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

February 11, 2023

Contents

B.1 Quantitative Model Set-up

The quantitative model is a multiple regions extension of the baseline theory of Section 3. For quantitative purposes, it also expands the baseline theory of Section 3 in the main text by considering a more general production function for the rural good, a more general specification for the commuting costs, by allowing for location-specific housing supply conditions and by considering a circular city on a surface. We also introduce an intertemporal utility function to pin down the path for the real interest rate. This model will then be used for parameter estimation and to generate the main quantitative results in Section B.2. Sensitivity analysis for some parameters values and extensions of the model around its baseline version are also developed in greater details.

B.1.1 Set-up Description

Multiple Regions. The economy is made up of K different regions, each endowed with area S. For quantitative purposes, we consider a surface instead of a line segment. 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 as in the baseline model.

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

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

[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, $\phi_k < \sqrt{S/\pi}$.

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

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

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,

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

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. $\sigma \geq 0$ is the elasticity of substitution between labor and land, $\sigma = 1$ corresponding to the baseline version.

Rural workers and land are paid their marginal productivities such that main text Equations (13) and (14) become for each region k,

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

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

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,

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

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 $1/\sigma$—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

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

to Eq. 4,

$$\tau_k(\ell_k) = f(\ell_k, m_k, w_{u,k}) + 2\zeta w_{u,k} \cdot \left(\frac{\ell_k}{m_k}\right), \tag{B.5}$$

where ℓ_k is the location/commuting distance in city k, m_k the mode in location ℓ_k and $w_{u,k}$ the urban wage in city k. 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.

An individual in location ℓ_k chooses the mode of transportation corresponding to speed m_k in order to minimize the commuting costs $\tau_k(\ell_k)$. 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 $\xi_\ell = \frac{\eta_m+\eta_l}{1+\eta_m} \in (0,1)$. Using Eqs. B.5-B.7, we get that equilibrium commuting costs satisfy,

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

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

Similarly to the baseline model, we denote $w_k(\ell_k) = w_{u,k} - \tau_k(\ell_k)$, the wage net of commuting costs in location ℓ_k of city k.

B.1.4 Preferences and Budget Constraint

Preferences. Consumption over urban and rural goods are non-homothetic as in the baseline. Consider a worker living in a location ℓ_k of region k. The composite consumption good is

$$C(\ell_k) = (c_r(\ell_k) - \underline{c})^{\nu(1-\gamma)} (c_u(\ell_k) + \underline{s})^{(1-\nu)(1-\gamma)} h(\ell_k)^{\gamma} \tag{B.9}$$

Budget constraint. The household earns a wage income net of spatial frictions $w_k(\ell_k)$ in location ℓ_k of region k. Given the spatial structure, $w_k(\ell_k) = w_u - \tau_k(\ell_k)$ for $\ell_k \leq \phi_k$ and $w_k(\ell_k) = w_{r,k}$ for $\ell_k \geq \phi_k$. The households also earns land rents, r. Land rents are redistributed lump-sum 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 satisfies

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

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,

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

B.1.5 Location Sorting

Mobility conditions. Consumption allocation across the different goods in each location ℓ_k remains identical to the baseline. Workers can freely move within each region k, as well as across regions. Within region, this gives the following mobility equation within each region k similar to the baseline model. For all location ℓ_k in region k,

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

These mobility conditions generate housing rental price gradients in each city k similar to Eq. 11 for each city k.

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

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

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 equilibrium

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

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

paid in units of the numeraire, where $1/\epsilon_k(\ell_k)$ 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

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

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,

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

where the parameter $\epsilon_k(\ell_k)$ 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,

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

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 (Equation (B.3)), where the condition holds with equality in the rural part of the economy, for $\ell_k \geq \phi_k$,

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

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. As in the baseline model, using the housing rental price gradient in each city k and the housing demand in each location ℓ_k, the density $D_k(\ell)$

follows for $\ell \leq \phi_k$,

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

where $w_k(\ell)$ is the wage net of commuting costs in location ℓ of city k, $\gamma_{\ell,k} = \frac{\gamma}{1+\epsilon_k(\ell)}$ 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, $D_k(\ell)$, across urban locations gives the total urban population of city k,

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

Note that with homogeneous supply conditions across locations, $\epsilon(\ell) = \epsilon_r = \epsilon$, Equation (B.13) simplifies into Equation (21) of the main text up to the circular city transformation,

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

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 (22), 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

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

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

1. $C_{u,k} = \left(\int_0^{\phi_k} c_{u,k}(\ell) D_k(\ell) 2\pi \ell d\ell + c_{u,k}(\ell_k \geq \phi_k) L_{r,k} \right)$ 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. $\mathbb{T}_k = \int_0^{\phi_k} \tau_k(\ell) D_k(\ell) 2\pi \ell d\ell$ 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. $\mathbb{H}_{u,k} = \left(\int_0^{\phi_k} \frac{\epsilon_k(\ell)}{1+\epsilon_k(\ell)} q_k(\ell) H_k(\ell) 2\pi \ell d\ell + \frac{\epsilon_r}{1+\epsilon_r} q_{r,k} H_{r,k} \right)$ 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 $C_{r,k} = \left(\int_0^{\phi_k} c_{r,k}(\ell) D_k(\ell) 2\pi \ell d\ell + c_{r,k}(\ell_k \geq \phi_k) L_{r,k} \right)$ 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 needs to be adjusted for the circular area with respect to the main text version, and becomes, by summing across regions,

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

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

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

B.1.8 Equilibrium Definition

The equilibrium remains the static one 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.

As in the main text, 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 clear 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 purpose of the dynamic model extension is to be able to compute purchase prices for urban and rural land in a certain location, which depend on discounted stream 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,

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

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. 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.10).

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

$$R_t = \frac{1}{\beta} \frac{\widehat{u'_t}}{\widehat{u'_{t+1}}}. \tag{B.21}$$

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

$$\widehat{u'} = \sum_{k=1}^{K} \frac{L_k}{L} \left[\frac{1}{L_k} \int_0^{\phi_k} \frac{2\pi \ell D_k(\ell)}{w_k(\ell) + r + \underline{s} - p\underline{c}} d\ell + \frac{L_{r,k}}{L_k} \frac{1}{w_{r,k} + r + \underline{s} - p\underline{c}} \right]. \tag{B.22}$$

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:

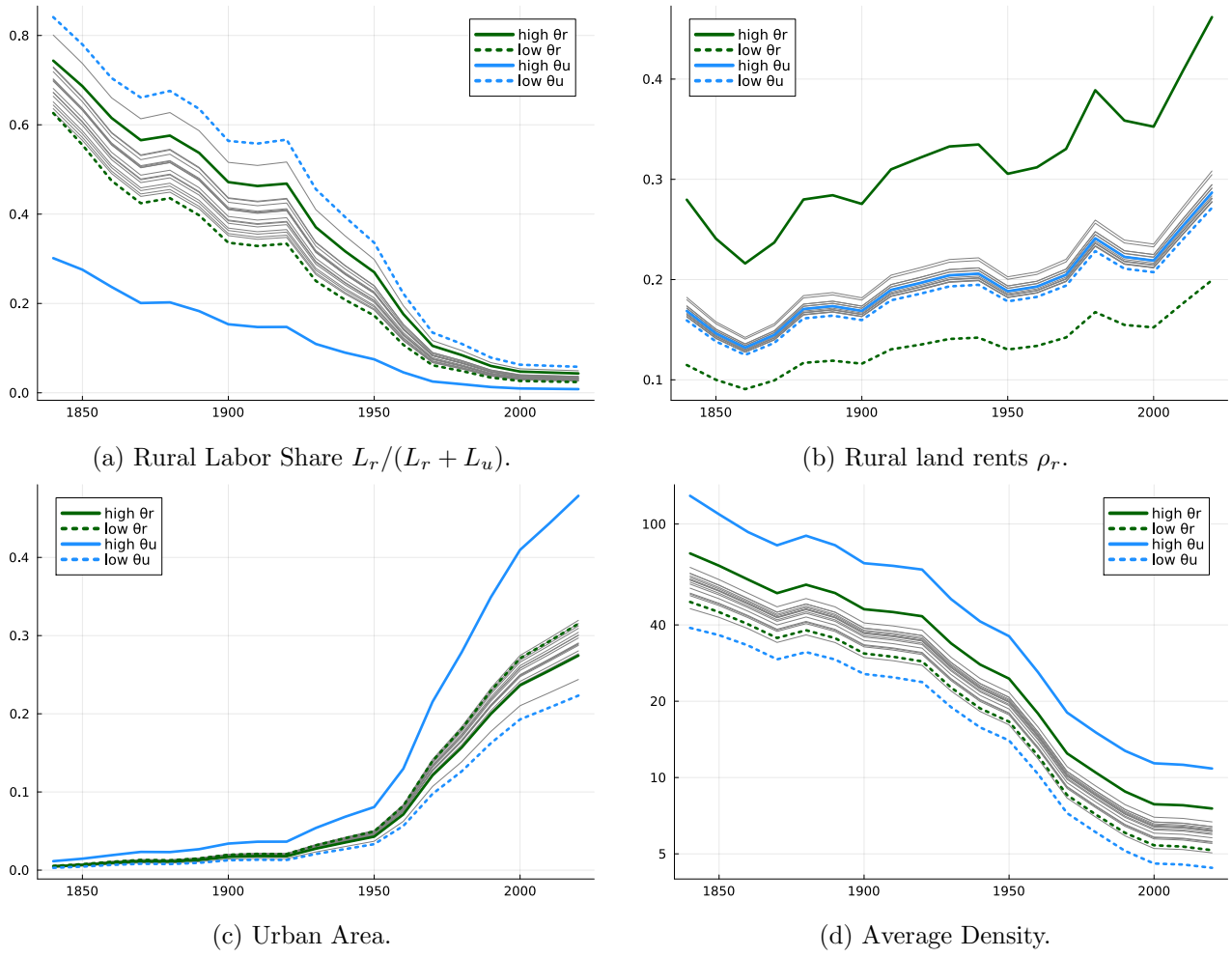
B.2.1 Multi-Region Numerical Illustration

Before jumping into the detailed estimation of the quantitative model for France, it is useful to plot an artificial economy with multiple regions that resembles the data and illustrates the cross-sectional properties of the model. To do so, we simulate an economy with the same path of sectoral aggregate productivity $(\theta_{u,t}, \theta_{r,t})$ as France (and further detailed below) and a set of parameter values $(a, \gamma, \nu, \underline{s}, \underline{c})$ close to the final baseline estimation. The externally calibrated parameters, land intensity in rural production, $1-\alpha$, elasticity between land and labour, σ and commuting costs elasticities, ξ_ℓ and ξ_w are set to the baseline values described in the main text.

Each region has a region-specific urban or rural productivity component, θ_u^k and θ_r^k constant over time: all regions face sectoral productivity improvements due to aggregate productivity changes and some are relatively more or less productive permanently, $\theta_{s,k,t} = \theta_s^k \cdot \theta_{s,t}$ in sector $s \in \{u, r\}$. The

distribution of regional-specific productivities is in a narrow range around the mean for most regions but four 'treated' regions have a region-specific sectoral productivity significantly above or below average. We denote respectively high/low θ_r and high/low θ_u these four regions. This generates cross-sectional heterogeneity that can be easily interpreted. The time-sequence starts in 1840 and we solve the model every ten years until 2020.

This multi-region numerical illustration, displayed in Figure B.1, sheds light on the identification of technology parameters in the cross-section, region-specific productivity parameters, θ_r^k and θ_u^k, and of the cross-sectional implication of the quantitative set-up described in Section B.1.



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

(b) Rural land rents ρ_r.

(c) Urban Area.

(d) Average Density.

Figure B.1: Illustration of regional heterogeneity effects.

Notes: We take the aggregate growth paths for θ_u and θ_k from the data and described below in Section B.2.2, but we artificially generate a regional component θ_s^k for $s \in u, r$. All but four cities have θ_s^k similar in a narrow range (grey lines). The four treated cities have either θ_u^k or θ_r^k significantly higher/lower from the rest.

Aggregate implications. The time-series evolution mimics the pattern of structural change for France due to aggregate productivity changes: labor moves away from the rural sector across all regions (Figure B.1a), city expands in population and area (Figure B.1c). However, cities expand

more in area than population and average density falls in all cities (Figure B.1d).

Cross-sectional implications. More novel relative to the baseline theory are the cross-sectional implications. Focusing the blue curves of Figure B.1, cities with a more productive urban sector are larger in population and area, and the rural employment share is lower in those regions. In the cross-section, more productive and larger cities are also denser—as in the data cross-sectionally. Importantly, for identification of the region-specific urban productivity component, there is a mapping between the (relative) region-specific urban productivities, θ_u^k, and the (relative) populations of cities. This mapping is at the heart of the identification of region-specific urban productivities in the cross-section—the $\theta_{u,t}^k$ will be identified to match the distribution of cities population at each date t.

Focusing the green curves of Figure B.1, regions with a more productive rural sector feature a higher employment share in the rural sector and higher rural land rents—rural land being more productive there. Therefore, as the opportunity cost of expanding the city at the fringe (a higher farmalnd rent) is higher, these cities will be denser (Figure B.1d). This latter prediction is at the heart of the mechanisms at play in our story and extensively discussed in Section 4.4. It forms the basis of our empirical investigation linking urban density and local farmland values in the cross-section of cities. Lastly, it is important to note that there is a mapping between the (relative) region-specific rural productivities, θ_r^k, and the (relative) regional farmland rents, or (relative) regional farmland prices (appropriately discounted sum of future rents as detailed in Section B.2.3). This mapping is at the heart of the identification of region-specific rural productivities in the cross-section—the $\theta_{r,t}^k$ will be identified to match the distribution of regional farmland prices at each date t.

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, $t \in \{1840, 1850, ..., T\}$.

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 Fig-

ure 10 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.
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 (implying 335 years of future years in simulation).

This procedure yields the smoothed series displayed in Figure B.2.

