The Costs of Agglomeration: House and Land Prices in French Cities

Pierre-Philippe Combes[*†]
University of Lyon and Sciences Po

Gilles Duranton[*‡]
University of Pennsylvania

Laurent Gobillon[*§]
Paris School of Economics, CNRS

Final version: October 2018

ABSTRACT: We develop a new methodology to estimate the elasticity of urban costs with respect to city population using French house and land price data. After handling a number of estimation concerns, we find that the elasticity of urban costs increases with city population with an estimate of about 0.03 for an urban area with 100,000 inhabitants to 0.08 for an urban area of the size of Paris. Our approach also yields a number of intermediate outputs of independent interest such as the share of housing in expenditure, the elasticity of unit house and land prices with respect to city population, and within-city distance gradients for house and land prices.

Key words: urban costs, house prices, land prices, land use, agglomeration

JEL classification: R14, R21, R31

[*]We thank four anonymous referees, the editor Stéphane Bonhomme, conference and seminar participants, Monica Andini, Fabien Candau, Morris Davis, Jan Eeckhout, Sanghoon Lee, François Ortalo-Magné, Gilles Orzoni, Henry Overman, Jean-Marc Robin, Stuart Rosenthal, Nathan Schiff, Daniel Sturm, and Yuichiro Yoshida for their comments and suggestions. We also thank Pierre-Henri Bono, Julian Gille, Giordano Mion, and Benjamin Vignolles for their help with the data. Finally, we are grateful to the *Service de l'Observation et des Statistiques (SOeS) - Ministère de l'Écologie, du Développement durable et de l'Énergie* for giving us on-site access to the data and to the CASD (Centre d'accès sécurisé aux données founded by the French National Research Agency, ANR-10-EQPX-17) for remote access to the French Family Expenditure Survey.

[†]University of Lyon, CNRS, GATE-LSE UMR 5824, 93 Chemin des Mouilles, 69131 Ecully, France and Sciences Po, Economics Department, 28, Rue des Saints-Pères, 75007 Paris, France (e-mail: ppcombes@gmail.com; website: https://www.gate.cnrs.fr/ppcombes/). Also affiliated with the Centre for Economic Policy Research.

[‡]Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA (e-mail: duranton@wharton.upenn.edu; website: https://real-estate.wharton.upenn.edu/profile/21470/). Also affiliated with the Centre for Economic Policy Research and the National Bureau of Economic Research.

[§]Paris School of Economics, CNRS, 48 Boulevard Jourdan, 75014 Paris, France (e-mail: laurent.gobillon@psemail.eu; website: http://laurent.gobillon.free.fr/). Also affiliated with the Centre for Economic Policy Research and the Institute for the Study of Labor (IZA).

© The Author(s) 2018. Published by Oxford University Press on behalf of The Review of Economic Studies Limited.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

1. Introduction

As a city's population grows, three major changes potentially occur. First, larger cities are expected to be more productive as agglomeration effects become stronger. Second, larger cities are expected to become more expensive as the cost of housing and urban transport rises. The price of other goods may also be affected. Third, larger cities may differ in how attractive they are in terms of amenities. From past research, we know a fair amount about agglomeration and we have some knowledge about urban amenities. However, we know virtually nothing about urban costs and how they vary with city population. Although high housing prices and traffic jams in Central Paris, London, or Manhattan are for everyone to observe, we know of no systematic evidence about urban costs and their magnitude. This paper seeks to fill that gap.

To that end, we develop a new methodology to estimate the elasticity of urban costs with respect to city population using French data about house and land prices and household expenditure. Our baseline estimates range from about 0.03 for an urban area with 100,000 inhabitants to 0.08 for an urban area of the size of Paris. Put differently, a 10% larger population in a small city leads to a 0.3% increase in expenditure for its residents to remain equally well off. For a city with the same population as Paris, the same 10% increase in population implies a 0.8% increase in expenditure. These figures are 'all else constant', including the urban area of cities. Allowing cities to increase their physical footprint as they grow in population reduces the magnitude of the elasticity of urban costs by a factor of about two. In the 'short run', we estimate instead larger elasticities in the 0.1-0.3 range as housing supply adjusts sluggishly to population increases. Our approach also yields a number of intermediate outputs of independent interest such as within-city distance gradients for land and house prices, the share of housing in expenditure, and the elasticities of land and house prices with respect to city population.

Plausible estimates for urban costs are important for a number of reasons. In many countries, urban policies attempt to limit the growth of cities by imposing barriers to labour mobility and stringent land use regulations that limit new constructions. The underlying rationale for these policies is that the population growth of cities imposes large costs to already established residents by bidding up housing prices and crowding out the roads. Our analysis shows that in the French case, the costs of having larger cities are modest for most cities and of about the same magnitude as agglomeration economies. This lends little support to the imposition of barriers to urban growth.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Quite the opposite, urban costs increase much faster when cities are prevented from adjusting their supply of housing.

More generally, households allocate a considerable share of their resources to housing and transport. In France, homeowners and renters in the private sector devote on average 33.4% of their expenditure to housing and 13.5% to transport.[1] As we document below, there are sizeable differences across cities in how much households spend on housing as its cost varies greatly across places. Understanding this variation is thus a first-order allocation issue.

Urban costs also matter for how we think about cities in theory. Following Henderson (1974) and Fujita and Ogawa (1982), cities are predominantly viewed as the outcome of a tradeoff between agglomeration economies and urban costs. Much of contemporary urban theory relies or builds on this tradeoff. Fujita and Thisse (2002) dub it the 'fundamental tradeoff of spatial economics'. The existence of agglomeration economies is now well established and much has been learnt about their magnitude.[2] To assess the fundamental tradeoff of spatial economics empirically, evidence about urban costs is obviously needed.

To measure how urban costs vary with city population, three challenges must be met. The first regards the definition and measurement of urban costs since they can take a variety of forms. Using consumer theory, we define the elasticity of urban costs with respect to city population as the percentage increase in expenditure that residents in a city must incur when population grows by one percent, keeping utility constant. At a simple spatial equilibrium, this elasticity is equal to the product of the share of housing in expenditure and the elasticity of housing prices with respect to city population, both taken at the city centre.[3] We also show that the elasticity of housing prices can be decomposed into the product of the share of land in housing construction and the population elasticity of land prices.

After this conceptual clarification, our second challenge is to gather data to implement our

[1]Our figure of 33.4% for housing is the mean between the figure for renters and the figure for homeowners for 2006-2011 in the French expenditure survey. It is higher than the aggregate share of housing in expenditure of 27% reported by CGDD (2015) because we exclude rural areas where housing is less expensive and renters living in public housing who often pay well below market price. The figure for transport is from 2010 and covers the entire country (CGDD, 2015). In the US, households devote 32.8% of their expenditure to housing and 17.5% to transport (US BTS, 2013). In both countries, transport is defined as all forms of personal transport but most of it is road transport. Air transport represents only 6% of transport expenditure in France and 5% in the US.

[2]See Puga (2010) and Combes and Gobillon (2015) for reviews. See also Combes, Duranton, and Gobillon (2008), Combes, Duranton, Gobillon, and Roux (2010), or Combes, Duranton, Gobillon, Puga, and Roux (2012) for some work on French cities.

[3]At the equilibrium, higher housing costs offset lower transport costs as we consider locations closer to the centre. Then, we work with prices at the centre because we can, to a first approximation, ignore travel costs for these locations.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

approach empirically. For housing prices, we rely on detailed price indices that are estimated for French municipalities between 2000 and 2012. For land prices, we exploit a unique record of transactions for land parcels with a development permit from 2006 to 2012. For housing expenditure we use a household expenditure survey. For the share of land in housing, we rely on the results obtained in our companion paper (Combes, Duranton, and Gobillon, 2016) which provides a detailed investigation of the production function for housing. Finally, we gathered a vast array of data at the level of municipalities and urban areas.

Our third challenge is the actual estimation of our key elasticities and shares. For the elasticity of both housing and land prices at the centre with respect to city population, we first need to estimate housing and land prices at the centre of each city. This first exercise poses one main difficulty, estimating an appropriate distance gradient for each city. We show that our results are robust to how we handle the distribution of heterogenous residents within cities and to our choices of functional form, specification, and city centres.

Next, when regressing housing and land prices at the centre on city population, our main worry is the endogeneity of city population. We employ a variety of approaches to assess the robustness of our baseline results, including extensive control variables at both the municipality and city level and instrumental variables. We also show that house and land prices both imply similar estimates for the elasticity of urban costs. Finally, we also address a number of related endogeneity concerns regarding the estimation of the share of housing in expenditure and how it varies with city population.

Tolley, Graves, and Gardner (1979), Thomas (1980), Richardson (1987), Henderson (2002), and Au and Henderson (2006) are the main antecedents to our research on urban costs.[4] To the best of our knowledge, this short list is close to exhaustive. Despite the merits of these works, none of their estimates has had much influence. We attribute this lack of credible estimate for urban costs and the scarcity of research on the subject to a lack of integrated framework to guide empirical work, a lack of appropriate data, and a lack of attention to a number of identification issues — the three main innovations of this paper.

[4]Thomas (1980) compares the cost of living for four regions in Peru focusing only on the price of consumption goods. Richardson (1987) compares 'urban' and 'rural' areas in four developing countries. Closer to the spirit of our work, Henderson (2002) regresses commuting times and rents to income ratio for a cross-section of cities in developing countries. Like us, Au and Henderson (2006) are interested in the tradeoff between agglomeration benefits and urban costs. They use nonetheless a very different approach and investigate the net productivity gains associated with city size instead of trying to separate the costs from the benefits of cities.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

There is also research that partly overlaps with what we do. Albouy (2008), Bleakley and Lin (2012), and Baum-Snow and Pavan (2012) estimate the elasticity of housing prices with respect to city population but do so less systematically and with very different objectives in mind. Davis and Heathcote (2007), Davis and Palumbo (2008), and Albouy, Ehrlich, and Shin (2019) measure land values for a broad cross-section of urban (and sometimes rural) areas but, again, they do so with very different objectives in mind.

2. Model

We want to estimate how the cost of living in cities increases with their population. To provide a rigourous definition of urban costs and some guidance about how to estimate them empirically, we consider a model where households choose in which city to live and work, where to reside in this city, and how much housing and other goods to consume at their chosen location.

The utility of a resident at location ℓ in city c with population N_c is given by $U(h(\ell),x(\ell),M_c)$ where M_c denotes the quality of amenities in the city, $h(\ell)$ is housing consumption, and $x(\ell)$ is the consumption of a composite good. Utility is increasing in all its arguments and is strictly quasi-concave. The budget constraint is,

$$W_c \geq P(\ell)\, h(\ell) + \tau(\ell) + Q_c\, x(\ell)\,, \tag{1}$$

where W_c is the wage that prevails in city c, $P(\ell)$ is the price of housing at location ℓ, $\tau(\ell)$ is the cost of transport at the same location, and Q_c is the city price of the composite consumption good.[5]

We can solve the consumer problem in steps. First, households choose a city. Then, they choose a residential location ℓ in their city. Finally, at their chosen location residents maximise their utility with respect to their consumption of housing $h(\ell)$ and their consumption of the composite good $x(\ell)$ subject to the budget constraint (1). We start with this last step and consider its dual. Omitting the city subscript c, we note the expenditure function for a resident at location ℓ as $E(P(\ell),\tau(\ell),Q,M,\overline{U}) = P(\ell)\, h(\ell) + \tau(\ell) + Q\, x(\ell)$. This function describes the minimum total expenditure on housing, transport, and the composite consumption good needed at location ℓ to achieve utility $\overline{U}$.

[5] A special case of our model is the monocentric model of Alonso (1964), Mills (1967), and Muth (1969). In this model, ℓ measures the distance to the central business district (CBD) where all the jobs are located. Residents must commute to this CBD at a cost $\tau(\ell) = \tau \times \ell$. The results that follow do not rely on these restrictions.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

We can now examine the effect of a marginal increase in city population on the resident located at location ℓ. Totally differentiating the expenditure function with respect to population leads to,

$$\begin{aligned}\frac{\mathrm{d}E(P(\ell),\tau(\ell),Q,M,\overline{U})}{\mathrm{d}N} &= \frac{\partial E(P(\ell),\tau(\ell),Q,M,\overline{U})}{\partial P(\ell)}\frac{\mathrm{d}P(\ell)}{\mathrm{d}N}+\frac{\mathrm{d}\tau(\ell)}{\mathrm{d}N} \\ &+ \frac{\partial E(P(\ell),\tau(\ell),Q,M,\overline{U})}{\partial Q}\frac{\mathrm{d}Q}{\mathrm{d}N}+\frac{\partial E(P(\ell),\tau(\ell),Q,M,\overline{U})}{\partial M}\frac{\mathrm{d}M}{\mathrm{d}N}. \qquad (2)\end{aligned}$$

This equation indicates that, for a given location ℓ, the change in expenditure that is needed to keep utility constant following a change in city population works through four channels: the change in expenditure that arises from the change in housing prices at location ℓ, the change in transport cost at location ℓ (e.g., more congestion), the change in expenditure due to the change in the price of the composite good, and the change in expenditure associated with the change in amenities.

Applying Shephard's lemma to equation (2) and omitting the arguments of the expenditure function to ease notations, we obtain,

$$\frac{\mathrm{d}E}{\mathrm{d}N}=h(P(\ell),Q,\overline{U})\frac{\mathrm{d}P(\ell)}{\mathrm{d}N}+\frac{\mathrm{d}\tau(\ell)}{\mathrm{d}N}+x(P(\ell),Q,\overline{U})\frac{\mathrm{d}Q}{\mathrm{d}N}+\frac{\partial E}{\partial M}\frac{\mathrm{d}M}{\mathrm{d}N}, \qquad (3)$$

where $h(P(\ell),Q,\overline{U})$ is the compensated demand for housing in ℓ and $x(P(\ell),Q,\overline{U})$ is the compensated demand for the composite good at the same location. To simplify the exposition, assume without loss of generality that we measure amenities so that the elasticity of expenditure with respect to amenities is minus one: $\frac{\partial E}{\partial M}=-\frac{E}{M}$.[6] More concretely, our choice of units for amenities is such that a 1% decrease in amenities requires a 1% increase in consumption expenditure to keep utility constant. Using this normalisation and dividing both sides by E/N, we can rewrite equation (3) more compactly as:

$$\epsilon_N^E=\epsilon_N^{UC}(\ell)-\epsilon_N^M \qquad (4)$$

where

$$\epsilon_N^{UC}(\ell)\equiv s_E^h(\ell)\epsilon_N^{P(\ell)}+s_E^\tau(\ell)\epsilon_N^{\tau(\ell)}+s_E^x(\ell)\epsilon_N^Q, \qquad (5)$$

ϵ_Y^X is the elasticity of X with respect to Y, and $s_E^X(\ell)$ is the expenditure share of X.

