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

February 6, 2023

Contents

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 Desriers (2007) until 2000 by decade and available on annual basis since 2000 on the website of the Ministry (Agreste)).

Before WW2, agricultural statistics on land use are also available but on a very irregular basis.[1] 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 Savoye and Comté de Nice in 1860 (see discussion in Augé-Laribé (1945)). This makes the across-time comparison difficult, even though we show our measure of the SAU as a share of the French territory at the time. A third difficulty

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

for the early periods (before 1882), detailed below, regards the treatment of pasture and grazing fields in a consistent way across years.

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

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

Nineteenth century. Before World War 1, we have measures in 1882 and 1892 (Mauguin (1890), Fléchey (1898), Hitier (1899) and Toutain (1993) for further details). Both measures are consistent across sources, including the main results of the 1892 Agricultural Census as a more primary source.[3] This gives an SAU of 34 882 thousands of ha in 1882 and 34 720 in 1892—slightly higher than the values in between the wars despite a smaller French territory. Figure 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', 'pâtis communaux/vaines pâtures') is excluded from the SAU. The non-cultivated part of agricultural land thus appears to be a much larger measured area than in all subsequent years.[5] As discussed in the results of the 1892 Agricultural Census, comparison across years is difficult due to the reallocation of grazing fields into the cultivated part of French land over the period 1840-1880. This reallocation is quite artificial—mostly a statistical artefact coming from the earlier exclusion of common pasture. Excluding entirely the measured non-cultivated part from the SAU in 1840 gives thus a lower bound,

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

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

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

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

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

Figure 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.

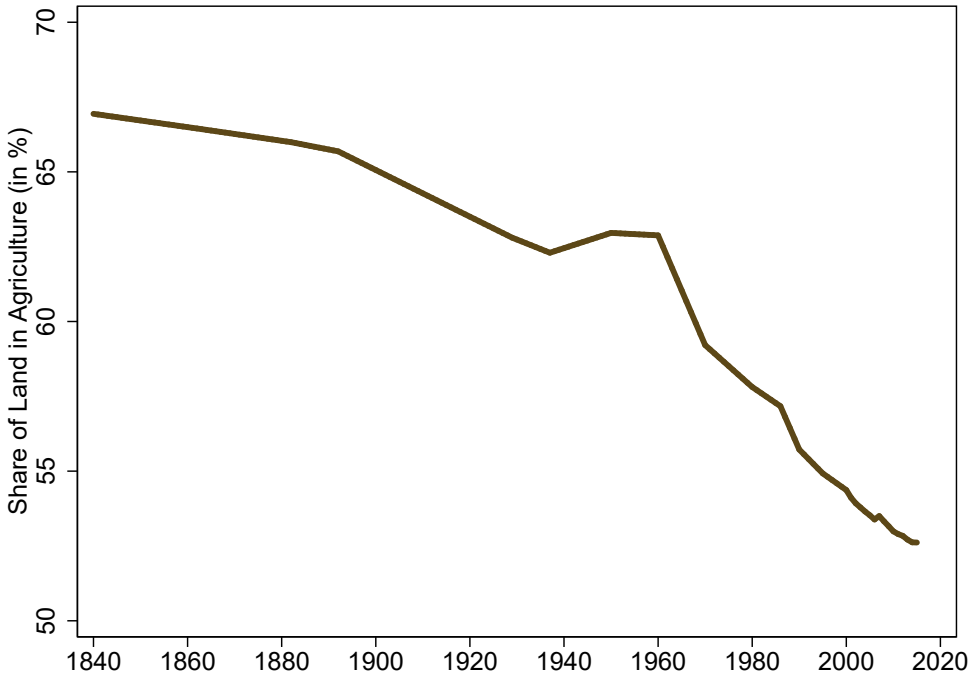


Figure A.2: Shares of Land used in Agriculture (1840-2015).

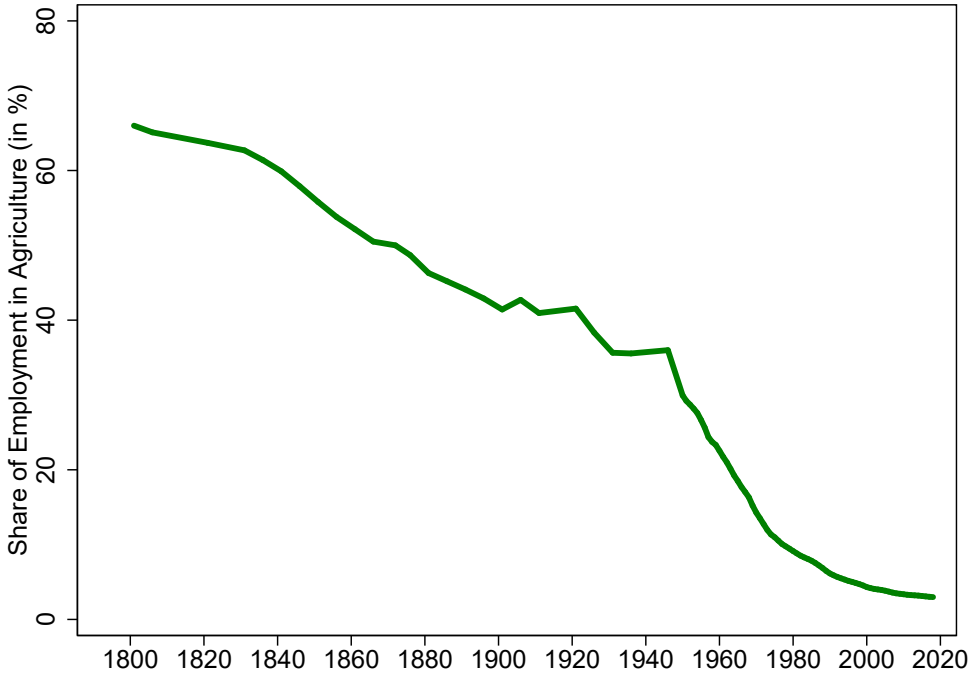


Figure A.3: Shares of Employment in Agriculture (1806-2018).

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 covers the entire period 1815-2019. The series are interrupted at war times: observations are missing for the periods 1914-1919 and 1939-1948.

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

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

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

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

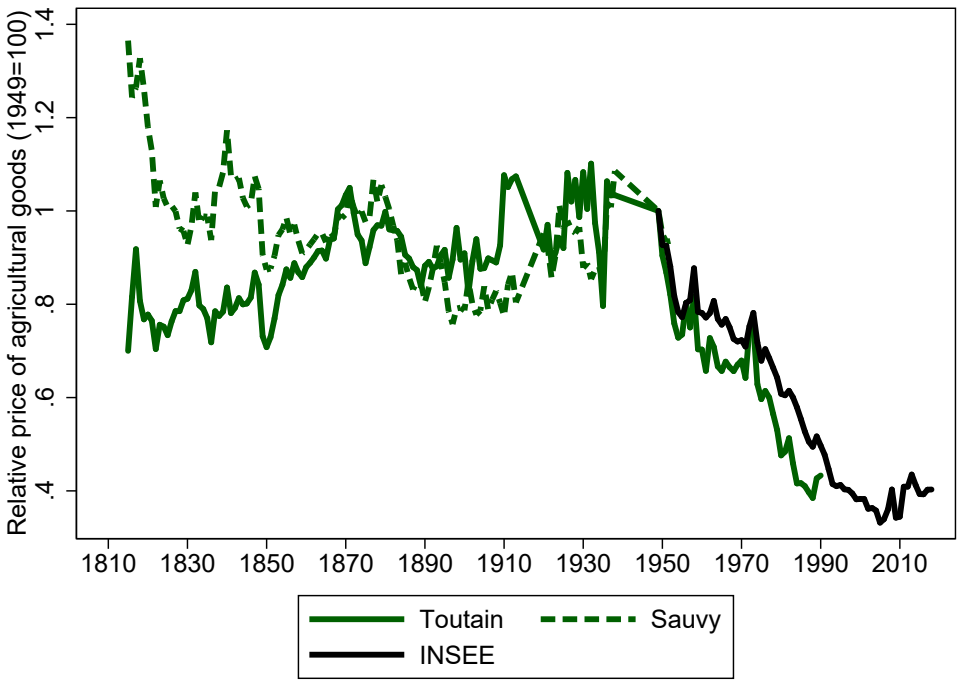


Figure A.4: Relative prices of agricultural goods, 1949=100 (1815-2019).

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

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

as

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

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

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

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

Sensitivity analysis for the price of agricultural goods. Before WW2, the Statistique Generale de France (the predecessor of INSEE), in particular thanks to the work of Alfred Sauvy, provides an alternative series for the price of agricultural goods: 'indice des prix de gros agricoles' which is constituted by a basket of 19 raw agricultural commodities (food related).[9] The series is retropolated back to 1810 by A. Sauvy (see Sauvy (1952)). This data includes some foreign commodities (e.g. English and US corn prices) and is in part computed using customs price data. For this reason, we use the price of agricultural goods computed using production data of Toutain pre WW2 as baseline. This said, the 'indice des prix de gros agricoles' still contains useful information regarding the price of agricultural goods in France before WW2. Comparison with the price computed using production data from Toutain indicates that the two series exhibit very similar patterns starting 1850. Prior to this date, the 'indice des prix de gros agricoles' from Sauvy (1952) exhibits a significant downward trend, while our baseline from Toutain stays roughly stable (see Figure 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 θs) up to a constant of normalization. Our measure of land use in agriculture necessary to estimate rural productivity starts in 1840. Thus, we compute 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

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

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

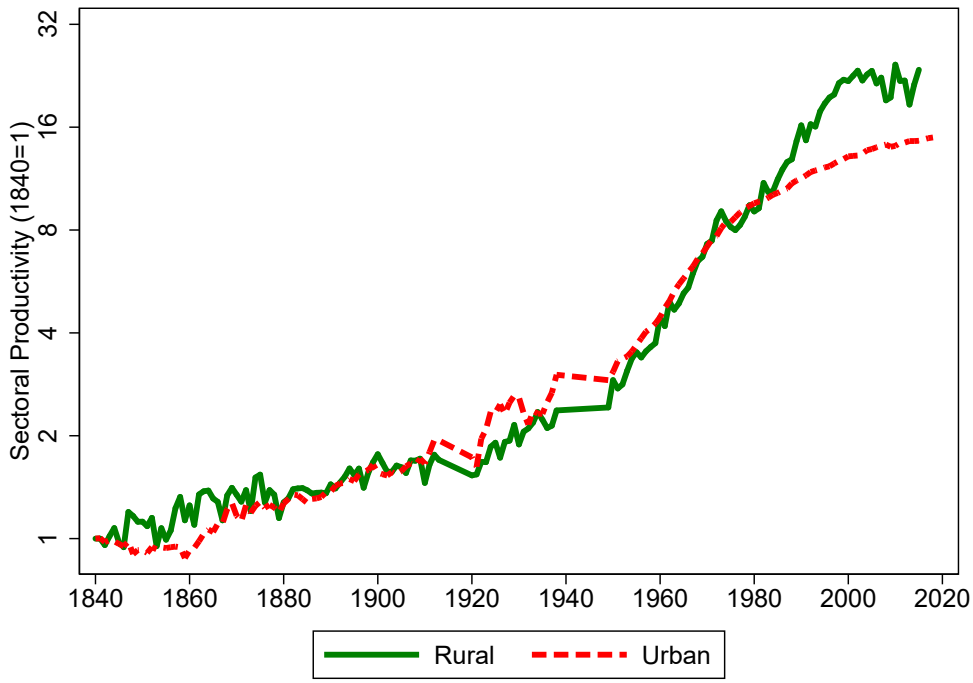


Figure A.5: Rural and Urban Aggregate Productivity, 1840=1 (1840-2019).

