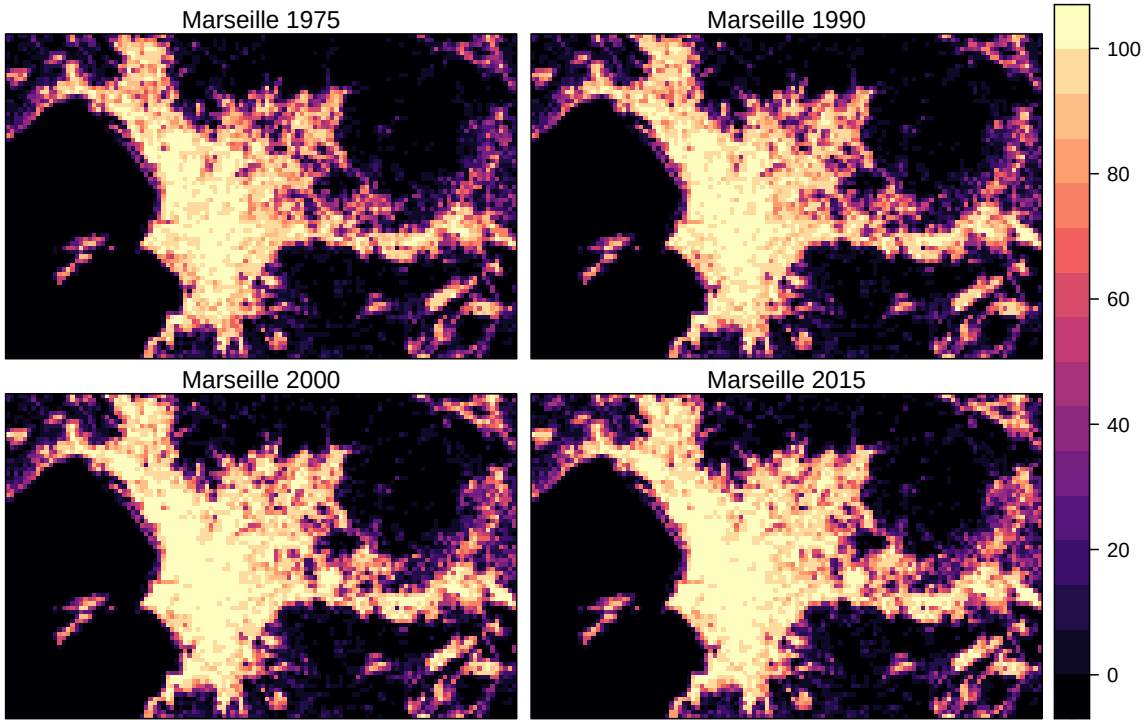


Figure XII: `GHS-BUILT` raster map of Marseille. The color scale represents percentage built-up in each grid cell.

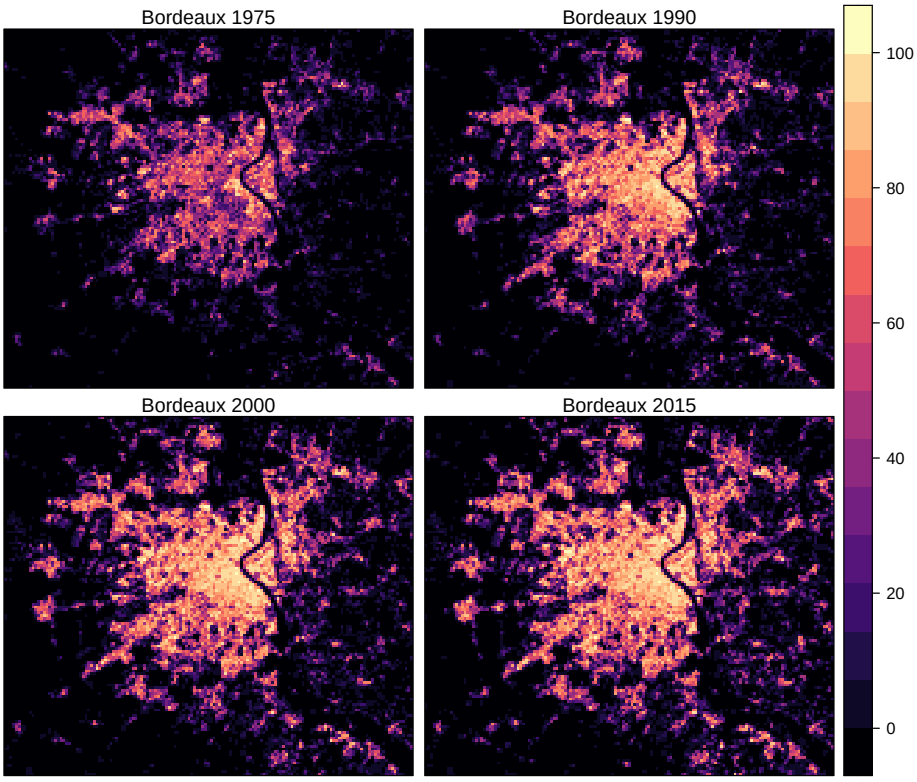


Figure XIII: `GHS-BUILT` raster map of Bordeaux.

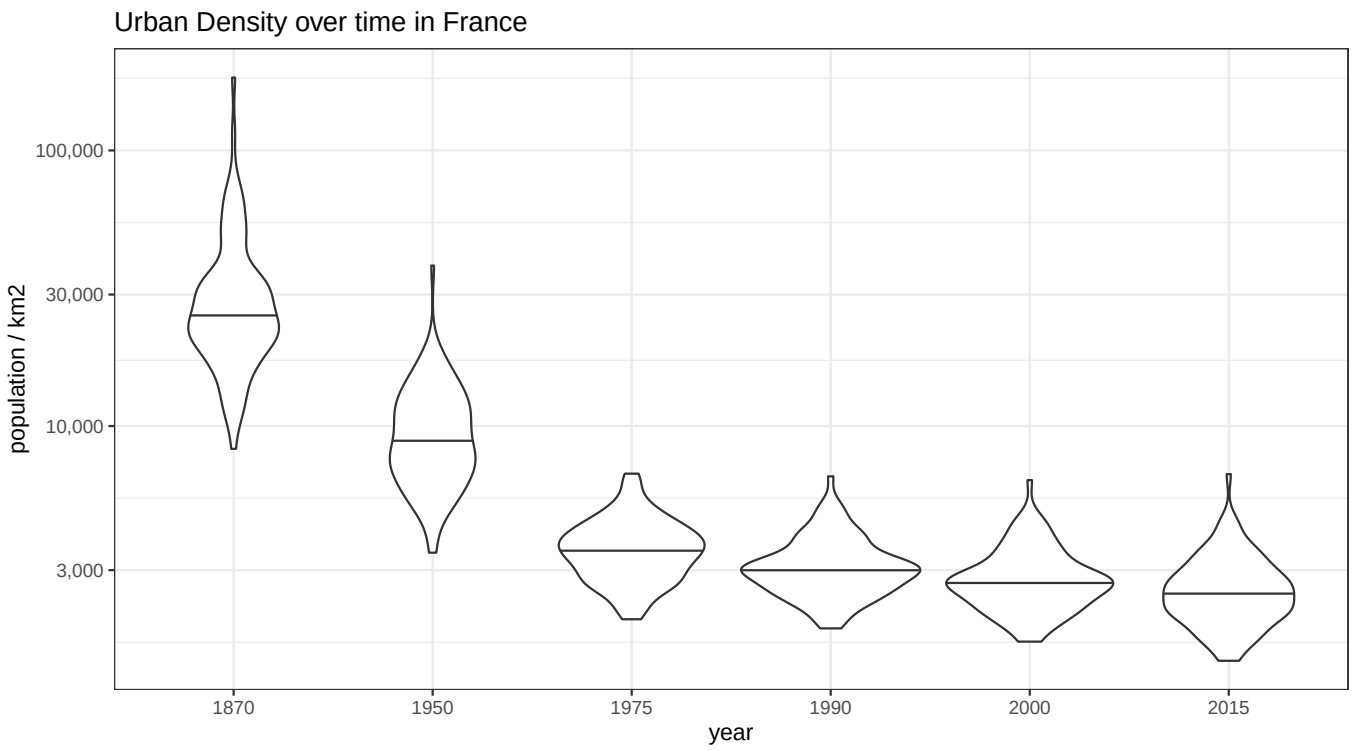


Figure XIV: Distribution of Urban Density over Time. This *violin plot* represents the distribution of densities at each date, labeling the extreme values. The horizontal line denotes the median value.

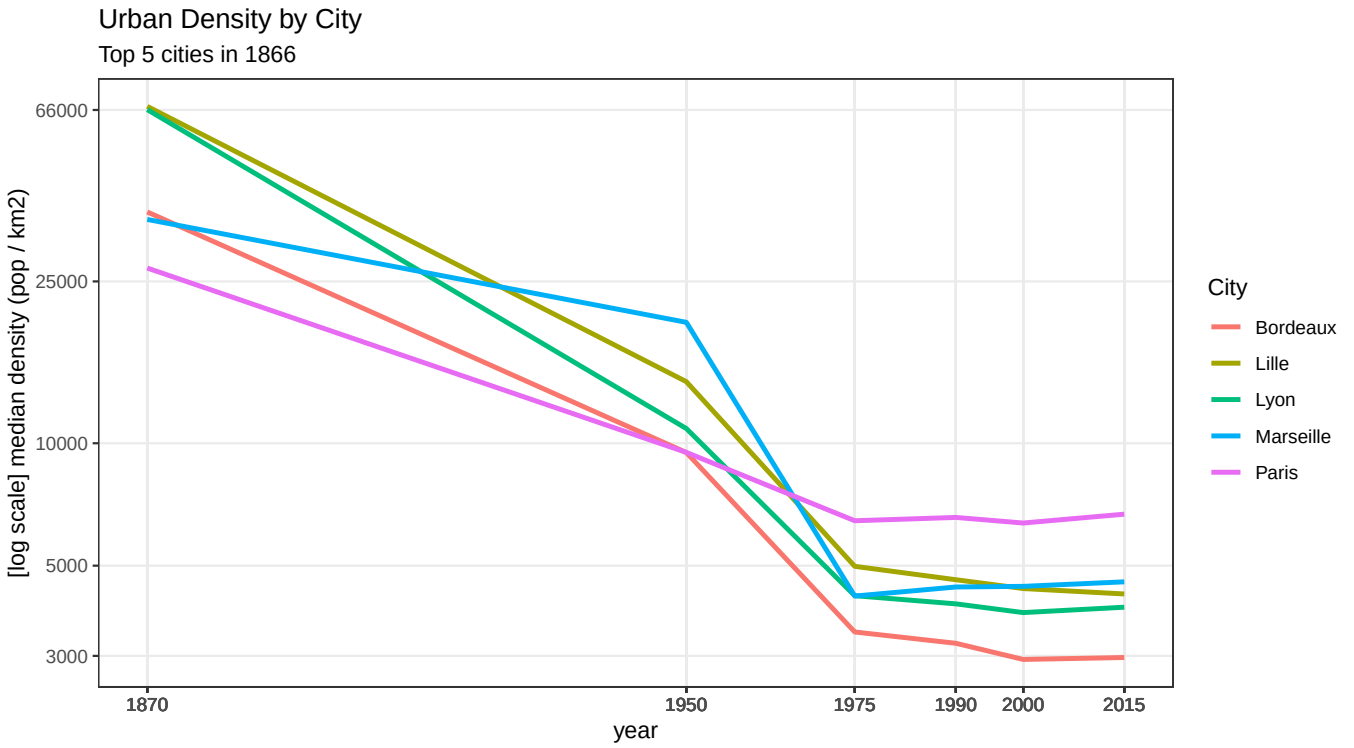


Figure XV: Density in the largest five cities in 1876 over time.

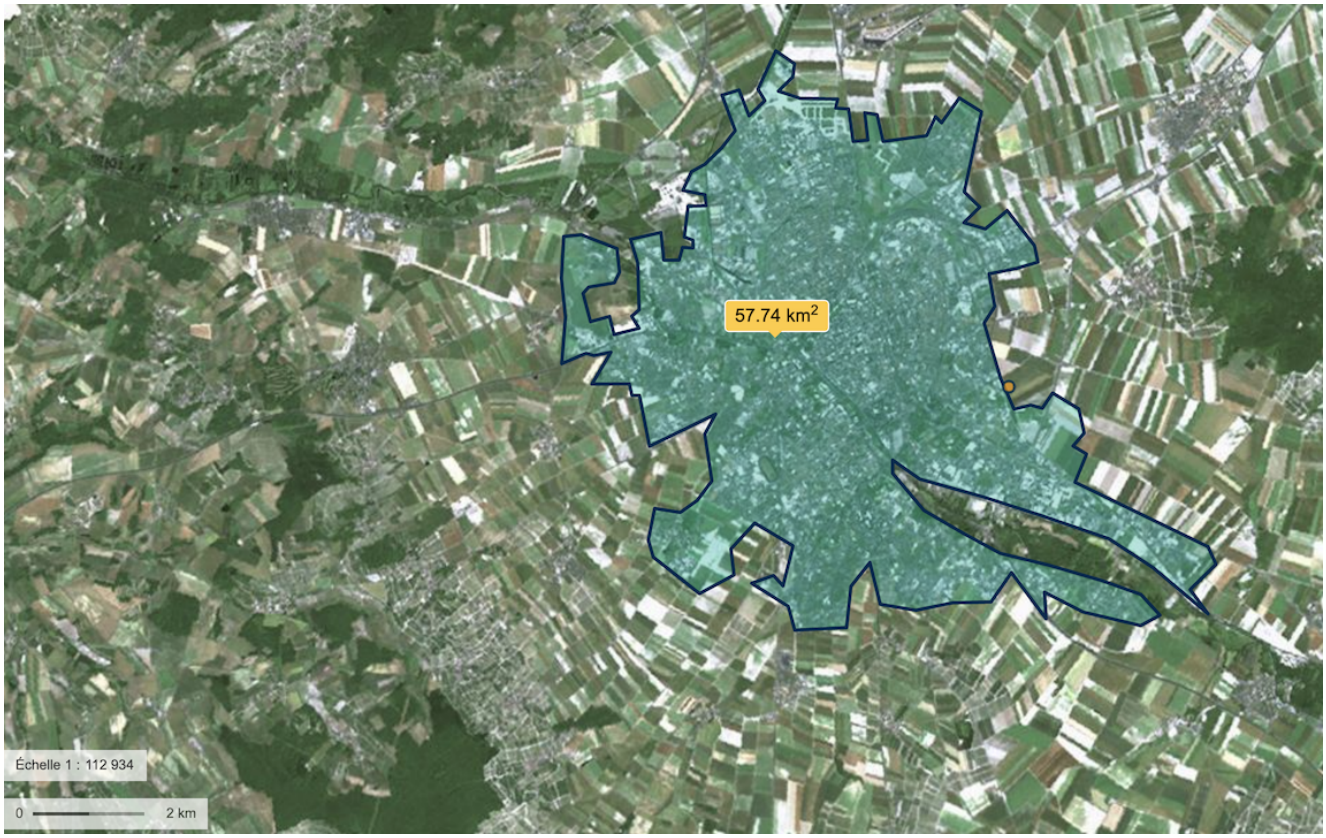


Figure XVI: Area measurement of Reims using modern day photograph - used only for cross-checking GHSL measures.

B.7.5 Consistency of Area Measures across Methods and Sources

We have aerial photography from 2016 available (see an example for the city of Reims in Figure XVI), which we use to also measure area of cities manually. The main purpose of this exercise is to show the consistency across methods (manual measurement and the automatic measure using satellite data). We report the relationship between manual 2016 and automatic 2015 measures in Figure XVII. Results are comforting. Both measures give similar estimates and are very highly correlated across cities.

Additionally, we can rely on historical data compiled by Shlomo Angel and co-authors for Paris (amongst many other cities), see Angel et al. (2012) and Angel et al. (2010). We report in Figure XVIII that our manual measures correspond closely to their obtained measures despite different measurement strategies.

B.7.6 GHSL `cutoff` Parameter Sensitivity

As mentioned earlier, we chose a cutoff of 30% built up in a grid cell to discriminate urban from rural area in terms of building density. The purpose of this parameter is to decide what type of

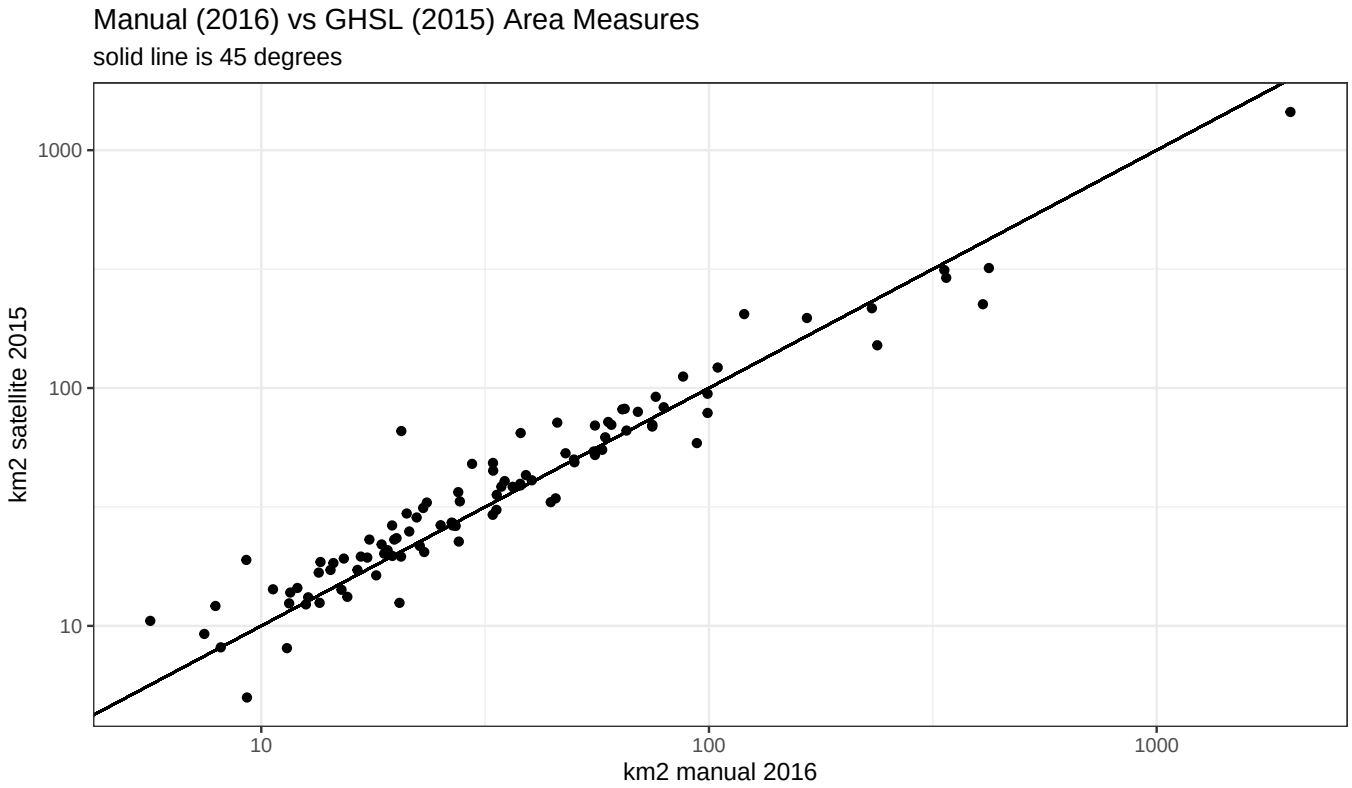


Figure XVII: Comparing manually obtained area measures for each of our cities with automatically obtained ones via GHSL data.

