

Structural Change, Land Use and Urban Expansion
Online Appendix A — Data and Measurement

Nicolas Coeurdacier
SciencesPo Paris, CEPR

Florian Oswald
SciencesPo Paris

Marc Teignier
Serra Húnter Fellow,
University of Barcelona

July 2, 2024

Contents

A.1	Aggregate Data	2
A.1.1	Agricultural Land Use	2
A.1.2	Sectoral employment	5
A.1.3	Sectoral National Accounts and Prices	7
A.1.4	Sectoral Productivities	9
A.1.5	Consumption expenditures	11
A.1.6	Land and Housing Wealth	13
A.2	Urban Area and Population Measurement	14
A.2.1	Manual Urban Area Measurements 1870 and 1950	14
A.2.2	Manual Population Measurements 1870 and 1950	17
A.2.3	Automatic Area and Population Measurement via GHSL	23
A.2.4	Density Measurement Results	24
A.2.5	Discussion and Sensitivity for Area Measurement	28
A.3	Spatial Data on Agricultural Land Use, Yields and Farmland Prices	33
A.3.1	Agricultural Land Use Around Cities	33
A.3.2	Data on local farmland values	34
A.3.3	Data on wheat yields and land use for wheat	35
A.4	Urban density and farmland values	37
A.4.1	Sample and Data	37
A.4.2	Results	38
A.5	Urban Individual Data	45
A.5.1	Individual Commuting Data from ENL	45
A.5.2	Individual Commuting Data from DADS	48
A.5.3	Urban Productivity and Wages	51
A.6	Historical Commuting Speed in Paris	53

A.1 Aggregate Data

A.1.1 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 [Desrires \(2007\)](#) 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 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 about 30% 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. This assumption deserves some discussion though. On one hand, one could consider exploited forests as agricultural land as this was the case in the 19th century. Forests produce primary (necessity) materials (used in particular for heating in the 19th century), subject to structural change. On the other hand, a significant fraction of French forests is not exploited and used for leisure as natural amenities—particularly so in the recent period. As the data do not allow to differentiate across forests’ uses, we stick on a narrower definition of agricultural land, which only includes land used to grow crops and feed cattle—corresponding to the current definition of the SAU.

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 Savoie 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

¹In 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.

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 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 This gives a 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 A.1 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', 'pâts 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,

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.

³Statistique 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>

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

⁵As shown in Figure A.1, 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 PROPORTIONS.
		hectares.	p. 100.
1^o TERRITOIRE AGRICOLE.			
	Céréales.....	1 827,085	28.06
	Grains autres que les céréales.....	319,705	0.60
	Pommes de terre.....	1,475,144	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
Superficie cultivée.	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.....	38,392	0.07
	TOTAUX de la superficie non cultivée.....	6,226,189	11.77
TOTAUX DU TERRITOIRE AGRICOLE.....		50,467,909	95.47
2^o TERRITOIRE NON AGRICOLE.....		2,389,290	4.53
<i>Totaux généraux du Territoire.....</i>		52,857,199	100.00

⁽¹⁾ Non compris les cultures dérobées.

⁽²⁾ Y compris les herbages alpestres.

Figure A.1: Land Use in the 1892 Recensement Agricole.

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 [A.2](#).

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 [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.

A.1.2 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 \(1991\)](#) 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

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.

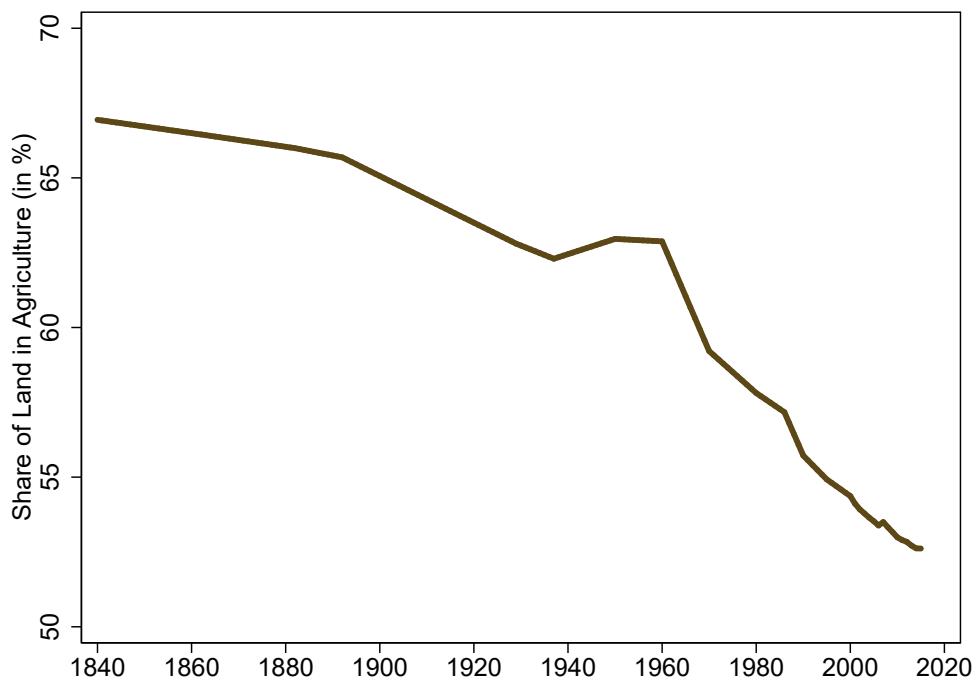


Figure A.2: Shares of Land used in Agriculture (1840-2015).

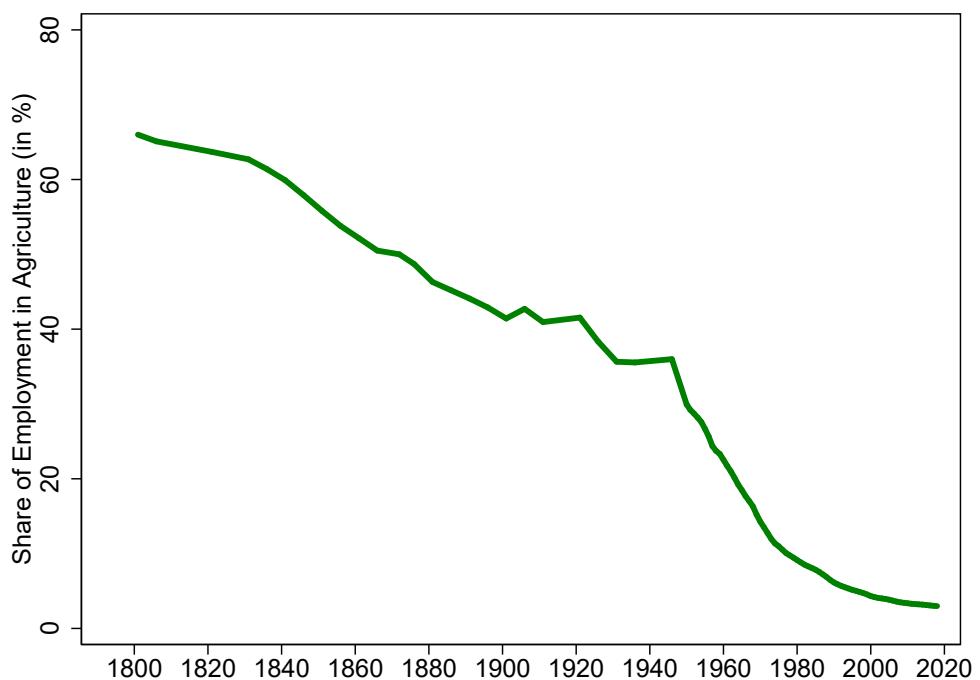


Figure A.3: Shares of Employment in Agriculture (1806-2018).

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 A.3. 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.

A.1.3 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, data cover 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

⁷The 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.

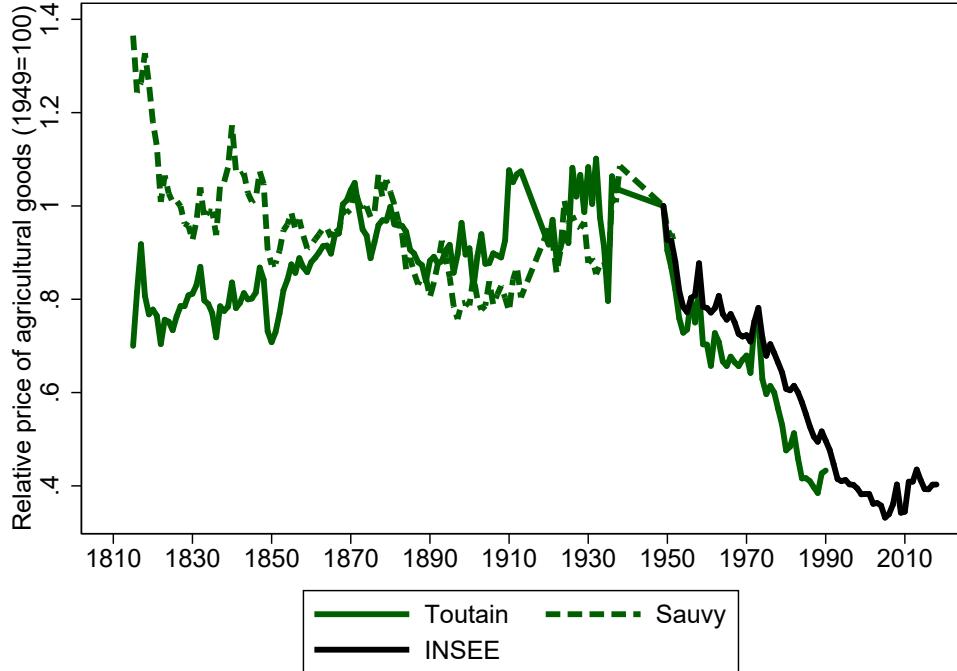


Figure A.4: Relative prices of agricultural goods, 1949=100 (1815-2019).

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 A.4. While the relative price of agricultural goods appears fairly stable until 1910, it exhibits later a clear downward trend over the twentieth century. 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

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

as

$$\frac{1}{P_{GDP,t}} = \frac{s_{agr,t}}{P_{agr,t}} + \frac{1 - s_{agr,t}}{P_{non-agr,t}}, \quad (\text{A.1})$$

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 $P_{non-agr,t}$, we can invert Eq. A.1 to back out a price index for non-agricultural goods (urban goods including manufacturing and services),

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

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 Générale 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 A.4).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.

A.1.4 Sectoral Productivities

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

⁹Details 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>

¹⁰We 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.

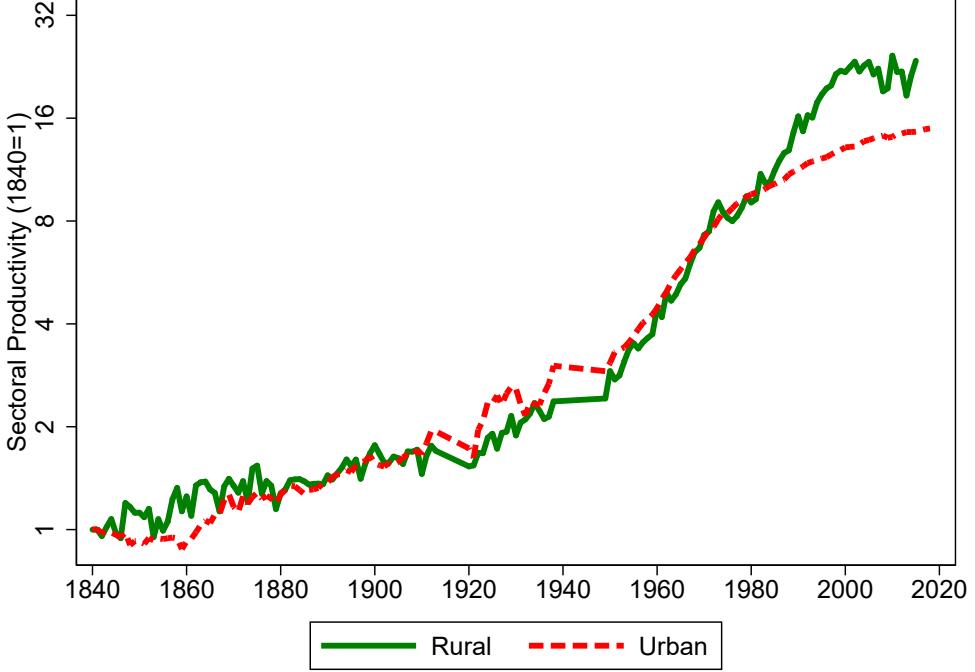


Figure A.5: Rural and Urban Aggregate Productivity, 1840=1 (1840-2019).

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. 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 A.5 (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,

$$\frac{Y_r}{L_r} = \theta_r \left(\alpha + (1 - \alpha) \left(\frac{S_r}{L_r} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} = \theta_r F \left(\frac{S_r}{L_r} \right). \quad (\text{A.2})$$

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. A.2, one can compute the rural productivity parameter, $\theta_{r,t}$, at each date,

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

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 A.5 (solid green line). This will be our baseline exogenous aggregate rural/agricultural productivity shifters.

Comments. Comparing aggregate 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.

A.1.5 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 in-

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

¹²Over 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.

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

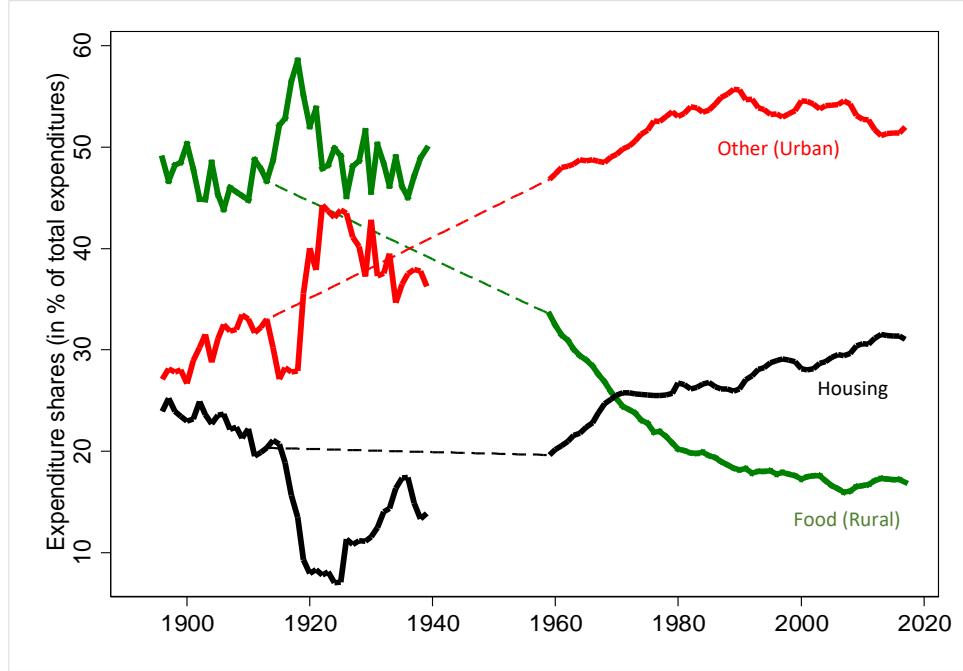


Figure A.6: Spending Shares for Rural, Urban and Housing goods.

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 A.6. 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 from 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

¹³We 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.

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 A.6, where the share of expenditures 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 A.6.

A.1.6 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 around 0.3 for this period).

¹⁴Rents 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 units with very cheap rents under the special regime of 1948 still subsist.

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

A.2 Urban Area and Population Measurement

As explained in Section 2.2 in the main text, we consider the 100 most populated cities in the 1876 Census as our sample. We constrain this list to contain only cities which are still independent entities nowadays (not part of a larger urban area).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 kms), and population count. Depending on the period, we use 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 maps and the 1954 population census. Due to the lack of other data sources, we regard 1866 and 1876 as well as 1950 and 1954 as the same points in time, and we refer to 1870 and 1950 for simplicity. In subsequent years, the Global Human Settlement Layer (GHSL) provides built up area and population data for 1975, 1990, 2000 and 2015. For these later years, we also expand the sample to 200 cities for the empirical specification of Section A.4.

A.2.1 Manual Urban Area Measurements 1870 and 1950

We rely on georeferenced maps provided via <https://www.geoportail.gouv.fr> to take area measures of 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 the satellite data for the later periods (see Section A.2.5).¹⁷ We use the tools on geoportail.fr to delineate the urban area of the EM and 1950 maps/aerial photos manually on screen, taking a screenshot of each measurement.

For the EM maps, the criteria to classify land as urban are fairly straightforward, thanks to the color coding used: red, rectangular shapes show buildings, whereas brown shading stands for rural land. Therefore the area where one observes contiguous buildings is classified as urban area. In this early period, classification is unambiguous, because there are almost no suburbs and the city ends abruptly. In many cases we even observe fortification walls which surround the city and help the task. We show examples for this time period in Figures A.7 and A.9 for two cities.

In the 1950s we also rely on manual classification. As for 1870, we aim at delineating the city with an abrupt change in the density of built-up area at the boundary of cities (marked by a color change in the map/aerial photos). The situation has however evolved at this point, and suburbs with low density housing are more prevalent. We need to take a clear stand on how to classify those. We try

¹⁶This 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_dhPD0Y80y67cU_L4Q2FHg0r711ffB3XYm/pubhtml?gid=0&single=true

¹⁷Contemporary photographs are taken between 2016 and 2020: https://www.geoportail.gouv.fr/depot/fiches/photographies-aerielles-RVB/geoportail_dates_des_prises_de_vues_aerielles-RVB.pdf

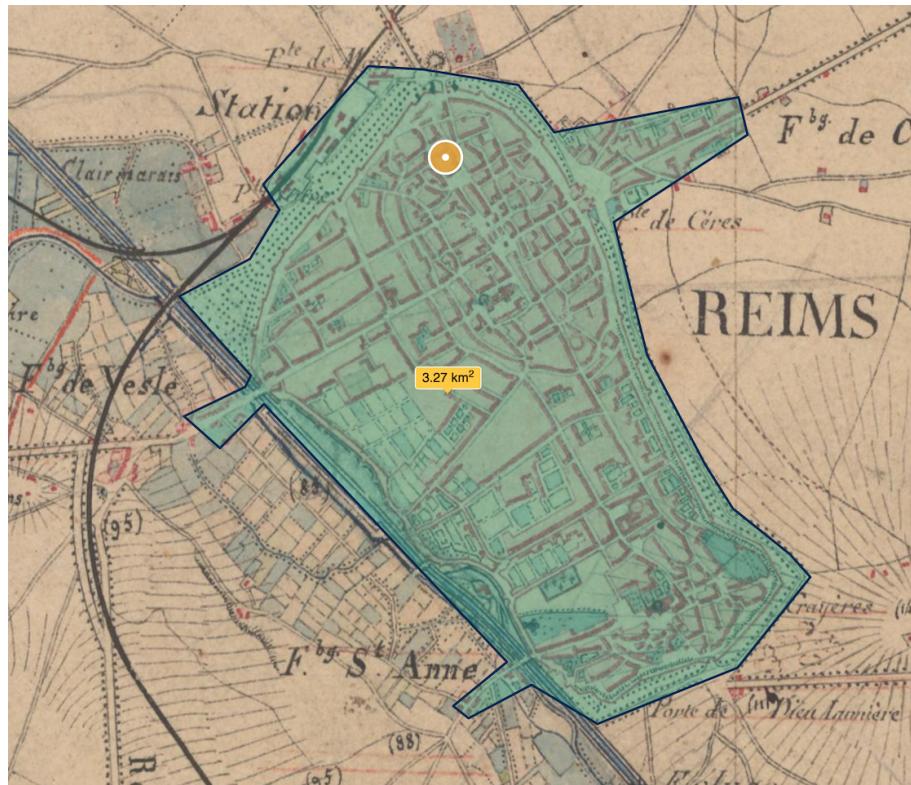


Figure A.7: Area measurement of Reims using Etat Major map

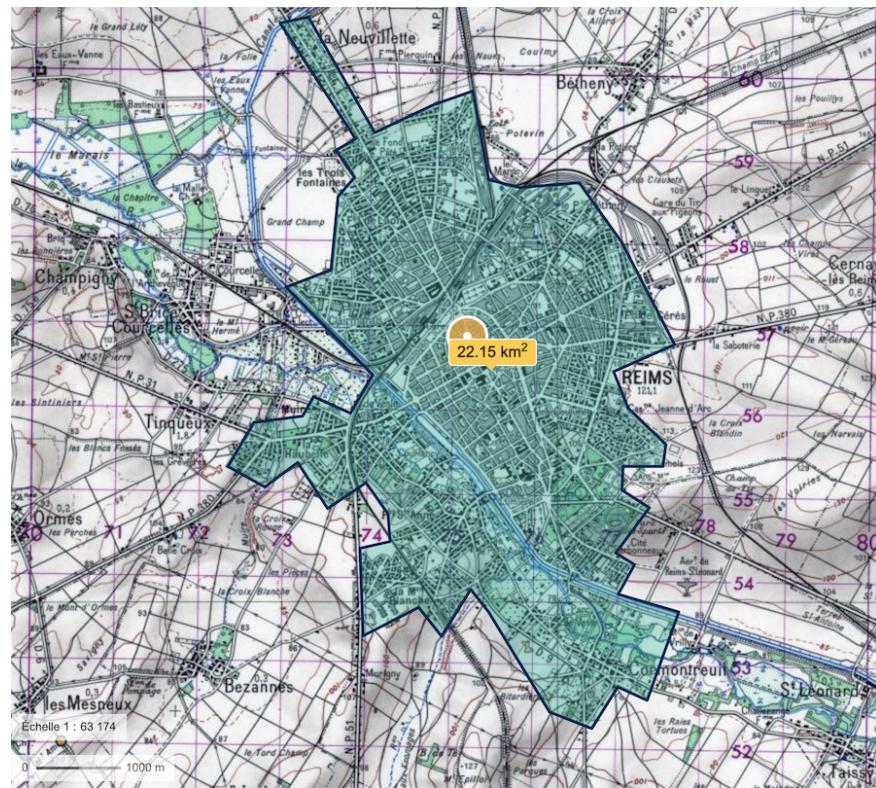


Figure A.8: Area measurement of Reims using 1950 map

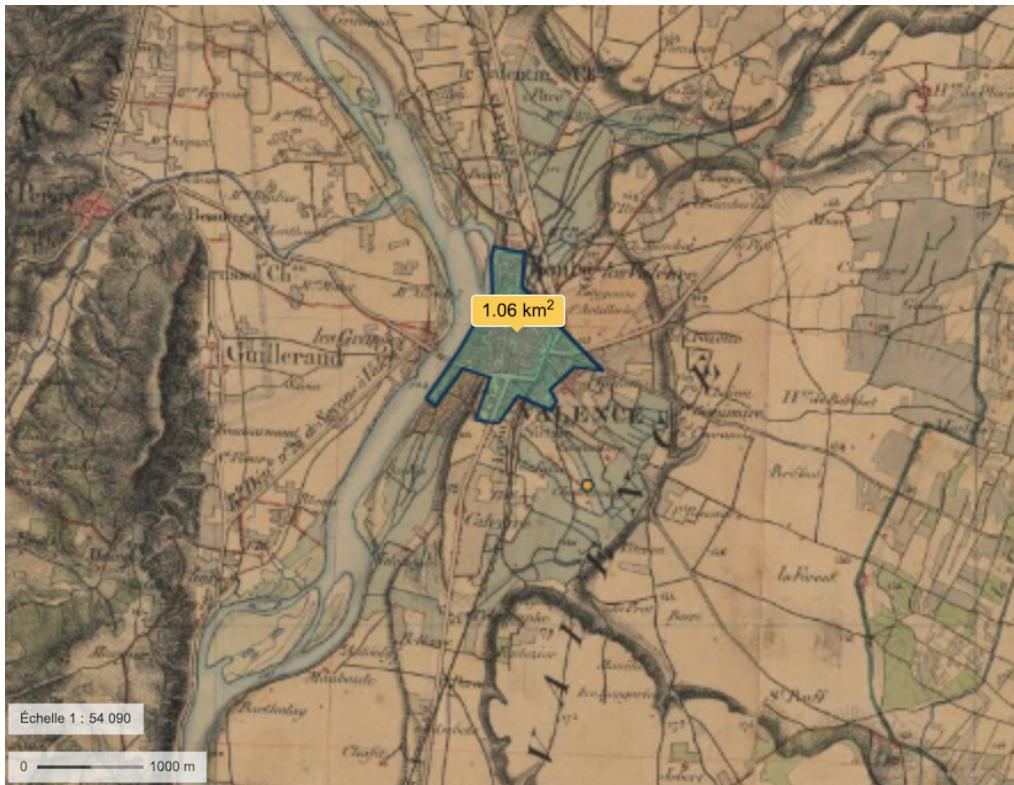


Figure A.9: Area measurement of Valence using Etat Major map

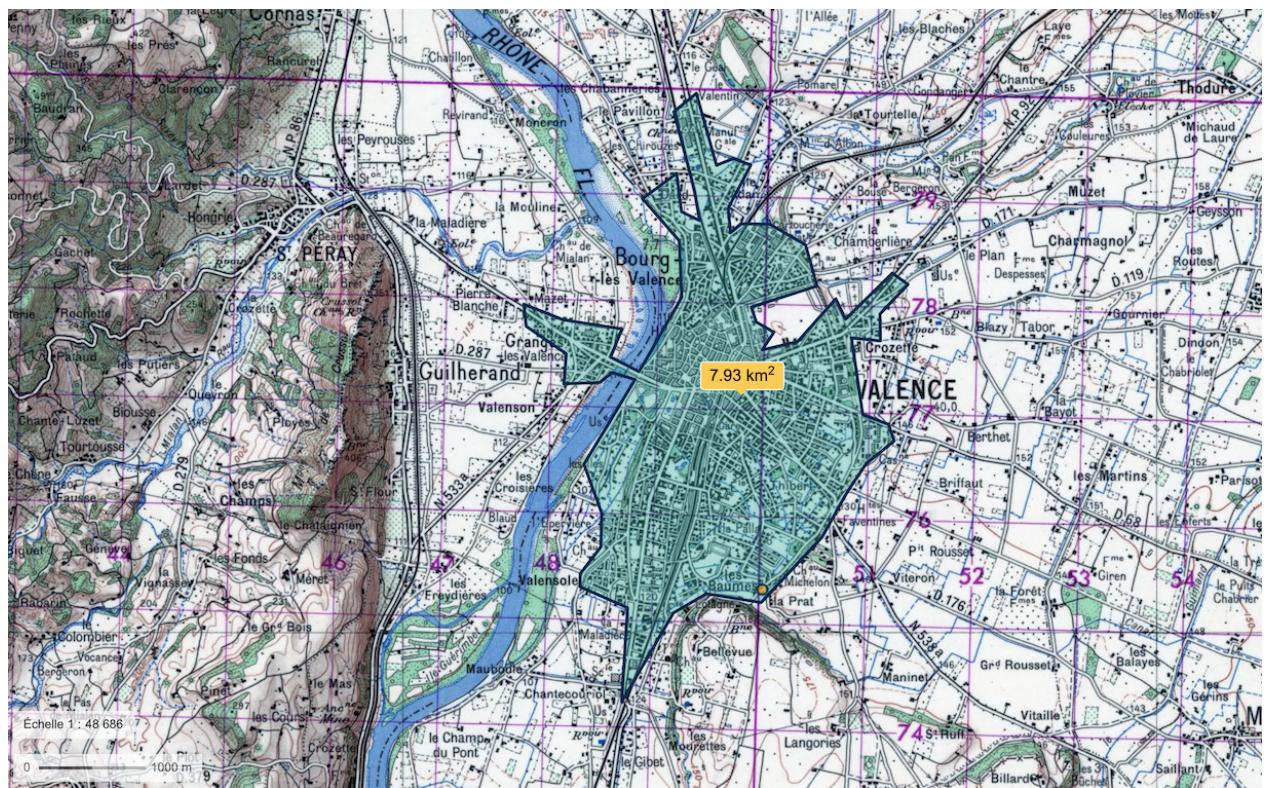


Figure A.10: Area measurement of Valence using 1950 map

to adopt criteria to classify as urban area that coincide with the criteria which will be applied to the automated satellite measures in later periods, and explained in greater detail below. In short, we manually classify an area as part of a city, if two conditions hold:

1. Contiguous built-up structure: We observe a contiguous housing structure (in an *imaginary* grid cell of 250m by 250m). The 1950s maps do not show such grid cells, so the analyst has to use the scale indication on the map to infer how large such a cell would be.
2. Density of built-up environment: we try to enforce a low built-up threshold (corresponding to the 30% threshold when dealing with satellite data) in each grid cell—excluding from the city areas with low density of built-up. The aim is to make the manual measure as close as possible to the automated approach in order to distinguish very low-density suburbs from proper city. This means that areas contiguous to the main city, but significantly less dense because of interspersed rural/agricultural land or gardens, are excluded from the city area.

Examples for this measurement exercise are in Figures A.8 and A.10 for the same cities as above. While measurement error when delineating the urban area is unavoidable at the city level (some farmland might be included in our measure or some urban buildings excluded), we believe that the measurement error should be averaged out when computing the main stylized facts of the paper in Section 2.2 for the average across the 100 cities.

A.2.2 Manual Population Measurements 1870 and 1950

In order to collect population counts for each city for the 1870 data point, 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 the sample are contained within their administrative boundaries in 1870. This is also true for Paris since the municipality of Paris was extended in 1860 to incorporate the main municipalities in the nearest surroundings—together with redesigning the Parisian districts ('arrondissements').¹⁸

The next data point for the 1950 cities is obtained from the Census in 1954. Given the area measure obtained for 1950 (described in Section A.2.1), 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/communes by 1950, and in particular Paris), we carefully check which administrative areas (i.e. former independent villages/communes) have now become part of our 1950 city area, and we sum the corresponding population counts for the concerned areas. The mapping of villages/communes to cities is given in Table A.1 and the one of Paris administrative areas is shown in Table A.2.

¹⁸The 1870 area measurement does incorporate a small part of Montreuil on the east and of Neuilly-sur-Seine on the west, both municipalities being contiguous to Parisian districts. However, the total population of these very rural municipalities, account for 1.7% of the population of Paris. Adding the total population of these communes to provide an upper-bound of the Parisian population and density in 1870 would not affect the main stylized facts of Section 2.2.

Table A.1: 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

Table A.2: 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
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

Table A.2: Paris 1950 Population Classification (*continued*)

CODGEO	REG	DEP	LIBGEO	year	population	date
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
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

Table A.2: Paris 1950 Population Classification (*continued*)

CODGEO	REG	DEP	LIBGEO	year	population	date
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
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

Table A.2: Paris 1950 Population Classification (*continued*)

CODGEO	REG	DEP	LIBGEO	year	population	date
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

Measurement issues. If not contiguous to the main city, closeby municipalities are considered as separate for our measure of urban area by definition. However, there are always a few low-density villages in the immediate surroundings of a large city. Their exclusion (or not) from the urban area would lead to different measurements for population and area. In principle, measurement error can go both ways. However, given that cities are measured as growing mostly out of their main municipality until post-1950 (with the clear exception of Paris), we might be slightly understating population and area of some cities in the earlier periods (resp. slightly overstating average density).

A related issue is that one cannot have a more precise population count with finer grid-cells than municipality level data for these two years of data (1870 and 1950). This forces us to incorporate the entire population of municipalities as part of the urban area, while, at the fringe of the urban area, some residents might be still working in the agricultural sector and should be in principle excluded from the population count. Arguably, this source of measurement error is likely to be quite minimal given that this concerns only the fringe of low-density municipalities at the boundary of each urban area. Note that this measurement issue does not apply to the later years, where we have finer grid-cells available thanks to satellite data.

We now turn to the measurement of urban areas and populations for the later years using satellite data.

A.2.3 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. \(2019\)](#)).¹⁹ 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 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. \(2016\)](#).

¹⁹https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2019.pdf

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 Section A.2.1) 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 A.2.5.
 - (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.

We show example output for built-up area classifications for two cities in Figures A.11 and A.12.

A.2.4 Density Measurement Results

Built-up and Density Measures. Equipped with area (built-up) and population measurements at dates 1870, 1950, 1975, 1990, 2000 and 2015 for each of the 100 cities, the average density of a given city is simply its population divided by its area at a given date. Example measures of resulting

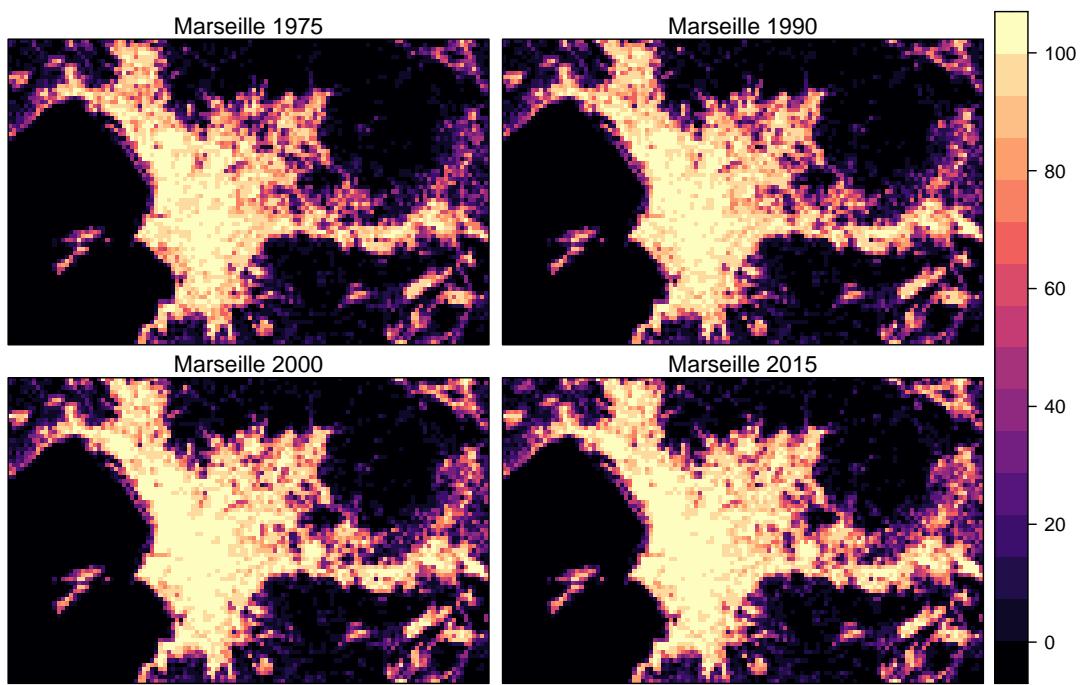


Figure A.11: GHS-BUILT raster map of Marseille. The color scale represents percentage built-up in each grid cell.

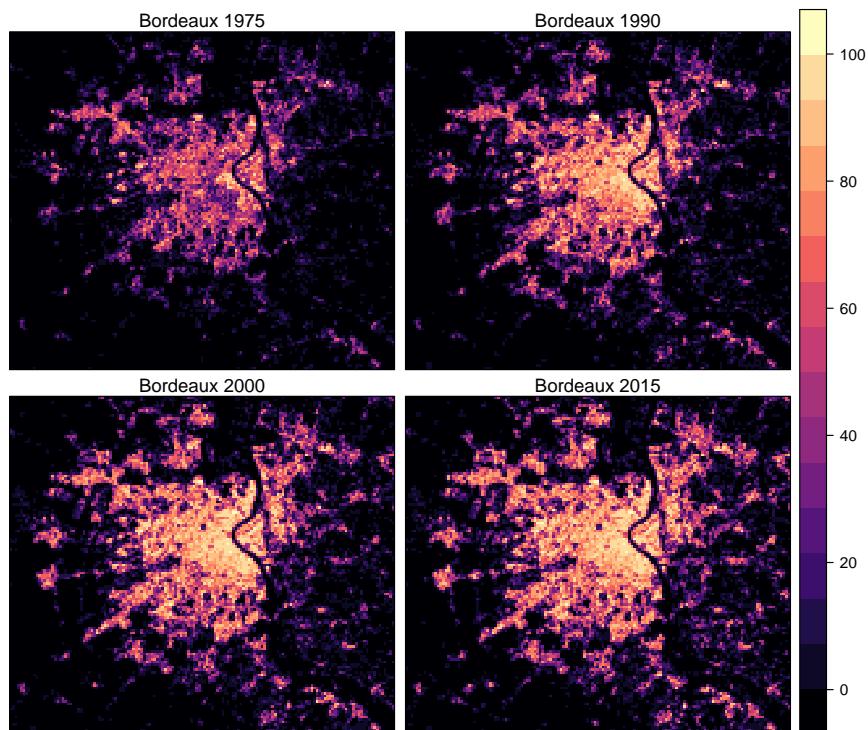


Figure A.12: GHS-BUILT raster map of Bordeaux.