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,

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

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,

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

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

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

Due to the mere presence of a price index, this series is defined up to a multiplicative constant. Like $\theta_{u,t}$, we normalize $\theta_{r,t}$ to unity in the first period (1840). This gives the time-series for $\theta_{r,t}$ plotted in Figure 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-

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

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

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]

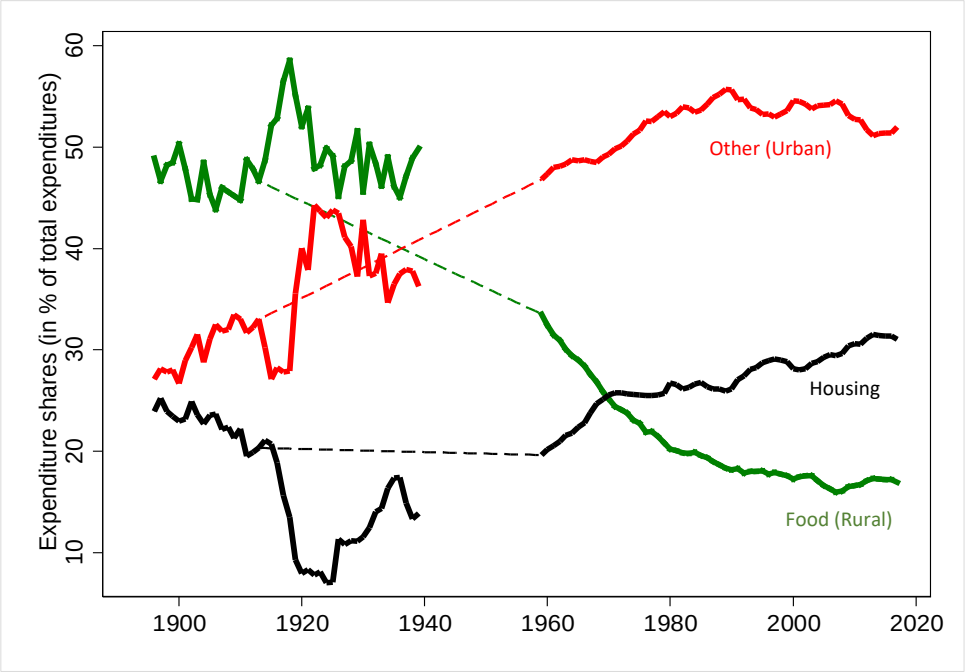


Figure A.6: Spending Shares for Rural, Urban and Housing goods.

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 to 23% to 31%; the share of expenditure on other goods increased as a consequence from 27% to more than 50%. This reallocation of expenditures away from rural goods towards housing and urban goods fits well with the process of structural transformation.

Rent control and the housing expenditure share. An important issue is the significant and persistent dip of the housing expenditure share starting at WW1. This evolution is largely due to the presence of rent controls that were put in place at the beginning of WW1 in France. As the French government wanted families to be able to afford their home during the war, it decreed that rents would be blocked (in nominal terms). As inflation picked up, this generated a large fall

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

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).

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

[15] Data 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 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 to

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

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

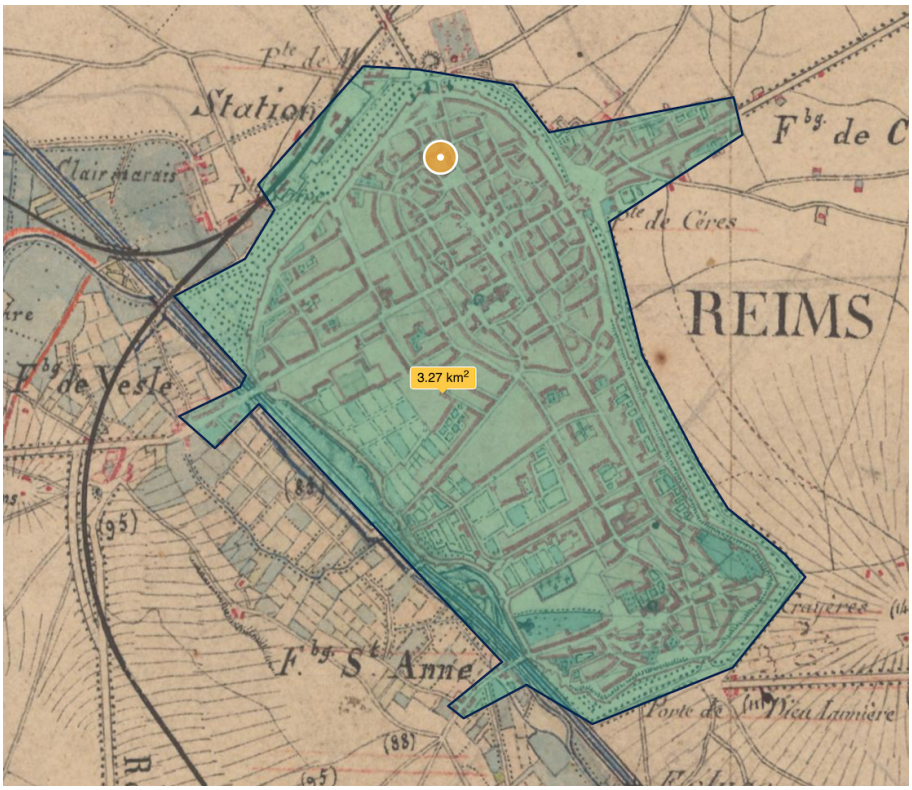


Figure A.7: Area measurement of Reims using Etat Major map

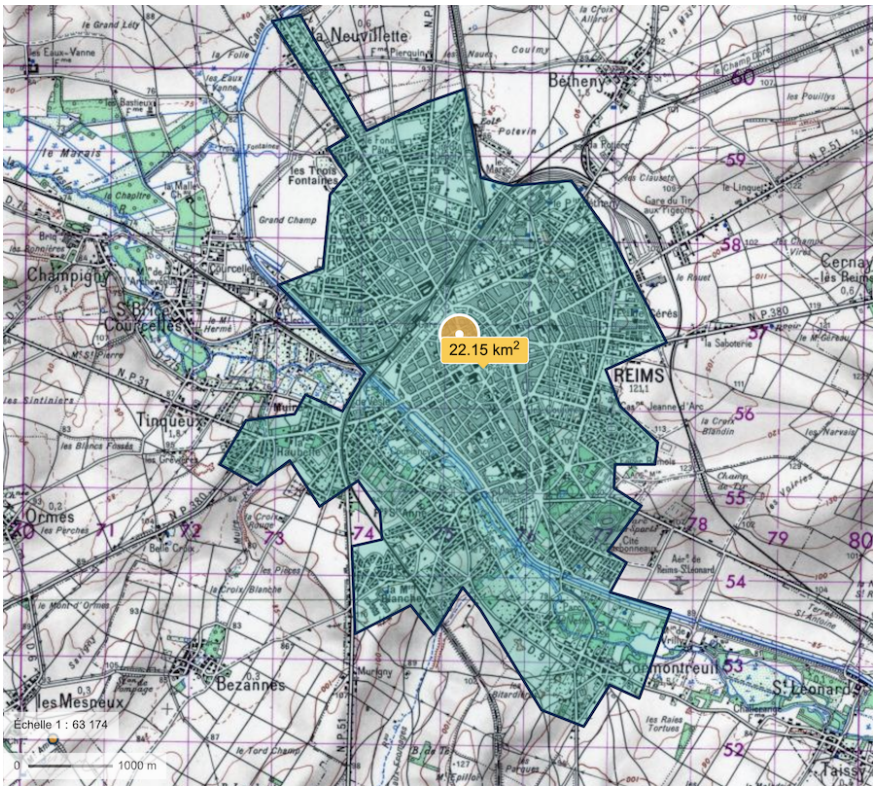


Figure A.8: Area measurement of Reims using 1950 map

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

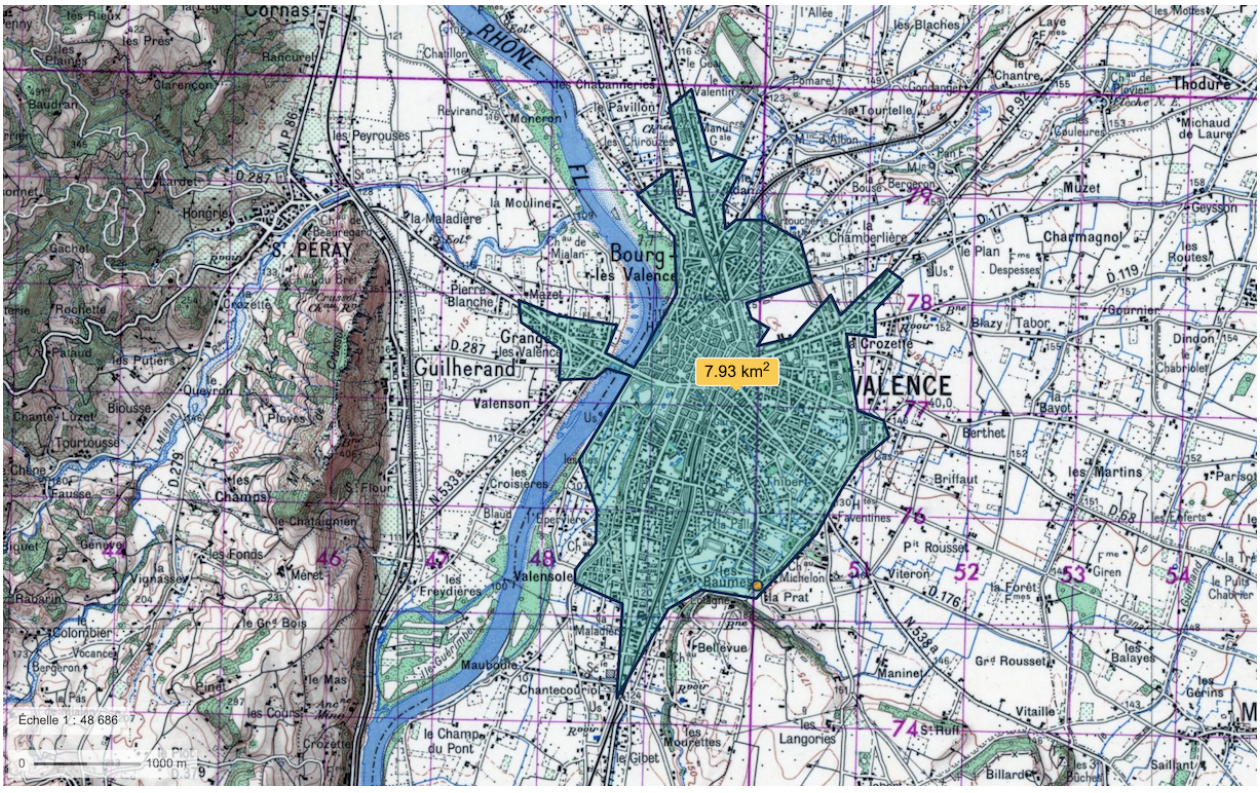


Figure A.10: Area measurement of Valence using 1950 map

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 very low density built-up. The aim is to make the manual measure as close as possible from the automated approach in order to distinguish very low-density suburbs from city proper. 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.

