Structural Change, Land Use and Urban Expansion
Supplementary Online Appendix C

Nicolas Coeurdacier
SciencesPo Paris, CEPR

Florian Oswald
Uni Turin ESOMAS,
Collegio Carlo Alberto

Marc Teignier
Serra Húnter Fellow,
University of Barcelona

November 7, 2025

Contents

C.1 References for this Appendix

This is a separate supplementary online appendix to our paper. It is hosted exclusively on our personal websites. The references to main text, and appendices A and B refer to the published paper in REStud and the associated online appendix. Accordingly, references to main text are plain numbers, references to Appendix A are prefixed with A, same for Appendix B and Appendix C.

You can find working paper versions of all 4 documents (main text, appendix A, B and C) on our websites e.g., https://floswald.github.io/publication/landuse/.

C.2 Land Use Around Cities

Using GHSL data on urban settlements only, one cannot check that cities are gaining on farmland when expanding their area. However, using recent satellite data, one can check if French cities are mostly surrounded by agricultural land. 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 C.1 and the resulting average in Figure C.2.

C.3 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 C.3.1 and C.3.2). 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 C.3.3). Cities are denoted by the index k, individuals by i and dates by t.

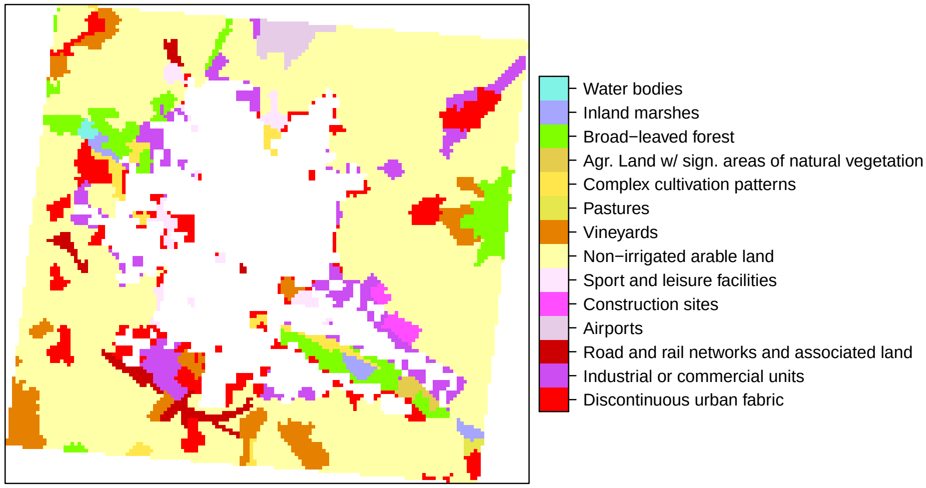


Figure C.1: Land use measures from CLC data for Reims. The white area represents our definition of the Reims Urban area in the last GHLS periods (2015), hence it is our definition of *inside vs outside* of the city. For instance, the red areas labelled *discontinuous urban fabric* are not part of our definition of the city.

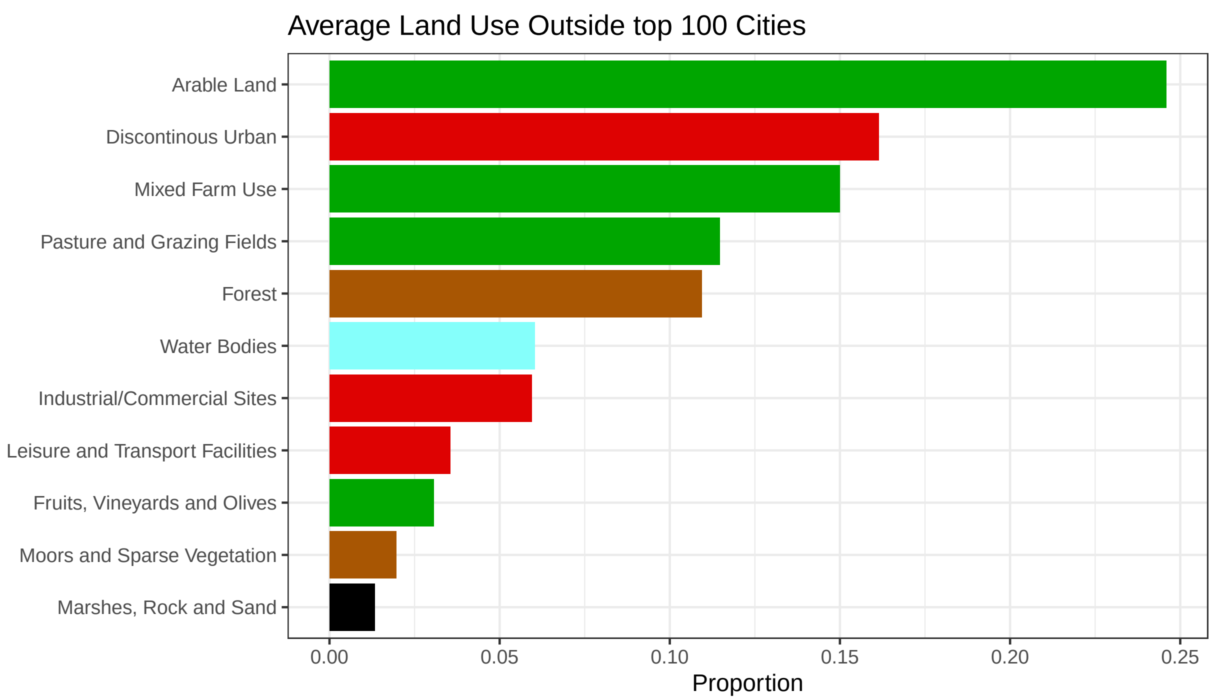


Figure C.2: Average land use measure from CLC data for our sample of top 100 French cities. This plot uses our own aggregation from 45 CLC labels into 11 exhaustive classes. We group all categories corresponding to agriculture into green bars.