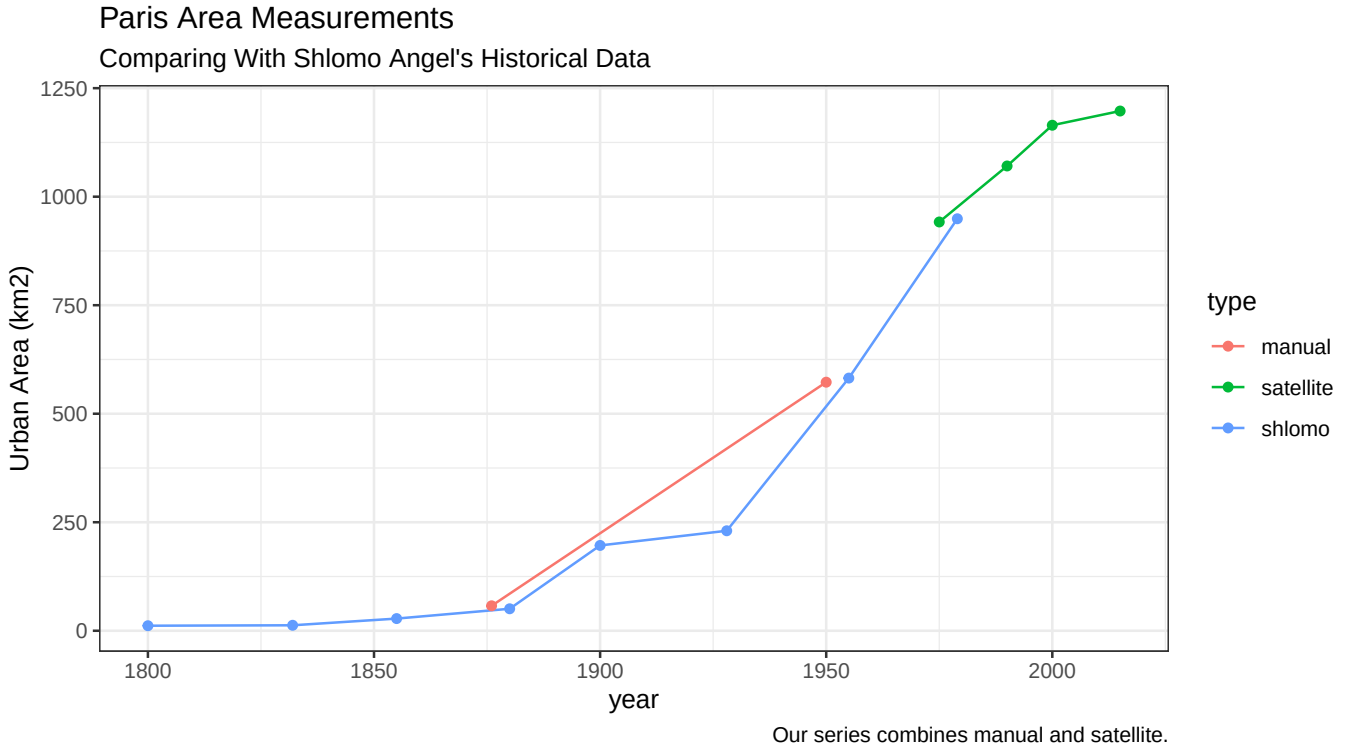


Figure XVIII: Comparing our measures with Shlomo Angel's data used in Angel et al. (2012) and Angel et al. (2010). We are reassured that our manual measurement exercise aligns closely with what they obtained. Also, their final data point is reassuringly close to our first satellite measure.

suburbanization should be considered to be still part of the city. In rough terms, our default setting would keep a property with 90 m^2 roofspace and 300 m^2 lot area (e.g. 210 m^2 garden) as part of the city. The criterion to classify an area as urban or not is necessarily subjective to some degree. We try to be as pragmatic as possible in choosing 30% and presenting how measured outcomes measure for a range of different cutoff values. De Bellefon et al. (2019) for example use a different approach which considers the distribution of grid cell density over the entire country and chooses a cutoff on that distribution as criterion (e.g. the 95-th percentile). The fundamental problem (value of the cutoff) remains. With this in mind, we present in figures XIX and XX our derived statistics about median and population-weighted average urban density, using different values for the cutoff parameter. We are reassured that towards the lower range of values, the density measure is rather stable. Very large values (less than half of a gridcell built up being excluded from *urban*) increase density more importantly. Our main data moment from this exercise – the ratio in (population-weighted) average urban density between 1876 and 2015 – is only minimally affected by the choice of `cutoff`.

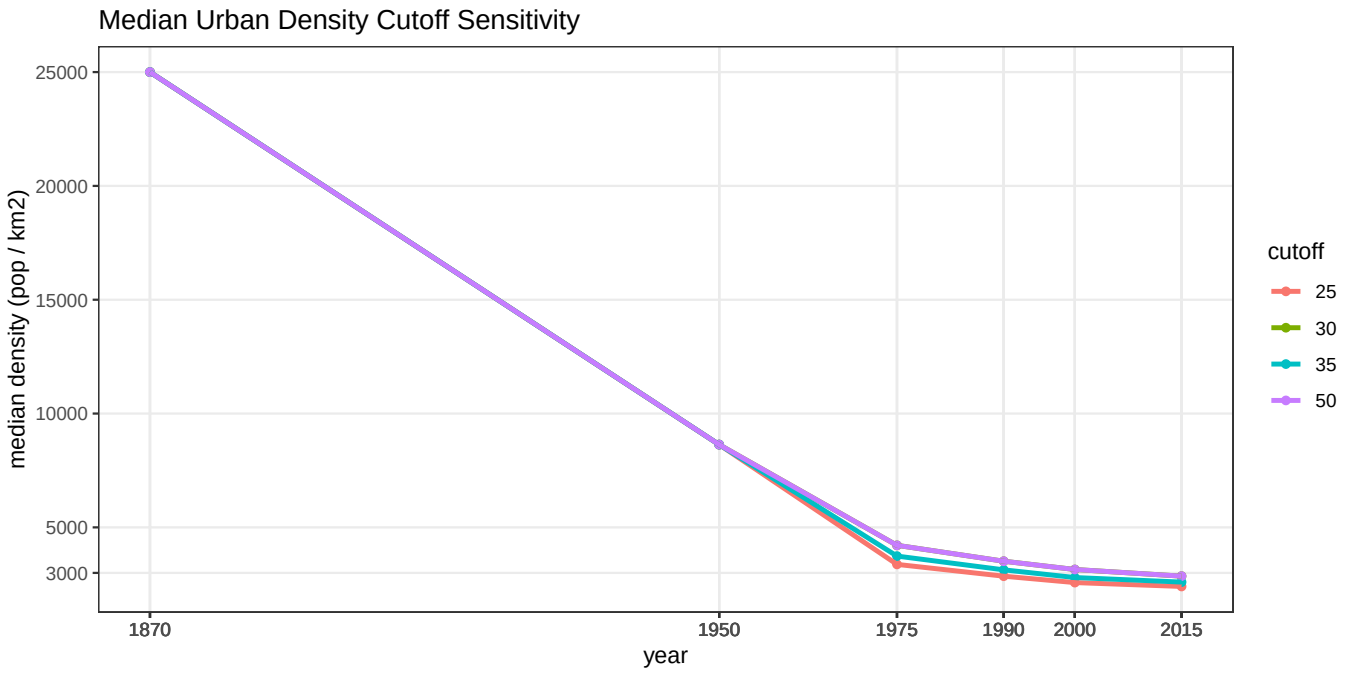


Figure XIX: Median urban density for different cutoff parameter values. The parameter indicates the percentage of a grid cell (250x250 meter) that has to be built-up in order to be classified as *urban area*.

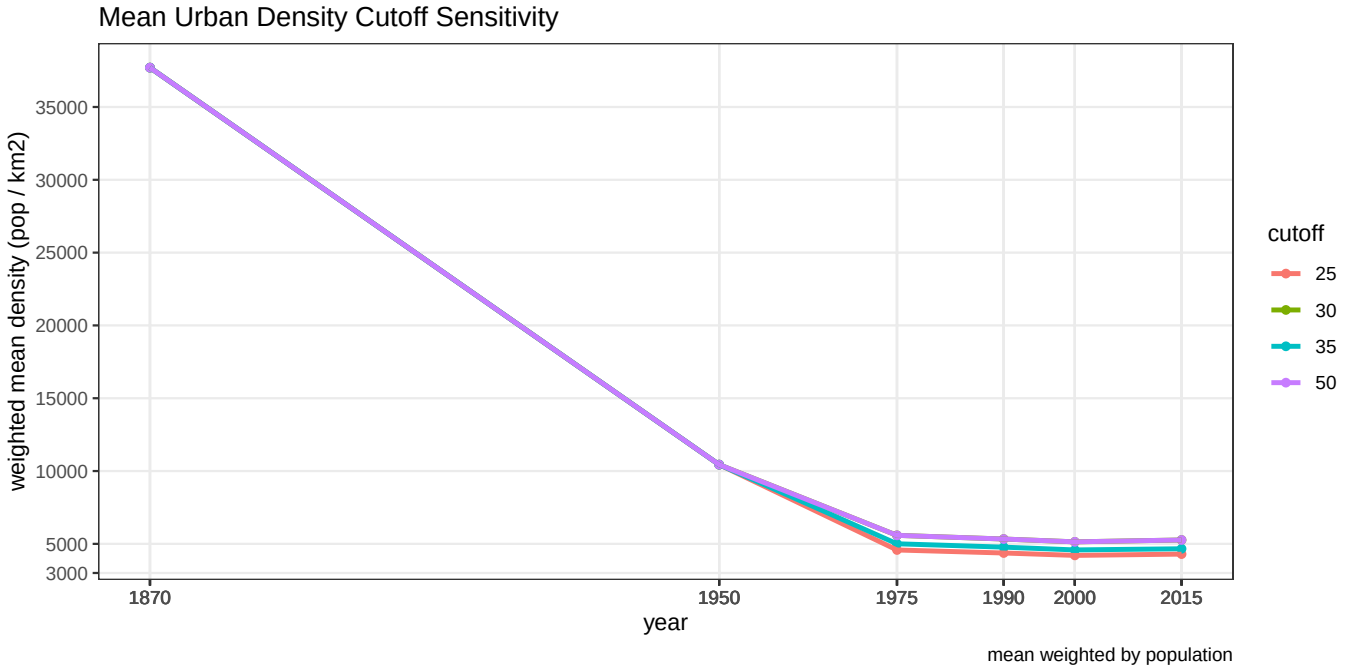


Figure XX: Weighted mean urban density for different cutoff parameter values. The parameter indicates the percentage of a grid cell (250x250 meter) that has to be built-up in order to be classified as *urban area*.

B.8 CORINE Land Cover (CLC) 2018 Data

We use CLC data to substantiate the claim made in Section 2.2 of the main text that land outside our top 100 French cities is to a large extent used for agricultural purpose nowadays. We rely on the 2018 edition of the European Land Monitoring Service called CORINE Land Cover (CLC) based on Sentinel-2 and Landsat satellite imagery European Union (n.d.). The geometric accuracy is better than 100m and the thematic accuracy is greater than 85%. We refer for all technical issues to the user manual of CLC available at https://land.copernicus.eu/user-corner/technical-library/clc-product-user-manual.

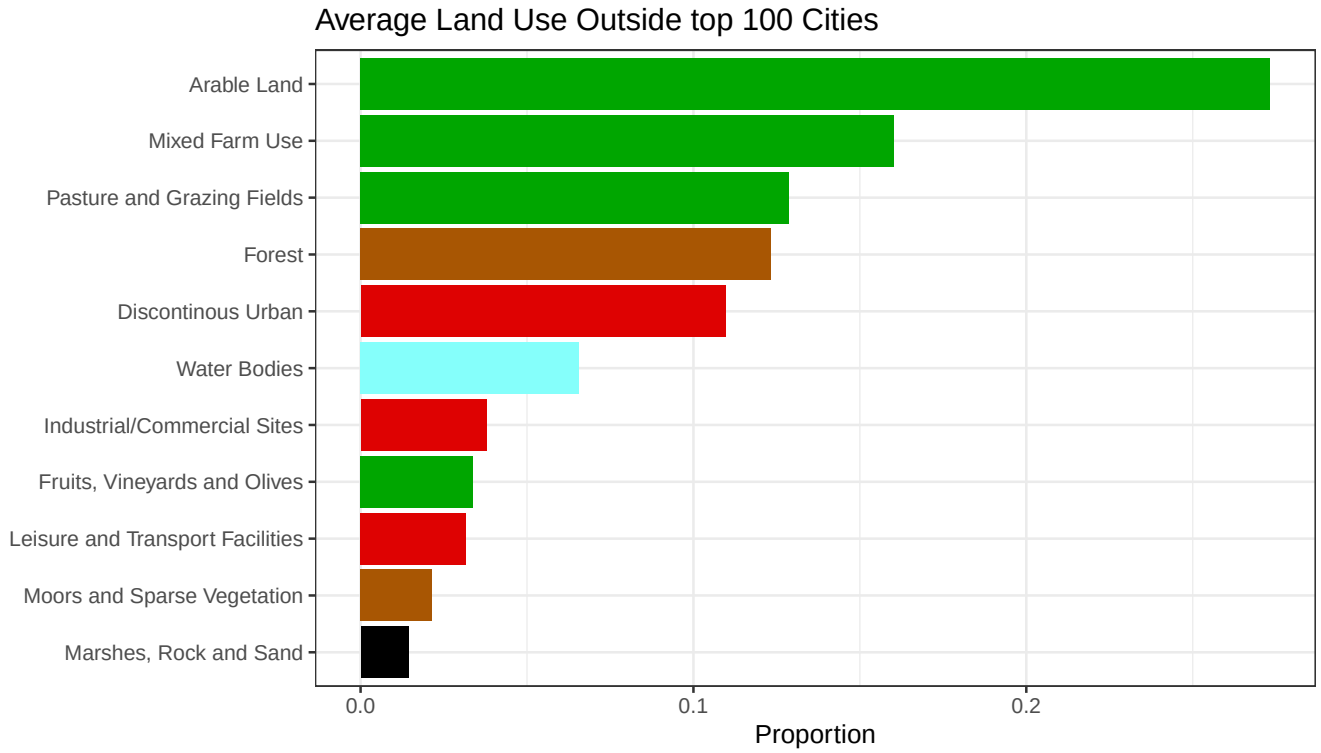


Figure XXI: Average land use measure from CLC data for our sample of top 100 French cities. This plot uses our own aggregation from 45 CLC labels into 11 exhaustive classes. We group all categories corresponding to agriculture into green bars.

Our usage of the data is very similar to the GHSL above. In fact, we crop CLC to a bounding box of continental France and then cut out the respective bounding boxes of our 100 cities. Care has to be taken to convert to the same coordinate reference system in this operation. Once the box around each city is contained, we report the proportion with which each of 41 land use types occurs. We show the resulting average in Figure XXI and an example for Reims in Figure XXII.

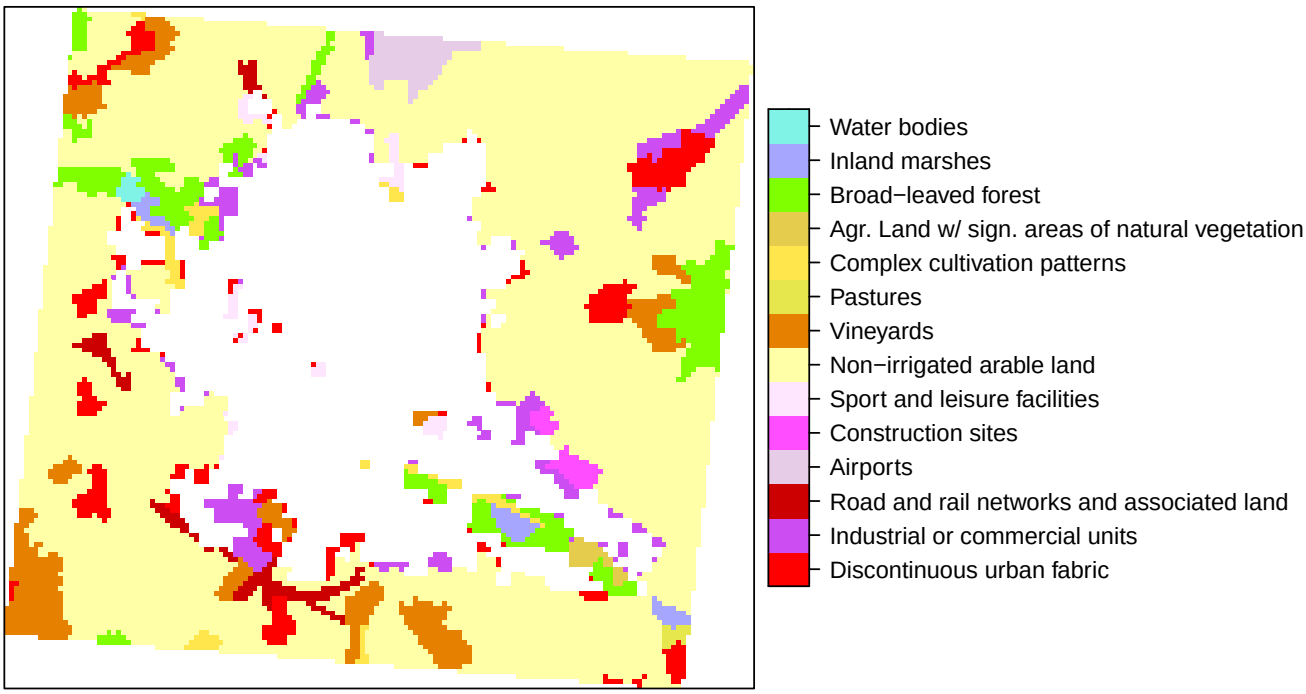


Figure XXII: Landuse measures from CLC data for Reims. The white area represents our definition of the Reims Urban area in the last GHLS periods (2015), hence it is our definition of *inside vs outside* of the city. For instance, the red areas labelled *discontinuous urban fabric* are not part of our definition of the city

B.9 Individual Commuting Data

We present here data from the "Enquête National du Logement (ENL)" as well as from the "Déclaration annuelle des données sociales" (DADS) in order to investigate individual commuting behaviour over space and time.

B.9.1 ENL: Enquête National du Logement

We obtain confidential access to the ENL and use it to measure commuting speed as a function of commuting distance. The ENL asks respondents questions about commuting behaviour, mode of commute, and importantly, duration of commute in minutes.

We use the waves 1984 (sample size $n = 9433$), 1988 ($n = 8910$), 2006 ($n = 12390$) and 2013 ($n = 7860$) where all required measures are observed. We subset the data to workers who work outside their home and to be the the reference person in each household. We observe workplace and residence at the commune level. We can therefore compute an approximation to commuting distance by taking the straight line distance between the central location of an individual's commune of residence and their commune of work. The central location is indicated by the IGN as *Chef Lieu* for each commune (most of the times the town hall). The variable *speed in km/h* is then implied by dividing our measure of commuting distance by each individual's commuting time (variable `GTT1`, reported in minutes) divided by 60. We drop all observations where reported commuting time or residence-workplace combination implies a commute of more than 100 km (or implied speeds of more than 100 km/h). We use the provided sampling weights for all computations. Figures XXIII and XXIV illustrate the distributions of our commuting distance variable.

We find that from 1984 to 2013, the average commuting distance increased by 3.2km, while the average commuting speed increased by 6km/h. Note that the increase in average speed over time is arguably the outcome of two forces: the use of faster commutes for a given commuting distance and an increasing importance of longer distance commutes for which workers use faster modes. The subsequent analysis aims at disentangling how speed changes over time for a given commuting distance and how speed varies with commuting distance at a given date.