[18]The 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, close by municipalities are considered as separate for our measure of urban area by definition. However, there are always 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).

[19] https://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

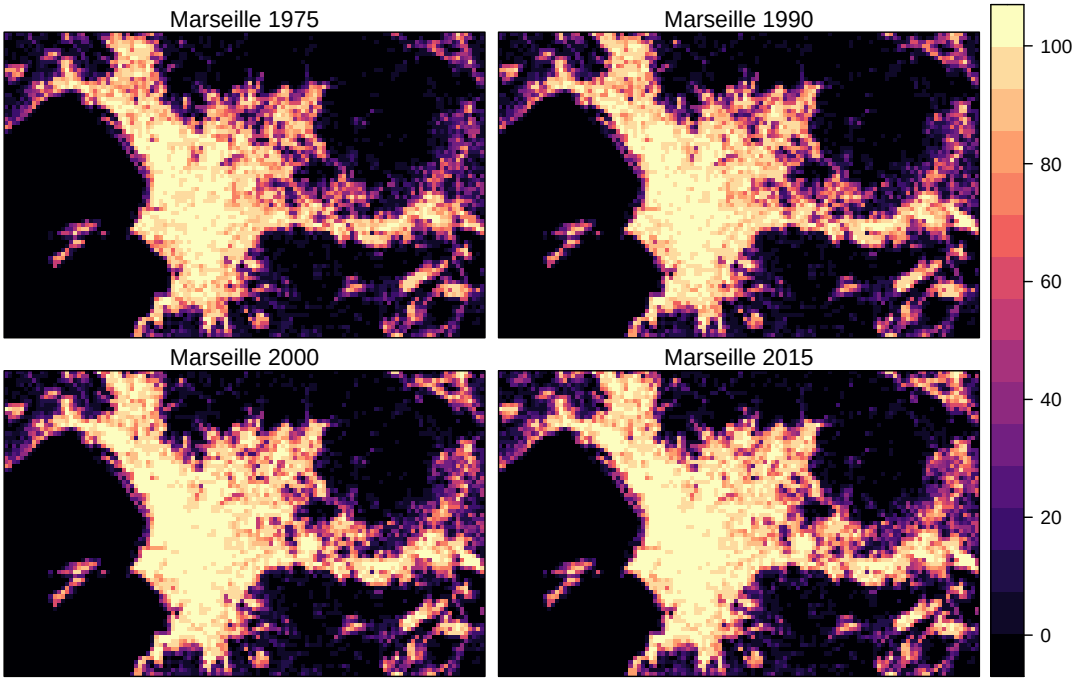


Figure A.11: `GHS-BUILT` raster map of Marseille. The color scale represents percentage built-up in each grid cell.

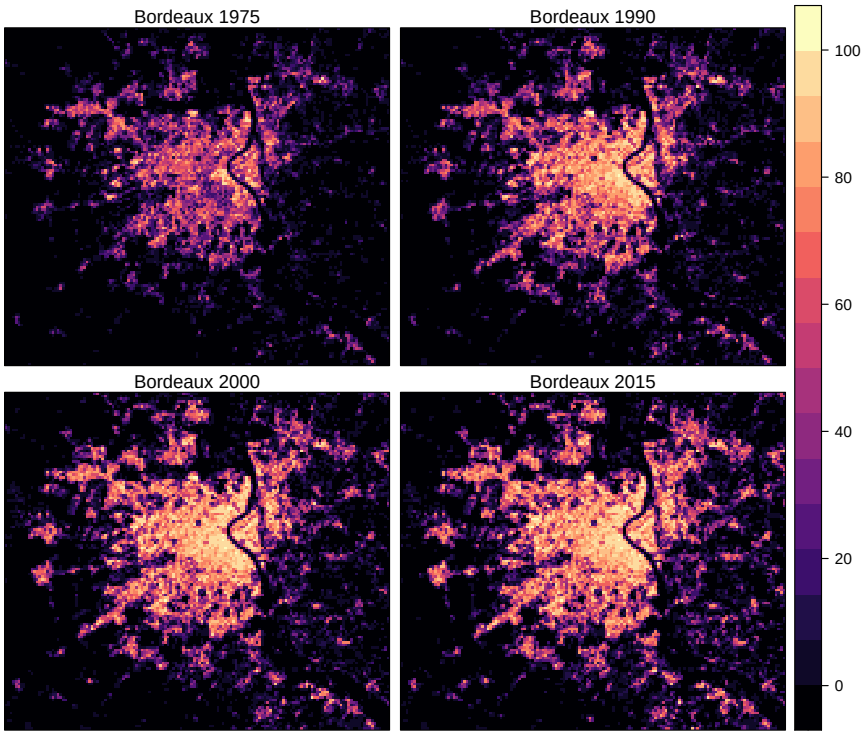


Figure A.12: `GHS-BUILT` raster map of Bordeaux.

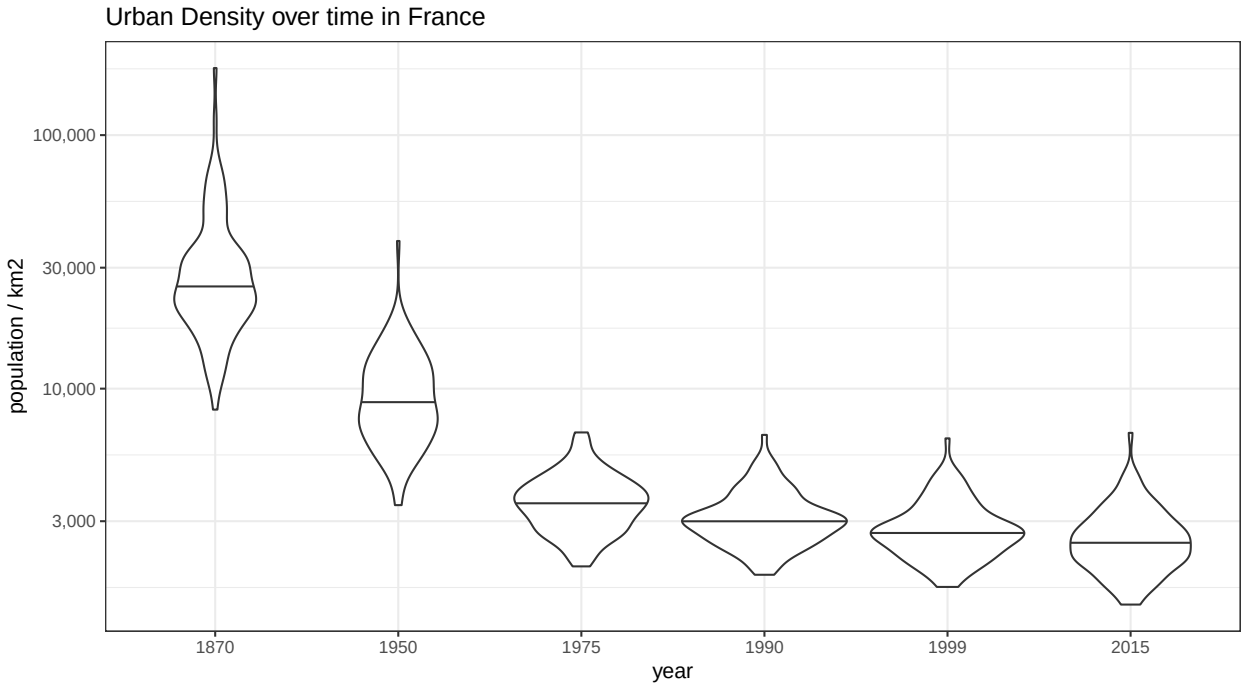


Figure A.13: Distribution of Urban Density over Time. This *violin plot* represents the distribution of densities at each date, labeling the extreme values. The horizontal line denotes the median value.

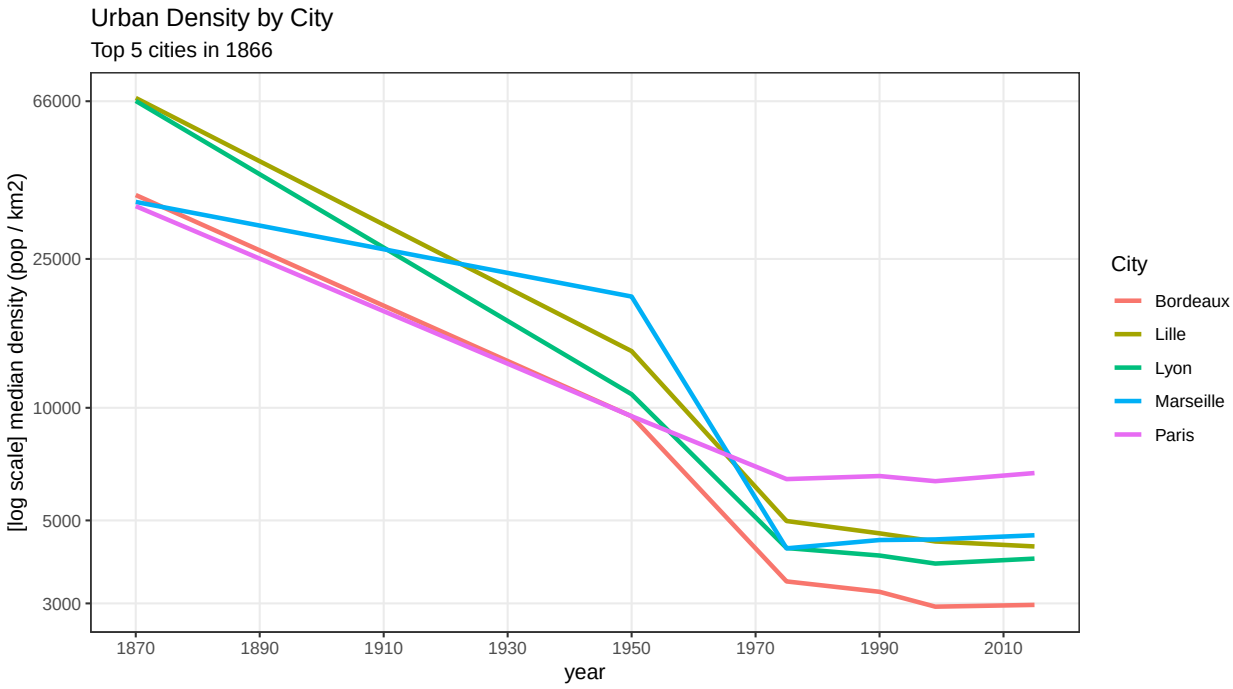
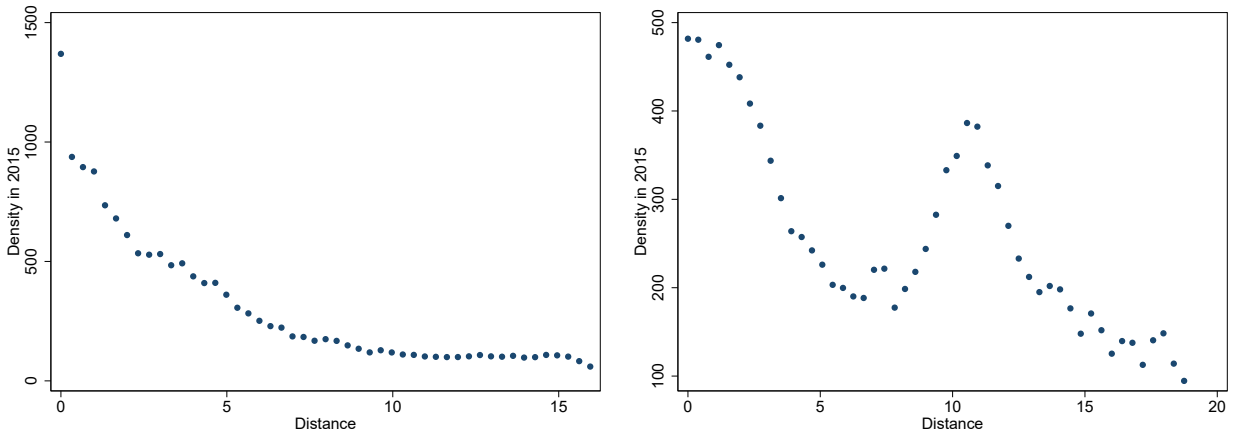


