Geographic Mobility and Redistribution

Daniele Coen-Pirani*

September 19, 2020

Abstract

I study the effect of progressive taxation on internal migration and welfare using a quantitative dynamic model of geographic mobility. The model, which is analytically tractable, predicts that a more progressive tax-transfer scheme reduces internal migration rates. The magnitude of this relationship is consistent with reduced-form evidence for OECD countries. The internal migration channel contributes to significantly lower optimal tax progressivity relative to the one-location version of the economy. The optimal sequence of tax progressivity along the transition features a relatively high degree of tax progressivity early on, followed by a declining path of tax progressivity over time.

Keywords: Geographic mobility, progressive taxation, redistribution, welfare, transitional dynamics.

*University of Pittsburgh. E-mail: coen@pitt.edu. First version: April 2017.
1 Introduction

A large literature studies the effect of taxes and tax progressivity on labor supply, both positively (Prescott, 2004) and normatively (Diamond and Saez, 2011). The common argument is that high or more progressive taxes reduce the incentives to participate in the workforce or work longer hours. In this paper I study the effect of progressive national (in the U.S. Federal) taxes and transfers, on the location - rather than the quantity - of a household’s labor supply. The basic idea is that, if migration is costly, a more progressive national tax-transfer system reduces individuals’ incentives to migrate internally to locations where they are idiosyncratically more productive. I embed this mechanism in a new tractable model of internal migration. The estimated model is used to quantify the empirical relevance of this distortion and to study the welfare implications of tax reform.

The new model of internal migration is rich, as it allows for both ex-ante and ex-post heterogeneity in terms of age, moving costs, and labor income, while delivering closed-form solutions for individual-level migration and labor supply behavior. I use these to show that a higher degree of income tax progressivity tends to reduce migration rates. A higher degree of tax progressivity reduces the returns from migration by taxing away a portion of the earnings growth associated with moving to a new location. This reduction in the after-tax benefit of moving reduces the incentives to migrate in the presence of a positive moving cost. The magnitude of the migration cost parameter, therefore, has important implications for the policy counterfactuals. The moving cost and other key parameters of the model are identified by the frequency of interstate migration by age, and by the evolution of the cross-sectional mean and variance of earnings over the life-cycle. These moments are computed using data from the American Community Survey.

I test the quantitative predictions of the model on moments that are not targeted in the estimation and that are computed using two different datasets. First, I compare its predictions with estimates of the growth in earnings associated with interstate migration. These estimated gains are based on household-level panel data from the Survey of Income and Program Participation. The model correctly predicts that earnings gains from internal
migration are proportionately larger for more educated households and tend to decline with age. Second, I evaluate the model’s ability to account for the magnitude of the cross-sectional relationship between newly constructed measures of tax progressivity and internal migration rates for OECD countries. In the data tax progressivity is negatively associated with geographic mobility. The model provides an accurate quantitative account of this relationship.

I characterize analytically the welfare function and the trade-offs associated with higher tax progressivity using a simplified version of the model. In principle, the internal migration channel has ambiguous implications for optimal tax progressivity. A more progressive tax system distorts incentives to move towards locations in which individuals are more productive. However, migration is associated with increased levels of earnings inequality and higher tax progressivity mitigates its effect on consumption inequality.

The quantitative analysis with the full model shows that optimal tax progressivity is smaller than in an equivalent model with one location and no scope for mobility. Keeping constant the average tax rate, the average income-weighted marginal tax rate is about 8 percentage points lower in the benchmark model than in an otherwise similar one-location economy. Taking into account the transition from the initial to the final steady state contributes to make the optimal degree of tax progressivity significantly higher than if the planner cared only about steady state welfare. This is because at the time of the reform the productivity level of old agents is fixed. Consequently, a relatively high level of tax progressivity is less distortionary in the short-run than in the long-run when the migration choices of all agents are based on the new policy regime.

This intuition suggests that an optimal time-varying tax progressivity policy would involve a relatively high degree of tax progressivity in the early phase of the tax reform - when old agent’s productivity levels are fixed - followed by a declining path towards the new steady state. I exploit the analytical tractability of the model to compute this policy and verify that this is indeed the case. The welfare gain of optimal time-varying policies is 1.09% percent in consumption equivalent units, relative to 0.71% for a tax reform involving a constant level
of tax progressivity.

This paper adds to the recent literature on the effects of tax progressivity and subsidies on human capital investment. Papers in this literature focus on investment in either formal schooling or training. Absent the geographic mobility element, my model would be similar to those in Benabou (2002) and Heathcote et al (2017, 2020). Their functional form assumptions allow me to derive closed-form solutions to the agents’ dynamic migration problem. Differently from papers in this literature, I focus on internal migration, a costly investment that can be observed in many publicly available datasets (Schultz, 1961).

Gentry and Hubbard (2004) explore empirically the effect of the level and progressivity of taxation on job-to-job transitions in the U.S. and find evidence that a more progressive tax system is associated with less job mobility. This is consistent with my empirical evidence across OECD countries. Hassler et al (2005) provide suggestive evidence that countries with more generous unemployment insurance systems are characterized by lower internal geographic mobility. They use a politico-economic model of endogenous unemployment insurance and migration to account for this fact, but, differently from my approach, their focus is not quantitative.

The importance of transitional dynamics considerations for the calculation of optimal (Ramsey) tax progressivity has been emphasized, in the context of different models, by a number of authors, e.g. Domeji and Heathcote (2004), Bakis et al (2015) and Krueger and Ludwig (2016) among others. My paper departs from these by explicitly considering the optimal path of tax progressivity over the transition. In a related paper, Dyrda and Pedroni (2018) study optimal time-varying linear capital and labor income taxes in the context of the Aiyagari model using a numerical approach similar to mine.

A number of recent papers study optimal spatial taxation in the context of static equilibrium models. Albouy (2009), Eeckout and Guner (2015) and Colas and Hutchinson (2020) point out that Federal taxation of nominal income might distort the spatial allocation of

workers across locations. The argument advanced in these papers hinges crucially on the fact that locations are ex-ante heterogeneous in terms of productivity and amenities. In this paper, instead, I focus on the dynamic effect of national tax progressivity on gross internal migration flows of workers across local labor markets, instead of its effect on the geographic distribution of population across heterogeneous locations. In my model agents make dynamic mobility choices and progressive Federal taxation reduces the net gain from moving even if all locations are ex-ante identical. The tax distortion I study is the effect of progressive taxation on workers’ incentives to locate in the labor market where they are idiosyncratically more productive.\footnote{Other related contributions on the topic of location-specific taxes are Ales and Sleet (2017), Fajgelbaum and Gaubert (2020) and Albov et al (2019).}

The paper is also related to the recent papers by Caliendo et al (2019) and Artuc et al (2010) on spatial and sectoral labor reallocation following trade shocks. These papers follow Kennan and Walker (2011) and generate gross flows of labor across sectors and regions through extreme-value location-specific shocks, which are usually interpreted as preference or moving cost shocks. While my model also relies on extreme-value shocks for analytical convenience, it interprets them as permanent shocks to households’ productivity growth in each location. Consequently, in the model as in the data, geographic mobility is primarily motivated by job-related reasons, such as prospect of income growth.\footnote{The paper is also related to the growing migration literature in many areas of economics, such as urban (Ahlfeldt et al (2015), Diamond, 2016), labor (Auray et al, 2017, Gemici 2017, Monras, 2017), macro (Coen-Pirani, 2010, Bayer and Juessen, 2012, Lkagvasuren, 2014, Kaplan and Schulhofer-Wohl, 2017), and development (Desmet and Rossi-Hansberg, 2017 and Lagakos et al (2017)).}

Finally, a number of recent papers focus on individuals’ incentives to migrate across jurisdictions in order escape from relatively high-tax locations (Akcigit et al (2016) and Moretti and Wilson (2017)) or take advantage of more generous state and local welfare systems (Gelbach, 2004).\footnote{See also Fajgelbaum et al (2015)’s contribution arguing that heterogeneity in taxes across states is a cause of misallocation.} The tax distortion I consider, instead, is due to the national – in the U.S. Federal – tax system instead of tax differences across political sub-units.

The rest of the paper is organized as follows. Section\footnote{2} introduces the model and Section
characterizes its analytical solution. The description of the model’s estimation and data fit is contained in Section 4. Section 5 tests some of the quantitative implications of the model using moments that were not targeted in the estimation. Optimal policies and welfare are discussed in Section 6. Finally, Section 7 concludes and discusses future work. All proofs, details about the data, and discussions of extensions of the model are contained in the Appendix.

2 Model

Geography, Technology, and Policy The economy is comprised of a finite number of local labor markets indexed by \( k = 1, 2, \ldots, K \). Local economies are assumed to be ex-ante identical in terms of their aggregate characteristics, such as amenities and productivity. Thus, the reasons for migration from one labor market to another are purely idiosyncratic. Each local economy produces an homogenous good using the same constant returns to scale production function. The only production input is represented by efficiency units of labor whose marginal product is normalized to one without loss of generality. Time is discrete and infinite, starting at \( t = 1 \). For the rest of the analysis, the only source of exogenous variation across time in the economy pertains to the aggregate policy variables \( \{\tau_t, \lambda_t, G_t\}_{t=1}^\infty \), where \( G_t \) denotes the government’s public good provision while \( \tau_t \) and \( \lambda_t \) are the parameters of the tax-transfer scheme that maps a household’s market income \( y \) into its after-tax and transfer income \( \tilde{y}_t \):

\[
\tilde{y}_t = \lambda_t y^{1-\tau_t},
\]

Appendix E.2 discusses the case of ex-ante heterogeneous locations and shows that the latter collapses to the economy with homogeneous locations under specific restrictions on how local prices and amenities vary in the cross-section of locations.

The presence of a public good does not affect the positive properties of the model because it enters additively in utility (see equation 3 below). However, its presence affects the socially optimal choice of tax progressivity \( \{\tau_t\}_{t=1}^\infty \), as shown by Heathcote et al (2017).
In this tax-transfer system, which has been employed by Benabou (2002) and Heathcote et al (2017) among others, the household pays net taxes $T_t(y) = y - \tilde{y}_t$ and the parameter $\tau_t$ indexes tax progressivity while $\lambda_t$ indexes the overall level of taxation. For example, when $\tau_t = 0$, then marginal and average taxes are the same ($T'_t(y) = T_t(y)/y$) and the tax system features a proportional income tax with rate $1 - \lambda_t$. By contrast, when $\tau_t = 1$ the government taxes away the entire marginal unit of income earned so $T'_t(y) = 1$.

Demographic Structure, Timing, and Migration Choices The economy is populated by a continuum of finitely-lived households of measure one. Households live from age $a = 1$ to $\bar{a}$. A measure $1/\bar{a}$ of households is born in each location at the beginning of each period. A household belongs to one of $\tau$ possible types, indexed by $r = 1, 2, \ldots, \tau$. Household type is determined at birth and does not change over time. There is a measure $\omega_r$ of households of type $r$, with $\sum_r \omega_r = 1$. The initial distribution of households by age and type across locations is symmetric, so that a measure $\omega_r/(\bar{a}K)$ of households of type $r$ and age $a$ resides in each location at $t = 1$. A household begins life at age $a = 1$ with initial productivity $\exp(\alpha_{1,r} + z)$, where $z$ is drawn from the density $f(z|1, r)$. The latter is exogenous, assumed to be the same in all locations and time periods, and allowed to vary by type $r$. Household productivity evolves exogenously with aging, according to the deterministic function $\alpha_{a,r}$ and endogenously, through migration.

At each age $a$ the timing of actions is as follows. At the beginning of each period, a household characterized by labor productivity $\exp(z + \alpha_{a,r})$ resides in one of the economy’s $K$ locations. Then, migration occurs. Some households are exogenously relocated while others are in the position of making a moving choice. Migration of either type involves a utility cost $\kappa_r$. The household then supplies labor $\ell$ and earns labor income $y = \ell \exp(\alpha_{a,r} + z')$, where $z'$ denotes the household’s idiosyncratic level of $z$ after migration. Then, redistribution takes place leaving the household with disposable income $\tilde{y}$, given by equation (1), which is assumed to be entirely consumed, $c = \lambda_t y^{1-\tau_t}$.

The assumption that an agent cannot save or borrow reflects a form of asset market incompleteness. Differently from Heathcote et al (2017), this assumption is not without loss of generality due to the explicit
Migration incentives are driven by the location-dependent evolution of household productivity. Specifically, a household with current idiosyncratic log productivity \( z \) observes, at the beginning of period \( t \), the evolution of its efficiency in all locations \( k \). Its log productivity in location \( k \) is assumed to evolve as follows during period \( t \):

\[
z_k' = \rho_r z + \eta_r \varepsilon_k \quad \text{for all } k = 1, \ldots, K,
\]

where the parameter \( \rho_r \leq 1 \) determines the persistence of productivity over time and the parameter \( \eta_r < 1 \) governs the importance of location-specific shocks \( \varepsilon \equiv \{ \varepsilon_k \}_{k=1}^K \). The shocks \( \varepsilon_k \) are assumed to be distributed according to a type-1 extreme value distribution (Rust, 1987). They are independently drawn both over time for a given household as well as in the cross-section of households. I allow shocks drawn from the household’s current home location, denoted by \( h \), to have a different distribution from shocks drawn elsewhere due, for example, to its superior information about the home location’s labor market. Specifically, the distribution of \( \varepsilon_h \) is assumed to be a type 1 extreme-value with unit scale parameter and location parameter \( \ln \delta_r \). Instead, the distribution of \( \varepsilon_k \) for \( k \neq h \) is assumed to be a type 1 extreme-value with unit scale parameter and location parameter equal to zero. A value \( \delta_r > 1 \) implies that, on average, a shock drawn from the home location is larger than shocks drawn from the other locations.

I distinguish between endogenous and exogenous migration in the following simple way.

---

8 I don’t include a household or age-specific subindex to the variables in equation (2) not to complicate the notation. It should be kept in mind, however, that both \( z_k \) and \( \varepsilon_k \) vary idiosyncratically by household and over time for each household.

9 It is straightforward to allow for some correlation among shocks, but their common component would not matter for migration decisions. The assumption that shocks are distributed as extreme-value is mostly for analytical convenience. It is not necessary to show that tax progressivity has a negative effect on geographic mobility.

10 An interpretation is that a household of type \( r \) makes \( \delta_r \) draws in its current location and only one in each of the away locations. The maximum of \( \delta_r \) independent extreme-value draws is still distributed as extreme-value, so the distribution of shocks from the home location for a type \( r \) agent is an extreme-value with mean \( \ln \delta_r + \gamma \), where \( \gamma \) is Euler’s number. A draw from each of the other locations has a mean equal to \( \gamma \).
With probability \( \theta_r \), upon observing the vector \( \varepsilon \), a household is able to select the location that provides the highest present discounted value of utility going forward. With probability \( 1 - \theta_r \), instead, the household is exogenously relocated to one of the other \( K - 1 \) locations with equal probability. If the household moves to a location \( k \), either by choice or exogenously, it pays a utility cost \( \kappa_r \) and its productivity at the beginning of next period becomes \( \exp (\alpha_{a+1,r} + z'_k) \).

**Households’ Preferences and Recursive Formulation**  Agents maximize their expected discounted utility net of moving costs. The discount factor is denoted by the parameter \( \beta < 1 \). The household’s static utility function is:

\[
u(c, \ell, G) = \ln c - \zeta^{-1} \ell^\zeta + \chi \ln G, \tag{3}\]

where \( \ell \) denotes labor supply, the parameter \( \zeta \) determines the Frisch elasticity of labor supply (which equals \( (1 - \zeta)^{-1} \)), and \( \chi \) is the utility weight of the public good \( G \), provided by the national government. The household’s optimal labor supply maximizes equation (3) subject to the budget constraint \( c = \lambda_t (\ell \exp (\alpha_{a,r} + z'))^{1-\tau_t} \). It is straightforward to show that it is given by\(^{11}\)

\[
\ell^*_t = (1 - \tau_t)^{\frac{1}{\zeta}}. \tag{4}
\]

Thus, a higher degree of tax progressivity reduces labor supply, as in Benabou (2002) and Heathcote et al (2017). Replacing \( \ell^*_t \) back into (3), the static indirect utility function takes the log-linear form:

\[
u^*_t (a, z'; r) = \overline{u}_t (a; r) + (1 - \tau_t) z', \tag{5}\]

\(^{11}\)The logarithmic specification of utility in (3) simplifies the dynamic programming problem faced by each agent. While convenient, it has some well-known limitations (see e.g. Heathcote et al (2017)). For example, due to offsetting income and substitution effects, the labor supply choice is independent of an agent’s productivity and of the overall level of taxation \( \lambda_t \). In Appendix A.1 I use a simplified version of the model to discuss the implications of a CRRA utility function for the effect of tax progressivity on migration.
where:

\[
\overline{u}_t^* (a; r) \equiv \ln \lambda_t + (1 - \tau_t) \ln \ell_t^* + (1 - \tau_t) \alpha a, r - \zeta (\ell_t^*)^\zeta + \chi \ln G_t. \tag{6}
\]

Each household faces a dynamic optimization problem that involves the choice of location. Let \( V_t (a, z', k; r) \) denote the household’s conditional value function. This is the maximum remaining lifetime utility attainable by a household of age \( a \), with idiosyncratic productivity \( z' \), who is located in \( k \) after migration:

\[
V_t (a, z', k; r) = \begin{cases} 
  u_t^* (a, z'; r) & \text{if } a = \bar{a} \\
  u_t^* (a, z'; r) + \beta E_{\varepsilon'} [P_{t+1} (a + 1, z', \varepsilon', k; r)] & \text{if } a < \bar{a},
\end{cases} \tag{7}
\]

where \( E_{\varepsilon'} [.] \) denotes the expectation taken with respect to the distribution of location-specific \( t + 1 \) shocks, \( \varepsilon' \). The function \( P_{t+1} \) on the right-hand side of (7) represents the unconditional value function in \( t + 1 \). The unconditional value function, denoted by \( P_t (a, z, h, \varepsilon; r) \), represents the lifetime utility attainable by a household of age \( a \) who, before migration, is located in \( h \), has idiosyncratic productivity \( z \), and observes the vector of location-specific shocks \( \varepsilon \):

\[
P_t (a, z, \varepsilon, h; r) = \theta_r \max_k \left\{ V_t (a, \rho_r z + \eta_r \varepsilon_k, k; r) - I_{h k} \right\} + \\
+ \frac{1 - \theta_r}{K - 1} \sum_{k \neq h} \left\{ V_t (a, \rho_r z + \eta_r \varepsilon_k, k; r) - \kappa_r \right\} \tag{8}
\]

where \( I_{h k} \) is an indicator function that takes a value of 1 if and only if \( k \neq h \) and zero otherwise. Notice that the term multiplying \( \theta_r \) in the first row of equation (8) represents the value of endogenous migration, while the term multiplying \( (1 - \theta_r) \) in the second row represents the expected value of exogenous migration. Let \( M_t (a, \varepsilon, h, k; r) \) denote the migration decision rule for households that are not exogenously relocated. In writing it in this way I anticipate the fact that it does not depend on \( z \) (see Proposition 2). Specifically, \( M_t (a, \varepsilon, h, k; r) = 1 \) if a household of age \( a \) with a vector of shocks \( \varepsilon \) moves voluntarily at the beginning of time \( t \) from its current location \( h \) to location \( k \), and \( M_t (a, \varepsilon, h, k; r) = 0 \) if the household chooses not move to \( k \). Notice that, by definition, \( \sum_{k=1}^K M_t (a, \varepsilon, h, k; r) = 1 \).
Competitive Equilibrium  Given the sequences \( \{\tau_t, G_t\}_{t=1}^{\infty} \) and the pre-migration distributions \( f(z|a,r) \) of household productivity by age and type at the beginning of \( t = 1 \), a competitive equilibrium for this economy is comprised of: a sequence \( \{\lambda_t\}_{t=1}^{\infty} \); sequences of post-migration densities of household labor productivity by age and type \( \{f^p_t(z'|a,r)\}_{t=1}^{\infty} \); sequences of unconditional and conditional value functions \( \{P_t(a,z,\varepsilon,h;r), V_t(a,z',k;r)\}_{t=1}^{\infty} \); sequences of decision rules \( \{M_t(a,\varepsilon, h, k;r)\}_{t=1}^{\infty} \) for geographic mobility; sequences of decision rules for labor supply \( \{\ell_t^r\}_{t=1}^{\infty} \) such that:

1) The value functions \( \{P_t(a,z,\varepsilon,h;r), V_t(a,z',k;r)\}_{t=1}^{\infty} \) and decision rules \( \{M_t(a,\varepsilon, h, k;r)\}_{t=1}^{\infty} \) represent the solution to the agent’s dynamic optimization problem (7)-(8), with \( \ell_t^r \) taking the form in (4).

2) The value of \( \lambda_t \) is consistent with the government’s balanced budget for all \( t \geq 1 \):

\[
G_t = \frac{1}{\pi} \sum_{a=1}^{\pi} \sum_{r=1}^{\rho} \omega_r \int \left( \ell_t^r \exp(\alpha_{a,r} + z') - \lambda_t (\ell_t^r \exp(\alpha_{a,r} + z'))^{1-\tau_t} \right) f^p_t(z'|a,r) dz'. \tag{9}
\]

3) The densities \( \{f^p_t(z'|a,r)\}_{t=1}^{\infty} \) are generated by the transition equation for productivity (2) combined with the optimal decision rules \( \{M_t(a,\varepsilon, h, k;r)\}_{t=1}^{\infty} \), starting from the initial densities \( f(z|a,r) \).

The definition of competitive equilibrium reflects the model’s symmetry in that all locations are characterized by the same distribution of population and productivity by age and type. Notice that the general equilibrium dimension of this economy stems from the government’s budget constraint (9). At each point in time \( t \), two of the variables in the triple \( (\tau_t, G_t, \lambda_t) \) are given, and the remaining one has to satisfy equation (9). A stationary equilibrium of the model refers to a situation in which \( (\tau, G) \) and all the endogenous objects that comprise a competitive equilibrium are constant over time. I will focus on a stationary equilibrium when estimating the model’s parameters.

Discussion of modeling choices  Before characterizing the households’ decision rules, it is useful to briefly discuss the environment outlined above. The model describes house-
holds’ geographic relocation in response to idiosyncratic productivity shocks. This formulation captures, realistically, the idea that an important portion of moves across local labor markets is due to job-related reasons. For example, the Current Population Survey asks respondents to select a reason for geographic moves. In 2000-2007 data, “new job or job transfer” is the single most-frequently selected answer among head of households ages 25–59 who moved across state lines in the previous year. A new job and other job-related reasons jointly account for about 42 percent of all interstate moves. In addition to being realistic, the presence of exogenous relocation allows the model to match the fact that the magnitude of income gains associated with migration is decreasing over the life cycle (see Section 5.1).

The model environment describes gross flows of workers across homogeneous locations. The assumption of homogeneous locations is a convenient simplification that allows me to focus on gross migration flows driven by idiosyncratic productivity shocks. This approach complements much of the existing literature on internal migration (e.g. Diamond, 2016), which emphasizes idiosyncratic preference shocks as a source of gross migration flows. In the latter, localities are ex-ante different in terms of aggregate productivity and amenities but all workers with the same observable skills (e.g. education) earn the same wage if they work in the same labor market. In reality, observable skills such as those associated with schooling explain only a relatively small fraction of wage differences among individuals in the same local labor market. In my model instead, all labor markets are ex-ante identical from an aggregate perspective, but there is a substantial heterogeneity in wages among workers with the same observable skills and age operating within a labor market.

While the model above might also describe job-to-job transitions within a local labor market, I favor the migration interpretation because the household suffers the moving cost

---

12 In the U.S. economy, gross migration flows are very large relative to net flows. For example, in the American Community Survey sample I use to estimate the model’s parameters, the average U.S. state experiences combined inflows and outflows of nine households for each household it gains or loses in net terms in the course of one year. Moreover, in Coen-Pirani (2010) I show that workers who enter (gross inflows) and exit (gross outflows) the same local labor market are observationally very similar in terms of age, industry and occupation.

13 In Appendix E.2 I show that the model with homogeneous locations may be derived from one with ex-ante heterogeneous locations in which housing rents fully capitalize differences in local wages, taxes and amenities.
when it relocates. The internal migration literature (e.g. Kennan and Walker, 2011) has estimated sizeable moving costs and the findings of my paper are consistent with this evidence (see Section 4.5).

One of the paper’s innovations is modelling the dynamics of location-specific productivity at the household level. In this specification, productivity gains in a location persist over time independently of whether the agent remains in that location or migrates to another one. In a way, locations provide idiosyncratic opportunities for a household to become more productive by moving there. This increase in productivity, which might depreciate over time depending on the persistence parameter $\rho_r$, is reflected in the household’s human capital, i.e. it is portable across locations. An advantage of this specification is that each household is characterized by only one state variable, instead of $K$ location-specific levels of productivity.

3 Analytical Derivation

In this section I characterize analytically the value function, the geographic mobility decision rule, and the dynamic evolution of household productivity. The results of this section provide valuable intuition on the main mechanisms of the model and form the basis for its structural estimation in Section 4.

3.1 Value Function and Decision Rules

Propositions 1 and 2 present, respectively, the value function that solves the Bellman equation (7)-(8) and the decision rule for migration. These closed-form solutions allow me to characterize analytically, in Proposition 3, how tax progressivity affects migration rates.

Proposition 1 (Value function) Given policies $\{\tau_t, G_t, \lambda_t\}_{t=1}^{\infty}$, the unique value function that solves the dynamic programming problem (7)-(8) takes the following form:

$$V_t (a, z', k; r) = v_t^0 (a; r) + v_t^1 (a; r) z' \text{ for all } k,$$

(10)
where for \( a < \bar{a} \):

\[
v_t^1 (a; r) = 1 - \tau_t + \sum_{k=1}^{\bar{a}-a} (\beta \rho_r)^k (1 - \tau_{t+k}),
\]

while \( v_t^1 (\bar{a}; r) = 1 - \tau_t \). The age-dependent term \( v_t^0 (a; r) \) is defined recursively by:

\[
v_t^0 (a; r) = \bar{\pi}_t^r (a; r) + \beta v_{t+1}^0 (a+1; r) - \beta (1 - \theta_r) \kappa_r + \beta v_{t+1}^1 (a+1; r) \eta_r (\gamma - \theta_r \ln (p_{t+1} (a+1; r) / \delta_r)),
\]

starting from \( v_0^0 (\bar{a}; r) = \bar{\pi}_r^r (\bar{a}; r) \). The term \( p_{t+1} (a+1; r) \) in equation (12) represents the probability of choosing to remain in the same location at age \( a+1 \) and is formally defined in equation (14) below.