We are interested in the cross-sectional elasticity of speed w.r.t commuting distance in a given year, controlling for as many individual characteristics as possible. This is reported in tables III and IV for the years 1984 and 2013 (we omit the 1988 and 2006 waves for brevity but results are very similar across years). We can see that across specifications with different control variables and across different years of data, the elasticity of speed with respect to distance is in the range of 0.438 (regional fixed effect specification in 2013) and 0.506 (simple regression of log speed on log distance in 1984). Since

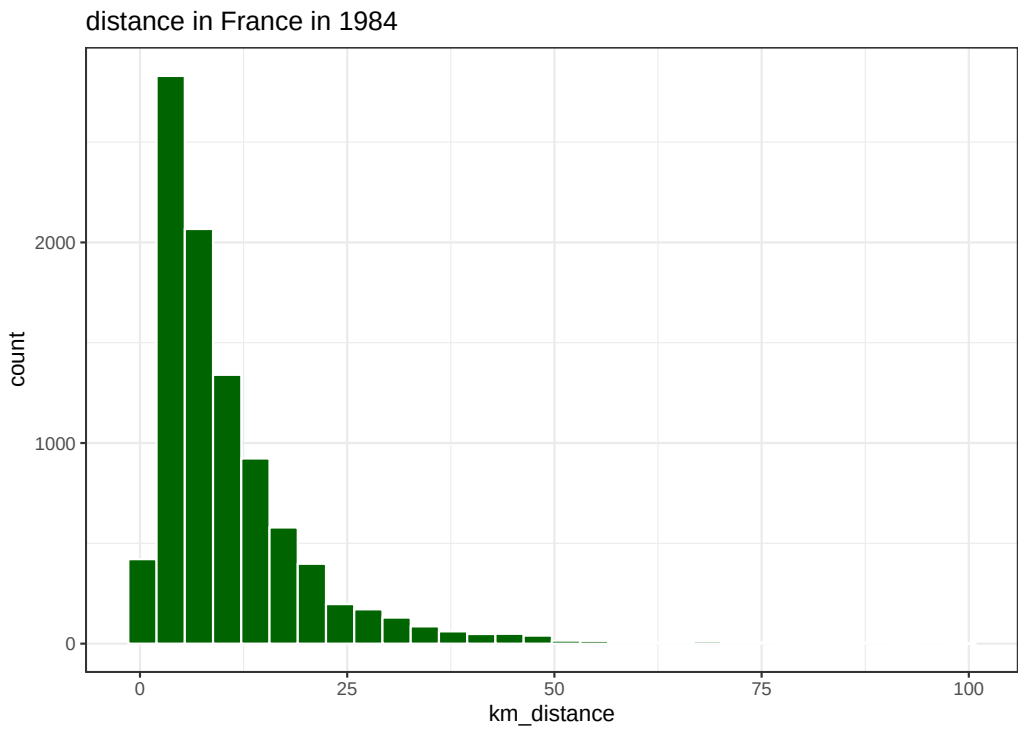


Figure XXIII: ENL: Distribution of Commuting Distances for a representative French Sample in 1984.

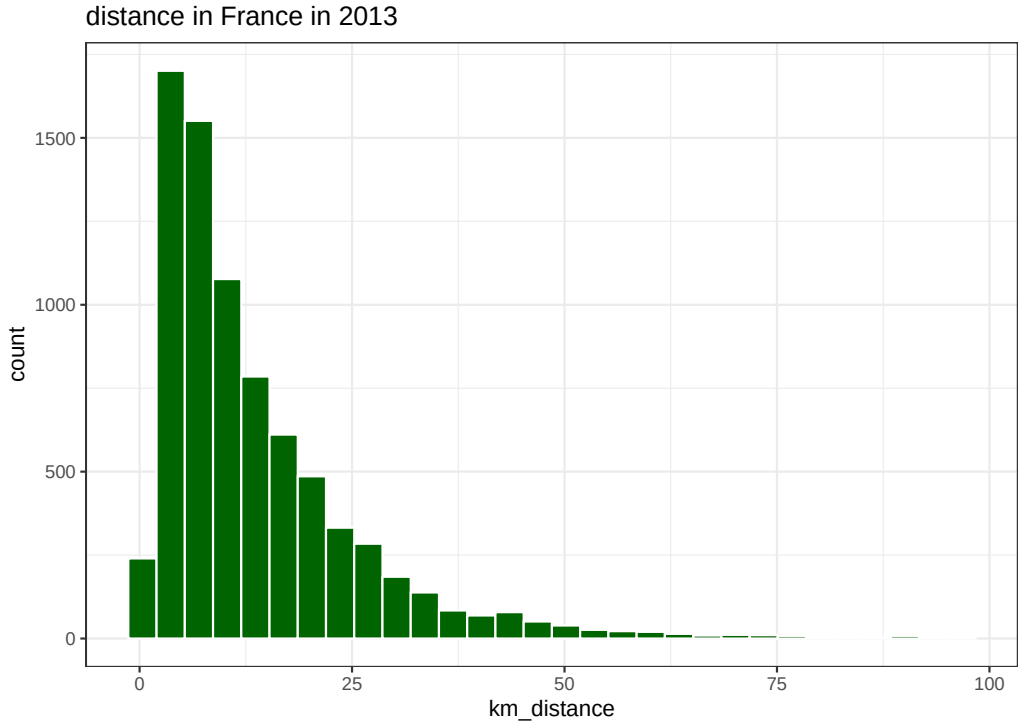


Figure XXIV: ENL: Distribution of Commuting Distances for a representative French Sample in 2013.

our preferred estimates with controls and regional fixed effects range from 0.43 to 0.47, we use 0.45 as baseline value to calibrate externally ξ_ℓ. This yields a value of $1 - 0.45 = 0.55$ for ξ_ℓ.

	log(speed) in 1984				
	(1)	(2)	(3)	(4)	(5)
log(km_distance)	0.506 ***	0.505 ***	0.498 ***	0.502 ***	0.470 ***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
r.squared	0.357	0.357	0.372	0.376	0.549
nobs	9199	9199	9189	9199	9189

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table III: Cross sectional regression of Speed on Commuting Distance using ENL 1984 data. Columns specify control variables as follows: Column (1) has no additional controls; (2) adds log income, (3) adds age and education class to (2), (4) adds adds age and SES to (2), and (5) adds age, education, SES and a regional fixed effect to (2).

	log(speed) in 2013				
	(1)	(2)	(3)	(4)	(5)
log(km_distance)	0.476 ***	0.478 ***	0.469 ***	0.474 ***	0.438 ***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
r.squared	0.361	0.362	0.397	0.410	0.570
nobs	7795	7795	7773	7795	7773

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table IV: Cross sectional regression of Speed on Commuting Distance using ENL 2013 data. Columns are specified as in table III.

We also want to know how the average commuting speed has evolved, controlling for commuting distance, between 1984 and 2013. To achieve this, we pool both cross sections from 1984 and 2013 and run a simple regression of the form

$$\ln \text{speed}_{it} = \beta_0 + \beta_1 \ln \text{dist}_{it} + \beta_2 \text{year}_{it} + u_{it}$$

This is reported in Table V. We group the data into 50 bins of log distance, shown in Figure XXV. Then we run a simple regression on this grouped data to measure the size of intercept shift, which will be our measure of *average increase in commuting speed at given commuting distance* between 1984 and 2013. We obtain a value of 0.109 on the dummy variable indicating `year == 2013`, hence the (approximate) marginal effect of being in year 2013 is given by a 10.9% increase in speed – controlling for commuting distance. This number is used in the quantitative model to calibrate parameter ξ_w as described in the calibration Section 4.1 in the main text.

	log_speed
(Intercept)	2.116 ***
	(0.027)
log_dist	0.457 ***
	(0.011)
factor(year)2013	0.109 ***
	(0.019)
r.squared	0.951
nobs	98

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table V: ENL Data. Measuring average increase in commuting speed between 1984 and 2013, controlling for commuting distance. This is done on data grouped into 50 bins of commuting distance. The coefficient of 'year==2013' is the size of the horizontal shift in figure XXV.

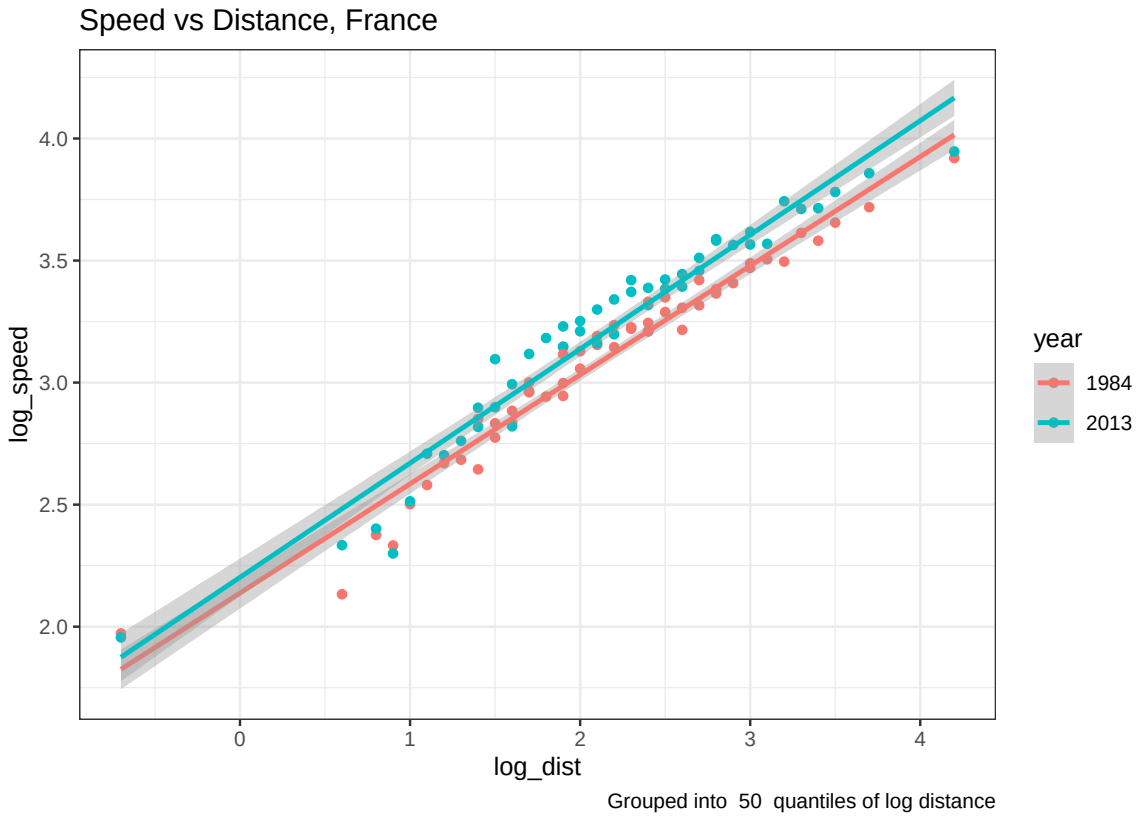


Figure XXV: ENL: Commuting speed for 50 bins of commuting distance (in log).

B.9.2 DADS: Déclaration annuelle des données sociales

The monocentric model implies that the location of residence ℓ maps one for one into commuting distance. Extension D.5 developed below (see Section 4.4 in the main text) relaxes this assumption. We will introduce in a reduced-form way the following relationship between commuting distance $d(\ell)$

and distance from the city center ℓ,

$$d(\ell) = d_0(\phi) + d_1(\phi) \cdot \ell \tag{IV}$$

where $d_0(\phi)$ and $d_1(\phi)$ are parametric functions of the city radius as detailed in extension D.5—with $d_0(\phi)$ increasing in ϕ and positive and $d_1(\phi)$ decreasing in ϕ and between 0 and 1. Data on residential and work locations are necessary to validate our reduced-from approach and discipline the calibration of $d_0(\phi)$ and $d_1(\phi)$.

We thus make use of confidential access to the DADS "Tous Salariés" (DADS-DSN) dataset for 2018 in order to investigate how commuting distance vary with residential location conditionally on city size in a large sample of the population. The DADS-DSN dataset contains all salaried workers in France, both private and public sector.

We start by reading the full dataset with 62 million records. We drop records which are in overseas territory, or which have as a residence or workplace identifier the code `75056`.[19] This reduces the sample to 60 million records. From this, we extract a 50% random sample. Next we obtain all unique pairs of residence and workplace communes (variables `COMR` and `COMT`) and compute straight-line distance for each pair. Then we add the distance of each commune to the center of their urban area. The urban area classification is officially given by INSEE and we use the AU2010 (Aire Urbaine 2010) classification. We end up with 18 million observations.

We aim to investigate how commuting distance varies with the distance from the center across different city sizes. We restrict our sample to individuals who do indeed conform to the INSEE definition of *aire urbaine* and whose workplace lies within their urban area, leaving us with 15 million observations. We also drop observations with commutes longer than 100 km, which concerns roughly 80000 workers. We have 15,317,995 observations left. Using the commuting distance (`distance_commute`) and the residential distance from the city center (`distance_center`) for each individual, we perform the following regression,

$$\texttt{distance_commute}_i = \gamma_{0,C(i)} + \gamma_{1,C(i)} \cdot \texttt{distance_center}_i + u_i \tag{V}$$

where i indexes an individual in DADS, $C(i)$ is the city (urban area) to which i belongs, and u_i is a mean-independent error term. $\gamma_{0,C(i)}$ and $\gamma_{1,C(i)}$ are city-specific coefficients (758 urban areas). We also perform the same regression by grouping cities into brackets of different sizes (with population above 3 millions, between 1 and 3 millions, between 50 000 and 1 million, ...).

[19] this stands for the entire commune of Paris and is the default value if Parisian Arrondissement is not available. This concerns only a small number of Parisian observations.

Figure XXVI plots the distribution of the intercept coefficient $\gamma_{0,C(i)}$ across all 758 urban areas. The mean across urban areas is 0.4 km and the mean weighted by the population of urban areas is 2.6 kms, significantly different from zero. Figure XXVII plots the distribution of the slope coefficient $\gamma_{1,C(i)}$ across all 758 urban areas. The distribution exhibits a mode around 0.7, while the population weighted mean is close to 0.5. Overall, residential distance from the city center is a very strong and robust predictor of commuting distance, even though commuting distance move less than one for one with residential distance from the center.

We also inspect the value of the estimates as a function of the size of the city. The intercept $\gamma_{0,C(i)}$ increases with city size, from about 0.2 km for the smallest urban areas to more than 4 kms for Paris. The slope coefficient $\gamma_{1,C(i)}$ decreases with city size—ranging from around 0.4 for Paris to more than 0.7 for small urban areas.

These results validate our reduced-form parametrization (Eq. IV), where commuting distance $d(\ell)$ increases less than proportionately with residential location ℓ, and less so for larger cities (larger radius ϕ). We use these findings in Section D.5 to parametrize $d_0(\phi)$ and $d_1(\phi)$.

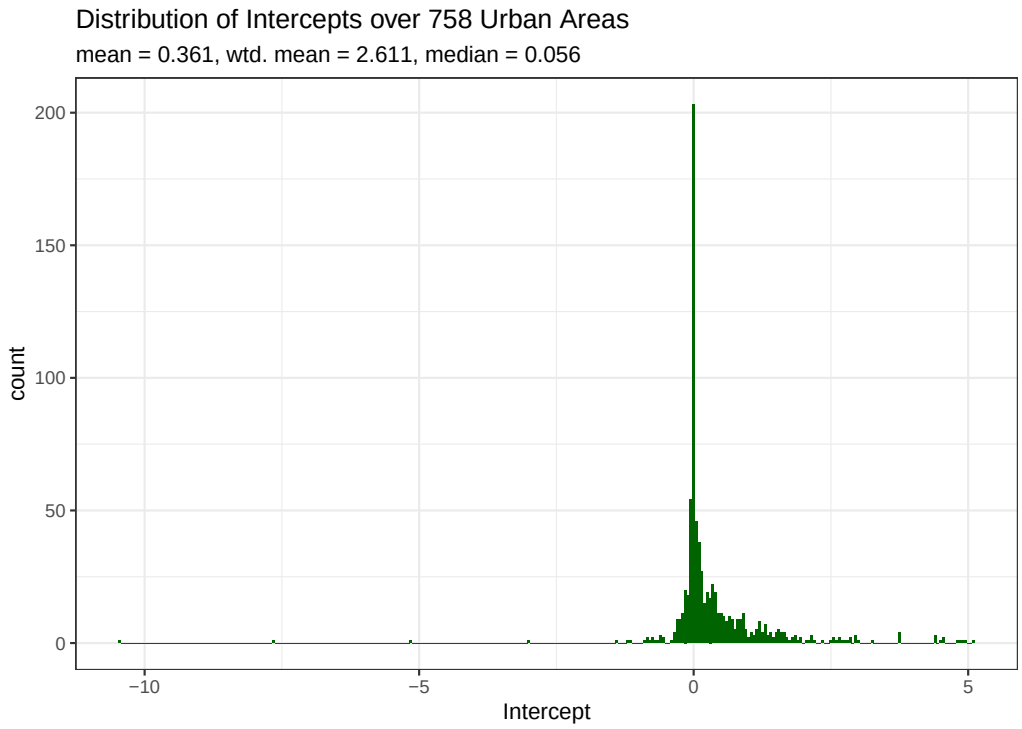


Figure XXVI: Distribution of DADS intercept estimates

Notes: City-specific intercepts $\gamma_{0,C(i)}$. *City* is defined as *Aire Urbaine (AU)* by INSEE. Results from individual level regression of commuting distance on distance from city center using DADS.

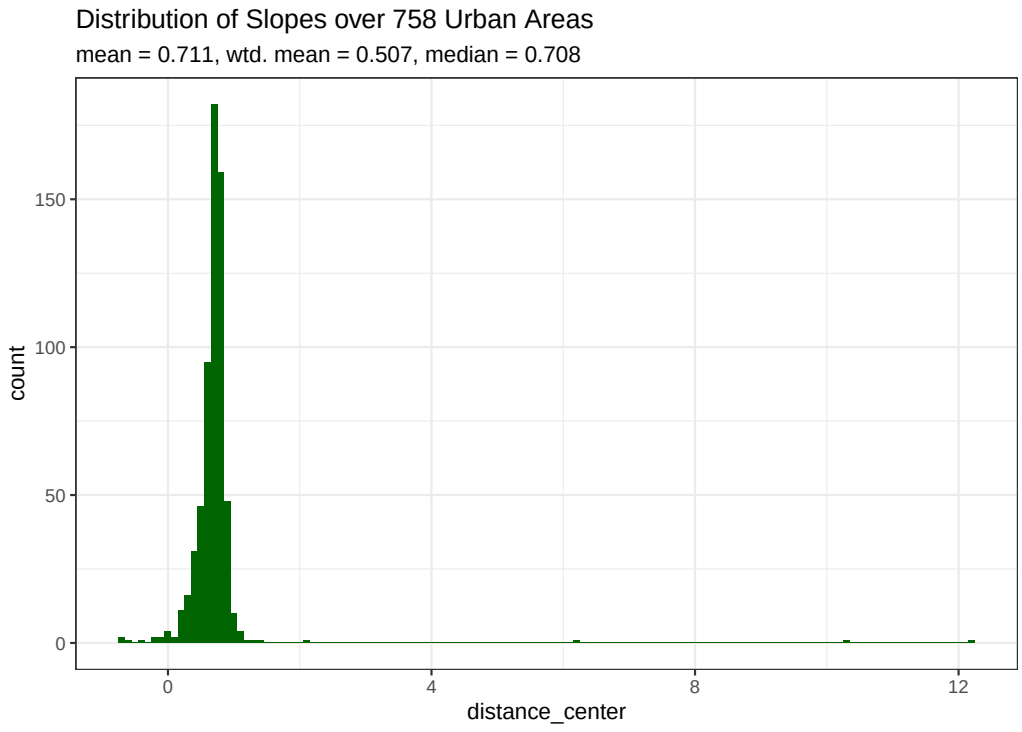


Figure XXVII: Distribution of DADS slope estimates

Notes: City-specific slopes $\gamma_{1,C(i)}$. *City* is defined as *Aire Urbaine (AU)* by INSEE. Results from individual level regression of commuting distance on distance from city center using DADS.

B.9.3 Historical commuting speed in Paris

We aim at providing estimates of the evolution of the average commuting speed for working trips in the Parisian urban area since 1840. These estimates are used to compare with the model's predictions (Figure 13a)). To do so, we use survey data (individual commuting data) in the Parisian urban area for the post-WW2 period. These data give the main mode used for working trips and well as the corresponding speed. Pre-WW2 (1840-1940), such individual surveys are not available. However, historical data on traffic by public transport modes and on registered private vehicles helps us to build estimates of the distribution of mode use over the whole period. Given estimates of the speed of each transportation mode, one can back out historical estimates of the average commuting speed.