Figure A.14: Density in the largest five cities in 1876 over time.

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.



(a) Lyon (b) Lille

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,

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

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 few polycentric ones such a 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. A slightly higher value of 0.152 is obtained if the gradient is computed only on the central part of the most polycentric ones (below 6kms distance from the center). If we consider only the first 10kms from the center for all cities in the sample, we get a gradient of 0.154.

Thus, according to our empirical estimates, we find a density gradient ranging from 0.14 to 0.18 for the average city in our sample and the value of 0.15 constitutes our baseline estimate. Note that beyond the value for this average density gradient, our empirical investigation also shows that the exponential shape of Eq. 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 excessive relative to 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 ex-

cessive 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 assumed for delimiting cities, the reason we perform robutness 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 gridcell 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.

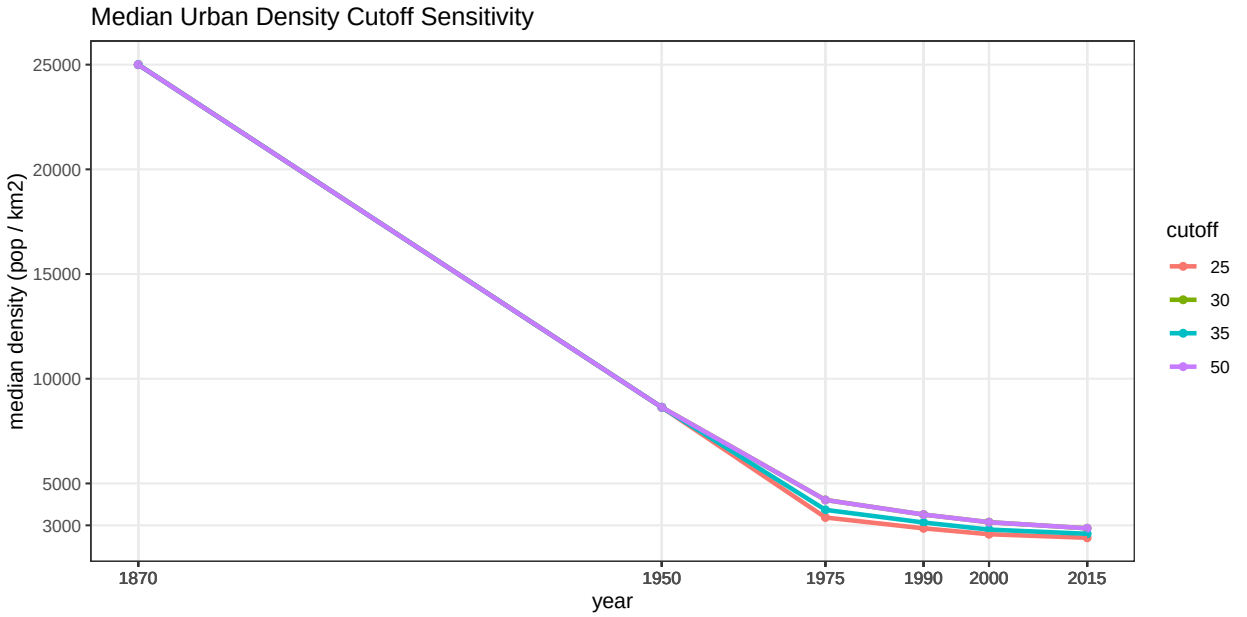


Figure A.16: Median urban density for different cutoff parameter values. The parameter indicates the percentage of a grid cell (250x250 meter) that has to be built-up in order to be classified as *urban area*.

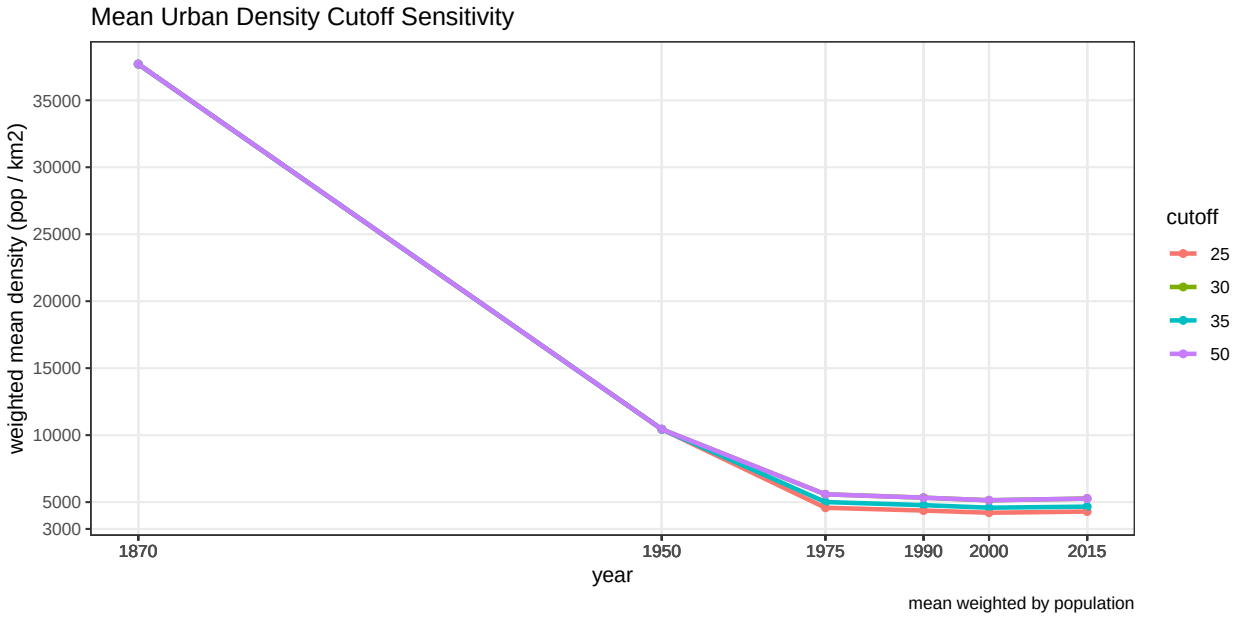


Figure A.17: Weighted mean urban density for different cutoff parameter values. The parameter indicates the percentage of a grid cell (250x250 meter) that has to be built-up in order to be classified as *urban area*.

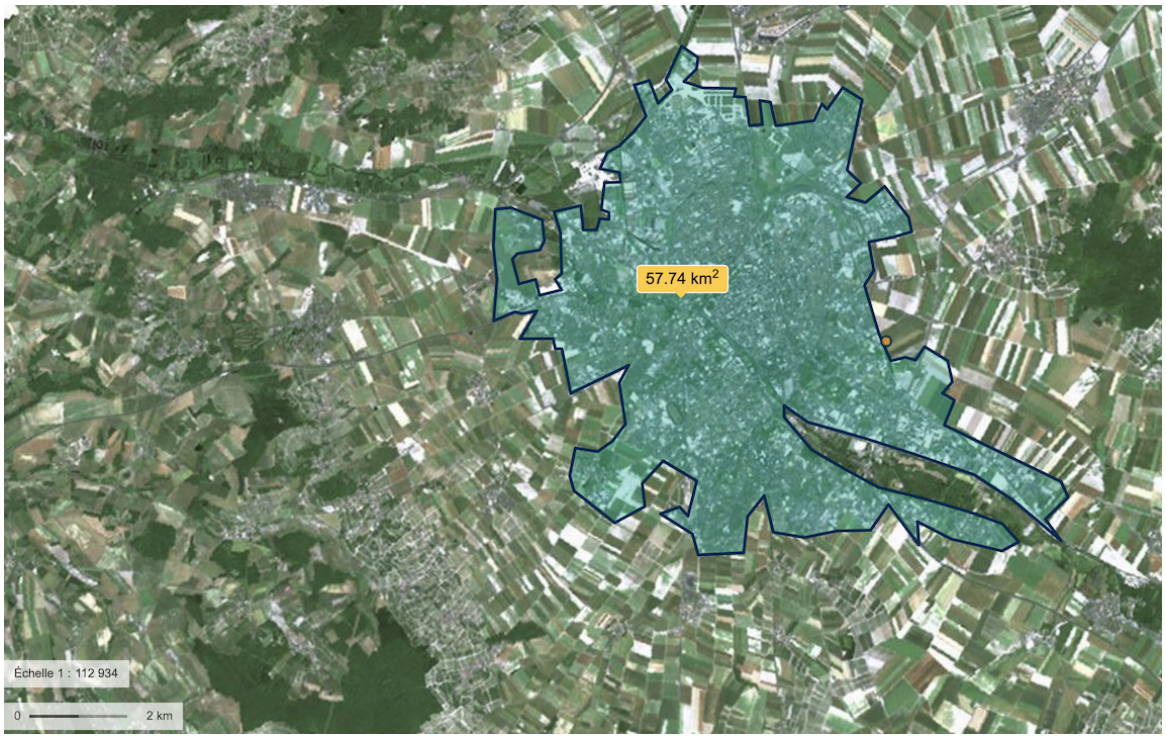


Figure A.18: Area measurement of Reims using modern day photograph - used only for cross-checking GHSL measures.

