Geographic Mobility and Redistribution

Daniele Coen-Pirani*

February 14, 2018

Abstract

I study the effect of progressive taxation on internal migration and welfare using a new dynamic Roy model. The model features an arbitrary number of labor markets and finitely-lived agents, yet is analytically tractable. It predicts that a more progressive tax-transfer scheme reduces internal migration rates. The magnitude of this relationship is consistent with evidence from the OECD countries. The optimal time-varying sequence of tax progressivity features relatively high degree of tax progressivity early on, and lower tax progressivity at later dates. There are substantial welfare gains from letting the optimal degree of tax progressivity vary over the transition.

Keywords: Geographic Mobility, Progressive Taxation, Redistribution, Welfare.

---

*University of Pittsburgh. E-mail: coen@pitt.edu. Thanks to seminar participants at New York University, Pittsburgh, Carnegie Mellon, Bologna, Haifa, Torino, IDC Hertzliya, European University Institute, and the 2017 SED meetings in Edinburgh for useful comments. Special thanks to Chris Flinn, Laurence Ales, Ali Shourideh, Giuseppe Bertola, Marla Ripoll, Omer Moav, Yona Rubinstein, Giulio Zanella, Stefania Albanesi, Fabian Lange, Sewon Hur, Douglas Hanley, Max Dvorkin, Arpad Abraham, Rodi Manuelli, Axelle Ferriere, Jean-Francois Richard, and Marco Casari for useful comments. The usual disclaimer applies.
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, if labor supply is elastic, high or more progressive taxes reduce the incentives to participate in the workforce or work longer hours. In this paper I focus on the effect of taxes, and in particular tax progressivity, on the location of an individual’s labor supply, instead of the quantity of labor supplied. The basic idea is that a more progressive tax system reduces an individual’s incentive to migrate internally within a country to locations in which she is idiosyncratically more productive. While it has been recognized since Sjaadstad (1962)’s early contribution that migration is a form of investment in human capital whose returns are affected by tax policy, there are very few papers, either empirical or theoretical, that seek to quantify the effect of progressive taxation and redistribution on internal migration.\footnote{Hassler et al (2005) had earlier shown that countries with more generous unemployment insurance systems are characterized by lower geographic mobility. They construct a general equilibrium politico-economic model of internal migration with endogenous unemployment insurance. However, their focus is not quantitative. Shaw (1986) studies empirically the effect of unemployment insurance and other fiscal policies on Canadians’ incentives to migrate internally from low to high income regions.}

In order to avoid confusions, it is worth clarifying at the onset the argument advanced here. The tax distortion I consider is due to the national – in the U.S. Federal – tax system instead of tax differences across political sub-units. In other words, I do not study individuals’ incentives to migrate across jurisdictions in order escape from 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.} In addition, 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. That is, the mechanism I study is different from Albouy (2009)’s point that since the Federal government taxes nominal income, it might distort the spatial allocation of workers across locations with heterogeneous productivity and amenities.\footnote{Similarly, to Albouy (2009), more recent papers by Eeckout and Guner (2015) and Colas and Hutchin-}
face dynamic mobility choices and progressive Federal taxation reduces the net gain from moving even if all locations are ex-ante identical from an aggregate point of view. 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. The focus on gross, rather than net, flows of workers is motivated by the fact that in the data gross migration flows are 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.

I begin by showing that OECD countries characterized by a higher degree of tax progressivity also tend to have lower rates of internal migration. I then introduce a new dynamic Roy model of internal migration. The model is rich, as it allows for both ex-ante and ex-post heterogeneity in terms of age, moving costs, and labor income, while being analytically tractable. I use the model’s closed-form solution to show that a higher degree of income tax progressivity reduces migration rates as long as migration costs are not zero or infinitely large. The intuition for this result is that 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 estimated using the Generalized Method of Moments to account for the frequency of interstate migration by age, the life-cycle profile of earnings, and the growth in earnings inequality as households age.

The quantitative version of the model is used to produce counterfactual estimates of the effect of tax progressivity on internal migration rates and forms the basis of welfare

son (2017) also analyze the role of Federal taxation in affecting the geographic allocation of labor across heterogeneous (from an aggregate perspective) labor markets.

In other words, excess reallocation – defined as the gross flows in excess of those needed to accommodate a location’s net flows – accounts for about 90 percent of overall gross flows of households across states. See also Coen-Pirani (2010, Table 1) for analogous calculations for the period 1970–2000.
analysis. The negative relationship between migration and tax progressivity predicted by the benchmark model is quantitatively consistent with the cross-country evidence mentioned above. From a normative perspective, tax progressivity represents a distortion to human capital investment. This induces a utilitarian social planner to select a less progressive tax system than under exogenous geographic mobility. This migration effect is shown to have a quantitatively large impact on the optimal tax system. Additionally, the tractability of the model environment allows me to compute the optimal time-varying sequence of tax progressivity, starting from the initial steady state. The planner chooses a relatively high degree of tax progressivity in the early phase of the reform, followed by a declining path towards relatively low levels of tax progressivity in the final steady state. This time-varying policy generates much larger welfare gains than a policy reform that entails a one-shot switch to a constant time independent policy.

This paper adds to the recent literature (Benabou (2002), Bovenberg and Jacobs (2005), Erosa and Koreshkova (2007), Krueger and Ludwig (2013, 1016), Guvenen et al (2014), Badel and Hugget (2015), Heathcote et al (2017), Stantcheva (2017)) on the effects of subsidies and taxes on human capital investment. These papers focus on investment in either formal schooling or training. Differently from these papers, I focus on internal migration, a form of investment in human capital that can be readily observed in publicly available data. 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. In my model it is tax progressivity as opposed to the level of taxation per se, that discourages migration. This is consistent with the empirical evidence across OECD countries discussed below.

5 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. Overall, this and other job-related reasons account for about 42 percent of all interstate moves. The prevalence of such motivation is suggestive of the idea that job-related reasons are a key driver of internal mobility of labor.

6 This point is related to, but different from, Boskin (1977), Heckman (1976), and Guvenen et al (2014)’s earlier demonstration that proportional taxes are neutral in the basic Ben Porath model, while progressive taxes are not.
The paper is also related to the recent papers by Caliendo et al (2015) and Artuc et al (2010) on spatial and sectoral labor reallocation following trade shocks. Ales and Sleet (2017) study the insurance benefits of location-specific taxes and transfers when individuals differ in terms of their attachment to various locations. 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 their analytical convenience, it interprets them as permanent shocks to households’ productivity 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, Lkhagvasuren, 2014, Kaplan and Schulhofer-Wohl, 2017), and development (Desmet and Rossi-Hansberg, 2017 and Lagakos et al (2017)).}

The rest of the paper is organized as follows. Section 2 presents the main theoretical argument using a static version of the model. It also introduces the empirical evidence on the relationship between tax progressivity and internal migration. Section 3 introduces the model and Section 4 characterizes its analytical solution. The description of the model’s estimation is contained in Section 5. Welfare analysis is discussed in Section 6. Finally, Section 7 concludes and discusses future work. All proofs and additional analytical details are contained in the Appendix at the end of the paper.

2 Mechanism and Empirical Evidence

2.1 Main Theoretical Argument

In this section I use a simple example to briefly explain the main mechanism of the paper linking tax progressivity and internal migration. A household faces a choice between staying in a location and earning income $y_s$ or moving to another location and earning income $y_m$. Relocation entails a utility cost $\kappa > 0$. The household derives utility from consuming its
after tax-transfer income \( \tilde{y} \), defined as:

\[
\tilde{y} = \lambda y^{1-\tau}.
\] (1)

Thus, the household pays net taxes \( T(y) = y - \tilde{y} \). This tax-transfer system has recently been used by Benabou (2002) and Heathcote et al (2017), among others. It is characterized by two parameters, \( \tau \) and \( \lambda \). The former indexes the degree of tax progressivity. To gain intuition, notice that \( \tau \) is related to the marginal and average tax rate by the following equation:

\[
\tau = 1 - \frac{1 - T'(y)}{1 - T(y)/y}.
\]

This expression shows that the larger the gap between marginal and average tax rates, the higher \( \tau \) is. For example, when the marginal tax rate equals the average tax rate, \( T'(y) = T(y)/y \), the tax system features a proportional income tax, so \( \tau = 0 \). By contrast, when the government taxes away the marginal unit of income earned, \( T'(y) = 1 \), we have \( \tau = 1 \). Therefore, in what follows, I refer to \( \tau \) as the degree of progressivity of the tax-transfer scheme in equation (1), with higher values of this parameter denoting a more progressive tax system.

The parameter \( \lambda \), instead, indexes the overall level of taxation. For analytical convenience, the utility function is assumed to be logarithmic, so the household chooses to migrate if:

\[
\ln \tilde{y}_m - \kappa > \ln \tilde{y}_s.
\] (2)

Replacing the definition of after-tax income (1) in equation (2) and rearranging it, the agent chooses to move if:

\[
(1 - \tau) \epsilon > \kappa,
\] (3)

where \( \epsilon \) denotes the proportional, before-tax, income gain from moving:

\[
\epsilon \equiv \ln y_m - \ln y_s.
\] (4)
The inequality in equation (3) makes it clear that higher tax progressivity reduces the after-tax proportional income gain from moving. A similar result applies if the agent only knows the distribution of $\epsilon$, instead of its realization, when making the migration choice. In this case equation (3) applies with the expected value of $\epsilon$ replacing $\epsilon$.

Denote by $Q(\epsilon)$ the cumulative distribution function of $\epsilon$ in the population. The economy’s migration rate is then:

$$1 - Q\left(\frac{\kappa}{1 - \tau}\right).$$

(5)

Thus, a larger $\tau$ reduces the aggregate migration rate. Moreover, the magnitude of the effect involved depends on the moving cost $\kappa$. If moving was costless ($\kappa = 0$), then the migration rate would be independent of the degree of tax progressivity $\tau$ because the household would find it optimal to always move to the location with the highest proportional gain from moving. On the other hand, if the moving cost was prohibitively large ($\kappa \to \infty$), then the migration rate would be zero and would not depend on the degree of tax progressivity either.

In both of these extreme cases, a utilitarian planner would find it optimal to redistribute all income by setting $\tau = 1$ as there would not be any distortions associated with migration. For intermediate levels of $\kappa$, instead, the planner would need to take migration distortions into account. For this reason, identification and estimation of the moving cost parameter $\kappa$ plays an important role in the analysis that follows.

In Section 3, I extend this static model to a dynamic setting in which households supply labor, and make migration choices over multiple periods and multiple locations. Even in this generalization, the basic intuition for the effect of tax progressivity on geographic mobility remains the same as in the simple case considered here.

---

8The expression in equation (3) also obtains with a more general utility function $u(\cdot)$ as long as the moving cost is a fraction $f$ of consumption. In this case the relevant comparison is

$$u((1 - f)\bar{y}_{\text{move}}) > u(\bar{y}_{\text{home}}),$$

which reduces to equation (3) with $\kappa \equiv -\ln(1 - f)$. 

7
2.2 Empirical Evidence

In this section I provide reduced-form cross-country evidence on the relationship between internal migration rates and the degree of progressivity of taxes and transfers. To measure the tax progressivity parameter $\tau$ in equation (1) I use OECD data on the Gini coefficient of household income before and after taxes and transfers. Assume that household market income $y$ is distributed lognormally in the population with variance parameter $v^2_y$. It follows that after-tax and transfers income $\tilde{y} = \lambda y^{1-\tau}$ is also lognormal with variance parameter $(1 - \tau)^2 v^2_y$.

For a lognormal distribution with variance parameter $v^2_y$, the Gini coefficient is given by:

$$\text{GINI}(y) = \text{erf}(0.5v_y), \quad (6)$$

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:

$$\text{GINI}(\tilde{y}) = \text{erf}(0.5(1 - \tau)v_y). \quad (7)$$

The OECD provides country-level data on $\text{GINI}(y)$ and $\text{GINI}(\tilde{y})$. The strategy I follow is therefore to invert equation (6) to obtain $v_y$ and use the latter together with the inverse of equation (7) to obtain a country-level estimate of $\tau$. Putting together these relationships we obtain the following estimator of $\tau$:

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

with $\Phi^{-1}(.)$ denoting the inverse of the cumulative distribution function. Intuitively, this

---

9 The assumption of lognormality is not necessary here although it facilitates some computations.
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 $\tau = 1$.

In order to implement this procedure I use data from the OECD’s Income Distribution Database. 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 data on $y$ represents market income, before taxes and transfers, while the data on $\bar{y}$ represents disposable income, i.e. market income received by a household less taxes paid plus transfers received. Since measures of tax progressivity may be subject to measurement error, I average them for a given country over time for the period 1980–2013, i.e. the same period on which internal migration rates are measured.\(^{10}\)

The data on internal migration also comes from the OECD.\(^{11}\) It refers to the post-1980 period all the way to 2010. The territorial unit in each country is what the OECD defines as Territorial Level 2 (TL2). This regional level corresponds to the first administrative tier of sub-national governments. The concept of TL2 corresponds to the state level for the U.S. (51 units), to the Lander level for Germany (16 Landers), to Comunidades Autonomas for Spain (19 Comunidades), etc.\(^{12}\) I focus on the 21 countries with more than 5 TL2s because 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. As for tax progressivity, I average the yearly migration data by country. Table 1 summarizes the information on migration rates and $\tau$ at the country level.

Figure 1 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.\(^{13}\) This

\(^{10}\)The estimated values of $\tau$ display a cross-country correlation of 0.69 with the estimates by Holter, Krueger, and Stepanchuk (2015, Table 1). Differently from my approach, their estimates are obtained using tax, but not transfers, data and only refer to the period 2000-07.

\(^{11}\)The 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.

\(^{12}\)A full list is provided in Appendix J. The correlation between the number of TL2 regions and the log of average population of the country is 0.74 (s.e. 0.00).

