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

Nicolas Coeurdacier
SciencesPo Paris, CEPR

Florian Oswald
SciencesPo Paris

Marc Teignier
Serra Húnter Fellow,
University of Barcelona

July 2, 2024

Contents

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

B.1 Model details

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

B.1.1 Set-up Description

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

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

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

B.1.2 Technology

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

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

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

$$w_{u,k} = \theta_{u,k}. \tag{B.1}$$

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

In the rural sector, we extend the model from the main text with a CES technology where the production of the rural good uses labor and land according to the following constant returns to scale technology in each region k,

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

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

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

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

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

where $w_{r,k}$ is the rural wage and $\rho_{r,k}$ the rental price of land in region k and p the relative price of the rural good in terms of the numeraire urban good. Note that it is useful to express the price of land relative to wages,

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

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

B.1.3 Commuting Costs

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

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

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

that m denotes both the mode and the speed of commute. The commuting time (both ways) is, therefore, $t(\ell) = \frac{2\ell}{m}$. The cost $f = f(\ell, m, w_u)$ depends on the transportation mode/speed m, the location ℓ and labor costs w_u.[2] Intuitively, beyond its speed, the pecuniary cost of a commuting mode depends on the distance traveled (e.g. cost of gasoline) as well as the level of wages (e.g. wage of the bus driver). Faster and longer commutes are more expensive and $f(\ell, m, w_u)$ is increasing in m and ℓ, with $\frac{\partial^2 f}{\partial^2 \ell} \leq 0$. The latter technical assumption makes sure that the importance of the cost f (relative to the opportunity cost of time) decreases as the commuting distance increases. The cost f also increases with the labor costs, w_u, with $\frac{\partial^2 f}{\partial^2 w_u} \leq 0$. For tractability, we will use the following functional form for f,

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

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

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

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

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

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

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

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

where $a = \left(\frac{1+\eta_m}{\eta_m} \right) c_\tau^{\frac{1}{1+\eta_m}} (2\zeta)^{\frac{\eta_m}{1+\eta_m}} > 0$. Commuting costs are falling with improvements in the commuting technology (a lower a), while they are increasing with the wage rate in each city (the opportunity cost of time) and the distance of commuting trips with constant elasticities. Expression (B.8) is the resulting commuting cost function which appears in the model solution. It is also important to note that the parameters ξ_w (resp. ξ_ℓ) directly map into elasticities of commuting

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

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

B.1.4 Preferences and Budget Constraint

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

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

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

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

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

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

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

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

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

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

B.1.5 Location Sorting

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

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

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

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

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

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

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

Location-specific housing supply. As shown in Baum-Snow and Han (2023), the elasticity of housing supply to prices is lower closer to the CBD than at the urban fringe. We allow in this extension for location-specific housing supply conditions. To do so, we assume that in each location ℓ_k of city k, land developers supply housing space $H(\ell_k)$ per unit of land with a convex cost

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

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

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

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

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

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

where the parameter $\epsilon(\ell_k)$ is the price elasticity of housing supply in location ℓ_k. More convex costs to build intensively on a given plot of land reduces the supply response of housing to prices.

In the rural area, the housing supply elasticity is assumed constant and identical across regions, $\epsilon_r = \epsilon(\ell_k \geq \phi_k)$.

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

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

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

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

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

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

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

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

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

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

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

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

B.1.7 Market Clearing

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

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

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

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

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

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

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

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

Labour Market Clearing. Labour must clear globally,

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

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

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

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

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

The market clearing condition for rural goods is

$$\sum_{k=1}^{K} C_{r,k} = \sum_{k=1}^{K} Y_{r,k}, \tag{B.17}$$

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

Aggregate Land Rents. The aggregate land rent definition is

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

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

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

B.1.8 Equilibrium Definition

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

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

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

B.1.9 Dynamic Optimization and the Real Interest Rate

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

We start by defining lifetime utility as follows:

$$U_t = \sum_{s=t}^{\infty} \beta^{s-t} \overline{u}_s, \tag{B.19}$$

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

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

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

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

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

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

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

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

$$R_t = \frac{1}{\beta} \frac{\widehat{u'_t}}{\widehat{u'_{t+1}}}. \tag{B.22}$$

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

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

B.2 Quantitative Evaluation

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

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

B.2.1 Multi-Region Numerical Illustration

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

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

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

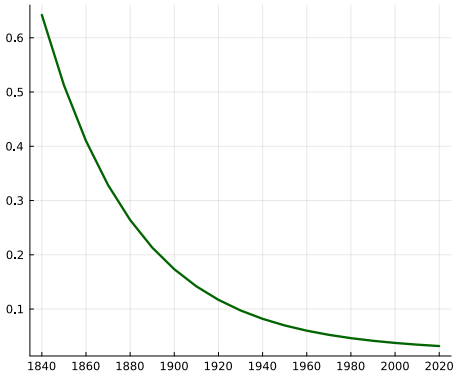
with $\theta_{s,t}$ an aggregate productivity shifter, normalized to 1 in the initial period, and θ_s^k a sector-region specific productivity constant through time. Aggregate productivity, $\theta_{s,t}$, is growing at the constant rate of 1.2% per annum in both sectors. Region-specific productivity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

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

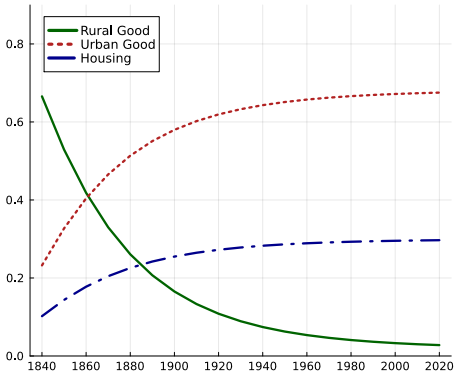
The bottom row shows outcomes that are more specific to our theory with endogenous land use. Plots B.1d and B.1e show aggregate urban area (compared to aggregate urban population) and urban density aggregated across cities (average, central and fringe). With structural change, urban area grows faster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

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

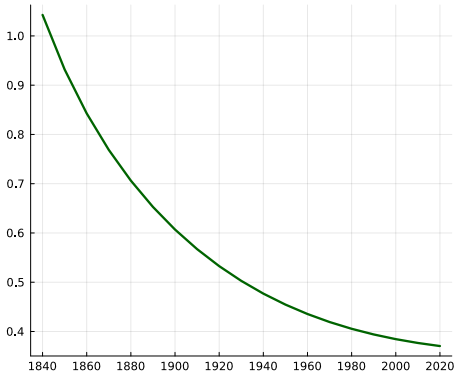
density. Regarding land rents, the reallocation of workers away from agriculture and the fall in the relative price of rural goods exerts downward pressure of the price of farmland. Thus, land rents are reallocated away from the rural part towards the urban part (plot B.1f).



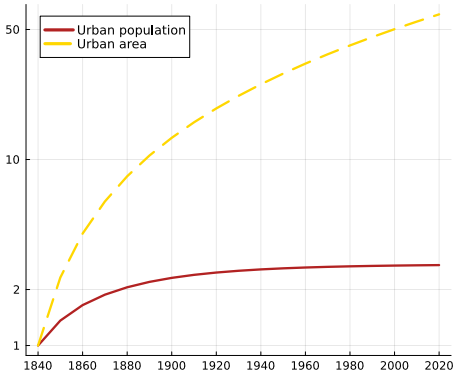
(a) Rural Labor Share $L_r/(L_r + L_u)$.



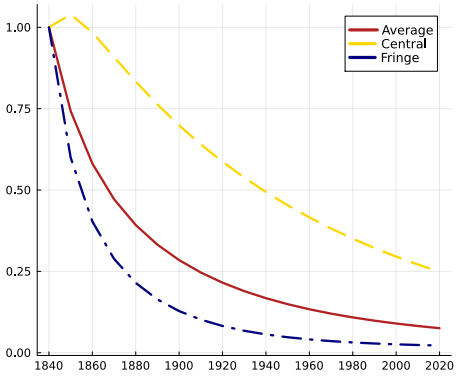
(b) Spending Shares.