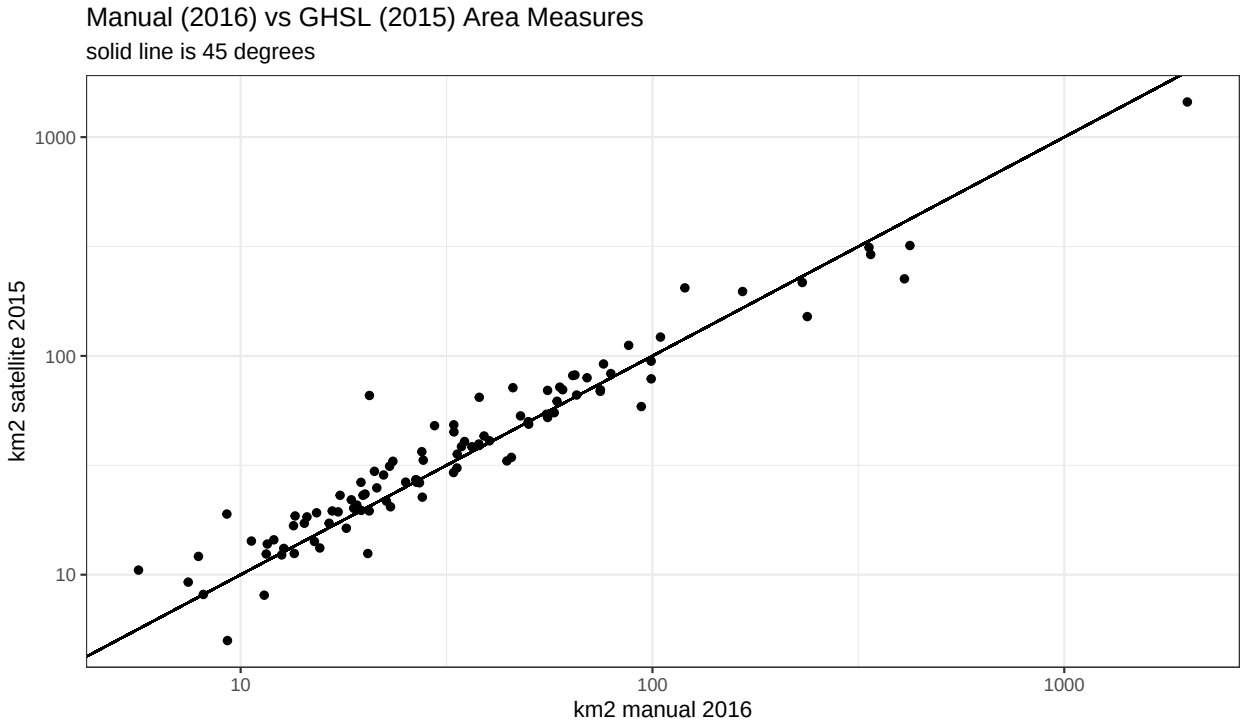


Figure A.19: Comparing manually obtained area measures for each of our cities with automatically obtained ones via GHSL data.

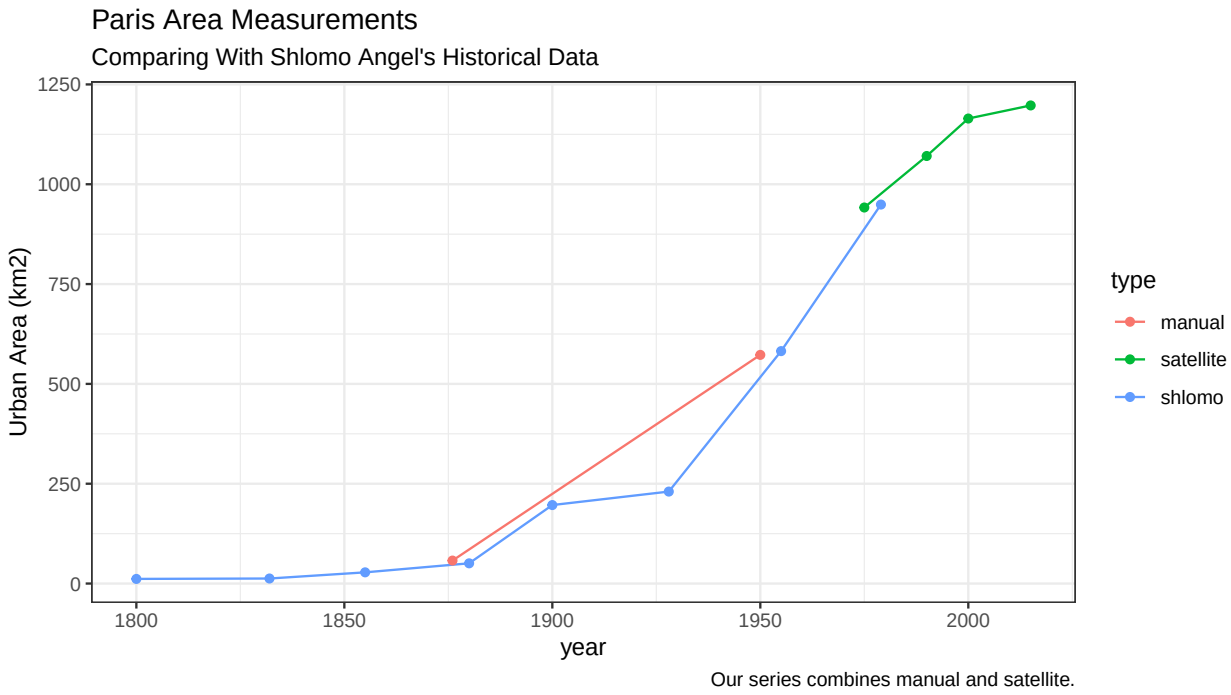


Figure A.20: Comparing our measures with Shlomo Angel's data used in Angel et al. (2012) and Angel et al. (2010). We are reassured that our manual measurement exercise aligns closely with what they obtained. Also, their final data point is reassuringly close to our first satellite measure.

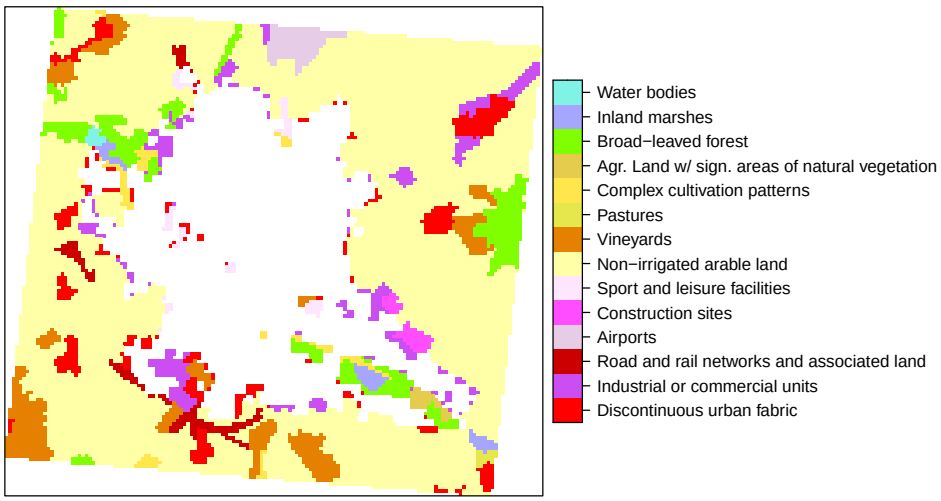


Figure A.21: Land use measures from CLC data for Reims. The white area represents our definition of the Reims Urban area in the last GHLS periods (2015), hence it is our definition of *inside vs outside* of the city. For instance, the red areas labelled *discontinuous urban fabric* are not part of our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

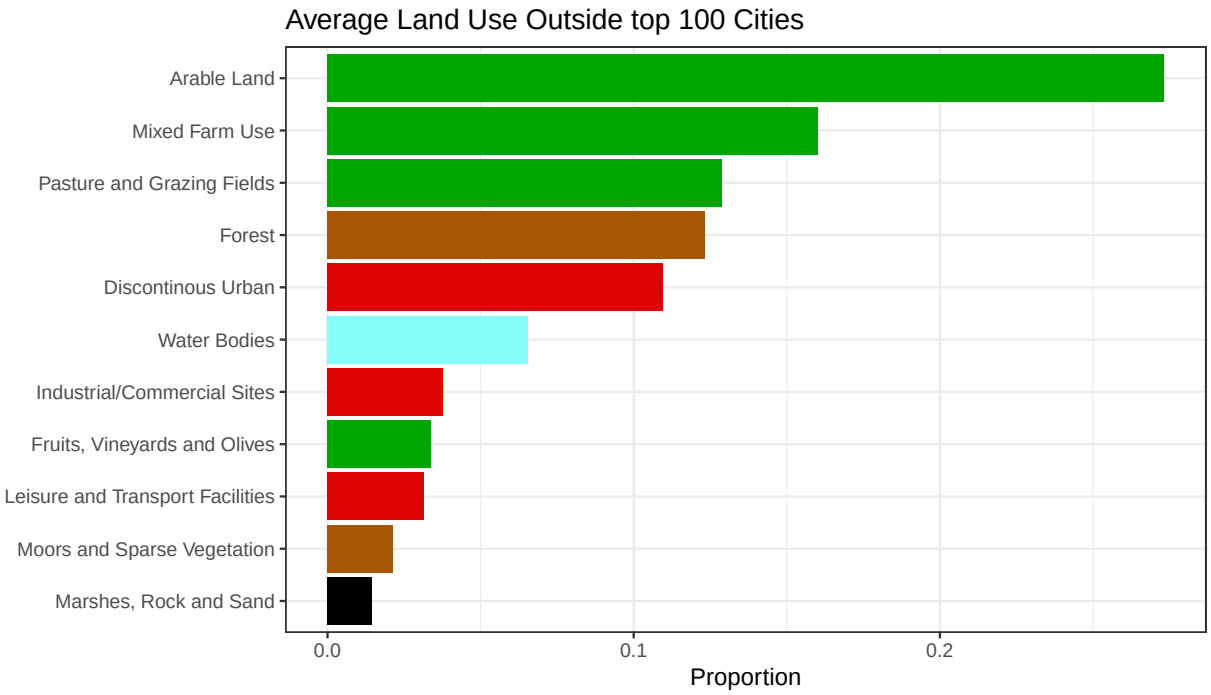


Figure A.22: Average land use measure from CLC data for our sample of top 100 French cities. This plot uses our own aggregation from 45 CLC labels into 11 exhaustive classes. We group all categories corresponding to agriculture into green bars.

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.

[20] The 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.

[21] See 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.

[22] In 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).

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

[24] A 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.