\(^{13}\)This estimate is robust to controlling for log population in the regression. In this case the slope is $-4.88$.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St. Dev</th>
<th>Min</th>
<th>Max</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration rate</td>
<td>1.65</td>
<td>.77</td>
<td>.42</td>
<td>3.17</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Poland)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Turkey)</td>
</tr>
<tr>
<td>Tax progressivity $\tau$</td>
<td>.27</td>
<td>.07</td>
<td>.13</td>
<td>.37</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Switzerland)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Austria)</td>
</tr>
</tbody>
</table>

Table 1: Summary statistics on country-level average (1980–2010) migration rates and the tax progressivity measure $\tau$.

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 0.36 percentage points, or slightly less than a half of the cross-sectional standard deviation of migration rates in Table 1.

Interestingly, the partial correlation between migration and tax progressivity remains roughly similar even controlling for the level of taxes. For example, including 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 sign in this regression. This evidence is consistent with the hypothesis advanced here that it is the degree of tax progressivity, rather, than the level of taxation per se that matters for migration.

In summary, there is suggestive evidence of a negative cross-sectional correlation between internal migration and tax progressivity across countries. What about 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 and p-value 0.29). While this is not (s.e. 2.06).
Figure 1: Scatterplot of average country-level migration rates against average country-level \( \tau \). The regression line has slope \(-5.15 \) (s.e. \( 2.06 \)) with an \( R^2 = 0.21 \).
statistically different from zero, the magnitude of the point estimate is comparable with the cross-sectional one.

Focusing on the U.S., it is interesting to notice that the measured $\tau$ is remarkably stable over time, in the range 0.19–0.23 for the entire period 1974 (when it was equal to 0.19) to 2012 when it was equal to 0.21. The data pattern shows measured $\tau$ increasing from 1974 to 2000 - when it reached 0.23 - and then falling slightly afterwards. In Section 5.1.2 when estimating the model’s parameters, I use Congressional Budget Office data to measure $\tau$ using a different procedure relative to the one described in this section. Using this alternative approach, I obtain a stable estimate of $\tau$ of about 0.19, without evidence of any time trend in this parameter for the period 1979–2007.

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 1. Countries may vary along other, potentially unobserved, dimensions in addition to the degree of progressivity of their tax system. For example, if a country’s internal migration rate is low for reasons other than taxes, its government might choose a more progressive tax system because it does not need to worry about its distortions to geographic mobility choices. In order to address this and related interpretation issues, in the next section I introduce a structural model of migration. The estimated model will then be used to study the causal effect of tax progressivity on migration.

3 Model

In this section I introduce the model economy that forms the basis of quantitative analysis.

Geography, Technology, and Time The economy is comprised of a finite number of local labor markets indexed by $j = 1, 2, ..., J$. Local economies are assumed to be 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 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 policy variables $\{\tau_t, \lambda_t, G_t\}_{t=1}^{\infty}$, where $\tau_t$ and $\lambda_t$ are the parameters of the tax-transfer scheme introduced in equation (1) and $G_t$ denotes the national government’s public good provision.\(^\text{14}\)

**Demographic Structure and Timing** The economy is populated by a continuum of finitely-lived households of measure one. Households live from age $a = 1$ to age $\bar{a}$. A measure $1/\bar{a}$ of households is born at age $a = 1$ in each location at the beginning of each period. A household belongs to one of $r$ possible types, indexed by $r = 1, 2, ..., r$. Household type is determined at birth and does not change over time. All the model’s parameters are indexed by $r$. There is a measure $\omega_r$ of households of type $r$, with $\sum_r \omega_r = 1$.

At each age $a$ the timing of actions is as follows. At the beginning of each period, the household resides in one of the economy’s $J$ locations. Let $h$ denote the household’s location of residence at the beginning of a period. In this location a household supplies labor $\ell$ and earns labor income $y = \ell \exp \left( z \right)$, where $\exp \left( z \right)$ denotes the household’s labor productivity at the beginning of the period. Then, redistribution takes place leaving the household with disposable income $\tilde{y}$, which the household consumes, so $c = \lambda_t \tilde{y}^{1-\tau_t}$. The assumption that an agent cannot save or borrow reflects an extreme form of asset market incompleteness.\(^\text{15}\)

After consumption, a household of age $a < \bar{a}$ faces the potential of migrating.\(^\text{16}\) Specifically, I assume that a household might move for either exogenous or endogenous reasons. In what follows I describe both, starting from the endogenous migration choice.

---

\(^{14}\)The presence of a public good does not affect the positive properties of the model because it enters additively in utility (see equation (10) below). However, its presence affects the socially optimal choice of tax progressivity $\{\tau_t\}_{t=1}^{\infty}$, as shown by Heathcote et al (2017).

\(^{15}\)Differently from Heathcote et al (2017), this assumption is not without loss of generality due to the explicit consideration of age as an important determinant of household choices. In my setting, on average, a young agent would like to save in anticipation of aging and a diminished scope for migration later in life. By contrast, Heathcote et al (2017) adopt Blanchard (1985)’s perpetual youth model, so all agents face the same time horizon going forward and this aging effect is not present. It can be shown that in the perpetual youth version of my model, agents would want to consume all of their income at the equilibrium interest rate.

\(^{16}\)In the last period of life the household simply consumes its earnings.
Evolution of Household’s Productivity and Migration Choice  The core of the model is the description of the household labor income process and the associated endogenous migration decisions. A household starts life at age $a = 1$ with initial productivity $\exp(z)$, 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 $r$. At the end of each period, before migration decisions, the household observes the potential evolution of its productivity in each location. Specifically, a household of age $a$, type $r$, with current productivity $\exp(z)$ observes, at the end of period $t$, its efficiency units in period $t+1$ in all locations $k$. Potential log productivity in location $k$ in $t+1$ is assumed to take the following form:

$$z_{ka+1} = \alpha_{a,r} + z_a + \eta_r \varepsilon_k \text{ for all } k = 1, \ldots, J,$$

(8)

where $\alpha_{a,r}$ determines the exogenous growth of household income as a function of age for a type $r$ household. The parameter $\eta_r$ governs the importance of location-specific shocks to efficiency $\{\varepsilon_k\}_{k=1}^J$. I assume that location-specific shocks are independently distributed both over time and in the cross-section.\(^{17}\) The shocks $\{\varepsilon_k\}_{k=1}^J$ are assumed to be distributed according to a type-1 extreme value distribution. I allow shocks drawn from the household’s current home location to have a different distribution from shocks drawn elsewhere due, for example, to its superior information about the home local labor market. The cumulative distribution function of $\varepsilon_k$, $Q_r(\varepsilon_k|h)$, may then depend on whether or not $k$ is the household’s current residence:

$$Q_r(\varepsilon_k|h) = \begin{cases} \exp(-\delta_r \exp(-\varepsilon_k)) & \text{if } k = h \\ \exp(-\exp(-\varepsilon_k)) & \text{if } k \neq h \end{cases},$$

(9)

\(^{17}\)It is straightforward to allow for some correlation among shocks, but their common component would not matter for migration decisions. For example, the $\{\varepsilon_k\}_{k=1}^J$ might be written as the sum of a common component and an idiosyncratic one:

$$\varepsilon_k = u + \varepsilon_k$$

where $u$ is a household-dependent random variable with mean zero and variance $\nu_u^2$, while the $\varepsilon_k$’s are independently distributed from one another, over time, and are orthogonal to $u$. Notice that the covariance between any two $\varepsilon_k$ and $\varepsilon_j$ is equal to $\nu_u^2$. It is straightforward to show that in this case $u$ would not matter for moving choices.
where $\delta_r \geq 1$ is a parameter. A larger $\delta_r$ implies that, on average, shocks drawn from the home locations are larger relative to shocks drawn from each of the other locations. I distinguish between endogenous and exogenous relocation in the following simple way. With probability $\theta_r$, upon observing the entire vector $\{\varepsilon_k\}_{k=1}^J$, a household of type $r$ is able to select the location that provides the highest present discounted value of utility going forward. With probability $1 - \theta_r$, instead, a household may be exogenously relocated to one of the other $J - 1$ locations. Let the random variable $x_{hk} = 1$ if a household who is exogenously relocated is sent to $k \neq h$ and zero otherwise. The household faces an equal probability of being sent to each of the other $J - 1$ locations and its productivity evolves accordingly to the $\varepsilon_k$ shock it draws there. Since an exogenously relocated household must reside somewhere it must be the case that $\sum_{k \neq h} x_{hk} = 1$ for all $h$. 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 $\exp (z_{ka+1})$.

**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 static utility function of an agent is

$$u (c, \ell, G) = \ln c - \zeta^{-1} \ell^\zeta + \chi \ln G,$$

(10)

where $\zeta$ is a parameter that 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. Given this utility function and the budget constraint $c = \lambda_t (\ell \exp z)^{1 - \tau_t}$, the

---

18 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 $\gamma$.

19 The logarithmic assumption may be relaxed and a more general CRRA family of static utilities considered (while keeping analytical tractability) if migration costs are assumed to be proportional to current period’s consumption.
household’s labor supply is given by:
\[ \ell_t^* = (1 - \tau_t)^{\frac{1}{\zeta}}. \]  

(11)

Thus, a higher degree of tax progressivity reduces labor supply, as in Benabou (2002) and Heathcote et al (2017). Replacing \( \ell_t^* \) back into (10), the static indirect utility function then takes the log-linear form:
\[ u_t^*(z) = \overline{u}_t^* + (1 - \tau_t) z \]

where:
\[ \overline{u}_t^* \equiv \ln \lambda_t + (1 - \tau_t) \ln \ell_t^* - \zeta^{-1}(\ell_t^*)^\zeta + \chi \ln G_t. \]

Each household faces a dynamic optimization problem that involves the choice of location. Define \( V_t(a, z, h; r) \) as the maximum remaining lifetime utility attainable by a household of type \( r \), age \( a \), log productivity \( z \), initially located in \( h \) at the beginning of period \( t \). For a household in the last period of life (\( a = \overline{a} \)) the value function is simply:
\[ V_t(\overline{a}, z, h; r) = u_t^*(z) \text{ for all } h, \]

while for \( a < \overline{a} \) it is defined recursively as:
\[
V_t(a, z, h; r) = u_t^*(z) + \theta_r E_{\varepsilon} \left[ \max_k \{ \beta V_{t+1}(a + 1, \alpha_{a,r} + z + \eta_x \varepsilon_k, k; r) - I_{hk} \kappa_r \} \right] + (1 - \theta_r) E_{\varepsilon,x} \left[ \sum_{k \neq h} x_{hk} \{ \beta V_{t+1}(a + 1, \alpha_{a,r} + z + \eta_x \varepsilon_k, k; r) - \kappa_r \} \right],
\]

where \( E_{\varepsilon,x}[.] \) denotes the expectation taken with respect to the distribution of the random variables \( \varepsilon_k \) and \( x_{hk} \), and \( I_{hk} \) 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 (12) represents the expected value of endogenous migration, while the term multiplying \((1 - \theta_r)\) in the second row represents the expected value of exogenous migration. Let \( M_t(\varepsilon, a, h, k; r) \) 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, \( M_t(\varepsilon, a, h, k; r) = 1 \) if a household of age \( a \) with a vector of shocks \( \varepsilon \) moves voluntarily at the end of time \( t \) from its current location \( h \) to location \( k \), and \( M_t(\varepsilon, a, h, k; r) = 0 \) if the household does not move to \( k \). Notice that, by definition, \( \sum_{k=1}^{J} M_t(\varepsilon, a, h, k; r) = 1 \).

**Symmetric Competitive Equilibrium**  Given the sequences \( \{\tau_t, G_t\}_{t=1}^{\infty} \) and distributions \( f_1(z|a, r) \) of household productivity by age and type at \( t = 1 \), a symmetric competitive equilibrium for this economy is comprised of: a sequence \( \{\lambda_t\}_{t=1}^{\infty} \); sequences of densities of household labor productivity by age and type \( \{f_t(z|a, r)\}_{t=2}^{\infty} \); sequences of value functions \( \{V_t(a, z, h; r)\}_{t=1}^{\infty} \); sequences of decision rules \( \{M_t(\varepsilon, a, h, k; r)\}_{t=1}^{\infty} \) for geographic mobility; sequences of decision rules for labor supply \( \{\ell^*_t\}_{t=1}^{\infty} \) such that:

1) The value functions \( \{V_t(a, z, h; r)\}_{t=1}^{\infty} \) and decision rules \( \{M_t(\varepsilon, a, h, k; r)\}_{t=1}^{\infty} \) represent the solution to the agent’s dynamic optimization problem (12), with \( \ell^*_t \) taking the form in (11).

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

\[
G_t = \frac{1}{\pi} \sum_{a=1}^{\pi} \sum_{r=1}^{r} \omega_r \int \left( \ell^*_t \exp z - \lambda_t (\ell^*_t \exp z)^{1-\tau_t} \right) f_t(z|a, r) \, dx.
\]  

(13)

3) The densities \( \{f_t(z|a, r)\}_{t=2}^{\infty} \) are generated by the transition equation for productivity (8) combined with the optimal decision rules \( \{M_t(\varepsilon, a, h, k; r)\}_{t=1}^{\infty} \).

Notice that the general equilibrium dimension of this economy comes from the government’s budget constraint (13). At each point in time \( t \), two of the variables in the triple \( (\tau_t, G_t, \lambda_t) \) are set exogenously, and the remaining one has to satisfy equation (13). 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.
4 Analytical Derivation

In this section I characterize analytically the value function, the geographic mobility decision rule, and the law of motion of households’ productivity.

4.1 Value Function and Decision Rules

The following proposition summarizes the form of the value function that solves the Bellman equation (12).