The empirical work that follows is concerned with the estimation the elasticity of urban costs with respect to city population $\epsilon_N^{UC}(\ell)$ described in equation (5). It essentially asks how much more costly it becomes to live at a location when city population increases. As made clear by equation (5), a change in urban costs includes three components: a change in house prices, a change in

[6]This equality will holds regardless of the choice of units when amenities enter the utility function in a multiplicatively separable way.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

transport costs, and a change in the price of the composite good. Each of these three component is an elasticity of an expenditure with respect to population weighted by its corresponding budget share. A complication is that equation (5) defines an elasticity of urban costs $\epsilon_N^{UC}(\ell)$ for each location ℓ within the city since five of the six terms that enter its calculation depend on location ℓ. The equilibrium choice of residential location within a city by residents allows us to simplify this elasticity, as we detail now.

At the spatial equilibrium, the rental price of housing within a city adjusts so that residents are indifferent across all occupied residential locations in the city: $U(h^*(\ell),\tau(\ell),x^*(\ell),M) = \overline{U}$. Because the expenditure is equal to the city wage in equilibrium and because amenities are not location-specific within a city, the urban costs elasticity must be the same for all locations within a city as per equation (4). We can thus measure the urban costs elasticity for an entire city using a single location. Given the data at hand, it is useful to consider the 'central' location of each city where the price of housing is the highest, $\overline{P}$. In equilibrium, this is also the location where the transport cost is the lowest, $\underline{\tau}$.

We now make two simplifications, which we discuss further below. First, as in many models of urban structure, we assume that $\underline{\tau} = 0$. In a monocentric urban model, this corresponds to the central resident who does not pay any commuting cost. Second, we assume free trade between cities for the composite good so that $\epsilon_N^Q = 0$. This allows us to simplify equation (5) and write the urban costs elasticity as:

$$\epsilon_N^{UC} = s_E^h \, \epsilon_N^{\overline{P}} \, . \tag{6}$$

The elasticity of urban costs with respect to city population is now the product of only two terms, the share of housing in expenditure and the elasticity of the price of housing with respect to city size. Both are measured at a 'central' location, $\overline{\ell}$, where the price of housing is the highest.

We finally turn to the first decision made by residents: the choice of a city. Under free mobility across cities, utility $\overline{U}$ is achieved in all cities in equilibrium, which allows us to infer the urban cost elasticity from comparisons across cities.[7]

[7]Returning to expression (4) and using again the fact that in equilibrium the city wage is equal to total expenditure, it is easy to see that the urban costs elasticity minus the wage elasticity is equal to the 'amenity' elasticity: $\epsilon_N^{UC}(\ell) - \epsilon_N^W = \epsilon_N^M$. As a city grows in population, we expect urban costs and wages to increase. At the spatial equilibrium between cities, if urban costs increase faster than wages, the difference must be made up by better amenities. Put differently, knowing about the agglomeration elasticity ϵ_N^W and the urban costs elasticity ϵ_N^{UC} and assuming a spatial equilibrium across cities, we can recover the amenities elasticity. This is consistent with the approach proposed by Roback (1982) and the large literature that followed, most notably Albouy (2016) who focuses on how urban amenities vary with city population. Our innovation lies in a more precise specification of urban costs and the development of an empirical strategy to measure them.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

In Appendix 1, we extend this model to consider a competitive housing production sector to show that the elasticity of housing price with respect to population can be decomposed into the product of the elasticity of land prices with respect to population and the share of land in housing production. We can thus rewrite equation (6) as $\epsilon_N^{UC} = s_E^h \, s_h^L \, \epsilon_N^{\overline{R}}$ where s_h^L is the share of land in housing and $\epsilon_N^{\overline{R}}$ is the population elasticity of land prices at the most expensive location in the city.

We acknowledge a number of limitations. First and as already mentioned, our final expression for the urban costs elasticity relies on two simplifications. Assuming zero minimum transport costs in the city is perhaps a reasonable first-order approximation in the centre of cities where a non-negligible share of residents report very low travel times for the trips they undertake.[8] Assuming constant prices for the composite consumption good is another empirically defensible first-order approximation. Work by Handbury and Weinstein (2015) strongly suggests that the price of individual varieties in groceries is mostly invariant with city population in the US.[9] Using broader product categories, Combes *et al.* (2012) confirm this result for French cities.

Second, we rely on a standard spatial equilibrium concept involving utility equalisation among homogeneous residents. We acknowledge the limitations of this type of approach but note that theoretical developments where the spatial equilibrium does not involve full utility equalisation are still in their infancy (e.g., Behrens, Duranton, and Robert-Nicoud, 2014) and empirical applications are also at early stages of development (Kline and Moretti, 2014). Empirically, we take two approaches to household heterogeneity within and across cities. We gather a lot of data about household characteristics at a fine spatial scale to condition out as much heterogeneity as we can in our estimations. We also experiment with specifications that allow for heterogeneous effects.

Third, our model is static and abstracts from housing tenure choices. Homeowners actually benefit when their house becomes more expensive as population grows. Our measure of urban costs is nonetheless the relevant one when residents need to choose a new location.[10]

Finally, we ignore fiscal issues. We expect them to affect location choices mostly through the

[8]For the US, we can use the same individual travel data as Duranton and Turner (2016). Among residents of US metropolitan areas with a million inhabitants or more who live within 2 kilometres of the CBD, 25% of them also live within one kilometre of their workplace and the median distance to work is 3 kilometres. For those living more than 20 kilometres away from their CBD, the 25th percentile of distance to work is above 5 kilometres and the median is 11 kilometres.

[9]They also find that larger cities offer a larger number of varieties, which we think of here as a consumption amenity.

[10]Then, tenure choice may be driven by a variety of factors. For instance residents may choose to buy instead of rent because they want to hedge themselves against future unforeseen changes in rents (Sinai and Souleles, 2005). We do not expect tenure choices to have a first-order effect on the choice of cities by residents (unlike house prices, amenities, and wages). Note also that we take tenure choice explicitly into account when estimating the share of housing in expenditure.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

agglomeration externality. In particular, the taxation of income implies that the agglomeration benefits of large cities are taxed which may distort location choices and lead to insufficient agglomeration (Albouy, 2009). However, the urban costs elasticity in expression (5) should not be directly affected.[11] A number of further issues including land use regulations and amenities that bear more directly on our estimations are discussed below.

To summarise, we develop a consumer-theoretic approach to define the elasticity of urban costs with respect to city population. This elasticity sums three price elasticities for housing, transport, and other goods, weighting them by their expenditure shares. We then rely on a free-trade assumption and a property of our spatial equilibrium for which we assume no commuting at the centre to simplify our expression of the urban costs elasticity into the product of the population elasticity of house prices at the most expensive location and the share of housing in expenditure at this location. In turn, the empirical estimation of the urban cost elasticity implies three separate empirical exercises. The first is to measure unit house prices consistently in cities at a central location. The second is to estimate the elasticity of house prices with respect to city population. The third is to estimate the share of housing in expenditure at the same central location. We conduct these three empirical exercises below. We also conduct our first two exercises for land prices in addition to house prices to check the consistency of our results.

3. Data

To estimates urban costs, we exploit three main sources of data for housing prices, land prices, and housing expenditure, which we describe in turn. We also use a broad range of municipal and urban area characteristics. We provide further details about the data in Appendix 2.

As main units of analysis, we use French urban areas. Our main sample contains 277 urban areas for which we can estimate housing price at the centre and have a complete set of characteristics.[12] Within urban areas, we work with municipalities. These municipalities are tiny. They correspond to a circle with a radius of 2.0 kilometres on average. Urban areas in our main sample contain on average 46 municipalities.

[11] A possible indirect effect relates to the fact that owner-occupiers are in general not taxed on their implicit housing rent, which may impact their capitalisation into property values. We leave this for future research.

[12] In total, 352 urban areas are delineated from the 1999 census in mainland France. The 75 urban areas that we lose (not urbanised in 1831, a necessary condition to compute our historical instruments), all have a population below 80,000 and 50 of them have a population below 25,000.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Housing prices

To measure housing prices, we use indices estimated at the municipality level from official transactions records. These transactions data are available from the Ministry of Sustainable Development for every even year over the 2000-2012 period. For each transaction, we know the type of dwelling (house or apartment), the number of rooms, floorspace area, and the construction period (before 1850, 1850-1913, 1914-1947, 1948-1959, 1960-1980, 1981-1991, after 1991), and a municipal identifier.

To construct municipal housing price indices, we regress the log of the price per square metre on indicator variables for the construction period and for the quarter of the transaction. We estimate a separate regression for every available year. We then compute housing price indices as the average of the residuals for each municipality and year after adding the regression constant. Since the explanatory variables are centred, we can interpret the resulting indices as a price per square metre of floorspace for a reference house or dwelling. Note that we first estimate housing price indices before using them as an input in our main analysis. As explained in Appendix 2, institutional constraints prevent us from merging data on property prices with other data. This stands in contrast to what we do with parcel prices, which we use directly into the analysis. We do not expect this difference to matter.

To allow for easier comparisons with our land price results, we mainly focus on price indices for single-family houses. In robustness checks, we duplicate our results using indices for all dwellings (houses and apartments). For houses, there are 184,371 municipality-year observations corresponding to 1,848,081 transactions that took place in mainland France. For our main sample with 277 urban areas, we end up with 74,621 observations corresponding to 1,199,506 transactions.

To measure distance to the centre of an urban area, our preferred metric is the log of the Euclidean distance between the centroid of the municipality of the transaction and the centroid of its urban area. To determine urban area centroids, we weigh municipalities by their population. In robustness checks, we use alternative distance metrics, definitions of urban area centres, and allow for more than one centre in each urban area.

Land prices

We use land price data extracted from the 2006-2012 Surveys of Developable Land Prices (*Enquête sur le Prix des Terrains à Bâtir*, EPTB) in France. An observation is a transaction record for a parcel

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

of land with a building or rebuilding permit for a detached house. Before 2010, around 2/3 of all building permits were surveyed. From 2010 onwards, all building permits are surveyed and the response rate is about 70%.[13] Overall, the land price data contain 662,060 observations with some fluctuations across years from 48,991 in 2009 to 127,479 in 2012. As discussed in Combes *et al.* (2016), this survey tracks the bulk of new constructions for single-family houses in France.

For each transacted parcel, we know its price, its municipality, its area, and a number of other characteristics. These characteristics include the type of transaction (purchase, donation, inheritance, other), whether the parcel was acquired through an intermediary (a broker, a builder, another type of intermediary, or none), and whether a parcel was 'serviced' (i.e., had access to water, sewerage, and electricity). We also know the construction cost of the house that was subsequently built.

We restrict our attention to purchases and ignore other transactions such as inheritances for which the price is unlikely to be informative. That leaves us with 394,818 observations for which detailed parcel characteristics are available. Of these observations, 204,656 took place in one of the 277 French urban areas from our main sample.

Family expenditure survey

To compute the share of housing in expenditure for French households, we exploit the 2006 and 2011 French Family Expenditure Surveys (*Budget des Familles*). This survey is managed by the French Statistical Institute (INSEE) and is designed to study the living conditions and consumption choices of households like the US consumer expenditure survey. This survey reports income and expenditure by category. It includes a municipality identifier. The 2006 wave includes 10,240 households while the 2011 wave contains 15,597 households.

There are three measures of housing expenditure that can be used. They correspond to two different samples: homeowners and renters. For homeowners, the survey reports a monthly rent-equivalent (or imputed rent) based on the market rental value assessed by homeowners. For private-sector renters, we know the monthly rent, both inclusive and exclusive of fees and taxes. At the sample mean, the difference between the two is modest, representing only 3.3% of

[13] We weigh land parcels transactions by their sample weight to mitigate possible selection problems here. This makes no difference to our results.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

expenditure.[14] We focus our analysis on rents inclusive of fees and taxes. In robustness checks, we verify that our results are not sensitive to this choice. The survey also reports information on household income, age, marital status, children, and seven levels of educational achievement.

We compute the shares of housing in expenditure by taking the ratio of the measure of monthly rents defined above for renters or imputed rents for homeowners to monthly household income. We delete observations with missing values (26.4% for imputed rents, 0.4% for rents inclusive of fees and taxes, and 8.0% for rents exclusive of fees and taxes). We also delete observations with missing values of explanatory variables and instruments, and trim the 1^{st} and 99^{th} percentiles to delete outliers. When pooling the two surveys, our final sample includes 2,464 observations for renters and 5,984 observations for homeowners.

Some descriptive statistics

Table 1 reports descriptive statistics for houses, parcels, housing expenditure, population, and land area. It is useful to keep in mind that a house in urban France has a mean floorspace area of 110 square metres and sells for 2,451 € per square metre (all prices in 2012 €). For land, a parcel has a mean area of 1,055 square metres and sells for 108 € per square metre.[15] French urban households devote on average 31 or 35% of their expenditure to housing, depending on their tenure choice.

Table 2 provides further descriptive statistics for four groups of urban areas, Paris, the next three large French urban areas, other large urban areas, and small urban areas. This table illustrates the cross-city variation in our variables of interest and shows that prices of both floorspace and land appear to increase with urban-area population. Households devote a smaller share of their expenditure to housing in smaller urban areas. The ordering is less clear for the next three size classes in the raw data.

To make the variation in house prices, land prices, and population easier to visualise, the three panels of figure 1 map mean house price per square metre, mean land price per square metre, and population for French urban areas. These maps confirm that there is a lot of variation across

[14]The difference includes local taxes, and management fees and utilities for the common parts for multi-family units. Local taxation in France is generally minimal as public goods are often provided directly by the central government and municipalities are mostly financed through grants. Residential taxation (paid by all residents) represents less than 250 euros per person per year. The revenue from property taxation paid by owners is about 25% larger but arises mainly from commercial properties.

[15]The transactions we observe cover a broad spectrum of prices and areas. This is because we use a systematic and compulsory survey based on administrative records. Unlike land transactions recorded by private real estate firms, ours are not biased towards large parcels.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 1: Descriptive statistics

Variable	Mean	St. Error	1st decile	Median	9th decile
Notary databases – houses					
Price (€ per m^2 of floorspace), sample mean)	2,451	1,187	1,321	2,185	3,820
Price (€ per m^2 of floorspace, urban area mean)	1,817	493	1,306	1,735	2,380
Dwelling area (m^2 of floorspace, sample mean)	110.4	18	92.9	108.2	130.2
Survey of developable land					
Price (€ per m^2 of land, sample mean)	107.7	104.1	25.1	81.5	215.8
Price (€ per m^2 of land, urban area mean)	78.6	53.0	26.7	64.4	150.1
Parcel area (m^2 of land, sample mean)	1,055	914	432	810	1,906
Family expenditure survey					
Housing expenditure share for homeowners	0.314	0.192	0.152	0.263	0.526
Housing expenditure share for renters	0.352	0.287	0.146	0.277	0.624
Population (urban area mean)	166,020	757,144	17,775	47,909	305,453
Land area (km^2, urban area)	597	1,036	99	349	1,324
Number of municipalities per urban area	45.8	104	6	24	90

Notes: All prices in 2012 €. 74,621 municipality price indices corresponding to 1,199,506 dwelling transactions for rows 1-3. 204,656 weighted parcel transactions for rows 4-6. 2,464 (resp. 5,984) households renting in the private sector (resp. owning their home) who correspond to 6.79 (resp. 14.1) million weighted observations for row 6 (resp. 7). 277 urban areas for rows 9-11.