Two main caveats are in order. First, the strategy developed only provide *estimates* since 1840 of the average commuting speed. These estimates depend on assumptions to convert historical data on traffic and registered vehicles into their modal use for work commutes and on assumptions regarding the speeds of the various modes. While some measurement error is unavoidable, our estimates provide a reasonable order of magnitude of the historical evolution of commuting speed in Paris. Second, due to historical data availability, we must focus on the Parisian urban area rather than France as a whole. Paris is arguably special. In the recent period, public transport is more widely used in Paris.[20] Paris might also be more congested than other French cities. Overall, one needs to be cautious with our estimates. However, it is clearly reassuring that estimates for Paris and model's predictions give very similar order of magnitude since the former were not targeted in the calibration.

Commuting data post-WW2. The first survey on commuting for work in the Parisian urban area was conducted in 1959 (on a representative sample of more than 20,000 individuals). While the original data are not available, secondary sources provide a detailed summary of the results (see Bertrand and Hallaire (1962)). For our purpose, this gives us the distribution of mode use in Parisian area in 1959. The majority of Parisian workers (50.2%) were using public transport (ventilated between metro, autobus and train); 21.5% were using a private mean of transportation (8.5% a private car, the rest for the most part a bicycle or a motorbike); the remaining 28.3% are walking.[21] The 1959 data do not provide the speed of each mode and we impute the speed measured in the later survey (1976) to compute the average commuting speed in the Parisian area in 1959. We use the 'Enquéte Global Transport (EGT)' for the years 1976, 1983, 1991, 2001 and 2010. The EGT provides individual commuting data for a representative sample of the Parisian urban area: distance of commuting trips, time, speed and modal use. We restrict our attention to trips to the work

[20]Note that the effect on commuting speed is however ambiguous. Cars are faster than public transport for longer distances but the large availability of public transports in Paris makes commuting easier for shorter distances.

[21]Note that less than 10% of surveyed individuals use a private car—reflecting the low level of car equipment in France in the 1950s. This number is up to 20.2% in 1967, 36.8% in 1976, 42.6% in 1983 and close to 50% since 1990.

location to extract the distribution of mode use and their respective speeds to compute the average commuting speed.[22] Note that the speed measured from these surveys is based on the distance as the crow flies and is measured using the time of the whole journey (including time to walk to the bus stop or metro/train station, time to park, ...). The implied speeds (around 9 km/h for the metro, 15 km/h for the train, 20 km/h for cars or motorbikes, ...) are thus significantly below the speed of the different modes when operating at full speed (see Figure XXIX).

Commuting data pre-WW2. Using traffic data for public transportation and numbers of registered private vehicles, we propose a strategy to estimate the distribution of workers across the different modes of transportation since 1840.

Public transportation. We investigate various secondary sources to measure the traffic of the different public transport modes at different dates (1835, 1856, 1876, 1890, 1910 and 1930). For the nineteenth century, we use data from Martin (1894) which provides very detailed statistics on transportation in the Parisian area across the various modes. Data for 1910 and 1930 are from Bertillon (1910), Brunet (1986), Merlin (1997), as well as the Annuaire statistique de la Ville de Paris in 1929, 1930 et 1931. Traffic is expressed in number of individual trips per year. Data for the Parisian urban area are available across the different modes: omnibus, tramway, metro, autobus, train and boat. The modes used depend on the time-period: only the horse-drawn omnibus initially, then appears the horse-drawn tramway in the late 1850s with 22 lines built between 1853 and 1873, followed by the electric tramway starting 1881 and motorized omnibus in 1905.[23] The network of the tramway is fully electric by the end of the nineteenth century and reaches its peak in the 1920s (122 lines) before slowly disappearing due to the development of the metro—being fully replaced later in the 1930s by the autobus. The first metro line opens in 1900—10 lines being built before WW1. Four more lines open in between the wars together with extensions of the existing ones. Suburban trains started post-1840 (with the exception of the line Paris-Saint Germain en Laye inaugurated in 1837) with major developments towards the late 1850s-early 1860s. Before WW2, it remains a mean of transportation much less used than the others. Lastly, boats were provided to the public to reach some specific destinations along the Seine before the offer was restricted to tourists post-WW2. This mean of transportation remained very anecdotal over the whole period.

We also collected similar data on traffic for public transportation post-WW2 at various dates (1955, 1975, 1990, 2000, 2010) using data from Bastié (1958), the Annuaire statistique de la Ville de Paris (1955), Merlin (1997), the Annual statistics of the Paris public transport entity RATP for 1975 and 1990 and data of the Observatoire de la mobilité en Ile-de-France (OMNIL) for 2000 and 2010

[22] The sample raw average commuting speed at each date gives very similar estimates.
[23] The horse-drawn omnibus disappears in 1913.

(annual traffic for all modes 2000-2020 from OMNIL). These more recent data help us to convert the traffic into a proportion of workers using the various modes to commute to work. To do so, we first compute, for a given mode m, the number $N_{m,t}$ of two-way trips per worker per working day in the Parisian urban area using employment at the various dates t from Census data.[24] The main issue arise since many of these measured trips are not made to commute to work but for other reasons (leisure, shopping, ...). Assuming that a fraction $x_{m,t} \in (0,1)$ of these trips are work commutes. By definition, the proportion of workers using mode m to commute to work, $p_{m,t}$, is the number of (two-way) working trips per worker (per working day) using mode m,

$$p_{m,t} = x_{m,t} \cdot N_{m,t}.$$

Thus, with some estimates of $x_{m,t}$, one can recover estimates of $p_{m,t}$ using traffic data. Note also that for the years post-WW2, $p_{m,t}$ and $N_{m,t}$ are both observed allowing us to back out $x_{m,t}$. However, some modes were abandoned post-WW2 (horse-drawn modes, tramways). Moreover, workers use sometimes more than one mode of public transportation (train + metro, ...). To avoid these issues, we assume for simplicity that $x_{m,t}$ is the same across modes. Under this assumption, the proportion p_t of workers using public transportation at date t is,

$$p_t = x_t \cdot \sum_m N_{m,t},$$

and $x_t = \frac{p_t}{\sum_m N_{m,t}}$ can be easily recover from the data for the years post-WW2—using measures of p_t in individual surveys and values for $(\sum_m N_{m,t})$ from traffic data. It is close to 1/3, relatively stable across years. Using EGT data which provides the motive for registered trips, 31% of non-walking trips in 1976 were between home and work. Such a value implies about 50% of people using public transport in 1955, in line with the corresponding survey data. Thus, prior to WW2, we set x to $\hat{x} = 31\%$.[25] This implies for each mode m at date $t = \{1835, 1856, 1876, 1890, 1910, 1930\}$,

$$p_{m,t} = \hat{x} \cdot N_{m,t}.$$

As summarized in Figure XXVIII, the estimated fraction of workers using public transportation, $p_t = \sum_m p_{m,t}$, starts from a very low value of 4.5% in 1835 and remains fairly low throughout the nineteenth century before picking up in the twentieth century. More than 50% of workers using public transportation by 1930. This proportion starts falling post WW2, largely due to the wider usage of automobiles. It is still around 40% in the recent years.

[24] We use all available censuses starting in 1835, initially considering the *Département de la Seine* as the Paris Urban Area; after 1975 we use INSEE's official definition of the Paris Urban Area.

[25] One could argue that commuting trips for leisure motives were perhaps less common in the 19th century, pushing towards setting a higher value for x. However, anecdotal evidence also emphasizes that public transportation, train in particular, were in the early years very often taken by the richer population for leisure activities.

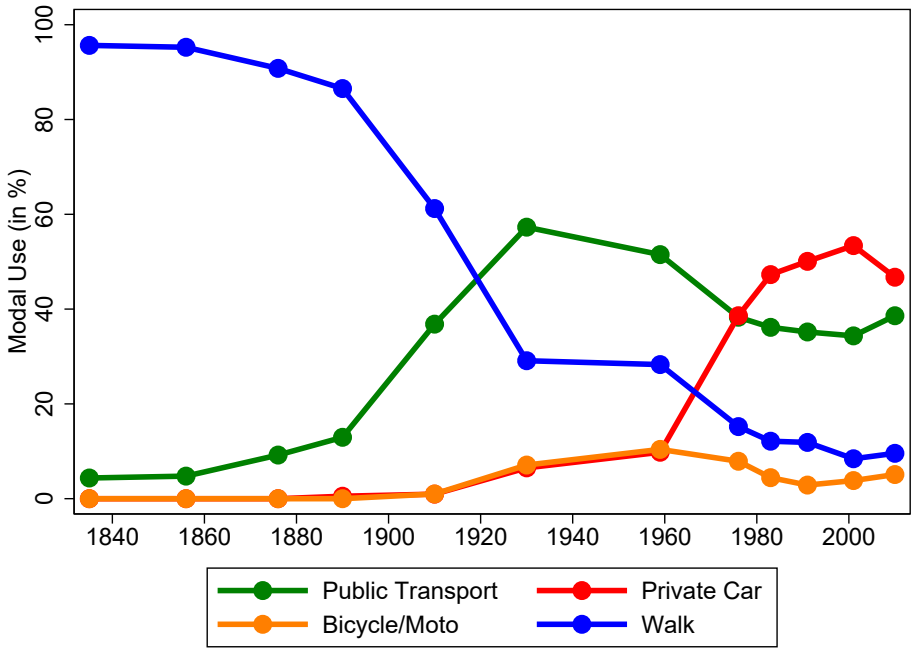


Figure XXVIII: Transportation mode use in the Parisian urban area.

Notes: Fraction of workers using the respective transportation mode over the period 1835-2010, in %. *Sources*: Data from secondary sources for the dates prior to WW2 (mostly traffic of the different public modes and registered private vehicles converted into modal use). Individual survey data on the main mode used for work commutes post-WW2 (Bertrand and Hallaire (1962) for 1959 and EGT data for 1976, 1983, 1991, 2001 and 2010).

Private transportation; Private transportation includes essentially private cars, bikes and motorbikes.[26] To evaluate the use of private cars pre-WW2, we use data on the number of registered vehicles, whether horse-drawn or motorized for years 1890, 1910 and 1930.[27] We also collected data for the number automobiles post-WW2 using Merlin (1997) and the annual statistics of the RATP for the years 2000 and 2010. Using these data and employment data, we compute the number of cars per worker (horse-drawn and motorized) since 1890. While the number of horse-drawn private cars per worker remained very small (below 1 for 200 before WW2), the number of automobiles per worker increases steadily until 1990 before reaching a plateau—about 1/100 in 1910, 11/100 in 1930, 22/100 in 1955, 61/100 in 1975 and 75/100 in 1990. However, many of these cars are not used on a daily basis for work commutes. To measure the proportion of workers using their car to go to work, we use survey data post-WW2 in the same vein as our strategy for public transportation. The ratio between the proportion of workers commuting to work by private cars and the number of cars per worker measures the fraction of cars used for work commutes. Post-WW2, this number is about 45% in 1959 and then hovers between 60% and 67%, with a mean across all observations of

[26]Pre-WW2, it also includes rented horse-drawn coaches with a driver. Post-WW2, it also includes other private means of transportation (taxis, private means provided by the employer, and recently scooters, ...). These remaining private means are either allocated to other categories according to their speed or neglected (employer buses considered as autobus, taxis as private cars, scooters as bikes...). Results are largely unaffected when omitting these categories.

[27]In 1899, 288 private automobiles were registered in Paris. We set the number of automobiles in 1890 to zero. In 1930, horse-drawn vehicles had almost disappeared in Paris and their number is also set to zero.

60%. Assuming a ratio pre-WW2 of 60% allows us to compute the fraction of workers commuting to work by private cars, less than 1% pre-WW1 and about 6% in 1930. Figure XXVIII summarizes the evolution of the proportion of workers using their private cars for work commutes.

The use of bikes and motorbikes was almost inexistent prior to 1890. The number of bikes in Paris is estimated to about 60 000 in 1891, 250 000 in 1901 and 285 000 in 1912 (Orselli (2008)). Unfortunately, such data are not available at a later dates and not readily available for motorbikes for the Parisian area.[28] Given the importance of bicycles for leisure and the lack of relevant data post-WW1, it is rather difficult to measure accurately the use of these means of transportation for work commutes. Prior to 1890, it seems reasonable to assume that these modes were not used. Given the low number of motorbikes registered in France as a whole pre-WW1 (about 27 000), we also assume that this means of transportation can be neglected in 1910. Thus, one needs to provide estimates in 1910 and 1930 for bikes and in 1930 for motorbikes. Based on a retrospective surveys provided by the ENTD2008 (Enquète nationale transports et déplacements) where people were asked their main mode of transport over their lifetime, on can assess the extent of bicycle/motorbike use relative to other means for 1930. Papon et al. (2010) provides such estimates by decades—reweighting observations to control for sample attrition due to survival: in 1930-1940, 9.9% of the population were using the bicycle as main mode of transportation in France, versus 2.3% for the 1920-1930 decade. We take the average between these values, 6.1%.[29] For the use of bikes in 1910, it is arguably very low and we set it to 1%, below their estimated value for the 1920s. For motorbikes, there are no survivors in the retrospective survey declaring using this mode for the decade 1930-1940, versus 4.8% for the following decade. While one cannot come up with a definitive estimate, motorbikes were most likely used by at most 2-3% of the workers. We set the share of workers using a motorcycle in 1930 to 1%.[30] Certainly, one might want to be cautious with these estimates due to the small sample size of survivors. Fortunately, given that motorcycles were barely used and bikes are not much faster than walking, the quantitative implications for the estimated average speed cannot be large. Figure XXVIII summarizes the estimates for the share of workers using bikes/motorbikes over the whole period.

Walking. The share of workers walking to their work location is estimated as a residual—made of workers using neither a public transportation nor a private one. Figure XXVIII summarizes the estimates for the share of workers walking to work over the whole period. In the early years, before

[28] Orselli (2008) provides data on the number of registered motorbikes for France over 1899-1914. This number is about 1/100 of the number of bikes—small enough to be neglected until WW1.

[29] For the following decades, 13% of people using bikes in 1940-1950, 13% in 1950-1960, 9.7% in 1960-1970—broadly in line with survey data for Paris available at the latest periods.

[30] Traffic data for France in 1934 (Orselli (2008)) shows that the share of traffic (per km per year) due to motorcycles is about 1/5 (resp. 1/10) of the one of bicycles (resp. automobiles)—broadly in line with the chosen value.

1840, Paris is a walkable city, public and private means of transportation are barely starting, and about 95% of the workers commute by feet. This share has been falling since reaching about 75% in the early twentieth century, 30% around WW2 and about 10% nowadays.

Average commuting speed. Average commuting speed is estimated as the weighted average of the speed of the various modes—weighted by their modal use. For modes of transportation still used in 1976 (first date for which the speed of the various modes can be measured), we set their speed at the earlier dates to the one observed in 1976. One caveat is that we ignore that current modes of transportation (public or private) have been faster through time. For the modes of transportation that disappeared (or have been replaced by more modern modes), we estimate speed based on anecdotal evidence related mostly in Martin (1894). Horse-drawn omnibus were not much faster than walking, about 7 to 8 kms per hour. When considering the time walking and waiting when using this mode, we set the horse-drawn omnibus speed to 6 kms per hour—in between walking speed and later measured metro speed (about 8.5 kms per hour). This is the value taken until 1890. Post-1890, we set the speed of omnibus to 7.5 kms per hour as a significant share of those were motorized. For tramways, we set the speed to 7.5 kms per hour when horse-drawn in 1876 and 8.5 kms per hour when fully electric in 1910. We use the average between these two values for 1890 since both were used. Boats were on average faster than ground transportation modes. We set their speed to 10 kms per hour but results are barely affected by this value within a reasonable range given that less than 1% of the Parisian population were using this mode when available. Lastly, we wet the speed of private horse-drawn cars to 8 kms per hour. Like for boats, results are barely sensitive to this value as this mode of transport for work commute was the privilege of few rich Parisians in the late nineteenth century. Figure XXIX summarizes the estimated speed of the different modes, by mode at different dates. Figures XXX shows the evolution over the whole period across broader mode categories—the speed of each category (public and private) is weighted by the modal use of the different modes within the category.

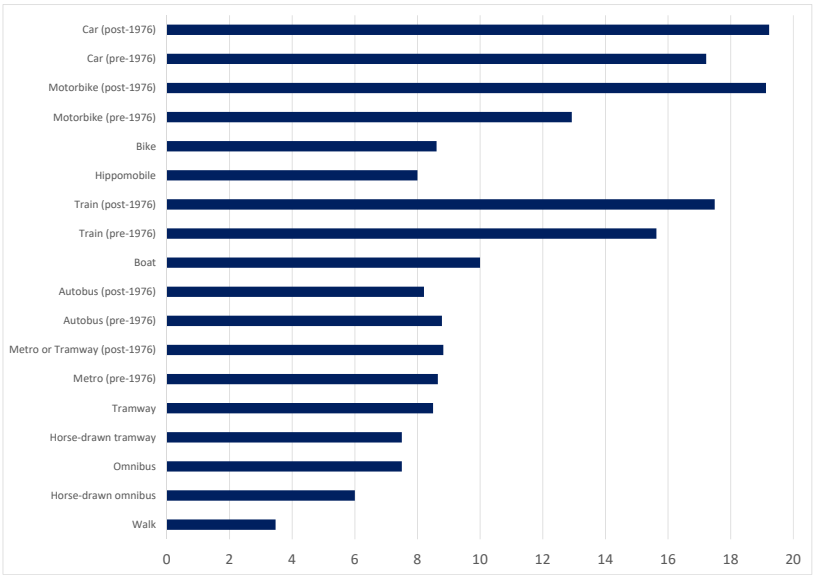


Figure XXIX: Speed across transportation modes.

Notes: Average speed of the different commuting modes. Measured using survey data in the Parisian urban area (EGT data) post-1976 (average over the 1983, 1991, 2001 and 2010 surveys). Values pre-1976 are based on the 1976-value from EGT data for modes still operating in 1976 and based on historical description for other modes.

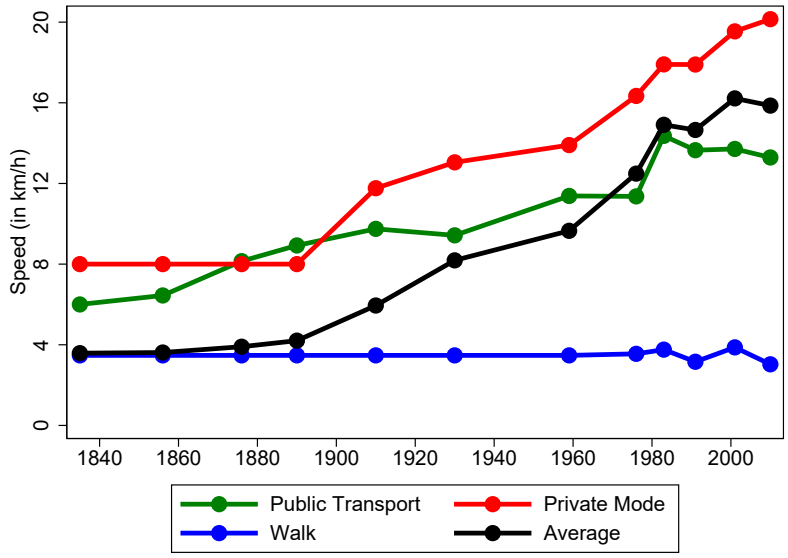


Figure XXX: Evolution of average speed across mode categories.

Notes: Public includes all public transportation modes. The speed for public transportation is a weighted average of the different public modes (weighted by their modal use). Private includes private car (horse-drawn and motorized), bikes and motorbikes. The speed for private transportation is a weighted average of the different private modes (weighted by their modal use). The average speed sums the speed of the different categories (walk, public, private) weighted by the computed modal use at the different dates. Average speed of the different commuting modes is measured using survey data in the Parisian urban area (EGT data) post-1976 (average over the 1983, 1991, 2001 and 2010 surveys). Values pre-1976 are based on the 1976-value from EGT data for modes still operating in 1976 and based on historical description for other modes.

C Quantitative Model, Numerical Solution and Estimation

C.1 Extensions to the baseline theory

The quantitative model expands the baseline theory of Section 3 in the main text by considering a more general production function for the rural good, a more general specification for the commuting costs, by allowing for location-specific housing supply conditions and by considering a circular city on a surface. We also introduce an intertemporal utility function to pin down the path for the real interest rate. Notice that each extension nests the previous one, such that the extension in Section C.1.4 contains all previous modifications and hence represents our full quantitative model for the single city case. This model will then be used for parameter estimation and main simulation results in Section C.2.5. Sensitivity analysis for some parameters values and extensions of the model around its baseline version are also developed in greater details.

