JOINT USC FBE FINANCE SEMINAR &
MACROECON & INT'L FINANCE WORKSHOP
presented by Pierre-Olivier Weill
FRIDAY, April 24, 2009
10:30 am – 12:00 pm, **Room: HOH-304**

Why Has House Price Dispersion Gone Up?[*]

Stijn Van Nieuwerburgh[†]   Pierre-Olivier Weill[‡]

April 15, 2009

Abstract

We set up and solve a spatial, dynamic equilibrium model of the housing market based on two main assumptions: households with heterogenous abilities flow in and out metropolitan areas in response to local wage shocks, and the housing supply cannot adjust instantly because of regulatory constraints. In our equilibrium, house prices compensate for cross-sectional productivity differences. We increase productivity dispersion in the calibrated model in order to match the 30-year increase in cross-sectional wage dispersion that we document based on metropolitan-level data. We show that the model quantitatively matches the observed 30-year increase in dispersion of house prices across U.S. metropolitan areas. It is consistent with several other features of the cross-sectional distribution of house prices and wages.

1 Introduction

This paper sets up and solves a spatial equilibrium model of the housing market. The model is a dynamic version of the canonical compensating-wage differential model of Rosen (1979) and Roback (1982). In contrast with the urban economics tradition of studying house prices in one region given some exogenous outside option of living in the countryside (the "reservation locale"), we model the entire distribution of regions. This has several advantages. First, the model provides predictions for the entire cross-section of income, house prices, and construction. This facilitates comparison with the data, which are realizations of that joint equilibrium distribution. Second, the outside option of living in the reservation locale is no longer exogenous: instead, its value is determined in equilibrium. This is important when studying the effect of changes in aggregate variables on house prices, because they operate precisely through endogenous changes in the value of the reservation locale. Third, our model has the benefit of numerical tractability. This is useful when we solve for transitional dynamics in order to evaluate the effect of changes in aggregate variables on house prices.

We model metropolitan areas as a collection of geographically separated islands, randomly hit by idiosyncratic and persistent productivity shocks in the non-housing sector. Construction firms can build new houses in any metropolitan area, but new construction is irreversible and is subject to supply regulation, implying that the local housing supply cannot adjust instantly in response to a local productivity shock. We assume households differ in their ability and are fully mobile: they can freely move across metropolitan areas, but are constrained to live in the same area they work. The equilibrium provides the endogenous joint dynamics of cross-sectional house prices, construction, wages and employment. In particular, more able households sort into more productive regions. House price differentials between metropolitan areas compensate for the wage differential of the marginal, lowest-ability household in the location, making him indifferent between staying and moving to the next best metropolitan area. Households also live in smaller and more expensive quarters if they choose to work in higher-wage metropolitan areas. Lastly, higher-wage metropolitan areas have, on average, a larger housing stock and a larger workforce.

We use this model to study the determinants of the evolution of the regional house price distribution since 1975. While the steep rise of house prices has received a lot of attention in the media and in the literature,[1] one salient feature of that distribution which has received much less attention is the steep rise in the *dispersion* of house prices across regions. Indeed,

---

[1] Case and Shiller (1987), Himmelberg, Mayer, and Sinai (2005), Campbell, Davis, Gallin, and Martin (2006), Gyourko, Mayer, and Sinai (2006) and many others document the historical evolution of house prices at the national and local level.

consider the cross-sectional coefficient of variation (CV), the ratio of the standard deviation to the mean, which is a scale-neutral measure of dispersion. In population-weighted terms, the CV of house prices increased from 0.15 in 1975 to 0.53 in 2007. As we explain below, an increase in the cross-sectional dispersion of regional productivity shocks generates an increase in the dispersion of house prices. Such an increase in the dispersion of productivity is consistent with the increase in the dispersion of labor income (henceforth "wages") that we document in our regional data. However, the increase in the CV of wages is much smaller and increases much less than that of house prices: from 0.08 in 1975 to 0.17 in 2007. Our main quantitative exercise shows that an increase in productivity dispersion can simultaneously generate the 8.6 point increase in the CV of wages and the 38 point increase in the CV of house prices. In fact, our benchmark model predicts a somewhat larger 51 point increase in the CV of house prices.

Our main transitional experiment is to feed in just enough of an increase in productivity dispersion to generate the observed increase in wage dispersion between 1975 and 2007, while keeping both the dispersion of ability and housing supply regulation constant. The increase in productivity dispersion creates flows of workers towards high-productivity metropolitan areas, driving local house prices up because of limited housing supply. Conversely, households flow out of low-productivity areas, driving local house prices down. This increases house price *dispersion.* In addition to explaining the increase in dispersion, our model also generates one third of the observed increase in the average house price level, despite keeping average productivity constant. The level effect arises because house prices are a *convex* function of productivity in the model, or equivalently, because price *differentials* increase with productivity. This effect arises for two reasons. First, as productivity increases, productivity differentials are compensated by housing expenditure differentials for smaller and smaller housing sizes, because households reduce their housing consumption in response to higher prices. Since the price differential is the housing expenditure differential *per unit* of housing consumption, it increases with productivity. A second effect arises because of assortative matching of ability with regional productivity. As productivity increases, the ability of the marginal household increases and so does the wage differential. Since the price differential compensates for the wage differential, the price differential increases with productivity.

The model endogenously generates several features of the joint wage-house price distribution. First, because of assortative matching, it features larger increases in wages at the top of the regional wage distribution. Second, it generates a large increase in the ratio of housing price to construction cost, consistent with the findings of Glaeser, Gyourko, and Saks (2007) and Davis and Heathcote (2005) that the non-structure component of house prices

has become more important over the last thirty years. Third, it is consistent with the observation that, in repeated cross-sectional regressions of house prices on wages, the coefficient on wages increased over the 1975-2007 period. In words, a one dollar wage differential became compensated by a larger and larger house price differential. This increasing price-wage sensitivity arises because we engineer the increase in wage dispersion through an increase in productivity dispersion. We would have obtained the opposite under the alternative assumption that the increase in wage dispersion reflects an increase in ability dispersion. Hence, the increased sensitivity provides indirect evidence for the mechanism through which we increase wage inequality. Fourth, the model generates an increasing concentration of people in highly productive regions; the fraction of people working in the highest wage quintile increases by 8.2 percentage points in model and data.

A second transition exercise illustrates that the increase in income inequality is an essential part of the explanation for increasing house price dispersion. To make that point, we attempt to generate the observed increase in house price level and dispersion in a counter-factual economy that experienced no increase in wage dispersion, but only an increase in housing supply regulation. Holding the dispersion of wages constant at its 1975 level, we tighten limits on construction over a thirty-three year period. In order to maximize the impact of supply regulation on prices, we assume that the tightening is more pronounced in high-productivity metropolitan areas. By 2007, the model does predict an increase in the level and dispersion of cross-sectional house prices, but the effects are quantitatively very small. Indeed, the negative impact of regulation on local housing supply is almost completely offset by the equilibrium response of households moving out of tightly regulated areas towards less regulated areas. Because this shifts the local demand down at the same time as the supply, the price impact of supply regulations ends up being quantitatively small.

Related Literature   Our model features wage differences across regions, which may reflect productivity gains from agglomeration effects (e.g., Glaeser, Scheinkman, and Schleifer, 1992, 1995). An alternative view in the urban literature is that house price differences reflect differences in amenities and other local traits (e.g, Roback, 1982). We note that our regional productivity process admits a broader interpretation that encompasses both productivity and amenities, which are then reflected into house prices.

Several authors have argued that housing supply regulation is an important determinant of house prices (Glaeser and Gyourko, 2003, 2005; Glaeser, Gyourko, and Saks, 2005, 2007; Quigley and Rosenthal, 2005; Quigley and Raphael, 2005). Both explanations for the increase in level and dispersion of house prices we investigate crucially rely on housing supply

regulation. Our quantitative exercise suggests that an initial level of regulation combined with an increase in wage dispersion go a long way towards accounting for the facts.

Our work is related to Gyourko, Mayer, and Sinai (2006), who also study the relationship between the U.S. income distribution and cross-sectional house prices. They provide a two-location model, in which regions differ by housing supply and households differ by income and preference for a particular location. A household lives in the low-supply location if it either has a strong preference for it or a high income. Our paper differs in terms of the economic mechanism -households move for productive rather than preference reasons-, and in terms of methodology. The upside of working with a dynamic and stochastic equilibrium model that is amenable to calibration is that it holds the promise of distinguishing between different mechanisms by looking at their quantitative implications. Spiegel (2001) also studies the link between wages, house prices, and construction in an equilibrium model with a moral hazard friction.

Our model of spatial allocation shares many features with labor search models (Lucas and Prescott, 1974; Alvarez and Veracierto, 1999, 2006) and the spatial allocation model of Shimer (2005). We complement this literature by focusing on a different friction. In our setup, households do not incur any cost when moving between islands. Instead, the flow of households between islands is limited by the supply of housing in each island. Coen-Pirani (2006) works with an island model for studying migration patterns between US states. Eeckhout (2004) uses similar techniques to explain the size distribution of cities.

Our approach to assortative matching of individual ability with regional productivity builds on the model of Sattinger (1993). Recent application of his assignment model to other markets are Gabaix and Landier (2008), Tervio (2008), and Costinot and Vogel (2008). Relative to these papers, we face the additional technical difficulty that the number of households matching with a given region is endogenous because of divisible housing.

Our work connects to the macroeconomics literature that documents increases in wage dispersion at the individual level (e.g. Hornstein, Krusell, and Violante, 2004) and studies its effects on risk-sharing (Krueger and Perri, 2006; Storesletten, Telmer, and Yaron, 2004; Heathcote, Storesletten, and Violante, 2005; Lustig and Van Nieuwerburgh, 2006) and on asset pricing (Constantinides and Duffie, 1996; Cogley, 2002; Lustig, 2003; Storesletten, Telmer, and Yaron, 2006; Lustig and Van Nieuwerburgh, 2007). Nakajima (2005) most closely connects these two strands of the literature. He sets up an incomplete markets OLG economy and studies a steady state with low (1967) and one with high individual income inequality (1996). He solves for portfolio allocations between housing and physical capital as well as for equilibrium prices. He finds that the increase in income inequality leads to increased

precautionary savings, lower interest rates, and 9% higher house prices. Our model studies the spatial dimension of income and house price inequality in the presence of housing supply restrictions in a complete markets economy.

Finally, our work is complementary to macro and asset pricing models that focus on the role of housing as a consumption good and/or a collateral asset (Iacoviello, 2005; Krueger and Fernández-Villaverde, 2006; Piazzesi, Schneider, and Tuzel, 2006; Lustig and Van Nieuwerburgh, 2005, 2007). In our model, the discount factor is constant across dates and states. An interesting avenue for future work is to incorporate the insights from the asset pricing literature. Recent work by Ortalo-Magné and Prat (2008) along these lines derives equilibrium house and stock prices in a spatial model and shows that a version of the CAPM holds.

The rest of the paper is organized as follows. Section 2 presents our island model. Section 3 calibrates a steady-state of the model to match features of 1975 data. Section 4 provides the quantitative impact of increasing wage dispersion and tightening regulation on prices and the distribution of population and construction. Section 5 concludes.

2 An Island Economy

2.1 The Economic Environment

We start by describing the stochastic environment as well as the technologies for producing housing and non-housing consumption. The next paragraph describes households.

2.1.1 Information and Technology

Time is taken to be discrete and runs forever. The economy is made up of a measure-one continuum of homogenous regions we call islands. At each time $t \in \{1, 2, \ldots\}$, an island's production function of non-housing consumption good is linear in labor with an idiosyncratic productivity $A_t \in \mathbb{R}_+$. We take the productivity process $\{A_t\}_{t=1}^{\infty}$ to be a first-order, N-states, Markov chain with possibly time-dependent support $A_{1t} < A_{2t} < \ldots < A_{Nt}$ and transition function $Q_t(A, \cdot)$. We assume that the productivity process is persistent in the sense that, if $A' > A$, then $Q_t(A', \cdot)$ first-order stochastically dominates $Q_t(A, \cdot)$.[2]

Each island starts at time zero with an initial state $s^0 \equiv (A_0, H_0)$, where A_0 is the initial productivity and $H_0 \in (0, H_{\max})$ the initial housing stock. Although we allow the initial housing stock of an island to be correlated with the initial productivity, we assume that,

---

[2]This definition of a persistent stochastic process is used, for instance, by Lucas and Prescott (1974).

conditional on A_0, it does not help predict the future path of the productivity.[3] We denote by $\mu_0(ds^0)$ the initial cross-sectional probability measure over initial states s^0. At subsequent times, an island is indexed by its history $s^t = (A^t, H_0)$, where $A^t \equiv (A_0, A_1, \ldots, A_t)$ is the productivity history and H_0 is the initial housing stock. As it is standard, starting from the initial measure μ_0, and using the transition functions $Q_t(A, \cdot)$, one constructs inductively the entire sequence of unconditional probability measures, $\mu_t(ds^t)$, over histories s^t.

Each period, firms purchase construction material in order to construct housing services in the islands of their choosing. A representative construction firm can transform Δ units of construction material into housing consumption according to the Leontief production function $\min\{\Delta, \Pi_t(A_t)\}$, where $\Pi_t(\cdot)$ is some strictly positive bounded function of the current productivity A_t of the island. This function is designed to represent not only technological and physical constraints on construction (such as the amount of constructible land) but also regulatory constraints. One may think of $\Pi_t(A_t)$ as the number of building permits in an island with current productivity A_t. Allowing permits to depend on time and productivity will allow us to capture the commonly held view that regulation became tighter over time, and even more so on highly productive metropolitan areas.

We assume that construction is irreversible and the stock of housing consumption depreciates at rate $\delta \in (0, 1)$. These assumptions are summarized by the constraints

$$\Delta_t(s^t) \geq 0 \tag{1}$$
$$\Delta_t(s^t) \leq \Pi_t(A_t) \tag{2}$$
$$H_t(s^t) = (1 - \delta)H_{t-1}(s^{t-1}) + \Delta_t(s^t), \tag{3}$$

where $\Delta_t(s^t)$ denotes the construction flow and $H_t(s^t)$ denotes the housing stock in island s^t. Inequality (1) is the irreversibility constraint, inequality (2) is imposed by the Leontief construction technology, and equation (3) is the law of motion for the housing stock. Lastly, the resource constraint for construction material is

$$\int \Delta_t(s^t)\mu_t(ds^t) \leq M, \tag{4}$$

where M denotes the per-period endowment of perishable construction material.

---

[3]Formally, $\Pr(A_t \mid A_{t-1} \ldots, A_0, H_0) = \Pr(A_t \mid A_{t-1}, \ldots, A_0)$. This will imply that, in an dynamic equilibrium, the housing stock does not Granger (1969)-cause productivity.

2.1.2 Preferences and Endowments

The economy is populated by a measure one continuum of infinitely-lived and fully mobile households with discount factor $\beta \in (0,1)$. Households have separable utility for non-durable consumption and housing services. Their flow utility for non-durable consumption is taken to be linear, while their flow utility over housing consumption is represented by some strictly increasing, strictly concave, bounded above and twice continuously differentiable function $v : (0, \infty) \rightarrow \mathbb{R}$. We assume in addition that $v(\,\cdot\,)$ is unbounded below, meaning that $v(h)$ goes to minus infinity as h goes to zero. Lastly, and without further loss of generality since $v(h)$ is bounded above, we assume that $v(h)$ goes to zero as h goes to infinity.[4,5]

We assume that households differ in terms of their ability. Namely, at each time a household supplies inelastically $e \in [\underline{e}, \overline{e}]$ effective units of labor in the island of its choosing, and we let $f(e)$ be the cross-sectional density of effective units of labor in the population. Given firms linear production function and competition, a household with ability e working in an island with productivity A earns a wage $e \times A$.

Although the assumptions of ability differences is not needed for the main qualitative results of the paper, it turns out to be crucial in our quantitative exercise. As will become clear later, it will allow us to address a standard self-selection problem when we use wages differentials between regions to infer productivity differentials. Indeed, ability may be imperfectly observable and high ability households may prefer to locate in high productivity areas. This implies that observed wage differentials partly reflect ability differentials, and may be larger than the underlying productivity differentials.

Letting $n_t(e, s^t)$ be the number of households with ability e who choose to live in island s^t, we have

$$\int n_t(e, s^t)\mu_t(ds^t) = f(e), \tag{5}$$

since the density of households with ability e must be equal to $f(e)$.

A key assumption of our model is that our fully mobile households are constrained to live in the same island they choose to work.[6] In other words, letting $h_t(e, s^t)$ be the housing

---

[4]An iso-elastic utility function $v(h) = h^{1-\gamma}/(1 - \gamma)$ satisfies these parametric assumptions when $\gamma > 1$. Lemma 6 of Appendix A.1 shows that these properties imply that the utility function $v(h)$ satisfies Inada (1963) conditions.

[5]The key implication of quasi-linearity is that the marginal utility of consumption is equated across islands and, in that sense, that our *ex-ante* identical households are fully insured. Appendix B.3 explains how to extend and keep our dynamic model tractable when households have a general convex utility function.

[6]This is how the Bureau of Economic Analysis defines a metropolitan statistical area (MSA), the empirical counterpart of islands in the model.

consumption per household of ability e in island s^t, we have the local resource constraint

$$\int_{\underline{e}}^{\overline{e}} n_t(e, s^t) h_t(e, s^t)\, de \leq H_t(s^t). \tag{6}$$

An *allocation* is a collection of measurable functions specifying, for each time $t \in \{1, 2, \ldots\}$, each ability e and each island s^t, the number $n_t(e, s^t)$ of households, the housing consumption $h_t(e, s^t)$ per household, the flow $\Delta_t(s^t)$ of construction, and the housing stock $H_t(s^t)$. An allocation is *feasible* if it satisfies the constraints (1)-(6).

2.2 Definition of a Competitive Equilibrium

Every period, a representative competitive construction firm is endowed with construction permits and purchases material at price C_t in order to produce and sell housing consumption in the islands of its choosing.[7] The price of housing consumption in island s^t is denoted $P_t(s^t)$. Hence, the representative construction firm problem is to choose quantities $\Delta_t(s^t)$ of construction material in order to maximize

$$\int \left(P_t(s^t) - C_t \right) \Delta_t(s^t) \mu_t(ds^t), \tag{7}$$

subject to (1)-(2).

We assume that competitive real estate firms purchase the stock of housing consumption in all islands and rent it to households.[8] The rent in island s^t is denoted by $R_t(s^t)$. Clearly, a real estate firm finds it optimal to supply its entire housing stock as long as the rent is strictly positive.

Competition among real estate firms implies that the current price of housing consumption is equal to the rent plus the present value of the price next period, net of depreciation:

$$P_t(s^t) = R_t(s^t) + \beta(1 - \delta) E_t \left[P_{t+1}(s^{t+1}) \,|\, s^t \right]. \tag{8}$$

Under the transversality condition

$$\lim_{T \to \infty} \beta^T E_t \left[P_{t+T} \left(s^{t+T} \right) \,|\, s^t \right] = 0, \tag{9}$$

---

[7]Because of linearity of the construction technology, the distribution of permits across construction firms does not matter. An alternative approach would be to assume that households are endowed with permits to construct on their land and sell them to construction firms. This would deliver the same results.

[8]This assumption is made for expositional simplicity. As it is standard with frictionless housing markets, the same equilibrium price would arise if households were purchasing their homes instead of renting them.

we obtain the Rosen and Topel (1988) result that a house price is equal to the expected present value of rents net of depreciation

$$P_t(s^t) = E_t \left[\sum_{j=0}^{\infty} \beta^j (1-\delta)^j R_{t+j}(s^{t+j}) \,\middle|\, s^t \right].$$ (10)

In Appendix B.2 we provide a complete treatment of the household's inter-temporal problem. Households choose in which location to live and work at any given time, as well as their housing and non-housing consumption in each island. Households receive all profits from the real estate sector: the profit of selling the endowment of construction material to construction firms, the profit of building houses, and the profit of renting out houses. Because of full mobility, the households' inter-temporal problem can be simplified dramatically: it reduces to a sequence of static problems. Every period, a household chooses in which island to work, and how much housing to rent in that island. Namely, given optimal housing consumption choice, a household with ability e who chooses to live in island s^t enjoys the value:

$$u_t(e, s^t) = e \times A_t + \sup_{h \geq 0} \left\{ v(h) - R_t(s^t) h \right\}.$$ (11)

And, of course, a household's optimal location choice is to work and live in any island that yields the maximum value:

$$U_t(e) = \sup_{s^t} u_t(e, s^t).$$ (12)

A *competitive equilibrium* is a price system and a feasible allocation such that: i) the price and the rent satisfy (10), ii) given the price $P_t(s^t)$ of housing consumption and the price C_t of construction material, the construction flow $\Delta_t(s^t)$ solves the construction firm's problem, iii) given the rent $R_t(s^t)$, housing consumption $h_t(s^t)$ solves the household's problem and the allocation of households across islands is individually optimal, that is

$$n_t(e, s^t) \geq 0 \qquad \text{if} \qquad u_t(e, s^t) = U_t(e)$$ (13)

$$n_t(e, s^t) = 0 \quad \text{otherwise.}$$ (14)

2.3 Equilibrium Characterization

In this subsection we characterize an equilibrium. Section 2.3.1 provides some elementary properties of rents and prices. Section 2.3.2 shows that an equilibrium features assortative matching of ability with productivity. Section 2.3.3 derives the dynamics of the housing stock. Finally, Section 2.3.4 proves that an equilibrium exists and is unique, and provides an efficient numerical algorithm for calculating it.

2.3.1 Rents, Prices, and Housing Consumption

In this section we derive elementary properties of an equilibrium. First, in an island without population, the rent must be equal to zero. Otherwise, a real-estate firm would find it optimal to supply a positive quantity of housing and the market would not clear.[9]

Second, the rent is a function of an island's current productivity, and does not depend on other idiosyncratic characteristics of the island, such as past productivity shocks or the local housing supply. Formally, if the current productivity of an island is equal to A_{it}, the ith element of the productivity grid, then its rents is equal to some R_{it}. Otherwise, if two islands with the same current productivity had different rents, the location choice of households living in the high-rent island would not be optimal: they would prefer to move to the low-rent island where they would earn the same wage (because of equal productivity) but pay a lower rent.

Third, the rent must be increasing in productivity, and strictly increasing across populated islands. Indeed consider two islands with productivity $A_{it} < A_{jt}$. If island A_{it} is populated, then $R_{it} < R_{jt}$. Otherwise the location choice of households in the low productivity island would not be optimal: they would prefer to move to the high-productivity where they would earn a higher wage and pay a lower rent. If island A_{it} is not populated, then $R_{it} = 0$ and, evidently, $R_{it} \leq R_{jt}$.

Plugging the rent back into the pricing equation (10) and using the Markov property shows that the price $P_t(s^t)$ is a function of the current productivity but does not depend on other idiosyncratic characteristics of an island. In addition, because the rent is increasing in productivity, and the productivity process is persistent the price is also increasing in productivity.

Now, going back to the quasi-linear household's problem of equation (11), it follows that a household's optimal housing consumption does not depend directly on its ability e.

---

[9]Note that the local housing supply is strictly positive in every island. Indeed, each island starts with a strictly positive housing stock, and the depreciation rate δ is strictly less than one.

Furthermore, because the rent is an increasing function of productivity, it immediately follows that housing consumption is a decreasing function of productivity.