C.3.1 Individual Commuting Data from ENL

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

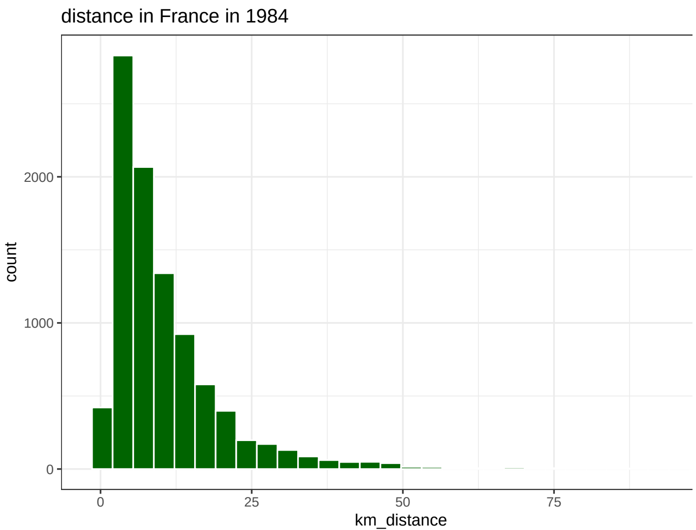
We use the waves 1984 (sample size $n = 9433$), 1988 ($n = 8910$), 2006 ($n = 12390$) and 2013 ($n = 7860$) where all required measures are observed. We subset the data to individuals working outside their home and to be the reference person in the household. We observe workplace and residence at the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

Figures C.3a and C.3b 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.

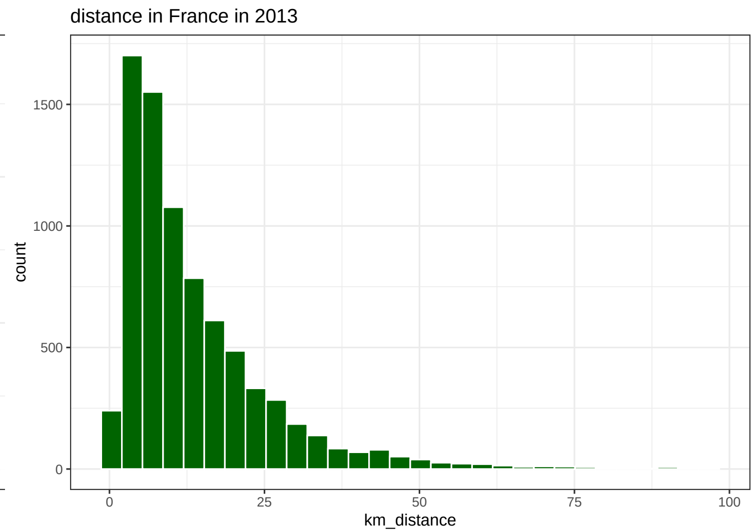
Elasticity of speed w.r.t commuting distance. We are interested in measuring the elasticity of speed w.r.t commuting distance in a given year. Grouping data into 50 bins of log distance, Figure C.4 illustrates the relationship between log of speed and log of commuting distance for the years 1984 and 2013. For each ENL wave (1984, 1988, 2006, 2013), we perform the following regression at the individual level,

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

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 C.1 and C.2 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



(a) Distribution of Commuting Distances in 1984..



(b) Distribution of Commuting Distances in 2013.

Figure C.3: 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.