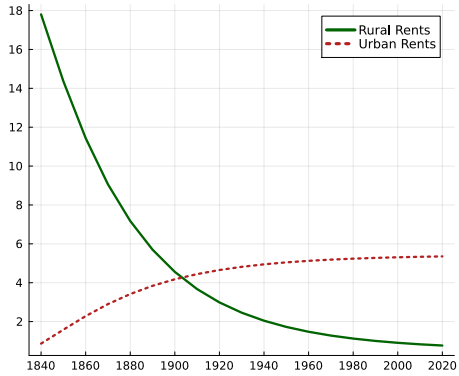
(c) Relative price p.



(d) Urban Area and Population.



(e) Average Urban Densities.



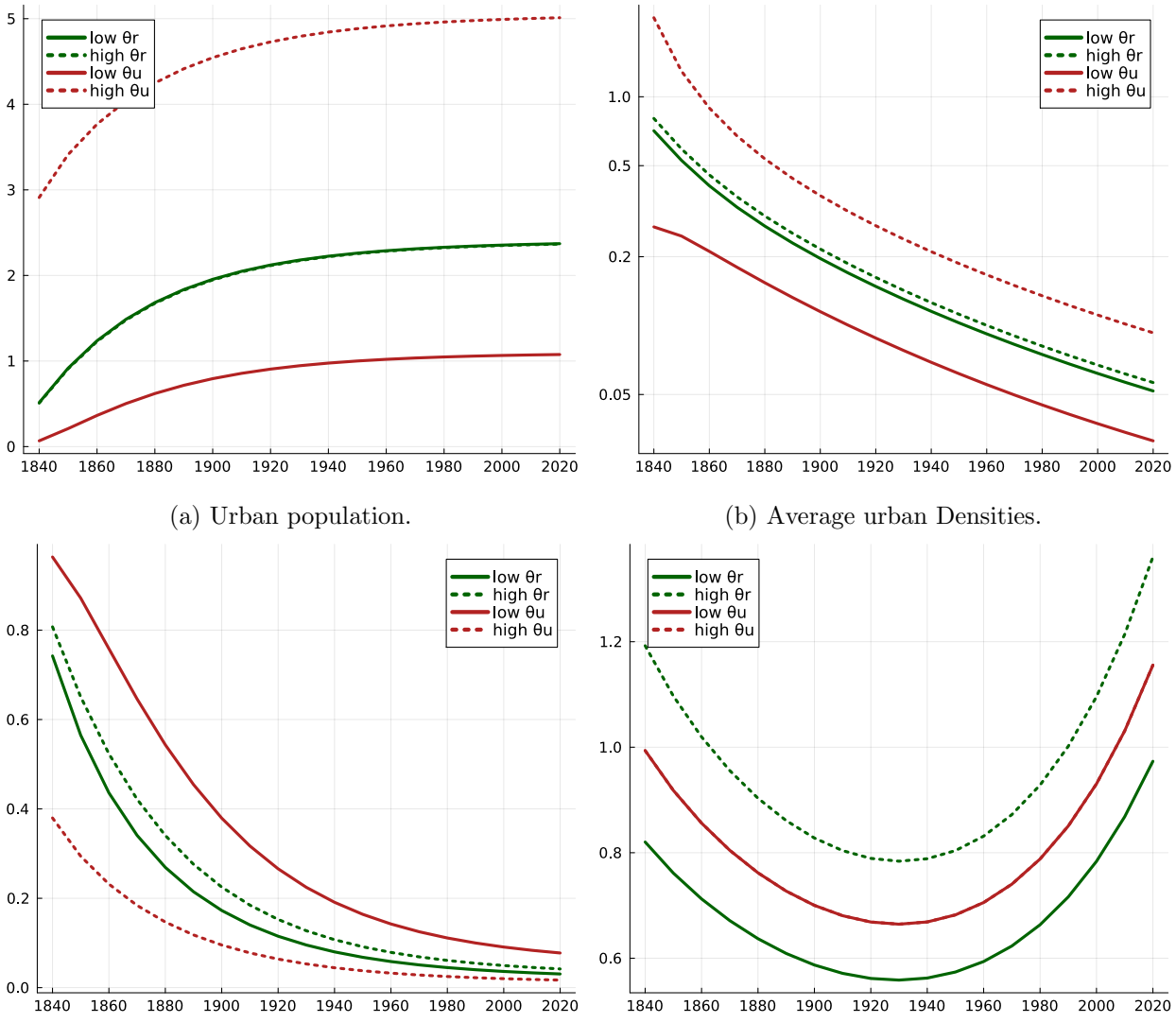
(f) Average land rents (% of income).

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

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

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

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



(a) Urban population.

(b) Average urban Densities.

(c) Rural employment share.

(d) Rural Land Rents.

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

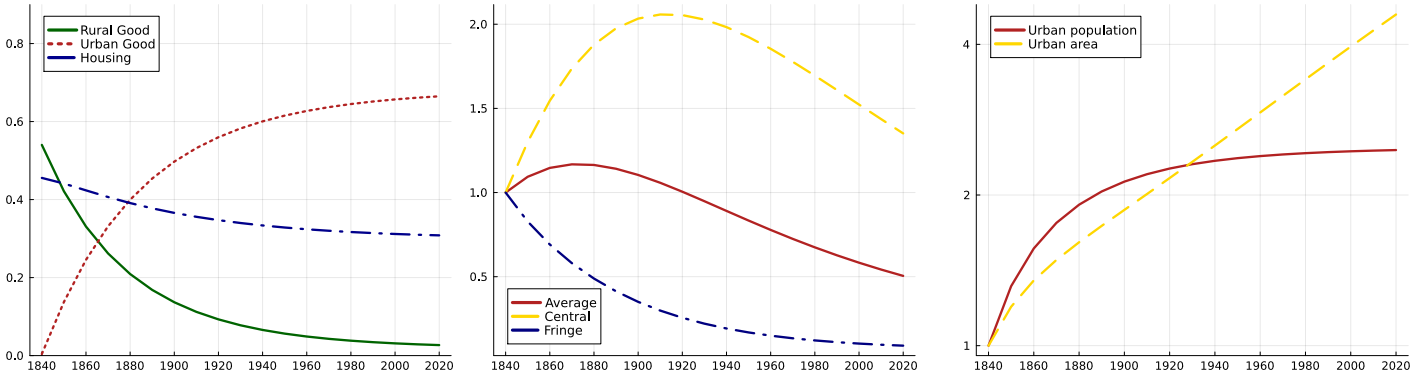
Notes: The setup in this experiment is identical to the one in Figure B.1. Each panel illustrates how variation in θ_s^k can be mapped to a data counterpart. Panels (a) and (d) shows how θ_u^k and θ_r^k are identified in the data.

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

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

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

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



(a) Spending Shares. (b) Average Urban Densities. (c) Urban Area and Population.

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

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

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

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

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

B.2.2 Data Inputs for the Model

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

B.2.2.1 Aggregate Data Inputs

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

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

1. We obtain the estimated series at annual frequency.
2. We subset both series to start in 1840 and end in 2015 (rural productivity ends in that year)
3. We linearly interpolate the missing interwar years.
4. Smoothing is done with a Hann window and a 15-year window size. We experimented with the window size until high-frequency oscillations disappear.

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

5. Our rural productivity series get very volatile starting at the 2000s. We abstract from this noise by growing the smoothed series forward with 1% annual growth from the year 2000 onwards until $T = 2350$, which is our approximation of $T = \infty$ in the model simulation.