According to Proposition 1, the value function is linear in the household’s log productivity.\(^{14}\) The utility value of higher productivity depends on the age and time-dependent coefficient \( v_t^1 (a; r) \). Intuitively, the latter declines with the household’s age and with the degree of tax progressivity that it faces in its remaining life span. Using the expression for the value function, it is possible to characterize the geographic decision rule of households in this economy.\(^{15}\)

**Proposition 2 (Geographic mobility)** A household of age \( a \) with shocks \( \varepsilon \) chooses to move from location \( h \) to a different location \( k \neq h \) \((M_t (a, \varepsilon, h, k; r) = 1)\) if and only if:

\[
\varepsilon_k - \varepsilon_h > \frac{\kappa_r}{v_t^1 (a; r) \eta_r}
\]

and \( \varepsilon_k = \max_{l \neq h} \varepsilon_l \).

\(^{14}\)The existence of a closed-form solution is due to the logarithmic specification of static utility and not to the extreme-value distribution of the \( \varepsilon_k \) shocks. A different distributional assumption on the \( \varepsilon_k \)’s would be reflected in the term \( v_t^0 (a; r) \).

\(^{15}\)Notice that the parameter \( \lambda \), which determines the level of taxes, has no effect on moving choices because of offsetting income and substitution effects associated with logarithmic utility. An analogous result holds, for example, in Heathcote et al (2017, Proposition 2) where the skill investment depends only on tax progressivity and not on the level of taxes.
Notice that, absent moving costs \((\kappa_r = 0)\), the household would simply pick the location with the highest growth rate of labor income, \(\varepsilon_k\). A positive moving cost implies that the household might choose to remain in the same location even if the growth rate of its productivity in that location is smaller than in the rest of the economy. Conditional on moving to a different location, a household chooses the location with the highest growth rate of its productivity.

When equation (13) does not hold a household chooses not to migrate. The probability that this happens can be computed using the distribution of the random variable \(\max_{t \neq h} \varepsilon_t - \varepsilon_h\). The assumption that the \(\varepsilon\) shocks are independently distributed as type 1 extreme-value random variables implies that \(\max_{t \neq h} \varepsilon_t - \varepsilon_h\) is logistic with location parameter \(\ln((K - 1) \delta_r^{-1})\) and unit scale parameter. It follows that the probability that a household of age \(a\) chooses to remain in the same location (when given the opportunity to choose) is:

\[
p_t(a; r) = \frac{1}{1 + (K - 1) \delta_r^{-1} \exp\left(-\kappa_r \eta_r v_t^1(a; r)\right)}.
\] (14)

Recall that a household might migrate for exogenous reasons at the rate \(1 - \theta_r\). Thus, the overall probability of moving away from a location is \(1 - \theta_r p_t(a; r)\). The following proposition presents some important implications of the theory for the rate of geographic mobility.

**Proposition 3 (Migration probabilities)** Assume that migration costs are positive and finite, \(\kappa_r \in (0, +\infty)\). Then:

1. A household’s probability of migration declines with its age \(a\).
2. Given a household’s age \(a\), its migration probability declines with the degree of tax progressivity \(\{\tau_{t+k}\}_{k=0}^{\pi-a}\) that it faces in its remaining working life.
3. The impact of tax progressivity on a household’s probability of migration first rises and then falls with the moving cost parameter \(\kappa_r\).
4. Exogenous relocations account for a growing fraction of all moves as an individual ages.

The intuition for these results is relatively straightforward. First, migration rates decline with age because the horizon over which the household can take advantage of the benefits of
moving shrinks as it gets older. Second, the effect of tax progressivity on mobility is negative because the moving cost $\kappa_r$ is positive and is not tax deductible. Thus, while the income gain from moving is subject to progressive taxation, the utility gain from not moving is not taxed. This asymmetry is at the core of the negative effect of $\tau_{t+k}$ on mobility.\footnote{Thus, the assumption that the mobility cost is a utility cost as opposed to an income cost is important. For example, one could specify the migration cost as a loss of a fraction of human capital, in which case $\tau_{t+k}$ would have no effect on mobility choices because redistribution would affect symmetrically both the choice of staying and that of moving.} Third, the effect of $\tau_{t+k}$ on mobility depends on the moving cost $\kappa_r$ in a non-monotonic fashion. With costless mobility ($\kappa_r = 0$) the household always moves to the location that provides the highest before-tax income growth, independently of the degree of tax progressivity. As the moving cost increases, tax progressivity becomes distortionary, but eventually, if the moving cost is very large, there is no mobility and tax progressivity is again non-distortionary. Thus, the estimate of the moving cost parameter bears important implications for evaluating the effect of tax policy on migration and welfare. Last, as an agent grows older, it becomes more likely that a geographic move is due to exogenous reasons instead of being a choice because the probability of exogenous relocation is age-independent while ageing reduces the incentives to move. This result plays an important role in accounting for the fact that in the data the difference in earnings growth between movers and stayers tends to decline with age (see Section 5.1).

Notice that, while in this model progressive taxation has the potential to increase social welfare by providing insurance against shocks and initial conditions, it does not increase the propensity to migrate because agents make moving choices after observing the vector $\varepsilon$. In principle, higher tax progressivity has the potential to increase migration by providing insurance against unanticipated negative income shocks. For example, following a geographic move, a household might experience a temporary decline in income or its income process may become more volatile.\footnote{For example, Manovskii (2002) shows that, in an incomplete markets framework, higher tax progressivity can increase human capital accumulation by reducing the cost of on-the-job training.} In Appendix C.1 I have tested these two hypothesis by analyzing the empirical properties of the income process upon migration and found no empirical support...
for either hypothesis. The panel data used to conduct this analysis is from the Survey of Income and Program Participation. It is described in more detail in Section 5.1 where it is used to measure the income gains associated with interstate migration.\footnote{My empirical analysis cannot entirely rule out the possibility of a positive effect of tax progressivity on migration. For example, even if the income process does not become more volatile upon migration, individuals might have better information about local than foreign shocks at the time of making moving choices. In Appendix E.3, I formalize this idea, using an Epstein-Zin utility function. I show that if the coefficient of relative risk-aversion is larger than one, this version of the model features a positive effect of tax progressivity on migration (in addition to the negative effect on which I focus).}

### 3.2 Households’ Idiosyncratic Productivity Growth

In this section I characterize the evolution of the idiosyncratic productivity process for movers and stayers. This characterization provides the basis for comparing the model’s predictions with its empirical counterparts in Section 5.1. I also rely heavily on the results of Proposition 4 to compute analytically the moments targeted by the model’s estimation and (numerically) the elements of the government’s budget constraint (equation (9)) and the various components of the welfare function.

Over time a household’s labor income changes because of the age-dependent evolution of its productivity, as captured by $\alpha_{a,r}$, and because of idiosyncratic shocks to productivity and migration. If the household is able to make a migration choice, its idiosyncratic productivity evolves as:

$$z' = \rho_r z + \eta_r g_t(\varepsilon, a; r),$$

where:

$$g_t(\varepsilon, a; r) = \sum_{k=1}^{K} M_t(a, \varepsilon, h, k; r) \varepsilon_k,$$

and $g_t(\varepsilon, a; r)$ does not depend on $h$ because of the symmetry of the model.\footnote{In other words, productivity growth does not depend on the specific “identity” $h$ of the location where the household resides at the beginning of the period because the aggregate characteristics of each location are the same.} If the household is exogenously relocated to $k \neq h$, instead, its productivity evolves according to equation (2). The following proposition characterizes the distribution of $g_t(\varepsilon, a; r)$. 

\begin{equation}
\end{equation}
Proposition 4 (Innovation to productivity growth with endogenous mobility) The innovation $g_t(\varepsilon, a; r)$ to the productivity process, conditional on choosing to stay in the same location, is distributed according to a type 1 extreme-value distribution with mean:

$$E_{\varepsilon} \left[ g_t(\varepsilon, a; r) \mid M_t(a, \varepsilon, h, h; r) = 1 \right] = \gamma + \ln \delta_r - \ln p_t(a; r),$$

(16)

and variance:

$$\text{VAR}_{\varepsilon} \left[ g_t(\varepsilon, a; r) \mid M_t(a, \varepsilon, h, h; r) = 1 \right] = \pi^2/6. \quad (17)$$

The innovation $g_t(\varepsilon, a; r)$ to the productivity process, conditional on choosing to migrate, is distributed according to a type 1 extreme-value distribution with mean:

$$E_{\varepsilon} \left[ g_t(\varepsilon, a; r) \mid M_t(a, \varepsilon, h, h; r) = 0 \right] = \gamma + \ln \delta_r - \ln p_t(a; r) + \frac{\kappa_r}{\eta, \nu^2_t(a; r)},$$

(18)

and the same variance as in (17). The parameter $\gamma$ in these equations denotes Euler’s constant.

Notice the difference between the evolution of productivity with exogenous and endogenous mobility. In the former case there is no selection, and the innovation to productivity experienced by an agent who is exogenously relocated follows the density of shocks drawn from one of the “away” locations. By assumption, the latter is distributed as a type 1 extreme-value distribution with mean $\gamma$ and variance $\pi^2/6$. For agents who can choose whether to migrate or not, movers are self-selected based on the evolution of productivity. In particular, the expressions in equations (16) and (18) reflect selection through migration. The conditional mean in equation (16) reflects selection by agents who choose to stay put. For the latter to be the optimal choice, stayers must experience higher productivity growth, on average, than the set of households that are exogenously reallocated. This is indeed the case since it is always the case that $p_t(a; r) < 1 \leq \delta_r$. In particular, the lower the ex-ante chance of staying put – the lower $p_t(a; r)$ – the higher the average conditional productivity growth.

\footnote{See Nakosteen and Zimmer (1980) for an early application of Heckman’s correction for selectivity bias to the income equations of internal migrants and non-migrants with normally-distributed disturbances.}
for stayers must be. For agents who choose to move out of a location, instead, average productivity growth in (18) exceeds, on average, productivity growth conditional on staying by an amount that reflects the cost of moving. This is, again, a reflection of selection. For costly migration to be preferable to staying put, agents must experience, on average, an additional gain in productivity. We can summarize the discussion above in the following corollary.

**Corollary 1** Average productivity growth is largest for households who choose to relocate and smallest for households who are forced to relocate. Households who choose to stay in the same location experience an intermediate level of productivity growth.

4 Empirical Implementation

In order to estimate the model’s parameters I focus on the stationary equilibrium of the model with constant tax progressivity parameter $\tau$. The following functional forms are assumed. The exogenous distribution of household productivity $f(z|1,r)$ at age one is taken to be lognormal with parameters $(\mu_r, \sigma_r^2)$. The exogenous component of the evolution of household productivity is a quadratic function of age:

$$\alpha_{a,r} = \alpha_{0,r}a + \alpha_{1,r}a^2. \quad (19)$$

The parameter vector is then:

$$\{\beta, \zeta, \bar{\alpha}, \bar{\tau}, K, \tau, \chi, (\omega_r, \kappa_r, \theta_r, \bar{\alpha}_{0,r}, \bar{\alpha}_{1,r}, \eta_r, \delta_r, \mu_r, \sigma_r^2, \rho_r)_{r=1,\tau}\}.$$

The empirical strategy has two parts. First, a number of parameters are set a-priori, including $\tau$. Second, the remaining parameters are estimated by the Generalized Method

---

21 A low probability of staying reflects a relatively small moving costs. Thus, agents who choose to stay must experience a high income growth in the home location relative to what they would have earned if they had chosen to move.
of Moments (Hansen, 1982). I postpone setting $\chi$ until Section 6 as it only plays a role in welfare analysis.

### 4.1 Parameters Set A-Priori

The frequency of the model is one year. The years of working life are $\pi = 35$, from age 25 to age 59 included. The annual discount factor $\beta = 0.97$. The labor supply parameter $\zeta = 3$, implying a Frisch elasticity of 0.5. These three numbers are the same as those selected by Heathcote et al (2020). The number of locations $K$ is set to 51 to capture mobility across U.S. states and the District of Columbia. The number of household types is set to $r = 2$: households whose head has less than a college degree ($r = 1$) and households whose head has a college degree or more ($r = 2$). The fraction of $r = 1$ types in the American Community Survey 2000-07 data used to estimate the model is $\omega_1 = 0.6751$. Finally, I set the parameters $\rho_r = 1$ a-priori, so that the idiosyncratic productivity process is a random walk. This is consistent with numerous empirical studies of the income process (e.g. Meghir and Pistaferri, 2011).

### 4.2 Measure of Tax Progressivity

The tax progressivity parameter $\tau$ is estimated using data on households’ market income and income post-Federal taxes and transfers, following Heathcote et al (2017). Recall that, from equation (1), $1 - \tau$ represents the elasticity of income post-Federal taxes and transfers to market income. This suggests an empirical specification of the form:

$$\ln \tilde{y}_{pt} = a_t + b \ln y_{pt} + u_{pt},$$

(20)

---

*Notowidigdo (2011), Lkhagvasuren (2014) and Amior (2020) study differences in geographic mobility between workers of different education level. Empirically, there is a monotonic relationship between age-specific migration rates and schooling.

*I have tried to estimate $\rho_r$ directly. In this case the estimation procedure converges to 1, so I have imposed it a-priori.*
where $a_t$ represents year fixed-effects, $b$ is the elasticity of post-redistribution income to market income (or $1 - \tau$), and $u_{pt}$ is an error term, assumed to be uncorrelated with $\ln y_{pt}$.

The subscript $p$ denotes percentiles of the household income distribution. The data used to estimate the parameter $b$ are, in fact, percentiles of the distribution of post and pre-government household income from the Congressional Budget Office (2011, Table A-1). Data is available for percentiles $p = 20, 40, 60, 81, 91, 96, 99$. The CBO market income measure includes all cash income (taxable and tax-exempt), taxes paid by businesses and imputed to households such as corporate taxes and the employer’s share of payroll taxes, and benefits, such as employer-paid health insurance premiums. The after-government income includes cash transfer payments (for example, unemployment insurance and welfare) and estimates of the value of in-kind benefits (Medicare, Medicaid, Children’s Health Insurance Program, Supplemental Nutrition Assistance Program). It subtracts federal individual and corporate income taxes, payroll taxes and excise taxes.$^{24}$

For consistency with the other data used in the paper, I restrict attention to the period 2000-2007. The estimated value of $b$ is equal to 0.808 with a standard error 0.011 and $R^2 = 0.995$. This leads to a value of $\tau$ equal to 0.192, close to Heathcote et al (2017)’s estimate of 0.18 using the PSID for the period 2000–2006 and to Bakis et al (2015)’s estimate of 0.17 using March CPS data (1979–2009) and the NBER tax simulator. The Congressional Budget Office data are available at an yearly frequency starting in 1979. Interestingly, there is no evidence for a trend in $\tau$ over this period.$^{25}$ In Section 5.2 I discuss a different approach to measuring $\tau$ that can be used for all OECD countries. This approach derives $\tau$ from a comparison of a country’s Gini coefficient of market income and the Gini coefficient of income after taxes and transfers. Following this alternative procedure yields an average estimate of $\tau$ equal to 0.21 for the U.S. in 2000–2007. Consistently with the evidence from the CBO

---

$^{24}$State and local income taxes are not included in the CBO calculations. These taxes are significantly less progressive than Federal taxes so their exclusion is unlikely to significantly affect the estimate of $\tau$. My benchmark estimate is, in fact, very close to Heathcote et al (2017)’s figure (see the main text). They include state and local income taxes in their calculations based on the PSID.

$^{25}$Adding a linear interaction term between year and the coefficient on $\ln y_{pt}$ in equation (20) and using the entire sample period 1979-2007, yields an estimate of $b$ equal to 0.817 and a statistically insignificant estimate of $-0.0003$ on the interaction term.
data, also these estimates of $\tau$ are remarkably stable over time.\footnote{The estimates are in the range 0.19–0.23 for the period 1974-2012. In 1974, $\tau$ is equal to 0.19 while in 2012 it is equal to 0.21. Ferriere and Navarro (2018) use U.S. federal tax (but not transfer) data to estimate the tax progressivity parameter $\tau$ over time starting in the early 20th century. Their estimate suggests that tax progressivity in the U.S. has declined in the early-mid 1980s and has been relatively stable since then. Guner et al (2014) also estimate versions of the tax function (20) using tax (but not transfers) data and estimate a smaller value of $b$.}

\section*{4.3 Estimation and Identification of the Remaining Parameters}

The remaining parameters are estimated by GMM. There are a total of 16 parameters to be estimated:

$$\phi_r = \{\mu_r, \sigma_r^2, \kappa_r, \theta_r, \delta_r, \eta_r, \overline{c}_{0,r}, \overline{c}_{1,r}\}_{r=1,2}.$$ \hfill (21)

The parameter vector $\phi_r$ is estimated separately by household type $r$. For each type, the estimation procedure targets the following 3$\bar{a}$ moments: 1) the average migration rate by age ($\bar{a}$ conditions); 2) the cross-sectional mean of the distribution of log household earnings by age ($\bar{a}$ conditions); 3) the cross-sectional variance of the distribution of log household earnings by age ($\bar{a}$ conditions). The analytical results of Section 3 allow me to compute all moment conditions analytically (see Appendix B.2), so the estimation procedure does not rely on the use of simulations. The GMM estimator of $\phi_r$ solves the following problem:

$$\hat{\phi}_r = \arg \min_{\phi_r} m'_r (\phi_r) W^{-1}_r m (\phi_r),$$

where $m_r (\phi_r)$ denote the moment conditions and $W^{-1}_r$ is Hansen (1982)'s optimal weighting matrix.

These moment conditions are sufficient to identify the parameter vector. The geographic mobility data identify the parameters $\kappa_r, \theta_r, \delta_r$. In order to understand the role played by each of these parameters, consider the migration rate $1 - \theta_r p(a; r)$ at age $a$, where $p(a; r)$ is formally defined in equation (14).\footnote{Appendix B.1 contains a formal illustration of this identification argument based on a version of the model with $\bar{a} = 3$. This is the minimum number of ages necessary to identify the three migration cost parameters.} As a household grows older, its incentives to voluntarily

\[22\]
migrate decline because migration is costly \((\kappa_r > 0)\) and its remaining time horizon \((\bar{a} - a)\) shrinks. Therefore, the parameter \(\theta_r\) is identified by the migration rate at older ages, when migration mostly occurs for exogenous reasons (see Proposition 3 point 4). To understand the different roles played by \(\delta_r\) and \(\kappa_r\), notice that if moving costs were zero \((\kappa_r = 0)\), the migration rate would be a constant independent of age, because \(p(a; r)\) would not depend on \(a\) (see equation 14). Thus, to account for the declining pattern of migration as a function of age, it is necessary that \(\kappa_r > 0\). Given the discount factor \(\beta\), the size of the moving cost relative to the importance of idiosyncratic shocks, \(\kappa_r/\eta_r\), regulates the dependence of migration rates on age. Thus, \(\kappa_r/\eta_r\) is identified by the observed difference in average migration rates at younger and middle ages \((a = 1\) and, approximately, \(a = 20\), based on Figure 1). Given \(\kappa_r/\eta_r\), the parameter \(\delta_r\) pins down the average migration rate of an agent through her lifecycle.

The age profile of mean wages for each education type identifies the parameters \((\pi_{0,r}, \pi_{1,r})\). The observed mean and variance of log income at age 25 \((a = 1)\) identify, respectively, the mean \(\mu_r\) and variance \(\sigma_r^2\) of the initial productivity distribution. As a cohort ages, the cross-sectional variance of log income in the model increases, due to the presence of idiosyncratic shocks. The slope of this relationship in the data identifies the parameter \(\eta_r\), which regulates the importance of the \(\varepsilon\) shocks for households’ productivity growth.

### 4.4 Sample Selection

The data set used to estimate the model is from the American Community Survey (ACS), years 2000-2007 (Ruggles et al (2017)). The main advantage of the ACS data is the relative large sample which allows one to accurately measure migration rates by age. The main disadvantage is that the ACS data is purely cross-sectional and does not allow one to measure wage growth for either migrants or non-migrants. The ACS data sample consists

---

28 Notice that, since the productivity scale is arbitrary, I have normalized, without loss of generality, average income to equal one at age 25 for \(r = 1\) agents. This normalization pins down \(\mu_1\).

29 For this purpose I use longitudinal data from the Survey of Income and Program Participation, as discussed in Section 5.1.
of households whose head is 25–59 years old, and is not institutionalized or in school. I define a household’s labor income as the sum of the wage, salary and business income of the household’s members. I drop observations in the bottom 10 percent of yearly earnings. The geographic unit of observation is a U.S. state and each model period represents a year. Households’ labor income and mobility data are purged of year, state, and sex effects by running regressions of each of these variables on year, state and sex dummies and using residuals to construct the moments of interest. The latter approach is consistent with the model’s assumption that locations are ex-ante identical.

4.5 Estimated Parameters and Model’s Fit for Targeted Moments

Table 1 reports the estimates of the model’s parameters and their standard errors. The mobility cost $\kappa_r$, converted into consumption-equivalent units ($1 - \exp(-\kappa_r)$), corresponds to 90 and 99 percent of yearly consumption for $r = 1, 2$ agents respectively, a relatively large number but comparable to the figure estimated by Kennan and Walker (2011). The parameters $\theta_r$ are such that the probabilities of an exogenous move are 1.1 and 1.8 percent per year for households of type $r = 1, 2$, respectively. Therefore, exogenous relocation accounts for 56 and 48 percent of all moves for these two types. These figures are consistent with households’ responses to the Current Population Survey’s question about reasons for interstate migration cited in Section 2. College educated agents are characterized by a higher moving cost ($\kappa_2 > \kappa_1$) and by a lower frequency of offers in the home location ($\delta_2 < \delta_1$). As discussed in the identification section above and shown formally in Appendix B.1 the lower frequency of home offers allows the model to account for the higher mobility of college-educated workers. Their higher moving cost gives rise to the faster decline between young and middle-ages in the geographic mobility rate observed for this group. Figure 1 shows the age pattern of migration. At younger ages the migration rate of college-educated labor is

Amior (2020) studies the gap in migration rates across skill groups in a search model of migration. While his model and data are different from mine, he finds that more educated workers are more mobile because the wage returns from interstate migration are larger for this group. His estimated migration costs are also larger for more educated workers.
Parameters set a priori

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J$ number of locations</td>
<td>51</td>
<td>U.S. states</td>
</tr>
<tr>
<td>$\xi$ Frisch elasticity parameter</td>
<td>3</td>
<td>Heathcote et al (2020)</td>
</tr>
<tr>
<td>$\beta$ yearly discount factor</td>
<td>.97</td>
<td>Heathcote et al (2020)</td>
</tr>
<tr>
<td>$\tau$ degree of tax progressivity</td>
<td>.192</td>
<td>CBO</td>
</tr>
<tr>
<td>$\pi$ duration of working life</td>
<td>35</td>
<td>Heathcote et al (2020)</td>
</tr>
<tr>
<td>$\omega_1$ measure of $r = 1$ (less than college) agents</td>
<td>0.6751</td>
<td>ACS, 2000-2007</td>
</tr>
<tr>
<td>$\omega_2$ measure of $r = 2$ (college and above) agents</td>
<td>0.3249</td>
<td>ACS, 2000-2007</td>
</tr>
<tr>
<td>$\rho_1$ autocorrelation of productivity, $r = 1$</td>
<td>1.00</td>
<td>n.a.</td>
</tr>
<tr>
<td>$\rho_2$ autocorrelation of productivity, $r = 2$</td>
<td>1.00</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Estimated Parameters for Type $r = 1$ Agents (Less than College Degree)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1$ mobility cost parameter</td>
<td>2.3095</td>
<td>0.1095</td>
</tr>
<tr>
<td>$\delta_1$ frequency of local offers parameter</td>
<td>98.501</td>
<td>14.7234</td>
</tr>
<tr>
<td>$\eta_1$ importance of idiosyncratic shocks</td>
<td>0.0444</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\alpha_{0,1}$ age profile of earnings (linear)</td>
<td>-0.2000</td>
<td>0.0063</td>
</tr>
<tr>
<td>$\alpha_{1,1}$ age profile of earnings (quadratic)</td>
<td>-0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\theta_1$ one minus probability of exogenous move</td>
<td>0.9893</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\mu_1$ mean of initial log productivity</td>
<td>-0.1558</td>
<td>0.0026</td>
</tr>
<tr>
<td>$\sigma_1^2$ variance of initial log productivity</td>
<td>0.3588</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

Estimated Parameters for Type $r = 2$ Agents (College Degree and Above)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_2$ mobility cost parameter</td>
<td>5.7771</td>
<td>0.1212</td>
</tr>
<tr>
<td>$\delta_2$ frequency of local offers parameter</td>
<td>2.2817</td>
<td>0.3074</td>
</tr>
<tr>
<td>$\eta_2$ importance of idiosyncratic shocks</td>
<td>0.0594</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\alpha_{0,2}$ age profile of earnings (linear)</td>
<td>-0.0556</td>
<td>0.0080</td>
</tr>
<tr>
<td>$\alpha_{1,2}$ age profile of earnings (quadratic)</td>
<td>-0.0006</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\theta_2$ one minus probability of exogenous move</td>
<td>0.9828</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\mu_2$ mean of initial log productivity</td>
<td>0.2074</td>
<td>0.0039</td>
</tr>
<tr>
<td>$\sigma_2^2$ variance of initial log productivity</td>
<td>0.2925</td>
<td>0.0028</td>
</tr>
</tbody>
</table>

Table 1: Summary of model parameters and standard errors.
Figure 1: Average migration rate by age in the model (line) and in the data (circles) for type $r = 1, 2$ agents.

about twice as large as the migration rate of those in the less skilled group. Such difference tends to disappear at older ages. Figure 2 represents the fit of the model with respect to the (log) earnings moments for each type of agents. Notice that the model captures very well the evolution of the cross-sectional pattern of the mean and the variance of earnings across the ages.