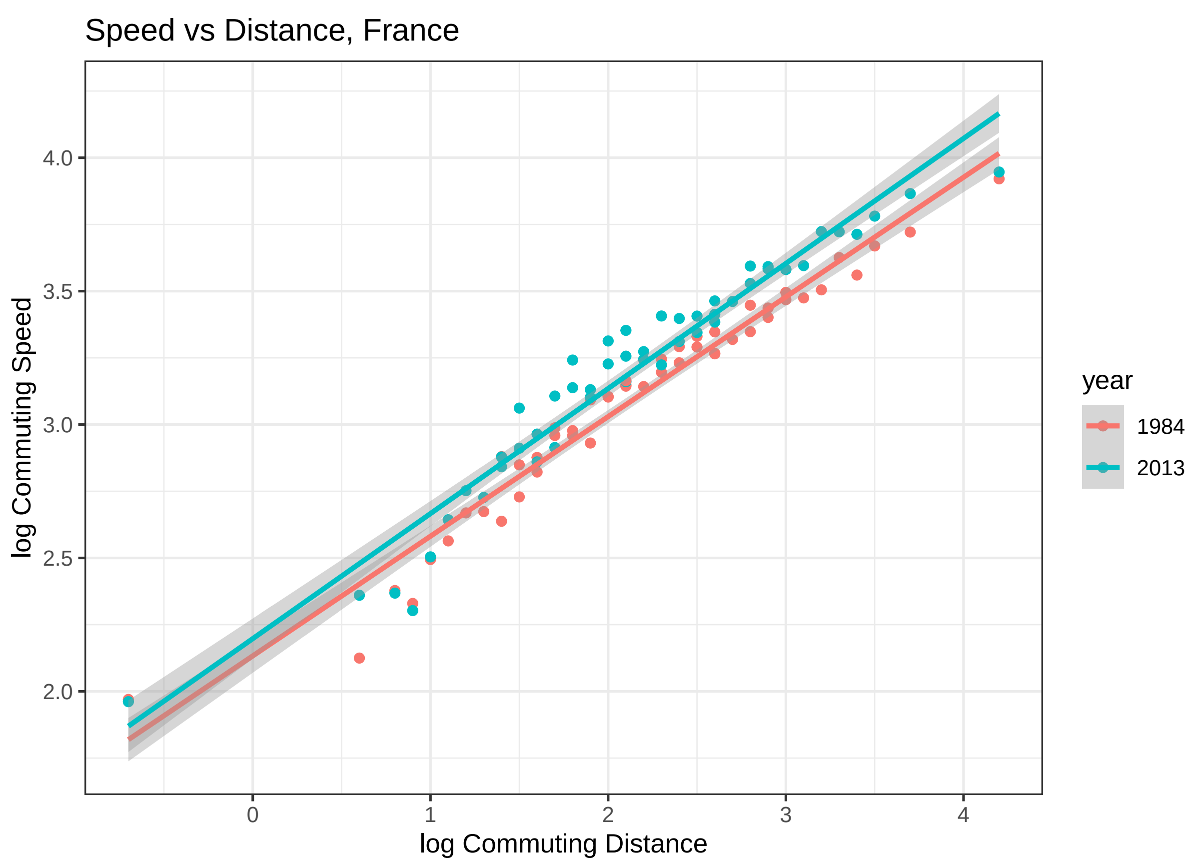


Figure C.4: Commuting speed and commuting distance (1984 and 2013).

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

$1 - 0.45 = 0.55$ 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 C.1: 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 C.2: Cross sectional regression of Speed on Commuting Distance using ENL 2013 data. Columns are specified as in table C.1.

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

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

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

Results are reported in Table C.3 (see Figure C.4 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.

	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 C.3: 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 C.4.

C.3.2 Individual Commuting Data from DADS

Data from Déclaration annuelle des données sociales (DADS). We make use of confidential access to the DADS "Tous Salariés" (DADS-DSN) dataset for 2018 in order to investigate how commuting distance vary with residential location conditionally on city size in a large sample of the population. The DADS-DSN dataset contains all salaried workers in France, both private and public sector and the large sample size allows to study the link between commuting distance and residential location at the city-level—the ENL sample being too small.

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

$$d_k(\ell_k) = d_0(\phi_k) + d_1(\phi_k) \cdot \ell_k \tag{C.1}$$

where $d_0(\phi)$ and $d_1(\phi)$ are parametric functions of the city radius ϕ as detailed in extension B.3.5—with $d_0(\phi)$ increasing in ϕ and positive and $d_1(\phi)$ decreasing in ϕ and between 0 and 1. Data on residential and work locations are necessary to validate our reduced-from approach and discipline the calibration of $d_0(\phi)$ and $d_1(\phi)$.

We start by reading the full dataset with 62 million records. We drop records which are in overseas territory, or which have as a residence or workplace identifier the code `75056`.[1] 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

[1] 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.

2010) classification. We end up with 18 million observations.

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

$$\texttt{distance\_commute}_i = \gamma_{0,k(i)} + \gamma_{1,k(i)} \cdot \texttt{distance\_center}_i + u_i \tag{C.2}$$

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 million, between 1 and 3 million, between 50 000 and 1 million, ...).

Figure C.5a 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 C.5b 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.

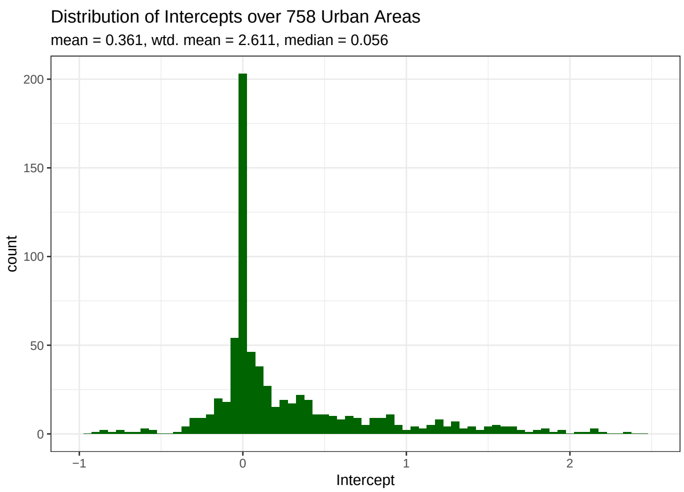
We also inspect the value of the estimates as a function of the size of the city. The intercept $\gamma_{0,k(i)}$ increases with city size, from about 0.2 km for the smallest urban areas to more than 4 kms for Paris. The slope coefficient $\gamma_{1,k(i)}$ decreases with city size—ranging from around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

These results validate our reduced-form parametrization (Eq. C.1), where commuting distance $d(\ell)$ increases less than proportionately with residential location ℓ, and less so for larger cities (larger radius ϕ_k). We use these findings in Section B.3.5 to parametrize $d_0(\phi)$ and $d_1(\phi)$.