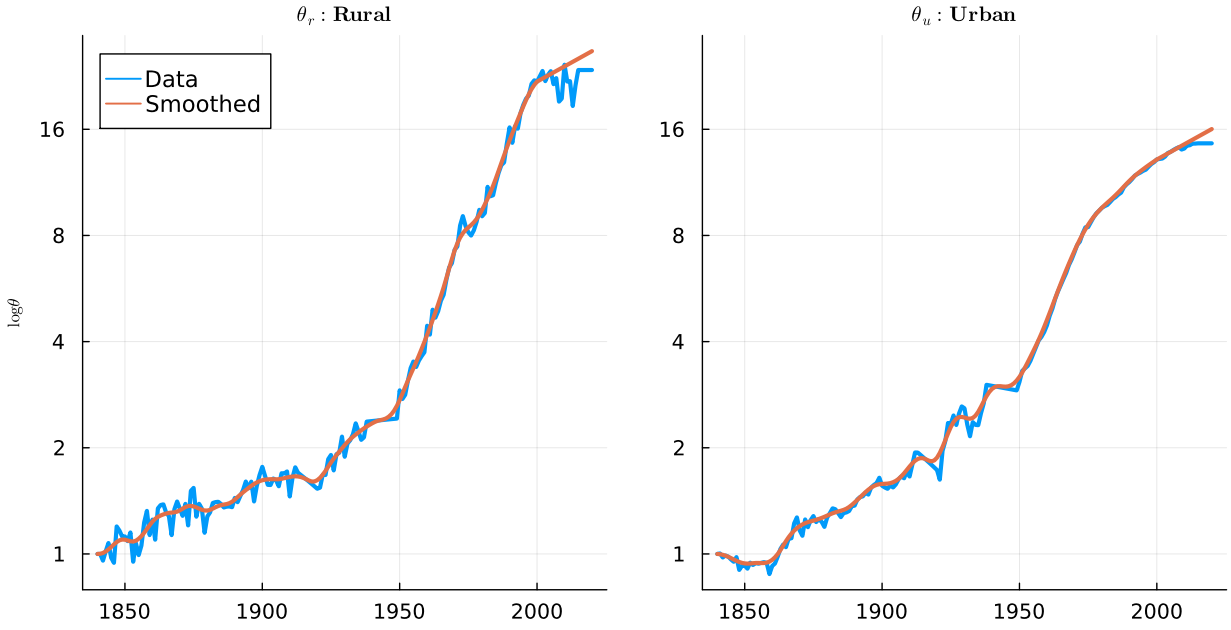


Figure B.2: Smoothing Procedure applied to aggregate sectoral productivity data.

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.3 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.

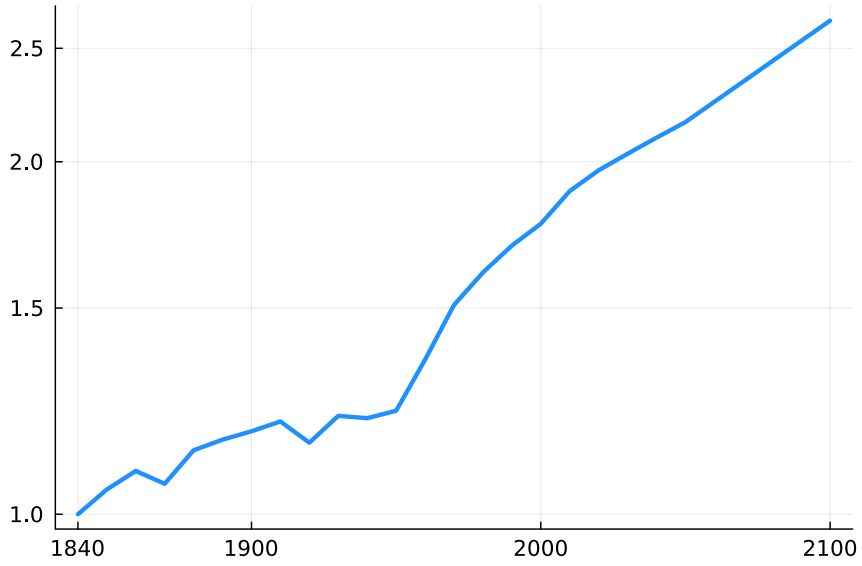


Figure B.3: Population data inputs (1840-2100)

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

Treatement of Cities 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. Data are not directly available at other years.

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

population 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.4.

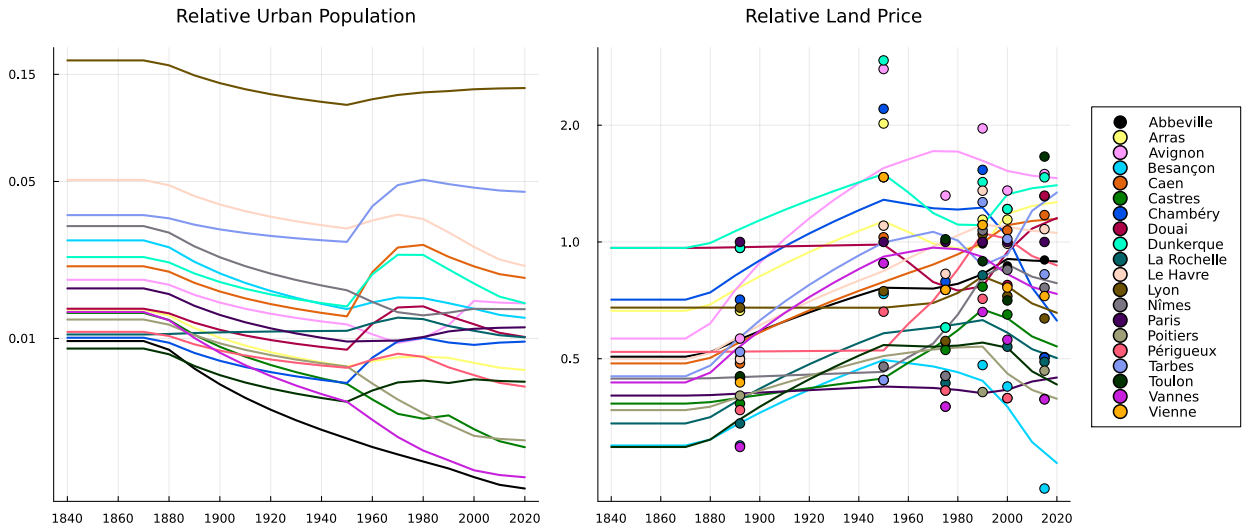


Figure B.4: Regional population and land price data inputs to the model for the sample of 20 cities. *Notes*: The left panel shows urban population relative to Paris, $\frac{\text{pop}_{k,t}}{\text{pop}_{1,t}}$, the right panel shows the value of rural land relative to Paris, $\frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}}$. 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).

Treatement 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 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 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.[2] 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.4.

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, σ, and the location-specific housing supply elasticities, $\epsilon_k(\ell)$,

[2] 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.

are calibrated externally based on standard values in the literature. Sensitivity is performed with respect to the elasticity of substitution σ and the housing supply elasticities $\epsilon_k(\ell)$ in Section B.3 and results are robust for parameters 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 parameters values, the equilibrium is independent on β, which only matters to compute equilibrium land/housing values beyond rents. It does impact slightly the estimation of parameters described below by affecting 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 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.