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

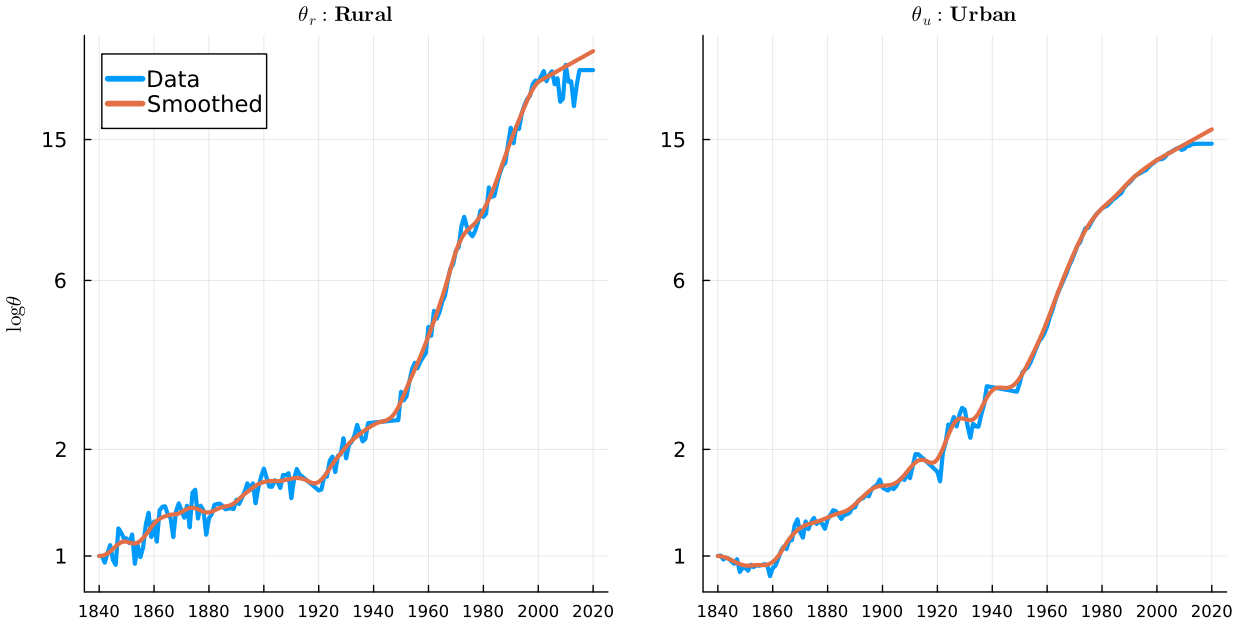


Figure B.4: Smoothing Procedure applied to aggregate sectoral productivity data.

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

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

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

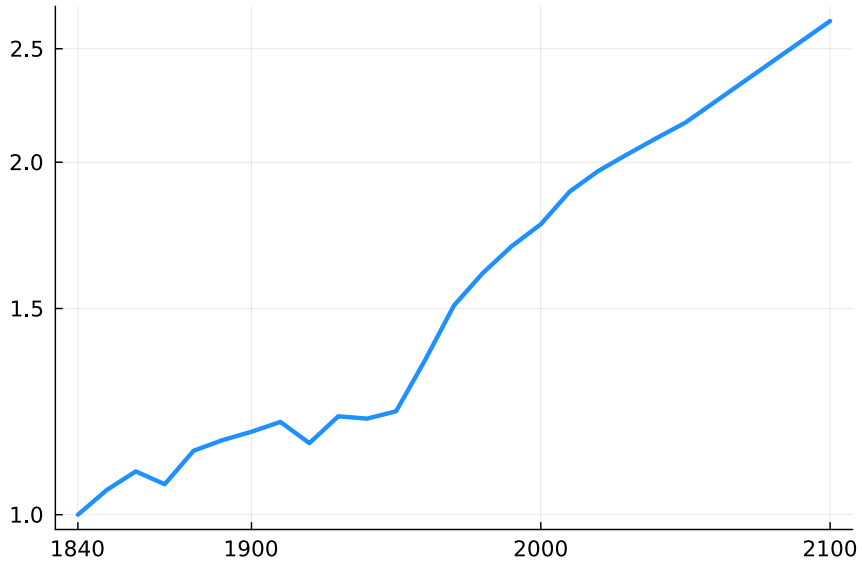


Figure B.5: Population data inputs (1840-2100)

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

B.2.2.2 Cross-Sectional Data Inputs

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

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

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

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

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

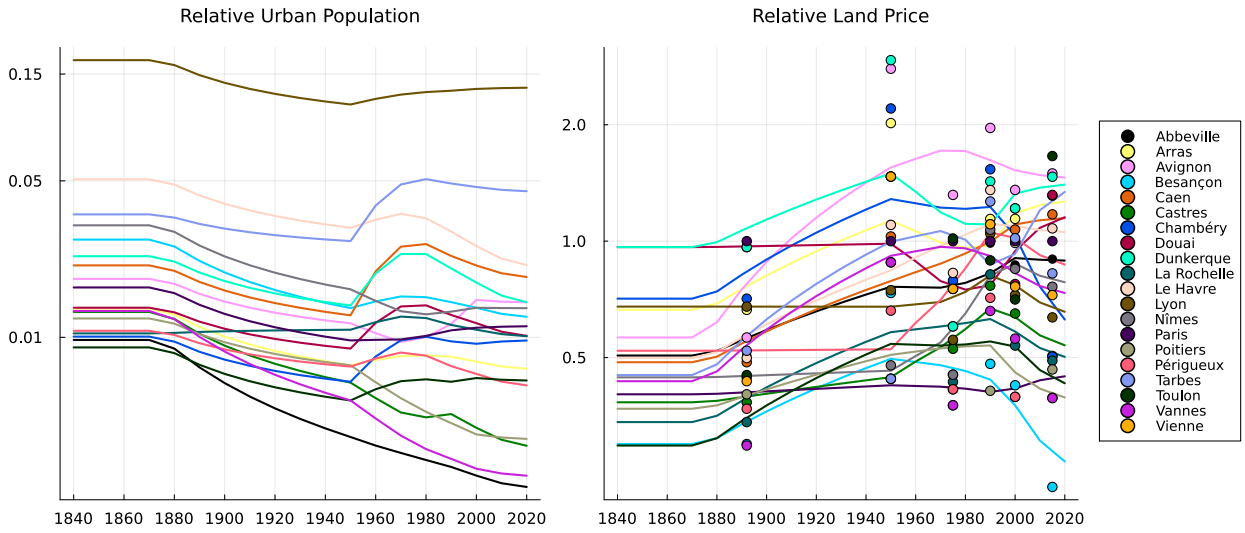


Figure B.6: Regional population and land price data inputs to the model for the sample of 20 cities. *Notes*: The left panel shows urban population relative to Paris, $\frac{\text{pop}_{k,t}}{\text{pop}_{1,t}}$, the right panel shows the value of rural land relative to Paris, $\frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}}$. In the right panel, the markers show the raw data values seen in the data. The displayed colored lines are the result of smoothing. In both panels the respective line for Paris would be constant at unity (but is omitted from the graph).

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

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

1. In each available year, we divide all farmland prices (per ha) by the one of the Parisian area. The main data input is thus a price of farmland relative to the Parisian area. This takes care of scale issues (different price levels or currencies in different periods), and it is consistent with our estimation strategy, targeting the distribution relative to a reference city.

2. We relabel the initial year 1892, when relative farmland prices are first observed, to 1870 corresponding to the first observation of urban population and areas. We assume that data are observed only in years $\mathcal{T} = \{1870, 1950, 1975, 1990, 2000, 2015\}$.

3. As the model requires input data for years $t \notin \mathcal{T}$, we perform linear interpolation to obtain the required value for the relative farmland price, using as interpolation nodes the closest two dates. Outside the range 1870-2015, we extrapolate using the closest observed value.

4. Finally, we smooth the obtained prices as above with a Hann window of window size 5. This is mostly to smooth extreme observations of farmland prices in some years of observation for computational purposes (mostly in 1950, where farmland values for few cities are not in line with other years). Doing so, we make sure that the 1870 data input remains identical to the first observation in the data for each city.

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

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

B.2.2.3 Additional Data Inputs

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

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

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

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

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

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

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

B.2.3 Mapping of Model Outputs to the Data Inputs

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

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

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

B.2.3.1 Cross-Sectional Model Outputs

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

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

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

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

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

$$\frac{\tilde{\rho}_{k,t}}{\tilde{\rho}_{1,t}} = \frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}} \tag{B.25}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B.2.3.2 Selection of City Subset

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

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

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

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

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

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

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

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

B.2.3.3 Aggregate Moment Function

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

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

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

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

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

expenditure (E_k) as

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

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

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

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

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

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

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

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

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

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

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

B.2.4 Solution and Estimation Algorithm

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

1. an outermost loop, where we search for a vector $\varsigma = (a, \gamma, \nu, \underline{s}, \underline{c}, \sigma)$ which is a member of set $\Xi \subset \mathbb{R}^6$ in order to optimize a GMM objective function with relevant aggregate data moment. This part is described in B.2.4.5.