5 Model’s Fit for Non-Targeted Moments

In this section I discuss the fit of the estimated model using two datasets that were not used to estimate its parameters. First, I ask whether the model can account for the association between migration and earnings growth at the household level. Second, I compare its prediction for the relationship between tax progressivity and aggregate migration rates with empirical evidence based on cross-country data.
Figure 2: Cross-sectional mean and variance of log earnings by age in the model (line) and in the data (circles) for type $r = 1, 2$ agents.
5.1 Earnings Growth and Migration

The model’s parameters have been estimated using cross-sectional moments on household earnings and migration patterns by age. In this section, I evaluate its implications for the difference in earnings growth between movers and stayers. This requires using longitudinal household-level data with a sufficiently large sample size because, on average, yearly migration rates are of the order of two percent per year. The data I use to measure gains in labor income associated with inter-state migration are from the Survey of Income and Program Participation (SIPP User Guide, 2001). The SIPP is a nationally representative longitudinal household survey. Starting with the 1996 redesign, every three or four years the SIPP selects a new panel of households and interviews them every four months for three or four years. In each interview (wave), the household provides information on a large number of variables, including earnings and state of residence, in each of the prior four months. I use the 1996, 2001, 2004, and 2008 panels of the SIPP.

I focus on households with heads aged between 25 and 59 years old. For each household $i$, I use the SIPP data to construct a panel of monthly observations on household earnings and inter-state moves. Given the short length of each panel, conditional on ever moving across state lines, the great majority of households moves only once. Let $M_i$ be an indicator that equals one if household $i$ ever moves across state lines and zero otherwise. The earnings gain associated with an inter-state move are estimated from the following regression equation:

$$\ln y_{i,a,s,m} = M_i \times Post_{i,m} \times (\xi_1 + \xi_2 (a - 25)) + \xi_3 \ln n_{i,m}^{18-64} + \xi_i + \xi_a + \xi_s + \xi_m + \epsilon_{i,a,s,m},$$

\[ (22) \]

31 I am aware of only a few empirical studies in this area. Yankow (2003, Table 8) uses panel data from the 1979 National Longitudinal Study of Youth to compare wage gains of internal migrants relative to stayers and finds wage gain gaps of around 9 percentage points, three to four years after the migration event for full-time workers with more than a high school degree. The corresponding gap estimated by Yankow for full-time workers with a high school degree or less is about 2 percentage points.

32 I thank Jose’ Mustre-del-Rio for pointing out the potential usefulness of the SIPP data to analyze the dynamics of gains in earnings upon migration. The advantage of the SIPP is that it has a much larger sample size than other longitudinal datasets such as the Panel Study of Income Dynamics or the National Longitudinal Study of Youth. Differently from the Current Population Survey, the SIPP attempts to locate original sample members even if they move to a new address.

33 Appendix C.1 provides further details on sample selection criteria.
where $y_{i,a,s,m}$ denotes household earnings and $a$, $s$ and $m$ are indices for household age, state-of-residence and month-year of the data. The variable $Post_{i,m}$ is an indicator that equals one if the household is observed in month-year $m$ after an interstate move and zero else. The regression controls for the (log) number of households members between 18 and 64 ($n_{i,m}^{18-64}$), and for fixed effects for household ($\zeta_i$), age ($\zeta_a$), state-of-residence ($\zeta_s$), and month-year of the data ($\zeta_m$).

The effect of geographic mobility on earnings is captured by the term $\xi_1 + \xi_2 (a - 25)$, which represents the within-household proportional change in earnings after an interstate move. Notice that this specification allows the effect of geographic mobility on earnings to vary by age. I estimate the regression separately for households headed by an individual with a college degree and for households headed by an individual with less than a college degree. The estimates of these parameters are reported in Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Less than college</th>
<th>College and above</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_1$</td>
<td>0.105**</td>
<td>0.174***</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>-0.008**</td>
<td>-0.006*</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1,209,340</td>
<td>652,242</td>
</tr>
<tr>
<td>Number of households</td>
<td>36,154</td>
<td>18,146</td>
</tr>
</tbody>
</table>

Table 2: Estimates of the impact of interstate migration on household earnings. Standard errors clustered at the household-level in parenthesis. Data source: SIPP 1996, 2001, 2004, 2008 panels. Regressions control for household, age, state, and month-year fixed effects. * denotes a p-value <0.10; ** <0.05 and *** <0.01.

Figure 3 provides a visual representation of these estimates by plotting the predicted gap in labor income growth between movers and stayers as a function of age in the data $(\hat{\xi}_1 + \hat{\xi}_2 (a - 25))$ and in the model.

The estimates in Table 2 are qualitatively and quantitatively consistent with two basic predictions of the structural model. First, earning gains of migration are proportionally larger
Figure 3: Average gap in labor income growth between movers and stayers for type $r = 1, 2$ agents as a function of age. Data generated by the model are denoted by the thick blue lines. Estimates based on the regression equation (22) are denoted by the dashed line (−−). 95% confidence intervals based on the regression estimates are also reported.

for more educated workers. Second, for both groups these gains tend to decline with age and may even be negative at older ages. The model is able to account for the second prediction due to the possibility of exogenous relocation. Migrants who choose to move are positively selected on earnings growth and the extent of positive selection increases in age because older households have a shorter remaining horizon to make up for the fixed cost of moving. Therefore, absent exogenous relocation, the model would incorrectly predict that the earnings gains of migration increased with age. In the model, however, exogenous relocation accounts for a growing fraction of all moves as an individual ages and incentives to migrate voluntarily decline (Proposition 3, part 4). Since those who are relocated exogenously experience smaller average earnings growth than workers who choose not to migrate (see Corollary 1), this effect accounts for the observation that the earnings gains of migration decline with age.

Formally, inspection of equations (16) and (18) in Proposition 4 reveals that the difference between earnings growth of movers and stayers (by choice) is given by the term $\kappa_r/v_t(a;r)$ which is increasing in age.

34Formally, inspection of equations (16) and (18) in Proposition 4 reveals that the difference between earnings growth of movers and stayers (by choice) is given by the term $\kappa_r/v_t(a;r)$ which is increasing in age.
5.2 Tax Progressivity and Internal Migration Across-Countries

The model predicts that economies characterized by a higher degree of tax progressivity $\tau$ will display smaller internal migration rates. In this section I test this prediction quantitatively using OECD data on tax progressivity and internal migration that was not used in the estimation of the model’s parameters. Data on internal migration comes from the OECD (2000, 2005, 2013) and refers to the period 1980–2010. The territorial unit in each country is the first administrative tier of sub-national governments, such as a state for the U.S. and a Lander for Germany.\footnote{These are called Territorial Level 2 (TL2) units by the OECD. I focus on the 21 countries with more than five TL2s because internal migration in countries with very few TL2s (Belgium, Denmark, Finland, Slovakia, Slovenia, Iceland) is likely to be qualitatively different from the rest of the sample. See Appendix C.2.2 for the full list of countries.} Since there is no readily-available estimate of the parameter $\tau$ across countries, I measure the latter using OECD data on the distribution of household income before ($y$) and after ($\bar{y}$) taxes and transfers for the period 1980–2013. Specifically, for each country and year I compute the value of $\tau$ that rationalizes the ratio between the Gini coefficient of income after taxes and transfers and the Gini coefficient of market income.\footnote{The income data is measured at the household level for the population 18–65 and is adjusted for household size with an equivalence scale. The procedure to compute $\tau$ is described in detail in Appendix C.2 Table A.3 in this appendix presents summary statistics on the estimates of tax progressivity across countries.} Since measures of tax progressivity and of internal migration may be subject to measurement error, I average each of these variables for a given country over the sample period.

Figure 4 represents a scatter plot of average migration rates and estimates of tax progressivity, $\tau$. The slope of the regression line is $-5.15$ with a standard error equal to $2.06$.\footnote{Appendix C.2 contains details on the regression results discussed in this section and on a number of robustness exercises.} This implies that an increase in tax progressivity by a cross-sectional standard deviation (0.07) is predicted to reduce a country’s average migration rate by about 0.35 percentage points, or slightly less than one-half of the cross-sectional standard deviation of migration rates (see Table A.3).

Interestingly, the partial correlation between migration and tax progressivity is robust to controlling for differences in the level of taxes across countries. For example, including
Figure 4: Scatterplot of country-level migration rates against country-level tax progressivity $\tau$. The regression line has slope $-5.15$ (s.e. 2.06) with an $R^2 = 0.21$.

average personal income taxes (net of cash transfers) relative to the average wage in the regression for internal migration yields a coefficient on $\tau$ equal to $-5.58$ (s.e. 1.38). The measure of average taxes enters with a positive and marginally statistically significant sign in this regression.\footnote{See the regression results in Appendix Table A.4. Taking into account the cross-sectional standard deviations of these variables and the size of the regression coefficients, the effect of the level of taxation on migration is, in absolute terms, about two-thirds of the effect of tax progressivity. Appendix E.1 shows that, with a general CRRA utility function, the effect of the level of taxes on the incentive to migrate might be either positive or negative.}

While the evidence presented in this section is suggestive of a negative effect of tax progressivity on internal migration, caution has to be used in interpreting the correlation in Figure 4. Countries may vary along other, potentially unobserved, dimensions in addition to the degree of progressivity of their tax system. For example, countries in which geographic mobility is low for cultural reasons might choose a more progressive tax system to
provide additional insurance against idiosyncratic income shocks (Hassler et al (2005)).\footnote{A related, but different, interpretation of the data is that in a country in which the internal migration rate is low for reasons other than taxes, a utilitarian government might choose a more progressive tax system because it does not need to worry about distorting geographic mobility choices. Such interpretation is consistent with the hypothesis advanced in this paper.}

Some of these concerns might be attenuated by exploiting the panel dimension of the data. Unfortunately, the available data is an unbalanced panel as neither observations on internal migration nor on Gini coefficients are collected on a yearly basis. There are only 32 country-year pairs of observations for the post-1980 period for 21 countries. A panel regression of internal migration rates on measured tax progressivity with year and country fixed effects yields an estimate of $-7.51$ (s.e. $6.94$). While this is not statistically different from zero, the magnitude of the point estimate is comparable with the cross-sectional one. Ultimately, questions of causality are difficult to settle without quasi-experimental variation in $\tau$ across countries and over time.

In what follows I sidestep the causality question and, instead, ask whether the quantitative model can account for this cross-country evidence. To do so, I compute the model-implied steady state migration rate for various degrees of tax progressivity $\tau$.\footnote{The steady state assumption is the counterpart of the fact that the cross-country empirical evidence is based on comparisons of country-average internal migration rates and tax progressivity over a period of about 30 years.} Figure 5 plots the steady state migration rate for the economy as a whole (solid line) and for each type of agent as the parameter $\tau$ varies. The benchmark calibration is denoted by the point “US” in the figure. For comparison, the figure also represents (dashed line) the cross-country regression represented in Figure 4. Notice that I have adjusted its intercept so that it passes through the point “US” in order to facilitate comparison with the model. The quantitative prediction of the model is consistent with the cross-country evidence. Specifically, an increase in $\tau$ from its benchmark value of 0.192 to 0.262 - a cross-country standard deviation - is predicted to reduce internal mobility by 0.41 percentage points in the model and by 0.35 percentage points in the data.
Figure 5: Aggregate and type-specific migration rates across steady states for different values of the tax progressivity parameter $\tau$. The dot denotes the model-predicted average migration rate for the U.S. in the benchmark calibration.

6 Welfare Analysis

In this section I use the model to analyze how a utilitarian planner would optimally set tax progressivity. I begin by considering a simplified version of the model to show analytically how internal migration affects the planner’s trade-offs concerning tax progressivity. I then consider the full quantitative model and compute optimal policies numerically.

6.1 Policy and Welfare: Analytical Results

A simplified version of the model allows me to study analytically the welfare implications of varying the degree of tax progressivity. In the simplified version of the model households live forever, $(\bar{u} \to +\infty)$ and the exogenous productivity term $\alpha_{\omega, r}$ is set to zero, without loss of generality. In addition, I restrict attention to economies with one type of agent ($\omega_1 = 1$); no exogenous mobility ($\theta_r = 1$); i.i.d. shocks ($\rho = 0$); two locations ($K = 2$); and symmetric shock distributions ($\delta_r = 1$). The lack of persistence in $z$ implies that, without loss of
generality, I can focus on a tax policy that is constant over time.\footnote{To discuss the welfare implications of fiscal policy in this economy, I assume that the social welfare function is utilitarian. The latter is equal to the sum of the lifetime utilities of all agents, taking moving costs into account.\footnote{Notice that I multiply the social welfare function by \((1 - \beta)\) to simplify the notation.}}

To discuss the welfare implications of fiscal policy in this economy, I assume that the social welfare function is utilitarian. The latter is equal to the sum of the lifetime utilities of all agents, taking moving costs into account.\footnote{Recall that given these two policy variables, the third one, \(\lambda\), is implied by the government’s budget constraint, equation (9).}

\[ W = (1 - \beta) \int P(\varepsilon) g(\varepsilon) d\varepsilon, \quad (23) \]

where \(P(\varepsilon)\) is the unconditional value function (see equation (8)).

Under these assumptions, \(W\) can be computed analytically as a function of the two policy variables available to the planner, i.e., tax progressivity and the supply of the public good.\footnote{Recall that given these two policy variables, the third one, \(\lambda\), is implied by the government’s budget constraint, equation (9).}

**Proposition 5 (Social welfare function)** Under the assumptions of this section, the social welfare function takes the form

\[ W(\phi, \tau) = W^{ra}(\phi, \tau) + \]
\[ + (1 + \chi) \ln \left(p^{1 - \eta} + (1 - p)^{1 - \eta}\right) - \kappa (1 - p) - \]
\[ - \left(\ln \Gamma(1 - (1 - \tau) \eta) - \eta \gamma (1 - \tau) + \ln (p + (1 - p) \exp(\kappa)) - \kappa (1 - p)\right), \]  

where \(W^{ra}(\phi, \tau)\) is welfare in the representative-agent economy, defined as:

\[ W^{ra}(\phi, \tau) = \ln (1 - \phi) + \chi \ln \phi + (1 + \chi) \ln \ell^* - \zeta^{-1} (\ell^*)^\zeta + (1 + \chi) \ln \Gamma (1 - \eta), \quad (25) \]

\(\phi\) denotes the share of public goods in aggregate output, \(\Gamma\) is the gamma function, and \(p\) is the probability of not migrating:

\[ p = \frac{1}{1 + \exp\left(-\kappa/((1 - \tau) \eta)\right)}. \quad (26) \]

\footnote{The assumption of i.i.d. shocks allows me to derive the welfare function analytically because it implies that the stationary distribution of productivity \(f_t^p(z')\) can be easily inferred from the distribution of shocks \(g_t(\varepsilon)\) characterized in Proposition 4.}
The first row of equation (24) represents the welfare of a representative-agent version of this economy, in which each agent’s productivity is exogenous and equal to the average level of productivity in the case of no migration. In this case, welfare depends only on the allocation of output between private and public consumption and on the quantity of output produced by exerting effort, net of the disutility of labor supply. As discussed by Heathcote et al. (2017, Proposition 5), and easily shown here as well, the welfare of the representative is maximized with a share of public good equal to \( \varphi^* = \chi / (\chi + 1) \) and by subsidizing labor supply by setting tax progressivity to \( \tau = -\chi \). The labor subsidy is needed because agents do not internalize the effect of their labor supply choices on the provision of public goods.

Consider now the second row of equation (24). It represents the net social welfare derived from the migration process. The latter has the potential to increase aggregate productivity by locating workers where they are idiosyncratively more productive. Specifically, the term multiplying \((1 + \chi)\) in the second row of equation (24) is the logarithm of average productivity in the economy. This term is multiplied by \((1 + \chi)\) because the planner, but not the agents, takes into account the effect of higher geographic mobility on the provision of public goods. The term \(\kappa (1 - p)\), instead, represents the aggregate utility costs of migration. The level of \(\tau\) that optimally trades-off the social benefits of the productivity gains from migration and aggregate mobility costs satisfies the following condition:

\[
(1 + \chi) (1 - \eta) \frac{(1 - p)^{-\eta} - p^{-\eta}}{p^{1-\eta} + (1 - p)^{1-\eta}} = \kappa.
\]

The term in the third row of equation (24) represents the welfare cost of consumption inequality. Consumption dispersion is due to both the direct effect of exogenous idiosyncratic shocks and to agents’ endogenous migration responses to these shocks. For example, the term \(\kappa (1 - p)\) in this row reflects the fact that the earnings growth of migrating workers is increasing in the cost of moving (see equation 18 in Proposition 4). To minimize the welfare costs of consumption inequality, the social planner would like to set tax progressivity at its maximum level, \(\tau = 1\), pooling consumption among all agents.
Putting all these effects together, the first-order condition of $W$ with respect to $\tau$ is:

$$\frac{\partial W (\varphi, \tau)}{\partial \tau} = \frac{\partial W^{ra} (\varphi, \tau)}{\partial \tau} - \eta \left\{ \frac{\Gamma (1 + (1 - \tau) \eta)}{\Gamma (1 - (1 - \tau) \eta)} + \gamma \right\} + \left\{ \frac{\exp (\kappa) - 1}{p + (1 - p) \exp (\kappa)} - (1 + \chi) (1 - \eta) \frac{(1 - p)^{-\eta} - p^{-\eta}}{p^{1-\eta} + (1 - p)^{1-\eta}} \right\} \frac{\partial p}{\partial \tau} = 0,$$

(27)

where $\partial p/\partial \tau > 0$ by equation (26). For ease of interpretation, I have re-arranged terms so that the first-row of equation (27) reflects the trade-offs associated with the one-location version of the model. If $K = 1$, there is no migration and agents experience purely exogenous variation in their productivity over time. The relevant trade-offs are those associated with the representative agent version of the model and the social benefits from reducing consumption inequality (term in curly brackets in the first row of equation (27)). Let $\hat{\tau}$ denote the level of tax progressivity that maximizes welfare in the one-location version of the model.

The second row of equation (27) represents, instead, the contribution that geographic mobility makes to the socially optimal level of $\tau$. Since the two terms $\kappa (1 - p)$ involving mobility costs in equation (24) cancel out, a marginal increase in $\tau$ produces only two effects. The first-term in curly brackets represents the welfare benefit of lowering consumption dispersion by reducing internal migration. The second term in the curly brackets represents the social cost associated with lower average productivity brought about by a higher $\tau$. An important question is whether these additional trade-offs associated with geographic mobility contribute to push the welfare-optimal $\tau$ above or below $\hat{\tau}$. The answer to this question depends on the model’s parameters as described by the following proposition.

**Proposition 6 (The role of geographic mobility)**  The welfare function $W (\varphi, \tau)$ is such that:

---

44 Formally, $\hat{\tau}$ is such that the first row of equation (27) is equal to zero. It can be shown that, if $K = 1$, the function $W$ is globally concave in $\tau$.

45 The reason the terms $\kappa (1 - p)$ cancel out is that the aggregate moving costs, a net welfare loss, are exactly offset by the increase in average earnings enjoyed by movers, a net gain. As pointed out in the text, the latter point emerges clearly in equation (18) of Proposition 4.

---
(a) If $\hat{\tau} \geq 0$ and $\chi/(1 + \chi) > \eta$, then:

$$\frac{\partial W(\varphi, \hat{\tau})}{\partial \tau} < 0.$$ 

(b) If $\hat{\tau} \leq 0$ and $\chi/(1 + \chi) < \eta$, then:

$$\frac{\partial W(\varphi, \hat{\tau})}{\partial \tau} > 0.$$ 

Proposition 6's sufficient conditions have an intuitive interpretation. Suppose that in the one-location version of the model the optimal tax system is progressive ($\hat{\tau} > 0$). If the share of public goods in consumption $\chi$ is large enough relative to the amount of risk faced by agents $\eta$, then in the two-location version of the model ($K = 2$) the planner has an incentive to reduce tax progressivity below $\hat{\tau}$. By reducing tax progressivity, the planner increases the incentives for labor mobility leading to higher output and aggregate private and public consumption. The reason why the planner is willing to reduce tax progressivity and accept a higher cross-sectional dispersion in consumption is that the income risk $\eta$ faced by the agents is small relatively to the benefits associated with public goods consumption. The same logic explains why, if the amount of risk is relatively large ($\eta$ large relative to $\chi$) and the one-location tax system system is regressive ($\hat{\tau} < 0$), the planner might want to reduce its regressivity when it takes into account the trade-offs associated with geographic mobility. 

The main take-away from Proposition 6 is that while tax progressivity has unambiguous effects on labor mobility, the new endogenous mobility channel highlighted in this paper might contribute to either reduce or increase the welfare-optimal $\tau$ relative to the one-location version of the model. In the next section I use the full quantitative model to study the optimal degree of tax progressivity and the contribution of the labor mobility channel to its determination.

\[46\] Since the value of $\hat{\tau}$ is itself a function of the model’s parameters, I have verified numerically that there exist different sets of parameter values such that cases (a) and (b) apply.
6.2 Quantitative Analysis

I now study optimal tax progressivity and public good provision policies in this economy starting from an initial steady state corresponding to a constant $\tau = 0.192$, the value estimated from U.S. data. At $t = 1$, the social planner chooses tax progressivity to maximize a utilitarian welfare criterion. The latter includes both the remaining lifetime utility of agents that are alive at $t = 1$ as well as the lifetime utility of all cohorts that are not born yet, discounting their utility using the agents’ own discount factor $\beta$. The welfare criterion $W$ takes the following form:

$$
W = (1 - \beta) \frac{1}{\omega} \sum_{r=1}^{\infty} \sum_{a=1}^{r} E_z \{ E_{\xi} [P_1 (a, z, \xi; r) | z, a, r] \} + (1 - \beta) \frac{1}{\omega} \sum_{r=1}^{\infty} \beta^{r-1} E_z \{ E_{\xi} [P_t (1, z, \xi; r) | z, 1, r] \}.
$$

(28)

Notice that social welfare is therefore computed taking into account the entire transition of the economy from its initial steady state to its final one. The planner is subject to the government’s budget constraint (9), which requires a period-by-period budget balance. Since the government’s budget imposes a relationship among the objects of the triple ($\tau_t, \lambda_t, G_t$) at each point in time, the planner needs to optimize only with respect to two of them, taking (9) into account. In what follows I express the problem as that of finding the optimal level or sequence of $\tau_t$ and the share of aggregate income spent on public goods, defined as $\varphi_t \equiv G_t / Y_t$. It turns out that, given the log utility specification (3), these two dimensions of optimization are independent of one another.\(^{47}\) In particular, the optimal share of public expenditures can be obtained analytically, as in Heathcote et al (2017).

Proposition 7 (Optimal public good provision) The welfare maximizing share of pub-

\(^{47}\)It follows that the optimal value of $\varphi_t$ is the same independently of whether the planner chooses a unique value of $\tau$ or the entire sequence $\{\tau_t\}_{t=1}^{\infty}$. 

39
lic goods in aggregate income is constant over time and given by:

\[ \varphi^* = \frac{\chi}{1 + \chi}. \] (29)

In order to set the parameter \( \chi \), I assume that public good provision in the U.S. is set optimally, according to (29). The data counterpart of \( G/Y \) for this economy is the share of the Federal government’s consumption expenditures in the sum of the latter and personal (non-durable) consumption expenditures. The average share for the period 2000-2007 is 0.082, which implies a value \( \chi = 0.089 \).

6.2.1 Optimal Tax Progressivity

I first consider the case in which the planner announces at \( t = 1 \) a degree of tax progressivity \( \tau \) which remains constant over time, while the economy goes into the transition phase towards its new steady state (e.g. Domeij and Heathcote (2004) and Krueger and Ludwig (2016)). The first column of Table 3 reports the welfare maximizing value \( \tau^* = 0.307 \) for the benchmark case.

<table>
<thead>
<tr>
<th>Benchmark version</th>
<th>One location version ((K = 1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau^* )</td>
<td>( \tau^{ss} )</td>
</tr>
<tr>
<td>0.307</td>
<td>0.113</td>
</tr>
</tbody>
</table>

Table 3: Optimal one-shot tax progressivity policy taking the transition into account \((\tau^*)\) and considering only steady state welfare \((\tau^{ss})\).

As discussed in Section 6.1, in selecting the optimal \( \tau \), the planner trades-off the costs and benefits of redistribution. On the benefits side, higher tax progressivity provides insurance against idiosyncratic location-specific shocks \( \varepsilon \) and the initial \((a = 1)\) productivity draw \( z \). On the cost side, a more progressive tax system reduces incentives to supply labor and to
move to locations and jobs with higher earnings, a distortion to a dynamic choice. There are two important features of the full model that are absent in the simplified version considered in Section 6.1: finite lifetimes and persistent idiosyncratic shocks. These two elements together imply that the incentives to migrate decline with age. The fact that older agents are relatively immobile reduces the distortions associated with redistribution in the early periods of a tax reform because the short-run elasticity of average productivity to changes in tax progressivity is smaller than the associated long-run elasticity.

The two plots in the top row of Figure 6 illustrate the impact of the one-shot reform on migration rates by age. The reform increases $\tau$ relative to its initial value, so migration rates for both types and for all model ages less than (about) $a = 20$ fall. The dynamic path of average productivity for each agent type is plotted in the bottom row of Figure 6. As the agents who are alive at $t = 1$ age and are replaced by young agents, the decline in age-specific migration rates produces a decline in average productivity over time until the economy settles down at a lower average productivity in the final steady state.