Urban Density over time in France

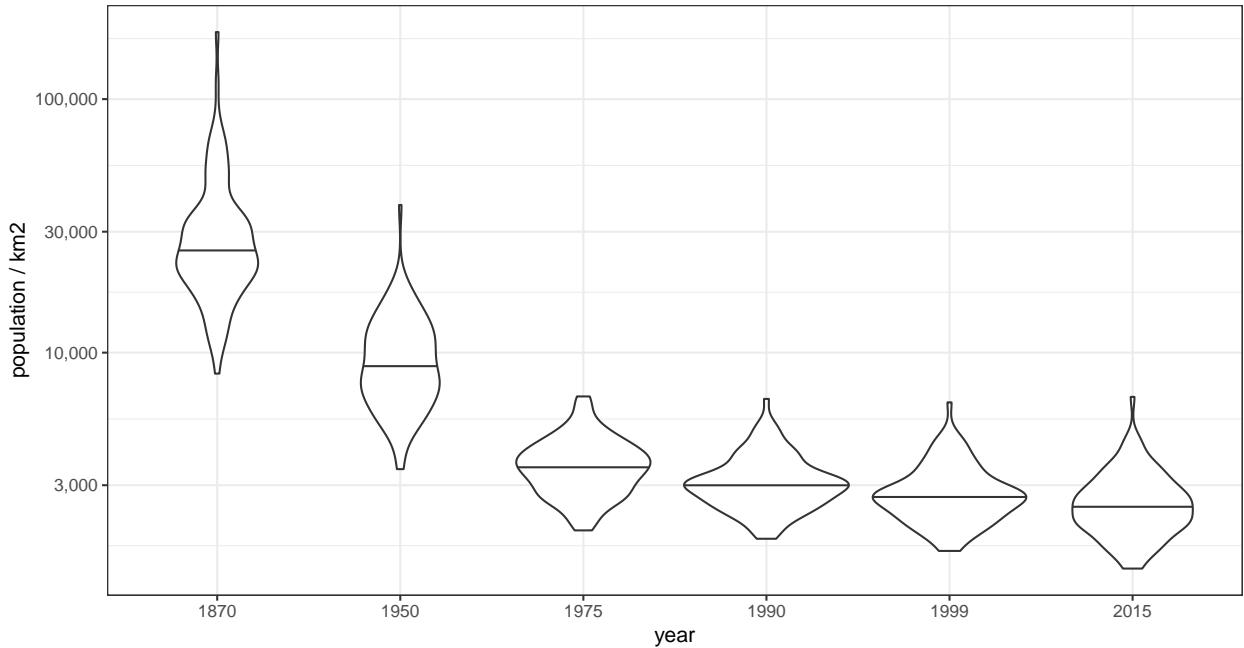


Figure A.13: Distribution of Urban Density over Time. This *violin plot* represents the distribution of densities at each date, labeling the extreme values. The horizontal line denotes the median value.

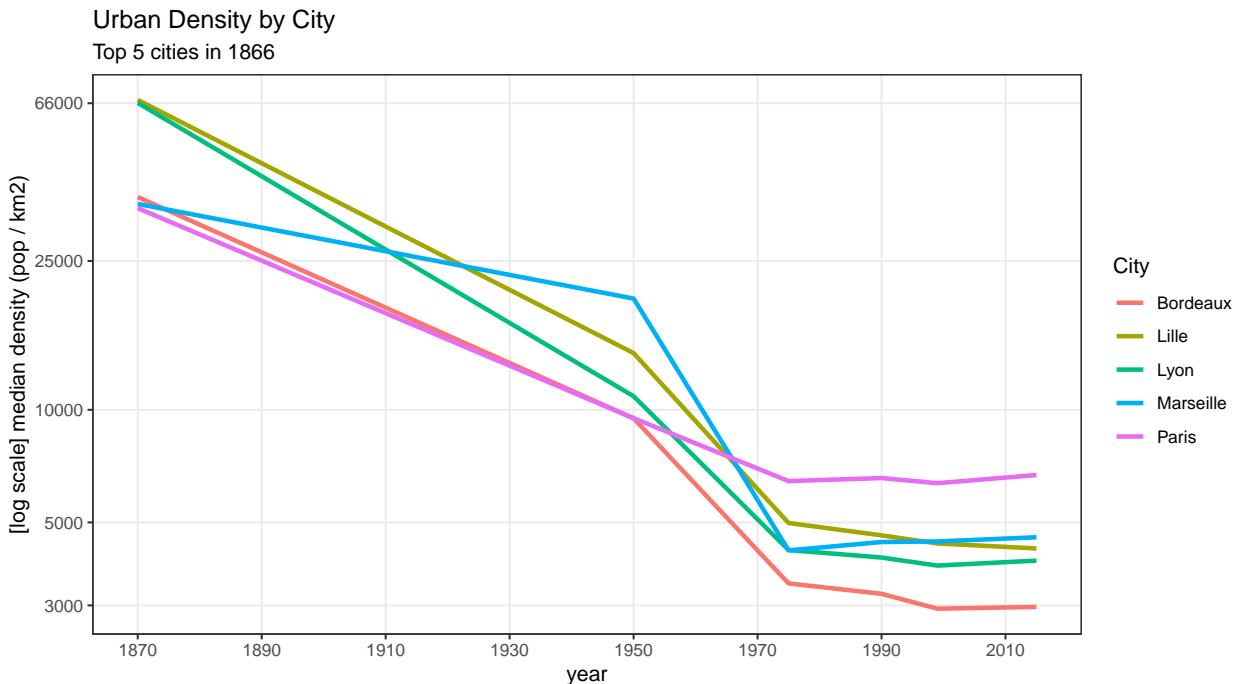


Figure A.14: Density in the largest five cities in 1876 over time.

urban densities can be seen in Figure A.13 for the entire distribution and in Figures A.14 for the top 5 cities.

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 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 k , we compute the density D_{k,ℓ_k} at distance ℓ_k (in kms) from the city center. The set of distances ℓ_k varies across cities, as bins are of different size.

Figure A.15a 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 A.15b.

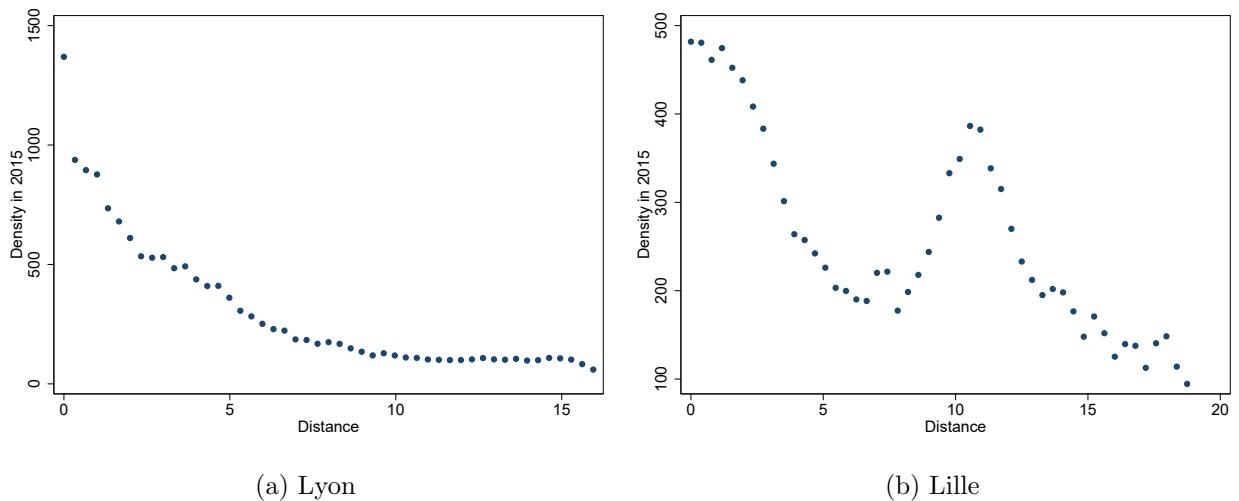


Figure A.15: Density gradients.

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 (or equivalently decay coefficient) of city k , b_k , as the value of slope coefficient of $\ln(D_{k,\ell_k})$ on ℓ_k ,

$$D_{k,\ell_k} \approx a_k \exp(-b_k \ell_k), \quad (\text{A.3})$$

where a_k and b_k are estimated (positive) numbers. We use an exponential decay model as it fits very well the data for all French cities—apart from a few polycentric ones such as Lille. This gradient can be computed for every city in our sample of 100 cities.

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 population-weighted mean of gradients, excluding large polycentric cities. This gives a value of 0.176 for the average population-weighted 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. 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. A.3 provides a very accurate description of the density data within cities.

A.2.5 Discussion and Sensitivity for Area Measurement

This Section briefly discusses the measurement of urban areas, how they relate to the model's predictions and how the different measurement tools ('manual' and 'automatic' using satellite data) are comparable. We also perform some related sensitivity analysis regarding these measurements.

Discussion. The relevant concept for the theory is land use. In the model, the city ends, when land starts being used for rural/agricultural production. In the data, land use is not directly observed and the land use change can happen less abruptly in some locations. Our strategy is to impose a threshold on built-up (not population) density, below which we no longer include a certain plot into the urban area—the built-up density informing us on the intensity of the use of land for residential purposes. When satellite data are not available, we implement a strategy which aims to get as close as possible to the model's definition and to the 'automatic' measure with satellite data. However, measurement error is unavoidable since some very low-density suburbs might be inappropriately excluded from the urban area. Vice-versa, some agricultural plots with housing units might be included.

This way of measuring urban area is different from the approach taken in [Combes et al. \(2021\)](#). Their delineation of cities is not directly comparable to ours as it is based on local density measurement—identifying on the maps the universe of buildings at a very granular level and, under some assumptions, allocating population to built parts at each date. Land is part of the city if local density is significantly in excess of the counterfactual density where people are randomly distributed on the French territory—above the 95th-percentile of this counterfactual distribution (see also [De Bellefon et al. \(2019\)](#)). This definition can lead to a fairly different measurement—particularly so in the nineteenth century where about two thirds of the population works in agriculture and some excess density might be observed in the surroundings of cities if farms are more densely located there. However, for both measurements, the measured urban area is dependent on the cut-off value as-

sumed for delimiting cities, the reason we perform robustness with alternative cut-offs when using satellite data.

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 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 (210 m^2 garden/agricultural plot) 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 measured outcomes for a range of different cutoff values. With this in mind, we present in Figures A.16 and A.17 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 grid cell built up being excluded from *urban*) increase density more significantly. 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`.

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 A.18), 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 A.19. Results are comforting. Both measures give similar estimates and are very highly correlated across cities. One should also note that there is no systematic bias in a specific direction.

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 A.20 that our manual measures correspond closely to their obtained measures despite different measurement strategies.

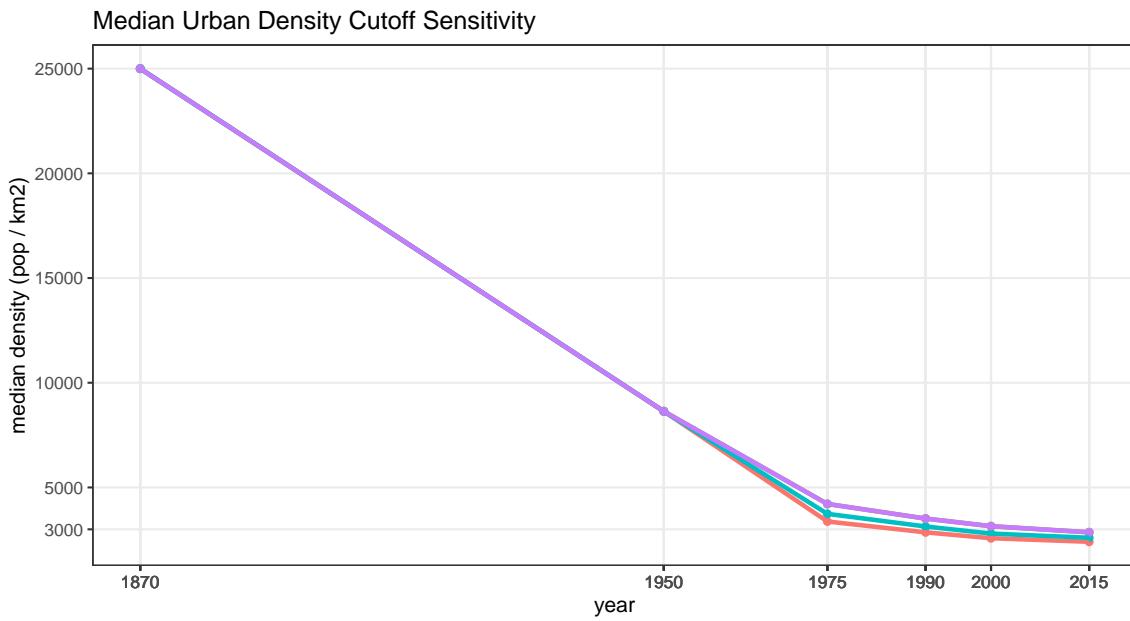


Figure A.16: Median urban density for different cutoff parameter values. The parameter indicates the percentage of a grid cell (250x250 meter) that has to be built-up in order to be classified as *urban area*.

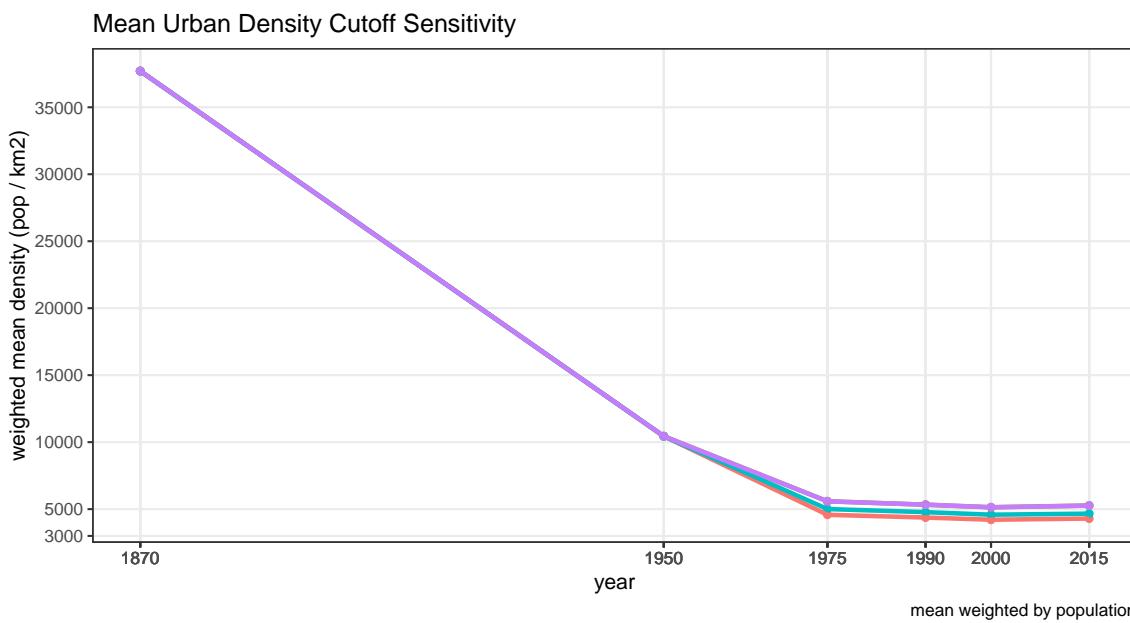


Figure A.17: Weighted mean urban density for different cutoff parameter values. The parameter indicates the percentage of a grid cell (250x250 meter) that has to be built-up in order to be classified as *urban area*.

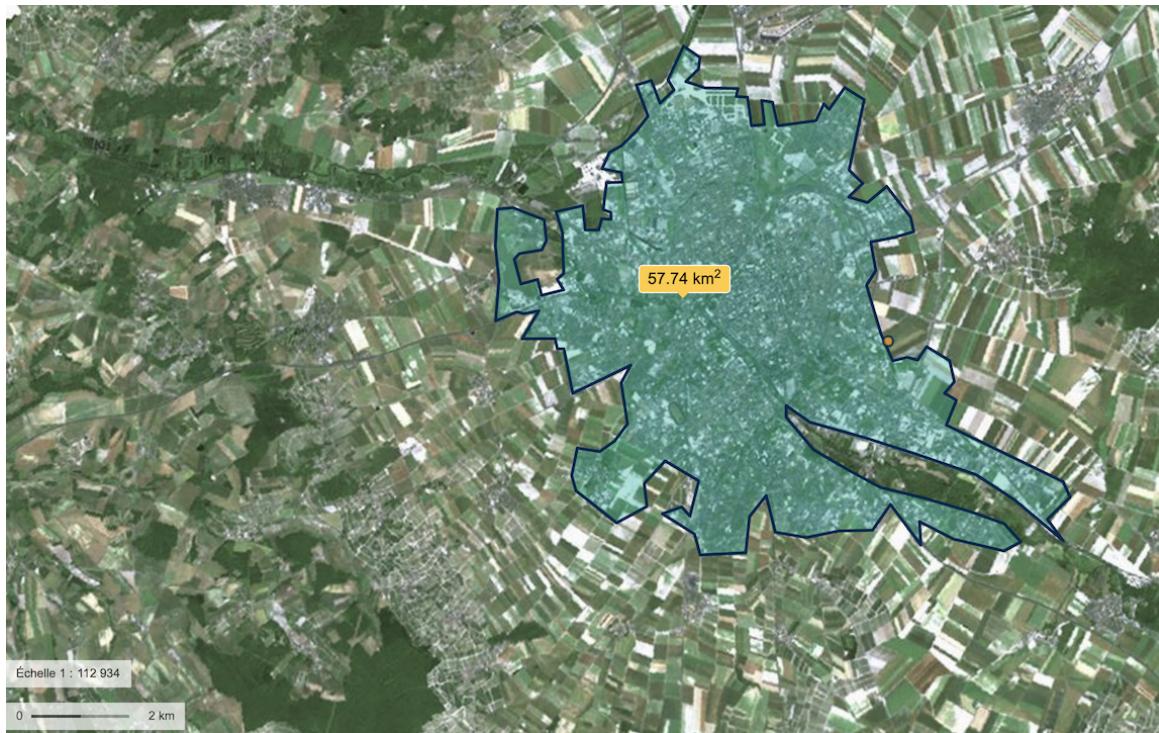


Figure A.18: Area measurement of Reims using modern day photograph - used only for cross-checking GHSL measures.

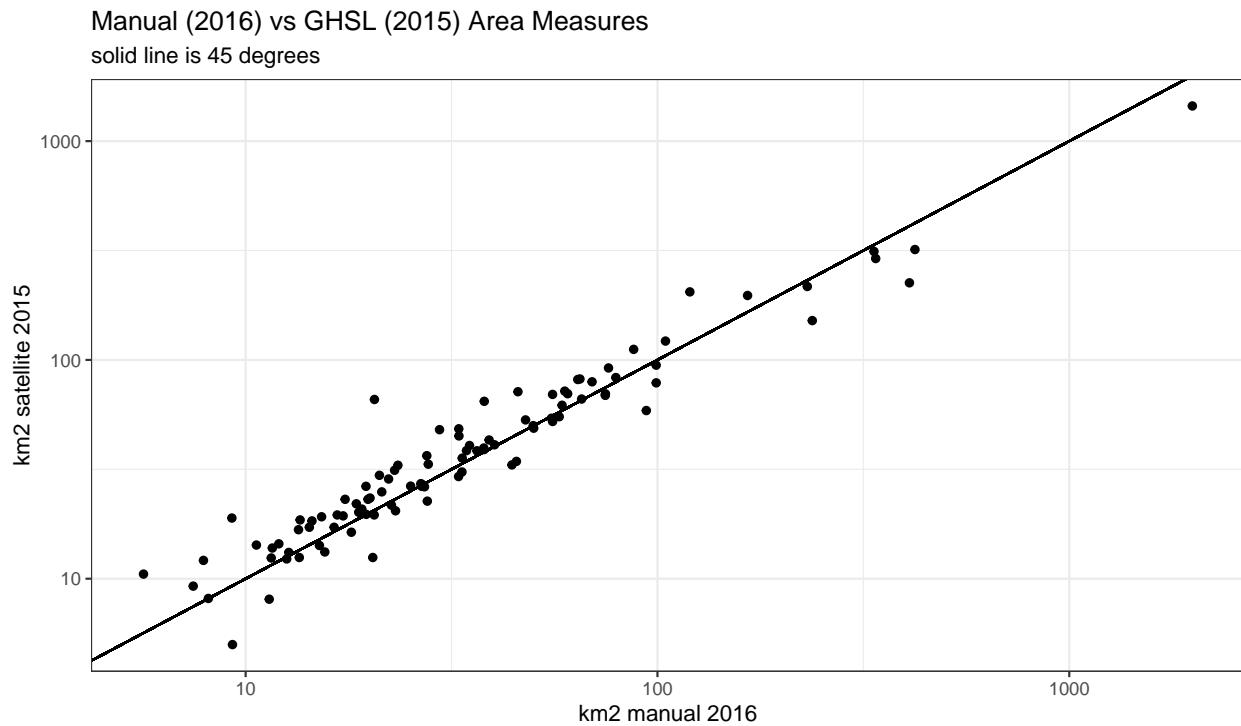


Figure A.19: Comparing manually obtained area measures for each of our cities with automatically obtained ones via GHSL data.

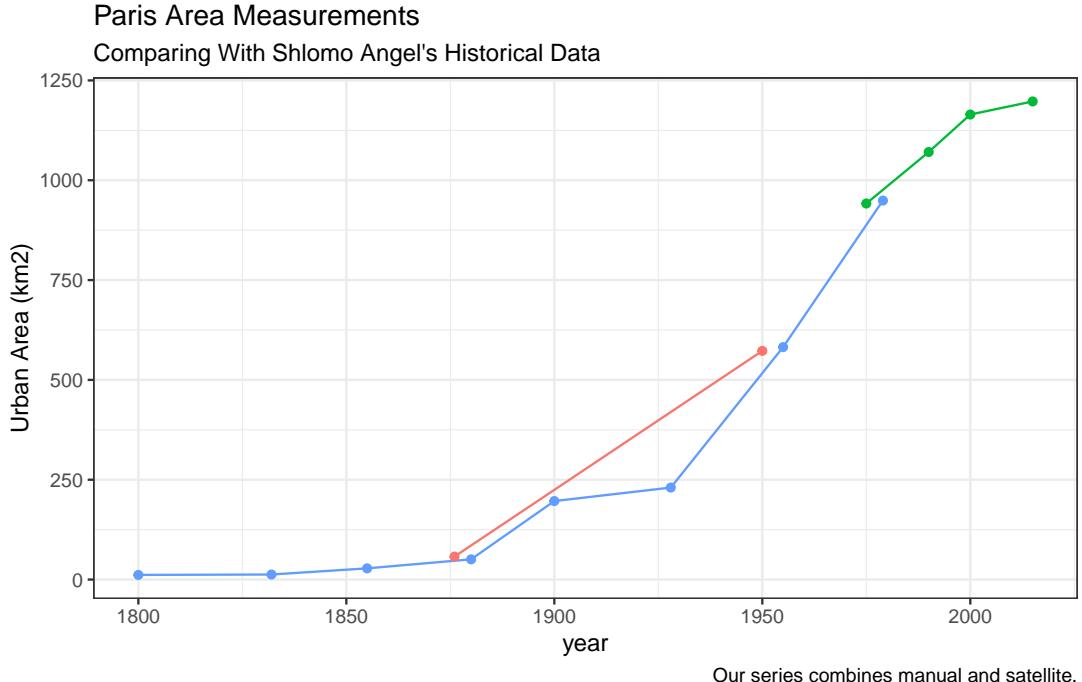


Figure A.20: Comparing our measures with Shlomo Angel's data used in [Angel et al. \(2012\)](#) and [Angel et al. \(2010\)](#). We are reassured that our manual measurement exercise aligns closely with what they obtained. Also, their final data point is reassuringly close to our first satellite measure.

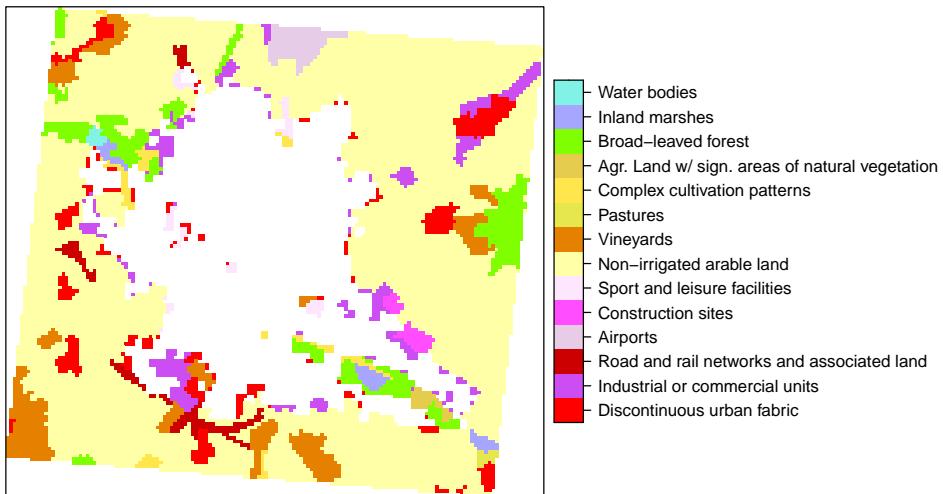


Figure A.21: Land use measures from CLC data for Reims. The white area represents our definition of the Reims Urban area in the last GHLS periods (2015), hence it is our definition of *inside vs outside* of the city. For instance, the red areas labelled *discontinuous urban fabric* are not part of our definition of the city.

A.3 Spatial Data on Agricultural Land Use, Yields and Farmland Prices

A.3.1 Agricultural Land Use Around Cities

We use CORINE Land Cover (CLC) data for 2018 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>.

The use of the data is very similar to the GHSL data in Section A.2.3. 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 an example for Reims in Figure A.21 and the resulting average in Figure A.22.

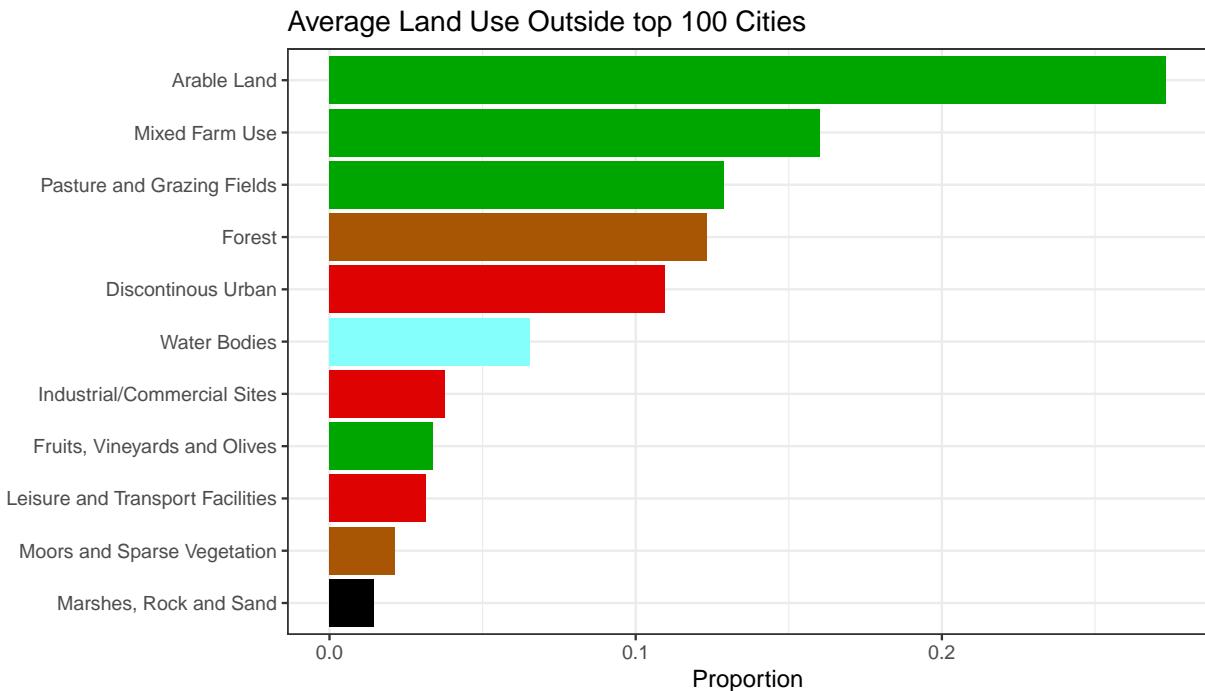


Figure A.22: Average land use measure from CLC data for our sample of top 100 French cities. This plot uses our own aggregation from 45 CLC labels into 11 exhaustive classes. We group all categories corresponding to agriculture into green bars.

A.3.2 Data on local farmland values

Local farmland values. We digitized data from the Recensement Agricole in 1892, which provides at the département level, the price of arable land ('valeur vénale des terres labourables, en francs/ha'). Data are available in the 'Statistique Agricole de la France: Résultats généraux de l'Enquête Décennale de 1892'.²⁰ Data do not include Alsace and Moselle, not part of the French territory until 1918 after the Franco-Prussian war of 1870.

Post 1950, we use data from the Ministry of Agriculture, which provides at the level of 'Petite Région Agricole (PRA)', the price of arable land (per ha) ('Prix des terres agricoles, terres labourables, libres, converted in francs/ha). PRA is a subdivision of a département, with more than 700 PRAs in Metropolitan France (versus 96 départements), providing a fairly local farmland price surrounding the different cities.²¹ Data are average market transaction prices for farmland in the different locations, weighted by area and filtered for extreme values. We digitized data until year 2000 and use data on local farmland values for years 1950, 1975, 1990, 2000 and 2015, dates at which cities' areas are measured (manually or with satellite data). Data are missing for few PRAs and we did our best to come back to the original source to fill the gaps.²² For the year 2015, due to a revision in the measurement of farmland values between 2007 and 2010, only a common price of farmland, including both arable land and grazing fields, is available.²³ This revision also led to a redefinition of the PRAs with some merging between PRAs existing before 2007. We made the different dates consistent by reallocating the new PRAs (in 2015) to their former definition based on the names—this had to be done for each PRA one by one given slight changes in names.²⁴ Equipped with data consistent across years at the PRA level, the geographical allocation on these PRAs on the French territory is made using a mapping between French 'communes' and their respective PRAs (using the geographical coordinates of the 'communes'). Figure A.23 shows the data on local farmland prices (PRA level) for years 1950, 1975, 1990, 2000 and 2015 together with departmental data for 1892. Data are available online at <https://floswald.github.io/LandUseR/articles/pra-check200.html>.

²⁰The online archives are available at: <https://gallica.bnf.fr/ark:/12148/bpt6k855121k/f1.item>. See p238-241 of the second volume with statistics for France.

²¹See classification in 2017 at <https://agreste.agriculture.gouv.fr/agreste-web/methodon/Z.1/!searchurl/listeTypeMethodon/> for the classification of PRAs. France counts 432 'régions agricoles' which can overlap multiple départements. PRAs are intersections of one département and one région agricole, 713 PRAs.

²²In the Parisian area (département 77), data are missing in 1990 and 2000 at the PRA level for the price of 'terres libres' but available for 'terres louées'—the latter being sold at a discount as occupied by a renter. We compute a price of arable land for 'terres libres' by rescaling proportionately the price of 'terres louées'—measuring the average percentage discount of 'terres louées' across the three départements of the Parisian area where both prices are available. For the PRAs where both prices are observed, this strategy gives a price fairly close to the one observed.

In the region of Nice, we use 'département' level data for Alpes-Maritimes due to missing data at the PRA level (lack of reliable transactions in two out of the three PRAs of the département—even 'département' data is missing in 2015).

²³Data and details available at <https://agreste.agriculture.gouv.fr/agreste-web/disaron/Chd21010/detail/>.

²⁴A typical example is the first three PRAs of Département 1 (Ain) pré-2007, 'VALLEE DE LA SAONE', 'DOMBES', 'COTEAUX EN BORDURE DES DOMBES', which become only one 'VALLÉE DE LA SAONE - DOMBES - COTEAUX' post-revision. For few PRAs more difficult to reallocate due to a change in the name, we searched on maps using the corresponding département and commune of the PRAs to allocate them to their previous definition. While some misallocation is unavoidable, this has very minor consequences given prices are spatially correlated and two neighboring PRAs have very similar prices.

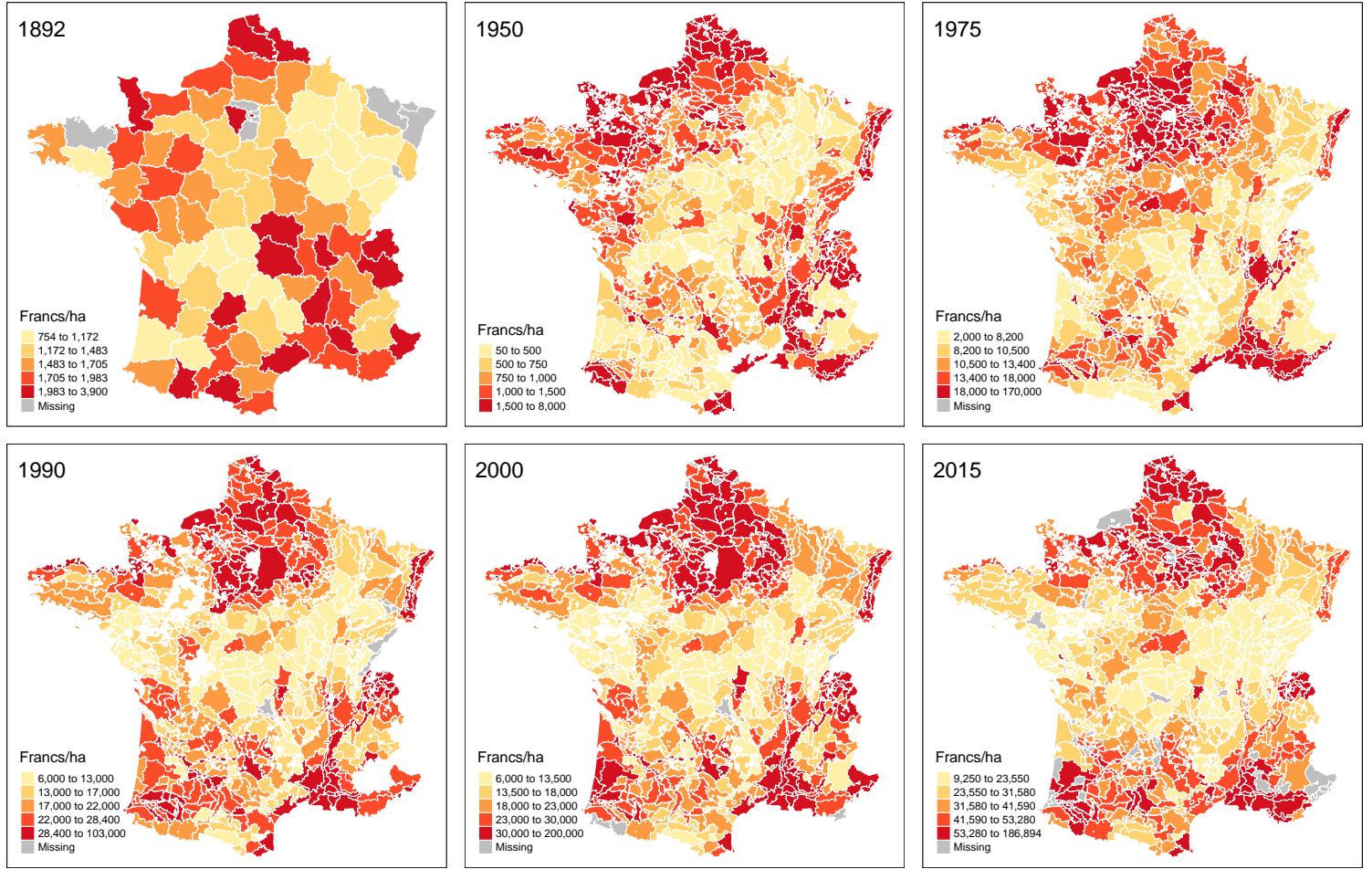


Figure A.23: Arable land value per hectare.