Table 2: Descriptive statistics (means by population classes of urban areas)

City class	Paris	Lyon, Lille, and Marseille	Population >200,000	Population ≤200,000
Notary databases – houses				
Price (€ per m^2 of floorspace)	3,455	2,558	2,310	1,777
Dwelling area (m^2 of floorspace)	107.9	111.4	112.1	110.1
Survey of developable land				
Price (€ per m^2 of land)	255.2	210.6	115.2	69.8
Parcel area (m^2 of land)	850	1,075	984	1,149
Family expenditure survey				
Housing expenditure share for homeowners	0.344	0.344	0.304	0.293
Housing expenditure share for renters	0.369	0.367	0.382	0.285
Population (urban area)	12,197,910	1,512,162	415,950	54,142
Land area (urban area, km^2)	14,598	2,380	1,486	361
Number of urban areas	1	3	40	233
Number of municipalities per urban area	1,565	172	112	26.2

Notes: See table 1. The figures in column 3 refer to all French urban areas with population above 200,000 excluding Paris, Lyon, Lille, and Marseille.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

urban areas with respect to their land area, population, and house and land prices. These maps also suggest strong correlations between these variables. Much of the rest of our work below will document these correlations more precisely and interpret them.

Finally, to illustrate the reality of the data within particular urban areas, the left panels of figure 2 plot municipal house prices per square metre and distance to the centre for four urban areas in 2012. The right panels of the same figure show analogous plots for land prices per square metre of individual parcels. The first urban area at the top of the figure is Paris, the largest French urban area with a population of 12.2 million. The second is Toulouse, the fifth largest French urban area with a population of 1.2 million. The third is Dijon, a mid-sized urban area, which ranks 25th with a population of 330,000. Finally, the last one is Arras, a smaller urban area, which ranks 68th with a population of 130,000.

These graphs demonstrate the importance of using comparable prices across urban areas as prices vary a lot within urban areas and observations are distributed differently. Mean house price in Paris is only 28% above the national mean whereas mean house price in Dijon is 17% below the national mean. By contrast, a house located at the centre of Paris is 187% more expensive than the national mean whereas a house at the centre of Dijon is just 1% below the national mean.[16] The difference between Paris and Dijon is thus about four times as large when looking at prices at the centre relative to mean prices. Hence, comparing mean house prices greatly understates true differences across cities because houses in Paris are on average much further away from the centre than houses in Dijon. For land, the contrast is even starker. Mean land price is 132% higher than the national mean in Paris and 13% higher in Dijon. Land price at the centre is instead a staggering 1080% higher than the national mean in Paris and only 37% higher in Dijon.

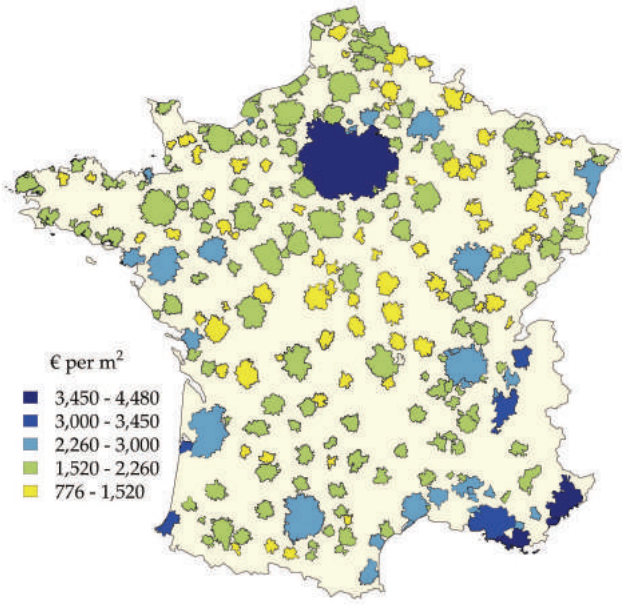
For land parcels, we also note that we observe transactions close to the centre, in close suburbs, and in remote suburbs. This is because French land use regulations encourage in-filling and try to limit expansions of the urban fringe.[17] The plots for land are helpful to alleviate the worry that

[16]With a slight abuse of language and because we use a log scale, we speak of "centre" for the origin which corresponds to a distance of one kilometre. Recall that we measure distances from the centroid of municipalities where a transaction takes place to the centroid of the entire urban area. The two do not coincide in general nor do they even come close in the data so that we do not face zero or very short distances transformed into logs.

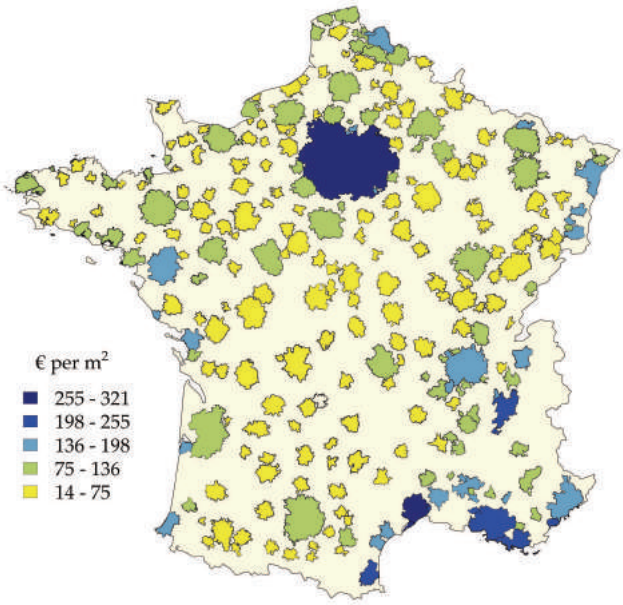
[17]French municipalities need to produce a planning and development plan (*plan local d'urbanisme*) which is subject to national guidelines and requires approval from the central government. Existing guidelines for municipalities or groups of municipalities insist on the densification or re-development of already developed areas to save on the provision of new infrastructure (usually paid for by higher levels of government) relative to expansions of the urban fringe.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

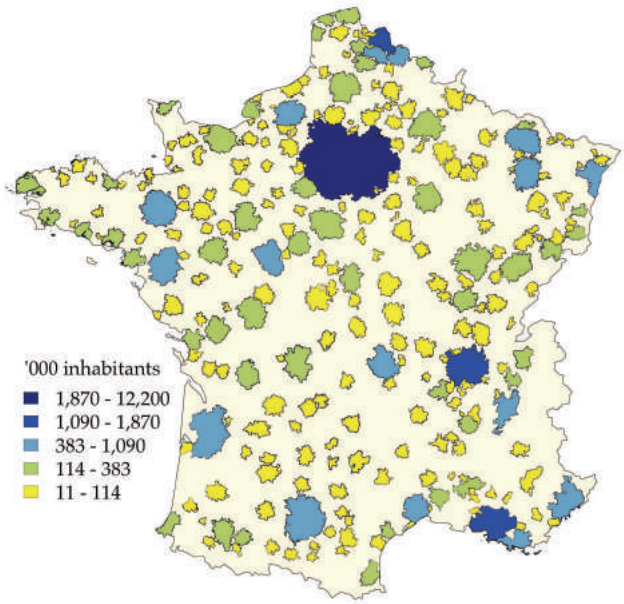
Figure 1: Mean house and land prices per square metre and population in French urban areas



Panel (a): Mean house prices, 2000-2012



Panel (b): Mean land prices, 2006-2012

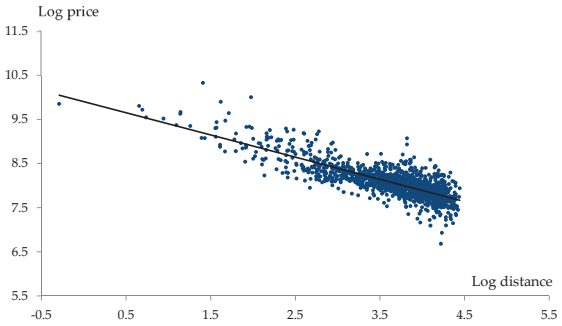


Panel (c): Population, 2000-2012

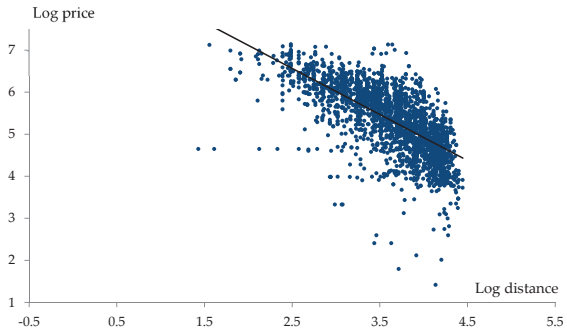
Notes: The classes on each map were created to include about 20% of the French population in each class. All prices in 2012 €.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

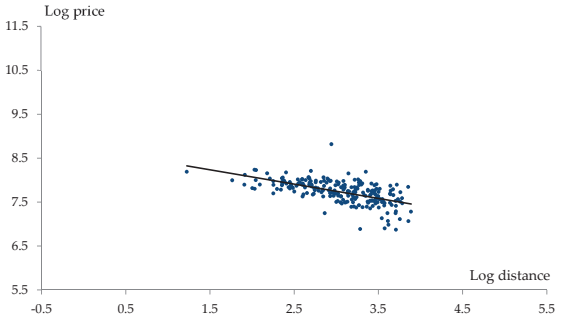
Figure 2: House and land prices per square metre and distance to their centre for four urban areas



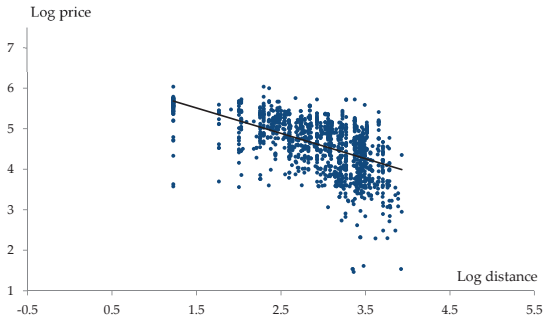
Panel (a.1): House prices in Paris



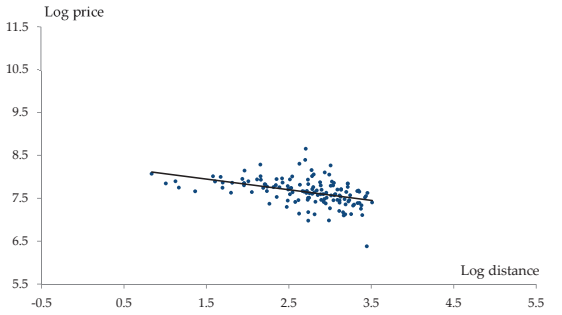
Panel (a.2): Land prices in Paris



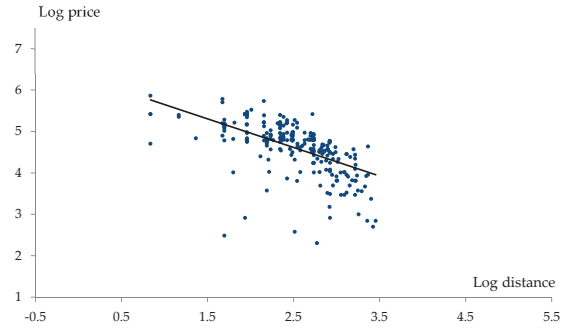
Panel (b.1): House prices in Toulouse



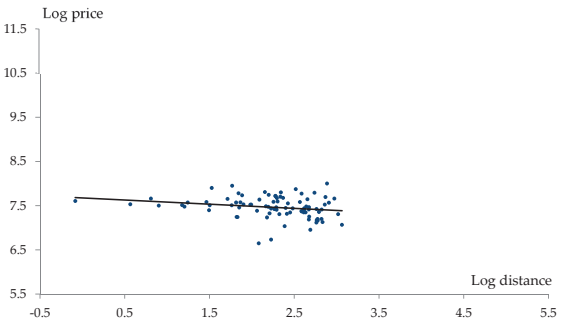
Panel (b.2): Land prices in Toulouse



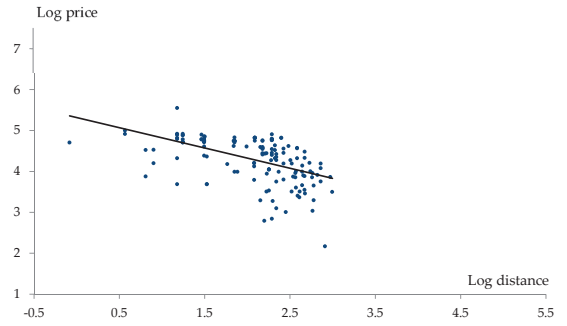
Panel (c.1): House prices in Dijon



Panel (c.2): Land prices in Dijon



Panel (d.1): House prices in Arras



Panel (d.2): Land prices in Arras

Notes: All panels represent 2012 data. The horizontal axis represents the log of the distance between a municipality centroid and the centre of its urban area. The vertical axis represents the log prices estimated from municipal means for house prices and from individual transactions for land prices.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

parcels sold with a building permit are geographically highly selected.

We draw a number of further conclusions from the plots of figure 2. The differences within urban areas in land prices are larger than for house prices. This is in part driven by the fact that house prices are aggregated by municipalities, but not only. The value of housing floorspace per square metre varies much less than the value of land per square metre. Consistent with this, in all four urban areas, the gradient is stronger for land prices. We also note that these gradients appear to differ across urban areas.

4. Comparable house and land prices across French urban areas

To compute the urban costs elasticity as in equation (6), we must, in a first-step, estimate the prices of housing at the centre of each urban area. Hence, from pooled cross-sections we estimate,

$$\log P_{mt} = C^P_{c(m)t} - \delta^P_{c(m)} \ln D_m + X_{mt}\,\alpha^P + \nu^P_{mt}\,, \tag{7}$$

where the dependent variable $\log P_{mt}$ is a (natural log) house price index for municipality m and year t, and our explanatory variable of interest, $C^P_{c(m)t}$ is a fixed effect for the urban area c of municipality m and year t. This fixed effect measures a house price index per square metre of floorspace at the centre of urban area c. In addition, D_m is the distance of municipality m to the centre of the urban area, $\delta^P_{c(m)}$ is a distance gradient for urban area c, and X_{mt} are controls for amenities and socio-economic characteristics in municipality m and year t.[18]

For the price of land parcels, the corresponding equation is,

$$\log R_i = C^R_{c(i)t(i)} - \delta^R_{c(i)} \ln D_{m(i)} + X_{m(i)t(i)}\,\alpha^R + Y_i\,\gamma^R + \nu^R_i\,, \tag{8}$$

where the dependent variable R_i is now the unit land price for parcel i and $C^R_{c(i)t(i)}$ is a fixed effect for the urban area $c(i)$ and year $t(i)$. This fixed effect now measures the unit price of land in year t at the centre of urban area $c(i)$, where parcel i is located and $m(i)$ is its municipality. Equation (8) also includes both parcel, Y, and municipality controls, X. Note that equations (7) and (8) are variants of urban gradient regressions that have often been estimated since Clark (1951).