This discussion is summarized in the following proposition:

Proposition 1. *At each time $t \in \{1, 2, \ldots\}$, housing consumption, rent, and price are functions of the islands' current productiveness, and do not depend on any other idiosyncratic characteristic of the island. In addition, housing consumption h_{it} is decreasing, rent R_{it} is increasing, and price P_{it} is increasing, with the island's current productivity, A_{it}.*

2.3.2 Assortative Matching

In this paragraph we show that our model formalizes the commonly held view that households with higher ability tend to work in higher productivity locations. This is because the households' objective function is super-modular: the cross-derivative between ability and productivity is positive. This implies that households' location decisions are weakly increasing in ability, i.e. high-ability households find it optimal to locate in higher productivity locations. Formally, we prove the following proposition:

Proposition 2. *At each time, there is an integer $p \in \{1, \ldots, N\}$ and a sequence of ability cutoffs*

$$e_{pt} = \underline{e} < e_{p+1t} < \ldots < e_{Nt} < e_{N+1t} \equiv \overline{e}$$

such that

1. *An island is populated if and only if its productivity is greater than A_{pt}.*

2. *For all $i \in \{p, \ldots, N\}$, households with ability levels $e \in (e_{it}, e_{i+1t})$ strictly prefer to live in islands with current productivity A_{it}.*

3. *For all $i \in \{p, \ldots, N-1\}$, households with ability $e = e_{i+1}$ are indifferent between living in islands with current productivity A_{it} or A_{i+1t}.*

In the proposition, we suppressed the dependence of the integer p on time to simplify notations. The results are illustrated in Figure 1: households with ability in the interval (e_{it}, e_{i+1t}) go work in islands with current productivity A_{it}.

We now provide a method for constructing the cutoffs e_{it} which has two main benefits: it leads to a simple existence proof and it provides an efficient algorithm for calculating the equilibrium assignment of households to islands. We start by letting U_{it} be, for all $i \geq p$, the
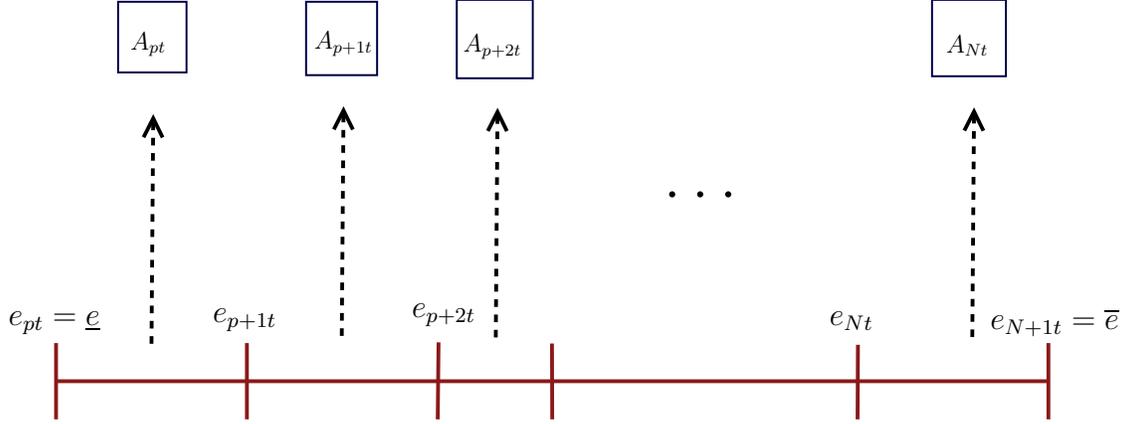
Figure 1: The assignment of abilities to islands

maximum attainable utility of a household with ability e_{it}. For all $i < p$, we let $e_{it} = e_{pt} = \underline{e}$ and $U_{it} = U_{pt}$. Then, we have

$$U_{it} \geq A_{it} e_{it} + v(h_{it}) - R_{it} h_{it}, \tag{15}$$

with an equality for $i \geq p$, and where h_{it} denote housing consumption in island i at time t. After plugging the first-order condition $R_{it} = v'(h_{it})$ and letting $w(h) \equiv h v'(h) - v(h)$, we obtain that housing consumption is:[10]

$$h_{it} = w^{-1} \left(\max\{ A_{it} e_{it} - U_{it}, 0 \} \right). \tag{16}$$

Keeping in mind that housing consumption h_{it} only depends on an island's current productivity, we add up the housing market clearing conditions (6) across all islands with current productivity A_{it} to find:

$$n_{it} h_{it} \leq H_{it},$$

with an equality for $i \geq p$, and where H_{it} denotes the *average* housing stock and n_{it} the *average* number of households per island with current productivity A_{it}.[11] Thus, the total

---

[10]Note that the formula implies that, for $i < p$, the argument of $w^{-1}(\cdot)$ is zero and the optimal housing consumption is $h_{it} = \infty$. Such housing demand is consistent with optimality because the rent is zero and $v(h)$ is strictly increasing. It is also consistent with market clearing because islands $i < p$ are not populated.

[11]Note that, in all islands s^t with current productivity A_{it}, the housing stock $H_t(s^t)$ and the population $n_t(s^t)$ will differ from the averages H_{it} and n_{it}. This is because, unlike housing consumption, the housing stock $H_t(s^t)$ and the population $n_t(s^t)$ depend on the initial housing stock and past productivity realizations.

number of households living in islands with productivity A_{it} is equal to

$$\mu_{it} n_{it} = \mu_{it} H_{it}/h_{it} = \mu_{it} H_{it} \Phi \left(\max\{A_{it} e_{it} - U_{it}, 0\} \right),$$

where μ_{it} denotes the number of islands with productivity A_{it} and $\Phi(x) \equiv 1/w^{-1}(x)$. From Proposition 2, it follows that the total number of households living in islands A_{it} must be equal to the number of households with ability in between cutoffs e_{it} and e_{i+1t}. Taken together, this gives the difference equation:

$$F(e_{i+1t}) - F(e_{it}) = \mu_{it} H_{it} \Phi \left(\max\{A_{it} e_{it} - U_{it}, 0\} \right), \tag{17}$$

where $F(e)$ is the cumulative distribution function of the ability distribution. This difference equation allows to calculate the sequence of equilibrium cutoffs given a sequence $U_{1t}, \ldots, U_{Nt}$ of maximum attainable utilities.

To calculate the maximum attainable utilities on the right of side of equation (17), we use the indifference conditions of households with abilities $e_{p+1t}, \ldots, e_{Nt}$. Indeed, because of point 3 in Proposition 2, we know that a household with ability e_{i+1t} is indifferent between the following two alternatives. He can work in an island with productivity A_{i+1t} and receive utility U_{i+1t}. Or, he can work in an island with productivity A_{it}, earning a wage $e_{i+1t} A_{it}$, enjoying a quantity h_{it} of housing consumption, and paying a rent $R_{it} h_{it}$. This adds up to a utility $e_{i+1t} A_{it} + v(h_{it}) - R_{it} h_{it} = U_{it} + A_{it} (e_{i+1t} - e_{it})$. Equating these two utilities and using equation (15), we obtain the difference equation:

$$U_{i+1t} = U_{it} + (e_{i+1t} - e_{it}) A_{it}. \tag{18}$$

Note that this difference equation also holds for $i < p$, given our convention that $e_{it} = \underline{e}$ and $U_{it} = U_{pt}$.

Taken together, the difference equations (17) and (18) suggest a simple "shooting" algorithm for calculating the equilibrium assignment of households to islands, given a distribution of housing stocks $H_{1t}, \ldots, H_{Nt}$. Given an initial condition U_{1t} and $e_{1t} = \underline{e}$, we calculate e_{2t} using equation (17), then U_{2t} using equation (18), then e_{3t} using (17), and so on until we obtain the entire sequence $U_{1t}, \ldots, U_{Nt}$ and $e_{1t}, \ldots, e_{Nt}, e_{N+1t}$.[12] Moreover, one can easily show that the terminal cutoff e_{N+1t} is a decreasing function of the initial condition U_{1t}. We use this monotonicity property to prove that there exists a unique U_{1t} such that $e_{N+1t} = \overline{e}$.

---

[12]To make this procedure well defined, we need to artificially extend the domain of the cdf $F(e)$ above the upper bound $\overline{e}$. We deal with this in detail in Appendix A.2.2.

This discussion is summarized in the following proposition:

Proposition 3 (Equilibrium Assignment). *Given a distribution $H_{1t}, \ldots, H_{Nt}$ of housing stocks, there exists a unique pair of sequences $\{e_{it}\}_{i=1}^{N+1} \in [\underline{e}, \overline{e}]^N$ and $\{U_{it}\}_{i=1}^{N} \in \mathbb{R}^N$ solving the difference equations (17)-(18) with initial condition $e_{1t} = \underline{e}$ and terminal condition $e_{N+1t} = \overline{e}$.*

Another reason why this procedure is computationally convenient is that we always start shooting at the lower bound of the productivity grid, $i = 1$, so that there is no need to guess-and-verify the cutoff p. Instead, p can be calculated in a second step, as the smallest grid point i such that $e_{it} A_{it} > U_{it}$.

2.3.3   Housing Stock

To complete our characterization of an equilibrium, we need to solve for the distribution $H_t(s^t)$ of housing stocks. To that end, we first note that the linearity of the construction firm's objective (7) implies that an optimal construction plan is simply to build $\Pi_t(A_{it}) \equiv \Pi_{it}$ units of housing consumption in every island where $P_t(s^t) > C_t$. Since we proved that $P_t(s^t)$ only depends on the current productivity A_{it} and is increasing, it follows that there is some productivity cutoff A_{ct} such that a construction firm builds a quantity $\Delta_{it} = \Pi_{it}$ of housing if $A_{it} > A_{ct}$, a quantity $\Delta_{ct} \in [0, \Pi_{ct}]$ if $A_{it} = A_{ct}$ and does not construct anything otherwise. Plugging this back into the resource constraint (4) for construction material, we obtain

$$\mu_{ct} \Delta_{ct} + \sum_{i=c+1}^{N} \mu_{it} \Pi_{it} \leq M, \tag{19}$$

with an equality if the following condition is satisfied

$$\sum_{i=1}^{N} \mu_{it} \Pi_{it} > M. \tag{20}$$

Condition (20), which we assume holds from now onwards, implies that there is a large supply of constructible land. That is, the amount of housing that could be built on all constructible land, on the left-hand side of (20), is greater than the amount of housing that can be built with the available supply M of construction material. Under condition (20), if at the cutoff $\Delta_{ct} < \Pi_{ct}$, then the representative construction firm is indifferent between constructing or not, implying that

$$C_t = P_{ct}, \tag{21}$$

at each time $t \in \{1, 2, \ldots\}$. If $\Delta_{ct} = \Pi_{ct}$, then $C_t = P_{ct}$ is also an equilibrium construction price.

2.3.4 An Algorithm

Taken together, the above paragraphs provide an algorithm for calculating an equilibrium:

1. First, one uses (19) to solve, at each time, for the construction cutoff c and the construction plan $\{\Delta_{it}\}_{i=1}^{N}$.

2. Given the construction cutoffs, one can use the difference equation (3) to calculate the distribution of housing stocks across islands. Note that, unlike rents, prices, and housing consumptions, the local housing stocks will depend on the entire productivity history s^t of an island. However, as Section 2.3.2 made clear, only N moments of the housing stock distribution matters: the average housing stock H_{it} per island with current productivity A_{it}. These N moments jointly solve the difference equation:

$$H_{it} = (1 - \delta) \sum_{j=1}^{N} \frac{\mu_{jt-1} Q_{t-1}(j, i)}{\mu_{it}} H_{jt-1} + \Delta_{it}, \tag{22}$$

where the first term on the right-hand side is the (depreciated) average housing stock last period in an island with current productivity A_{it}.

3. Given the distribution of housing stocks H_{it}, one solves for the ability cutoffs e_{it} and the maximum attainable utilities using (17) and (18).

4. Finally, given the ability cutoffs and the maximum attainable utilities, one solves for the housing consumption using (16), for population using $n_{it} = H_{it}/h_{it}$, for rents using $R_{it} = v'(h_{it})$ and for prices using the present value formula (10).

Based on these four steps, we first show:

Proposition 4 (Existence and Uniqueness)**.** *There exists an equilibrium. The equilibrium is unique in the sense that all equilibria share the same rents R_{it}, housing consumptions h_{it}, population n_{it}, housing stocks H_{it}, and ability cutoffs e_{it}.*

Some equilibrium objects are not uniquely determined: for instance, if $\Delta_{ct} = \Pi_{ct}$, then all construction prices $C_t \in [P_{c-1t}, P_{ct}]$ are consistent with optimality. Also, households at the ability cutoffs e_{it} are indifferent and of measure zero, so their island assignment is

15

indeterminate. These dimensions of indeterminacy, however, do not change the answers of the questions at hand.

Note that the algorithm translates into an efficient computational procedure, because the distribution of housing stocks can be characterized before calculating prices. Also, given that the other objects of interest only depend on the current productivity of an island, we do not need to calculate the entire population distribution, $n_t(s^t)$, to calculate population-weighted moments. Instead, it is enough to calculate $n_{it} = \mathbb{E}\left[n_t(s^t)\,|\,A_{it}\right]$, the average population per island A_{it}. Our discrete state space allows us to calculate expectations and present value quickly, while approximating standard continuous-state processes using the quadrature methods of Tauchen and Hussey (1991). Lastly, the discrete state also speeds up the calculation of households' equilibrium assignment, relative to the ordinary differential equations arising in a continuous-state model. Our algorithm results in quick calculations of transitional dynamics, without using any linearization technique. This turns out to be important for our results, because the price impact of wage dispersion stems from a non-linear convexity effect.

2.4   Convexity

We now show an important property of our model: that increasing productivity dispersion increases house price levels.

Proposition 5 (Convexity). *At each time $t \in \{1, 2, \ldots\}$, the rent R_{it} is a convex function of an island's current productivity, in that:*

$$\frac{R_{i+1t} - R_{it}}{A_{i+1t} - A_{it}} \tag{23}$$

is increasing in i.

The following back-of-the-envelope calculation provides intuition. Consider a household of ability e_{i+1t}, who is indifferent between island A_{i+1t} and island A_{it}:

$$e_{i+1t}A_{i+1t} + v(h_{i+1t}) - R_{i+1t}h_{i+1t} = e_{i+1t}A_{it} + v(h_{it}) - R_{it}h_{it}$$
$$\Rightarrow \quad R_{i+1t}h_{i+1t} - R_{it}h_{it} = e_{i+1t}\left(A_{i+1t} - A_{it}\right) + v(h_{i+1t}) - v(h_{it}).$$

The equation says that the housing expenditure differential between island i and $i+1$ compensates for the wage differential of the *marginal* household, as well as for the utility differential arising from differential housing consumptions. Now use the last equation to calculate the

16

housing expenditure differential, holding housing consumption constant:

$$
(R_{i+1t} - R_{it})\, h_{it} = e_{i+1t} (A_{i+1t} - A_{it}) + v(h_{i+1t}) - v(h_{it}) - R_{it+1} (h_{i+1t} - h_{it})
$$
$$
\Rightarrow \quad (R_{i+1t} - R_{it})\, h_{it} \simeq e_{i+1t} (A_{i+1t} - A_{it}) + (v'(h_{i+1t}) - R_{i+1t}) (h_{i+1t} - h_{it}),
$$
$$
\Rightarrow \quad \frac{R_{i+1t} - R_{it}}{A_{i+1t} - A_{it}} \simeq \frac{e_{i+1t}}{h_{it}}, \tag{24}
$$

where the second line follows from a first-order approximation of $v(h_{i+1t}) - v(h_{it})$ and the last line because $R_{i+1t} = v'(h_{i+1t})$.[13] Equation (24) shows that convexity arises for two reasons. The first reason is that, as productivity increase, households respond to higher house prices by reducing their housing consumption: thus, productivity differentials $A_{i+1t} - A_{it}$ are compensated by housing expenditure differentials for smaller and smaller a housing consumption. Since the rent differential, $R_{i+1t} - R_{it}$, is the housing expenditure differential *per unit* of housing consumption, it becomes larger and larger. The second effect arises because ability increases with productivity. Intuitively, the rent differential compensates for the wage differential of the marginal household. But since the ability of the marginal household increases with productivity, the wage differential and the corresponding rent differential becomes larger and larger.

The house price implications of a productivity-induced increase in wage dispersion follow immediately from Proposition 5. Consider a (mean-preserving) increase in productivity dispersion, holding the mapping from productivity to price the same. This mechanically increases wage dispersion. Also, because R_{it} is an increasing function of A_{it}, the rent increases in high-productivity islands and decreases in low-productivity islands. Hence, the cross-sectional dispersion of rents increases. Now, convexity means that the rent increases by more in high-productivity islands than it decreases in low-productivity islands. This creates two level effects. First, the cross-sectional average rent goes up. Second, the house price level increases *in every island*. To understand this second effect, consider the example of an independent and identically distributed productivity process. That is, every period, the productivity in an island is an independent draw from the cross-sectional distribution. Our pricing equation (8) implies that the price in an island with current productivity A is

$$
P_{it} = R_{it} + \frac{E[R_{jt+1}]}{1 - \beta(1 - \delta)}, \tag{25}
$$

where the expectation is taken with respect to the cross-sectional distribution of productivity.

---

[13]Our formal proof does not rely on any approximation. Note also that, if the productivity state were continuous instead of discrete, then equality (24) would hold exactly.

Convexity implies that an increase in productivity dispersion increases the second term in the price equation (25).[14] In words, the house price increases because households anticipate that the rent will increase by more when the island draws a high productivity than it will decrease when it draws a low productivity.

3   Calibration Parameters and Targets

In this section we calibrate our model so that the initial steady state of the model matches key moments of the wage and population distribution in 1975. We then engineer an increase in the dispersion of wages of the observed magnitude, and ask whether the model can account for some key features (what we call "targets") of the post-1975 house price distribution. The results of this exercise are in the next section. We start here by describing the moments of the data we are trying to match.

3.1   Targets in the Data

Our goal is to account for the joint distribution of wages and house prices across U.S. metropolitan areas over the last 33 years. First, we briefly describe the wage and price data; details on all data definitions, sources, and construction are relegated to Appendix D. Our sample consists of 330 U.S. metropolitan statistical areas with annual data from 1975 to 2007.

Raw Data   Wages are measured using nominal wage per job data available from the Bureau of Economic Analysis Regional Economic Information System (REIS). This is a measure of the average annual earnings per employed worker in that region. We also obtain the number of jobs for each metropolitan area from REIS, and use them to calculate population-weighted moments. To calculate the *real* wage per job we deflate the nominal wage per job by a regional cost-of-living index which excludes housing. The index combines data from the Bureau of Labor Statistics to compute year-to-year changes in each MSA, together with data on relative non-housing prices across MSAs from the private data vendor COLI. The base year is 1983-84, when the *average* region has a non-housing price level normalized to 100.

House prices are measured as the nominal median home value. We combine the median single-family home values from the 2000 Census with the Freddie Mac Conventional Mortgage

---

[14]If the wage process is persistent, then the same effect operates in the long run. Indeed, by ergodicity, the distribution of the wage T periods ahead converges to the cross-sectional distribution as T goes to infinity.

Home Price Index (CMHPI), a repeat-sale house-price index available from 1975 until 2007.[15] Proceeding as with nominal wages, we deflate nominal home values by the non-housing price index to obtain real prices. A balanced panel of prices is only available for a subset of 81 regions. The sample with house price data gradually increases from 81 MSAs in 1975 to 330 MSAs in 1994, and stays constant thereafter. We refer to this growing sample as the unbalanced panel. Figure 2 plots the population-weighted cross-sectional average and coefficient of variation of real wages for the balanced panel of 81 regions (top row) and the unbalanced panel of 330 regions (bottom row). Figure 3 does the same for real home prices. The figures indicates that changes in cross-sectional level and dispersion are similar for both samples.[16] In what follows, we will focus on the unbalanced panel of 330. Both the CV and the level of real wages increase moderately, while the CV and the level of real prices increase strongly.

Figure 2: First and Second Moments of Real Wages in the Data

The top row of the figure plots the population-weighted cross-sectional average, cross-sectional standard deviation, and cross-sectional coefficient of variation of the real wage per job for a balanced panel of 81 metropolitan statistical areas. The bottom panel reports the same moments for an unbalanced panel of regions that grows over time from 81 to 330 metropolitan statistical areas. The real wage per job is calculated as the nominal wage per job divided by the regional non-housing price index.


---
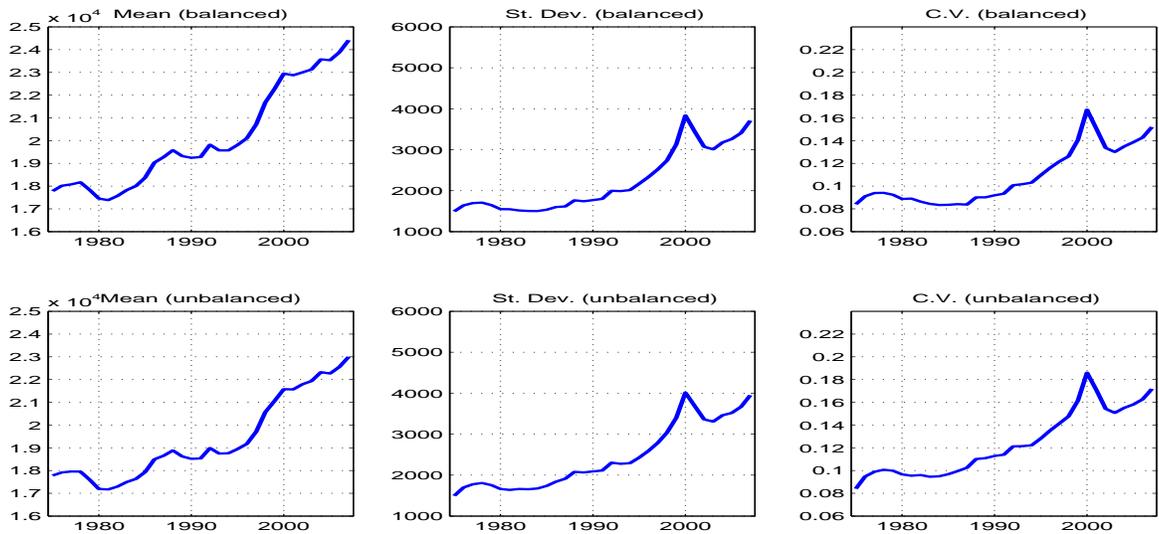
[15]The CMHPI is a constant quality house price index. It pertains to single-family properties financed with a mortgage below the conforming loan limit. See Case and Shiller (1987). We use fourth quarter values.

[16]Non-population-weighted moments (not reported here) also display similar increases in level and dispersion.