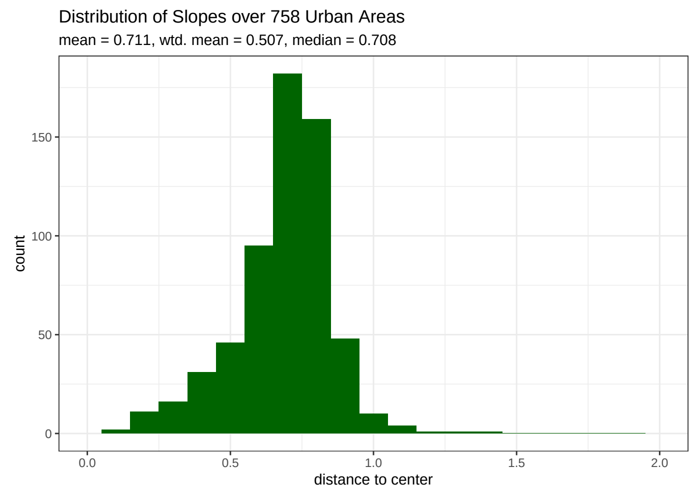
C.3.3 Urban Productivity and Wages

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

Sample selection. As we do not observe hours worked, we first implement a procedure on the



(a) Distribution of DADS city-specific intercept estimates.



(b) Distribution of DADS city-specific slope estimates.

Figure C.5: Distribution of DADS city-specific estimates.

Notes: *City* is defined as *Aire Urbaine (AU)* by INSEE. Results from individual level regression of commuting distance on distance from city center using DADS.

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 C.4. 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 C.4: DADS Panel 2019-EDP Subsetting Procedure

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

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

Remarks. To estimate the average urban productivity of a given city, we would like to control for the composition of the workforce across cities and compute an urban area fixed effect for each city, controlling for various worker-level observables (education and age). Unfortunately, the sample size is too small in the earlier years to reliably compute a fixed effect for each urban area. For the year 2015, the sample is significantly larger and we are able to estimate city fixed-effects when controlling for observables (age and education). We find that our raw measure, the log of unconditional mean of net salaries across workers, yields a measure very highly correlated with city fixed-effects (correlation of 0.76). This is reassuring that our raw measure of $w_{u,k,t}$ is a reasonable proxy for a cities fixed effects (e.g. the city-specific urban productivity). We also expect a positive relationship in the city-wide average net salary and the population of the urban area. For each year t, we regress the log of the measured average wage, $\log w_{u,k,t}$, on the log of the population of the urban area, $\log L_{u,k,t}$. We find a highly significant positive relationship, robust across all years.

C.4 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 11a)). To do so, we use survey data (individual commuting data) in the Parisian urban area for the post-WW2 period. These data give the main mode used for working trips as well as the corresponding speed. Pre-WW2 (1840-1940), such individual surveys are not available. However, historical data on traffic by public transport modes and on registered private vehicles helps us to build estimates of the distribution of mode use over the whole period. Given estimates of the speed of each transportation mode, one can back out historical estimates of the average commuting speed.

Two main caveats are in order. First, the strategy developed only provides *estimates* since 1840 of the average commuting speed. These estimates depend on assumptions to convert historical data on traffic and registered vehicles into their modal use for work commutes and on assumptions regarding the speeds of the various modes. While some measurement error is unavoidable, our estimates provide a reasonable order of magnitude of the historical evolution of commuting speed in Paris. Second, due to historical data availability, we must focus on the Parisian urban area rather than France as a whole. Paris is arguably special. In the recent period, public transport is more widely used in Paris.[2] Paris might also be more congested than other French cities. Overall, one needs to be cautious with our estimates. However, it is clearly reassuring that estimates for Paris and model's predictions give very similar order of magnitude since the former were not targeted in the calibration.

Commuting data post-WW2. The first survey on commuting for work in the Parisian urban area was conducted in 1959 (on a representative sample of more than 20,000 individuals). While the original data are not available, secondary sources provide a detailed summary of the results (see Bertrand and Hallaire (1962)). For our purpose, this gives us the distribution of mode use in Parisian area in 1959. The majority of Parisian workers (50.2%) were using public transport (distributed between metro, autobus and train); 21.5% were using a private mean of transportation (8.5% a private car, the rest for the most part a bicycle or a motorbike); the remaining 28.3% are walking.[3] 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

[2] 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.

[3] 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.

the average commuting speed.[4] 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 C.7a).

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.[5] 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, 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 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.[6] The main issue arise since many

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

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

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

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

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

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

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

and $x_t = \frac{p_t}{\sum_m N_{m,t}}$ can be easily recovered from the data for the years post-WW2—using measures of p_t in individual surveys and values for $(\sum_m N_{m,t})$ from traffic data. It is close to 1/3, relatively stable across years. Using EGT data which provides the motive for registered trips, 31% of non-walking trips in 1976 were between home and work. Such a value implies about 50% of people using public transport in 1955, in line with the corresponding survey data. Thus, prior to WW2, we set x to $\hat{x} = 31\%$.[7] 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 C.6, 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.[8] 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.[9] We also collected data

[7] 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.

[8] 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.

[9] 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.