2. A nested loop, described in B.2.4.2, which chooses sequences $\{\theta_{ukt}, \theta_{rkt}\}$ in order to optimize an objective function which minimizes the distance between model and data in terms of relative farmland prices and population distributions. Notice that the solution proceeds period by period (see below), hence in practice the choice involves two vectors of length K in each period t, i.e. $\{\theta_{uk}, \theta_{rk}\}_{k=1}^{K}$. Implied land prices from model need to be built up iteratively,

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

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

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

3. A final innermost loop, which each time solves the system of equations that constitutes an equilibrium and which is described in B.2.4.1.[6]

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

B.2.4.1 Solving a Sequence of Equilibria given parameters

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

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

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

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

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

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

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

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

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

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

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

constraints) described in Su and Judd (2012).

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

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

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

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

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

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

$$\min_{\mathbf{x}_t} \quad g(\mathbf{x}_t) = 1, \quad t = 2030, \ldots, 2350 \tag{B.32}$$

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

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

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

B.2.4.3 Computation of Prices from Rents

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

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

B.2.4.4 Starting Values

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

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

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

B.2.4.5 Estimation

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

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

Estimation involves solving the standard GMM optimization problem

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

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

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

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

B.2.5 Untargeted Model Outputs

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

B.2.5.1 Urban Area and Density

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

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

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

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

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

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

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

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

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

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

weights,

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

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

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

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

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

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

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

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

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

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

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

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

in the ballpark of the data although slightly higher.

Moment	Data	Model
density_decay_MSE	–	0.631
density_decay_coef	0.15	0.184
avg_density_change	−0.883	−0.825
max_mode_increase	4.524	4.776

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

B.2.5.2 Commuting Speed and Agricultural Productivity Gap

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

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

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

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

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

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

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

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

population, is

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

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

B.2.5.3 Land Values and Housing Price Indices

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and, similarly, in the rural part,

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

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

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

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

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

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

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

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

$$RHPI_{k,t} = \frac{HPI_{k,t}}{\tilde{P}_t} = \frac{\mathbb{W}^h_{k,t}}{\mathcal{H}_{k,t}} \frac{1}{\tilde{P}_t}, \tag{B.38}$$

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

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

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

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

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

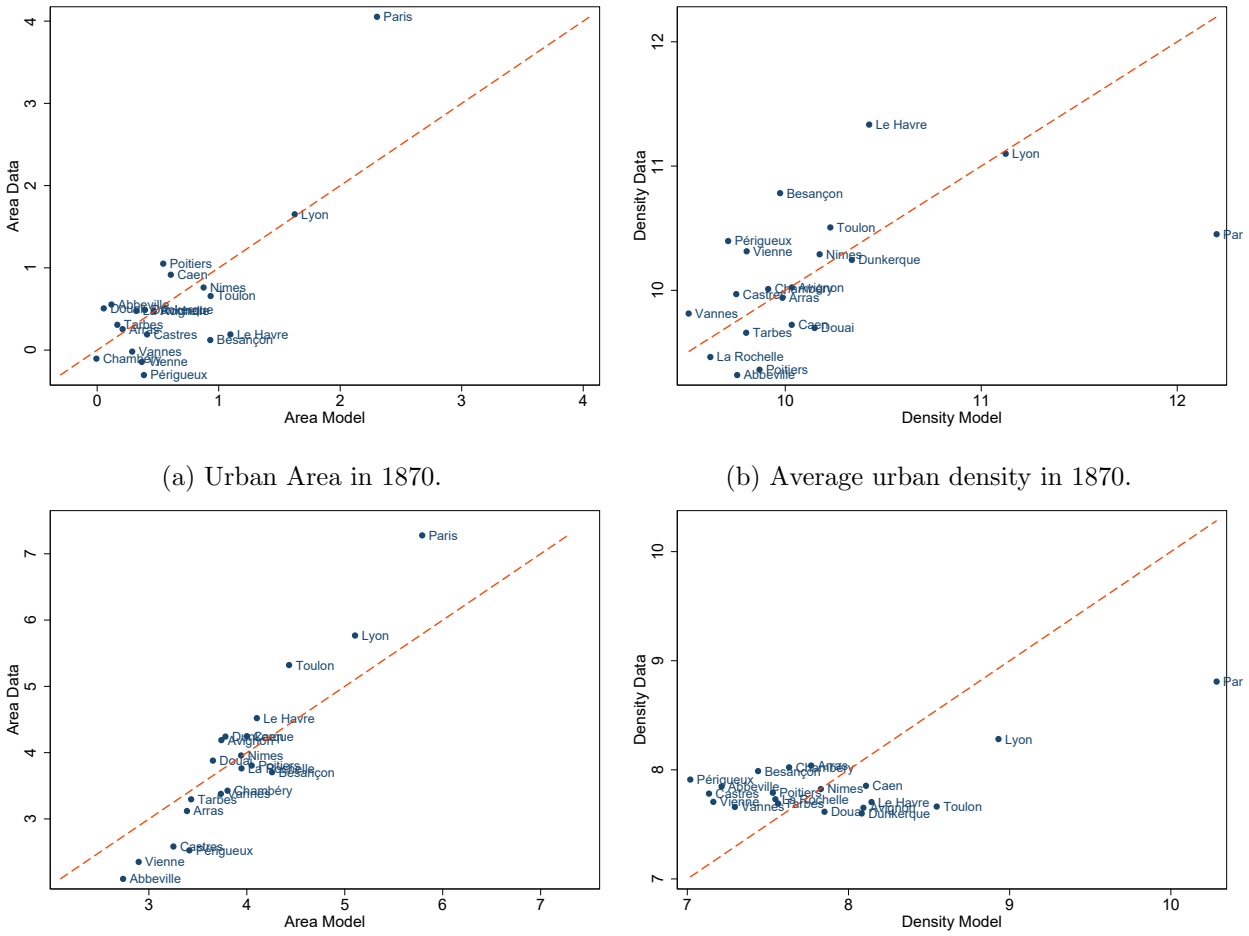
To compute a real house price index for a different set of locations, e.g. the center of a city, we proceed in the same fashion, adjusting the upper integration limits in expressions (B.35),(B.36), and (B.37) appropriately.

B.2.5.4 Additional Untargeted Model Cross-Sectional Outputs

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

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

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



(a) Urban Area in 1870.

(b) Average urban density in 1870.

(c) Urban Area in 2015.

(d) Average urban density in 2015.

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

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

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

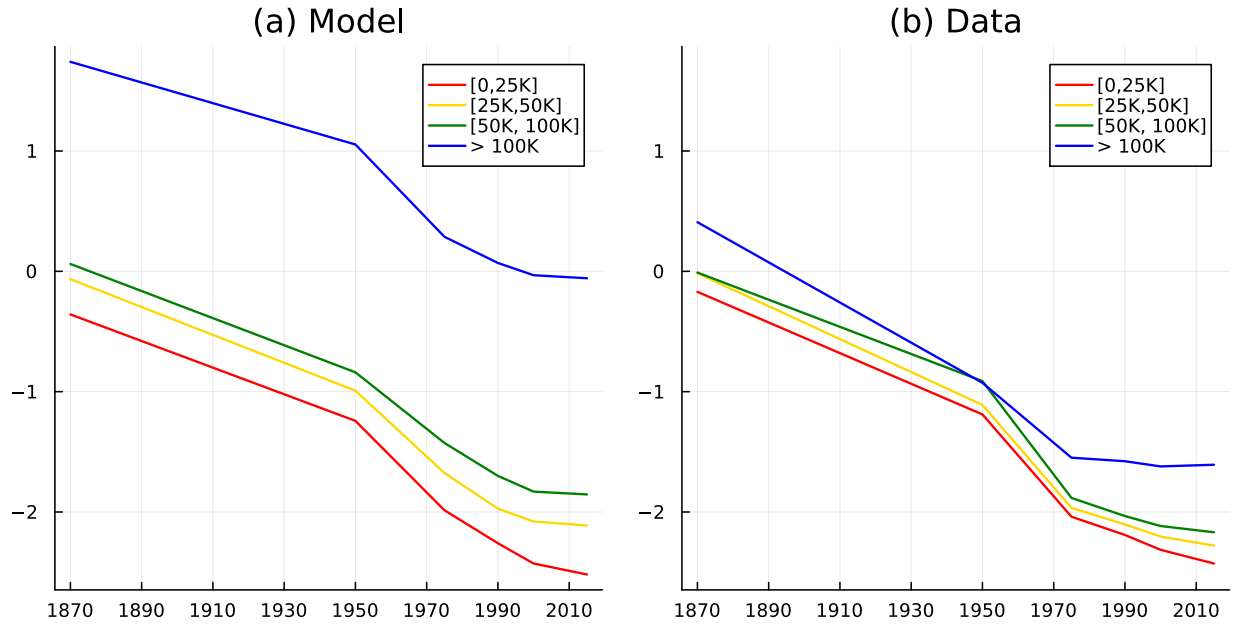


Figure B.8: Average urban density by size-bins: model versus data.