Figure 3: First and Second Moments of Real Home Prices in the Data

The top row of the figure plots the population-weighted cross-sectional average, cross-sectional standard deviation, and cross-sectional coefficient of variation of real single-family home values in the data for a balanced panel of 81 metropolitan statistical areas. The bottom panel reports the same moments for an unbalanced panel of regions that grows over time from 81 to 330 metropolitan statistical areas. the real home value is computed as the nominal home value divided by the
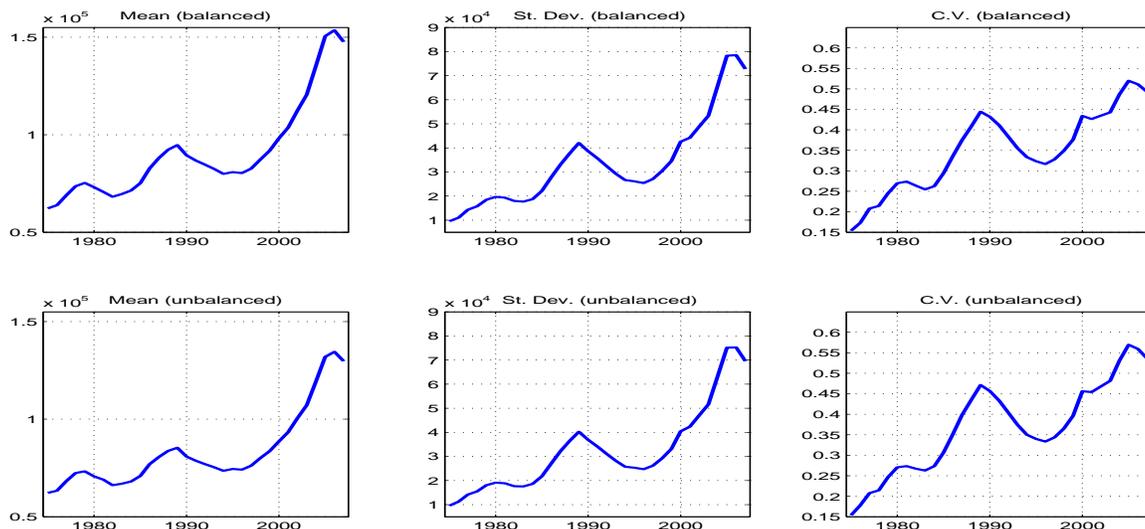
While the house price series is a constant-quality series, it does not correct for the increase in the quantity of housing services that a typical single-family house provides. We measure this quantity as the average square foot of completed single-family units, for sale inside metropolitan areas. The Census's construction statistics indicate that house size has grown from 1,715 to 2,563 square feet between 1975 and 2007, an average growth rate of 1.256% per year. As explained below, we de-trend house price by size to remove the mechanical increase in house prices that is due to the increase in house size.

Data in the De-Trended Economy. Appendix E explains that, because of the quasi-linear preferences, our model is not consistent with balanced growth in productivity and in the quantity of housing services.[17] At the same time, the data do not seem consistent with balanced growth either: over our sample period, population-weighted average real wages grew at $g_w = 0.80\%$ per year in the unbalanced panel whereas house size grew at the higher rate of

---

[17]The model is, however, consistent with population growth, i.e. growth in the number of households. Using a standard argument, Appendix E shows how relative prices and per-household quantities are the same in a model with population growth and in an appropriately transformed model without growth.

$g_H = 1.256\%$ per year. These observations suggest that our model is better suited to explain de-trended data. This leads us to feed in and confront the model with de-trended data.

In the data, we deflate the house price series by the observed size of a house, i.e. by $(1 + g_H)^t$ in year t. This generates a constant-size house price series. We also remove a trend from real wage data. It is important not to remove the entire growth rate g_w because, as will become clear later, the model endogenously generates a trend in wages even when productivity has no trend. If g_m denotes the (endogenous) growth rate of wages in the de-trended model, then we de-trend the real wage at a rate $g_w - g_m$. This guarantees that the de-trended wage grows at the same rate g_m per year in the calibrated model and in the de-trended data.

The resulting de-trended, population-weighted average real wage increases from \$17,782 in 1975 to \$19,489 in 2007, an annual change of $g_m = 0.29\%$. The population-weighted CV of real wages increases from 0.084 in 1975 to 0.172 in 2007. The de-trended, population-weighted average real house price increases from \$62,212 in 1975 to \$87,013 in 2007, an annual change of 1.05%. Note in particular that, without the de-trending by house size, house prices go up to about \$150,000 in 2007 (see Figure 3), so that a large fraction of the run-up in levels is indeed accounted for by the increase in house size. The population-weighted CV of real de-trended house prices increases from 0.154 in 1975 to 0.536 in 2007, which is similar to the increase shown in Figure 3. Our main quantitative exercise is to feed in the observed wages into the model and to ask what fraction of the observed increase in level and especially in dispersion of home prices it can explain. In particular, can a small increase in wage dispersion of 8.8 points generate a large increase in house price dispersion of 38.2 points?

Price-Wage Sensitivity In reality, several factors outside of our model presumably contribute to the observed house price dispersion. For example, in addition to productivity differentials, amenity differentials may matter. To quantify the importance of wages differentials in creating price differentials, we compute the following R^2 statistic:

$$R^2 = 1 - \frac{\text{var}(p_{it} - \hat{p}_{it}^d)}{\text{var}(p_{it})}, \tag{26}$$

where p_{it} denotes the (real de-trended) house price in region i and period t and $\hat{p}_{it}^d$ denotes a linear projection of real house prices on real wages in the data. This projection employs all available year-MSA observations, i.e., the entire unbalanced panel, and takes into account the population size of each region. We find that 26.5% of variation in house prices can be explained by variation in wages. This R^2 value is an important target for our model to

match.[18] Namely, we will feed in the model observed wage, obtain the model-predicted prices $\hat{p}_{it}^{m}$, and calculate their R^2 with in equation (26) replacing the linear projection $\hat{p}_{it}^{d}$ by $\hat{p}_{it}^{m}$.

In addition, we study the slope b_p from a repeated cross-sectional regression of house prices on wages. These regressions again weigh the importance of each region by its population. In the data, the slope of this regression increases from 0.81 in 1975 to 7.89 in 2007, suggesting that the sensitivity of house prices to wages has increased substantially over time. Explaining the increase in this slope coefficient is another target for our model.

3.2 Calibration

This section discusses the calibration. Table 1 summarizes the parameters and their benchmark values. Our calibration strategy has three components. Five parameters, indicated with an "E" in the first column of the table, are chosen so that the initial steady state of our model replicates key moments of the 1975 data. One parameter, indicated with a "E^{*}" superscript, is chosen so that we replicate the observed increase in wage dispersion. In order to pick these six parameters, we solve the model repeatedly until the six endogenously generated moments exactly match their counterpart in the data. All other parameters are set "externally" to conventional values. We now describe these choices in more detail.

3.2.1 Preferences

The model is calibrated at annual frequency. We set the households' time discount factor to $\beta = .951$ in order to match the average real interest rates of 5.15% on the conforming 30-year fixed rate mortgage between 1975 and 2007. This is the most relevant interest rate to use in the present-value formula that pins down the house price. We let households have an iso-elastic utility function $v(h) = \kappa h^{1-\gamma}/(1-\gamma)$ over housing consumption, implying that the price elasticity of housing demand is equal to $-1/\gamma$. Because the micro-level evidence suggests an elasticity of about -0.5, we set $\gamma = 2$.[19] The parameter κ governs the housing expenditure share. We choose κ so that the housing expenditure to income ratio in the model (first averaged across regions, then across time) matches the value of 0.12 in the 2000 Census data.

---

[18]In order to focus on the lower-frequency relationship, we run this regression on price and wage data that have been averaged over five-year periods. The R^2 is similar for other horizons. For example, it is 25.6% for annual observations (no averaging) and 27.8% based on ten-year averages.

[19]Hanushek and Quigley (1980) exploit a natural experiment where a subgroup of 586 low income renters in Phoenix and 799 households in Pittsburgh received rent subsidies ranging from 30-60%, whereas a control group received nothing. They estimate long-run elasticities of -.45 for Phoenix and -.64 for Pittsburgh, based on estimates of how fast the housing demand adjusts towards an equilibrium level in the two years of data.

3.2.2 Productivity and Ability

Productivity Regions differ in their productivity. We assume that the regional productivity process follows a geometric random walk with exponential lifetime. As in Yaari (1965), this guarantees the existence of a stationary distribution. Every period, a measure $\lambda \in (0, 1)$ of new region is created with an initial log productivity $a_t = \log(A_t)$ drawn from a normal distribution with mean μ_{bt} and standard deviation σ_{bt}. In every subsequent period, a region either disappears with probability λ or survives with probability $1 - \lambda$. In case of survival, it draws a new log productivity

$$a_t = a_{t-1} + \sigma_{at}\varepsilon_t, \tag{27}$$

where ε_t is a standard normally distributed shock.[20] are Appendix C.1 explains that the cross-sectional distribution of log productivity across islands is not known in closed form (although it can be easily written as a mixture of normal densities) and behaves like a Pareto distribution in its two tails. However, the first and second moment of the cross-sectional productivity distribution can be calculated easily. As explained in the appendix, and in keeping with the discrete productivity grid of our theoretical setup, we also discretize the productivity process on $N = 190$ Gaussian quadrature points using the quadrature methods of Tauchen and Hussey (1991).

Thus, in the initial steady state, the productivity process is characterized by four parameters: μ_b, σ_{a0}^2, σ_b^2, and λ. We choose the parameter μ_b to match the 1975 population-weighted average wage per job. We set the variance of productivity innovation σ_{a0}^2 and the variance of productivity at birth σ_b^2 to match the population weighted coefficient of variation of wage per job, with the identifying assumption that initial conditions, represented by σ_b^2, explain half of the variance in productivity. The results turn out to be rather insensitive to this identifying assumption. Lastly, we exogenously fix the death rate λ at 1% per year, which delivers an autocorrelation of wages that is statistically indistinguishable from the data; see Section 4.1 below.

Ability Households differ in their effective units of labor $e \in [\underline{e}, \overline{e}]$. Given that the assumed cross-regional productivity distribution exhibits Pareto behavior in its tails, we chose an ability distribution $f(e)$ with the same properties. Namely, we assume that ability is distributed

---

[20]Although our theoretical section did not consider such exogenous entry and exit of productive locations, it is in fact a straightforward extension. There are only things that need to be adjusted. First the average housing stock per region of type A_i depreciates faster by a factor $(1 - \lambda)$. Second the discount factor for the present value formula is also scaled down by $(1 - \lambda)$.

according to a double-Pareto distribution. Appendix C.1 provides the details. We choose the parameters of this distribution so that, first, ability has a mean normalized to 1. Second, the Pareto coefficient k_e, which relates inversely to the cross-sectional standard deviation of ability, is such that we match the 1975 sensitivity of prices to wages b_{p0} of 0.81. To understand the relationship between ability differentials and the sensitivity of house prices to wages, note that the price differential between two regions is determined by the wage differential of the *marginal household*, i.e. the wage drop that household would incur by moving from its current regions to the next-highest productivity one (holding of course its ability the same). In short, the price differential reflects a constant-ability wage differential. The key observation is that the *observed* wage differential may be larger than the price differential because it not only reflects productivity differentials, but also the ability differentials of households. In particular, the more cross-sectional dispersion in ability there is (lower k_e), the smaller price differentials are relative to wage differentials. Cross sectionally, this results in a lower sensitivity, b_p, of house prices to wages (see Appendix B.1). Of course, when there are no ability differences (k_e is very large), the sensitivity of prices to wages is at its highest. The 1975 data suggest a sensitivity of prices to wages requiring a Pareto coefficient $k_e = 17.89$, which implies a cross-sectional standard deviation of ability of 0.079.

The variance $0.0062 = 0.079^2$ of ability of we use is not excessive. Appendix B.4 spells out an argument which shows that the difference between the overall cross-sectional variance of individual wages and the variance of wages that are averaged at the regional level is an upper bound on the cross-sectional variance of ability. Based on micro data from Heathcote, Storesletten, and Violante (2008) and our regional data, we find that our ability variance is thirteen times smaller than this upper bound.

Transition Exercise   In our main exercise below, we engineer an increase in the cross-regional wage dispersion of the same magnitude as in the data. The discipline in our transition experiment comes from assuming that the entire increase is generated by an increase in the log productivity dispersion over time. The dispersion of ability (i.e., k_e) and all other parameters stay constant during the transition. To keep things simple, we assume a linear increase in the productivity dispersion between the initial steady state (1975) and period 32 of the transition (2007). From period 33 (2008) onwards, we assume that productivity dispersion stays constant at its final steady state value. We choose the final steady state value so that we exactly hit the coefficient of variation of the real wage per job in 2007.

3.2.3 Construction Technology

We set the housing depreciation rate $\delta = 0.016$. This is the average depreciation rate over the last 35 years, calculated as the ratio of depreciation at current cost and the current cost net stock of residential fixed assets from the Fixed Asset Tables provided by the Bureau of Economic Analysis. See also Davis and Heathcote (2005).

We set the yearly endowment M_t of construction material so that, year-by-year, the aggregate housing supply per household in the model, H_t, matches the de-trended house size per household we observe in the data. For the years 1975-2007, we feed in the observed de-trended size. After 2007, the de-trended size equals the initial steady state value of 1,715 square feet. This procedure amounts to exogenously fixing the total quantity of square feet in the economy and letting the equilibrium endogenously allocate these square feet across regions.

The last object we need to calibrate is the permit function $\Pi_t(A_t)$ which measures the maximum amount of construction per period in a region with productivity A_t. We start by assuming a constant permit function, i.e. $\Pi_t(A_t) = \pi_a$. This captures the notion that housing supply regulation is no tighter in some metropolitan areas than in others.[21] Because the parameter π_a determines the distribution of housing across islands, it indirectly governs the distribution of households across islands. Indeed, a larger π_a allows firms to construct more housing in high-wage areas, which in turns increases the population in these areas. This observation motivates us to choose π_a in order to match the 1975 concentration of jobs in high-wage metropolitan areas, as follows.[22] Each year, we sort the largest sample of regions we can find, into (equal-sized) wage quintiles and compute the fraction of jobs in each quintile. The data indicate an increase in the fraction of jobs that are concentrated in the highest wage quintile (Q5) from 64.9% in 1975 to 73.1% in 2007 (see Appendix D.5). We choose the parameter π_a in order to match the 1975 Q5 number.

---

[21]While this assumption is uncontroversial as a description of the early 1970s, some have argued that housing supply restrictions have become tighter and more widespread over time. In Section 4.6, we allow for such a change and find that the quantitative effects of tightening regulation had negligible effects on equilibrium house prices.

[22]The strategy of calibrating $\Pi(A_t)$ directly to regulation data, instead of relying on its indirect impact on the population distribution, is not feasible. While there exist indices of housing supply constraints at the metropolitan level (e.g., Malpezzi (1996) and Saks (2005)), they have no time-series dimension. In addition, there is no natural mapping between such ordinal measures and our quantity constraint $\Pi(A_t)$.

3.2.4 Demographics

In order to control for the effect of demographics on house prices, we feed into our model the observed 1975-2007 data for the growth rate in the number of households g_{Nt} as well as for the number of jobs per household. Appendix E.3 provides the details. The growth rate in the number of households enters in the depreciation rate of the de-trended housing stock: $(1-\delta)/(1+g_{Nt})/(1+g_H)$.[23] Finally, the number of jobs is relevant for household earnings, which is the product of the real wage per job and the number of jobs per household.

4   Quantitative Results

In this section we investigate the effects of feeding in the model the progressive increase in wage dispersion we observed in the data. We study the economy's transition from 1975 until 2007, and ultimately towards the new steady-state. In the figures we present below, the red dashed line denotes the initial 1975 steady state, the green dashed line denotes the final steady state, the blue solid line denotes the transition path, and the solid red line with dots denotes the data.

4.1   Wages

Our calibration parameters are picked so that, in the initial steady state of the model, the population-weighted cross-sectional mean and CV of the real wage per job is the same as in the 1975 data. Then, along the transition path, we linearly increase the dispersion of productivity by increasing the innovation variance of productivity shocks over 32 periods; see Appendix C.1.1 for the details. We pick the path of productivity dispersion so that the population-weighted CV of real wages in period 32 of the model's transition is the same as in the 2007 data. The procedure is summarized in Figure 4. The left panel plots the exogenous cross-sectional standard deviation of productivity we feed in the model: it increases from 0.0063 in the initial steady state to 0.1029 in period 32, and then stays constant until the final steady state. The right panel shows that this indeed allows to match the 8.8 point increase in the CV in the data (red solid line). The middle panel shows that we also match the increase in the cross-sectional average wage between 1975 and 2007 as part of our de-

---

[23]In the data, the number of households has grown faster than the population (1.53% average growth per year versus 1.05%) because the number of persons per household has declined (-0.46% growth). We feed in the faster growth rate in the number of households and thus capture its effect on house prices.

Figure 4: Increasing the Wage Dispersion

The left panel plots the cross-sectional standard deviation of log productivity that arises from our calibration (exogenous). The middle panel plots the equilibrium population-weighted average of the (endogenous) real wage while the right panel plots the equilibrium population-weighted coefficient of variation. The red dashed line is the initial steady-state, the green dashed line is the final steady-state, and the solid line denotes the first 200 years along the transition path. This figure is for our benchmark calibration. The solid red line with dotted markers plots the data from 1975 until 2007.
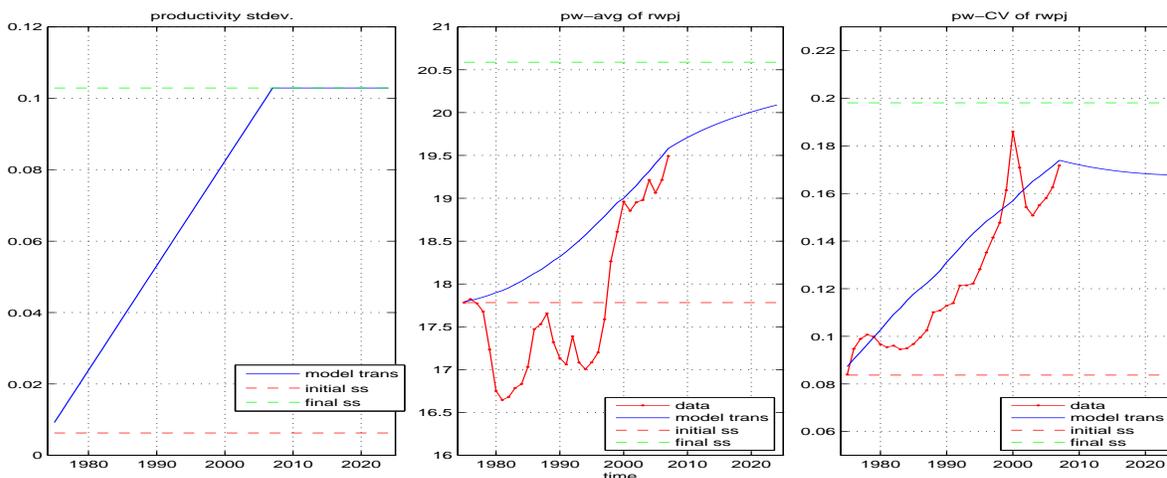
trending procedure.[24] Note that, even though the standard deviation of log productivity is held fixed after 2007 (period 32 of the transition), the mean and CV of wages continue to rise as the economy converges towards the final steady state. As more construction continues to take place in the newly productive regions, the population continues to relocate there. Table 2 shows the key moments in the data (Panel A) and in our benchmark model (Panel B). Rows 1 and 2 contain the moments for wages.

Before moving to the house price implications, we show that the model fits additional moments of the wage data. First, it fits the evolution of the cross-sectional wage distribution beyond the mean and CV. We compare the tenth, fiftieth, and ninetieth percentiles of the wage distribution in model and in data (on a population-weighted basis). These percentile cutoffs are 16.0 ($16,000 real wage per job in 1983 dollars), 17.7, and 19.6 in the model's initial steady state and 15.9, 17.9, and 20.0 in the 1975 data. Thirty-two periods into the transition, these same percentile cutoffs are 15.6, 19.3, and 23.5 in the model. This compares to 15.9, 19.1, and 25.0 and in the 2007 data. In 2007 (period 32 of the transition), the

---

[24]Since we do not attempt to explain the cross-sectional (co-)movement of wages and house prices at business cycle frequencies, we do not try to match the entire time series of the average or CV of wages. In reality, other factors such as unemployment or interest rates, whose dynamics our model abstracts from, undoubtedly affect house prices.

cross-sectional skewness of real wages is 0.86 in the data and 0.82 in the model. Second, we ask whether the model's wage dynamics are consistent with the data using a specification analysis. In the spirit of indirect inference, we argue that an econometrician would have a hard time telling apart the data generating process for wages in our model and in the data. We envision an econometrician who estimates an autoregressive process with fixed effects using dynamic panel data on log real wages. Han and Phillips (2006) develop a double-difference least squares (DDLS) estimator which is consistent even when wages follow a unit root and contain a deterministic time trend. This estimator is appropriate for our model because of endogenous growth in wages. For our 1975-2007 panel of 330 regions, the DDLS estimation indicates a unit root. We then simulate 250 wage panels for 330 regions and 33 years each from the model. The DDLS estimation on simulated data generates point estimates within one standard error from the point estimate in the data, so that the model also generates unit root behavior.[25]

4.2 House Prices

Our main object of interest is the post-1975 evolution of house prices. In particular, the central question of our paper is whether the modest increase in wage dispersion (pw-CV increases by 8.8 points) can generate a large increase in house price dispersion (pw-CV increases by 38.2 points)? Figure 5 shows the model's predictions for the pw-average (left panel) and pw-CV (right panel) of house prices in the initial and final steady states (dashed lines), as well as along the transition path (solid line). Both the level and the CV of house prices are predicted to continue rising towards the new steady state after 2007.

The results show that our model features enough amplification to turn the modest increase in productivity dispersion into a large increase in house price dispersion. Our benchmark calibration generates an increase in the pw-CV of house prices of 51 points from 0.022 in the initial steady-state to 0.532 thirty-two periods into the transition, see Table 2, Panel B, Row 3. In the data, the CV increases from 0.154 in 1975 to 0.536 in 2007 (Panel A, Row 3). Thus the model is able to account for the observed cross-sectional dispersion in house prices in 2007. Because its initial steady state CV of house prices is lower than the observed 1975

---

[25]The DDLS procedure estimates $\hat{\theta}$ which relates to the persistence parameter estimate $\hat{\rho}$ of log wages via $\hat{\rho} = 0.5 \left(2 + \hat{\theta} - (\hat{\theta}^2 - 8\hat{\theta})^{.5} \right)$ if $\hat{\theta} \in (-1, 0]$ and $\hat{\rho} = 1$ if $\hat{\theta} \geq 0$. The procedure calls to truncate θ at zero if the estimate $\hat{\theta} > 0$. We estimate $\hat{\theta} = 0.26$ with a standard error of .02 in the data and $\hat{\theta} = 0.24$ with a standard error of 0.03 in the model (averaged across 250 simulations). The null hypothesis that the point estimates for $\hat{\theta}$ in model and data are the same cannot be rejected at conventional significance levels. Therefore, both model and data suggest that the persistence $\rho = 1$.

value, the increase exceeds that in the data. This low initial CV is a direct consequence of the low observed 1975 price-wage sensitivity we match in our initial steady-state as part of the calibration. Section 4.5.1 contains a version of the model and Panel C contain calibration results when we match the 1975 CV of house prices instead.

Because of the convexity effect discussed in Section 2.4, the increase in dispersion also generates a moderate increase in the population-weighted average house price, from \$55,719 in the initial steady state to \$62,571 after 32 periods of transition (Row 4). In the data, average house prices increase from \$62,212 to \$87,013. Thus, while the model accounts for all of the increase in the dispersion of house prices, it can only account for one-third of the increase in house prices (11% vs 33% increase). The 20% fall in average house prices observed in 2008 will further narrow the distance between model predictions and data.