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 cities population. 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),

$$\frac{L_{u,k,t}}{L_{u,1,t}} = \frac{\text{pop}_{k,t}}{\text{pop}_{1,t}} \tag{B.23}$$

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 relative farmland prices, relative to the reference city $k = 1$ (Paris)

$$\frac{\tilde{\rho}_{k,t}}{\tilde{\rho}_{1,t}} = \frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}} \tag{B.24}$$

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, $\mathbb{R}_{k,t}(\ell) = \sum_{s=t}^{\infty} \frac{\rho_{k,s}(\ell)}{(R_s)^{(s-t)}}$, where the infinite sum is

approximated for a T large enough relative to t,

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

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,

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

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

$$\tilde{\rho}_{k,t} = \frac{\mathbb{W}^l_{r,k,t}}{S_k - \phi^2_{k,t}\pi} = \frac{1}{S_k - \phi^2_{k,t}\pi} \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \mathbb{R}_{k,t}(\ell) 2\pi\ell d\ell. \tag{B.25}$$

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.25 as follows,

$$\begin{aligned} \tilde{\rho}_{k,t} &= \rho_{r,k,t} + \frac{1}{S_k - \phi^2_{k,t}\pi} \sum_{s=t+1}^{T} \frac{\int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \rho_{k,s}(\ell) 2\pi d\ell}{(R_s)^{(s-t)}} \\ &= \rho_{r,k,t} + \mathbb{S}_{k,t}\left(\phi_{k,t}, \sqrt{S_k/\pi}\right), \end{aligned} \tag{B.26}$$

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.21). 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 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.

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 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

City	Area	Population	Rural Land Price	Département
Paris	1397.94	8898707.0	1.00	11
Lyon	298.81	1145494.1	0.77	84
Toulon	196.12	417663.3	0.93	93
Le Havre	83.81	227594.2	1.09	28
Caen	64.62	186321.4	1.10	28
Dunkerque	69.00	156273.3	1.33	32
Avignon	61.31	130705.6	1.52	93
Besançon	38.19	120628.4	0.38	27
Nîmes	46.56	120585.1	0.88	76
Douai	46.62	102944.2	0.94	32
Poitiers	38.50	98203.5	0.42	75
La Rochelle	39.75	96235.7	0.58	75
Chambéry	27.25	83291.7	1.03	84
Arras	21.75	69290.0	1.18	32
Tarbes	23.56	61073.9	1.02	76
Vannes	26.19	58532.3	0.53	53
Castres	12.44	35094.1	0.64	76
Périgueux	9.56	32778.1	0.46	75
Vienne	9.38	23030.1	0.83	84
Abbeville	7.69	21463.7	0.90	32

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.

City	Area	Population	Rural Land Price	Département
Paris	1397.94	8898707.0	1.00	11
Bordeaux	206.00	605708.1	0.74	75
Montpellier	80.69	279285.5	1.13	76
Tours	75.94	229875.1	0.47	24
Mulhouse	74.50	208798.6	1.01	44
Dijon	55.25	205932.1	0.66	27
Brest	58.25	173505.0	0.72	53
Pau	45.50	122734.8	1.03	75
Troyes	49.69	121934.0	1.13	44
Chalon-sur-Saône	28.50	64985.3	0.33	27
Roanne	24.81	61905.0	0.45	84
Béziers	16.44	58099.8	1.13	76
Quimper	24.31	57372.3	0.58	53
Châteauroux	21.88	53116.9	0.79	24
Nevers	20.44	50740.8	0.44	27
Niort	23.75	50371.9	0.39	75
Armentières	12.00	43496.2	1.15	32
Moulins	16.88	33243.4	0.36	84
Rochefort	13.50	27265.6	0.48	75
Morlaix	10.44	17412.3	1.28	53

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.

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	Land-Labor Elasticity of Substitution	1.0	1.0
ν	Preference Weight for Rural Consumption Good	0.02	0.02
γ	Utility Weight of Housing	0.301	0.302
$\underline{c}$	Rural Consumption Good Subsistence Level	0.704	0.704
$\underline{s}$	Initial Urban Good Endowment	0.191	0.199
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	1.693	1.685
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.32): the baseline (resp. alternative) achieves a value of 0.0135 (resp. 0.0252).

Moment	Data	Model	Weight
housing_share_2010	0.306	0.3019	10.0
housing_share_1900	0.237	0.2417	10.0
rel_city_area_2010	0.173	0.1738	15.0
rural_emp_1840	0.6019	0.6772	1.0
rural_emp_1850	0.5625	0.6145	1.0
rural_emp_1860	0.5248	0.5382	1.0
rural_emp_1870	0.5018	0.4847	1.0
rural_emp_1880	0.4677	0.4942	1.0
rural_emp_1890	0.4433	0.4545	1.0
rural_emp_1900	0.4172	0.3907	0.01
rural_emp_1910	0.413	0.3821	0.01
rural_emp_1920	0.4149	0.3872	0.01
rural_emp_1930	0.3618	0.2977	0.01
rural_emp_1940	0.3573	0.2508	0.01
rural_emp_1950	0.2994	0.2112	0.01
rural_emp_1960	0.2255	0.1337	0.01
rural_emp_1970	0.1427	0.0783	0.01
rural_emp_1980	0.0914	0.0623	0.01
rural_emp_1990	0.0615	0.0445	0.01
rural_emp_2000	0.0432	0.0355	0.01
rural_emp_2010	0.0337	0.0339	0.01
rural_emp_2020	0.0313	0.0324	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

variables

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

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

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

1. The aggregate rural employment share, at each date t, $\frac{\sum_{k=1}^{K} L_{r,k,t}}{L_t}$.
2. The share of total urban land over total rural land, at each date t, $\frac{\sum_{k=1}^{K} \pi\phi_{k,t}^2}{\sum_{k=1}^{K}(S_k - \pi\phi_{k,t}^2)}$.
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 $\varsigma = (a, \gamma, \nu, \underline{s}, \underline{c})$ which is a member of set $\Xi \subset \mathbb{R}^5$ 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, 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 $t \in \{1840, 1850, \ldots, T\}$, where $T = 2350$.

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.[3]

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

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

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

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

$$\mathbf{x} = \left(\{S_{r,k}\}_{k=1}^{K}, \{L_{r,k}\}_{k=1}^{K}, \{L_{u,k}\}_{k=1}^{K}, r, p\right) \tag{B.27}$$

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 land 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 (27) 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.27)) 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).

$$\min_{\mathbf{x}_t, \{\theta_{u,j,t}\}_{j=1}^{K}, \{\theta_{r,j,t}\}_{j=1}^{K}} \quad \sum_{k=1}^{K} \left(\frac{L_{u,k,t}}{L_{u,1,t}} - \frac{\text{pop}_{k,t}}{\text{pop}_{1,t}} \right)^2 + \omega_{p,t} \sum_{k=1}^{K} \left(\frac{\tilde{\rho}_{k,t}}{\tilde{\rho}_{1,t}} - \frac{\bar{\rho}_{k,t}}{\bar{\rho}_{1,t}} \right)^2 \tag{B.28}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \frac{\text{pop}_{k,t}}{\sum_{j=1}^{K} \text{pop}_{j,t}} \theta_{u,k,t} = \theta_{u,t}, \tag{B.29}$$

$$\sum_{k=1}^{K} \frac{L_{r,k,t}}{\sum_{j=1}^{K} L_{r,j,t}} \theta_{r,k,t} = \theta_{r,t}, \tag{B.30}$$

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

This is a constrained optimization problem where the objective function (B.28) 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. It is important to notice two aggregation constraints which are added to this problem. Equation (B.29) 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.30) imposes the same on rural productivities. In other words, region-specific productivity parameters are constrained to generate

a path of sectoral aggregate productivity in line with aggregate data inputs described in Section B.2.2.1.

