Online appendix of

# "Why Has House Price Dispersion Gone Up?"

Stijn Van Nieuwerburgh and Pierre-Olivier Weill

January 23, 2010

# 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:

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

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:

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

Plugging the approximation for $\Delta W$ back into (1), we find:

$$\Delta P = \frac{e\Delta A}{e\Delta A + \Delta e A} \times \frac{\Delta W}{1-\beta(1-\delta)} \Rightarrow \Delta P = \underbrace{\frac{\Delta A/A}{\Delta A/A + \Delta e/e}}_{\text{less than 1}} \times \frac{\Delta W}{1-\beta(1-\delta)}. \quad (2)$$

The coefficient in front of the present-value discount factor is the "correction" that must be applied to the observed wage differential, $\Delta W$, 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:

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

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:

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

where it is anticipated that the price of time $t$ consumption at time 0 is equal to $\beta^{t-1}$, 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:

$$\bar{W}_0 = \sum_{t\geq 1} \beta^{t-1} \int R_t(s^t) h_t(s^t) n_t(e, s^t)\, de\, \mu_t(ds^t) = \sum_{t\geq 1} \beta^{t-1} \int R_t(s^t) H_t(s^t) \mu_t(ds^t), \qquad (3)$$

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

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

$h_t^*(e)$ 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:

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

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 $\{A_1, A_2, \dots, A_N\}$, 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 $\mu_i$ 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 $n(e \,|\, s)$ 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:

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

subject to the family budget constraint:

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

and the constraint that:

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

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 $\{n(e \,|\, s), c(e, s), h(e, s)\}$ 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 1.** *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:*

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

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 $eA_i$, in marginal utility units $\lambda$. 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 1, one can show the same positive assortative matching result as in Proposition 2 for the present non-separable utility model: the sets $\varepsilon_i$ 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

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

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

$$\begin{aligned} \frac{R_{i+1} - R_i}{A_{i+1} - A_i} &= \frac{R_{i+1} - R_i}{\theta(R_{i+1}) - \theta(R_i)} \times \frac{\theta(R_{i+1}) - \theta(R_i)}{A_{i+1} - A_i} \\ &= \frac{R_{i+1} - R_i}{\theta(R_{i+1}) - \theta(R_i)} \times e_{i+1}, \end{aligned} \tag{4}$$

because

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

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 (4) that the slope of the rent is an increasing function of productivity.

#### B.3.5 Proof of Lemma 1

**Necessity** We start by proving the necessity of the condition in Lemma 1. 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:

$$\begin{aligned} v_i(e) &\equiv \lambda e A_i + u(c_i^*, h_i^*) - \lambda \left[c_i^* + R_i h_i^*\right] \\ \mathcal{I}(e) &\equiv \arg \max_{i \in \{1,2,\ldots,N\}} v_i(e) \\ \varepsilon_i &\equiv \{e \in [\underline{e}, \overline{e}] \,:\, i \in \mathcal{I}(e)\}. \end{aligned}$$

In words, $\mathcal{I}(e)$ is the set of locations that maximize the marginal value $v_i(e)$, and $\varepsilon_i$ 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 2.** *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 2, 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:

$$\begin{aligned} \hat{c}_i &= c_i + \frac{\eta}{1 - \eta F(\mathcal{E})} \left(\Delta Y - \Delta X\right) \\ &= c_i + \eta \left(\Delta Y - \Delta X\right) + o(\eta^2). \end{aligned}$$

The resulting change of utility for the family is:

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

where the second line follows from substituting in the formula for $\Delta X$ and $\Delta Y$. 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 $\eta$ is small enough.

**Sufficiency.** To prove sufficiency, consider a consumption and location plan satisfying the conditions of Lemma 1, 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:

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

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:

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

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:

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

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

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 (2009). 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 $\upsilon$ of households match assortatively with regions. We assume that the complementary fraction $1 - \upsilon$ 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

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

**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:

$$w_{ij} = h_j^m + (1+\zeta)e_i + (h_j - h_j^m) \text{ with probability } \upsilon, \tag{5}$$

$$w_{ij} = h_j^m + \zeta e_i + k_j + (h_j - h_j^m) \text{ with probability } 1-\upsilon, \tag{6}$$

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

$$\begin{aligned} V(w_{ij} - h_j^m) &= \upsilon\left((1+\zeta)^2 V(e) + V(h_j - h_j^m)\right) + (1-\upsilon)\left((1+\zeta^2)V(e) + V(h_j - h_j^m)\right) \\ &= V(h_j - h_j^m) + \left(1 + 2\upsilon\zeta + \zeta^2\right) V(e) \qquad (7) \\ &\geq \left(1 + 2\upsilon\zeta + \zeta^2\right) V(e) \geq V(e) \end{aligned}$$

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 (5) and (6), we obtain that average wage within region $i$ is

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

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

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

Together with (7), this last equation implies that the difference between the overall cross-sectional variance of the Mincerian residual and the cross-regional variance is:

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

We thus obtain the second upper bound:

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

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 $\upsilon = 0$, 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.