The final extension in this section (see Section C.2) will introduce a multiple city version, for which we will use the estimated parameters from the single city model.

C.1.1 CES production in the rural sector

We perform sensitivity analysis with respect to the elasticity of substitution between land and labor in the rural sector. We hence upgrade the model from the main text with a CES technology where the production of the rural good uses labor and land according to the following constant returns to scale technology

$$Y_r = \theta_r \left(\alpha (L_r)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)(S_r)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}},$$

where L_r denotes the number of workers working in the rural (agricultural) sector, S_r the amount of land used for production and θ_r a Hicks-neutral productivity parameter. $0 < \alpha < 1$ is the intensity of labor use in production. $\sigma \geq 0$ is the elasticity of substitution between labor and land, $\sigma = 1$ corresponding to the baseline version. Rural workers and land are paid their marginal productivities such that main text equations (2) and (3) become

$$w_r = \alpha p \theta_r \left(\alpha + (1-\alpha) \left(\frac{S_r}{L_r} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} \tag{VI}$$

$$\rho_r = (1-\alpha) p \theta_r \left(\alpha \left(\frac{L_r}{S_r} \right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha) \right)^{\frac{1}{\sigma-1}} \tag{VII}$$

where w_r is the rural wage and ρ_r the rental price of land anywhere in the rural sector and p the relative price of the rural good in terms of the numeraire urban good. Note that it is useful to express

the price of land relative to wages,

$$\rho_r = \left(\frac{1-\alpha}{\alpha}\right) w_r \left(\frac{L_r}{S_r}\right)^{\frac{1}{\sigma}}. \tag{VIII}$$

Note that due to the CES technology, the rental price of land increase with (rural) wages with a unitary elasticity and with population working in the rural sector L_r with an elasticity $1/\sigma$—stronger complementarities between land and labor implying a larger fall of land prices if workers are reallocated to urban production.

C.1.2 Commuting Costs

We adopt a more general specification for the commuting costs f. The cost $f = f(\ell, m, w_u)$ depends on the transportation mode/speed, the location ℓ and labor costs w_u. Faster and longer commutes are more expensive and $f(\ell, m, w_u)$ is increasing in m and ℓ, with $\frac{\partial^2 f}{\partial^2 \ell} \leq 0$. The latter technical assumption makes sure that the importance of the cost f (relative to the opportunity cost of time) decreases as the commuting distance increases. The cost f also increases with the labor costs, w_u, with $\frac{\partial^2 f}{\partial^2 w_u} \leq 0$. This gives the following expression for the commuting costs, similar to Eq. 5,

$$\tau(l) = f(\ell, m, w_u) + 2\zeta w_u \cdot \left(\frac{\ell}{m}\right). \tag{IX}$$

For tractability, we will use the following functional form for f,

$$f(\ell, m, w_u) = \frac{c_\tau}{\eta_m} \cdot m^{\eta_m} \cdot l^{\eta_\ell} \cdot w_u^{\eta_w}, \tag{X}$$

with $\eta_m > 0$, $0 \leq \eta_\ell < 1$, $0 \leq \eta_w < 1$ and c_τ a cost parameter measuring the efficiency of the commuting technology.

An individual in location ℓ chooses the mode of transportation corresponding to speed m in order to minimize the commuting costs $\tau(\ell)$. This equalizes the marginal cost of a higher speed m to its marginal benefits in terms foregone wage,

$$\frac{\partial f}{\partial m} = 2\zeta \cdot w_u \left(\frac{\ell}{m^2}\right).$$

Using Eq. X, the optimal chosen mode/speed satisfies

$$m = \left(\frac{2\zeta}{c_\tau}\right)^{\frac{1}{1+\eta_m}} \cdot w_u^{1-\xi_w} \cdot \ell^{1-\xi_\ell}, \tag{XI}$$

where $\xi_w = \frac{\eta_m+\eta_w}{1+\eta_m} \in (0,1]$ and $\xi_\ell = \frac{\eta_m+\eta_l}{1+\eta_m} \in (0,1]$. Using Eqs. IX-XI, we get that equilibrium commuting costs satisfy,

$$\tau(\ell) = a \cdot w_u^{\xi_w} \cdot \ell^{\xi_\ell}, \tag{XII}$$

where $a = \left(\frac{1+\eta_m}{\eta_m}\right) c_\tau^{\frac{1}{1+\eta_m}} (2\zeta)^{\frac{\eta_m}{1+\eta_m}} > 0$. Commuting costs are falling with improvements in the commuting technology (a lower a). They are increasing with the wage rate (the opportunity cost of time) and the distance of commuting trips with constant elasticities. Expression (XII) is the resulting commuting cost function which appears in the model solution. It is also important to note that the parameters ξ_w (resp. ξ_ℓ) directly maps into elasticities of commuting speed to income (resp. commuting distance) through Equation XI. We use this link to directly parametrize both ξ_w and ξ_ℓ.

C.1.3 Location-specific housing supply conditions

As shown in Baum-Snow and Han (2019), the elasticity of housing supply to prices is lower closer to the CBD than at the urban fringe. We allow in this extension for location-specific housing supply conditions. To do so, we assume that in each location ℓ, land developers supply housing space $H(\ell)$ per unit of land with a convex cost

$$\frac{H(\ell)^{1+1/\epsilon(\ell)}}{1+1/\epsilon(\ell)}$$

paid in units of the numeraire, where $1/\epsilon(\ell)$ can depend on the location. This is meant to capture that it might be more costly for developers to build closer to the city center than in the suburbs or the rural part of the economy. Profits per unit of land of the developers are

$$\pi(\ell) = q(\ell)H(\ell) - \frac{H(\ell)^{1+1/\epsilon(\ell)}}{1+1/\epsilon(\ell)} - \rho(\ell),$$

where $\rho(\ell)$ is the rental price of a unit of land in location ℓ. Similarly to the housing price $q(\ell)$ above, for locations beyond the fringe ϕ, the land rent is constant, hence $\rho_r = \rho(\ell \geq \phi)$.

Maximizing profits gives the following supply of housing $H(\ell)$ in a given location ℓ,

$$H(l) = q(\ell)^{\epsilon(\ell)}, \tag{XIII}$$

where the parameter $\epsilon(\ell)$ is the price elasticity of housing supply in location ℓ. More convex costs to build intensively on a given plot of land reduces the supply response of housing to prices. In the rural area, the housing supply elasticity is assumed constant, $\epsilon_r = \epsilon(\ell \geq \phi)$.

Lastly, free entry imply zero profits of land developers. This pins down land prices in a given location,

$$\rho(\ell) = \frac{q(\ell)H(\ell)}{1+\epsilon(\ell)} = \frac{q(\ell)^{1+\epsilon(\ell)}}{1+\epsilon(\ell)}, \tag{XIV}$$

Arbitrage across land usage imply that the latter land price must be in equilibrium above the marginal productivity of land for production of the rural good (Equation (VII)), where the condition holds

with equality in the rural part of the economy, for $\ell \geq \phi$,

$$\rho_r = \frac{(q_r)^{1+\epsilon_r}}{1+\epsilon_r} = (1-\alpha)p\theta_r \left(\alpha \left(\frac{L_r}{S_r} \right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha) \right)^{\frac{1}{\sigma-1}}. \tag{XV}$$

This last equation shows that a fall in the relative price of rural goods and/or a reallocation of workers away from the rural sector lowers the price of urban land at the fringe of cities.

Consider first locations within the city, $\ell \leq \phi$. Market clearing for housing in each location implies $H(\ell) = D(\ell)h(\ell)$, where $D(\ell)$ denotes the density (number of urban workers) in location ℓ. Within the city, the density $D(\ell)$ follows,

$$D(\ell) = \left(\frac{q_r^{1+\epsilon(\ell)}}{1+\epsilon(\ell)} \right) \frac{1}{\gamma_\ell} (w(\phi) + r + \underline{s} - p\underline{c})^{-1/\gamma_l} (w(\ell) + r + \underline{s} - p\underline{c})^{1/\gamma_l - 1}, \tag{XVI}$$

where $\gamma_\ell = \frac{\gamma}{1+\epsilon(\ell)}$ represents the spending share on housing adjusted for the supply elasticity in location ℓ and the fringe housing price q_r satisfies $\rho_r = \frac{(q_r)^{1+\epsilon_r}}{1+\epsilon_r}$. Integrating density defined in Eq. XVI across urban locations gives the total urban population,

$$L_u = \int_0^\phi D(\ell)d\ell = \int_0^\phi \left(\frac{q_r^{1+\epsilon(\ell)}}{1+\epsilon(\ell)} \right) \frac{1}{\gamma_\ell} (w(\phi) + r + \underline{s} - p\underline{c})^{-1/\gamma_\ell} (w(\ell) + r + \underline{s} - p\underline{c})^{1/\gamma_\ell - 1} d\ell \tag{XVII}$$

Note that with homogeneous supply conditions across locations, $\epsilon(\ell) = \epsilon_r = \epsilon$, Equation (XVII) simplifies into (22) of the main text, which we give again:

$$L_u = \int_0^\phi D(\ell)d\ell = \rho_r \int_0^\phi \frac{1+\epsilon}{\gamma} (w(\phi) + r + \underline{s} - p\underline{c})^{-\frac{1+\epsilon}{\gamma}} (w(\ell) + r + \underline{s} - p\underline{c})^{\frac{1+\epsilon}{\gamma} - 1} d\ell.$$

In the rural area, $\ell \geq \phi$, market clearing for residential housing imposes

$$L_r \gamma \left(w_r + r + \underline{s} - p\underline{c} \right) = S_{hr} \left(q_r \right)^{1+\epsilon_r} = S_{hr}(1+\epsilon_r)\rho_r,$$

where S_{hr} is the amount of land demanded in the rural area for residential purposes. This leads to the following demand of land for residential purposes in the rural area,

$$S_{hr} = \frac{L_r \gamma_r \left(w_r + r + \underline{s} - p\underline{c} \right)}{\rho_r}, \tag{XVIII}$$

where $\gamma_r = \frac{\gamma}{1+\epsilon_r}$.

The market clearing condition for land, Equation (24), becomes

$$S_r = S - \phi - \frac{L_r \gamma_r \left(w_r + r + \underline{s} - p\underline{c} \right)}{\rho_r}. \tag{XIX}$$

The last modification regards the market clearing for urban goods. The amount of urban good used to produce housing becomes location specific. Hence, Equation (28) becomes

$$c_u + \frac{1}{L} \int_0^\phi \tau(\ell) D(\ell) d\ell + \frac{1}{L} \int_0^1 \frac{\epsilon(\ell)}{1+\epsilon(\ell)} q(\ell) H(\ell) d\ell = y_u, \tag{XX}$$

where $y_u = \frac{Y_u}{L}$ denotes the production per worker of the urban good.

C.1.4 Single City On Circular Surface

We now take model C.1.3 and extended it to a surface (instead of a line), where the city is circular around its center—ϕ denotes the radius of the city. The modifications concerns for the most part the integrals used. We start with the city size Equation (XVII), which becomes

$$L_u = \int_0^{\phi} D(\ell) 2\pi \ell d\ell = \int_0^{\phi} \left(\frac{q_r^{1+\epsilon(\ell)}}{1+\epsilon(\ell)} \right) \frac{1}{\gamma_\ell} (w(\phi) + r - p\underline{c})^{-1/\gamma_\ell} (w(\ell) + r - p\underline{c})^{1/\gamma_\ell - 1} 2\pi \ell d\ell \quad \text{(XXI)}$$

Notice again that with homogeneous supply conditions across locations, $\epsilon(\ell) = \epsilon_r = \epsilon$, this simplifies into the equivalent of the baseline model, if this were defined on a circle:

$$L_u = \int_0^{\phi} D(\ell) 2\pi \ell d\ell = \rho_r \int_0^{\phi} \frac{1+\epsilon}{\gamma} (w(\phi) + r + \underline{s} - p\underline{c})^{-\frac{1+\epsilon}{\gamma}} (w(\ell) + r + \underline{s} - p\underline{c})^{\frac{1+\epsilon}{\gamma} - 1} 2\pi \ell d\ell.$$

Demand of land for residential purposes in the rural area is unchanged and given in (XVIII). The market clearing condition for land (24), however, needs to be adjusted to read

$$S_r = S - \pi \phi^2 - \frac{L_r \gamma_r \left(w_r + r + \underline{s} - p\underline{c} \right)}{\rho_r}. \quad \text{(XXII)}$$

The aggregate land rent definition also needs to be adjust for the circular area and becomes

$$rL = \int_0^{\phi} \rho(\ell) 2\pi \ell d\ell + \rho_r \times (S_r + S_{hr}). \quad \text{(XXIII)}$$

The aggregate per capita income y net of spatial frictions in the economy that is spent on both goods becomes,

$$y = r + \frac{L_r}{L} w_r + \frac{1}{L} \int_0^{\phi} w(\ell) D(\ell) 2\pi \ell d\ell. \quad \text{(XXIV)}$$

The market clearing condition for rural goods is unchanged. The last modification regards the market clearing for urban goods,

$$c_u + \frac{1}{L} \int_0^{\phi} \tau(\ell) D(\ell) 2\pi \ell d\ell + \frac{1}{L} \int_0^{\phi} \frac{\epsilon(\ell)}{1+\epsilon(\ell)} q(\ell) H(\ell) 2\pi \ell d\ell + \frac{1}{L} \frac{\epsilon_r}{1+\epsilon_r} q_r H_r = y_u, \quad \text{(XXV)}$$

where $y_u = \frac{Y_u}{L}$ denotes the production per worker of the urban good.

C.1.5 Dynamic Model

The value of the real interest rate in the economy is obtained from the intertemporal optimization of ex-ante identical, infinitely-lived agents. Their intertemporal utility is given by

$$U_t = \sum_{s=t}^{\infty} \beta^{s-t} \overline{u}_s,$$

where β is the discount factor and $\overline{u}_t$ denotes the expected utility flow at period t. The expected utility flow at period t is derived from the agents' optimization, in which they maximize

$$\frac{1}{L}\int_0^{\phi} 2\pi\ell D(\ell)\log(C_t(\ell))d\ell + \frac{L_r}{L}\log(C_{r,t}),$$

subject to

$$p_t c_{r,t}(\ell) + c_{u,t}(\ell) + q_t(\ell)h_t(\ell) = w_t(\ell) + r_t + R_t B_{t-1} + B_t, \forall \ell,$$

where R_t denotes the real interest rate at period t, B_t denotes the amount of risk-free asset purchased at period t, and the other variables are defined in Section 3. Using the optimal expenditures described in equations (10)–(12) and the fact that in equilibrium $B_t = 0\ \forall t$, the interest rate is given by

$$R_t = \frac{1}{\beta}\frac{\widehat{u'_t}}{\widehat{u'_{t+1}}},$$

where $\widehat{u'} = \frac{1}{L}\int_0^{\phi}\frac{2\pi\ell D(\ell)}{w(\ell)+r+\underline{s}-p\underline{c}}d\ell + \frac{L_r}{L}\frac{1}{w_r+r+\underline{s}-p\underline{c}}$.

Once we have the value of the real interest rate for every period, we compute land and housing values at each location as the discounted sum of future land rents. We simulate the model until the year 2350 to have at least 120 future periods in the computation. To simulate the model forward, we assume a productivity growth rate after 2000 of 1% in both sectors and a population growth rate after 2040 (last year for INSEE population projections) equal to 0.4%, which corresponds to the average growth in the period 1990-2040.

The housing price index is computed as the total value of housing in the economy divided by the total units of housing, which corresponds to average housing purchasing price. The total value of housing, W_t^h, is equal to the value of housing in the city plus the value of housing outside the city, and it is computed as

$$W_t^h = \int_0^{\phi} 2\pi\ell H_t(\ell)Q_t(\ell)d\ell + \int_{\phi}^{\sqrt{S/\pi}} 2\pi\ell\frac{S_{hr}}{S_{hr}+S_r}H_t(\ell)Q_t(\ell)d\ell,$$

where the purchasing price of a housing unit in location ℓ, $Q_t(\ell)$, is computed as the discounted sum of future rental prices until a final period T:

$$Q_t(\ell) = \sum_{s=t}^{T}\frac{q_s(\ell)}{R_s}.$$

The total units of housing, HH_t, is equal to the housing units in the city plus the housing units outside the city, which is computed as

$$HH_t = \int_0^{\phi} 2\pi\ell H_t(\ell)d\ell + \int_{\phi}^{\sqrt{S/\pi}} 2\pi\ell\frac{S_{hr}}{S_{hr}+S_r}H_t(\ell)d\ell.$$

Finally, the Housing Price Index is computed as

$$HPI_t = \frac{W_t}{HH_t}.$$

We then deflate this value using an implied GDP-deflator that takes the geometric average of the Laspeyres and the Paasche price index.

C.1.6 Numerical Integration over Circular City

In this subsection we define some integrals over space. We limit ourselves to the definition for the full model on a circular surface described in C.1.4, the integrals simplify accordingly for the city on a line segment.

We use Gaussian quadrature to compute numerical approximations to our integrals. To this end we use $N_j = 64$ Gauss-Legendre integration weights $\omega \in (0, 1)$ and nodes $\hat{z}$, which are defined in $[-1, 1]$. We map the nodes into $z \in [0, \phi]$. By way of example, consider the city size Equation (XXI), where we want to approximate the integral of residential density D at location ℓ over a circle with radius ϕ:

$$\begin{aligned} L_u &= \int_0^{\phi} D(\ell) 2\pi \ell d\ell \\ &\approx \frac{\phi - 0}{2} \sum_{j=1}^{N_j} \omega_j D(z_j) 2\pi z_j \end{aligned}$$

We use the second line in our numerical implementation – notice the initial fraction as a result of mapping nodes into $[0, \phi]$. The approach to integrate other expressions (integral of urban good consumption, for example) is identical. In most cases, $D(\ell)$ above is the weight attached to a given model outcome. A slight difference occurs if we compute urban-population averages – for example the average commuting speed as mentioned in the main text. Define function $m(\ell, L_u)$ as (XI). To obtain the average, we have to divide by the number of urban residents to obtain the correct per-capita representation:

$$\begin{aligned} \overline{m} &= \frac{1}{L_u} \int_0^{\phi} m(\ell, L_u) D(\ell) 2\pi \ell d\ell \\ &\approx \frac{\phi}{2L_u} \sum_{j=1}^{N_j} \omega_j m(z_j, L_u) D(z_j) 2\pi z_j \end{aligned}$$

C.2 Numerical Model Solution and Parameter Estimation

In this subsection we describe the numerical solution techniques to solve the single city model.

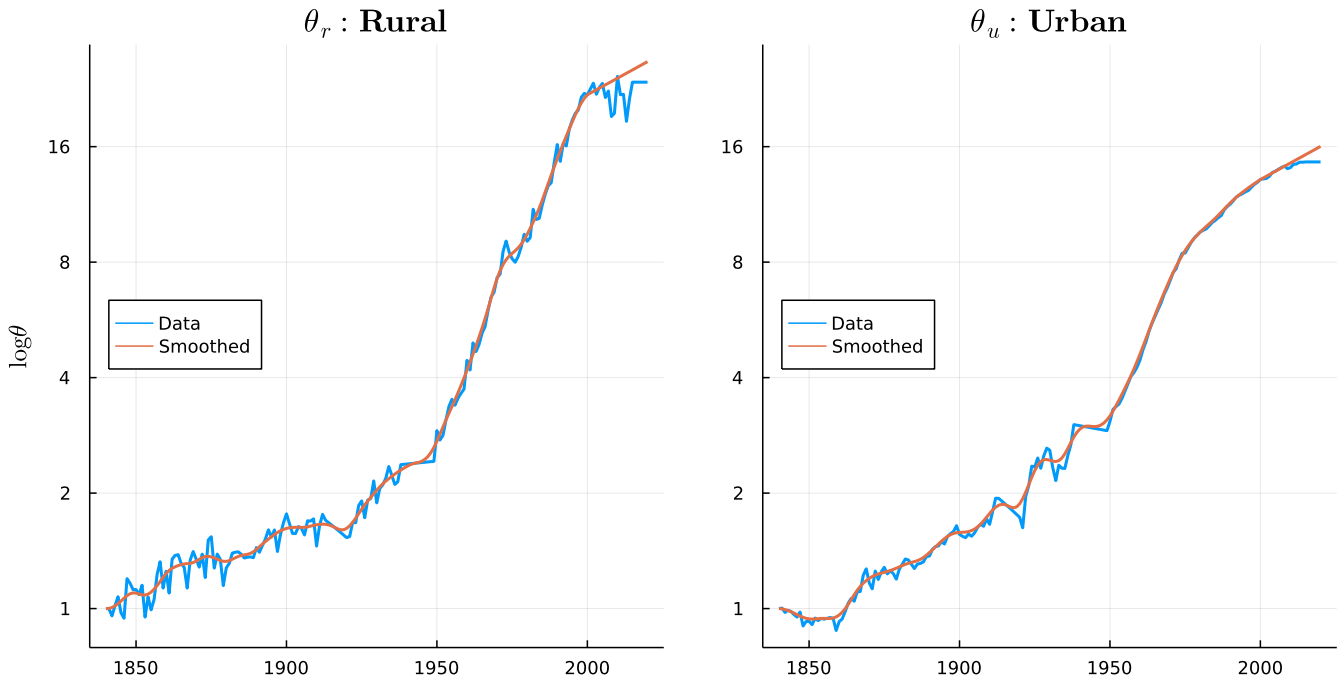


Figure XXXI: Smoothing Procedure applied to estimated productivity data. The rightmost panel are the series for θ which are given to the quantitative model in Section 4. Notice how the data are extrapolated from 2000 onwards assuming constant 1% growth.