Proposition 1 (Value function) \( \{\tau_t, G_t, \lambda_t\}_{t=1}^{\infty} \), the unique value function that solves the dynamic programming problem (12) takes the following form:

\[
V_t(a, z, h; r) = v_t^0(a; r) + v_t^1(a) z \quad \text{for all } h, \tag{14}
\]

where

\[
v_t^1(a) = \sum_{k=0}^{\pi-a} \beta^k (1 - \tau_{t+k}). \tag{15}
\]

The age-dependent term \( v_t^0(a; r) \) is defined recursively by:

\[
v_t^0(a; r) = \bar{\pi}_t^* + \beta v_{t+1}^0(a + 1; r) - (1 - \theta_r) \kappa_r + \beta v_{t+1}^1(a + 1) \left[ \alpha_{a,r} + \eta_r (\gamma + \theta_r \ln \delta_r) - \eta_r \theta_r \ln p_t(a; r) \right], \tag{16}
\]

starting from \( v_t^0(\bar{a}; r) = \bar{\pi}_t^* \). The term \( p_t(a; r) \) in equation (16) represents the probability of choosing to remain in the same location and is formally defined in equation (20) below.

The following proposition characterizes the geographic decision rule.

Proposition 2 (Geographic mobility) A household of age \( a \) with shocks \( \varepsilon \) chooses to stay in the same location \( h \) in period \( t \) \( (M_t(\varepsilon, a, h, h; r) = 1) \) if and only if:

\[
\frac{\kappa_r}{v_{t+1}^1(a + 1) \beta \eta_r} > \max_{i \neq h} \varepsilon_i - \varepsilon_h. \tag{17}
\]
The household chooses to move from $h$ to $k \neq h$ ($M_t(\varepsilon, a, h, k; r) = 1$) if and only if:

$$\frac{\kappa_r}{\nu_{t+1}(a+1) \beta \eta_r} < \max_{l \neq h} \varepsilon_l - \varepsilon_h$$

(18)

and

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

(19)

Notice that absent moving costs ($\kappa_r = 0$), the household would simply pick the location with the highest growth rate of labor income. A positive moving cost introduces a bias towards staying in the same location. Therefore, an agent might remain in the same location even if the growth rate of household income in that location is smaller than in the rest of the economy (equation (17)). Conditional on moving to a different location (18), a household chooses the location with the highest growth rate of its productivity (equation (19)).

Equation (17) gives the condition under which an agent of age $a$ who is given the opportunity to choose whether to migrate or stay in the same location, chooses the latter option. Therefore, the probability that this happens can be computed using the distribution of the random variable $\max_{l \neq h} \varepsilon_l - \varepsilon_h$. The assumption that the $\varepsilon_k$ shocks are distributed as type 1 extreme-value implies that $\max_{l \neq h} \varepsilon_l - \varepsilon_h$ is logistic with location parameter $\ln{(J - 1)}$ and unit scale parameter. It follows that the probability that a household chooses to remain in the same location (when given the opportunity to choose) is:

$$p_t(a; r) = \frac{1}{1 + (J - 1) \delta_r^{-1} \exp\left(-\kappa_r \left(\beta \eta_r \sum_{k=0}^{\pi-a-1} \beta^k (1 - \tau_{t+1+k})\right)^{-1}\right)}.$$  

(20)

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.\(^{20}\)

**Proposition 3 (Migration rates)** Assume that migration costs are positive and finite,\(^{20}\)

\(^{20}\)The results in this proposition can generalized to the case in which the $\varepsilon_k$ shocks are not distributed as type-1 extreme values. This result is available from the author upon request.
Then, the migration rate of a household:

1. Declines with age \( a \).

2. Given age \( a \), it is lower the higher the degree of tax progressivity \( \tau_{t+1+k} \) that the household faces in its remaining working life \( k \in [0, \bar{a} - a - 1] \).

The intuitions for these results are 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. The intuition for this result is due to the fact that 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 \( \tau_{t+1+k} \) on mobility. Last, the effect of \( \tau_{t+1+k} \) 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. If moving is prohibitively costly, there is no mobility and tax progressivity is also non-distortionary. Thus, the estimate of the moving cost parameter bears important implications for evaluating the effect of tax policy on migration and welfare.

### 4.2 Households’ Productivity Growth

Over time a household’s income changes because of the age-dependent evolution of its productivity, as captured by \( \alpha_{a,r} \), and because of the shocks to productivity and migration. Observed productivity growth across periods \( t \) and \( t + 1 \) for a household with location-specific shocks \( \varepsilon \) depends on whether it is free to choose whether to migrate or it is exogenously

\[ \kappa_r \in (0, +\infty) \].
relocated. If the household is free to choose, then:

\[ z' = z + g^*_t (\varepsilon, a; r), \]

where the growth rate of productivity is:

\[ g^*_t (\varepsilon, a; r) = \sum_{k=1}^{J} M_t (\varepsilon, a, h, k; r) (\alpha_{a,r} + \eta_r \varepsilon_k), \]

and \( g^*_t (\varepsilon, a; r) \) does not depend on \( h \) because of the symmetry of the model. If the household is exogenously relocated, instead, its productivity growth evolves as:

\[ z' = z + \sum_{k \neq h} x_{hk} (\alpha_{a,r} + \eta_r \varepsilon_k). \]

The following propositions characterize the distribution of \( g^*_t (\varepsilon, a; r) \) when the agent is able to choose her location next period, while subsequently I focus on the case of exogenous migration.

**Proposition 4 (Productivity growth with endogenous mobility)** Productivity growth conditional on choosing to stay in the same location is distributed according to a type 1 extreme-value distribution with mean:

\[ E_{\varepsilon} [g^*_t (\varepsilon, a; r) | M_t (\varepsilon, a, h, h; r) = 1] = \alpha_{a,r} + \eta_r (\gamma + \ln \delta_r - \ln p_t (a; r)), \]

and variance:

\[ \text{VAR}_{\varepsilon} [g^*_t (\varepsilon, a; r) | M_t (\varepsilon, a, h, h; r) = 0] = \eta_r^2 \pi^2 / 6. \]

Productivity growth conditional on choosing to migrate is distributed according to a type 1 extreme-value distribution with mean:

\[ E_{\varepsilon} [g^*_t (\varepsilon, a; r) | M_t (\varepsilon, a, h, h; r) = 0] = \alpha_{a,r} + \eta_r (\gamma + \ln \delta_r - \ln p_t (a; r)), \]

and variance:

\[ \text{VAR}_{\varepsilon} [g^*_t (\varepsilon, a; r) | M_t (\varepsilon, a, h, h; r) = 0] = \eta_r^2 \pi^2 / 6. \]

22 In other words, productivity growth does not depend on the specific “identity” \( h \) of the location where the household is initially residing because the aggregate characteristics of each location are the same.
extreme-value distribution with mean:

\[ E_{\varepsilon} [g_t^* (\varepsilon, a; r) | M_t (\varepsilon, a, h, h; r) = 0] = \alpha_{a,r} + \eta_r (\gamma + \ln \delta_r - \ln p_t (a; r)) + \frac{\kappa_r}{\beta \eta_{t+1}^2 (a + 1)}, \]  

(23)

and the same variance as in (22). The parameter \( \gamma \) in these equations denotes Euler’s constant.

Consider now productivity growth for an agent who is forced (i.e. cannot choose) to move. This agent is randomly reallocated to one of the remaining \( J - 1 \) locations where she will experience productivity growth \( \alpha_{a,r} + \eta_r \varepsilon_k \). Therefore, for this agent productivity growth follows the density of shocks drawn from one of the “foreign” locations.

**Proposition 5 (Productivity growth with exogenous mobility)** Productivity growth for a household who is forced to move (exogenous move) is distributed according to a type 1 extreme-value distribution with mean \( \alpha_{a,r} + \eta_r \gamma \) and variance \( \eta_r^2 \pi^2 / 6 \).

The previous result is useful in order to interpret the expressions in Proposition 4. With random mobility, average productivity growth reflects age effects \( \alpha_{a,r} \) and the term \( \eta_r \gamma \) associated with the extreme-value shocks drawn in the away-locations. The expressions in equations (21) and (23) reflect selection through migration.\(^{23}\) The conditional mean in equation (21) reflects selection by agents who choose to stay put. For this to be the optimal choice, they should receive 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 for stayers must be.\(^{24}\) For agents who choose to move out of a location, instead, average productivity growth in (23) exceeds, on average, productivity growth conditional on staying by an amount that reflects

\(^{23}\)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.

\(^{24}\)A low probability of staying reflects a relatively small moving costs. Thus, agents who choose to stay must experience a high income growth relative to what they would have experienced if they had chosen to move.
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. Finally, the variance of productivity growth in (22) is increasing in the parameter $\eta_r$, which indexes the importance of idiosyncratic shocks for households’ productivity growth.

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.

Propositions 4 and 5 are important because they allow me to compute analytically the moments targeted by the model’s estimation procedure as well as the various components of the welfare function.

## 5 Empirical Implementation

### 5.1 Empirical Strategy and Parameter Identification

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 distribution of household productivity $f(x|1,r)$ at age $a = 1$ is taken to be lognormal with parameters $(\mu_r, \sigma_r^2)$. The exogenous component of the evolution of household productivity is a linear function of age:

$$\alpha_{a,r} = \alpha_{0,r} + \alpha_{1,r} a.$$ 

The parameter vector is:

$$\left\{ \beta, \xi, \omega, \tau, J, \chi, \omega_r, \kappa, \theta_r, \tau_0, \tau_1, \eta_r, \delta_r, \mu_r, \sigma_r^2 \right\}_{r=1,2}.$$
The calibration 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 of Moments (Hansen, 1982). I postpone setting \( \chi \) until Section 6 as it only plays a role in welfare analysis.

5.1.1 Parameters Set A-Priori

The frequency of the model is one year. The years of working life are \( \bar{a} = 35 \), from age 25 to age 59 included. The annual discount factor \( \beta = 0.96 \). 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 (2017). The number of locations \( J \) 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 \( \omega_1 = 0.6751 \) (see Section 5.1.4).

5.1.2 Measure of Tax Progressivity

The tax progressivity parameter \( \tau \) is estimated using data on households’ market income and income post-Federal taxes and transfers. 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:

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

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.\(^{26}\)

For consistency with the other data used in the paper, I restrict attention to the period 2000-2007. However, as discussed below, the estimate of \( \tau \) is remarkably robust to considering a longer time period. 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 the Heathcote et al (2017)’s estimate of 0.181. 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.\(^{27}\)

5.1.3 Estimation

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

\[
\phi_r = \{ \mu_r, \sigma_r^2, \kappa_r, \delta_r, \eta_r, \alpha_{0,r}, \alpha_{1,r} \}_{r=1,2}.
\]

Moment Conditions The parameter vector \( \phi_r \) is estimated separately by household type \( r \). For each type, the estimation procedure targets the following \( 3\bar{a} - 1 \) moments: 1) the

\(^{26}\)State and local income taxes are not included in the CBO calculations. These taxes tend to be 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.

\(^{27}\)Adding a linear interaction term between year and the coefficient on \( \ln y_{pt} \) in equation \(^{24}\) 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.
migration rate by age ($\bar{a} - 1$ 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). Parameter identification is discussed in the following section. The analytical results of Section 4 allow me to compute all moment conditions analytically, so the estimation procedure does not rely on the use of simulations.

The migration rate by age for a type $r$ household is:

$$1 - \theta_r p(a; r) = 1 - \frac{\theta_r}{1 + (J - 1) \delta_r^{-1} \exp \left( -\kappa_r (1 - \beta) \left[ \beta \eta_r (1 - \tau) (1 - \beta^{a - a}) \right]^{-1} \right)},$$

(25)

where $p(a; r)$ is found by imposing a constant $\tau$ in equation (20).

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

$$m_{\ln y}(a; r) = m_z(a; r) + \ln \ell^*,$$

where $m_z(a; r)$ is the mean of the distribution of log productivity $z$ at age $a$. Using the results in Propositions 4 and 5 it can be shown (see Appendix H.1) to be recursively determined as follows:

$$m_z(a + 1; r) = m_z(a; r) + \alpha_{a,r} + \eta_r (\gamma + \ln \delta_r + \Delta(a; r)),$$

(26)

starting from $m_z(1; r) = \mu_r$. The term $\alpha_{a,r} + \eta_r (\gamma + \ln \delta_r)$ represents the growth rate of labor productivity in the one-location version of the model. The term $\Delta(a; r)$ in (26) represents the average growth rate of log income at age $a$ for a household of type $r$ that can be attributed to voluntary and involuntary migration:

$$\Delta(a; r) \equiv -\theta_r \ln p(a; r) + \theta_r \frac{(1 - p(a; r)) \kappa_r}{\beta \eta_r v^1(a + 1)} - (1 - \theta_r) \ln \delta_r.$$

Finally, the cross-sectional variance of log labor income for a household of type $r$ at age $a$ is equal to the cross-sectional variance of log productivity $v_{\ln y}(a; r) = v_z(a; r)$ because
labor supply is age-independent. Using again Propositions 4 and 5, it can be shown that the cross-sectional variance of log productivity evolves recursively as follows (see Appendix H.2):

\[ v_z(a + 1; r) = v_z(a; r) + \eta_r^2 \left( \pi^2 / 6 + \Psi(a; r) \right), \]

starting from \( v_z(1; r) = \sigma_r^2 \). The term \( \eta_r^2 \pi^2 / 6 \) in (27) represents the “within” component of the variance of income growth, which, by Propositions 4 and 5 is the same for all households independently of whether they choose to stay put, to migrate, or are exogenously relocated. The term \( \Psi(a; r) \) in equation (27) is formally defined in equation (A.19), Appendix H.2. It represents the “between” component of the variance of household productivity growth and it refers to differences between: (1) voluntary movers vs voluntary stayers, whose income growth differs because of the mobility cost \( \kappa_r \) (Proposition 4); (2) voluntary stayers vs involuntary movers, whose income growth is the lowest (Proposition 5).