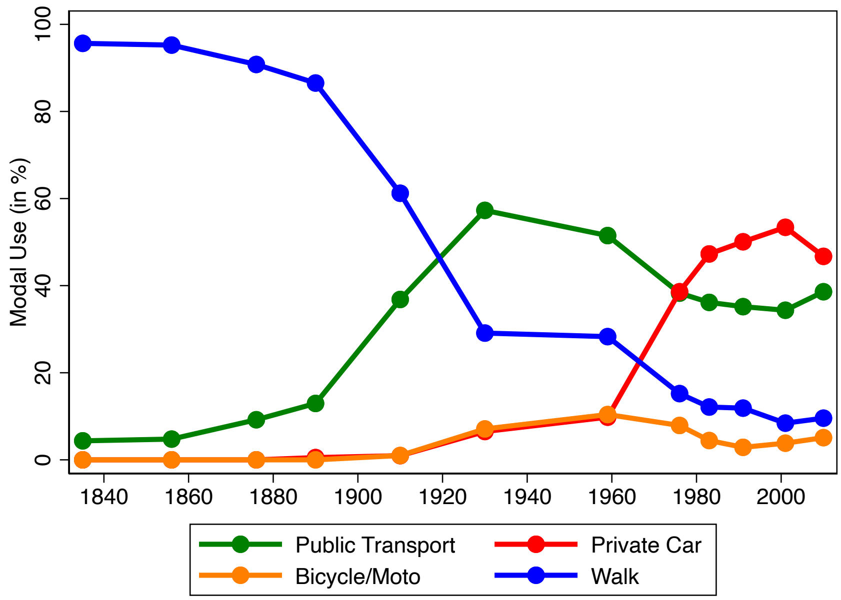


Figure C.6: Transportation mode use in the Parisian urban area.
Notes: Fraction of workers using the respective transportation mode over the period 1835-2010, in %. *Sources*: Data from secondary sources for the dates prior to WW2 (mostly traffic of the different public modes and registered private vehicles 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 C.6 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.[10] Given the importance of bicycles for leisure and the lack of relevant data post-

[10] 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%.[11] 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%.[12] 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 C.6 summarizes the estimates for the share of workers using bikes/motorbikes over the whole period.

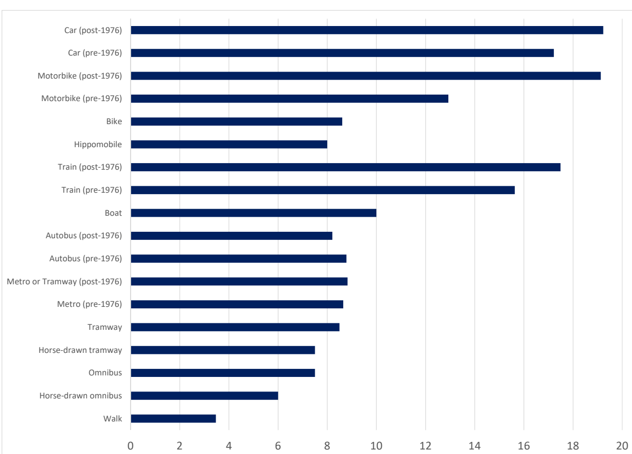
Walking. The share of workers walking to their work location is estimated as a residual—made of workers using neither a public transportation nor a private one. Figure C.6 summarizes the estimates for the share of workers walking to work over the whole period. In the early years, before 1840, Paris is 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

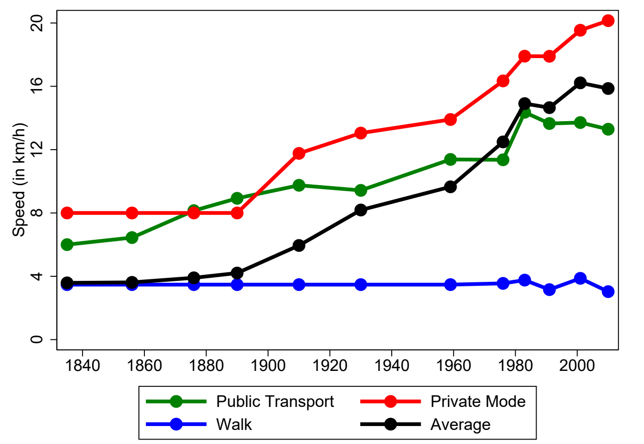
[11] For the following decades, 13% of people using bikes in 1940-1950, 13% in 1950-1960, 9.7% in 1960-1970—broadly in line with survey data for Paris available at the latest periods.

[12] 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 C.7a summarizes the estimated speed of the different modes, by mode at different dates. Figures C.7b 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. When comparing to the model, the average commuting speed (black line) is normalized to unity in 1835 (1840 for model comparison).



(a) Speed across transportation modes.



(b) Evolution of average speed across mode categories.

Figure C.7: Evolution of speed across transportation modes.

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

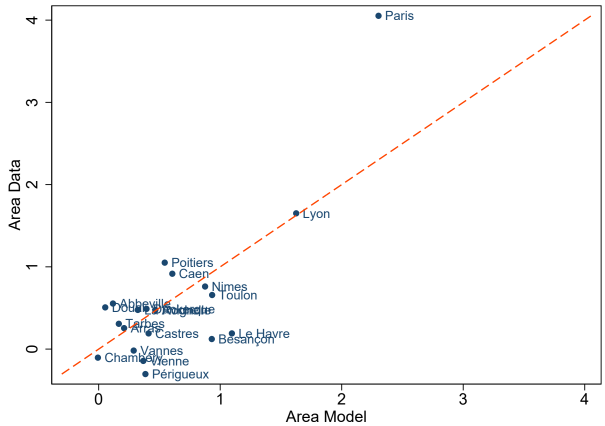
C.5 Additional Untargeted Model Cross-Sectional Outputs