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

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

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

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

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

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

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

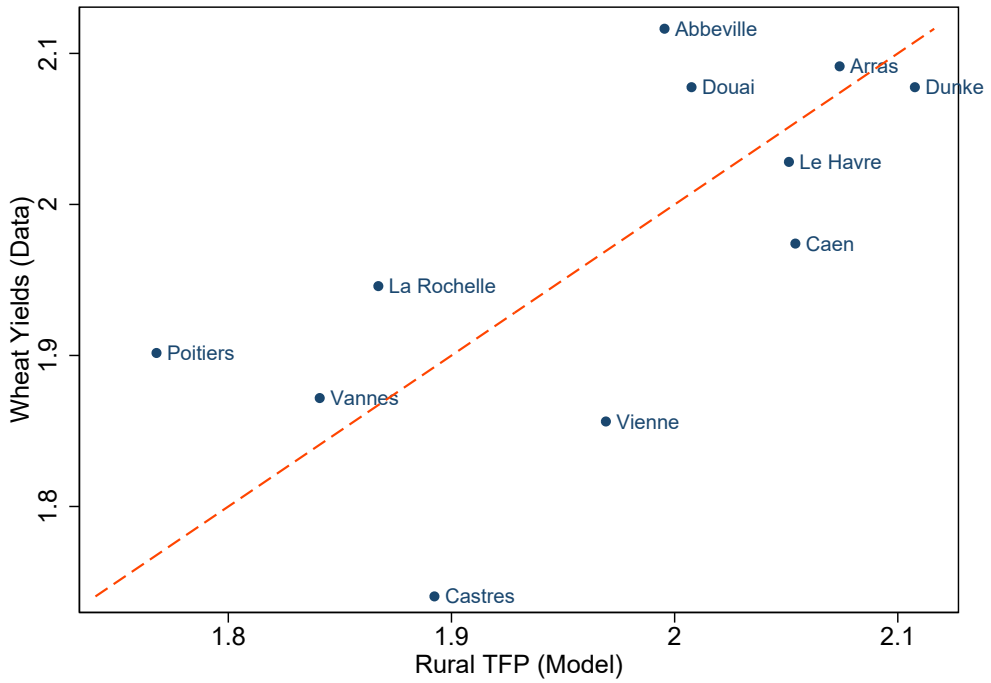


Figure B.9: Wheat yields in 2000 and Rural Regional Productivity in the model, $\theta_{r,k,2000}$.

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

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

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

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

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

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

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

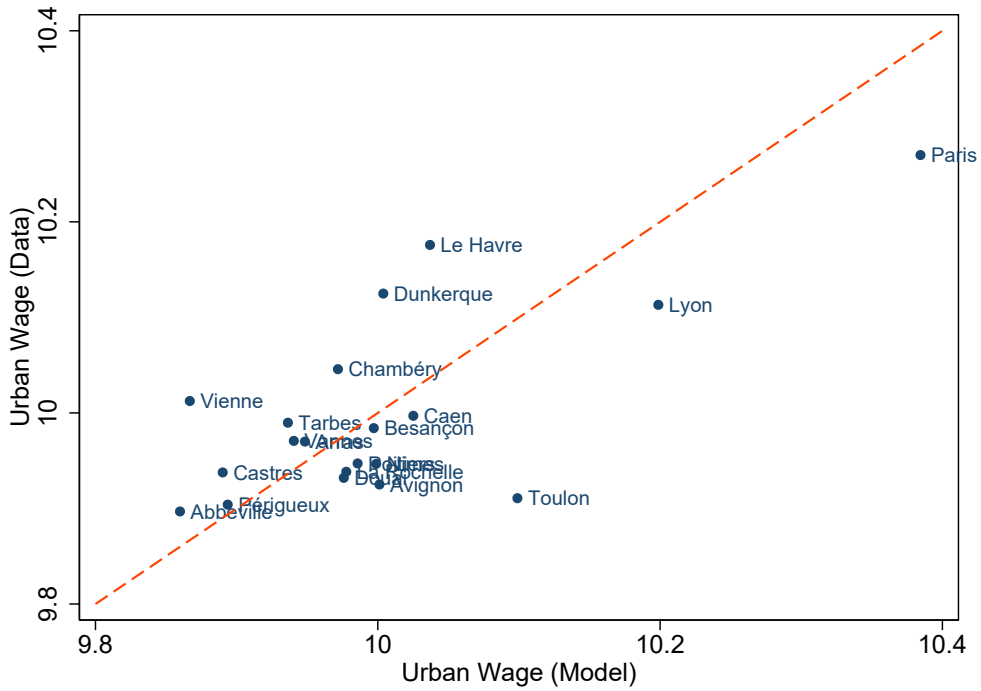


Figure B.10: Urban wage in 2015 in the data and in the model, $\theta_{u,k,2015}$.

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

B.3 Sensitivity Analysis and Extensions

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

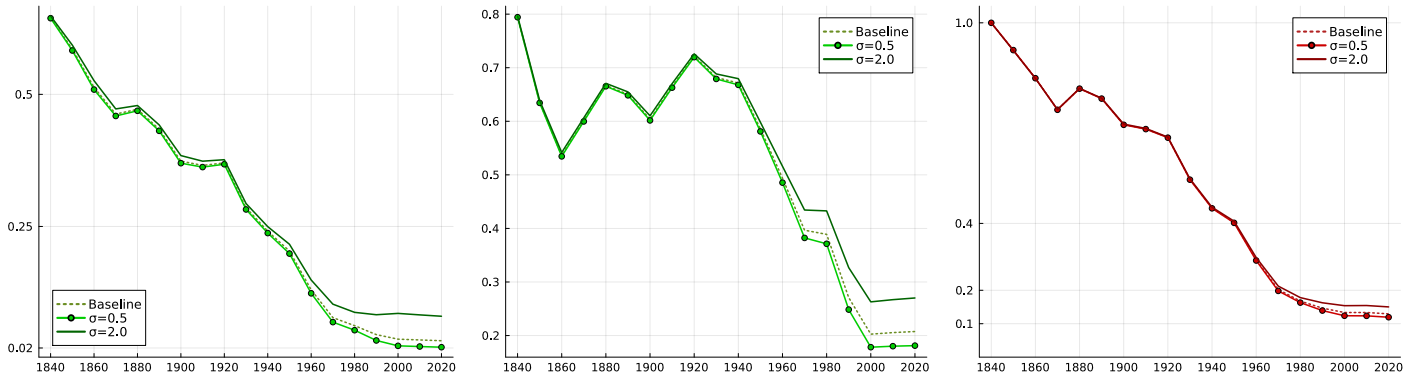
For sensitivity analysis, all model parameters but the one(s) on which sensitivity is performed are kept identical to their baseline values for comparison.