Notes: Arable land value (in francs/ha, ancien francs/ha in 1892) in each Département in 1892 and ‘Petite Région Agricole (PRA)’ post 1950. Polygons delimit the PRA. Data are from the 1892 Recensement agricole and the Ministry of Agriculture post-1950.

A.3.3 Data on wheat yields and land use for wheat

Wheat yields data. We use data from Schaubberger et al. (2022), which provides yield data for France over the period 1900-2018 for ten different crops at the département level. Yields are expressed in tons/ha using data from the Ministry of Agriculture (‘Statistique agricole annuelle’ or ‘Annuaire de statistique agricole’). Yields are spatially very correlated across the main crops and we focus on wheat, the main cereal cultivated in France. For a city k , we denote $\text{Yield}_{k,t}$, the yield of wheat in the département of city k at date t . Figure A.24 (left panel) summarizes the spatial variations in wheat yields across French départements, ranging from 2.5 tons/ha to 8.6 tons/ha.

Land use for wheat. We use data at the département level of land use by crop. Data are available from the Ministry of Agriculture in 2000, 2010 and 2016-2022. Data provides the area by crop in each département. Land use by crop is very persistent and we focus on year 2000. For each département of city k , we denote $(S_{r,wheat}/S_r)_{k,t}$ the fraction of agricultural land in the département

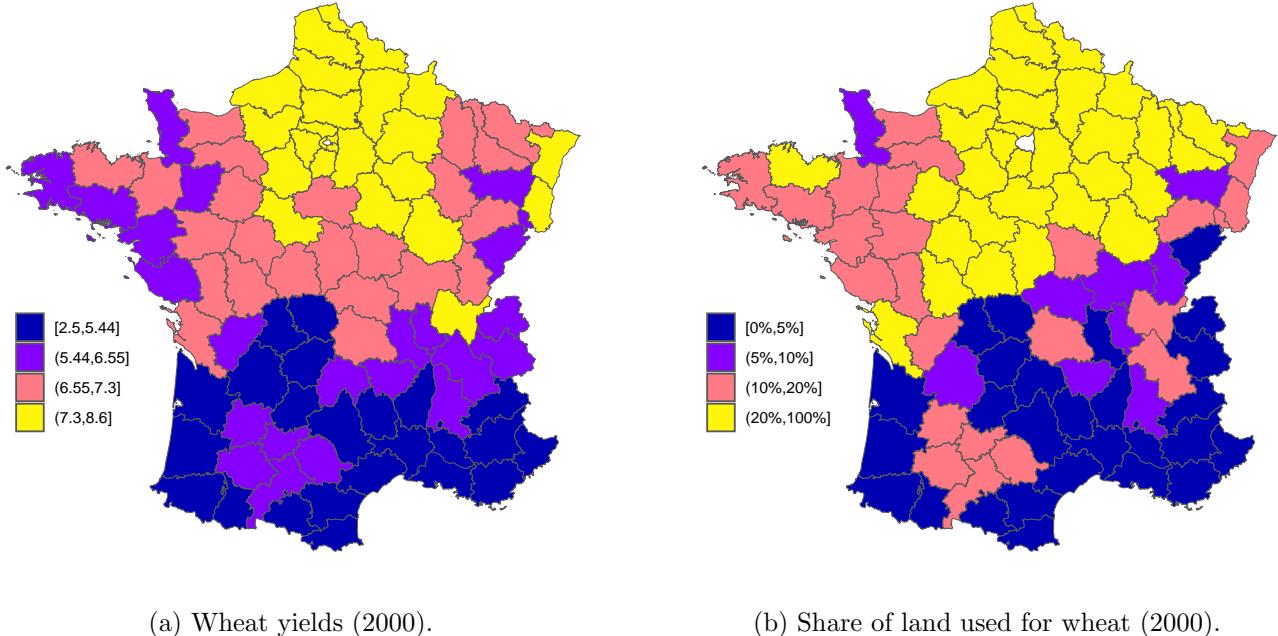


Figure A.24: Yields and land use for wheat in France

Notes: The left-panel shows the yields for wheat (tons/ha) across French départements in 2000. The right-panel shows the fraction of agricultural land used for wheat ('blé tendre' only). The scales in both panels represent quartiles. Data on yields are from [Schauberger et al. \(2022\)](#) and data on land use by crop are from the Ministry of Agriculture.

of city k that is dedicated to wheat (soft wheat, 'blé tendre'), the main cereal grown in France.²⁵ Across French départements, this ratio ranges from 0.03% to 45% (with a mean across département of 16%). Figure A.24 (right panel) summarizes the large spatial variations in land use for wheat—soft wheat being largely produced in regions surrounding Paris and towards the north and east of France.²⁶ Not surprisingly, comparing the right and the left panel of Figure A.24, one can see that land use for wheat is significantly higher in départements where wheat yields are higher.

²⁵We focus on 'blé tendre' abstracting from 'blé dur' (durum wheat). 'Blé tendre' accounts for more than half of all cereals grown in France and durum wheat is a very small fraction of wheat production—only significantly present in few southern départements as it resists better the lack of water. It is sold at a different price and, to us, it is like a different cereal—requiring to adjust yields for the relative price of both cereals in départements producing 'blé dur'. For simplicity and to preserve homogeneity in the data, we focus on the land use for soft wheat. Most départements producing wheat, produce soft wheat and barely durum wheat. Results are however unchanged if one select départements based on the land use of both types of wheat.

²⁶The 'Bassin Parisien' and specifically the Beauce region in the south of Paris, are known historically for being the breadbasket of France.

A.4 Urban density and farmland values

A.4.1 Sample and Data

Sample of cities. We extend the sample of 100 cities to a sample of 200 cities using GHS data for years $t \in \{1975, 1990, 2000, 2015\}$. The methodology to measure urban population and urban area on the extended sample is identical to the one described in Appendix A.2. We add the 100 largest cities in population in 1975 that are not in the initial sample of 100 cities.²⁷ The extension of the sample is done for statistical power when performing the IV-strategy—the IV-strategy being performed on a sub-sample of cities in départements where wheat is one of the main crop as detailed below.

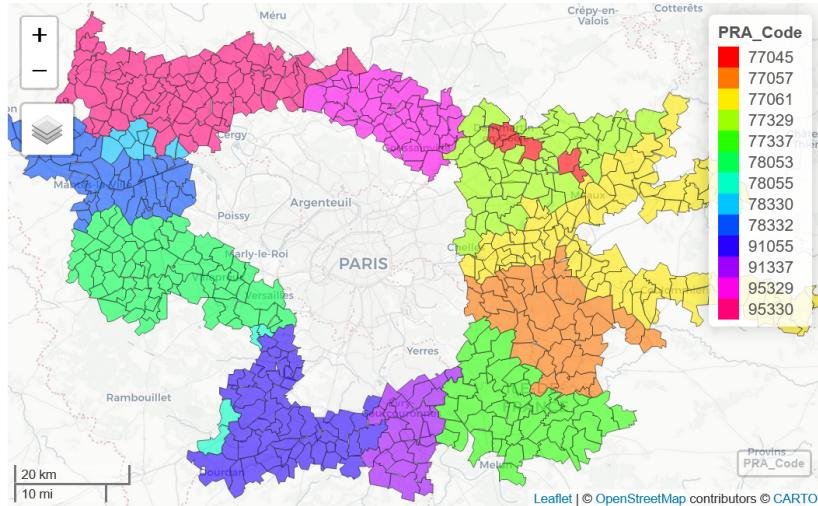


Figure A.25: Petites Régions Agricoles (PRAs) around the Parisian Urban Area.

Notes: PRAs around the Parisian urban area used to compute the farmland price of Paris, $\bar{\rho}_{r,Paris,t}$. Source: Ministry of Agriculture

Local farmland prices by city. For $t \in \{1975, 1990, 2000, 2015\}$, the observed local farmland value for each city k at date t , $\bar{\rho}_{r,k,t}$, is the corresponding price of arable land in the PRA of city k (see Appendix A.3.2 for a description of data on local farmland prices). Almost all cities can be allocated to a unique PRA²⁸ but a few large cities (Paris, Lyon and Nantes) are surrounded by multiple PRAs—the Parisian urban area being surrounded by 13 PRAs as displayed in Figure A.25 (see online version at <https://floswald.github.io/LandUseR/articles/pra-check-paris.html>). For those, we take the average of the farmland price in the different PRAs surrounding the urban area.²⁹

²⁷The list of additional cities is published online at https://docs.google.com/spreadsheets/d/e/2PACX-1vS02WpT0e7YTiS6f-svIXR3sURjiMRw7kBgfH1XF8LRre_dhPD0Y80y67cU_L4Q2FHg0r71ffB3XYm/pubhtml?gid=816431754&single=true

²⁸For a couple of observations with a missing price (Bruay-la-Buissière and Béthune in 2000 and Epernay in 2015), we use the price of farmland in a PRA located few kilometers away from the city—checking that in other years this price is very close to the one of the PRA of the city. The area around Nice and Menton (département Alpes-Maritimes) and Manosque (Alpes de Haute Provence) also do not provide data in 2015. We left them as missing due to the touristic nature of these locations in Provence for which the price in the neighboring PRAs is quite different. None of the results depend on the way missing values are adjusted.

²⁹In 2015, some PRAs are merged and we average across the remaining PRAs. Due to the spatial correlation of

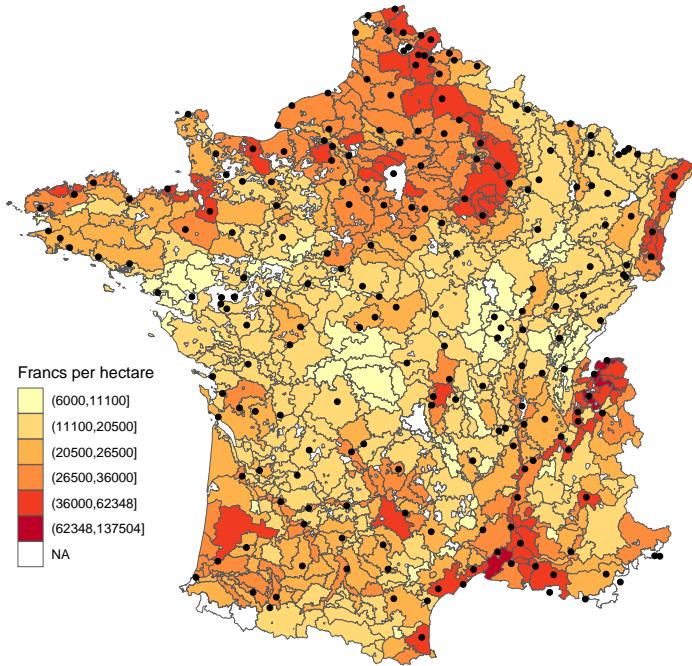


Figure A.26: Arable land value per hectare (2000).

Notes: Arable land value (in francs/ha) in each ‘Petite Région Agricole (PRA)’. Polygons delimit the PRA, and black dots mark the location of cities in the sample of 200 cities. Data are from the Ministry of Agriculture.

The sample of the 200 cities are represented by the black dots spread throughout France on Figure A.26 together with the local price of arable land.

A.4.2 Results

Empirical specification. In light of the model’s predictions, we investigate if a city is denser when the value of farmland around it is higher (holding everything else constant). To study the link between average urban density and local farmland values, we perform the following regression for years $t \in \{1975, 1990, 2000, 2015\}$ using the sample of 200 cities,

$$\log \text{density}_{k,t} = a_t + b \cdot \log \bar{\rho}_{r,k,t} + c \cdot Z_{k,t} + u_{k,t}, \quad (\text{A.4})$$

where $\text{density}_{k,t}$ is the average urban density of city k measured using Satellite (GHSL) data, $\bar{\rho}_{r,k,t}$ the farmland price around city k , a_t a time-effect and $Z_{k,t}$ region/city-specific controls. In the baseline specification shown in Table 2 in the main text, we only control for the log of the average wage in city k , $\log(w_{u,k,t})$, to be as close as possible from the theoretical model where we control

prices, the farmland price for these cities is not very sensitive to the weighting scheme across PRAs. For these cities, we do not include in the average the farmland price in the central municipality available only for the earlier years (Ceinture de Paris, zone maraîchère de Lyon, Région urbaine et maraîchère de Nantes) since these PRAs became urban later.

for urban wage (productivity) in city k .³⁰ The average wage (for full-time workers) at the urban area level, $w_{u,k,t}$, is from DADS panel EDP version 2019 which goes back until 1976 (see Appendix A.5.3). As sensitivity analysis, we also control for regional dummies (19 French regions) to capture time-invariant regional amenities.

OLS-estimates. Table A.3 (columns (1) to (3)) shows the OLS-estimates with and without controls for the sample of 200 cities. Across the different specifications, we do find that cities surrounded by a higher price of arable land are denser.

	log Urban Density					
	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
log $\bar{\rho}_{r,k,t}$	0.134*** (0.033)	0.126*** (0.026)	0.088** (0.038)	0.414*** (0.111)	0.346*** (0.098)	0.279** (0.117)
Controls	-	log $w_{u,k,t}$	log $w_{u,k,t}$	-	log $w_{u,k,t}$	log $w_{u,k,t}$
Num.Obs.	797	766	766	324	314	314
R2	0.237	0.253	0.312	0.226	0.272	0.432
FE: Year	X	X	X	X	X	X
FE: Région			X			X

Table A.3: Urban density and rural land values.

Notes: Results of Regression Eq. A.4 for years $t \in \{1975, 1990, 2000, 2015\}$. Data on local farmland value $\bar{\rho}_{r,k,t}$ is the price of arable land in the ‘Petite Region Agricole (PRA) of city k . Average urban density is measured using GHSL data for a sample of 200 cities. For IV-Regressions (columns (4) to (6)), local farmland values are instrumented by wheat yields on the restricted sample of cities in départements with wheat as one of the main crop in 2000 ($(S_{r,wheat}/S_r)_{k,2000} > 20\%$). Controls are urban wages (in log), $w_{u,k,t}$, in city k (Column (2) and (5)) together with Regional dummies (Column (3) and (6)). Standard errors are clustered at the département level. Signif. Codes: ***=0.01, **=0.05, *=0.1.

Endogeneity and IV-estimates. Regarding the OLS-estimates, results should be taken with extreme care given measurement and endogeneity concerns. First, the local price of farmland is arguably measured with some errors as data regards the price of arable land. While this might be a good measure for cities surrounded by arable land with cereals as the main crop, some French cities are arguably surrounded by vineyards, land growing fruit trees or grazing fields. For these cities, the price of arable land might not be the best measure of local farmland values. Beyond measurement issues, the regression of Eq. A.4 faces endogeneity concerns. Beyond possible reverse causality whereby land is more valuable close to more productive and denser cities, estimates of b can be biased due to unobservable local characteristics: possible confounding factors like local amenities, land use regulations and others, might simultaneously affect the local price of farmland and the

³⁰We control for the log of urban wage instead of the log of the population of the urban area. In theory, the effect of higher farmland prices would be a reduced population $L_{u,k,t}$ and controlling for the urban population would capture part of the tested mechanisms. However, results are not much affected when controlling for log of urban population instead of log of urban wage. As expected, the estimated coefficient b is found to be smaller.

size/density of cities.³¹ Note that the bias could go either way: while local amenities might increase both farmland prices and urban density, land use regulations prevent cities to expand further at their fringe and might increase density and lower farmland prices—increasing locally the supply of farmland.

In any case, the OLS-estimates must be treated with extreme caution and we develop an IV-strategy. To do so, we dig for variations in farmland values arguably exogenous to the density of cities. In line with the theoretical model, an obvious candidate is the productivity of farmland (measured by yields per unit of land). However, as noticed above, the productivity of farmland depends on the crops grown on it (some land might be better suited for cereals, some land for vineyards). This makes it challenging to measure the farmland productivity without modeling the crop choice and analyzing relative prices for different crops—a task beyond this paper’s objectives. To circumvent this difficulty, we focus on homogenous regions, which grow very similar crops, cereals and more specifically wheat. To do so, we isolate départements, for which the share of land used for wheat, is above 20% (on average, in these départements the share of cereals’ land use is close to 50%). These are the départements in yellow on Figure A.24 (right panel), covering mostly the ‘Bassin Parisien’ and about a third of the French territory (35 départements)—81 cities of the 200 sample belong to these départements. Then, for these locations, we instrument city-level values of arable land using local wheat yields with the following first-stage,

$$\log \bar{\rho}_{r,k,t} = \tilde{a}_t + \tilde{b} \cdot \log \text{Yield}_{k,t} + \tilde{c} \cdot Z_{k,t} + u_{k,t}, \quad (\text{A.5})$$

where $\text{Yield}_{k,t}$ is the wheat yield at date t in the département of city k , \tilde{a}_t a time-effect and $Z_{k,t}$ the same set of region/city-specific controls. The first-stage is very strong as shown in Table A.4.

Results of the second-stage (Eq. A.4) are shown in Table A.3, columns (4) to (6). The elasticity b is close to 0.3—a 10% increase in the local price of arable land reduces urban density by about 3% (3.5% in the baseline specification of column (5)). Results are robust across specifications. While the coefficient is less significant once we control for region fixed-effects, this is not a major concern since this is driven by the important variations of yields across the boundaries of purely administrative regions.

Discussion. Our IV-strategy is valid to the extent that the instrument is a good predictor of farmland values (as validated by the first-stage), while not affecting urban density through other channels. One could for instance argue that the high productivity of larger and denser cities benefits agricultural productivity in the surrounding département. While one cannot exclude other confounding factors, we do not find any significant relationship between city size or wages and wheat yields in the département. Related to this, our instrument uses mostly (permanent) cross-sectional variations in wheat yields. As a consequence, one cannot identify the effects in the time-series,

³¹It is also important to note that the price of land close to cities might be particularly valuable in cities that are expected to grow fast in the future as this land might be converted into valuable urban land. In periods where growth is biased towards larger cities, this might also bias the coefficient.

	$\log \bar{\rho}_{r,k,t}$		
	(1)	(2)	(3)
$\log \text{Yield}_{k,t}$	1.984*** (0.257)	2.025*** (0.205)	1.654*** (0.339)
Controls	-	$\log w_{u,k,t}$	$\log w_{u,k,t}$
Num.Obs.	324	314	314
R2	0.750	0.762	0.806
FE: Year	X	X	X
FE: Région			X

Table A.4: First-Stage. Arable land values and wheat yields.

Notes: Results of the first-stage Regression Eq. A.5 for years $t \in \{1975, 1990, 2000, 2015\}$. Data on local farmland value $\bar{\rho}_{r,k,t}$ is the price of arable land in the ‘Petite Region Agricole (PRA) of city k . Data on wheat yield is the yield (per ha) in the département of city k . Restricted sample of 81 cities in départements with wheat as one of the main crop in 2000 ($(S_{r,wheat}/S_r)_{k,2000} > 20\%$). Controls are urban wages (in log), $w_{u,k,t}$, in city k (Column (2)) together with Regional dummies (Column (3)). Standard errors are clustered at the département level. Signif. Codes: ***=0.01, **=0.05, *=0.1.

without running into weak instruments. In particular, we cannot control for more granular FE than regional ones with this IV-strategy. Ideally, one would like to provide instruments which capture exogenous variations over time in farmland prices to identify the effect of farmland prices on urban density in the time-series—controlling for more granular city FE. As a robustness check detailed below, we lever up on the availability of different crops in the data of Schäuberger et al. (2022) and develop an alternative IV-strategy using spatial variations in land use for different crops interacted with national changes in yields of each crop (shift-share instruments). Lastly, the validity of our IV-strategy based on a selected sample is also threatened in the presence of heterogenous effects. Performing sensitivity analysis on the subsample of wheat producers partly addresses this concern.

Sensitivity with the same IV-strategy. We perform sensitivity for the selection of the sample using different thresholds for the fraction of land used for wheat. Results are robust for a fairly wide range of values for the selection threshold. Lowering the threshold below the baseline of 20% weakens the first-stage as expected—départements for which wheat yields are not measuring accurately land productivity are added. Results are robust for the sample of cities in départements for which wheat land use is above 10% ($(S_{r,wheat}/S_r)_{k,t} > 10\%$, keeping about 130 cities, the yellow and pink areas on Figure A.24, right panel). An estimated elasticity b very similar with a less stringent selection is also suggestive that the effect is not very much heterogeneous across space—a possible concern when running an IV-methodology on a restricted sample. Increasing the threshold is at the expense of a smaller sample of cities. Results hold for a higher threshold up to 30%. For a threshold strictly above 31% (twice the mean across départements), the sample of cities becomes very small (about 30 cities at most, almost all in the same Northern region ‘Picardie’) and the second-stage loses statistical power due to lack of variations in yields and farmland values.

We also perform sensitivity analysis controlling for urban population instead of urban wages—one could, for instance, argue that in presence of mobility frictions across France, cities are smaller close to the most productive agricultural land as people prefer working in agriculture (or, to the opposite innovation and/or a more skilled labor force in larger cities also benefit the productivity of farmland close by). Results are robust (the population of cities does not seem related to agricultural land yields).

Robustness checks with a different IV-strategy. We use the same data source on agricultural yields from Schaubberger et al. (2022) but levered up on the availability of different crops in these data. We have data since 1900 on yields for different crops (wheat, maize, oats, barley, potatoes, sunflower, sugar beet, rape, vineyards) together with the land area in the département used for the same crops. When available, $\text{Yield}_{j,k,t}$ denotes the yield of crop j and $s_{j,k,t}$ the share of land use for crop j , in département k at date t . We focus on five crops (wheat, maize, oats, barley, potatoes) which are present in almost all départements to avoid missing values for yields.

Based on these data, a strategy is to build shift-share instruments that could identify the effect in the time-series. The idea of the instrument is to use spatial variations in land use for different crops interacted with national changes in yields of each crop. To do so, we first compute for each date $t \in \{1975, 1990, 2000, 2015\}$ and each crop $j \in \{\text{wheat, maize, oats, barley, potatoes}\}$, the national aggregate yield, $\text{Yield}_{j,t}^a$ as the land use weighted-average of yields $\text{Yield}_{j,k,t}^a$ in each département k ,

$$\text{Yield}_{j,t}^a = \sum_k \bar{s}_{j,k,t} \cdot \text{Yield}_{j,k,t},$$

where $\bar{s}_{j,k,t} = \left(\frac{s_{j,k,t}}{\sum_k s_{j,k,t}} \right)$ is the land use weight for crop $j \in \{\text{wheat, maize, oats, barley, potatoes}\}$ in département k . Then, to build shift-share instruments, the aggregate yield of crop j , the common *shift* $\text{Yield}_{j,t}^a$, is interacted with the land use *share* of crop j in département k at an early date, 1960, $s_{j,k,1960}$. The set of shift-share instruments $\mathbf{IV}_{k,t}$ is the vector of aggregate yields (in log) interacted with the land use share in 1960 in département k ,

$$\mathbf{IV}_{k,t} = \{s_{j,k,1960} \cdot \log(\text{Yield}_{j,t}^a)\}_{j \in \{\text{wheat, maize, oats, barley, potatoes}\}} \quad (\text{A.6})$$

The main advantage of this instrument, which uses time-series variations as opposed to our baseline IV-strategy based on cross-sectional variations, is the possibility of controlling for more granular FE in the regressions.

Results of the second stage with département/city FE are shown in Table A.5 on the selected sample of cities in départements that are significant producers of cereals—cereals constituting most of the land use of the crops used in the first-stage (sample restricted to département where land use for cereals is above 30% in 2000). Results show that the estimated effect of land prices on urban density is very similar to our baseline estimates identified in the cross-section and robust with more granular city-FE—only slightly higher than our baseline estimates.

While this shift-share strategy provides instruments along the time-dimension for each location (each département where the city is located), this comes at the cost of not providing immediate economic interpretation regarding coefficients in the first-stage. Estimated coefficient in the first-stage are shown in Table A.6. They are mostly negative but could in principle go either way depending on conflicting forces: on one side, specialization in the crops with the largest increase in yields tend to appreciate farmland values; on the other side, a faster increase in aggregate yields of a given crop would decrease its relative price, exerting downward pressure on farmland prices in the locations specialized in such a crop. One would need a model of crop choice with well-defined preferences across crops to shed light on estimated coefficients in the first-stage—a task beyond our paper. This strategy also comes at the cost of an overall weaker instrument and weaker first-stage compared to the baseline strategy using cross-sectional variations—the time-span (1975-2015) with only 4 dates of observation being fairly limited.

	log Urban Density		
	(1)	(2)	(3)
$\log \bar{\rho}_{r,k,t}$	0.4078** (0.1730)	0.4270** (0.1792)	0.4296** (0.1888)
Controls	-	$\log w_{u,k,t}$	$\log w_{u,k,t}$
FE: Year	X	X	X
FE: Département	X	X	
FE: City			X
Num.Obs.	444	430	430
R2	0.450	0.449	0.849

Table A.5: Urban density and rural land values: robustness with alternative IV.

Notes: Results of IV-Regressions (shift-share IV) for Eq. A.4 with more granular FE (département or city-FE) for years $t \in \{1975, 1990, 2000, 2015\}$. Data on local farmland value $\bar{\rho}_{r,k,t}$ is the price of arable land in the ‘Petite Region Agricole (PRA) of city k . Average urban density is measured using GHSL data for a sample of 200 cities. Local farmland values are instrumented by the shift-share instruments (A.6) on the restricted sample of cities in départements with cereals as main crops in 2000 ($(S_{r,cereal}/S_r)_{k,2000} > 30\%$). Regressions include Département FE (Columns (1) and (2)), city-FE (column (3)) controlling for urban wages (in log), $w_{u,k,t}$, in city k (Columns (2) and (3)). Standard errors are clustered at the département level. Signif. Codes: ***=0.01, **=0.05, *=0.1.

Lastly, we perform additional sensitivity analysis using related identification strategies using the data on yields/land use by crops across locations. This sensitivity analysis (not reported) gives similar results. In particular, we add to the shift-share IV by crops, local yields for these crops at each date (instead of national ones). Results were very similar.

	$\log \bar{\rho}_{r,k,t}$		
	(1)	(2)	(3)
Shift-SHare Wheat	-2.468** (1.029)	-2.353** (1.022)	-2.286** (1.052)
Shift-SHare Maize	-2.559*** (0.9456)	-2.409** (0.9111)	-2.342** (0.9180)
Shift-SHare Oats	-2.855 (2.251)	-2.912 (2.251)	-2.750 (2.195)
Shift-SHare Barley	-3.071*** (1.081)	-3.114*** (1.101)	-3.124*** (1.102)
Shift-SHare Potatoes	-0.5852 (1.954)	-0.6411 (1.913)	-0.6588 (1.879)
Controls	-	$\log w_{u,k,t}$	$\log w_{u,k,t}$
FE: Year	X	X	X
FE: Département	X	X	
FE: City			X
Num.Obs.	444	430	430
R2	0.841	0.840	0.907

Table A.6: First-Stage: robustness with alternative IV.

Notes: Results of the first-stage regression (shift-share IV) with more granular FE (département or city-FE) for years $t \in \{1975, 1990, 2000, 2015\}$. Data on local farmland value $\bar{\rho}_{r,k,2000}$ is the price of arable land in the ‘Petite Region Agricole (PRA) of city k . Shift-share instruments (A.6) on the restricted sample of cities in départements with cereals as main crops in 2000 ($(S_{r,cereal}/S_r)_{k,t} > 30\%$). Regressions include Département FE (Columns (1) and (2)), city-FE (column (3)) controlling for urban wages (in log), $w_{u,k,t}$, in city k (Columns (2) and (3)). Standard errors are clustered at the département level. Signif. Codes: ***=0.01, **=0.05, *=0.1.

A.5 Urban Individual Data

We use individual data from the ‘Enquête National du Logement (ENL)’ and from the ‘Déclaration annuelle des données sociales’ (DADS) in order to investigate individual commuting behavior in urban areas over space and time (Sections A.5.1 and A.5.2). These data allow to measure the commuting elasticities necessary for the calibration of the quantitative model. We also compute the average wage by city using the DADS panel EDP (version 2019) (Section A.5.3). Cities are denoted by the index k , individuals by i and dates by t .

A.5.1 Individual Commuting Data from ENL

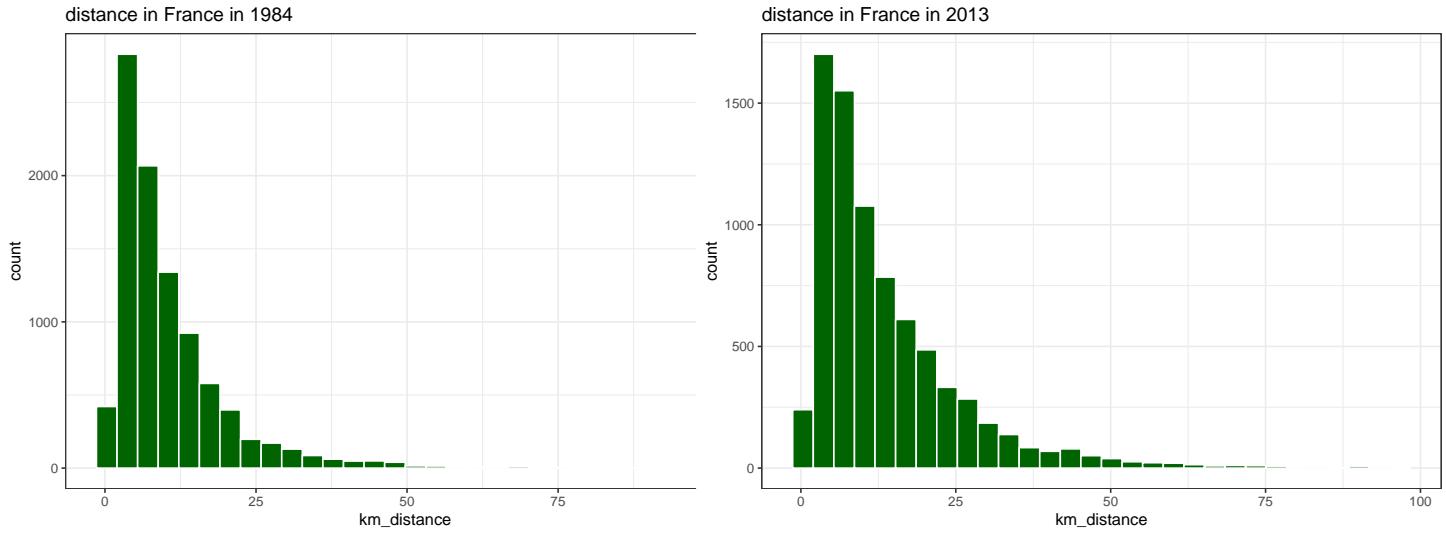
Data from Enquête National du Logement (ENL). We obtain confidential access to the ENL and use it to measure commuting speed as a function of commuting distance. The ENL asks respondents questions about commuting behavior, mode of commute, and importantly, duration of commute in minutes.

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 individuals working outside their home and to be the reference person in the 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 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 A.27a and A.27b illustrate the distributions of the commuting distance variable in 1984 and 2013. 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.

Elasticity of speed w.r.t commuting distance. We are interested in measuring the elasticity of speed w.r.t commuting distance in a given year. Grouping data into 50 bins of log distance, Figure A.28 illustrates the relationship between log of speed and log of commuting distance for the years 1984 and 2013. For each ENL wave (1984, 1988, 2006, 2013), we perform the following regression at the individual level,

$$\ln \text{speed}_i = \beta_0 + \beta_1 \ln \text{dist}_i + \beta_2 \cdot Z_i + u_i,$$



(a) Distribution of Commuting Distances in 1984..

(b) Distribution of Commuting Distances in 2013.

Figure A.27: Distribution of Commuting Distances in 1984 and 2013

Notes: Distribution of Commuting Distances for a representative French Sample in 1984 and 2013 from ENL data.

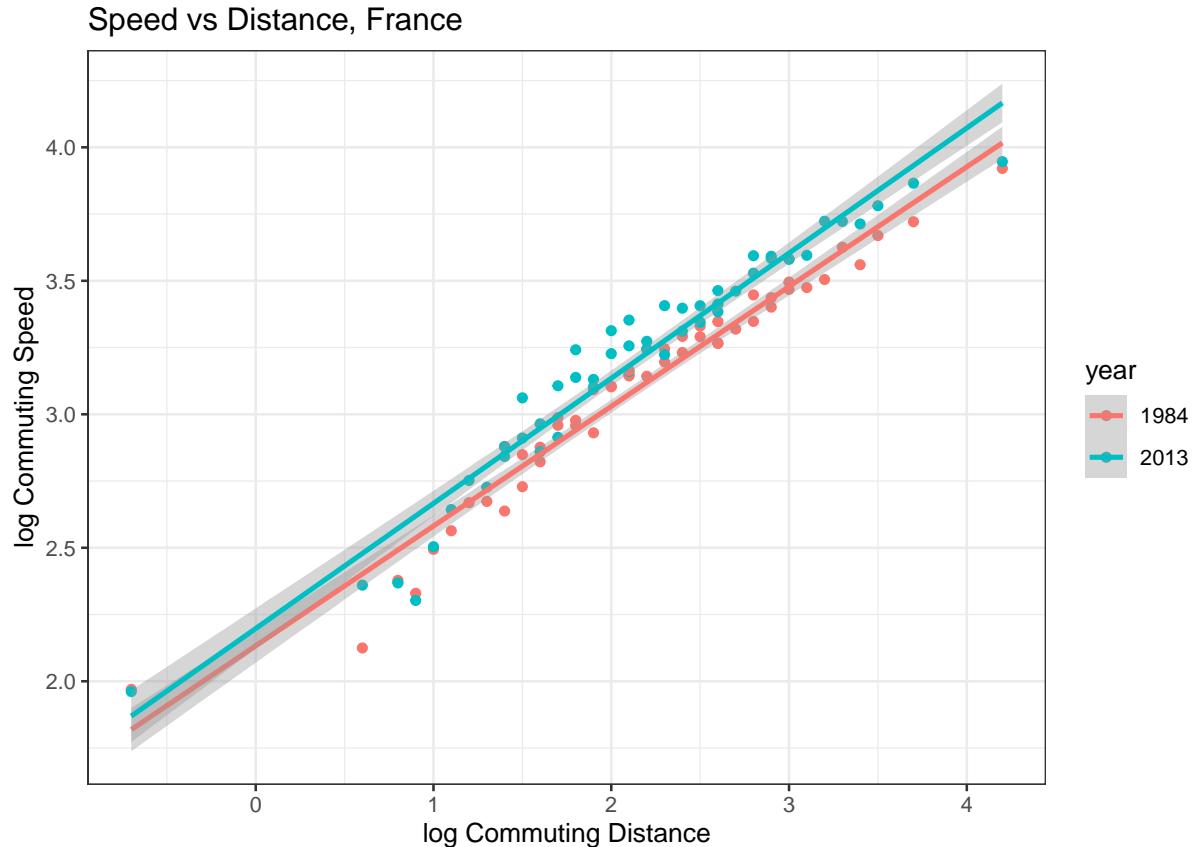


Figure A.28: Commuting speed and commuting distance (1984 and 2013).