[18]Formally, our intercept corresponds to $\ln D_m = 0$, that is to a distance to the centroid of the urban area equal to 1 kilometre. Keeping in mind that we measure distances from the centroid of each municipality, there is obviously some measurement error for short distances. We perform a number of robustness checks below to verify that our results are not sensitive to this choice.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Main first-step results

Panel A of table 3 reports summary results for house prices using equation (7). Panel B of the same table reports corresponding results for land prices using equation (8). Column 1 includes only house or parcel characteristics. In panel A, mean house characteristics have little explanatory power because we work with municipal price indices that already condition out individual house characteristics. In panel B, parcel characteristics, especially log parcel area and its square, explain 48% of the variance of land prices per square metre.[19]

Column 2 of table 3 no longer includes house or parcel characteristics and estimates only fixed effects for urban areas. Urban area effects explain about two thirds of the variance of our municipal house price index and more than half of the variance of the unit price of individual parcels. The lower R^2 for land parcels is due to the more disaggregated nature of the land data.

It would be cumbersome to report 277 urban areas fixed effects over 7 years of data. We report instead moments of their distribution after averaging across years. It is interesting to look at the interquartile range, which is three times as wide for land prices as for house prices at the centre. Normalising the mean of all urban area fixed effects to zero, the bottom quartile is at -0.173 for house prices (about 16% below the mean) and at -0.469 for land prices (37% below the mean). The top quartile of house prices is at 0.152 (16% above the mean) and at 0.513 for land prices (67% above the mean).

Column 3 enriches the specification of column 2 with a distance effect specific to each urban area. Column 4 further includes house or parcel characteristics. While distance gradients differ across urban areas, they are in most cases negative. Like for the four cities of figure 2, land price gradients are in general much steeper than house price gradients. In column 4, the median land price gradient is four times as large as the median house price gradient. This feature is closely related to the greater dispersion of prices at centre for land parcels relative to houses we just highlighted.

Amenities make some municipalities more desirable and their spatial distribution differs across

[19]The other characteristics we include are whether a parcel is serviced and three indicator variables that relate to the type of intermediary through whom the parcel was purchased. Although we do not report the details of the coefficients for parcel characteristics in table 3, some interesting features are to be noted. Most importantly, smaller parcels fetch a higher price per square metre. Then, a serviced parcel is more than 50% more expensive than a parcel with no access to basic utilities. Parcels sold by real estate agencies, builders, or other intermediaries are also more expensive since real estate professionals are likely to specialise in the sale of more expensive parcels.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 3: Summary statistics from the first step estimation regressions, 277 urban areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Log house prices per square metre									
Urban area effect									
1st quartile		-0.173	-0.207	-0.209	-0.207	-0.208	-0.204	-0.200	-0.198
3rd quartile		0.152	0.156	0.153	0.154	0.181	0.156	0.156	0.172
Log distance effect									
1st quartile			-0.0884	-0.0869	-0.0812	-0.0805	-0.0705	-0.0726	-0.0417
Median			-0.0374	-0.0374	-0.0378	-0.0397	-0.0251	-0.0268	-0.0088
3rd quartile			-0.0006	0.0016	0.0089	-0.0054	0.0163	0.0145	0.0242
Observations	74,621	74,621	74,621	74,621	74,621	74,621	74,621	74,621	74,621
R^2	0.01	0.66	0.79	0.80	0.81	0.85	0.80	0.81	0.86
Panel B. Log land prices per square metre									
Urban area effect									
1st quartile		-0.467	-0.565	-0.505	-0.502	-0.452	-0.484	-0.487	-0.443
3rd quartile		0.513	0.482	0.369	0.357	0.388	0.387	0.381	0.410
Log distance effect									
1st quartile			-0.411	-0.239	-0.244	-0.218	-0.199	-0.233	-0.143
Median			-0.263	-0.148	-0.145	-0.145	-0.116	-0.140	-0.087
3rd quartile			-0.153	-0.066	-0.063	-0.085	-0.047	-0.068	-0.032
Observations	204,656	204,656	204,656	204,656	204,656	204,656	204,656	204,656	204,656
R^2	0.48	0.52	0.63	0.82	0.82	0.83	0.82	0.82	0.83
Controls									
House/Parcel charac.	Y			Y	Y	Y	Y	Y	Y
Geography and geology					Y				Y
Income, education						Y			Y
Land use							Y		Y
Consumption amenities								Y	Y

Notes: OLS regressions in all columns. For house prices, we weigh municipalities by the number of transactions. All reported R^2 are within-year. Reported urban area effects are averaged over time weighting each year by its number of observations.

For house price indices, house characteristics include log mean area and its square for each municipality. For land prices, parcels characteristics include log area and its square and indicator variables for whether the parcel is serviced and three types of intermediaries through whom the parcel may have been bought. Geography and geology characteristics for municipalities include maximum and minimum altitude, dummies for presence of each of the five main rivers (Seine, Loire, Garonne, Rhône, Rhin), dummies for contiguity to each neighbouring country (Spain, Italy, Switzerland, Germany, Belgium/Luxemburg), dummies for contiguity to each major body of water (British Channel, Atlantic Ocean, and Mediterranean Sea), four geology variables (erodability, hydrogeological class, dominant parent material for two main classes). Income and education variables of a municipality include the logarithm of mean income and of income standard deviation, and the share of population with a university degree. Land use variables of a municipality include the share of land that is build-up and the average height of buildings. Consumption amenities for each municipality are all normalised per unit of population and include the number of restaurants, supermarkets, primary, secondary, and high schools, medical establishments, doctors, cardiologists, medical laboratory, and cinemas. All municipal controls are centred relative to their urban area mean.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

urban areas. The spatial distribution and relative population sizes of socio-economic groups also differs across urban areas. In models of urban structure, amenities and residential heterogeneity will affect both gradients and prices at the centre (Duranton and Puga, 2015). We may also worry about differences in land use regulations.[20]

To address these concerns, columns 5 to 8 further introduce different sets of control variables that pertain to the geography and geology of municipalities (20 variables in total), to their socioeconomic characteristics (including log mean income, its standard deviation, and the share of university-educated residents), to their land use (including the share of land that is built and average height of building), and to their consumption amenities (9 variables in total). These explanatory variables are all centred relative to their urban area mean to condition out municipality effects within each urban area.

Column 9 includes all house/parcel and municipality controls at the same time. It is our preferred first-step estimation because it controls for many sources of heterogeneity within urban areas. Relative to column 2 where only urban area fixed effects are included, the R^2 is much higher, well above 80% for both house and land prices per square metre.

Importantly, the values of the top and bottom quartiles of urban area fixed effects do not fluctuate much across our specifications for neither house nor land prices. To provide more direct evidence of the stability of our first-step results, we compute the correlation between the urban area fixed effects estimated in column 2 with no further controls and those estimated in column 9 with a full set of controls (house or parcel characteristics and 34 municipal controls). The correlation is 0.95 for house prices and 0.90 for parcel prices. The corresponding Spearman rank correlations are similarly high. We also have high correlations between the urban area fixed effects for house prices and those for land prices. For our preferred specification this correlation is equal to 0.88 . This high correlation is reassuring because our model (like most models of land development) establishes a tight link between land and house prices.

Further robustness checks

A number of further concerns about our first-step estimation must be discussed. The first is about our choice of functional form for the distance gradients. Ultimately, the appropriate functional

[20]This concern may not be as important as it seems because, in simple models of spatial structure, differences in house prices within urban areas are determined by differences in accessibility, not by differences in relative local housing supply.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

form should depend on accessibility and transport costs, which we know little about. As illustrated by the four cities represented in figure 2, measuring distance to the centre in log seems appropriate in practice.[21] In further robustness checks, we estimate equations (7) and (8) with alternative functional forms, including measuring distance in levels, mixing logs and levels, or estimating a separate gradient for each urban area and year of data.[22] To explore the issue of sorting within urban areas further, we also experiment with specifications for which we additionally include interaction terms between distance to the centre and municipal income for all urban areas.

Then, the geography we impose to urban areas with a unique centre is perhaps questionable. In response, we estimate equations (7) and (8) allowing for two different centres. We also experiment with alternative definitions for the centre of urban areas. Instead of defining the centre of an urban area as its population centroid across all municipalities, we can take as centre, the geographic centroid of the core municipality by population. Because of this ambiguity about the definition of centres, measurement error is possibly worse for short distances. As a check, we also duplicate our preferred estimation after eliminating the 25% of observations closest to the centre in each urban area. This last check is also helpful to address the issue that in some urban areas, central municipalities may be special in terms of unobserved amenities, unobserved characteristics of their residents, or unobserved land use regulations. Additionally, we duplicate our preferred estimation after eliminating the 25% of observations with the lowest prices in each urban area.[23]

Finally, note that for consistency with the land parcels results our preferred estimation considers a price index for housing that only relies on transactions of single-family houses. We duplicate our first-step estimation for housing prices using an index that includes both houses and apartments. The results are reported in separate appendix A.[24]

[21] Beyond our four illustrative cities, the relationship between house prices and population is generally well described by a log specification. The fit is less good for land prices but after experimenting with various functional forms, we concluded that no simple functional form is obviously better.

[22] The urban area fixed effects estimated with our preferred estimation in column 9 of table 3 and panel A have a correlation of 0.98 with those estimated from a similar specification which uses distance in levels instead of logs. The correlation between our preferred fixed effects and those estimated using year-specific gradients is 0.99. We do not report first-step results systematically for these robustness checks because endless duplications of table 3 are of limited interest. Below, we report second-step results using the supplementary first-step estimations mentioned in this section.

[23] The urban area fixed effects estimated with our preferred estimation of column 9 in panel A of table 3 are generally highly correlated with those estimated from the alternatives mentioned in this paragraph and the previous one. The two relative exceptions are when we allow for two centres (correlation of 0.63 with our preferred fixed effects for house prices) and when we eliminate 25% municipalities closest to the centre (correlation 0.76). We also verify below that our second-step results are robust to these alternative first-step estimates.

[24] The Spearman rank correlation with the house price fixed effects from our preferred estimation is again high at 0.91.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

5. Estimating the elasticity of house and land prices with respect to population

In the second step, we use the prices of houses and land at the centre estimated in the first step as dependent variables to estimate the elasticity of these prices with respect to urban-area population. For housing prices, from the pooled cross-sections we estimate,

$$\hat{C_{ct}^P} = Z_{ct}\,\beta^P + \phi_t^P + \xi_{ct}^P\,, \tag{9}$$

where the dependent variable, the (log) price of houses at the centre of urban area c at time t, is obtained from the estimation of equation (7). The explanatory variables are a vector of urban area characteristics Z_{ct} and year fixed effects ϕ_t^P. For land prices, we estimate,

$$\hat{C_{ct}^R} = Z_{ct}\,\beta^R + \phi_t^R + \xi_{ct}^R\,, \tag{10}$$

which mirrors equation (9) but the dependent variable is now obtained from the estimation of equation (8).

In both equations (9) and (10), the explanatory variable of interest is the log of urban area population included in Z_{ct}. Our main concern with equations (9) and (10) is the endogeneity of population. More specifically, we worry about possible missing variables that are correlated with both population and land or house prices at the centre. We also worry about potential reverse causation leading more expensive cities to end up smaller. Before instrumenting or relying on the longitudinal dimension of the data, our first strategy is to consider an exhaustive set of control variables to alleviate doubts about missing variables.

Pooled cross-section results

Table 4 reports results for a number of OLS regressions. Panel A uses the estimated (log) unit price of houses at the centre of urban areas as dependent variable while panel B uses the estimated (log) unit price of land. The specifications are otherwise identical across both panels.

Columns 1 to 3 use house and land prices estimated in column 2 of table 3 in the first step as dependent variable. Aside from year effects, column 1 only includes log urban area population and log of land area as explanatory variables.[25] The estimated population elasticity is 0.217 for house

[25]We generally include the log of land area in our regressions. Besides being a major determinant of the availability of land and housing, we also think that the relevant question about urban costs regards their increase following an increase in population, keeping land area constant. French land use regulations make the expansion of urban boundaries extremely difficult. Below, we nonetheless contrast the results we obtain for urban costs with constant land areas to estimates that allow urban boundaries to adjust.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 4: The determinants of unit house and land prices at the centre, OLS regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First-step	Only fixed effects			Basic controls			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.	N	Y	Ext.
Panel A. Houses									
Log population	0.217^a	0.176^a	0.224^a	0.259^a	0.215^a	0.305^a	0.252^a	0.208^a	0.304^a
	(0.0210)	(0.0142)	(0.0283)	(0.0276)	(0.0187)	(0.0378)	(0.0262)	(0.0179)	(0.0368)
Log land area	-0.151^a	-0.153^a	-0.224^a	-0.114^a	-0.122^a	-0.242^a	-0.143^a	-0.152^a	-0.276^a
	(0.0219)	(0.0136)	(0.0293)	(0.0250)	(0.0189)	(0.0379)	(0.0241)	(0.0174)	(0.0382)
R^2	0.35	0.65	0.72	0.44	0.67	0.73	0.40	0.66	0.73
Observations	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937
Panel B. Land parcels									
Log population	0.774^a	0.707^a	0.871^a	0.678^a	0.604^a	0.702^a	0.662^a	0.597^a	0.738^a
	(0.0464)	(0.0435)	(0.122)	(0.0464)	(0.0362)	(0.0865)	(0.0432)	(0.0360)	(0.0875)
Log land area	-0.676^a	-0.676^a	-0.881^a	-0.344^a	-0.363^a	-0.505^a	-0.437^a	-0.453^a	-0.630^a
	(0.0527)	(0.0448)	(0.133)	(0.0464)	(0.0379)	(0.0905)	(0.0445)	(0.0372)	(0.0934)
R^2	0.54	0.64	0.69	0.63	0.75	0.79	0.61	0.73	0.77
Observations	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933

Notes: The dependent variable is an urban area-year fixed effect estimated in the first step. Columns 1 to 3 use the output of column 2 of table 3. Columns 4 to 6 use the output of column 4 of table 3. Columns 7 to 9 use the output of column 9 of table 3. All regressions include year effects. All reported R^2 are within-time. The superscripts *a*, *b*, and *c* indicate significance at 1%, 5%, and 10% respectively. Standard errors clustered at the urban area level are between brackets. For second-step controls, N, Y, and Ext. stand for no further explanatory variables beyond population, land area, and year effects, a set of explanatory variables, and a full set, respectively. Second-step controls include population growth of the urban area (as log of 1 + annualised population growth over the period), income and education variables for the urban area (log mean income, log standard deviation, and share of university degrees). Extended controls additionally include the urban-area means of the same 20 geography and geology controls as in table 3 and the same two land use variables (share of built-up land and average height of buildings) used in the same table.

prices and 0.774 for land prices. Column 2 also includes population growth, log mean income, log standard deviation of income, and the share of university educated workers. Including these controls marginally lowers the coefficient on log population, to 0.176 for house prices and to 0.707 for land prices. Column 3 enriches the regression further with 20 geography and geology variables and two important land use variables, the share of built-up area and the log of the average height of buildings. Adding these extra controls leads to a slight increase of the coefficient on population in both panels.