## B.5 Additional Moments of Wage Data

In this appendix, 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 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 (2009) 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.

Specifically, 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 $\theta$ 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 $\rho = 1$.

## B.6 Fixed Supply of Material

In order to get a better sense of the effect of the amount of construction material on the steady state allocation and prices, we conduct experiments where we either increase or lower the amount of housing material by 10%. We recalibrate the model so that we continue to match the same features of the 1975 wage distribution as in the benchmark model, the observed housing expenditure share, as well as the population in the highest-wage quintile. The resulting allocations and prices are *identical* to those in the benchmark model. The reason is that the calibration increases/lowers the number of permits $\pi_a$ by the same 10% so that the fraction of households in the 20% most productive regions (Q5) continues to match the 1975 data and changes the preference parameter $\kappa$ so that the housing expenditure share continues to match the data. This equivalence result follows directly from the homogeneity of the utility function for housing. The bottom line is that the results remain unchanged when the amount of material is changed, as long as the calibration matches the same six moments as in the benchmark model.

Alternatively, we can fix the parameters at their benchmark value, give up on the 1975 Q5 and housing expenditure share, and increase/lower the amount of housing material by 10%. If we lower $M$ without lowering the permits, fewer construction material is available and construction will only take place in the more productive regions. In steady state, the population must live there as well. Hence, we observe a counter-factually high fraction of people living in the most productive regions. A 10% lower M results in a Q5 of 70.62% versus 64.88% in the benchmark, a 10% increase. It results in average house prices of $61,905, 10% higher than the $55,720 in the benchmark. The CV of house prices in the initial steady state is 0.020 versus 0.022 in the benchmark, also 10% higher. If we then consider the transition with increasing productivity dispersion but with an amount of material that stays constant at its initial steady state level, then average house prices are $67,188 by 2007 and their CV is 0.486. These are similar increases as in our benchmark model. Increasing the construction material by 10% leads to the opposite changes in Q5 and house prices as lowering material by 10%.

# 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 $\mu_{bt}$ and standard deviation $\sigma_{bt}$.

In every subsequent period, an island either disappears with probability $\lambda$, or survives with probability $1-\lambda$, in which case it draws the new log productivity

$$a_t = a_{t-1} + \sigma_{at}\varepsilon_t, \tag{8}$$

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 $\mu_{bt-k}$ and variance $\sigma^2_{bt-k} + \sigma^2_{at-k+1} + \ldots + \sigma^2_{at}$. We are not aware of a closed form formula for this distribution, even in the stationary case where $\mu_{bt}$, $\sigma_{bt}$ and $\sigma_{at}$ 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:

$$\begin{aligned}
\text{avg}(a_t) &= \lambda\mu_{bt} + (1-\lambda)\text{avg}(a_{t-1}) && (9)\\
\text{avg}(a_t^2) &= \lambda\left(\mu_{bt}^2 + \sigma_b^2\right) + (1-\lambda)\left(\text{avg}(a_{t-1}^2) + \sigma_{at}^2\right), && (10)\\
\text{disp}(a_t)^2 &= \text{avg}(a_t^2) - \text{avg}(a_t)^2, && (11)
\end{aligned}$$

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 $t = T$:

$$\text{disp}(a_t) = \text{disp}(a_0) + \frac{t}{T}\left(\text{disp}(a_T) - \text{disp}(a_0)\right). \tag{12}$$

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:

$$\text{avg}(a_t) \simeq \log(\text{avg}(A_t)) - \frac{\text{disp}(a_t)^2}{2}. \tag{13}$$

Then given a constant time path for $\text{avg}(A_t)$, and given the time path (12) for dispersion, equation (13) provides the time path for $\text{avg}(a_t)$. Since, on the other hand, $\text{avg}(a_t)$ evolves according to equation (9), this pins down the time path of $\mu_{bt}$:

$$\mu_{bt} = \frac{\text{avg}(a_t) - (1-\lambda)\text{avg}(a_{t-1})}{\lambda}, \tag{14}$$

the mean log productivity of newly created islands. In our numerical computation, approximation (13) 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 (12) of dispersion can be calculated with equations (10) and (11):

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

given that we know $\text{disp}(a_t)$ from (12), $\text{avg}(a_t)$ from (13), $\mu_{bt}$ from (14), 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:

$$a_t = \text{avg}(a_{t-1}) + \rho_t\left(a_{t-1} - \text{avg}(a_{t-1})\right) + \sigma_{at}\varepsilon_t. \tag{15}$$

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 (9), while the second moment is

$$\mathrm{avg}(a_t^2) \;=\; \lambda\left(\mu_{bt}^2+\sigma_b^2\right)+(1-\lambda)\left(\mathrm{avg}(a_{t-1})^2+\rho_t^2\mathrm{disp}(a_{t-1})^2+\sigma_{at}^2\right), \tag{16}$$

As in the previous paragraph, we impose a deterministic linear increase in the cross-sectional productivity dispersion. As before, this imposes to choose $\mu_{bt}$ according to formula (14).