Figure 5: House Prices

The left panel plots the population-weighted cross-sectional mean of the real median home value, which re refer to as house price. It is calculated as the square foot housing price multiplied by the detrended housing size. The right panel plots the corresponding population-weighted coefficient of variation, the ratio of standard deviation to the mean. In all panels, the red dashed line denotes the initial 1975 steady state. The green dashed line denotes the final steady state which is reached well beyond 2007. The blue solid line denotes the transition from the initial steady state to the final steady state. The solid red line with dotted markers plots the data from 1975 until 2007.
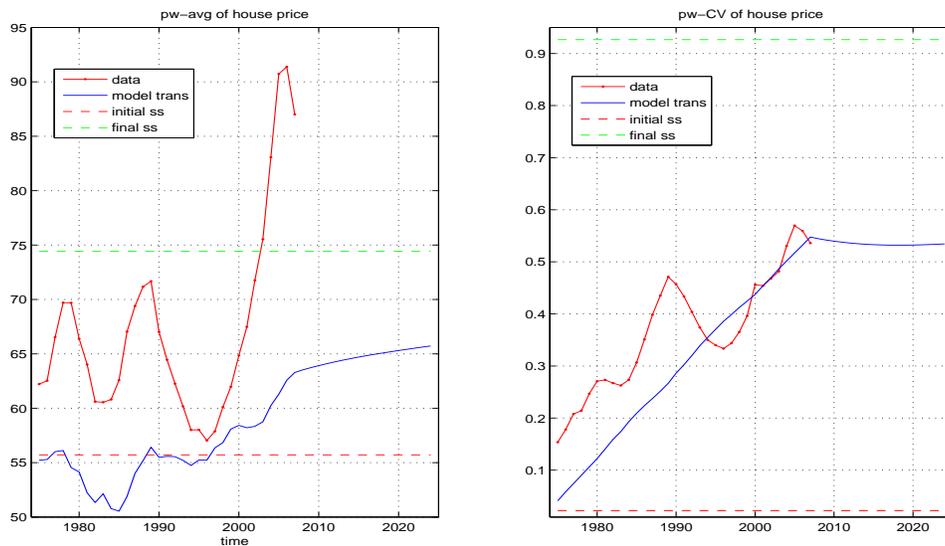
The above results are *population* statistics of house prices and wages because they are derived from and calculated for a model with a continuum of regions. While our continuum of regions model matches the key features of the observed wage distribution, we consider the additional exercise of feeding in the observed wage data from our unbalanced panel of

330 regions and computing *sample* statistics. More precisely, we evaluate the equilibrium price-wage function at the observed wage data for each region and each period. Having fed in the observed wages, we can then recompute the model's implications for house prices at the observed region-year observations. The results are almost identical to the population moments from the previous paragraph. First, the CV of house prices increases from 0.022 in the initial steady state to 0.545 thirty-two periods into the transition. Second, the average house price level increases from \$55,755 to \$62,478. The reason for this close correspondence between population and sample moments is that the population distribution for wages does a good job describing its sample counterpart; see Section 4.1.

Our benchmark model generates a strong increase in the ratio of the average house price relative to its construction cost.[26] This ratio increases from 1.75 in the initial steady state to 2.62 thirty-two periods into the transition, and continues to climb to its final steady state value of 3.65 afterwards. A similarly strong increase in the non-structure component of house prices is present in national and regional data (Davis and Heathcote (2005) and Glaeser et al. (2007)). Davis and Heathcote (2005) emphasize the value of land; similarly Glaeser, Gyourko, and Saks (2007) emphasize the value of the right to build on that land. The empirical evidence suggests that the non-structure component accounts for about half of the value of the U.S. housing stock, and it has risen much faster than the structure component since the 1970s. In our model, this non-structure component is the shadow price of an additional construction permit. As a region becomes more productive and thus more attractive to households, this shadow price increases. Hence, our model is able to account for the increase in the house price to construction cost ratio despite the fact that the quantity of construction permits stays constant.

4.3   Sensitivity of Prices to Wages

The model captures several features of the observed relationship between wages and house prices. Our calibration is designed to match the population-weighted sensitivity coefficient $b_{p0} = 0.81$ that arises from regressing the cross-section of house prices on the cross-section of wages in 1975. Second, the increase in productivity dispersion generates an increase in that sensitivity coefficient over time from 0.81 in the initial steady state to 9.82 thirty-two periods into the transition. The wage-feeding exercise described above generates an increase from

---

[26]The assumption of a centralized market for construction material implies that construction costs are the same in every island in the model. Although this implication of the model is violated in the data, Davis and Palumbo (2006) show that it is a reasonable approximation: very little of cross-sectional variation in housing prices is due to variation in construction costs.

Figure 6: Price-Wage Sensitivity

Each period we run a cross-sectional regression of real house prices on real wages. The slope coefficient b_p is computed as the population-weighted covariance of prices and wages, divided by the population-weighted variance of wages. The solid line (red, with dotted markers) denotes the time series for b_p in the data, while the solid line (blue with no markers) denotes the same slope in the model. The model's slope is computed by feeding in the observed wage data, and evaluating them at the equilibrium price function.
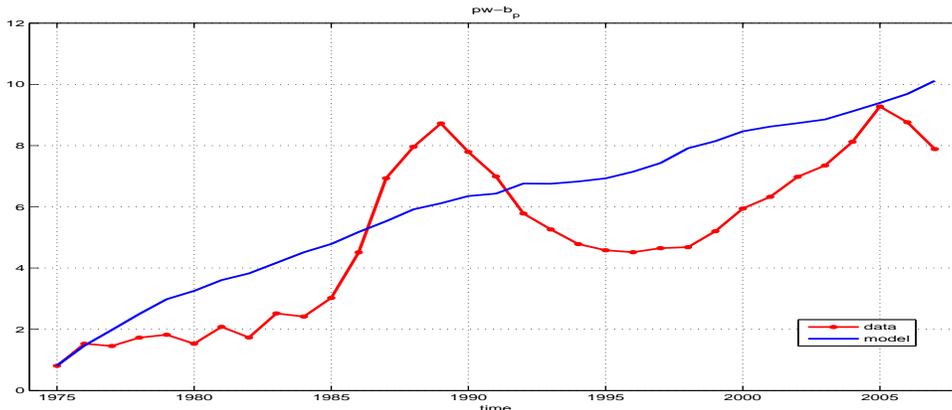
0.81 to 10.10. In the data, the increase in sensitivity is similar: from 0.81 in 1975 to 7.89 in 2007. Figure 6 plots the population-weighted sensitivity coefficient b_p in the data (solid red line with dots) and in the model (solid blue line).

The increasing sensitivity provides indirect evidence for an increasing productivity dispersion as the root cause for the increase in the wage dispersion. An alternative explanation would have been that the increase in wage dispersion was due to highly skilled workers becoming more productive. In our model, this would amount to an increase in ability dispersion. However, as explained in the second paragraph of Section 3.2.2, an increase in wage dispersion engineered through an increase in ability dispersion would result in a *reduction* in the cross-sectional sensitivity of house prices to wages.

The wage-feeding exercise suggests an important specification test for our model: Does the model predict the right amount of sensitivity of house prices to wages as measured by the R^2 statistic in equation (26)? In our benchmark model, we obtain an R^2 statistic of 31.0%, close to the value of 26.5% in the data. So, wages account for around 30% of the variation in house prices across regions in both model and data. In contrast, a model without ability dispersion would lead to model-implied prices that are extremely sensitive to wages and, therefore, vary too much in the cross-section. The R^2 in such a model is -31.5%. The slope b_p from a period by period cross-sectional regression of house prices on wages, averaged over time, is similar in our benchmark model and in data. It is 5.39 in the model and 3.22 in the

31

data. In contrast, in a model without ability dispersion, the average slope is 12.22. In short, (constant) ability dispersion is instrumental in avoiding excessive price-wage sensitivity.

Finally, our model predicts a convex relationship between houses prices and productivity. In our calibrated model, this translates into a convex relationship between house prices and wages. We find direct evidence for a similar convex relationship in the data: a cross-sectional regression of house prices on demeaned wages and squared demeaned wages generates not only significant linear but also quadratic coefficients.

4.4   Population Dynamics

Our calibration guarantees we match the 64.88% share of population that lives and works in the 20% highest-wage regions (Q5). In the benchmark model, Q5 drops from 64.88% to 46.95% in the first period of the transition. It then rises gradually to 55.18% over the next 32 periods. Although the initial drop is counterfactual (we explain why it occurs below and propose a resolution in Section 4.5.2) the 8.2% increase between 1976 and 2007 is similar (identical) to the 9.4% (8.2%) increase in the data between 1976 (1975) and 2007. The increase in population concentration after 1976 is made possible by increased construction in high-productivity regions. In contrast, low-productivity regions see no construction, a declining housing stock because of depreciation, and they lose population. This construction pattern facilitates population concentration in the highest-wage quintile. Figure 7 shows that the population further concentrates towards the highest productivity regions as the economy moves towards the final steady-state, at which point Q5 is 80.50%. Even though there are no more exogenous changes to the productivity process after 2007 and the construction threshold has reached its steady state value, the housing distribution continues to adjust towards its steady state. This continued population concentration towards high-wage regions explains why the population-weighted average and CV of wages and house prices in Figures 4 and 5, respectively, continue to increase in the model after 2007.

The increasing dispersion in productivity causes migration from previously productive to newly productive regions. Appendix F shows that the migration pattern and the magnitude of the migration rate predicted are similar in model and data. In both, there is annual out-migration at a rate of 7% per year in the 10% lowest-wage regions and in-migration at a rate of 1.5% in the 10% highest-wage regions. Hence, the assumption of frictionless reallocation does not lead to excessive migration. This is due to the high persistence of wages. Indeed, consider the equilibrium of a version of our model with constant wages: in that extreme case, nobody would find it optimal to move despite perfect mobility.

Figure 7: Population Distribution

This figure plots the population distribution by wage quintile in the benchmark model. We use the model with an increasing productivity dispersion to generate a population time-series for each MSA. We sort the MSAs into five equally sized wage bins and calculate the ratio of the number of people in each quintile to the number of people in the economy (normalized to 1). The graph shows the distribution in the initial steady state (1975), after 32 years (2007) and in the final steady state.
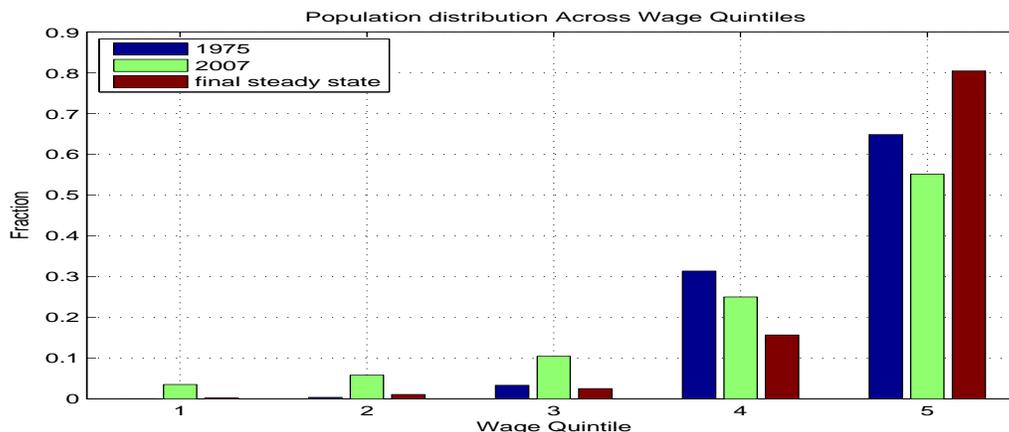
As mentioned above, one problem with the benchmark calibration is the large initial drop in Q5 between the initial steady state and the first period of the transition. This drop is an artefact of the specific mechanics driving the increase in cross-sectional wage dispersion. Indeed, in order to obtain a gradual, linear increase in the dispersion of log productivity, plotted in the left panel of Figure 4, we need a large jump in the *innovation* variance of log productivity in period 1, σ_{a1}. This high innovation variance acts as a big shock to the wage distribution between the initial steady state and period 1. Previously high-productivity, high-population regions may draw a very negative productivity innovation which puts them in a lower wage quintile and vice versa. This breaks the strong association between population and wages, resulting in the initial drop in Q5.[27] One consequence of these productivity dynamics is that the *rank correlation of wages* between adjacent years is counter-factually low. For instance, the rank correlation between the initial steady state and the first period of the transition is 68.8% compared to a rank correlation of 96.8% in the 1975-76 wage data. This rank correlation gradually increases along the transition path and exceeds 95% only twenty years into the transition. In the data it is above 95% throughout. In Section 4.5.2

---

[27]Since we assume that the innovation variance reaches its steady state value from period 33 onwards, the linear increase in the standard deviation of log productivity also requires that the innovation variance increases between period 1 and period 32, and that it overshoot its final steady state value. Precisely, the time path for the innovation standard deviation of log productivity in our benchmark calibration is as follows: 0.0004 (initial steady state), 0.0068 (period 1), 0.008 (period 2), gradually increasing to 0.0263 (period 32), then constant at 0.0103 (from period 33 until final steady state).

below, we consider an alternative calibration that increases productivity dispersion in a rank-preserving way. Its prediction for the fraction of jobs in the highest wage quintile matches the data.

4.5 Robustness

In this section, we discuss two alternative calibrations and their implications for wages and house prices.

4.5.1 Alternative 1: Calibrating to Initial CV of House Prices

First, we explore a calibration in which the initial steady state CV of house prices matches the 1975 value of 0.154 in the data. See Table 2, Panel C. This is an alternative to matching the 1975 sensitivity of prices to wages b_p in Panel B. The calibration, which continues to match the level and CV of wages in 1975 and 2007, features less ability dispersion and more regional productivity dispersion than the benchmark. In particular, the cross-sectional standard deviation of ability is 0.056 compared to 0.079 in the benchmark calibration (k_e is 25.28 compared to 17.89), the initial productivity dispersion is higher (0.039 compared to 0.006), and it rises to a higher value in the final steady state (0.135 versus 0.103). This calibration predicts a rise in the population-weighted CV of house prices of 53 points, similar to the 51 point increase in the benchmark. It generates a substantially larger increase in house price levels: a 23% increase compared to an 11% increase in the benchmark and a 33% increase in the data. The downside to matching the initial CV of house prices is that we overstate the initial sensitivity coefficient: $n_{p0} = 5.37$ versus 0.81 in the 1975 data. Both models feature a similar increase in this sensitivity coefficient. Hence, this alternative calibration introduces excess sensitivity of house prices to wages: a year-by-year cross-sectional regression of house prices on wages delivers a slope coefficient of 8.2 (averaged over time) inside the model and 3.2 in the data. The excess sensitivity problem is still less pronounced than in the model without ability dispersion though. The R^2 for house prices in equation (26) is 19% in this model compared to 26.5% in the data and 31% in the benchmark model.

4.5.2 Alternative 2: Rank-Preserving Increase in Wage Dispersion

Second, we explore a calibration in which the increase in the standard deviation of productivity dispersion is engineered in a different way. As explained above, in order to generate a linear increase in dispersion, we need that the innovation variance jumps from the initial steady state to the first period of the transition, then gradually rises until period 33, and

then jumps back down to the final steady state value. The initial jump in innovation variance introduces too much "mixing" in the cross-sectional productivity distribution and leads to a counter-factually low rank correlation between city-specific wages in adjacent periods. As an alternative, we consider a lower time path for the innovation standard deviation of log productivity: a linear rise from its initial steady state to its final steady state value over 33 periods. Because this lower time path of productivity shocks, by itself, generates a smaller increase in productivity dispersion than in our benchmark, we add a second engine of productivity dispersion. We deterministically increase the productivity of regions above the average and, vice versa, decrease the productivity of regions below the average. Importantly, unlike random productivity shocks, these deterministic changes preserve the rank of regions in the productivity distribution. Formally, the law of motion for log productivity becomes:

$$a_t = \text{avg}(a_{t-1}) + \rho_t \left(a_{t-1} - \text{avg}(a_{t-1})\right) + \sigma_{at}\varepsilon_t,$$

where $\text{avg}(a_{t-1})$ denotes the cross-sectional average productivity. Setting $\rho_t > 1$ allows us to increase dispersion in a rank-preserving way; Appendix C.1 explains the mechanics in detail. As in the benchmark model, we match the 1975 sensitivity coefficient b_{p0}. The calibration is essentially identical to the benchmark model.

The main moments of interest are discussed in Panel D of Table 2. First, the model continues to generate a large increase in the CV of house prices for a given increase in the CV of wages. The increase is 60 points compared to 51 points in the benchmark. Second, it generates a larger increase in house prices: 19% increase compared to 11% in the benchmark. Third, the sensitivity coefficient increases strongly from 0.81 in 1975 to 11.95 in 2007. Fourth, and most significantly, this calibration generates population dynamics close to those observed in the data. The fraction of people working and living in the top-20% regions in terms of wage is 64.88% in the initial steady state and rises to 74.77% after 32 periods. This is close to the 73.09% in the 2007 data. This calibration avoids the steep drop in $Q5$ in the first period of the transition, which we noted for the benchmark model. Instead, the 1976 value for $Q5$ in the model is 64.62%, close to the initial steady state value. The population then gradually relocates towards the newly productive regions. In the final steady state, $Q5$ reaches a value of 80.37%, similar to the benchmark model. The key difference with the benchmark model, therefore, is the transition path of $Q5$. Because it avoids the initial drop in population, this model generates a higher increase in population-weighted house prices and a higher population-weighted sensitivity of prices to wages. By the same token, this version of our model matches the rank correlation of wages between adjacent years. It is 99.55% on

average in the model and 99.22% on average in the data. Finally, the model generates an R^2 statistic of 25.1%, close to the 26.5% number in the data.

4.6 Increase in Regulation

In this section, we use our model to pursue an alternative explanation for the increase in the level and dispersion of house prices: housing supply regulation has become tighter over time and has gradually spread geographically (from the coastal areas inland). To keep matters as simple as possible, we hold all parameters of the benchmark calibration fixed with one exception. Instead of increasing the dispersion of productivity over time, we hold it fixed at its initial steady state level. The resulting initial steady state wage distribution is the same as in our benchmark model and matches the observed 1975 wage distribution as before. Instead, we tighten regulation. We let the number of permits at time t in a region with productivity A be

$$\Pi_t(A) = \pi_a \left(\frac{A}{A_{\min}} \right)^{\phi_t}.$$

The elasticity parameter $\phi_t = 0$ in 1975 as in our benchmark model. We let it decrease from 0 to $\phi_t = -0.5$ between periods one and thirty-two of the transition. We keep it constant at -0.5 after 2007 (periods thirty-three and later). Figure 8 illustrates how lowering the supply elasticity parameter ϕ reduces the number of permits, and more so in highly productive regions. This rotation captures the stylized fact that supply regulation gradually tightened over the last three decades, especially among more productive regions such as the coastal metropolitan areas.

Panel E of Table 2 shows the results of the regulatory tightening exercise. We find that decreasing building permits has quantitatively minor effects on average house prices and on the dispersion of house prices. While both increase, the increases are quantitatively small compared to those we found in our benchmark exercise. The same is true for the price-wage sensitivity which increases only slightly over time. The population in Q5 is also almost constant. Finally, our R^2 goodness-of-fit metric for house prices is only 5%, one-fifth of its value in the data and one-sixth of its value in the benchmark model. The intuition for the small impact of regulation is simple. While tighter regulation reduces the supply of houses in high-wage metropolitan areas, the equilibrium response of labor is to move out, thereby effectively reducing the housing demand in those same areas. The net effect is a tiny increase in price. A similar intuition is at work in the closed city model of Arnott and MacKinnon

Figure 8: Tightening Housing Supply Regulation

This figure plots the permit function $\Pi(A) = \pi_a (A/A_{\min})^\phi$. The top line denotes the situation in 1975 when $\pi_a = .30026$ and $\phi = 0$. The bottom line denotes the situation in 2007 and beyond when $\pi_a = .30026$ and $\phi = -0.5$. In the years between 1975 and 2007, ϕ decreases linearly from 0 to -0.5, so that the permit function gradually rotates from the top line to the bottom line.
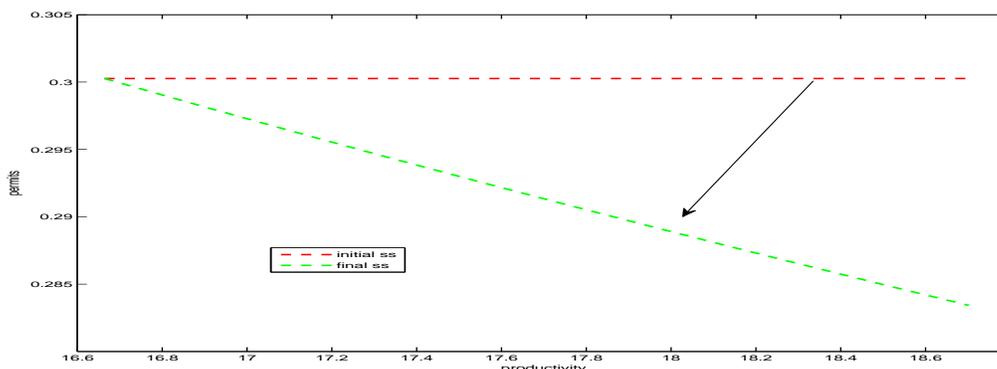
(1977) and the open city model of Aura and Davidoff (2008). We have explored alternative values for the supply elasticity parameter ϕ (-3, -1, -0.1, and even +1). The results were quantitatively similar across cases because of the endogenous response of mobility to the various regulatory changes. In the same vein, tightening regulation alongside an increase in wage dispersion delivers the same quantitative results as in our benchmark calibration with constant regulation. Impediments to labor mobility, absent from the model, may slow down the reduction in housing demand, but are unlikely to reverse it. These results suggest that an increase in wage dispersion is a necessary ingredient to generate a quantitatively meaningful increase in house price level and dispersion.

4.7 Rental Prices

In the model, the price of a house equals the present discounted value of the rents. An alternative to testing the model's implications for house prices would be to test its implications for rents. After all, the spatial equilibrium model also predicts a relationship between rents and wages. We collected nominal rent data from the Fair Market Rents database, as detailed in Appendix D.3. As we did with nominal house prices, we deflate them by the regional non-housing CPI as well as by the trend in house size. Census data suggest that the size of multi-family homes, which is likely to be rental housing, grew at the same rate as single-family housing, which is likely to be owner-occupied. The rental data are only available in 1982 and from 1984 until 2007. As the last two rows of Table 2, panel A show,

de-trended real rents seem to have fallen from $4,220 per year in 1982 to $3,420 in 2007. The CV increased from 0.153 in 1982 to 0.190 in 2007. Our model generates a moderate increase in average rents and a strong increase in the CV of rents, similar to its predictions for house prices. Hence, our model is unable to account simultaneously for the measured distribution of house prices and rents.

The likely explanation for this divergence is that the cash flows entering the present value formula for house prices are unlikely to be the rents we measure in the data. Glaeser (2007a) make several compelling empirical observations suggesting that house price and rent series can be best understood as the costs of two different types of housing, reflecting different demands on two related, but not directly comparable, markets. This market segmentation introduces severe measurement error in a hypothetical present-value relationship relating prices to rents. Our motivation to focus on the ownership market instead of the rental market is motivated by the following facts: (i) twice as many households own their home than rent, (ii) there exists no regional panel data set for the US with units that are both for rent and for sale, (iii) single-family ownership price data are of higher quality, and (iv) our key driving force is the increase in income inequality, which is most pronounced in the top half of the income distribution, a group that is more likely to be composed of homeowners. Rents in our model are then to be interpreted as the per-period user cost of owner-occupied housing. Since there are no regional data on the user cost, it seems natural to test the model using house price data instead. Like us, the bulk of the spatial location literature derives implications for (implicit) rents, but almost always tests them on owner-occupied house price data (recent examples are Gyourko, Mayer, and Sinai, 2006; Glaeser, 2007b). Needless to say, it would be interesting to consider a segmented-markets model with an explicit decision of owning versus renting. Such a model could then be used to confront price and rent data simultaneously. This extension is beyond the scope of the present paper.