Columns 4 to 6 repeat the same pattern of estimation as columns 1 to 3 but use as dependent variable the fixed effects estimated in column 4 of table 3, a more complete first-step regression,

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

which includes house or parcel characteristics and a distance effect specific to each urban area in addition to urban area fixed effects and year fixed effects. Columns 7 to 9 repeat again the same pattern of estimation but use this time the output of the most complete first-step regression from column 9 of table 3. In these three columns, the urban area fixed effects are estimated at the first step conditional on house or parcel characteristics and 34 municipality characteristics, including their socioeconomic composition, geography, geology, land use, and amenities.

Our preferred OLS estimates are in column 8. They suggest an elasticity of house prices with respect to population of 0.208 and an elasticity of land prices with respect to population of 0.597. We are interested in estimating the elasticity of house and land prices with respect to population, all else equal. The estimates of column 7 do not condition out the socio-economic characteristics of cities. They thus fail to account for the possibility that, among others, larger cities are also more skilled. We also prefer the estimates of column 8 to those of column 9, which additionally control for share of land that is built-up and the average height of buildings. While we think that these two land-use controls are useful proxies for land-use regulations, it may be too extreme to think of an increase in population in a city that would keep both land use and land area constant as the most relevant thought experiment.

Although we do not report the coefficients on all the control variables in the table, some results are worth a brief mention. Most notably, we introduce population growth in the regression to separate rents today and expectations of future rent increases which are driven by population growth. Both are included in house prices. A one percentage point of annual population growth is typically associated with about 10% higher prices for houses. Despite this large effect, including population growth does not affect the coefficient on population because population and population growth are only weakly correlated, in keeping with Gibrat's law. Many of our geographic controls including the distance to the main rivers and various borders have a significant effect. They capture broad regional trends in land and housing prices in France. Finally, the estimated coefficient on log mean income is always significant and equal to 1.57 in column 8.

In column 8, the elasticity of land prices is nearly three times as high as the elasticity of house prices. This is consistent with our findings above that the interquartile range for land prices at the centre in our preferred first-step estimation is also about two and half times as large as the interquartile range for house prices at the centre.

Recall that, when we extend our model to allow for a housing construction sector, the popula-

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

tion elasticity of the price of housing is the product of the population elasticity of the price of land and the share of land in construction. In the data, the average share of land in the total cost of a new house is 36% and roughly constant across urban areas and parcel size (Combes *et al.*, 2016). Using our model, the estimates of column 8 imply an implicit share of land of 35% for old houses. With the caveat that we compare new constructions with old houses, this is extremely close.

We document in separate appendix B that the distance gradients for urban areas with greater population are steeper. This appendix duplicates table 4 but uses the distance gradient estimated in the first stage instead of the urban area fixed effect as dependent variable. While prices at the fringe do not differ much across urban areas, the higher prices at the centre that we observe in urban areas with greater population are associated with both a greater distance to the urban fringe and a steeper distance gradient.

Robustness checks

Before implementing alternative estimation strategies, we further explore the robustness of our second-step OLS results.

First, household heterogeneity across urban areas may affect our results.[26] Empirical evidence suggests that more skilled households sort into larger cities in France (Combes *et al.*, 2008). We expect the price premium of central locations to be determined by both city population and the socioeconomic characteristics of this population. While in table 4 we control for a wide range of socioeconomic characteristics, more complicated interactions may be at work. To assess this possibility, we duplicate the specifications of table 4 and include interactions between city population and income or education in separate appendix C. This leads to modestly smaller population elasticities. For house prices, adding an interaction term between population and income to our preferred estimation marginally lowers the population elasticity from 0.208 to 0.199. For parcel prices, the elasticity is 0.572 when introducing a similar interaction instead of 0.597 without.

Second, we also duplicate the estimations of panel A of table 4 for housing prices that pertain to all dwellings instead of only houses. The results are reported in separate appendix D. The estimated elasticities of the price of central dwellings with respect to city population are modestly

[26] In the first step of our estimation, we condition out various socio-economic characteristics of municipalities within urban areas given our worry that the spatial distribution of heterogeneous households within the urban area may affect the estimation of gradients and thus of prices at the centre. However, municipal characteristics are measured relative to the city mean and only condition out household heterogeneity within cities, not differences between cities. We need to address heterogeneity both within and between cities.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

lower than in table 4. This is likely caused by the lower land intensity of apartments relative to houses.

Third, we also consider a number of further variants for our preferred specification of column 8 in table 4 in separate appendix E. In particular, we experiment with dependent variables estimated in the first step with alternative functional forms for distance to the centre, alternative definitions of a centre, the inclusion of a second centre, separate gradients for each urban area and year, and interactions between municipal income and distance to the centre. We also use alternative samples which exclude the 25% cheapest municipalities or the 25% closest municipalities to the centre in the first step to deal with potential selection problems for transactions. We also consider alternative weighting schemes in the estimation and alternative second-step samples that eliminate observations with negative growth. Because we rely in our second step on a dependent variable that is estimated (with error) in a first step, we also experiment with FGLS and WLS techniques to explicitly account for this measurement error (see separate appendix F for further explanations). Finally, instead of using a two-step procedure, we can also estimate everything in one step. While we estimate sometimes smaller or larger population elasticities, the magnitudes are in general close and supportive of our baseline findings.

Instrumental-variable estimates

To repeat, when estimating equations (9) and (10) we are concerned with the endogeneity of population. We expect the main source of endogeneity to arise from the existence of missing variables that are correlated with population and affect land or house prices through some other channel. Another possible source of endogeneity is reverse causation: population may become larger in cheaper cities. Both sources of endogeneity can be addressed through instrumental variables. Because land area is highly correlated with population, we need to instrument both variables.

We use two sets of instruments. Our first set of instruments is suggested by our model where exogenous amenities in a city attract population without otherwise affecting the demand or supply of housing in this city. More specifically, we use a measure of temperatures in January, a count of hotel rooms, and the share of budget hotel rooms. Our measure of climate is motivated by the literature on urban growth. This literature shows that January temperatures is a strong predictor of urban growth and thus of urban population in the long run (Duranton and Puga, 2014). A count

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

of hotel rooms is in the spirit of Carlino and Saiz (2008) who argue that tourism visits provide a summary proxy for all amenities in a city. We prefer to focus on budget hotels since higher-end hotels in France arguably cater predominantly to the needs of business travellers.

Our second set of instruments consists of long lags of urban population and density constructed from population and area data from 1831, 1851, and 1881. This instrumental strategy follows a long tradition in the urban literature where city population is instrumented with past values of the same variable to estimate agglomeration effects (Ciccone and Hall, 1996, Combes *et al.*, 2008, Combes and Gobillon, 2015). The use of long lags for population, area, or density is motivated by the idea that the factors that made an urban area a particularly cheap (or expensive) place to live nearly two centuries ago differ from the factors that drive the demand or supply of housing today. We expect these predictors of city population to be immune from reverse causation and from the effects of more recent shocks affecting both population and prices.

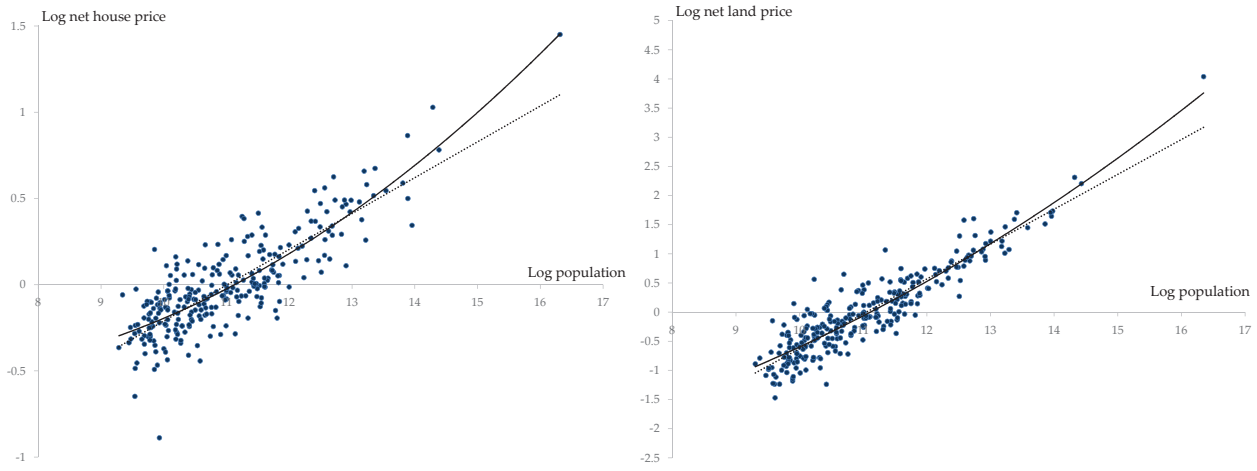
While we can make the case that these instruments are strong enough predictors of contemporaneous city population, they might still be correlated with land or housing prices through some other demand or supply channels. For instance, amenities may induce residents to consume more (or less) housing. To address this worry, we can control extensively for the characteristics of municipalities and urban areas to preclude these sources of correlation with the error term.[27] We also note that long population lags and amenities rely on different sources of variation in the data to predict contemporaneous populations. For instance, the correlation between January temperatures and the other instruments is always below 0.10. Obtaining statistically similar coefficients from these different instruments is reassuring.

In Appendix 3, we provide further details about our IV strategy and report detailed results for both house prices in table 8 and for land prices in table 9. For house prices, most of our estimates of the population elasticity are between 0.20 and 0.27 with a few exceptions above or below. For land prices, most of the estimates of the population elasticity are between 0.60 and 0.80. In both cases, this is moderately larger than our preferred OLS estimates of 0.208 and 0.597 but comparable to other OLS estimates reported in table 4 and in the external appendices. We conclude that our IV results are supportive of our baseline OLS results.

[27] A possible issue with introducing more controls is that these controls may themselves be endogenous and correlated with city population. In Appendix 3, we report results for different combinations of instruments and different specifications that include fewer or more controls.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Figure 3: Log house and land prices (component plus residual) and log city population



Panel (a): House prices

Panel (b): Land prices

Notes: The horizontal axis in both panels represents log urban area population. The vertical axis represents the residual of the regression of column 8 of table 4 plus log urban area population multiplied by its estimated coefficient and then averaged over all years. The dependent variable is house prices at the centre of urban areas in panel (a) and the corresponding land prices in panel (b). The plain continuous curve is a quadratic trend line. The dotted line is a linear trend. Mean prices across all urban areas are normalised to zero in both panels.

Non-constant population elasticities

Given that we are interested in how the elasticity of urban costs varies with city population, we now examine whether the elasticity of house or land prices with respect to city population is constant for all cities regardless of their population size. In panel A of figure 3, we provide a 'component plus residual' plot for our preferred OLS estimation. We represent log urban area population on the horizontal axis and the price of housing after conditioning out explanatory variables other than population on the vertical axis. In panel B of figure 3, we provide a similar plot for land prices. Each plot also contains two trend lines, linear and quadratic.

In panel A, for log population below 14 (which corresponds to about 1.2 million inhabitants) the two trend lines are extremely close but they diverge for the largest cities, in particular Paris which is unusually expensive for its population relative to a log linear trend. A similar but milder convexity is also apparent for land prices.

To explore this issue further, separate appendix G reports results for a series of regressions where we introduce terms of higher order for log population. Adding a quadratic term for log population to our preferred specification of column 8 of table 4 implies an elasticity of house prices

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

with respect to population of 0.205 for an urban area with 100,000 inhabitants, an elasticity of 0.288 for an urban area with a million inhabitants, and 0.378 for an urban area with the same population as Paris. The other specifications yield roughly similar estimates. This said, we must remain cautious about this non-linearity because it is driven only by the three or four largest cities.

To summarise our findings so far, our preferred estimate for the elasticity of house prices at the centre of urban areas with respect to population is 0.208. Alternative OLS and IV estimates for this elasticity reported in tables 4 and 8 are mostly in the 0.15-0.30 range. We also find that this elasticity possibly increases with population for the largest urban areas. The estimates for land prices are equally stable and consistent with those for house prices.

Estimates for alternative time horizons

All our specifications so far include land area as a control. Given the current institutional framework in France, which strongly encourages in-filling but discourages the expansion of the urban fringe, we view the population elasticities of land and house prices conditional on urban area as the relevant benchmarks to think about urban costs.

In the very long-run, the current institutional framework may change and allow urban areas to expand physically with population. In separate appendix H, we duplicate table 4 and estimate the same population elasticity as previously without including land area. We find much smaller coefficients for population equal to or slightly larger than the sum of the population coefficients and the (negative) land area coefficients estimated in table 4. This is consistent with an estimated coefficient of about 0.7 for log population when we regress log land area on log population. For our preferred specification but without including land area, we estimate a population elasticity of house prices equal to 0.109 instead of 0.208 previously.

At the other extreme, it is also interesting to estimate urban costs over a short time horizon, perhaps before the housing stock fully adjusts to population changes.[28] For that purpose, we can estimate equation (10) in the within dimension using observations every odd year between 2000 and 2012. We can also estimate this equation in difference using 2012 and 2000.[29] These two

[28] A change in demand may take time to be perceived by house builders. Obtaining a building permit takes time and building a house also takes time. Beyond this, new housing often requires a change in the zoning designation (conversion from agricultural to residential or from commercial/manufacturing to residential). These zoning changes are infrequent in France – every 20 years or so, see the example of Lyon discussed at `https://www.grandlyon.com/fileadmin/user_upload/media/pdf/espace-presse/dp/2017/20170911_dp_pluh.pdf` (consulted on 22 December 2017).

[29] We do not use land price data here because they are only available for a short time period (2006-2012) instead of 2000-2012 for house price data.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 5: The determinants of unit house prices at the centre, Within and 2000-2012 difference regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Within area				2000-2012 difference			
First-step	Only fixed effects		Full set of controls		Only fixed effects		Full set of controls	
Controls	N	Y	N	Y	N	Y	N	Y
Log population	0.400^a	0.324^b	0.409^a	0.342^b	0.681^a	0.742^a	0.703^a	0.780^a
	(0.0871)	(0.144)	(0.0877)	(0.0978)	(0.140)	(0.183)	(0.114)	(0.174)
Observations	1,937	1,937	1,937	1,937	275	275	275	275
Within R^2	0.02	0.03	0.02	0.03	0.11	0.12	0.12	0.14

Notes: The dependent variable is an urban area-time fixed effect estimated in the first step. Columns 1, 2 and 5 and 6 use the output of column 2 of table 3. Columns 3, 4 and 7 and 8 use the output of column 9 of table 3. Columns 1, 3, 5, and 7 only include population. Columns 2, 4, 6, and 8 also include population growth, log mean municipal income, its standard deviation, and the share of university graduates which all vary over time. Columns 1 to 4 are within area estimates. The R^2 are within urban area. Columns 5 to 8 are 2000-2012 difference estimates. White-robust standard errors between brackets. The superscripts *a*, *b*, and *c* indicate significance at 1%, 5%, and 10% respectively.

estimation approaches use higher-frequency variation and difference out permanent unobserved urban area effects.