C.2.1 Timing and Productivity Series Smoothing

We run the model in time steps of 10 years, starting in 1840 and up until 2020.

A key input for the model is the estimated series of sectoral productivities $\{\theta_{u,t}, \theta_{r,t}\}_{t=1840}^{2020}$, displayed in Figure V. In order to be useable in the model, we need to smooth the displayed data. We explain the necessary steps briefly here:

1. We obtain the estimated series at annual frequency.
2. We subset both series to start in 1840 and end in 2015 (rural productivity ends in that year)
3. We linearly interpolate the missing interwar years.
4. Smoothing is done with a Hann window and a 15-year window size. We experimented with the window size until high-frequency oscillations disappear.
5. Our rural productivity series get very volatile starting at the 2000s. We abstract from this noise by growing the smoothed series forward with 1% annual growth from the year 2000 onwards.

This procedure yields the smoothed series displayed in Figure XXXI.

C.2.2 Single City Model Solution

In this section we describe the solution of the system of equations that describe the model with all extensions as enumerated in Section C.1, hence we give the equation system of model C.1.4 as

$$\mathcal{S} = \begin{cases} \text{(VI)} \quad w_r & = & \alpha p \theta_r \left(\dfrac{S_r}{L_r}\right)^{1-\alpha} \\ \text{(VII)} \quad \rho_r & = & (1-\alpha) p \theta_r \left(\dfrac{L_r}{S_r}\right)^{\alpha} \\ \text{(XXI)} \quad L_u & = & \displaystyle\int_0^{\phi} D(\ell) 2\pi \ell d\ell \\ \text{(XXIII)} \quad rL & = & \displaystyle\int_0^{\phi} \rho(\ell) d\ell + \rho_r \times (1-\phi) \\ \text{(XXII)} \quad S_r & = & S - \pi\phi^2 - \dfrac{L_r \gamma_r \left(w_r + r + \underline{s} - p\underline{c}\right)}{\rho_r} \\ \text{(XXV)} \quad y_u & = & c_u + \dfrac{1}{L}\displaystyle\int_0^{\phi} \tau(\ell) D(\ell) 2\pi \ell d\ell + \dfrac{1}{L}\int_0^{\phi} \dfrac{\epsilon(\ell)}{1+\epsilon(\ell)} q(\ell) H(\ell) 2\pi \ell d\ell + \dfrac{1}{L}\dfrac{\epsilon_r}{1+\epsilon_r} q_r H_r \end{cases}$$

The solution proceeds by choosing values for $x = \{\rho_r, \phi, r, L_r, p, S_r\}$ which, given parameters and two values for sectoral productivities (θ_u, θ_r) solves the system of equations $\mathcal{S}$. An important consideration concerns valid starting values x_0: once a valid starting point for the first period is found, we supply the solution x_{t-1} as a starting point for period t's algorithm. We explain generation of starting values in C.2.3 below.

C.2.3 Starting Values

We generate valid starting values for the single city model in the following way.

1. Given parameters $(\alpha, \theta_u, \theta_r, \gamma, \nu, \epsilon_r, \underline{s}, \underline{c})$, specify a two-sector model (rural and agricultural production) but without commuting costs. We search over rural land rent ρ_r and rural workforce L_r in order to satisfy a land market clearing condition and a feasibility constraint on the economy. We obtain thus $\left(\rho_r^{(0)}, L_r^{(0)}\right)$.
2. We can compute the remaining entries of starting vector $x^{(0)}$ with those values in hand.
3. We return $\phi/10$ to ensure the initial city is not too big to aid the first period solution.

This procedure is sufficient to run the baseline model and to explore a limited range of parameter values. For estimation of the model, however, we are confronted with convergence issues when moving too far away from the initial value generated by this simple procedure. We therefore upgrade the procedure in the following section.

C.2.4 Estimation Procedure

For estimation, we choose the vector $\Gamma \in \Theta$ with following elements and spaces:

$$\Gamma = \begin{cases} \underline{c} & \in (0.7, 0.9) \\ \underline{s} & \in (0.2, 0.26) \\ \nu & \in (0.02, 0.03) \\ a & \in (2.0, 3.0) \\ \gamma & \in (0.28, 0.33) \end{cases}$$

We create a Cartesian grid Θ over this five-dimensional space and evaluate the model at each parameter value using the procedure described in C.2.3. The solution encounters invalid points in a highly irregular fashion - in particular, non-monotonic in any particular parameter's space. We therefore employ a deep learning procedure to impute the starting values for combinations of Γ which result in invalid starting values.

We train a neural network with 4 dense layers, where the first 3 have a RELU activation function and the final layer is linear, in order to map a 5-dimensional parameter vector Γ into a 6-dimensional starting point $x^{(0)}$, which produces a valid model solution. Our data are all valid starting points obtained from our grid evaluation mentioned above. We split data into training (70%) and test samples and we use gradient descent to optimize a MSE loss function.

We can use the resulting neural network to generate starting values which allow evaluation of the model anywhere inside the above described parameter space Θ. Estimation involves solving the standard GMM optimization problem

$$\min_{\Gamma \in \Theta} L(\Gamma) = \min_{\Gamma \in \Theta} [m - m(\Gamma)]^T W [m - m(\Gamma)]$$

where m is a data moment and $m(\Gamma)$ is its model-generated counterpart. Both sets of values are displayed in Section C.2.5. We optimize this loss function with a differential evolution optimizer from the `BlackBoxOptim.jl` package.[31] Notice that we focus here solely on achieving the best fit of the model to our main data moments (leaving aside considerations related to optimal weighting for inference purposes), hence we set the weights on the diagonal of W in order to ensure that those moments do not vanish in the gradient of the moment function.

C.2.5 Estimation Results

The estimated parameter Table VI implies the following moment fits of both targeted (Table VII and non-targeted (Table VIII) moments.

[31] https://github.com/robertfeldt/BlackBoxOptim.jl

Parameter	Description	Value
S	Total Space	1.0
L_0	Total Population in 1840	1.0
θ_0	Initial Productivity in 1840	1.0
α	Labor Weight in Rural Production	0.75
σ	Land-Labor Elasticity of Substitution	0.9999
ν	Preference Weight for Rural Consumption Good	0.0296
γ	Utility Weight of Housing	0.3017
$\underline{c}$	Rural Consumption Good Subsistence Level	0.7368
$\underline{s}$	Initial Urban Good Endowment	0.2106
β	Discount Factor	0.935
ξ_l	Elasticity of commuting cost wrt location	0.55
ξ_w	Elasticity of commuting cost wrt urban wage	0.75
a	Commuting Costs Base Parameter	2.2492
ϵ_r	Housing Supply Elasticity in rural area	5.0
$\epsilon(0)$	Housing Supply Elasticity at city center	2.0

Table VI: Parameter values

Moment	Data	Model	Weight
rural_emp_1840	0.6019	0.7029	0.01
rural_emp_1850	0.5625	0.6392	0.01
rural_emp_1860	0.5248	0.5612	0.01
rural_emp_1870	0.5018	0.5059	0.01
rural_emp_1880	0.4677	0.5158	0.01
rural_emp_1890	0.4433	0.4756	0.01
rural_emp_1900	0.4172	0.41	0.01
rural_emp_1910	0.413	0.4008	0.01
rural_emp_1920	0.4149	0.4053	0.01
rural_emp_1930	0.3618	0.3129	0.01
rural_emp_1940	0.3573	0.2644	0.01
rural_emp_1950	0.2994	0.2235	0.01
rural_emp_1960	0.2255	0.1447	0.01
rural_emp_1970	0.1427	0.0877	0.01
rural_emp_1980	0.0914	0.0713	0.01
rural_emp_1990	0.0615	0.0527	0.01
rural_emp_2000	0.0432	0.043	0.01
rural_emp_2010	0.0337	0.0413	0.01
rural_emp_2020	0.0313	0.0396	0.01
rel_city_area_2010	0.173	0.1741	10.0
housing_share_1900	0.237	0.2377	5.0
housing_share_2010	0.306	0.3028	5.0

Table VII: Moment function at optimal parameter vector in Table VI.

Moment	Data	Model
avg_density_fall	7.9	6.0405
max_mode_increase	5.0	4.8597
density_decay_coef	−0.16	−0.1641
density_decay_MSE	-	0.653

Table VIII: Non-targeted moments at optimal parameter vector in Table VI. Notice that the implied city structure (fall in density as measure by the `density_decay_coef`) matches very well with the data. The `density_decay_MSE` provides a measure of goodness-of-fit of the exponential decay model.

D Sensitivity Analysis and Extensions

D.1 Estimation Results with $\epsilon(0) = 2.5$

In our baseline results we used a value $\epsilon(0) = 2$ at the city center. In this section we report results if estimation is performed setting instead a value of $\epsilon(0) = 2.5$. We show resulting parameter values in Table IX, targeted moments in Table X as well as non-targeted moments in Table XI. Parameter estimates are close to identical w.r.t. to the baseline values in Table VI, except for the value of a, which is lower. This makes intuitive sense because a greater supply elasticity in the center would imply higher buildings in the central area, hence the city would become smaller overall. In order to satisfy the urban area data moment, the model thus decreases the base parameter of commuting costs in order to offset this effect. The resulting moments are virtually identical to the baseline model. We see from the non-targeted moments in Table XI that by choosing $\epsilon(0) = 2.5$, we create a city that is too dense in the center: the coefficient of the exponential decay model is too large (density falls too fast, moving away from the center).

Parameter	Description	Value
ν	Preference Weight for Rural Consumption Good	0.03
γ	Utility Weight of Housing	0.3
$\underline{c}$	Rural Consumption Good Subsistence Level	0.74
$\underline{s}$	Initial Urban Good Endowment	0.21
a	Commuting Costs Base Parameter	2.14

Table IX: Estimated parameter values when $\epsilon(0) = 2.5$. Most estimates differ after the second digit, the main difference is the value of commuting cost base parameter a.

D.2 Sensitivity with constant $\epsilon = 3$

We perform sensitivity analysis assuming a housing supply elasticity ϵ equal to 3 in all locations. We use the estimated parameter values from Table VI for this exercise. Results are displayed in Figure XXXII for the some variables pertaining to urban expansion and density. Outcomes in the baseline

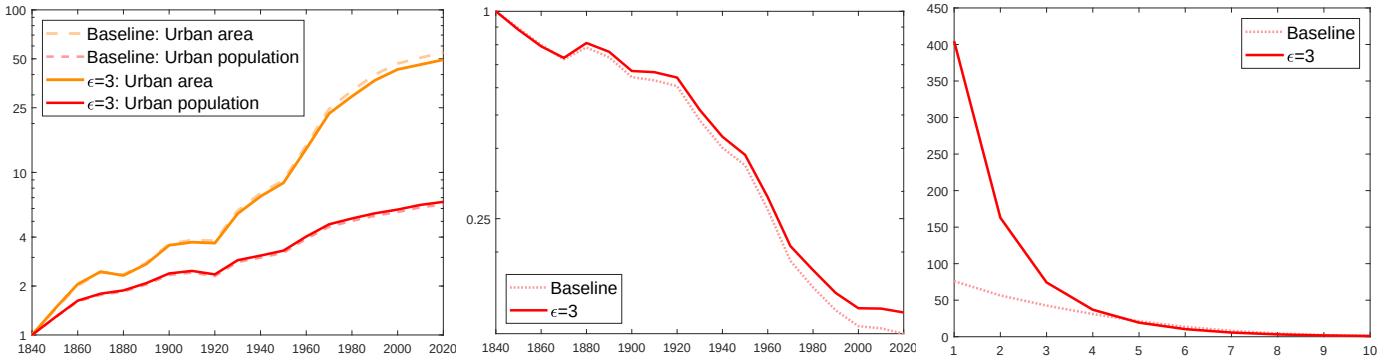
Moment	Data	Model	Weight
rural_emp_1840	0.6019	0.6998	0.01
rural_emp_1850	0.5625	0.6357	0.01
rural_emp_1860	0.5248	0.5573	0.01
rural_emp_1870	0.5018	0.5021	0.01
rural_emp_1880	0.4677	0.5122	0.01
rural_emp_1890	0.4433	0.4719	0.01
rural_emp_1900	0.4172	0.4062	0.01
rural_emp_1910	0.413	0.397	0.01
rural_emp_1920	0.4149	0.4017	0.01
rural_emp_1930	0.3618	0.3091	0.01
rural_emp_1940	0.3573	0.2604	0.01
rural_emp_1950	0.2994	0.2194	0.01
rural_emp_1960	0.2255	0.1405	0.01
rural_emp_1970	0.1427	0.0836	0.01
rural_emp_1980	0.0914	0.0672	0.01
rural_emp_1990	0.0615	0.0487	0.01
rural_emp_2000	0.0432	0.0392	0.01
rural_emp_2010	0.0337	0.0375	0.01
rural_emp_2020	0.0313	0.0358	0.01
rel_city_area_2010	0.173	0.1733	10.0
housing_share_1900	0.237	0.2367	5.0
housing_share_2010	0.306	0.3032	5.0

Table X: Target moments with $\epsilon(0) = 2.5$. The values here are very similar to the ones in Table VII.
.

Moment	Data	Model
avg_density_fall	7.9	6.2571
max_mode_increase	5.0	4.8519
density_decay_coef	-0.16	-0.2268
density_decay_MSE	-	0.257

Table XI: Non-targeted moments with $\epsilon(0) = 2.5$. The important difference to the baseline value $\epsilon(0) = 2$ comes from the too large value of the decay coefficient – here we create a city that is too dense in the center and falls stronger than what we see in the data..

simulation are shown for comparison purposes. With a housing supply elasticity ϵ equal to 3 in all locations the model again generates a city that is too dense in the center.



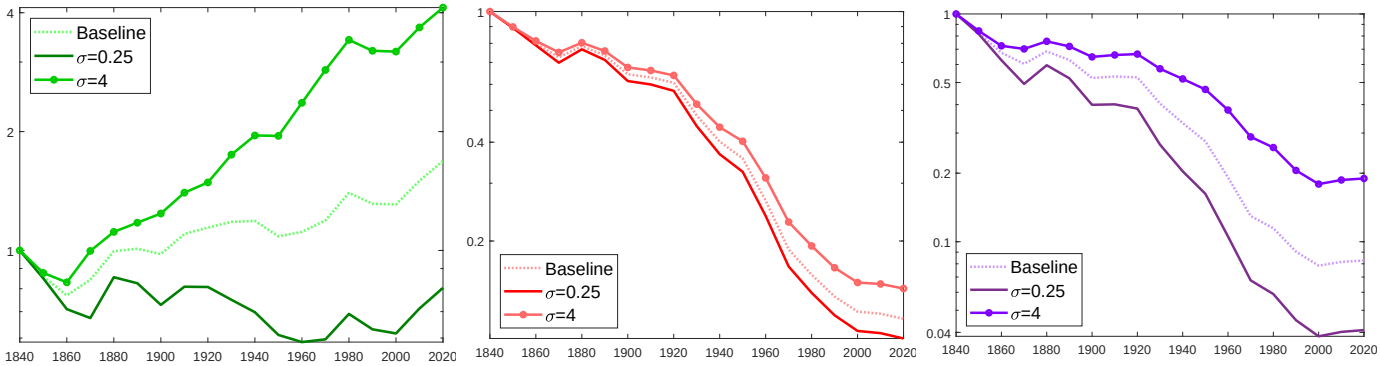
(a) City size (Index, 1840=1). (b) Average density (1840=1). (c) Density by decile (Year 2010).

Figure XXXII: Constant housing supply elasticity, $\epsilon = 3$.

Notes: The housing supply elasticity ϵ is set to 3 in all locations. Other parameters set to their baseline value of Table VI. Outcomes of interest with constant elasticity, $\epsilon = 3$, are displayed with a solid line. The baseline simulation is shown with a dashed line for comparison.

D.3 Sensitivity w.r.t the elasticity of substitution between land and labor

Our baseline assumes a unitary elasticity of substitution between land and labor in the rural sector, $\sigma = 1$. Details of the derivation with a CES production in the rural sector are shown above in C.1.1. We perform sensitivity analysis with a lower value of σ equal to 0.25 and for a higher value of 4. Results for variables of interest discussed in the main text are displayed on Figure XXXIII.



(a) Rental value of farmland (1840=1). (b) Average urban density (1840=1). (c) Density at the urban fringe (1840=1).

Figure XXXIII: Sensitivity to the elasticity of substitution between land and labor σ.

Notes: The elasticity of substitution between land and labor σ is set to a low value of 0.25 (resp. a high value of 4). All other parameters are kept to their baseline value of Table VI. Simulation for the baseline calibration shown in dashed for comparison.

D.4 Agglomeration and Congestion

Agglomeration. We introduce urban agglomeration forces by assuming that the urban productivity increases externally with urban employment, $\theta_u(L_u) = \theta_u \cdot L_u^{\lambda}$. Urban wages, $w_u = \theta_u . L_u^{\lambda}$, thus increase faster relative to our baseline when the city expands. Apart from the adjustment of urban wages, the model is identical to our baseline. So is the numerical solution. It is important to note though that the faster increase in urban wages makes urban workers use faster commutes as the opportunity cost of time increases faster.