5   Conclusion

Our paper provides a new general equilibrium framework for analyzing the joint dynamics of regional income, house prices, and housing quantities. It extends the Rosen-Roback spatial equilibrium model along several dimensions in order to establish closer contact with the data. We used our framework to study the quantitative effect of wage dispersion and housing supply regulation for the regional house price level and its dispersion. The model accounts for several features of the joint price-wage distribution. Faced with an increase in the productivity dispersion across metropolitan areas, households choose to reallocate from lower towards

higher-productivity metropolitan areas. This pushes up house prices in high-wage areas. The observed increase in wage dispersion is sufficient to generate the observed increase in the house price dispersion across metropolitan areas.

The same thirty years since 1975 also saw a tightening of housing supply regulation, especially in the coastal areas. One might think that the supply effect induced by this regulatory tightening could, in and of itself, account for the increase in house price level and dispersion. However, because the equilibrium response of households to move out of the more tightly regulated areas, the house price effects of tightening supply restrictions are small. So, while supply constraints are important, the increase in wage dispersion is an essential part of the explanation.

The model's prediction of an increasingly strong cross-sectional sensitivity of house prices to wages is consistent with the data. It suggests that increasing dispersion of regional productivity, as opposed to an increasing dispersion in the ability of households, underlies the changes in spatial location, wage, and house price patterns we have observed over the last three decades. In future research, we plan to further investigate the relative importance of productive and consumptive motives in location decisions of households and firms. Finally, endogenizing the housing supply constraints in our model, along the lines of Ortalo-Magné and Prat (2005), would be interesting.

Table 1: Benchmark Calibration

The six parameters with a notation "E" next to them are determined by repeatedly solving the model until six moments in the data, listed in the last column of the corresponding rows, are matched exactly. One of the six parameter has a "E^{*}" next to indicate that it governs a 2007 moment of the data, The other ones, without a star, govern features of the 1975 data. While none of the parameters solely pin down the moment mentioned on the same row, that parameter is in practice the key parameter for matching that moment. the abbreviation "pw-CV" stands for population-weighted coefficient of variation and "ss" for steady state.

	Parameter	Description	Benchmark	Source/Matches
		Preferences		
	β	time discount factor	0.9510	historical avg. 30-yr fixed-rate mortgage rate
	γ	inverse elasticity of housing demand	2	Hanushek & Quigley (1980)
E	κ	weight on housing	1.5056	average housing expenditure share
		Productivity and Ability Processes		
E	$\bar{A}$	mean productivity level	17.654	1975 pw-avg. of real wage per job
E	σ_{a0}	productivity dispersion in initial ss	0.0063	1975 pw-CV of real wage per job
E^{*}	σ_{aT*}	productivity dispersion in final ss	0.1029	2007 pw-CV of real wage per job
	λ	death rate of regions	0.01	consistent with AC of wages
	σ_{bt}	productivity dispersion at birth	see text	
E	k_e	governs ability dispersion	17.889	1975 pw- sensitivity of house prices to wages
		Technology		
	δ	depreciation rate of housing stock	0.0160	Bureau of Economic Analysis
	M_t	construction material	see text	house size from Census of Construction
E	π_a	building permits	0.3003	1975 fraction in the top-20% wage regions

Table 2: Summary Statistics Data and Model

The table reports the cross-sectional average and coefficient of variation, defined as the ratio of the standard deviation to the mean, of the real wage per job (Rows 1 and 2) and the real median home value (Rows 3 and 4) across US metropolitan areas. Row 5 reports the sensitivity of real house prices to real wages, b_p, measured as the time-series average of the slope coefficient of a cross-sectional linear regression of real house prices on real wages. Row 6 reports the fraction of jobs in the highest wage quintile, $Q5$. Rows 7 and 8 report the cross-sectional mean and CV of the real rent. Row 9 reports the R^2 statistics of equation 26. In the data panel (A), this is simply the R^2 of a regression of observed prices on wages, measuring the fraction of observed house prices variation accounted for by wages in the data. In the model panel (B to E), the R^2 measures the empirical variation that can be accounted for by model-generated prices, i.e. the prices obtained by feeding in the model the wages we observed in the data. The cross-sectional mean and CV moments are population-weighted, as well as the sensitivity coefficient b_p. Panel A is for the data. Real wages, prices, and rents are obtained by dividing the nominal amounts by the regional CPI ex-shelter. The means represent thousands of 1983 dollars; the coefficient of variation is unit-free. The sample is the unbalanced panel of 330 metropolitan statistical areas for which we have house price data. Population weights are computed as the number of jobs in that region relative to the full sample. House prices and wages are detrended as explained in the main text. Panel B is for our benchmark model. The first column (1975) refers to the initial steady state, the second column (2007) to period 32 of the transition path towards the new steady-state. Panel C is for an alternative calibration that targets the 1975 CV of house prices instead of the 1975 sensitivity coefficient b_p. Panel D is for an alternative calibration that engineers the increase in wage dispersion in a rank-preserving way. Panel E investigates a tightening of housing supply regulation. The numbers denoted by a star are values in 1982 instead of 1975; 1982 is the first observation on rental data. The corresponding numbers in the models denote period 8 of the transition instead of the initial steady state.

	A: Data		B: Benchmark		C: Altern. 1		D: Altern. 2		E: Regulation	
	1975	2007	1975	2007	1975	2007	1975	2007	1975	2007
Mean wage	17.78	19.49	17.78	19.49	17.78	19.49	17.78	19.49	17.78	17.78
CV wage	0.084	0.172	0.084	0.172	0.084	0.172	0.084	0.172	0.084	0.084
Mean hp	62.21	87.01	55.72	62.57	55.67	70.15	57.65	69.77	55.72	55.76
CV hp	0.154	0.536	0.022	0.532	0.154	0.671	0.021	0.620	0.022	0.025
b_p	0.81	7.89	0.81	9.82	5.61	13.60	0.81	11.95	0.81	0.91
Q_5	64.88	73.09	64.88	55.18	64.88	61.73	64.88	74.77	64.88	64.92
Mean rent	4.22*	3.42	4.51*	5.00	4.60*	5.51	4.63*	5.03	4.45*	4.09
CV rent	0.153*	0.190	0.141*	0.515	0.266*	0.651	0.106*	0.602	0.021*	0.025
R^2	26.46		30.96		19.06		25.08		5.60	

References

Fernando Alvarez and Marcelo Veracierto. Labor market policies in an equilibrium search model. *NBER Macroeconomic Annual*, pages 265–303, 1999.

Fernando Alvarez and Marcelo Veracierto. Fixed-term employment contracts in an equilibrium search model. Working Paper, University of Chicago, 2006.

Richard J. Arnott and James G. MacKinnon. Measuring the cost of height restrictions with a general equilibrium model. *Regional Science and Urban Economics*, pages 359–375, 1977.

Saku Aura and Thomas Davidoff. Supply constraints and housing prices. *Economic Letters*, 99: 275–277, 2008.

Marigee Bacolod, Bernardo S. Blum, and William C. Strange. Skills in the city. Working Paper, University of Toronto, June 2007.

Christopher R. Berry and Edward L. Glaeser. The divergence of human capital levels across cities. *Papers in Regional Science*, 84(3):407–444, 2005.

Sean Campbell, Morris Davis, Joshua Gallin, and Robert Martin. What moves housing markets: A trend and variance decomposition of the rent-price ratio. Working Paper, Federal Reserve Board of Governors, 2006.

Karl E. Case and Robert J. Shiller. Prices of single-family homes since 1970: New indexes for four cities. *New England Economic Review*, pages 46–56, September/October 1987.

Daniele Coen-Pirani. Understanding gross worker flows across u.s. states. Working Paper, Carnegie Mellon University, August 2006.

Tim Cogley. Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey. *Journal of Monetary Economics*, 49:309–334, 2002.

George M. Constantinides and Darrell Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy*, 104:219–240., 1996.

Arnaud Costinot and Jonathan Vogel. Matching and inequality in the world economy. Working Paper, MIT and Columbia University, 2008.

Morris Davis and Jonathan Heathcote. The price and quantity of residential land in the united states. Working Paper, CEPR, October 2005.

Morris Davis and Michael Palumbo. The price of residential land in large u.s. cities. Working Paper, Federal Reserve Board, March 2006.

Jan Eeckhout. Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451, 2004.

Paul L. Fackler and Mario J. Miranda. *Applied Computational Economics and Finance*. The MIT Press, Cambridge, MA, first edition, 2002.

Xavier Gabaix and Augustin Landier. Why has ceo pay increased so much? *Quarterly Journal of Economics*, 123:49–100, 2008.

Edward L. Glaeser and Joseph Gyourko. The impact of zoning on housing affordability. *Economic Policy Review*, 9(2):21–39, 2003.

Edward L. Glaeser and Joseph Gyourko. Urban decline and durable housing. *Journal of Political Economy*, 113(2):345–375, 2005.

Edward L. Glaeser, José Scheinkman, and Andrei Schleifer. Growth in cities. *Journal of Political Economy*, 100(6):331–370, 1992.

Edward L. Glaeser, José Scheinkman, and Andrei Schleifer. Economic growth in a cross-section of cities. *Journal of Monetary Economics*, 36:117–143, 1995.

Edward L. Glaeser, Joseph Gyourko, and Raven E. Saks. Why is manhattan so expensive? *Journal of Law and Economics*, 48(2):331–370, 2005.

Edward L. Glaeser, Joseph Gyourko, and Raven E. Saks. Why have house prices gone up? *American Economic Review Papers and Proceedings*, 95(2):329–333, 2007.

Joseph Glaeser, Edward L. Gyourko. Arbitrage in the housing market. Working Paper, Harvard University and Wharton School of Business, December 2007a.

Joseph Glaeser, Edward L. Gyourko. Housing dynamics. Working Paper, Harvard University and Wharton School of Business, May 2007b.

Clive W. J. Granger. Investing causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(4):424–438, 1969.

Joseph Gyourko, Christopher Mayer, and Todd Sinai. Superstar cities. Working Paper, NBER, July 2006.

Chirok Han and Peter C. B. Phillips. Gmm estimation for dynamic panels with fixed effects and strong instruments at unity. Working Paper, Cowles Foundation, 2006.

Eric A. Hanushek and John M. Quigley. What is the price elasticity of housing demand? *The Review of Economics and Statistics*, 62(3):449–454, 1980.

Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Insurance and opportunities: The welfare implications of rising wage dispersion. Working Paper, CEPR No. 5200, August 2005.

Jonathan Heathcote, Kjetil Storesletten, and Gianluca Violante. The macroeconomic implications of rising wage inequality in the u.s. Working Paper, Federal Reserve Bank of Minneapolis, University of Oslo, and NYU, 2008.

Charles Himmelberg, Christopher Mayer, and Todd Sinai. Assessing high house prices: Bubbles, fundamentals, and misperceptions. Working Paper, NBER No. 11643, September 2005.

Andreas Hornstein, Per Krusell, and Giovanni L. Violante. The effects of technical change on labor market inequalities. In Philippe Aghion and Steven Durlauf, editors, *Handbook of Economic Growth*. Elsevier Science, Amsterdam, 2004.

Matteo Iacoviello. House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review*, 95(3), June 2005.

Ken-Ichi Inada. On a two-sector model of economic growth: Comments and generalization. *Review of Economic Studies*, 30(2):119–127, 1963.

Dirk Krueger and Jesús Fernández-Villaverde. Consumption over the life cycle: Facts from the consumer expenditure survey. *Review of Economics and Statistics*, 89:552–565, 2006.

Dirk Krueger and Fabrizio Perri. Does income inequality lead to consumption inequality? evidence and theory. *The Review of Economic Studies*, Forthcoming, 2006.

Robert E. Lucas and Edward C. Prescott. Equilibrium search and unemployment. *Journal of Economic Theory*, 7(2):188–209, 1974.

Hanno Lustig. The wealth distribution and aggregate risk. Mimeo, University of California at Los Angeles, 2003.

Hanno Lustig and Stijn Van Nieuwerburgh. Housing collateral, consumption insurance and risk premia: An empirical perspective. *Journal of Finance*, 60(3):1167–1219, 2005.

Hanno Lustig and Stijn Van Nieuwerburgh. How much does household collateral constrain regional risk sharing. Working Paper NYU, Stern and UCLA, February 2006.

Hanno Lustig and Stijn Van Nieuwerburgh. Can housing collateral explain long-run swings in asset returns? Working Paper, NYU Stern and UCLA, June 2007.

Stephen Malpezzi. Housing prices, externalities, and regulation in u.s. metropolitan areas. *Journal of Housing Research*, 7(2):209–241, 1996.

Makoto Nakajima. Rising earnings instability, portfolio choice, and housing prices. Working Paper, University of Illinois, Urbana-Champaign, 2005.

Francois Ortalo-Magné and Andrea Prat. The political economy of housing supply. Working Paper, University of Wisconsin and LSE, December 2005.

Francois Ortalo-Magné and Andrea Prat. Spatial asset pricing: A first step. Working Paper, University of Wisconsin and LSE, 2008.

Monika Piazzesi, Martin Schneider, and Selale Tuzel. Housing, consumption and asset pricing. *Journal of Financial Economics*, Forthcoming, 2006.

John Quigley and Steven Raphael. Regulation and the high cost of housing in california. *American Economic Review Papers and Proceedings*, 95(2):323–328, 2005.

John Quigley and Larry Rosenthal. The effects of land use regulation on the price of housing: What do we know? what can we learn? *Cityscape*, 8(1):69–138, 2005.

William J. Reed. The pareto, zipf and other power law. *Economics Letters*, 74:15–19, 2001.

William J. Reed. The pareto law of incomes - an explanation and an extension. *Physica A*, 319: 469–486, 2003.

Jennifer Roback. Wages, rent, and the quality of life. *Journal of Political Economy*, 90(2):191–229, 1982.

Sherwin Rosen. Wage-based indexes of urban quality of life. In Mieszkowski and Straszheim, editors, *Current Issues in Urban Economics*. Johns Hopkins University Press, Baltimore, 1979.

Sherwin Rosen and Robert Topel. Housing investment in the united states. *Journal of Political Economy*, 96(4):718–740, 1988.

Raven E. Saks. Job creation and housing construction: Constraints on metropolitan area employment growth. Federal Reserve Board Finance and Economics Discussion Series No. 2005-49, September 2005.

Michael Sattinger. Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31:831–880, 1993.

Robert Shimer. Mismatch. Working Paper University of Chicago, November 2005.

Matthew Spiegel. Housing returns and construction cycles. *Real Estate Economics*, 29(4):521–551, 2001.

Kjetil Storesletten, Chris Telmer, and Amir Yaron. Consumption and risk sharing over the life cycle. *The Journal of Monetary Economics*, 51:609–633, 2004.

Kjetil Storesletten, Chris Telmer, and Amir Yaron. Asset pricing with idiosyncratic risk and overlapping generations. Working Paper, University of Oslo, February 2006.

George Tauchen and Robert Hussey. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, 59(2):371–396, 1991.

Marko Tervio. The dierence that ceos make: An assignment model approach. *American Economic Review*, 2008. Forthcoming.

Menahem E. Yaari. Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, 32:137–150, 1965.

A Proofs

A.1 Preliminary Results

The following Lemma compiles technical results which are used in the following subsection.

Lemma 6. *Consider some strictly increasing strictly concave, and twice continuously differentiable function $v : (0, \infty) \to \mathbb{R}$. Suppose that $v(h)$ goes to minus infinity as h goes to zero, and that $v(h)$ goes to zero as h goes to infinity. Then*

1. *The derivative $v'(h)$ goes to infinity as h goes to zero, and goes to zero as h goes to infinity.*

2. *The function $hv'(h)$ goes to zero as h goes to infinity.*

3. *The function $w(h) \equiv hv'(h) - v(h)$ is continuous and strictly decreasing, goes to zero as h goes to infinity, and goes to infinity as h goes to zero.*

4. *The function $\Phi(x) = 1/w^{-1}(x)$ is continuous and strictly increasing. It can be extended by continuity at zero with $\Phi(0) = 0$. It goes to infinity as x goes to infinity.*

5. *The function $R(x) \equiv v' \circ w^{-1}(h)$ is increasing, convex, continuous, goes to zero as x goes to zero and goes to infinity as x goes to infinity.*

6. *Consider any density $g(A)$ such that, for all $x \in \mathbb{R}$,*

$$G(x) = \int_{A_{\min}}^{A_{\max}} \Phi\left(\max\{A - x, 0\}\right) g(A) \, dA < \infty.$$

Then, the function $G(x)$ is continuous.

Proof.

1. For any $h_1 > h_2$, concavity implies that $v'(h_2)(h_1 - h_2) \geq v(h_1) - v(h_2)$. Therefore, $v'(h_2)h_1 \geq v'(h_2)h_2 + v(h_1) - v(h_2) \geq v(h_1) - v(h_2)$. Letting h_2 go to zero in the inequality implies that $v'(h_2)$ goes to infinity as h_2 goes to zero. Second, since $v'(h)$ is positive and decreasing, it has some positive limit $\underline{v}'$ as h goes to infinity. Since $v(h)$ is concave, then for all $h_1 > h_2$, $0 \geq v(h_1) \geq v(h_2) + v'(h_1)(h_1 - h_2) \geq v(h_2) + \underline{v}'(h_1 - h_2)$. Letting h_1 go to infinity shows that $\underline{v}' = 0$. Therefore, $v'(h)$ goes to zero as h goes to infinity.

2. Rearranging the previous inequality implies that

$$v(h_1) + h_2 v'(h_1) - v(h_2) \geq h_1 v'(h_1) \geq 0.$$

Letting h_1 go to infinity shows that $-v(h_2) \geq \limsup_{h \to \infty} hv'(h) \geq 0$ for all h_2. Letting h_2 go to infinity shows that $hv'(h)$ also goes to zero as h goes to infinity.

3. Consider the function $w(h) \equiv hv'(h) - v(h)$. The above results show that $w(h)$ goes to zero as h goes to infinity. Because $w'(h) = hv''(h) < 0$, it follows that $w(h) \geq 0$. Lastly, since $w(h) \geq -v(h)$, letting h go to zero shows that $w(h)$ goes to infinity as h goes to zero.

4. The previous paragraph implies that the function $\Phi(x) = 1/w^{-1}(x)$ is well defined. It is continuous, increasing, goes to zero as x goes to zero, and to infinity as x goes to infinity. Lastly, consider the function $R(x)$ is increasing because both $v'(x)$ and $w^{-1}(x)$ are decreasing. Point 1 and 3 of the Lemma imply that it is goes to zero as x goes to zero, and to infinity as x goes to infinity.

5. In order to prove that $R(x)$ is convex, note that

$$
\begin{aligned}
R'(x) &= \frac{v'' \circ w^{-1}(x)}{w' \circ w^{-1}(x)} \\
&= \frac{v'' \circ w^{-1}(x)}{w^{-1}(x) \times v'' \circ w^{-1}(x)} \\
&= \frac{1}{w^{-1}(x)},
\end{aligned}
$$

where the second line follows from the fact that $w'(h) = hv''(h)$. Since $w^{-1}(x)$ is decreasing, it follows that $R'(x)$ is increasing, which establishes convexity.

6. Pick any $x \in \mathbb{R}$ and some $\eta > 0$. Then that, for all $y \in [x - \eta, x + \eta]$

$$
\begin{aligned}
|G(x) - G(y)| &\leq \int_{A_{\min}}^{\alpha} \big|\Phi\left(\max\{A - x, 0\}\right) - \Phi\left(\max\{A - y, 0\}\right)\big| g(A)\, dA \\
&\quad \int_{\alpha}^{A_{\max}} \big|\Phi\left(\max\{A - x, 0\}\right) - \Phi\left(\max\{A - y, 0\}\right)\big| g(A)\, dA \\
&\leq \int_{A_{\min}}^{\alpha} \big|\Phi\left(\max\{A - x, 0\}\right) - \Phi\left(\max\{A - y, 0\}\right)\big| g(A)\, dA \\
&\quad 2 \int_{\alpha}^{A_{\max}} \Phi\left(\max\{A - x + \eta, 0\}\right) g(A)\, dA
\end{aligned}
$$

where the second inequality follows because $\Phi(x)$ is decreasing. Now, because $G(x - \eta) < \infty$, it follows that for all $\varepsilon > 0$ there exists some $\alpha > 0$ such that the second integral on the right-hand side is less than $\varepsilon/2$. Since the function $\Phi\left(\max\{z, 0\}\right)$ is uniformly continuous over the compact $[0, \alpha - x + \eta]$, there exists some $\eta' < \eta$, such that $|x - y| < \eta'$ implies that $|\Phi\left(\max\{A - x, 0\}\right) - \Phi\left(\max\{A - y, 0\}\right)| < \varepsilon/2$. Plugging this back into the first integral on the right-hand side shows that $|x - y| < \eta'$ implies that $|G(x) - G(y)| < \varepsilon$.

$\square$

A.2 Proofs of the results in the text

A.2.1 Proof of Proposition 2

We proceed in five steps. First, we let $\varepsilon_{it} \in [\underline{e}, \overline{e}]$ be the set of households who find it optimal to locate in an island with current productivity A_{it}. Then, we have

Result 1. *If, for some j, $F(\varepsilon_{jt}) = 0$, then $\varepsilon_{it} = \emptyset$ for all $i < j$.*

To prove the first statement, consider an island A_{it} with a measure $F(\varepsilon_{it}) = 0$ of households. Then the rent must be $R_{it} = 0$. For all $j < i$, we then have that $R_{jt} \geq 0 = R_{it}$ and $A_{jt} < A_{it}$. Thus, island j is strictly

less attractive to household than island i and, consequently, is not populated, i.e. $\varepsilon_{jt} = \emptyset$. We then show that:

Result 2. *Consider $i \neq j$. Then $\varepsilon_{it} \cap \varepsilon_{jt}$ is either empty or is a singleton.*

Indeed, household $e \in \varepsilon_{it} \cap \varepsilon_{jt}$ is indifferent between island i and j if and only if $eA_{it} + v(h_{it}) - R_{it}h_{it} = eA_{jt} + v(h_{jt}) - R_{jt}h_{jt}$. Since there is a unique e solving this equation, the result follows. We then have:

Result 3. *If, for some i, $F(\varepsilon_{it}) > 0$, then $F(\varepsilon_{jt}) > 0$ for all $j > i$.*

Note that, because of result 2, all households e in the set ε_{it} live in islands of type i. Thus, islands of type i must be inhabited by a positive measure of households, so housing demand is non zero and $R_{it} > 0$. Now, if an island $j > i$ were populated by a measure zero of households, then $R_{jt} = 0$ and households $e \in \varepsilon_{it}$ would strictly prefer it over island i because of its lower rent and strictly higher wage, which would contradict optimality. Let p be the smallest integer such that $F(\varepsilon_{it}) > 0$ and let e_{it} be the infimum of ε_{it} for all $i \geq p$. This infimum is well defined since, by the previous result then ε_{it} is not empty for all $i \geq p$. For $i = N + 1$, we define $e_{N+1t} \equiv \overline{e}$. Next, we show:

Result 4. *Suppose $A_{it} < A_{jt}$. If household $e \in [\underline{e}, \overline{e}]$ weakly prefers island j to i, then all households $e' > e$ strictly prefer j to i.*

This results follows because the household's objective function is super-modular. Since:

$$eA_{jt} + v(h_{jt}) - R_{jt}h_{jt} > eA_{it} + v(h_{it}) - R_{it}h_{it} \geq 0 \Leftrightarrow e(A_{jt} - A_{it}) \geq v(h_{it}) - v(h_{jt}) - R_{it}h_{it} + R_{jt}h_{jt}, \quad (28)$$

and $A_{jt} > A_{it}$, then the inequality is strict for all $e' > e$. Equipped with this result, we obtain:

Result 5. *The lowest ability cutoff is $e_{pt} = \underline{e}$ and, for all $i \geq p$, $\varepsilon_{it} = [e_{it}, e_{i+1t}]$.*

First note that, the sets ε_{it} are increasing. Otherwise, suppose we had $j > i$, $e \in \varepsilon_{it}$, $e' \in \varepsilon_{jt}$ and $e' < e$. Then e' weakly prefers j to i and so, by result 4, e would strictly prefer j to i, a contradiction. It then follows that the sequence e_{it} is weakly increasing. Now all $e \in (e_{it}, e_{i+1t})$ must belong to ε_{it}. Otherwise, if some $e \in (e_{it}, e_{i+1t})$ belonged to ε_{jt} for some $j < i$, then we could find some $e' \in \varepsilon_{it} < e$. By result 4, e would strictly prefer i to j, which is a contradiction. Also, if e belonged to ε_{jt} for $j > i$, then $e_{jt} \geq e_{i+1t} > e$ which contradicts the fact that e_{jt} is the infimum of ε_{jt}. By a similar line of argument, we can show that any $e \notin (e_{it}, e_{i+1t})$ cannot belong to ε_{it}. This shows that $(e_{it}, e_{i+1t}) \subseteq \varepsilon_{it} \subseteq [e_{it}, e_{i+1t}]$. It then follows that $e_{it} < e_{i+1t}$ because otherwise ε_{it} would have measure zero. Since, in an equilibrium, all household must live in some location, we must also have that $\varepsilon_{pt} = \underline{e}$. Lastly, letting $e \to e_{i+1t}$ from the left and from the right in equation (28) for i and $i + 1$, we obtain that household e_{i+1t} is indifferent between island i and island $i + 1$, so both e_{it} and e_{i+1t} belong to the set ε_{it}.