Table 5 reports results for a series of estimations exploiting the variation in house prices and in urban area population over time. Columns 1 to 4 of table 5 report within estimates of the population elasticity of house prices. These estimates vary between 0.324 and 0.409 and are larger than our preferred estimate of 0.208 above. We interpret these larger elasticities in light of the slow adjustment of housing supply.

Columns 5 to 8 report estimates of the same population elasticity of housing prices using 2000-2012 differences. The estimates are even larger, between 0.681 and 0.780. We suspect that the difference between the within and 2000-2012 difference estimates is due to measurement error for population over two-year intervals in the within estimation.

Just like population may be endogenous in our cross-section estimations above, changes in population may also be endogenous here, perhaps even more so. To address this, we can instrument population changes in the spirit of the approach first developed by Bartik (1991). This approach is described in greater details in Appendix 2 and separate appendix I. In appendix I, we also report some instrumented results. The results for these IV estimations in difference are hard to interpret

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

since we estimate both large coefficients and large standard errors.

6. The share of housing in expenditure

Estimating the share of housing in expenditure

After the population elasticity of the price of housing, the share of housing in expenditure is the second key input into the computation of the urban costs elasticity. To be consistent with our estimations above, we want to estimate the share of housing at a central location and assess how it depends on urban area population.[30] Using data from the French Family Expenditure Survey, we estimate variants of the following regression,

$$s_i^h = \bar{s}^h + X_{m(i)\,t(i)}\,\alpha^S + Y_i\gamma^S + Z_{c(i)\,t(i)}\,\beta^S + \phi^S_{t(i)} + \mu_i\,, \tag{11}$$

where the dependent variable is the share of housing in expenditure for household i, $\bar{s}^h$ is a constant, Y_i is a set of socio-demographic characteristics and housing tenure indicators for household i, $X_{m(i)\,t(i)}$ is a set of explanatory variables for municipality $m(i)$ where household i lives in year $t(i)$, $Z_{c(i)\,t(i)}$ is a corresponding set of explanatory variables for urban area $c(i)$, and $\phi^S_{t(i)}$ is a year fixed effect (as we pool two waves of data for 2006 and 2011). The main explanatory variable of interest is again log urban area population. Household control variables include demographic characteristics, and income. As previously, municipal variables include distance to the city centre and various socioeconomic characteristics.

Although we estimate the semi-elasticity of the housing share with respect to population in a single step, our approach mirrors our estimation of the population elasticity above.[31] We thus face essentially the same identification issues regarding potential missing variables and various forms of spatial heterogeneity within and between urban areas. We handle those concerns in the same way.

There is an additional concern because we include household characteristics in equation (11), as we expect them to play an important role in the demand for housing. In particular, we expect

[30] Unless the demand for housing is unit price elastic, the share of housing in expenditure will in general vary with distance to the centre within urban areas. Unless the demand for housing is also unit income elastic, it will vary across income groups. The literature often assumes that housing enters utility in a Cobb-Douglas manner so that the share of housing in expenditure can be taken to be the same everywhere for everyone. While this may be a reasonable first-order approximation for many purposes, this is problematic here because modest deviations from this assumption can have a sizeable effect on our estimates of urban costs given the large variation in housing prices across French urban areas.

[31] We perform a single-step estimation because there is less to be learnt from a two-step estimation and because we are more limited in terms of statistical power. In this respect, note that we estimate a single coefficient common to all urban areas for the distance to the centre.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

housing decisions to be driven by permanent income, while we only observe current income. Because income and population are possibly related (be it only because of agglomeration effects), this may affect the estimates of our coefficient of interest. Like previous literature (e.g., Glaeser, Kahn, and Rappaport, 2008), we instrument household income by education.

Baseline results

Table 6 reports results for the pooled sample of homeowners and renters in the French Family Expenditure Surveys for 2006 and 2011. Column 1 regresses the share of housing in expenditure on household demographic characteristics, (log) household income, and (log) urban area population. We estimate a coefficient on city population of 0.028. Column 2 also includes distance to the city centre. Columns 3 and 4 further enrich the regression by including log land area, population growth, and a number of further controls to condition out the socioeconomic characteristics of urban areas. The coefficient on population increases slightly to 0.039.[32] Column 5 duplicates column 4 but instruments for income using five indicator variables for educational achievement. This lowers the magnitude of the coefficient on income but does not appear to affect the rest of the regression. In particular, the coefficient on population in column 5 differs only marginally from its counterpart in column 4.

Column 6 of table 6 instruments contemporaneous urban area population by urban area population in 1831. The point estimate on population modestly rises from 0.039 with OLS in column 4 to 0.047. These two coefficients are only about one standard deviation apart. Column 7 instruments population with urban area amenities. More specifically, we use, as previously, the overall number of hotel rooms and the number of low-end hotel rooms per capita.[33] This leads to a slightly higher coefficient on city population of 0.067. While this larger coefficient does not really affect our conclusions as we show below, we should keep in mind that the instruments are weaker in that case. Finally, column 8 uses both amenities and past population as instruments to estimate a coefficient of 0.048 for population.

[32]Most of the change in the coefficient on city population between columns 2 and 3 of table 6 is due to the inclusion of land area into the regression. Recall that land area is strongly positively correlated with city population.

[33]When using amenities as instruments at the urban area level, we include a measure of the same variables at the municipal level as explanatory variables in the regression. All our municipal explanatory variables are centred relative their urban area means. Moreover, we keep in mind that the regressions of table 6 exploit data from only 197 urban areas instead of 277 previously when estimating the elasticity of house and land prices with respect to population.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 6: The share of housing in expenditure for homeowners and renters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log population	0.028[a]	0.031[a]	0.037[a]	0.039[a]	0.036[a]	0.047[a]	0.067[a]	0.048[a]
	(0.001)	(0.001)	(0.005)	(0.007)	(0.007)	(0.011)	(0.010)	(0.008)
Log land area			-0.011	-0.017[b]	-0.020[a]	-0.025[b]	-0.043[a]	-0.025[a]
			(0.007)	(0.007)	(0.006)	(0.010)	(0.010)	(0.008)
Population growth			2.767[a]	2.694[a]	2.503[a]	2.521[a]	2.121[a]	2.502[a]
			(0.562)	(0.640)	(0.679)	(0.665)	(0.692)	(0.649)
Log distance to city centre		-0.008[c]	-0.008	-0.006[b]	-0.003	-0.008[a]	-0.013[a]	-0.008[a]
		(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log income	-0.282[a]	-0.284[a]	-0.283[a]	-0.286[a]	-0.170[a]	-0.286[a]	-0.286[a]	-0.286[a]
	(0.013)	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)
First-stage statistic					158.0	112.5	6.6	17.2
Overidentification p-value					0.09		0.03	0.00
Instruments								
Educational level (degree)					X			
Urban population in 1831						X		X
Consumption amenities							X	X
Local controls	No	No	No	Yes	Yes	Yes	Yes	Yes
R^2	0.56	0.56	0.56	0.57				

Note All R^2 are within time. 8,446 observations in each regression corresponding to 197 urban areas. Standard errors are clustered at the urban area level. [a]: significant at 1% level; [b]: significant at 5% level; [c]: significant at 10% level. All variables are centred and the estimated constant, which corresponds to the expenditure share in a city of average size (2.99 million inhabitants, 3.17 million with weights), takes the value 0.325 in all specifications (weighted and unweighted). Regressions are weighted with sampling weights and include: age and indicator variables for year 2011 (ref. 2006), homeowner (ref. renter), living in couple within the dwelling (ref. single), one child, two children, three children and more (ref. no child). Local controls include the same geography variables for urban areas as in table 4 and the same geology, land use, and amenity variables at the municipality level as in table 3. OLS for columns (1) to (4). IV estimated with limited information maximum likelihood (LIML) in columns (5) (income instrumented), (6) and (7) (population instrumented) and (8) (income and population instrumented). The first-stage statistics is the Kleibergen-Paap rk Wald F. The critical value for 10% maximal LIML size of Stock and Yogo (2005) weak identification test is 4.45 for column (5), 16.38 for column (6), 3.50 for column (7), and 3.42 for column (8). The education instruments are five indicator variables corresponding to PhD and elite institution degree, master, lower university degree, high school and technical degree, lower technical degree, and primary school (reference). Amenities instruments are: January temperature, the log number of hotel rooms and the share of one-star hotel rooms.

These small variations in the coefficient for urban area population make no economically meaningful difference to our final results. With a mean share of housing in expenditure of 0.325 for a mean urban area of 3.17 million inhabitants, our preferred coefficient of 0.048 from column 8 implies a share of housing in expenditure of 0.390 for a city with the same population as Paris and a share of 0.159 for an urban area with only 100,000 inhabitant. Retaining a population coefficient of 0.028 as in column 1 rather than 0.048 implies a share of housing in expenditure of 0.363 for a city with the same population as Paris. At the other extreme, a population coefficient of 0.067 as in column 7 implies a housing share of 0.415 for the same hypothetical city.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Robustness checks

In separate appendix J, we report results for a number of robustness checks. In particular, we replicate the results of table 6 for homeowners and renters separately. For our preferred estimation, we find modest differences for the estimated coefficient on city population for renters and homeowners of about 0.02 apart. This is small and statistically insignificant. We also discuss results for a range of further supplementary estimations when also instrumenting for land area in addition to population or controlling directly for household education in reduced form instead of using it as an instrument for income. We also provide evidence to alleviate worries about possible non-linearities in the relationship between the share of income in housing and urban area population.

7. The elasticity of urban costs with respect to population

With both the population elasticity of house prices at the centre and the share of housing in household expenditure now at hand, we can compute their product to obtain the elasticity of urban costs with respect to city population, as per equation (6). Because both quantities possibly vary with city population, the elasticity of urban costs will also vary with population. To illustrate our results, we consider three hypothetical cities. A small city with 100,000 inhabitants, a larger city with a million inhabitants, and a large city with a population equal to that of Paris, slightly above 12 million.

Starting with the elasticity of house prices with respect to city population, we consider four different situations in panel A of table 7. First, we use our preferred OLS estimate of 0.208 from column 8 of table 4 for our baseline calculation. Among all the OLS cross-sectional estimates reported in the rest of table 4 and the separate appendix, the smallest is equal to 0.134 and the largest is 0.306. These extreme values, which are respectively 36% smaller and 47% larger than our baseline, provide useful bounds.[34] Second, we also use estimates for which we allow the population elasticity of house prices to vary with city population. These estimates imply a population elasticity of house prices of 0.205 for a small city, an elasticity of 0.288 for a city with a million inhabitants, and an elasticity of 0.378 for a large city like Paris. Finally, we consider two more extreme cases that rely on values of 0.780 and 0.109 for the population elasticity of house prices.

[34] Alternatively, if we consider the 92 estimates for the coefficient on log population in all the specifications reported in tables 4 and 8 and in the separate appendix which include log population and log area, their mean is 0.224 and the standard deviation is 0.052. Considering two standard deviations around this average comes reasonably close to the values of 0.134 and 0.306 retained in our bounding exercise.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

The former elasticity is estimated using the 2000-2012 house price difference as dependent variable in column 8 of table 5 while the latter is estimated in a specification in the separate appendix that does not include land area as a control. These two values aim to capture a situation where we do not allow for the housing stock to adjust to changes vs., at the other extreme, a situation where we allow for a full adjustment of the urban fringe.

Turning to the share of housing in expenditure, it is equal to 0.325 at the sample mean (which corresponds to a city of 3.17 million inhabitants). We use our preferred estimate for the coefficient on log city population of 0.048. This value predicts a share of housing in expenditure of $0.325 + 0.048 \log(0.1/3.17) = 0.159$ for a city with 100,000 inhabitants, a share of 0.269 for a city with one million inhabitants, and a share of 0.390 for a city like Paris. We focus on these values here. In separate appendix K, we also use alternative predictions arising from estimated coefficients on log population from other columns of table 6.

The urban costs elasticities computed for the four scenarios we consider regarding the population elasticity of house prices are reported in panel C of table 7. Our first finding is that the elasticity of urban costs increases with population size. In the first, third, and fourth scenario, this finding is driven by the larger housing share in expenditure in larger cities. In the second scenario in panel C, the higher urban costs elasticity in larger cities is also explained by the higher population elasticity of house prices in larger cities, which we uncovered some evidence of for the very largest cities in France.

This increase in urban costs with city population is consistent with the 'fundamental tradeoff of spatial economics' (Fujita and Thisse, 2002). Extent literature about agglomeration effects usually regresses log wages or other productivity outcomes on log city population or density and never highlighted much evidence of a deviation from log linearity (Combes and Gobillon, 2015). This is in particular the case for agglomeration effects in France (Combes *et al.*, 2008, 2010). Some convexity for urban costs is thus consistent with a bell shape for the net gains from city population where agglomeration effects may initially dominate but eventually get trumped by urban costs.

We now turn to the differences across rows in panel C of table 7. While the elasticities reported in this panel appear to differ greatly, we must keep in mind that they reflect different thought experiments. The first row is our baseline. The urban cost elasticity is 0.033 for a city with 100,000 inhabitants, 0.056 for a city with one million inhabitants, and 0.081 for a city like Paris. When

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 7: The elasticity of urban costs

	City 1 (pop. 100,000)	City 2 (pop. 1m)	City 3 (pop. Paris)
Panel A. Population elasticity of prices			
Baseline (preferred OLS)	0.208	0.208	0.208
Non-linear population elasticity	0.205	0.288	0.378
12-year adjustment	0.780	0.780	0.780
Allowing for urban expansion	0.109	0.109	0.109
Panel B. Housing share			
Slope of the housing share	0.048	0.048	0.048
Share of housing in expenditure	0.159	0.269	0.390
Panel C. Urban costs elasticity			
Baseline	0.033	0.056	0.081
	(0.007)	(0.005)	(0.007)
Non-linear population elasticity	0.032	0.078	0.147
	(0.007)	(0.007)	(0.017)
12-year adjustment	0.124	0.210	0.304
	(0.036)	(0.047)	(0.069)
Allowing for urban expansion	0.017	0.029	0.043
	(0.004)	(0.003)	(0.005)

Notes: In panel A, row 1, the estimate of 0.208 is our preferred OLS estimate from column 8 of table 4. In row 2, the three estimates are marginal effects computed from column 4 of appendix table 8 in separate appendix G. In row 3, the estimate of 0.780 is for the 2000-2012 difference from column 8 of table 5. In row 4, we use the elasticity of 0.109 estimated in column 8 of appendix table 9 in separate appendix H, which does not include land area as a control. In panel B, for the coefficient on log population in the housing share equation we use our preferred estimate from column 8 of table 6. From these coefficients and the constant of the regression, we compute the predicted housing share in expenditure for our three hypothetical cities. Panel C reports the urban cost elasticity for the all combinations of housing share in expenditure and population elasticity of house prices. Standard errors in brackets are computed from the estimated coefficients and their variances using the following formula for the variance of their product: $var(XY) = var(X)var(Y) + var(X)E(Y)^2 + var(Y)E(X)^2$.

allowing the population elasticity of prices to change with city population in the second row, we find roughly similar urban costs elasticities for the two smaller hypothetical cities but a higher urban cost elasticity of 0.147 for a city the size of Paris. It is difficult to make a definitive choice between our baseline and this higher number for Paris given that we lack power in the estimation given the scarcity of large cities in France.