This discussion suggests that taking the transition into account plays a potentially important role in the planner’s trade-offs between incentives and redistribution. In the short-run, at the time the reform is implemented, agents older than $a = 20$ have basically stopped moving geographically by choice. Their productivity is therefore unaffected by the tax reform. As a consequence, the planner has an incentive to set $\tau$ at a relatively high value because it understands that this margin will not be distorted. To provide a sense of the importance of the transition, consider the alternative exercise in which the planner maximizes steady state welfare, ignoring the transition all together (as in, e.g., Erosa and Koreshkova, 2007)). The optimal degree of tax progressivity in this case is $\tau^{ss} = 0.113$ (see Table 3, second column), a much smaller degree of progressivity than obtained when taking the transition into account.

---

48 Notice that there is no variation over time in migration rates by age because $\tau$ is a constant in the one-shot reform considered in this section.

49 Average productivity at a point in time is the average of individual wages $\exp(z' + \alpha_{a,r})$ across all ages for a given type $\tau$ and time period $t$.

50 This comparison is consistent with the findings of Heathcote et al (2017, Section 6.3), who also find a smaller optimal degree of tax progressivity in an economy with reversible skill investments (a steady state...
The last two columns of Table 3 show the welfare-maximizing \( \tau \) in the one-location version of the model \( (K = 1) \). Notice that in this case taking or not the transition into account is irrelevant for the optimal \( \tau \) because, absent migration, productivity evolves exogenously over time. In the one-location version of the model, the optimal \( \tau \) is 9 percentage points higher than in the benchmark model. This implies that while the two economies share the same average tax rate \( \varphi^* \), the average income-weighted marginal tax rate is about 8 percentage points higher in the one-location economy than in the benchmark. \(^{51}\) Thus, the endogenous geographic mobility mechanism has a considerable effect on the optimal \( \tau \). These results are in accordance with those in Proposition 6, part (a), according to which, if \( \chi \) is sufficiently larger than \( \eta \), optimal tax progressivity in the benchmark model is smaller than in the one-location version. Notice, however, that, as discussed above, transitional dynamics considerations reduce the distortions associated with high tax progressivity relative to the steady state analysis of Section 6.1.

6.2.2 Optimal Path of Tax Progressivity

In this section I extend the analysis by considering the optimal policy sequence \( \{\tau_t\}_{t=1}^{\infty} \) and the additional welfare gains from a time-varying policy. The existence of such gains had been hypothesized, but not verified, by Krueger and Ludwig (2016, p.96), among others.\(^{52}\) The thought experiment now is as follows. Starting from the steady state with the benchmark \( \tau = 0.192 \), at \( t = 1 \) the planner select, unexpectedly, a policy sequence \( \{\tau_t, \lambda_t, G_t\}_{t=1}^{\infty} \) to maximize utilitarian welfare as defined in equation \( (28) \).

The optimal sequence \( \{\tau^*_t\}_{t=1}^{\infty} \) is represented in Figure 7 by the solid (purple) line. The main feature of the optimal time-varying policy is that at \( t = 1 \) the planner finds it optimal to set tax progressivity at a relatively high level, with \( \tau^*_1 \) approximately equal to 0.43, while

---

\(^{51}\) As shown in Appendix D.1, the average income-weighted marginal tax rate in the economy is \( (1 - (1 - \tau)(1 - \varphi^*)) \) while the average net tax rate is simply \( \varphi^* \).

\(^{52}\) Meghir (2016, p. 101) proposes an approach along these lines in his discussion of Krueger and Ludwig (2016)’s paper. Appendix D.3 describes the numerical algorithm used to solve for the optimal time-varying policy. Dyrda and Pedroni (2018) use a similar method to study optimal time-varying taxation in the context of the Aiyagari model.
announcing a declining path of $\tau_t^*$ for subsequent periods. The optimal path declines until it reaches its final steady state value around $t = 40$. The intuition behind this declining path hinges crucially on the observation that high degrees of tax progressivity in the early periods are less distortionary, in terms of reduced incentives for geographic mobility, than the expectation of high degrees of tax progressivity in subsequent periods. Higher levels of $\tau_t$ increase taxes on high income households and increase subsidies to low income ones. Since income is the product of labor effort and productivity, relatively rich households may respond in the short-run by reducing their work effort but cannot escape higher taxes by reducing their productivity. In other words, at $t = 1$, the distribution of productivity in the population is fixed and redistribution from high to low productivity households has a lump-sum component akin to unexpectedly taxing physical capital in the neoclassical growth model. The planner however understands that keeping high levels of $\tau_t$ over time is suboptimal because it induces households not to pursue job opportunities that would increase their income over time. This reduced incentive to move would lead to a decline in average productivity and therefore in
Figure 7: Optimal tax progressivity reform in various scenarios. The solid (purple) line represents the optimal time-varying policy $\{\tau_t\}_{t \geq 1}$. The dotted (dark blue) line represents the optimal one-shot policy. The solid (light-blue) line represents the optimal one-shot and time-varying policies (they are the same) for the version of the model with one location ($K = 1$). The dashed (red) line represents the initial steady state $\tau = 0.192$. 

\[ \begin{array}{c}
\text{t (time period)} \\
0 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 \\
\text{One-shot} & \text{Time varying} & \text{One location} & \text{Initial steady state} \\
\end{array} \]
the economy’s tax base. Hence, the declining pattern of \( \tau_t^* \) over time.

As already discussed in the previous section, the intuition for this result is confirmed if one considers the version of the model with one location \((K = 1)\), shutting down all migration opportunities. In this circumstance, the optimal policy reform calls for a constant \( \tau \) because the evolution of productivity is exogenous and independent of the degree of progressivity. The planner does not face any dynamic trade-off and its optimization problem becomes a static choice between the costs of labor supply distortions and the benefits of redistribution. This case is represented by the solid horizontal line around 0.4 in Figure 7 (labelled “One location”). For reference, the figure also presents the one-shot optimal policy \( \tau^* = 0.307 \) (dotted line) discussed in the previous section and the calibrated \( \tau = 0.192 \) in the initial steady state (dashed line).

Figure 8 represents the evolution of life-cycle profiles of migration for a number of cohorts over the transition with time-varying tax progressivity. For reference I have included the profile corresponding to the initial steady state as a dashed line starting at \( t = 1 \). Notice how the reform reduces, on impact, mobility rates of the cohort born in period \( t = 1 \), especially type 2 agents. Smaller mobility translates into a lower life-cycle profile of productivity and wages. Over time as optimal tax progressivity falls migration rates increase at each age and so do lifecycle profiles of wages. However, they remain depressed relative to the initial steady state because optimal tax progressivity remains asymptotically larger than its value in the original steady state.

\footnote{In a similar way, in the neo-classical growth model physical capital is more elastic in the long-run than in the short-run so a planner might want to tax it more heavily earlier than later. See Straub and Werning (2015) for a recent discussion of the optimal capital taxation literature. Notice that in the optimal taxation literature in the Ramsey tradition (Jones et al (1993)), upper bounds on capital tax rates have to be imposed early on to prevent the planner from effectively taxing capital in a lump-sum fashion in the early period of the reform. In my model, the presence of elastic labor supply endogenously reduces the planner’s incentives to redistribute in the early periods of the reform.}

\footnote{Figure A.1 in Appendix D.2.2 presents the optimal tax progressivity sequence for different values of the model’s parameters. In all these cases, the endogenous migration mechanism is active and the shape of optimal tax progressivity over time is qualitatively similar to the one in the benchmark, although it might be shifted upwards or downwards according to the specific parameter that is changed.}
Figure 8: Average migration rates and average wages by age and cohort in the transition following the optimal time-varying tax progressivity reform. The figures represent profiles over the life cycle for selective cohorts born in periods $t = 1, 10, 20, 30, 40$. For reference I present the pattern corresponding to the initial steady state as a dotted (––) line.
6.2.3 Welfare Effects

In this section I discuss the welfare gains from tax progressivity reform. Welfare gains in consumption equivalent units are defined as the proportional increase in households’ period consumption in the benchmark economy (with $\tau = 0.192$) that makes social welfare the same as under the optimal policy. Formally, given the logarithmic specification of utility, the equivalent variation is

$$s^* = \exp(W^* - W^{\text{bench}})$$

where $W^*$ denotes the welfare level corresponding to the optimal degree of tax progressivity under consideration and $W^{\text{bench}}$ welfare in the model’s original steady state. The first row of Table 4 reports $s^*$ corresponding to the one-shot, time-varying, and steady-state-optimal tax progressivity reform, respectively.

<table>
<thead>
<tr>
<th>Optimal Policy</th>
<th>One-shot</th>
<th>Time-varying</th>
<th>Steady-state</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau^* = 0.307$</td>
<td></td>
<td></td>
<td>$\tau^{ss} = 0.113$</td>
</tr>
<tr>
<td>$s^*$</td>
<td>0.706%</td>
<td>1.090%</td>
<td>-1.505%</td>
</tr>
<tr>
<td>$s^{ss}$</td>
<td>-1.775%</td>
<td>-0.352%</td>
<td>0.508%</td>
</tr>
</tbody>
</table>

Table 4: Summary of welfare effects of tax progressivity reform. The equivalent variation $s^*$ is defined in equation (30). The equivalent variation $s^{ss}$ is computed in a similar way, except that it only compares social welfare across initial and final steady states.

In order to compare with the previous literature, the second row of Table 4 also reports the equivalent variation, $s^{ss}$, obtained by comparing initial and final steady states, ignoring the transition.

The welfare gain from the optimal one-shot policy is 0.706 percent. By contrast, the welfare gain of the optimal time-varying policy is equivalent to 1.090 percent of lifetime consumption, or about 54 percent larger than the gain from the optimal one-shot policy.
Consistent with the previous literature (Krueger and Ludwig, 2016), adopting a policy that is optimal when comparing steady states may lead to welfare losses once the transition is taken into account. The welfare loss in this case is $1.505$ percent of lifetime consumption. As shown in the second row of Table 4, this same policy increases welfare by $0.508$ percent when comparisons are restricted to steady states. The table also illustrates how, adopting policies that are optimal taking the transition into account may lead to welfare losses when comparing steady states.

Who gains and who loses from the optimal policy during the transition? In order to answer this question Proposition A.1 in Appendix D.4 shows how to decompose the expression for $s^*$ into a weighted average of type-age-cohort specific equivalent variations. This decomposition reveals that, relative to the benchmark economy, both optimal policies (one-shot and sequence of tax progressivity) favor type 1 agents and hurt type 2 agents independently of their age and birth cohort. In addition, the one shot reform tends to produce more evenly distributed welfare gains and losses across various age groups and cohorts than the reform that optimizes over the transition. The latter produces larger gains and losses for the cohorts who are alive at the time of the reform and smaller welfare effects for the cohorts that are born after period 10.

### 6.3 A Migration Subsidy

In this economy the planner uses tax progressivity to affect three welfare-relevant margins: labor supply, migration choices, and consumption inequality. The negative impact of tax progressivity on migration choices might, in principle, be mitigated by supplying the planner with a second policy instrument, a migration subsidy. For example, until the 2017 Tax Cuts and Jobs Act, the U.S. fiscal code allowed a household to deduct moving expenses from its gross income for the purpose of determining its taxable income. Intuitively, a migration subsidy might operate similarly to an education subsidy in an economy in which distortionary taxes reduce education investment (Krueger and Ludwig, 2013). In Appendix E.4 I formally introduce a migration subsidy $x$ in the simplified version of the model discussed in Section 6.1.
With such subsidy, an agent’s consumption is $c = \lambda y^{1-\tau}$ if she stays put and $c = \lambda \exp(x) y^{1-\tau}$ if she migrates. The new policy vector is $(\tau, x, \varphi, \lambda)$, subject to the government’s budget constraint (9). In Proposition A.2 I show that the availability of the subsidy induces the planner to set tax progressivity in the same way as it would in the one-location version of the model. The welfare subsidy, instead, is set optimally to equalize the marginal social benefits and costs of migration.

7 Conclusions and Future Work

I have constructed an analytically tractable, yet rich, dynamic model of internal migration to study quantitatively the relationship between tax progressivity and the geographic mobility of labor. Higher degrees of tax progressivity reduce households’ incentives to undertake costly migration to take advantage of higher earnings opportunities. I use the model to compute the optimal tax progressivity policy and show that the migration channel plays a quantitatively important role in determining both the level and the time-path of optimal tax progressivity relative to a one-location economy.

A more general contribution of the paper is to introduce a new dynamic model of internal migration that is analytically tractable. The framework proposed here can be extended in a relatively straightforward fashion to allow for heterogeneity in productivity and amenities across locations, and for additional geographic details, such as moving costs that depend on distance. This flexibility makes the framework suitable to investigate the implications of other policies and shocks that lead to geographic reallocation of labor. A particularly interesting application of this framework would be to study the aggregate implications of heterogeneity in the level of taxes across U.S. states (see Fajgelbaum et al, 2015).

It would be interesting, but not straightforward, to relax some of the assumptions that make the model analytically tractable. For example, one could consider a non-logarithmic utility function, allow agents to save and borrow, or choose whether to participate in the workforce. The last extension would allow one to evaluate the impact on migration of policies,
such as disability payments, that are tied to non-participation in the labor market. I leave these and other extensions to future research.
Bibliography


Online Appendix (Not meant for publication)

A  Proofs

This section contains the proofs of the Propositions that appear in the text.

A.1 Proof of Proposition 1

The proof proceeds by guessing and verifying that the solution takes the form in Proposition 1. Replace the guess (10) into (8):

\[
P_t(a, z, \varepsilon, h; r) = \theta_r \max_k \left\{ v_t^0(a; r) + v_t^1(a; r) (\rho_r z + \eta_r \varepsilon_k - I_{hk} \kappa_r) \right\} + \\
+ \frac{1 - \theta_r}{K - 1} \sum_{k \neq h} \left\{ v_t^0(a; r) + v_t^1(a; r) (\rho_r z + \eta_r \varepsilon_k - \kappa_r) \right\}
\]

Simplify this expression:

\[
P_t(a, z, \varepsilon, h; r) = v_t^0(a; r) + v_t^1(a; r) \rho_r z + \theta_r v_t^1(a; r) \eta_r \max_k \left\{ \varepsilon_k - I_{hk} \frac{\kappa_r}{v_t^1(a; r) \eta_r} \right\} + \\
+ \frac{1 - \theta_r}{K - 1} v_t^1(a; r) \eta_r \sum_{k \neq h} \varepsilon_k - (1 - \theta_r) \kappa_r.
\]

For \( a < \bar{a} \) we need to compute

\[
E_{\varepsilon'} \left[ P_{t+1}(a + 1, z', \varepsilon', k; r) \right] = v_{t+1}^0(a + 1; r) + v_{t+1}^1(a + 1; r) \rho_r z' \\
+ \theta_r v_{t+1}^1(a + 1; r) \eta_r E_{\varepsilon'} \left[ \max_k \left\{ \varepsilon'_k - \frac{I_{hk} \kappa_r}{v_{t+1}^1(a + 1; r) \eta_r} \right\} \right] \\
+ (1 - \theta_r) v_{t+1}^1(a + 1; r) \eta_r E_{\varepsilon'} \left[ \varepsilon'_k \right] - (1 - \theta_r) \kappa_r.
\]

Consider the distribution of the random variable

\[
\max_k \left\{ \varepsilon'_k - \frac{I_{hk} \kappa_r}{v_{t+1}^1(a + 1; r) \eta_r} \right\}.
\]
It can be computed as
\[
\Pr \left[ \max_k \left\{ \varepsilon_k - \frac{I_{hk}\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right\} < y \right] \\
= \Pr [\varepsilon_h < y] \Pr \left[ \max_{k \neq h} \left\{ \varepsilon_k - \frac{I_{hk}\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right\} < y \right] \\
= \exp (-\delta_r \exp (-y)) \exp \left( -(K - 1) \exp \left( -\left( y + \frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right) \right) \\
= \exp \left( -\left[ \delta_r + (K - 1) \exp \left( -\frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right] \exp (-y) \right).
\]

This is a type 1 extreme-value with location parameter:
\[
\ln \left[ \delta_r + (K - 1) \exp \left( -\frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right],
\]
and unit scale parameter. Thus, its mean is:
\[
E_{\varepsilon'} \left[ \max_k \left\{ \varepsilon'_k - \frac{I_{hk}\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right\} \right] = \ln \left[ \delta_r + (K - 1) \exp \left( -\frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right] + \gamma.
\]

The expectation \(E_{\varepsilon'} [\varepsilon'_k]\) for \(k \neq h\) is simply \(\gamma\), Euler’s constant. Thus:
\[
E_{\varepsilon'} [P_{t+1} (a + 1, z', \varepsilon', k; r)] = v_{t+1}^0 (a + 1; r) + v_{t+1}^1 (a; r) \rho_r z' + v_{t+1}^1 (a; r) \eta_r \gamma - (1 - \theta_r) \kappa_r \\
+ \theta_r \eta_r v_{t+1}^1 (a; r) \ln \left[ \delta_r + (K - 1) \exp \left( -\frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right].
\]

It follows from (7) that for \(a < \bar{a}\):
\[
V_i (a, z', k; r) = \bar{u}_t (a) + (1 - \tau_t) z' + \beta v_{t+1}^0 (a + 1; r) + \beta v_{t+1}^1 (a + 1; r) (\eta_r \gamma + \rho_r z') \\
- \beta (1 - \theta_r) \kappa_r + \beta \theta_r v_{t+1}^1 (a + 1; r) \eta_r \ln \left[ \delta_r + (K - 1) \exp \left( -\frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right].
\]
Collect the terms in $z'$:

$$V_t (a, z', k; r) = \pi^*_t (a) + [(1 - \tau_t) + \beta \rho_r v_{t+1}^1 (a + 1; r)] z' + \beta v_{t+1}^0 (a + 1; r) + \beta v_{t+1}^1 (a; r) \eta_r \gamma$$

$$- \beta (1 - \theta_r) \kappa_r + \beta \theta_r v_{t+1}^1 (a; r) \eta_r \ln \left[ \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{\eta_r v_{t+1}^1 (a; r)} \right) \right].$$

Impose consistency with the initial guess for $a < \bar{a}$:

$$v_t^1 (a; r) = (1 - \tau_t) + \beta \rho_r v_{t+1}^1 (a + 1; r), \quad (A.1)$$

$$v_t^0 (a; r) = \pi^*_t (a) + \beta v_{t+1}^0 (a + 1; r) + \beta v_{t+1}^1 (a + 1; r) \eta_r \gamma + \beta (1 - \theta_r) \kappa_r + \beta \theta_r v_{t+1}^1 (a + 1; r) \eta_r \ln \left[ \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right].$$

Notice that for $a = \bar{a}$ the conditional value function is just given by (5) so:

$$v_t^1 (\bar{a}; r) = (1 - \tau_t),$$

$$v_t^0 (\bar{a}; r) = \pi^*_t (a)$$

Equation (A.1) can be solved to obtain:

$$v_t^1 (a; r) = 1 - \tau_t + \sum_{k=1}^{a-\bar{a}} (\beta \rho_r)^k (1 - \tau_{t+k}) \text{ for } a < \bar{a}.$$  

Now, impose consistency with the constant term for $a < \bar{a}$:

$$v_t^0 (a; r) = \pi^*_t (a) + \beta v_{t+1}^0 (a + 1; r) - \beta (1 - \theta_r) \kappa_r + \beta v_{t+1}^1 (a + 1; r) \eta_r \gamma + \beta \theta_r \eta_r v_{t+1}^1 (a + 1; r) \ln \left[ \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right]. \quad (A.2)$$

This can be simplified even further using the definition of $p_t (a; r)$:

$$p_t (a; r) \equiv \frac{1}{1 + (K - 1) \delta_r^{-1} \exp \left( - (\kappa_r / \eta_r) v_t^1 (a; r)^{-1} \right)}.$$  

(A.3)
Thus:

\[
\ln \left[ \delta_r + (K - 1) \exp \left( -\frac{\kappa_r}{\eta_r v_{t+1}^1 (a + 1; r)} \right) \right] = - \ln \left( p_{t+1} (a + 1; r) / \delta_r \right).
\]

Replacing it into (A.2):

\[
v_0^0 (a; r) = \pi_t (a) + \beta v_{t+1}^0 (a + 1; r) - \beta (1 - \theta_r) \kappa_r + \beta v_{t+1}^1 (a + 1; r) \eta_r \left( \gamma - \theta_r \ln \left( p_{t+1} (a + 1; r) / \delta_r \right) \right).
\]

Q.E.D.

A.2 Proof of Proposition 2

Consider an agent of age \(a\), located in \(h\) at the beginning of the period, with shock vector \(\varepsilon\) and productivity \(z\). Comparing value functions, a choice to remain in location \(h\) \((M_t (a, \varepsilon, h, h; r) = 1)\) requires that:

\[
v_1^1 (a; r) (\rho_r z + \eta_r \varepsilon_h) > v_0^1 (a; r) (\rho_r z + \eta_r \varepsilon_l) - \kappa_r \text{ for all } l \neq h.
\]

The inequality above is equivalent to:

\[
\varepsilon_h > \max_{l \neq h} \left( \varepsilon_l - \frac{\kappa_r}{v_1^1 (a; r) \eta_r} \right).
\]

Rearranging this gives:

\[
\frac{\kappa_r}{v_1^1 (a; r) \eta_r} > \max_{l \neq h} \varepsilon_l - \varepsilon_h.
\]

If the agent moves to a location \(k \neq h \) \((M_t (a, \varepsilon, h, k; r) = 1)\), it must be the case that:

\[
\varepsilon_k > \varepsilon_l \text{ for } l \neq h
\]

and

\[
v_1^1 (a; r) (\rho_r z + \eta_r \varepsilon_h) < v_1^1 (a; r) (\rho_r z + \eta_r \varepsilon_k) - \kappa_r.
\]
Simplifying these expressions we obtain:

\[
\frac{\kappa_r}{v_t^1(a; r) \eta_r} < \max_{l \neq h} \varepsilon_l - \varepsilon_h
\]

\[
k = \arg \max_{l \neq h} \varepsilon_l.
\]

Q.E.D.

A.3 Proof of Proposition 3

1. Migration declines with age. The migration rate is \(1 - \theta_r p_t(a; r)\), where \(p_t(a; r)\) is defined in (A.3). The latter is increasing in age, \(a\), because \(v_t^1(a; r)\) is decreasing in \(a\). Thus, the migration rate declines with age.

2. A higher \(\tau_{t+k}\) reduces \(v_t^1(a; r)\) and increases \(p_t(a; r)\), so it reduces migration.

3. The impact of a higher \(\tau_{t+k}\) depends on the moving cost parameter \(\kappa_r\). Specifically, it first rises and then fall with \(\kappa_r\). Consider the derivative:

\[
\frac{\partial p_t(a; r)}{\partial \tau_{t+k}} = \frac{(K - 1) \delta_r^{-1} \exp \left(-\kappa_r (\eta_r v_t^1(a; r))^{-1}\right) \left(\kappa_r / \eta_r\right) \left(1/v_t^1(a; r)^2\right)}{\left[1 + (K - 1) \delta_r^{-1} \exp \left(-\kappa_r (\eta_r v_t^1(a; r))^{-1}\right)\right]^{2}} (\beta \rho_r)^k.
\]

Using the definition of \(p_t(a; r)\), this expression simplifies to

\[
\frac{\partial p_t(a; r)}{\partial \tau_{t+k}} = p_t(a; r) (1 - p_t(a; r)) \left(\kappa_r / \eta_r\right) \left(1/v_t^1(a; r)^2\right) (\beta \rho_r)^k
\]

where \(p_t(a; r)\) is monotonically increasing in \(\kappa_r\) and is such that \(p_t(a; r) \to 1\) if \(\kappa_r \to \infty\) and

\[
p_t(a; r) \to \frac{1}{1 + (K - 1) \delta_r^{-1}} < 1
\]
if \( r \rightarrow 0 \). Take the derivative with respect to \( \kappa_r \) ignoring the constant term \( \left( \frac{1}{\eta_r v^1_t (a; r)^2} \right) (\beta \rho_r)^k \):

\[
\frac{\partial p_t (a; r)}{\partial \tau_{t+k} \partial \kappa_r} = p_t (a; r) (1 - p_t (a; r)) + \kappa_r (1 - p_t (a; r)) \frac{\partial p_t (a; r)}{\partial \kappa_r} - \kappa_r p_t (a; r) \frac{\partial p_t (a; r)}{\partial \kappa_r}
\]

\[
= p_t (a; r) (1 - p_t (a; r)) + \kappa_r (1 - 2p_t (a; r)) \frac{\partial p_t (a; r)}{\partial \kappa_r}
\]

where

\[
\frac{\partial p_t (a; r)}{\partial \kappa_r} = \frac{(K - 1) \delta_r^{-1} \exp \left( -\kappa_r (\eta_r v^1_t (a; r))^{-1} \right) (\eta_r v^1_t (a; r))^{-1}}{\left[ 1 + (K - 1) \delta_r^{-1} \exp \left( -\kappa_r (\eta_r v^1_t (a; r))^{-1} \right) \right]^{2}}
\]

\[
= p_t (a; r) (1 - p_t (a; r)) \left( 1 + \kappa_r (\eta_r v^1_t (a; r))^{-1} (1 - 2p_t (a; r)) \right)
\]

Thus:

\[
\frac{\partial p_t (a; r)}{\partial \tau_{t+k} \partial \kappa_r} = p_t (a; r) (1 - p_t (a; r)) \left( 1 + \kappa_r (\eta_r v^1_t (a; r))^{-1} (1 - 2p_t (a; r)) \right)
\]

and its sign depends on the sign of the term:

\[
1 + \kappa_r (\eta_r v^1_t (a; r))^{-1} (1 - 2p_t (a; r)).
\]

If \( \kappa_r = 0 \) this sign is positive. Since \( p_t (a; r) \) is monotonically increasing in \( \kappa_r \) and tends to 1 as \( \kappa_r \rightarrow \infty \), as \( \kappa_r \) increases eventually \( p_t (a; r) \) will become larger than 0.5 and the second portion of the above term becomes negative and arbitrarily large in absolute value as \( \kappa_r \) grows.