The moment conditions implied by the model are then:

\[
E [M_i | i \in \Omega(a; r)] - (1 - \theta_r p(a; r)) = 0, \quad \text{for } a = 1, 2, \ldots, \bar{a} - 1
\]

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

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

where \( M_i = 1 \) if household \( i \) migrates and zero otherwise, and \( \Omega(a; r) \) is the set of households of age \( a \) and type \( r \). Let \( m_r(\varphi_r) \) denote the sample counterpart of the moments above. The GMM estimator of \( \varphi_r \) solves the following problem:

\[
\hat{\varphi}_r = \arg \min_{\varphi_r} m_r'(\varphi_r) W_r^{-1} m_r(\varphi_r),
\]

where \( W_r^{-1} \) is Hansen (1982)’s optimal weighting matrix.
**Identification**  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 at age $a$ for an individual of type $r$ (equation 25). As a household grows older, its incentives to voluntarily migrate decline because migration is costly ($\kappa_r > 0$) and its remaining time horizon $(\bar{a} - a)$ shrinks. Specifically, as $a \to \bar{a}$, the household migration rate converges to $1 - \theta_r$. Therefore, the parameter $\theta_r$ is identified by the migration rate at older ages. To understand the different roles played by $\delta_r$ and $\kappa_r$, consider the case in which moving costs are zero ($\kappa_r = 0$). In this case the migration rate is a constant independent of age:

$$\frac{1 - \theta_r + (J - 1) \delta_r^{-1}}{1 + (J - 1) \delta_r^{-1}}.$$  (28)

From this observation it follows that: i) In order to account for the declining pattern of migration as a function of age, it is necessary that $\kappa_r > 0$. Inspection of the expression for the migration rate in equation (25) suggests that the size of the moving cost relative to the importance of idiosyncratic shocks, or $\kappa_r/\eta_r$ and the discount factor $\beta$ regulate the dependence of migration rates on age. ii) Given $\kappa_r/\eta_r$, the parameter $\delta_r$ pins down the average migration rate for an agent of type $r$. Finally, notice that conditional on matching the migration data, changing the number of locations $J$ only results in a rescaling of the parameter $\delta_r$. The age profile of mean wages for each education type identifies the parameters $(\bar{a}_0, \bar{a}_1)$. The observed dispersion of log income at age 25 ($a = 1$) identifies the initial variance of log income, $\sigma^2$. Over time the cross-sectional variance of log income increases due to the presence of idiosyncratic shocks, whose importance is regulated by the parameter $\eta_r$ (see equation 27). A larger $\eta_r$ leads to a faster increase in income dispersion as a cohort of households ages. Finally, the mean parameter $\mu_r$ is identified by the magnitude of average log income at age 25.$^{28}$

$^{28}$Notice that, since the income 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 $\mu_1$.  

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$. 

---

28Notice that, since the income 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 $\mu_1$.
5.1.4 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 of households whose head is ages 25–59, 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. Among the households selected above, 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. The household labor income and mobility data are purged of year and state effects by running regressions of each of these variables on year and state dummies and using residuals to construct the moments of interest.

5.2 Estimated Parameters and Model’s Fit for Targeted Moments

Table 2 reports the estimates of the model’s parameters and their standard errors. Notice that, since the constraint $\delta_r \geq 1$ was binding for households of type $r = 2$, I imposed a-priori $\delta_2 = 1$ and estimated the remaining parameters for type 2 agents. The mobility cost $\kappa_r$, converted into consumption-equivalent units $(1 - \exp(-\kappa_r))$, corresponds to 88 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 parameter $\theta_r$ is such that the probability of an exogenous move for a type $r = 1$ is about 1.09 percent per year and so exogenous moves account, on average, for about 58 percent of all moves for this type. By contrast, the probability of a move due to exogenous reasons is 1.74 percent per year, or 50 percent of all moves for this type of agent. These figures are consistent with households’ responses to the Current Population Survey’s question about reasons for interstate migration (see footnote 5). 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 ($\delta_2 < \delta_1$).
<table>
<thead>
<tr>
<th>Parameters set a priori</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>HSV</td>
</tr>
<tr>
<td>( \beta ) yearly discount factor</td>
<td>.96</td>
<td>HSV</td>
</tr>
<tr>
<td>( \tau ) degree of tax progressivity</td>
<td>.192</td>
<td>see text</td>
</tr>
<tr>
<td>( \bar{\alpha} ) duration of working life</td>
<td>35</td>
<td>HSV</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>0.3248</td>
<td>ACS, 2000-2007</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>2.1123</td>
<td>0.1019</td>
</tr>
<tr>
<td>( \delta_1 ) frequency of local offers parameter</td>
<td>73.0442</td>
<td>12.0000</td>
</tr>
<tr>
<td>( \eta_1 ) importance of idiosyncratic shocks</td>
<td>0.0444</td>
<td>0.0004</td>
</tr>
<tr>
<td>( \bar{\alpha}_{0,1} ) age profile of earnings (constant)</td>
<td>-0.1880</td>
<td>0.0071</td>
</tr>
<tr>
<td>( \bar{\alpha}_{1,1} ) age profile of earnings (linear)</td>
<td>-0.0011</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \theta_1 ) one minus probability of exogenous move</td>
<td>0.9891</td>
<td>0.0002</td>
</tr>
<tr>
<td>( \mu_{1,1} ) mean of log earnings at ( a = 1 )</td>
<td>-0.1242</td>
<td>0.0023</td>
</tr>
<tr>
<td>( \phi_{1,1} ) variance of log earnings at ( a = 1 )</td>
<td>0.3625</td>
<td>0.0013</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>5.2246</td>
<td>0.0417</td>
</tr>
<tr>
<td>( \delta_2 ) frequency of local offers parameter</td>
<td>1.0000</td>
<td>n.a.</td>
</tr>
<tr>
<td>( \eta_2 ) importance of idiosyncratic shocks</td>
<td>0.0566</td>
<td>0.0004</td>
</tr>
<tr>
<td>( \bar{\alpha}_{0,2} ) age profile of earnings (constant)</td>
<td>-0.0101</td>
<td>0.0006</td>
</tr>
<tr>
<td>( \bar{\alpha}_{1,2} ) age profile of earnings (linear)</td>
<td>-0.0011</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \theta_2 ) one minus probability of exogenous move</td>
<td>0.9826</td>
<td>0.0002</td>
</tr>
<tr>
<td>( \mu_{2,1} ) mean of log earnings at ( a = 1 )</td>
<td>0.2626</td>
<td>0.0028</td>
</tr>
<tr>
<td>( \phi_{2,1} ) variance of log earnings at ( a = 1 )</td>
<td>0.3006</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Table 2: Summary of model parameters and standard errors.
As discussed in the identification section above, the lower frequency of home offers allows the model to account for the higher mobility of college-educated workers, while their higher moving cost gives rise to the faster decline in the geographic mobility rate for this group. Figure 2 shows the age pattern of migration. At younger ages the migration rate of college-educated labor is about twice as large as the migration rate of those in the less skilled group. Such difference tends to disappear at older ages. Figure 3 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.
Figure 3: Mean and variance of log earnings by age in the model (line) and in the data (circles) for type $r = 1, 2$ agents.
5.3 Implications for Moments not Targeted by the Estimation

In this section I discuss the model’s implications for moments that were not targeted in the GMM estimation. I focus on two: the observed gap in earnings growth between movers and stayers and the long-run relationship between tax progressivity and internal migration documented in Section 2.2.

5.3.1 Earnings Growth and Migration

The model’s parameters have been estimated using cross-sectional moments on household earnings and migration patterns by age. Since the model has implications for the magnitude of average earnings growth associated with migration, it is natural to ask whether the latter is consistent with the available evidence. Consider first the model’s predictions. According to Proposition 4, average earnings growth among households who choose to migrate is larger than average growth of households who choose to stay put. In other words, migrants who are in a position to choose to move are positively selected on earnings growth. On the other hand, not all migrants move intentionally in the model. Some are relocated exogenously, in which case Proposition 5 and Corollary 1 suggest that they should experience smaller average earnings growth than workers who choose not to migrate. It follows that the observed difference in average earnings growth between migrating and non-migrating households can, in principle, be either positive or negative depending on the incidence of exogenous relocations. It turns out that, using the estimated model, this difference is positive for both groups of households early on in the life cycle. It eventually shrinks and becomes negative as agents age and the relative incidence of exogenous relocations increases. Figure 4 plots the average gap in labor income growth between movers and stayers as a function of age.

Averaging over ages and taking into account that the probability of migration declines with age, the model implies that the average wage growth gap between movers and stayers is 3.3 percentage points for households whose head has less than a college degree and 20.4 percentage points for households with a college degree or more. In other words, for households with less than a college degree an interstate move leads, on average, to a drop
Figure 4: Average gap in labor income growth between movers and stayers for the two types of agents as a function of age. Data generated by the model.

in labor income, while more highly educated households experience an average gain. There are not many empirical studies that provide a counterpart of these numbers in the data. Yankow (2003, Table 8) uses panel data from the National Longitudinal Study of Youth 1979 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 amount is about 2 percentage points for full-time workers with a high school degree or less. My estimates have the same order of magnitude as these, but refer to populations with higher average schooling and concern household, rather than individual, earnings.

5.3.2 Internal Migration and Tax Progressivity

In Section 2.2 I presented reduced-form evidence of a negative effect of tax progressivity on internal migration rates. In order to interpret this evidence from the perspective of

\footnote{Yankow (1999, Table 2) finds wage growth gaps of about 5 percentage points between movers and stayers but does not condition this measure on education levels.}
Figure 5: Aggregate and type-specific migration rates in the steady state against values of the tax progressivity parameter $\tau$. The dot denotes the model-predicted average migration rate for the U.S. in the benchmark calibration.

In the model, I computed the model-implied steady state migration rate for various degrees of tax progressivity $\tau$. The steady state assumption is the counterpart of the fact that the empirical evidence of Section 2.2 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 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 line derived in Section 2.2. 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 remarkably consistent with the cross-country evidence. Specifically, an increase in $\tau$ from its benchmark value of 0.192 to 0.262 (a standard
deviation of the cross-sectional dispersion in Table 1) is predicted to reduce internal mobility by 0.45 percentage points in the model and by 0.36 percentage points in the data. Notice that, according to the model, the mobility rate of college educated workers is more sensitive to changes in tax progressivity than that of workers with less than a college degree.

6 Welfare Analysis

6.1 Welfare Function and Optimal Public Good Provision

In this section I use the model to study the welfare effects of tax progressivity. The main thought experiment is as follows. At some time denoted by convention as \( t = 1 \) the planner announces, unexpectedly, a policy sequence \( \{ \tau_t, \lambda_t, G_t \}_{t=1}^\infty \) to which it commits. It then computes the resulting competitive equilibrium and evaluates the welfare implications of the plan. The welfare criterion is utilitarian and therefore equal to the sum of the utilities of all cohorts, including those who are not alive at \( t = 1 \). The planner discounts the utility of future generations using the agents’ discount factor \( \beta \). Formally, the welfare criterion \( W \) is such:

\[
W = \pi^{-1} \sum_{r=1}^\infty \omega_r \sum_{a=1}^\pi \int V_1^* (a, z; r) f_1 (a|z, r) \, dz + \\
\pi^{-1} \sum_{r=1}^\infty \omega_r \sum_{k=2}^\infty \beta^{k-1} \int V_k^* (1, z; r) f_k (1|z, r) \, dz,
\]

where \( V_t^* (a, z; r) \equiv V_t (a, z, h; r) \) since the value function does not depend on the current residence indicator \( h \) due to the model’s symmetry.

The planner is subject to the government’s budget constraint (13), which requires a period-by-period budget balance. Since the government’s budget imposes a relationship among the

\[^{30}\]The 95% confidence interval associated with the regression’s prediction is \([0.07, 0.65]\).
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 \([13]\) into account. In what follows I express the problem as that of finding the optimal sequence of \(\tau_t\) and of 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 \([10]\), the two dimensions of optimization are independent of one another. The optimal share of public expenditures can be obtained analytically and is the same as in Heathcote et al (2017). This is summarized in the following proposition.

**Proposition 6 (Optimal public good provision)** The welfare maximizing share of public goods in aggregate income is constant over time and given by:

\[
\varphi_t^* = \frac{\chi}{1 + \chi} \text{ for all } t = 1, 2, ...
\]

In order to set the parameter \(\chi\), I follow Heathcote et al (2017) and assume that public good provision in the U.S. is set optimally, according to \([30]\). 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 of \(\chi = 0.089\).