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

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

We then pick the time path of persistence so as to satisfy equation (16), i.e.:

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

**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 $\mu_{bt}$ and variance $\sigma_{bt}^2$: 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):

$$F(x) = \frac{1}{2}\left(\frac{x}{X_e}\right)^{k_e} \text{ if } x \leq X_e \tag{17}$$

$$= \frac{1}{2} + \frac{1}{2}\left[1 - \left(\frac{x}{X_e}\right)^{-k_e}\right] \text{ if } x > X_e. \tag{18}$$

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 $\alpha$ 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 (17) and (18), 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 1. 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 $\sigma_e$, the tail parameter $k_e$ must be

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

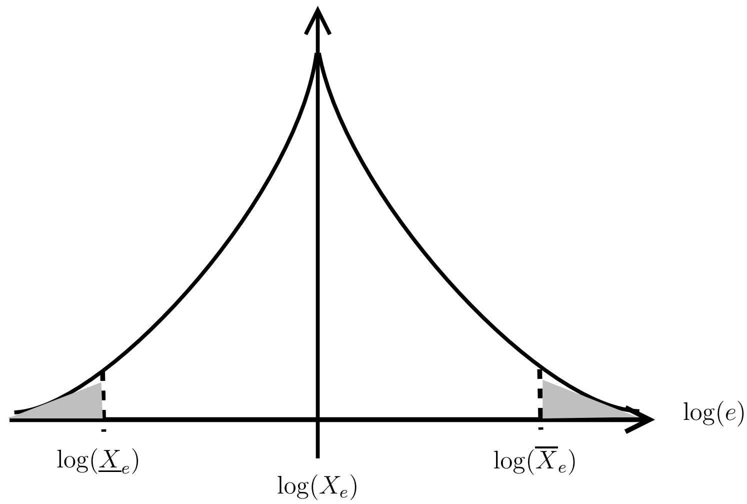


Figure 1: In level, the distribution of ability is a double Pareto. As a results the logarithm of ability is distributed according to a double exponential that is symmetric around $X_e$. To keep a compact support, we need to truncate the distribution in both tails. The figure shows the log of our truncation 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.

## C.2 Mobility in the Model

This appendix describes in details how we calculate net migration rates between regions 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

$$\frac{gH_s^{is}}{h_{is}}, \tag{19}$$

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:

$$\begin{aligned} gH^{is}_{jt} &= (1-\delta)\sum_{k=1}^{N} G^{kt-1}_{jt} gH^{is}_{kt-1} + g\Delta^{is}_{jt}, \\ \Leftrightarrow \quad gH^{is}_{t} &= (1-\delta)(G^{t-1}_{t})' gH^{is}_{t-1} + g\Delta^{is}_{t}, \end{aligned}$$

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

$$\Delta_t \times g^{is}_t,$$

where $\Delta_t$ 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 $g^{is}_t$ is the number of islands of type $j$ at time $t$ who started of type $i$ at time $s$:

$$g^{is}_t \quad = \quad (G^{t-1}_t)' g^{i}_{t-1}s = (G^{s}_t)' g^{is}_s.$$

Taken together, the recursion for $gH^{is}_t$ is:

$$gH^{is}_t = (1-\delta)(G^{t-1}_t)' gH^{is}_{t-1} + \Delta_t (G^{s}_t)' g^{is}_s.$$

By induction, we find that:

$$gH^{is}_t = A^s_t gH^{is}_s + B^s_t g^{is}_s,$$

for some matrices $A^s_t$ and $B^s_t$ that do not depend on $i$, and which solve the recursions:

$$\begin{aligned} A^s_t &= (1-\delta)(G^{t-1}_t)' A^s_{t-1} \\ B^s_t &= (1-\delta)(G^{t-1}_t)' B^s_{t-1} + \Delta_t (G^s_t)'. \end{aligned}$$

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:

$$\frac{gH^{is}_{ju}}{h_{ju}}.$$

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 $\lambda$ 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.

# 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 $40^{th}$ 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 $40^{th}$ 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 $45^{th}$ percentile until 1993 for all regions, at which point it switches to the $40^{th}$ percentile for all regions. In 2000, it switches to the $50^{th}$ percentile in 40 regions, and in 2005 it switches back to the $40^{th}$ 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 $(\Phi^{-1}(0.45) - \Phi^{-1}(0.40)) * c$, 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 $\sigma_r$, 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.

# References

Marigee Bacolod, Bernardo S. Blum, and William C. Strange. Skills in the city. *Journal of Urban Economics*, 65:136 – 153, 2009.

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

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.

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

Chirok Han and Peter C. B. Phillips. GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory*, 2009. Forthcoming.

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.

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.

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

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