A.2.2 Proof of Proposition 3

In this proof we suppress time subscripts to simplify notation. We proceed in two steps. First, we show that the system of difference equations (17)and (18) has a unique solution with $e_{1t} = \underline{e}$ and $e_{N+1t} = \overline{e}$. Then, we show that the unique solution of the difference equation is indeed the basis of an equilibrium assignment of households to islands.

Step 1. To solve the system of difference equations, it is useful to consider the following change of variables:

$$e = F^{-1}\left(\min\{1 - q, 1\}\right) \equiv \psi(q), \tag{29}$$

where $q_i \in (-\infty, 1]$. The variable q is a "generalized percentile," such that when $q \leq 0$, then $e = \bar{e}$ and when $q = 1$, $e = \underline{e}$. In terms of q_i, the system of difference equation becomes:

$$q_i - q_{i+1} = \mu_i H_i \Phi\left(\max\{\psi(q_i)A_i - U_i, 0\}\right) \tag{30}$$
$$U_{i+1} = \left(\psi(q_{i+1}) - \psi(q_i)\right)A_i + U_i \tag{31}$$

One first sees that a pair of sequences $\{U_i\}_{i=1}^N$ and $\{e_i\}_{i=1}^{N+1}$ solves (17)-(18) with $e_1 = \underline{e}$ and $e_{N+1} = \bar{e}$ if and only if the pair of sequences $\{U_i\}_{i=1}^N$, $q_i = \{1 - F(e_i)\}_{i=1}^{N+1}$ solves (30) and (31) with $q_1 = 1$ and $q_{N+1} = 0$.

Now, given an initial condition U_1 and $q_1 = 1$, we show that $U_i - \psi(q_i)A_i$ is strictly increasing in the initial condition U_1, q_i is increasing in the initial condition U_1, and strictly increasing if $q_i < 1$. We proceed by induction. Clearly, the property is true for $i = 1$. Now suppose it is true for all $j \leq i$. We have that

$$q_{i+1} = q_i - \mu_i H_i \Phi\left(\max\{\psi(q_i)A_i - U_i, 0\}\right).$$

Then there are three cases to consider. If $q_i < 1$, then q_{i+1} is strictly increasing in U_1 given that q_i and $U_i - \psi(q_i)A_i$ are strictly increasing in U_1, and $\Phi(\cdot)$ is an increasing function. The second case is if $q_i = 1$ and $q_{i+1} < 1$, then $\psi(q_i)A_i - U_i > 0$ and so q_{i+1} is strictly increasing in U_1. Lastly, if $q_i = 1$ and $q_{i+1} = 1$, then $\psi(q_i)A_i - U_i \leq 0$, and we obtain that q_{i+1} is increasing in U_1. Now turn to:

$$U_{i+1} - \psi(q_{i+1})A_{i+1} = U_i - \psi(q_i)A_i + (A_i - A_{i+1})\psi(q_{i+1}). \tag{32}$$

The result follows by the induction hypothesis and because $A_i - A_{i+1} < 0$ and $\psi(q)$ is decreasing.

We thus have that q_{N+1} is increasing, and strictly increasing if $q_{N+1} < 1$. Moreover, for all $U_1 > \bar{e}A_N$, $U_i \geq U_1 > \bar{e}A_N \geq e_i A_i$ for all i, so $q_i = 1$ for all i. On the other hand

$$q_{i+1} - q_i < - \min_{j \in \{1,\dots,N\}}\{\mu_j H_j\} \times \Phi\left(\max\{\psi(q_1)A_1 - U_1, 0\}\right),$$

because, from equation (32), the sequence $\psi(q_i)A_i - U_i$ is increasing in i. Since $\psi(q_1) = \underline{e}$, it follows that, as U_1 goes to minus infinity, $q_{i+1} - q_i$ goes to minus infinity, and so does q_{N+1} Taken together, these properties show that there exists a unique U_1 such that $q_{N+1} = 0$.

Step 2. We now verify that the solution we constructed is indeed the basis of an equilibrium. That is, given $\{U_i\}_{i=1}^N$ and $\{e_i\}_{i=1}^{N+1}$ we let:

$$p = \min\{i \in \{1, \dots, N\} : e_i A_i - U_i > 0\}$$
$$n_i = H_i \Phi\left(\max\{e_i A_i - U_i, 0\}\right)$$
$$h_i = H_i / n_i$$
$$R_i = v'(h_i).$$

The labor market clears by construction of n_i and the housing market clears by construction of h_i. The rent R_i makes it optimal for a household in island i to consume h_i. All we need to verify is that, for all $i \geq p$, households $e \in [e_i, e_{i+1}]$ find it indeed optimal to live in island i. We start by noting that:

$$
\begin{aligned}
U_i(e) &\equiv eA_i + v(h_i) - R_i h_i = eA_i + v(h_i) - v'(h_i)h_i = eA_i - w(h_i) \\
&= eA_i - \max\{e_i A_i - U_i, 0\} = (e - e_i)A_i + \min\{e_i A_i, U_i\}.
\end{aligned}
$$

For $\underline{e} = e_p$, we have that, for $i < p$, $U_i(\underline{e}) = 0 + \underline{e}A_i$, since $U_i = U_p \geq \underline{e}A_p > \underline{e}A_i$. For $i = p$, then $U_p(\underline{e}) \geq \underline{e}A_p$ by definition of p. Thus, $\underline{e}$ prefers island p to any island $i < p$, and so does any household $e \geq \underline{e}$ because location choices are monotonic in ability. Thus, islands $i < p$ are not populated. Now, for any $i \geq p$, we have that $U_i \leq e_i A_i$ because of (17) and the fact that $e_{i+1} > e_i$. Thus:

$$
\begin{aligned}
U_{i+1}(e) &= (e - e_{i+1})A_{i+1} + U_{i+1} = (e - e_{i+1})A_{i+1} + U_i + (e_{i+1} - e_i)A_i \\
&= U_i(e) + (e - e_{i+1})(A_{i+1} - A_i) = U_p(e) + \sum_{j=p}^{i}(e - e_{j+1})(A_{j+1} - A_j),
\end{aligned}
$$

where the second equality follows by equation (18), the third equality by definition of $U_i(e)$, and the last equality by iterating backward until $j = p$. The terms of the sum are positive if and only if $e \geq e_{j+1}$. It thus follows that a household finds it optimal to locate in the largest j such that $e \geq e_j$. In other words, households in $[e_i, e_{i+1}]$ find it optimal to locate in islands of type i.

A.2.3 Proof of Proposition 5

We start from the definition

$$
\begin{aligned}
U_{it} &= e_{it}A_{it} + \max_{h \geq 0}\{v(h) - R_i h\} \\
&\equiv e_{it}A_{it} - \theta(R_i)
\end{aligned}
\tag{33}
$$

where $\theta(R) \equiv \min_{h \geq 0}\{Rh - v(h)\}$. Note that, for each h, the function $h \mapsto Rh - v(h)$ is positive, increasing, and affine. Being the upper envelope of such a family of functions, $\theta(R)$ is increasing, and concave. We then write:

$$
\frac{R_{i+1t} - R_{it}}{A_{i+1t} - A_{it}} = \frac{R_{i+1t} - R_{it}}{\theta(R_{i+1t}) - \theta(R_{it})} \times \frac{\theta(R_{i+1t}) - \theta(R_{it})}{A_{i+1t} - A_{it}}.
\tag{34}
$$

The first term is increasing in i because, as argued above, the function $\theta(R)$ is increasing and concave. The second terms is also positive and increasing. Indeed, using (33), we have that $\theta(R_{it}) = U_{it} - e_{it}A_{it}$. Therefore:

$$
\begin{aligned}
\theta(R_{i+1t}) - \theta(R_{it}) &= e_{i+1t}A_{i+1t} - U_{i+1t} + U_{it} - e_{it}A_{it} \\
&= e_{i+1t}A_{i+1t} - U_{it} - (e_{i+1t} - e_{it})A_{it} + U_{it} - e_{it}A_{it} \\
&= e_{i+1t}(A_{i+1t} - A_{it}),
\end{aligned}
$$

where the second line follows from the indifference equation (18). Thus, the second term of (34) is simply equal to e_{i+1t} which is increasing because of assortative matching of ability with productivity.

B Supplementary Material

B.1 The sensitivity of house prices to wages: a simple example

Our model is about the impact of *productivity differentials* on house price differentials. A commonly held view is that the *wage differentials* we observe in the data are much larger than the true productivity differentials, because high-ability households tend to self select in high-productivity regions. For example, high-ability lawyers tend to take jobs in New York City rather than in Fargo. This creates a measurement problem because the component of ability that matters for self-selection may be unobserved, i.e. not captured by standard Mincerian human capital proxies.

A bit more formally, consider the following example. Suppose that two regions with different productivity $A > A'$ attract households with different abilities $e > e'$, measured in effective units of labor. Suppose that wages in the two regions are given by $W = e \times A$ and $W' = e' \times A'$ respectively. For this assignment of ability to be individually optimal, the *marginal* household in the high-productivity region must be indifferent between staying or moving to the low-productivity region. That is:

$$eA + \bar{u} - R = eA' + \bar{u} - R' \Rightarrow \Delta R = e\Delta A \Rightarrow \Delta P = \frac{e\Delta A}{1 - \beta(1-\delta)}. \tag{35}$$

What this means is that the price differential compensates for a *constant ability* wage differential, $e\Delta A$. The compensation is thus smaller that the observed wage differential which reflects both ability and productivity differentials:

$$\Delta W \simeq \Delta(e \times A) \simeq e\Delta A + \Delta eA > e\Delta A.$$

Plugging the approximation for ΔW back into (35), we find:

$$\Delta P = \frac{e\Delta A}{e\Delta A + \Delta eA} \times \frac{\Delta W}{1 - \beta(1-\delta)} \Rightarrow \Delta P = \underbrace{\frac{\Delta A/A}{\Delta A/A + \Delta e/e}}_{\text{less than 1}} \times \frac{\Delta W}{1 - \beta(1-\delta)}. \tag{36}$$

The coefficient in front of the present-value discount factor is the "correction" that must be applied to the observed wage differential, ΔW, in order to calculate the constant-ability wage differential that ultimately determines the equilibrium house-price differential. When there are no ability differentials, $\Delta e = 0$, the correction coefficient is equal to one, and the sensitivity of house prices to wages is equal to $1/(1 - \beta(1-\delta))$. When there are ability differentials, $\Delta e > 0$, the correction coefficient is less than one, and the sensitivity of prices to wages is lower.

B.2 Households' inter-temporal problems

In this appendix we formulate and solve the inter-temporal problem of a household. The goals of the analysis are twofold. First, to show that the household's inter-temporal problem reduces to the sequence of static location problems described in the text. Second, to make clear that landlords are not "absentees landlords." In our model, the profits of the real estate sector are rebated to the households.

The household's problem. Perhaps the easiest way to set-up the household's inter-temporal problem is with time-zero markets where, at time zero, a household of ability $e \in [\underline{e}, \overline{e}]$ chooses its entire location plan and consumption plan. The household of type e's problem is then to maximize:

$$\sum_{t=1}^{\infty} \beta^{t-1} [c_t + v(h_t)],$$

with respect to some positive consumption plan $\{c_t, h_t\}_{t=1}^{\infty}$ and some location plan $\{A_t\}_{t=1}^{\infty}$, specifying the productivity of the islands to be visited at each point in time, and subject to the inter-temporal budget constraint:

$$\sum_{t=1}^{\infty} \beta^{t-1} [c_t + R_t(A_t)h_t - eA_t] \le W_0(e),$$

where it is anticipated that the price of time t consumption at time 0 is equal to β^{t-1}, and that the rent in a given location only depends of the local productivity. Lastly, in the budget constraint, $W(e)$ represents the (non-human) wealth of a household of type e.

Aggregate non-human wealth. In our setup *without* absentee landlords, aggregate non-human wealth,

$$\bar{W}_0 = \int W_0(e) f(e) \, de,$$

must be equal to the present value of revenues of the real estate sector, rebated lump sum to the households. These profits are obtained by adding up

1. the profits of the representative households who sells its endowment of construction material,

2. the profits of the representative construction firm who buys construction material and builds houses,

3. the profits of real estate firms who buy houses from the construction firm and rent them out to households.

Clearly, when adding up 1 and 2, the revenue of selling and the cost of purchasing construction material cancel out. The same is true when adding up 2 and 3 up with the revenue and cost of selling and buying houses. Hence, after adding up 1, 2, and 3, the only thing that remains is the present value of the rents generated by the housing stock in each island. In the notation of the paper:

$$\bar{W}_0 = \sum_{t \ge 1} \beta^{t-1} \int R_t(s^t) h_t(s^t) n_t(e, s^t) \, de \, \mu_t(ds^t) = \sum_{t \ge 1} \beta^{t-1} \int R_t(s^t) H_t(s^t) \mu_t(ds^t), \tag{37}$$

where the second line follows from using the local housing market clearing condition.

An equilibrium. Now consider the equilibrium housing consumption and location plans described in the paper, $h_t^*(e)$ and $A_t^*(e)$, as well as the rent function $R_t^*(A_t)$. We verify that, as long as

$$W_0(e) + \sum_{t=1}^{\infty} [eA_t^*(e) - R_t^*(A_t) h_t^*(e)] \ge 0,$$

53

$h_t^*(e)$ and $A_t^*(e)$ solve the household's inter-temporal problem described above, together with some positive consumption plan $c_t^*(e)$ which is consistent with market clearing. Note that the inequality above can be satisfied in many different ways. One possibility is, for instance, to let $W_0(e)$ be the present value of rental payments of a household of ability e who follows the prescribed location and consumption plan $A_t^*(e)$ and $h_t^*(e)$.

To see this, denote aggregate output at time t in the equilibrium of the paper by Y_t^*, and consider the consumption plan $c_t^*(e) = \omega_t(e)Y_t^*$, where

$$\omega_t(e) = \frac{W_0(e) + \sum_{t=1}^{\infty}\left[eA_t^*(e) - R_t^*(A_t)h_t^*(e)\right]}{\sum_{t=1}^{\infty}\beta^{t-1}Y_t^*}.$$

Since the $W_0(e)$ add up to the present value of aggregate rents, and since the present value of wages add up to the present value of aggregate output, then by construction the shares $\omega_t(e)$ add up to one. Therefore, the candidate optimal consumption plan is consistent with market clearing.

Denote by $h_t^*(e)$ and $A_t^*(e)$ the housing and location plan of a type-e household in the equilibrium of the paper. We now verify that this plan solves the inter-temporal maximization problem above. Indeed, consider any other plan $A_t(e)$, $c_t(e)$ and $h_t(e)$. Suppressing the dependence on e to simplify notation, we find that the utility difference between the candidate optimal plan and this plan is:

$$\sum_{t=1}^{\infty}\beta^{t-1}\left(c_t^* - c_t + v(h_t^*) - v(h_t)\right)$$
$$\geq \sum_{t=1}^{\infty}\left(eA_t^* + v(h_t^*) - R_t(A_t^*)h_t^* - eA_t - v(h_t) + R_t(A_t)h_t\right)$$
$$\geq 0,$$

where the second line follows by substituting in the inter-temporal budget constraint, which holds with an equality at the candidate optimal plan, and with an inequality at the other plan. The third line follows because by construction at each time, the candidate optimal plan maximizes $eA + v(h) - R_t(A)h$, with respect to location A and housing consumption h.

B.3 Convexity in a Static Setting with General Utility

In this appendix we show that if agents have a general concave non-separable utility over non-housing and housing consumption bundles, the rent remains a convex function of productivity.

B.3.1 Equilibrium Definition

Consider the following static environment. As in the paper, there is a continuum of locations and in each location there is a representative competitive firm operating a linear technology $n \mapsto An$, where A denotes the location-specific productivity and n the number of effective units of labor employed.

Productivity and housing stock can differ across locations. We assume that productivity can take on finitely many values $\{A_1, A_2, \ldots, A_N\}$, and that the housing stock belongs to some interval $(0, \bar{H}]$. We let $\mu(ds)$ denote the joint distribution of productivity and housing stock $s = (A, H)$ across locations, and μ_i denote the measure of locations with productivity A_i.

There is a representative family made up of a continuum of members with heterogenous abilities. The density of family members with ability $e \in [\underline{e}, \overline{e}]$ is denoted by $f(e) > 0$. The head of the family decides where to locate each of its members, subject to the constraint that a family member has to work and live in the same place. Importantly, family members utility for non-housing and housing consumption is represented by some general strictly increasing and concave function $u(c, h)$, satisfying Inada conditions.

The head of the family chooses the number $n(e \mid s)$ of family members with ability e per location of type s, as well as their non-housing and housing consumption, $c(e, s)$ and $h(e, s)$, in order to maximize the family utility:

$$\int u(c(e, s), h(e, s)) n(e \mid s) \mu(ds),$$

subject to the family budget constraint:

$$\int \left(c(e, s) + R(s) h(e, s) - eA \right) n(e \mid s) \mu(ds) \le B,$$

and the constraint that:

$$\int n(e \mid s) \mu(ds) = f(e),$$

meaning that the number of family members with ability e must add up to $f(e)$. An equilibrium is a family location and consumption plan $\{n(e \mid s), c(e, s), h(e, s)\}$ and a rent function $R(s)$ such that: given the rent, the family plan solves the family's optimization problem, the right-hand side B of the family budget constraint is equal to the aggregate profit of real estate firms, and the housing market clears in every location, i.e.

$$\int n(e \mid s) h(e, s) \, de = H.$$

Following the same argument as in Section 2.3.1 in the paper, one shows easily that the following elementary properties must hold:

- If a location is not populated, then its rent must be equal to zero.

- The rent in a location only depends on the location productivity, A_i, and does not depend on the local housing stock H.

- The rent is an increasing function of productivity.

B.3.2   Optimal Consumption and Location Plan

Simplifying the family problem.   Note first that the head of the family finds it optimal to give the same consumption bundle (c_i, h_i) to all family members living in locations with productivity A_i, regardless of their ability, and regardless of the housing stock in these locations. Indeed, suppose that consumption was heterogenous among family members living in locations A_i. Then, it would be budget feasible to replace these heterogenous consumptions by the average consumption bundle among these family members. Because of concave utility, the average utility of family members living in locations A_i would increase.

This remark allows us to simplify the family's problem as follows. The head of the family chooses the consumption bundle (c_i, h_i) as well as the density $n_i(e)$ of members with ability e in locations i, in order to maximize:

$$\sum_{i=1}^{N} \mu_i \int n_i(e) u(c_i, h_i)\, de,$$

subject to

$$\sum_{i=1}^{N} \mu_i \int n_i(e) \left(c_i + R_i h_i - e A_i\right) de \leq B,$$

and:

$$\sum_{i=1}^{N} \mu_i n_i(e) = f(e).$$

Note that, because the price of non-housing consumption is the same across locations, it follows that the family head finds it optimal to smooth the marginal utility of its members across locations. That is, there is a $\lambda > 0$ such that, for all populated locations:

$$u_c(c_i, h_i) = \lambda.$$

The housing consumption in populated location i, on the other hand, must satisfy:

$$u_h(c_i, h_i) = \lambda R_i.$$

Characterizing the optimal location plan.   We now turn to the characterization of the optimal location plan. We show that, with concave utility, an optimal assignment of households across locations features positive assortative matching, and can be characterized using a method analogous to the one developed in the paper for a quasi-linear utility. The main result of this paragraph is

Lemma 7. *A budget feasible consumption and location plan* $\{c_i^*(e), h_i^*(e), n_i^*(e)\}$ *is optimal if and only if the budget constraint holds with equality and there exists some* $\lambda > 0$ *such that:*

$$
\begin{aligned}
(c_i^*, h_i^*) &= \arg\max \left\{ u(c, h) - \lambda(c + R_i h) \right\} \\
n_i^*(e) &= f(e) \mathbb{I}_{\{i \in \mathcal{I}(e)\}} \quad \textit{almost surely, where } \mathcal{I}(e) = \arg\max_{e \in [\underline{e}, \overline{e}]} \lambda e A_i + u(c_i^*, h_i^*) - \lambda(c_i^* + R_i h_i^*).
\end{aligned}
$$

The Lemma, proved in Section B.3.5, shows that an optimal location plan maximizes $\lambda e A_i + u(c_i^*, h_i^*) - \lambda(c_i^* + R_i h_i^*)$, interpreted as the marginal utility of allocating family member e to location i. Indeed, the first term is the utility value of earning the wage $e A_i$, in marginal utility units λ. The second term is the maximum utility of living in the location, net of the utility cost of non-housing and consumption. A special case of this condition is the one we used in the text when households have quasi-linear utility. In that case, indeed, $u(c, h) = c + v(h)$ so the marginal utility of non-housing consumption is $\lambda = 1$. Then, non-housing consumption drops out of the equation and we are left with the condition that an optimal location plan

maximizes $eA_i + v(h_i) - R_i h_i$, and $R_i = v'(h_i)$, which is the same as in the text.

B.3.3 Assortative Matching

Equipped with Lemma 7, one can show the same positive assortative matching result as in Proposition 2 for the present non-separable utility model: the sets ε_i are empty for all $i < p$, and increasing closed intervals $[e_i, e_{i+1}]$ for all $i \geq p$. Family members with ability e_{i+1} are indifferent between location i and $i + 1$, in that

$$v_i(e_i) + \lambda (e_{i+1} - e_i) A_i = v_{i+1}(e_{i+1}),$$

where $v_i(e) = \lambda e A_i + u(c_i^*, h_i^*) - \lambda(c_i^* + R_i h_i^*)$.