### 6.2 Optimal Policy Reform

The optimal path \(\{\tau_t\}_{t=1}^{\infty}\) cannot, in general, be computed analytically and has to be found numerically. In what follows I approximate the optimal dynamic solution by postulating that the welfare-maximizing tax progressivity takes the flexible form:

\[
\tau_t = \overline{\tau} + \frac{1}{1 + \exp (\nu_0 + \nu_1 t)},
\]

where \(\nu_0, \nu_1,\) and \(\overline{\tau}\) are parameters. The algorithm chooses these three parameters to maximize the welfare function \([29]\) with \(\varphi_t^*\) given by \([30]\). The parameter \(\overline{\tau}\) can be computed \(^{31}\) Appendix I describes the numerical algorithm in more detail.

\[^{31}\]Appendix I describes the numerical algorithm in more detail.
a-priori because it represents the limit of the sequence of optimal policies. In the limit, the planner cares only about new-born generations and the economy is in the steady state, so $\pi$ can be computed by maximizing their expected utility:

$$\sum_{\tau=1}^{\pi} \omega_\tau \int V^*(1, z; r) f(1|z, r) dz,$$

where the absence of time subscripts in the value function is due to the fact that in the limit the economy is in steady state. The optimal sequence $\{\tau_t\}$ is represented in Figure 6 by the solid (blue) line. Interestingly, the optimal path $\{\tau_t\}$ converges over time to a value $\tau = 0.202$, which is close to the value of $\tau = 0.192$ measured in the data and used in the estimation of the model. The latter is represented in Figure 6 by a dashed (red) line. The welfare gain of the optimal policy - defined as the proportional increase in all ages’ and cohorts’ consumption necessary to make social welfare the same in the initial steady state as under the reform - is equivalent to 1.26% of lifetime consumption.\textsuperscript{32}

It is interesting to compare the optimal path of tax progressivity with 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 a transition phase towards its new steady state. The optimal constant tax progressivity is $\tau = 0.326$, represented in Figure 6 as a dash-dotted (yellow) line. Figure 7 reports the average migration rates by agent type associated with the one-shot and time-varying reforms. The decline in migration rates, especially for the cohorts that are young in the early period of high redistribution, leads to a decline in human capital accumulation. This can be seen in Figure 8 which represents the effect of the time-varying policy reform on the cross-section of average human capital by age in different time periods. Period $t = 1$ corresponds to the model’s initial steady state. By period $t = 20$, the reform (which goes into effect in $t = 1$) has reduced the average human capital, especially of college educated middle-aged individuals who have been exposed to high degrees of redistribution for their entire lives. By period $t = 50$ the optimal path of $\tau_t$ is close to its initial steady

\textsuperscript{32}The welfare gain in consumption units is $\exp((1 - \beta) (W^R - W^B)) - 1$, where $W^R$ is social welfare under the reform and $W^B$ is welfare in the benchmark steady state.
Figure 6: Optimal tax progressivity reform. The solid (blue) line represents the optimal path of time-varying $\tau_t$ according to the relationship in equation (31). The dashed (red) line represents $\tau = 0.192$ in the initial steady state. The dash-dotted (yellow) line represents the constant (i.e. non-time varying) $\tau$ that maximizes welfare. The dotted (purple) line represents the $\tau$ that maximizes steady state welfare.
Figure 7: Migration rates for type $r = 1$ (left plot) and $r = 2$ (right plot) in the initial steady state (dash-dotted yellow line), the one-shot policy reform (dashed blue line), and the time-varying reform (solid red line).
Figure 8: Effect of the time-varying tax progressivity reform on the cross-section of human capital by age. Each line represents the average efficiency at a certain age (x-axis) of agents who are alive in periods $t = 1$, $t = 20$, and $t = 50$.

state value and the human capital profile of the young cohorts is close to the corresponding levels of young households in $t = 1$. Notice that the effect of the reform on the human capital of $r = 1$ agents is, instead, largely muted.

In order to better understand the contribution of endogenous geographic mobility to the results in Figure 6, I have solved for the optimal path $\{\tau_t\}_{t=1}^\infty$ while keeping migration behavior the same as in the original steady state of the model (corresponding to the estimated model of Section 5). In practice, I adjust the migration cost in each period and for each age so that the migration rate of agents in that age group and period equals their corresponding migration rate in the initial steady state. Formally, this means replacing the moving cost $\kappa_r$ with $\tilde{\kappa}_{rt}(a)$ such that:

$$\frac{\tilde{\kappa}_{rt}(a)}{v^1_{t+1}(a + 1)} = \frac{\kappa_r}{v^1(a + 1)}$$  \hspace{1cm} (32)
where $v^1(a + 1)$ is the version of equation (15) in the initial steady state and $v^1_{t+1}(a + 1)$ refers to its counterpart in the economy with time-varying policy. The ratio in equation (32) determines a household’s migration rate according to the expression in equation (25). Therefore, adjusting the moving cost according to equation (32) guarantees that the migration rate does not change over time in response to alternative sequences of tax progressivity. As a result, the economy-wide distribution of human capital remains constant over time. Since the planner does not need to take into account the negative effect of redistribution on human capital accumulation, it will choose higher degrees of redistribution than in the benchmark version of the model. Figure 9 plots both the optimal sequence $\{\tau_t\}$ for the benchmark model (the one already reported in Figure 6) and the optimal sequence when migration choices are not affected by tax progressivity.
The figure shows that at the time of the reform \((t = 1)\) the optimal degrees of tax progressivity are approximately the same in the benchmark model and in the version in which migration is kept constant. However, over time, optimal tax progressivity declines substantially in the former model while it remains constant in the latter scenario. This exercise shows that the optimal dynamic path of tax progressivity is driven by the effect of tax progressivity on internal migration rates.

### 6.3 Discussion of welfare effects

In the previous section I have discussed the implications of the model for optimal tax progressivity reform. One important result is that there are substantial welfare gains from a time-dependent path of tax progressivity. The latter displays a declining path of redistribution over time. High degrees of redistribution early on during the reform are due to the fact that the human capital of the cohorts that are alive at the time of the reform is fixed. With inelastic labor supply, there would not be any distortion associated with setting \(\tau_1\) close to one.\(^{33}\) However, labor supply is elastic and therefore the planner finds it optimal not to set \(\tau_1\) at the highest level. The optimal reform involves a declining path of \(\tau_t\) over time because of the distortions of future tax progressivity on future migration choices. As pointed out by many authors (e.g. Domeij and Heathcote (2004)), analyzing the entire transitional dynamics of the economy after a tax reform, as opposed to comparing steady states, has important implications for assessing the welfare effects of the reform. In my model, the welfare gain from a one-shot optimal reform is 2.37 percent of consumption when comparing steady states and 0.78 percent when considering the entire transition. In the former case, the planner selects a slightly progressive tax system while in the latter the planner chooses a higher degree of tax progressivity than in the benchmark (Figure 6). The intuition is that a relatively small \(\tau\) gives rise to a much larger increase in average productivity in

\(^{33}\)In the optimal taxation a la Ramsey literature (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.
the distant new steady state than in the early phases of the transition. 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 comparison) than in one where investment is irreversible (i.e., taking the transition into account). Krueger and Ludwig (2016) study optimal education subsidies and taxes in a rich quantitative model of schooling choices with intergenerationally-linked households. They find that explicitly taking the transition into account is not only quantitatively, but also qualitatively, important when measuring the welfare effects of education and tax policies. Specifically, in their model implementing the steady state optimal policy substantially lowers social welfare (relative to the pre-reform benchmark) once the economy’s transitional path is taken into account. Similarly to Krueger and Ludwig (2016), I find that welfare declines by 3.47 percent, as measured by lifetime consumption, under a reform that lowers \( \tau \) from its benchmark steady state value of 0.192 to the value \( \tau = 0.048 \) that maximizes steady state welfare. In other words, due to the presence of transitional dynamics, social welfare is higher in the status quo than under a policy that is only optimal in the steady state.

While these papers explicitly consider transitional dynamics, they do not compute the optimal time-varying policy during the transition from the initial to the final steady state.\footnote{The reason is that it is numerically very hard to compute time-varying optimal policies in models such as those mentioned in the text, featuring a large amount of cross-sectional heterogeneity and general equilibrium effects.} Interestingly, Krueger and Ludwig (2016, p.) hypothesize that “given the important differences between steady state optimal and transition-optimal policy it is conceivable, in fact likely, that policies that are time-varying over the transition provide further welfare gains.” This conjecture is proven correct in the context of my model. As I mentioned in the previous section, optimizing policy over the transition leads to a welfare gain of 1.26 percent of lifetime consumption, while the welfare gain from the one-shot optimal policy is about forty percent smaller. A summary of the welfare effects discussed here is contained in Table 3.
Table 3: Summary of welfare effects. Each entry represents an equivalent variation measured as a percent of period consumption that has to be added to each household to make social welfare in the initial steady state (where $\tau = 0.192$) equal to social welfare under the policy reform.

### 7 Conclusions and Future Work

In this paper I study the effect of tax progressivity on internal migration, a form of human capital investment. I first document empirically that internal migration rates are lower in countries characterized by higher degrees of tax progressivity. I then construct an analytically tractable, yet rich, dynamic model of internal migration. The model predicts that higher degrees of tax progressivity reduce migration rates. In order to quantify the importance of this effect, the model’s parameters are estimated using data on migration and wage dynamics. The effect of tax progressivity on migration is found to be negative and quantitatively consistent with the cross-country data. The model is also used to compute the optimal degree of tax progressivity and to quantify the contribution of the migration channel to the optimal policy. The migration channel plays a quantitatively important role in determining the degree of tax progressivity chosen by the social planner. The model’s tractability allows me to compute the optimal dynamic path of tax progressivity and not only the one-shot reform, as common in the literature. The dynamic path features a higher degree of tax progressivity than in the original steady state in the early periods, followed by declining tax progressivity over time. High degrees of tax progressivity in the early periods allow the planner to reap the benefits of redistribution while minimizing its distortions since
the human capital stock is a slow-moving state variable.

A more general contribution of the paper is to introduce a new dynamic model of 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 local or national policies that lead to a geographic reallocation of labor. I plan to pursue this extension in future work.
Bibliography


Online Appendix (Not meant for publication)

A  Proposition 1

The proof proceeds by guessing and verifying that the solution takes the form in Proposition 1. The Bellman equation is:

\[
V_t(a, z, h; r) = \bar{w}_t + (1 - \tau_t) z + \theta_r E_{\bar{\epsilon}} \left[ \max_k \{ \beta V_{t+1}(a + 1, \alpha_{a,r} + z + \eta_r \bar{\epsilon}_k, k; r) - I_{hh} \kappa_r \} \right] + (1 - \theta_r) E_{\bar{\epsilon}, x} \left[ \sum_{k \neq h} x_{hk} \{ \beta V_{t+1}(a + 1, \alpha_{a,r} + z + \eta_r \bar{\epsilon}_k, k; r) - \kappa_r \} \right].
\]

Guess and verify that the value function takes the general form:

\[
V_t(a, z, h; r) = v^0_t(a; r) + v^1_t(a) z \text{ for all } h. \tag{A.1}
\]

Replace the guess:

\[
V_t(a, z, h; r) = \bar{w}_t + (1 - \tau_t) z + \theta_r E_{\bar{\epsilon}} \left[ \max_k \{ \beta v^0_{t+1}(a + 1; r) + \beta v^1_{t+1}(a + 1) (\alpha_{a,r} + z + \eta_r \bar{\epsilon}_k) - I_{hh} \kappa_r \} \right] + (1 - \theta_r) E_{\bar{\epsilon}, x} \left[ \sum_{k \neq h} x_{hk} \{ \beta v^0_{t+1}(a + 1; r) + \beta v^1_{t+1}(a + 1) (\alpha_{a,r} + z + \eta_r \bar{\epsilon}_k) - \kappa_r \} \right].
\]

Simplify the equation above:

\[
V_t(a, z, h; r) = \bar{w}_t + (1 - \tau_t) z + \beta v^0_{t+1}(a + 1; r) + \beta v^1_{t+1}(a + 1) (\alpha_{a,r} + z) - (1 - \theta_r) \kappa_r + \theta_r \beta \eta_r v^1_{t+1}(a + 1) E_{\bar{\epsilon}} \left[ \max_k \{ \bar{\epsilon}_k - \frac{I_{hh} \kappa_r}{\beta \eta_r v^1_{t+1}(a + 1)} \} \right] + (1 - \theta_r) \beta v^1_{t+1}(a + 1) \eta_r \bar{\gamma}. \tag{A.2}
\]
The distribution of the random variable

\[
\max_k \left\{ \varepsilon_k - \frac{I_{hk} \kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right\}
\]

can be computed as

\[
\Pr \left[ \max_k \left\{ \varepsilon_k - \frac{I_{hk} \kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right\} < y \right] = \Pr [\varepsilon_h < y] \Pr \left[ \max_{k \neq h} \left\{ \varepsilon_k - \frac{I_{hk} \kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right\} < y \right] = \exp \left(-\delta_r \exp(-y)\right) \exp \left(- (J-1) \exp \left(- \left( y + \frac{\kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right) \right) \right) \exp (-y).
\]

This is a type 1 extreme-value with location parameter:

\[
\ln \left[ \delta_r + (J-1) \exp \left(- \frac{\kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right) \right],
\]

and unit scale parameter. Thus, its mean is:

\[
E \varepsilon \left[ \max_k \left\{ \varepsilon_k - \frac{I_{hk} \kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right\} \right] = \ln \left[ \delta_r + (J-1) \exp \left(- \frac{\kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right) \right] + \gamma.
\]

Replace the latter into (A.2):

\[
V_t(a, z, h; r) = \overline{u}_t^* + (1 - \tau_t + \beta v^1_{t+1} (a+1)) z + \beta v^0_{t+1} (a+1; r) + \beta v^1_{t+1} (a+1) (\alpha_{a,r} + \eta_r \gamma) - (1 - \theta_r) \kappa_r + \theta_r \beta \eta_r v^1_{t+1} (a+1) \ln \left[ \delta_r + (J-1) \exp \left(- \frac{\kappa_r}{\beta \eta_r v^1_{t+1} (a+1)} \right) \right].
\]

Impose consistency with the original guess (A.1):

\[
v^1_t(a) = 1 - \tau_t + \beta v^1_t (a+1)
\]
for $a < \bar{a}$. This can be solved to obtain:

$$v_t^1 (a) = \sum_{k=0}^{\bar{a}-a} \beta^k (1 - \tau_{t+k}).$$

Now, impose consistency with the constant term. In the last period of life, $v_0^0 (\bar{a}; r) = \bar{u}_t^*.$ For $a < \bar{a}$ instead:

$$v_t^0 (a; r) = \bar{u}_t^* + \beta v_{t+1}^0 (a+1; r) - (1 - \theta_r) \kappa_r + \beta v_{t+1}^1 (a+1) (\alpha_{a,r} + \eta_r \gamma) +$$

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

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

$$p_t (a; r) = \frac{\delta_r}{\delta_r + (J - 1) \exp \left( -\frac{\kappa_r}{\beta \eta_r v_{t+1}^1 (a+1)} \right)}.$$ (A.4)

Thus:

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

Replacing it into (A.3):

$$v_t^0 (a; r) = \bar{u}_t^* + \beta v_{t+1}^0 (a+1; r) - (1 - \theta_r) \kappa_r$$

$$+ \beta v_{t+1}^1 (a+1) (\alpha_{a,r} + \eta_r \gamma + \theta_r \ln \delta_r) - \theta_r \eta_r \ln p_t (a; r).$$

Q.E.D.