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

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

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

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



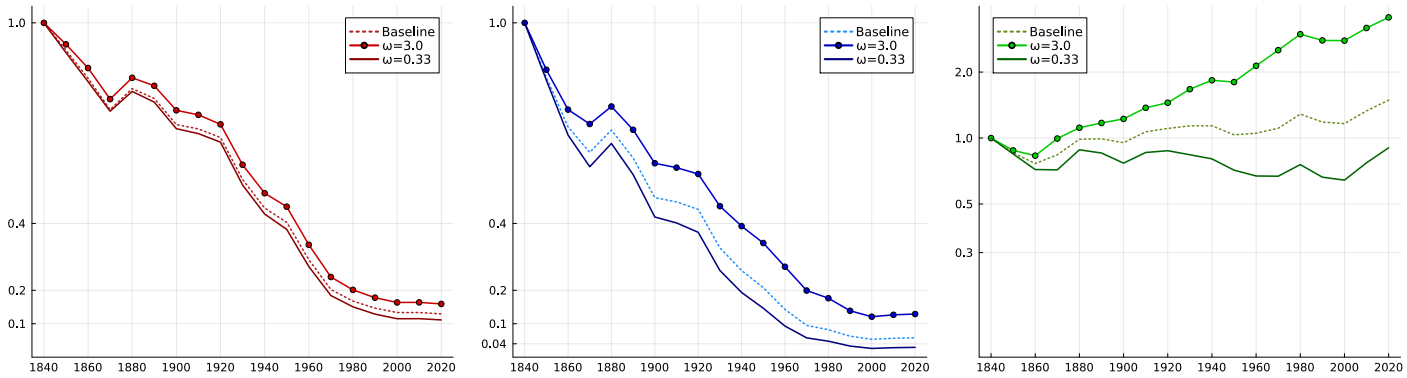
(a) Rural employment share. (b) Relative price of rural good. (c) Average urban density (1840=1).

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

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

B.3.2 Elasticity of substitution between land and labor ω

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



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

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

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

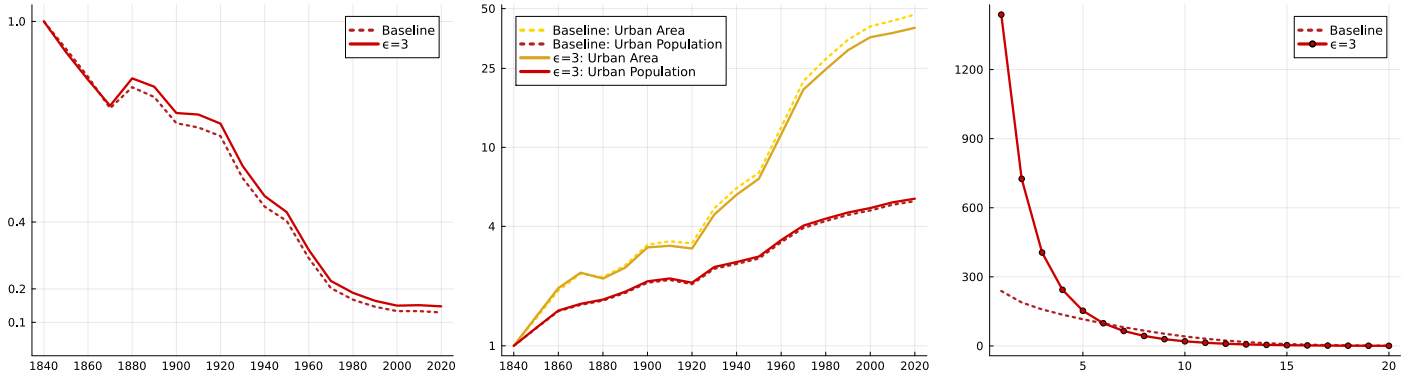
Results are displayed in Figure B.12 for variables of interest, where we focus on aggregate moments

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

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

B.3.3 Housing Supply Elasticity ϵ

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



(a) Average urban density (1840=1). (b) City size (1840=1). (c) Density by ventile (2020).

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

Notes: The housing supply elasticity ϵ is set to 3 in all locations (within and across regions). All other parameters set to their baseline value. Outcomes of interest with constant elasticity, $\epsilon = 3$, are displayed with a solid line. The baseline simulation is shown with a dashed line for comparison.

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

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

B.3.4 Congestion and Agglomeration

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

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

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

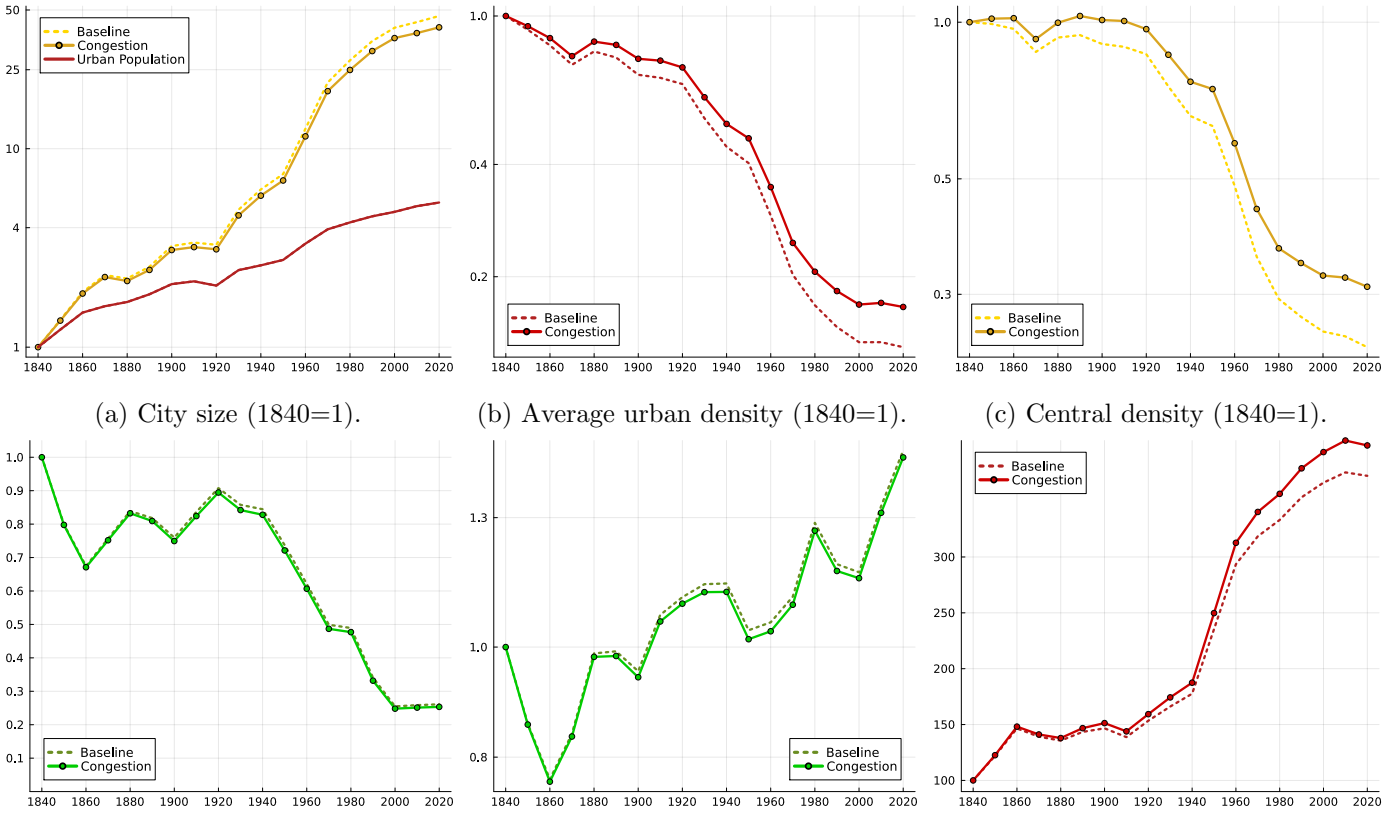
We set externally $\mu = 0.05$. As explained above, under this specification of commuting costs, we re-estimate the commuting cost parameter a and the region-specific sectoral productivities, $\theta_{k,s,t}$—making sure that land used for urban purposes remains reasonable relative to the data in the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

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

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

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

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



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

(d) Rel. price of rural good (1840=1). (e) Rental price of farmland (1840=1). (f) Housing Price Index (1840=100).