4. The incidence of exogenous relocations increases with age. Compute the probability of exogenous relocation relative to the probability of moving by age:

\[
1 - \frac{\theta_r}{1 - \theta_r p_t (a; r)}
\]

Since, \( p_t (a; r) \) increases with age, the incidence of exogenous relocations increases too.
A.4 Proof of Proposition 4

A.4.1 Household chooses to stay

Suppose that the agent chooses to stay in same location \( h \), so \( M_t(a, \varepsilon, h; h; r) = 1 \). The CDF of productivity growth conditional on staying in the same location is:

\[
Pr(g_t(\varepsilon, a; r) \leq s| M_t(a, \varepsilon, h, h; r) = 1) = \frac{Pr(g_t(\varepsilon, a; r) \leq s \text{ and } M_t(a, \varepsilon, h, h; r) = 1)}{Pr(M_t(a, \varepsilon, h; h; r) = 1)},
\]

where by definition:

\[
Pr(M_t(a, \varepsilon, h; h; r) = 1) = p_t(a; r).
\]

Consider now the numerator in equation (A.5):

\[
Pr(g_t(\varepsilon, a; r) \leq s \text{ and } M_t(a, \varepsilon, h, h; r) = 1) = Pr(\varepsilon_h \leq s \text{ and } \varepsilon_h > \max_{l \neq h} \varepsilon_l - \frac{\kappa_r}{v^1_t(a; r) \eta_r}).
\]

The equation above can be rewritten as:

\[
Pr\left(\max_{l \neq h} \varepsilon_l - \frac{\kappa_r}{v^1_t(a; r) \eta_r} < \varepsilon_h \leq s\right).
\]

Taking into account the fact that \( \max_{l \neq h} \varepsilon_l \) is type-1 extreme value with location parameter \( \ln(K - 1) \), (A.6) can be rewritten as:

\[
\int Pr\left(x - \frac{\kappa_r}{v^1_t(a; r) \eta_r} < \varepsilon_h \leq s|x\right) (K - 1) \exp(-x) \exp(-(K - 1) \exp(-x)) dx. \quad \text{(A.7)}
\]

We know that:

\[
Pr\left(x - \frac{\kappa_r}{v^1_t(a; r) \eta_r} < \varepsilon_h \leq s|x\right) = 0
\]
if

\[ x > s + \frac{\kappa_r}{v_t^1(a; r) \eta_r}. \]  \hspace{1cm} (A.8)

Also, since \( \varepsilon_h \) is type-1 extreme value with location parameter \( \ln \delta_r \), if (A.8) does not hold:

\[
\Pr \left( x - \frac{\kappa_r}{v_t^1(a; r) \eta_r} < \varepsilon_h \leq s | x \right) 
= \exp (-\delta_r \exp (-s) - \exp \left( -\delta_r \exp \left( -\left( x - \frac{\kappa_r}{v_t^1(a; r) \eta_r} \right) \right) \right)).
\]

Replace the previous expressions into (A.7):

\[
\int \Pr \left( x - \frac{\kappa_r}{v_t^1(a; r) \eta_r} < \varepsilon_h \leq s | x \right) \times 
(K - 1) \exp (-x) \exp (- (K - 1) \exp (-x)) \, dx 
= \int_{-\infty}^{s+ \frac{\kappa_r}{v_t^1(a; r) \eta_r}} \Pr \left( x - \frac{\kappa_r}{v_t^1(a; r) \eta_r} < \varepsilon_h \leq s | x \right) \times 
(K - 1) \exp (-x) \exp (- (K - 1) \exp (-x)) \, dx 
= \int_{-\infty}^{s+ \frac{\kappa_r}{v_t^1(a; r) \eta_r}} \exp (-\delta_r \exp (-s)) \times 
(K - 1) \exp (-x) \exp (- (K - 1) \exp (-x)) \, dx 
- \int_{-\infty}^{s+ \frac{\kappa_r}{v_t^1(a; r) \eta_r}} \exp (-\delta_r \exp \left( -\left( x - \frac{\kappa_r}{v_t^1(a; r) \eta_r} \right) \right)) \times 
(K - 1) \exp (-x) \exp (- (K - 1) \exp (-x)) \, dx.
\]
The first component is:

\[
\int_{-\infty}^{s+} \frac{\kappa_r}{v^l_t(a; r) \eta_r} \exp(-\delta_r \exp(-s)) (K - 1) \exp(-x) \exp(-(K - 1) \exp(-x)) \, dx
\]

\[= \exp(-\delta_r \exp(-s)) \exp(- (K - 1) \exp(- \left(s + \frac{\kappa_r}{v^l_t(a; r) \eta_r}\right))) \]

\[= \exp(-\delta_r \exp(-s)) \exp(- (K - 1) \exp(- \left(\frac{\kappa_r}{v^l_t(a; r) \eta_r}\right)) \exp(-s)) \]

\[= \exp\left(-\left((K - 1) \exp\left(- \left(x - \frac{\kappa_r}{v^l_t(a; r) \eta_r}\right)\right) + \delta_r \right) \exp(-s) \right) \exp(-s) \]
The integral is therefore:
\[
\exp \left( - \left[ \delta_r \exp \left( \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right) + (K - 1) \right] \exp \left( - \left( s + \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right) \right) \right) \\
= \exp \left( - \left[ \delta_r \exp \left( \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right) + (K - 1) \right] \exp \left( - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \exp (-s) \right) \right) \\
= \exp \left( - \left[ \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right) \right] \exp (-s) \right)
\]

It follows that the second term (A.10) is:
\[
\int_{-\infty}^{s+} \exp \left( - \delta_r \exp \left( - \left( x - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right) \right) \right) \times \\
(K - 1) \exp (-x) \exp (- (K - 1) \exp (-x)) \; dx \\
= \frac{(K - 1)}{(K - 1) + \delta_r \exp \left( - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right)} \times \\
\exp \left( - \left[ \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right) \right] \exp (-s) \right).
\]

Putting all together equation (A.9) becomes:
\[
\int \Pr \left( x - \frac{\kappa_r}{v_l^1(a;r) \eta_r} < \varepsilon_h \leq s \mid x \right) (K - 1) \exp (-x) \exp (- (K - 1) \exp (-x)) \; dx \\
= \frac{1}{1 + (K - 1) \delta_r^{-1} \exp \left( - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right)} \times \exp \left( - \left( \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right) \right) \exp (-s) \right)
\]

Since
\[
\Pr \left( M_t(a, \varepsilon, h, \varepsilon; r) = 1 \right) = \frac{1}{1 + (K - 1) \delta_r^{-1} \exp \left( - \frac{\kappa_r}{v_l^1(a;r) \eta_r} \right)}.
\]
it follows that equation (A.5) becomes:

\[
\Pr (g_t (\varepsilon, a; r) \leq s \mid M_t (a, \varepsilon, h, h; r) = 1) = \exp \left( - \left( \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \right) \right) \exp (-s) \\
= \exp \left( - \exp \left( \ln \left( \delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \right) \right) \right) \exp (-s) \\
= \exp \left( - \exp \left( - \{ s - \ln (\delta_r + (K - 1) \exp \left( - \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \} \} \right) \right) \\
= \exp \left( - \exp \left( - \{ s - (\ln \delta_r - \ln (p_t (a; r))) \} \right) \right)
\]

This is the CDF of a type-1 extreme value distribution with location parameter

\[
\ln (\delta_r / p_t (a; r))
\]

and scale parameter 1. The mean of this distribution is:

\[
\ln (\delta_r / p_t (a; r)) + \gamma
\]

and its variance is \(\pi^2/6\).

**A.4.2 Household chooses to move**

Consider now the case in which the agent chooses to move away from \(h\). The agent may move to any location \(k \neq h\). I am interested in the distribution of productivity growth conditional on the event “move” away from \(h\). Formally, I am interested in the CDF:

\[
\Pr (g_t (\varepsilon, a; r) \leq s \mid M_t (a, \varepsilon, h, h; r) = 0).
\]

Since the agent may move to one of the remaining \(K - 1\) locations and these events are
disjoint, I can write:

\[
Pr(g_t(\varepsilon, a; r) \leq s | M_t(a, \varepsilon, h, h; r) = 0) = \frac{Pr(g_t(\varepsilon, a; r) \leq s \text{ and } M_t(a, \varepsilon, h, h; r) = 0)}{Pr(M_t(a, \varepsilon, h, h; r) = 0)},
\]

where

\[
Pr(g_t(\varepsilon, a; r) \leq s \text{ and } M_t(a, \varepsilon, h, h; r) = 0) = Pr(K_X \prod_{k=1}^K M_t(a; h, k; h) \text{ and } r \leq s) = Pr(K_X \prod_{k \neq h} M_t(a; h, k; h) \text{ and } r \leq s, \sum_{k \neq h} M_t(a, \varepsilon, h, k; r) = 1).
\]

Since the events \( M_t(a, \varepsilon, h, k; r) = 1 \) are disjoint we can further write:

\[
Pr \left( \sum_{k \neq h} M_t(a, \varepsilon, h, k; r) \varepsilon_k \leq s \text{ and } \sum_{k \neq h} M_t(a, \varepsilon, h, k; r) = 1 \right)
= \sum_{k \neq h} Pr(\varepsilon_k \leq s \text{ and } M_t(a, \varepsilon, h, k; r) = 1).
\]

Recall that for \( k \neq h \):

\[
Pr(\varepsilon_k \leq s \text{ and } M_t(a, \varepsilon, h, k; r) = 1) = Pr(\varepsilon_k \leq s \text{ and } \frac{K_r}{v_t^1(a; r) \eta_r} < \varepsilon_k - \varepsilon_h \text{ and } \varepsilon_k = \max_{l \neq h} \varepsilon_l)
= Pr(\varepsilon_k \leq s \text{ and } \frac{K_r}{v_t^1(a; r) \eta_r} < \varepsilon_k - \varepsilon_h | \varepsilon_k = \max_{l \neq h} \varepsilon_l) \times Pr(\varepsilon_k = \max_{l \neq h} \varepsilon_l).
\]

Notice that due to the symmetry of the distributions:

\[
Pr(\varepsilon_k = \max_{l \neq h} \varepsilon_l) = \frac{1}{K-1}.
\]
The distribution of $\varepsilon_k$ conditional on being the maximum of $K-1$ i.i.d. type 1 extreme-value random variables, is type 1 extreme-value with location parameter $\ln (K-1)$. Write:

$$\Pr \left( \varepsilon_k \leq s \text{ and } \frac{\kappa_r}{v^1_l (a; r) \eta_r} < \varepsilon_k - \varepsilon_h | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right)$$

$$= \Pr \left( \frac{\kappa_r}{v^1_l (a; r) \eta_r} + \varepsilon_h < \varepsilon_k \leq s | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right)$$

$$= \int \Pr \left( \frac{\kappa_r}{v^1_l (a; r) \eta_r} + u < \varepsilon_k \leq s | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) g(u) \, du,$$

where $g(u)$ is a type 1 extreme-value with location parameter $\ln \delta_r$ and:

$$\Pr \left( \frac{\kappa_r}{v^1_l (a; r) \eta_r} + u < \varepsilon_k \leq s | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) = 0$$

if

$$\frac{\kappa_r}{v^1_l (a; r) \eta_r} + u > s$$

and

$$\exp \left( -(K-1) \exp (-s) \right) - \exp \left( -\left(\exp \left( -(K-1) \exp \left(-\left(\frac{\kappa_r}{v^1_l (a; r) \eta_r} + u\right)\right)\right)\right) \right)$$

if

$$u < s - \frac{\kappa_r}{v^1_l (a; r) \eta_r}.$$

Thus:

$$\int \Pr \left( \frac{\kappa_r}{v^1_l (a; r) \eta_r} + u < \varepsilon_k \leq s | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) g(u) \, du$$

$$= \int_{-\infty}^{s - \frac{\kappa_r}{v^1_l (a; r) \eta_r}} \exp \left( -(K-1) \exp (-s) \right) - \exp \left( -\left(\exp \left(-\left(\frac{\kappa_r}{v^1_l (a; r) \eta_r} + u\right)\right)\right) \right) g(u) \, du$$

$$= \int_{-\infty}^{s - \frac{\kappa_r}{v^1_l (a; r) \eta_r}} \exp \left( -(K-1) \exp (-s) \right) g(u) \, du -$$

$$- \int_{-\infty}^{s - \frac{\kappa_r}{v^1_l (a; r) \eta_r}} \exp \left( -\left(\exp \left(-\left(\frac{\kappa_r}{v^1_l (a; r) \eta_r} + u\right)\right)\right) \right) g(u) \, du.$$
The first integral is:

\[
\int_{-\infty}^{s} \exp \left( -\left( K - 1 \right) \exp (-s) \right) g(u) \, du
= \exp \left( -\left( K - 1 \right) \exp (-s) \right) \int_{-\infty}^{s} \exp \left( -\left( s - \frac{K_r}{v_t^1 (a; r) \eta_r} \right) \right) g(u) \, du
= \exp \left( -\left( K - 1 \right) \exp (-s) \right) \exp \left( -\delta_r \exp \left( -\left( s - \frac{K_r}{v_t^1 (a; r) \eta_r} \right) \right) \right)
= \exp \left( -\left( K - 1 + \delta_r \exp \left( \frac{K_r}{v_t^1 (a; r) \eta_r} \right) \right) \exp (-s) \right).
\]

The second integral is:

\[
\int_{-\infty}^{s} \exp \left( -\left( K - 1 \right) \exp \left( -\left( \frac{K_r}{v_t^1 (a; r) \eta_r} + u \right) \right) \right) g(u) \, du
= \int_{-\infty}^{s} \exp \left( -\left( K - 1 \right) \exp \left( -\frac{K_r}{v_t^1 (a; r) \eta_r} \exp (-u) \right) \right) g(u) \, du
= \int_{-\infty}^{s} \exp \left( -\left( K - 1 \right) \exp \left( -\frac{K_r}{v_t^1 (a; r) \eta_r} \exp (-u) \right) \right) \left[ \delta_r \exp (-u) \exp (-\delta_r \exp (-u)) \right] \, du
= \delta_r \int_{-\infty}^{s} \exp (-u) \exp \left( -\left( K - 1 \right) \exp \left( -\frac{K_r}{v_t^1 (a; r) \eta_r} \right) + \delta_r \right) \exp (-u) \, du
= \frac{\delta_r}{(K - 1) \exp \left( -\frac{K_r}{v_t^1 (a; r) \eta_r} \right) + \delta_r} \times 
\exp \left( -\left( K - 1 \right) \exp \left( -\frac{K_r}{v_t^1 (a; r) \eta_r} \right) + \delta_r \right) \exp \left( -\left( s - \frac{K_r}{v_t^1 (a; r) \eta_r} \right) \right).
\]
Simplify the result:

\[
\delta_r \times (K - 1) \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} + \delta_r \right) \times \exp \left( - \left[ (K - 1) \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} + \delta_r \right) \exp \left( - s - \frac{\kappa_r}{v_t^1(a; r) \eta_r} \right) \right) \right) = \delta_r \times (K - 1) \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} + \delta_r \right) \times \exp \left( - \left[ K - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_t^1(a; r) \eta_r} \right) \right] \exp (-s) \right).
\]

Put together the two integrals to get:

\[
\int \Pr \left( \frac{\kappa_r}{v_t^1(a; r) \eta_r} + u < \varepsilon_k \leq s | \varepsilon_k = \max_{i \neq h} \varepsilon_i \right) g(u) \, du = \frac{(K - 1) \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} \right)}{(K - 1) \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} + \delta_r \right) \times \exp \left( - \left[ K - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_t^1(a; r) \eta_r} \right) \right] \exp (-s) \right)}.
\]

It follows that

\[
\Pr \left( g_t(\varepsilon, a; r) \leq s | M_t(a, \varepsilon, h, h; r) = 0 \right) = \frac{(K - 1) \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} \right)}{(K - 1) \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} + \delta_r \right) \times \exp (-s)} \left[ 1 - p_t(a; r) \right].
\]

Notice that

\[
1 - p_t(a; r) = 1 - \frac{1}{1 + (K - 1) \delta_r^{-1} \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} \right)} = \frac{(K - 1) \delta_r^{-1} \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} \right)}{1 + (K - 1) \delta_r^{-1} \exp \left( -\frac{\kappa_r}{v_t^1(a; r) \eta_r} \right)}.
\]
Replace:

\[
\Pr \left( g_t (\epsilon, a; r) \leq s \mid M_t (a, \epsilon, h, h; r) = 0 \right) = \frac{(K - 1) \delta_r^{-1} \exp \left( -\frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right)}{(K - 1) \delta_r^{-1} \exp \left( -\frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) + 1} \exp \left( -\left[ K - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \right] \right) \]

\[
= \exp \left( -\exp \left( -s - \ln \left[ K - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \right] \right) \right)
\]

Apply the definition of \( p_t (a; r) \):

\[
\exp \left( -\exp \left( -s - \ln \left[ \delta_r \exp \left( \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \right] \right) \right)
\]

\[
= \exp \left( -\exp \left( -s + \ln p_t (a; r) - \ln \delta_r - \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \right)
\]

\[
= \exp \left( -\exp \left( -\left\{ s - \left( -\ln p_t (a; r) + \ln \delta_r + \frac{\kappa_r}{v_t^1 (a; r) \eta_r} \right) \right\} \right) \right)
\]

This is the CDF of a type-1 extreme value distribution with scale parameter 1 and location parameter:

\[-\ln p_t (a; r) + \ln \delta_r + \frac{\kappa_r}{\eta_r v_t^1 (a; r)} \].

The moments of this distribution are:

\[
\text{mean} = \gamma + \ln \delta_r - \ln p_t (a; r) + \frac{\kappa_r}{\eta_r v_t^1 (a; r)} ,
\]

\[
\text{variance} = \frac{\pi^2}{6} .
\]

Q.E.D.
A.5 Proof of Corollary 1

The claim is that:

\[ \eta_r (\gamma + \ln \delta_r - \ln p_t(a; r)) + \frac{\kappa_r}{v^1_t(a; r)} > \eta_r (\gamma + \ln \delta_r - \ln p_t(a; r)) \geq \eta_r \gamma, \]

where the first expression from the left represents equation (18) multiplied by \( \eta_r \), the middle expression is equation (16) multiplied by \( \eta_r \), and the right-most expression denotes the mean of the unconditional distribution of \( \varepsilon_k \) multiplied by \( \eta_r \). The first inequality holds if \( \kappa_r > 0 \). The second inequality holds if \( \delta_r \geq p_t(a; r) \), which is the case since \( \delta_r \geq 1 \geq p_t(a; r) \).

Q.E.D.

A.6 Proof of Proposition 5

Under the assumption that \( \sigma \to \infty, \omega_1 = 1, K = 2, \theta_r = 1, \delta_r = 1, \alpha_{a,r} = 0 \), and \( \rho_r = 0 \), the value function in equation (10) takes the form:

\[ V(z') = \underbrace{u^* + (1 - \tau) z'}_{\text{current utility}} + \underbrace{\frac{\beta}{1 - \beta} [u^* + (1 - \tau) \eta (\gamma - \ln \rho)]}_{\text{expected future utility}} (A.12) \]

where the probability of staying in the same location is also a constant:

\[ p = \frac{1}{1 + \exp \left(-\kappa/((1 - \tau) \eta)\right)}. \quad (A.13) \]

Equation (A.12) helps to illustrate the various effects of fiscal policy on agents’ utility. The term \( u^* \), whose form is in equation (6), captures both the partial equilibrium effects of \( \tau \) on individual labor supply and leisure as well as the general equilibrium effects that operate through the government’s budget constraint (9). The term \( (1 - \tau) z' \) reflects the differential
impact of tax progressivity on utility, depending on an agent’s (log) productivity. Finally, the term in square brackets in (A.12) reflects the impact of tax progressivity on the agent’s expected utility starting from the following period and onward. A higher $\tau$ reduces expected future utility by increasing the probability $p(\tau)$ of not moving and by reducing the net gains from undertaking a move.

Proposition 1 implies that the expected value function is

$$
\int P(\epsilon) g(\epsilon) d\epsilon = v^0 + v^1 \eta - v^1 \eta \ln p \tag{A.14}
$$

where

$$
v^1 = 1 - \tau,
$$

$$
v^0 = \frac{\overline{u}^*}{1 - \beta} + \frac{\beta}{1 - \beta} (1 - \tau) \eta (\gamma - \ln p),
$$

and $p$ as defined in equation (A.13). From equation (6), the constant $\overline{u}^*$ is:

$$
\overline{u}^* = \ln \lambda + (1 - \tau) \ln \ell^* - \zeta^{-1} (\ell^*)^\zeta + \chi \ln G, \tag{A.15}
$$

with $G = \varphi Y$ and from the balanced-budget equation (9):

$$
\ln \lambda = \ln (1 - \varphi) + \tau \ln \ell^* + \ln \int \exp z' f^p (z') dz' - \ln \int \exp ((1 - \tau) z') f^p (z') dz'.
$$

Replace it in (A.15), taking into account $G = \varphi Y$ to obtain:

$$
\overline{u}^* = \ln (1 - \varphi) + \chi \ln \varphi + (1 + \chi) \ln \ell^* - \zeta^{-1} (\ell^*)^\zeta + (1 + \chi) \ln \int \exp z' f^p (z') dz' - \ln \int \exp ((1 - \tau) z') f^p (z') dz'. \tag{A.16}
$$

Now, develop the two terms with the integrals. Due to the fact that $\rho_x = 0$, we have:

$$
z' = \eta g(\epsilon).
$$
The form of the integrals is (for an arbitrary constant $b$):

\[
E \left[ \exp (b \eta g (\varepsilon)) \right] = \exp (-\eta \gamma b + \ln \Gamma (1 - b \eta)) \times \\
\left\{ p \exp (b \eta E [g (\varepsilon) | \text{stay}]) + (1 - p) \exp (b \eta E [g (\varepsilon) | \text{move}]) \right\} .
\]

This result is proved in Lemma A1. Proposition 4 guarantees that

\[
E [g (\varepsilon) | \text{move}] = E [g (\varepsilon) | \text{stay}] + \frac{\kappa}{\eta \nu}.
\]

Replace this relationship into (A.17) and collect to obtain

\[
E \left[ \exp (b \eta g (\varepsilon)) \right] = \exp (-\eta \gamma b + \ln \Gamma (1 - b \eta)) \exp (b \eta E [g (\varepsilon) | \text{stay}]) \left( p + (1 - p) \exp \left( \frac{bk}{\nu^2} \right) \right) .
\]

Taking into account the expression for $E [g (\varepsilon) | \text{stay}]$ in Proposition 4, it follows that:

\[
\ln \int \exp z' f^p (z') \, dz' = \ln \Gamma (1 - \eta) - \eta \ln p + \ln \left( p + (1 - p) \exp \left( \frac{\kappa}{1 - \tau} \right) \right),
\]
\[
\ln \int \exp ((1 - \tau) z') f^p (z') \, dz' = \ln \Gamma (1 - (1 - \tau) \eta) - (1 - \tau) \eta \ln p + \ln (p + (1 - p) \exp (k)).
\]

Replacing these expressions into (A.16) and the latter into the social welfare function

\[
W = (1 - \beta) \int P (\varepsilon) g (\varepsilon) \, d\varepsilon
\]

yields

\[
W = \ln (1 - \varphi) + \chi \ln \varphi + (1 + \chi) \ln \ell^* \cdot \zeta^{-1} (\ell^*)^\zeta + (1 + \chi) \ln \Gamma (1 - \eta) + \left(\ln \Gamma (1 - \tau) \eta - (1 - \tau) \eta \gamma + \ln (p + (1 - p) \exp (k))\right).
\]

75
From the definition of $p$ in equation (A.13) notice that

\[
\exp \left( \frac{\kappa}{1 - \tau} \right) = \left( \frac{p}{1 - p} \right)^{\eta}.
\]  

(A.20)

Replace this relationship into (A.19). Notice that the first row of equation (A.19) represents welfare of the representative agent, $W^{ra}(\varphi, \tau)$, as defined in equation (25). Then, equation (A.19) can be written as:

\[
W(\varphi, \tau) = W^{ra}(\varphi, \tau) + \left( 1 + \chi \right) \ln \left( p^{1 - \eta} + (1 - p)^{1 - \eta} \right) - \\
- \left( \ln \Gamma (1 - (1 - \tau) \eta) - \eta \gamma (1 - \tau) + \ln (p + (1 - p) \exp(\kappa)) \right) \\
\]

Subtracting and adding the moving costs $\kappa (1 - p)$ to this equation yields equation (24) in the text of the Proposition.

Q.E.D.

A.7 Proof of Proposition 6

Notice that $\partial p / \partial \tau > 0$, so I can focus on the term inside the curly brackets. Using the relationship (A.20), the latter can be written as:

\[
H(p, \kappa) - \left( \frac{1 - \eta}{1 - \varphi^*} \right) H \left( p, \frac{\kappa}{1 - \tau} \right)
\]

(A.22)

where I define:

\[
H(p, x) \equiv \frac{\exp(x) - 1}{p + (1 - p) \exp(x)}
\]

(A.23)

and $\varphi^*$ is defined in equation (29). Notice that the function $H(p, x)$ is increasing in $x$. It follows that if $\tau \in (0, 1)$ it is the case that:

\[
H \left( p, \frac{\kappa}{1 - \tau} \right) > H(p, \kappa).
\]
Thus, if $\varphi^* > \eta$ and $\hat{\tau} \in (0, 1)$, the expression (A.22) is negative. By contrast, if $\hat{\tau} < 0$ and $\varphi^* < \eta$, then the expression (A.22) is positive.

Q.E.D.