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

$$\begin{aligned} \min_{\mathbf{x}_t} \quad & g(\mathbf{x}_t) = 1, \quad t = 2030, \ldots, 2350 \\ \text{subject to} \quad & \text{(B.12)}, \text{(B.13)}, \text{(B.14)}, \text{(B.15)}, \text{(B.16)}, \text{(B.18)} \end{aligned} \tag{B.31}$$

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.26) 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:

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

We create a cartesian grid Ξ over this five-dimensional space and evaluate the model at each parameter value using the procedure described in B.2.4.4. The solution encounters infeasible points in a highly irregular fashion - in particular, non-monotonic in any particular parameter's space. We therefore employ a deep learning proceedure to impute the starting values for combinations of ς which result in infeasbilities.

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

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

$$\min_{\varsigma \in \Xi} L(\varsigma) = \min_{\varsigma \in \Xi} [m - m(\varsigma)]^T W [m - m(\varsigma)] \tag{B.32}$$

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 differential evolution optimizer.[4] 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.

[4] We use method `adaptive_de_rand_1_bin`, see 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. We focus on the model outcomes, which involve some computations and that are shown in the main Figures of the baseline simulation. The main non-targeted moments are displayed in Table B.5.

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 12a 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 12b and Figure 13a, 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 3 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 ratio $\frac{\bar{D}_{1870}}{\bar{D}_{2020}}$ and compare it to its data counterpart over the period 1870-2015—variable *avg_density_fall* displayed in Table B.5.

We proceeed in a similar fashion to compute central density and fringe density displayed in Figure 13a. We compute the central density 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 13a 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 13a 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 13a, 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.33 in Appendix A.2.4),

$$D_{k,n} \approx a_k \exp(-b_k \cdot \tilde{\ell}_{k,n}), \tag{B.33}$$

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	0.0	0.613
density_decay_coef	0.15	0.1855
avg_density_fall	8.57	6.4223
max_mode_increase	4.5244	4.707

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 14a 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,

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

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 14b.

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,

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

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 the city $\mathbb{W}^l_{u,k,t}$ 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 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 15a 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

$$\mathbb{W}^h_{u,k,t} = \int_0^{\phi_{k,t}} H_{k,t}(\ell)\mathbb{Q}_{k,t}(\ell)2\pi\ell d\ell, \tag{B.34}$$

and, similarly, in the rural part,

$$\mathbb{W}^h_{r,k,t} = \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \frac{S_{hr,k}}{S_{hr,k}+S_{r,k}} H_{k,t}(\ell)\mathbb{Q}_{k,t}(\ell)2\pi\ell d\ell, \tag{B.35}$$

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

$$\mathcal{H}_{k,t} = \int_0^{\phi} H_{k,t}(\ell)2\pi\ell d\ell + \int_{\phi_{k,t}}^{\sqrt{S_k/\pi}} \frac{S_{hr,k}}{S_{hr,k}+S_{r,k}} H_{k,t}(\ell)2\pi\ell d\ell, \tag{B.36}$$

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 15b, 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

$$RHPI_{k,t} = \frac{HPI_{k,t}}{\tilde{P}_t} = \frac{\mathbb{W}^h_{k,t}}{\mathcal{H}_{k,t}} \frac{1}{\tilde{P}_t}, \tag{B.37}$$

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:

$$\begin{aligned}
\mathbf{P}_{0,t} &= \frac{p_t Y_{r,t-1} + Y_{u,t-1}}{p_{t-1} Y_{r,t-1} + Y_{u,t-1}} \\
\mathbf{P}_{1,t} &= \frac{p_t Y_{r,t} + Y_{u,t}}{p_{t-1} Y_{r,t} + Y_{u,t}} \\
\Delta \mathbf{P}_t &= \sqrt{\mathbf{P}_{0,t} \mathbf{P}_{1,t}} \\
\tilde{P}_t &= \tilde{P}_{t-1} \Delta \mathbf{P}_t, t > 1 \\
\tilde{P}_1 840 &= p_1 840
\end{aligned} \tag{B.38}$$

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 15b, 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.34),(B.35), and (B.36) appropriately.

B.3 Sensitivity Analysis and Extensions

This section contains details of sensitivity analysis with respect to the elasticity of substitution between land and labor, σ 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 distance and commuting costs. For sensitivity analysis, all model's parameters but the one(s) on which sensitivity is performed are kept identical to their baseline values for comparison with the baseline quantitative evaluation. For the extensions, in order to give the best chances to these model's variations to fit the data, the model is entirely re-estimated (following the strategy detailed in Appendix B.2.4), including the estimation of all the region-specific productivity parameters.

B.3.1 Elasticity of substitution between land and labor σ

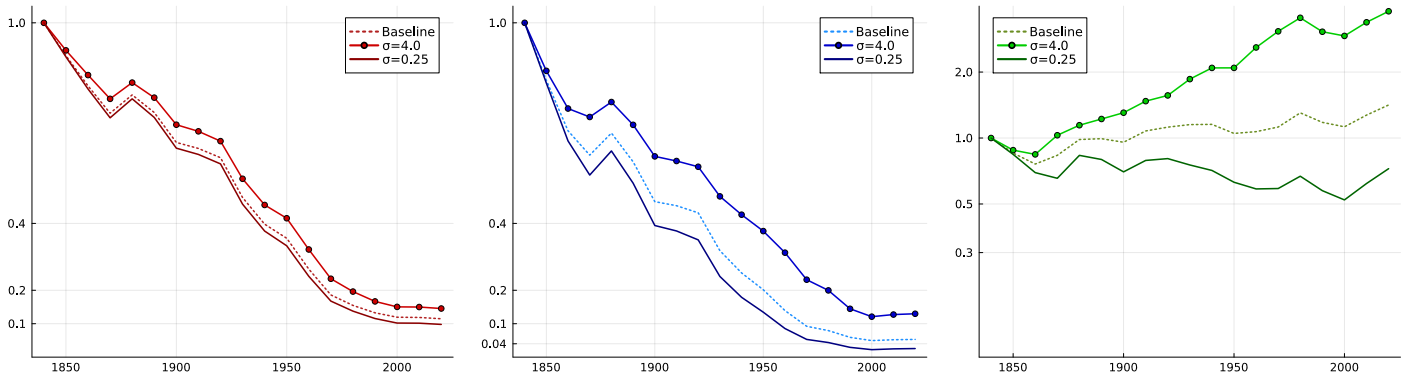
Our baseline simulation assumes a unitary elasticity of substitution between land and labor, $\sigma = 1$. Values used in the literature typically range between 0 and 1 (Bustos et al. (2016) and Leukhina and Turnovsky (2016)). We perform sensitivity analysis with a lower value of 0.25. We also show results for a high value of 4 to enlighten further the quantitative importance of the adjustment of land values at the fringe of the city for our results.[5] For this sensitivity analysis, we only change σ and keep all other parameters to their baseline values, in particular the estimated sequences $\{\theta_{u,k,t}, \theta_{r,k,t}\}$. Then, we re-solve the model system of equations without any data fitting efforts involved.

Results are displayed in Figure B.5 for variables of interest, where we focus on aggregate moments 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.5c). As the opportunity cost of expanding the city is lower, the urban area increases more and the average urban density falls more (Figure B.5a). This is driven by a larger fall of density in the cheaper suburban parts (Figure B.5b). With $\sigma = 0.25$, the model matches 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 ($\sigma = 4$), 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.3.2 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 $\epsilon(\phi_k) = 5$.