**B Proposition 2**

Consider an agent of age $a < \bar{a}$, located in $j$, with shock vector $\varepsilon$. Comparing value functions, a choice to remain in location $h$ ($M_t (\varepsilon, a, h, h; r) = 1$) requires that:
\[ \beta v^1_{t+1} (a + 1) (\alpha_{a,r} + z_a + \eta_r \xi_h) > \beta v^1_{t+1} (a + 1) (\alpha_{a,r} + z_a + \eta_r \xi_l) - \kappa_r \text{ for all } l \neq h. \]

The inequality above is equivalent to:

\[ \varepsilon_h > \max_{l \neq h} \left( \varepsilon_l - \frac{\kappa_r}{v^1_{t+1} (a + 1) \beta_r \eta_r} \right). \]

Rearranging this gives:

\[ \frac{\kappa_r}{v^1_{t+1} (a + 1) \beta_r \eta_r} > \max_{l \neq h} \varepsilon_l - \varepsilon_h. \]

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

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

and

\[ \beta v^1_{t+1} (a + 1) (\alpha_{a,r} + z_a + \eta_r \xi_k) - \kappa_r > \beta v^1_{t+1} (a + 1) (\alpha_{a,r} + z_a + \eta_r \xi_h). \]

Simplifying these expressions we obtain:

\[ \frac{\kappa_r}{v^1_{t+1} (a + 1) \beta_r \eta_r} < \max_{l \neq h} \varepsilon_l - \varepsilon_h \]

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

Q.E.D.

C 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.4). This variable is increasing in age, \( a \), because \( v^1_{t+1} (a + 1) \) is decreasing in \( a \). Thus, the migration rate declines with age.

2. A higher \( \tau_{t+k+1} \) reduces \( v^1_{t+1} (a + 1) \) and increases \( p_t (a ; r) \), so it reduces migration.
With a zero moving cost $\kappa_r$ both effects in 1. and 2. go away because becomes:

$$p_t(a;r) = \frac{\delta_r}{\delta_r + J - 1}.$$ 

As $\kappa_r \to \infty$, $p_t(a;r) \to 1$. Q.E.D.

## D Proposition 4

### D.1 Household chooses to stay

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

$$\Pr(g^*_t(\varepsilon, a; r) \leq s | M_t(\varepsilon, a, h, h; r) = 1) = \frac{\Pr(g^*_t(\varepsilon, a; r) \leq s \text{ and } M_t(\varepsilon, a, h, h; r) = 1)}{\Pr(M_t(\varepsilon, a, h, h; r) = 1)},$$  

(A.5)

where by definition:

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

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

$$\Pr(g^*_t(\varepsilon, a; r) \leq s \text{ and } M_t(\varepsilon, a, h, h; r) = 1) = \Pr(\alpha_{a,r} + \eta_r \varepsilon_h \leq s \text{ and } \varepsilon_h > \max_{l \neq h} \varepsilon_l - \frac{\kappa_r}{\nu_{l+1}^1(a + 1) \beta \eta_r}).$$

(A.5)

The equation above can be rewritten as:

$$\Pr\left(\max_{l \neq h} \varepsilon_l - \frac{\kappa_r}{\nu_{l+1}^1(a + 1) \beta \eta_r} < \varepsilon_h \leq \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r}\right).$$

(A.6)

Taking into account the fact that $\max_{l \neq h} \varepsilon_l$ is type-1 extreme value with location para-
\[ \int \Pr \left( x - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} < \varepsilon_h \leq \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} | x \right) (J - 1) \exp (-x) \exp (- (J - 1) \exp (-x)) \, dx. \]  

(A.7)

We know that:

\[ \Pr \left( x - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} < \varepsilon_h \leq \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} | x \right) = 0 \]

if

\[ x > \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r}. \]  

(A.8)

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

\[ \Pr \left( x - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} < \varepsilon_h \leq \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} | x \right) = \exp \left( - \delta_r \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) \right) \right) - \exp \left( - \delta_r \exp \left( - \left( x - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right) \right) \right). \]

Replace the previous expressions into (A.7):

\[ \int \Pr \left( x - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} < \varepsilon_h \leq \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} | x \right) \times (J - 1) \exp (-x) \exp (- (J - 1) \exp (-x)) \, dx \]

(A.9)
The first component is:

\[
\begin{align*}
\int_{-\infty}^{\frac{z}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \exp \left( -\delta_r \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) \right) \right) (J - 1) \exp (-x) \exp (- (J - 1) \exp (-x)) \right) \right) dx \\
= \exp \left( -\delta_r \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) \right) \right) \exp \left( - (J - 1) \exp \left( - (J - 1) \exp (-x) \right) \right) \right) \right) dx
\end{align*}
\]

The second term is:

\[
\begin{align*}
\int_{-\infty}^{\frac{z}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \exp \left( -\delta_r \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) \right) \right) (J - 1) \exp (-x) \exp (- (J - 1) \exp (-x)) \right) \right) \right) dx \\
= \int_{-\infty}^{\frac{z}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \exp \left( -\delta_r \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) \right) \right) \exp \left( - (J - 1) \exp (-x) \right) \right) \right) \right) dx \\
= \int_{-\infty}^{\frac{z}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \exp \left( -\delta_r \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) \right) \right) (J - 1) \exp (-x) \exp \left( - \left[ \delta_r \exp \left( \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \right) + (J - 1) \right] \exp (-x) \right) \right) \right) \right) \right) dx \\
= \frac{(J - 1)}{\delta_r \exp \left( \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \right) + (J - 1)} \int_{-\infty}^{\frac{z}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \exp \left( -\delta_r \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) \right) \right) (J - 1) \exp (-x) \exp \left( - \left[ \delta_r \exp \left( \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \right) + (J - 1) \right] \exp (-x) \right) \right) \right) \right) \right) \right) dx.
\end{align*}
\]

The term inside the integral is a density, so its CDF is:

\[
\exp \left( - \left[ \delta_r \exp \left( \frac{\kappa_r}{v_{t+1}^1(a+1)\beta_{r}} \right) + (J - 1) \right] \exp (-x) \right).
\]

58
The integral is therefore:

\[
\exp \left( - \left[ \delta_r \exp \left( \frac{\kappa_r}{v_{t+1} v_{t+1} (a + 1) \beta \eta_r} \right) + (J - 1) \right] \exp \left( - \left( \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1} v_{t+1} (a + 1) \beta \eta_r} \right) \right) \right)
\]

\[
= \exp \left( - \left[ \delta_r \exp \left( \frac{\kappa_r}{v_{t+1} v_{t+1} (a + 1) \beta \eta_r} \right) + (J - 1) \right] \exp \left( \frac{\alpha_{a,r} - \kappa_r}{\eta_r} \right) \right) \exp \left( - \left( \frac{s}{\eta_r} \right) \right)
\]

It follows that the second term (A.10) is:

\[
\int_{-\infty}^{\frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} + \frac{\kappa_r}{v_{t+1} v_{t+1} (a + 1) \beta \eta_r}} \exp \left( - \delta_r \exp \left( - \left( x - \frac{\kappa_r}{v_{t+1} (a + 1) \beta \eta_r} \right) \right) \right) \times (J - 1) \exp (-x) \exp(- (J - 1) \exp (-x)) \, dx
\]

\[
= \frac{(J - 1) \exp (-x) \exp(- (J - 1) \exp (-x))}{(J - 1) + \delta_r \exp \left( \frac{\kappa_r}{v_{t+1} v_{t+1} (a + 1) \beta \eta_r} \right) \exp \left( - \exp \left( \frac{\alpha_{a,r}}{\eta_r} \right) \delta_r + (J - 1) \exp \left( - \frac{\kappa_r}{v_{t+1} v_{t+1} (a + 1) \beta \eta_r} \right) \right) \exp \left( - \frac{s}{\eta_r} \right) \right).}
\]

Putting all together equation (A.9) becomes:

\[
\int \Pr \left( x - \frac{\kappa_r}{v_{t+1} (a + 1) \beta \eta_r} < \varepsilon_h \leq \frac{s}{\eta_r} - \frac{\alpha_{a,r}}{\eta_r} \right) (J - 1) \exp (-x) \exp(- (J - 1) \exp (-x)) \, dx
\]

\[
= \frac{1}{1 + (J - 1) \delta_r^{-1} \exp \left( - \frac{\kappa_r}{v_{t+1} (a + 1) \beta \eta_r} \right) \exp \left( - \exp \left( \frac{\alpha_{a,r}}{\eta_r} \right) \delta_r + (J - 1) \exp \left( - \frac{\kappa_r}{v_{t+1} (a + 1) \beta \eta_r} \right) \right) \exp \left( - \frac{s}{\eta_r} \right) \right).}
\]

Since

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

\[
\Pr (g^*_t (\varepsilon, a; r) \leq s | M_t (\varepsilon, a, h, h; r) = 1) \\
= \exp \left( - \exp \left( \frac{\alpha_{a,r}}{\eta_r} \left( \delta_r + (J - 1) \exp \left( - \frac{\kappa_r}{\nu_{i+1} (a + 1) \beta \eta_r} \right) \right) \right) \right) \exp \left( - \frac{s}{\eta_r} \right) \\
= \exp \left( - \exp \left( \frac{\alpha_{a,r} + \eta_r \ln \delta_r + (J - 1) \exp \left( - \frac{\kappa_r}{\nu_{i+1} (a + 1) \beta \eta_r} \right) \right)}{\eta_r} \right) \exp \left( - \frac{s}{\eta_r} \right) \\
= \exp \left( - \exp \left( - \eta^{-1}_r \left\{ s - \left[ \alpha_{a,r} + \eta_r \ln \left( \delta_r + (J - 1) \exp \left( - \frac{\kappa_r}{\nu_{i+1} (a + 1) \beta \eta_r} \right) \right) \right] \right) \right) \right) \exp \left( - \frac{s}{\eta_r} \right) \\
= \exp \left( - \exp \left( - \eta^{-1}_r \left\{ s - \left[ \alpha_{a,r} + \eta_r \ln \delta_r - \eta_r \ln (p_{t, a; r}) \right] \right) \right) \right) \\
\]

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

\[
\alpha_{a,r} + \eta_r \ln \left( \frac{\delta_r}{p_{t, a; r}} \right)
\]

and scale parameter \(\eta_r\). The mean of this distribution is:

\[
\alpha_{a,r} + \eta_r \ln \left( \frac{\delta_r}{p_{t, a; r}} \right) + \eta_r \gamma \\
= \alpha_{a,r} + \eta_r \left( \gamma + \ln \delta_r - \ln p_{t, a; r} \right).
\]

Its variance is:

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

**D.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:

$$\Pr (g_t^* (\varepsilon, a; r) \leq s| M_t (\varepsilon, a, h; r) = 0). \tag{A.11}$$

Since the agent may move to one of the remaining $J - 1$ locations and these events are disjoint, I can write:

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

$$= \frac{\Pr (g_t^* (\varepsilon, a; r) \leq s \text{ and } \cup_{k \neq h} (M_t (\varepsilon, a, h, k; r) = 1))}{\Pr (M_t (\varepsilon, a, h; r) = 0)}$$

$$= \frac{\Pr (\cup_{k \neq h} (g_t^* (\varepsilon, a; r) \leq s \text{ and } M_t (\varepsilon, a, h, k; r) = 1)))}{\Pr (M_t (\varepsilon, a, h; r) = 0)}.$$

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

$$\Pr (\cup_{k \neq h} (g_t^* (\varepsilon, a; r) \leq s \text{ and } (M_t (\varepsilon, a, h, k; r) = 1)))$$

$$= \sum_{k \neq h} \Pr \left( \sum_{k=1}^J M_t (\varepsilon, a, h, k; r) g_k (\varepsilon_k, a; r) \leq s \text{ and } M_t (\varepsilon, a, h, k; r) = 1 \right)$$

$$= \sum_{k \neq h} \Pr (g_k (\varepsilon_k, a; r) \leq s \text{ and } M_t (\varepsilon, a, h, k; r) = 1),$$

where $g_k (\varepsilon_k, a; r) \equiv \alpha_{a,r} + \eta_r \varepsilon_k$. Recall that for $k \neq h$:

$$\Pr (g_k (\varepsilon_k, a; r) \leq s \text{ and } M_t (\varepsilon, a, h, k; r) = 1)$$

$$= \Pr \left( \alpha_{a,r} + \eta_r \varepsilon_k \leq s \text{ and } \frac{\kappa_r v_{l+1} (a + 1) \beta \eta_r}{\varepsilon_k - \varepsilon_h} < \varepsilon_k - \varepsilon_h \text{ and } \varepsilon_k = \max_{l \neq h} \varepsilon_l \right)$$

$$= \Pr \left( \alpha_{a,r} + \eta_r \varepsilon_k \leq s \text{ and } \frac{\kappa_r v_{l+1} (a + 1) \beta \eta_r}{\varepsilon_k - \varepsilon_h} < \varepsilon_k - \varepsilon_h | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) \times \Pr (\varepsilon_l = \max_{l \neq h} \varepsilon_l).$$

Notice that:

$$\Pr \left( \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) = \frac{1}{J - 1}.$$

61
The distribution of $\varepsilon_k$ conditional on being the maximum of $J - 1$ i.i.d. type 1 extreme-value random variables, is type 1 extreme-value with location parameter $\ln(J - 1)$. Write:

$$ \Pr\left( \alpha_{a,r} + \eta_r \varepsilon_k \leq s \text{ and } \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} < \varepsilon_k - \varepsilon_h | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) $$

$$ = \Pr\left( \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} + \varepsilon_h < \varepsilon_k \leq \frac{s - \alpha_{a,r}}{\eta_r} | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) $$