Notes: Commuting speed for 50 bins of commuting distance (in log). Data source: ENL

where speed_i is the speed of individual i , dist_i its commuting distance and Z_i a set of individual controls (income, education, age, ...) and regional dummies. Regression results are reported in Tables A.7 and A.8 for the years 1984 and 2013 (we omit the 1988 and 2006 waves for brevity but results are very similar across years). Across specifications with different control variables and 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 (regression without controls of log speed on log distance in 1984). Since 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 *** (0.007)	0.505 *** (0.007)	0.498 *** (0.007)	0.502 *** (0.007)	0.470 *** (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 A.7: 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 *** (0.007)	0.478 *** (0.007)	0.469 *** (0.007)	0.474 *** (0.007)	0.438 *** (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 A.8: Cross sectional regression of Speed on Commuting Distance using ENL 2013 data. Columns are specified as in table A.7.

Evolution of speed at a given commuting distance. We investigate how the average commuting speed has evolved, controlling for commuting distance, between 1984 and 2013. To achieve this, we pool two cross sections a date $t = 1984$ and $t = 2013$ and run the following regression by bins b of commuting distance:

$$\ln \text{speed}_{b,t} = \beta_0 + \beta_1 \ln \text{dist}_b + \beta_2 \text{year}_t + u_{b,t},$$

where $\text{speed}_{b,t}$ is the average speed of households in distance-bin b at date t , dist_b the average commuting distance in distance bin b , and year_t a dummy equal to one in 2013.

Results are reported in Table A.9 (see Figure A.28 for the graphical representation). We use the

	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 A.9: 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 A.28.

regression results to measure the magnitude of the shift over time in the intercept—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 time dummy for `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 4.2 in the main text.

A.5.2 Individual Commuting Data from DADS

Data from Déclaration annuelle des données sociales (DADS). We 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 and the large sample size allows to study the link between commuting distance and residential location at the city-level—the ENL sample being too small.

Commuting distance and residential location. The monocentric model implies that the location of residence ℓ maps one for one into commuting distance. Extension B.3.5 (see Section 4.6 in the main text) relaxes this assumption. We introduce in a reduced-form way the following relationship between commuting distance $d_k(\ell_k)$ and distance from the city center ℓ_k in city k of radius ϕ_k ,

$$d_k(\ell_k) = d_0(\phi_k) + d_1(\phi_k) \cdot \ell_k \quad (\text{A.7})$$

where $d_0(\phi)$ and $d_1(\phi)$ are parametric functions of the city radius ϕ as detailed in extension B.3.5—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 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 75056.³² 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 between center and residence locations 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 ($\text{distance}_{\text{commute}_i}$) and the residential distance from the city center ($\text{distance}_{\text{center}_i}$) for each individual i in city k , we perform the following regression,

$$\text{distance}_{\text{commute}_i} = \gamma_{0,k(i)} + \gamma_{1,k(i)} \cdot \text{distance}_{\text{center}_i} + u_i \quad (\text{A.8})$$

where i indexes an individual in DADS, $k(i)$ is the city k (urban area) to which i belongs, and u_i is a mean-independent error term. $\gamma_{0,k(i)}$ and $\gamma_{1,k(i)}$ 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, ...).

Figure A.29 plots the distribution of the intercept coefficient $\gamma_{0,k(i)}$ 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 A.30 plots the distribution of the slope coefficient $\gamma_{1,k(i)}$ 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 $\gamma_{0,k(i)}$ increases with city size, from about 0.2 km for the smallest urban areas to more than 4 kms for Paris. The slope coefficient $\gamma_{1,k(i)}$ decreases with city size—ranging from around 0.4 for Paris to more than 0.7 for the small urban areas.

These results validate our reduced-form parametrization (Eq. A.7), where commuting distance $d(\ell)$ increases less than proportionately with residential location ℓ , and less so for larger cities (larger radius ϕ_k). We use these findings in Section B.3.5 to parametrize $d_0(\phi)$ and $d_1(\phi)$.

³²This 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.

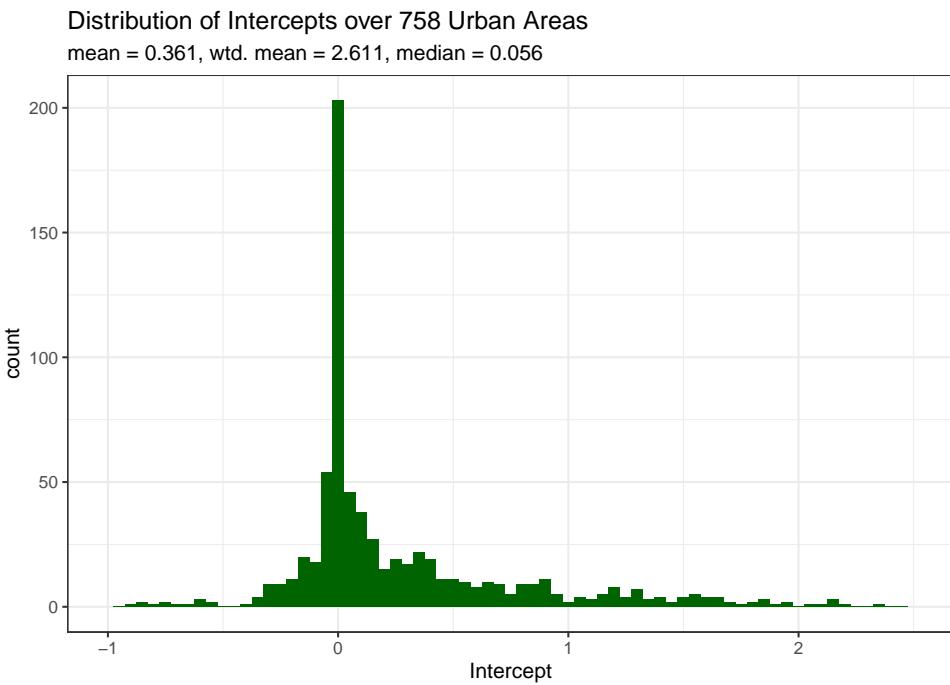


Figure A.29: Distribution of DADS intercept estimates

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

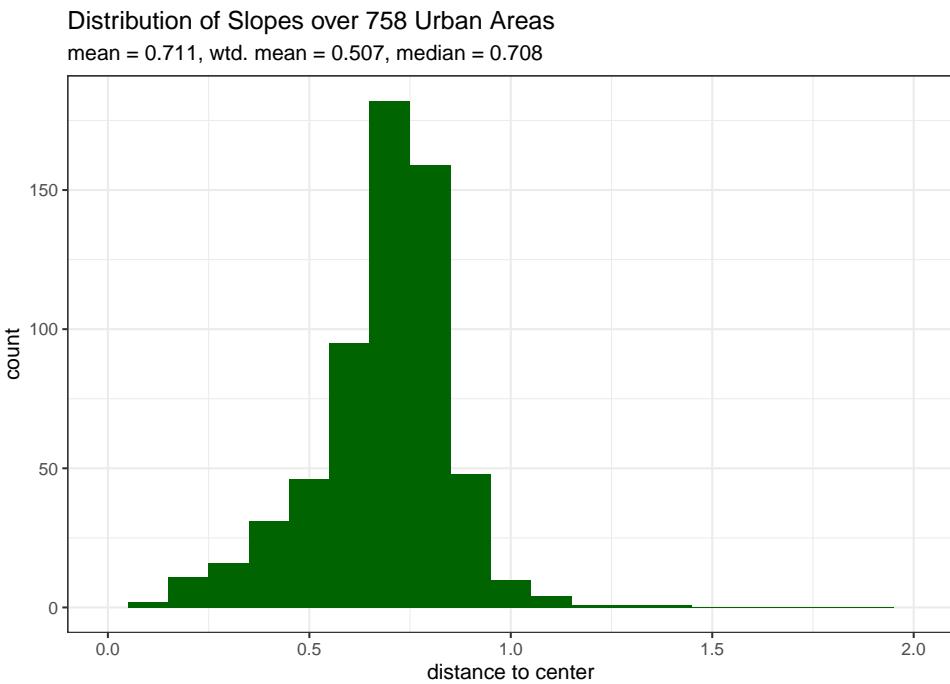


Figure A.30: Distribution of DADS slope estimates

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

A.5.3 Urban Productivity and Wages

Data. In Appendix A.4, we need to control for the urban productivity (urban wage) at the city level, $w_{u,k,t}$, for each city k and date $t \in \{1975, 1990, 2000, 2015\}$. In order to measure city-level urban wages, we use the DADS panel EDP version 2019 which goes back until 1976. We assign 1976 to the year 1975. Notice that there is no wage data available before 1976. The data provide the net salary for a representative sample of workers in each urban area.

Sample selection. As we do not observe hours worked, we first implement a procedure on the panel to select the sample of observations and get as close as possible to the notion of a *full time worker* in the private sector. The sample of cities considered is the sample of 200 cities considered in Appendix A.4. We follow the labor literature (Schmutz and Sidibé (2019)) to select the sample of workers. The sample selection is shown in Table A.10. The number of observations by year ranges from about 30,000 in years $t \in \{1975, 1990, 2000\}$ and about 270,000 in 2015.

Table A.10: DADS Panel 2019-EDP Subsetting Procedure

Sample	Criterion
2,147,723	Full Sample
1,061,697	Males Only
1,057,428	Metropolitan France Only
976,187	Part of unique Urban Area
670,551	Full Time Workers
582,651	Workers not in Public Sector
575,299	Not Postal Office or Telecom
575,219	No Distance to UA center available
558,889	Positive Salary
553,298	Salary below 99-th %-ile by year
551,083	Age 15-65
417,620	Workers with Single Job by year
365,747	In relevant Urban Area (200 cities sample)

Measurement. For each city k and date $t \in \{1975, 1990, 2000, 2015\}$, we compute the mean net salary (across full-time male private workers) to measure $w_{u,k,t}$. These data are used as control for the regressions in Section A.4. Data are available for most cities in the sample of 200 cities used in Section A.4, with few missing observations for small cities due to an insufficient number of individual observations.

Remarks. To estimate the average urban productivity of a given city, we would like to control for the composition of the workforce across cities and compute an urban area fixed effect for each city, controlling for various worker-level observables (education and age). Unfortunately, the sample size is too small in the earlier years to reliably compute a fixed effect for each urban area. For the year 2015, the sample is significantly larger and we are able to estimate city fixed-effects when controlling for observables (age and education). We find that our raw measure, the log of unconditional mean of

net salaries across workers, yields a measure very highly correlated with city fixed-effects (correlation of 0.76). This is reassuring that our raw measure of $w_{u,k,t}$ is a reasonable proxy for a cities fixed effects (e.g. the city-specific urban productivity). We also expect a positive relationship in the city-wide average net salary and the population of the urban area. For each year t , we regress the log of the measured average wage, $\log w_{u,k,t}$, on the log of the population of the urban area, $\log L_{u,k,t}$. We find a highly significant positive relationship, robust across all years.

A.6 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 10a)). 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 as 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 provides *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.³³ 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 (distributed 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.³⁴ 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 'Enquête 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 location to extract the distribution of mode use and their respective speeds to compute the

³³Note 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.

³⁴Note 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.

average commuting speed.³⁵ 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 A.32).

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 digitized 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 1905.³⁶ 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 (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.³⁷ The main issue

³⁵The sample raw average commuting speed at each date gives very similar estimates.

³⁶The horse-drawn omnibus disappears in 1913.

³⁷We 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.

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 recovered 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 $\hat{x} = 31\%$.³⁸ 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 A.31, 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 use of automobiles. It is still around 40% in the recent years.

Private transportation. Private transportation includes essentially private cars, bikes and motor-bikes.³⁹ 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 1930.⁴⁰ We also collected data

³⁸One 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.

³⁹Pre-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.

⁴⁰In 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.

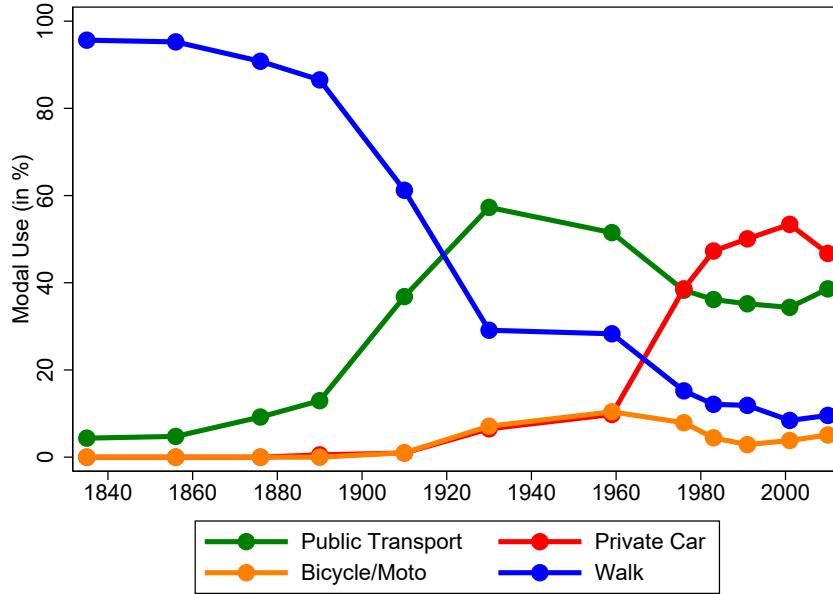


Figure A.31: Transportation mode use in the Parisian urban area.

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).

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 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 A.31 summarizes the evolution of the proportion of workers using their private cars for work commutes.

The use of bikes and motorbikes was almost nonexistent 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.⁴¹ Given the importance of bicycles for leisure and the lack of relevant data post-

⁴¹[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.

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 (Enquête nationale transports et déplacements) where people were asked their main mode of transport over their lifetime, one 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%.⁴² 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%.⁴³ 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 A.31 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 A.31 summarizes the estimates for the share of workers walking to work over the whole period. In the early years, before 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 current modes of transportation (public or private) might 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

⁴²For 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.

⁴³Traffic 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.

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 set 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 A.32 summarizes the estimated speed of the different modes, by mode at different dates. Figures A.33 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.

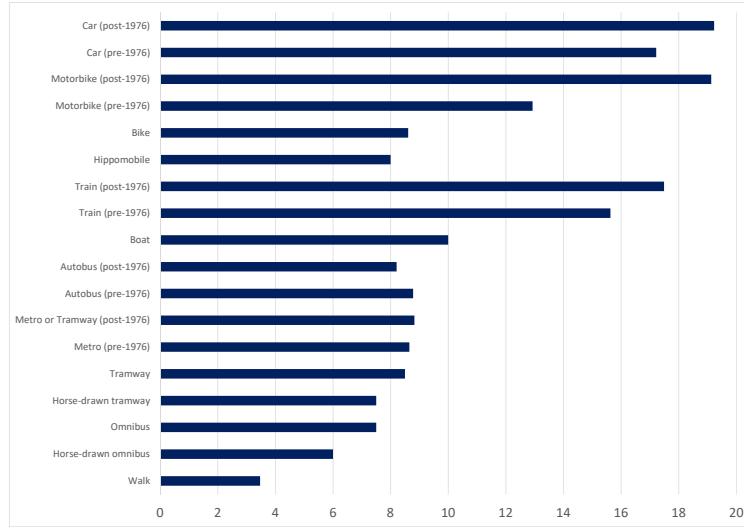


Figure A.32: Speed across transportation modes.

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.

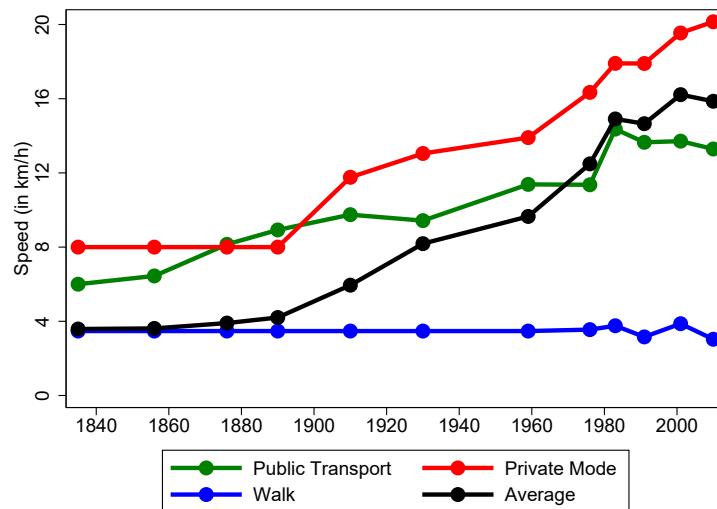


Figure A.33: Evolution of average speed across mode categories.

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.

Bibliography

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.

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, Laurent Gobillon, Gilles Duranton, and Clement Gorin, “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).

Orsellie, 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.

Schauberger, Bernhard, Hiromi Kato, Tomomichi Kato, Daiki Watanabe, and Philippe Ciais, “French crop yield, area and production data for ten staple crops from 1900 to 2018 at county resolution,” *Scientific Data*, 2022, 9 (1), 38.

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

Schmutz, Benoît and Modibo Sidibé, “Frictional labour mobility,” *The Review of Economic Studies*, 2019, 86 (4), 1779–1826.

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.

Structural Change, Land Use and Urban Expansion
Online Appendix B — Quantitative Model

Nicolas Coeurdacier
SciencesPo Paris, CEPR

Florian Oswald
SciencesPo Paris

Marc Teignier
Serra Húnter Fellow,
University of Barcelona

July 2, 2024

Contents

B.1	Model details	2
B.1.1	Set-up Description	2
B.1.2	Technology	2
B.1.3	Commuting Costs	3
B.1.4	Preferences and Budget Constraint	5
B.1.5	Location Sorting	6
B.1.6	Housing market description with location-specific housing supply	6
B.1.7	Market Clearing	8
B.1.8	Equilibrium Definition	9
B.1.9	Dynamic Optimization and the Real Interest Rate	10
B.2	Quantitative Evaluation	12
B.2.1	Multi-Region Numerical Illustration	12
B.2.2	Data Inputs for the Model	17
B.2.3	Mapping of Model Outputs to the Data Inputs	21
B.2.4	Solution and Estimation Algorithm	27
B.2.5	Untargeted Model Outputs	33
B.3	Sensitivity Analysis and Extensions	43
B.3.1	Elasticity of substitution between rural and urban goods σ	43
B.3.2	Elasticity of substitution between land and labor ω	44
B.3.3	Housing Supply Elasticity ϵ	45
B.3.4	Congestion and Agglomeration	46
B.3.5	Commuting distance and residential location	51

B.1 Model details

In this Section we present the quantitative model in detail. For sake of space, some of the elements were not discussed in the main text and are introduced here. In particular, we add the discussion of the commuting choice model that provides a micro-foundation to the commuting costs functional form described in the main text; we extend the model from the main text with a CES technology in the rural sector; we also describe the housing market equilibrium with location-specific housing supply and, finally, we introduce an intertemporal utility function to pin down the path for the real interest rate.

B.1.1 Set-up Description

Multiple Regions. The economy is made up of K different regions, each endowed with area S . Each region $k \in \{1, \dots, K\}$ is made up of urban and rural land, with only one city per region – we will use "city" and "region" interchangeably if unambiguous. Regions are heterogeneous in their urban and rural productivities. $\theta_{u,k}$ is the urban productivity in city k and is $\theta_{r,k}$ rural productivity in region 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. Land rents per worker r are redistributed equally.

Circular City. Regions are assumed to be circular of radius $\sqrt{S/\pi}$ and the city in each region k is centrally located and circular around its center with endogenous radius ϕ_k and area $\pi\phi_k^2$.¹ We denote ℓ_k a location in a region k . Due to symmetry, the location $\ell_k \in (0, \phi_k)$ in city k also denotes the commuting distance to the center of city k .

Time Sequence. The lifetime utility of ex-ante identical, infinitely-lived consumers is the discounted flow of instantaneous utilities with geometric discounting. Optimal choices of agents over time pin down a path for the real interest rate. For most of the Appendix, we abstract from t indices as the spatial equilibrium remains static due to perfect mobility at each date t . The dynamic formulation serves the sole purpose of determining a time path of the real interest rate, and thus the appropriate discount factor needed to compute land *values* (instead of only land *rents*).

B.1.2 Technology

Production and Factor Payments. Given regional urban productivity parameters $\{\theta_{u,k}\}$, the regional production in sector u is

$$Y_{u,k} = \theta_{u,k} L_{u,k}$$

where $L_{u,k}$ denotes the urban workers in region k . Urban workers are paid their marginal productivity such that,

$$w_{u,k} = \theta_{u,k}. \quad (\text{B.1})$$

¹Regions are assumed large enough in area such that cities do not expand in neighboring regions. S is large enough such that for all cities, $\phi_k < \sqrt{S/\pi}$.

In the rural sector, we extend 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 in each region k ,

$$Y_{r,k} = \theta_{r,k} \left(\alpha (L_{r,k})^{\frac{\omega-1}{\omega}} + (1-\alpha) (S_{r,k})^{\frac{\omega-1}{\omega}} \right)^{\frac{\omega}{\omega-1}},$$

where $L_{r,k}$ denotes the number of workers working in the rural sector in region k , $S_{r,k}$ the amount of land used for production and $\theta_{r,k}$ a Hicks-neutral productivity parameter. $0 < \alpha < 1$ is the intensity of labor use in production. $\omega \geq 0$ is the elasticity of substitution between labor and land, $\omega = 1$ corresponding to the baseline version.

Rural workers and land are paid their marginal productivities such that main text Equations (11) and (12) become for each region k ,

$$w_{r,k} = \alpha p \theta_{r,k} \left(\alpha + (1-\alpha) \left(\frac{S_{r,k}}{L_{r,k}} \right)^{\frac{\omega-1}{\omega}} \right)^{\frac{1}{\omega-1}} \quad (\text{B.2})$$

$$\rho_r = (1-\alpha)p \theta_{r,k} \left(\alpha \left(\frac{L_{r,k}}{S_{r,k}} \right)^{\frac{\omega-1}{\omega}} + (1-\alpha) \right)^{\frac{1}{\omega-1}} \quad (\text{B.3})$$

where $w_{r,k}$ is the rural wage and ρ_r the rental price of land in region k 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,

$$\rho_{r,k} = \left(\frac{1-\alpha}{\alpha} \right) w_{r,k} \left(\frac{L_{r,k}}{S_{r,k}} \right)^{\frac{1}{\omega}}. \quad (\text{B.4})$$

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

B.1.3 Commuting Costs

As mentioned in the main text, commuting costs are partly endogenous in our framework because urban households adjust their mode of commuting m depending on their location ℓ and opportunity cost of time (wage rate w_u). In particular, commuting costs in location ℓ_k , $\tau(\ell_k)$, are the sum of spending on commuting using transport mode m_k , $f = f(\ell_k, m_k, w_{u,k})$, and time-costs proportional to $w_{u,k} \cdot t(\ell_k)$, where $t(\ell_k)$ denotes the time spent on daily commutes of an individual located in ℓ_k , such that

$$\tau(\ell_k) = f(\ell_k, m_k, w_{u,k}) + 2\zeta w_{u,k} \cdot t(\ell_k), \quad (\text{B.5})$$

where $0 < \zeta \leq 1$ represents the valuation of commuting time in terms of foregone wages. Transportation modes m are continuously ordered by their speed, as in DeSalvo and Huq (1996), such

that m denotes both the mode and the speed of commute. The commuting time (both ways) is, therefore, $t(\ell) = \frac{2\ell}{m}$. The cost $f = f(\ell, m, w_u)$ depends on the transportation mode/speed m , the location ℓ and labor costs w_u .² Intuitively, beyond its speed, the pecuniary cost of a commuting mode depends on the distance traveled (e.g. cost of gasoline) as well as the level of wages (e.g. wage of the bus driver). 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 $\frac{\partial^2 f}{\partial^2 w_u} \leq 0$. For tractability, we will use the following functional form for f ,

$$f(\ell, m, w_u) = \frac{c_\tau}{\eta_m} \cdot m^{\eta_m} \cdot \ell^{\eta_\ell} \cdot w_u^{\eta_w}, \quad (\text{B.6})$$

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, common across regions.

At any given moment in time, prevailing technology offers different transportation modes ordered by their respective speed m . An individual in location ℓ_k chooses the mode of transportation corresponding to speed m_k in order to minimize the commuting costs $\tau(\ell_k)$. This equalizes the marginal cost of a higher speed m_k to its marginal benefits in terms foregone wage,

$$\frac{\partial f}{\partial m_k} = 2\zeta \cdot w_{u,k} \left(\frac{\ell_k}{m_k^2} \right).$$

Using Eq. B.6, the optimal chosen mode/speed satisfies

$$m_k = \left(\frac{2\zeta}{c_\tau} \right)^{\frac{1}{1+\eta_m}} \cdot w_{u,k}^{1-\xi_w} \cdot \ell_k^{1-\xi_\ell}, \quad (\text{B.7})$$

where $\xi_w = \frac{\eta_m + \eta_w}{1 + \eta_m} \in (0, 1)$ and $\xi_\ell = \frac{\eta_m + \eta_\ell}{1 + \eta_m} \in (0, 1)$. Individuals living further away choose faster commuting modes. The speed of commuting also increases with the wage rate as a higher wage increases the opportunity cost of time. Using Eqs. B.5-B.7, we get that equilibrium commuting costs satisfy,

$$\tau(\ell_k) = a \cdot w_{u,k}^{\xi_w} \cdot \ell_k^{\xi_\ell}, \quad (\text{B.8})$$

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), while they are increasing with the wage rate in each city (the opportunity cost of time) and the distance of commuting trips with constant elasticities. Expression (B.8) 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 map into elasticities of commuting

²The cost $f(\ell, m, w)$ has several possible interpretations. At a more macro level, it can represent the fixed cost of installing public transportation, where a faster mode is more expensive (a train line versus the horse drawn omnibus). At a more individual level, it represents the cost of buying an individual mean of transportation—a bike being cheaper than an automobile. However, this reduced-form approach sets aside the possibility that the implemented commuting technologies and their speed depend in a more sophisticated way on the equilibrium allocation in the city (e.g. traffic congestion or the construction of transport infrastructures may depend on the spatial allocation of urban residents).

speed to income (resp. commuting distance) through Equation B.7. We use this link to directly parametrize both ξ_w and ξ_ℓ .

B.1.4 Preferences and Budget Constraint

Preferences. Consumption over urban and rural goods are non-homothetic. Consider a worker living in a location ℓ_k of region k . Denote $c_r(\ell_k)$ the consumption of rural (agricultural) goods, $c_u(\ell_k)$ the consumption of urban goods (used as numeraire) and $h(\ell_k)$ the consumption of housing. The composite consumption good is

$$C(\ell_k) = \mathcal{C}(c_r(\ell_k), c_u(\ell_k))^{1-\gamma} h(\ell_k)^\gamma, \quad (\text{B.9})$$

where the housing preference parameter γ belongs to $(0, 1)$ and the consumption composite \mathcal{C} over rural and urban goods is a CES aggregate with substitution elasticity σ ,

$$\mathcal{C}(c_r(\ell), c_u(\ell)) = \left[\nu^{1/\sigma} (c_r(\ell) - \underline{c})^{\frac{\sigma-1}{\sigma}} + (1 - \nu)^{1/\sigma} (c_u(\ell) + s)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}.$$

\underline{c} denotes the minimum consumption level for the rural (subsistence) good, s stands for the initial endowment of the urban (luxury) good and the preference parameter ν belongs to $(0, 1)$. Preferences are Stone-Geary for $\sigma = 1$. Workers derive utility only from consumption. The utility of a household in location ℓ_k is equivalent to $C(\ell_k)$.

Budget constraint. The household earns a wage income net of spatial frictions $w(\ell_k)$ in location ℓ_k of region k . Given the spatial structure, $w(\ell_k) = w_{u,k} - \tau(\ell_k)$ for $\ell_k \leq \phi_k$ and $w(\ell_k) = w_{r,k}$ for $\ell_k \geq \phi_k$. The households also earn land rents, r , redistributed lump-sum and equally across workers.

Workers can borrow and lend at the risk-free gross interest rate R and the budget constraint of a worker in location ℓ_k of region k satisfies

$$pc_r(\ell_k) + c_u(\ell_k) + q(\ell_k)h(\ell_k) = w(\ell_k) + r + RB - B', \quad (\text{B.10})$$

where B (resp. B') are inherited (resp. next period) bond holdings and $q(\ell_k)$ the rental price per unit of housing in location ℓ_k of region k . Given that all workers are ex-ante identical, there is no borrowing and lending in equilibrium, $B = B' = 0$ and the budget constraint remains the static one,

$$pc_r(\ell_k) + c_u(\ell_k) + q(\ell_k)h(\ell_k) = w(\ell_k) + r.$$

B.1.5 Location Sorting

Mobility conditions. Workers can freely move within each region k , as well as across regions. Within region k , this gives the following mobility equation. For all location ℓ_k in region k ,

$$\overline{C}_k = \kappa \frac{w(\ell_k) + r + \underline{s} - p\underline{c}}{q(\ell_k)^\gamma}. \quad (\text{B.11})$$

These mobility conditions generate housing rental price gradients in each city k .

Workers can freely move across regions k equalizing consumption of the urban and rural worker at the fringe across the different regions. For all regions $k \in \{1, \dots, K\}$,

$$\overline{C}_k = \overline{C} = \kappa \frac{w_{u,k} - \tau(\phi_k) + r + \underline{s} - p\underline{c}}{(q_{r,k})^\gamma} = \kappa \frac{w_{r,k} + r + \underline{s} - p\underline{c}}{(q_{r,k})^\gamma}, \quad (\text{B.12})$$

where $q_{r,k}$ is the housing rental price at the fringe of city k , equal to the rental price for all locations $\ell_k \geq \phi_k$ in region k .

B.1.6 Housing market description with location-specific housing supply

Location-specific housing supply. As shown in Baum-Snow and Han (2023), 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_k of city k , land developers supply housing space $H(\ell_k)$ per unit of land with a convex cost

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

paid in units of the numeraire, where $1/\epsilon(\ell_k)$ can depend on the location ℓ_k . 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

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

where $\rho(\ell_k)$ is the rental price of a unit of land in location ℓ_k of city k . Similarly to the housing price $q(\ell_k)$ above, for locations beyond the fringe ϕ_k , the land rent is constant, hence $\rho_{r,k} = \rho(\ell_k \geq \phi_k)$.

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

$$H(\ell_k) = q(\ell_k)^{\epsilon(\ell_k)},$$

where the parameter $\epsilon(\ell_k)$ is the price elasticity of housing supply in location ℓ_k . 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 and identical across regions, $\epsilon_r = \epsilon(\ell_k \geq \phi_k)$.

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

$$\rho(\ell_k) = \frac{q(\ell_k)H(\ell_k)}{1 + \epsilon(\ell_k)} = \frac{q(\ell_k)^{1+\epsilon(\ell_k)}}{1 + \epsilon(\ell_k)}.$$

Arbitrage across land usage implies that the latter land rental price $\rho(\ell_k)$ is in equilibrium above the marginal productivity of land for production of the rural good, where the condition holds with equality in the rural part of the economy, for $\ell_k \geq \phi_k$,

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

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.

Urban Housing Market Equilibrium. Consider first locations within city k , $\ell \leq \phi_k$. Market clearing for housing in each location implies $H(\ell_k) = D_k(\ell_k)h(\ell_k)$, where $D_k(\ell_k)$ denotes the density (number of urban workers) in location ℓ_k of city k . Using the housing rental price gradient in each city k and the housing demand in each location ℓ_k , the density $D_k(\ell_k)$ follows for $\ell \leq \phi_k$,

$$D_k(\ell_k) = \left(\frac{q_{r,k}^{1+\epsilon(\ell_k)}}{1 + \epsilon(\ell_k)} \right) \frac{1}{\gamma_{\ell_k}} (w(\phi) + r + \underline{s} - p\underline{c})^{-1/\gamma_{\ell_k}} (w(\ell_k) + r + \underline{s} - p\underline{c})^{1/\gamma_{\ell_k}-1},$$

where $w(\ell_k)$ is the wage net of commuting costs in location ℓ_k of city k , $\gamma_{\ell_k} = \frac{\gamma}{1+\epsilon(\ell_k)}$ represents the spending share on housing adjusted for the supply elasticity in location ℓ of city k and the fringe housing price $q_{r,k}$ satisfies $\rho_{r,k} = \frac{(q_{r,k})^{1+\epsilon_r}}{1 + \epsilon_r}$.

Integrating density, $D_k(\ell_k)$, across urban locations gives the total urban population of city k ,

$$L_{u,k} = \int_0^{\phi_k} D_k(\ell_k) 2\pi d\ell_k = \int_0^{\phi_k} \left(\frac{q_{r,k}^{1+\epsilon(\ell_k)}}{1 + \epsilon(\ell_k)} \right) \frac{1}{\gamma_{\ell_k}} (w(\phi_k) + r + \underline{s} - p\underline{c})^{-1/\gamma_{\ell_k}} (w(\ell_k) + r + \underline{s} - p\underline{c})^{1/\gamma_{\ell_k}-1} 2\pi d\ell_k \quad (\text{B.13})$$

Note that with homogeneous supply conditions across locations, $\epsilon(\ell) = \epsilon_r = \epsilon$, Equation (B.13) simplifies into Equation (19) of the main text.

$$L_{u,k} = \int_0^{\phi_k} D_k(\ell_k) 2\pi d\ell_k = \rho_{r,k} \int_0^{\phi_k} \frac{1 + \epsilon}{\gamma} (w(\phi_k) + r + \underline{s} - p\underline{c})^{-\frac{1+\epsilon}{\gamma}} (w(\ell_k) + r + \underline{s} - p\underline{c})^{\frac{1+\epsilon}{\gamma}-1} 2\pi d\ell_k.$$

B.1.7 Market Clearing

The land market clears locally in each region k , while labor and goods markets clear globally.

Land Market Clearing. In the rural area, $\ell_k \geq \phi_k$, market clearing for residential housing imposes

$$q_{r,k} H_{r,k} = L_{r,k} \gamma (w_{r,k} + r + \underline{s} - p\underline{c}) = S_{hr,k} (q_{r,k})^{1+\epsilon_r} = S_{hr,k} (1 + \epsilon_r) \rho_r,$$

where $H_{r,k}$ is the total rural housing and $S_{hr,k}$ the amount of land demanded in the rural area for residential purposes in region k . This leads to the following demand of land for residential purposes in the rural area of region k ,

$$S_{hr,k} = \frac{L_{r,k} \gamma_r (w_{r,k} + r + \underline{s} - p\underline{c})}{\rho_{r,k}},$$

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

The market clearing condition for land from the main text, Equation (20), becomes for each region k ,

$$S_{r,k} = S - \pi \phi_k^2 - \frac{L_{r,k} \gamma_r (w_{r,k} + r + \underline{s} - p\underline{c})}{\rho_{r,k}}. \quad (\text{B.14})$$

Labour Market Clearing. Labour must clear globally,

$$\sum_{k=1}^K L_k = \sum_{k=1}^K (L_{r,k} + L_{u,k}) = L. \quad (\text{B.15})$$

Goods Market Clearing. Rural and urban goods clear globally. By summing demand for urban goods across all locations, the market clearing condition for urban goods is

$$\sum_{k=1}^K (C_{u,k} + \mathbb{T}_k + \mathbb{H}_{u,k}) = \sum_{k=1}^K Y_{u,k}, \quad (\text{B.16})$$

where the terms of the summation in brackets denote, in order:

1. $C_{u,k} = \left(\int_0^{\phi_k} c_{u,k}(\ell_k) D_k(\ell_k) 2\pi \ell_k d\ell_k + c_{u,k}(\ell_k \geq \phi_k) L_{r,k} \right)$ denoting total consumption of urban goods by urban workers (its first term) and rural workers (second term of $C_{u,k}$) of region k ;
2. $\mathbb{T}_k = \int_0^{\phi_k} \tau(\ell_k) D_k(\ell_k) 2\pi \ell_k d\ell_k$ denoting urban good used to pay for commuting costs. Notice that the amount of urban good used for commuting purpose or to produce housing is region-specific.
3. $\mathbb{H}_{u,k} = \left(\int_0^{\phi_k} \frac{\epsilon(\ell_k)}{1+\epsilon(\ell_k)} q(\ell_k) H(\ell_k) 2\pi \ell_k d\ell_k + \frac{\epsilon_r}{1+\epsilon_r} q_{r,k} H_{r,k} \right)$ denotes the total demand of urban goods for urban housing (the first term) and rural housing (the second term) in region k .