The third row of panel C of table 7 reports urban costs elasticities that rely on the 2000-2012 variations in house prices and population. The much higher point estimates for the elasticity of house prices with respect to population lead to much higher estimates for the urban costs elasticity: 0.124 for a city with 100,000 inhabitants, 0.210 for a city with a million inhabitants, and 0.304 for a city with the same population as Paris. Although the standard errors are larger than for the other

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

rows of results in the table, these figures are suggestive of large urban cost elasticities in the 'short run' before the supply of housing can adjust (which may take many years in the French context). In turn, these findings are indicative of potentially large frictions in the housing market. When population takes extremely long to adjust following the economic shocks that affect cities, workers may end up residing where housing is affordable and not where they are the most economically productive or where amenities are the highest.

Finally, the last row of panel C of table 7 allows for a full adjustment of cities to population growth, including a physical expansion. With this scenario, the elasticity of urban costs with respect to city population is 0.017 for a city with 100,000 inhabitants, 0.029 for a city with a million inhabitants, and 0.043 for a city of the size of Paris. These figures indicate that when cities can adjust their physical footprint, the costs of urban expansion are low. With an elasticity of wages with respect to city population of about 0.02-0.03 (Combes *et al.*, 2008), our results indicate that in the bell shape associated with the fundamental tradeoff of spatial economics is relatively flat in that case. Cities appear to operate close to net constant returns when they can fully adjust.

If we take seriously the notion of a spatial equilibrium across cities as described in the model of section 2, the difference between the urban cost elasticity and the agglomeration elasticity should be equal to the change in willingness to pay for amenities as city population increases. This difference is negative for small cities and becomes positive for large cities. In a spatial equilibrium framework, we should interpret our results as indicating that amenities are getting mildly better as cities of a larger size are considered (as wages increase less fast than urban costs). The key is nonetheless the small size of these effects, an interpretation consistent with the results of Albouy (2016) for US cities.

8. Conclusion

This paper develops a new methodology to estimate the elasticity of urban costs with respect to city population. Our model derives this elasticity as the product of two terms: the share of housing in consumer expenditure and the elasticity of the price of houses at the centre of cities with respect to city population.

Using data for French urban areas, our preferred estimate of the elasticity of house prices with respect to city population is 0.208 with most alternative estimates being between 0.15 and 0.30 in pooled cross section. Finally, we estimate that the share of housing in expenditure varies from

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

0.159 in small urban areas with 100,000 inhabitants to 0.409 in a city with more than 12 million inhabitants like Paris.

These findings imply elasticities of urban costs from about 0.033 for an urban area with 100,000 inhabitants to 0.081 for an urban area of the size of Paris. These figures refer to the effect of an increase in population, keeping land area constant (i.e., higher density). We think these are the relevant magnitudes to consider in France during our study period as planning regulations strongly discourage urban expansion. Allowing land area to adjust following population increases in cities leads to urban costs elasticities which are smaller by a factor of about two. Looking at changes within cities over time leads instead to larger estimates of the urban cost elasticity as housing supply takes long to adjust.

Given the existence of agglomeration benefits with apparently a constant elasticity of urban wages with respect to city population at around 0.02-0.03 for France, higher elasticities of urban costs in larger cities are consistent with the 'fundamental tradeoff of spatial economics' according to which cities face a region of increasing returns where agglomeration gains dominate urban costs followed by a region of decreasing returns as we consider larger population sizes. This tradeoff may play nonetheless only a minor role in explaining the future evolution of French cities. In the short run, the adjustment of housing supply is expected to play a major role as house prices are fairly sensitive to population changes over a period of a decade or so. In the long run, the bell shape of net urban gains as a function of population is relatively flat so that cities may deviate from their efficient size without leading to large economic losses.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

References

Albouy, David. 2008. Are big cities really bad places to live? Improving quality-of-life estimates across cities. Working Paper 14472, National Bureau of Economic Research.

Albouy, David. 2009. The unequal geographic burden of federal taxation. *Journal of Political Economy* 117(4):635–667.

Albouy, David. 2016. What are cities worth? Land rents, local productivity, and the total value of amenities. *Review of Economics and Statistics* 98(3):forthcoming.

Albouy, David, Gabriel Ehrlich, and Minchul Shin. 2019. Metropolitan land values. *Review of Economics and Statistics* 101:forthcoming.

Alonso, William. 1964. *Location and Land Use; Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.

Au, Chun-Chung and J. Vernon Henderson. 2006. Are Chinese cities too small? *Review of Economic Studies* 73(3):549–576.

Bartik, Timothy. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo (MI): W.E. Upjohn Institute for Employment Research.

Baum-Snow, Nathaniel and Ronni Pavan. 2012. Understanding the city size wage gap. *Review of Economic Studies* 79(1):88–127.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy* 122(3):507–553.

Bleakley, Hoyt and Jeffrey Lin. 2012. Portage and path dependence. *Quarterly Journal of Economics* 127(2):587–644.

Carlino, Gerald A. and Albert Saiz. 2008. Beautiful city: Leisure amenities and urban growth. Federal Reserve Bank of Philadelphia Working Paper No. 08-22.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1):54–70.

Clark, Colin. 1951. Urban population densities. *Journal of the Royal Statistical Association Series A* 114(4):490–496.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2):723–742.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2016. The production function for housing: Evidence from France. Processed, Wharton School, University of Pennsylvania.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80(6):2543–2594.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration economies with history, geology, and worker effects. In Edward L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge (MA): National Bureau of Economic Research, 15–65.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Combes, Pierre-Philippe and Laurent Gobillon. 2015. The empirics of agglomeration economies. In Gilles Duranton, Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5A. Amsterdam: Elsevier, 247–348.

Commissariat Général au Développement Durable. 2015. *RéférenceS: Les Comptes des Transports en 2014*. Paris: Ministère de l'Ecologie, du Développement Durable, des Transports et du Logement.

Davis, Morris A. and Jonathan Heathcote. 2007. The price and quantity of residential land in the United States. *Journal of Monetary Economics* 54(8):2595–2620.

Davis, Morris A. and Michael G. Palumbo. 2008. The price of residential land in large US cities. *Journal of Urban Economics* 63(1):352–384.

Duranton, Gilles and Diego Puga. 2014. The growth of cities. In Philippe Aghion and Steven Durlauf (eds.) *Handbook of Economic Growth*, volume 2. Amsterdam: North-Holland, 781–853.

Duranton, Gilles and Diego Puga. 2015. Urban land use. In Gilles Duranton, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5A. Amsterdam: North-Holland, 467–560.

Duranton, Gilles and Matthew A. Turner. 2016. Urban form and driving: Evidence from US cities. Processed, Wharton School, University of Pennsylvania.

Fujita, Masahisa and Hideaki Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2):161–196.

Fujita, Masahisa and Jacques-François Thisse. 2002. *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge: Cambridge University Press.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport. 2008. Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics* 63(1):1–24.

Guerin-Pace, France and Denise Pumain. 1990. 150 ans de croissance urbaine. *Economie et Statistiques* 0(230):5–16.

Handbury, Jessie and David E. Weinstein. 2015. Goods prices and availability in cities. *Review of Economic Studies* 82(1):258–296.

Henderson, J. Vernon. 1974. The sizes and types of cities. *American Economic Review* 64(4):640–656.

Henderson, Vernon. 2002. Urban primacy, external costs, and the quality of life. *Resource and Energy Economics* 24(1):95–106.

Kline, Patrick and Enrico Moretti. 2014. People, places and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics* 6(1):629–662.

Mills, Edwin S. 1967. An aggregative model of resource allocation in a metropolitan area. *American Economic Review (Papers and Proceedings)* 57(2):197–210.

Muth, Richard F. 1969. *Cities and Housing*. Chicago: University of Chicago Press.

Puga, Diego. 2010. The magnitude and causes of agglomeration economies. *Journal of Regional Science* 50(1):203–219.

Richardson, Harry W. 1987. The costs of urbanization: A four-country comparison. *Economic Development and Cultural Change* 35(3):561–580.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Roback, Jennifer. 1982. Wages, rents and the quality of life. *Journal of Political Economy* 90(6):1257–1278.

Sinai, Todd and Nicholas S. Souleles. 2005. Owner-occupied housing as a hedge against rent risk. *Quarterly Journal of Economics* 120(2):763–789.

Stock, James H. and Motohiro Yogo. 2005. Testing for weak instruments in linear IV regression. In Donald W.K. Andrews and James H. Stock (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 80–108.

Thomas, Vinod. 1980. Spatial differences in the cost of living. *Journal of Urban Economics* 8(1):108–122.

Tolley, George S., Philip E. Graves, and John L. Gardner. 1979. *Urban Growth Policy in a Market Economy*. New York: Academic Press.

United States Bureau of Transportation Statistics. 2013. *Transportation Statistics Annual Report 2013*. Washington, DC: US Government printing office.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Appendix 1. Extending the model to housing construction

We extend our model to encompass the role of land. Housing is produced using land L and non-land K inputs, available at prices $R(\ell)$ and r respectively. To produce an amount of housing $H(\ell)$ at location ℓ, competitive builders face a cost function $C(\ell) \equiv C(r,R(\ell),H(\ell))$. Since free entry among builders at location ℓ implies $P(\ell)\,H(\ell) = C(\ell)$, we can rewrite the elasticity of housing prices with respect to city population as,

$$\epsilon_N^{P(\ell)} \equiv \frac{\mathrm{d}P(\ell)}{\mathrm{d}N}\frac{N}{P(\ell)} = \frac{\mathrm{d}\frac{C(\ell)}{H(\ell)}}{\mathrm{d}N}\frac{N}{P(\ell)} = \frac{N}{P(\ell)H^2(\ell)}\left(H(\ell)\frac{\mathrm{d}C(\ell)}{\mathrm{d}N} - C(\ell)\frac{\mathrm{d}H(\ell)}{\mathrm{d}N}\right). \tag{A1}$$

Since we assume that the cost of non-land inputs remains constant within and between cities, i.e., $\frac{\mathrm{d}r}{\mathrm{d}N} = 0$, totally differentiating the cost function leads to,

$$\frac{\mathrm{d}C(\ell)}{\mathrm{d}N} = \frac{\partial C(\ell)}{\partial R(\ell)}\frac{\mathrm{d}R(\ell)}{\mathrm{d}N} + \frac{\partial C(\ell)}{\partial H(\ell)}\frac{\mathrm{d}H(\ell)}{\mathrm{d}N}. \tag{A2}$$

From the builders' first-order condition for profit maximisation, we have, $P(\ell) = \frac{\partial C(\ell)}{\partial H(\ell)}$. After substituting for $P(\ell)$ using the zero-profit condition, this condition can be rewritten as $C(\ell) = H(\ell)\frac{\partial C(\ell)}{\partial H(\ell)}$. In turn, we can use this expression and equation (A2) to simplify equation (A1) and obtain,

$$\epsilon_N^{P(\ell)} = \frac{N}{C(\ell)}\frac{\partial C(\ell)}{\partial R(\ell)}\frac{\mathrm{d}R(\ell)}{\mathrm{d}N}. \tag{A3}$$

Applying Shephard's lemma, equation (A3) can be written as,

$$\epsilon_N^{P(\ell)} = L(\ell)\frac{N}{C(\ell)}\frac{\partial R(\ell)}{\partial N} = s_h^L(\ell)\epsilon_N^{R(\ell)}, \tag{A4}$$

where $\epsilon_N^{R(\ell)}$ is the elasticity of land prices at location ℓ with respect to city population and $s_h^L(\ell) \equiv \frac{R(\ell)\,L(\ell)}{C(\ell)}$ is the share of land in construction costs at the same location.

We can take expression (A4) at the central location and substitute for ϵ_N^P in equation (6) to obtain

$$\epsilon_N^{UC} = s_E^h\, s_h^L\, \epsilon_N^{\overline{R}}. \tag{A5}$$

where $\overline{R}$ is the price of land at the central location. Instead of using the elasticity of house price to estimate the urban costs elasticity, we can use instead the product of share of land in housing and the elasticity of land prices with respect to housing. Again, these quantities need to be measured at the city centre. This alternative approach relies additionally on the existence of a competitive supply of housing. We implement both approaches in our empirical analysis.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Appendix 2. Further data description

Notary database. Regional notary associations conduct an annual census of all transactions of non-new dwellings. Although reporting is voluntary, about 65% of transactions appear to be recorded. The coverage is higher in Greater Paris (80%) than in the rest of the country (60%). We could not legally append housing prices to the rest of our data directly as the secured data access for house prices is not compatible with the secured access we use for the rest of our data. We could only append price indices for each municipality and year. We are grateful to Benjamin Vignolles for his help with this process.

In addition, note that floorspace area is missing for 25.7% of dwellings that appear in the data. It can be imputed from the FILOCOM repository, which is constructed from property and income tax records. This repository contains information about all buildings in France. For dwellings with missing floorspace area, our imputation attributes the average floorspace area of all dwellings with the same number of rooms in FILOCOM and in the same cadastral section which were involved in a transaction during the same year.[35] This imputation is conducted separately for houses and apartments. It reduces the number of observations with missing floor area to 5.1% (but not to zero as the match with FILOCOM is not perfect). Dwellings for which the floor area cannot be recovered are dropped from the sample. With about 270,000 cadastral sections in France, this imputation is fairly accurate. We can assess this formally by imputing a floor area to all dwellings, including those for which this quantity is observed. Comparing actual and imputed floor areas, the average error is around 5%, and the R^2 of the regression of actual floor areas on imputed ones is about 0.75. Note that accuracy is higher for apartments than for houses since the average error is 2% for the former and 15% for the latter.

Enquête sur le Prix des Terrains à Bâtir (EPTB). While the data is put together by the French Ministry of Sustainable Development, the sample is composed of land parcels originally drawn from Sitadel, the official registry which covers the universe of all building permits for a detached house. Houses must include only one dwelling. Permits for extensions to existing houses are excluded.

Over the 2006-2009 period, parcels were drawn randomly from each municipal strata (about 3,700 of them) which corresponds to a group of municipalities (about 36,000 in France). Overall, two thirds of the permits were surveyed. Some French regions paid for an exhaustive survey: Alsace,

[35] In addition to a municipal identifier, the data contain a cadastral section identifier. Cadastral sections comprise on average less than 100 housing units.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Champagne-Ardennes, Île-de-France, Poitou-Charentes and Pays de la Loire (for Loire-Atlantique and Vendée départements). From 2010 onwards, the survey is exhaustive for the entire country.