### A.8 Proof of Proposition 7

The indirect utility function takes the form:

$$u_t^* (a, z' ; r) = \bar{u}_t^* (a; r) + (1 - \tau_t) z'$$  \hspace{1cm} (A.24)

where:

$$\bar{u}_t^* (a; r) \equiv \ln \lambda_t + (1 - \tau_t) \ln \ell_t^* + (1 - \tau_t) \alpha_{a,r} - \zeta^{-1} \left( \ell_t^* \right)^\zeta + \chi \ln G_t,$$

and where $G_t = \varphi_t Y_t$. By definition, $\lambda_t$ is given by:

$$\lambda_t = (1 - \varphi_t) \frac{a^{-1} \sum_{a=1}^{A} \sum_{r=1}^{R} \omega_r \int \ell_t^* \exp \left( \alpha_{a,r} + z' \right) f_t^p \left( z'|a, r \right) dz'}{a^{-1} \sum_{a=1}^{A} \sum_{r=1}^{R} \omega_r \int \left( \ell_t^* \exp \left( \alpha_{a,r} + z' \right) \right)^{1-\tau_t} f_t^p \left( z'|a, r \right) dz'}.$$

Take logs:

$$\ln \lambda_t = \ln (1 - \varphi_t) + \ln a^{-1} \sum_{r=1}^{R} \omega_r \sum_{a=1}^{A} \exp \alpha_{a,r} \int \ell_t^* \exp z' f_t^p \left( z'|a, r \right) dz'$$

$$- \ln a^{-1} \sum_{r=1}^{R} \omega_r \sum_{a=1}^{A} \exp \left( (1 - \tau_t) \alpha_{a,r} \right) \int \left( \ell_t^* \exp z' \right)^{1-\tau_t} f_t^p \left( z'|a, r \right) dz'.$$  \hspace{1cm} (A.25)

Replace $\ln \lambda_t$ from (A.25) into (A.24), taking into account $G_t = \varphi_t Y_t$. Notice that $\varphi_t$ only appears in this expression. Maximizing the latter with respect to $\varphi_t$ implies that the welfare-maximizing provision of public good is always equal to a constant fraction of aggregate output:

$$\varphi_t^* = \frac{\chi}{1 + \chi},$$

for all $t$.

Q.E.D.
A.9 Lemma A1

In this section I show how to compute an integral of the form

\[ E_t \left[ \exp (bz') | a \right] = \int \exp (bz') f_t^p (z' | a) \, dz' \]

recursively, where for simplicity of notation I have dropped the index \( r \). Notice that, based on equation (15) with \( \rho_r = 1 \) we can write that

\[ bz' = bz + b\eta g_t (\varepsilon, a) \rightarrow \]

\[ \exp (bz') = \exp (bz) \exp (b\eta g_t (\varepsilon, a)). \]

Therefore:

\[ E_t \left[ \exp (bz') | a \right] = E_{t-1} \left[ \exp (bz) | a \right] E_t \left[ \exp (b\eta g_t (\varepsilon, a)) | a \right] \]

where

\[ E_t \left[ \exp (b\eta g_t (\varepsilon, a)) | a \right] = \theta E_t \left[ \exp (b\eta g_t (\varepsilon, a)) | a, \text{choose} \right] + \]

\[ + (1 - \theta) E_t \left[ \exp (b\eta g_t (\varepsilon, a)) | a, \text{exogenous reloc} \right], \tag{A.26} \]

so the two expectations are conditional on choosing to move or being exogenously relocated. For agents who can choose to relocate:

\[ E_t \left[ \exp (b\eta g_t (\varepsilon, a)) | a, \text{choose} \right] = p_t (a) E_t \left[ \exp (b\eta g_t (\varepsilon, a)) | a, M_t (a) = 0 \right] \]

\[ + (1 - p_t (a)) E_t \left[ \exp (b\eta g_t (\varepsilon, a)) | a, M_t (a) = 1 \right]. \]

To move forward, we need to use the following fact. Suppose that a random variable \( X \) is distributed as \( \text{EV} (\mu, \sigma) \), where \( \mu \) is the location parameter and \( \sigma \) the scale parameter. The
CDF of this random variable is
\[
\exp \left( - \exp \left( - \left( \frac{x - \mu}{\sigma} \right) \right) \right).
\]

Consider now the random variable \(\exp(X)\). Its CDF is
\[
\Pr (\exp(X) \leq z) = \Pr (X \leq \ln z) = \exp \left(- \exp \left(- \left( \frac{\ln z - \mu}{\sigma} \right) \right) \right) \\
= \exp \left(- \exp \left(- \left( \frac{\ln z^{\frac{1}{\sigma}}}{\exp(\mu)} - \ln \exp(\mu/\sigma) \right) \right) \right) \\
= \exp \left(- \left( \frac{z}{\exp(\mu)} \right)^{-\frac{1}{\sigma}} \right).
\]

This is the CDF of a Frechet with shape parameter \(\frac{1}{\sigma}\) and scale parameter \(\exp(\mu)\). Its mean is
\[
E [\exp(X)] = \exp (\mu + \ln \Gamma (1 - \sigma)).
\]

Consider each in isolation.

- First, conditional on an exogenous relocation, \(g_t (\varepsilon, a)\) is distributed as an EV(0, 1) with scale parameter 1 and location parameter 0, so that \(b \eta g_t (\varepsilon, a)\) is EV(0, \(b \eta\)). It follows that
  \[
  E_t [\exp (b \eta g_t (\varepsilon, a)) | a, \text{exogenous reloc}] = \Gamma (1 - b \eta).
  \]

- Second, to compute \(E_t [\exp (b \eta g_t (\varepsilon, a)) | a, M_t (a)]\), notice that, since \(g_t (\varepsilon, a)\) conditional on \(M_t (a) = 0\) is distributed as an EV with scale parameter 1 and location parameter \(\ln (\delta/p_t (a))\), \(b \eta g_t (\varepsilon, a)\) is EV with scale parameter \(b \eta\) and location parameter \(b \eta \ln (\delta/p_t (a))\). Thus:
  \[
  E_t [\exp (b \eta g_t (\varepsilon, a)) | a, M_t (a) = 0] = \exp (b \eta \ln (\delta/p_t (a)) + \ln \Gamma (1 - b \eta)).
  \]

- Third, since \(g_t (\varepsilon, a)\) conditional on \(M_t (a) = 1\) is distributed as an EV with scale parameter 1 and location parameter \(\ln (\delta/p_t (a)) + \kappa_r/ (\eta_r v_t^1 (a; r))\), \(b \eta g_t (\varepsilon, a)\) is EV
with scale parameter $b\eta$ and location parameter $b\eta [\ln (\delta/p_t (a)) + \kappa/ (\eta v^1_t (a))]$. Thus:

$$E_t [\exp (b\eta g_t (\varepsilon, a)) | a, M_t (a) = 1] = \exp (b\eta \ln (\delta/p_t (a)) + b\kappa/ v^1_t (a; r) + \ln \Gamma (1 - b\eta)) .$$

Thus, putting all together, \[A.26\] becomes:

$$E_t [\exp (b\eta g_t (\varepsilon, a)) | a] = \theta E_t [\exp (b\eta g_t (\varepsilon, a)) | a, \text{choose}] + (1 - \theta) E_t [\exp (b\eta g_t (\varepsilon, a)) | a, \text{exogenous reloc}]$$

$$= \theta \exp (\ln \Gamma (1 - b\eta)) + (1 - \theta) p_t (a) \exp (b\eta \ln (\delta/p_t (a)) + \ln \Gamma (1 - b\eta)) + (1 - \theta) (1 - p_t (a)) \exp (b\eta \ln (\delta/p_t (a)) + b\kappa/ v^1_t (a; r) + \ln \Gamma (1 - b\eta)) .$$

Collect:

$$E_t [\exp (b\eta g_t (\varepsilon, a)) | a] = \exp (\ln \Gamma (1 - b\eta)) \times \{\theta + (1 - \theta) p_t (a) \exp (b\eta \ln (\delta/p_t (a))) + (1 - \theta) (1 - p_t (a)) \exp (b\eta \ln (\delta/p_t (a)) + b\kappa/ v^1_t (a; r))\} .$$

Collect again:

$$E_t [\exp (b\eta g_t (\varepsilon, a)) | a] = \exp (\ln \Gamma (1 - b\eta)) \times \{\theta + \exp (b\eta \ln (\delta/p_t (a))) ((1 - \theta) p_t (a) + (1 - \theta) (1 - p_t (a)) \exp (b\kappa/ v^1_t (a; r)))\} .$$

Q.E.D.

\section*{B Details on Identification and Estimation}

In Section \[B.1\] I first show how the moving cost parameters are identified using a simple version of the model. Then, in Section \[B.2\] I show how to derive analytically the moment conditions used to estimate the model’s parameters.
B.1 Identification of Moving Cost Parameters

In order to illustrate the intuition behind identification of the moving cost parameters \((\kappa, \delta, \theta)\), consider the following example with three ages, \(\alpha = 3\). These correspond roughly to young \((a = 1)\), middle-age \((a = 2)\), and old \((a = 3)\). Assume that the data we observe are migration rates by age, denoted by \(\text{migdata}(a)\), for \(a = 1, 2, 3\). Therefore, there are three parameters and three data moments. Consistently with the full model, I assume that at age \(a = 3\), agents do not move if given the opportunity to choose, so \(p(3) = 1\). It follows that at \(a = 3\) all mobility is exogenous and equal to \(1 - \theta\). Thus, the parameter \(\theta\) is identified by \(\text{migdata}(3)\) according to

\[
1 - \theta^* = \text{migdata}(3) \rightarrow \theta^* = 1 - \text{migdata}(3).
\]

Consider now the model-implied migration rates at ages 1 and 2:

\[
1 - \theta p(1) = 1 - \frac{\theta}{1 + \delta^{-1} \exp\left(-\tilde{\kappa} (\beta^2 + \beta + 1)^{-1}\right)}, \quad (A.27)
\]

\[
1 - \theta p(2) = 1 - \frac{\theta}{1 + \delta^{-1} \exp\left(-\tilde{\kappa} (\beta + 1)^{-1}\right)}. \quad (A.28)
\]

These reflect the expression for \(p(a)\) in equation \([14]\). To simplify the notation and without loss of generality I have set \(K = 2\) and defined the modified moving cost parameter

\[
\tilde{\kappa} \equiv \frac{\kappa}{\eta (1 - \tau)}.
\]

In what follows I provide intuition on how to identify \(\tilde{\kappa}\). Once it is identified, it is straightforward to see how \(\kappa\) is identified.

Now, equalize the data moments with the model-implied moments in equation \((A.27)\).
(A.28) and rearrange them to obtain:

\[
\frac{\theta^*}{1 - \text{migdata}(1)} - 1 = \delta^{-1} \exp \left( -\frac{\tilde{\kappa}}{\beta^2 + \beta + 1} \right), \tag{A.29}
\]

\[
\frac{\theta^*}{1 - \text{migdata}(2)} - 1 = \delta^{-1} \exp \left( -\frac{\tilde{\kappa}}{\beta + 1} \right), \tag{A.30}
\]

where I have also denoted \( \theta \) by \( \theta^* \) to indicate that it is already identified (and so it can be treated as known). Taking the ratio of these two equations gives

\[
\frac{\frac{\theta^*}{1 - \text{migdata}(1)} - 1}{\frac{\theta^*}{1 - \text{migdata}(2)} - 1} = \exp \left( \tilde{\kappa} \left( \frac{1}{\beta + 1} - \frac{1}{\beta^2 + \beta + 1} \right) \right),
\]

where

\[
\frac{1}{\beta + 1} > \frac{1}{\beta^2 + \beta + 1}.
\]

It follows that, given \( \beta \) (which is set a-priori), the composite parameter \( \tilde{\kappa} \) is identified by the difference between migdata(1) and migdata(2). Younger individuals have a higher migration rate than older ones, so in the data migdata(1) > migdata(2). The larger the gap between these two data moments the larger \( \tilde{\kappa} \) has to be for the model to account for this pattern.

Last, the parameter \( \delta \) is identified by either equation (A.29) or (A.30) (since we have already used their ratio to identify \( \tilde{\kappa} \)). In either of them, it is straightforward to realize that a higher migration rate in the data has to be associated with a lower value of \( \delta \).

These intuitions explain why college-educated workers in my estimation (see Table 1) are characterized by a larger \( \kappa \) and a lower \( \delta \) than workers without a college degree. The reason is that they experience a larger decline in migration from younger to older ages (hence the higher \( \kappa \)) and a higher overall migration rate.
B.2 Moment Conditions

In this appendix I derive the moment conditions analytically. To keep the notation simple, I don’t include the index $r$ in the expressions of this section. The moment conditions implied by the model are:

\[
E \left[ M_i | i \in \Omega (a) \right] - (1 - \theta p (a)) = 0, \text{ for } a = 1, 2, \ldots, \bar{a},
\]

\[
E \left[ \ln y_i | i \in \Omega (a) \right] - m_{\ln y} (a) = 0, \text{ for } a = 1, 2, \ldots, \bar{a},
\]

\[
E \left[ (\ln y_i - E [\ln y_i | i \in \Omega (a)])^2 | i \in \Omega (a) \right] - \nu_{\ln y} (a) = 0, \text{ for } a = 1, 2, \ldots, \bar{a},
\]

where $M_i = 1$ if household $i$ migrates and zero otherwise, and $\Omega (a)$ is the set of households of age $a$.

B.2.1 Migration rate

The migration rate by age for a household is:

\[
1 - \theta p (a) = 1 - \frac{\theta}{1 + (K - 1) \delta^{-1} \exp \left( - (\kappa / (\eta (1 - \tau))) (1 - \beta) (1 - \beta^{\pi - a + 1})^{-1} \right)}, \quad (A.31)
\]

where $p (a)$ is found by imposing a constant $\tau$ in equation (14).

B.2.2 Cross-Sectional Mean of Log Income by Age

Result The cross-sectional mean of the distribution of log labor income ($\ln y = z' + \alpha_a + \ln \ell^*$) for type $r$ households at age $a$ is, by definition, equal to:

\[
m_{\ln y} (a) = m_{z'} (a) + \alpha_a + \ln \ell^*,
\]

where $m_{z'} (a)$ is the mean of the distribution of log productivity $z'$ at age $a$ after migration. Using the results in Proposition 4, it can be shown (see the derivation in the next subsection) that $m_{z'} (a)$ denote by $m_z (a)$ the average log productivity at the beginning of the period (before migration) among agents of age $a$. 

83
to be recursively determined as follows for \( a \geq 1 \):

\[
m_{z^*(a)} = \rho m_{z^*(a-1)} + \eta (\gamma + \ln \delta + \Delta (a)) ,
\]

starting from \( m_{z^*(0)} = \mu \). The term \( \alpha_a + \eta (\gamma + \ln \delta) \) represents the growth rate of labor productivity in the one-location version of the model. The term \( \Delta (a) \) in (A.32) represents the average growth rate of log income at age \( a \) that can be attributed to voluntary and involuntary migration:

\[
\Delta (a) \equiv -\theta \ln p (a) + \theta \left( \frac{1 - p (a)}{\eta \nu^1 (a)} \right) - (1 - \theta) \ln \delta.
\]

**Derivation** The log income of a household of age \( a \) is denoted by \( \ln y_a = z_a + \ln \ell^* \). The age \( a = 1 \) income is observed after the agent’s initial moving choices. The agent draws its initial \( z \) from \( f(z) \), observes \( \varepsilon \) and then makes its initial moving choice producing \( z_0' \). The latter is such that at \( a = 1 \):

\[
z_1' = \rho z + \eta g (\varepsilon, 1) .
\]

(A.33)

For subsequent ages \( a > 1 \), the law of motion of log income is:

\[
z_a' = \rho z_{a-1} + \eta g^* (\varepsilon, a)
\]

(A.34)

for a household who can choose her location and

\[
z_a' = \rho z_{a-1} + \eta \hat{g} (\varepsilon, a),
\]

(A.35)

for a household who is relocated exogenously, where I use the “hat” symbol to denote exogenous mobility. Let \( Z \) be a dummy variable for being able to choose whether to migrate or not \( (Z = 1) \) or being relocated exogenously \( (Z = 0) \). Thus:

\[
z_a' = \rho z_{a-1} + Z \eta g^* (\varepsilon, a) + (1 - Z) \eta \hat{g} (\varepsilon, a).
\]
The cross-sectional mean evolves as follows:

\[
E[z'_a] = \rho E[z_{a-1}] + \eta \theta, E[Z^g (\varepsilon, a) | Z = 1] + \eta (1 - \theta, r) E[(1 - Z) \hat{g} (\varepsilon, a) | Z = 0] \\
= \rho E[z_{a-1}] + \eta \theta, E[g^* (\varepsilon, a)] + \eta (1 - \theta, r) E[\hat{g} (\varepsilon, a)].
\]

Taking into account Proposition 4:

\[
E[z'_a] = \rho E[z_{a-1}] + \eta \theta, p(a) (\gamma + \ln \delta - \ln p(a)) + \\
+ \eta \theta, (1 - p(a)) \left( \gamma + \ln \delta - \ln p(a) + \frac{\kappa_r}{\eta v^1(a)} \right) + \eta (1 - \theta) \gamma.
\]

This can be simplified as:

\[
E[z'_a] = \rho E[z_{a-1}] + \eta \theta, (\gamma + \ln \delta - \ln p(a)) + \theta (1 - p(a)) \frac{\kappa_r}{v^1(a)} + \eta (1 - \theta) \gamma \\
= \rho E[z_{a-1}] + \eta \gamma + \eta \theta, (\ln \delta - \ln p(a)) + \theta (1 - p(a)) \frac{\kappa_r}{v^1(a)}.
\]

It is convenient to write it as:

\[
E[z'_a] = \rho E[z_{a-1}] + \eta (\gamma + \ln \delta) - \eta (1 - \theta) \ln \delta - \eta \theta \ln p(a) + \theta (1 - p(a)) \frac{\kappa_r}{v^1(a)}.
\]

Let \( \Delta (a) \) :

\[
\Delta (a) = -\theta \ln p(a) + \theta \frac{(1 - p(a)) \kappa}{\eta v^1(a)} - (1 - \theta) \ln \delta
\]

and write:

\[
E[z'_a] = \rho E[z_{a-1}] + \eta (\gamma + \ln \delta + \Delta (a)). \tag{A.36}
\]

At \( a = 1 \), the mean of \( z_1 \) is given by

\[
E[z_1] = \rho \mu + \eta (\gamma + \ln \delta + \Delta (1)),
\]

where \( \mu = E[z] \).
B.2.3 Cross-Sectional Variance of Log Income by Age

Result  The cross-sectional variance of log labor income for households of type \( r \) at age \( a \) is equal to the cross-sectional variance of log productivity \( v_{lny}(a) = v_{\omega}(a) \) because labor supply is age-independent. Using again Proposition 4, it can be shown that the cross-sectional variance of log productivity evolves recursively as follows (see the derivation in the next subsection):

\[
v_{\omega}(a) = \rho^2 v_{\omega}(a-1) + \eta^2 \left( \pi^2 / 6 + \Psi(a) \right),
\]

starting from \( v_{\omega}(0) = \sigma^2 \). The term \( \eta^2 \pi^2 / 6 \) in (A.37) represents the “within” component of the variance of income growth, which, by Proposition 4 is the same for all households independently of whether they choose to stay put, to migrate, or are exogenously relocated. The term \( \Psi(a) \) in equation (A.37) is formally defined in equation (A.40). It represents the “between” component of the variance of household productivity growth and it refers to differences between voluntary movers, stayers, and involuntary movers in this dimension (see Proposition 4).

Derivation  The evolution of log productivity is given by:

\[
\begin{align*}
z'_a &= \rho z_{a-1} + Z \eta g^*(\varepsilon, a) + \eta (1 - Z) \hat{g}(\varepsilon, a) \\
&= \rho z_{a-1} + \eta Z M(\varepsilon, a) g^*(\varepsilon, a) + \eta Z (1 - M(\varepsilon, a)) g^*(\varepsilon, a) + \eta (1 - Z) \hat{g}(\varepsilon, a)
\end{align*}
\]

The variance at subsequent ages is therefore:

\[
V[z'_a] = V[\rho z_{a-1} + \eta Z M(\varepsilon, a) g^*(\varepsilon, a) + \eta Z (1 - M(\varepsilon, a)) g^*(\varepsilon, a) + (1 - Z) \eta \hat{g}(\varepsilon, a)]
= \rho^2 V[z_{a-1}] + \eta^2 V[Z M(\varepsilon, a) g^*(\varepsilon, a) + Z (1 - M(\varepsilon, a)) g^*(\varepsilon, a) + (1 - Z) \hat{g}(\varepsilon, a)].
\]
From the law of total variance:

\[ V[ZM(\varepsilon, a)g^*(\varepsilon, a) + Z(1 - M(\varepsilon, a))g^*(\varepsilon, a) + (1 - Z)\hat{g}(\varepsilon, a)] = \text{(A.38)} \]

\[
E_{Z,M}\{V_\varepsilon[ZM(\varepsilon, a)g^*(\varepsilon, a) + Z(1 - M(\varepsilon, a))g^*(\varepsilon, a) + (1 - Z)\hat{g}(\varepsilon, a)|a, Z, M]\}
\]

\[
+ V_{Z,M}\{E_\varepsilon[ZM(\varepsilon, a)g^*(\varepsilon, a) + Z(1 - M(\varepsilon, a))g^*(\varepsilon, a) + (1 - Z)\hat{g}(\varepsilon, a)|a, Z, M]\},
\]

where I shortened \( M(\varepsilon, a) \) to \( M \).

Consider the term inside \( V_\varepsilon[.] \). Notice that for all combinations of \((Z, M(\varepsilon, a))\):

\[ V_\varepsilon[ZM(\varepsilon, a)g^*(\varepsilon, a) + Z(1 - M(\varepsilon, a))g^*(\varepsilon, a) + (1 - Z)\hat{g}(\varepsilon, a)|a, Z, M] = \eta^2\pi^2/6 \]

because of Proposition 4. Thus, the average is also \( \eta^2\pi^2/6 \):

\[
E_{Z,M}\{V_\varepsilon[ZM(\varepsilon, a)g^*(\varepsilon, a) + Z(1 - M(\varepsilon, a))g^*(\varepsilon, a) + (1 - Z)\hat{g}(\varepsilon, a)|a, Z, M]\}
\]

\[ = \eta^2\pi^2/6. \]

Consider now the second term in the decomposition \( \text{(A.38)} \), which I call the “between-variance”, \( V_{Z,M}\{.\} \). Let

\[ X(Z, M) \equiv E_\varepsilon[ZMg^*(\varepsilon, a) + Z(1 - M(\varepsilon, a))g^*(\varepsilon, a) + (1 - Z)\hat{g}(\varepsilon, a)|a, Z, M]. \]

We need to compute \( V_{Z,M}[X(Z, M)] \). To compute the latter consider that there are 3 possible mutually exclusive combinations of \((Z, M)\):

\[ Z = 1, M = 0, \]
\[ Z = 1, M = 1, \]
\[ Z = 0. \]

Then:
• 1. $Z = 1, M = 0$ occurs with probability $\theta (1 - p (a))$:

$$X (1, 0) = \gamma - \ln (p (a) / \delta) + \frac{\kappa}{\eta v^1 (a)}$$

• 2. $Z = 1, M = 1$ occurs with probability $\theta p (a)$:

$$X (1, 1) = \gamma - \ln (p (a) / \delta)$$

• 3. $Z = 0$ occurs with probability $1 - \theta$:

$$X (0, M) = \gamma.$$

The mean can be written as in equation (A.36):

$$\theta (1 - p (a)) \left( \gamma - \ln (p (a) / \delta) + \frac{\kappa}{\eta v^1 (a)} \right) + \theta p (a) (\gamma - \ln (p (a) / \delta)) + (1 - \theta) \gamma$$

$$= \gamma + \ln \delta - \ln \delta + \theta \ln \delta - \theta \ln p (a) + \theta \frac{\kappa (1 - p (a))}{\eta v^1 (a)}$$

$$= \gamma + \ln \delta + \Delta (a).$$

(A.39)

The variance $V_{Z, M} [X (Z, M)]$ is then:

$$V_{Z, M} [X (Z, M)] = \theta (1 - p (a)) \left( \gamma - \ln p (a) + \ln \delta + \frac{\kappa}{\eta v^1 (a)} - (\gamma + \ln \delta + \Delta (a)) \right)^2 +$$

$$+ \theta p (a) (\gamma - \ln p (a) + \ln \delta - (\gamma + \ln \delta + \Delta (a)))^2 +$$

$$+ (1 - \theta) (\gamma - (\gamma + \ln \delta + \Delta (a)))^2$$

Simplify:

$$V_{Z, M} [X (Z, M)] = \theta (1 - p (a)) \left( \ln p (a) + \Delta (a) - \frac{\kappa}{\eta v^1 (a)} \right)^2 + \theta p (a) (\ln p (a) + \Delta (a))^2 +$$

$$+ (1 - \theta) (\ln \delta + \Delta (a))^2.$$
Define:

\[ \Psi(a) \equiv \theta (1 - p(a)) \left( \ln p(a) + \Delta(a) - \frac{\kappa}{\eta \psi^2(a)} \right)^2 + \theta p(a) \ln p(a) + \Delta(a)^2 + \theta (1 - \theta) (\ln \delta + \Delta(a))^2. \]

Putting everything together, we can write (A.38) as:

\[ V[z_{a-1}] = \sigma^2 V[z] + \eta^2 \pi^2 / 6 + \eta^2 \Psi(a). \]

The cross-sectional variance of log productivity at age \( a = 1 \) is obtained from equation (A.33):

\[ V[z_1] = \sigma^2 V[z] + \eta^2 V[g(\epsilon, 1)], \]

where \( V[z] = \sigma^2 \) and

\[ V[g(\epsilon, 1)] = \pi^2 / 6 + \Psi(1). \]

Q.E.D.

C Data Appendix

This data appendix contains information about the SIPP data and how I used it to analyze the properties of the income process upon migration (Section C.1). In Section C.2 I provide details related to the measurement of \( \tau \) across countries using OECD data and on the estimates of the cross-country correlation between \( \tau \) and internal migration rates.

C.1 SIPP Data

C.1.1 Sample Selection and Estimation

I use monthly household-level data from the 1996, 2001, 2004 and 2008 panels of the SIPP. I apply the following selection criteria to select the sample of households in a way that is
consistent with the sample used to estimate the model and that makes sure that interstate moves are not due to the split of a households (due to say, a divorce) relative to its initial composition:

- 1. consider only households where the number of families is equal to one;
- 2. the household reference person must be of ages 25-59 as of the first wave and reference month of the survey;
- 3. the household reference person must be present in the household at all periods in which the household is interviewed;
- 4. longitudinal panel weights must be present in the household record;
- 5. the individuals of ages 25-59 as of the first wave and reference month of the survey must be present in the household at all periods in which the household is interviewed.