The market clearing condition for rural goods is

$$\sum_{k=1}^K C_{r,k} = \sum_{k=1}^K Y_{r,k}, \quad (\text{B.17})$$

where $C_{r,k} = \left(\int_0^{\phi_k} c_{r,k}(\ell_k) D_k(\ell_k) 2\pi \ell_k d\ell_k + c_{r,k}(\ell_k \geq \phi_k) L_{r,k} \right)$ denotes the total consumption of rural goods by urban workers (the first term) and rural workers (the second term) of region k .

Aggregate Land Rents. The aggregate land rent definition is

$$rL = \sum_{k=1}^K \left(\int_0^{\phi_k} \rho(\ell_k) 2\pi \ell_k d\ell_k + \rho_{r,k} \times (S_{r,k} + S_{hr,k}) \right). \quad (\text{B.18})$$

This is equivalent to, using Eq. B.14,

$$rL = \sum_{k=1}^K \left(\int_0^{\phi_k} \rho(\ell_k) 2\pi \ell_k d\ell_k + \rho_{r,k} \times (S - \pi \phi_k^2) \right).$$

B.1.8 Equilibrium Definition

The equilibrium is static for all variables but the real rate of interest. We focus on the static equilibrium, the path for the real interest is pinned down in the following Section B.1.9. An equilibrium with multiple regions is defined as follows,

Definition 1. In an economy with K regions with heterogeneous sectoral productivities $\{\theta_{u,k}, \theta_{r,k}\}$, an equilibrium is, in each region $k \in \{1, \dots, K\}$, a sectoral labor allocation, $(L_{u,k}, L_{r,k})$, a city fringe ϕ_k and rural land used for production $S_{r,k}$, sectoral wages $(w_{u,k}, w_{r,k})$, a rental price of farmland $(\rho_{r,k})$ together with a relative price of rural goods p and land rents (r), such that:

- Factors are paid the marginal productivity in each region $k \in \{1, \dots, K\}$, Eqs. B.1-B.3.
- Workers are indifferent in their location decisions, within and across regions, Eqs. B.11 and B.12 for all $k \in \{1, \dots, K\}$.
- The demand for urban residential land (or the city fringe ϕ_k) satisfies Eq. B.13 in each region $k \in \{1, \dots, K\}$.
- The land market clears in each region $k \in \{1, \dots, K\}$, Eq. B.14.
- The labor market clears globally, Eq. B.15.
- Rural and urban goods markets clear globally, Eqs. B.16 and B.17.
- Land rents satisfy Eq. B.18.

B.1.9 Dynamic Optimization and the Real Interest Rate

The objective of the dynamic model extension is to be able to compute purchasing prices for urban and rural land in each location, which depend on discounted streams of future rents. For this purpose, we assume log utility over instantaneous consumption, which simplifies the consumption-savings problem.

We start by defining lifetime utility as follows:

$$U_t = \sum_{s=t}^{\infty} \beta^{s-t} \bar{u}_s, \quad (\text{B.19})$$

where β is the discount factor in annual terms and \bar{u}_t denotes the expected utility flow at period t . It is important to note that, thanks to the assumption of no moving costs and perfect residential mobility, agents behave like static optimizers, that is, optimal choices are independent of β . All locations yield identical utility but at different consumption baskets, so we must compute a weighted average of region-location specific utilities which constitute overall attainable utility. Therefore we cast utility at the start of period t , \bar{u}_t , as a draw from a lottery over regions as follows,

$$\bar{u}_t = \sum_{k=1}^K \frac{L_k}{L} \left[\frac{1}{L_k} \int_0^{\phi_k} 2\pi \ell_k D_k(\ell_k) \log(C_t(\ell_k)) d\ell_k + \frac{L_{r,k}}{L_k} \log(C_{r,k,t}) \right]. \quad (\text{B.20})$$

Here the intuition is that via full information, every agent is informed about which population shares each region k is going to attain in each period t , hence they weight attainable utility in each region by the respective population share. Inside the square bracket we have expected per-capita utility in urban and rural areas of region k .

Workers can borrow and lend at the risk-free gross interest rate R and the budget constraint of a worker in location ℓ_k satisfies

$$pc_r(\ell_k) + c_u(\ell_k) + q(\ell_k)h(\ell_k) = w(\ell_k) + r + RB - B', \quad (\text{B.21})$$

where B (resp. B') are inherited (resp. next period) bond holdings and $q(\ell_k)$ the rental price per unit of housing in location ℓ_k of region k . Given that all workers are ex-ante identical, there is no borrowing and lending in equilibrium, $B = B' = 0$ and the budget constraint remains the static one defined in Eq. 2.

We then posit a consumption-savings problem, where a representative agent aims to optimize lifetime utility (Eq. B.19) subject to the budget constraint previously defined in Eq. (B.21). Using the expressions for optimal expenditures from the main text, and the fact that in equilibrium $B_t = 0$

$\forall t$, the interest rate is given by the standard Euler Equation

$$R_t = \frac{1}{\beta} \frac{\widehat{u}'_t}{\widetilde{u}'_{t+1}}. \quad (\text{B.22})$$

where β is the ten years discount rate to account for the ten-year period length in the model and, and where marginal utility at the start of period t is defined as

$$\widehat{u}' = \sum_{k=1}^K \frac{L_k}{L} \left[\frac{1}{L_k} \int_0^{\phi_k} \frac{2\pi\ell_k D_k(\ell_k)}{w(\ell_k) + r + \underline{s} - p\underline{c}} d\ell_k + \frac{L_{r,k}}{L_k} \frac{1}{w_{r,k} + r + \underline{s} - p\underline{c}} \right]. \quad (\text{B.23})$$

B.2 Quantitative Evaluation

This section is a detailed description of all required data inputs and their treatment, as well as numerical solution algorithms in order to perform solution and estimation of the model. The section is structured according to this outline:

B.2.1	Multi-Region Numerical Illustration	12
B.2.2	Data Inputs for the Model	17
B.2.2.1	Aggregate Data Inputs	17
B.2.2.2	Cross-Sectional Data Inputs	18
B.2.2.3	Additional Data Inputs	20
B.2.3	Mapping of Model Outputs to the Data Inputs	21
B.2.3.1	Cross-Sectional Model Outputs	22
B.2.3.2	Selection of City Subset	24
B.2.3.3	Aggregate Moment Function	25
B.2.4	Solution and Estimation Algorithm	27
B.2.4.1	Solving a Sequence of Equilibria given parameters	28
B.2.4.2	Optimal Choice of $\{\theta_{ukt}, \theta_{rkt}\}$	29
B.2.4.3	Computation of Prices from Rents	30
B.2.4.4	Starting Values	31
B.2.4.5	Estimation	31
B.2.5	Untargeted Model Outputs	33
B.2.5.1	Urban Area and Density	33
B.2.5.2	Commuting Speed and Agricultural Productivity Gap	35
B.2.5.3	Land Values and Housing Price Indices	36
B.2.5.4	Additional Untargeted Model Cross-Sectional Outputs	38

B.2.1 Multi-Region Numerical Illustration

Before jumping into the detailed estimation of the quantitative model for France, it is useful to plot an artificial economy with multiple regions that qualitatively resembles the data and illustrates the cross-sectional properties of the model. This allows us to evaluate the ability of a simple version of the model to replicate the stylized facts of Section 2. This experiment also sheds light on data moments that can be used to identify the model's parameters.

Parameters. We consider an economy as described in Section 3 and made of $K = 4$ regions, each region is endowed with land and labor, both normalized to 1. At each date t , the productivity of each region k in sector $s \in \{r, u\}$, is

$$\theta_{s,k,t} = \theta_{s,t} \cdot \theta_s^k,$$

with $\theta_{s,t}$ an aggregate productivity shifter, normalized to 1 in the initial period, and θ_s^k a sector-region specific productivity constant through time. Aggregate productivity, $\theta_{s,t}$, is growing at the constant rate of 1.2% per annum in both sectors. Region-specific productivity is either high or low in a given sector s , and normalized to 1 in the other sector—in the high (resp. low) productivity region in sector s , $\theta_s^k = 1.05$ (resp. $\theta_s^k = 0.95$). We set values for other parameters in a reasonable range with respect to the data. We set the land intensity in rural production to 25% ($\alpha = 0.75$) and the elasticity of housing supply ϵ to 4. Preferences towards the different goods are set to roughly match the employment share in agriculture and the housing spending share in the recent period in France— $\nu = 2.5\%$ and $\gamma = 30\%$. With rising productivity, structural change emerges due to the presence of subsistence needs for rural goods, $c = 0.7$. As we focus on income effects driven by subsistence needs, we set \underline{s} to zero and σ to unity. With such preferences, the employment share in the rural sector is about 2/3 at start. The commuting costs parameter a is set to 3 to preserve a small urban area relative to land used in agriculture. Elasticities of commuting costs to urban income and commuting distance, ξ_w and ξ_ℓ , are set to 0.8.

Aggregate implications. Figure B.1 summarizes the aggregate dynamics with rising productivity in both sectors—starting at a period labeled 1840 for illustration. The top panels show the evolution of aggregate employment, spending shares and relative prices. As well known in the literature, due to low initial productivity, the share of workers needed to produce rural goods is high at start to satisfy subsistence needs. The demand for rural goods for subsistence makes them initially relatively expensive and households spend a disproportionate share of income on rural goods. With rising productivity solving the ‘food problem’, workers move away from the rural to the urban sector, the relative price of rural goods falls, as well as the spending share towards rural goods.

The bottom row shows outcomes that are more specific to our theory with endogenous land use. Plots B.1d and B.1e show aggregate urban area (compared to aggregate urban population) and urban density aggregated across cities (average, central and fringe). With structural change, urban area grows faster than urban population, leading to a fall in the average urban density. This is the outcome of two different forces. On the one hand, this is the natural consequence of *rural* productivity growth: higher rural productivity frees up farmland for cities to expand, lowering farmland rents relative to income. Moreover, as workers spend less on rural goods, they can afford larger homes and spend relatively more on housing. The city expands outwards at a fast rate. With land at the city fringe getting cheaper (relative to income), the city expands by adding a less and less dense suburban fringe over time. On the other hand, rising *urban* productivity leads to a reallocation of workers away from the dense center towards the fringe—contributing further to the fall in average urban density. With rising urban income, the share of income devoted to commuting costs falls ($\xi_w < 1$) and workers move towards the suburbs to enjoy larger homes despite a rising opportunity cost of commuting time.³ Thus, although the mechanisms are entirely different, both rural and urban productivity growth contribute to urban sprawl and falling urban

³According to the micro-foundation of commuting costs in Appendix B.1.3, this is so because urban workers optimally choose faster commuting modes when moving towards the suburbs, implying $\xi_w < 1$.

density. Regarding land rents, the reallocation of workers away from agriculture and the fall in the relative price of rural goods exerts downward pressure of the price of farmland. Thus, land rents are reallocated away from the rural part towards the urban part (plot B.1f).

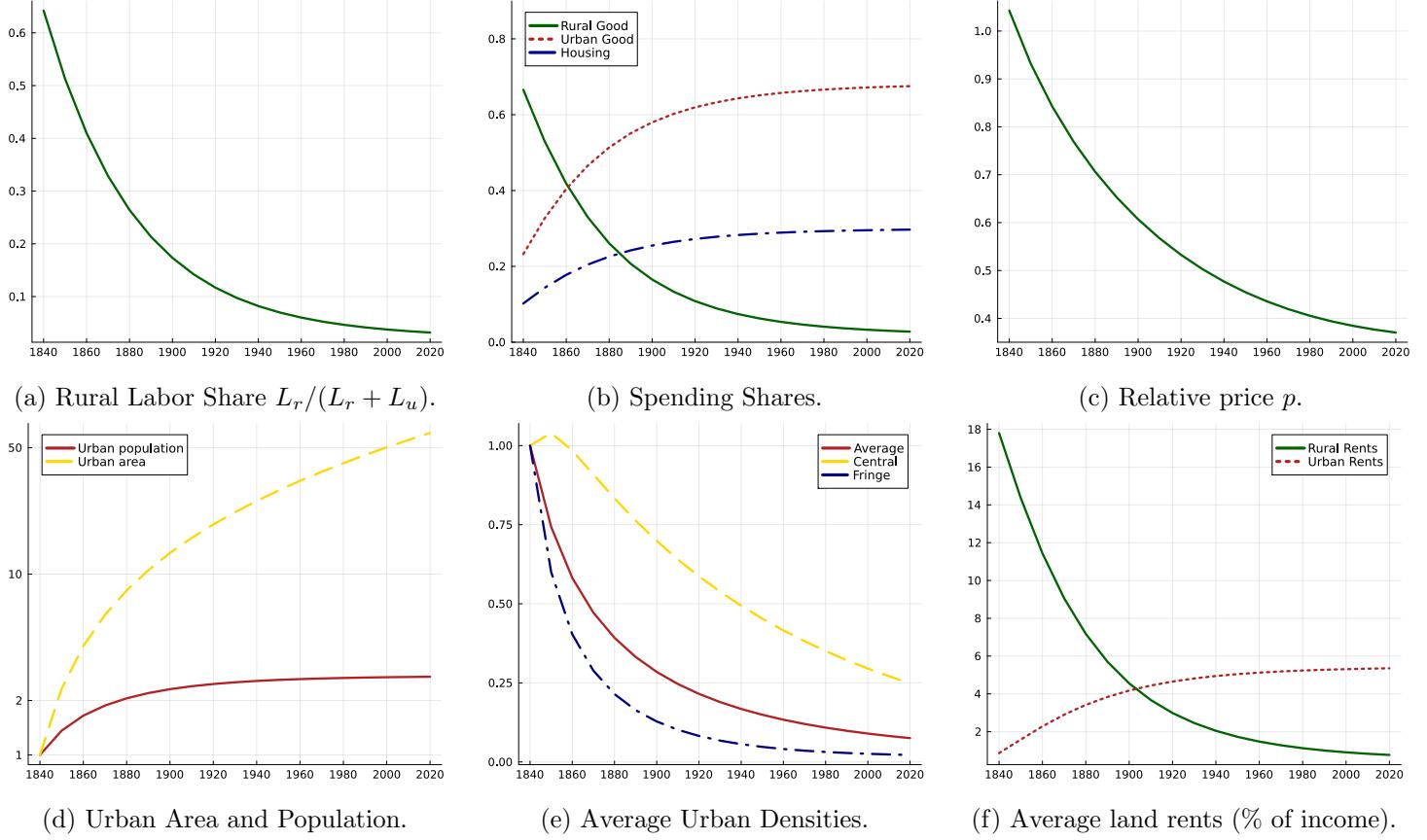


Figure B.1: Aggregate results in an artificial economy with 4 regions.

Notes: Four regions/cities which differ by a constant sectoral productivity shifter θ_r^k or θ_u^k . Aggregate productivity $\theta_{s,t}$ grows at a constant rate of in both sectors. The structural parameters $\{\underline{c}, \underline{s}, \nu, \gamma, \sigma, \alpha\} = \{0.7, 0, 0.025, 0.30, 0.75\}$ are set in a reasonable range to approximately match aggregate moments for France. Commuting parameters, $a = 3$, $\xi_w = \xi_e = 0.8$. The top row illustrates aggregate structural change outcomes, the bottom row shows aggregate implications for city structure and land rents. Aggregate outcomes are summed across the 4 regions/cities.

Cross-sectional implications. The plots in Figure B.2 show the model's implications across regions/cities, whereby the cross-sectional productivity differences trigger dispersion in employment and land use across regions. The city with permanently higher urban productivity attracts more workers and is more populated (Figure B.2a). This mapping between cross-sectional urban productivity differences and urban population is used to discipline the cross-sectional heterogeneity in urban productivity with urban population data—the θ_u^k will be identified to match the distribution of cities population at each date. Despite a larger area, more productive cities are denser in the cross-section—the opposite of the evolution in the time-series (Figure B.2b).

Rural productivity differences across regions triggers variations in rural employment and land values: the high rural productivity region has higher rural employment (Figure B.2c) and higher farmland

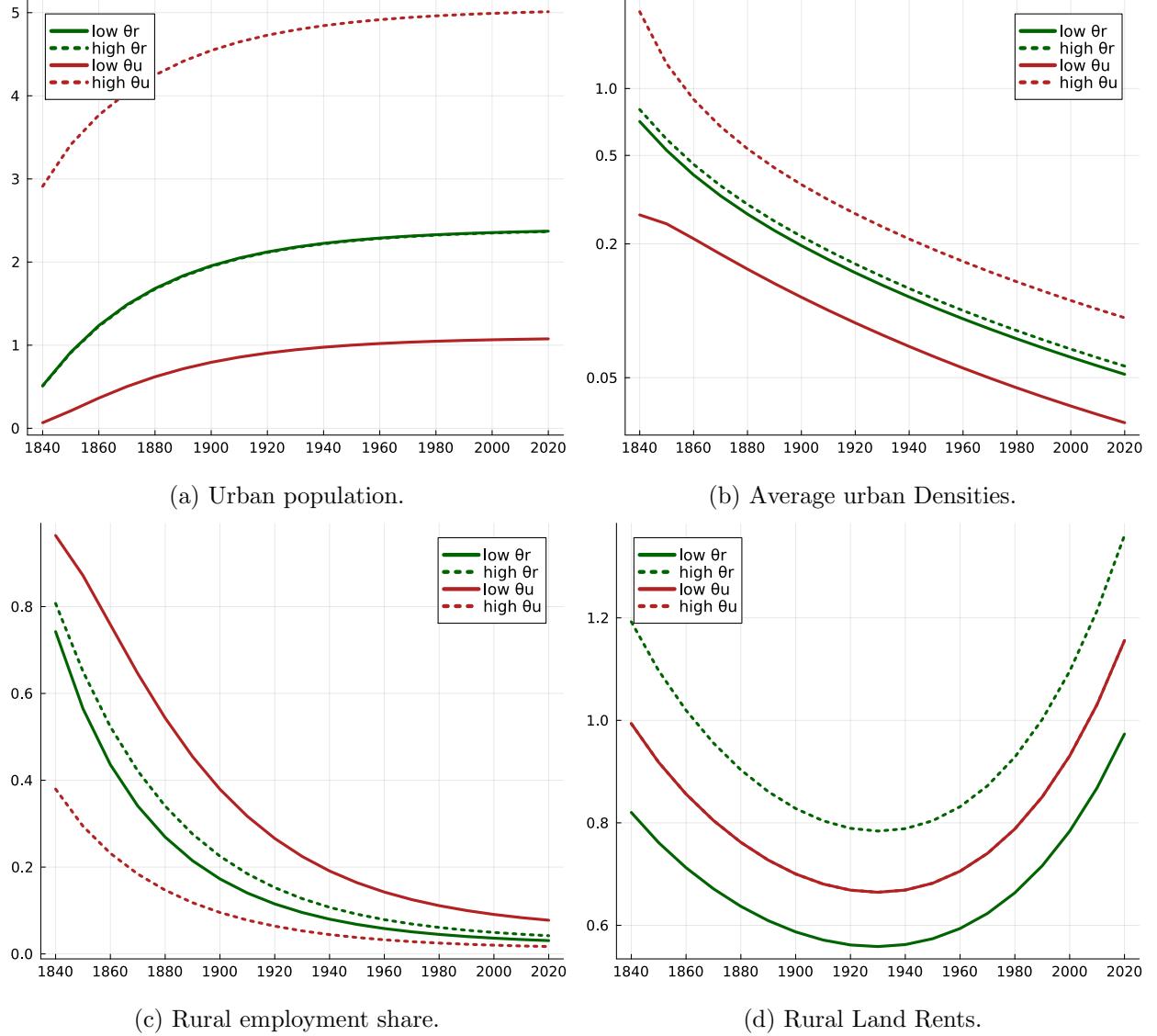


Figure B.2: Cross-sectional results in an artificial model with 4 regions.

Notes: The setup in this experiment is identical to the one in Figure B.1. Each panel illustrates how variation in θ_s^k can be mapped to a data counterpart. Panels (a) and (d) shows how θ_u^k and θ_r^k are identified in the data.

rental price (Figure B.2d). This mapping between cross-sectional rural productivity differences and farmland prices is used to identify the cross-sectional heterogeneity in rural productivity with farmland price data—the θ_r^k will be identified to match the distribution of regional farmland prices at each date. Importantly, the model predicts that cities in high rural productivity regions are denser. With higher farmland prices at the urban fringe, urban land and housing prices are higher, lowering urban area and increasing density. This cross-sectional prediction is the mirror of the mechanism emphasized in the time-series, whereby downward pressure on farmland prices (relative to income) due to aggregate rural productivity growth triggers a fall in urban density. Note that the U-shape evolution of farmland rents over time mirrors the structural change mechanisms at play.

With constant aggregate productivity growth and in the absence of demographic growth, farmland rents evolve due to two conflicting forces: on one hand, structural change puts downward pressure on farmland rents and on the other hand, rising income increases the demand for land. With faster structural change at start, the first channel dominates initially, while the second one dominates when structural change slows down.

To sum up, beyond well-known predictions regarding sectoral employment, the theory is equipped to reproduce the salient facts described in Section 2 of the main text for France regarding the expansion of the urban area, the evolution of urban density and land values. Beyond aggregate implications, the numerical exercise uncovers novel testable cross-sectional predictions linking farmland prices and urban density and sheds light on the mapping between regional sectoral productivity differences and urban population and farmland prices—at the heart of the identification strategy in the quantitative evaluation detailed below.

Sensitivity with a high \underline{s} relative to \underline{c} . In the baseline illustration above, the driver of structural change is rural productivity growth combined with subsistence needs for rural goods—a model where rising productivity frees up resources for the urban sector to expand ('rural labor push'). An alternative view would emphasize a rising demand for (luxury) urban goods as income rises ('urban labor pull'). In our set-up, this would correspond to a high \underline{s} relative to \underline{c} .

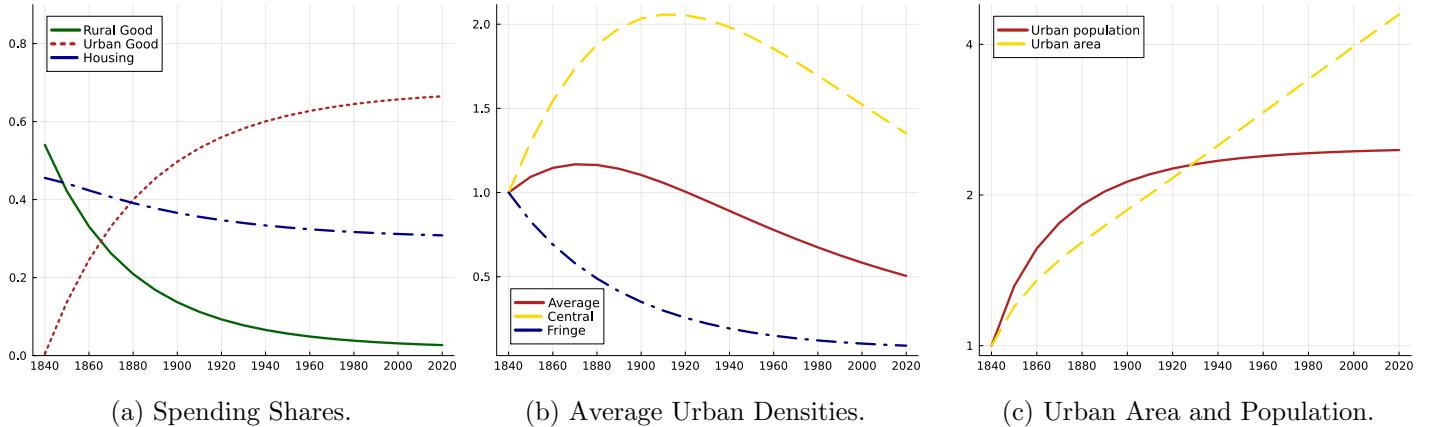


Figure B.3: Sensitivity with a high \underline{s} relative to \underline{c} : aggregate results.

Notes: Parameters are the same as in the previous illustration shown in Figure B.1 with the exception of \underline{s} and \underline{c} : $\underline{s} = 0.9 = 1.5\underline{c}$.

For comparison, we simulate the economy with a value for \underline{s} significantly larger than \underline{c} ($\underline{s} = 1.5\underline{c} = 0.9$), such that, keeping all other parameters to their baseline values, the initial share of employment in the rural sector remains close to 60%. Under such preferences, Figure B.3 shows the model dynamics following rising aggregate productivity in both sectors as in the previous numerical illustration. Cross-sectional differences in productivity are also identical to the previous experiment. While such a calibration can generate employment shares broadly in line with the evidence, it cannot generate the observed fall in urban density. As income increases, the spending share on housing falls as the income elasticity of housing demand is low: workers are willing to reduce their housing

size to consume more of the urban good (Figure B.3a). Thus, the city does not expand much in area to host more numerous urban workers and urban density does not fall—at least in the first decades. Urban density tends to increase due to the reallocation of workers towards the urban center (Figure B.3b): as they shrink their housing size, urban workers relocate towards central locations, increasing central density—the opposite of the data.⁴ A high enough subsistence need is thus important for urban density to decline as it leads to an increase in the housing spending share following structural change. Note also that the evolution of the spending share on housing is informative regarding the relative magnitude of c and s (comparing Figures B.1b and B.3a). An increasing share of housing spending, as in the data points towards a calibration where c is significantly larger than s as discussed in the main text.

We now turn to the estimation of the quantitative model on French data since 1840—starting with the data inputs necessary to estimate the model’s parameters.

B.2.2 Data Inputs for the Model

Solution of the equilibrium requires numerical values for all structural parameters, as well as for sectoral productivities in each region, $\theta_{u,k,t}, \theta_{r,k,t}$ and aggregate population L_t . We describe the data inputs used for the estimation of all the parameters, starting with aggregate variables, sectoral productivity, sectoral employment and population before describing cross-sectional data on urban population and farmland prices. The time sequence for the quantitative model starts in 1840 with steps of 10 years until a final period T far away in the future, $t \in \{1840, 1850, \dots, T\}$. We set $T = 2350$, implying 335 years of future in the simulations.

B.2.2.1 Aggregate Data Inputs

Smoothing of Sectoral Aggregate Productivities. Estimation of sectoral aggregate productivity series $\theta_{u,t}, \theta_{r,t}$ has been described previously in Appendix A.1.4, here we describe an additional smoothing and extrapolation step.

We start with estimated series of aggregate sectoral productivity $\{\theta_{u,t}, \theta_{r,t}\}_{t=1840}^{2020}$, displayed in Figure 6 in the main text. Given their high variability, we smooth this data to remove short-term fluctuations and focus on long-term evolutions. The involved steps are as follows:

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.

⁴Suburban (fringe) density does fall in this experiment (plot (b) of Figure B.3). The same mechanisms as in the baseline illustration play a role: structural change makes farmland cheaper at the city fringe.

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 until $T = 2350$, which is our approximation of $T = \infty$ in the model simulation.

This procedure yields the smoothed series displayed in Figure B.4.

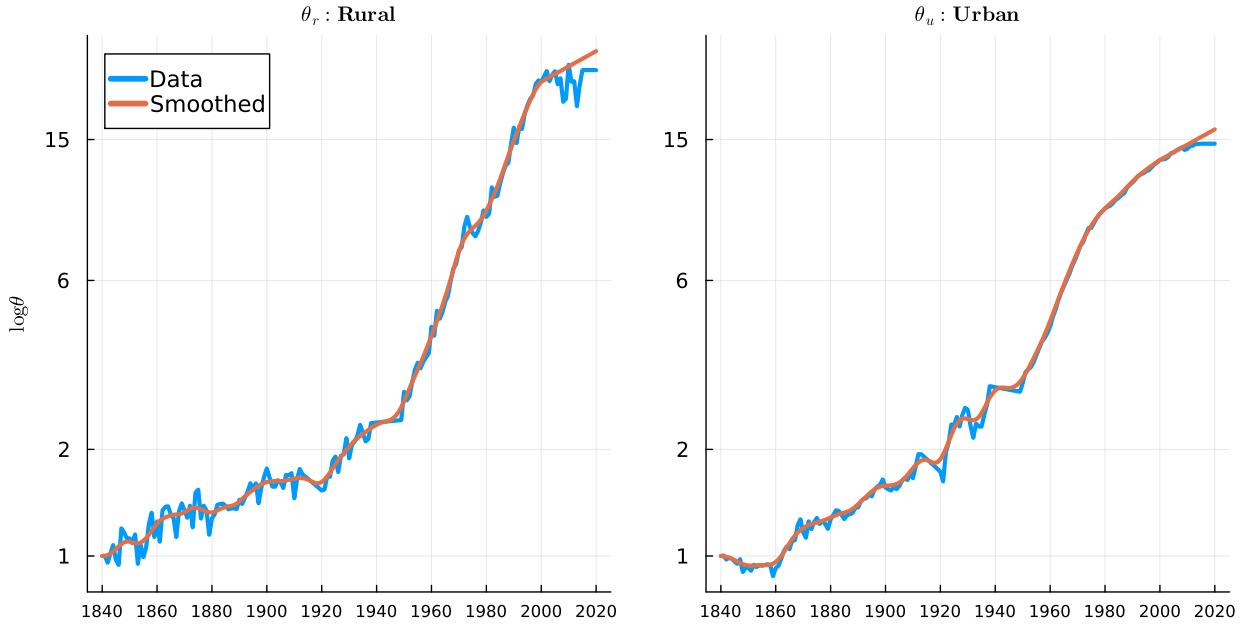


Figure B.4: Smoothing Procedure applied to aggregate sectoral productivity data.

Notes: The left panel shows aggregate rural productivity, the right one shows urban productivity. Both series are normalized to unity in 1840. The red lines show the smoothed series $\{\theta_{u,t}, \theta_{r,t}\}$ used as model inputs. The model inputs are extrapolated from 2000 onwards assuming constant 1% growth. The blue lines are estimated using national accounts data as described in Appendix A.1.4.

Sectoral Employment Share Data. We use data on sectoral employment shares described in Appendix A.1.2 as data inputs that will be targeted in the estimation. More specifically, from 1840 onwards, we use the agricultural employment share shown in Figure A.3. The agricultural employment share is not available at all years and is interpolated between observation dates to provide data inputs at each date $t \in \{1840, 1850, \dots, 2020\}$.

Population Data and Forecasts. The model requires a value for total population L_t in each period. We use official French population counts from the Census for all periods until 2015, and we append the central growth scenario forecast of INSEE for 2050, obtained [here](#). We linearly interpolate 2016–2049 using those data. Then we extrapolate population forward until the year $T = 2350$, assuming a constant growth rate of 0.4% (pre-2050 average growth rate). The resulting series for aggregate population is shown in Figure B.5 for the period 1840–2100, where the data are normalized to 1 in the first period to exhibit the population change over long-period. In the model, the population in 1840 is also normalized, equal to K , the number of regions.

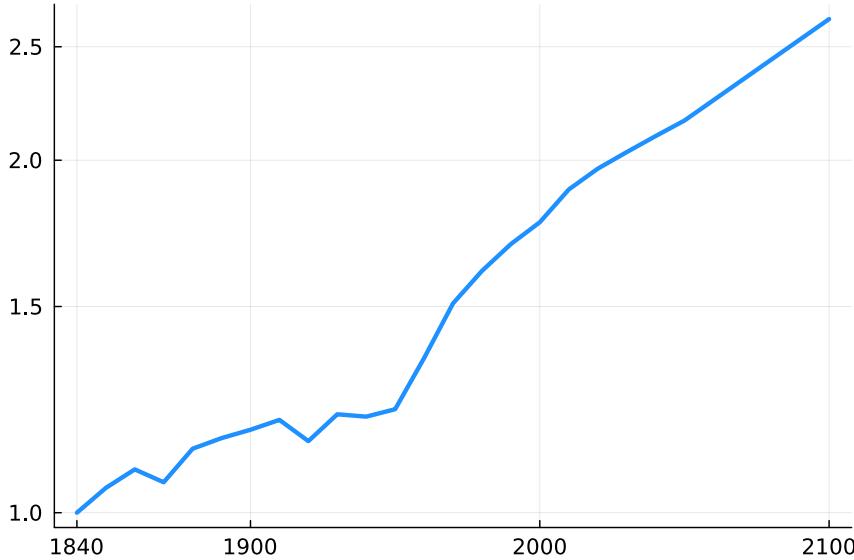


Figure B.5: Population data inputs (1840-2100)

Notes: Population normalized to 1 in 1840. Data until 2015 are from the Census and from INSEE forecasts until 2049. Post-2050, population is assumed to grow at a constant rate of 0.4%.

B.2.2.2 Cross-Sectional Data Inputs

Treatment of City Population Input Data. We describe the population of urban areas in Appendix A.2 for the sample of 100 cities. For the city k in region k , urban area population data, $\text{pop}_{k,t}$, are available at years $\mathcal{T} = \{1870, 1950, 1975, 1990, 2000, 2015\}$ —using Census data in 1876 for 1870.

To estimate the model, and more specifically city-specific urban productivities, $\theta_{u,k,t}$, at each date $t \in \{1840, 1850, \dots, 2020\}$, we need urban area population data (relative a reference city chosen to be Paris) at all dates in each city. For years $t \notin \mathcal{T}$, we perform a linear interpolation on the data to obtain the required value for estimation, using as interpolation nodes the closest two dates. Outside the range 1870–2015, we assume the values are unchanged to the closest observed date.

We are now equipped at all dates $t \in \{1840, 1850, \dots, 2020\}$ with the population of each city k relative to Paris, $\frac{\text{pop}_{k,t}}{\text{pop}_{1,t}}$, where $k = 1$ denotes the Parisian region. The resulting relative urban populations in all cities (but Paris) for the sample of cities used in our quantitative evaluation (described below in Section B.2.3.2) are shown in Figure B.6.

Treatment of Farmland Price Input Data. We describe the local level farmland price data in Appendix A.4. The data inputs for city/region k are the local farmland prices at the ‘département’ level in 1892 and at the PRA level at dates 1950, 1975, 1990, 2000 and 2015. As described in Appendix A.4, a unique farmland price is allocated to each city k of our sample of 100 cities at these dates. We denote the farmland price in region/city k used as input in the model as $\bar{\rho}_{k,t}$.