Figure B.14: Congestion forces.

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

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

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

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

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

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

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

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

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

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

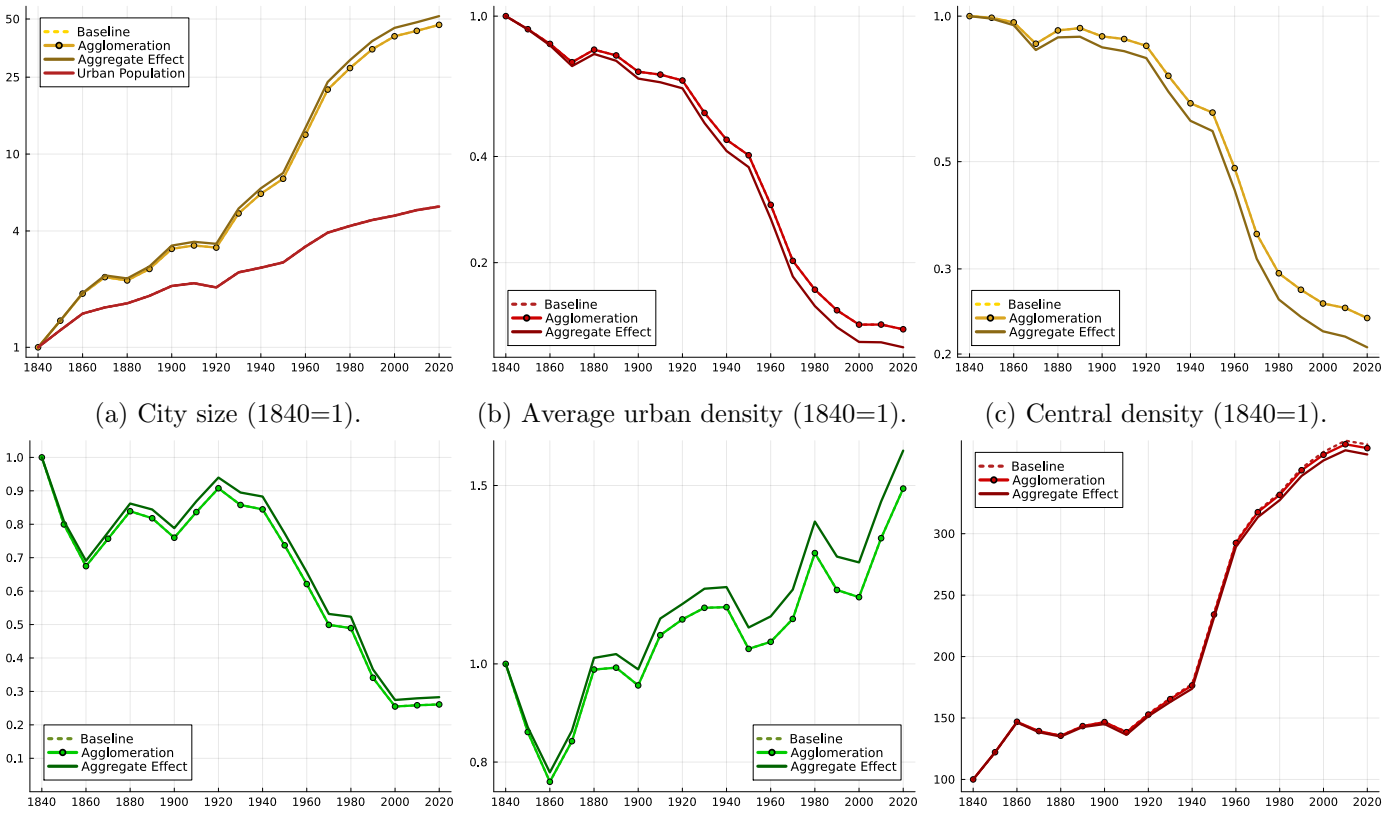
glomeration forces following an aggregate urban expansion along the process of structural change.[10] To do so, we re-estimate the model fitting as aggregate productivity a modified version of Equation (B.42),

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

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

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

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



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

(d) Rel. price of rural good (1840=1). (e) Rental price of farmland (1840=1). (f) Housing Price Index (1840=100).

Figure B.15: Agglomeration forces.

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

B.3.5 Commuting distance and residential location

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

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

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

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

In the quantitative evaluation, we make the following parametric assumptions: $d_0(\phi) = d_0 \cdot \phi$, with d_0 small and positive and $d_1(\phi) = \frac{1}{1+d_1 \cdot \phi}$, with $d_1 \geq 0$. Across cities, $d_0 \cdot \phi$ corresponds to the intercept of Eq. B.45, ranging from 0.2 km for the smaller cities to more than 4 kms for Paris. Given that further away residential locations are typically at 5 kms of the center in smaller urban areas and up to 50 kms away from the center of Paris, d_0 should range within 4% and 8%. We calibrate d_0 externally to 5% in our quantitative experiment. For a radius of about 20 kms (close to the population weighted-mean of our sample of 100 urban areas), a person living in the city center ($\ell = 0$) would commute on average 1 km. Across cities, $d_1(\phi_{k,t}) = \frac{1}{1+d_1 \cdot \phi_{k,t}}$ corresponds to the slope of Eq. B.45—with an estimated mode across urban areas close 0.7 in the data. We calibrate $d_1 = 2$ externally. This yields after model's estimation a slope coefficient that varies across cities, ranging

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

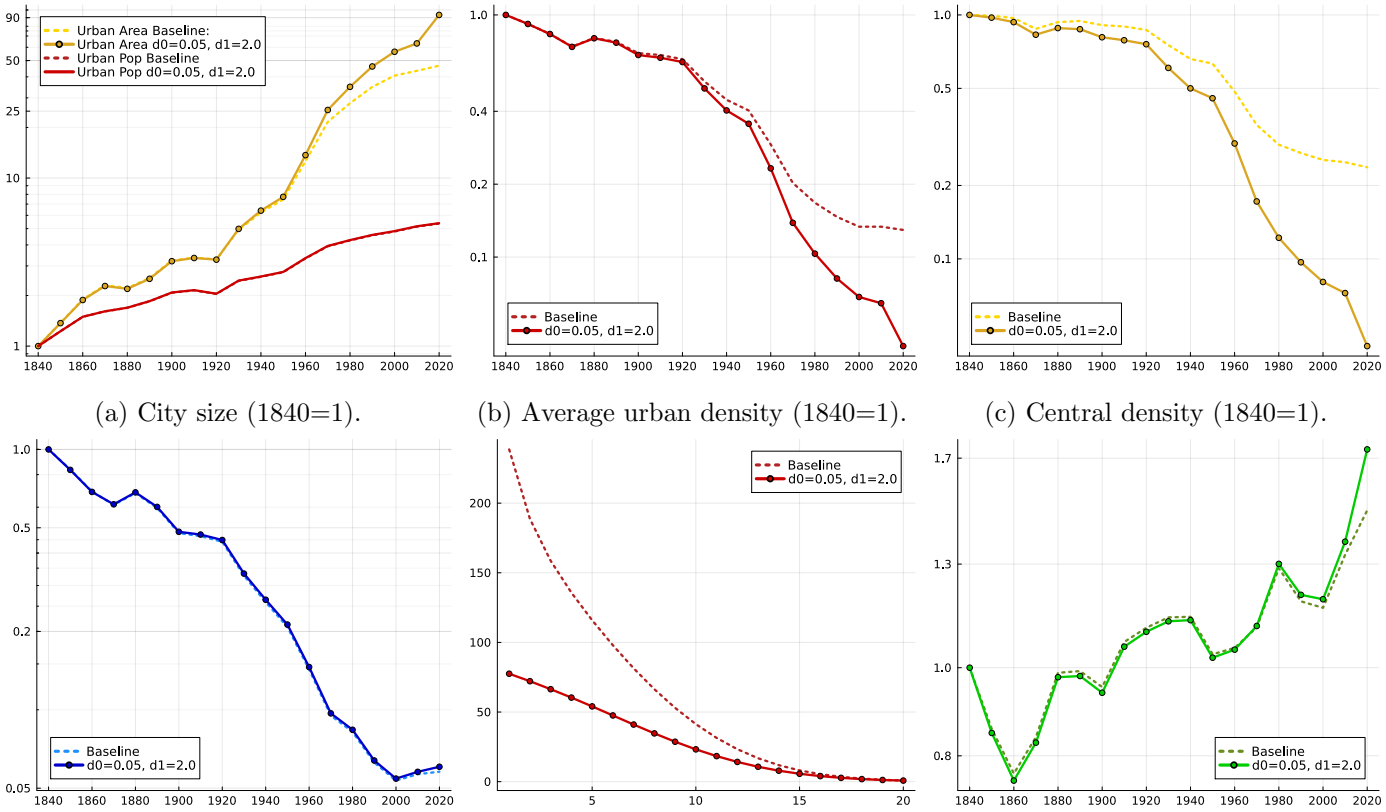
[12] This remains consistent with our calibrated value of ξ_ℓ estimated using commuting distance. The elasticity of speed $m(\ell)$ to commuting distance $d(\ell)$ being $1 - \xi_\ell$.