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

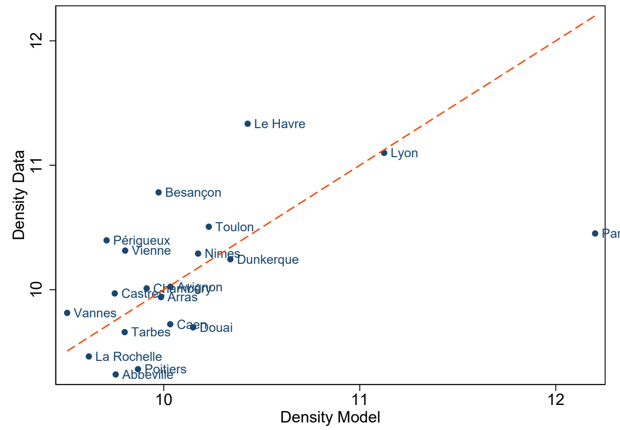
Urban Area and Density. In the cross-section at each date, more populated cities (i.e. more productive) are larger in area and, due to higher housing prices, more populated cities are also denser. While the model can reproduce these facts qualitatively, it does not match them quantitatively as discussed in the main text and displayed in Figure C.8 in the first and last dates of observation for the restricted sample of 20 cities. The urban area in the model does not increase enough with city size compared to the data—equivalently, large cities are (relatively) too dense in the model compared to the data. Overall, the cross-sectional fit for urban density is not very good, as cross-regional heterogeneity is fairly limited in the model, restricted to heterogeneous sectoral productivities to generate a reasonable dispersion in urban population and farmland prices at the urban fringe. Many other city-specific factors possibly influence the density of individual cities in the cross-section (different natural constraints, different housing supply conditions/commuting infrastructure, different amenities, different protected areas/land use regulations in the more recent period,...) that our model cannot possibly account for.

To visualize the cross-section in the time dimension, one can bin cities into size-groups and take averages within bins—mitigating the idiosyncrasies of individual cities as well as concerns regarding their density measurement (Appendix A.2.5). With a restricted number of bins relative to the number of cities, outcomes in model and data become readable in the time-dimension. This is done in Figure C.9 in the model and in the data (for the whole initial sample of 100 cities to avoid idiosyncrasies in the random sample of 20 cities). Density is averaged by bins of size in the initial period, 1870 (above 100,000; between 50,000 and 100,000, between 25,000 and 50,000; below 25,000) and normalized by the first period (1870) median density in the sample to visualize the cross-sectional and time series variations. While the model performs well relative to the data in the time-series, qualitatively and quantitatively, the cross-sectional dispersion of densities is too large in the model at each date. Larger cities are denser, in the model and in the data, but an order of magnitude denser in the model relative to smaller ones—the problem being particularly severe for Paris. This said, this is to us not a major concern given that our main focus is the evolution in the time-series, which barely interacts with cross-sectional heterogeneity.

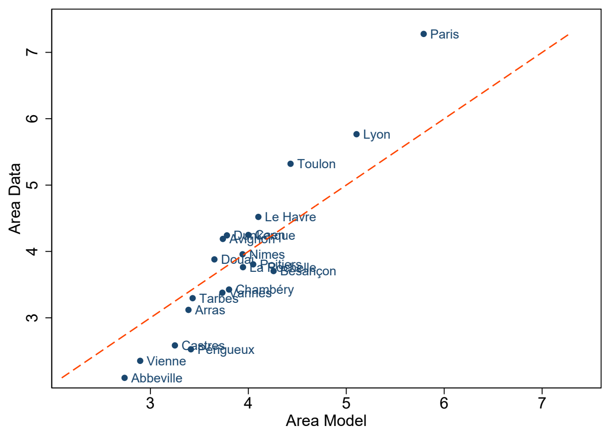
Rural Productivity and Wheat Yields. The cross-sectional dispersion in farmland prices is driven by rural productivity differences in the model. As a sanity check, we compare the model estimated cross-sectional differences in rural productivity to observed measures of rural productivity: wheat yields. In our restricted sample of 20 regions, due to specialization in crops, not all cities are in regions producing wheat (only 7 cities are in départements which devote more than 20% of their land use for wheat in 2000, 11 above 10% with the mean wheat land use in this sample below 15%). This said, we investigate the link between the estimated region-specific rural productivity and wheat yields on the sample of simulated cities in regions producing wheat (with a low (resp.



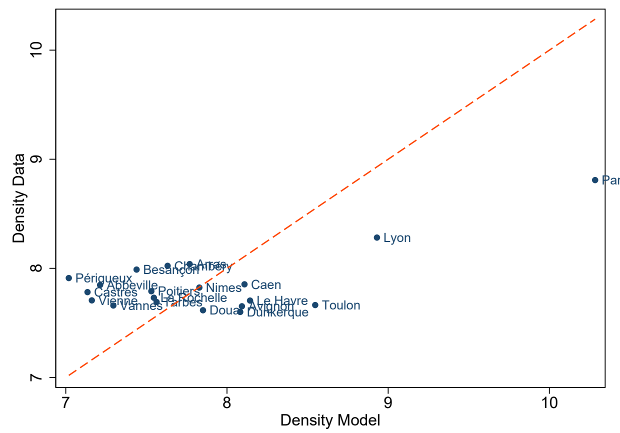
(a) Urban Area in 1870.