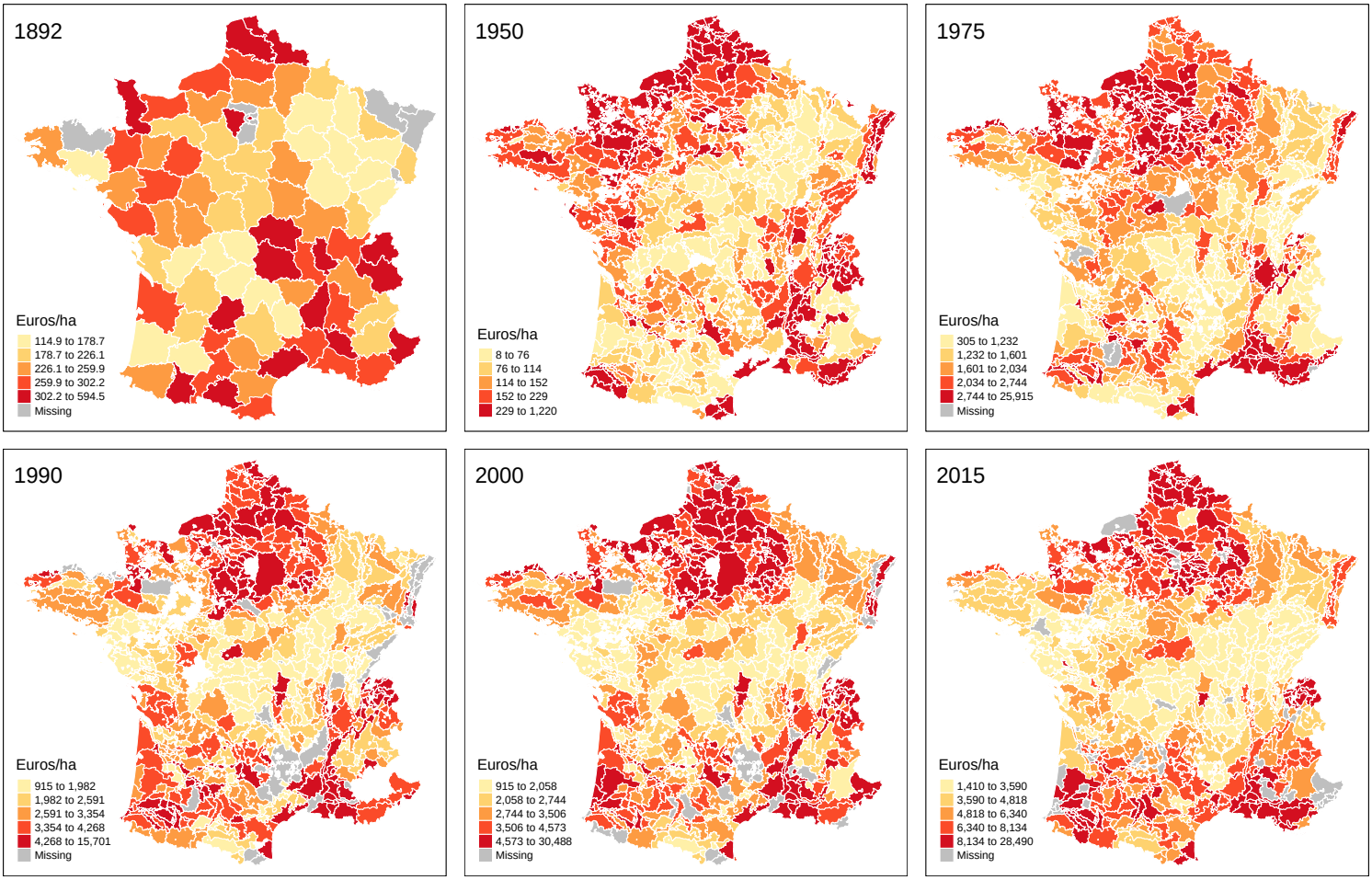


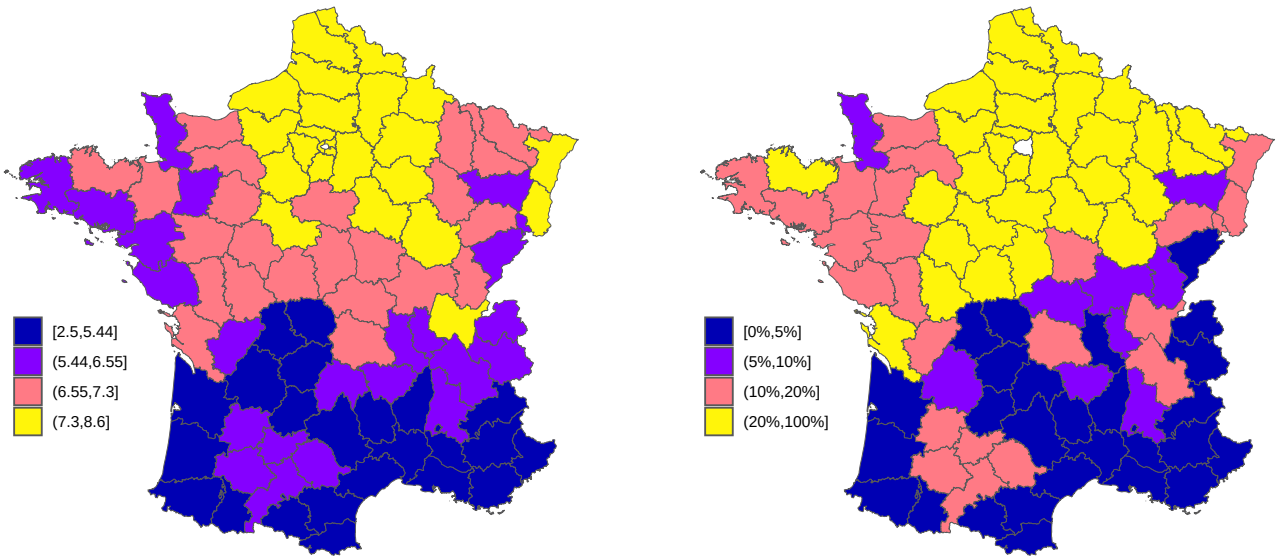
Figure A.23: Arable land value value per hectare.

Notes: Arable land value (in francs/ha) 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 Schauberger 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 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



(a) Wheat yields (2000).

(b) Share of land used for wheat (2000).

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

Notes: The left-panels 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.

[25] We 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 alike 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.

[26] The '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 GHSL 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. 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.

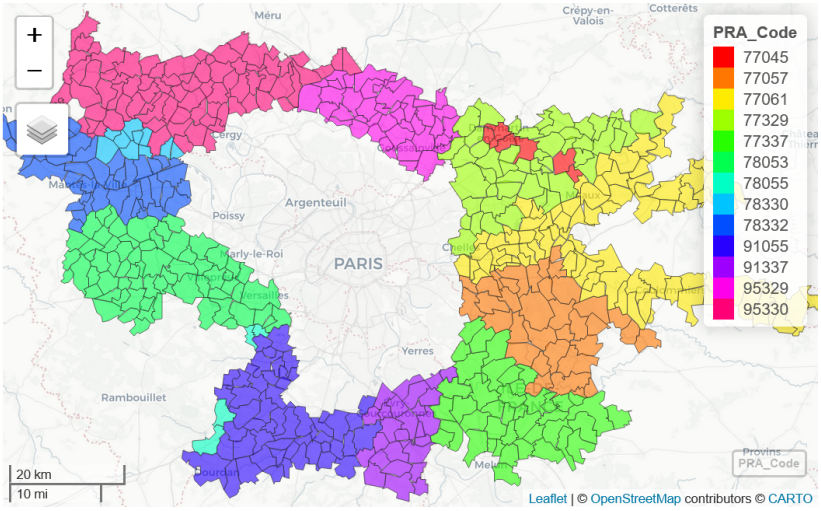


Figure A.25: Petites Régions Agricoles (PRAs) around the Parisian Urban Area.

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 PRA[27] but 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.[28] The

[27]For 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.

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

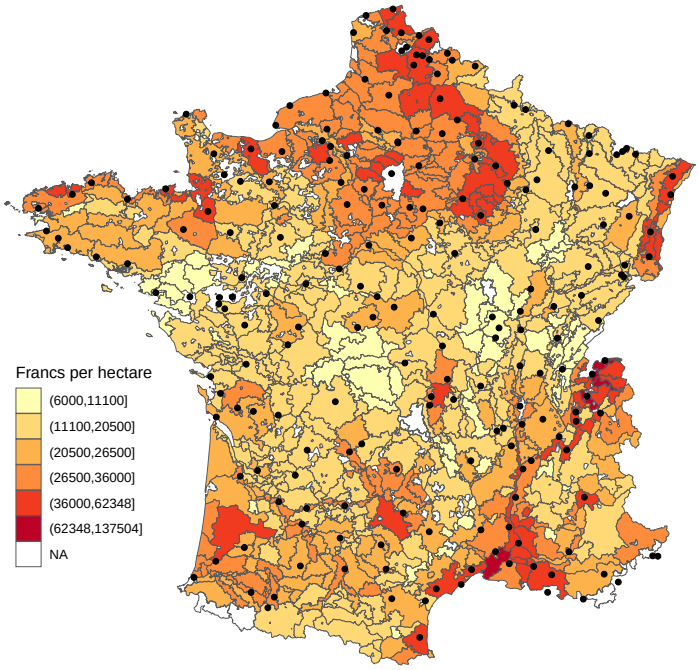


Figure A.26: Arable land value value per hectare (2000).

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.

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,

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

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

(Ceinture de Paris, zone maraichère de Lyon, Région urbaine et maraichère de Nantes) since these PRAs became urban later.

for urban wage (productivity) in city k.[29] 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		
$\log \overline{\rho}_{r,k,t}$	0.134***	0.126***	0.0877**	0.407***	0.346***	0.295**
	(0.033)	(0.026)	(0.038)	(0.103)	(0.101)	(0.140)
Num.Obs.	797	766	766	332	322	322
R2	0.237	0.253	0.312	0.310	0.323	0.432
Controls	/	$w_{u,k,t}$	$w_{u,k,t}$	/	$w_{u,k,t}$	$w_{u,k,t}$
FE: year	X	X	X	X	X	X
FE: region			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 culture in 2000 ($(S_{r,wheat}/S_r)_{k,t} > 20\%$). Controls are urban wages, $w_{u,k,t}$, in city k (Column (2) and (5)) together with Regional dummies (Column (3) and (6)). Standard errors are clustered at the département level.

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 size/density of cities.[30] Note that the bias could go either way: while local amenities might increase

[29] We 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. However, as expected, the estimated coefficient b is found to be smaller.

[30] It 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.

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 modelling 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)—83 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,

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

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%. Results are robust across specifications. While the coefficient is less significant once we control for region fixed-effects, this is not 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 benefit agricultural productivity in the surrounding département. While one cannot exclude other counfounding factors, we do not find any significant relationship between city size or wages and wheat yields in the département. The validity of our IV-strategy based on a selected sample is also threatened in presence of heterogenous effects. Performing sensitivity analysis on the subsample of wheat producers partly addresses this concern.

Sensitivity. 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

	$\log \bar{\rho}_{r,k,t}$		
$\log \text{Yield}_{k,t}$	1.577***	1.580***	1.547***
	(0.316)	(0.314)	(0.340)
Num.Obs.	332	322	322
R2	0.717	0.728	0.804
Controls		$w_{u,k,t}$	$w_{u,k,t}$
FE: year	X	X	X
FE: region			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 83 cities in départements with wheat as one of the main culture in 2000 ($(S_{r,wheat}/S_r)_{k,t} > 20\%$). Controls are urban wages, $w_{u,k,t}$, in city k (Column (2)) together with Regional dummies (Column (3)). Standard errors are clustered at the département level.