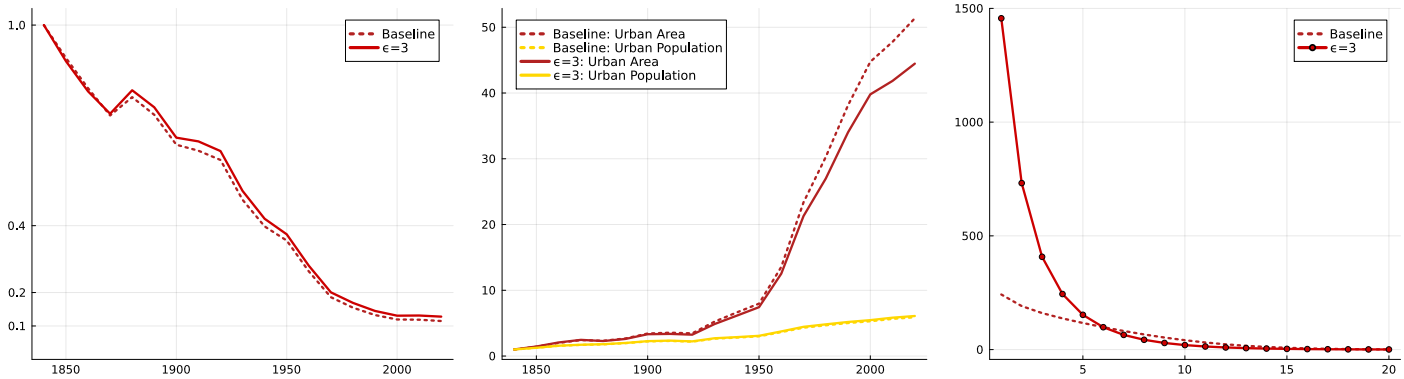
[5] A higher σ limits the fall of farmland values at the fringe of cities when workers move towards the urban sector.



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

Figure B.5: Sensitivity to the elasticity of substitution between land and labor σ.

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



(a) Average urban density (1840=1). (b) Urban Area and Population. (c) Density by decile (2020).

Figure B.6: 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.

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, including region-specific productivities.

After solving the model, results are displayed in Figure B.6 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 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.6a). The same implication can be seen in Figure B.6b, where urban area is shown to be 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.6c 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.3.3 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 $\mu = 0.05$ and fully re-estimate the model using the same strategy as for the baseline described previously in Section B.2.4, including the region-specific productivity parameters aiming at matching the distribution of urban populations and local farmland prices.

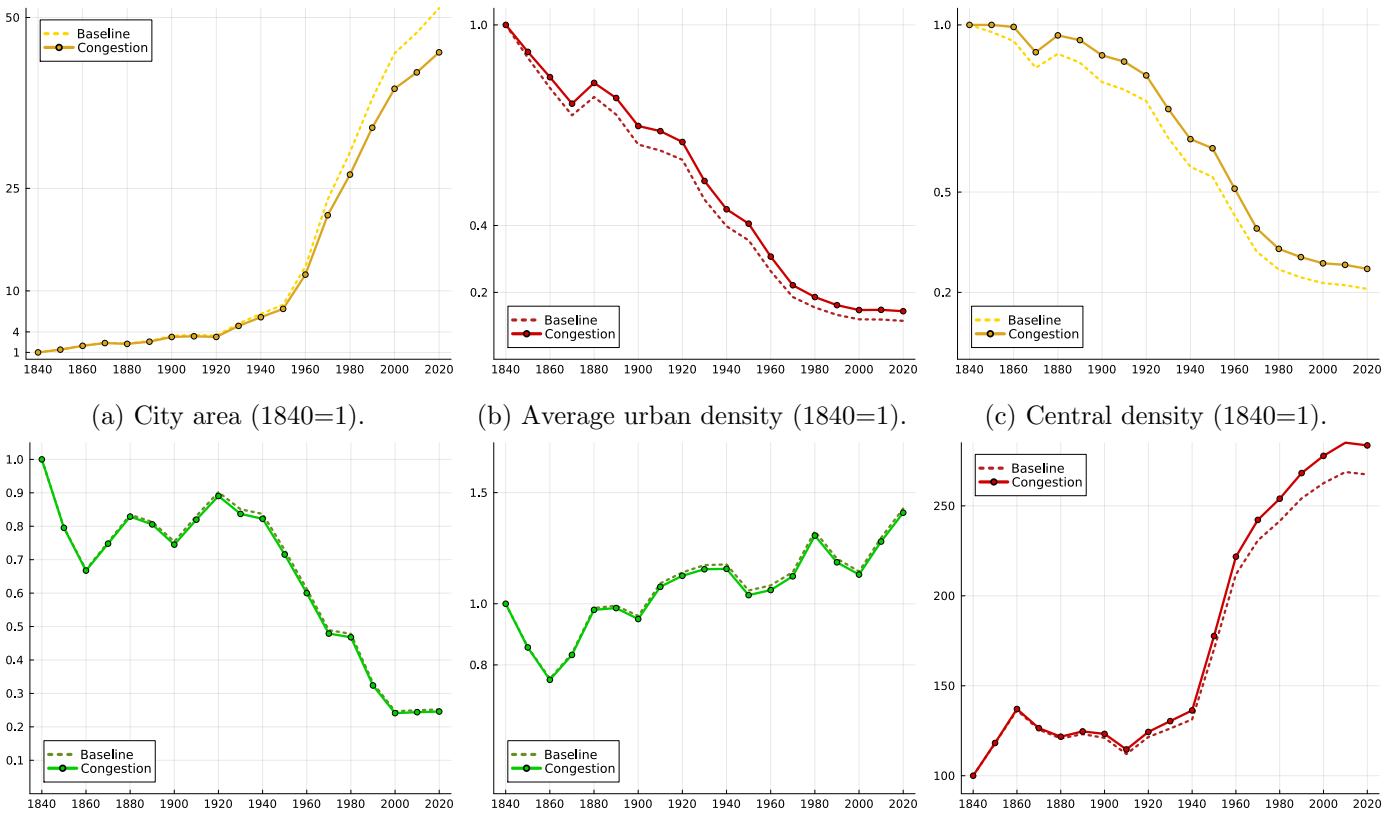
Focusing on the aggregate implications, the evolution of the variables of interest is shown in Figure B.7 together with the baseline results for comparison. Congestion forces reduce the expansion in area and the extent of suburbanization (Figure B.7a). By increasing commuting costs, they also increase urban housing prices (Figure B.7f). 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)). Then, we fully re-estimate the model using the same strategy described in Section B.2.4, including the region-specific productivity parameters.

Irrelevance Result. Using $\lambda = 0.05$, we re-estimate the model using the same strategy as for the baseline described in B.2.4.5. 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 targets and matches almost perfectly their aggregate



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

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

Figure B.7: 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 based on the re-estimation of all the parameters following the methodology described in Section B.2.4.

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.[6] This is so because, despite agglomeration forces, the model implied aggregate urban productivity,

$$\theta_{u,t} = \sum_{k=1}^{K} \left(\theta_{u,k,t} \cdot L_{u,k,t}^{\lambda} \cdot \left(\frac{L_{u,k,t}}{L_{u,t}} \right) \right) = \left(\sum_{k=1}^{K} \theta_{u,k,t} \cdot \left(\frac{L_{u,k,t}}{L_{u,t}} \right)^{1+\lambda} \right) \cdot L_{u,t}^{\lambda}, \tag{B.39}$$

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 model 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 due to agglomeration forces. While it is well known that in the cross-section, larger cities are more productive, which make 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' agglomeration forces following an aggregate urban expansion along the process of structural change.[7]

[6] They are not exactly identical because we do not match perfectly aggregate urban employment and the aggregate population of cities (see Figure B.11).