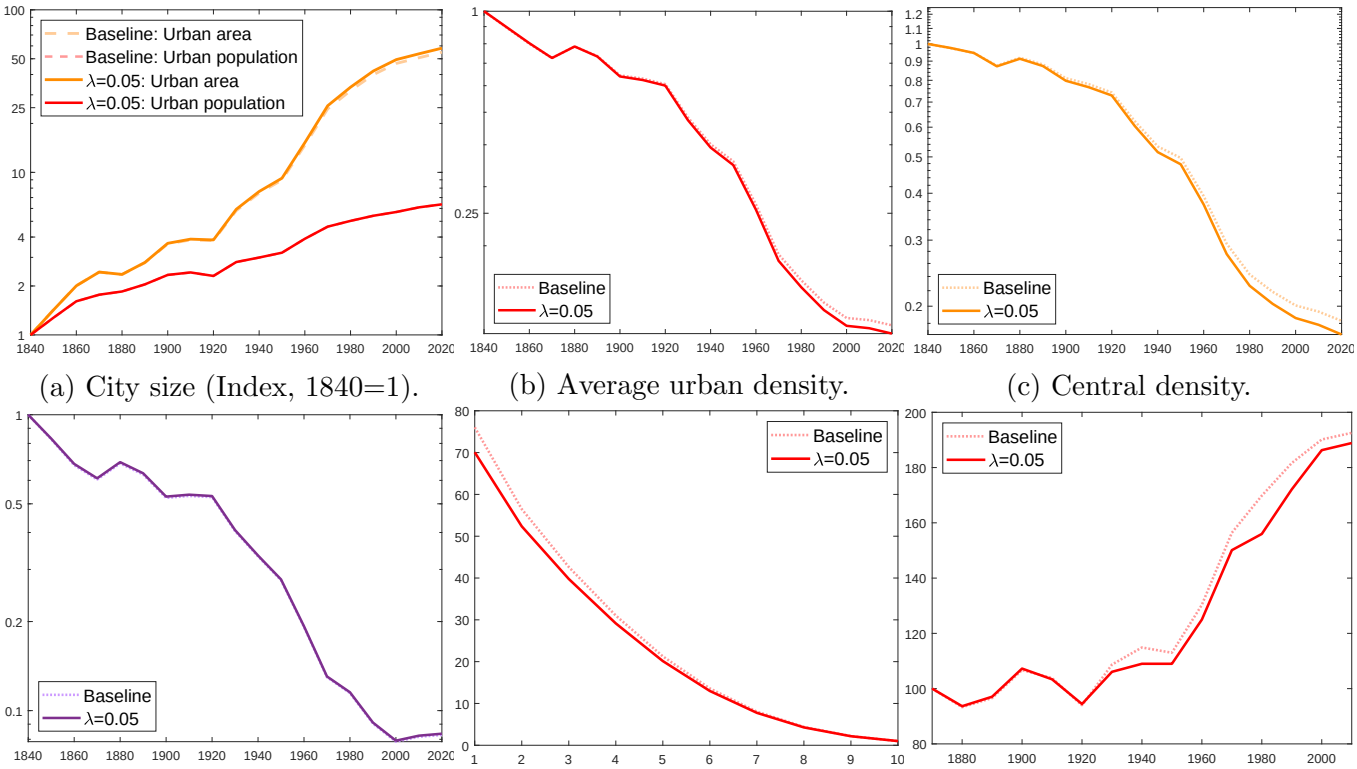
We set $\lambda = 0.05$, in the range of empirical estimates for France (Combes et al. (2010)). Other parameters are left identical to the baseline for comparison, adjusting the initial value of θ_u to have the same initial urban productivity. For variables of interest, results in the presence of agglomeration forces are displayed in Figure XXXIV together with the baseline. While the city expands slightly more in area, there is barely no effect of agglomeration forces on urban population. The faster increase in the urban wage due to agglomeration forces increases urban housing demand and reduces urban commuting costs (as a share of income). This relocates urban households towards the suburbs where they can enjoy larger homes and the city sprawls more. However, a higher urban income makes also rural goods more valuable increasing (almost equally) rural workers' wage. General equilibrium forces thus prevent workers' reallocation towards cities. They also make rural land more valuable—mitigating the area expansion of the city. As a consequence, despite higher incomes driven by urban expansion, the economy with agglomeration forces behave quantitatively similarly to our baseline.

Congestion. We consider additional urban congestion costs by assuming that commuting costs are increasing with urban population, $a(L_u) = a \cdot L_u^{\mu}$. This summarizes the potential channels through which larger cities might involve longer and slower commutes for a given commuting distance. Under such a formulation, commuting costs in location ℓ satisfies,

$$\tau(\ell) = a \cdot L_u^{\mu} \cdot w_u^{\xi_w} \cdot (\ell)^{\xi_\ell}.$$

Apart from the adjustment of commuting costs, which depend on the urban population, the model is identical to our baseline. So, is the numerical solution.

We set $\mu = 0.05$ and we re-scale the constant a to have the same initial value for the commuting costs. Other parameters are set to their baseline values. The evolution of the variables of interest is shown on Figure XXXV with the baseline for comparison. Congestion forces move the equilibrium in the opposite direction of agglomeration forces. They reduce the expansion in area and the extent of suburbanization. By increasing commuting costs, they also increase urban housing prices. However, via general equilibrium forces, they also make rural goods and rural land less valuable—severely mitigating the direct effect of congestion costs on urban expansion.

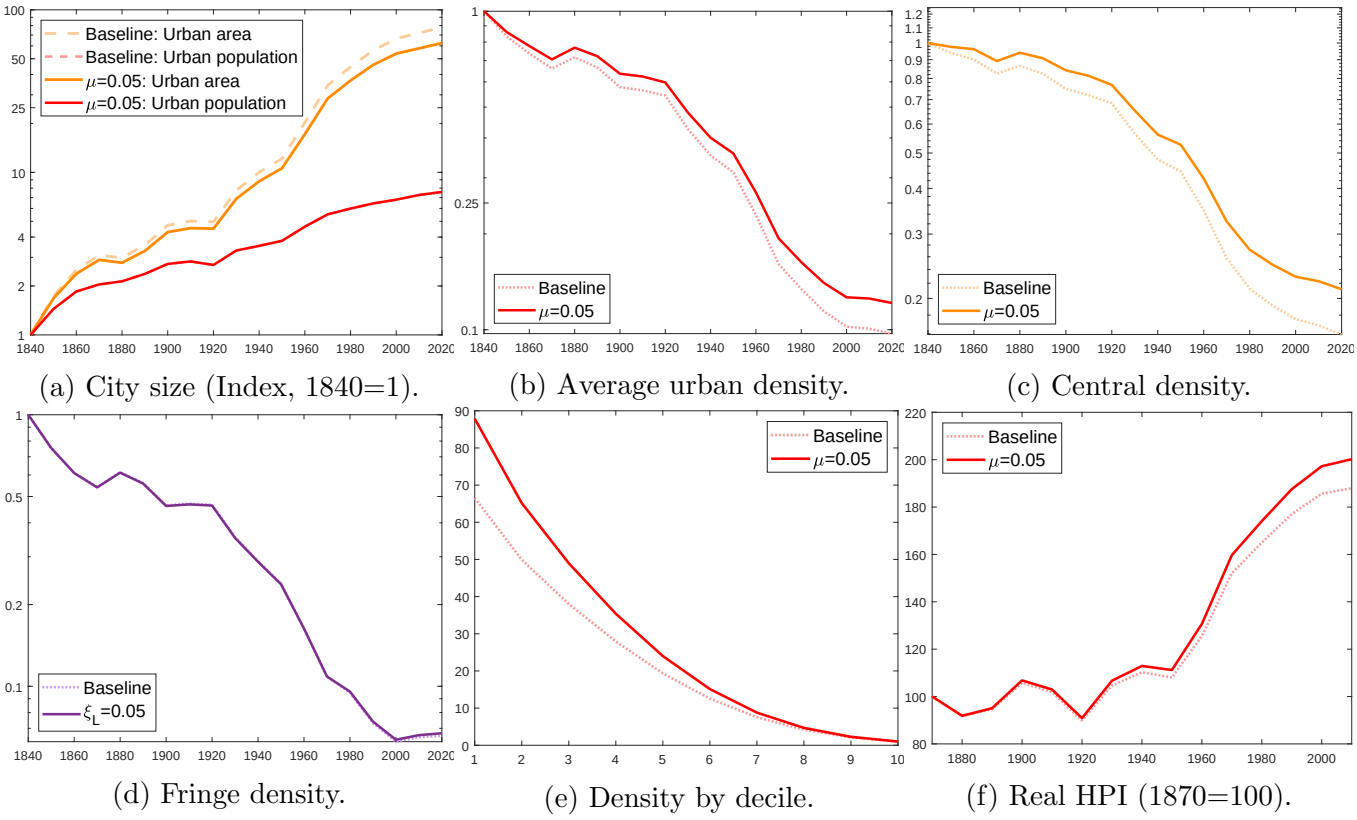


(a) City size (Index, 1840=1). (b) Average urban density. (c) Central density.

(d) Fringe density. (e) Density by decile. (f) Real HPI (1870=100).

Figure XXXIV: Positive agglomeration forces, $\lambda = 0.05$.

Notes: The urban agglomeration forces parameter λ is set to 0.05. Other parameters set to their baseline value of Table VI, while θ_u is adjusted to have the same initial urban productivity. Outcomes of interest with agglomeration forces are displayed with a solid line. The baseline simulation is shown with a dashed line for comparison.



(a) City size (Index, 1840=1). (b) Average urban density. (c) Central density.

(d) Fringe density. (e) Density by decile. (f) Real HPI (1870=100).

Figure XXXV: Congestion costs, $\mu = 0.05$.

Notes: The congestion cost parameter η is set to 0.05. Other parameters set to their baseline value of Table VI, and we re-scale the constant a to have the same initial value for the commuting costs. Outcomes of interest with congestion costs are displayed with a solid line. The baseline simulation is shown with a dashed line for comparison.

D.5 Commuting Distance and Residence Location

Guided by the structure of French cities, our baseline results hinge on the assumption of a monocentric model where urban individuals commute to the city center to work. While endogeneizing firms location across space is beyond the scope of the paper, one can still partly relax the monocentric assumption by assuming that commuting distance, $d(\ell)$, does not map one for one with residential distance ℓ from the central location. In line with empirical observations in Section B.9.2, we model commuting distance $d(\ell)$ in a reduced-form way as follows,

$$d(\ell) = d_0(\phi) + d_1(\phi) \cdot \ell,$$

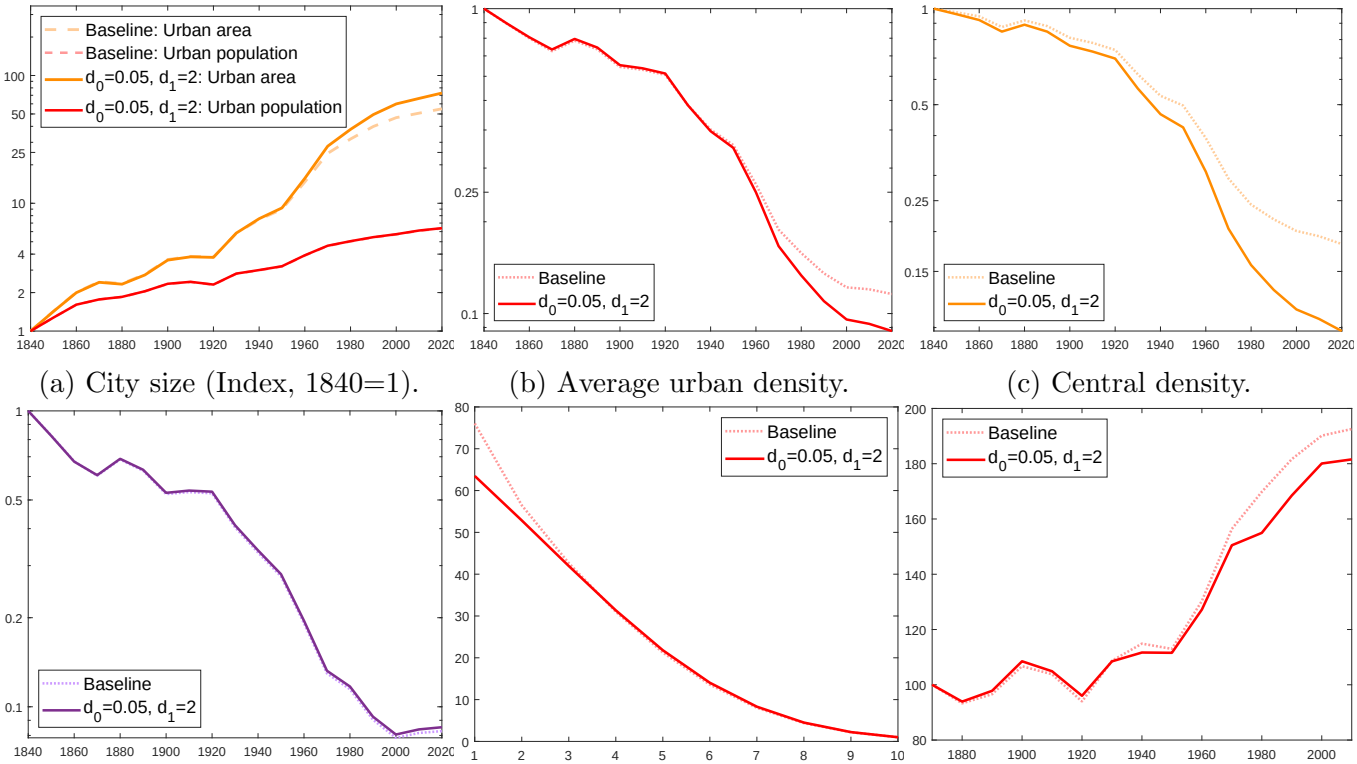
with $d_0(\phi)$ a positive and increasing function of ϕ, satisfying $\lim_{\phi \to 0} d_0(\phi) = 0$, and $d_1(\phi)$ a decreasing function belonging to $(0, 1)$ with $\lim_{\phi \to 0} d_1(\phi) = 1$. d_0 represents the (minimum) commuting distance travelled by an individual living in the center, while d_1 is the slope between commuting distance and residential distance from the center. This specification is consistent with the data (see Section B.9.2). It also makes sure that at the limit of $\phi \to 0$, the city is monocentric as all the jobs must be centrally located. It is important to note that commuting costs are now defined as,[32]

$$\tau(\ell) = a \cdot w_u^{\xi_w} \cdot (d(\ell))^{\xi_\ell}. \tag{XXVI}$$

In the quantitative simulations, we make the following parametric assumptions: $d_0(\phi) = d_0 \cdot \phi$, with d_0 small and positive and $d_1(\phi) = \frac{1}{1+d_1 \cdot \phi}$, with $d_1 \geq 0$. The parameters d_0 and d_1 are guided by the data (Section B.9.2). Across cities, $d_0 \cdot \phi$ corresponds to the intercept of Eq. V, ranging from 0.2 km for the smaller cities to more than 4 kms for Paris. Given that further away residential locations are typically at 5 kms of the center in smaller urban areas and up to 50 kms away from the center of Paris, d_0 should range within 4 and 8%. We set d_0 to 5% in our quantitative experiment. For a radius of about 20 kms (corresponding to the population weighted-mean of our sample of 100 urban areas), a person living in the city center ($\ell = 0$) would commute on average 1 km. Across cities, $d_1(\phi) = \frac{1}{1+d_1 \cdot \phi}$ corresponds to the slope of Eq. V—with an estimated mode across urban areas close 0.7. Given that the model implied radius of our representative city is $\phi = 0.24$, we set $d_1 = 2$. This yields a slope coefficient of $\frac{1}{1+2 \cdot 0.24} = 0.68$ close to the empirical moment.

With these parameters values and the corresponding commuting costs, we simulate the quantitative model. Other parameters are left unchanged with the exception of the commuting costs parameter a. Changing commuting costs according to Eq. XXVI implies a city area in equilibrium different from the targeted one in the baseline version of the model. Thus, for comparison purposes, a is recalibrated

[32] Note that this remains consistent with our calibrated value of ξ_ℓ estimated using commuting distance. The elasticity of speed $m(\ell)$ to commuting distance $d(\ell)$ being $1 - \xi_\ell$.



(a) City size (Index, 1840=1). (b) Average urban density. (c) Central density.

(d) Fringe density. (e) Density by decile. (f) Real HPI (1870=100).

Figure XXXVI: Alternative specification of commuting distance in the single city case, $d(\ell) = d_0 \cdot \phi + \frac{\ell}{1+d_1 \cdot \phi}$.

Notes:The solid line represents outcomes under the alternative specification of commuting distance $d(\ell)$, with parameter $d_0 = 5\%$ and $d_1 = 2$. Other parameters set to their baseline value of Table VI up to a recalibration of the commuting cost parameter $a = 2.67$ to preserve the share of urban land in 2015. For comparison, outcomes of the baseline simulation are shown with a dashed line ($d_0 = d_1 = 0$).

to make sure that the urban footprint still occupies 18% of rural land in the recent period. Results for the variables of interest are displayed in Figure XXXVI.

We find that our results are not much affected under this alternative specification of the commuting costs. Quantitatively, the city expands more in area in the last decades under this specification of the commuting distance, bringing the model closer to the data. As a consequence of this larger sprawling, the average urban density falls more. This is driven by a larger fall of central density, the most noticeable difference relative to our baseline monocentric model. With urban expansion, residents in central locations end up commuting larger distances—implicitly due to the reallocation of jobs away from the center—, this makes central locations less attractive relative to suburban ones.

D.6 Multiple City Model

In this section we extend the model in Section C.1.4, which describes a single region, to allow for K different regions. The spatial structure in each region k is identical to our baseline one-city version: each region k of area S_k is made of urban and rural land, with only one (potential) city per region.[33] Regions are heterogeneous only in their urban productivity—with $\theta_{u,k}$ the urban productivity in region/city k. Workers are freely mobile within and across regions and labor markets clear globally. Urban and rural goods are freely traded within and across regions and goods markets clear globally. For quantitative evaluation, we consider 20 regions, corresponding to the 20 largest French cities in 1870. City-specific urban productivities, $\theta_{u,k}$, are set to match the distribution of population of the different cities over the period 1870-2015, while keeping aggregate urban productivity equal to its baseline value estimated from French historical national accounts.[34] Each region is endowed with the same land area as our baseline. Aggregate population is multiplied by $K = 20$ relative to the baseline to preserve the endowment of land per head.[35] Other parameters are set to their baseline values of Table VI.

D.6.1 Multiple City Model: Setup

We will introduce index x_k to refer to variable x in region k, adding to the end of any existing indices as for example in $\theta_{u,k}$. We will make a distinction between *region* k and *location* ℓ, there the former is meant to denote one of K different subsets of total space S, and the latter to index a location *within a city* k.

Consider thus K different regions, each endowed with land area S_k, where total space is defined as $\sum_{k=1}^{K} S_k = S$. In each location, the productivity of the rural sector is $\theta_{r,k}$. Each region contains a potential city characterized by a level of productivity in the urban sector $\theta_{u,k}$.[36]

We suppose that workers can freely move across regions k. Denote ϕ_k the radius of city k. Mobility across regions k gives the following mobility equations, equalizing the real wage of the urban and rural worker at the fringe across fringe workers in different regions,

$$\overline{C} = \kappa \frac{w_{u,k} - \tau_k(\phi_k) + r + \underline{s} - p\underline{c}}{(q_{r,k})^{\gamma}} = \kappa \frac{w_{r,k} + r + \underline{s} - p\underline{c}}{(q_{r,k})^{\gamma}},$$

[33] For each region, the city center (CBD) is centrally located within each region and regions are assumed large enough in area such that cities do not expand in neighbouring regions.

[34] Denote Ω_k the share of urban population in city k, $\Omega_k = L_{u,k} / \sum_{j=1}^{K} L_{u,j}$. Measured aggregate urban productivity from historical national accounts at each date satisfies in equilibrium, $\theta_u = \sum_{j=1}^{K} \Omega_j \theta_{u,j}$.

[35] This makes sure that, with homogeneous urban productivity, the version with multiple regions behaves like the one-city model in each region.