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. For a threshold stricly above 31% (twice the mean across départements), the sample of cities becomes very small (less than 30 cities almost all in the same Northern region 'Picardie') and the second-stage looses 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).

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.

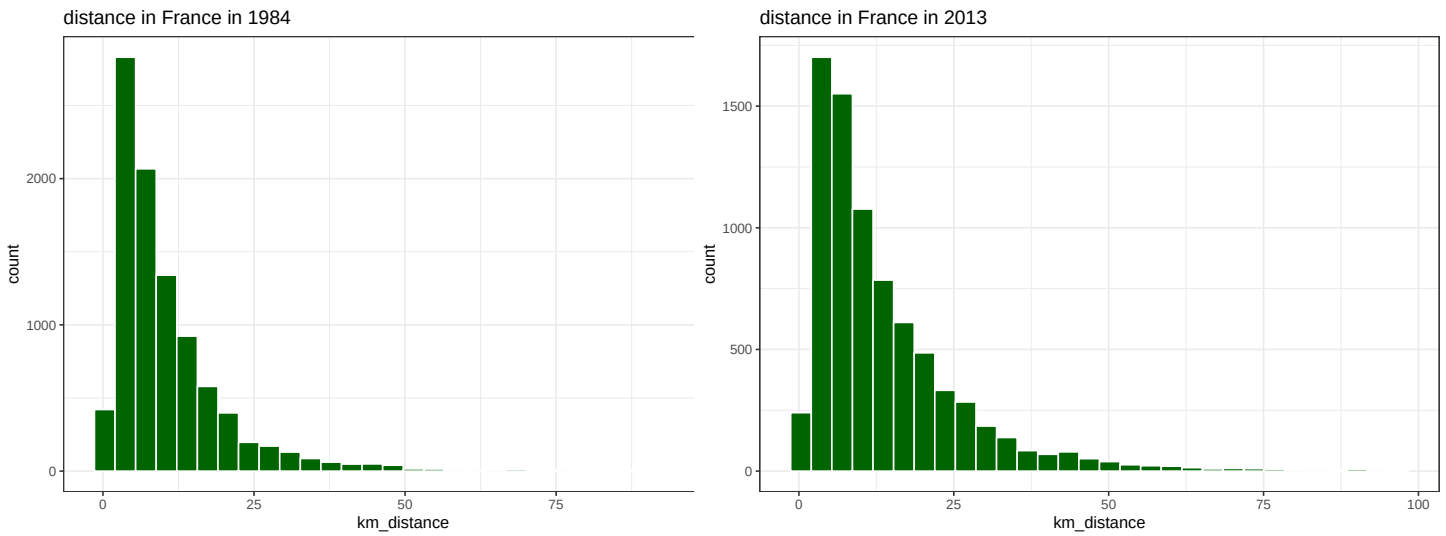
A.5.1 Individual Commuting Data from ENL

Data 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 behaviour, mode of commute, and importantly, duration of commute in minutes.

We use the waves 1984 (sample size $n = 9433$), 1988 ($n = 8910$), 2006 ($n = 12390$) and 2013 ($n = 7860$) where all required measures are observed. We subset the data to 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 and log of commuting distance for the years 1984 and 2013. For each ENL wave (1984, 1988, 2006, 2013), we perform the following



(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.

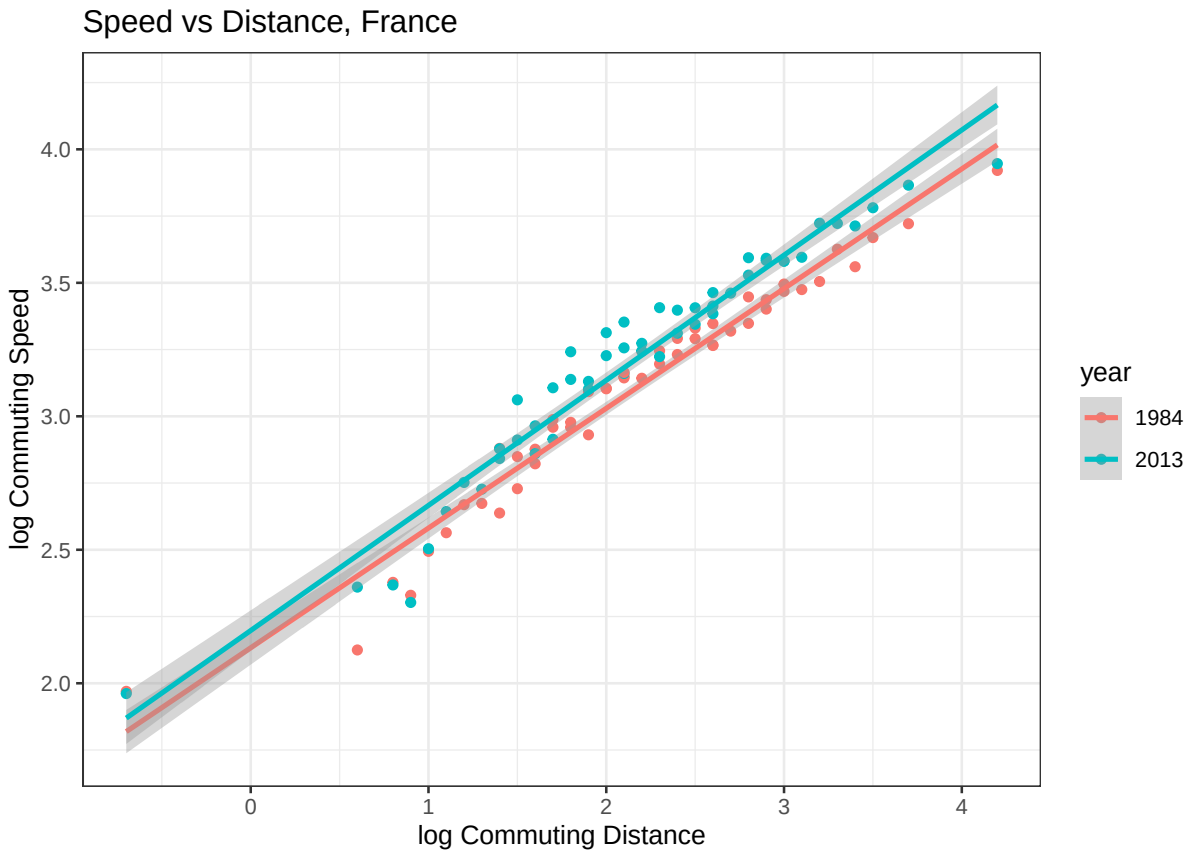


Figure A.28: Commuting speed and commuting distance (1984 and 2013).

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

regression at the individual level,

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

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.5 and A.6 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 ***	0.505 ***	0.498 ***	0.502 ***	0.470 ***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
r.squared	0.357	0.357	0.372	0.376	0.549
nobs	9199	9199	9189	9199	9189

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

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

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

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

Table A.6: Cross sectional regression of Speed on Commuting Distance using ENL 2013 data. Columns are specified as in table A.5.

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,

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

	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.7: 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.

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

Results are reported in Table A.7 (see Figure A.28 for the graphical representation). We use the 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.3.4 (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,

$$d_k(\ell_k) = d_0(\phi_k) + d_1(\phi_k) \cdot \ell_k \tag{A.6}$$

where $d_0(\phi)$ and $d_1(\phi)$ are parametric functions of the city radius ϕ as detailed in extension B.3.4—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`.[31] This reduces the sample to 60 million records. From this, we extract a 50% random sample. Next we obtain all unique pairs of residence and workplace communes (variables `COMR` and `COMT`) and compute straight-line distance for each pair. Then we add the distance of each commune to the center of their urban area. The urban area classification is officially given by INSEE and we use the AU2010 (Aire Urbaine 2010) classification. We end up with 18 million observations.

We aim to investigate how commuting distance varies with the distance from the center of the residence 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 ($\texttt{distance\_commute}_i$) and the residential distance from the city center ($\texttt{distance\_center}_i$) for each individual i in city k, we perform the following regression,

$$\texttt{distance\_commute}_i = \gamma_{0,k(i)} + \gamma_{1,k(i)} \cdot \texttt{distance\_center}_i + u_i \tag{A.7}$$

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.6), where commuting distance $d(\ell)$ increases less than proportionately with residential location ℓ, and less so for larger cities (larger

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

radius ϕ_k). We use these findings in Section B.3.4 to parametrize $d_0(\phi)$ and $d_1(\phi)$.

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 slection. 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.8. 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.8: 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 $w_{u,k,t}$.

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

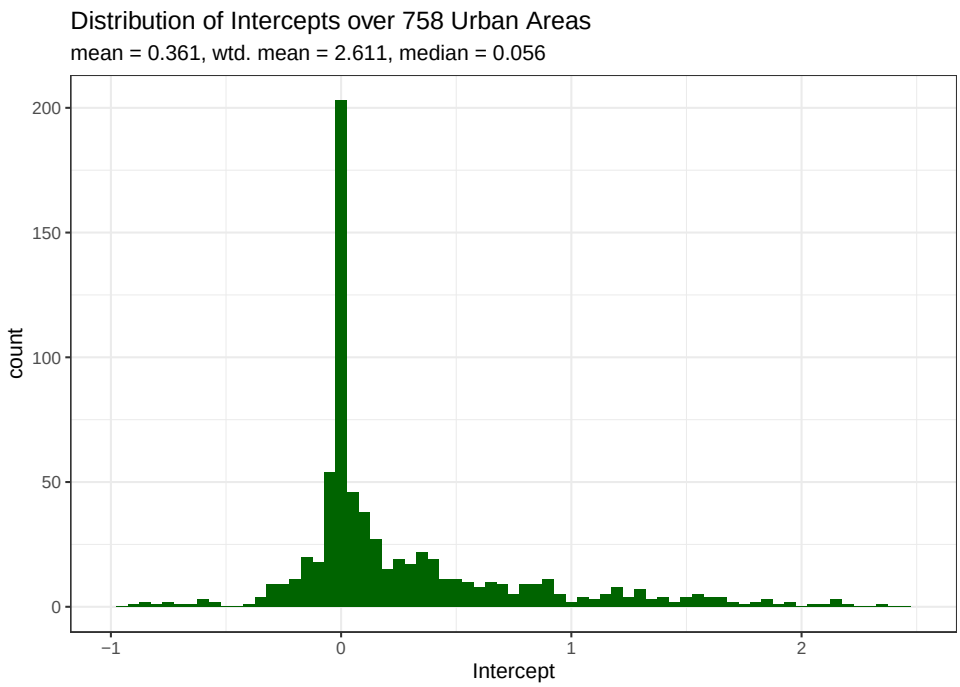


Figure A.29: Distribution of DADS intercept estimates

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.