B.3.4 Convexity

To prove convexity of the rent with respect to productivity, we note that

$$
\begin{aligned}
v_i(e_i) &= \lambda e_i A_i + \max_{(c,h)} \{u(c, h) - \lambda(c + R_i h)\} \\
&= \lambda e_i A_i - \theta(R_i),
\end{aligned}
$$

where $\theta(R) = \min_{(c,h)} \{\lambda(c + Rh) - u(c, h)\}$ is an increasing and concave function, because it is the lower envelope of a collection of increasing affine functions of $R \mapsto \lambda(c + Rh) - u(c, h)$, for all (c, h). Now

$$
\begin{aligned}
\frac{R_{i+1} - R_i}{A_{i+1} - A_i} &= \frac{R_{i+1} - R_i}{\theta(R_{i+1}) - \theta(R_i)} \times \frac{\theta(R_{i+1}) - \theta(R_i)}{A_{i+1} - A_i} \\
&= \frac{R_{i+1} - R_i}{\theta(R_{i+1}) - \theta(R_i)} \times e_{i+1},
\end{aligned}
\tag{38}
$$

because

$$
\begin{aligned}
\theta(R_{i+1}) - \theta(R_i) &= e_{i+1} A_{i+1} - v_{i+1}(e_{i+1}) - e_i A_i + v_i(e_i) \\
&= e_{i+1} A_{i+1} - v_i(e_i) - (e_{i+1} - e_i) A_i - e_i A_i + v_i(e_i) \\
&= e_{i+1} (A_{i+1} - A_i)
\end{aligned}
$$

where the second line follows from the fact that workers e_{i+1} are indifferent between location i and location $i + 1$. Because e_i is increasing and $\theta(R)$ is concave, it follows from (38) that the slope of the rent is an increasing function of productivity.

B.3.5 Proof of Lemma 7

Necessity   We start by proving the necessity of the condition in Lemma 7. First, an optimal non-housing and housing consumption plan must maximize $u(c, h) - \lambda(c + R_i h)$ in every location i. Turning to the optimal

location plan, we let:

$$v_i(e) \equiv \lambda e A_i + u(c_i^*, h_i^*) - \lambda \left[c_i^* + R_i h_i^* \right]$$

$$\mathcal{I}(e) \equiv \arg \max_{i \in \{1,2,\ldots,N\}} v_i(e)$$

$$\varepsilon_i \equiv \{ e \in [\underline{e}, \overline{e}] : i \in \mathcal{I}(e) \}.$$

In words, $\mathcal{I}(e)$ is the set of locations that maximize the marginal value $v_i(e)$, and ε_i is the set of abilities who maximize their marginal value in a particular location i. Proceeding as in Result 2, in the Proof of Proposition 3, we find that:

Lemma 8. *For every $i \neq j$, $\varepsilon_i \cap \varepsilon_j$ is either empty or a singleton. Moreover, there are finitely many ability types whose $\mathcal{I}(e)$ has more than one element.*

If the second property were not satisfied, one island type would be visited by more than one ability type, which would contradict the first property. We now show that $\mathcal{I}(e)$ is the basis of an optimal assignment. That is, in an optimal location plan, the family head assigns member e to location in $\mathcal{I}(e)$, except perhaps for a measure-zero set of ability types. The proof is by contradiction: suppose that there is a positive measure set of ability types, $\mathcal{E}$, that are not assigned to a location in their $\mathcal{I}(e)$. Because of Lemma 8, we can assume that for all ability types $e \in \mathcal{E}$, $\mathcal{I}(e)$ is a singleton.

Now consider the following deviation. Move a fraction $\eta > 0$ of the ability types in $\mathcal{E}$ to location $\mathcal{I}(e)$, and give them the optimal local consumption bundle $(c_{\mathcal{I}(e)}^*, h_{\mathcal{I}(e)}^*)$. This results in a change in output of:

$$\eta \Delta Y = \eta \int_{e \in \mathcal{E}} \left(e A_{\mathcal{I}(e)} f(e) \, de - \sum_{i=1}^{N} \mu_i e A_i n_i(e) \, de \right)$$

The change in non-housing and housing consumption expenditure is:

$$\eta \Delta X = \eta \int_{e \in \mathcal{E}} \left(\left[c_{\mathcal{I}(e)} + R_{\mathcal{I}(e)} h_{\mathcal{I}(e)} \right] f(e) - \sum_{i=1}^{N} \mu_i \left(c_j + R_j h_j \right) n_i(e) \right) de.$$

To meet the budget constraint after moving these family members, one needs to adjust the consumption of the remaining family members so some level $\hat{c}_i$ in each populated region:

$$\hat{c}_i = c_i + \frac{\eta}{1 - \eta F(\mathcal{E})} (\Delta Y - \Delta X)$$

$$= c_i + \eta (\Delta Y - \Delta X) + o(\eta^2).$$

The resulting change of utility for the family is:

$$\eta \int_{e \in \mathcal{E}} \left[u(c_{\mathcal{I}(e)}, h_{\mathcal{I}(e)}) f(e) \, de - \sum_{j=1}^{N} \mu_j n_j(de) u(c_j, h_j) \right] + \lambda \eta (1 - \eta F(\mathcal{E})) (\Delta Y - \Delta X) + o(\eta^2)$$

$$= \eta \int_{e \in \mathcal{E}} \left[v_{\mathcal{I}(e)} f(e) \, de - \sum_{j=1}^{N} \mu_j n_j(de) v_j(e) \right] + o(\eta^2)$$

where the second line follows from substituting in the formula for ΔX and ΔY. Since $\mathcal{I}(e)$ achieves the strict maximum of $v_i(e)$, it follows that the change in utility is strictly positive, as long as η is small enough.

Sufficiency. To prove sufficiency, consider a consumption and location plan satisfying the conditions of Lemma 7, and compare it to any other budget feasible consumption and location plan. Without loss of generality, we can assume that this other plan prescribes the same consumption to all family members living in locations A_i. The utility difference between the two plans can be written as:

$$\sum_{i=1}^{N} \mu_i \int \left[n_i^*(e) u(c_i^*, h_i^*) - n_i(e) u(c_i, h_i) \right] de$$

$$= \sum_{i=1}^{N} \mu_i \int \left[\left(n_i^*(e) - n_i(e) \right) u(c_i^*, h_i^*) + n_i(e) \left(u(c_i^*, h_i^*) - u(c_i, h_i) \right) \right] de$$

$$\geq \sum_{i=1}^{N} \mu_i \int \left[\left(n_i^*(e) - n_i(e) \right) u(c_i^*, h_i^*) + n_i(e) \left(u_c(c_i^*, h_i^*)(c_i^* - c_i) + u_h(c_i^*, h_i^*)(h_i^* - h_i) \right) \right] de,$$

where the last line follows by concavity of the utility function. Plugging in the first order conditions $u_c(c_i^*, h_i^*) = \lambda$ and $u_h(c_i^*, h_i^*) = \lambda R_i$, we obtain:

$$\sum_{i=1}^{N} \mu_i \int \left[\left(n_i^*(e) - n_i(e) \right) u(c_i^*, h_i^*) + \lambda n_i(e) \left((c_i^* - c_i) + R_i(h_i^* - h_i) + eA_i - eA_i \right) \right] de$$

$$= \sum_{i=1}^{N} \mu_i \int \left[\left(n_i^*(e) - n_i(e) \right) \left(u(c_i^*, h_i^*) - \lambda(c_i^* + R_i h_i^* - eA_i) \right) \right] de$$

$$+ \lambda \sum_{i=1}^{N} \mu_i \int n_i^*(e) \left(c_i^* - R_i h_i^* - eA_i \right) de - \lambda \sum_{i=1}^{N} \mu_i \int n_i(e) \left(c_i - R_i h_i - eA_i \right) de.$$

Now using the fact that the budget constraint holds with equality at the candidate optimal plan, and with inequality for the other plan, we note that the sum of the last two terms is positive. Thus the change in utility is greater than:

$$\sum_{i=1}^{N} \mu_i \int \left[\left(n_i^*(e) - n_i(e) \right) \left(u(c_i^*, h_i^*) - \lambda(c_i^* + R_i h_i^* - eA_i) \right) \right] de$$

$$= \int \left[v_{\mathcal{I}(e)} f(e) - \sum_{i=1}^{N} \mu_i n_i(e) v_i(e) \right] de \geq 0,$$

where the last line follows because $v_{\mathcal{I}(e)} \geq v_i(e)$ for all i and e, by the definition of $\mathcal{I}(e)$.

B.4 Why our Ability Dispersion is Conservative

Our benchmark model calibration implies a cross-sectional standard deviation of ability of 0.079, corresponding to a variance of $0.079^2 = 0.0062$. In this appendix we argue that this number is conservative.

Ability has many dimensions and it is not obvious which component of ability matters for the assortative matching process of households with regions. Some components seem to matter little. For instance, the

component of ability that is measured by Mincerian human capital proxies such as education and cognitive skills is quite evenly distributed across regions and relatively constant over time; see Berry and Glaeser (2005) and Bacolod, Blum, and Strange (2007). Therefore, the dispersion of that ability component that matters for assortative matching should be smaller than the variance of the Mincerian *residual*, which itself is about 0.10 according to the individual level data of Heathcote, Storesletten, and Violante (2008). This is 16 times larger than the 0.0062 variance we feed in the model. By this benchmark, we are feeding in a mild amount of ability dispersion.

In what follows, we formally develop the argument. It results in a tighter upper bound on ability dispersion by exploiting information coming from the cross-regional variance of wages. The tighter bound is 13 times larger than the variance of 0.0062 we feed in the model.

A simple statistical model

There is a continuum of regions indexed by i and within each region there is a continuum of households indexed by j. Households differ in terms of their ability. We assume that ability is measured in effective units of labor and has two independent components h_j and k_j. The first ability component has no complementarity with the local productivity a_i, while the second component does. Because of the complementary between k_j and a_i, we assume that a fraction v of households match assortatively with regions. We assume that the complementary fraction $1 - v$ matches randomly, capturing the fact that some location decisions are driven by non-productive motives (e.g., demographic reasons such as divorce). All random variables have mean zero.

If the match is assortative, the log wage is $w_{ij} = a_i + e_i + h_j$, where e_i is the average ability k_j of households j who locate in region i. Because of assortative matching, the log productivity of the region is perfectly correlated with this component of ability and we can write: $a_i = \zeta e_i$. If they are matched randomly, their log wage is $w_{ij} = a_i + k_j + h_j$, where $a_i = \zeta e_i$ but k_j is the random ability of the household and is independent from a_i. We assume that e_i and k_j are independently and identically distributed with variance $V(e)$. Taken together, the model implies that the wage of a randomly chosen household is

$$
\begin{aligned}
w_{ij} &= (1 + \zeta)e_i + h_j \text{ with probability } v \\
w_{ij} &= \zeta e_i + k_j + h_j \text{ with probability } 1 - v
\end{aligned}
$$

Mincerian regression.

Now suppose an econometrician runs a Mincerian regression on a representative sample of households. Empirical evidence cited above suggests that Mincerian human capital proxies are quite evenly distributed across regions. In the context of our model, this means that the Mincerian regressors are correlated with h_j, but are orthogonal to e_i and k_j. Thus, the wage can be written as:

$$
\begin{aligned}
w_{ij} &= h_j^m + (1 + \zeta)e_i + (h_j - h_j^m) \text{ with probability } v, &(39) \\
w_{ij} &= h_j^m + \zeta e_i + k_j + (h_j - h_j^m) \text{ with probability } 1 - v, &(40)
\end{aligned}
$$

60

where h_j^m denote the fitted value of a Mincerian regression. The variance of the Mincerian residual is therefore:

$$
\begin{aligned}
V(w_{ij} - h_j^m) &= v\left((1+\zeta)^2 V(e) + V(h_j - h_j^m)\right) + (1-v)\left((1+\zeta^2)V(e) + V(h_j - h_j^m)\right) \\
&= V(h_j - h_j^m) + \left(1 + 2v\zeta + \zeta^2\right)V(e) \\
&\geq \left(1 + 2v\zeta + \zeta^2\right)V(e) \geq V(e)
\end{aligned}
\tag{41}
$$

This implies a first upper bound:

$$
V(e) \leq V(w_{ij} - h_j^m).
$$

Based on the analysis of Heathcote, Storesletten, and Violante (2008) for individuals in the U.S., the part of log wage variance that is unexplained by Mincerian controls and shocks is about 0.10. By this measure, the upper bound on the variance of ability is 0.10. This is sixteen times bigger than the variance of 0.0062 we use in our calibration.

A Tighter Bound.

Next, we derive a tighter upper bound by exploiting information coming from the cross-sectional variance of *regional* wages. Using (39) and (40), we obtain that average wage within region i is

$$
\bar{w}_i = v(1+\zeta)e_i + (1-v)\zeta e_i = (v + \zeta)e_i,
$$

after averaging the zero-mean random variables k_j and h_j. Thus, the cross-sectional variance of the regional wages is:

$$
V(\bar{w}_i) = (v + \zeta)^2 V(e) = V(w_{ij} - h_j^m) - V(h_j - h_j^m) - (1 - v^2)V(e).
$$

Together with (41), this last equation implies that the difference between the overall cross-sectional variance of the Mincerian residual and the cross-regional variance is:

$$
V(w_{ij} - h_j^m) - V(\bar{w}_i) = V(h_j - h_j^m) + (1 - v^2)V(e).
$$

We thus obtain the second upper bound:

$$
V(e) \leq \frac{V(w_{ij} - h_j^m) - V(\bar{w}_i)}{1 - v^2} \simeq \frac{0.10 - 0.02}{1 - v^2},
$$

because, in our data, the variance of log wage across regions is around 0.020 (on average over time). After making the bound as tight as possible by setting $v = 0$, we obtain an upper bound of 0.080. This is thirteen times the ability variance we use in our calibration. In short, we use a conservative value for the amount of dispersion in ability across households.

C Computations

In this Appendix we explain how we compute an equilibrium.

C.1 Distributional Assumptions on Productivity and Ability

We start by describing our specification of the productivity process and the distribution of ability.

C.1.1 Productivity

Our discrete Markov chain for productivity approximates a geometric random walk with exponential lifetime. As in Yaari (1965), this guarantees the existence of a stationary distribution.

The continuous-state process we seek to approximate. We assume that, every period $t \in \{1, 2, \ldots\}$, a measure $\lambda \in (0, 1)$ of new islands are exogenously created, with an initial log productivity drawn from a normal distribution with mean μ_{bt} and standard deviation σ_{bt}. In every subsequent period, an island either disappears with probability λ, or survives with probability $1 - \lambda$, in which case it draws the new log productivity

$$a_t = a_{t-1} + \sigma_{at}\varepsilon_t, \tag{42}$$

where $\{\varepsilon_t\}_{t=0}^{\infty}$ is a sequence of independent standard normal random variables. It follows from the above specification that the steady state cross-sectional distribution of log productivity across islands is an infinite mixture of normal distributions. Namely, at time t, there is a fraction $\lambda(1 - \lambda)^k$ of islands born at time $t - k$. The distribution of log productivity among these islands is normal, with mean μ_{bt-k} and variance $\sigma_{bt-k}^2 + \sigma_{at-k+1}^2 + \ldots + \sigma_{at}^2$. We are not aware of a closed form formula for this distribution, even in the stationary case where μ_{bt}, σ_{bt} and σ_{at} are time independent. The stationary distribution is known, however, for the continuous-time counterpart of this stochastic process. In particular, Reed (2001) and (2003) show that the stationary distribution behaves like a Pareto distribution in its two tails, with parameters he characterizes in closed form. Our numerical computations suggests that, in the discrete-time case, the stationary distribution exhibits a similar Pareto behavior in both tails.

Note that, although we don't have closed form solution for the distribution, its first and second moments can be calculated easily:

$$
\begin{align}
\text{avg}(a_t) &= \lambda\mu_{bt} + (1 - \lambda)\text{avg}(a_{t-1}) \tag{43}\\
\text{avg}(a_t^2) &= \lambda\left(\mu_{bt}^2 + \sigma_b^2\right) + (1 - \lambda)\left(\text{avg}(a_{t-1}^2) + \sigma_{at}^2\right), \tag{44}\\
\text{disp}(a_t)^2 &= \text{avg}(a_t^2) - \text{avg}(a_t)^2, \tag{45}
\end{align}
$$

where the $\text{avg}(x)$ and $\text{disp}(x)$ denote, respectively, the cross-sectional average and standard deviation of some random variable x.

Increasing dispersion through innovation. In this paragraph we explain how our benchmark calibration engineers the increase in productivity dispersion.

We start with values for the initial $(t = 0)$ and final $(t = T)$ steady state dispersion $\text{disp}(a_0)$ and $\text{disp}(a_T)$ and we impose that

1. the dispersion increases linearly between $t = 0$ and $t = T$:

$$\text{disp}(a_t) = \text{disp}(a_0) + \frac{t}{T}\left(\text{disp}(a_T) - \text{disp}(a_0)\right). \tag{46}$$

2. the cross-sectional average of productivity *levels*, $\text{avg}(A_t)$, remains constant over time.

Note that the increase in dispersion creates a Jensen inequality bias that we need to correct in order to keep the cross-sectional average of productivity levels constant. Our bias correction is based on the approximation:

$$\text{avg}(a_t) \simeq \log(\text{avg}(A_t)) - \frac{\text{disp}(a_t)^2}{2}. \tag{47}$$

Then given a constant time path for $\text{avg}(A_t)$, and given the time path (46) for dispersion, equation (47) provides the time path for $\text{avg}(a_t)$. Since, on the other hand, $\text{avg}(a_t)$ evolves according to equation (43), this pins down the time path of μ_{bt}:

$$\mu_{bt} = \frac{\text{avg}(a_t) - (1 - \lambda)\text{avg}(a_{t-1})}{\lambda}, \tag{48}$$

the mean log productivity of newly created islands. In our numerical computation, approximation (47) works well in the sense that the time path of $\log(\text{avg}(A_t))$ is indeed approximately constant over time.

We engineer the increase in dispersion by progressively increasing the volatility of productivity innovations. The time path of innovation producing the linear time path (46) of dispersion can be calculated with equations (44) and (45):

$$\sigma_{at}^2 = \frac{\text{disp}(a_t)^2 + \text{avg}(a_t)^2 - \lambda\left(\mu_{bt}^2 + \sigma_b^2\right)}{(1 - \lambda)} - \text{avg}(a_{t-1}^2),$$

given that we know $\text{disp}(a_t)$ from (46), $\text{avg}(a_t)$ from (47), μ_{bt} from (48), and that $\text{avg}(a_{t-1}^2) = \text{disp}(a_{t-1})^2 + \text{avg}(a_{t-1})^2$.

Increasing dispersion through innovation and persistence. In the alternative calibration of Section 4.5.2, we propose to increase productivity dispersion with smaller productivity shocks, by deterministically increasing the productivity of regions above the average productivity and, vice versa, decreasing productivity of regions below the average. This approach has the benefit of making the *rank* of islands in the productivity distribution much more persistent, in line with what we observe in the data. Precisely, we assume that, during the transition $t \in \{1, 2, \ldots, T\}$, the productivity process of a region follows the first-order auto-regression:

$$a_t = \text{avg}(a_{t-1}) + \rho_t\left(a_{t-1} - \text{avg}(a_{t-1})\right) + \sigma_{at}\varepsilon_t. \tag{49}$$

The increase in dispersion is obtained by setting $\rho_t > 1$, i.e. by scaling up productivity deviations from the mean. The first moment of the cross-sectional productivity distribution is still given by (43), while the

second moment is

$$\text{avg}(a_t^2) \quad = \quad \lambda\left(\mu_{bt}^2 + \sigma_b^2\right) + (1 - \lambda)\left(\text{avg}(a_{t-1})^2 + \rho_t^2 \text{disp}(a_{t-1})^2 + \sigma_{at}^2\right), \tag{50}$$

As in the previous paragraph, we impose a deterministic linear increase in the cross-sectional productivity dispersion. As before, this imposes to choose μ_{bt} according to formula (48).

As for the time path of innovation volatility, σ_{at}, we now assume that it increases linearly from the old steady state σ_{a0} to the new one σ_{aT}:

$$\sigma_{at} = \sigma_{a0} + \frac{t}{T}\left(\sigma_{aT} - \sigma_{a0}\right).$$

We then pick the time path of persistence so as to satisfy equation (50), i.e.:

$$\rho_t = \sqrt{\frac{\text{avg}(a_t^2) - (1 - \lambda)\left(\text{avg}(a_{t-1}^2) + \sigma_{at}^2\right) - \lambda\left(\mu_{bt}^2 + \sigma_b^2\right)}{(1 - \lambda)\text{disp}(a_{t-1})^2}}.$$

The discrete state approximation. Our approximation works as follows. At each time $t \in \{1, 2, \ldots\}$, given the cross sectional mean and variance implied by the above stochastic process, we pick $N = 190$ Gaussian quadrature points $a_{1t} < a_{2t} < \ldots < a_{Nt}$ using Fackler and Miranda's (2002) MATLAB routine `qnwnorm.m`. We take these quadrature points to be the N possible states of log-productivity at time t. A newborn island draws a state at random, according to the probability distribution $\{g_{it}\}_{i=1}^N$ which is chosen to approximate a normal with mean μ_{bt} and variance σ_{bt}^2: specifically, we let g_{it} be proportional to the normal probability density of being in state a_{it}. An existing island draws a new productivity state at random, according to the probability transition matrix G_{t-1} constructed using the method of Tauchen and Hussey (1991).

C.1.2 Ability

Given that the productivity distribution across islands exhibits Pareto behavior in its tail, we chose to pick an ability distribution with the same tail properties. Numerically, we found it to be an important property for our model to have reasonable behavior in its tails. Intuitively, it ensures that the ratio of ability and productivity differentials between islands stays roughly constant.

Given that we know from Reed (2001) and (2003) that the steady state productivity distribution exhibits Pareto behavior in its tail, we assume that ability is distributed according a double-Pareto distribution, with cumulative density function (cdf):

$$F(x) \quad = \quad \frac{1}{2}\left(\frac{x}{X_e}\right)^{k_e} \quad \text{if } x \leq X_e \tag{51}$$

$$= \quad \frac{1}{2} + \frac{1}{2}\left[1 - \left(\frac{x}{X_e}\right)^{-k_e}\right] \quad \text{if } x > X_e. \tag{52}$$

Because our results require that the ability distribution has a compact support, we truncate a small probability mass, $\alpha/2$, in both tails. As long as α is small, our results are not sensitive to this procedure (we set it equal to 10^{-8}). Precisely, we pick two truncation points $\underline{X}_e < \overline{X}_e$ such that $F(\underline{X}_e) = \alpha/2$ and $1 - F(\overline{X}_e) = \alpha/2$.