As for urban area population, we need farmland price data at all dates (relative a region of reference

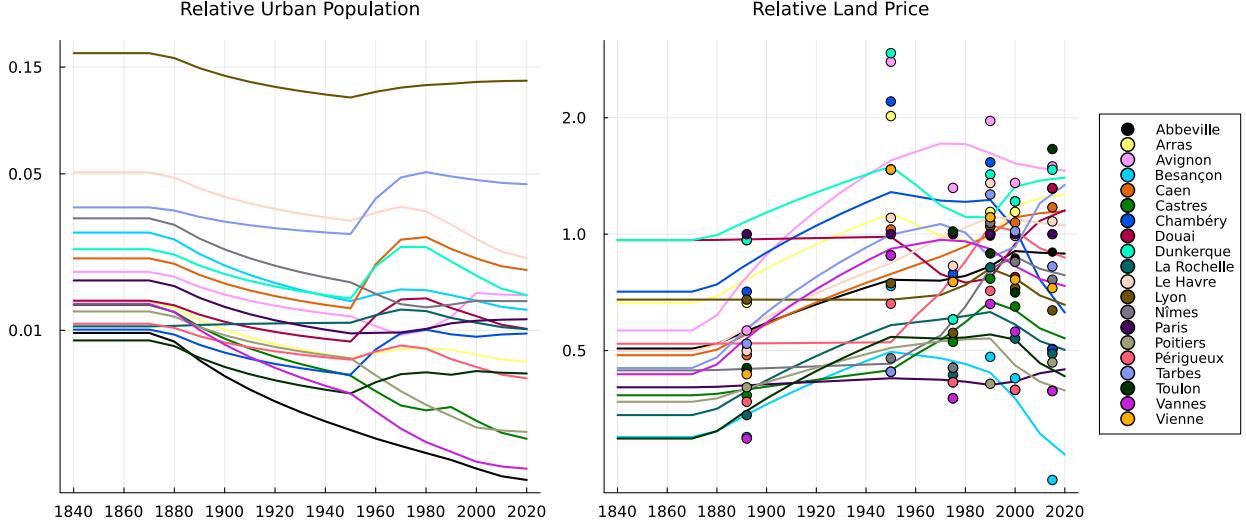


Figure B.6: Regional population and land price data inputs to the model for the sample of 20 cities.

Notes: The left panel shows urban population relative to Paris, $\frac{\text{pop}_{k,t}}{\text{pop}_{1,t}}$, the right panel shows the value of rural land relative to Paris, $\frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}}$. In the right panel, the markers show the raw data values seen in the data. The displayed colored lines are the result of smoothing. In both panels the respective line for Paris would be constant at unity (but is omitted from the graph).

chosen to be the Parisian region, $k = 1$). This is necessary to estimate the model, and more specifically region-specific rural productivities, $\theta_{r,k,t}$, at each date $t \in \{1840, 1850, \dots, 2020\}$.

We apply the following transformations to the raw farmland price data,

1. In each available year, we divide all farmland prices (per ha) by the one of the Parisian area. The main data input is thus a price of farmland relative to the Parisian area. This takes care of scale issues (different price levels or currencies in different periods), and it is consistent with our estimation strategy, targeting the distribution relative to a reference city.
2. We relabel the initial year 1892, when relative farmland prices are first observed, to 1870 corresponding to the first observation of urban population and areas. We assume that data are observed only in years $\mathcal{T} = \{1870, 1950, 1975, 1990, 2000, 2015\}$.
3. As the model requires input data for years $t \notin \mathcal{T}$, we perform linear interpolation to obtain the required value for the relative farmland price, using as interpolation nodes the closest two dates. Outside the range 1870-2015, we extrapolate using the closest observed value.
4. Finally, we smooth the obtained prices as above with a **Hann window** of window size 5. This is mostly to smooth extreme observations of farmland prices in some years of observation for computational purposes (mostly in 1950, where farmland values for few cities are not in line with other years). Doing so, we make sure that the 1870 data input remains identical to the first observation in the data for each city.

We are now equipped with farmland prices relative to Paris at all dates $t \in \{1840, 1850, \dots, 2020\}$

in regions k , $\frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}}$, where $k = 1$ denotes the Parisian region. Data on farmland prices are however missing for Strasbourg in the initial period and Nice in the later periods.⁵ The resulting smoothed relative farmland prices in all cities (but Paris) for the sample of cities used in our quantitative evaluation (see Section B.2.3.2) are shown in Figure B.6.

B.2.2.3 Additional Data Inputs

Land use data. Recent data over the period 2000-2015 from the Ministry of Agriculture (Agreste) provides the land used for agriculture (SAU) as described in Appendix A.1.1 but also estimates of the amount of land that is artificialized ('Sols artificialisés'). In 2010, the SAU is 29096 thousands of ha for 5029 thousands of ha of artificialized land—the amount of artificialized land is 17.3% of land used in agriculture. This value, corresponding to the share of urban land over agricultural land, is targeted in the estimation. Essentially, this will pin down the commuting technology parameter a in the model's estimation—a better commuting technology implying a larger fraction of urban land.

Housing spending share. The aggregate spending share on housing in the data is described in Appendix A.1.5. We obtained values of 0.237 for 1900 (with a 5-year average around 1900) and 0.306 for the year 2010. These targeted values will allow to pin down the housing spending share parameter γ and the degree of non-homotheticity towards the urban good s .

Commuting Data. Individual commuting data described in Appendix A.5.1 are used to provide estimates for the elasticity of commuting speed with respect to commuting distance (which maps to a model value for $1 - \xi_\ell$) and with respect to income (which maps to a model value for $1 - \xi_w$). These two elasticities are calibrated externally following the estimation in Appendix A.5.1: $\xi_\ell = 0.55$ and $\xi_w = 0.75$. As described in the main text, the former is based on the elasticity of commuting speed to commuting distance in the data and the latter is based on the percentage change in commuting speed in a given location over the period 1984-2013—a 11% increase for an increase in aggregate urban productivity of 44% ($\xi_w = 1 - 11/44$).

Note the remaining technology parameters, the land intensity in agriculture $1 - \alpha$, the elasticity of substitution between land and labor, ω , and the location-specific housing supply elasticities, $\epsilon(\ell)$, are calibrated externally based on standard values in the literature. Sensitivity is performed with respect to the elasticity of substitution ω and the housing supply elasticities $\epsilon(\ell)$ in Section B.3 and results are robust for parameter values within the range of the estimates in the literature.

The value of the discount rate β is also calibrated externally to 0.96 on an annual basis. For given parameter values, the equilibrium is independent of β , which only matters to compute equilibrium land/housing values beyond rents. It does impact slightly the estimation of parameters described below by affecting the model's implied (relative) regional farmland values (see Section B.2.3). The effect is however extremely small (targeting relative regional farmland values implied that both the

⁵Strasbourg was not part of France from 1870 to 1918 following the Franco-Prussian war. Data for Nice are missing due to the lack of farmland transactions in the PRA of Nice in the recent period.

numerator and the denominator are discounted). Results are thus not affected for alternative values of β within the range of admissible values.

B.2.3 Mapping of Model Outputs to the Data Inputs

Mapping model outcomes to data inputs used to for estimation involves two main difficulties. First, the model is solved in 10-year steps, while the data are observed at irregularly spaced time intervals. Second, we have two different layers of geographic resolution of moments which we want to capture in the model, regional and country level.

In terms of timing, we start to solve the model in year 1840, the first year we have reliable macro input data series. As described above, data for relative urban area populations and farmland prices have been extrapolated to this starting date and made available at the different dates through interpolation. A similar procedure applies when we want to compare model predictions to data. For table 2 and figure 12 in the main text, which compare model predictions to data in terms of urban density, we linearly interpolate model outcomes in order to get predictions for years 1975 and 2015, which lie off our time grid spaced by 10-year intervals.

With regards of different levels of geographic resolution of moments, at the regional level we fit the distribution of urban populations and farmland prices in order to capture regional heterogeneity, while at the aggregate level we fit a series of moments relating to sectoral employment shares and productivities, population and land use at the country level. The aggregate fitting exercise is standard and is described below in Section B.2.3.3. The mapping between model's outcomes for aggregate variables and aggregate data inputs is also quite straightforward. Therefore, we focus in this section on the mapping between model's outcomes and data inputs in the cross-section, fitting cities population and price distributions.

B.2.3.1 Cross-Sectional Model Outputs

Distribution of city populations. Equipped with data inputs on the relative population of cities at each date (Section B.2.2.2), we impose at each date t the following constraint on the model implied size of urban populations, relative to the reference city $k = 1$ (Paris),

$$\frac{L_{u,k,t}}{L_{u,1,t}} = \frac{\text{pop}_{k,t}}{\text{pop}_{1,t}} \quad (\text{B.24})$$

where $\text{pop}_{k,t}$ is the population count for city k in period t in the data inputs and $L_{u,k,t}$ is the model counterpart. This constraint identifies the distribution of regional urban productivities $\{\theta_{u,k,t}\}$ —more productive cities being relatively more populated.

Distribution of farmland prices. Equipped with data inputs on the relative farmland prices at each date t (Section B.2.2.2), one can similarly aim at fitting these relative prices $\frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}}$ —relative to the reference city $k = 1$ (Paris). One difficulty arises though: data are purchase prices of farmland (per ha) in a given region k and not farmland rents. Thus, one needs to compute the model implied

regional farmland price as the appropriately discounted sum of future farmland rents in a given region. We describe below how this is done but let us assume that one can compute in each period t , a model implied farmland price (per unit of land) in each region k , $\tilde{\rho}_{k,t}$. Then, we impose at each date t the following constraint on the model implied farmland prices relative to the reference city $k = 1$ (Paris)

$$\frac{\tilde{\rho}_{k,t}}{\tilde{\rho}_{1,t}} = \frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}} \quad (\text{B.25})$$

where object $\bar{\rho}_{k,t}$ is the data counterpart to the rural land price in the model described in Section B.2.2.2. Similar to above, this constraint identifies the distribution of regional rural productivities $\{\theta_{r,k,t}\}$ — more productive farmland being relatively more expensive.

Computation of the Model Implied Purchase Price of Rural Land. The model solution delivers a value for land *rents* at each region k and date t , $\rho_{k,t}(\ell)$, with $\rho_{k,t}(\ell) = \rho_{r,k,t}$ in all rural locations. We observe in the data purchase prices of farmland based on transaction data and we need to map the model implied rents to those price data.

For rural land values, a central difficulty is that certain rural locations in the vicinity of current urban land will likely be urban in the future, so their purchase price should reflect this—hence prices differ not only because of current differential rents, but because future rents might change when these locations become urban. Moreover, the price data is not reflecting land values at a given point (e.g. at the fringe of the city), but in a wider region outside the city (e.g. rural). Our aim is therefore to compute a compatible measure of rural land value in the model, providing land values as an average over a range of locations, which in period t , are all rural. Some of those locations will remain rural forever, some will be converted to urban space in the future.

We denote S_k the circular area of region k and $\sqrt{S_k/\pi}$ its radius, where $S_k = S$ is constant across regions in our quantitative evaluation. In practice, to compute the value of rural land in period t , we will consider the average of values of all rural locations at date t , i.e. all locations between two concentric rings of radius $\phi_{k,t}$ and $\sqrt{S_k/\pi}$, respectively. We now define the model implied rural land values (per unit of land) in each region k at all dates t , $\tilde{\rho}_{k,t}$.

We define land *values* from discounted future rents at a given location ℓ . Let $\mathbb{R}_{k,t}(\ell)$ denote the land purchase price in region k in year t in location ℓ . It is defined as the discounted sum of future land rents to be collected at this location, $\mathbb{R}_{k,t}(\ell) = \sum_{s=t}^{\infty} \frac{\rho_s(\ell_m)}{(R_s)^{(s-t)}}$, where the infinite sum is approximated for a T large enough relative to t ,

$$\mathbb{R}_{k,t}(\ell) = \sum_{s=t}^T \frac{\rho_s(\ell_k)}{(R_s)^{(s-t)}}.$$

Integrating across all rural locations, for locations $\ell \in [\phi_{k,t}, \sqrt{S_k/\pi}]$, the corresponding land value

in the rural part of each region k , $\mathbb{W}_{r,k,t}^l$, is defined as,

$$\mathbb{W}_{r,k,t}^l = \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \mathbb{R}_{k,t}(\ell) 2\pi \ell d\ell,$$

Dividing by the rural area, $S_k - \phi_{k,t}^2 \pi$, leads to the definition of the purchase price of rural land per unit of land in region k at date t ,

$$\tilde{\rho}_{k,t} = \frac{\mathbb{W}_{r,k,t}^l}{S_k - \phi_{k,t}^2 \pi} = \frac{1}{S_k - \phi_{k,t}^2 \pi} \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \mathbb{R}_{k,t}(\ell) 2\pi \ell d\ell. \quad (\text{B.26})$$

Remark. One should notice that the rural land rent is homogenous across rural locations of region k in period t , equal to $\rho_{r,k,t}$, such that one can rewrite Eq. B.26 as follows,

$$\begin{aligned} \tilde{\rho}_{k,t} &= \rho_{r,k,t} + \frac{1}{S_k - \phi_{k,t}^2 \pi} \sum_{s=t+1}^T \frac{\int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \rho_s(\ell_k) 2\pi \ell d\ell_k}{(R_s)^{(s-t)}} \\ &= \rho_{r,k,t} + \mathbb{S}_{k,t} \left(\phi_{k,t}, \sqrt{S_k/\pi} \right), \end{aligned} \quad (\text{B.27})$$

where $\mathbb{S}_{k,t} \left(\phi_{k,t}, \sqrt{S_k/\pi} \right)$ represents the summation of discounted average values for future periods until a final date T . Discounting uses the real interest rate R_t obtained from the dynamic model in expression (B.22). Notice that the concerned area in all future periods $s = t+1, \dots, T$ is always starting at *today's* fringe, i.e. at $\phi_{k,t}$. This expression is useful for the numerical solution, because it provides an immediate updating rule in a loop that aims a finding land values. Both objects on the right hand side are computable at any given iteration, as further explained in Section B.2.4.3.

B.2.3.2 Selection of City Subset

The problem is computationally challenging because the system of equations grows fast with number of regions K . We settled for a value of $K = 20$ as a reasonable tradeoff in generating heterogeneity and achieving computational performance which remains feasible. We proceed as follows to create a subset of K cities out of our sample of 98 (we excluded Strasbourg and Nice due to missing farmland price data as explained above in B.2.2.2).

We select Paris as city $k = 1$ by default. The remaining $K - 1$ cities are chosen in a random procedure, which aims at preserving the distribution of urban populations found in the data. Notice that cities with similar populations in the data can have very different surrounding agricultural land values, which is precisely the feature we want to capture.

We proceed by splitting the population distribution of the 97 remaining cities at its median. For the group with population above the median of population, we create 9 bins of population, and from each bin we draw exactly one city. For the group below the median of population, we form

a single bin and draw from it 10 times without replacement. Population sizes are very similar for this group (all are relatively small), hence this procedure ensures better mixing of cities of different cities.

The resulting set of cities for the baseline results are given in Table B.1. To guard against any concerns that the selected subset of cities might in any way be driving some of the obtained results, we choose a different subset by resampling with the above procedure, shown in Table B.2, and we re-estimate the parameters. They are reported in Table B.3. The estimated parameters do differ slightly across samples, as one would expect, given heterogeneity in the data. Estimations using both samples achieve comparably good fits to the targeted moments such that we are not concerned about bias arising from the selection of this city subset.

City	Area	Population	Rural Land Price	Departement
Paris	1397.94	8898707.0	1.00	75
Lyon	298.81	1145494.1	0.77	69
Toulon	196.12	417663.3	0.93	83
Le Havre	83.81	227594.2	1.09	76
Caen	64.62	186321.4	1.10	14
Dunkerque	69.00	156273.3	1.33	59
Avignon	61.31	130705.6	1.52	84
Besançon	38.19	120628.4	0.38	25
Nîmes	46.56	120585.1	0.88	30
Douai	46.62	102944.2	0.94	59
Poitiers	38.50	98203.5	0.42	86
La Rochelle	39.75	96235.7	0.58	17
Chambéry	27.25	83291.7	1.03	73
Arras	21.75	69290.0	1.18	62
Tarbes	23.56	61073.9	1.02	65
Vannes	26.19	58532.3	0.53	56
Castres	12.44	35094.1	0.64	81
Périgueux	9.56	32778.1	0.46	24
Vienne	9.38	23030.1	0.83	38
Abbeville	7.69	21463.7	0.90	80

Table B.1: Baseline subset of $K = 20$ cities. Data are for year 2000.

Notes Rural Land Price is relative to the Parisian rural land price in 2000.

B.2.3.3 Aggregate Moment Function

Aggregate moments. Remember that we have K instances of cities/regions which differ in most outcomes, but we want to map an aggregation of those outcomes to aggregate French data to target some aggregate data moments.

Abstracting from t indices for simplicity, we define total regional consumption expenditures of urban goods ($C_{u,k}$), rural goods ($p \times C_{r,k,t}$) and housing goods ($E_{h,k}$) as well as total consumption

City	Area	Population	Rural Land Price	Departement
Paris	1397.94	8898707.0	1.00	75
Bordeaux	206.00	605708.1	0.74	33
Montpellier	80.69	279285.5	1.13	34
Tours	75.94	229875.1	0.47	37
Mulhouse	74.50	208798.6	1.01	68
Dijon	55.25	205932.1	0.66	21
Brest	58.25	173505.0	0.72	29
Pau	45.50	122734.8	1.03	64
Troyes	49.69	121934.0	1.13	10
Chalon-sur-Saône	28.50	64985.3	0.33	71
Roanne	24.81	61905.0	0.45	42
Béziers	16.44	58099.8	1.13	34
Quimper	24.31	57372.3	0.58	29
Châteauroux	21.88	53116.9	0.79	36
Nevers	20.44	50740.8	0.44	58
Niort	23.75	50371.9	0.39	79
Armentières	12.00	43496.2	1.15	59
Moulins	16.88	33243.4	0.36	3
Rochefort	13.50	27265.6	0.48	17
Morlaix	10.44	17412.3	1.28	29

Table B.2: Alternative subset of $K = 20$ cities. Data are for year 2000.

Notes Rural Land Price is relative to the Parisian rural land price in 2000.

expenditure (E_k) as

$$\begin{aligned}
C_{u,k,t} &= \int_0^{\phi_k} c_{u,k}(\ell) D_k(\ell) 2\pi \ell d\ell + L_{r,k} c_{u,k}(\ell_k \geq \phi_k), \\
p \times C_{r,k} &= p \times \left(\int_0^{\phi_k} c_{r,k}(\ell) D_k(\ell) 2\pi \ell d\ell + L_{r,k} c_{r,k}(\ell_k \geq \phi_k) \right), \\
E_{h,k} &= \int_0^{\phi_k} q_k(\ell) h_k(\ell) D_k(\ell) 2\pi \ell d\ell + q_k(\phi_k) h_k(\ell_k \geq \phi_k) L_{r,k}, \\
E_k &= C_{u,k} + p \times C_{r,k} + E_{h,k},
\end{aligned}$$

We simply add up across regions several key variables to represent an aggregate quantity for the variables

$$v_{k,t} \in \{L_{u,k,t}, L_{r,k,t}, \pi\phi_{k,t}^2, S_{r,k,t}, S_{hr,k,t}, C_{r,k,t}, C_{u,k,t}, E_{h,k,t}, E_{k,t}\}.$$

The relevant aggregation in this case is $\sum_{k=1}^K v_{k,t}$.

We use it to compute the following aggregate moments of the model at each date t :

1. The aggregate rural employment share, at each date t , $\frac{\sum_{k=1}^K L_{r,k,t}}{L_t}$.
2. The share of total urban land over total rural land, at each date t , $\frac{\sum_{k=1}^K \pi\phi_{k,t}^2}{\sum_{k=1}^K (S_k - \pi\phi_{k,t}^2)}$.

Parameter	Description	Baseline	Alternative
S	Total Space	1.0	1.0
L_0	Total Population in 1840	1.0	1.0
θ_0	Initial Productivity in 1840	1.0	1.0
α	Labor Weight in Rural Production	0.75	0.75
σ	Elasticity of Substitution Urban and Rural Good	1.009	0.982
ω	Land-Labor Elasticity of Substitution	1.0	1.0
ν	Preference Weight for Rural Consumption Good	0.022	0.023
γ	Utility Weight of Housing	0.301	0.3
\underline{c}	Rural Consumption Good Subsistence Level	0.678	0.676
\underline{s}	Initial Urban Good Endowment	0.171	0.17
β	Annual Discount Factor	0.96	0.96
ξ_l	Elasticity of commuting cost wrt location	0.55	0.55
ξ_w	Elasticity of commuting cost wrt urban wage	0.75	0.75
a	Commuting Costs Base Parameter	1.688	1.71
ϵ_r	Housing Supply Elasticity in rural area	5.0	5.0
$\epsilon(0)$	Housing Supply Elasticity at city center	2.0	2.0

Table B.3: Comparing optimal estimates (baseline vs. alternative subset of cities).

Notes: Baseline sample of cities listed in Table B.1, the alternative one in Table B.2. Both estimation runs achieve a similar fitness of the loss function (B.33): the baseline (resp. alternative) achieves a value of 0.0135 (resp. 0.0252).

3. The aggregate housing spending share, at each date t , $\frac{\sum_{k=1}^K C_{h,k,t}}{\sum_{k=1}^K E_{k,t}}$.

Aggregate Moment Function. The moment function computes the squared distance between model and data aggregate moments. We target the aggregate moments described in Section B.2.2: the spending share on housing in 1900 and 2010, the aggregate urban area as a fraction of agricultural area in 2010, and aggregate rural employment shares in all dates t from 1840 to 2020. We display the elements of the moment function for aggregate variables in Table B.4.

B.2.4 Solution and Estimation Algorithm

In this subsection we describe numerical solution and estimation of the quantitative model, which can be thought of as having a nested structure:

1. an outermost loop, where we search for a vector $\varsigma = (a, \gamma, \nu, \underline{s}, \underline{c}, \sigma)$ which is a member of set $\Xi \subset \mathbb{R}^6$ in order to optimize a GMM objective function with relevant aggregate data moment. This part is described in B.2.4.5.
2. A nested loop, described in B.2.4.2, which chooses sequences $\{\theta_{ukt}, \theta_{rkt}\}$ in order to optimize an objective function which minimizes the distance between model and data in terms of relative farmland prices and population distributions. Notice that the solution proceeds period by period (see below), hence in practice the choice involves two vectors of length K in each period t , i.e. $\{\theta_{uk}, \theta_{rk}\}_{k=1}^K$. Implied land prices from model need to be built up iteratively,

Moment	Data	Model	Weight
housing_share_2010	0.306	0.3021	10.0
housing_share_1900	0.237	0.2426	10.0
rel_city_area_2010	0.173	0.1726	15.0
rural_emp_1840	0.6019	0.6454	1.0
rural_emp_1850	0.5625	0.5858	1.0
rural_emp_1860	0.5248	0.5133	1.0
rural_emp_1870	0.5018	0.4627	1.0
rural_emp_1880	0.4677	0.4717	1.0
rural_emp_1890	0.4433	0.4341	1.0
rural_emp_1900	0.4172	0.3736	0.01
rural_emp_1910	0.413	0.3655	0.01
rural_emp_1920	0.4149	0.3703	0.01
rural_emp_1930	0.3618	0.2854	0.01
rural_emp_1940	0.3573	0.241	0.01
rural_emp_1950	0.2994	0.2035	0.01
rural_emp_1960	0.2255	0.13	0.01
rural_emp_1970	0.1427	0.0774	0.01
rural_emp_1980	0.0914	0.0622	0.01
rural_emp_1990	0.0615	0.0453	0.01
rural_emp_2000	0.0432	0.0367	0.01
rural_emp_2010	0.0337	0.0352	0.01
rural_emp_2020	0.0313	0.0337	0.01

Table B.4: Components of the moment function at the optimal parameter values. The weights have no econometric interpretation and are chosen as tuning parameters to ensure that because of different scaling, some moments do not vanish in the gradient of the moment function

hence the need for a loop. This part is described in Section B.2.4.3. Notice that this step needs to be performed at each period $t \in \{1840, 1850, \dots, 2020\}$. For future periods, we extrapolate the distributions $\{\theta_{ukt}, \theta_{rkt}\}$ based on the final estimates in 2020 and the aggregate forecasts for $\{\theta_{ut}, \theta_{rt}\}$

3. A final innermost loop, which each time solves the system of equations that constitutes an equilibrium and which is described in B.2.4.1.⁶

We start the description with the lowest level and will work our way upwards.

B.2.4.1 Solving a Sequence of Equilibria given parameters

Given values for ς and $\{\theta_{ukt}, \theta_{rkt}\}$, solution of the model proceeds in standard fashion to find values for endogenous variables such that the system of equations set out in Section B.1.8 is satisfied. Given the Definition (1) of the equilibrium, in given period t , the system is defined as

⁶In practice, steps 2 and 3 are a single step in the implementation.

$$\mathcal{S} = \begin{cases} \text{(B.12)} & \bar{C} - \bar{C}_k, \quad k = 1, \dots, K \\ \text{(B.13)} & L_{u,k} - \int_0^{\phi_k} D_k(\ell) 2\pi d\ell, \quad k = 1, \dots, K \\ \text{(B.14)} & S_{r,k} - \left(S - \pi \phi_k^2 - \frac{L_{r,k} \gamma_r (w_{r,k} + r + s - p\underline{c})}{\rho_{r,k}} \right), \quad k = 1, \dots, K \\ \text{(B.15)} & L - \sum_{k=1}^K (L_{r,k} + L_{u,k}) \\ \text{(B.16)} & \sum_{k=1}^K Y_{u,k} - \sum_{k=1}^K (C_{u,k} + \mathbb{T}_k + \mathbb{H}_{u,k}) \\ \text{(B.18)} & rL - \sum_{k=1}^K \left(\int_0^{\phi_k} \rho(\ell_k) 2\pi \ell d\ell + \rho_{r,k} \times (S_{r,k} + S_{hr,k}) \right) \end{cases}$$

The solution to this system is sought by choosing a vector of values

$$\mathbf{x} = (\{S_{r,k}\}_{k=1}^K, \{L_{r,k}\}_{k=1}^K, \{L_{u,k}\}_{k=1}^K, r, p) \quad (\text{B.28})$$

such that $\mathcal{S}(\mathbf{x}) = \mathbf{0}$. Starting at an initial guess for the first period, which we generate from a single city version of the model, we supply the solution \mathbf{x}_{t-1} as a starting point for period t 's algorithm. A collection of consecutive solutions for periods $t = 1, \dots, T$ is the result of this innermost loop.

B.2.4.2 Optimal Choice of $\{\theta_{ukt}, \theta_{rkt}\}$

Immediately above the step described before in Section B.2.4.1, we want to choose sequences

$$\{\theta_{ukt}, \theta_{rkt}\}_{k=1}^K, t = 1840, 1850, \dots, 2020$$

such that model and data for a set of K cities are close in terms of the distributions of farmland values and urban population sizes. From 2020 onwards we extrapolate both sequences $\{\theta_{ukt}, \theta_{rkt}\}_{k=1}^K, t = 2030, \dots, 2350$, using the extrapolations on aggregate θ_u, θ_r and L_t described above in Section B.2.2.1. In doing so, we keep fixed the distribution of regional components $\theta_{s,2020}^k, s \in \{r, u\}$ – defined in Equation (24) in the main text – going forward.

We formalize the problem as follows in a certain period $t \leq 2020$. Notice that we are nesting the preceding step, i.e. we are choosing optimal \mathbf{x} (see (B.28)) at the same time as we choose $\{\theta_{ukt}, \theta_{rkt}\}_{k=1}^K$. This procedure is a version of MPEC (Mathematical programming with equality

constraints) described in [Su and Judd \(2012\)](#).

$$\min_{\mathbf{x}_t, \{\theta_{u,j,t}\}_{j=1}^K, \{\theta_{r,j,t}\}_{j=1}^K} \sum_{k=1}^K \left(\frac{L_{u,k,t}}{L_{u,1,t}} - \frac{\text{pop}_{k,t}}{\text{pop}_{1,t}} \right)^2 + \Omega_{p,t} \sum_{k=1}^K \left(\frac{\tilde{\rho}_{k,t}}{\tilde{\rho}_{1,t}} - \frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}} \right)^2 \quad (\text{B.29})$$

$$\text{subject to } \sum_{k=1}^K \frac{\text{pop}_{k,t}}{\sum_{j=1}^K \text{pop}_{j,t}} \theta_{u,k,t} = \theta_{u,t}, \quad (\text{B.30})$$

$$\sum_{k=1}^K \frac{L_{r,k,t}}{\sum_{j=1}^K L_{r,j,t}} \theta_{r,k,t} = \theta_{r,t}, \quad (\text{B.31})$$

$$\text{and } (\text{B.12}), (\text{B.13}), (\text{B.14}), (\text{B.15}), (\text{B.16}), (\text{B.18})$$

This is a constrained optimization problem where the objective function (B.29) measures the distance of model-implied price and population distributions to their empirical counterparts. $\Omega_{p,t}$ is a tuning parameter which is allowed to take values less than one in selected periods where convergence in the price finding loop (see B.2.4.3) is particularly challenging – this concerns 2 periods in practice. It is important to notice two aggregation constraints which are added to this problem. Equation (B.30) constrains the distribution of regional urban productivities $\theta_{u,k,t}$ to add up to the estimate aggregate time series of the urban sector, $\theta_{u,t}$. Similarly for the rural sector, where Equation (B.31) imposes the same on rural productivities. In other words, region-specific productivity parameters are constrained to generate a path of sectoral aggregate productivity in line with aggregate data inputs described in Section B.2.2.1.

For periods in the future, i.e. $t > 2020$, we have the series of productivities given, and can drop both the objective function and adding up constraints. The problem collapses to the standard solution of the model system of equations:

$$\begin{aligned} \min_{\mathbf{x}_t} \quad & g(\mathbf{x}_t) = 1, \quad t = 2030, \dots, 2350 \\ \text{subject to} \quad & (\text{B.12}), (\text{B.13}), (\text{B.14}), (\text{B.15}), (\text{B.16}), (\text{B.18}) \end{aligned} \quad (\text{B.32})$$

where $g(\mathbf{x}_t) = 1$ defines a constant function (i.e. nothing to be optimized as objective) – which is of course identical to solving system \mathcal{S} described above for optimal \mathbf{x}_t .

It is worth noting that we use automatic differentiation to compute the gradient to the implied Lagrangian of this problem, which delivers greater accuracy and speed than finite difference-based solution methods (we use the excellent `JuMP.jl` package together with the `Ipopt` solver backend for the julia language to implement this, see [Dunning et al. \(2017\)](#)).

B.2.4.3 Computation of Prices from Rents

The algorithm just described in B.2.4.2 has one shortcoming, in that it does not deliver the required target value $\tilde{\rho}_{k,t}$, but only $\rho_{k,t}$ – i.e. the model delivers rents, not prices. In order to obtain prices,

therefore, we need to iterate on the solution from B.2.4.2, where we start in the objective function with $\rho_{k,t}$ instead of $\tilde{\rho}_{k,t}$. From this sequence of length T , we can compute an implied first set of land prices $\tilde{\rho}_{k,t}^{(1)}$. Then, Equation (B.27) proposes an updating equation, in that it defines $\tilde{\rho}_{k,t}$ as $\rho_{r,k,t} + \mathbb{S}_{k,t}(\phi_{k,t}, \sqrt{S_k/\pi})$. Therefore, we now put $\rho_{r,k,t} + \mathbb{S}_{k,t}(\phi_{k,t}, \sqrt{S_k/\pi})$ into the objective function, and keep iterating until the resulting price vector $\tilde{\rho}_{k,t}^{(s)}$ at iteration s has converged.

B.2.4.4 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 $(\rho_r^{(0)}, L_r^{(0)})$.
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 thus generated intial value. We therefore upgrade the proceedure in the following section.

B.2.4.5 Estimation

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

$$\Xi = \begin{cases} \underline{c} & \in (0.65, 0.71) \\ \underline{s} & \in (0.17, 0.2) \\ \nu & \in (0.02, 0.029) \\ a & \in (1.6, 1.73) \\ \sigma & \in (0.7, 2.0) \\ \gamma & \in (0.28, 0.31) \end{cases}$$

Estimation involves solving the standard GMM optimization problem

$$\min_{\varsigma \in \Xi} L(\varsigma) = \min_{\varsigma \in \Xi} [m - m(\varsigma)]^T W [m - m(\varsigma)] \quad (\text{B.33})$$

where m is an aggregate data moment and $m(\varsigma)$ is its model-generated counterpart, described in Section B.2.3.3. Both sets of values are displayed in Section B.2. We optimize this loss function with

a distance-weighted exponential natural evolution strategy (see Fukushima et al. (2011)).⁷ Notice that we focus here solely in achieving the best fit of the model to our main data moments (leaving aside other considerations related to optimal weighting for inference purposes), hence we set the weights on the diagonal of W in order to ensure that this moment does not vanish in the gradient of the moment function. We evaluate the objective function for 4240 times, which amounts to 106 steps of the `dxnes` algorithm, each step evaluating 40 candidate vectors $\hat{\zeta}$ in parallel on a suitable computing environment (40 CPUs with at least 4G RAM each).

⁷We use method `dxnes`, from package <https://github.com/robertfeldt/BlackBoxOptim.jl>

B.2.5 Untargeted Model Outputs

Beyond the model outputs used to estimate the model and described in Section B.2.3, this section is concerned with describing the necessary steps to generate additional model outputs, some of which are confronted to untargeted data moments. First, we focus on the model aggregate outcomes, which involve some computations and that are shown in the main Figures of the baseline simulation. Second, we provide additional untargeted cross-sectional model outputs together with their data counterparts.

B.2.5.1 Urban Area and Density

Aggregate Urban Area and Population. The urban area of city k at each date t is $\pi\phi_{k,t}^2$. The aggregate urban area of all cities is simply the sum of each urban area, $\sum_{k=1}^K \pi\phi_{k,t}^2$. Its evolution is displayed in Figure 8a of the main text, normalizing to unity the 1870 aggregate urban area for comparison to the data. The corresponding aggregate urban population is, $\sum_{k=1}^K L_{u,k,t}$, at each date t .

Urban Density. We are interested in time series as well as spatial implications of urban density at a given date t . The average density of a city k at date t , $\text{density}_{k,t}$, is defined as,

$$\text{density}_{k,t} = \frac{L_{u,k,t}}{\pi\phi_{k,t}^2} = \int_0^{\phi_{k,t}} D_{k,t}(\ell) 2\pi\ell d\ell / \pi\phi_{k,t}^2.$$

Average urban density across regions/cities used for Figures 8b and Figure 9a, both in the main text, is defined with urban population weights,

$$\bar{D}_t = \sum_k \left(\frac{L_{u,k,t}}{\sum_j L_{u,j,t}} \right) \cdot \text{density}_{k,t}.$$

Note that in Figure 3a we normalize the 1870 value to unity for comparison to data, plotting $\frac{\bar{D}_t}{\bar{D}_{1870}}$.

One can look at the overall fall predicted by the model, computing the %-change in density since 1870, $\frac{\bar{D}_{2020} - \bar{D}_{1870}}{\bar{D}_{1870}}$ and compare it to its data counterpart over the period 1870-2015—variable *avg_density_change* displayed in Table B.5.

We proceeded in a similar fashion to compute central density and fringe density displayed in Figure 9a. We compute the central density in city k as,

$$\text{central density}_{k,t} = \int_0^{\phi_{k,c}} D_{k,t}(\ell) 2\pi\ell d\ell / \pi\phi_{k,c}^2.$$

where the radius of the central part of city k , $\phi_{k,c}$, is kept constant and equal to $15\% \cdot \phi_{k,1840}$. Average central urban density across regions/cities used for Figure 9a is defined with urban population

weights,

$$\bar{D}_t^{\text{central}} = \sum_k \left(\frac{L_{u,k,t}}{\sum_j L_{u,j,t}} \right) \cdot \text{central density}_{k,t}.$$

The fringe density is simply equal to the local density at the fringe of city k at each date t , $D_{k,t}(\phi_{k,t})$ and the average fringe urban density across regions/cities used for Figure 9a is defined with urban population weights,

$$\bar{D}_t^{\text{fringe}} = \sum_k \left(\frac{L_{u,k,t}}{\sum_j L_{u,j,t}} \right) \cdot D_{k,t}(\phi_{k,t}).$$

Note that in Figure 9a, central, average and fringe densities are normalized to unity in the 1840 initial period to focus on their respective evolutions.

Density Gradients. At a given point in time, we want to know how fast and in which way urban density falls as one moves away from the center. To this end, we estimate an exponential decay model of urban density over distance in a given year in both model (for 2020) and data (for 2015) for each city. In the data, the urban population-weighted average of density decay coefficients was estimated between 0.14 and 0.18, with 0.15 as baseline estimate (see Appendix A.2.4).