(b) Average urban density in 1870.



(c) Urban Area in 2015.



(d) Average urban density in 2015.

Figure C.8: Cross-Sectional Predictions: Area and Urban Density in 1870 and 2015.

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

higher) threshold of land use for wheat of at least 10% (resp. 20%) of wheat land use, 11 cities (resp. 7 cities)). To do so, we perform the following regression for readily available dates $t \in \{1975, 1990, 2000, 2015\}$ (see Appendix A.3.2 for data description on wheat yields and land use for wheat),

$$\log \text{Yield}_{k,t} = a_t + b \cdot \log \theta_{r,k,t} + u_{k,t}, \tag{C.3}$$

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

Results of regression C.3 on the restricted sample give an estimated b is very close to unity and

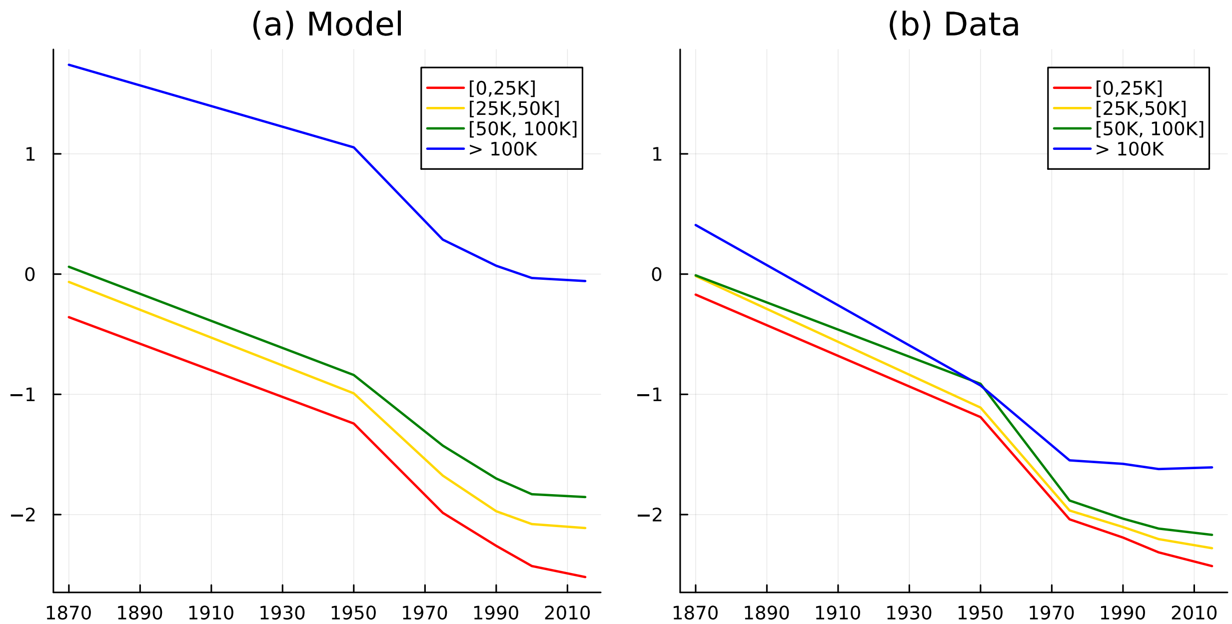


Figure C.9: Average urban density by size-bins: model versus data.

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

highly significant in both sub-samples, particularly so when focusing on locations more specialized in wheat ($b = 1.05$ with T-stat of 4.0 on the larger sample of 11 cities = 44 obs., corresponding to a low threshold of 10% of land use for wheat; $b = 0.96$ with T-stat of 8.4 on the restricted sample of 7 cities = 28 obs., corresponding to a higher threshold of 20% of land use for wheat). Alternatively, one can run regression C.3 on the whole sample of cities in the model but weighting observations by their share of land use for wheat in 2000: results are the same ($b = 1.03$ with T-stat of 5.1). This is illustrated by the scatter plot of Figure C.10 for the sample cities of wheat producers in 2000 (land use for wheat above 10%).

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

Urban Productivity and Wages. Similarly to our attempt to compare the model implied rural productivity to wheat yields, one can compare model implied urban productivity/wage, $\theta_{u,k,t}$, to urban wages in the data for the sample of 20 cities in our simulations. This is a useful sanity check to see if the dispersion of wages across cities implied by the model is broadly in line with the data. Data are readily available from DADS for years $t \in \{1975, 1990, 2000, 2015\}$ (see Appendix C.3.3). We