[7] 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

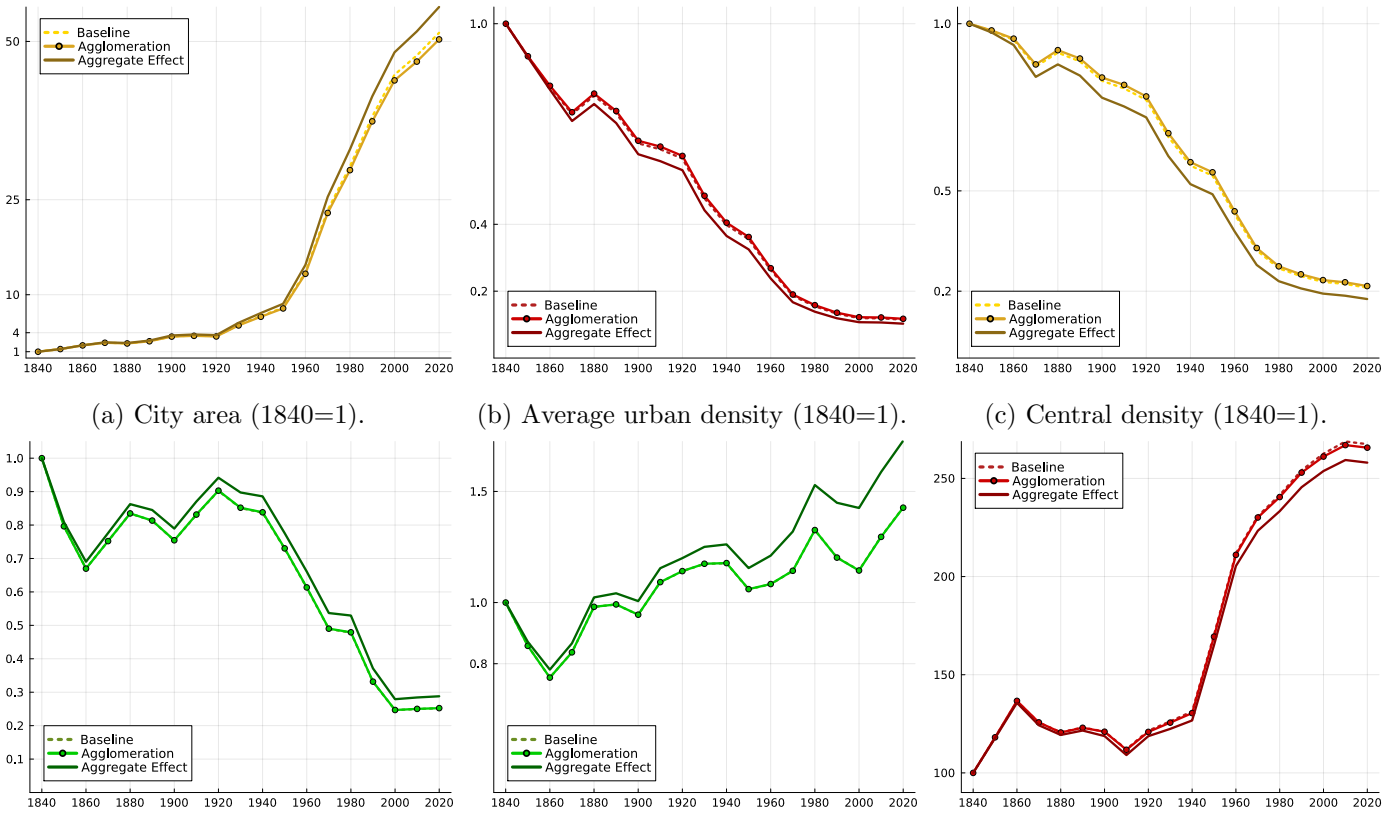
To do so, we re-estimate the model fitting as aggregate productivity a modified version of Equation (B.39),

$$\theta_{u,t} = \sum_{k=1}^{K} \theta_{u,k,t} \cdot \left(\frac{L_{u,k,t}}{L_{u,t}} \right)^{1+\lambda}, \tag{B.40}$$

In other words, all the model's parameters are re-estimated in the same way as described in B.2.4 but targeting for French aggregate productivity $\theta_{u,t}$ in Eq. (B.40), 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.8 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 no 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.8a—B.8c). However, a higher urban income makes also rural goods more valuable increasing rural workers' wage almost one for one (Figure B.8d). 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.8e). 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.

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 area (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.8: Agglomeration forces.

Notes: The solid lines represents outcomes in presence of agglomeration forces under both specification of aggregate productivity (Eq. B.39 and Eq. B.40). The line with circles corresponds to the "irrelevance result', where aggregate productvity matches the data, and we generate cross sectional differences in urban productivity—which are basically undone by a new set of $\theta_{u,k,t}$. The line without markers shows equilibrium effects of 'aggregate' agglomeration forces relative to the baseline, which is shown as dotted line. Model outcomes with agglomeration forces are based on the re-estimation of all parameters following the methodology described in Section B.2.4.

B.3.4 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.[8] 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,

$$d_{k,t}(\ell_k) = d_0(\phi_{k,t}) + d_1(\phi_{k,t}) \cdot \ell_k, \tag{B.41}$$

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,[9]

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

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.3.4, 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.3.4—with an estimated mode across urban areas close 0.7 in the data. We calibrate $d_1 = 2.25$ externally. This yields after model's estimation a slope coefficient of that varies