The model counterpart is obtained as follows. First, we convert the distance ℓ_k for each city in the model into kms. For this, we need a data counterpart to the radius of cities. We compute the radius of the average city in 2020 as the urban population weighted-sum of the cities radius,

$$\bar{\phi}_{2020} = \sum_k \left(\frac{L_{u,k,t}}{\sum_j L_{u,j,t}} \right) \cdot \phi_{k,2020}.$$

Using data described in Appendix A.2.4, one can compute the counterpart as the urban population-weighted mean of the largest distance bin in each of the 100 cities. This gives a value of $\bar{\phi}_{2020}^d = 21.43$ kms. A location ℓ_k in city is thus assumed to be at distance $\tilde{\ell}_k$ kms from the center of city k , where,

$$\tilde{\ell}_k = \left(\frac{\bar{\phi}_{2020}^d}{\bar{\phi}_{2020}} \right) \cdot \ell_k \text{ (in kms)}$$

Second, we run for each city k an exponential decay model by dividing each city k in date $t = 2020$ into 20 intervals of same length, $\phi_{k,2020}/20$ (equal to $\frac{\bar{\phi}_{2020}^d}{\bar{\phi}_{2020}} \cdot \phi_{k,2020}/20$ kms). Denoting $\tilde{\ell}_{k,n}$ the distance between the midpoint of each interval $n \in \{1, 2, \dots, 20\}$ and the city center in city k , we compute the corresponding model implied density in each interval, $D_{k,n}$, and estimate the following equation for each city k (similar to Eq. B.34 in Appendix A.2.4),

$$D_{k,n} \approx a_k \exp(-b_k \cdot \tilde{\ell}_{k,n}), \quad (\text{B.34})$$

This provides decay coefficients b_k for each city k at date $t = 2020$. As for the data, we compute the urban population weighted average of decay coefficients, $\sum_k \left(\frac{L_{u,k,t}}{\sum_j L_{u,j,t}} \right) \cdot b_k$. This gives a value of 0.18 as shown in Table B.5 together with the baseline data counterpart. The obtained value is

in the ballpark of the data although slightly higher.

Moment	Data	Model
density_decay_MSE	—	0.631
density_decay_coef	0.15	0.184
avg_density_change	-0.883	-0.825
max_mode_increase	4.524	4.776

Table B.5: Non-targeted aggregate moments at the optimal parameter values

B.2.5.2 Commuting Speed and Agricultural Productivity Gap

Commuting Speed. We derived optimal mode or speed choice $m_k(w_u, \ell)$ as a function of urban wage and location of residence in Equation (B.7) above. We compute for each period the urban population-weighted average speed in each city k ,

$$\bar{m}_{k,t} = \frac{1}{L_{u,k,t}} \int_0^{\phi_{k,t}} m_k(w_{u,k,t}, \ell) D_{k,t}(\ell) 2\pi \ell d\ell$$

The national average commuting speed, population-weighted average across all cities, is defined as,

$$\bar{m}_t = \sum_k \left(\frac{L_{u,k,t}}{\sum_j L_{u,j,t}} \right) \bar{m}_{k,t},$$

and plotted in main text Figure 10a together with the data counterpart for the Parisian urban area. The overall change in average mode/speed in the model, $\frac{\bar{m}_{2020}}{\bar{m}_{1840}}$, is displayed in Table B.5 (variable *max_mode_increase*) together with the data counterpart for Paris. Note that the overall increase in the model for the city of Paris is also similar to the data: on one side, Parisian have faster modes at a given distance due to a higher opportunity cost of time (higher wages), but the fraction of population at short distance and lower speed is also higher as the city is denser due to higher housing costs. Both effects seem to roughly cancel out in the model such that the evolution of speed in Paris in the model mimics the aggregate one.

Agricultural Productivity Gap. For the agricultural productivity gap (APG), we define the APG in region k at date t , as a monotonic transformation of the urban-rural wage gap in each region k (as in Gollin et al. (2014)),

$$\text{Raw-APG}_{k,t} = \alpha \frac{w_{u,k,t}}{w_{r,k,t}} = \left(\frac{L_{r,k,t}/L_{u,k,t}}{VA_{r,k,t}/VA_{u,k,t}} \right),$$

where $L_{s,k,t}$ and $VA_{s,k,t}$ denotes the employment and value added in sector s of region k at date t . In line with the definition in the main text, the national average of the APG, weighting by regional

population, is

$$\text{Raw-APG}_t = \sum_{k=1}^K \left(\frac{L_{k,t}}{L_t} \right) \cdot \text{Raw-APG}_{k,t},$$

where $L_{k,t} = L_{u,t} + L_{r,t}$ is the population of region k at date t . The model implied Raw-APG $_t$ is plotted in Figure 10b.

B.2.5.3 Land Values and Housing Price Indices

Define land *values* from discounted future rents at a given location ℓ . As before, let $\mathbb{R}_{k,t}(\ell)$ denote the land purchase price in region k in year t in location ℓ , defined as the discounted sum of future land rents to be collected at this location until final period T ,

$$\mathbb{R}_{k,t}(\ell_k) = \sum_{s=t}^T \frac{\rho(\ell_k)}{(R_s)^{(s-t)}}.$$

Value of Urban and Rural Land. As an accounting identity at a given time t , we want to compute the total current value of urban and rural land. Proceeding in a similar fashion as above, we define first the discounted sum of future urban land rents in city k , $\mathbb{W}_{u,k,t}^l$, as follows

$$\mathbb{W}_{u,k,t}^l = \int_0^{\phi_{k,t}} \mathbb{R}_{k,t}(\ell) 2\pi \ell d\ell,$$

and the corresponding land value in the rural part of each region k as $\mathbb{W}_{r,k,t}^l$:

$$\mathbb{W}_{r,k,t}^l = \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \mathbb{R}_{k,t}(\ell) 2\pi \ell d\ell,$$

The total value of land in period t in region k is thus

$$\mathbb{W}_{k,t}^l = \mathbb{W}_{u,k,t}^l + \mathbb{W}_{r,k,t}^l$$

Figure 11a plots for each date t , the model implied aggregate share of land value in the rural area, $(\sum_{k=1}^K \mathbb{W}_{r,k,t}^l) / (\sum_{k=1}^K \mathbb{W}_{k,t}^l)$, and the model implied aggregate share of land value in the urban area, $(\sum_{k=1}^K \mathbb{W}_{u,k,t}^l) / (\sum_{k=1}^K \mathbb{W}_{k,t}^l)$. This is plotted against data from [Piketty and Zucman \(2014\)](#), where the share of land value in the rural area is the share of land value in agriculture and the share of urban land value is obtained from aggregate French housing wealth, assuming a constant land share of 0.32 (average over the period 1979-2019 in the data).

Value of Urban and Rural Housing. To compute housing price indices at city level (or in other locations, like the center of a city), we also need to know the value of housing in a given location. This value takes the form of *quantity times price*, where the purchase price is similarly to above the discounted future *housing* rent q , and the quantity is given by the housing supply function H . We

focus here on the task of computing housing values for an entire region k .

We define first the purchasing price of a housing unit in location ℓ of city k at each date t , $\mathbb{Q}_{k,t}(\ell)$, as the discounted sum of future rental prices until a final period T large enough relative to t (the infinite sum being truncated at T),

$$\mathbb{Q}_{k,t}(\ell) = \sum_{s=t}^T \frac{q_{k,s}(\ell)}{(R_s)^{(s-t)}}.$$

We can compute the total value of housing in the urban part of region k at date t as

$$\mathbb{W}_{u,k,t}^h = \int_0^{\phi_{k,t}} H_{k,t}(\ell) \mathbb{Q}_{k,t}(\ell) 2\pi \ell d\ell, \quad (\text{B.35})$$

and, similarly, in the rural part,

$$\mathbb{W}_{r,k,t}^h = \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \frac{S_{hr,k}}{S_{hr,k} + S_{r,k}} H_{k,t}(\ell) \mathbb{Q}_{k,t}(\ell) 2\pi \ell d\ell, \quad (\text{B.36})$$

where the ratio in this expression adjusts for the fact that only a fraction of land in the rural part is used for housing (the rest being used for rural production). The total value of housing (in terms of the numeraire urban good) is thus,

$$\mathbb{W}_{k,t}^h = \mathbb{W}_{u,k,t}^h + \mathbb{W}_{r,k,t}^h.$$

The total number units of housing, $\mathcal{H}_{k,t}$, is equal to the housing units in the city plus the housing units outside the city, which is computed as

$$\mathcal{H}_{k,t} = \int_0^{\phi_{k,t}} H_{k,t}(\ell) 2\pi \ell d\ell + \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \frac{S_{hr,k}}{S_{hr,k} + S_{r,k}} H_{k,t}(\ell) 2\pi \ell d\ell, \quad (\text{B.37})$$

The Housing Price Index in terms of the numeraire (urban good) for region k is computed as the total housing value per housing units,

$$HPI_{k,t} = \frac{\mathbb{W}_{k,t}^h}{\mathcal{H}_{k,t}}$$

In Figure 11b, we take into account that the GDP-deflator evolves over time due to the sectoral reallocation and changes in the relative price p and compute a real housing price index in each region k , $RHPI_{k,t}$, defined as

$$RHPI_{k,t} = \frac{HPI_{k,t}}{\tilde{P}_t} = \frac{\mathbb{W}_{k,t}^h}{\mathcal{H}_{k,t}} \frac{1}{\tilde{P}_t}, \quad (\text{B.38})$$

where \tilde{P}_t is a model implied GDP-deflator that takes the geometric average of the Laspeyres and the Paasche price index, defined as follows:

$$\begin{aligned}\mathbf{P}_{0,t} &= \frac{p_t Y_{r,t-1} + Y_{u,t-1}}{p_{t-1} Y_{r,t-1} + Y_{u,t-1}} \\ \mathbf{P}_{1,t} &= \frac{p_t Y_{r,t} + Y_{u,t}}{p_{t-1} Y_{r,t} + Y_{u,t}} \\ \Delta \mathbf{P}_t &= \sqrt{\mathbf{P}_{0,t} \mathbf{P}_{1,t}} \\ \tilde{P}_t &= \tilde{P}_{t-1} \Delta \mathbf{P}_t, t > 1840 \\ \tilde{P}_{1840} &= p_{1840}\end{aligned}\tag{B.39}$$

The national real housing price index is computed as a population-weighted average of real price indices across regions,

$$RHPIt = \sum_{k=1}^K \left(\frac{L_{k,t}}{L_t} \right) \cdot RHPIk,t,$$

and is displayed in Figure 11b, normalizing to 100 the index in 1840.

To compute a real house price index for a different set of locations, e.g. the center of a city, we proceed in the same fashion, adjusting the upper integration limits in expressions (B.35),(B.36), and (B.37) appropriately.

B.2.5.4 Additional Untargeted Model Cross-Sectional Outputs

This Section provides additional non-targeted cross-sectional outputs of the model together with their data counterparts, not included in the main text for sake of space.

Urban Area and Density. In the cross-section at each date, more populated cities (i.e. more productive) are larger in area and, due to higher housing prices, more populated cities are also denser. While the model can reproduce these facts qualitatively, it does not match them quantitatively as discussed in the main text and displayed in Figure B.7 in the first and last dates of observation for the restricted sample of 20 cities. The urban area in the model does not increase enough with city size compared to the data—equivalently, large cities are (relatively) too dense in the model compared to the data. Overall, the cross-sectional fit for urban density is not very good, as cross-regional heterogeneity is fairly limited in the model, restricted to heterogeneous sectoral productivities to generate a reasonable dispersion in urban population and farmland prices at the urban fringe. Many other city-specific factors possibly influence the density of individual cities in the cross-section (different natural constraints, different housing supply conditions/commuting infrastructure, different amenities, different protected areas/land use regulations in the more recent period,...) that our model cannot possibly account for.

To visualize the cross-section in the time dimension, one can bin cities into size-groups and take averages within bins—mitigating the idiosyncrasies of individual cities as well as concerns regarding

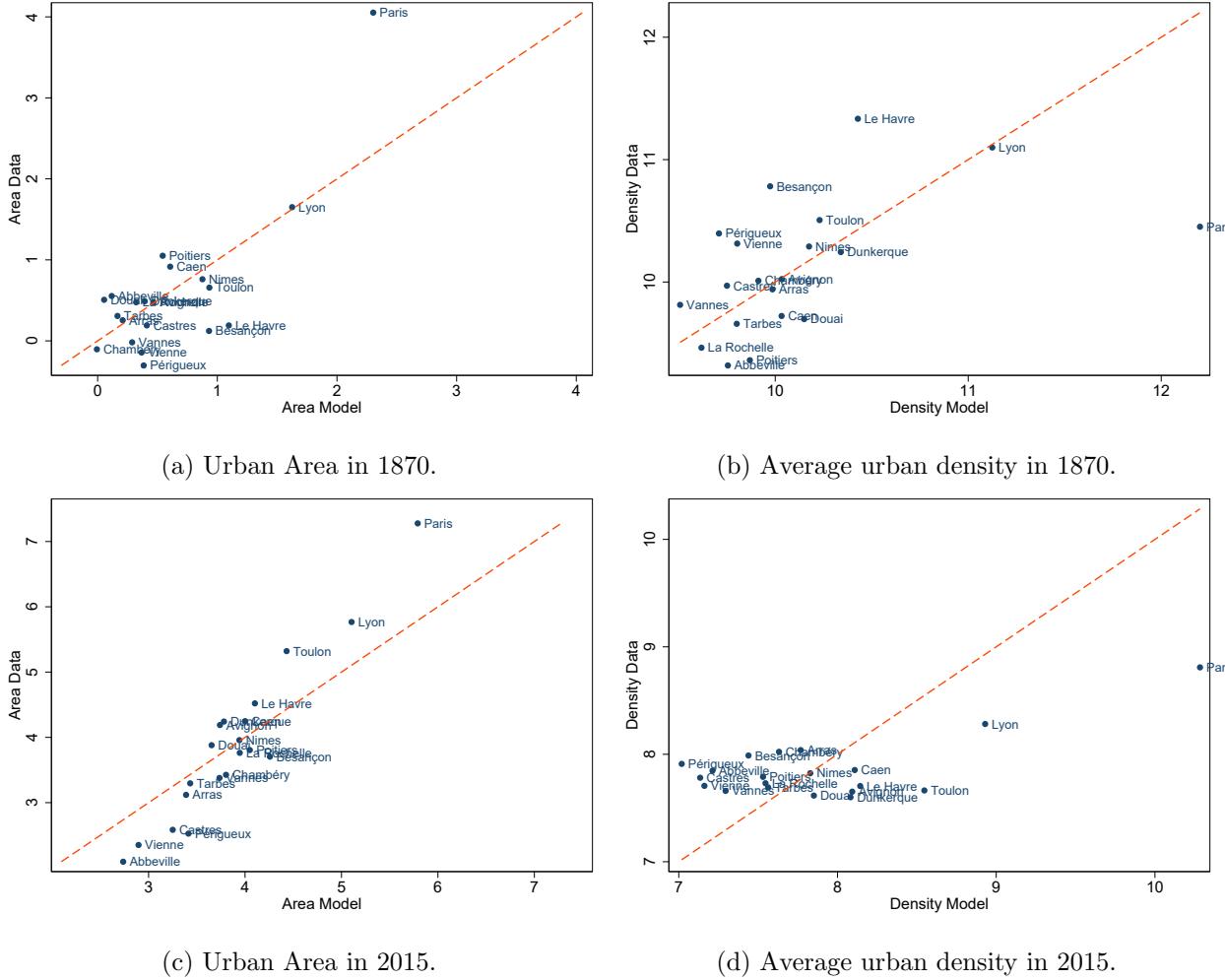


Figure B.7: Cross-Sectional Predictions: Area and Urban Density in 1870 and 2015.

Notes: This plots the cross-sectional predictions for urban area and urban density (in log) of the 20 individual cities against their data counterpart. Model counterparts in log are normalized such that the cross-sectional mean matches the data. For 2015, we interpolate the model's outcomes between 2010 and 2020.

their density measurement (Appendix A.2.5). With a restricted number of bins relative to the number of cities, outcomes in model and data become readable in the time-dimension. This is done in Figure B.8 in the model and in the data (for the whole initial sample of 100 cities to avoid idiosyncrasies in the random sample of 20 cities). Density is averaged by bins of size in the initial period, 1870 (above 100,000; between 50,000 and 100,000, between 25,000 and 50,000; below 25,000) and normalized by the first period (1870) median density in the sample to visualize the cross-sectional and time series variations. While the model performs well relative to the data in the time-series, qualitatively and quantitatively, the cross-sectional dispersion of densities is too large in the model at each date. Larger cities are denser, in the model and in the data, but an order of magnitude denser in the model relative to smaller ones—the problem being particularly severe for Paris. This said, this is to us not a major concern given that our main focus is the evolution in the

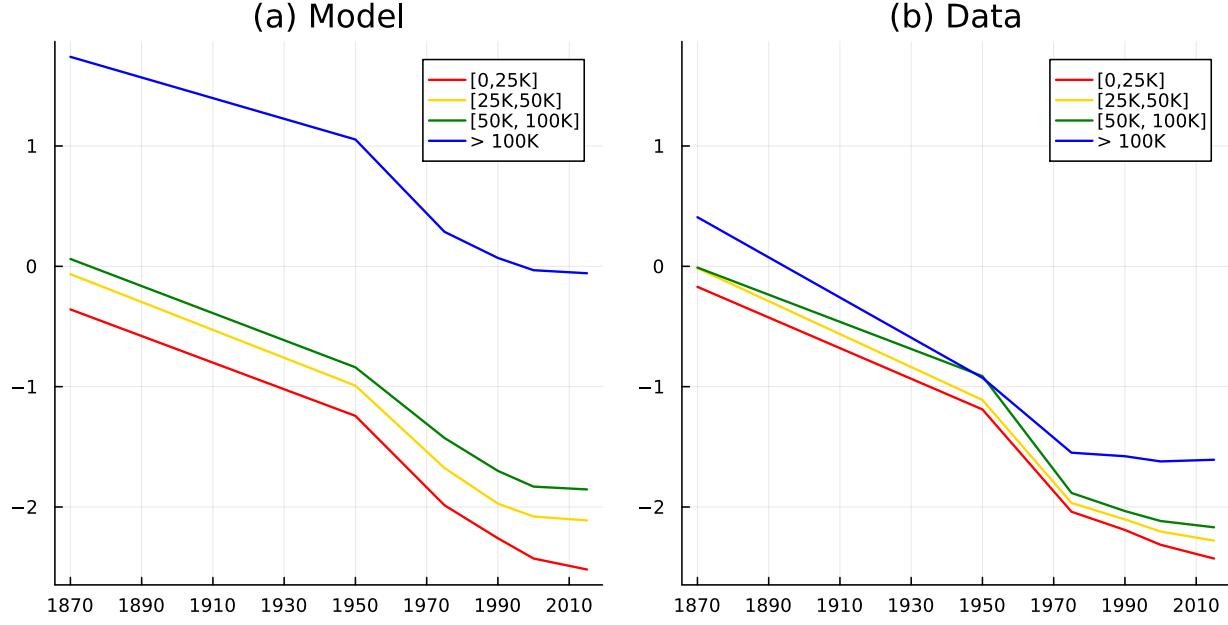


Figure B.8: Average urban density by size-bins: model versus data.

Notes: This plots show the evolution of average urban density in log by bins of city-size by population in 1870 in the model (sample of 20 cities) and in the data (sample of 100 cities): 4 bins by 1870 population, above 100,000; between 50,000 and 100,000; between 25,000 and 50,000; below 25,000 (in the model: top 3 cities; cities ranked 4 to 6; 7 to 14; bottom 6). The unit is normalized in the data and in the model by the median urban density in 1870.

time-series, which barely interacts with cross-sectional heterogeneity.

Rural Productivity and Wheat Yields. The cross-sectional dispersion in farmland prices is driven by rural productivity differences in the model. As a sanity check, we compare the model estimated cross-sectional differences in rural productivity to observed measures of rural productivity: wheat yields. In our restricted sample of 20 regions, due to specialization in crops, not all cities are in regions producing wheat (only 7 cities are in départements which devote more than 20% of their land use for wheat in 2000, 11 above 10% with the mean wheat land use in this sample below 15%). This said, we investigate the link between the estimated region-specific rural productivity and wheat yields on the sample of simulated cities in regions producing wheat (with a low (resp. higher) threshold of land use for wheat of at least 10% (resp. 20%) of wheat land use, 11 cities (resp. 7 cities)). To do so, we perform the following regression for readily available dates $t \in \{1975, 1990, 2000, 2015\}$ (see Appendix A.3.3 for data description on wheat yields and land use for wheat),

$$\log \text{Yield}_{k,t} = a_t + b \cdot \log \theta_{r,k,t} + u_{k,t}, \quad (\text{B.40})$$

where $\text{Yield}_{k,t}$ is the observed wheat yield at date $t \in \{1975, 1990, 2000, 2015\}$ in the département of region/city k , $\theta_{r,k,t}$ is the rural productivity of region/city k at date t in the model and a_t a time-effect which controls for aggregate productivity changes common across regions. The sample is restricted to regions/cities in département for which land use for wheat is above 10% (resp. 20%)

of agricultural land use in 2000. Standard errors are clustered at the city/region k level.

Results of regression B.40 on the restricted sample give an estimated b is very close to unity and highly significant in both sub-samples, particularly so when focusing on locations more specialized in wheat ($b = 1.05$ with T-stat of 4.0 on the larger sample of 11 cities = 44 obs., corresponding to a low threshold of 10% of land use for wheat; $b = 0.96$ with T-stat of 8.4 on the restricted sample of 7 cities = 28 obs., corresponding to a higher threshold of 20% of land use for wheat). Alternatively, one can run regression B.40 on the whole sample of cities in the model but weighting observations by their share of land use for wheat in 2000: results are the same ($b = 1.03$ with T-stat of 5.1). This is illustrated by the scatter plot of Figure B.9 for the sample cities of wheat producers in 2000 (land use for wheat above 10%).

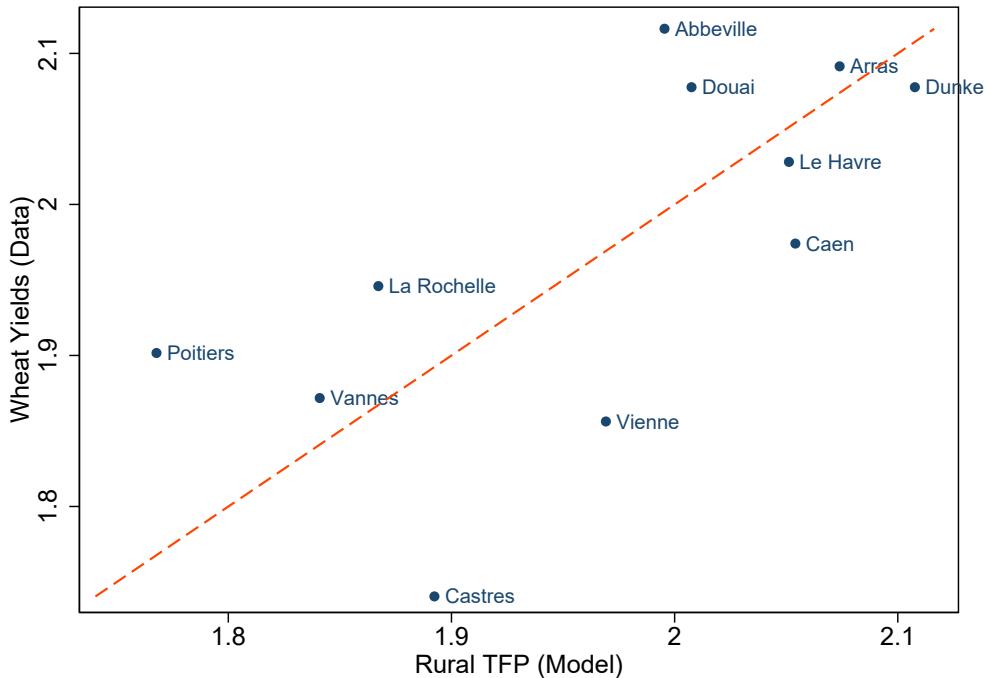


Figure B.9: Wheat yields in 2000 and Rural Regional Productivity in the model, $\theta_{r,k,2000}$.

Notes: For regions/cities producing wheat (11 cities with land use wheat share above 10%), this scatter plot shows wheat yields in city/region k against the rural regional productivity in the model, $\theta_{r,k,t}$. Model counterparts in log are normalized such that the cross-sectional mean matches the data. *Data Source:* See Appendix A.3.3.

This strongly suggests that our regional model estimates of rural productivity reflect effective land productivity in the data, at least for locations producing wheat. Unfortunately, data limitations to estimate local productivity in agriculture independently of the crop specialization prevent us to do further sanity checks on the whole sample of cities.

Urban Productivity and Wages. Similarly to our attempt to compare the model implied rural productivity to wheat yields, one can compare model implied urban productivity/wage, $\theta_{u,k,t}$, to urban wages in the data for the sample of 20 cities in our simulations. This is a useful sanity check to see if the dispersion of wages across cities implied by the model is broadly in line with the data. Data

are readily available from DADS for years $t \in \{1975, 1990, 2000, 2015\}$ (see Appendix A.5.3). We can perform a similar regression to regression B.40 as sanity check for $t \in \{1975, 1990, 2000, 2015\}$,

$$\log w_{u,k,t} = a_t + b \cdot \log \theta_{u,k,t} + u_{k,t}, \quad (\text{B.41})$$

where $w_{u,k,t}$ is the observed urban wage at date t in city k , $\theta_{u,k,t}$ is the estimated urban productivity of city k at date t in the model and a_t a time-effect which controls for aggregate changes common across cities. Standard errors are clustered at the city k level.

In regression B.41, b is estimated to 0.65, highly statistically significant (T-stat close to 5.5). It is robust across years, with cross-sectional estimates of b always highly significant and hovering between 0.5 and 0.9 depending on the date t . This is best illustrated by Figure B.10 for the 2015 cross-section, for which one can see that the model implied wage differential between large cities (e.g. Paris or Lyon) and smaller ones is broadly in line with the data. It is important to note that data on urban wages across cities have not been used to estimate $\theta_{u,k,t}$, targeted to match the distribution of urban population. In other words, the wage gap between large and small cities necessary to induce migrations across cities/regions in line with the data matches relatively well the wage gap observed in the data—despite the latter not being among the data moments used for the estimation.

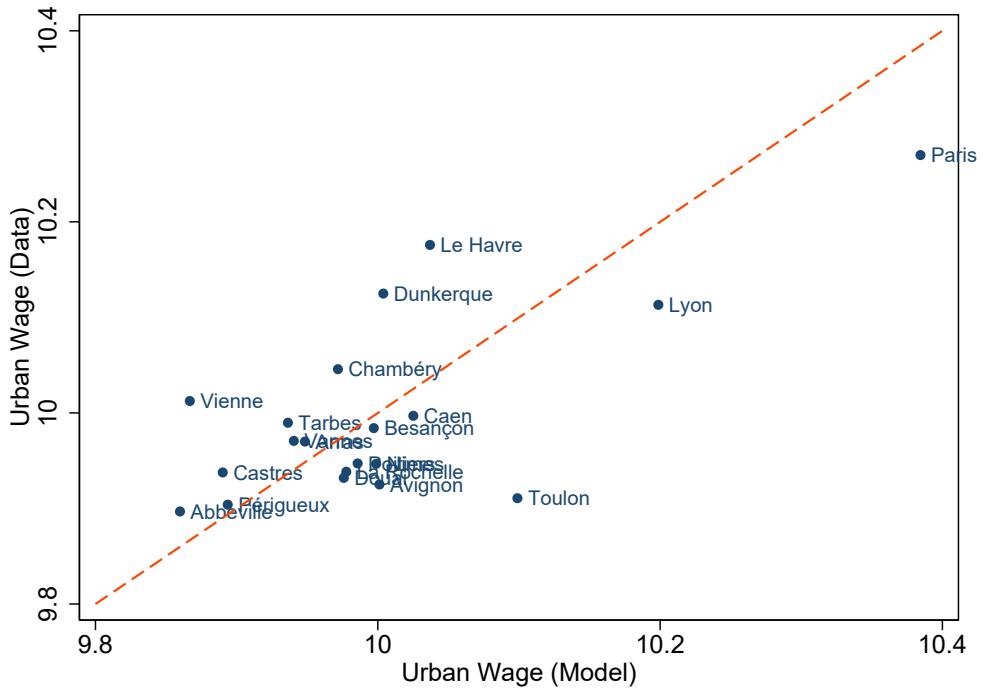


Figure B.10: Urban wage in 2015 in the data and in the model, $\theta_{u,k,2015}$.

Notes: This plots show the average urban wage (in log) in 2015 in the model and in the data (sample of 20 cities). Model counterparts in log are normalized such that the cross-sectional mean matches the data. Model outcomes in 2015 are obtained by linear interpolation between 2010 and 2020. *Data source:* see Appendix A.5.3.

B.3 Sensitivity Analysis and Extensions

This section contains details of sensitivity analysis with respect to the elasticity of substitution between rural and urban goods, σ , the elasticity of substitution between land and labor, ω and the housing supply elasticity, $\epsilon(\ell)$ discussed in Section 4.6 in the main text. We also provide details of the extensions discussed in the same section, where we introduce congestion/agglomeration forces and consider an alternative specification of commuting costs to partly relax the monocentric assumption.

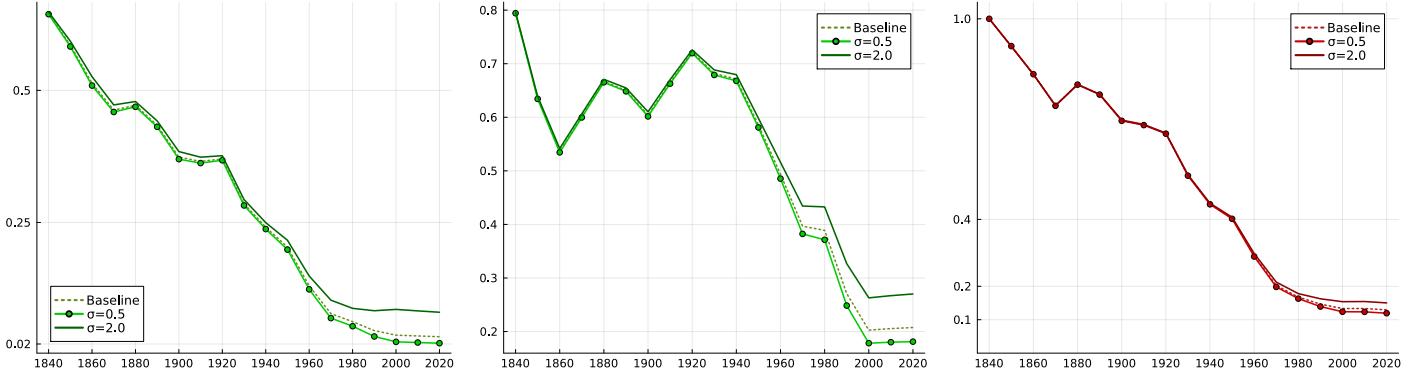
For sensitivity analysis, all model parameters but the one(s) on which sensitivity is performed are kept identical to their baseline values for comparison.

For the extensions, we keep aggregate parameters fixed with the exception of the commuting cost parameter a in order to make sure that we do not shift significantly commuting costs in all regions (particularly relevant when introducing congestion or changing the specification of commuting costs). As in the baseline estimation, the commuting cost parameter a is estimated to make the total urban area, $\sum_k \pi \phi_k^2$, still represents 17% of rural land in the recent period. However, following our baseline estimation (Appendix B.2.3)), regional sectoral productivities, $\theta_{k,s,t}$, are re-estimated to match the relative population of cities and relative local farmland prices while still matching aggregate sectoral productivity observed in the data. This makes sure that aggregate sectoral productivity is unchanged in these extensions compared to the baseline, while still giving the best chances of these extensions to match cross-sectional outcomes. Importantly, such a strategy allows us to compare aggregate outcomes to our baseline and to the data, keeping fixed the aggregate parameters that matter for structural change—including aggregate sectoral productivity.

B.3.1 Elasticity of substitution between rural and urban goods σ

Our baseline estimate relies on an elasticity of substitution between rural and urban goods, σ very close to unity. While there is no clear consensus in the literature on such value ([Storesletten et al. \(2019\)](#)), we perform sensitivity analysis with a lower value of 0.5 and a higher value of 2. This accounts for a wide range of substitution patterns between both goods. We only change σ and keep all other parameters to their baseline values. Results are displayed in Figure B.11 for variables of interest, where we focus on aggregate moments and show the baseline for comparison.

As productivity is growing at a similar rate in both sectors over most of the period, different substitution patterns have very little implications for structural change (Figure B.11a) or the relative price of rural goods (Figure B.11b). As a consequence, and most importantly, the evolution of average urban density is largely unaffected (Figure B.11c). In the second half of the twentieth century, the fall in the relative price of the rural good leads to slightly larger rural consumption and employment shares when the goods are substitutes ($\sigma = 2$) and to slightly smaller ones when the goods are complements ($\sigma = 0.5$), but this has little effect on the average urban density.



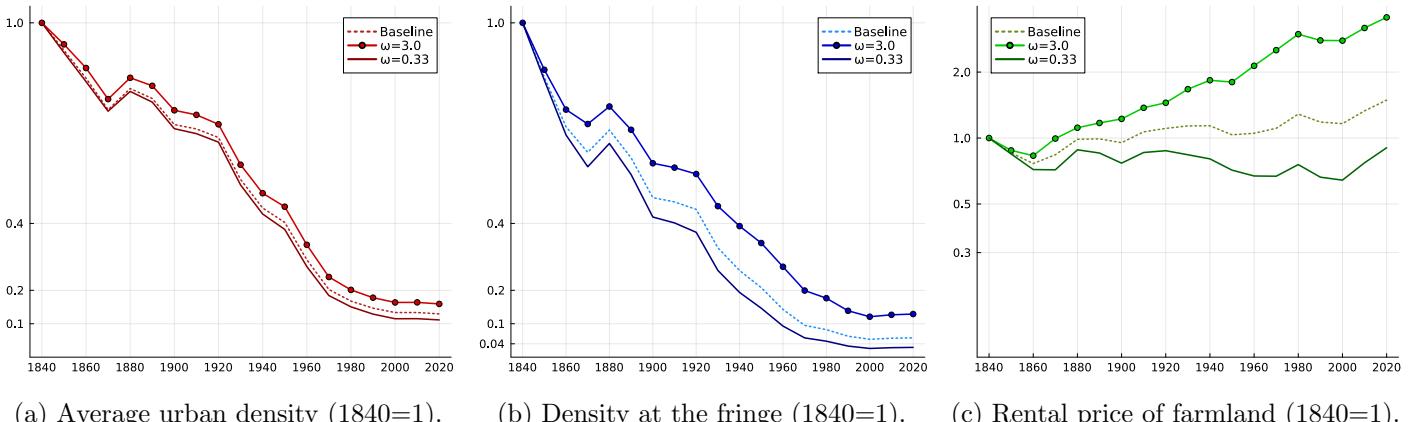
(a) Rural employment share. (b) Relative price of rural good. (c) Average urban density (1840=1).

Figure B.11: Sensitivity to the elasticity of substitution between rural and urban goods σ .

Notes: The elasticity of substitution between rural and urban goods σ is set to a low value of 0.5 (resp. a high value of 2). All other parameters are kept to their baseline value. Simulation for the baseline calibration shown in dashed for comparison.

B.3.2 Elasticity of substitution between land and labor ω

Our baseline simulation assumes a unitary elasticity of substitution between land and labor, $\omega = 1$, in the rural production. Values used in the literature typically range between 0 and 1 (Bustos et al. (2016) and Leukhina and Turnovsky (2016)). We perform sensitivity analysis with a lower value of 0.33. We also show results for a high value of 3 to enlighten further the quantitative importance of the adjustment of land values at the fringe of the city for our results.⁸ We only change ω and keep all other parameters to their baseline values.



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

Figure B.12: Sensitivity to the elasticity of substitution between land and labor ω .