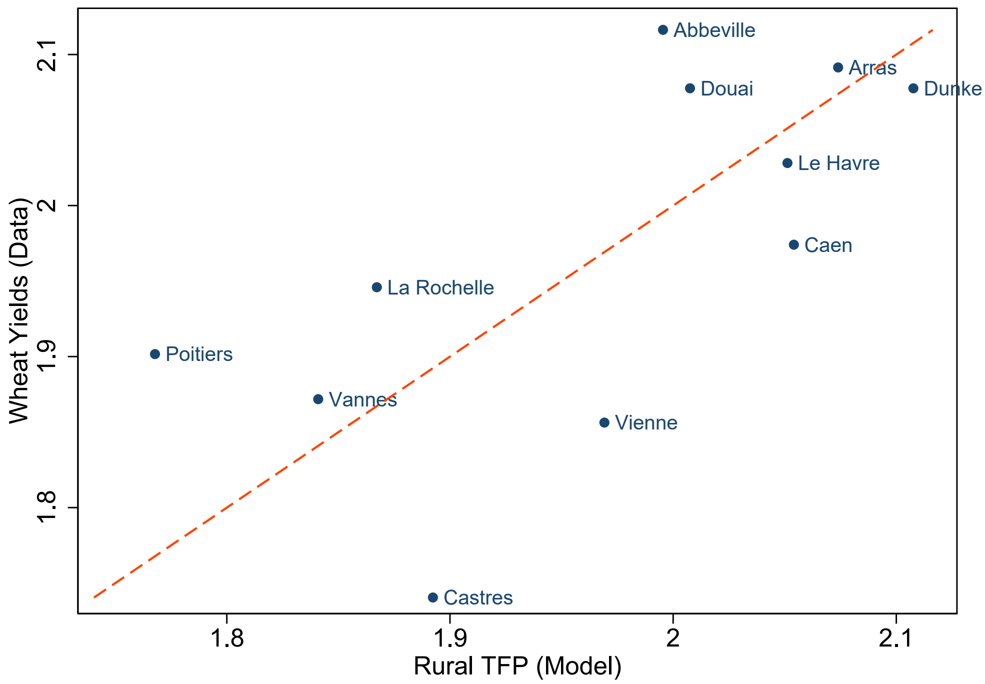


Figure C.10: Wheat yields in 2000 and Rural Regional Productivity in the model, $\theta_{r,k,2000}$.
Notes: For regions/cities producing wheat (11 cities with land use wheat share above 10%), this scatter plot shows wheat yields in city/region k against the rural regional productivity in the model, $\theta_{r,k,t}$. Model counterparts in log are normalized such that the cross-sectional mean matches the data. *Data Source*: See Appendix A.3.2.

can perform a similar regression to regression C.3 as sanity check for $t \in \{1975, 1990, 2000, 2015\}$,

$$\log w_{u,k,t} = a_t + b \cdot \log \theta_{u,k,t} + u_{k,t}, \tag{C.4}$$

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

In regression C.4, b is estimated to 0.65, highly statistically significant (T-stat close to 5.5). It is robust across years, with cross-sectional estimates of b always highly significant and hovering between 0.5 and 0.9 depending on the date t. This is best illustrated by Figure C.11 for the 2015 cross-section, for which one can see that the model implied wage differential between large cities (e.g. Paris or Lyon) and smaller ones is broadly in line with the data. It is important to note that data on urban wages across cities have not been used to estimate $\theta_{u,k,t}$, targeted to match the distribution of urban population. In other words, the wage gap between large and small cities necessary to induce migrations across cities/regions in line with the data matches relatively well the wage gap observed in the data—despite the latter not being among the data moments used for the estimation.

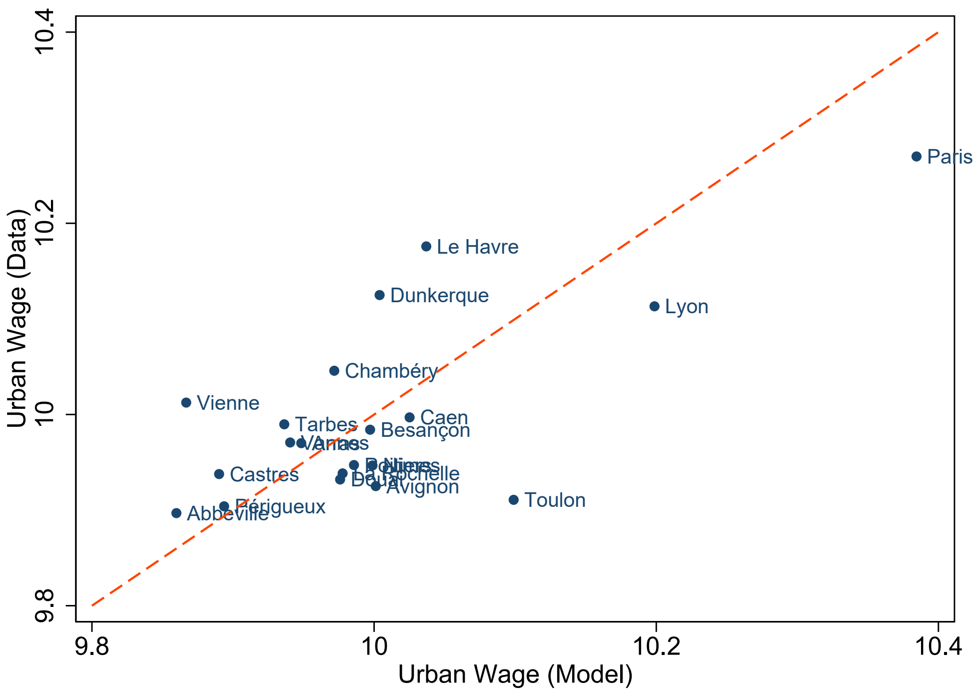


Figure C.11: Urban wage in 2015 in the data and in the model, $\theta_{u,k,2015}$.

Notes: This plots show the average urban wage (in log) in 2015 in the model and in the data (sample of 20 cities). Model counterparts in log are normalized such that the cross-sectional mean matches the data. Model outcomes in 2015 are obtained by linear interpolation between 2010 and 2020. *Data source:* see Appendix C.3.3.

Bibliography

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

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.

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

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

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.

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