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

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

$$ \Pr\left( \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} + u < \varepsilon_k \leq \frac{s - \alpha_{a,r}}{\eta_r} | \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) = 0 $$

if

$$ \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} + u > \frac{s - \alpha_{a,r}}{\eta_r} $$

and

$$ \exp\left( - (J - 1) \exp\left( - \left( \frac{s - \alpha_{a,r}}{\eta_r} \right) \right) \right) - \exp\left( - (J - 1) \exp\left( - \left( \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} + u \right) \right) \right)$$

if

$$ u < \frac{s - \alpha_{a,r}}{\eta_r} - \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r}. $$
Thus:

\[
\int \Pr \left( \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} + u < \varepsilon_k \leq \frac{s - \alpha_{a,r}}{\eta_r} \mid \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) g(u) \, du
\]

\[
= \int_{-\infty}^{s - \alpha_{a,r}} - \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} \left\{ \exp \left( - (J - 1) \exp \left( - \left( \frac{s - \alpha_{a,r}}{\eta_r} \right) \right) \right) - \exp \left( - (J - 1) \exp \left( - \left( \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} + u \right) \right) \right) \right\} g(u) \, du
\]

\[
= \int_{-\infty}^{s - \alpha_{a,r}} - \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} \exp \left( - (J - 1) \exp \left( - \left( \frac{s - \alpha_{a,r}}{\eta_r} \right) \right) \right) g(u) \, du - \int_{-\infty}^{s - \alpha_{a,r}} - \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} \exp \left( - (J - 1) \exp \left( - \left( \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} + u \right) \right) \right) g(u) \, du.
\]

The first integral is:

\[
\int_{-\infty}^{s - \alpha_{a,r}} - \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} \exp \left( - (J - 1) \exp \left( - \left( \frac{s - \alpha_{a,r}}{\eta_r} \right) \right) \right) g(u) \, du
\]

\[
= \exp \left( - (J - 1) \exp \left( - \left( \frac{s - \alpha_{a,r}}{\eta_r} \right) \right) \right) \int_{-\infty}^{s - \alpha_{a,r}} - \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} g(u) \, du
\]

\[
= \exp \left( - (J - 1) \exp \left( - \left( \frac{s - \alpha_{a,r}}{\eta_r} \right) \right) \right) \exp \left( - \delta_r \exp \left( - \left( \frac{s - \alpha_{a,r}}{\eta_r} - \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} \right) \right) \right)
\]

\[
= \exp \left( - \left[ J - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_{l+1}^1 (a + 1) \beta \eta_r} \right) \right] \exp \left( \frac{\alpha_{a,r}}{\eta_r} \right) \exp \left( - \left( \frac{s}{\eta_r} \right) \right) \right).
\]
The second integral is:

\[
\int_{-\infty}^{s-a_r \nu_{\nu_r}} \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \exp \left( - (J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) + u \right) g(u) \, du \\
= \int_{-\infty}^{s-a_r \nu_{\nu_r}} \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \exp \left( - (J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) \exp (-u) \right) g(u) \, du \\
= \int_{-\infty}^{s-a_r \nu_{\nu_r}} \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \exp \left( - (J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) \exp (-u) \right) \\
\left[ \delta_r \exp (-u) \exp (-\delta_r \exp (-u)) \right] \, du \\
= \delta_r \int_{-\infty}^{s-a_r \nu_{\nu_r}} \exp (-u) \exp \left( - \left[ (J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) + \delta_r \right] \exp (-u) \right) \\
= \frac{\delta_r}{(J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) + \delta_r} \times \\
\exp \left( - \left[ (J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) + \delta_r \right] \exp \left( - \left( \frac{s - \alpha_{a_r}}{\eta_r} - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) \right) \right) .
\]

Simplify the result:

\[
= \frac{\delta_r}{(J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) + \delta_r} \times \\
\exp \left( - \left[ (J-1) \exp \left( - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) + \delta_r \right] \exp \left( - \left( \frac{s - \alpha_{a_r}}{\eta_r} - \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) \right) \right) \\
= \delta_r \exp \left( - \left[ J - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_{t+1}^1 (a+1) \beta \eta_r} \right) \right] \exp \left( \frac{\alpha_{a_r}}{\eta_r} \right) \exp \left( - \frac{s}{\eta_r} \right) \right) .
\]
Put together the two integrals to get:

\[
\int \Pr \left( \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} + u < \varepsilon_k \leq \frac{s - \alpha_{a,r}}{\eta_r} \mid \varepsilon_k = \max_{l \neq h} \varepsilon_l \right) g(u) \, du
\]

\[
= \frac{(J - 1) \exp \left( - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right)}{(J - 1) \exp \left( - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right) + \delta_r} \times \exp \left( - \left[ J - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right) \right] \exp \left( \frac{\alpha_{a,r}}{\eta_r} \right) \exp \left( - \frac{s}{\eta_r} \right) \right).
\]

It follows that

\[
\Pr \left( g_t^* (\varepsilon, a; r) \leq s \mid M_t (\varepsilon, a, h, h; r) = 0 \right)
\]

\[
= \frac{(J - 1) \exp \left( - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right) \exp \left( - \left[ J - 1 + \delta_r \exp \left( \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right) \right] \exp \left( \frac{\alpha_{a,r}}{\eta_r} \right) \exp \left( - \frac{s}{\eta_r} \right) \right)}{(J - 1) \exp \left( - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right) + \delta_r}.
\]

Notice that

\[
1 - p_t (a; r) = 1 - \frac{1}{1 + (J - 1) \delta_r^{-1} \exp \left( - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right)}
\]

\[
= \frac{(J - 1) \delta_r^{-1} \exp \left( - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right)}{1 + (J - 1) \delta_r^{-1} \exp \left( - \frac{\kappa_r}{v_{t+1}^l (a + 1) \beta \eta_r} \right)}.
\]
Replace:

\[
\Pr(g_t^*(\varepsilon, a; r) \leq s|\mathbf{M}_t(\varepsilon, a, h, h; r) = 0) = (J - 1) \delta_r^{-1} \exp \left(-\frac{\eta_r}{v_{t+1}(a+1)\beta \eta_r}\right) \exp \left(-J - 1 + \delta_r \exp \left(\frac{\kappa_r}{v_{t+1}(a+1)\beta \eta_r}\right)\right) \exp \left(\frac{\alpha_{a,r}}{\eta_r}\right) \exp \left(-\frac{s}{\eta_r}\right)
\]

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

\[
\exp \left(-\exp \left(-\frac{s - \alpha_{a,r} - \eta_r \ln \left(J - 1 + \delta_r \exp \left(\frac{\kappa_r}{v_{t+1}(a+1)\beta \eta_r}\right)\right)}{\eta_r}\right)\right)
\]

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

\[
\alpha_{a,r} - \eta_r \ln p_t(a; r) + \eta_r \ln \delta_r + \frac{\kappa_r}{v_{t+1}(a+1)\beta}.
\]

The moments of this distribution are:

\[
\text{mean} = \alpha_{a,r} + \eta_r \left(\gamma + \ln \delta_r - \ln p_t(a; r)\right) + \frac{\kappa_r}{v_{t+1}(a+1)\beta},
\]

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

Q.E.D.
E Proposition 5

If an agent is forced to move from a location $j$ and reallocated randomly across the remaining $J-1$ locations, her productivity growth conditional on being reallocated to $k$ is $g_k (\varepsilon_k, a; r) = \alpha_{a,r} + \eta_r \varepsilon_h$. An agent is reallocated to each $k$ with probability $(J-1)^{-1}$. The CDF of this distribution is:

$$
\Pr \left( \sum_{k \neq h} x_k g_k (\varepsilon_k, a; r) < z \right)
$$

$$
= \sum_{k \neq h} (J-1)^{-1} \Pr \left( \sum_{k \neq h} x_k g_k (\varepsilon_k, a; r) < z | x_k = 1 \right)
$$

$$
= \sum_{k \neq h} (J-1)^{-1} \Pr (g_k (\varepsilon_k, a; r) < z)
$$

$$
= \Pr (g_k (\varepsilon_k, a; r) < z).
$$

It follows that its CDF is simply the CDF of $g_k (\varepsilon_k, a; r)$:

$$
\Pr (g_k (\varepsilon_k, a; r) < s) = \Pr (\alpha_{a,r} + \eta_r \varepsilon_h < s)
$$

$$
= \Pr \left( \varepsilon_h < \frac{s - \alpha_{a,r}}{\eta_r} \right)
$$

$$
= \exp \left( - \exp \left( - \frac{s - \alpha_{a,r}}{\eta_r} \right) \right).
$$

This is the CDF of a type-1 extreme value distribution with scale parameter $\eta_r$ and location parameter $\alpha_{a,r}$. Its moments are:

$$
\text{mean} = \alpha_{a,r} + \eta_r \gamma,
$$

$$
\text{variance} = \frac{\eta_r^2 \pi^2}{6}.
$$

Q.E.D.
Corollary 1

Compare the means of the three extreme-value distributions derived in Propositions 4 and 5. The claim is that:

\[
\alpha_{a,r} + \eta_r (\gamma + \ln \delta_r - \ln p_t(a;r)) + \frac{\kappa_r}{\nu_{t+1} (a+1) \beta} > \alpha_{a,r} + \eta_r (\gamma + \ln \delta_r - \ln p_t(a;r)) \geq \alpha_{a,r} + \eta_r \gamma.
\]

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.

Proposition 6

The indirect utility function takes the form:

\[
u_t^* (z) = \overline{\nu}_t^* + (1 - \tau_t) z \tag{A.12}
\]

where:

\[
\overline{\nu}_t^* \equiv \ln \lambda_t + (1 - \tau_t) \ln \ell_t^* - \zeta^{-1} (\ell_t^*)^\zeta + \chi \ln G_t,
\]

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

\[
\lambda_t = (1 - \varphi_t) \frac{\sum_{a=1}^\pi \sum_{r=1}^\rho \omega_r \int \ell_t^* \exp z f_t(z | a, r) \, dz}{\sum_{a=1}^\pi \sum_{r=1}^\rho \omega_r \int (\ell_t^* \exp z)^{1 - \tau_t} f_t(z | a, r) \, dz}.
\]
Take logs:

\[
\ln \lambda_t = \ln (1 - \varphi_t) + \ln \bar{\alpha}^{-1} \sum_{a=1}^{\bar{\alpha}} \sum_{r=1}^{\bar{r}} \omega_r \int \ell_t^* \exp z f_t(z|a, r) \, dz \\
- \ln \bar{\alpha}^{-1} \sum_{a=1}^{\bar{\alpha}} \sum_{r=1}^{\bar{r}} \omega_r \int (\ell_t^* \exp z)^{1-\tau_t} f_t(z|a, r) \, dz.
\]

(A.13)

Replace \(\ln \lambda_t\) from (A.13) into (A.12), 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.

**H  Moment Conditions**

The migration rate is derived in the text. Here I derive the recursions of the mean and variance of log earnings.

**H.1  Cross-Sectional Mean of Log Income by Age**

The log income of a household is simply \(\ln y = \ln z + \ln \ell^*\). The law of motion of log productivity is:

\[
\ln z_{a+1} = \ln z_a + g^* (\varepsilon, a; r)
\]

(A.14)

for a household who can choose her location and

\[
\ln z_{a+1} = \ln z_a + \tilde{g} (\varepsilon, a; r)
\]

(A.15)
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:

\[
\ln z'_{a+1} = \ln z_a + Zg^* (\varepsilon, a; r) + (1 - Z) \hat{g} (\varepsilon, a; r).
\]

The initial \( a = 1 \) distribution of log income for a type \( r \) household has mean:

\[
E [\ln z_a | a = 1, r] = \mu_r.
\]

The cross-sectional mean evolves as follows:

\[
E [\ln z'_{a+1} | a+1, r] = E [\ln z_a | a, r] + \theta_r E [Zg^* (\varepsilon, a; r) | Z = 1] + (1 - \theta_r) E [(1 - Z) \hat{g} (\varepsilon, a; r) | Z = 0]
= E [\ln z_a | a, r] + \theta_r E [g^* (\varepsilon, a; r)] + (1 - \theta_r) E [\hat{g} (\varepsilon, a; r)].
\]

Taking into account Propositions 4 and 5:

\[
E [\ln z'_{a+1} | a+1, r] = E [\ln z_a | a, r] + \theta_r p(a; r) (\alpha_{a,r} + \eta_r \gamma - \eta_r \ln p(a; r)/\delta_r) + \\
+ \theta_r (1 - p(a; r)) \left( \alpha_{a,r} + \eta_r \gamma - \eta_r \ln p(a; r)/\delta_r + \frac{\kappa_r}{\beta v^1(a+1)} \right) + \\
+ (1 - \theta_r) (\alpha_{a,r} + \eta_r \gamma).
\]

This can be simplified as:

\[
E [\ln z'_{a+1} | a+1, r] = E [\ln z_a | a, r] \\
+ \theta_r \left[ \alpha_{a,r} + \eta_r \gamma - \eta_r \ln p(a; r)/\delta_r + \frac{(1 - p(a; r)) \kappa_r}{\beta v^1(a+1)} \right] + \\
+ (1 - \theta_r) (\alpha_{a,r} + \eta_r \gamma).
\]
Or:

\[
E [\ln z'_{a+1}|a+1, r] = E [\ln z_a|a, r] + \alpha_{a,r} + \eta_r \gamma + \theta_r \left(-\eta_r \ln (p(a; r) / \delta_r) + \frac{(1 - p(a; r)) \kappa_r}{\beta v^1 (a + 1)}\right) + \eta_r \ln \delta_r
\]

\[
= E [\ln z_a|a, r] + \alpha_{a,r} + \eta_r \gamma + \theta_r \eta_r \left[\ln \delta_r - \ln p(a; r) + \frac{(1 - p(a; r)) \kappa_r}{\beta \eta_r v^1 (a + 1)}\right] + \theta_r \eta_r \left(-\ln p_t (a; r) + \frac{(1 - p(a; r)) \kappa_r}{\beta \eta_r v^1 (a + 1)}\right)
\]

Let \( \Delta (a, r) : \)

\[
\Delta (a, r) = \theta_r \left(-\ln p(a; r) + \frac{(1 - p(a; r)) \kappa_r}{\beta \eta_r v^1 (a + 1)}\right) - (1 - \theta_r) \ln \delta_r
\]

represent the average growth of income due to migration (as a choice and endogenous). Then we obtain:

\[
E [\ln z'_{a+1}|a+1, r] = E [\ln z_a|a, r] + \alpha_{a,r} + \eta_r (\gamma + \ln \delta_r) + \eta_r \Delta (a, r).
\]  \hfill (A.16)

Q.E.D.