Population. We have access to data on population at the municipality level from the 1990 and 1999 general censuses. For every other year from 2000 to 2012, we use the FILOCOM repository that is managed by the *Direction Générale des Finances Publiques* of the French Ministry of Finance. This repository contains a record of all housing units and their occupants. This is a better source of 'high-frequency' population data than the permanent rotating census of population, which replaced the general census in 2004 and surveys 20% of the population of large municipalities every year and smaller municipalities every five years.

Labour force administrative records. We use detailed information from the 1/4 sample of the 1990 census and the 1/20 sample of the 1999 census to construct measures of employment (by municipality of residence) by 4-digit occupational category and by 4-digit sector for each urban area (weighting by survey rates for the data to be representative of the whole population of occupied workers). We also use similar data for 2006 and 2011. The resulting aggregates are used to construct Bartik instruments.

Bartik instruments. To ease the exposition, we index the final year by t and the initial year by $t-1$. Denote N_{jst} employment in urban area j in the four-digit sector s, N_{jt} employment in urban area j, and $N_{(-j)st}$ employment in sector s nationally outside of urban area j. The Bartik sectoral instrument that predicts growth in urban area j between $t-1$ and t is:

$$B_{jt}^{sec} = \sum_{s} \left(\frac{N_{(-j)st}}{N_{(-j)st-1}} \right) \frac{N_{jst-1}}{N_{jt-1}} \tag{B1}$$

A similar computation is applied to construct the Bartik occupation instrument that relies on changes in the four-digit occupational structure of national employment interacted with initial shares of occupations in urban areas.

Income. Mean household income and its standard deviation by municipality and urban area can be constructed using information from each cadastral section (about 100 housing units on average) contained in the FILOCOM repository, which is matched to income tax records.

Land use. We compute the fraction of land that is built up in each municipality and the average height of buildings from the *BD Topo* (version 2.1) from the French National Geographical Institute. This dataset is originally produced using satellite imagery combined with the French land registry. It reports information for more than 95% of buildings in the country including their footprint,

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

height, and use (residential, production, commerce, public sector, religious, etc) with an accuracy of one metre.

Amenity data. We use data from the *French Permanent Census of Equipments* aggregated at the municipality level and maintained by the French Institute of Statistics. The original sources are: the French Ministry for Education for primary, middle, and high schools, the French Ministry of Health for medical doctors, hospitals and other medical services, the registry of establishments (*siren*) for retail establishments, restaurants, and movie theaters, and various other administrative sources.

Historical population data. We use a file containing some information on population by municipality for 27 censuses covering the 1831-1982 period (Guerin-Pace and Pumain, 1990). Over 1831-1910, the data contain only information on "urban municipalities" which are defined as municipalities with at least 2,500 inhabitants. The population of municipalities varies over time. Municipalities appear in the file when their population goes above the threshold and disappear from the file when their population goes below the threshold. Data are aggregated at the urban area level to construct our historical instruments.

Tourism data. These data at the municipality level are constructed by the French Institute of Statistics (INSEE) since 2002 from the census and a survey of hotels. It contains some information on the number of hotels depending on their quality (from zero star to four stars) and the number of rooms in these hotels. We construct our instruments, the number of hotel rooms and the share of 1-star rooms, by aggregating the data for 2006 at the urban area level.

Climate measures The original data come from the ATEAM European project as a high-resolution grid of cells of 10 minutes (approximately 18.6 km) per 10 minutes. These data came to us aggregated at the département level. The value of a climate variable for a département was computed as the average of the cells whose centroid is located in that département. The main climate variables we use is January temperature (in C). We attribute to each municipality the value of its département. The value of an urban area is computed as the average of its municipalities, weighting by the area.

Soil variables We use the European Soil Database compiled by the European Soil Data Centre. The data originally come as a raster file with cells of 1 km per 1 km. We aggregated it at the level of each municipality and urban area. See Combes *et al.* (2010) for further description.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Appendix 3. Second-step: IV results

The four panels of table 8 report results for a series of IV regressions that use house prices as dependent variable. The specifications of panel A include the same set of control variables as our preferred OLS regressions while those of panel B do not include second-step controls beyond time indicators and the variables for which we report coefficients. Panels C and D duplicate the first two panels but consider a dependent variable estimated without first-stage controls. We first note that historical instruments are in general strong whereas amenities tend to be weaker even though they pass weak instrument requirements. Interestingly, including controls appears to matter little for the strength of the instruments. We made the choice of using exactly the same sets of instruments for all panels to allow for more meaningful comparisons of points estimates between panels.

Turning to the analysis of the coefficients, in panel A where controls are included in both steps, the population elasticity remains between 0.215 and 0.266, ranging from marginally above our preferred OLS estimate to about 25% larger. With the IV coefficients being less precisely identified, these differences between IV and OLS are statistically insignificant. We nonetheless keep this variation in mind when computing the urban cost elasticity in section 7. As for the slight increase of the population elasticity, we can only speculate about what might drive it. Although unlikely, our instruments may correct for measurement error. A more plausible explanation is that our OLS estimates suffer from a minor reverse causation bias where urban areas with higher urban costs end up with a smaller population. Another possibility is that our instruments have more bite for larger cities with a larger population elasticity (as shown in separate appendix G).

The estimates of the population elasticity in panels B to D are very close to those of panel A. The main exceptions are the much higher elasticities when using only amenities. These higher coefficients are nonetheless imprecisely estimated so that it is hard to draw conclusions here.

Table 9 duplicates table 8 for land prices instead of house prices. In particular, we use the same instruments. In substance, the results are very similar. The presence or absence of first or second step controls makes only modest differences to the strength of the instruments and the estimated coefficients. The specifications that use only amenities are more fragile and estimations often lead to sizeably higher coefficients for population. With historical instruments, the estimated population elasticities are modestly above our preferred OLS estimate.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 8: The determinants of unit house prices at the centre, IV estimations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Log house prices per m^2, with first-step and second-step controls								
Log population	0.247[a]	0.253[a]	0.247[a]	0.226[a]	0.227[a]	0.267[a]	0.215[a]	0.266[a]
	(0.0281)	(0.0279)	(0.0281)	(0.0248)	(0.0249)	(0.0557)	(0.0226)	(0.0563)
Log land area	-0.170[a]	-0.175[a]	-0.170[a]	-0.140[a]	-0.142[a]	-0.217[a]	-0.150[a]	-0.216[a]
	(0.0239)	(0.0236)	(0.0237)	(0.0204)	(0.0203)	(0.0677)	(0.0213)	(0.0684)
First-stage statistic	111.9	99.1	74.5	119.1	120.1	9.3	101.3	6.2
Overidentification p-value	.	0.18	0.98	0.95	0.20	.	0.29	0.79
Panel B. Log house prices per m^2, with first-step controls and without second-step controls								
Log population	0.237[a]	0.246[a]	0.236[a]	0.211[a]	0.214[a]	0.392[a]	0.237[a]	0.400[a]
	(0.0398)	(0.0386)	(0.0399)	(0.0351)	(0.0351)	(0.0759)	(0.0302)	(0.0768)
Log land area	-0.119[a]	-0.128[a]	-0.119[a]	-0.0858[a]	-0.0891[a]	-0.276[a]	-0.0789[b]	-0.287[a]
	(0.0343)	(0.0334)	(0.0343)	(0.0308)	(0.0305)	(0.0927)	(0.0334)	(0.0941)
First-stage statistic	116.3	117.8	76.7	122.8	129.2	9.9	155.0	7.1
Overidentification p-value	.	0.25	0.60	0.60	0.05	.	0.02	0.72
Panel C. Log house prices per m^2, without first-step controls and with second-step controls								
Log population	0.204[a]	0.210[a]	0.202[a]	0.188[a]	0.189[a]	0.243[a]	0.187[a]	0.249[a]
	(0.0216)	(0.0215)	(0.0216)	(0.0187)	(0.0187)	(0.0498)	(0.0164)	(0.0608)
Log land area	-0.170[a]	-0.175[a]	-0.168[a]	-0.147[a]	-0.149[a]	-0.223[a]	-0.151[a]	-0.231[a]
	(0.0174)	(0.0172)	(0.0174)	(0.0148)	(0.0146)	(0.0610)	(0.0158)	(0.0753)
First-stage statistic	111.9	99.1	74.5	119.1	120.1	9.3	101.3	6.2
Overidentification p-value	.	0.14	0.13	0.15	0.18	.	0.21	0.12
Panel D. Log house prices per m^2, without first-step and second-step controls								
Log population	0.194[a]	0.206[a]	0.191[a]	0.173[a]	0.177[a]	0.353[a]	0.205[a]	0.420[a]
	(0.0298)	(0.0281)	(0.0305)	(0.0259)	(0.0255)	(0.0687)	(0.0230)	(0.0949)
Log land area	-0.126[a]	-0.137[a]	-0.123[a]	-0.0994[a]	-0.103[a]	-0.280[a]	-0.0905[a]	-0.364[a]
	(0.0281)	(0.0267)	(0.0286)	(0.0255)	(0.0250)	(0.0854)	(0.0280)	(0.119)
First-stage statistic	116.3	117.8	76.7	122.8	129.2	9.9	155.0	7.1
Overidentification p-value	.	0.12	0.08	0.08	0.05	.	0.02	0.13
Instruments								
Urban population in 1831	Y	Y	Y	Y	Y	N	N	N
Urban pop. density in 1851	Y	Y	Y	N	N	N	N	N
Urban area in 1881	N	Y	N	N	N	N	N	N
Urban pop. density in 1881	N	N	N	Y	Y	N	Y	N
January temperature	N	N	Y	Y	N	N	N	Y
Number of hotel rooms	N	N	N	N	N	Y	Y	Y
Share of one-star hotel rooms	N	N	N	N	Y	Y	Y	Y
Observations	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937

Notes: [a]: significant at 1% level; [b]: significant at 5% level; [c]: significant at 10% level. Standard errors are clustered at the urban area level. The first-step controls are the same as in column 9 of table 3. The second-step controls correspond to the controls used in columns 2, 5, and 8 of table 4. All estimations are performed with LIML. The critical value for 10% maximal LIML size of Stock and Yogo (2005) weak identification test is 7.03 for columns (1) and (6) and 5.44 for other columns. These critical values do not depend on control variables because the role of those is first conditioned out before the estimation. This conditioning does not affect the estimates and their standard error for population and area but it is required due to multi-collinearity arising from a few urban areas with too few observations. The first-stage statistics is the Kleibergen-Paap rk Wald F.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018

Table 9: The determinants of unit land prices at the centre, IV estimations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Log land prices per m^2, with first-step and second-step controls								
Log population	0.684^a	0.713^a	0.697^a	0.650^a	0.647^a	0.776^a	0.627^a	0.920^a
	(0.0588)	(0.0570)	(0.0609)	(0.0522)	(0.0512)	(0.125)	(0.0467)	(0.264)
Log land area	-0.507^a	-0.530^a	-0.517^a	-0.453^a	-0.455^a	-0.661^a	-0.469^a	-0.845^b
	(0.0528)	(0.0512)	(0.0542)	(0.0467)	(0.0457)	(0.157)	(0.0477)	(0.336)
First-stage statistic	101.9	97.3	66.9	110.8	111.2	9.7	76.3	6.5
Overidentification p-value	.	0.03	0.00	0.00	0.11	.	0.17	0.03
Panel B. Log land prices per m^2, with first-step controls and without second-step controls								
Log population	0.676^a	0.699^a	0.677^a	0.621^a	0.625^a	0.905^a	0.651^a	0.888^a
	(0.0657)	(0.0634)	(0.0659)	(0.0575)	(0.0574)	(0.155)	(0.0510)	(0.175)
Log land area	-0.439^a	-0.461^a	-0.441^a	-0.366^a	-0.373^a	-0.687^a	-0.363^a	-0.664^a
	(0.0604)	(0.0586)	(0.0606)	(0.0546)	(0.0539)	(0.194)	(0.0564)	(0.220)
First-stage statistic	112.2	115.8	74.6	118.3	121.5	8.8	150.1	6.2
Overidentification p-value	.	0.07	0.28	0.31	0.09	.	0.06	0.21
Panel C. Log land prices per m^2, without first-step controls and with second-step controls								
Log population	0.729^a	0.766^a	0.738^a	0.716^a	0.713^a	0.752^a	0.719^a	0.781^a
	(0.0676)	(0.0663)	(0.0692)	(0.0594)	(0.0583)	(0.150)	(0.0533)	(0.273)
Log land area	-0.690^a	-0.718^a	-0.697^a	-0.668^a	-0.667^a	-0.707^a	-0.664^a	-0.744^b
	(0.0624)	(0.0606)	(0.0633)	(0.0549)	(0.0537)	(0.186)	(0.0566)	(0.346)
First-stage statistic	101.9	97.3	66.9	110.8	111.2	9.7	76.3	6.5
Overidentification p-value	.	0.02	0.01	0.01	0.85	.	0.82	0.01
Panel D. Log land prices per m^2, without first-step and second-step controls								
Log population	0.729^a	0.765^a	0.729^a	0.696^a	0.697^a	0.843^a	0.738^a	0.832^a
	(0.0676)	(0.0648)	(0.0674)	(0.0584)	(0.0578)	(0.175)	(0.0564)	(0.177)
Log land area	-0.629^a	-0.663^a	-0.630^a	-0.586^a	-0.588^a	-0.702^a	-0.568^a	-0.687^a
	(0.0713)	(0.0686)	(0.0710)	(0.0643)	(0.0632)	(0.221)	(0.0662)	(0.223)
First-stage statistic	112.2	115.8	74.6	118.3	121.5	8.8	150.1	6.2
Overidentification p-value	.	0.02	0.85	0.87	0.77	.	0.54	0.76
Instruments								
Urban population in 1831	Y	Y	Y	Y	Y	N	N	N
Urban pop. density in 1851	Y	Y	Y	N	N	N	N	N
Urban area in 1881	N	Y	N	N	N	N	N	N
Urban pop. density in 1881	N	N	N	Y	Y	N	Y	N
January temperature	N	N	Y	Y	N	N	N	Y
Number of hotel rooms	N	N	N	N	N	Y	Y	Y
Share of one-star hotel rooms	N	N	N	N	Y	Y	Y	Y
Observations	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933

Notes: ^{a}: significant at 1% level; ^{b}: significant at 5% level; ^{c}: significant at 10% level. Standard errors are clustered at the urban area level. The first-step controls are the same as in column 9 of table 3. The second-step controls correspond to the controls used in columns 2, 5, and 8 of table 4. All estimations are performed with LIML. The critical value for 10% maximal LIML size of Stock and Yogo (2005) weak identification test is 7.03 for columns (1) and (6) and 5.44 for other columns. These critical values do not depend on control variables because the role of those is first conditioned out before the estimation. This conditioning does not affect the estimates and their standard error for population and area but it is required due to multi-collinearity issues arising from a few urban areas with too few observations. The first-stage statistics is the Kleibergen-Paap rk Wald F.

Downloaded from https://academic.oup.com/restud/advance-article-abstract/doi/10.1093/restud/rdy063/5135829 by guest on 15 November 2018