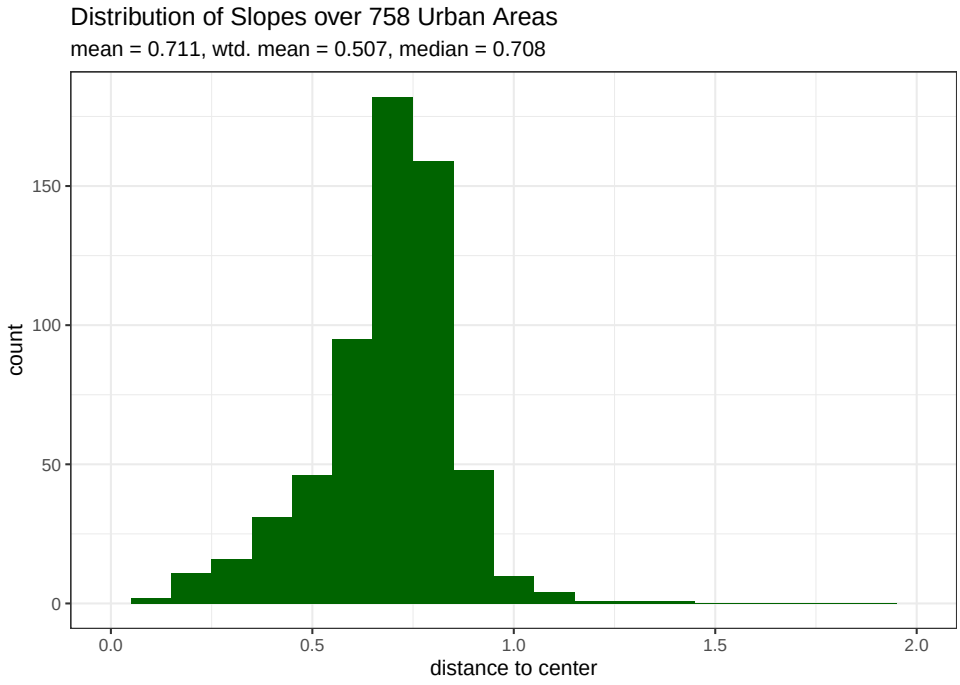


Figure A.30: Distribution of DADS slope estimates

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.

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 14a)). To do so, we use survey data (individual commuting data) in the Parisian urban area for the post-WW2 period. These data give the main mode used for working trips and well as the corresponding speed. Pre-WW2 (1840-1940), such individual surveys are not available. However, historical data on traffic by public transport modes and on registered private vehicles helps us to build estimates of the distribution of mode use over the whole period. Given estimates of the speed of each transportation mode, one can back out historical estimates of the average commuting speed.

Two main caveats are in order. First, the strategy developed only provide *estimates* since 1840 of the average commuting speed. These estimates depend on assumptions to convert historical data on traffic and registered vehicles into their modal use for work commutes and on assumptions regarding the speeds of the various modes. While some measurement error is unavoidable, our estimates provide a reasonable order of magnitude of the historical evolution of commuting speed in Paris. Second, due to historical data availability, we must focus on the Parisian urban area rather than France as a whole. Paris is arguably special. In the recent period, public transport is more widely used in Paris.[32] Paris might also be more congested than other French cities. Overall, one needs to be cautious with our estimates. However, it is clearly reassuring that estimates for Paris and model's predictions give very similar order of magnitude since the former were not targeted in the calibration.

Commuting data post-WW2. The first survey on commuting for work in the Parisian urban area was conducted in 1959 (on a representative sample of more than 20,000 individuals). While the original data are not available, secondary sources provide a detailed summary of the results (see Bertrand and Hallaire (1962)). For our purpose, this gives us the distribution of mode use in Parisian area in 1959. The majority of Parisian workers (50.2%) were using public transport (ventilated between metro, autobus and train); 21.5% were using a private mean of transportation (8.5% a private car, the rest for the most part a bicycle or a motorbike); the remaining 28.3% are walking.[33] The 1959 data do not provide the speed of each mode and we impute the speed measured in the later survey (1976) to compute the average commuting speed in the Parisian area in 1959. We use the 'Enquéte Global Transport (EGT)' for the years 1976, 1983, 1991, 2001 and 2010. The EGT provides individual commuting data for a representative sample of the Parisian urban area: distance of commuting trips, time, speed and modal use. We restrict our attention to trips to the work location to extract the distribution of mode use and their respective speeds to compute the

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

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

average commuting speed.[34] 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.[35] 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.[36] The main issue

[34] The sample raw average commuting speed at each date gives very similar estimates.

[35] The horse-drawn omnibus disappears in 1913.

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

arise since many of these measured trips are not made to commute to work but for other reasons (leisure, shopping, ...). Assuming that a fraction $x_{m,t} \in (0,1)$ of these trips are work commutes. By definition, the proportion of workers using mode m to commute to work, $p_{m,t}$, is the number of (two-way) working trips per worker (per working day) using mode m,

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

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

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

and $x_t = \frac{p_t}{\sum_m N_{m,t}}$ can be easily recover from the data for the years post-WW2—using measures of p_t in individual surveys and values for $(\sum_m N_{m,t})$ from traffic data. It is close to 1/3, relatively stable across years. Using EGT data which provides the motive for registered trips, 31% of non-walking trips in 1976 were between home and work. Such a value implies about 50% of people using public transport in 1955, in line with the corresponding survey data. Thus, prior to WW2, we set x to $\hat{x} = 31\%$.[37] 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 motorbikes.[38] 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.[39] We also collected data

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

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

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

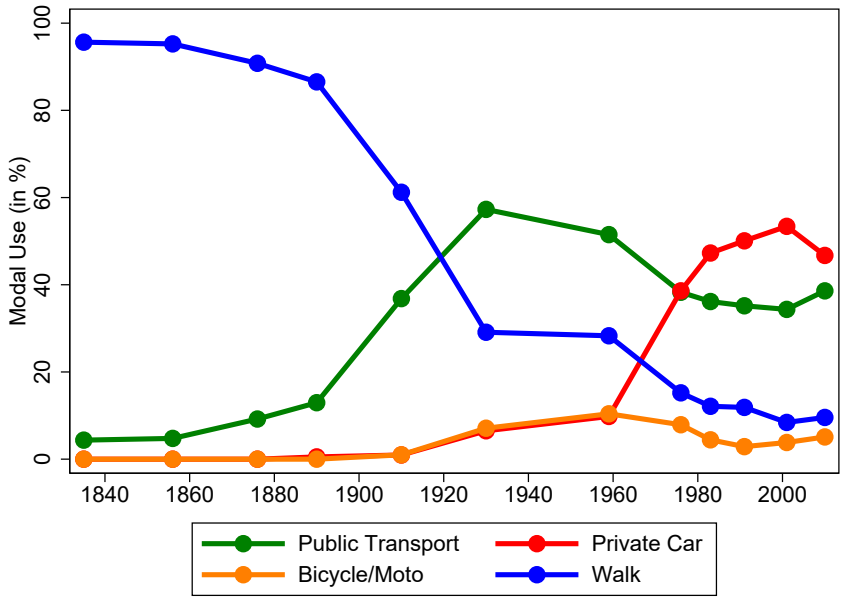


Figure A.31: Transportation mode use in the Parisian urban area.

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 inexistent prior to 1890. The number of bikes in Paris is estimated to about 60 000 in 1891, 250 000 in 1901 and 285 000 in 1912 (Orselli (2008)). Unfortunately, such data are not available at a later dates and not readily available for motorbikes for the Parisian area.[40] Given the importance of bicycles for leisure and the lack of relevant data post-

[40] 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, on can assess the extent of bicycle/motorbike use relative to other means for 1930. Papon et al. (2010) provides such estimates by decades—reweighting observations to control for sample attrition due to survival: in 1930-1940, 9.9% of the population were using the bicycle as main mode of transportation in France, versus 2.3% for the 1920-1930 decade. We take the average between these values, 6.1%.[41] 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%.[42] 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

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

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

metro speed (about 8.5 kms per hour). This is the value taken until 1890. Post-1890, we set the speed of omnibus to 7.5 kms per hour as a significant share of those were motorized. For tramways, we set the speed to 7.5 kms per hour when horse-drawn in 1876 and 8.5 kms per hour when fully electric in 1910. We use the average between these two values for 1890 since both were used. Boats were on average faster than ground transportation modes. We set their speed to 10 kms per hour but results are barely affected by this value within a reasonable range given that less than 1% of the Parisian population were using this mode when available. Lastly, we wet the speed of private horse-drawn cars to 8 kms per hour. Like for boats, results are barely sensitive to this value as this mode of transport for work commute was the privilege of few rich Parisians in the late nineteenth century. Figure 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.

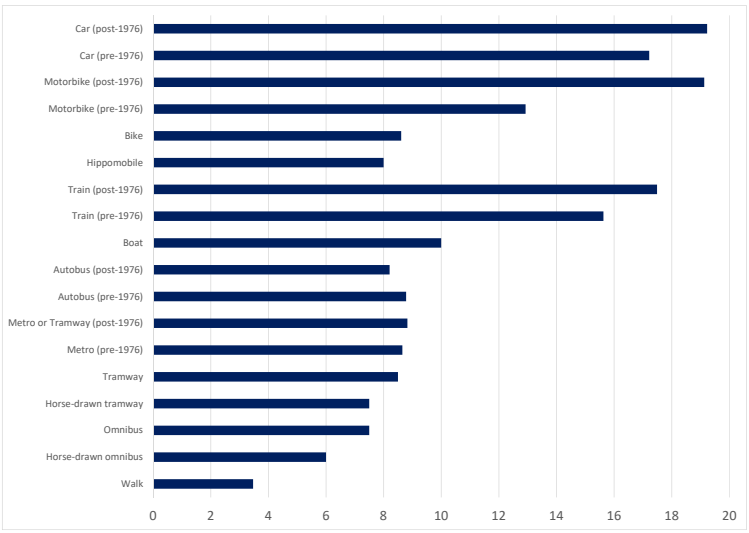


Figure A.32: Speed across transportation modes.

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.

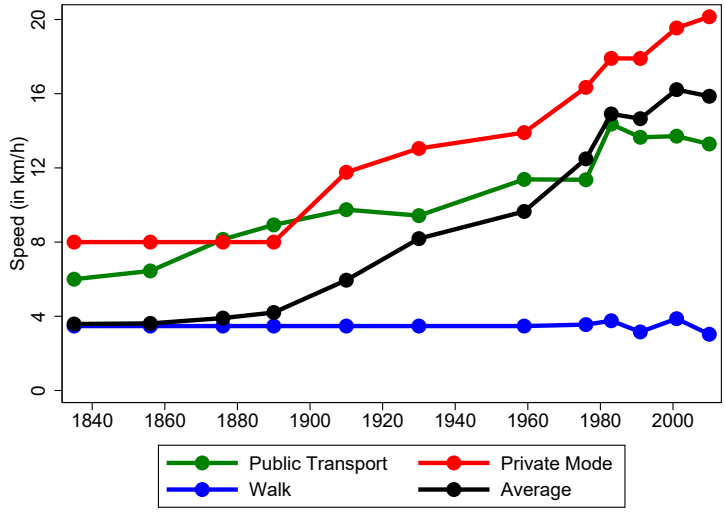


Figure A.33: Evolution of average speed across mode categories.

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 Extracting Land Use from Historical Maps Using Machine Learning: The Emergence and Disappearance of Cities in France," *mimeo*, 2021.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 multitemporal (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.