H.2 Cross-Sectional Variance of Log Income by Age

For sake of space, in what follows I shorten \( M (\varepsilon, a, h; h; r) \) to \( M (\varepsilon, a; r) \) and define the conditioning term:

\[
C (a, \varepsilon, r) \equiv a, r, Z, M (\varepsilon, a; r).
\]

The cross-sectional variance of log income at age \( a = 1 \) is

\[
V [\ln z_a|a = 1, r] = \phi_r^2.
\]
The evolution of log income is given by:

$$\ln z'_{a+1} = \ln z_a + Z g^* (\varepsilon, a, d; r) + (1 - Z) \hat{g}(\varepsilon, a; r)$$
$$= \ln z_a + Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r).$$

The variance at subsequent ages is therefore:

$$V[\ln z'_{a+1}|a + 1, r]$$
$$= V[\ln z_a + Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r)|a, r]$$
$$= V[\ln z_a|a, r] + V[Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r)|a, r].$$

From the law of total variance:

$$V[Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r)|a, r] =$$
$$E_{Z,M} \{V_{\varepsilon} [Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r)|C(a, \varepsilon, r), Z]\}$$
$$+ V_{Z,M} \{E_{\varepsilon} [Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r)|C(a, \varepsilon, r), Z]\}.$$

Consider the term inside $V_{\varepsilon}$. Notice that for all combinations of $(Z, M (\varepsilon, a; r))$:

$$V_{\varepsilon} [Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r)|C(a, \varepsilon, r), Z]$$
$$= \eta_r^2 \pi^2 / 6$$

because of Propositions 4 and 5. Thus, the average is also $\eta_r^2 \pi^2 / 6$:

$$E_{Z,M} \{V_{\varepsilon} [Z M (\varepsilon, a; r) g^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g^* (\varepsilon, a; r) + (1 - Z) \hat{g}(\varepsilon, a; r)|C(a, \varepsilon, r), Z]\}$$
$$= \eta_r^2 \pi^2 / 6.$$

Consider now the second term in the decomposition (A.17), which I call the “between-variance”. To compute the latter consider that there are 3 possible mutually exclusive
combinations of \((Z, M)\):

\[
Z = 1, M (\varepsilon, a; r) = 0, \\
Z = 1, M (\varepsilon, a; r) = 1, \\
Z = 0.
\]

The inside portion is:

\[
E_\varepsilon [Z M (\varepsilon, a; r) g_i^* (\varepsilon, a; r) + Z (1 - M (\varepsilon, a; r)) g_i^* (\varepsilon, a; r) + (1 - Z) \tilde{g}_j (\varepsilon, a; r) | C (a, \varepsilon, r), Z]
\]

which corresponds to:

- 1. \(Z = 1, M (\varepsilon, a; r) = 0\):

\[
\alpha_{a,r} + \eta_r \gamma - \eta_r \ln (p (a; r) / \delta_r) + \frac{\kappa_r}{\beta v_{l+1}^1 (a + 1)}
\]

- 2. \(Z = 1, M (\varepsilon, a; r) = 1\):

\[
\alpha_{a,r} + \eta_r \gamma - \eta_r \ln (p (a; r) / \delta_r)
\]

- 3. \(Z = 0\):

\[
\alpha_{a,r} + \eta_r \gamma.
\]

The variance \(V_{Z,M}\) of these is computed using the distribution of \((Z, M (\varepsilon, a; r))\):

- 1. \(Z = 1, M (\varepsilon, a; r) = 0\) occurs with probability:

\[
\theta_r (1 - p (a; r))
\]
• 2. $Z = 1, M(\varepsilon, a; r) = 1$ occurs with probability:

$$\theta_r p(a; r)$$

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

The mean can be written as in equation (A.16):

$$\text{mean growth } (a, r) \equiv \alpha_{a,r} + \eta_r (\gamma + \ln \delta_r) + \eta_r \Delta (a, r).$$

(A.18)

The variance is then:

$$\text{between variance } (a, r)$$

$$\equiv \theta_r (1 - p(a; r)) \left( \alpha_{a,r} + \eta_r \gamma - \eta_r \ln (p(a; r)/\delta_r) + \frac{\kappa_r}{\beta v_{t+1}^1 (a + 1)} - \text{mean growth } (a, r) \right)^2 +$$

$$+ \theta_r p(a; r) (\alpha_{a,r} + \eta_r \gamma - \eta_r \ln (p(a; r)/\delta_r) - \text{mean growth } (a, r))^2 +$$

$$+ (1 - \theta_r) (\alpha_{a,r} + \eta_r \gamma - \text{mean growth } (a, r))^2.$$  

Replace (A.18) and simplify:

$$\text{between variance } (a, r) = \theta_r (1 - p(a; r)) \left( -\eta_r \ln p(a; r) + \frac{\kappa_r}{\beta v_{t+1}^1 (a + 1)} - \eta_r \Delta (a, r) \right)^2 +$$

$$+ \theta_r p(a; r) (-\eta_r \ln (p(a; r)) - \eta_r \Delta (a, r))^2 + (1 - \theta_r) (-\eta_r \ln \delta_r - \eta_r \Delta (a, r))^2.$$  

Collect outside $\eta^2_r$:

$$\text{between variance } (a, r) = \theta_r (1 - p(a; r)) \eta^2_r \left( \frac{\kappa_r}{\beta \eta_r v_{t+1}^1 (a + 1)} - \ln p(a; r) - \Delta (a, r) \right)^2 +$$

$$+ \theta_r p(a; r) \eta^2_r (\ln (p(a; r)) + \Delta (a, r))^2 + (1 - \theta_r) \eta^2_r (\ln \delta_r + \Delta (a, r))^2.$$
Define:

$$
\Psi (a, r) = \theta_r (1 - p(a; r)) \left( \ln p(a; r) + \Delta (a, r) - \frac{\kappa_r}{\beta \eta_v v_{t+1}^1 (a + 1)} \right)^2 + \theta_r p(a; r) (\ln (p(a; r)) + \Delta (a, r))^2 + (1 - \theta_r) (\ln \delta_r + \Delta (a, r))^2. \tag{A.19}
$$

Putting everything together, we can write (A.17) as:

$$
V \left[ \ln z_{a+1}' | a + 1, r \right] = V \left[ \ln z_a | a, r \right] + \eta_r^2 \pi^2 / 6 + \eta_r^2 \Psi (a, r).
$$

Q.E.D.

I Numerical Algorithm to Compute Optimal Policy Path

In order to solve for the optimal policy path I first postulate a flexible function form for $\tau_t$:

$$
\tau_t = \bar{\tau} + \frac{1}{1 + \exp (\nu_0 + \nu_1 t)}, \tag{A.20}
$$

where $(\bar{\tau}, \nu_0, \nu_1)$ are parameters. As long as $d > 0$, the path will converge to $\bar{\tau}$ over time. This observation allows me to set $\bar{\tau}$ a-priori because it represents the degree of tax progressivity that the planner would select if its objective was to maximize the welfare of young cohorts. The latter is the only portion of the welfare function that is relevant as $t \to \infty$. Thus, I can find $\bar{\tau}$ by maximizing the objective:

$$
\sum_{r=1}^{\bar{\tau}} \omega_r \int V^* (1, z; r) f (1 | z, r) dz,
$$

subject to the government budget constraint. In accounting for the latter, keep in mind that it only depends on $\bar{\tau}$ either directly or through the distribution of human capital of agents alive in the distant future. The rest of the algorithm describes how to find $(\nu_0, \nu_1)$
to maximize the overall welfare function. I first show how to derive the welfare function and then describe the algorithm.

I.1 Welfare Function

The welfare function is defined as follows:

\[
\bar{\alpha}W_1 = \sum_{r=1}^{\tau} \omega_r \sum_{a=1}^{\pi} \int V_1 (a, z; r) f_1 (a|z, r) \, dz + \sum_{r=1}^{\tau} \omega_r \sum_{k=2}^{\infty} \beta^{k-1} \int V_k (1, z; r) f (1|z, r) \, dz.
\]  

(A.21)

The value function in period \( t \) is:

\[
V_t (a, z; r) = v_0^t (a; r) + v_1^t (a) z.
\]

Replacing into (A.21):

\[
\bar{\alpha}W_1 = \sum_{r=1}^{\tau} \omega_r \sum_{a=1}^{\pi} \int [v_0^1 (a; r) + v_1^1 (a) z] f_1 (a|z, r) \, dz + \sum_{r=1}^{\tau} \omega_r \sum_{k=2}^{\infty} \beta^{k-1} \int [v_0^k (1; r) + v_1^k (1) z] f (1|z, r) \, dz.
\]

It can be simplified to:

\[
\bar{\alpha}W_1 = \sum_{a=1}^{\pi} \sum_{r=1}^{\tau} \omega_r v_0^1 (a; r) + \sum_{a=1}^{\pi} v_1^1 (a) \sum_{r=1}^{\tau} \omega_r m_{z_1} (a; r) + \sum_{r=1}^{\tau} \omega_r \sum_{k=2}^{\infty} \beta^{k-1} v_0^k (1; r) + \sum_{r=1}^{\tau} \omega_r \sum_{k=2}^{\infty} \beta^{k-1} v_1^k (1) \mu_r,
\]

where \( m_{z_1} (a; r) \) is defined as mean log productivity at age \( a \) in the initial steady state of the model:

\[
m_{z_1} (a; r) = \int zf (a|z, r) \, dz.
\]
In the numerical solution, I find it convenient to re-write welfare as the sum of the utilities of various cohorts, indexed by $c$, with $c = 1$ denoting the cohort who was born in $t = -33$ and is of age $\bar{a} = 35$ in period $t = 1$ and cohort $c = 35$ the cohort who is of age $a = 1$ in $t = 1$. Recall that the relationship between time, age, and cohort is:

$$c = t - a + \bar{a}.$$ 

Let $v^0_c(a; r)$ and $v^1_c(a; r)$ represent the “cohort” notation counterparts of $v^0_t(a; r)$ and $v^1_t(a; r)$. For example, for a household of age $a$ that is part of cohort $c$:

$$v^1_c(a; r) = \sum_{k=0}^{\pi-a} \beta^k (1 - \tau_{c+a+\bar{a}+k}), \quad \text{(A.22)}$$

and similarly for $v^0_c(a; r)$. Then, the welfare function can be written as:

$$aW_1 = \sum_{r=1}^{\pi} \omega_r \sum_{c=1}^{\bar{a}} v^0_c(1 - c + \bar{a}; r) + \sum_{r=1}^{\pi} \omega_r \sum_{c=1}^{\bar{a}} v^1_c(1 - c + \bar{a}; r) m_{z1} (1 - c + \bar{a}; r) \mu_r. \quad \text{(A.23)}$$

I.2 Algorithm

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 human capital at $t = 1$, $m_{z1} (a; r)$ by age $a$. Recall that the economy is in steady state in $t = 1$ and since human capital is a state variable it does not respond to changes in $\tau$ that occur at $t = 1$.

- 2. Assume that the cohorts that are born after cohort $n$ face the same constant $\tau$ as faced by cohort $n$. This allows me to compute the infinite sums that appear on the
second row of equation (A.23). This approximate welfare function $\tilde{W}_1$ is defined as:

$$\tilde{W}_1 = \sum_{r=1}^{\tau} \omega_r \sum_{c=1}^{\pi} \bar{p}_c^0 (1 - c + \bar{a}; r) + \sum_{r=1}^{\tau} \omega_r \sum_{c=1}^{\pi} p_c^1 (1 - c + \bar{a}; r) m_{c+1} (1 - c + \bar{a}; r) +$$

$$\sum_{r=1}^{\tau} \omega_r \left[ \sum_{c=\pi+1}^{n} \beta^{(c-\pi)} \bar{p}_c^0 (1; r) + \sum_{c=n+1}^{\infty} \beta^{(c-\pi)} \bar{p}_n^0 (1; r) \right]$$

$$+ \sum_{r=1}^{\tau} \omega_r \mu_r \left[ \sum_{c=\pi+1}^{n} \beta^{(c-\pi)} \bar{p}_c^1 (1; r) + \sum_{c=n+1}^{\infty} \beta^{(c-\pi)} \bar{p}_n^1 (1; r) \right],$$

with

$$\sum_{c=n+1}^{\infty} \beta^{(c-\pi)} \bar{p}_c^0 (1; r) = \beta^{n+1-\pi} \sum_{c=n+1}^{\infty} \beta^{(c-n-1)} \bar{p}_n^0 (1; r) = \bar{p}_n^0 (1; r) \frac{\beta^{n+1-\pi}}{1 - \beta},$$

$$\sum_{c=n+1}^{\infty} \beta^{(c-\pi)} \bar{p}_c^1 (1; r) = \beta^{n+1-\pi} \sum_{c=n+1}^{\infty} \beta^{(c-n-1)} \bar{p}_n^1 (1; r) = \bar{p}_n^1 (1; r) \frac{\beta^{n+1-\pi}}{1 - \beta}.$$ 

- 3. Use the “Amoeba” algorithm (Numerical Recipes) to minimize $-\tilde{W}_1 (\nu_0, \nu_1)$ with respect to the parameters $(\nu_0, \nu_1)$.

To compute the one-shot constant optimal policy I simply maximize the $\tilde{W}_1$ with respect to a constant $\tau$.

### J Data Appendix

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](http://www.oecd.org/social/income-distribution-database.htm). 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.