Since the 1996 and 2001 panels do not distinguish between Maine and Vermont and between North Dakota, South Dakota, and Wyoming, I have recoded the 2004 and 2008 panels to reflect this classification of states.

In estimating equation (22) I construct a household-level weight equal to a household’s longitudinal panel weight $\omega_i$ multiplied by a factor that adjusts for differences between the age distribution in the data and in the model. Specifically, I count the relative frequency $p_a$ of observations where household heads have age $a$ in the data. A household weight in regression (22) is therefore $\omega_i/p_a$ if household $i$ is of age $a$.

C.1.2 Earning gains from migration at different horizons

In order to determine whether, upon migration, a household might first experience a drop in earnings and then, at a longer horizons, a gain, I re-estimate equation (22) replacing the main effect of the variable $Post_{i,m}$ with two variables that denote an horizon of a year ($Post_{i,m}^{12-}$) and an horizon longer than a year ($Post_{i,m}^{12+}$). Notice that the SIPP panels track
households for three years in panels 1996 and 2004, two years in the 2001 panel and 4 years in the 2008 panel. In particular, the regression equation becomes:

\[
\ln y_{i,a,s,m} = M_i \times \text{Post}_{i,m}^{12-} \times \xi_1 + M_i \times \text{Post}_{i,m}^{12+} \times \xi_2 + M_i \times \text{Post}_{i,m} \times \xi_3 (a - 25) + \\
+ \xi_4 \ln n_{i,m}^{18-64} + \zeta_i + \zeta_a + \zeta_s + \zeta_m + \epsilon_{i,a,s,m}.
\]

The estimates are reported in Table A.1. There are two notable results. First, on average, a household experiences a gain in earnings within a year of an interstate move. Second, the gain is larger at a longer horizon and the difference is economically large for households with at least a college degree. I find no evidence of short-term declines in earnings associated with migration.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Less than college</th>
<th>College and above</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\xi_1)</td>
<td>0.100</td>
<td>0.149***</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>(\xi_2)</td>
<td>0.119***</td>
<td>0.226***</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>(\xi_3)</td>
<td>-0.008**</td>
<td>-0.007**</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1,209,340</td>
<td>652,242</td>
</tr>
<tr>
<td>Number of households</td>
<td>36,154</td>
<td>18,146</td>
</tr>
</tbody>
</table>

Table A.1: Estimates of the impact of interstate migration on household earnings based on equation (A.41). Standard errors clustered at the household-level in the parenthesis. Data source: SIPP 1996, 2001, 2004, 2008 panels. Regressions control for household, age, state, and month-year fixed effects. * denotes a p-value <0.10; ** <0.05 and *** <0.01.

C.1.3 Volatility of income process

I have used the SIPP data to test the hypothesis that the volatility of the income process might be higher after a migration event. To do so, I computed the estimated residuals
\( \hat{\varepsilon}_{i,a,s,m} \) from equation \([22]\) for each household that ever moves in the sample period. For each household \(i\) in the sample, I then computed the standard deviation of \( \hat{\varepsilon}_{i,a,s,m} \) before and after an interstate move:

\[
\hat{sd}_{i}^{\text{post}} = \sqrt{\sum_{m} Post_{i,m} \left( \hat{\varepsilon}_{i,a,s,m} - \hat{\varepsilon}_{i,a,s,m}^{\text{post}} \right)^2},
\]

\[
\hat{sd}_{i}^{\text{pre}} = \sqrt{\sum_{m} (1 - Post_{i,m}) \left( \hat{\varepsilon}_{i,a,s,m} - \hat{\varepsilon}_{i,a,s,m}^{\text{pre}} \right)^2},
\]

where \(Post_{i,m}\) is defined in the main text, \(\hat{\varepsilon}_{i,a,s,m}^{\text{post}}\) and \(\hat{\varepsilon}_{i,a,s,m}^{\text{pre}}\) are sample means of \(\hat{\varepsilon}_{i,a,s,m}\) for a given household after and before an interstate move. This procedure gives me two cross-sectional (across households) distributions of standard deviations (before and after geographic mobility) for each education group (less than college and college and above). The distribution of these estimates has properties summarized in Table A.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Education group</th>
<th>Mean</th>
<th>Std. Err</th>
<th># of obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{sd}_{i}^{\text{pre}})</td>
<td>less than college</td>
<td>0.333</td>
<td>0.011</td>
<td>686</td>
</tr>
<tr>
<td>(\hat{sd}_{i}^{\text{post}})</td>
<td>less than college</td>
<td>0.326</td>
<td>0.012</td>
<td>686</td>
</tr>
<tr>
<td>(\hat{sd}<em>{i}^{\text{post}} - \hat{sd}</em>{i}^{\text{pre}})</td>
<td>less than college</td>
<td>-0.007</td>
<td>0.014</td>
<td>686</td>
</tr>
<tr>
<td>(\hat{sd}_{i}^{\text{pre}})</td>
<td>college and above</td>
<td>0.334</td>
<td>0.011</td>
<td>822</td>
</tr>
<tr>
<td>(\hat{sd}_{i}^{\text{post}})</td>
<td>college and above</td>
<td>0.312</td>
<td>0.012</td>
<td>822</td>
</tr>
<tr>
<td>(\hat{sd}<em>{i}^{\text{post}} - \hat{sd}</em>{i}^{\text{pre}})</td>
<td>college and above</td>
<td>-0.028</td>
<td>0.014</td>
<td>822</td>
</tr>
</tbody>
</table>

Table A.2: Properties of the distribution of estimated household-level standard deviations of residuals.

I then use a t-test to test the null hypothesis that

\[ E[\hat{sd}_{i}^{\text{post}}] = E[\hat{sd}_{i}^{\text{pre}}], \]
against the alternative that the standard deviation of the innovations is higher after a move:

\[ E [sd_i^{\text{post}}] > E [sd_i^{\text{pre}}]. \]

For both groups (college and non-college), the test fails to reject the null at any conventional level of statistical significance.

C.2 Tax Progressivity and Internal Migration Across Countries

C.2.1 Computing tax progressivity

Assume that household market income \( y \) is distributed lognormally in the population with variance parameter \( v_y^2 \). It follows that after-tax and transfers income \( \tilde{y} = \lambda y^{1-\tau} \) is also lognormal with variance parameter \( (1 - \tau)^2 v_y^2 \). For a lognormal distribution with variance parameter \( v_y^2 \), the Gini coefficient is given by:

\[
\text{GINI} (y) = \text{erf} \left( 0.5v_y \right),
\]

(A.42)

where \( \text{erf}(x) \) denotes the error function. The latter is defined as:

\[
\text{erf} (x) = 2\Phi \left( \sqrt{2}x \right) - 1,
\]

where \( \Phi (x) \) is the cumulative distribution function of the normal distribution. Similarly, the Gini coefficient of post-redistribution income \( \tilde{y} = \lambda y^{1-\tau} \) is given by:

\[
\text{GINI} (\tilde{y}) = \text{erf} \left( 0.5 \left( 1 - \tau \right) v_y \right).
\]

(A.43)

The OECD provides country-level data on GINI\((y)\) and GINI\((\tilde{y})\). The strategy I follow is therefore to invert equation (A.42) to obtain \( v_y \) and use the latter together with the inverse of equation (A.43) to obtain a country-level estimate of \( \tau \). Putting these relationships together

\[ \text{The assumption of lognormality is not necessary here although it facilitates some computations.} \]
I obtain the following estimator of $\tau$:

$$\tau = 1 - \frac{\Phi^{-1}[0.5 (1 + \text{GINI}(\bar{y}))]}{\Phi^{-1}[0.5 (1 + \text{GINI}(y))]},$$

with $\Phi^{-1}(.)$ denoting the inverse of the normal cumulative distribution function. Intuitively, this procedure assigns to a country a higher level of $\tau$ the smaller its Gini coefficient for post-redistribution income is relative to the Gini for market income. Interesting special cases are $\text{GINI}(\bar{y}) = \text{GINI}(y)$ in which case $\tau = 0$ and $\text{GINI}(\bar{y}) = 0$ in which case $\tau = 1$.

The data used to estimate $\tau$ are from the OECD’s Income Distribution Database, available at http://www.oecd.org/social/income-distribution-database.htm.

C.2.2 Country-Level Internal Migration Data

The 21 countries in my sample are: Australia, Austria, Canada, Czech Republic, France, Germany, Greece, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, UK, USA. The countries with less than 5 TL2 are Belgium, Denmark, Finland, Slovakia, Slovenia, Iceland. The data for Estonia refer to a lower geographic level, TL3, so I don’t include it in the analysis. The internal migration data sources are OECD Employment Outlook 2000 (Table 2.12) for the period 1980-1998, OECD Employment Outlook (2005, Table 2.12) for the period around 2003, and Regions at a Glance (2013, Figure 4.10) for the period around 2010.

C.2.3 Summary Statistics

Table A.3 summarizes the information on migration rates, $\tau$ and average personal income tax (net of cash transfers) on wage income at the country level. The cross-country correlation between tax progressivity and average net taxes relative to wage income is 0.12 (p-value 0.61).
Table A.3: Summary statistics on country-level average (1980–2010) migration rates, the tax progressivity measure $\tau$, and the average personal income taxes (net of cash transfers) relative to wage income. The sample contains 21 countries.

C.2.4 Regression Results and Robustness Checks

Table A.4 reports the results of various specifications of the regression equation that underlies the discussion around Figure 1 in the main text:

\[
\text{migration rate}_j = \alpha_0 + \alpha_1 \tau_j + \alpha_2 \text{average tax rate}_j + \alpha_3 X_j + u_j \tag{A.44}
\]

where $j$ denotes a country, $X_j$ denotes the control variables (if any), and $u_j$ is the error term. The tax variables are country averages for the period 1980–2013. The migration rate is a country average for 1980-2010. The data does not represent a balanced panel.

D Welfare Analysis

In this section I present additional details regarding the computation of optimal policies discussed in Section 6 of the paper. Section D.1 shows the relationship between $\tau$ and the average marginal tax rates in the economy. Sensitivity analysis for the optimal tax progressivity results with respect to the model’s parameters is contained in Section D.2. Section D.3 describes the algorithm used to compute optimal policy, and Section D.4 presents
Table A.4: Estimates of the parameters $\alpha_1$ and $\alpha_2$ in equation (A.44). Column (1) is the benchmark reported in Figure [1]. Column (2) weights each country by its population. Column (3) includes as controls log population and the number of TL2 regions. Column (4) controls for the country’s average personal income tax (net of cash transfers) relative to wages. Column (5) is a panel regression with country-specific fixed effects. Significance levels: *** $p < 0.01$, ** $p < 0.05$. In column (5), standard errors are clustered at the country-level.

the results from decomposing the aggregate welfare effects into group-specific effects.

D.1 Average Marginal Tax Rates

In this section I show that the average income-weighted marginal tax rate is such that

$$
\int T'(y) \frac{y}{Y} f(y) \, dy = 1 - (1 - \tau) (1 - \varphi),
$$

(A.45)

where $f(y)$ is an arbitrary cross-sectional density of income and

$$
Y = \int y f(y) \, dy
$$

is aggregate income. Recall from the definition of $T(y) = y - \lambda y^{1-\tau}$ that

$$
T'(y) = 1 - \lambda (1 - \tau) y^{-\tau}.
$$
Thus:

\[
\int T'(y) y f(y) \, dy = Y - (1 - \tau) \int \lambda y^{1-\tau} f(y) \, dy
\]

\[
= Y - (1 - \tau) \int (y - T(y)) f(y) \, dy
\]

\[
= Y - (1 - \tau) (Y - G),
\]

where

\[
G = \int T(y) f(y) \, dy
\]

from the government’s budget constraint. Since \( G = \varphi Y \), replacing into (A.46) we get the result in equation (A.45).

\[\text{(A.46)}\]

\[\text{D.2 Sensitivity Analysis for Optimal Tax Progressivity}\]

In this section I present and discuss the impact of varying a number of parameters for the optimal degree of tax progressivity. Section \[\text{D.2.1}\] focuses on the one-shot case, while Section \[\text{D.2.2}\] focuses on the optimal tax progressivity sequence.

\[\text{D.2.1 Optimal Tax Progressivity (One-shot)}\]

Table \[\text{A.5}\] focuses on the one-shot policy, discussed in Section \[\text{6.2.1}\]. In the version of the model with multiple locations and the opportunity to migrate, a smaller weight of the public good in utility increases the optimal \( \tau \) relative to the benchmark calibration because it reduces the magnitude of the externality associated with the fact that agents take \( \{G_i\}^{\infty}_{i=1} \) as given when they choose how much to work and whether to migrate or not. Similarly, a higher elasticity of labor tends to give rise to smaller optimal degrees of tax progressivity. The last two columns of Table \[\text{A.5}\] consider the case in which the economy is populated exclusively by either type 1 or type 2 agents. Interestingly, in an economy populated by type 1 agents only, the planner would select a higher level of tax progressivity than in an economy populated only by type 2 agents. The reason for this is that type 1 agents are much less
mobile than type 2 ones, so tax progressivity is relatively less distortionary in a type 1 agent economy. This intuition also explains why transitional dynamics considerations are much less relevant for the socially optimal \( \tau \) in a type 1, relative to a type 2, agent economy. In a type 1 agent economy, taking into account the transitional dynamics induces the planner to reduce tax progressivity by 4 percentage points, while in a type 2 agent economy the drop in tax progressivity associated with the transition is 25 percentage points (computed as \( 0.25 = 0.17 - (-0.08) \)).

<table>
<thead>
<tr>
<th>Bench.</th>
<th>No exog. mobility</th>
<th>Small ( G )</th>
<th>High labor el.</th>
<th>One type only</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_1 = \theta_2 = 1 )</td>
<td>( \chi = 0.01 )</td>
<td>( \zeta = 2 )</td>
<td>( \omega_1 = 1 )</td>
<td>( \omega_2 = 1 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Taking transition into account</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau^* )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Steady state only</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau^{ss} )</td>
</tr>
</tbody>
</table>

Table A.5: Optimal policy taking the transition into account (\( \tau^* \)) and considering only steady state welfare (\( \tau^{ss} \)). Each exercise varies one feature of the model at a time, while keeping all other parameters at their estimated value.

**D.2.2 Optimal Path of Tax Progressivity**

In this section I study the sensitivity of the optimal tax progressivity sequence to the parameter configurations considered in Section D.2.1. The results are in Figure A.1. With a smaller weight on the public good (\( \chi = 0.01 \)), the optimal policy calls for a higher degree of tax progressivity at each point in time, relative to the benchmark (see Figure A.1 “Small \( G \)” ). The intuition is analogous to the one discussed in Section D.2.1. Eliminating exogenous mobility by setting \( \theta_r = 1 \) for \( r = 1, 2 \), while keeping all other parameters the same as in
Figure A.1: Sensitivity analysis for optimal tax progressivity reform. The solid (purple) line represents the optimal path \( \{ \tau_t \} \). “Small \( G \)” refers to the case \( \chi = 0.01 \), a smaller value than in the benchmark. “No exogenous mobility” sets \( \theta_r = 1 \) for \( r = 1, 2 \). “High labor elasticity” sets \( \zeta = 2 \), (Frisch elasticity equal to 1). “Only type 1” is an economy populated only by type \( r = 1 \) agents. “Only type 2” is an economy populated only by type \( r = 2 \) agents.

the benchmark model has basically no impact on the time path of optimal tax progressivity (Figure A.1 “No exogenous mobility”). With a higher labor elasticity, the planner chooses a smaller degree of tax progressivity than in the benchmark through the transition and in the final limit (Figure A.1 “High labor elasticity”).

D.3 Numerical Algorithm to Compute Optimal Policy Path

The optimal time-varying path \( \{ \tau_t \} \) cannot, in general, be computed analytically and has to be found numerically. To do so, I use a piece-wise linear function to approximate the dependence of tax progressivity \( \tau_t \) on time \( t \):

\[
\tau_t = \tau^j + \left( \frac{t - t^j}{t^{j+1} - t^j} \right) (\tau^{j+1} - \tau^j) \quad \text{for} \quad t \in [t^j, t^{j+1}],
\]

\( j \)
where \( \{t^j\}_{j=1}^J \) are pre-specified time periods and \( \{\tau^j\}_{j=1}^J \) the corresponding values of the tax progressivity variable. After period \( t^J \), tax progressivity is assumed to be constant, so \( \tau_t = \tau^J \) for \( t \geq t^J \). The numerical algorithm searches over the \( J \) values \( \{\tau^j\}_{j=1}^J \) in order to maximize the welfare function \( W \), subject to the government’s budget constraint, and with \( \varphi^* \) given by equation (29). In implementing the numerical algorithm I choose 21 nodes (\( K = 21 \)) between periods \( t^1 = 1 \) and \( t^{21} = 100 \). The economy is assumed to be at its final steady state, corresponding to \( \tau^{21} \), in period \( t = 150 \). Notice that while \( \tau_t \) is constant after period \( t^{21} \), the economy takes longer to converge to the final steady state.

The rest of the algorithm describes how to find \( \{\tau^j\}_{j=1}^J \) in order to maximize the welfare function.

I proceed as follows:

1. Compute the initial steady state of the model given the benchmark value of \( \tau = 0.192 \) used in estimating the model’s parameter. The initial steady state provides the initial distribution of log productivity at \( t = 1 \), \( m_z(a;r) \) by age \( a \). Recall that the economy is in steady state in \( t = 1 \) and since productivity is a state variable it does not respond to changes in \( \tau \) that occur at \( t = 1 \).

2. To solve the model numerically, consider the welfare function (28). Assume that after \( T > t^J \) periods, the economy has converged to a new steady state characterized by time-invariant distributions of productivity by age and type of agents. This means that the welfare function can be approximated as:

\[
W = a^{-1} \sum_{r=1}^{\pi} \omega_r \sum_{a=1}^{\pi} E_z \{ E_{\varepsilon} [P_1(a,z,\varepsilon;r) \mid z,a,r] \} + \sum_{r=1}^{T} \sum_{a=1}^{\pi} \beta^{t-1} E_z \{ E_{\varepsilon} [P_t(1,z,\varepsilon;r) \mid z,1,r] \} + \sum_{r=1}^{\pi} \sum_{a=1}^{\pi} \beta^{t-1} E_{fss} \{ E_{\varepsilon} [P(1,z,\varepsilon;r) \mid z,1,r] \} , \tag{A.47}
\]

where \( E_{fss} \{ \cdot \} \) denotes an expectation computed with respect to the (endogenous) final

100
steady state’s distribution of productivity. Notice that the last term in (A.47) is equal to
\[
\bar{\sigma}^{-1} \frac{\beta T}{1 - \beta} \sum_{r=1}^{\bar{\tau}} \omega_r E_s^{\text{FSS}} \{ E_{\varepsilon} [P (1, z, \varepsilon; r) | z, 1, r] \}
\]
since the expectation is time-independent in the final steady state.

3. Define for simplicity of notation
\[
e_t (a, r) \equiv E_z \{ E_{\varepsilon} [P_t (a, z, \varepsilon; r) | z, a, r] \}
\]
so that
\[
W \approx \bar{\sigma}^{-1} \sum_{r=1}^{\bar{\tau}} \omega_r \sum_{a=1}^{\bar{\tau}} e_1 (a, r) + \sum_{r=1}^{\bar{\tau}} \omega_r \sum_{t=2}^{T} \beta^{t-1} e_t (1, r) + \frac{\beta T}{1 - \beta} \sum_{r=1}^{\bar{\tau}} \omega_r e_\infty (1, r).
\]
It is convenient to write the welfare function as the discounted sum of the flow utilities of all agents who are alive at a given point in time \(t\). From Proposition 1, the expected value function can be written as:
\[
e_t (a, r) = v^0_t (a; r) + v^1_t (a; r) \rho_r E_t [z | a, r] + v^1_t (a; r) \eta_r (\gamma - \theta_r \ln (p_t (a; r) / \delta_r)) - (1 - \theta_r) \kappa_r.
\]
Replace \(v^0_t (a; r)\) from equation (A.4) and replace the resulting term \(v^0_{t+1} (a + 1; r)\) as a function of \(e_{t+1} (a + 1, r)\), to obtain:
\[
e_t (a, r) = \pi^*_t (a) + \beta e_{t+1} (a + 1, r) + v^1_t (a; r) \rho_r E_t [z | a, r] + v^1_t (a; r) \eta_r (\gamma - \theta_r \ln (p_t (a; r) / \delta_r)) \\
- \beta v^1_{t+1} (a + 1; r) \rho_r E_{t+1} [z' | a + 1, r] - (1 - \theta_r) \kappa_r
\]
for \(a = 1, ..., \bar{\tau} - 1\), whereas at age \(\bar{\tau}\) we have
\[
e_t (\bar{\tau}, r) = v^0_t (\bar{\tau}; r) + v^1_t (\bar{\tau}; r) \rho_r E_t [z | \bar{\tau}, r] + v^1_t (\bar{\tau}; r) \eta_r (\gamma - \theta_r \ln (p_t (\bar{\tau}; r) / \delta_r)) - (1 - \theta_r) \kappa_r,
\]
with \(v^0_t (\bar{\tau}; r) = \pi^*_t (a)\) and \(v^1_t (\bar{\tau}; r) = (1 - \tau_t)\).
4. Use the “Amoeba” algorithm from Numerical Recipes to minimize \(-W\left(\{\tau^j\}_{j=1}^J\right)\)
with respect to the vector \(\{\tau^j\}_{j=1}^J\).

To compute the one-shot constant optimal policy I simply maximize the objective \(W\) with respect to a constant \(\tau\).

### D.4 Decomposition of Welfare Effects

Who gains and who loses from the optimal policy during the transition? In order to answer this question it is useful to decompose the expression for \(s^*\) into its disaggregated components. It is useful to distinguish between agents who are alive at time \(t = 1\), when the new policy is announced and agents who are born in later years \(t > 1\). Consider first the former group and notice that the initial distribution of productivity \(z\) among agents who are alive at \(t = 1\) is the same in the benchmark steady state and in the economy under the optimal policy. Let \(s^*_1(a, r)\) denote the equivalent variation for an agent of type \(r\) who, at \(t = 1\), is \(a\)-years old, and has average (for her age group and type) beginning-of-the-period log productivity \(z\). Consider now agents born in \(t > 1\). These agents’ initial productivity at age 1 is drawn from the exogenous density \(f(z|1, r)\), which is the same across the two economies (benchmark and reform). Denote by \(s^*_t(1, r)\) the equivalent variation for an age 1 agent with average log productivity born in period \(t\). The disaggregated equivalent variations \(s^*_1(a, r)\) and \(s^*_t(1, r)\) represent the proportional increase in remaining lifetime’s consumption that makes this average-productivity agent indifferent between the allocation in the benchmark economy and the one under the optimal policy. Proposition [A.1] provides the basis for the decomposition of \(s^*\) into its components for various demographic groups.

**Proposition A.1 (Decomposition of equivalent variation)** The expression for the equivalent variation \(s^*\) may be decomposed as follows:

\[
s^* \approx \sum_{r=1}^R \sum_{a=1}^\Pi \sum_{t=1}^\infty w_t(a, r) s^*_t(a, r), \tag{A.48}
\]
where

\[ w_t(a, r) = \begin{cases} \frac{(1 - \beta^{a+1})}{\omega_r/\bar{a}} & \text{if } t = 1 \\ \frac{(1 - \beta^a)}{\beta^{t-1} \omega_r/\bar{a}} & \text{if } t > 1 \text{ and } a = 1 \\ 0 & \text{otherwise} \end{cases} \]  

(A.49)

and \( \sum_{r=1}^{T} \sum_{a=1}^{\bar{a}} \sum_{t=1}^{\infty} w_t(a, r) = 1. \)

Proof:

Start from equation (30) and replace \( W^* \) and \( W^{\text{bench}} \):

\[
\ln (1 + s^*) = 
\frac{(1 - \beta)}{\bar{a}} \sum_{r=1}^{T} \omega_r \sum_{a=1}^{\bar{a}} \left\{ E_z \{ E_e [P^*_t(a, z, \epsilon; r) | z, a, r] \} - E_z \{ E_e [P^{\text{bench}}(a, z, \epsilon; r) | z, a, r] \} \right\} -
\frac{(1 - \beta)}{\bar{a}} \sum_{r=1}^{T} \sum_{a=1}^{\bar{a}} \sum_{t=2}^{\infty} \beta^{t-1} \left\{ E_z \{ E_e [P^*_t(1, z, \epsilon; r) | z, 1, r] \} - E_z \{ E_e [P^{\text{bench}}(1, z, \epsilon; r) | z, 1, r] \} \right\}.
\]

By definition, \( s^*_t(a, r) \) satisfies the following condition:

\[
E_z \{ E_e [P^*_t(a, z, \epsilon; r) | z, a, r] \} + \left( \frac{1 - \beta^{a+1}}{1 - \beta} \right) \ln (1 + s^*_t(a, r)) = E_z \{ E_e [P^*_t(a, z, \epsilon; r) | z, a, r] \}
\]

(A.50)

where \( P^{\text{bench}} \) is the value function in the benchmark steady state and \( P^*_t \) the value function along the transition. The term that multiplies the logarithm in the previous equation reflects the fact that agents have different remaining lifetimes according to their age.

Apply definition (A.50) so that

\[
\ln (1 + s^*) = \frac{(1 - \beta)}{\bar{a}} \sum_{r=1}^{T} \omega_r \sum_{a=1}^{\bar{a}} (1 - \beta^{a+1}) \ln (1 + s^*_1(a, r)) -
\frac{(1 - \beta)}{\bar{a}} \sum_{r=1}^{T} \omega_r \sum_{t=2}^{\infty} \beta^{t-1} (1 - \beta^t) \ln (1 + s^*_1(1, r)) .
\]

Use the approximation \( \ln (1 + x) \approx x \) for \( x \) small to verify that the equation above leads to equation (A.48), taking the definition of the weights (A.49) into account.
Q.E.D.