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

Results. We find that our results are not much affected (Figure B.16). Quantitatively, the city expands more in area in the last decades under this specification of the commuting distance, bringing the model closer to the data (Figure B.16a). As a consequence of this larger sprawling, the average urban density falls more (Figure B.16b). This is driven by a larger fall of central density, the most noticeable difference relative to our baseline monocentric model (Figure B.16c). With urban expansion, residents in central locations end up commuting larger distances—implicitly due to the reallocation of jobs away from the center—, this makes central locations less attractive relative to suburban ones. As a consequence, the within city density gradient is less steep (Figure B.16e). Due to the larger area expansion of cities, rural land gets scarcer and more valuable relative to the baseline (Figure B.16f).

This specification provides also a slightly better fit of the data across cities (Figure B.17). Relative to the baseline monocentric model (in Figure B.18 for comparison), commuting distances in the center (resp. at the fringe) are larger (resp. lower) in larger cities. This, in turn, increases the area of more populated cities in the cross-section at a given date, reducing their average density and bringing the model closer to the data. More populated cities in the model are still noticeably denser than in the data, but less so compared to the baseline monocentric model. The improvement comes from the relative urban area distribution, which fits the data better with the exception of few small cities in the most recent period (2015).[13]

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

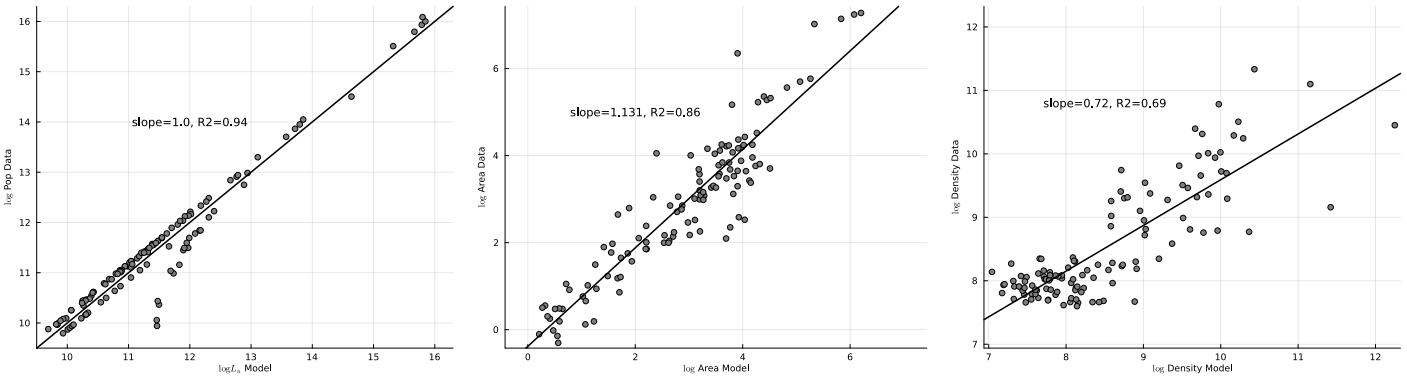


(a) City size (1840=1). (b) Average urban density (1840=1). (c) Central density (1840=1).

(d) Density at the fringe (1840=1). (e) Density by ventile (2020). (f) Rental price of farmland (1840=1).

Figure B.16: Relaxing monocentricity. Aggregate Moments.

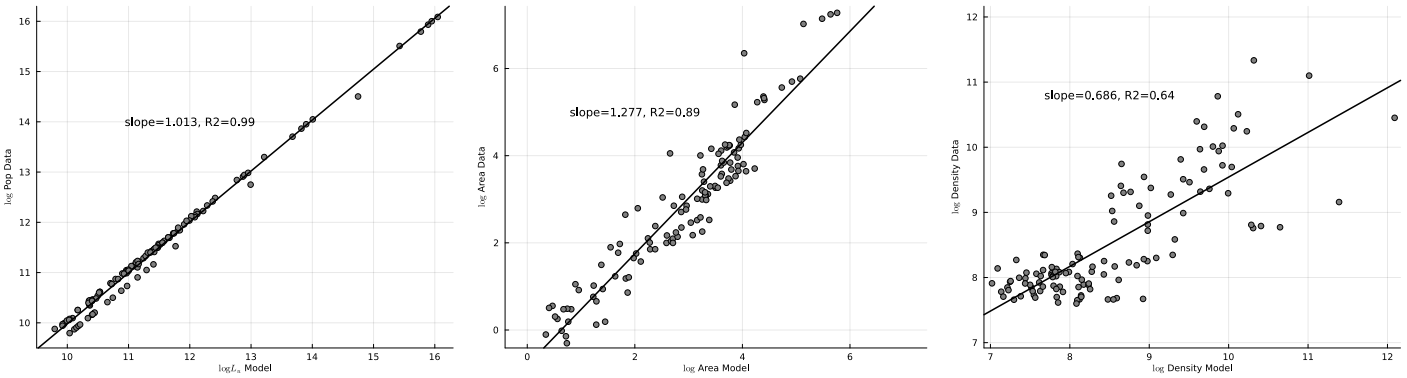
Notes: The solid line represents outcomes in the extended model with alternative commuting costs ((d_0, d_1) extension). For comparison, outcomes of the baseline simulation are shown with a dotted line. Model's outcomes under this alternative specification of commuting costs are obtained re-restimating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



(a) Urban Population. (b) Urban Area. (c) Urban Density.

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

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



(a) Urban Population. (b) Urban Area. (c) Urban Density.

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

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

Bibliography

Baum-Snow, Nathaniel and Lu Han, "The Microgeography of Housing Supply," *Journal of Political Economy*, 2023, *forthcoming.*

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli, "Agricultural productivity and structural transformation: Evidence from Brazil," *American Economic Review*, 2016, *106* (6), 1320–65.

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

DeSalvo, Joseph S and Mobinul Huq, "Income, residential location, and mode choice," *Journal of Urban Economics*, 1996, *40* (1), 84–99.

Dunning, Iain, Joey Huchette, and Miles Lubin, "JuMP: A Modeling Language for Mathematical Optimization," *SIAM Review*, 2017, *59* (2), 295–320.

Fukushima, Nobusumi, Yuichi Nagata, Sigenobu Kobayashi, and Isao Ono, "Proposal of distance-weighted exponential natural evolution strategies," in "2011 IEEE Congress of Evolutionary Computation (CEC)" 2011, pp. 164–171.

Gollin, Douglas, David Lagakos, and Michael E Waugh, "Agricultural productivity differences across countries," *American Economic Review*, 2014, *104* (5), 165–70.

Leukhina, Oksana M. and Stephen J. Turnovsky, "Population Size Effects in the Structural Development of England," *American Economic Journal: Macroeconomics*, July 2016, *8* (3), 195–229.

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

Storesletten, Kjetil, Bo Zhao, and Fabrizio Zilibotti, "Business Cycle during Structural Change: Arthur Lewis' Theory from a Neoclassical Perspective.," Technical Report, National Bureau of Economic Research 2019.

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