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

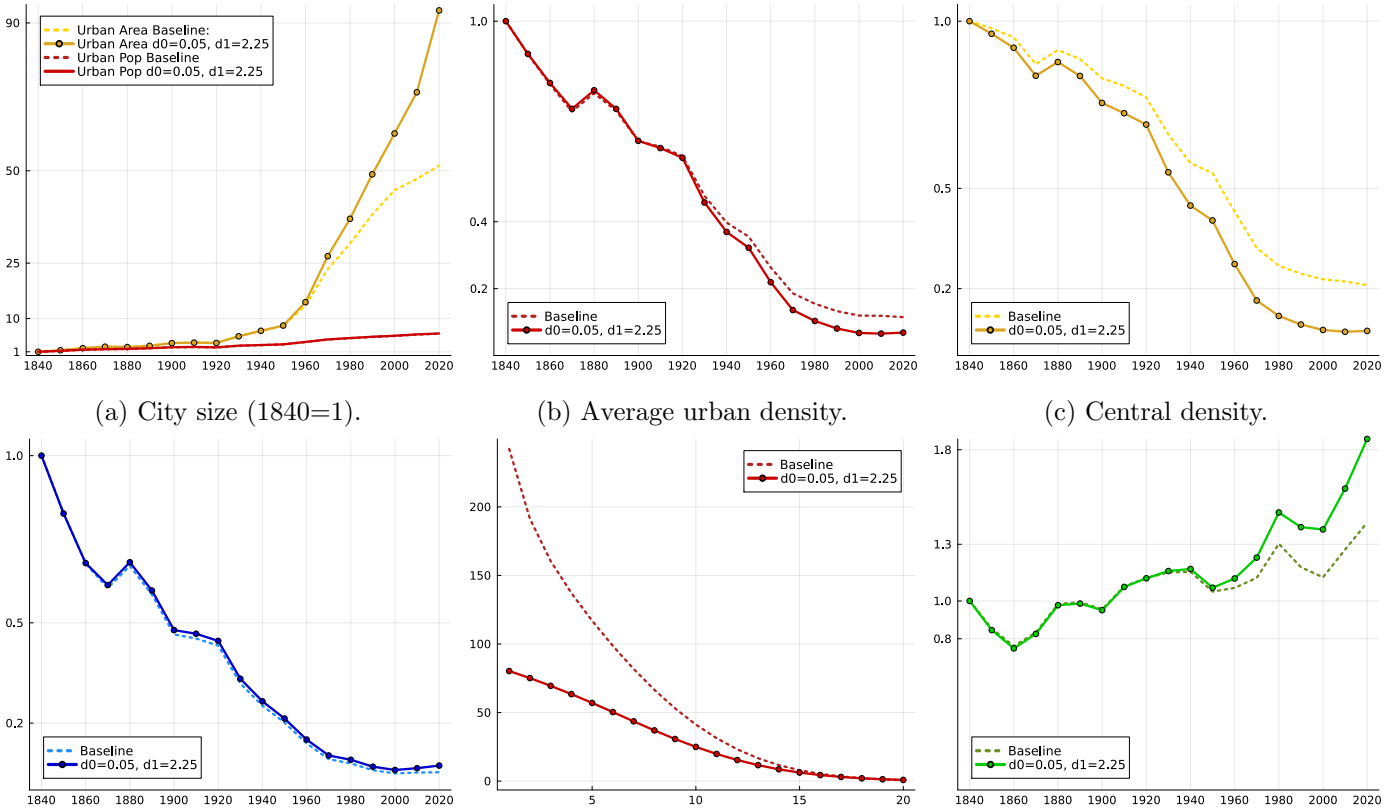
[9] This remains consistent with our calibrated value of ξ_ℓ estimated using commuting distance. The elasticity of speed $m(\ell)$ to commuting distance $d(\ell)$ being $1 - \xi_\ell$.

across cities, ranging from 0.44 for Paris to an average amongst the remaining cities of 0.67, which is reasonably close to the corresponding empirical moment.

Estimation. The model is fully re-estimated in order to fit aggregate and cross-sectional moments as in the baseline version, i.e. the procedure is identical to Section B.2.4, up to the fact that we provide a different starting value and search range for parameter a, which needs to be substantially higher in order to match the aggregate area extent of cities in 2010.

Results. We find that our results are not much affected (Figure B.9). 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.9a). As a consequence of this larger sprawling, the average urban density falls more (Figure B.9b). This is driven by a larger fall of central density, the most noticeable difference relative to our baseline monocentric model (Figure B.9c). 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.9e).

This specification provides also a better fit of the data across cities (Figure B.10). Relative to the baseline monocentric model (in Figure B.11 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.

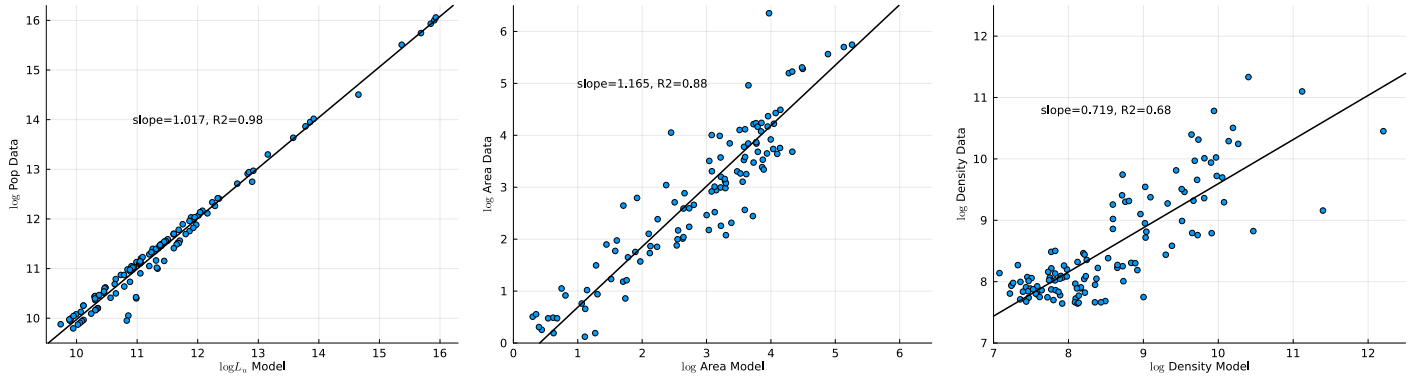


(a) City size (1840=1). (b) Average urban density. (c) Central density.

(d) Fringe density. (e) Density Gradient (2020). (f) Rural Rents.

Figure B.9: Relaxing monocentricity. Aggregate Moments.

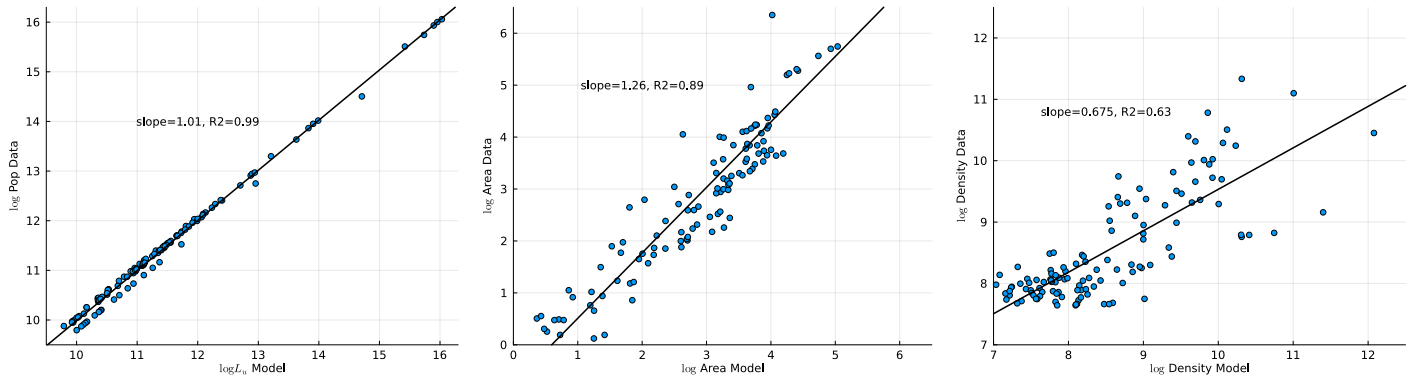
Notes: The solid line represents outcomes in the extended model with alternative commuting costs ((d_0, d_1) extension). For comparison, outcomes of the baseline simulation are shown with a dotted line. Model's outcomes under this alternative specification of commuting costs based on the re-estimation of all the parameters following the methodology described in Section B.2.4.



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

Figure B.10: 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 $t \in \{1870, 1950, 1975, 1990, 2000, 2015\}$. Model's outcomes based on the re-estimation of all the parameters following the methodology described in Appendix B.2.



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

Figure B.11: 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 $t \in \{1870, 1950, 1975, 1990, 2000, 2015\}$. 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," *Work in progress, University of Toronto*, 2019.

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.

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

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.

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