Notes: The elasticity of substitution between land and labor ω is set to a low value of 0.33 (resp. a high value of 3). All other parameters are kept to their baseline value. Simulation for the baseline calibration shown in dashed for comparison.

Results are displayed in Figure B.12 for variables of interest, where we focus on aggregate moments

⁸A higher ω limits the fall of farmland values at the fringe of cities when workers move towards the urban sector.

and show the baseline for comparison. With a lower elasticity of substitution, the rental price of farmland falls more (increases less) following structural change as land and labor are more complement in the rural sector (Figure B.12c). As the opportunity cost of expanding the city is lower, the urban area increases more and the average urban density falls more (Figure B.12a). This is driven by a larger fall of density in the cheaper suburban parts (Figure B.12b). With $\omega = 0.33$, the model matches better the expansion in area and the corresponding decline in average density observed in French cities since 1870. To the opposite, if land and labor are more substitutes ($\omega = 3$), the reallocation of workers away from agriculture puts less downward pressure on the value of farmland, limiting the expansion of the urban area and the decline in density, which falls short of the data. These experiments further illustrate the importance of the farmland price adjustment at the urban fringe to understand the reallocation of land use.

B.3.3 Housing Supply Elasticity ϵ

Our baseline simulation features location-specific housing supply elasticities with a lower elasticity at the city center relative to the fringe, where the values increase linearly from $\epsilon(0) = 2.0$ to $\epsilon(\phi_k) = 5$. As sensitivity analysis, we set the elasticity to 3 in all locations, in the mid-range of empirical estimates. This value corresponds to a land share in housing of 25%, slightly lower than the average in the data over the period 1979-2019. For this sensitivity analysis, we change only the housing supply elasticities, keeping all other parameters to their baseline values.

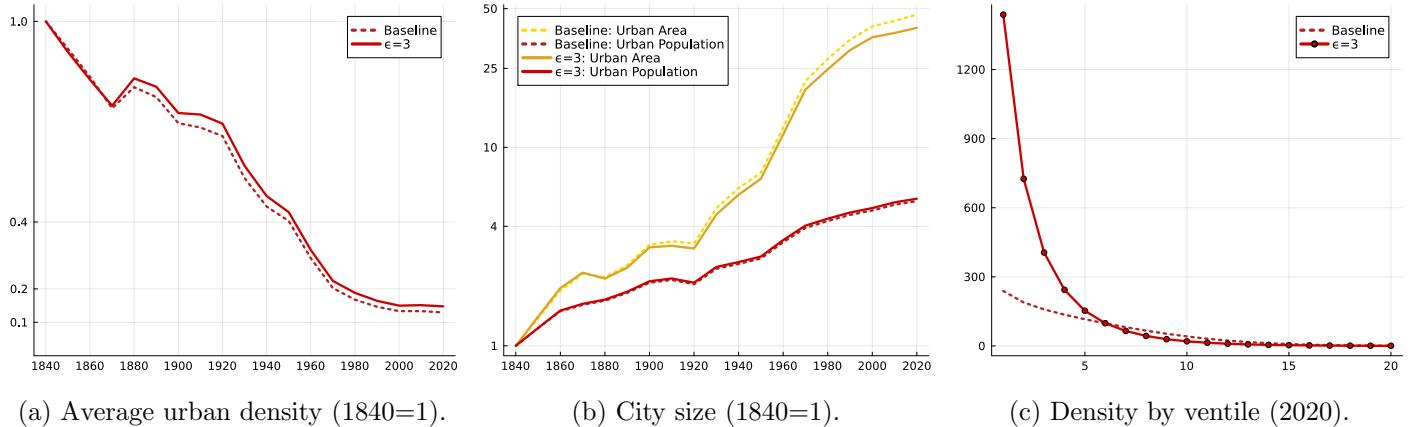


Figure B.13: Sensitivity analysis to setting $\epsilon(\ell) = \epsilon_r = 3$.

Notes: The housing supply elasticity ϵ is set to 3 in all locations (within and across regions). All other parameters set to their baseline value. 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.

After solving the model, results are displayed in Figure B.13 for variables of interest (with the baseline for comparison). Results regarding the time evolution of the aggregate variables of interest—employment, relative price of rural goods, urban area, average urban density and land values—are barely affected and most of them are not displayed. The most noticeable difference is that a constant housing supply elasticity generates a city center much denser relative to the suburban part.

Compared to our baseline simulation, a more elastic housing supply at the center leads to a larger provision of housing in these locations. As a consequence, average urban density is higher in all but the initial periods (see Figure B.13a). The same implication can be seen in Figure B.13b, where urban area is shown to be a bit smaller with constant supply elasticity: the city needs to expand less to host more numerous urban workers. Striking in this regard is how the density structure of the average city at a given time point changes, as shown in Figure B.13c for 2020. With a constant housing supply elasticity, the model generates an extremely dense center, and the fall in density as we move away from the center is much faster than seen in the baseline. The within-city density gradient becomes much steeper, much more than in the data.

B.3.4 Congestion and Agglomeration

Congestion. We consider additional urban congestion costs by assuming that commuting costs are increasing with urban population,

$$a(L_{u,k}) = a \cdot L_{u,k}^\mu.$$

This summarizes the potential channels through which larger cities might involve longer and slower commutes for a given commuting distance.

We set externally $\mu = 0.05$. As explained above, under this specification of commuting costs, we re-estimate the commuting cost parameter a and the region-specific sectoral productivities, $\theta_{k,s,t}$ —making sure that land used for urban purposes remains reasonable relative to the data in the recent period and that aggregate sectoral productivity is unchanged relative to the baseline.

Focusing on the aggregate implications, the evolution of the variables of interest is shown in Figure B.14 together with the baseline results for comparison. Congestion forces reduce the expansion in area and the extent of suburbanization (Figure B.14a). By increasing commuting costs, they also increase urban housing prices (Figure B.14f). However, via general equilibrium forces, they also make rural goods and rural land slightly less valuable—mitigating the direct effect of congestion costs on urban expansion. Overall, the effect of congestion forces on the equilibrium remain relatively mild.

Agglomeration. We introduce urban agglomeration forces by assuming that the urban productivity increases externally with urban employment in city k at date t ,

$$\theta_{u,k,t}(L_{u,k,t}) = \theta_{u,k,t} \cdot L_{u,k,t}^\lambda.$$

We set $\lambda = 0.05$ externally. This value is in the range of empirical estimates for France (Combes et al. (2010)). As discussed above, using $\lambda = 0.05$, we re-estimate the commuting cost function parameter a as well as the sectoral regional productivities to make sure we do not change the level of the commuting costs and aggregate sectoral productivity, while still fitting the cross-sectional data.

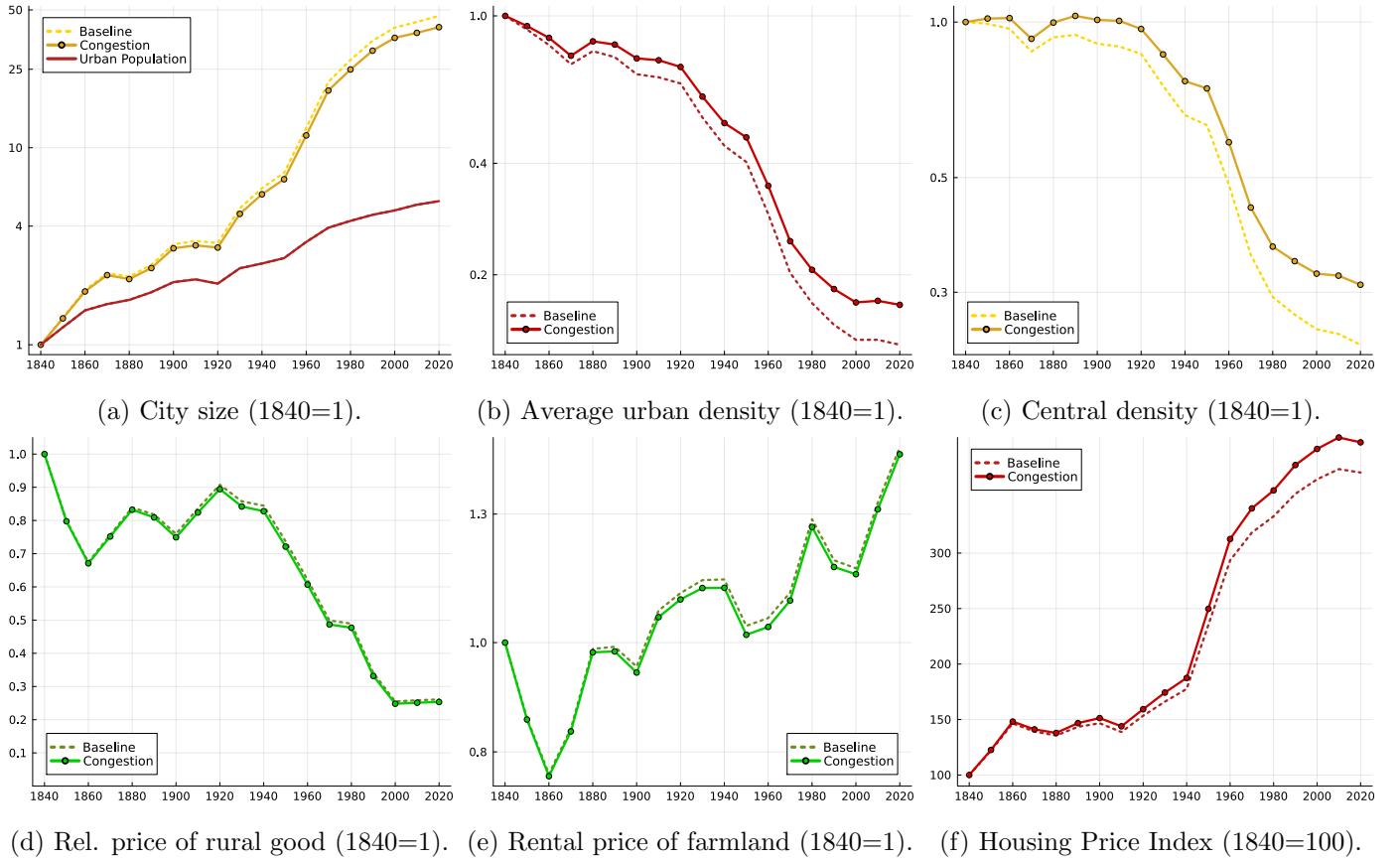


Figure B.14: Congestion forces.

Notes: The solid line represents outcomes in presence of congestion forces, with parameter $\mu = 0.05$. For comparison, outcomes of the baseline simulation are shown with a dotted line. Model's outcomes with congestion are obtained re-estimating the commuting cost parameter a and region-specific productivity parameters as described in the introduction of Section B.3. Other aggregate parameters are left unchanged relative to the baseline. For urban population, outcomes in baseline and counterfactual are indistinguishable.

Irrelevance Result. It is important to remember that the estimation has the relative population size of each city k , $L_{u,k,t}/L_{u,1,t}$, as a target and it matches almost perfectly their urban population by matching relatively well the aggregate employment in the urban sector, $L_{u,t}$. Thus, when re-estimating the model, it should not come at a surprise that outcomes are (almost) identical to our baseline.⁹ This is so because, despite agglomeration forces, the model implied aggregate urban productivity,

$$\theta_{u,t} = \sum_{k=1}^K \left(\theta_{u,k,t} \cdot L_{u,k,t}^\lambda \cdot \left(\frac{L_{u,k,t}}{L_{u,t}} \right) \right) = \left(\sum_{k=1}^K \theta_{u,k,t} \cdot \left(\frac{L_{u,k,t}}{L_{u,t}} \right)^{1+\lambda} \right) \cdot L_{u,t}^\lambda, \quad (\text{B.42})$$

is set to match the French aggregate data, while aiming at generating to same population size distribution of cities (term $\left(\frac{L_{u,k,t}}{L_{u,t}} \right)$ in the previous summation) and the same aggregate urban employment (term $L_{u,t}$).

In other words, the re-estimation of the parameters will essentially adjust the exogenous region-specific productivity parameters, $\theta_{u,k,t}$, relative to the baseline to preserve the targeted moments regarding urban populations. Roughly speaking, larger cities will have a lower exogenous component, $\theta_{u,k,t}$, relative to the baseline estimation, to prevent agglomeration forces from generating counterfactual population size distribution of cities. While quite intuitive, this shows that our identification strategy and the resulting model's output are robust to the presence of agglomeration forces.

The latter irrelevance result makes it however difficult to assess how agglomeration forces affect the equilibrium. More specifically, one cannot assess how the increase in the urban employment share due to structural change further expands cities when agglomeration forces are present. While it is well known that, in the cross-section, larger cities are more productive, which makes them even larger as a consequence (see Combes et al. (2010)), the impact in presence of agglomeration forces of an increase *over time* of aggregate urban employment due to structural change on urban outcomes is much less studied. This is the purpose of the following counterfactual experiment.

Sensitivity to Aggregate Agglomeration Forces. Agglomeration forces have intuitively two possible effects in our framework,

1. In the cross-section, larger cities are more productive. Agglomeration increases the productivity of relatively larger cities ('cross-sectional' agglomeration forces).
2. Over time, due to structural change, all cities are growing in size and becoming more productive. Agglomeration forces increase aggregate urban productivity (labeled as 'aggregate' agglomeration forces).

The objective of this alternative experiment is to study the equilibrium effects of 'aggregate' ag-

⁹They are not exactly identical because we do not match perfectly aggregate urban employment and the aggregate population of cities (see Figure B.18).

glomeration forces following an aggregate urban expansion along the process of structural change.¹⁰ To do so, we re-estimate the model fitting as aggregate productivity a modified version of Equation (B.42),

$$\theta_{u,t} = \sum_{k=1}^K \theta_{u,k,t} \cdot \left(\frac{L_{u,k,t}}{L_{u,t}} \right)^{1+\lambda}, \quad (\text{B.43})$$

In other words, the sectoral productivities are re-estimated but targeting for French aggregate productivity $\theta_{u,t}$ in Eq. (B.43), while the effective model-implied aggregate productivity is $\theta_{u,t} \cdot L_{u,t}^\lambda$. Note that the distribution of (relative) urban populations is targeted in the estimation, such that we abstract from equilibrium effects due to ‘cross-sectional’ agglomeration forces—equivalently the exogenous city-specific urban productivity will adjust in the re-estimation to preserve the relative population size of cities in presence of agglomeration forces. This strategy allows us to disentangle the equilibrium effects of ‘aggregate’ agglomeration forces, relative to the baseline estimation (corresponding to $\lambda = 0$). Note that it is quite immediate to see that if there were only one city (abstracting from cross-sectional implications), this counterfactual would be equivalent to performing sensitivity w.r.t λ —equivalently focusing on the equilibrium effect of agglomeration forces (only present in the ‘aggregate’ with only one city).

For variables of interest, results of this counterfactual experiment (labeled ‘Aggregate Effect’) are displayed in Figure B.15 together with the baseline simulation. We focus on aggregate moments for the ‘average’ city, since all cities are similarly impacted. While cities expand slightly more in area, there is barely any effect of ‘aggregate’ agglomeration forces on urban population. The faster increase in the urban wage across all cities 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 (Figures B.15a—B.15c). However, a higher urban income makes also rural goods more valuable increasing rural workers’ wage almost one for one (Figure B.15d). General equilibrium forces thus prevent workers’ reallocation towards cities. They also make rural land more valuable—mitigating the area expansion of the city (Figure B.15e). As a consequence, despite higher incomes driven by urban expansion, the equilibrium effects of ‘aggregate’ agglomeration forces are very small and the economy behaves quantitatively similarly to our baseline. Thus, while agglomeration effects are potentially important for the cross-sectional allocation of employment, these effects remain small for the expansion of the urban sector and urbanization in the aggregate following structural change.

¹⁰We focus on the equilibrium effects of [2] (‘aggregate’ agglomeration forces) but abstract from [1] (‘cross sectional’ agglomeration forces). We do so for two reasons. First, with equilibrium effects of ‘cross sectional’ agglomeration forces, aggregate productivity will be affected (more productive cities being larger) and it will be difficult to disentangle both effects. We believe that the equilibrium effects of ‘aggregate’ agglomeration forces are more novel, justifying our decision. Second, abstracting from 1 by targeting the same distribution of relative urban populations, simplifies the numerical procedure as, otherwise, agglomeration forces might make some small cities disappear due to the endogenous reallocation of employment across regions, leading to corner solutions.

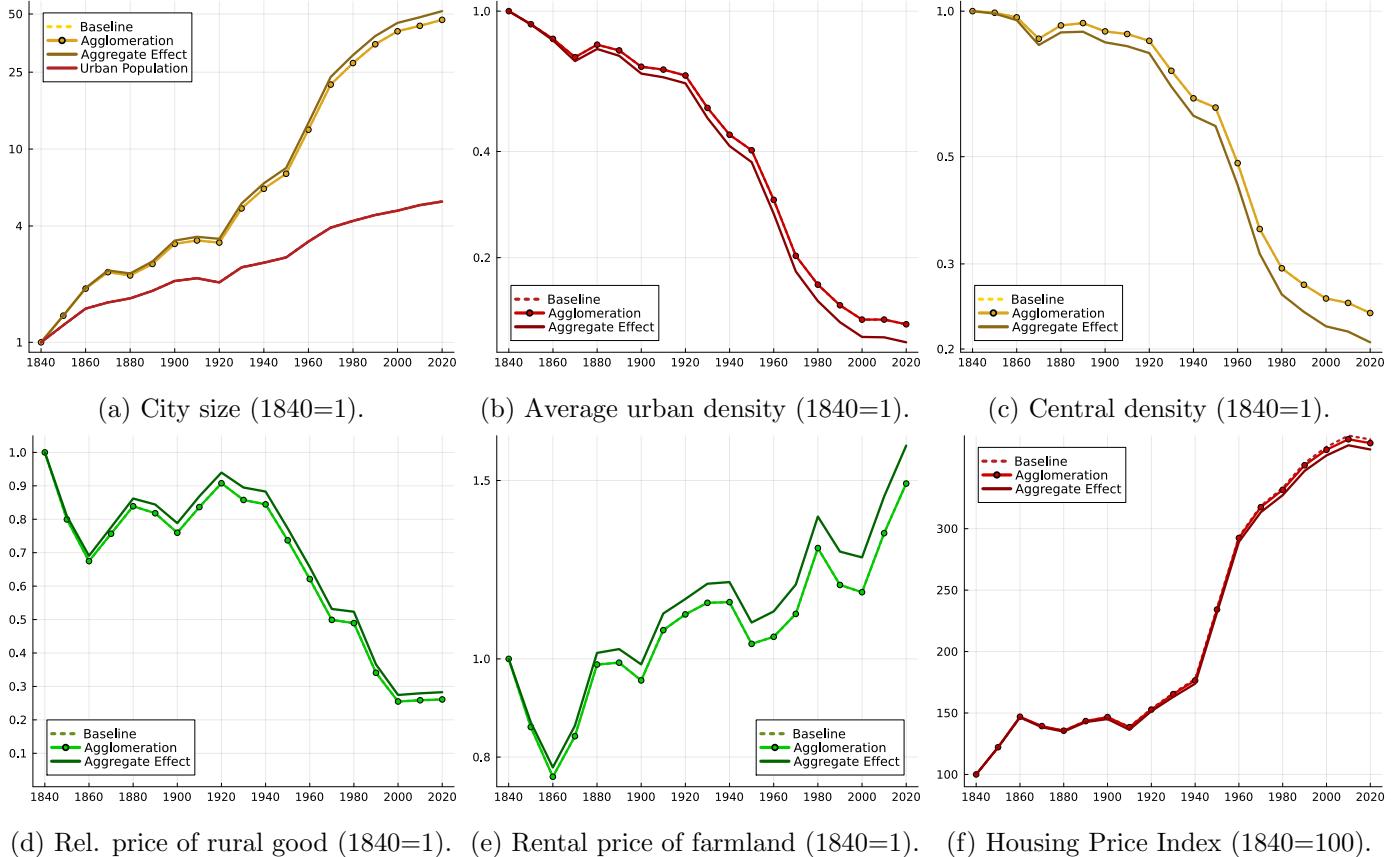


Figure B.15: Agglomeration forces.

Notes: The solid lines represent outcomes in presence of agglomeration forces under both specifications of aggregate productivity (Eq. B.42 and Eq. B.43). The line with circles corresponds to the “irrelevance result”: aggregate productivity matches the data and cross sectional differences in urban productivities, $\theta_{u,k,t}$, matches the relative population of cities. The line without markers shows equilibrium effects of ‘aggregate’ agglomeration forces relative to the baseline, which is shown as dotted line. For urban population, outcomes in baseline and counterfactuals are indistinguishable. Model’s outcomes with agglomeration forces are obtained re-estimating the commuting cost parameter a and region-specific productivity parameters under both specifications of aggregate productivity (Eq. B.42 and Eq. B.43). Other aggregate parameters are left unchanged relative to the baseline.

B.3.5 Commuting distance and residential location

Set-up. 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 endogenizing firms location across space is beyond the scope of the paper, one can still partly relax the monocentric assumption by assuming that commuting distance at location ℓ_k in city k , $d_k(\ell_k)$, does not map one for one with residential distance ℓ_k from the central location. Using data available for the recent period to investigate the link between commuting distance and residential location (see Appendix A.5.2 for details), we find that households residing further away do commute longer distances on average. However, commuting distance increases less than one for one with the distance of residence from the city center. Moreover, individuals residing very close to the center commute longer distances than the distance of their home from the central location. Lastly, data show that commuting distance increases less with the distance of residence from the center in larger cities.¹¹ Based on these observations, we model commuting distance, in location ℓ_k of city k , $d_{k,t}(\ell_k)$ in a reduced-form way as follows,

$$d_{k,t}(\ell_k) = d_0(\phi_{k,t}) + d_1(\phi_{k,t}) \cdot \ell_k, \quad (\text{B.44})$$

with $d_0(\phi)$ being a positive and increasing function of ϕ satisfying $\lim_{\phi \rightarrow 0} d_0(\phi) = 0$, and $d_1(\phi)$ being a decreasing function belonging to $(0, 1)$ with $\lim_{\phi \rightarrow 0} d_1(\phi) = 1$. d_0 represents the (minimum) commuting distance traveled by an individual living in the center, while d_1 is the slope between commuting distance and residential distance from the center. This specification fits recent data well. It also makes sure that at the limit of $\phi \rightarrow 0$, the city is monocentric as all the jobs must be centrally located. The parameters d_0 and d_1 are guided by the data (Section A.5.2) as detailed below. It is important to note that commuting costs are now defined as,¹²

$$\tau_{k,t}(\ell_k) = a \cdot w_{u,k,t}^{\xi_w} \cdot (d_{k,t}(\ell_k))^{\xi_\ell}. \quad (\text{B.45})$$

In the quantitative evaluation, 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 $d_1 \geq 0$. Across cities, $d_0 \cdot \phi$ corresponds to the intercept of Eq. B.45, 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 calibrate d_0 externally to 5% in our quantitative experiment. For a radius of about 20 kms (close 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, $d_1(\phi_{k,t}) = \frac{1}{1+d_1 \cdot \phi_{k,t}}$ corresponds to the slope of Eq. B.45—with an estimated mode across urban areas close 0.7 in the data. We calibrate $d_1 = 2$ externally. This yields after model's estimation a slope coefficient that varies across cities, ranging

¹¹This points towards a larger dispersion of employment away from the center in larger cities. See Appendix A.5.2.

¹²This 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 $1 - \xi_\ell$.

from 0.44 for Paris to an average amongst the remaining cities of 0.66, which is reasonably close to the corresponding empirical moments. Beyond these new parameters, other aggregate parameters with the exception of the commuting cost parameter a are left unchanged but region-specific sectoral productivities, $\theta_{k,s,t}$, are re-estimated while keeping aggregate sectoral productivity unchanged (as discussed in the introduction of Section B.3).

Results. We find that our results are not much affected (Figure B.16). 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 (Figure B.16a). As a consequence of this larger sprawling, the average urban density falls more (Figure B.16b). This is driven by a larger fall of central density, the most noticeable difference relative to our baseline monocentric model (Figure B.16c). 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. As a consequence, the within city density gradient is less steep (Figure B.16e). Due to the larger area expansion of cities, rural land gets scarcer and more valuable relative to the baseline (Figure B.16f).

This specification provides also a slightly better fit of the data across cities (Figure B.17). Relative to the baseline monocentric model (in Figure B.18 for comparison), commuting distances in the center (resp. at the fringe) are larger (resp. lower) in larger cities. 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 the exception of few small cities in the most recent period (2015).¹³

¹³Starting the most recent period (2015), the urban population of the smallest cities expand too much relative to data due to lower commuting costs relative to the baseline (Figure B.17a). This is mostly due to the numerical solution which puts little weight on fitting small cities relative to fitting larger cities and the rural employment share. For consistency with the rest of the paper, we do not change weights in the objective functions for the counterfactuals given that this faster expansion of small cities relative to the data only appears post-2010.

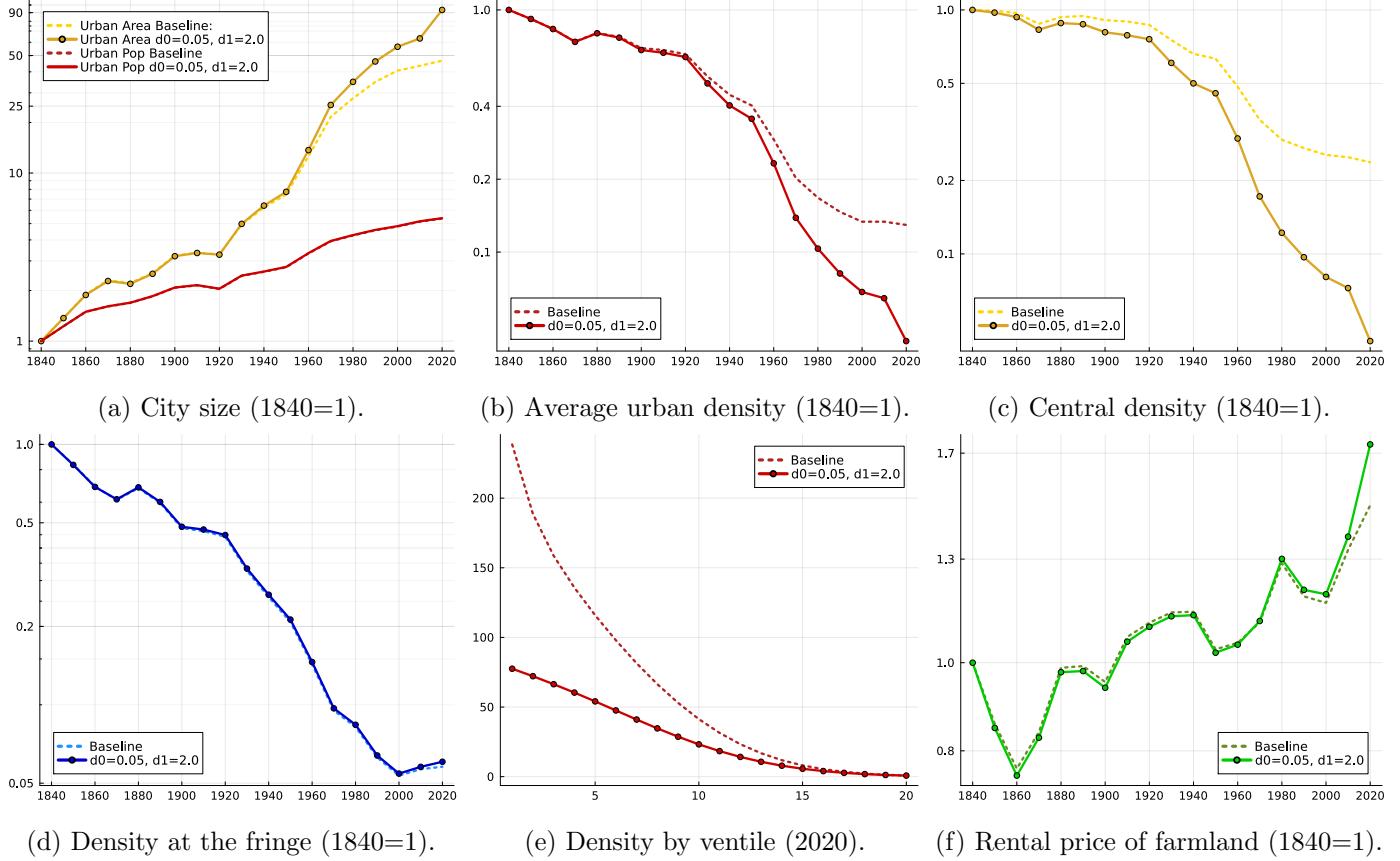


Figure B.16: Relaxing monocentricity. Aggregate Moments.

Notes: The solid line represents outcomes in the extended model with alternative commuting costs ((d_0, d_1) extension). For comparison, outcomes of the baseline simulation are shown with a dotted line. Model's outcomes under this alternative specification of commuting costs are obtained re-estimating the commuting cost parameter a and region-specific productivity parameters as described in the introduction of Section B.3. Other aggregate parameters are left unchanged relative to the baseline. For urban population, outcomes in baseline and counterfactual are indistinguishable.

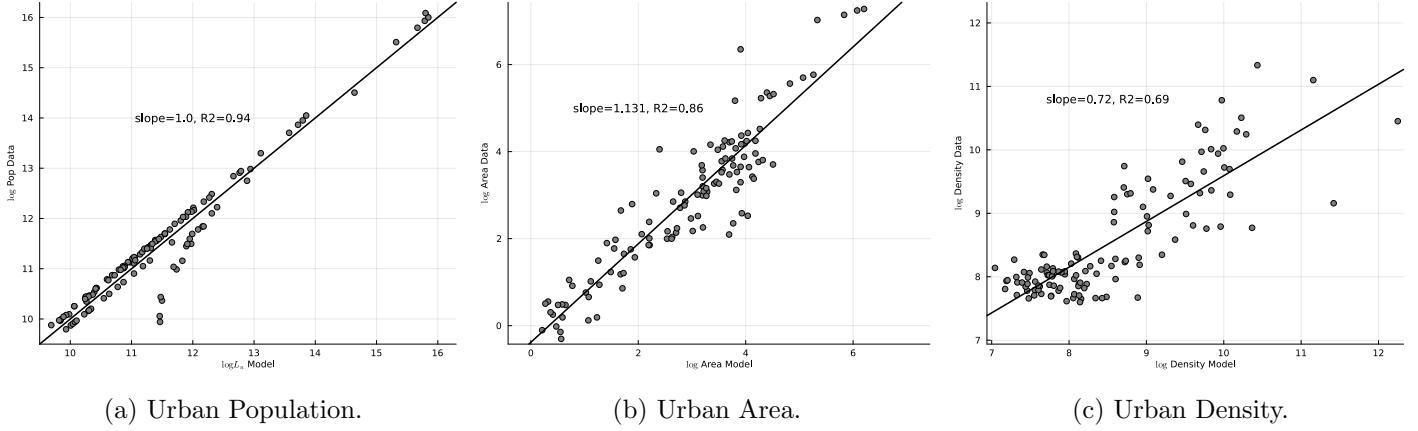


Figure B.17: Relaxing monocentricity. Regional Urban Moments.

Notes: Here we illustrate the impact of relaxing monocentricity on the distribution of urban area in the extended model with alternative commuting costs ((d_0, d_1) extension). We plot the log of model population/areas/density vs the log of population/areas/density in the data for all observed dates. Variable are centered such that the mean in the data across observations match the model's counterpart. Data and model's outcomes are for the dates $t \in \{1870, 1950, 1975, 1990, 2000, 2015\}$, with the model interpolated for 1975 and 2015. Model's outcomes under this alternative specification of commuting costs are obtained re-estimating the commuting cost parameter a and region-specific productivity parameters as described in the introduction of Section B.3. Other aggregate parameters are left unchanged relative to the baseline.

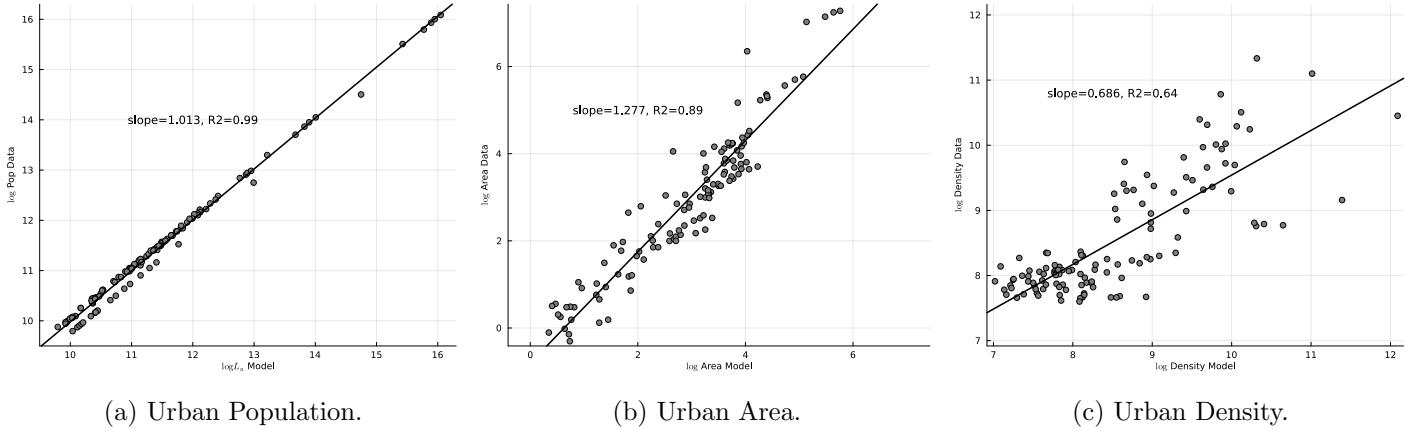


Figure B.18: Baseline Model. Regional Urban Moments.

Notes: We plot the log of model population/areas/density vs the log of population/areas/density in the data for all observed dates in the baseline model. Variable are centered such that the mean in the data across observations match the model's counterpart. Data and model's outcomes are for the dates $t \in \{1870, 1950, 1975, 1990, 2000, 2015\}$, with the model interpolated for 1975 and 2015. Outcomes of the baseline simulation of the quantitative model where parameters are set to the values of Table 1.

Bibliography

- Baum-Snow, Nathaniel and Lu Han**, “The Microgeography of Housing Supply,” *Journal of Political Economy*, 2023, forthcoming.
- Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli**, “Agricultural productivity and structural transformation: Evidence from Brazil,” *American Economic Review*, 2016, 106 (6), 1320–65.
- 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.
- DeSalvo, Joseph S and Mobinul Huq**, “Income, residential location, and mode choice,” *Journal of Urban Economics*, 1996, 40 (1), 84–99.
- Dunning, Iain, Joey Huchette, and Miles Lubin**, “JuMP: A Modeling Language for Mathematical Optimization,” *SIAM Review*, 2017, 59 (2), 295–320.
- Fukushima, Nobusumi, Yuichi Nagata, Sigenobu Kobayashi, and Isao Ono**, “Proposal of distance-weighted exponential natural evolution strategies,” in “2011 IEEE Congress of Evolutionary Computation (CEC)” 2011, pp. 164–171.
- Gollin, Douglas, David Lagakos, and Michael E Waugh**, “Agricultural productivity differences across countries,” *American Economic Review*, 2014, 104 (5), 165–70.
- Leukhina, Oksana M. and Stephen J. Turnovsky**, “Population Size Effects in the Structural Development of England,” *American Economic Journal: Macroeconomics*, July 2016, 8 (3), 195–229.
- 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.
- Storesletten, Kjetil, Bo Zhao, and Fabrizio Zilibotti**, “Business Cycle during Structural Change: Arthur Lewis’ Theory from a Neoclassical Perspective.,” Technical Report, National Bureau of Economic Research 2019.

Su, Che-Lin and Kenneth L Judd, “Constrained optimization approaches to estimation of structural models,” *Econometrica*, 2012, 80 (5), 2213–2230.