Given the cdf in equations (51) and (52), we find that:

$$F(\underline{X}_e) = \alpha/2 \quad \Rightarrow \quad \underline{X}_e = X_e \alpha^{1/k_e}$$
$$1 - F(\overline{X}_e) = \alpha/2 \quad \Rightarrow \quad \overline{X}_e = X_e \alpha^{-1/k_e}.$$

After truncating, we re-normalize the cdf. This amounts to using the *conditional* distribution on the interval $[\underline{X}_e, \overline{X}_e]$ whose cdf is:

$$F^T(x) \;\; = \;\; \frac{F(x) - \alpha/2}{1 - \alpha} \text{ for } x \in [\underline{X}_e, \overline{X}_e],$$

where the $\alpha/2$ term comes about because $F(\underline{X}_e) = \alpha/2$ by construction. The procedure is summarized in Figure 9. The last parameter to pick is X_e. We choose it so that the mean of the truncated distribution is normalized to one. Simple calculations show that this imposes

$$X_e = 2(1 - \alpha) \left[\frac{k_e}{k_e + 1} \left(1 - \alpha^{\frac{k_e+1}{k_e}} \right) + \frac{k_e}{k_e - 1} \left(1 - \alpha^{\frac{k_e-1}{k_e}} \right) \right]^{-1}.$$

Lastly, note tat the parameter k_e governs the standard deviation of log ability. Simple calculations show that, in order for the standard deviation of log ability to be equal to σ_e, the tail parameter k_e must be

$$k_e = \frac{\sqrt{2}}{\sigma_e} \times \sqrt{\frac{1 - \alpha/2 - \alpha/2(1 + \log(\alpha))^2}{1 - \alpha}}.$$
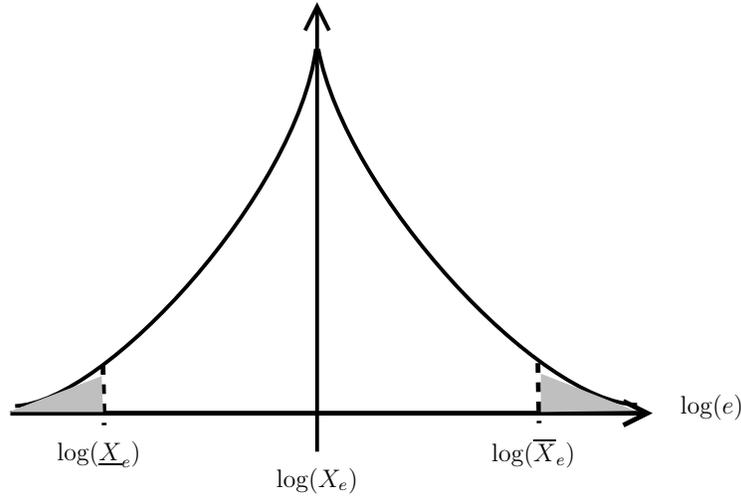

Figure 9: In level, the distribution of ability is a double Pareto. As a results the logarithm of ability is distributed according to a double exponential that is symmetric around X_e. To keep a compact support, we need to truncate the distribution in both tails. The figure shows the log of our truncation points, $\log(\underline{X}_e)$ and $\log(\overline{X}_e)$, which are chosen so to leave a probability mass of $\alpha/2$ in both tails.

D   Data and Calibration

This appendix provides details on data sources, definitions, and calculations. Our unit of observation is a core-based statistical area (metropolitan statistical area or MSA). We use the 2006 MSA definitions. In an effort to recognize size differences between MSAs, for the largest MSAs we replace the MSA by its constituent metropolitan divisions (MSAD), whenever these are defined in the data. There are 11 such instances in which are replaced by 29 constituent divisions (included divisions): Boston (Boston-Quicy, Cambridge-Newton-Framingham, Essex County, and Rockingham County-Strafford County), Chicago (Chicago-Naperville-Joliet, Gary, Lake County-Kenosha County), Dallas (Dallas-Plano-Irving and Forthworth-Arlington), Detroit (Detroit-Livonia-Dearborn and Warren-Farmington Hills-Troy), Miami (Fort Lauderdale-Pompano Beach-Deerfield Beach, Miami-Miami Beach-Kendall, and West Palm Beach-Boca Raton-Boynton Beach), Los Angeles (Los Angeles-Long Beach-Glendale and Santa Ana-Anaheim-Irvine), New York (Edison, Nassau-Suffolk, Newark-Union, and New York-White Plains-Wayne), Philadelphia (Camden, Philadelphia, and Wilmington), San Francisco (Oakland-Fremont-Hayward and San Francisco-San Mateo-Redwood City), Seattle (Seattle-Bellevue-Everett and Tacoma), and Washington (Washington-Arlington-Alexandria and Bethesda-Gaithersburg-Frederick). Forty-nine MSAs were not used because of missing house price data. But none of them have more than 200,000 jobs in 2007. Our resulting sample consists of 330 regions, which we refer to as MSAs even though 29 of them are actually MSADs. These 330 metropolitan areas account for 83.9% of all jobs in the U.S. in 2007.

D.1   Nominal Wages and Number of Jobs

The regional wage data are from the Regional Economic Information System (REIS) compiled by the Bureau of Economic Analysis (BEA, Table CA34). REIS reports total wage and salary income as well as the number of employees for each of the MSAs. We calculate the average wage per job in a region as total wage and salary disbursements divided by the total wage and salary employment. Wage and salary disbursements is a measure of total earnings. It consists of the monetary remuneration of employees, including the compensation of corporate officers; commissions, tips, and bonuses; and receipts in kind, or pay-in-kind.

The underlying micro data are constructed based primarily on data from quarterly unemployment insurance (UI) contribution reports that are filed with state employment security agencies by employers in industries that are covered by, and subject to state UI laws. The employment and security agencies summarize the data by county. The data from all states are then published as the Quarterly Census of Employment and Wages (QCEW) by the Bureau of Labor Statistics (BLS) of the Department of Labor. The QCEW data account for 95 percent of wage and salary disbursements as estimated by BEA.

Whenever we calculate population-weighted moments across regions, we use the number of jobs from the REIS as population weights.

D.2   Nominal House Prices

We form a time series of home prices by combining level information from the 2000 Census with time series information from Freddie Mac. From the 2000 Census, we use the nominal home value for the median single-family home value. We use the Freddie Mac Conventional Mortgage Home Price Index (CMHPI) from 1975 until 2007. The CMHPI is a repeat-sale house-price index (e.g., Case and Shiller (1987)) which pertains to

single-family properties purchased or refinanced with a mortgage below the conforming loan limit ($417,000 in 2007). As a repeat-sale, it is a constant quality house price index.

In 1975, CMHPI data are only available for 81 regions. We refer to this sample as the balanced panel of 81 MSAs. Over time, more MSAs enter the sample. By 1985, there are 235 MSAs, and by 1996, all 330 regions have house price data. From 1996 until 2007, the sample coverage stays constant at 330 metropolitan areas. We refer to the complete sample as the unbalanced panel; the number of regions gradually increases from 81 to 330 over time. The balanced panel of 81 regions accounts for 58.5% of jobs, while we recall that the unbalanced panel of 330 regions accounts for 83.9% of jobs in 2007.

Since the 2000 Census data are based on an earlier MSA classification, we carefully map the median home value in the Census to the corresponding MSA in the CMHPI data. For the large areas, such as San Francisco or New York, we match up the primary metropolitan statistical areas with the metropolitan divisions. When Census data is missing, we manually construct the median home value by population-weighting the home values in the constituent counties.

D.3 Nominal Rents

We use rental data from the Fair Market Rents database (FMR), published annually by the U.S. Department of Housing and Urban Development (HUD) for 530 metropolitan areas. The FMR are gross rents, including utilities, and are used to determine payment amounts in various government housing subsidy programs. The FMR reports the 40th percentile of the housing rent distribution, the dollar amount below which 40 percent of the standard-quality rental housing units are rented in a given area. The 40th percentile rent is drawn from the distribution of rents of all units occupied by recent movers, who moved to their present residence within the past 15 months. Social housing units are excluded from the distribution. In the calculation, the FMR uses base-year rent levels from the decennial Census, which are trended forward using regional CPI data for rents and utilities from the Bureau of Labor Statistics for 102 regions (not publicly available from the BLS). For the other regions, coarser regional data are used for the updating. The data are available only for 1982 and 1984-2007.

The FMR data are reported for finer regions than the 330 metropolitan areas in our wage and house price data. For example, for the metropolitan area Atlanta, there are 23 sub-divisions in the FMR data. We use the 2000 Census population in each of these sub-divisions to construct population weights that sum to one in each MSA. We use these population weights to form a time series of rents for each MSA. For some large MSAs which we replaced by their constituent MSADs (e.g., Boston, Detroit, and Philadelphia), we have the rent at the MSA level but not at the MSAD level. We assign the same rent to each MSAD.

One problem with the FMR data is that the reported rent percentile changes over time. The reported percentile is the 45th percentile until 1993 for all regions, at which point it switches to the 40th percentile for all regions. In 2000, it switches to the 50th percentile in 40 regions, and in 2005 it switches back to the 40th percentile for some of them. Typically, the regions with switchers are large, populous regions. For example, the reported rent percentile for Atlanta is 45 until 1993, 40 from 1994 until 1999, 50 from 2000 until 2004 and 40 afterwards. We try to correct for these switches in two ways. First, to deal with the switch from 45 to 40 in 1993-1994, we adjust the 1994 rent level up by a constant $\left(\Phi^{-1}(0.45) - \Phi^{-1}(0.40)\right) * c$, where $\Phi(\cdot)$ is the normal cdf, and the constant c is chosen so that the resulting average rent increase between 1993 and 1994 matches the nationwide rent increase. It is easy to show that if rents are normally distributed with standard

deviation σ_r, then $c = \sigma_r$. We then apply the growth rate from the 40% rent series to update the 45% series after 1994. We do a second, partial correction for the post-1994 switches: when available, we use BLS data to update the FMR rents from 1999 until 2007. The BLS constructs a price index for housing (gross rents) for a sample of 25 large metropolitan areas (most of them are combined metropolitan statistical areas or CMSAs, which account for 41 of our regions). We use this price index for housing to roll forward the 1998 FMR rents to form the 1999-2007 rent data for these 41 regions. These are the regions where many of the percentile switches are concentrated.

To obtain the level for each regions, we use the median gross rent from the 2000 Census. This is the same data source we used to impute the 2000 median home values.

D.4 Regional Non-Housing Price Indices

For each region, we construct a non-housing price index, which reflects the local cost of living on all goods and services excluding gross rent, which is the sum of rent and utilities. We form a time-series of non-shelter prices by combining level information from the 2000 COLI survey with time series information from the Bureau of Labor Statistics (BLS).

The BLS provides regional price indices for the 23 largest metropolitan areas from 1975 to 2006. The base year is 1983-1984. We use their "all items less shelter" index. Since these 23 areas are consolidated metropolitan areas (e.g., the San Francisco CMSA contains the San Francisco MSAD, the San Jose MSA, and the Oakland MSAD), we impute the index to all constituent MSAs and MSADs. This delivers observations for 43 out of our 330 areas. For all other metropolitan areas, the BLS does not provide detailed ex-shelter price index data. Instead, it publishes ex-shelter price indices as a function of the city population and Census region. There are 3 size bins and 4 Census regions. The size bins are: greater than 1500000 (class A), between 50000 and 1500000 (class B/C) and below 50000 (class D). The Census regions are Northeast, Midwest, South, and West. We assign each of the remaining 292 regions a CPI ex-shelter index based on its Census region and its population. For the Northeast and the West, the class D series are missing, in which case we use the overall regional index.

The resulting price index is 100 for all regions in the base year 1983-84, by construction. To obtain the cross-sectional variation in the price level, we purchased data from COLI (formerly, ACCRA). This private firm conducts surveys to gauge the cost of living in various metropolitan areas. They collect data on 75 goods and services, which are aggregated into six main categories: transportation (10%), housing (28%), utilities (8%), groceries (16%), health care (5%), and miscellaneous goods and services (33%). We use data for the year 2000, construct a non-housing price index by aggregating the transportation, groceries, health care, and miscellaneous goods and services (using the above weights), then average across the four quarters of 2000. This delivers a non-housing price index for 303 regions. We are able to match 230 of these regions to MSAs in our sample of 330. For the other 100 regions in our sample, we use a transformation of a cost of living index from Boyer and Savageu's "Places Rated Almanac, 1989". The latter ranks the cost of living in 333 regions from 1 to 333. We calculate the percentile of each region in the "Places Rated" distribution. For each of the 100 regions with missing COLI price data, we impute a non-housing price index equal to the price of that same percentile but from the COLI distribution of 230 regions. Finally, we scale down the resulting 2000 COLI price index so that it has a mean of 100 and a coefficient of variation (CV) which is smaller by a factor of 2.034. The reason for scaling down the CV is that i) rent data suggests that, relative to Census,

COLI imputes too much cross-sectional variation in prices across regions and ii) our house price levels are based on Census data. Our scaling factor of 2.034 is obtained as the ratio of the CV of gross rent in COLI to the CV of gross rent in the 2000 Census. That is, we use the gross rent series, which we have available from both Census and COLI, to infer the extent to which COLI overstates regional dispersion in non-housing prices. The resulting non-housing price index has a cross-sectional correlation of 0.72 with the housing price index. This is very close to the 0.69 correlation between housing and non-housing price indices in the COLI data.

We deflate nominal wages, house prices, and rents of any region by its CPI ex-shelter index.

D.5   Population in highest-wage quintile

The fraction of jobs (of the population) that resides in the top wage quintile of regions (Q5) is the last moment of interest. As the next paragraph illustrates, numbers can differ substantially depending on the sample we use. Given the model's emphasis on the entire cross-section of regions, it seems best to use the largest available sample of regions. We have data on wages and number of jobs for a balanced panel of 955 metropolitan and micropolitan areas. Together, they cover 95% of the US population in 2007. For this large sample, the fraction of jobs in the top wage quintile increases from 64.88% in 1975 to 73.09% in 2007.

For comparison, for a balanced panel of 330 regions, we find an increase of the population in Q5 from 41.2% to 55.5%, while for the balanced panel of 81 regions, there is a small decline from 33.6% to 31.7%. This decrease is not surprising: most of these 81 regions would be in the top-wage quintile in the larger samples of 330 and 955 regions. It merely reflects faster growth in the next-biggest regions relative to the biggest regions. Finally, for the increasing sample for which we have house price data (increasing from 81 to 330 regions), we find the largest increase from 33.6% to 55.5%. This is not the right target for our model because almost half of this increase is a pure sample composition effect. This suggests that the increase from 64.88% to 73.09% we target is similar to that in the balanced sample of 330; the higher level of concentration in the sample of 956 is more representative of the entire distribution of regions we are trying to capture in the model.

E   Growth and Detrending

This appendix describes how we deal with population growth and growth in house size.

E.1   Adjusting Model for Growth

Our model is consistent with growth in the number of households. That is, let us start an economy where the number of households, construction material and permits grow at possibly time-varying rate g_{Nt} between year t and year $t+1$. Then, the relative prices and per-household quantities are the same as in an economy without growth. The only change is that the depreciation discount factor for housing is $(1-\delta)/(1+g_{Nt})$. Note that population growth lowers the effective depreciation rate.

We propose to de-trend the housing stock not only by the growth rate in the number of *households* but also by the growth rate in house size. De-trending by household population growth is necessary because our model considers a unit measure of households, so all variables have to be expressed in *per-household* units.

De-trending by housing size is reasonable because our preference specification is not consistent with balanced growth in productivity and housing size, so it is not equipped to deal with the price impact of trends in the size of housing services (a feature of housing "quality"). Further motivation comes from the fact that the data do not seem consistent with balanced growth either: over our sample period, population-weighted average real wages grew at $g_w = 0.80\%$ per year in the unbalanced panel whereas house size grew at the higher rate of $g_H = 1.256\%$ per year. A reasonable approach is to remove the size trend from the data before calibrating the model because the above observations suggest that our model is better suited to explain de-trended data. After de-trending the law of motion for the de-trended housing stock becomes:

$$\hat{H}_{it} = \frac{(1-\delta)}{(1+g_{Nt})(1+g_H)}\hat{H}_{it-1} + \hat{M}_{it},$$

where g_{Nt} is the growth rate of the households' population between $t-1$ and t and g_H is the growth rate of house size.

In the calibration, we pick the aggregate supply $\hat{M}_t$ of construction material along the transition so that, year by year, the aggregate housing supply per household, $\hat{H}_t$, matches the de-trended house size per household we observe in the data. This amounts to fix exogenously the total quantity of square feet in the economy. Then, the equilibrium endogenously allocate these square feet across regions.

E.2 Adjusting Data for Growth

We deflate housing prices by the growth rate of house size. This allows us to capture the fact that part of the increase in the price of single-family homes is simply due to the fact that houses have been getting bigger over time. We obtain data from the Census on the average square footage of completed single-family housing for sale inside metropolitan areas. The data are annual from 1975 until 2007. Over this period, the average house size grew from 1,715 to 2,563 square feet. This represents a growth rate of $g_H = 1.2555\%$ per year. We form a detrended house size time series by dividing the house size in year $t > 1975$ by $(1+g_H)^{t-1975}$. We feed in this detrended house size series into the model: the initial steady state house size is 1,715, the house size in 1976 until 2007 is assigned to periods 1 through 32 in the transition, and from period 33 onwards, we assume the de-trended housing size remains at 1,715 square feet.

We need to de-trend wages in the data before we can feed in observed wages into the model. One should be careful, however, not to remove the entire wage trend observed in the data. Indeed, even without a trending productivity, our model endogenously generate a trend in wage. The reason is that, when productivity dispersion goes up, households concentrate in high productivity regions so the wage in the model mechanically trends up. We thus chose to remove the part of the wage trend that our model fails to explain. That is, if g_w is the growth rate of wages in the data, and g_m is the growth rate of wage in the calibrated model, then we de-trend the observed wage at the rate $g_w - g_m$ (for the wage-feeding exercise). In practise, a given calibration of the model implies a wage growth rate that is different from the wage growth rate in the data. We iterate on the parameter g_m until wage data deflated by $(1+g_m)^{t-1975}$ have the same growth rate as wages in the (detrended) model.

E.3   Demographics

We feed into the model a time series for the growth rate in the number of households g_{Nt}. We set $g_{N0} = 1.81\%$ in the initial steady state, which is the observed growth rate in the number of households between 1974 and 1975 in the U.S. Census data. For periods 1 through 32 of the transition, g_{Nt} is taken from the 1976-2007 data. There are no data available that forecast the growth rate in the number of households for 2008 and beyond. We make the following assumptions to obtain reasonable numbers. For periods 33 through 75 of the transition we apply Census *population* projections for 2008-2050, and combine them with the evolution of the number of people per household. The number of people per household for periods 33 through 75 is obtained by fitting a quadratic spline through the 1975-2007 data. After period 75, we assume that population growth stays at its 2050 value of 0.79% for the rest of the transition. Given that we assume that the number of people per household stays constant after that same date, the growth rate in the number of households in the final steady state is $g_{NT} = 0.79\%$.

We also feed into the model a time series for the number of jobs per household. This is relevant for household earnings, which is the product of the real wage per job and the number of jobs per household. The number of jobs per household is set to 1.19 in the initial steady state, its value in 1975 U.S. data, and to 1.26 in the final steady state. The latter number is obtained by fitting a quadratic spline through the observed time series. The last observed value is 1.25 for 2007. For periods 1 through 32 along the transition path, we use the observed values for 1976-2007. The values along the transition path after period 32 are exponentially adjusting towards their steady-state value.

F   Addendum on Mobility

This addendum describes in detail how we calculate net migration rates between regions in the model and in the data. It also shows that the benchmark calibration does a good job matching the migration pattern and magnitude of the observed net migration rate.

F.1   Mobility in the Model

The following describes how we calculate net migration rates in the model.

Given two times $s < u$, consider the set of regions that started in productivity bin i at time s. Let g_s^{is} be a $N \times 1$ vector with zeros everywhere except in entry i where it is equal to the number of regions of type i at time s. Similarly, let gH_s^{is} be a $N \times 1$ vector with zeros everywhere except in entry i where it is equal to the total housing stock in regions of type i at time s. The total population of regions of type i at time s can be written as

$$\frac{gH_s^{is}}{h_{is}}, \tag{53}$$

the total housing stock divided by the housing consumption per household.

We seek to calculate the total population of these same regions at time u, conditional on survival. To that end, we need to i) calculate the total housing stocks of these regions at time u, for each productivity realization, conditional on survival and then ii) divide the housing stock by the housing consumption per

household to get the population. To calculate the housing stock of these regions at time u for each productivity realization, we use the recursion:

$$gH_{jt}^{is} = (1 - \delta) \sum_{k=1}^{N} G_{jt}^{kt-1} gH_{kt-1}^{is} + g\Delta_{jt}^{is},$$

$$\Leftrightarrow \quad gH_t^{is} = (1 - \delta)(G_t^{t-1})' gH_{t-1}^{is} + g\Delta_t^{is},$$

where G_t^{t-1} is the transition probability matrix between time $t - 1$ and time t. The vector $g\Delta_t^{is}$ is the total amount of housing constructed in these regions at time t, which we can write:

$$\Delta_t \times g_t^{is},$$

where Δ_t is a diagonal matrix whose diagonal element j is the amount of construction *per island* of type j at time t, and the $N \times 1$ vector g_t^{is} is the number of islands of type j at time t who started of type i at time s:

$$g_t^{is} \quad = \quad (G_t^{t-1})' g_{t-1}^i s = (G_t^s)' g_s^{is}.$$

Taken together, the recursion for gH_t^{is} is:

$$gH_t^{is} = (1 - \delta)(G_t^{t-1})' gH_{t-1}^{is} + \Delta_t (G_t^s)' g_s^{is}.$$

By induction, we find that:

$$gH_t^{is} = A_t^s gH_s^{is} + B_t^s g_s^{is},$$

for some matrices A_t^s and B_t^s that do not depend on i, and which solve the recursions:

$$A_t^s \quad = \quad (1 - \delta)(G_t^{t-1})' A_{t-1}^s$$
$$B_t^s \quad = \quad (1 - \delta)(G_t^{t-1})' B_{t-1}^s + \Delta_t (G_t^s)'.$$

The recursions are initiated at $A_s^s = I_d$ and $B_s^s = 0$. This delivers, for all regions of type i at time s, how the total housing stocks at time u is distributed across types. Given the housing consumption per household, we obtain that:

$$\frac{gH_{ju}^{is}}{h_{ju}}.$$

This calculation provides, for all regions of type i at time s, how the total population stocks at time u is distributed across types j. Of course, the total population at time u of regions who where of type i at time s is:

$$\sum_{j=1}^{N} \frac{gH_{ju}^{is}}{h_{ju}}.$$

For comparison with the data, we sort our continuum of regions into 25 real wage bins and compute net

migration rates for each 1995 wage bin as the difference between the population in that bin in 2000 (period 25 of the transition) and the population in 1995 (period 20 in the transition) divided by the population in 1995.

In the model, regions die at the exogenous rate λ and loose their entire housing stock. Their population is absorbed by the surviving regions. Since we are looking at a sample of regions that survives, this death of regions makes migration rates mechanically high. For instance, after 5 years, given a death rate of 1%, we should expect the surviving regions to grow (overall) by 5%. It suggests a calculation that nets it out this mechanical population growth in surviving regions by normalizing the total population to one in the surviving regions. Hence, in the calculation of the net migration rate, we normalize the 2000 population so that it is the same as the 1995 population. As explained below, we normalize the data in the same way.

F.2 Mobility in the Data

We use U.S. Census data for the in-migration and out-migration between 1995 and 2000 for each of 271 MSAs. Net migration is defined as the difference between in- and out-migration. We have those data for the population at large, as well as for the sub-population of young (25-39), single, college-educated. This latter group is interesting because mobility is more likely to be for productive reasons; hence, we focus on that group. We sort regions into 25 wage-per-job bins and compute net migration rates for each bin. As we did in the model, the calculation adjusts for growth in the population by scaling the population in 2000 so that it is the same as in 1995. The data show negative net migration to the tune of 7% per year in the lowest wage bins, which gradually increases, and turns into positive 1.5% net migration for the top-2 wage bins.

F.3 Comparing Model and Data

When we compare model to data, we find a similar pattern: out-migration from low-wage areas and in-migration into high wage areas. The top-10% lowest-wage regions see an out-migration of about 7% in both model and data. The top-10% highest-wage regions see an in-migration of about 1.5% in both model and data. Figure 10 shows the annual net migration rate in the benchmark model and in the data. The root mean squared error between the net migration rates of model and data across the 25 wage bins is 1.7%. We conclude that the model does a good job in capturing the pattern and the magnitude of migration rates.

Figure 10: Net Migration Rates

The model plots net migration rates in the benchmark model and in the data. Each bar represents the net migration (in-migration minus out-migration) between 1995 and 2000 of a group of regions with similar real wages. In particular, we form 25 groups of regions, sorted by their real wage from lowest to highest. The data is from the U.S. Census for young, single, college-educated persons. The model computes migration rates in the same way as in the data.

74