[36] One could also consider different region specific parameters ϵ_k, τ_k and a region specific amenity ζ_k

where the housing rental value at the fringe of region k satisfies

$$\rho_{r,k} = \frac{(q_{r,k})^{1+\epsilon_{r,k}}}{1+\epsilon_{r,k}} \tag{XXVII}$$

Wages and land values are such that,

$$w_{u,k} = \theta_{u,k}, \tag{XXVIII}$$

$$\rho_{r,k} = (1-\alpha)p\theta_{r,k}\left(\alpha\left(\frac{L_{r,k}}{S_{r,k}}\right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right)^{\frac{1}{\sigma-1}}. \tag{XXIX}$$

Land must clear in every region such that,

$$S_k = S_{r,k} + \phi_k + \frac{L_{r,k}\gamma_{r,k}\left(w_{r,k} + r + \underline{s} - p\underline{c}\right)}{\rho_{r,k}}. \tag{XXX}$$

where $\gamma_{r,k} = \frac{\gamma_k}{1+\epsilon_{r,k}}$. Labour must clear globally,

$$\sum_{k=1}^{K} L_k = \sum_{k=1}^{K}(L_{r,k} + L_{u,k}) = L. \tag{XXXI}$$

Goods market must also clear globally

$$\sum_{k=1}^{K}\int_0^{S_k} c_{u,k}(l)D_k(l)2\pi l dl + w_{u,k}\int_0^{\phi_k}\tau_k(l)D_k(l)2\pi l dl + \int_0^{\phi_k}\frac{\epsilon_k(l)}{1+\epsilon_k(l)}q_k(l)H_k(l)2\pi l dl = \sum_{k=1}^{K} Y_{u,k} \tag{XXXII}$$

We consider the case where housing supply conditions/rural productivity are identical across regions. In this case, one can show that $\frac{L_{r,k}}{S_{r,k}}$, $q_{r,k}$, $\rho_{r,k}$ and $w_{r,k}$ are equalized across regions (homogeneous rural sector). Then, using the mobility condition at the fringe of the different cities, one gets that the difference $\theta_{u,k} - \tau_k(\phi_k)$ is constant and equal to w_r across regions: more productive cities will be larger in area— the reason they will host a larger population in equilibrium. This can be seen using Equation (XXXIII) in region k,

$$L_{u,k} = \int_0^{\phi_k} D(\ell)2\pi\ell d\ell = \int_0^{\phi_k}\left(\frac{q_r^{1+\epsilon(\ell)}}{1+\epsilon(\ell)}\right)\frac{1}{\gamma_\ell}(w(\phi_k) + r + \underline{s} - p\underline{c})^{-1/\gamma_\ell}(w(\ell) + r + \underline{s} - p\underline{c})^{1/\gamma_\ell - 1}2\pi\ell d\ell \tag{XXXIII}$$

where q_r denotes the rural housing price equalized in all regions. This last equation pins down the relative population of cities. Together with the equalization of $\frac{L_{r,k}}{S_{r,k}}$ and the indifference condition at the fringe across regions,

$$\theta_{u,k} - \tau_k(\phi_k) = w_r, \forall k \tag{XXXIV}$$

$$\frac{L_{r,k}}{S_{r,k}} = b, \forall k \tag{XXXV}$$

Land market clearing in all regions (XXX) and the city size equations (XXXIII) pin down $\{L_{u,k}, L_{r,k}, \phi_k, S_{r,k}\}$ for given aggregate prices common across regions ρ_r and p. These two aggregate prices are determined using goods and labor market clearing.

D.6.2 Multiple City Model: Numerical Solution

Solution of the multi city model is very similar to the single city case outlined in section C.2.2 with a slight modification on the sequence of urban productivities $\theta_{u,k,t}$ chosen in each period and for each city k. We formalize the problem as follows in a certain period t:

$$\min_{\{\theta_{u,j,t}\}_{j=1}^{K}} \quad \sum_{j=1}^{K} \left(\frac{L_{u,j,t}}{L_{u,1,t}} - \zeta_{j,t} \right)^2 \tag{XXXVI}$$

$$\text{subject to}$$

$$\sum_{j=1}^{K} \Omega_{j,t} \theta_{u,js,t} = \theta_{u,t} \tag{XXXVII}$$

$$\text{and subject to} \quad (\text{XXVIII}), (\text{XXIX}), (\text{XXX}), (\text{XXXI}), (\text{XXXV})$$

where $\zeta_{k,t}$ is the empirical ratio of urban population between city k in our dataset and the largest city (indexed by $k = 1$), and where $\Omega_{k,t}$ is the urban population weight of city k in the same dataset. The objective function (XXXVI) tries to generate a population distribution across cities that is similar to what we observe in the data, by choosing K values of urban productivity $\theta_{u,k,t}$. This choice is subject to constraint (XXXVII) saying that the weighted summarize of city-specific productivities $\theta_{u,k,t}$ needs to replicate the level of productivity which was given to the *average city* in the baseline model.

D.6.3 Multiple City Model: Results

Here we present results for the multi-city extension introduced in Section D.6.1.

For aggregate variables of interest (aggregate sectoral employment, aggregate land use, relative price of rural goods/farmland, ...), the model with multiple regions behaves quantitatively very similarly to the baseline with only one city. Intuitively, the one-city model describes well the dynamics of aggregate variables for a representative 'average' city. We illustrate in Figure XXXVII how the time series of the single baseline city evolves next to the time series of an average over our 20 cities for urban density, relative price of rural good, and the rural employment share.

We focus next on the dispersion of city size and density, which are novel outcomes relative to our baseline. This is shown in Figure XXXVIII. The first panel shows how urban population relative to the largest city evolves in city and model. We achieve a perfect fit with our solution. This is a mechanical result as a consequence of how we set up the solution as a constrained optimization problem (we set the population distribution as a constraint, as explained in Section D.6.2).

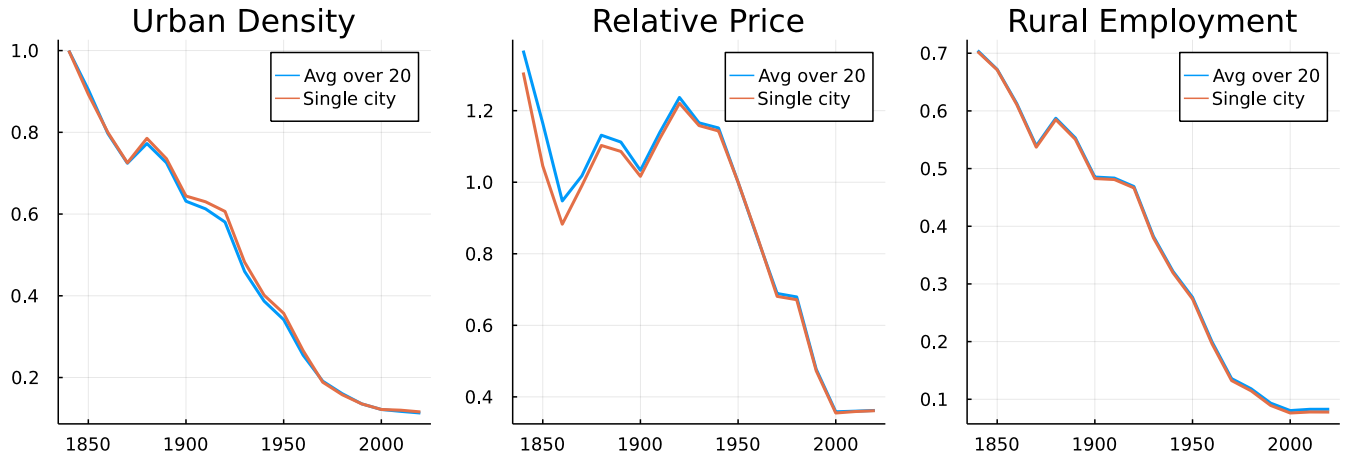


Figure XXXVII: Comparing the average of the top 20 French cities with the single city from the baseline model. We normalize average urban density to its first value (1840), and we normalize the relative price series to its value in 1950.

Panel 2 of Figure XXXVIII shows the dispersion of area (in log) in the model and in the data for dates 1870, 1950, 1975, 1990 and 2015.[37] As visible from this second panel, there are significant cross sectional variations in urban area that the model does not reproduce. While the correlation between model and data is very high, area in the model and in the data do not move one for one—the elasticity being significantly above unity. The model does not generate a large enough area in more populated cities relative to smaller ones, particularly so in a given year. As written in the main text, large cities are too dense in the model's cross-section although the model performs relatively well in the time-series. Note that we abstract from the use of land of economic activities but also from reallocation of jobs away from central locations due to the monocentric structure. To the extent that large cities reallocate more jobs away from the center, our model under-predicts their area. The cross-sectional variations in area also relates to city-specific land use regulations, which are not considered in our theory.

Panel 3 of Figure XXXVIII plots the log of average urban density in a given city against its data counterpart for 1870, 1950, 1975, 1990 and 2015.[38] It shows that the model implied cross sections of urban densities provides a good fit to the data. The elasticity is reasonably close to unity. The model predicts that, over time, for a given city, urban density falls as urban population increases—in line with the predictions of our baseline one-city model. In the cross-section, more populated cities are however denser as they feature higher housing prices. At a given date, the reallocation of workers towards a more productive city does not imply the strong general equilibrium effects on rural (good and farmland) relative prices, which lies at the heart of the time series evolution of urban density

[37] We use the model's predictions for 1970 and 2010 as counterparts for 1975 and 2015 as our model provides solutions on 10-year intervals.

[38] As for area, we take 1970 and 2010 instead because the model is run in 10-year steps.

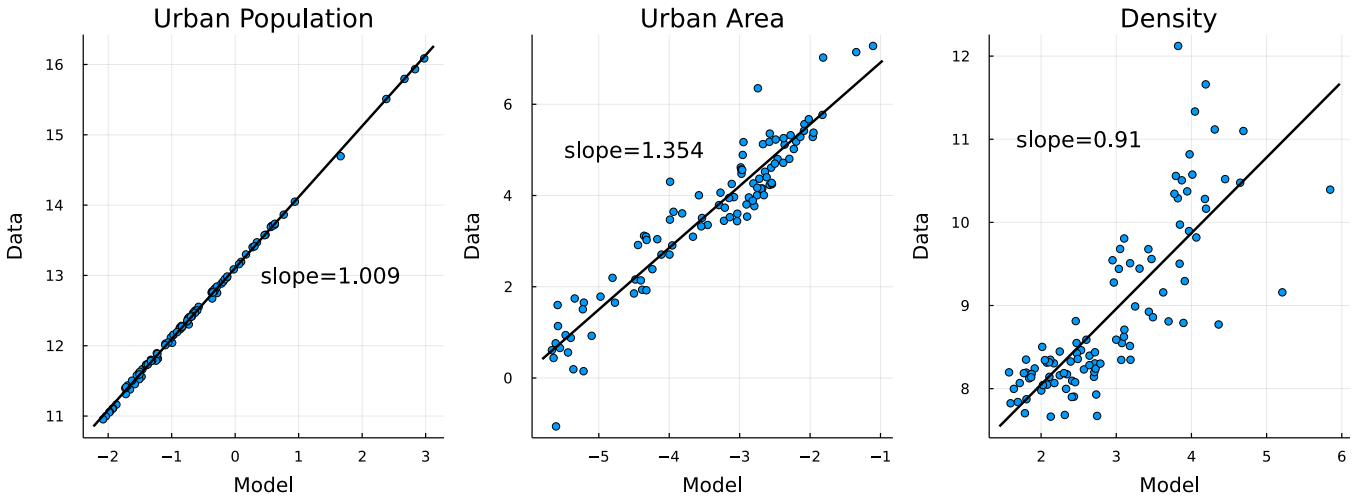


Figure XXXVIII: Cross sections of model vs data

Notes: We plot log model vs log data outcomes, combined across dates 1870, 1950, 1975, 1990 and 2015. An ideal model would yield an elasticity of one. The first panel shows the fit of the population distribution across cities and time. Panel two shows how urban area lines up between model and data. Panel three finally shows the urban density cross section.

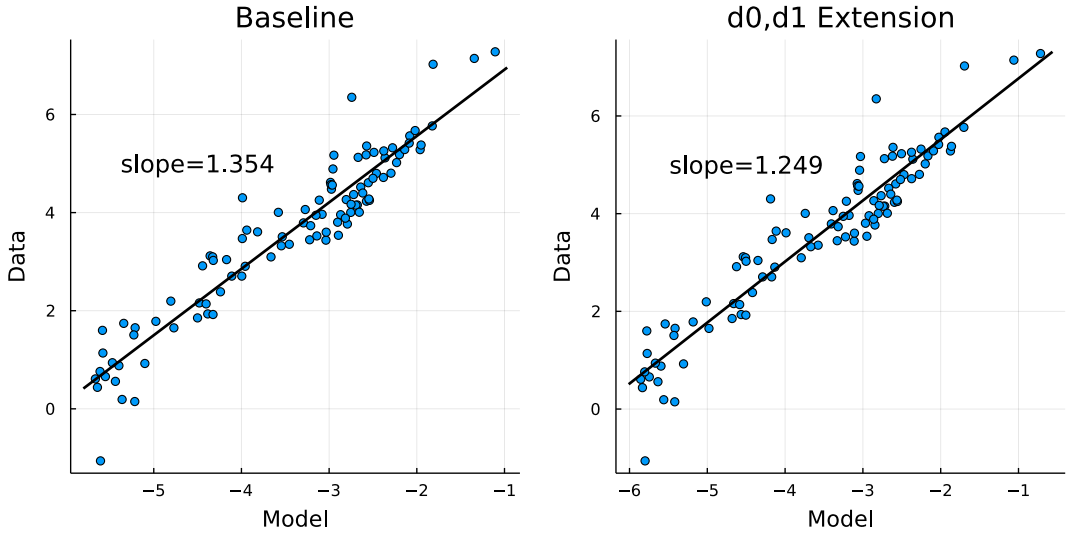


Figure XXXIX: Urban Area Cross Sections. Sensitivity

Notes: Here we illustrate the impact of relaxing monocentricity on the distribution of urban area with the d_0, d_1 extension. We plot the log of model areas vs the log of areas in the data. The corresponding elasticity gets closer to one as we implement the extension, meaning the model is better able to generate larger cities.

when cities grow in size. It is important to note that qualitatively, both predictions, over time and in the cross-section, are in line with the data. Quantitatively, the model does notably better in the time-series than in the cross-section. As discussed above, in a given year, more populated cities are not large enough in area relative to smaller ones—they are thus significantly denser in the model than they are in the data.

As last sensitivity analysis, we also consider the model with multiple-cities where, in each city, residential location does not map one for one into commuting distance. Using the same parametric assumptions, we extend Eq. XXVI for cities of different areas,

$$d_k(\ell) = d_0 \cdot \phi_k + \left(\frac{1}{1 + d_1 \cdot \phi_k}\right) \cdot \ell,$$

where ϕ_k is the equilibrium radius of city k.

The fit between model and data improves under this specification for commuting distance. In line with the data presented in Appendix B.9, commuting distances in the center (resp. at the fringe) are larger (resp. lower) in larger cities in this specification relative to the baseline monocentric model. This, in turn, increases the area of more populated cities in the cross-section at a given date, reducing their average density and bringing the model closer to the data. More populated cities in the model are still noticeably denser than in the data, but less so compared to the baseline monocentric model. The improvement comes from the relative urban area distribution, which fits the data better with this last extension. When looking at areas (in log), the slope between model and data gets closer to unity as illustrated by Figure XXXIX.

References

Angel, Shlomo, Alejandro M Blei, Daniel L Civco, and Jason Parent, *Atlas of urban expansion*, Lincoln Institute of Land Policy Cambridge, MA, 2012.

_ , Jason Parent, Daniel L Civco, and Alejandro M Blei, *Persistent Decline in Urban Densities: Global and Historical Evidence of 'Sprawl'*, Lincoln Institute of Land Policy., 2010.

Augé-Laribé, Michel, "Les statistiques agricoles," in "Annales de géographie," Vol. 53 JSTOR 1945, pp. 81–92.

Bairoch, Paul, "Les trois révolutions agricoles du monde développé: rendements et productivité de 1800 à 1985," *Annales*, 1989, pp. 317–353.

Bastié, Jean, "La population de l'agglomération parisienne," in "Annales de géographie," Vol. 67 JSTOR 1958, pp. 12–38.

Baum-Snow, Nathaniel and Lu Han, "The microgeography of housing supply," *Work in progress, University of Toronto*, 2019.

Bellefon, Marie-Pierre De, Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin, "Delineating urban areas using building density," *Journal of Urban Economics*, 2019, p. 103226.

Bertillon, Jacques, "L'accroissement de la circulation à Londres et à Paris," *Journal de la société française de statistique*, 1910, *51*, 381–397.

Bertrand, Pierre and Jean Hallaire, "Une enquête sur les déplacements journaliers des personnes actives de la région parisienne ou migrations alternantes," *Journal de la société française de statistique*, 1962, *103*, 186–217.

Brunet, Jean-Paul, "Le mouvement des migrations journalières dans l'agglomération parisienne au cours de l'entre-deux-guerres," *Villes en Parallèle*, 1986, *10* (1), 250–269.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux, "Estimating agglomeration economies with history, geology, and worker effects," in "Agglomeration economics," University of Chicago Press, 2010, pp. 15–66.

_ , Laurent Gobillon, Gilles Duranton, and Clement Gorin, "Extracting Land Use from Historical Maps Using Machine Learning: The Emergence and Disappearance of Cities in France Extracting Land Use from Historical Maps Using Machine Learning: The Emergence and Disappearance of Cities in France," *mimeo*, 2021.

Corbane, Christina, Aneta Florczyk, Martino Pesaresi, Panagiotis Politis, and Vasileios Syrris, "GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A. European Commission, Joint Research Centre (JRC) doi:10.2905/jrc-ghsl-10007," 2018.

_ , Martino Pesaresi, Thomas Kemper, Panagiotis Politis, Aneta J. Florczyk, Vasileios Syrris, Michele Melchiorri, Filip Sabo, and Pierre Soille, "Automated global delineation of human settlements from 40 years of Landsat satellite data archives," *Big Earth Data*, 2019, *3* (2), 140–169.

Desriers, Maurice, "L'agriculture française depuis cinquante ans: des petites exploitations familiales aux droits à paiement unique," *Agreste cahiers*, 2007, *2*, 3–14.

European Union, "CORINE Land Cover Data: EU Land Monitoring Service 2018, Copernicus , European Environment Agency (EEA)."

Fléchey, Edmond, "La statistique agricole décennale de 1892," *Journal de la société française de statistique*, 1898, *39*, 321–333.

Florczyk, Aneta J, Christina Corbane, Daniele Ehrlich, Sergio Freire, Thomas Kemper, Luca Maffenini, Michele Melchiorri, Martino Pesaresi, Panagiotis Politis, Marcello Schiavina et al., "GHSL data package 2019," *Publications Office of the European Union, Luxembourg*, doi:10.2760/290498 2019, *29788* (10.2760), 290498.

Freire, Sergio, Erin Doxsey-Whitfield, Kytt MacManus, Jane Mills, and Martino Pesaresi, "Development of new open and free multi-temporal global population grids at 250 m resolution," in "https://agile-online.org/conference_paper/cds/agile_2016/shortpapers/152_Paper_in_PDF.pdf" Association of Geographic Information Laboratories in Europe (AGILE) 2016.

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi, "Growth and structural transformation," in "Handbook of economic growth," Vol. 2, Elsevier, 2014, pp. 855–941.

Hitier, Henri, "La statistique agricole de la France," in "Annales de Géographie," Vol. 8 JSTOR 1899, pp. 350–357.

Marchand, Olivier and Claude Thélot, "Deux siècles de travail en France: population active et structure sociale, durée et productivité du travail," 1991.

Martin, Alfred, *Étude historique et statistique sur les moyens de transport dans Paris, avec plans, diagrammes et cartogrammes*, Imprimerie nationale, 1894.

Mauco, Georges, "Les modes d'exploitation agricole en France," in "Annales de Géographie," Vol. 46 JSTOR 1937, pp. 485–493.

Mauguin, Ch, "Statistique comparée de l'agriculture française en 1790 et en 1882," *Journal de la société française de statistique*, 1890, *31*, 200–213.

Merlin, Pierre, "Les transports en région parisienne," *Notes et études documentaires (Paris)*, 1997, (5052).

Orselli, Jean, "Usages et usagers de la route: pour une histoire de moyenne durée (1860-2008)." PhD dissertation, Paris 1 2008.

Papon, Francis, Marina Marchal, Sophie Roux, Philippe Marchal, and Jimmy Armoogum, "Parcours individuels et histoire de la mobilité. Analyse du volet "biographie" de l'Enquête Nationale sur les Transports et les Déplacements 2007-2008," 2010.

Piketty, Thomas and Gabriel Zucman, "Capital is back: Wealth-income ratios in rich countries 1700–2010," *The Quarterly Journal of Economics*, 2014, *129* (3), 1255–1310.

Sauvy, Alfred, "Variations des prix de 1810 à nos jours," *Journal de la société française de statistique*, 1952, *93*, 88–104.

Schiavina, Marcello, Sergio Freire, and Kytt MacManus, "GHS population grid multitemporal (1975-1990-2000-2015), R2019A. European Commission, Joint Research Centre (JRC) [Dataset] doi:10.2905/0C6B9751- A71F-4062-830B-43C9F432370F," 2019.

Toutain, Jean-Claude, *La production agricole de la France de 1810 à 1990: départements et régions: croissance, productivité, structures* number 17, Presses universitaires de Grenoble, 1993.

_ **and Jean Marczewski**, "Le produit intérieur brut de la France de 1789 à 1982," *Economies et sociétés (Paris)*, 1987, *21* (15), 3–237.

Villa, Pierre, "Productivité et accumulation du capital en France depuis 1896," *Revue de l'OFCE*, 1993, *47* (1), 161–200.