Proposition A.1 states that $s^*$ is approximately equal to the average of the age, type and calendar date equivalent variations $s^*_t(a,r)$. The weights $w_t(a,r)$ reflect two distinct factors. The first is the remaining lifetime of an agent. For example, for those agents alive at $t = 1$, the component $\left(1 - \beta^{\pi-a+1}\right)$ of its weight reflects the fact that, say, an agent of age 1 has $\pi$ periods to live, while an agent of age $\pi$ only has one period to live. The second component of a weight $w_t(a,r)$ reflects the agent’s weight in the planner’s social welfare function. In addition to the relative size $\omega_r/\pi$ of each demographic group $(a,r)$, the planner discounts the utilities of agents born at later dates. Hence, the term $\beta^{t-1}$ in the weight received by an age 1 agent in $t > 1$.

Table A.6 uses the result in Proposition A.1 to decompose $s^*$ into a number of subcomponents by age, type and period. Specifically, the entries in the table denote weighted averages of equivalent variations $s^*_t(a,r)$ over the specified combinations of $(a,r,t)$. The weighted sum of all entries corresponds to the equivalent variation reported in Table 4 ($s^* = 0.706\%$ and $1.09\%$ for the one-shot and the progressivity path reforms, respectively). For example, the entry 3.24% corresponding to $r = 1$, $t = 1$ and $a$ in $1 - 10$ for the one-shot reform represents the weighted average of the equivalent variations of type 1 agents, of ages between 1 and 10 at $t = 1$.

This decomposition reveals a number of interesting results. First, relative to the benchmark economy, both optimal policies (one-shot and sequence of tax progressivity) favor type 1 agents and hurt type 2 agents. The former group experiences positive gains both in period 1 as well as in subsequent periods. Second, the one shot reform tends to produce more evenly distributed welfare gains and losses across various age groups and time periods than the reform that optimizes over the transition. The latter produces larger gains and losses for the cohorts who are alive at the time of the reform and smaller welfare effects for the cohorts that are born after period 10. Third, the impact of each reform is non-monotonic across various demographic groups. For example, the one-shot reform produces larger welfare losses for the middle-aged group (ages 11-20) of type 2 agents than for the younger (1-10) and older
<table>
<thead>
<tr>
<th>time (t) periods</th>
<th>1</th>
<th>2-10</th>
<th>11-20</th>
<th>21-35</th>
<th>36+</th>
</tr>
</thead>
<tbody>
<tr>
<td>age (a) groups</td>
<td>1-10 11-20 21-35</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>steady state welfare maximization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 1</td>
<td>-3.20</td>
<td>-3.03</td>
<td>-3.37</td>
<td>-3.18</td>
<td>-2.78</td>
</tr>
<tr>
<td>r = 2</td>
<td>1.21</td>
<td>1.25</td>
<td>0.73</td>
<td>1.23</td>
<td>1.65</td>
</tr>
<tr>
<td>optimal one-shot reform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 1</td>
<td>3.24</td>
<td>3.04</td>
<td>3.67</td>
<td>3.06</td>
<td>2.21</td>
</tr>
<tr>
<td>optimal sequence of tax progressivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 1</td>
<td>4.75</td>
<td>4.40</td>
<td>5.61</td>
<td>4.03</td>
<td>1.84</td>
</tr>
<tr>
<td>r = 2</td>
<td>-3.06</td>
<td>-5.38</td>
<td>-5.54</td>
<td>-1.47</td>
<td>-1.43</td>
</tr>
<tr>
<td>welfare weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 1</td>
<td>0.12</td>
<td>0.09</td>
<td>0.06</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>r = 2</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table A.6: Decomposition of equivalent variation $s^*$. Entries for the first three panels are average equivalent variations $\sum_{a,r,t} w_t(a,r) s^*_t(a,r) / \sum_{a,r,t} w_t(a,r)$ for the stated combinations of $(a,r,t)$ (expressed in %). The welfare weights in the fourth panel are the relevant sum of weights $\sum_{a,r,t} w_t(a,r)$. 

105
(21-35) groups. This is due to the fact that the middle-aged group is, for reasons associated with the exogenous process $\alpha_{a,r}$, at the peak of earnings capacity when the reform goes into effect.

## E Extensions of the Model

This section consider four extensions of the model that are discussed in the main text.

### E.1 CRRA utility function

The model in the paper assumes a logarithmic utility function. This assumption allows me to solve the model analytically. In order to convey some intuition about what would happen under a more general CRRA utility function, I focus on a simple static version of the model that captures its basic economic intuition. A household faces a choice between staying in a location and earning income $y$ or moving to another location and earning income $y \exp \epsilon$, where $\epsilon$ denotes the proportional, before-tax, income gain from moving. Relocation entails a utility cost $\kappa > 0$. The household maximizes utility from consumption, which equals after-tax income. The utility function is assumed to be CRRA:

$$u(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma},$$

for $\sigma > 0$. A household’s after tax-transfer income $\tilde{y}$ is given by:

$$\tilde{y} = \lambda y^{1-\tau}.$$

In this environment the household chooses to migrate if:

$$\frac{(\lambda y^{1-\tau} \exp [(1 - \tau) \epsilon])^{1-\sigma}}{1 - \sigma} - \frac{(\lambda y^{1-\tau})^{1-\sigma}}{1 - \sigma} > \kappa,$$  

(A.51)
or

\[
\frac{(\lambda y^{1-\tau})^{1-\sigma}}{1-\sigma} \left[ \exp \left[ (1 - \tau) (1 - \sigma) \epsilon \right] - 1 \right] > \kappa.
\]

Consider the approximation

\[
\exp \left[ (1 - \tau) (1 - \sigma) \epsilon \right] \approx 1 + (1 - \tau) (1 - \sigma) \epsilon
\]

so that the migration condition becomes (approximately):

\[
\lambda^{1-\sigma} y^{(1-\tau)(1-\sigma)} (1 - \tau) \epsilon > \kappa.
\]

There are three cases to consider (assuming \( \tau \in [0, 1) \)):

- **If \( \sigma = 1 \):**
  - the migration choice is independent of the level of income \( y \) and of the tax parameter \( \lambda \).
  - a linear income tax (\( \tau = 0 \) and \( \lambda \in (0, 1) \)) has no impact on migration choices.
  - the propensity to migrate is independent of an agent’s income \( y \).

- **If \( \sigma < 1 \):**
  - a higher linear income tax (a lower \( \lambda \) for given \( \tau \)) makes migration less likely to happen.
  - higher income (higher \( y \)) agents are more likely to move.
  - a higher rate of tax progressivity \( \tau \) would unambiguously reduce the incentives to move of high income agents.
  - a higher rate of tax progressivity \( \tau \) has ambiguous effects on the incentives to move of low income agents.

- **If \( \sigma > 1 \):**
- a higher linear income tax (a lower $\lambda$ for given $\tau$) makes migration more likely to happen.

- higher income (higher $y$) agents are less likely to move.

- a higher rate of tax progressivity $\tau$ would unambiguously reduce the incentives to move of low income agents.

- a higher rate of tax progressivity $\tau$ has ambiguous effects on the incentives to move of high income agents.

E.2 Heterogeneous locations

In the main text, I discussed a version of the model with ex-ante heterogeneous locations and argued that the model with homogeneous locations may be interpreted as the reduced-form of such model under an appropriate restriction relating housing rents to wages and local amenities (including taxes). In this section I sketch this extension of the model and derive this restriction. To keep the notation simple I don’t make differences in type and age explicit in describing the model. The results are robust to including this kind of heterogeneity as long as preferences are homogeneous across $r$ types.

In the more general model, agents care about consumption of goods $c$, housing $h$, leisure, and local amenities $\phi_k$:

$$u_k (c, h, \ell, G) = (1 - \psi) \ln c + \psi \ln h - \zeta^{-1} \ell^\zeta + \chi \ln G + \phi_k.$$  

Locations are ex-ante heterogeneous in terms of housing prices $p_{kt}$, rental rates per unit of productivity $w_{kt}$, and state-level proportional tax rates and the budget constraint takes the form:

$$c + p_{kt}h = \lambda_{kt} y_k^{1-\tau},$$

$$y_k = w_{kt} \ell \exp z,$$
with $\lambda_{kt}$ capturing any linear state-level tax scheme. It is straightforward to show that the optimal labor supply in this version of the model takes the same form as in equation (4) and that the indirect utility function takes the form:

$$u_{kt}^* (z') = \overline{u}_{kt}^* + (1 - \tau_t) z', \quad \text{(A.52)}$$

with

$$\overline{u}_{kt}^* = \text{constants} + \ln \lambda_{kt} + (1 - \tau) \ln w_{kt} - \psi \ln p_{kt} + \phi_k + (1 - \tau_t) \ln \ell_t^* - \zeta^{-1} (\ell_t^*)^\zeta + \chi \ln G_t.$$

Notice that equation (A.52) takes a form similar to the benchmark version of the model (5), with the only difference being that the intercept is now location-dependent. The indirect utility function in equation (A.52) collapses to the one in the benchmark version of the model if the following condition holds:

$$\ln \lambda_{kt} + (1 - \tau) \ln w_{kt} - \psi \ln p_{kt} + \phi_k = \text{constant}. \quad \text{(A.53)}$$

The economic intuition is straightforward: if housing prices in each location $k$ fully reflect local wages, taxes, and amenities, then individuals will be indifferent across all locations and the model with ex-ante homogeneous locations is a valid representation of their migration behavior. Of course, condition (A.53) involves a number of variables which we might view as endogenous, such as housing rents and wages. To fully close the model we might want to impose equilibrium on local labor and housing markets. I leave this extension to future research.

### E.3 Asymmetric information about home and foreign locations - version with Epstein-Zin utility

In this section I describe how to modify the benchmark model in order to illustrate the possibility that progressive taxation might increase geographic mobility. To show this I
modify the benchmark model in two dimensions. First, the agent should observe the shock in the current location, but not in the foreign ones. Uncertainty about foreign productivity might open the door for tax progressivity to increase geographic mobility through increased provision of insurance. It turns out, however, that this asymmetry is not sufficient per se due to the unit coefficient of relative risk aversion implied by logarithmic utility. The second ingredient would therefore be a coefficient of relative risk-aversion larger than one. In order to preserve the tractability associated with logarithmic utility and consider this extension one needs to consider an Epstein-Zin utility function. In this section I consider such modification of the benchmark model, in which agents make moving decisions after observing the home idiosyncratic shock $\varepsilon_h$, but before observing the shocks in the other locations. In this version, the decision problem today (setting for simplicity $\theta_r = 1$, one type of agent, and considering only two locations $K = 2$, home $h$ and foreign $k$) is:

\[
P_t(a, z, \varepsilon, h) = \max \left\{ \exp V_t(a, \alpha_a + \rho z + \eta \varepsilon_h, h), E_{\varepsilon_h} \left[ \left( \exp V_t(a, \alpha_a + \rho z + \eta \varepsilon_k, k) \right)^{1-\phi} \right]^{\frac{1}{1-\phi}} \exp (-\kappa) \right\},
\]

where $\phi$ denotes the coefficient of relative risk-aversion, which in the Epstein-Zin utility function is different from the inverse of the intertemporal elasticity of substitution. The latter remains equal to one (log utility). The case considered in the main text of the paper features unit risk-aversion, or $\phi = 1$.

The conditional value function is:

\[
V_t(a, z', k) = \begin{cases} 
(1 - \beta) u_t^* (z') & \text{if } a = \bar{a} \\
(1 - \beta) u_t^* (z') + \beta \ln E_{\varepsilon'} \left[ P_{t+1} (a + 1, z', \varepsilon', k)^{1-\phi} \right]^{1/(1-\phi)} & \text{if } a < \bar{a}
\end{cases}.
\]

Guess that the solution takes the form:

\[
V_t(a, z', k) = v_t^0 (a) + v_t^1 (a) z'.
\]
Replace the guess in (A.54) and simplify so that (A.55) becomes:

\[
V_t(a, z'; k) = (1 - \beta) \bar{u}_t^1 + (1 - \beta) (1 - \tau_t) z' + \beta v_{t+1}^0 (a + 1) + \beta v_{t+1}^1 (a + 1) \rho z' + \\
+ \beta \ln E_{\epsilon'} \left[ \max \left\{ \exp (v_{t+1}^1 (a + 1) \eta \varepsilon'_h), E_{\varepsilon_k} \left[ \exp (1 - \phi) v_{t+1}^1 (a + 1) \eta \varepsilon'_k \right] \frac{1}{1 - \phi} \exp (-\kappa) \right\} \right].
\]

Collect and impose consistency with the guess so that:

\[
v_t^1 (a) = (1 - \beta) (1 - \tau_t) + \beta \rho v_{t+1}^1 (a + 1).
\]

This gives rise to the same solution for \( v_t^1 (a) \) as in the benchmark model, equation (11). The migration choice problem is then

\[
\max \left\{ \exp (v_t^1 (a) \eta \varepsilon'_h), E_{\varepsilon_k} \left[ \exp (1 - \phi) v_t^1 (a) \eta \varepsilon'_k \right] \frac{1}{1 - \phi} \exp (-\kappa) \right\},
\]

so the agent does not migrate if

\[
\varepsilon_h \geq \ln E_{\varepsilon_k} \left[ \exp (1 - \phi) v_t^1 (a) \eta \varepsilon'_k \right] \frac{1}{1 - \phi} (1 - \phi) v_t^1 (a) \eta - \frac{\kappa}{v_t^1 (a) \eta}.
\] (A.56)

If \( \varepsilon_h \) and \( \varepsilon_k \) are distributed as extreme-value of type 1 with location parameter 0 and scale parameter 1, then

\[
(1 - \phi) v_t^1 (a) \eta \varepsilon'_k
\]

is also extreme-value with location parameter 0 and scale parameter \( (1 - \phi) v_t^1 (a) \eta \). Moreover, the random variable \( \exp ((1 - \phi) v_t^1 (a) \eta \varepsilon'_k) \) is Frechet with shape parameter \( ((1 - \phi) v_t^1 (a) \eta)^{-1} \).

Thus, we can compute

\[
E_{\varepsilon_k} \left[ \exp (1 - \phi) v_t^1 (a) \eta \varepsilon'_k \right] = \Gamma \left( 1 - (1 - \phi) v_t^1 (a) \eta \right),
\]

where \( \Gamma(x) \) is the gamma function. Replacing this expression into (A.56) and simplifying,
we obtain that the agent does not migrate if
\[ \varepsilon_h > \frac{1}{(1 - \phi) v^1_t(a) \eta} \ln \Gamma (1 - (1 - \phi) v^1_t(a) \eta) - \frac{\kappa}{v^1_t(a) \eta}. \]

The probability of not-migrating is
\[ p_t(a) = 1 - \exp \left( -\exp \left( -\frac{1}{(1 - \phi) v^1_t(a) \eta} \ln \Gamma (1 - (1 - \phi) v^1_t(a) \eta) + \frac{\kappa}{v^1_t(a) \eta} \right) \right). \]

Notice that when \( \phi = 1 \) (unit risk aversion), \( p_t(a) \) is the same as in the benchmark model, equation (14) because if \( \phi = 0 \)
\[ \lim_{\phi \to 1} \frac{-1}{(1 - \phi) v^1_t(a) \eta} \ln \Gamma (1 - (1 - \phi) v^1_t(a) \eta_r) = \frac{\Gamma'(1)}{\Gamma(1)} = -\gamma. \]

The question of interest is the effect of \( \tau_t \) on the propensity to migrate. It can be shown that if \( \phi \neq 1 \), in addition to the benchmark negative effect of \( \tau_t \) on migration, there is another effect that is either positive or negative, depending on the magnitude of \( \phi \). Specifically, with risk-aversion smaller than one \( (\phi < 1) \), a higher \( \tau_t \) reduces mobility, while with risk aversion larger than one \( (\phi > 1) \), it increases mobility.

### E.4 A Migration Subsidy and Welfare Maximization

In order to consider the implications of a migration subsidy, assume that an agent’s consumption is \( c = \lambda y^{1-\tau} \) if the agent stays put and \( c = \lambda \exp(x) y^{1-\tau} \) if she migrates, with \( x \) representing the proportional difference between the consumption of movers and stayers. Notice that in principle \( x \) may be positive or negative. The social welfare function of the version of the model with a migration subsidy takes exactly the same form as in equation (24), with the probability of staying in the same locality being:
\[ p(\tau, x) = \frac{1}{1 + \exp (- (\kappa - x) / ((1 - \tau) \eta))}. \]
Notice that a subsidy \((x > 0)\) plays the same role as a lower moving cost in inducing geographic mobility. Proposition A.2 shows that the parameters for which a migration subsidy is welfare-improving are exactly the same as those identified in Proposition 6.

**Proposition A.2** Under the conditions of Section 6.1, in the economy with a migration subsidy the welfare function \(W(\varphi, \tau, x)\) takes the same form as in equation (24) with \(p(\tau, x)\) given by equation (A.57). Moreover, if \(W(\varphi, \tau, x)\) is concave in \((\tau, x)\), the optimal degree of tax progressivity is the same as in the economy with \(K = 1\). The optimal migration subsidy satisfies:

\[
H(p, \kappa) = (1 + \chi)(1 - \eta)H\left(p, \frac{\kappa - x}{1 - \tau}\right)
\]

where the function \(H\) is defined in equation (A.23).

Proof: I show that the welfare function \(W\) in the economy with a subsidy is the same as that in the economy without, except for the fact that \(p\) depends on \(x\) as well as on \(\tau\) as in equation (A.57). First, it is straightforward to show that the solution of the individual decision problem with a subsidy is the same as in equation (A.14) with \(p\) defined as in equation (A.57). Consider now the budget constraint and the expression for \(\ln \lambda\). The balanced-budget equation is now

\[
(1 - \varphi) Y = \int_{z': \text{stay}} \lambda [\ell^* \exp (z')]^{1 - \tau} f^p (z') \, dz' + \int_{z': \text{move}} \lambda \exp (x) [\ell^* \exp (z')]^{1 - \tau} f^p (z') \, dz'.
\]

Collecting terms yields

\[
(1 - \varphi) Y = \lambda (\ell^*)^{1 - \tau} \left\{ \int_{z': \text{stay}} \exp ((1 - \tau) z') f^p (z') \, dz' + \exp (x) \int_{z': \text{move}} \exp ((1 - \tau) z') f^p (z') \, dz' \right\}.
\]

Notice that

\[
\int_{z': \text{stay}} \exp ((1 - \tau) z') f^p (z') \, dz' = p \int_{z'} \exp ((1 - \tau) z') f^p (z' | \text{stay}) \, dz',
\]

\[
\int_{z': \text{move}} \exp ((1 - \tau) z') f^p (z') \, dz' = (1 - p) \int_{z'} \exp ((1 - \tau) z') f^p (z' | \text{move}) \, dz',
\]
where \( f^p(z'|\text{stay}) \) and \( f^p(z'|\text{move}) \) are the densities of the shocks \( z' \), conditional on staying or moving to the other location. From Proposition 4, I know the distribution of \( \varepsilon' = z'/\eta \), conditional on staying or moving. Since, conditional on staying, \( \varepsilon' \) is \( \text{EV}(\eta \ln (1/p), \eta) \), and \( (1 - \tau)z' \) is \( \text{EV}(\eta (1 - \tau) \ln (1/p), \eta (1 - \tau)) \). It follows from Lemma A1 that

\[
\int_{z'} \exp ((1 - \tau) z') f^p(z'|\text{stay}) \, dz' = \exp (\eta (1 - \tau) \ln (1/p) + \ln \Gamma (1 - \eta (1 - \tau))),
\]

\[
\int_{z'} \exp ((1 - \tau) z') f^p(z'|\text{move}) \, dz' = \exp (\eta (1 - \tau) \ln (1/p) + \ln \Gamma (1 - \eta (1 - \tau)) + \kappa - x).
\]

It follows that

\[
(1 - \varphi) Y = \lambda \ell^* (1 - \tau) \exp (-\eta (1 - \tau) \ln p + \ln \Gamma (1 - \eta (1 - \tau))) \left( p + (1 - p) \exp (\kappa) \right).
\]

Taking logs and solving for \( \ln \lambda \) yields:

\[
\ln \lambda = \ln (1 - \varphi) + \ln Y - (1 - \tau) \ln \ell^* + \eta (1 - \tau) \ln p - \ln \Gamma (1 - \eta (1 - \tau)) - \ln (p + (1 - p) \exp (\kappa)).
\]

Replace \( \ln \lambda \) into the definition of \( \pi^* \):

\[
\pi^* \equiv \ln \lambda + (1 - \tau) \ln \ell^* - \zeta^{-1} (\ell^*)^c + \chi \ln [\varphi Y]
\]

to obtain

\[
\pi^* \equiv \ln (1 - \varphi) + \chi \ln \varphi + (1 + \chi) \ln Y - \zeta^{-1} (\ell^*)^c +
\]
\[+ \eta (1 - \tau) \ln p - \ln \Gamma (1 - \eta (1 - \tau)) - \ln (p + (1 - p) \exp (\kappa)).
\]

Replace

\[
Y = \ell^* \int \exp (z') f^p(z') \, dz'
\]
where, using steps similar to before:

\[
\int \exp (z') f^p (z') \, dz' = p \int_{z'} \exp (z') f^p (z'|\text{stay}) \, dz' + (1 - p) \int_{z'} \exp (z') f^p (z'|\text{move}) \, dz' \\
= p \exp (-\eta \ln p + \ln \Gamma (1 - \eta)) + (1 - p) \exp \left(-\eta \ln p + \ln \Gamma (1 - \eta) + \frac{x}{1 - \tau}\right) \\
= \exp (-\eta \ln p + \ln \Gamma (1 - \eta)) \left( p + (1 - p) \exp \left(\frac{x}{1 - \tau}\right) \right)
\]

It follows that

\[
\pi^* \equiv \ln (1 - \varphi) + \chi \ln \varphi + (1 + \chi) (-\eta \ln p + \ln \Gamma (1 - \eta)) + \\
+ (1 + \chi) \ln \left[ p + (1 - p) \exp \left(\frac{x}{1 - \tau}\right) \right] + \\
+ (1 + \chi) \ln \ell^* - \zeta^{-1} (\ell^*)^\zeta + \eta (1 - \tau) \ln p - \ln \Gamma (1 - \eta (1 - \tau)) - \ln (p + (1 - p) \exp (\kappa)).
\]

Recall that the welfare function is as before:

\[
W = (1 - \beta) \int P (\varepsilon) g (\varepsilon) \, d\varepsilon \\
= \bar{u}^* + (1 - \tau) \eta \gamma - (1 - \tau) \eta \ln p.
\]

Replace \(\pi^*\):

\[
W = (1 - \gamma) \eta \gamma - (1 - \gamma) \eta \ln p + \\
\ln (1 - \varphi) + \chi \ln \varphi + (1 + \chi) (-\eta \ln p + \ln \Gamma (1 - \eta)) + \\
+ (1 + \chi) \ln \left[ p + (1 - p) \exp \left(\frac{x}{1 - \tau}\right) \right] + (1 + \chi) \ln \ell^* - \zeta^{-1} (\ell^*)^\zeta + \\
+ \eta (1 - \tau) \ln p - \ln \Gamma (1 - \eta (1 - \tau)) - \ln (p + (1 - p) \exp (\kappa)).
\]
Simplify it, ignoring terms that do not depend on policy variables:

\[
W = \ln (1 - \varphi) + \chi \ln \varphi + (1 + \chi) \ln \ell^* - \zeta^{-1} (\ell^*)^\zeta + (1 + \chi) \left( \ln \left[ p^{1-\eta} + p^{-\eta} (1 - p) \exp \left( \frac{\kappa - x}{1 - \tau} \right) \right] \right) - (\ln \Gamma (1 - \eta (1 - \tau)) - (1 - \tau) \eta \gamma + \ln (p + (1 - p) \exp (\kappa))).
\]

Notice that from equation (A.57) we obtain

\[
\exp \left( \frac{(\kappa - x)}{(1 - \tau)} \right) = \left( \frac{p}{1 - p} \right)^\eta.
\]

Replacing it into the welfare function and using the definition of \( W^{ra} (\varphi, \tau) \) from equation (25) yields:

\[
W (\varphi, \tau, x) = W^{ra} (\varphi, \tau) + (1 + \chi) \ln \left( p^{1-\eta} + (1 - p)^{1-\eta} \right) - \kappa (1 - p) - (\ln \Gamma (1 - (1 - \tau) \eta) - \eta \gamma (1 - \tau) + \ln (p + (1 - p) \exp (\kappa))) - \kappa (1 - p),
\]

which is the same as equation (24), except that \( p \) is now a function of both \( \tau \) and \( x \). If the function \( W \) is concave in \((\tau, x)\), the first-order conditions are sufficient for an optimum (the optimal \( \varphi \) being as in Proposition 7). Take the first-order condition with respect to \( \tau \):

\[
\frac{\partial W}{\partial \tau} = \frac{\partial W^{ra} (\varphi, \tau)}{\partial \tau} - \eta \gamma \left( \frac{\Gamma (1 - (1 - \tau) \eta)}{\Gamma (1 - (1 - \tau))} \right) - \eta \gamma + \left\{ H (p, \kappa) - (1 + \chi) (1 - \eta) H \left( p, \frac{\kappa - x}{1 - \tau} \right) \right\} \frac{\partial p}{\partial \tau} = 0, \tag{A.58}
\]

where \( H \) is defined in (A.23). Now, take the derivative of \( W \) with respect to \( x \):

\[
\frac{\partial W}{\partial x} = H (p, \kappa) - (1 + \chi) (1 - \eta) H \left( p, \frac{\kappa - x}{1 - \tau} \right) = 0.
\]
Replacing the latter into equation (A.58) simplifies it to:

\[
\frac{\partial W}{\partial \tau} = \frac{\partial W^{ra}(\varphi, \tau)}{\partial \tau} - \eta \frac{\Gamma'(1 - (1 - \tau) \eta)}{\Gamma(1 - (1 - \tau) \eta)} - \eta \gamma = 0.
\]

This corresponds to the first-order condition with respect to \( \tau \) in a version of the model with one location \((K = 1)\).

Q.E.D.

In summary, in this situation the planner sets \( x \) optimally in order equalize the social marginal benefits and costs of migration, while it sets \( \tau \) in order to trade-off the benefits and costs of redistribution in the one-location version of the model.